Estatística aplicada à pesquisa agrícola

2ª edição revista e ampliada



Francisco José Pfeilsticker Zimmermann

Embrapa

Empresa Brasileira de Pesquisa Agropecuária
Embrapa Arroz e Feijão
Ministério da Agricultura, Pecuária e Abastecimento

Estatística aplicada à pesquisa agrícola

2ª edição revista e ampliada

Francisco José Pfeilsticker Zimmermann

Embrapa
Brasília, DF
2014

Exemplares desta publicação podem ser adquiridos na:

Embrapa Arroz e Feijão
Rodovia Goiânia/Nova Veneza, no 462,
Km 12, Fazenda Capivara, Zona Rural
Caixa Postal 179
CEP 75375-000 Santo Antônio de Goiás, GO
Fone: (62) 3533-2110
Fax: (62) 3533-2100
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Unidade responsável pelo conteúdo
Embrapa Arroz e Feijão

Comitê de Publicações da Embrapa Arroz e Feijão

Presidente
Roselene de Queiroz Chaves

Secretário-executivo
Luiz Roberto Rocha da Silva

Membros
Camilla Souza de Oliveira
Flávia Aparecida de Alcântara
Luís Fernando Stone
Ana Lúcia Delalibera de Faria
Heloisa Célis de Paiva Breseghello
Márcia Gonzaga de Castro Oliveira
Fábio Fernandes Nolêto

Embrapa Informação Tecnológica
Parque Estação Biológica (PqEB)
Av. W3 Norte (final)
CEP 70770-901 Brasília, DF
Fone: (61) 3448-4236
Fax: (61) 3448-2494
www.embrapa.br/livraria
livraria@embrapa.br

Unidade responsável pela edição
Embrapa Informação Tecnológica

Coordenação editorial
Selma Lúcia Lira Beltrão
Lucilene Maria de Andrade
Nilda Maria da Cunha Sette

Supervisão editorial
Erika do Carmo Lima Ferreira

Revisão de texto
Jane Baptistone de Araújo

Normalização bibliográfica
Celina Tomaz de Carvalho

Ilustração da capa
Ronaldo Cordeiro

Projeto gráfico e capa
Carlos Eduardo Felice Barbeiro

1ª edição
1ª impressão (2004): 1.000 exemplares

2ª edição
1ª impressão (2014): 1.000 exemplares

Todos os direitos reservados
A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei n° 9.610).

Dados Internacionais de Catalogação na Publicação (CIP).
Embrapa Informação Tecnológica

Zimmermann, Francisco José Pfeilsticker.
Estatística aplicada à pesquisa agrícola / Francisco José Pfeilsticker Zimmermann. – 2. ed. – Brasília, DF : Embrapa, 2014.
582 p. ; 18,5 cm x 25,5 cm.

ISBN 978-85-7035-344-3

1. Análise estatística. 2. Método estatístico. I. Embrapa Arroz e Feijão. II. Título.

CDD 519.5

© Embrapa 2014

Autor

Francisco José Pfeilsticker Zimmermann

Engenheiro-agrônomo, Ph.D. em Estatística Aplicada, pesquisador aposentado da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO

Apresentação

A estatística experimental oferece uma ampla gama de ferramentas para análise de conjuntos de dados obtidos a partir de experimentos de campo e laboratório. Como todas as ferramentas, elas são úteis apenas nas mãos de quem sabe usá-las. A aplicação correta de métodos estatísticos é essencial para chegar a conclusões válidas e interpretáveis em termos probabilísticos.

O livro *Estatística Aplicada à Pesquisa Agrícola* oferece os conhecimentos necessários para que pesquisadores e estudantes planejem seus experimentos e analisem seus dados da forma mais adequada para cada caso. Esta segunda edição, para atender às solicitações de vários usuários da primeira, foi totalmente revista e significativamente ampliada, com novos procedimentos estatísticos.

A programação SAS (*Statistical Analysis System*) para a análise de todos os delineamentos e procedimentos é apresentada. O Capítulo 12 inclui uma série de modelos, tanto na sua expressão matemática quanto gráfica, e dá mais opções aos leitores na escolha de modelos mais adequados.

Três novos capítulos foram acrescentados. O Capítulo 16 trata da análise exploratória de dados, o capítulo 17 apresenta técnicas de amostragem e o Capítulo 18 é dedicado a uma breve introdução ao uso do SAS. Dados de produtividade média municipal de feijão do Estado de Santa Catarina, para os anos de 2008 e 2009, foram utilizados para exemplificar a análise exploratória e as técnicas de amostragens.

Este livro apresenta-se como um importante aliado à pesquisa agrícola. Com ele, o leitor poderá tirar o máximo proveito de dados experimentais, testando hipóteses, comparando tratamentos e validando conclusões.

Flávio Breseghello
Chefe-Geral da Embrapa Arroz e Feijão

Prefácio à 1ª edição

Nos idos de 1968, recém-formado engenheiro-agrônomo, comecei a trabalhar como experimentador agrícola na estação experimental da Secretaria de Agricultura, no Parque Estadual de Vila Velha em Ponta Grossa, Paraná, como responsável pela condução dos ensaios com cevada, no inverno, e com soja e feijão, no verão. Incentivado por meu primeiro chefe, Giampiero Baldanzi, dediquei-me a analisar os dados colhidos naqueles ensaios, usando como guia o livro de Pimentel Gomes, *Curso de Estatística Experimental*. Foi então que senti a "mordida do bichinho" da Estatística, da qual nunca mais me afastei, tornando-se, ela, uma das grandes paixões da minha vida. Mais tarde, tive o privilégio de ter Pimentel Gomes como professor em três disciplinas no mestrado em Experimentação e Estatística na Esalq, Piracicaba, SP (1972–1973). Em 1975, tornei-me responsável pela estatística experimental da Embrapa Arroz e Feijão, onde, nestes quase 30 anos, já analisei mais de 8 mil experimentos, não apenas com feijão e arroz, mas também com milho, sorgo, soja, trigo, girassol, pasto, gado leiteiro e de carne, além das bactérias, fungos, vírus e outras tantas "pragas" da agricultura. O doutorado, realizado na Califórnia (1979–1983), agregou uma outra área da estatística ao meu universo de pesquisa, a não paramétrica.

Em 2001, em licença sabática na Colômbia, onde atuei como professor visitante nas Universidades de La Sabana, Faculdade de Engenharia, em Chia, Colômbia, e na Nacional de Colômbia, Departamento de Estatística, em Bogotá, dediquei a maior parte do meu tempo a escrever este livro, que em breve deverá ser publicado também em espanhol na Colômbia. Nele procurei reunir toda a experiência adquirida trabalhando com meus colegas pesquisadores de arroz e feijão, com os quais muito aprendi e que muito me desafiaram no planejamento, análise e interpretação dos resultados. Foi nesse ambiente que desenvolvemos os ensaios centrais, puros ensaios fatoriais no que diz respeito à estatística, mas inovadores no que tange aos objetivos; que descobrimos os ensaios comerciais de Chacin Lugo; que exercitamos o emprego do delineamento em blocos aumentados de Federer nas culturas anuais; e que começamos a descobrir as parcelas divididas dentro de um experimento em faixa, delineamento muito pouco conhecido e que, em 2003, foi tema de um trabalho final de curso de um estudante de estatística na Nacional da Colômbia.

O livro contém as informações que considero mais importantes e necessárias a um pesquisador agrícola, e abrange desde o mais simples dos delineamentos, inteiramente

casualizado, até os estudos de superfície de resposta, análise de grupos de experimentos, testes de médias, contrastes e a aplicação da estatística não paramétrica aos delineamentos experimentais, regressão e correlação. Todos os exemplos de aplicação apresentados são oriundos de experimentos reais, conduzidos pela equipe técnica da Embrapa Arroz e Feijão, a qual agradeço pela cessão do uso de seus dados. Em particular, sem desmerecer qualquer um de meus colegas, cito alguns deles pelo constante e maravilhoso desafio que sempre me propuseram: o doutor Evane Ferreira, que mostrou a possibilidade de emprego dos quadrados latinos 12x12, normalmente não recomendados pelos demais autores, e pelas análises conjuntas de quadrado latino com um tratamento comum; o doutor Paulo Hideo pelos blocos aumentados de Federer e as comparações com os reticulados; os doutores Luis Stone e Pedro Marques, pelos ensaios comerciais e experimentos em faixa com parcelas divididas; o doutor Cleber Guimarães e a doutora Beatriz Pinheiro, fisiologistas; e os doutores Orlando Peixoto, Emilio da Maia de Castro e Joaquim Geraldo Cáprio da Costa, melhoristas de plantas, pelo aprimoramento do uso de amostragens dentro de parcelas experimentais na realização das análises estatísticas; e os doutores Nand Kumar Fagéria e Morel Barbosa, por me fazerem dedicar um bom tempo aos contrastes para fatores não igualmente espaçados.

Espero que este livro seja útil aos pesquisadores brasileiros e agradeço, antecipadamente, as críticas e sugestões que os leitores considerarem pertinentes. Aprimorar é sempre possível, especialmente na ciência em que o fim nunca está a vista e, ainda mais nesta obra, que não é exaustiva na exploração do assunto e nem pretendeu ser.

O autor

Prefácio à 2ª edição

Pouco depois da publicação da primeira edição deste livro, decidi aposentar-me e voltei para a Colômbia, onde continuei meu trabalho como professor da Faculdade de Engenharia da Universidad de La Sabana, até o ano de 2007. Nesse ano, acompanhando minha esposa, fomos morar em Maputo, Moçambique, onde permanecemos até o ano de 2010.

No período de 2007 a 2010, trabalhei como professor visitante na Universidade Eduardo Mondlane (UEM), ministrando cursos de estatística experimental nos mestrados em Desenvolvimento Rural, treinamentos específicos em estatística aplicada à segurança alimentar para os técnicos do Secretariado Técnico em Segurança Alimentar do Ministério da Agricultura daquele país.

Para esses treinamentos, elaborei o livro *Estatística Aplicada à Segurança Alimentar*, uma publicação conjunta entre a UEM e a Organização das Nações Unidas para a Agricultura e Alimentação (FAO). Durante esse mesmo período, fui informado de que os dois livros haviam se esgotado. Também recebi imensos pedidos e sugestões de colegas estatísticos, e de professores e estudantes de várias universidades no Brasil, e fora dele, o que, adicionado a uma indagação feita pela Embrapa Arroz e Feijão, me levaram a escrever esta segunda edição, revisada e ampliada.

Devo aqui agradecer a todas as pessoas da Embrapa Informação Tecnológica, pelo apoio na revisão e correção do texto, em especial a Lucilene Maria de Andrade, Jane Baptistone de Araújo e Celina Tomaz de Carvalho, sem as quais este livro não seria publicado. Quaisquer erros nele encontrados são de exclusiva responsabilidade do autor.

Continuo esperando novas críticas e sugestões, que possam no futuro levar a uma terceira edição.

O autor

Sumário

Introdução 13

Capítulo 1 Conceitos gerais e definições úteis 19

Capítulo 2 Delineamento inteiramente casualizado 47

Capítulo 3 Blocos completos ao acaso 69

Capítulo 4 Delineamento em quadrado latino 95

Capítulo 5 Delineamento em blocos incompletos 125

Capítulo 6 Delineamentos em reticulado 149

Capítulo 7 Blocos aumentados de Federer (BAF) 181

Capítulo 8 Ensaios fatoriais 195

Capítulo 9 Experimentos em parcelas divididas e suas variações 235

Capítulo 10 Confundimento em ensaios fatoriais 269

Capítulo 11 Análise de grupo de experimentos 313

Capítulo 12 Regressão e correlação 345

Capítulo 13 Análise de covariância 381

Capítulo 14 Experimentos fatoriais fracionados: análise de variância e uso para estudo de superfícies de resposta 399

Capítulo 15 Estatística não paramétrica 423

Capítulo 16 Análise exploratória de dados 481

Capítulo 17 Princípios e técnicas de amostragem 499

Capítulo 18 Usando o SAS na análise de dados 521

Referências 547

Anexos 553

Introdução

A agricultura surgiu nos primórdios da humanidade e foi o agente fixador do homem, isto é, com a agricultura o homem começou a abandonar o nomadismo, pois a prática agrícola o obrigava a permanecer numa mesma área durante o período necessário para o completo desenvolvimento de suas culturas.

Junto com a agricultura, surgiu a investigação ou a pesquisa de novas tecnologias agrícolas. Evidentemente surgiu de forma empírica, observacional, sem qualquer rigorismo científico e, de forma geral, assim continua a ser praticada pelos agricultores até os dias de hoje.

Uma das primeiras tentativas de pesquisa agrícola registrada deu-se há três séculos, quando Johan Baptista van Helmont, de Bruxelas, plantou uma muda de salgueiro em um vaso ao qual adicionou apenas água durante 5 anos, findos os quais obteve uma árvore com 76,7 kg enquanto ao solo faltavam apenas umas poucas gramas de terra. Sua conclusão foi de que as árvores eram constituídas apenas por água, desprezando assim o solo faltante (KALCKMANN, 1958).

Na segunda metade do século 19, começaram a surgir por todo o mundo as primeiras estações experimentais que, no entanto, executavam investigação sem uma metodologia definida e procedimentos estatísticos adequados. A partir dos primeiros anos do século 20, com os trabalhos pioneiros de Sir Ronald Fisher na Estação Experimental de Rothamsted, Inglaterra, começou a ser estabelecida a estatística experimental baseada nos conhecimentos matemáticos e de probabilidade desenvolvidos ao longo dos séculos 18 e 19, principalmente.

Durante a primeira metade do século passado, a maioria dos métodos estatísticos baseou-se na suposição de que os dados seguiam a distribuição normal, ou Gaussiana, e compõe-se o que se convencionou chamar de estatística paramétrica.

A partir do final da década de 1940, os trabalhos pioneiros de Wilcoxon (1945) e de Mann e Whitney (1947) lançaram os fundamentos teóricos da estatística não paramétrica, que teve um grande desenvolvimento nos anos que se seguiram, motivada pela insatisfação de um grupo de pesquisadores com a estatística baseada na distribuição normal, dada sua convicção de que ela não ocorre na natureza. Na estatística não paramétrica, nenhuma

pressuposição é feita sobre a distribuição dos dados. Precisa-se somente estar certo de que as amostras aleatórias que se quer comparar pertencem à mesma população básica (WALTER, 1964, citado por SACHS, 1982).

Como dito anteriormente, a estatística experimental mais utilizada nos dias de hoje tem por base a distribuição normal. Essa distribuição possui fantásticas propriedades matemáticas e os testes nela baseados são gerais para os mais diversos procedimentos de implantação de uma pesquisa. Numa rápida incursão à teoria da probabilidade, verifica-se que a soma de variáveis normais tem distribuição normal; que o quadrado de uma variável normal ou a soma dos quadrados de uma série de variáveis normais tem distribuição de qui-quadrado, e que a razão entre duas variáveis de distribuição qui-quadrado tem distribuição de F. Além disso, a lei dos grandes números estabelece que, dada uma amostra grande de valores de uma distribuição qualquer que possua pelo menos média e variância, ocorre uma grande aproximação da distribuição normal, mesmo que a distribuição original seja do tipo discreto.

No entanto, a maioria dos testes paramétricos apresenta como principais defeitos a falta de robustez quando não ocorre uniformidade de variância entre as variáveis e a restrição de seu uso a dados cuja escala de medida é de intervalo ou de proporção, segundo a classificação apresentada por Stevens (1946), que os agrupa em quatro tipos de escala.

1) Escala nominal: o número assinalado à observação serve somente como um nome à categoria à qual a observação pertence. São exemplos: números de placas de automóveis e código de endereçamento postal; estado civil; classificação ocupacional e cor.

2) Escala de posição: o valor numérico da medida é usado somente como meio de arranjar os elementos medidos, em ordem, do menor para o maior. Podem ser citados nesta categoria os graus escolares, a ordem de produção e a intensidade da cor.

3) Escala de intervalo: o ponto zero é definido por convenção, e os intervalos podem ter significado empírico, como datas, quociente de inteligência, temperatura em graus Celsius ou Fahrenheit. O número assinalado à observação considera como informação pertinente não somente a ordem relativa das medidas, mas também o tamanho do intervalo entre as medidas.

4) Escala de proporção: o ponto zero é verdadeiro. Exemplos são as temperaturas medidas em graus Kelvin e as quantidades físicas medidas em unidades tais como metro, quilo, segundo. Nesta escala, a única diferença em relação à anterior é aquela relativa ao ponto zero de medida. Podem também ser incluídas nesta

escala as observações derivadas da ocorrência de algum fenômeno quantificado por meio de contagem.

Os métodos não paramétricos são também chamados de distribuição livre, já que os parâmetros dificilmente desempenham qualquer papel (hipóteses não paramétricas), tanto que se define um teste não paramétrico como aquele cujo modelo não especifica condições sobre os parâmetros da população da qual a amostra foi obtida (CAMPOS, 1983). Eles são, na maior parte dos casos, muito facilmente trabalhados numericamente. Além do mais, esses procedimentos podem também ser aplicados a dados oriundos de escalas nominais e de posição.

Para tornar mais claras as diferenças entre os requerimentos necessários para a aplicação desses dois tipos de estatísticas, as comparações de médias são um bom exemplo.

Um método clássico de comparação de médias, o teste t de Student, somente pode ser aplicado sob as seguintes condições:

- Os dados devem ser independentes (amostras aleatórias).
- A característica tem de ser mensurável em unidades de uma escala métrica.
- As populações devem ser, pelo menos aproximadamente, normalmente distribuídas.
- As variâncias têm de ser iguais.

Os procedimentos de distribuição livre para o mesmo problema simplesmente requerem a condição 1, a qual pode ser deduzida a partir do conhecimento da forma de como eles foram obtidos. Portanto, nós somente precisamos assumir que todos os dados ou pares de dados são retirados casual e independentemente um do outro e de uma mesma população básica, e isso tem de ser garantido pela estrutura de realização do experimento.

No entanto, segundo Pitman (1949), um teste de distribuição livre, quando aplicado a dados de uma família de distribuição conhecida, é sempre mais fraco que o teste paramétrico correspondente.

No uso dos testes não paramétricos as seguintes vantagens podem ser citadas (CAMPOS, 1983):

- São menos exigentes que os paramétricos.
- As probabilidades das afirmativas obtidas na maioria dos testes não paramétricos são, no geral, exatas, salvo quando se usam aproximações para grandes amostras.
- Independem da forma da população da qual a amostra foi obtida.

- São, em geral, de mais fácil aplicação e exigem quase sempre menor volume de cálculos.
- Permitem trabalhar com dados de diferentes populações, o que não é possível com os paramétricos.
- Podem ser aplicados a dados que não possuam escala de valores quantitativos.
- São mais eficientes do que os paramétricos quando os dados da população não têm distribuição normal. E, quando a população é normalmente distribuída, sua eficiência é levemente inferior à de seus competidores.

Algumas restrições podem ser apontadas:

- Em geral, os testes de distribuição livre não levam em consideração a magnitude dos dados. Em muitos casos isso se traduz num desperdício de informações.
- Quando todas as exigências do modelo estatístico são satisfeitas, os testes paramétricos são mais poderosos.
- Em geral, não permitem testar interações, salvo sob condições especiais de aditividade.
- A obtenção, utilização e interpretação das tabelas são, em geral, mais complexas.
- Para cada método ou procedimento experimental, existe um teste não paramétrico específico, isto é, eles não são de uso geral.

Neste livro, são abordados os delineamentos experimentais mais difundidos e em uso corrente, procedimentos e técnicas para sua implantação, coleta de dados e sua análise, e a aplicação dos testes de hipóteses e de comparação de médias. Faz-se também o estudo de correlação e de regressão, desde o mais simples modelo matemático (linha reta) até o estudo de superfícies de resposta. Para todos os procedimentos discutidos, faz-se a apresentação dos processos paramétricos e, quando for o caso, também os testes não paramétricos pertinentes que se encontram no Capítulo 15. A escolha entre qual desses dois aspectos – paramétrico e não paramétrico – deve ser usado cabe ao pesquisador. Essa decisão leva em conta o tamanho do experimento e o tipo de dados coletados bem como sua escala de medida.

Nesta segunda edição, incorporaram-se os capítulos 16, 17 e 18, referentes a uma breve introdução à amostragem, à análise exploratória de dados e à programação em Statistical Analysis System (SAS), respectivamente. No Capítulo 12, intitulado *Regressão e correlação*, foram adicionados alguns modelos especiais não lineares. Na maioria dos capítulos, foram introduzidas as programações SAS correspondentes.

Os capítulos podem ser agrupados em quatro grandes tópicos: introdução aos processos estatísticos, amostragem e delineamentos experimentais, estudos de regressão e aplicações não paramétricas. Aos leitores interessados no uso do SAS, a respeito do qual já possuam algum conhecimento, a leitura sequencial dos capítulos é rápida e sem dificuldades; porém, aos que não detêm conhecimento do SAS, recomenda-se o estudo do Capítulo 18 antes de se aventurar neste livro a partir do Capítulo 2.

Capítulo 1

Conceitos gerais e definições úteis

A estatística intervém na pesquisa e/ou no método científico por meio da experimentação e da observação. O uso da estatística como ferramenta deve estar presente desde o planejamento geral do projeto de pesquisa. Se um projeto de pesquisa vai produzir dados que devem ser analisados, um método estatístico deve fazer parte do delineamento do ensaio.

Projeto de pesquisa

De uma forma simplificada, as seguintes etapas devem ser seguidas na realização de qualquer investigação:

- Formulação do problema: a partir do conhecimento existente, deve-se definir muito claramente o problema ou aspecto desconhecido a ser investigado. Além disso, devem-se formular perguntas e hipóteses bem claras, decidir o que se vai estudar e que tipos de dados ou medidas deverão ser tomadas, bem como definir o tempo e o custo que será envolvido na sua realização.
- Delinear o experimento: definir o tamanho da amostra, os tratamentos a serem aplicados, a quantidade de dados a ser coletada, bem como as épocas de coletá-los, onde e como o ensaio será instalado, os métodos estatísticos a serem aplicados.
- Condução do ensaio e coleta dos dados: implantação do ensaio segundo a metodologia ou desenho estabelecido e coleta de dados de acordo com o cronograma definido. Essa é a parte que mais tempo consome do pesquisador, e a mais importante, pois, da precisão na coleta dos dados, dependem os resultados da etapa seguinte.

- Tabulação e análise de dados: os dados devem ser ordenados e analisados segundo o delineamento preconizado e os testes estatísticos aplicados como programado.
- Discussão dos resultados e conclusão: o investigador deve discutir os resultados obtidos de acordo com as análises efetuadas, segundo o conhecimento existente ou adquirido e elaborar as respostas às perguntas e hipóteses formuladas durante a fase de planejamento.

Definições gerais

No decorrer deste livro, uma série de termos serão empregados e, para simplicidade de entendimento, eles são definidos a seguir:

- Pesquisa e experimentação: para muitos, pesquisa e experimentação são sinônimos e, em certo sentido, assim podem ser considerados, mas, para ser mais preciso, o termo pesquisa deve sempre ser empregado quando se investigam coisas novas, enquanto o termo experimentação deve ser usado ao se verificar a adaptação de conhecimentos ou tecnologias a situações diversas daquelas nas quais foram criadas ou desenvolvidas. Assim, a criação de novas cultivares deve ser considerada como pesquisa, mas a realização de um ensaio de competição de linhagens e/ou cultivares em ambiente diverso àquele no qual foram criadas é uma experimentação.
- Fator: aquilo que se aplica em um ensaio de forma não homogênea. Por exemplo: cultivar, quando se testam várias delas; adubação, ao se compararem diversas formulações, etc.
- Nível: as diferentes manifestações de um fator. Por exemplo: as doses de adubação empregadas, os espaçamentos utilizados, as linhagens que são testadas, diferentes temperaturas de cocção, marcas de pneus, etc.
- Tratamento: cada um dos níveis de um fator ou cada uma das combinações dos níveis dos fatores quando testando mais de um fator.
- Testemunha: tratamento padrão de comparação. Pode ser a ausência do fator (dose zero de um adubo, por exemplo), ou a aplicação usual do fator (cultivar recomendada para cultivo na região, espaçamento adotado pelos agricultores, etc.).
- Ensaio ou experimento: o conjunto de todos os tratamentos, aplicados de forma repetida. Quando mais de um fator estiver sendo estudado, o ensaio é chamado de ensaio fatorial.

- Delineamento: o esquema adotado para a distribuição dos tratamentos.
- Unidade experimental: sujeito ao qual se aplica um dos tratamentos. Pode também ser chamada de parcela ou canteiro. Pode ser uma área de solo, um vaso, um animal, uma placa de Petri, um indivíduo, a posição de montagem de um pneu, etc.
- Área útil: porção da unidade experimental efetivamente utilizada na avaliação do tratamento.
- Bordadura: parte da unidade experimental não coletada para avaliação do efeito do tratamento. É empregada para evitar efeito de competição ou de contaminação entre parcelas vizinhas. Normalmente é constituída pelo mesmo material da área útil. Em alguns casos pode ser montada com materiais distintos, como no caso de alguns experimentos em fitopatologia, por exemplo, em que as bordaduras de blocos são semeadas antecipadamente para que sirvam de fonte de inóculo para o ensaio, como forma de garantir maior uniformidade de infecção.
- Repetição: cada uma das aplicações de um tratamento.
- Bloco: conjunto ambiental homogêneo que contém todos os tratamentos ou parte deles. No delineamento em quadrado latino, em que se apresentam dois tipos de blocos, eles recebem as denominações de colunas e linhas.

Em qualquer ensaio ou experimento conduzido, medidas são tomadas para representar o efeito dos tratamentos em estudo. Essas medidas são os diversos valores que as variáveis associadas aos objetos sendo testados podem assumir. Elas estão sempre sujeitas a erro ou variações, que são devidas à atuação de fatores não controlados sobre as unidades experimentais, como, por exemplo, pequenas variações no ambiente ou material ambiental, as quais passam despercebidas ou não são conhecidas pelo pesquisador. Além disso, erros podem ser cometidos na tomada dos dados, como na leitura de um equipamento, balança ou trena, por exemplo. Neste último caso, são chamados de erros grosseiros. Não importa o tipo de erro que se comete, o seu efeito sobre as variáveis representam o erro experimental. Os delineamentos experimentais, então, devem ser escolhidos de forma que esses erros associados às condições ambientais sejam minimizados. Por sua vez, em relação aos erros grosseiros, apenas um maior cuidado ou treinamento na tomada dessas medidas é requerido.

Os delineamentos para a minimização dos efeitos das variações que ocorrem no ambiente em que se conduz um ensaio são estruturados segundo alguns princípios básicos da experimentação, que são a casualização, a repetição e o controle local.

- Casualização: é a atribuição dos tratamentos às unidades experimentais de forma aleatória, sem nenhuma interferência por parte do pesquisador. Para cada tipo de delineamento experimental, existe uma forma adequada de se fazer essa

alocação dos tratamentos, que será vista adiante à medida que for apresentado cada um dos desenhos experimentais.

- Repetição: refere-se à implantação de cada tratamento em mais de uma unidade experimental. Sua utilização está vinculada à necessidade de avaliação da variabilidade do material experimental.
- Controle local: está associado ao conhecimento do ambiente experimental e consiste na divisão das parcelas experimentais em subconjuntos homogêneos, quando é sabido que o total das unidades experimentais não possui a homogeneidade exigida. No geral, o conjunto de parcelas homogêneas constitui o que se convencionou chamar de bloco. Quando o bloco contém todos os tratamentos uma única vez, ele é chamado de bloco completo, e os delineamentos experimentais que apresentam essa característica são ditos em blocos completos. Se os blocos comportam apenas parte dos tratamentos, os delineamentos são chamados de delineamentos em blocos incompletos. Evidentemente, variações nessas situações são permitidas, como aquela em que, ao serem testados poucos tratamentos, todos podem ser aplicados mais de uma vez no mesmo bloco (PIMENTEL-GOMES, 2000); ou, como no caso do chamado delineamento em blocos aumentados, em que alguns tratamentos (testemunhas) aparecem em todos os blocos enquanto outros em apenas um dos blocos (FEDERER, 1955).

A cada experimento pode e deve ser associado um modelo matemático, o qual deve conter, dependendo do contexto e do tratamento estatístico a ser empregado, os efeitos dos fatores e dos erros que afetam cada observação, ou os parâmetros e outras variáveis que lhes são associadas, ou ainda uma combinação dessas duas situações. No primeiro caso, trata-se de uma simples análise de variância de acordo com o delineamento adotado; no segundo, de um estudo de regressão; e o último, de uma análise de covariância, em que o delineamento e um estudo de regressão para uma ou mais variáveis adicionais estão associados. A definição correta do modelo matemático é essencial para a realização da análise de variância e para a utilização de softwares estatísticos. Esses modelos serão apresentados e discutidos ao longo dos capítulos.

Estatística experimental paramétrica: pressuposições básicas

Para a realização das análises estatísticas paramétricas, algumas pressuposições básicas são necessárias, e elas devem ser obedecidas, pelo menos aproximadamente, sob pena de interpretações ou conclusões erradas serem tomadas.

As principais delas são: os erros são independentes, com distribuição normal, com média $\mu = 0$, variância σ^2, e o modelo matemático é linear e aditivo. Ademais, exige-se que os erros tenham todos a mesma variância e que os dados sejam quantitativos, isto é, que se enquadrem nas escalas de proporção ou de intervalo.

Deve o pesquisador ter em conta que os erros nunca podem ser eliminados, mas tão somente minimizados, e isso se consegue por meio de um conjunto de atividades ou procedimentos, entre as quais se destacam:

- Uniformização das parcelas experimentais.
- Parcela experimental de tamanho adequado.
- Uso de bordaduras.
- Utilização de um número adequado de repetições e de preferência igual para todos os tratamentos.
- Manejo das unidades experimentais de forma homogênea, no que diz respeito a todos os fatores não envolvidos no estudo.
- Uso do delineamento estatístico adequado para as condições de realização do experimento, de forma a obter o melhor aproveitamento dos resultados.

Quando as variâncias não são iguais ou aproximadamente iguais, ou o modelo não é absolutamente aditivo, ou os dados não são do tipo quantitativo, algumas transformações nos dados podem ser aplicadas para que essas condições se verifiquem (ver item *Transformação de dados*). Porém, em muitos casos, isso não é possível ou efetivo e então os processos não paramétricos devem ser aplicados.

Estatísticas descritivas

A cada conjunto de dados levantado em uma pesquisa, ocorre a necessidade de resumir a informação contida em cada um deles. Isso pode ser feito por processos gráficos ou por estatísticas elementares, tais como as medidas de tendência central e de dispersão. As distintas formas de representação gráfica serão apresentadas e discutidas no Capítulo 16, enquanto as estatísticas descritivas serão apresentadas a seguir, devendo-se considerar que algumas são de grande interesse na estatística experimental e outras de maior aplicação na estatística amostral.

Medidas de tendência central ou de posição

As medidas de tendência central mais conhecidas e utilizadas são a média, a mediana e a moda, que, numa distribuição simétrica, são coincidentes. Além dessas, os quantis são

de grande importância na interpretação de resultados em trabalhos de amostragem e/ou censo. Os quantís compreendem valores que representam certas porcentagens dentro de uma amostra ordenada. Em geral, é possível dividir os quantís em três grupos, segundo seu uso: quartis (frações de 25%), decis (frações de 10%) e percentis (porções variáveis de porcentagens).

A moda é o valor mais frequente em uma série de valores. Se existir apenas um valor mais frequente, a amostra ou população é dita unimodal. Se mais de um valor apresenta maior frequência, a amostra será bimodal (dois valores) ou multimodal (mais de dois valores). Quando todos os valores aparecem uma única vez, fala-se em amostra ou população amodal.

Exemplo 1:

1 2 3 3 4 4 4 5 6 7

Trata-se de uma série de dados considerada unimodal, pois o valor 4 aparece com frequência 3, e é o único valor com essa frequência.

Exemplo 2:

1 2 3 3 4 4 5 5 5 6 6 6 7 8

Aqui a série deve ser considerada bimodal, pois dois valores apresentam frequência maior que os demais, isto é, os valores 5 e 6 são as duas modas da amostra.

A mediana é o valor que divide uma amostra ou população finita ordenada em duas partes. É o valor central em uma série com número ímpar de elementos ou a média aritmética dos dois valores centrais em uma série com número par de elementos. É chamada também de quartil 2, decil 5 ou percentil 50.

$$\tilde{x} = \begin{Bmatrix} x_{([n+1]/2)}, & \text{ímpar} \\ \dfrac{x_{(n/2)} + x_{([n/2]+1)}}{2}, & \text{par} \end{Bmatrix} \qquad [1]$$

No caso de censo, substitui-se o valor de n, tamanho da amostra, por N tamanho da população.

Exemplo 3:

1 2 3 4 5 6 7 9 12

A mediana é o número 5, pois tem quatro elementos à sua esquerda e quatro à sua direita.

Exemplo 4:

1 2 3 4 6 7 8 10 11 12

Nesse caso, a mediana é a média aritmética de 6 e 7, que são, respectivamente, o quinto e sexto elementos da amostra ordenada com dez valores; portanto, a mediana é igual a 6,5.

A média ou média aritmética é o valor obtido pelo quociente entre a soma dos valores de uma amostra ou população e o seu número.

No caso da população:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad [2]$$

No caso de amostra:

$$\overline{X} = m = \frac{\sum_{i=1}^{n} X_i}{n} \qquad [3]$$

Exemplo 5:

1 2 3 3 4 4 4 5 6 7

Para essa amostra, tem-se que:

$$\overline{X} = \frac{1+2+3+3+4+4+4+5+6+7}{10} = \frac{39}{10} = 3{,}9$$

Verifica-se pelas definições dadas para moda e mediana que a existência de valores extremos não afeta o seu valor, o que já não ocorre com a média aritmética, que é

fortemente afetada pela existência de valores extremos na amostra ou série de valores sendo analisada. Para o exemplo dado anteriormente, se o valor máximo, 7, fosse na realidade 27, por exemplo, o valor da média seria 5,9, bem diverso do obtido (3,9), enquanto a moda e a mediana permaneceriam constantes em 4. Apesar disso, a média aritmética é o estimador de mínimo quadrado, não viciado, da média de uma população.

Os quartis, por sua vez, representam os pontos que dividem a série de valores em quatro partes iguais, ou seja, quartil 1 ou 25%, quartil 2 ou 50% (mediana) e quartil 3 ou 75%. O quartil 2, a mediana, divide o conjunto de dados em duas partes que contêm cada uma 50% dos valores, como se viu anteriormente. Os quartis 1 e 3 podem ser considerados as medianas dos 50% inferiores e dos 50% superiores dos dados, e assim devem ser calculados. Quando o número de elementos na amostra é ímpar, a mediana deve ser incluída nas duas metades, inferior e superior. Outras definições existem e em grande número.

Os decis dividem o conjunto de valores da amostra em dez partes iguais, e são designados como decil 1, decil 2, até decil 9, que representam os 10%, 20%, e assim sucessivamente até 90%.

Os percentis, por sua vez, dividem o conjunto de dados em 100 partes iguais, sendo considerados os mais importantes os percentis 1, 5, 10, 25, 50, 75, 90, 95 e 99, que representam os valores correspondentes a 1%, 5%, 10%, etc.

Os outros valores que quantificam uma amostra e servem também como medidas de posição são os valores mínimo e máximo, os quais, logicamente, representam o menor e o maior valor da série de dados. Algumas vezes, eles são também chamados de quartil 0 e quartil 4, ou de decil 0 e decil 10, assim como percentil 0 e percentil 100.

Medidas de dispersão

A variabilidade das medidas de uma variável é obtida com o uso das medidas de dispersão, sendo as mais empregadas, a variância, o desvio padrão e o coeficiente de variação.

A variância é obtida pelo quociente entre a soma dos quadrados dos desvios de cada medida em relação à sua média e o número de graus de liberdade (amostra) ou pelo número de indivíduos (população).

No caso de amostras, emprega-se a fórmula 4; e no caso de populações, a fórmula 5.

$$s^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n - 1} \qquad [4]$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} \left(y_i - \bar{y} \right)^2}{N} \quad [5]$$

Formas de mais fácil computação são possíveis de ser obtidas, como se mostra nas equações 6 e 7, a seguir.

$$s^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i \right)^2}{n - 1} \quad [6]$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i \right)^2}{N} \quad [7]$$

O desvio padrão é a raiz quadrada da variância, enquanto o erro típico, também chamado de erro padrão da média (*EEM*), é igual à raiz quadrada do quociente entre a variância da amostra e o número de valores existentes na amostra (*n*), ou seja,

$$EEM = \sqrt{\frac{s^2}{n}} \quad [8]$$

Coeficiente de variação (*CV*) é quociente entre o desvio padrão e a média, expresso em porcentagem.

$$CV = \frac{s}{m} * 100 \quad [9]$$

Dada a amostra anterior de dez elementos, podem-se obter estas três medidas de dispersão.

1 2 3 3 4 4 4 5 6 7

A variância então é

$$s^2 = \frac{\left(1^2 + 2^2 + 3^2 + 3^2 + 4^2 + 4^2 + 4^2 + 5^2 + 6^2 + 7^2 \right) - {(39)^2}/{10}}{9} = 3{,}2111$$

O desvio padrão é

$$dp = \sqrt{3,2111} = 1,791957$$

e o coeficiente de variação é

$$CV = \frac{1,791957}{3,9} * 100 = 45,95\% .$$

O coeficiente de variação é a expressão da precisão de um experimento. Cada variável possui uma escala própria de coeficientes de variação que está relacionada à sua variabilidade intrínseca. Vários autores têm procurado estabelecer tabelas que associem os valores do CV aos conceitos: baixo, médio, alto e muito alto (AMARAL, et al 1997; COSTA et al., 2002; ESTEFANEL et al., 1987; GARCIA, 1989; PIMENTEL-GOMES, 2000; SCAPIM et al., 1995). Para cada produto e variável, uma tabela pode ser construída com limites bem diversos de uma para a outra. Para a produção de grãos, a tabela apresentada por Pimentel-Gomes (2000) parece bem adequada, mas não para outras variáveis, como massa de 1.000 grãos, altura de plantas, sobrevivência ao ataque de pragas, etc. Para massa de 100 grãos, por exemplo, valores adequados (baixos) de CV estão situados entre 3% e 8%, enquanto valores acima de 10% devem ser considerados altos. Já para sobrevivência de plantas ao ataque de pragas e incidência de brusone, valores em torno de 20% podem ser considerados médios (COSTA et al., 2002).

A amplitude é a diferença entre os valores máximo e mínimo da amostra e representa a amplitude de variação dos dados. Dessa forma, a distância interquartílica (DI) é a diferença entre o quartil 3 e o quartil 1 e representa a amplitude de dados que contém 50% dos valores da amostra. Valores pequenos para a amplitude indicam pequena dispersão dos dados, logo pequenos valores para a distância interquartílica indicam alta concentração dos dados ao redor de seu centro (média ou mediana).

A assimetria ou coeficiente de assimetria (CA) é a estatística que indica a forma de distribuição dos dados com respeito à média. Dados igualmente distribuídos em ambos os lados da média têm alta simetria. O valor do coeficiente de assimetria é zero ou perto de zero, assim como o conjunto de dados com valores de CA diferentes de zero indicam valores mais concentrados à esquerda da média com valores do coeficiente de assimetria positiva (assimétricos à direita) ou valores mais concentrados à direita da média (assimétricos à esquerda), portanto com assimetria negativa. A fórmula de obtenção de assimetria é:

$$CA = \frac{n}{(n\text{-}1)*(n\text{-}2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3 \qquad [10]$$

A curtose representa o achatamento ou a elevação da curva dos dados quando comparada com uma distribuição normal. Curvas muito elevadas apresentam curtose positiva e curvas achatadas, curtose negativa.

$$Curtose = \left\{ \frac{n(n+1)}{(n-1)\,(n-2)\,(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{n(n-1)^2}{(n-2)\,(n-3)} \qquad [11]$$

Hipóteses, testes de hipóteses e nível de significância

Ao executar um experimento, o pesquisador busca comprovar alguma hipótese que estabeleceu. Dois tipos de hipóteses são formuladas: a hipótese nula (H_0), segundo a qual se declara a não existência de efeito do que se está testando e a hipótese alternativa (H_1), que declara a existência desse efeito e pode ser do tipo bilateral (diferente) ou unilateral (maior que ou menor que). É evidente que essas hipóteses são formuladas segundo o tipo de pressuposição adequada à análise a que se propõe.

Na aplicação de um teste estatístico qualquer, associam-se dois tipos de erros: erro tipo I e erro tipo II. A Tabela 1 resume a situação.

O erro tipo I ocorre ao rejeitar-se uma hipótese nula que é verdadeira, e o erro tipo II ao se aceitar uma hipótese nula falsa. O erro tipo I é também chamado de significância de um teste, não havendo nome específico para o erro tipo II. No entanto, $1 - \beta$ é chamado de poder de um teste.

Num estudo de regressão linear, por exemplo, as hipóteses nula e alternativa poderiam ser: $H_0 : b = 0$ e $H1 : b \neq 0$.

Ou seja, a hipótese nula declara que o valor do coeficiente de regressão, ou coeficiente angular, é igual a zero, enquanto a alternativa que se quer comprovar estabelece que o valor de b é diferente de zero.

Tabela 1. Tipos de erros na tomada de decisões.

Hipótese	Decisão	
	Aceitar	Rejeitar
H_0	Correta	Erro tipo I (α)
H_1	Erro tipo II (β)	Correta (poder)

Já num experimento tradicional, assumindo I tratamentos e seus efeitos como fixos, as hipóteses nulas e alternativas podem ser formuladas por: $H_0 : t_i = 0$ para todo i, em que $i = 1, 2, 3,...,I$; ou $H_0 : t_i = t_j$ para todo $i \neq j$; e $H_1 : t_i \neq 0$ para algum i; ou $H_1 : t_i \neq t_j$ para algum $i \neq j$.

O teste F é, neste caso, o indicado para a comprovação ou não da validade das hipóteses. No caso da regressão, nenhum outro teste necessita ser aplicado, mas, para o caso de experimentos, se existem mais de dois tratamentos, o fato de o teste F ser significativo apenas indica que existe alguma diferença entre os tratamentos, mas não especifica onde se encontra(m) a(s) diferenças(s). Dessa forma, outros testes devem ser empregados para detectá-la(s). São os testes para comparação de médias, discutidos na seção intitulada *Testes para comparação de médias*.

O teste F é calculado obtendo-se o quociente entre o quadrado médio do fator que se está testando (regressão ou tratamentos) e o quadrado médio do erro experimental, usando as expressões comuns de um quadro de análise de variância que será discutido a partir do Capítulo 2. Os quadrados médios representam a estimativa das variâncias respectivas, isto é, variância do fator e variância do erro. Sua significância se verifica ao comparar-se o valor calculado com os valores tabelados (Anexos 3 e 4), nos níveis de probabilidade disponíveis, segundo os graus de liberdade associados a cada variância.

Por convenção, assume-se que um teste é significativo quando a probabilidade a ele associada é igual ou menor que 5%. É lógico que esse nível de significância pode ser modificado de acordo com o tipo de trabalho que se executa, com os riscos envolvidos na aplicação dos resultados ou por outros motivos. Num experimento com drogas para cura de alguma doença, por exemplo, em que pode haver risco de morte do paciente por aplicação de dosagem errada, o nível de significância deve ser muito menor que 5%, por exemplo, 0,01%, enquanto em trabalhos com cultivares, nos quais pode ocorrer a substituição de um material por outro equivalente, o nível de significância poderia ser maior que 5%. O que se quer afirmar aqui é que o pesquisador deve definir o nível de significância adequado para o seu trabalho, segundo seus conhecimentos, mas esse nível deve ser definido a priori, isto é, antes da execução do ensaio e da coleta de qualquer dado.

Testes para comparação de médias

Verificada a existência da significância do teste F, em um experimento em que foram testados mais de dois tratamentos do tipo qualitativo, ocorre a necessidade de definir-se onde estão ocorrendo as diferenças entre eles, o que deve ser feito pela aplicação de um teste para comparação de médias ou pelo uso de contrastes. Existem hoje mais de uma dezena desses testes, mas os mais comuns são: t ou DMS, Tukey ou DHS, Duncan, Student-Newman-Keuls ou SNK, Scheffé e Dunnett. Cada um deles apresenta vantagens

e desvantagens quando comparados entre si e podem ser empregados nas comparações entre todos os pares de tratamentos (Tukey, Duncan, SNK e DMS), ou entre grupos de tratamentos (Scheffé, DMS e Scott-Knott), ou na comparação de cada tratamento contra um deles (testemunha), que é o caso do teste de Dunnett. A escolha do teste a ser empregado depende única e exclusivamente do pesquisador com base no tipo de hipótese formulada. As condições descritas anteriormente não são fixas, sendo possível utilizar Tukey para comparar grupos de médias ou Scheffé para comparar médias duas a duas, como foi apontado por Scheffé (1959). No uso de contrastes para comparação dos efeitos de tratamentos, por combinação de grupo de médias ou pares de médias, o teste F é o indicado.

Uso de contrastes para comparação de médias

Define-se contraste como um vetor de médias em que a soma dos coeficientes é igual a zero. Dois ou mais contrastes são ditos ortogonais quando a soma dos produtos dos coeficientes é igual a zero. O emprego de contrastes ortogonais é interessante para o caso da aplicação dos testes DMS e F para comparação de médias.

Dadas quatro médias ($m1$, $m2$, $m3$ e $m4$), por exemplo, um vetor explicitado por $1*m1 + (-1)*m2 + 0*m3 + 0*m4$ é um contraste, pois a soma dos coeficientes $1 -1 + 0 + 0 = 0$ permite a comparação das médias de dois tratamentos (1 x 2).

Da mesma forma o vetor $1*m1 + 1*m2 + (-2)*m3 + 0*m4$ é um contraste, pois $1 + 1 - 2 + 0 = 0$, em que se compararam a média dos tratamentos 1 e 2 contra o tratamento 3.

Ademais, esses dois contrastes são ortogonais, pois $1*1 + (-1)*1 + 0*(-2) + 0*0 = 0$.

No caso de uso do teste F, a soma de quadrados de um contraste (SQC) é calculada segundo a expressão

$$SQC = \frac{\left(\sum_i c_i * T_i\right)^2}{r * \sum_i c_i^2} \qquad [12]$$

em que r é o número de repetições de cada tratamento, c_i são os i coeficientes do contraste e T_i os totais dos i tratamentos. A cada contraste corresponde um grau de liberdade.

A variância de um contraste (V_c) é dada por

$$V_c = \frac{QME}{r} \sum_i c_i^2 \qquad [13]$$

assumindo todos os tratamentos com número igual de repetições (r), QME o quadrado médio do erro e c_i os coeficientes do contraste.

Exemplo: Suponha-se um ensaio de arroz, com seis cultivares e quatro repetições em blocos completos. O erro possuía, portanto, 15 graus de liberdade, e o valor do seu quadrado médio foi igual a 189.656,62. O CV do ensaio foi de 19%. Os seis tratamentos são compostos por três linhagens, CNA 8548 (1), CNA 8674 (4) e CNA 8676 (6), e por três cultivares em uso geral, Xingu (2), Progresso (3) e Maravilha (5), e as médias dos tratamentos 1 a 6, em kg ha^{-1}, foram iguais a 2.991,5, 2.499,5, 2.258,3, 2.070,5, 1.995,5 e 1.937,3, respectivamente.

Assumindo que um contraste de interesse seria o de comparação entre as linhagens e as cultivares, ter-se-ia: $C1 = -T1 + T2 + T3 - T4 + T5 - T6$, em que os valores de c_i são -1, 1, 1, -1, 1, -1 e a soma de quadrados desse contraste é igual a:

$$SQC1 = (-2.991,5 + 2.499,5 + 2.258,3 - 2.070,5 + 1.995,5 - 1.937,3)^2/(4 \times 6)$$

$$SQC1 = (-246)^2/24 = 2.521,5$$

Logo, o teste de F seria o quociente de $2.521,5/189.656,62 = 0,01$ com 1 e 15 graus de liberdade e, portanto, não significativo, pois se verifica, no Anexo 3, que o valor tabelado é igual a 4,54, muito maior que o calculado. Isso indica que não se pode rejeitar a hipótese de igualdade entre linhagens e cultivares.

Teste t ou diferença mínima significativa (DMS)

Este teste apresenta a seguinte forma

$$DMS = t_{\alpha,GLE} * s_d \qquad [14]$$

em que $t_{\alpha,GLE}$ representa o valor tabelado do teste t (Anexo 2), para um nível de probabilidade α e um dado número de graus de liberdade do erro (GLE). S_d é a raiz quadrada da variância da diferença entre duas médias. No caso de um experimento em que todos os tratamentos têm o mesmo número de repetições, s_d é dada por

$$s_d = \sqrt{\frac{2 * QME}{r}} \qquad [15]$$

em que QME representa o quadrado médio do erro e r o número de repetições.

Recordando o estudo de contrastes, visto anteriormente, percebe-se claramente que s_d é a raiz quadrada da variância do contraste entre duas médias. Se as médias que são comparadas apresentam número de repetições diferentes, a fórmula adequada é dada por

$$s_d = \sqrt{QME * \left(\frac{1}{r1} + \frac{1}{r2} \right)} \qquad [16]$$

em que $r1$ e $r2$ representam esses números de repetições.

Exemplo: para o ensaio de arroz, anteriormente discutido, a aplicação do teste de t será como mostrado a seguir: o valor de t, para 5% de probabilidade e 15 GL, é igual a 2,13, tendo-se então

DMS = 2,13*307,942 = 655,92

já que

$$s_d = \sqrt{\frac{2*189.656,62}{4}} = 307,942$$

Isso mostra que apenas as diferenças entre quaisquer duas médias só serão consideradas significativas se foram maiores que 655,92 e lembrando que as médias dos tratamentos 1 a 6 foram iguais a: 2.991,5, 2.499,5, 2.258,3, 2.070,5, 1.995,5 e 1.937,3, respectivamente. Verifica-se que os tratamentos 2, 3, 4, 5 e 6 devem ser considerados iguais entre si e que o tratamento 1 difere dos tratamentos 3, 4, 5 e 6, pois apenas essas diferenças superaram o valor do DMS de 655,92, como segue:

$t1$ - $t2$ = 2.991,5 - 2.499,5 = 492

$t1$ - $t3$ = 2.991,5 - 2.258,3 = 733,2

$t1$ - $t4$ = 2.991,5 - 2.070,5 = 921

$t1$ - $t5$ = 2.991,5 - 1.995,5 = 996

$t1$ - $t6$ = 2.991,5 - 1.937,3 = 1.054,2

$t2$ - $t3$ = 2.499,5 - 2.258,3 = 241,2

$t2$ - $t4$ = 2.499,5 - 2.070,5 = 429

$t2$ - $t5$ = 2.499,5 - 1.995,5 = 504

$t2$ - $t6$ = 2.499,5 - 1.937,3 = 562,2

*t*3 - *t*4 = 2.258,3 - 2.070,5 = 187,8

*t*3 - *t*5 = 2.258,3 - 1.995,5 = 262,8

*t*3 - *t*6 = 2.258,3 - 1.937,3 = 321

*t*4 - *t*5 = 2.070,5 - 1.995,5 = 75

*t*4 - *t*6 = 2.070,5 - 1.937,3 = 133,2

*t*5 - *t*6 = 1.995,5 - 1.937,3 = 58,2

A representação mais tradicional dos resultados de uma comparação de médias é feita por meio da aposição de letras após o valor da média, dando-se letras iguais para médias que não são consideradas diferentes, a partir da maior média. No caso presente, tem-se:

*t*1 - 2.991,5 a

*t*2 - 2.499,5 ab

*t*3 - 2.258,3 b

*t*4 - 2.070,5 b

*t*5 - 1.995,5 b

*t*6 - 1.937,3 b

Teste de Tukey ou DHS

Este teste, desenvolvido por J. W. Tukey (1955), é baseado na amplitude total estudentizada. É considerado um dos testes mais conservadores para a comparação de médias por manter constante seu erro tipo I não importando o número de médias sendo testadas. A sigla DHS é a tradução da original HSD, do inglês *Honestly Significant Difference* (Diferença Honestamente Significante). Scheffé (1959), considerou o DHS superior ao seu próprio teste para as comparações de médias, quando tomadas duas a duas, mas inferior para outros tipos de comparações.

Apresenta a forma de cálculo

$$DHS = q_{\alpha,n,GLE} * \sqrt{\frac{QME}{r}} \qquad [17]$$

em que $q_{\alpha,n,GLE}$ é o valor tabelado da amplitude total estudentizada a um nível de probabilidade α (Anexos 9 e 10), para n médias e número de graus de liberdade do erro (GLE); ou seja a constante é multiplicada pela raiz quadrada da metade da variância do contraste entre duas médias e, como será visto a seguir, isso também se verifica para os testes de Duncan e de Student-Newman-Keuls.

Para o exemplo anterior em blocos ao acaso com seis tratamentos e quatro repetições, com 15 GL para o erro e $QME = 189.656,62$, o valor do teste, a 5% de probabilidade, é

$$DHS = 4,60\sqrt{\frac{189.656,62}{4}} = 1.001,64$$

Utilizando-se as diferenças anteriormente calculadas, verifica-se que apenas uma diferença ultrapassa o valor de DHS calculado, que é aquela entre os tratamentos 1 e 6. Adotando o mesmo princípio de representação com letras, tem-se:

$t1$ - 2.991,5 a

$t2$ - 2.499,5 ab

$t3$ - 2.258,3 ab

$t4$ - 2.070,5 ab

$t5$ - 1.995,5 ab

$t6$ - 1.937,3 b

No caso de os tratamentos apresentarem número distinto de repetições, pode ser empregada uma aproximação deste teste, fazendo-se a substituição do valor r da fórmula, pelo fator $2*(r1*r2)/(r1 + r2)$, em que $r1$ e $r2$ são os respectivos números de repetição dos dois tratamentos sendo comparados.

Teste de Duncan ou teste de amplitudes múltiplas

Em 1955, Duncan apresentou este teste para comparação de médias adotando o princípio de múltiplas amplitudes, isto é, existem tantos valores de diferença mínima quantas forem as médias menos 1, e esses valores são tanto maiores quanto mais médias estiverem envolvidas na comparação. Esse critério adotado por Duncan implica que, ao se compararem duas médias quaisquer, devem ser consideradas não apenas aquelas médias, mas também todas as demais que se situarem entre elas, isto é, ao se comparar a primeira média com a quarta, deve-se considerar que a comparação envolve quatro médias, pois a segunda e terceira médias se localizam entre elas.

A forma do teste de Duncan é

$$D_i = z_{\alpha,i,GLE} * \sqrt{\frac{QME}{r}} \quad [18]$$

em que $z_{\alpha,n,GLE}$ é o valor tabelado da amplitude estudentizada (Anexo 11), no qual i indica o número de médias envolvidas na comparação e varia de 2 até n, o número total de médias; GLE representa o número de graus de liberdade do erro e α o nível de significância adotado; QME e r têm o mesmo significado anterior. Os valores D_i são crescentes, isto é, quanto maior o valor de i maior o valor de D_i. O teste de Duncan é menos conservador que o teste de Tukey. Portanto, pode apontar mais diferenças significativas que este último.

Na sua aplicação, começam-se as comparações entre os valores extremos, isto é, primeiro a maior média com a menor delas, depois com a segunda menor, e assim sucessivamente até que uma das diferenças seja não significativa. A seguir, toma-se a segunda maior para comparação com a menor e repete-se o processo até que não mais diferenças significativas possam ser encontradas. Deve-se ter em mente que, se uma diferença entre médias extremas é não significativa, não é permitida a busca de diferenças significativas entre os valores intermediários.

Para o exemplo usado anteriormente, seis tratamentos, quatro repetições e erro com 15 GL e quadrado médio igual a 189.656,62, os valores do teste de Duncan a 5% de probabilidade, para 2, 3, 4, 5 e 6 médias envolvidas nas comparações são:

$$D_6 = 3{,}36\sqrt{\frac{189.656{,}62}{4}} = 731{,}63$$

$$D_5 = 3{,}31\sqrt{\frac{189.656{,}62}{4}} = 720{,}75$$

$$D_4 = 3{,}25\sqrt{\frac{189.656{,}62}{4}} = 707{,}68$$

$$D_3 = 3{,}16\sqrt{\frac{189.656{,}62}{4}} = 688{,}08$$

$$D_2 = 3{,}01\sqrt{\frac{189.656{,}62}{4}} = 655{,}42$$

A diferença entre a maior média (1) e a menor (6) é igual a 1.054,2, portanto maior que D_6; logo, significativa. Da mesma forma, as diferenças entre a média 1 e as médias 5,

4 e 3 devem ser consideradas significativas por serem maiores que D_5, D_4 e D_3, respectivamente, enquanto a diferença entre 1 e 2 é menor que o valor do teste de Duncan para 2 médias (D_2); portanto, não significativo. Na sequência, faz-se a comparação entre as médias 2 e 5, sendo a diferença entre elas menor que D_5, logo nenhuma outra comparação do tratamento 2 deve ser executada. O processo deve ser então repetido para a média 3 e assim sucessivamente, até não haver mais possibilidades de comparação. Utilizando-se as letras para identificar as diferenças significativas, ter-se-ia o resultado abaixo, e nesse caso, idêntico ao obtido com o LSD.

*t*1 - 2.991,5 a

*t*2 - 2.499,5 ab

*t*3 - 2.258,3 b

*t*4 - 2.070,5 b

*t*5 - 1.995,5 b

*t*6 - 1.937,3 b

Neste teste, havendo número diferente de repetições para as diversas médias, pode ser feita a substituição de r pelo quociente dado por $(r1 + r2)/(r1*r2)$, em que $r1$ e $r2$ são os números de repetição dos tratamentos considerados na comparação. No entanto, deve ser lembrado que existe a possibilidade de ter-se mais de um valor para cada tipo de comparação, segundo o número de repetições que tenham os tratamentos considerados. Por exemplo, se no caso acima o tratamento 2 tivesse 1 repetição a menos que os demais, seria necessário calcular dois valores para D_5, D_4, D_3 e D_2.

Teste de Student-Newman-Keuls ou SNK

Assim como no teste de Duncan, este teste faz a aplicação de múltiplas amplitudes na comparação de médias. A diferença básica entre eles é o uso da amplitude q (Anexos 9 e 10), em substituição ao valor z (Anexo 11) nas mesmas condições do teste de Duncan.

O teste SNK apresenta a forma

$$SNK_i = q_{\alpha,i,GLE} * \sqrt{\frac{QME}{r}} \qquad [19]$$

em que os símbolos têm a mesma conotação dada anteriormente. Note-se que o valor deste teste para comparação entre os valores extremos é idêntico ao do teste de Tukey.

Para o exemplo que se vem discutindo, têm-se os seguintes valores para o teste SNK:

1	2	3	4	5	6
$q_{0,05,\ 15}$	3,01	3,67	4,08	4,37	4,60
SNK	655,42	799,13	888,41	951,56	1.001,64

A sua aplicação de forma similar à feita no teste de Duncan leva ao seguinte resultado:

t1 - 2.991,5 a

t2 - 2.499,5 ab

t3 - 2.258,3 ab

t4 - 2.070,5 b

t5 - 1.995,5 b

t6 - 1.937,3 b

Percebe-se claramente que seu resultado é intermediário entre aqueles obtidos com os testes de Tukey e o de Duncan, como seria de se esperar.

Teste de Scheffé

Este teste é similar ao de Tukey, mas dele difere por utilizar os valores do teste F (Anexos 3 e 4), e não os valores de amplitude estudentizada (q) (Anexos 9 e 10), mas, como afirmado por Scheffé (1959), este teste é superior ao de Tukey para comparar mais de duas médias, e inferior nesse caso particular.

O teste apresenta a forma

$$S = \sqrt{(n\text{-}1)\mathrm{F}_{\alpha,n-1,GLE}}\ \sqrt{\frac{2QME}{r}} \qquad [20]$$

em que n é o número de tratamentos, r o número de repetições, GLE e QME são, respectivamente, o número de graus de liberdade e o quadrado médio do erro ou resíduo, e α o nível de significância desejado.

No caso de comparação entre mais de duas médias, substitui-se a expressão $\sqrt{\frac{2QME}{r}}$, pela expressão correspondente ao erro padrão do contraste desejado, lembrando que esse valor é a raiz quadrada da variância do contraste, como visto anteriormente.

Continuando com o exemplo que se está empregando até agora, com 15 GLE, 189.656,62 de QME, seis tratamentos e quatro repetições, o valor do teste de Scheffé é igual a:

$$S = \sqrt{(6\text{-}1)*2{,}90}\sqrt{\frac{2*189.656{,}62}{4}} = 1.172{,}61$$

Verifica-se então que, por esse teste, não haveria qualquer diferença significativa entre as seis médias.

Aplicando-se a um contraste mais complexo, do tipo $T1 + T2 - 2*T3$, por exemplo, o teste de Scheffé teria o seguinte valor:

$$S = \sqrt{(6\text{-}1)*2{,}90}\sqrt{\frac{189.656{,}62}{4}*(1+1+4)} = 2.031{,}02$$

O valor do contraste é igual a: $2.991{,}5 + 2.499{,}5 - 2*2.258{,}3 = 974{,}4$; logo, não significativo.

Teste de Scott-Knott

Este teste, idealizado por Scott e Knott (1974), de forma diferente dos demais discutidos até aqui, tem sua base teórica na análise de conglomerados, e sua significância é formulada sobre o teste de χ^2. A grande vantagem em sua utilização advém do fato de que nenhuma média pode pertencer a mais de um agrupamento como ocorre nos anteriores, ou seja, o teste determina a constituição de grupos de médias disjuntos, sempre que haja sido encontrada significância na aplicação do teste F na análise de variância. Não apresenta uma fórmula básica de obtenção de um ou mais valores para comparação das médias, mas sim estabelece os grupos de acordo com a variabilidade entre esses grupos de médias.

A sua aplicação, portanto, deve ser feita segundo as regras a seguir:

- Regra 1 – ordene as k médias em ordem decrescente.

- Regra 2 – crie k - 1 partições de grupos de médias, da seguinte forma:

  Partição 1: grupo do tratamento 1 e grupo formado pelos tratamentos 2, 3, ..., k.

  Partição 2: grupo dos tratamentos 1 e 2 e grupo dos tratamentos 3, 4, ..., k, e assim sucessivamente até a próxima partição.

  Partição k - 1: grupo formado por tratamentos 1, 2, ..., k - 1 e grupo do tratamento k.

- Regra 3 – calcule a soma de quadrado entre grupos de cada partição, usando a fórmula

$$SQgrupos = \frac{(G1)^2}{k1} + \frac{(G2)^2}{k2} - \frac{(GT)^2}{k} \quad [21]$$

em que $G1$, $G2$ e GT são, respectivamente, os totais das médias componentes do grupo 1, grupo 2 e soma total; e $k1$ e $k2$ são o número de tratamentos em cada grupo na partição. O máximo valor de $SQgrupos$ é denominado de β^2_0.

- Regra 4 – calcule

$$\sigma^2_0 = \frac{SQmédias + \upsilon(\frac{QME}{r})}{k + \upsilon} \quad [22]$$

em que r é o número de repetições de cada tratamento e υ é o número de graus de liberdade do resíduo.

- Regra 5 – a partição de soma de quadrado máxima será válida se a estatística λ for significativa, sendo essa estatística dada por

$$\lambda = \frac{\pi}{2(\pi\text{-}2)} * \frac{\beta^2_0}{\sigma^2_0} \quad [23]$$

com λ tendo uma distribuição de χ^2 com w graus de liberdade, para $w = k/(\pi - 2)$.

- Regra 6 – se λ for significativo, aplicam-se as regras 2 até 5 dentro de cada grupo da partição, e assim sucessivamente, até que não seja mais possível dividir as médias em grupos. Se λ for não significativo não haverá formação de grupos de médias.

Para os dados sendo utilizados até aqui com seis tratamentos, é possível a criação de 5 partições iniciais, a saber:

Partição 1: grupo 1 (tratamento 1) e grupo 2 (tratamentos 2, 3, 4, 5 e 6).

Partição 2: grupo 1 (tratamentos 1 e 2) e grupo 2 (tratamentos 3, 4, 5 e 6).

Partição 3: grupo 1 (tratamentos 1, 2 e 3) e grupo 2 (tratamentos 4, 5 e 6).

Partição 4: grupo 1 (tratamentos 1, 2, 3 e 4) e grupo 2 (tratamentos 5 e 6).

Partição 5: grupo 1 (tratamentos 1, 2, 3, 4 e 5) e grupo 2 (tratamento 6).

Lembrando que as médias dos tratamentos 1 a 6 foram, respectivamente, iguais a 2.991,5; 2.499,5; 2.258,3; 2.070,5; 1.995,5 e 1.937,3, tem-se para a partição 1 a seguinte soma de quadrado:

$$SQPartição1 = \frac{(2.991,5)^2}{1} + \frac{(10.761,1)^2}{5} - \frac{(13.752,6)^2}{6} = 586.992,432$$

De forma semelhante, calculam-se as outras somas de quadrados em cada uma das partições e tem-se, respectivamente, 586.992,432; 616.714,68; 508.086,0; 318.241,47 e 151.059,648.

Como a partição 2 possui a maior soma de quadrado entre grupos, a soma de quadrados associada a essa partição será o primeiro valor de β^2_0. Sendo o valor de σ^2_0 igual a 71.785,29, tem-se que a estatística λ é igual a 11,82, como calculado abaixo, com aproximadamente 5 graus de liberdade, o que a torna significativa a 5% de probabilidade, com a consequente formação inicial de dois grupos de médias, um formado pelos tratamentos 1 e 2 e o outro pelos tratamentos 3, 4, 5 e 6.

$$SQMédias = (2.991,5^2 + 2.499,5^2 + 2.258,3^2 + 2.070,5^2 + 1.995,5^2 + 1.937,3^2) - \frac{(13.752,6)^2}{6}$$

$$SQMédias = 796.278,72$$

$$w = \frac{6}{\pi\text{-}2} = 5,26 \approx 5$$

$$\sigma_0^2 = \frac{796.278,72 + 15(\frac{189.656,62}{4})}{6+15} = 71.785,29$$

$$\lambda = \frac{\pi}{2(\pi\text{-}2)} * \frac{616.714,68}{71.785,29} = 11,82$$

A cada um desses grupos aplicam-se, a seguir, as regras 2 a 5, o que resulta na inexistência de qualquer outra partição significativa.

Esses resultados podem ser resumidos assim:

*t*1 - 2.991,5 a

*t*2 - 2.499,5 a

*t*3 - 2.258,3 b

*t*4 - 2.070,5 b

*t*5 - 1.995,5 b

*t*6 - 1.937,3 b

Teste de Dunnett

Muitas vezes o interesse do pesquisador não está centrado em descobrir diferenças entre os tratamentos, mas em descobrir se existem tratamentos que são diferentes, melhores ou piores, que um deles em particular, que é considerado o padrão ou a testemunha. Esse conjunto de comparações não forma um conjunto independente, o que levou Dunnett (1955) a apresentar uma alternativa, e esse procedimento requer um único valor para julgar a significância ou não das diferenças existentes. O procedimento tanto pode ser aplicado de forma unilateral (melhor ou pior que o padrão) ou bilateral (diferente do padrão) e os Anexos 8 e 9 são empregados.

Voltando ao mesmo exemplo que está sendo utilizado para os demais testes, e assumindo o tratamento 3 como padrão de comparação, poder-se-ia então aplicar o teste de Dunnett bilateral, ou seja, verificar se existem diferenças entre ele e os demais tratamentos.

O teste apresenta a expressão

$$Du = t_{p,\alpha,GLE} * s_d \qquad [24]$$

em que $t_{p,\alpha,GLE}$ representa o valor tabelado do teste (Anexos 5 e 6, unilateral, ou Anexos 7 e 8, bilateral) para *p* tratamentos (excluído o padrão), nível de probabilidades α e número de graus de liberdade do erro (GLE). S_d é a raiz quadrada da variância da diferença entre duas médias. No caso de um experimento em que todos os tratamentos têm o mesmo número de repetições, s_d é dada por

$$s_d = \sqrt{\frac{2 * QME}{r}} \qquad [25]$$

e tem-se, então,

$$Du = 2{,}90 * 307{,}94 = 893{,}03.$$

Como as diferenças entre os tratamentos 1, 2, 4, 5 e 6 com relação ao padrão (3) são: 733,2, 241,2, 187,8, 262,8 e 321, respectivamente, e todas menores que o valor calculado do teste (893,03), pode-se afirmar que nenhum tratamento difere do padrão.

Em alguns tipos de ensaios, como aqueles de controle químico de insetos, é recomendável usar o tratamento testemunha (sem aplicação de defensivos) com um número maior de repetições do que os demais tratamentos, pois ocorre com certa frequência uma maior incidência de insetos nesse tratamento, ocasionada mais pela fuga de insetos das parcelas tratadas do que propriamente pela ausência de tratamento, o que mascara o efeito real dos

tratamentos. Nesse caso, a fórmula de cálculo do teste de Dunnett deve ser modificada, e tem-se então que o valor do erro padrão da diferença entre a média da testemunha com r_0 repetições e qualquer outro tratamento com r repetições é igual a:

$$s_d = \sqrt{QME\left(\frac{1}{r} + \frac{1}{r_0}\right)} \qquad [26]$$

Número de repetições

A determinação do tamanho do experimento está associada ao número de repetições que devem ser empregadas, já que o número de tratamentos é definido pelo pesquisador na busca das informações necessárias para comprovar suas hipóteses. Lembrando que o valor do erro padrão de uma média é inversamente proporcional ao número de observações (repetições) envolvidas na sua obtenção, percebe-se a grande importância que se deve dar à escolha desse número.

Diversos procedimentos têm sido apresentados para se definir o número de repetições necessárias, e todos os métodos se baseiam no valor das diferenças mínimas que se deseja detectar e, portanto, são todos eles de aplicação iterativa (COCHRAN; COX, 1983; HARRIS et al., 1948; HARTER, 1957; TUKEY, 1949a). Infelizmente a maior parte das vezes o número de repetições é definido pelo pesquisador de acordo com a disponibilidade financeira, tempo requerido para as avaliações, área ou mão da obra disponível.

Um processo mais prático associado ao teste de Tukey requer o conhecimento do CV mais provável de ser observado no experimento, associado a um valor de diferença mínima significativa entre tratamentos expressa em termos de porcentagem da média do experimento. Em geral, os pesquisadores têm conhecimento do CV que se deve esperar em um experimento e podem definir o valor máximo aceitável para uma diferença mínima entre tratamentos. Associando-se as fórmulas de cálculo do CV ($CV = s/m$) e do teste de Tukey ($DHS = q*s*r^{1/2}$) e fazendo $DHS = p*m$, em que s é a raiz quadrada do quadrado médio do erro, m a média do experimento, q o valor de tabela para o teste de Tukey, r o número de repetições e $0 < p \leq 1$, verifica-se que

$$r = \frac{(CV)^2 q^2}{p^2} \qquad [27]$$

A Tabela 2 mostra alguns resultados da aplicação desse procedimento.

Tabela 2. Valores máximos da diferença mínima, em porcentagem da média do experimento, para dois números de tratamentos e números de repetição.

Número de tratamentos	Número de repetições	Coeficiente de variação		
		0,10	0,20	0,25
12	4	0,25	0,50	0,62
12	6	0,20	0,39	0,49
12	8	0,17	0,34	0,42
12	10	0,15	0,30	0,37
25	2	0,41	0,82	1,03
25	4	0,27	0,54	0,67
25	6	0,22	0,43	0,54
25	8	0,20	0,40	0,50

Uma análise dessa tabela mostra a grande dificuldade em se detectar diferenças pequenas entre tratamentos com baixo número de repetições e a enorme influência do coeficiente de variação nesses valores. Por exemplo, em um experimento com 12 tratamentos e CV esperado de 10%, se a diferença mínima desejada pelo pesquisador for de 20% do valor da média do ensaio, será necessária a utilização de seis repetições. Uma grande parte dos pesquisadores gosta de usar quatro repetições, quase um número mágico para eles, e essa tabela mostra que o mínimo que se pode detectar nesse caso como diferença mínima entre dois tratamentos corresponde a 25% da média do ensaio. Considere-se o caso do arroz irrigado, em que é comum a obtenção de médias de produção de grãos superiores a 5.000 kg ha^{-1}. Nesse caso, e aceitando esse número como base, será impossível determinar que tratamentos diferindo por menos de 1.250 kg ha^{-1} sejam diferentes.

Transformação de dados

As pressuposições básicas de erros independentes, normalmente distribuídos, com variância comum e aditividade no modelo matemático são também necessárias para a aplicação dos testes para comparação de médias. A independência dos erros é, normalmente, assegurada pelo processo de distribuição aleatória dos tratamentos às parcelas experimentais, segundo a metodologia prevista para cada delineamento, enquanto a aditividade pode ser associada à escala de medidas. A heterocesdaticidade dos erros (erros diferentes) é talvez o problema mais grave a ser considerado, e ela pode estar associada aos tratamentos ou às variáveis utilizadas na sua avaliação. Zimmermann (1987) mostrou o efeito das diferentes distribuições de probabilidade e da presença de variâncias diferentes

entre os tratamentos sobre o nível real de significância e sobre o poder do teste F, e esses resultados sem dúvida podem ser admitidos para os testes de comparações de médias.

Segundo Steel e Torrie (1981), a heterogeneidade dos erros pode ser de dois tipos: irregular e regular. A do tipo irregular está associada ao fato de alguns tratamentos possuírem maior variabilidade que outros, sem qualquer aparente associação entre média e variância, e essa diferença pode ou não ser esperada, já que muitas vezes elas fazem parte da resposta do tratamento. Um caso de típico de diferença em variância esperada é o estudo dos efeitos de inseticidas no controle de insetos. A testemunha, sem inseticida, apresenta maior presença de insetos e de grande variabilidade quando comparada com os tratamentos com aplicação de inseticidas nos quais a presença de insetos é bem reduzida. Casos semelhantes podem ser encontrados em experimentos com fungicidas, herbicidas, fertilizantes ou ainda com técnicas de irrigação. Nesse caso, os autores sugerem duas alternativas de solução, uma de eliminar da análise os tratamentos com variância muito discrepante e outra, mais elegante, a de dividir o erro em componentes aplicáveis aos distintos tipos de comparações desejadas, uma vez que as transformações de dados, usualmente, não resolvem esse problema.

Heterogeneidade de variância, do tipo regular, surge quando ocorre algum tipo de falta de ajuste à normalidade, com a variabilidade dos tratamentos associada à sua média de alguma forma. De uma maneira geral, as diversas transformações de dados disponíveis estão associadas às distribuições de probabilidade e procuram tornar média e variância independentes. Se não for possível encontrar uma transformação que torne as médias e variâncias individuais independentes e estáveis, outro procedimento deve ser buscado, tal como uma análise ponderada, ou deve-se adotar um procedimento não paramétrico.

Ao proceder a uma análise com dados transformados, deve-se tomar o cuidado, no momento de se reportarem os resultados, de informar sempre tal fato e, como muito comumente as médias dos valores transformados não farão sentido ao leitor, elas devem ser retransformadas à escala original. No entanto, as variâncias, intervalos de confiança ou valores de diferença mínima não podem passar por esse processo, pois seus valores não terão significado nem coerência com os resultados apresentados.

As transformações mais comuns são discutidas a seguir.

Raiz quadrada [$\sqrt{(X + k)}$] – dados que seguem uma distribuição de Poisson, isto é, que possuam média e variância iguais, são mais bem analisados após a aplicação da transformação por raiz quadrada. Alguns exemplos de dados que seguem essa distribuição são as contagens de números de plantas daninhas ou de insetos de uma dada espécie por unidade de área, por exemplo, onde as contagens são sempre em números inteiros e baixos. Dados em porcentagem, baseados em contagem e com denominador comum, situados entre 0% e 20% ou entre 80% e 100%, também podem ser transformados por raiz quadrada.

A constante k adicionada ao dado pode usualmente assumir três valores, 0, 0,5 e 1. Quando os valores são muito pequenos, e principalmente se zeros estão presentes, a simples raiz quadrada ($k = 0$) tende a sobrecorrigir e, portanto, $k = 0{,}5$ ou $k = 1$ devem ser empregados.

Logaritmo [log(X + k)] – três são as situações em que esta transformação é adequada: variâncias proporcionais ao quadrado da média; efeitos multiplicativos na escala original, que portanto se tornam aditivos com a transformação, e medidas em valores inteiros que cobrem uma amplitude muito grande. Teoricamente qualquer base logarítmica pode ser adotada, e até o surgimento dos computadores a base 10 era a mais usada por conveniência ou tabelas mais facilmente encontráveis. Nesse tipo de transformação, a constante k pode assumir, no geral, dois valores, 0 e 1. O valor 1 é empregado obrigatoriamente quando da ocorrência de zeros ou optativamente com a ocorrência de valores menores que 10.

Arco-seno (arcsen√X) – especialmente recomendada para dados expressos em porcentagem e principalmente para casos em que os dados estejam nas faixas entre 0% e 20% ou entre 80% e 100%. Os valores transformados tanto podem ser expressos em graus como em radianos. Se os valores em porcentagem estiverem todos entre 20% e 80%, pouco ou nada se ganhará ao fazer a transformação. A variância dos novos valores é aproximadamente constante, sendo igual a $821/n$ quando expressos em graus e igual a $0{,}25/n$ se em radianos (STEEL; TORRIE, 1981). Isso mostra que se pressupõe um denominador constante, mas essa transformação é comumente empregada quando os denominadores são diferentes.

Teste de aditividade

É possível isolar-se da soma de quadrados do erro um componente com 1 grau de liberdade referente à aditividade do modelo. Tukey (1949) introduziu o método para ensaios de dupla classificação ou blocos completos casualizados, e o mesmo autor (TUKEY, 1955) o fez para o delineamento em quadrado latino. Esses procedimentos serão discutidos nos capítulos correspondentes a esses delineamentos.

Capítulo 2

Delineamento inteiramente casualizado

Este delineamento apresenta como características principais a necessidade de homogeneidade de ambiente para todas as unidades experimentais, bem como a possibilidade de diferentes números de repetições para cada tratamento em teste. Assumindo um número igual de repetições para todos os tratamentos, é o delineamento que apresenta o maior número de graus de liberdade para o erro, e essa é a sua principal vantagem comparativa em relação aos demais delineamentos. Por não apresentar controle local, é considerado o mais simples de todos os delineamentos experimentais, mas sua aparente simplicidade é, na realidade, o fator que obriga o pesquisador a ter um enorme conhecimento do ambiente experimental no qual conduzirá seu trabalho. Essa ausência de controle local exige que o ambiente seja o mais homogêneo possível, pois toda e qualquer diversidade presente no ensaio, que não puder ser atribuída diretamente ao efeito de tratamento, será considerada como efeito aleatório e, portanto, incluída no erro ou resíduo experimental.

Outra fonte de cuidado diz respeito ao uso de número de repetições diferentes para os tratamentos. Essa aparente vantagem, de maior liberdade no planejamento, na realidade faz que as variâncias de cada tratamento possam ser diferentes e, mesmo que isso não ocorra, tem como consequência principal a diferença das variâncias das comparações entre dois tratamentos segundo os números de repetições que apresentam cada uma das médias envolvidas nas diversas comparações.

Com base nisso, experimentos usando esse delineamento só poderiam ser realizados quando as unidades experimentais estivessem alocadas em um ambiente único no que diz respeito à homogeneidade dos fatores ambientais. De forma mais particular, no caso da agricultura, essas condições homogêneas são mais facilmente encontráveis em laboratórios, casas de vegetação ou telados, câmaras de crescimento e em canteiros; e, na pecuária, em maternidades, aviários ou criatórios com controle ambiental.

A aplicação de cada tratamento às parcelas experimentais é feita, como afirmado antes, de forma totalmente aleatória. De uma maneira geral, identifica-se cada parcela por um número, sequencialmente, bem como os tratamentos. Após a casualização, os tratamentos são aplicados nas parcelas segundo sua numeração. Cada sorteio realizado tem probabilidade igual, e esse valor é dado pela fórmula (assumindo todos os tratamentos com igual número de repetições): $r!*k/n!$, em que k é o numero de tratamentos, r o de repetições e n é o produto $n*k$. Por exemplo, se $k = 4$ e $r = 4$, a probabilidade será 4!*4/16! Ou seja, 24*4/ 20.922.789.888.000 ou $4{,}5883*10^{-12}$; e esse experimento poderia na realidade ter o desenho descrito na Tabela 1, em que os números 101 a 116 identificam as parcelas; e os números entre 1 e 4, os tratamentos.

Tabela 1. Esquema para um experimento completamente ao acaso com quatro tratamentos e quatro repetições.

101	102	103	104	105	106	107	108
4	1	1	3	4	4	3	2
116	115	114	113	112	111	110	109
1	3	3	2	2	4	2	1

Modelo matemático e análise de variância

As pressuposições envolvidas neste delineamento são de que os erros são independentes, com distribuição normal, com média zero e variância homogênea (σ^2); e o modelo matemático associado é linear e aditivo e tem a forma

$$Y_{ij} = m + t_i + e_{ij}\text{, com } i = 1, 2, \ldots, I \text{ e } j = 1, 2, \ldots, J_i \qquad [1]$$

em que Y_{ij} representa a observação feita na unidade experimental de repetição j do tratamento i; m a média geral; t_i o efeito do i-ésimo tratamento; e_{ij} o erro associado à observação Y_{ij} e J_i o número de repetições do tratamento i. É recomendável que cada J_i seja pelo menos igual a 2, pois, se ele for igual a 1, o tratamento não contribuirá de forma alguma na estimação do erro experimental e apresentará um valor mais alto de erro padrão do que os demais tratamentos. Nesse caso, a análise de variância será calculada como dada na Tabela 2.

Se todos os tratamentos tiverem o mesmo número de repetições, a Tabela 2 se modifica. Então, têm-se as fórmulas dadas na Tabela 3.

Tabela 2. Análise de variância para o delineamento completamente ao acaso. Tratamentos com número diferente de repetições.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Tratamentos (T)	$I - 1$	$\sum_i \frac{T_i^2}{J_i} - \frac{1}{\sum_i J_i}\left(\sum_{ij} Y_{ij}\right)^2$	SQT/GLT
Erro ou resíduo (E)	$\sum_i J_i - I$	$SQTo - SQT$	SQE/GLE
Total (To)	$\sum_i J_i - I$	$\sum_{ij} Y_{ij}^2 - \frac{1}{\sum_i J_i}\left(\sum_{ij} Y_{ij}\right)^2$	

Tabela 3. Análise de variância para o delineamento completamente ao acaso. Tratamentos com o mesmo número de repetições.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Tratamentos (T)	$I - 1$	$\frac{1}{J}\sum_i T_i^2 - \frac{1}{IJ}\left(\sum_{ij} Y_{ij}\right)^2$	SQT/GLT
Erro ou resíduo (E)	$IJ - 1$	$SQTo - SQT$	SQE/GLE
Total (To)	$IJ - 1$	$\sum_{ij} Y_{ij}^2 - \frac{1}{IJ}\left(\sum_{ij} Y_{ij}\right)^2$	

No modelo matemático, a única alteração refere-se ao fato de que todos os J_i serão constantes e iguais a J.

É fácil ver que a perda de observações não afeta, portanto, a análise de variância, mas deveria haver a preocupação de se ter pelo menos duas unidades experimentais para cada tratamento. Tratamentos com uma só repetição deveriam ser eliminados, e a análise efetuada como se o experimento tivesse um tratamento a menos. No caso de serem mantidos, deve-se ter em mente os problemas anteriormente discutidos de que eles não contribuirão em nada para a estimação do erro além das modificações nos testes de comparação de médias.

Para quaisquer das duas situações, as hipóteses lançadas sobre os efeitos de tratamentos serão testadas sempre em relação ao erro ou resíduo. Ao assumir-se o modelo matemático como fixo, os efeitos de tratamentos são considerados fixos, isto é, os t_i's serão fixos, portanto $\Sigma t_i = 0$; e ao presumir-se o modelo como aleatório, assume-se que os t_i's são uma amostra da população de t's, com média 0 e variância σ_t^2.

Os valores esperados dos quadrados médios estão mostrados na Tabela 4. Depreende-se claramente que, em qualquer situação, o teste F a ser aplicado corresponde ao quociente entre o quadrado médio de tratamentos e o quadrado médio do erro ou resíduo.

Tabela 4. Valores esperados dos quadrados médios de tratamentos (fixos ou aleatórios) e do erro para números iguais e diferentes de repetições. No quadro: $G^2 = \Sigma J_i t_i^2 - (\Sigma J_i t_i)^2$ e $J_0 = (\Sigma J_i - \Sigma J_i^2/\Sigma J_i)/(I - 1)$.

	Valor esperado dos quadrados médios			
	Número igual de repetições		Número diferente de repetições	
	Modelo fixo	Modelo aleatório	Modelo fixo	Modelo aleatório
Tratamento	$\sigma^2 + J\Sigma T_i^2/(I - 1)$	$\sigma^2 + J\sigma_t^2$	$\sigma^2 + G^2\Sigma J_i/(I - 1)$	$\sigma^2 + J_0\sigma_t^2$
Erro	σ^2	σ^2	σ^2	σ^2

Exemplos de aplicação

Todos os tratamentos com o mesmo número de repetições

Em 2001, o pesquisador Alberto Baeta dos Santos, em um estudo sobre adubação nitrogenada na cultura de arroz irrigado, testou quatro formas de aplicação desse adubo: 1 = 80 kg ha^{-1} no plantio; 2 = 40 kg ha^{-1} no plantio e 40 kg ha^{-1} 40 dias após a emergência (DAE); 3 = 13,2 kg ha^{-1} no plantio e 66,8 kg ha^{-1} aos 40 DAE e 4 = 13,2 kg ha^{-1} no plantio e 33,4 kg ha^{-1} aos 40 e 60 DAE. O experimento tinha oito repetições, e os dados de produção de grãos em kg ha^{-1} encontram-se na Tabela 5.

Os tratamentos apresentaram os seguintes totais de produção: 1 = 49.000; 2 = 53.212; 3 = 49.976 e 4 = 54.958, sendo o total geral igual a 207.146.

O uso das fórmulas antes mostradas leva a:

$Fator\ de\ correção\ (FC) = 207.146^2/32 = 1.340.920.791{,}125$

$SQTotal = (6.276^2 + 6.035^2 + ... + 7.270^2) - FC = 6.573.562{,}875$

$SQTratamentos = (49.000^2 + 53.212^2 + 49.976^2 + 54.958^2)/8 - FC = 2.891.619{,}375$

$SQResíduo = SQTotal - SQTratamentos = 3.681.943{,}500$

Tabela 5. Dados de produção de grãos de arroz irrigado, em kg ha^{-1}, no delineamento inteiramente ao acaso, com quatro tratamentos e oito repetições.

Tratamento	Repetição							
	1	2	3	4	5	6	7	8
1	6.276	6.035	6.086	5.594	6.321	6.746	5.751	6.191
2	7.199	6.890	6.586	7.149	6.657	6.210	6.128	6.393
3	6.457	6.174	6.612	6.087	5.797	5.865	6.498	6.486
4	7.202	7.173	7.169	6.590	6.444	6.740	6.370	7.270

A análise de variância pode então ser construída (Tabela 6).

Tabela 6. Análise de variância para o ensaio completamente ao acaso com quatro tratamentos e oito repetições. Modelo considerado como fixo.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	Teste F
Tratamentos	3	2.891.619,375	963.873,125	7,33
Erro ou resíduo	28	3.681.943,500	131.497,982	
Total	31	6.573.562,875		

Considerando-se que a média geral do experimento foi 6.473,3 kg ha^{-1}, o coeficiente de variação tem um valor de 5,60%, o que demonstra a alta precisão com que foi conduzido o ensaio. O teste F para tratamentos de 7,33 é significativo com uma probabilidade de ocorrência menor que 1% (na realidade 0,09%), indicando, portanto, uma grande diferença nas respostas do arroz irrigado às distintas formas de aplicação do adubo nitrogenado. Para identificar essas diferenças em resposta, deve-se aplicar um teste de comparação de médias. Se um dos tratamentos fosse identificado como a forma usual de aplicação (testemunha), poder-se-ia aplicar o teste de Dunnett, mas, se esse não for o caso, então outros testes podem ser aplicados; neste exemplo, vamos empregar o teste de Tukey. O valor desse teste como foi visto no Capítulo 1, é dado por:

$$DHS = q_{\alpha,n,GLE} * \sqrt{\frac{QME}{r}}$$

Portanto, tem-se nesse caso que

DHS = 3,86*362,6265 / 2,8284 = 494,87,

e quaisquer dois tratamentos cujas médias difiram por mais do que esse valor devem ser considerados diferentes. Resumindo, tem-se:

4 (13,2 kg ha^{-1} no plantio e 33,4 kg ha^{-1} aos 40 e 60 DAE)	6.869,8	a
2 (40 kg ha^{-1} no plantio e 40 kg ha^{-1} aos 40 DAE)	6.651,5	a b
3 (13,2 kg ha^{-1} no plantio e 66,8 kg ha^{-1} aos 40 DAE)	6.247,0	b c
1 (80 kg ha^{-1} no plantio)	6.125,0	c

Uso do SAS: tratamentos com o mesmo número de repetições

Após a entrada dos dados no SAS, a análise de variância se faz com o uso do procedimento ANOVA, da seguinte forma:

```
PROC ANOVA;
CLASS TRAT;
MODEL PROD=TRAT;
MEANS TRAT/TUKEY;
```

O resultado obtido se vê na Figura 1. Caso houvesse interesse em outro teste de comparação de médias, bastaria substituir a palavra TUKEY por uma equivalente ao teste desejado e disponível no SAS, como, por exemplo: DMS, DUNCAN, SNK, etc. Se o interesse fosse a modificação do nível de significância do teste, uma vez que o padrão do SAS é de 5%, dever-se-ia especificar após a palavra TUKEY a opção ALPHA=sign, em que sign seria o valor da probabilidade desejada. Por exemplo, para 1% de probabilidade, o comando passa a ter a seguinte forma:

```
MEANS TRAT/TUKEY ALPHA=0.01;
```

Tratamentos com números distintos de repetições

Como se viu anteriormente, é perfeitamente possível a análise de ensaios neste delineamento mesmo quando o número de repetições não é o mesmo para todos os tratamentos, e essa diferença tanto pode ser devida à perda de dados como ao planejamento. No exemplo anterior, admite-se que os dados referentes às parcelas do tratamento 4, na repetição 8, e do tratamento 3, na repetição 4, tivessem sido perdidos de forma que a tabela de dados teria agora a forma dada Tabela 7.

Os totais de tratamentos e o total geral são então: 1 = 49.000; 2 = 53.212; 3 = 43.889 e 4 = 47.688, sendo o total geral igual a 193.789.

```
                          The ANOVA Procedure
                       Class Level Information
                    Class         Levels    Values
                    TRAT               4    1 2 3 4
                      Number of observations    32
                          The ANOVA Procedure
Dependent Variable: PROD
                              Sum of
 Source                DF       Squares    Mean Square   F Value   Pr > F
 Model                  3    2891619.375     963873.125      7.33   0.0009
 Error                 28    3681943.500     131497.982
 Corrected Total       31    6573562.875

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.439886      5.601869      362.6265      6473.313

Source              DF       Anova SS    Mean Square   F Value   Pr > F
 TRAT                3    2891619.375     963873.125      7.33   0.0009
                          The ANOVA Procedure
                Tukey's Studentized Range (HSD) Test for PROD

NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.

                  Alpha                                      0.05
                  Error Degrees of Freedom                     28
                  Error Mean Square                        131498
                  Critical Value of Studentized Range     3.86125
                  Minimum Significant Difference           495.04

          Means with the same letter are not significantly different.

      Tukey Grouping          Mean      N    TRAT

                   A        6869.8      8    4
                   A
              B    A        6651.5      8    2
              B
              B    C        6247.0      8    3
                   C
                   C        6125.0      8    1
```

Figura 1. Resultados da análise de variância e comparação de médias de um experimento no delineamento inteiramente casualizado com o uso do SAS.

Tabela 7. Dados de produção de grãos em arroz irrigado, em kg ha^{-1}, no delineamento completamente ao acaso, com quatro tratamentos e perda de duas parcelas experimentais.

Tratamento	Repetição							
	1	2	3	4	5	6	7	8
1	6.276	6.035	6.086	5.594	6.321	6.746	5.751	6.191
2	7.199	6.890	6.586	7.149	6.657	6.210	6.128	6.393
3	6.457	6.174	6.612	5.797	5.865	6.498	6.486	
4	7.202	7.173	7.169	6.590	6.444	6.740	6.370	

O uso das fórmulas antes mostradas leva a:

Fator de correção (*FC*) = $193.789^2/30 = 1.251.805.884{,}033$

SQTotal = $(6.276^2 + 6.035^2 + ... + 6.3702)$ - *FC* = 5.784.000,966

SQTratamentos = $(49.000^2/8 + 53.212^2/8 + 49.976^2/7 + 54.958^2/7)$ - *FC* = 2.314.400,395

SQResíduo = *SQTotal* - *SQTratamentos* = 3.469.600,571

A análise de variância pode então ser construída (Tabela 8).

Tabela 8. Análise de variância para o delineamento completamente ao acaso, com quatro tratamentos e duas parcelas perdidas. Modelo considerado como fixo.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	Teste F
Tratamentos	3	2.314.400,395	771.466,798	5,78
Erro ou resíduo	26	3.469.600,571	133.446,176	
Total	29	5.784.000,966		

Considerando-se que agora a média geral do experimento é de 6.459,6 kg ha^{-1}, o coeficiente de variação tem um valor de 5,66%, não muito diferente do valor anterior. O teste F, apesar de ter seu valor reduzido para 5,78, continua significativo com uma probabilidade de ocorrência menor que 1% (na realidade 0,36%), com a mesma interpretação anterior. Para identificar essas diferenças ainda usando o teste de Tukey, agora apenas de forma aproximada, é necessário calcular três valores para esse teste, um para comparar os tratamentos com oito repetições, outro para os de sete repetições e um terceiro para comparar médias com sete e oito repetições. Para este último caso, o número de repetições ponderado é dado por 2*(7*8)/(7 + 8) = 7,47.

Os valores serão então:

a) Tratamentos com oito repetições:

$$DHS = 3{,}88 * \sqrt{\frac{133.446{,}176}{8}} = 501{,}12$$

b) Tratamentos com sete repetições:

$$DHS = 3{,}88 * \sqrt{\frac{133.446{,}176}{7}} = 535{,}72$$

c) Tratamentos com sete e oito repetições:

$$DHS = 3{,}88 * \sqrt{\frac{133.446{,}176}{7{,}47}} = 518{,}59$$

Quaisquer dois tratamentos cujas médias difiram por mais do que esses valores, respeitando os respectivos números de repetições, devem ser consideradas diferentes.

A comparação entre o tratamento 4 (com sete repetições) e o tratamento 2 (com oito repetições) mostra que suas médias diferem em 161,1 kg ha^{-1} (6.812,6 - 6.651,5), valor menor que a DHS correspondente (518,59), indicando não haver diferença significativa entre eles. Já a diferença entre os tratamentos 4 e 3, ambos com sete repetições, que é igual a 542,7, supera o valor do teste (DHS = 535,72), o que indica serem esses tratamentos diferentes em suas respostas.

Em resumo, tem-se:

4 (13,2 kg ha^{-1} no plantio e 33,4 kg ha^{-1} aos 40 e 60 DAE)	6.812,6	a
2 (40 kg ha^{-1} no plantio e 40 kg ha^{-1} aos 40 DAE)	6.651,5	a b
3 (13,2 kg ha^{-1} no plantio e 66,8 kg ha^{-1} aos 40 DAE)	6.269,9	b c
1 (80 kg ha^{-1} no plantio)	6.125,0	c

Vê-se, então, que não houve qualquer modificação, nesse caso, nas comparações de médias.

Uso do SAS: tratamentos com números distintos de repetições

Para este delineamento, considerado o mais simples de todos, em que a perda de parcelas não implica ajuste das médias, o mesmo procedimento SAS pode ser usado, com apenas uma modificação, que é a necessidade de incluir um segundo comando MEANS para obter a impressão das médias dos tratamentos. A comparação das médias pelo teste de TUKEY é feita da mesma forma, mas a procedimento não mostra as médias, apenas as diferenças entre elas, os limites inferiores e superiores do intervalo de confiança para essas diferenças e a colocação de asteriscos após os intervalos para mostrar quais são significativamente diferentes de zero (0) no nível de probabilidade especificado. O programa para a análise, portanto, tem a forma:

```
PROC ANOVA;
CLASS TRAT;
MODEL PROD=TRAT;
MEANS TRAT;
MEANS TRAT/TUKEY;
```

A Figura 2 mostra os resultados obtidos para os dados do exemplo que se está estudando.

```
                          The ANOVA Procedure
                       Class Level Information
                    Class         Levels    Values
                    TRAT               4    1 2 3 4
                      Number of observations    30
                          The ANOVA Procedure
Dependent Variable: PROD
                              Sum of
 Source             DF        Squares    Mean Square   F Value   Pr > F
 Model               3    2314400.395     771466.798      5.78   0.0036
 Error              26    3469600.571     133446.176
 Corrected Total    29    5784000.967

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.400138      5.655164      365.3029      6459.633

 Source             DF       Anova SS       Mean Square   F Value  Pr > F
 TRAT                3     2314400.395     771466.798      5.78   0.0036
                          The ANOVA Procedure
                  Level of          -------------PROD------------
                  TRAT         N             Mean          Std Dev
                  1            8       6125.00000       355.099986
                  2            8       6651.50000       404.309288
                  3            7       6269.85714       328.517594
                  4            7       6812.57143       364.033816
                          The ANOVA Procedure
              Tukey's Studentized Range (HSD) Test for PROD
       NOTE: This test controls the Type I experimentwise error rate.
                  Alpha                                     0.05
                  Error Degrees of Freedom                    26
                  Error Mean Square                       133446.2
                  Critical Value of Studentized Range     3.87964
       Comparisons significant at the 0.05 level are indicated by ***.
                          Difference      Simultaneous
                  TRAT       Between     95% Confidence
                 Comparison   Means         Limits
                 4    - 2      161.1     -357.6   679.7
                 4    - 3      542.7        7.0  1078.4  ***
                 4    - 1      687.6      168.9  1206.2  ***
                 2    - 4     -161.1     -679.7   357.6
                 2    - 3      381.6     -137.0   900.3
                 2    - 1      526.5       25.4  1027.6  ***
                 3    - 4     -542.7    -1078.4    -7.0  ***
                 3    - 2     -381.6     -900.3   137.0
                 3    - 1      144.9     -373.8   663.5
                 1    - 4     -687.6    -1206.2  -168.9  ***
                 1    - 2     -526.5    -1027.6   -25.4  ***
                 1    - 3     -144.9     -663.5   373.8
```

Figura 2. Resultados da análise de variância e comparação de médias de um experimento no delineamento inteiramente casualizado, com o uso do SAS, no caso de tratamentos com números distintos de repetição.

Seria possível também ter utilizado para este caso o procedimento General Linear Models (GLM) e todos os demais comandos usados com o ANOVA, como foi visto anteriormente. Outra opção seria a de substituir os dois comandos MEANS pelo comando LSMEANS, mas, nesse caso, o teste de Tukey já não é utilizado pelo GLM e deveria ser substituído pelo

teste de t ou DMS, usando-se a opção PDIFF. Lembre-se que este teste já não é tão preciso quanto o Tukey.

```
PROC GLM;
CLASS TRAT;
MODEL PROD=TRAT;
LSMEANS TRAT/PDIFF;
```

A Figura 3 mostra o resultado dessa nova programação.

```
                          The GLM Procedure
                       Class Level Information
                    Class         Levels    Values
                    TRAT               4    1 2 3 4
                      Number of observations    30
                          The GLM Procedure
Dependent Variable: PROD
                              Sum of
Source                 DF       Squares    Mean Square   F Value   Pr > F
 Model                  3    2314400.395     771466.798      5.78   0.0036
 Error                 26    3469600.571     133446.176
Corrected Total        29    5784000.967

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.400138      5.655164      365.3029      6459.633

Source              DF      Type I SS    Mean Square   F Value   Pr > F
 TRAT                3    2314400.395     771466.798      5.78   0.0036

Source              DF    Type III SS    Mean Square   F Value   Pr > F
 TRAT                3    2314400.395     771466.798      5.78   0.0036
                          The GLM Procedure
                        Least Squares Means
                                                    LSMEAN
                    TRAT      PROD LSMEAN           Number

                    1          6125.00000                1
                    2          6651.50000                2
                    3          6269.85714                3
                    4          6812.57143                4

                  Least Squares Means for effect TRAT
                  Pr > |t| for H0: LSMean(i)=LSMean(j)

                      Dependent Variable: PROD
          i/j              1             2             3             4
             1                        0.0078        0.4505        0.0012
             2        0.0078                        0.0540        0.4020
             3        0.4505        0.0540                      0.0100
             4        0.0012        0.4020        0.0100
```

Figura 3. Resultados da análise de variância e comparação de médias de um experimento no delineamento inteiramente casualizado, com o uso do SAS, no caso de tratamentos com números distintos de repetição e o procedimento GLM.

Mais de uma observação por unidade experimental

É muito comum na pesquisa a obtenção de mais de um valor para algumas variáveis em cada unidade experimental. Isso ocorre, por exemplo, na tomada de altura de plantas, na leitura da incidência de doenças ou de pragas, na análise química de algum componente do produto sendo trabalhado, como teor de açúcar, ou de N, P ou outro elemento essencial, e muitas outras situações.

De uma maneira geral, verifica-se que muitos cientistas utilizam a média dessas observações para proceder à análise de variância, desperdiçando dessa maneira um potencial enorme de informação contida nas medidas individuais, além de perder a oportunidade de ter uma comparação de médias mais eficiente, que lhes permita, no geral, detectar menores diferenças significativas. O uso das informações individuais permite que a análise de variância seja modificada com a estimação de dois tipos de erros: o experimental e o amostral. O erro amostral dá uma medida da variabilidade interna existente dentro de cada tratamento, e o erro experimental dá variabilidade entre as unidades experimentais. A maior eficiência na comparação de médias está associada ao fato de que cada tratamento terá muito mais repetições; portanto, será obtida melhor estimativa do erro padrão das médias.

É interessante notar que o número de observações em cada uma das unidades experimentais deve ser sempre que possível o mesmo, pois a diversidade nesse número causa a impossibilidade de obtenção de um teste F simples e adequado como veremos mais adiante.

As pressuposições envolvidas são as mesmas de antes, só que agora se têm dois erros, ambos com distribuição normal com média zero e variância homogênea (σ^2 e δ^2, respectivamente, para o erro amostral e o erro experimental), e o modelo matemático associado passa a ser

$$Y_{ijk} = m + t_i + ee_{ij} + ea_{ijk} \qquad [2]$$

com $i = 1, 2, ..., I$; $j = 1, 2, ..., J_i$ e $k = 1, 2, \ldots, K$, em que Y_{ijk} representa a k-ésima observação feita na unidade experimental de repetição j do tratamento i, m a média geral, t_i o efeito do i-ésimo tratamento, ee_{ij} o erro experimental associado às observações Y_{ij}, ea_{ijk} o erro associado às k observações e J_i o número de repetições do tratamento i. O quadro da análise de variância é modificado de igual forma, representando as mudanças do modelo, como

se verifica nas Tabelas 9 e 10, nas quais se assume um número igual de observações por parcela.

Se todos os tratamentos tiverem o mesmo número de repetições, a Tabela 9 se modifica e tem-se então a Tabela 10.

Tabela 9. Análise de variância do delineamento ao acaso com k amostras por unidade experimental. Número diferente de repetições por tratamento.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Tratamentos (T)	$I - 1$	$\sum_i \frac{T_i^2}{kJ_i} - \frac{1}{k\sum_i J_i}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQT/GLT
Erro experimental (EE)	$\sum_i J_i - I$	$\frac{1}{k}\sum_{ij} Y_{ij}^2 - \frac{1}{k\sum_i J_i}\left(\sum_{ijk} Y_{ijk}\right)^2 - SQT$	$SQEE/GLEE$
Erro amostral (EA)	Diferença	Diferença	$SQEA/GLEA$
Total (To)	$K\sum_i J_i - I$	$\sum_{ijk} Y_{ijk}^2 - \frac{1}{k\sum_i J_i}\left(\sum_{ijk} Y_{ijk}\right)^2$	

Tabela 10. Análise de variância do delineamento ao acaso com k amostras por unidade experimental. Número igual de repetições para os tratamentos.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Tratamentos (Tr)	$I - 1$	$\frac{1}{JK}\sum_i T_i^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQT/GLT
Erro experimental (EE)	$IJ - 1$	$\frac{1}{K}\sum_{ij} Y_{ij.}^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2 - SQTr$	$SQEE/GLEE$
Erro amostral (EA)	Diferença	Diferença	$SQEA/GLEA$
Total (To)	$IJK - 1$	$\sum_{ijk} Y_{ijk}^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2$	

Se o número de observações por unidade experimental não for único, as fórmulas da Tabela 10 devem ser modificadas, lembrando que, nesse caso, tem-se $k_{ij} = 1,2,...,K_{ij}$.

Os valores esperados dos quadrados médios, assumindo um modelo aleatório, são mostrados na Tabela 11.

Tabela 11. Valores esperados dos quadrados médios em uma análise de variância do delineamento completamente ao acaso com k amostras por unidade experimental.

	Valor esperado dos quadrados médios			
	Número igual de repetições		Número diferente de repetições por tratamento	
	Número igual de observações	Número diferente de observações	Número igual de observações	Número diferente de observações
Tratamentos	$\sigma^2 + K\delta^2 + JK\sigma_t^2$	$\sigma^2 + K_2\delta^2 + K_3\sigma_t^2$	$\sigma^2 + K\delta^2 + K_3\sigma_t^2$	$\sigma^2 + K_2\delta^2 + K_3\sigma_t^2$
Erro experimental	$\sigma^2 + K\delta^2$	$\sigma^2 + K_1\delta^2$	$\sigma^2 + K\delta^2$	$\sigma^2 + K_1\delta^2$
Erro amostral	σ^2	σ^2	σ^2	σ^2

Neste caso, J é o número de repetições e K o número de amostras ou observações feitas em cada unidade experimental.

No caso de um modelo de efeito de tratamentos fixos, a expressão do componente de variância de tratamentos é substituída por uma forma quadrática que expressa o efeito desses tratamentos, como foi visto no caso de uma única observação por parcela.

Como a componente variância experimental está presente tanto no valor esperado do erro experimental quanto no de tratamentos, deve-se primeiro fazer o teste F para o erro experimental (quociente entre os dois erros). Se esse teste for significativo, deve-se proceder ao teste F para tratamentos como a razão entre o QM de tratamentos e o QM do erro experimental, quando se tem um mesmo número de observações por unidade experimental.

No caso de números diversos de observações, como se pode ver pelas expressões dadas na Tabela 11, não existe uma forma simples para o teste F, pois os coeficientes da variância experimental são distintos ($K1$ e $K2$). Uma aproximação possível, nesse caso, é o ajuste do valor do erro experimental por meio da seguinte expressão:

$$(K_2/K_1)*QM(\text{Erro experimental}) + (1 - K_2/K_1)*QM(\text{Erro amostral}).$$

Caso o teste F para o erro experimental não seja significativo, isto é, pode-se assumir que o componente δ^2 não difere de zero, sugere-se que a análise seja refeita, fazendo a agregação dos dois erros em um único, cujos graus de liberdade e soma de quadrados serão dados pela soma dos graus de liberdade dos dois erros existentes e das respectivas somas de quadrados. Portanto, o teste F para tratamentos ocorrerá pelo quociente entre o QM de tratamentos e o QM do novo erro.

Os três coeficientes (K_1, K_2 e K_3) referentes às esperanças com número diferente de amostras, dados no quadro anterior, apresentam as seguintes fórmulas de cálculo [3], em que k_{ij} corresponde ao número de observações no tratamento i e repetição j:

$$K_1 = \frac{k_{..} - \sum_i \left(\frac{\sum_j k_{ij}^2}{k_{..}} \right)}{GLEE}, \quad K_2 = \frac{\sum_i \left(\frac{\sum_j k_{ij}^2}{k_{..}} \right) - \frac{\sum_{ij} k_{ij}^2}{k_{..}}}{GLT} \quad \text{e} \quad K_3 = \frac{k_{..} - \frac{\sum_i k_{i.}^2}{k_{..}}}{GLT} \qquad [3]$$

Exemplo de aplicação: tratamentos com igual número de repetições e de amostras por parcela

Os dados na Tabela 12, em gramas de matéria seca por planta, constituem parte da informação obtida em uma pesquisa realizada pelo pesquisador Pedro Marques da Silveira em Santo Antônio de Goiás, em 2000. O experimento foi conduzido em casa de vegetação. Os cinco tratamentos foram aplicados em cinco vasos (repetição), que continham cada um três plantas de arroz (amostras). Os tratamentos eram:

1) Solo de fazenda Rodoarte sem cultivo prévio de arroz, com uso de Furadan.
2) Solo de fazenda Rodoarte sem cultivo prévio de arroz, sem uso de Furadan.
3) Solo de fazenda Rodoarte sem cultivo prévio de arroz, com uso de herbicida.
4) Solo de fazenda Rodoarte sem cultivo prévio de arroz, sem uso de herbicida.
5) Solo de fazenda da Embrapa com cultivo prévio de arroz, com uso de herbicida.

As somas de quadrados são obtidas utilizando as formas mostradas anteriormente de maneira que:

$SQTo = (2{,}17^2 + 0{,}99^2 + ... + 2{,}01^2 + 2{,}69^2) - (133{,}39)^2/75 = 94{,}3273$

$SQTrat = (41{,}13^2 + 29{,}92^2 + ... + 21{,}05^2)/15 - (133{,}39)^2/75 = 30{,}7415$

$SQEe = (8{,}49^2 + 5{,}24^2 + ... + 4{,}40^2 + 5{,}84^2)/3 - (133{,}39)^2/75 - SQTrat = 28{,}0585$

$SQEa = SQTo - SQTrat - SQEe = 35{,}5273$

Tabela 12. Dados de matéria seca por planta, em gramas, em um experimento no delineamento completamente ao acaso, com três amostras por parcela.

Tratamento	Planta	Vaso 1	Vaso 2	Vaso 3	Vaso 4	Vaso 5	Soma
1	1	2,17	0,99	2,32	2,51	3,36	
	2	1,63	2,21	1,49	3,76	3,63	
	3	4,69	2,04	3,63	4,33	2,37	
	Soma	8,49	5,24	7,44	10,60	9,36	41,13
2	1	3,26	2,19	3,22	1,34	3,23	
	2	1,27	2,44	2,95	2,10	1,15	
	3	1,21	1,13	1,92	1,37	1,14	
	Soma	5,74	5,76	8,09	4,81	5,52	29,92
3	1	0,99	1,37	0,64	0,87	0,34	
	2	0,32	1,27	0,58	0,47	0,86	
	3	0,02	1,10	0,98	0,55	2,00	
	Soma	1,33	3,74	2,20	1,89	3,20	12,36
4	1	0,83	0,36	3,57	2,18	0,55	
	2	0,93	2,38	1,62	3,51	0,43	
	3	0,58	3,97	4,50	2,91	0,61	
	Soma	2,34	6,71	9,69	8,60	1,59	28,93
5	1	1,57	1,30	1,88	1,39	1,35	
	2	1,07	1,11	1,13	1,00	1,86	
	3	0,48	1,35	0,92	2,01	2,63	
	Soma	3,12	3,76	3,93	4,40	5,84	21,05

A análise de variância encontra-se na Tabela 13.

O teste de F para o erro experimental (1,97), significativo a 2,66%, confirma a existência deste componente (δ^2); portanto, o teste adequado para a hipótese de efeito de tratamentos deve ser realizado por meio do erro experimental. Nesse caso, o teste mostra haver efeito significativo de tratamentos. Esse mesmo erro (experimental) deve ser empregado em qualquer prova de médias que se deseje empregar para verificar onde ocorrem as diferenças entre os tratamentos. Se utilizássemos o teste de Tukey, por exemplo, o valor de *DHS* seria:

$$DHS = 4{,}23 * \sqrt{\frac{1{,}4029}{15}} = 1{,}29$$

No entanto, se o componente erro experimental não houvesse sido significativo, a análise poderia ser refeita, considerando-se a existência de apenas um erro, que seria chamado de erro combinado, ou simplesmente erro, e a análise de variância teria a forma dada na Tabela 14.

Tabela 13. Análise de variância para os dados da Tabela 12.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Probabilidade F)
Tratamentos	4	30,7415	7,6854	5,48(0,0038)
Erro experimental	20	28,0585	1,4029	1,97(0,0266)
Erro amostral	50	35,5273	0,7105	
Total (To)	74	94,3273		

Tabela 14. Análise de variância dos dados da Tabela 12, ao assumir que o erro experimental não foi significativo.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Probabilidade F)
Tratamentos	4	30,7415	7,6854	8,46 (< ,0001)
Erro	70	63,5859	0,9084	
Total (To)	74	94,3273		

A aplicação do teste de Tukey teria agora valor

$$DHS = 3{,}96 * \sqrt{\frac{0{,}9084}{15}} = 0{,}97$$

e a diferença mínima para significância entre duas médias seria, portanto, 25% menor, pois o erro passa a ter muitos mais graus de liberdade e um valor consideravelmente reduzido, o que indica maior precisão das estimativas.

Uso do SAS: tratamentos com igual número de repetições e amostras por parcela

Usando os dados da Tabela 12, e chamando tratamento de TRAT e matéria seca por planta de PESO, a análise de variância será realizada com a seguinte programação:

```
PROC ANOVA;CLASS VASO PLANTA TRAT;
MODEL PESO=TRAT TRAT(VASO);
TEST H=TRAT E=TRAT(VASO);
MEANS TRAT/TUKEY E=TRAT(VASO);
```

No modelo especificado, a expressão TRAT(VASO) permite a estimação do erro experimental, enquanto o erro estimado pelo procedimento ANOVA corresponde ao erro amostral. Como já se sabia, pelo exemplo citado, que o erro experimental era significativo, foi então especificado pelo comando TEST que o efeito de tratamentos deveria ser testado pelo erro experimental, assim como no comando MEANS foi feita essa mesma definição com o uso da opção E=.

Na realidade, dever-se-ia proceder a essa análise em duas corridas do SAS, a fim de verificar, na primeira execução, se o erro experimental era ou não significativo, para, dessa forma, na segunda execução, ter-se a análise completa e correta, não havendo então necessidade de especificar o termo TRAT(VASO) no modelo matemático. A Figura 4 mostra o resultado desse procedimento. Cabe ao leitor o exercício de, a partir dos dados da figura, montar o quadro resumo da análise de variância.

```
                          The ANOVA Procedure
                       Class Level Information
                  Class          Levels    Values
                  VASO                5    1 2 3 4 5
                  PLANTA              3    1 2 3
                  TRAT                5    1 2 3 4 5

                      Number of observations    75
                         The ANOVA Procedure
Dependent Variable: PESO
                                Sum of
 Source                 DF      Squares    Mean Square   F Value   Pr > F
 Model                  24   58.80000533     2.45000022      3.45   0.0001
 Error                  50   35.52733333     0.71054667
 Corrected Total        74   94.32733867

              R-Square     Coeff Var      Root MSE     PESO Mean
              0.623361      47.39519      0.842939      1.778533

 Source                 DF       Anova SS    Mean Square   F Value   Pr > F
 TRAT                    4    30.74145867     7.68536467     10.82   <.0001
 TRAT(VASO)             20    28.05854667     1.40292733      1.97   0.0266

    Tests of Hypotheses Using the Anova MS for TRAT(VASO) as an Error Term
 Source                 DF       Anova SS    Mean Square   F Value   Pr > F
 TRAT                    4    30.74145867     7.68536467      5.48   0.0038
                         The ANOVA Procedure
                Tukey's Studentized Range (HSD) Test for PESO
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.

                 Alpha                                      0.05
                 Error Degrees of Freedom                     20
                 Error Mean Square                      1.402927
                 Critical Value of Studentized Range    4.23186
                 Minimum Significant Difference          1.2942

          Means with the same letter are not significantly different.
              Tukey Grouping          Mean      N    TRAT
                           A        2.7420     15    1
                           A
                      B    A        1.9947     15    2
                      B    A
                      B    A        1.9287     15    4
                      B
                      B             1.4033     15    5
                      B
                      B             0.8240     15    3
```

Figura 4. Análise de variância para um delineamento inteiramente ao acaso com várias observações por parcela com o uso do SAS.

Exemplo de aplicação: tratamentos com mesmo número de repetições e número diferente de amostras por parcela

Para o mesmo experimento anterior, suponha-se que sete plantas, ao acaso, morreram e, portanto, não se tem delas a massa de matéria seca. A tabela de dados tem então o seguinte aspecto que se encontra na Tabela 15.

O quadro de análise de variância, obtido com o uso das fórmulas apresentadas anteriormente, é visto na Tabela 16.

Como o erro experimental foi significativo, o teste F para tratamentos deve ser feito utilizando-o como denominador. No entanto, em virtude dos distintos números de amostras por parcela, as esperanças dos quadrados médios respectivos são dadas por

$E(QMTrat) = \sigma^2 + K_2\delta^2 + K_3\sigma_t^2 = \sigma^2 + 2{,}6039\delta^2 + 13{,}019\sigma_t^2$ e

$E(QMEe) = \sigma^2 + K_1\delta^2 = \sigma^2 + 2{,}6998\delta^2$,

em que os três coeficientes foram calculados usando as fórmulas dadas anteriormente.

A relação K_2/K_1 é igual a $2{,}6039/2{,}6998 = 0{,}9645$, de forma que o QM do erro experimental ajustado será dado por: $0{,}9645*1{,}3564 + (1 - 0{,}9645)*0{,}6613 = 1{,}3317$. Com o devido ajuste ao número de graus de liberdade, utilizando-se a aproximação de Satterthwaite (1946) dada por

$$n' = \frac{\left[\left(\frac{K_2}{K_1}\right) * QMEe + \left(1 - \frac{K_2}{K_1}\right) * QMEa\right]^2}{\frac{\left(\frac{K_2}{K_1}\right)^2 * (QMEe)^2}{GLEe} + \frac{\left(1 - \frac{K_2}{K_1}\right)^2 * (QMEa)^2}{GLEa}} \quad [4]$$

o novo erro possui 20,722, ou aproximadamente 20 GL, obtendo-se um valor de F igual a 3,98, cuja probabilidade é 0,02.

Uso do SAS: tratamentos com mesmo número de repetições e número diferente de amostras por parcela

Para esta análise de número diferente de amostras por repetição e tratamentos, o procedimento GLM é o adequado, com a seguinte programação:

Tabela 15. Dados de matéria seca por planta, em gramas, em um ensaio no delineamento completamente ao acaso, com três amostras por parcela, mas com sete dados perdidos por morte de plantas.

Tratamento	Planta	Vaso 1	Vaso 2	Vaso 3	Vaso 4	Vaso 5	Soma
1	1	2,17	0,99	2,32	2,51	3,36	
	2	1,63	2,21	1,49	3,76	3,63	
	3	4,69	2,04	3,63	4,33	2,37	
	Soma	8,49	5,24	7,44	10,60	9,36	41,13
2	1	3,26	2,19	3,22	1,34	3,23	
	2	1,27	2,44	2,95	2,10	1,15	
	3	1,21	1,13	1,92	1,37	1,14	
	Soma	5,74	5,76	8,09	4,81	5,52	29,92
3	1	0,99	1,37	0,64	0,87		
	2		1,27	0,58		0,86	
	3		1,10	0,98	0,55	2,00	
	Soma	0,99	3,74	2,20	1,42	2,86	11,21
4	1	0,83		3,57	2,18	0,55	
	2	0,93	2,38	1,62	3,51		
	3	0,58	3,97	4,50	2,91	0,61	
	Soma	2,34	6,35	9,69	8,60	1,16	28,14
5	1	1,57	1,30	1,88	1,39	1,35	
	2	1,07	1,11	1,13	1,00	1,86	
	3		1,35	0,92	2,01	2,63	
	Soma	2,64	3,76	3,93	4,40	5,84	20,57

Tabela 16. Análise de variância para os dados da Tabela 15.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Probabilidade F)
Tratamentos	4	21,1780	5,2945	
Erro experimental	20	27,1279	1,3564	2,05 (0,0243)
Erro amostral	43	28,4354	0,6613	
Total (To)	67	78,3296		

```
PROC GLM;CLASS VASO PLANTA TRAT;
MODEL PESO=TRAT TRAT(VASO);RANDOM TRAT TRAT(VASO);
TEST H=TRAT E=TRAT(VASO);
```

O comando RANDOM calcula a esperança dos quadrados médios mostrando os valores dos coeficientes $K1$, $K2$ e $K3$. Para esse caso, o ajuste do quadrado médio do erro experimental e de seus graus de liberdade ajustados deve ser feito manualmente como mostrado no item *Exemplo de aplicação: tratamentos com mesmo número de repetições e número diferente de amostras por parcela*. A Figura 5 mostra o resultado desse procedimento.

```
                    The GLM Procedure
                Class Level Information
Class          Levels    Values
VASO                5    1 2 3 4 5
PLANTA              3    1 2 3
TRAT                5    1 2 3 4 5

   Number of observations    68
                    The GLM Procedure
Dependent Variable: PESO
                                Sum of
Source                    DF       Squares    Mean Square   F Value   Pr > F
Model                     24    49.89424461     2.07892686      3.14   0.0005
Error                     43    28.43538333     0.66128798
Corrected Total           67    78.32962794

              R-Square     Coeff Var      Root MSE     PESO Mean
              0.636978      42.22138      0.813196      1.926029

Source                    DF      Type I SS    Mean Square   F Value   Pr > F

TRAT                       4    22.76630776     5.69157694      8.61   <.0001
TRAT(VASO)                20    27.12793684     1.35639684      2.05   0.0243

Source                    DF    Type III SS    Mean Square   F Value   Pr > F

TRAT                       4    21.17797192     5.29449298      8.01   <.0001
TRAT(VASO)                20    27.12793684     1.35639684      2.05   0.0243

                    The GLM Procedure
Source                  Type III Expected Mean Square
TRAT                Var(Error) + 2.6039 Var(TRAT(VASO)) + 13.019 Var(TRAT)
TRAT(VASO)          Var(Error) + 2.6998 Var(TRAT(VASO))

                    The GLM Procedure
Dependent Variable: PESO
   Tests of Hypotheses Using the Type III MS for TRAT(VASO) as an Error Term
Source                    DF    Type III SS    Mean Square   F Value   Pr > F
TRAT                       4    21.17797192     5.29449298      3.90   0.0168
```

Figura 5. Resultado da análise de variância de um delineamento inteiramente ao acaso, com número diferente de amostras, com o uso do SAS.

Capítulo 3

Blocos completos ao acaso

Este delineamento talvez seja mais utilizado em todo o mundo. Em relação ao delineamento inteiramente casualizado, apresenta como modificação importante o agrupamento das unidades experimentais em subgrupos que são designados de blocos. Nesse caso, cada bloco contém cada um dos tratamentos, uma única vez. A homogeneidade ambiental é agora requerida para cada um desses blocos, que, por sua vez, podem ser tão distintos quanto possível, sem que isso implique problemas na análise de variância, ou de estimação das médias, ou dos componentes da análise.

A definição de bloco é crucial para a aplicação deste delineamento, pois o requerido é que cada bloco seja ambientalmente homogêneo, como visto anteriormente, e esse conceito é extremamente dependente do tipo de ensaio e da área de conhecimento envolvida. Na agricultura, por exemplo, um bloco pode ser uma área de terra, no caso de ensaios com fertilizantes, irrigação, cultivares e outros tópicos; pode ser um germinador na pesquisa sobre tecnologias de sementes; pode ser uma câmara de crescimento ou laboratório em pesquisas sobre fisiologia vegetal ou insetos e, até uma semana, por exemplo, se houver impossibilidade de se aplicarem todos os tratamentos de uma só vez, como em estudos de processos de inoculação de doenças ou em estudos de biologia de insetos. Na pecuária, um bloco pode ser uma leitegada; um lote de terneiras de mesma idade e filhas do mesmo pai; um tipo de pasto e muitas outras condições.

Muito se tem escrito e discutido sobre o número máximo de tratamentos a ser empregado neste delineamento, bem como sobre a forma dos blocos e a distribuição das unidades experimentais dentro deles (DE-LA-LOMA, 1966; FEDERER, 1955; KEMPTHORNE, 1975; GOMEZ; GOMEZ, 1976; PIMENTEL-GOMES, 2000). Já Steel e Torrie (1981) afirmam que não há restrição nem no número de tratamentos nem no de blocos. A experiência tem demonstrado, no entanto, que o número máximo de tratamentos é extremamente variável de acordo com os seguintes aspectos: tamanho de cada unidade experimental; definição física de bloco; amplitude do ambiente homogêneo em que será alocado cada bloco; capacidade de trabalho da equipe envolvida no trabalho ou disponibilidade de recursos financeiros ou materiais.

A forma do bloco e a distribuição das parcelas dentro dele podem ser importantes nos casos em que o bloco é constituído por uma área de solo, e geralmente se aceita que a forma quadrada é a mais adequada. Na realidade, é possível a um bloco apresentar forma retangular, de um U ou até mesmo de um quadrado com um espaço vazio em seu interior. Estas duas últimas formas são especialmente importantes em áreas de desmatamento recente, onde permanecem tocos ou restos vegetais, e o bloco é então montado ao redor desses materiais. Em experimentos com animais, onde um bloco pode ser uma leitegada ou um grupo de vacas em um mesmo período de lactação, essas preocupações evidentemente não existem.

A locação dos tratamentos às parcelas experimentais é feita de forma aleatorizada em cada um dos blocos, individualmente. Em um experimento com K tratamentos, a sua disposição dentro de cada bloco pode se dar de $K!$ maneiras, cada uma delas com igual probabilidade de $1/K!$. Isso quer dizer que, se houver $K = 10$ tratamentos, cada distribuição dos tratamentos em um bloco tem probabilidade de ocorrência igual a 1/10! ou $2{,}7557*10^{-7}$. Assumindo a existência de quatro blocos, um exemplo de tal casualização seria aquele descrito na Tabela 1, em que as centenas identificam as parcelas segundo o bloco, isto é, 101 a 110 correspondem ao bloco 1, e assim sucessivamente; e os números de 1 a 10, os tratamentos.

Tabela 1. Esquema de campo para um ensaio em blocos completos ao acaso, com dez tratamentos e quatro repetições (blocos).

410 4	409 10	408 8	407 6	406 7	405 1	404 3	403 9	402 2	401 5
301 10	302 1	303 6	304 9	305 2	306 3	307 5	308 8	309 7	310 4
210 8	209 10	208 9	207 3	206 6	205 2	204 1	203 5	202 4	201 7
101 10	102 4	103 2	104 3	105 6	106 7	107 1	108 9	109 8	110 5

Neste delineamento, a perda de parcelas pode ocasionar problemas mais ou menos sérios e cabe ao pesquisador tomar as decisões necessárias. No entanto, a perda de todas ou da maioria das parcelas relativas a um tratamento ou a um bloco não ocasiona danos à análise de variância, pois, nesses dois casos, a análise será executada como se o experimento tivesse um tratamento ou um bloco a menos do que o planejado.

No entanto, quando se perdem algumas parcelas aleatoriamente entre os tratamentos e os blocos, o que implica perda de graus de liberdade no erro igual ao número de parcelas perdidas, pode haver necessidade de estimação dos dados perdidos ou de execução da análise segundo um delineamento em blocos incompletos generalizados, como será visto no item *Mais de uma observação por parcela* e no Capítulo 5.

Modelo matemático e análise de variância

O modelo matemático para este delineamento apresenta a expressão

$$Y_{ij} = m + b_j + t_i + e_{ij} \quad [1]$$

em que Y_{ij}, m, t_i e e_{ij} têm o mesmo significado dado anteriormente e b_j é o efeito do j-ésimo bloco. A pressuposição acerca dos erros se mantém, isto é, são independentes, de distribuição normal e com variância homogênea. O modelo é evidentemente também linear e aditivo. Esta última condição implica que não há interação entre tratamentos e blocos, ou seja, que o comportamento dos tratamentos é similar, independentemente de quão diferentes sejam os blocos. Caso isso não ocorra, esse efeito, por não ser considerado no modelo, será incorporado à estimação do erro ou resíduo com possíveis graves consequências na interpretação dos resultados obtidos, já que o modelo perde sua condição de aditividade (ver o item Teste de aditividade).

Por causa do modelo matemático, o quadro de análise de variância tem a forma dada na Tabela 2.

Nesse caso, o modelo matemático pode ser aleatório, fixo ou misto, dependendo das suposições feitas sobre os efeitos de blocos e tratamentos, sendo o mais comum a aceitação dos efeitos de blocos como aleatórios. Em qualquer uma das situações, entretanto, o teste F para tratamentos e blocos é sempre obtido pelo quociente entre o quadrado médio da fonte de interesse e o QM do erro ou resíduo. Os valores esperados dos quadrados médios são dados na Tabela 3.

Tabela 2. Análise de variância, delineamento em blocos completos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Blocos (B)	$J - 1$	$\frac{1}{I}\sum_j B_j^2 - \frac{1}{IJ}\left(\sum_{ij} Y_{ij}\right)^2$	SQB/GLB
Tratamentos (T)	$I - 1$	$\frac{1}{J}\sum_i T_i^2 - \frac{1}{IJ}\left(\sum_{ij} Y_{ij}\right)^2$	SQT/GLT
Erro ou resíduo (E)	$(I - 1)(J - 1)$	$SQTo - SQB - SQT$	SQE/GLE
Total (To)	$IJ - 1$	$\sum_{ij} Y_{ij}^2 - \frac{1}{IJ}\left(\sum_{ij} Y_{ij}\right)^2$	

Tabela 3. Esperança matemática dos quadrados médios, em um delineamento em blocos completos ao acaso, para os modelos aleatório, fixo e misto.

Fonte de variação	Grau de liberdade	Valor esperado dos quadrados médios		
		Modelo aleatório	Modelo fixo	Modelo misto
Blocos (B)	J - 1	$\sigma^2 + I\sigma_b^2$	$\sigma^2 + I\Sigma B_j^2/(J - 1)$	$\sigma^2 + I\sigma_b^2$
Tratamentos (T)	I - 1	$\sigma^2 + J\sigma_t^2$	$\sigma^2 + J\Sigma B_j^2/(I - 1)$	$\sigma^2 + J\Sigma T_i^2/(I - 1)$
Erro (E)	$(I - 1)(J - 1)$	σ^2	σ^2	σ^2

Exemplos de aplicação

O grupo de pesquisadores, formado por Maria José del Peloso, Geraldo Estevam Carneiro e Dino M. Soares, conduziu em Santo Antônio de Goiás, o Ensaio Regional de Feijão-Preto, no ano de 1999, que contemplava 15 cultivares e/ou linhagens, em blocos ao acaso com quatro repetições. Entre as variáveis mensuradas, a produção de grãos em kg ha^{-1} encontra-se detalhada na Tabela 4.

O fator de correção é: $FC = (100.975,0)^2/60 = 169.932.510,42$; enquanto a soma de quadrados total é dada por: $SQTo = 1.112,5^2 + 650,0^2 + ... + 2.247,5^2 - FC = 12.030.614,58$.

Tabela 4. Dados de produção de grãos de feijão, em kg ha^{-1}, de um ensaio de competição de cultivares, em blocos completos ao acaso.

Tratamento	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Total
CNFP 8015	1.112,5	650,0	1.037,5	1.362,5	4.162,5
CNFP 8016	1.850,0	1.637,5	1.787,5	1.387,5	6.662,5
CNFP 8017	987,5	1.537,5	1.600,0	1.262,5	5.387,5
CNFP 8018	1.587,5	2.412,5	2.100,0	1.912,5	8.012,5
CNFP 8019	2.012,5	2.075,0	2.087,5	1.862,5	8.037,5
CNFP 8020	1.975,0	1.787,5	2.225,0	1.950,0	7.937,5
CNFP 8021	1.837,5	2.137,5	2.050,0	2.587,5	8.612,5
CNFP 8022	1.825,0	2.100,0	1.375,0	1.450,0	6.750,0
CNFP 8023	2.250,0	2.062,5	1.950,0	1.637,5	7.900,0
CNFP 8024	562,5	1.437,5	1.550,0	925,0	4.475,0
CNFP 8025	1.650,0	1.625,0	1.187,5	1.325,0	5.787,5
CNFP 8026	1.362,5	862,5	1.225,0	962,5	4.412,5
Diamante Negro	1.600,0	1.725,0	1.750,0	1.337,5	6.412,5
FT Nobre	2.162,5	1.825,0	1.950,0	1.550,0	7.487,5
G. Brilhante	2.112,5	2.137,5	2.450,0	2.237,5	8.937,5
Total	24.887,5	26.012,5	26.325,0	23.750,0	100.975,0

A soma de quadrados de tratamento é obtida da seguinte forma: $SQTr = (4.162,5^2 + 6.662,5^2 + ... + 8.937,5^2)/4 - FC = 8.754.052,08$; e a soma de quadrados de blocos, segundo as fórmulas apresentadas antes é: $SQB = (24.887,5^2 + 26.012,5^2 + ... + 23.750,0^2)/15 - FC = 274.552,08$; de forma que a soma de quadrados do erro, obtida por diferença é: $SQE = 12.030.614,58 - 8.754.052,08 - 274.552,08 = 3.002.010,42$.

Dessa forma, o quadro de análise de variância é apresentado na Tabela 5.

Este ensaio apresentou um coeficiente de variação de 15,88%, o que indica uma boa precisão experimental. Verifica-se também que não houve efeito de repetições (blocos), mas um efeito muito marcado de linhagens e cultivares (tratamentos). Dessa forma, cabe a aplicação de um teste de médias. A aplicação do teste de Tukey, a uma probabilidade de 5%, para 15 tratamentos e 42 GL para o erro, apresenta um valor de $DHS = 5,09*71.476,44/2 = 680,49$, o que leva à comparação de médias mostrada na Tabela 6.

Tabela 5. Análise de variância para os dados da Tabela 4. Modelo considerado como fixo.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Blocos (B)	3	274.552,083	91.517,361	1,28 (0,2935)
Tratamentos (Tr)	14	8.754.052,083	625.289,435	8,75 (< ,0001)
Erro (E)	42	3.002.010,420	71.476,440	
Total (To)	59	12.030.614,58		

Tabela 6. Quadro de comparação de médias pelo teste de Tukey, 5% de probabilidade, para os tratamentos do ensaio da Tabela 4.

G. Brilhante	2.234,4	A		
CNFP 8021	2.153,1	A		
CNFP 8019	2.009,4	A	B	
CNFP 8018	2.003,1	A	B	
CNFP 8020	1.984,4	A	B	
CNFP 8023	1.975,0	A	B	
FT Nobre	1.871,9	A	B	
CNFP 8022	1.687,5	A	B	C
CNFP 8016	1.665,6	A	B	C
Diamante Negro	1.603,1	A	B	C
CNFP 8025	1.446,9		B	C
CNFP 8017	1.346,9		B	C
CNFP 8024	1.118,8			C
CNFP 8026	1.103,1			C
CNFP 8015	1.040,6			C

Uso do SAS: blocos completos casualizados

Para o caso de um experimento em blocos completos casualizados com todos os seus dados, o procedimento ANOVA do SAS é o indicado. A programação é a seguinte:

```
PROC ANOVA;CLASS BLOCO TRATAMENTO;
MODEL PROD=BLOCO TRATAMENTO;
MEANS TRATAMENTO/TUKEY;
```

No comando MEANS, foi solicitado o teste de comparação de médias de Tukey a 5% de probabilidade, como feito no item anterior (*Exemplos de aplicação*). O resultado se vê na Figura 1.

```
                          The ANOVA Procedure
                       Class Level Information
Class         Levels  Values
BLOCO              4  1 2 3 4
TRATAMENTO        15  CNFP8015 CNFP8016 CNFP8017 CNFP8018 CNFP8019 CNFP8020
                      CNFP8021 CNFP8022 CNFP8023 CNFP8024 CNFP8025 CNFP8026
                      D.NEGRO FTNOBRE G.BRILHANTE

                      Number of observations    60
                          The ANOVA Procedure
Dependent Variable: PROD
                                  Sum of
 Source                   DF       Squares    Mean Square   F Value   Pr > F
 Model                    17    9028604.17     531094.36      7.43   <.0001
 Error                    42    3002010.42      71476.44
 Corrected Total          59   12030614.58

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.750469      15.88616      267.3508      1682.917

 Source                   DF      Anova SS    Mean Square   F Value   Pr > F
 BLOCO                     3     274552.083     91517.361      1.28   0.2935
 TRATAMENTO               14    8754052.083     625289.435      8.75   <.0001
                          The ANOVA Procedure
                Tukey's Studentized Range (HSD) Test for PROD
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.

                  Alpha                                      0.05
                  Error Degrees of Freedom                     42
                  Error Mean Square                       71476.44
                  Critical Value of Studentized Range     5.09058
                  Minimum Significant Difference           680.49

          Means with the same letter are not significantly different.

     Tukey Grouping          Mean      N    TRATAMENTO

              A            2234.4      4    G.BRILHANTE
              A            2153.1      4    CNFP8021
         B    A            2009.4      4    CNFP8019
         B    A            2003.1      4    CNFP8018
         B    A            1984.4      4    CNFP8020
         B    A            1975.0      4    CNFP8023
         B    A            1871.9      4    FTNOBRE
         B    A    C       1687.5      4    CNFP8022
         B    A    C       1665.6      4    CNFP8016
         B    A    C       1603.1      4    D.NEGRO
         B         C       1446.9      4    CNFP8025
         B         C       1346.9      4    CNFP8017
                   C       1118.8      4    CNFP8024
                   C       1103.1      4    CNFP8026
                   C       1040.6      4    CNFP8015
```

Figura 1. Análise de variância de experimento em blocos completos ao acaso com o uso do procedimento ANOVA do SAS

Parcelas perdidas – estimação e análise de variância

Como afirmado anteriormente, a ocorrência de uma ou umas poucas parcelas no delineamento em blocos completos, implica a perda equivalente em graus de liberdade no erro e na necessidade de estimação dos valores perdidos. A estimação do valor ou dos valores perdidos é feita por meio dos métodos de mínimos quadrados. O procedimento adotado é a montagem da soma dos quadrados dos erros, com a substituição do(s) dado(s) perdido(s) por uma letra (x, y, z, etc.). A partir dela, é possível obter a(s) derivada(s) parcial(is) com relação ao(s) dado(s) perdido(s), igualando-a(s) a zero e resolver a equação ou sistema de equações resultantes. O procedimento será exemplificado a seguir para o caso de duas parcelas perdidas.

No caso de uma parcela perdida, o procedimento resulta na fórmula de estimação de seu valor

$$X = \frac{IT + JB - G}{(I - 1)(J - 1)} \qquad [2]$$

em que T corresponde ao total disponível do tratamento com dado perdido, B ao total disponível do bloco com parcela perdida, G ao total disponível no experimento e I e J referem-se ao número de tratamentos e de blocos, respectivamente.

Para duas parcelas perdidas, três situações são possíveis: de um único tratamento, de um único bloco ou de tratamentos e blocos distintos. Suponha-se que se perderam os valores correspondentes ao tratamento 1 no bloco 1 (x) e ao tratamento 2 no bloco 3 (y). Nesse caso, a soma de quadrados para erro normalmente obtida por diferença ($SQE = SQTo - SQB - SQT$) terá a expressão

$$SQE = W^2 + x^2 + y^2 - FC - \left(\frac{B + (B_x + x)^2 + (B_y + y)^2}{I} - FC \right) - \left(\frac{T + (T_x + x)^2 + (T_y + y)^2}{J} - FC \right) \qquad [3]$$

em que W^2 e G correspondem à soma dos quadrados e à soma simples de todos os elementos disponíveis, B é a soma do quadrado dos totais de blocos sem dados perdidos, B_x e B_y são os totais disponíveis dos blocos com dados perdidos, T é a soma dos quadrados dos totais de tratamentos sem dados perdidos e T_x e T_y os totais disponíveis dos tratamentos com dados perdidos. O fator de correção (FC) é dado por:

$$FC = \frac{(G + x + y)^2}{IJ} \quad [4]$$

Os passos seguintes são: obter as derivadas parciais com relação a x e y, igualar as expressões a 0 e simplificá-las. O sistema de equações assim obtido é:

$$X = \frac{(I - 1)(J - 1)(JB_x + IT_x - G) - (JB_y + IT_y - G)}{(I - 1)^2 (J - 1)^2 - 1} \quad [5]$$

e

$$Y = \frac{(I - 1)(J - 1)(JB_y + IT_y - G) - (JB_x + IT_x - G)}{(I - 1)^2 (J - 1)^2 - 1} \quad [6]$$

Ao aplicar o mesmo procedimento para duas parcelas perdidas de um mesmo tratamento, tem-se

$$X = \frac{(J - 1)(JB_x + IT_{xy} - G) + (JB_y + IT_{xy} - G)}{(I - 1)\ [(J - 1)^2 - 1]} \quad [7]$$

e

$$Y = \frac{(J - 1)(JB_y + IT_{xy} - G) - (JB_x + IT_{xy} - G)}{(I - 1)\ [(J - 1)^2 - 1]} \quad [8]$$

em que T_{xy} representa o total disponível do tratamento com as parcelas perdidas. Para duas parcelas perdidas em um mesmo bloco, tem-se

$$X = \frac{(I - 1)(JB_{xy} + IT_x - G) + (JB_{xy} + IT_y - G)}{(J - 1)\ [(I - 1)^2 - 1]} \quad [9]$$

e

$$Y = \frac{(I - 1)(JB_{xy} + IT_y - G) + (JB_{xy} + IT_x - G)}{(J - 1)\ [(I - 1)^2 - 1]} \quad [10]$$

em que B_{xy} é o total disponível do bloco com as parcelas perdidas. Os demais símbolos têm o mesmo significado dado anteriormente.

Para mais de duas parcelas perdidas, esse processo de estimação pode ser empregado, porém recomendam-se ao pesquisador alguns cuidados especiais, tais como: nenhum tratamento deveria ter menos de duas observações disponíveis (dados de duas repetições), assim como os blocos não deveriam ter menos de dois tratamentos com informação disponível. Quando isso ocorrer, deve-se efetuar a análise eliminando o(s) tratamento(s) e/ou bloco(s); verificar detalhadamente a razão das perdas de informações e, caso as parcelas perdidas se distribuam de forma casual entre tratamentos e blocos, provavelmente a análise desses dados deveria ser executada segundo o delineamento em blocos incompletos como afirmado anteriormente.

Exemplo de aplicação

Ao utilizar o mesmo ensaio da Tabela 4 e eliminar o valor de produção do tratamento 1 na repetição 4, os valores são aqueles apresentados na Tabela 7.

Tabela 7. Dados de produção de grãos em um experimento de cultivares de feijão, aceitando a perda de uma informação (linhagem CNFP 8015, repetição 4).

Tratamento	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Total
CNFP 8015	1.112,5	650,0	1.037,5		2.800,0
CNFP 8016	1.850,0	1.637,5	1.787,5	1.387,5	6.662,5
CNFP 8017	987,5	1.537,5	1.600,0	1.262,5	5.387,5
CNFP 8018	1.587,5	2.412,5	2.100,0	1.912,5	8.012,5
CNFP 8019	2.012,5	2.075,0	2.087,5	1.862,5	8.037,5
CNFP 8020	1.975,0	1.787,5	2.225,0	1.950,0	7.937,5
CNFP 8021	1.837,5	2.137,5	2.050,0	2.587,5	8.612,5
CNFP 8022	1.825,0	2.100,0	1.375,0	1.450,0	6.750,0
CNFP 8023	2.250,0	2.062,5	1.950,0	1.637,5	7.900,0
CNFP 8024	562,5	1.437,5	1.550,0	925,0	4.475,0
CNFP 8025	1.650,0	1.625,0	1.187,5	1.325,0	5.787,5
CNFP 8026	1.362,5	862,5	1.225,0	962,5	4.412,5
Diamante Negro	1.600,0	1.725,0	1.750,0	1.337,5	6.412,5
FT Nobre	2.162,5	1.825,0	1.950,0	1.550,0	7.487,5
G. Brilhante	2.112,5	2.137,5	2.450,0	2.237,5	8.937,5
Total	24.887,5	26.012,5	26.325,0	22.387,5	99.612,5

Para estimar o valor perdido, deve-se empregar a fórmula 2. Tem-se, então,

$$X = \frac{IT + JB - G}{(I - 1)(J - 1)} = \frac{15 * 2.800,0 + 4 * 22.387,5 - 99.612,5}{(15 - 1) * (4 - 1)} = 760,41$$

Esse valor é utilizado então em lugar do valor perdido, de forma que agora o total do tratamento 1 passa a ser $2.800,0 + 760,41 = 3.560,41$, o que leva a uma estimativa de sua média igual a 890,1; o total do bloco 4 é $22.387,5 + 760,41 = 23.147,91$ e o total geral $100.372,91$ $(99.612,5 + 760,41)$. Esses novos valores são então empregados na análise de variância e tem-se que: o fator de correção é $FC = (100.372,91)^2/60 = 167.911.984,24$; a soma de quadrados total é dada por $SQTo = 1.112,5^2 + 650,0^2 + ... + 2.247,5^2 - FC = 12.772.942,67$; a soma de quadrados de tratamento é obtida pela forma $SQTr = (3.560,4^2 + 6.662,5^2 + ... + 8.937,5^2)/4 - FC = 9.612.088,74$; e a soma de quadrados de blocos, segundo as fórmulas apresentadas antes é $SQB = (24.887,5^2 + 26.012,5^2 + ... + 23.147,91^2)/15 - FC = 412.596,55$, de forma que a soma de quadrados do erro, obtida por diferença é $SQE = 12.772.942,67 - 9.612.088,74 - 412.596,55 = 2.748.357,38$.

Esse dado perdido, apesar de estimado, implica a perda de um grau de liberdade total e, consequentemente, do erro, de forma que se tem a análise de variância que se encontra na Tabela 8.

Tabela 8. Análise de variância para os dados da Tabela 7, quando o valor perdido de uma parcela foi estimado.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Blocos (B)	3	412.596,55	137.532,18	2,05 (0,1290)
Tratamentos (T)	14	9.612.088,74	686.577,77	10,24 (< 0,0001)
Erro (E)	41	2.748.257,38	67.030,67	
Total (To)	58	12.772.942,67		

Esse ensaio apresenta agora um coeficiente de variação de 15,29%, pouco diferente do obtido sem dados perdidos. Verifica-se também que continua não havendo efeito de repetições (blocos) e o mesmo efeito acentuado de linhagens e cultivares (tratamentos).

Uso do SAS: blocos completos casualizados com dado perdido

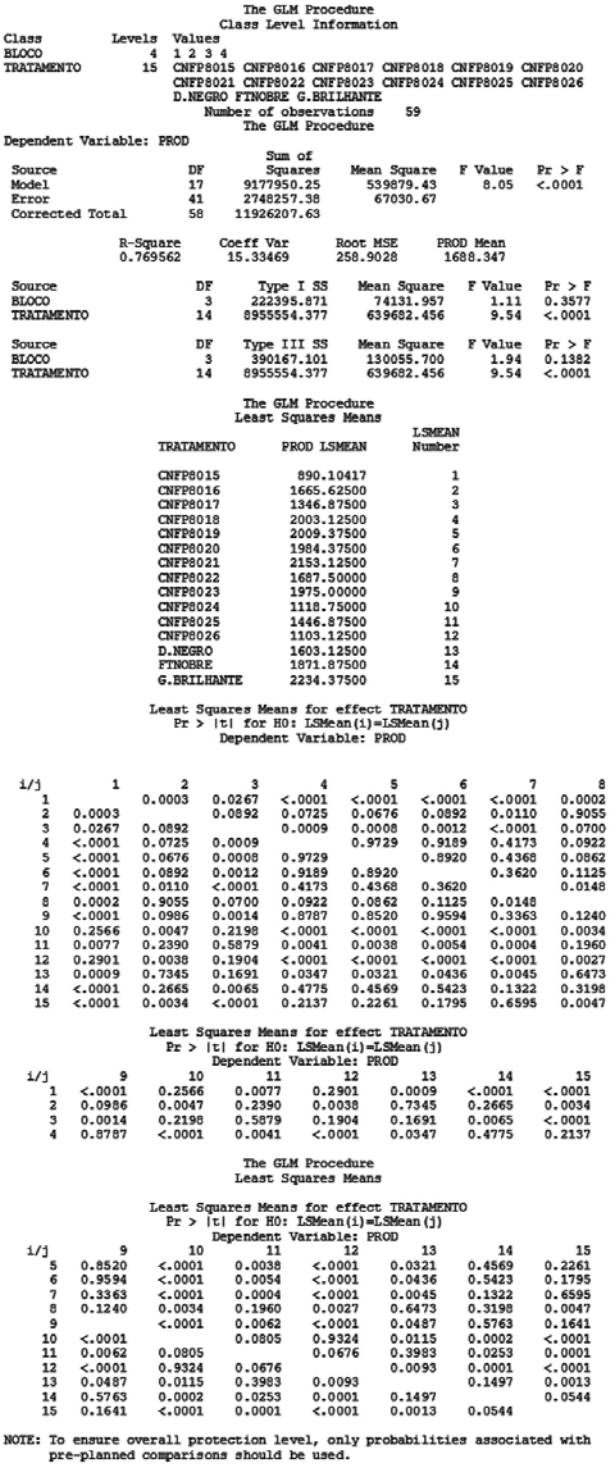
No caso de experimentos em blocos completos casualizados com parcelas perdidas, ao usar o SAS, não existe a necessidade de calcular os valores perdidos como mostrado na seção intitulada *Parcelas perdidas – estimação e análise de variância*, uma vez que o procedimento GLM do SAS faz os ajustes automaticamente tanto para a análise de variância quanto nas médias. O único inconveniente é que, nesse caso, o único teste de médias disponível é o de t, invocado com a opção PDIFF dentro do comando LSMEANS, que mostra a probabilidade, pelo teste t, de diferença entre duas médias para todos os pares de média. Poderia alternativamente ser solicitada a opção TDIFF, que permite, além da probabilidade do teste t, também imprimir o valor desse teste para todos os pares de médias. Os comandos para essa análise são:

```
PROC GLM;CLASS BLOCO TRATAMENTO;
MODEL PROD=BLOCO TRATAMENTO;
LSMEANS TRATAMENTO/PDIFF;
```

Os resultados são vistos na Figura 2, onde se pode observar que a média do tratamento com parcela perdida, CNFP 8015, é 890,1, como estimado na seção anterior. Entretanto, uma observação mais detalhada da análise de variância mostra valores bem distintos das somas de quadrados para blocos, tratamentos e total, mas com resultado idêntico para os graus de liberdade, para a soma de quadrado do erro e estimação das médias. Isso se dá em virtude da forma com que as duas análises de variância foram calculadas, uma usando a estimação do valor perdido e a outra com o uso de matrizes e com uma observação a menos (o valor perdido). Essas diferenças, no entanto, não devem causar preocupação ao investigador. Ao publicarem-se os resultados, devem ser usados os valores da ANOVA obtidos com o TYPE III SS, no qual os valores foram obtidos por ortogonalização das matrizes.

Mais de uma observação por parcela

Também neste desenho é possível a obtenção de mais de uma medida por parcela, a exemplo do que ocorre no desenho inteiramente casualizado. É extremamente comum, na pesquisa agrícola, que os investigadores tomem medidas de altura de plantas, de incidência de doenças, de pragas ou de ervas daninhas, repetidamente em cada parcela, tendo por base diversas plantas ou algumas pequenas áreas de amostragem. É também muito comum que eles, ao fazerem a análise de variância, não utilizem toda a informação coletada, e sim o valor médio ou, às vezes, o total das diversas amostras. Isso implica perda de informação e de precisão na análise de seus dados, como se verá mais adiante. O modelo



The GLM Procedure
Class Level Information

Class	Levels	Values
BLOCO	4	1 2 3 4
TRATAMENTO	15	CNFP8015 CNFP8016 CNFP8017 CNFP8018 CNFP8019 CNFP8020 CNFP8021 CNFP8022 CNFP8023 CNFP8024 CNFP8025 CNFP8026 D.NEGRO FTNOBRE G.BRILHANTE

Number of observations 59

The GLM Procedure

Dependent Variable: PROD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	9177950.25	539879.43	8.05	<.0001
Error	41	2748257.38	67030.67		
Corrected Total	58	11926207.63			

R-Square	Coeff Var	Root MSE	PROD Mean
0.769562	15.33469	258.9028	1688.347

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCO	3	222395.871	74131.957	1.11	0.3577
TRATAMENTO	14	8955554.377	639682.456	9.54	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCO	3	390167.101	130055.700	1.94	0.1382
TRATAMENTO	14	8955554.377	639682.456	9.54	<.0001

The GLM Procedure
Least Squares Means

TRATAMENTO	PROD LSMEAN	LSMEAN Number
CNFP8015	890.10417	1
CNFP8016	1665.62500	2
CNFP8017	1346.87500	3
CNFP8018	2003.12500	4
CNFP8019	2009.37500	5
CNFP8020	1984.37500	6
CNFP8021	2153.12500	7
CNFP8022	1687.50000	8
CNFP8023	1975.00000	9
CNFP8024	1118.75000	10
CNFP8025	1446.87500	11
CNFP8026	1103.12500	12
D.NEGRO	1603.12500	13
FTNOBRE	1871.87500	14
G.BRILHANTE	2234.37500	15

Least Squares Means for effect TRATAMENTO
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: PROD

i/j	1	2	3	4	5	6	7	8
1		0.0003	0.0267	<.0001	<.0001	<.0001	<.0001	0.0002
2	0.0003		0.0892	0.0725	0.0676	0.0892	0.0110	0.9055
3	0.0267	0.0892		0.0009	0.0008	0.0012	<.0001	0.0700
4	<.0001	0.0725	0.0009		0.9729	0.9189	0.4173	0.0922
5	<.0001	0.0676	0.0008	0.9729		0.8920	0.4368	0.0862
6	<.0001	0.0892	0.0012	0.9189	0.8920		0.3620	0.1125
7	<.0001	0.0110	<.0001	0.4173	0.4368	0.3620		0.0148
8	0.0002	0.9055	0.0700	0.0922	0.0862	0.1125	0.0148	
9	<.0001	0.0986	0.0014	0.8787	0.8520	0.9594	0.3363	0.1240
10	0.2566	0.0047	0.2198	<.0001	<.0001	<.0001	<.0001	0.0034
11	0.0077	0.2390	0.5879	0.0041	0.0038	0.0054	0.0004	0.1960
12	0.2901	0.0038	0.1904	<.0001	<.0001	<.0001	<.0001	0.0027
13	0.0009	0.7345	0.1691	0.0347	0.0321	0.0436	0.0045	0.6473
14	<.0001	0.2665	0.0065	0.4775	0.4569	0.5423	0.1322	0.3198
15	<.0001	0.0034	<.0001	0.2137	0.2261	0.1795	0.6595	0.0047

Least Squares Means for effect TRATAMENTO
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: PROD

i/j	9	10	11	12	13	14	15
1	<.0001	0.2566	0.0077	0.2901	0.0009	<.0001	<.0001
2	0.0986	0.0047	0.2390	0.0038	0.7345	0.2665	0.0034
3	0.0014	0.2198	0.5879	0.1904	0.1691	0.0065	<.0001
4	0.8787	<.0001	0.0041	<.0001	0.0347	0.4775	0.2137

The GLM Procedure
Least Squares Means

Least Squares Means for effect TRATAMENTO
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: PROD

i/j	9	10	11	12	13	14	15
5	0.8520	<.0001	0.0038	<.0001	0.0321	0.4569	0.2261
6	0.9594	<.0001	0.0054	<.0001	0.0436	0.5423	0.1795
7	0.3363	<.0001	0.0004	<.0001	0.0045	0.1322	0.6595
8	0.1240	0.0034	0.1960	0.0027	0.6473	0.3198	0.0047
9		<.0001	0.0062	<.0001	0.0487	0.5763	0.1641
10	<.0001		0.0805	0.9324	0.0115	0.0002	<.0001
11	0.0062	0.0805		0.0676	0.3983	0.0253	0.0001
12	<.0001	0.9324	0.0676		0.0093	0.0001	<.0001
13	0.0487	0.0115	0.3983	0.0093		0.1497	0.0013
14	0.5763	0.0002	0.0253	0.0001	0.1497		0.0544
15	0.1641	<.0001	0.0001	<.0001	0.0013	0.0544	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.

Figura 2. Análise de variância para um experimento em blocos completos casualizados, com perda de uma parcela, usando o procedimento GLM do SAS.

matemático, nesses casos, deve ser alterado, passando agora a conter dois tipos de erros, o experimental (*ee*) e o amostral (*ea*), e tem a forma

$$Y_{ijk} = m + b_j + t_i + ee_{ij} + ea_{ijk} \quad [11]$$

em que $k = 1, 2, ..., K$ e simboliza cada amostra ou cada uma das observações tomadas em cada parcela. Assume-se que os dois erros têm distribuição normal, são independentes, com média 0 e variância respectivamente iguais a δ^2 e σ^2. Da mesma forma, o modelo pode ser considerado aleatório, fixo ou misto. O número de observações pode ser variado em cada parcela, mas, sempre que possível, deve ser coletado em igual número, para maior simplicidade de análise e facilidade de aplicação dos testes F e de comparação de médias. Para estes últimos, sua aplicação seria apenas aproximada. O quadro de análise de variância apresenta agora a forma dada na Tabela 9.

Os valores esperados dos quadrados médios, para os três casos de modelo, estão mostrados na Tabela 10.

Tabela 9. Análise de variância no delineamento em blocos completos ao acaso, com k amostras por parcela.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Blocos (B)	$J - 1$	$\frac{1}{IK}\sum_j B_j^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQB/GLB
Tratamentos (T)	$I - 1$	$\frac{1}{JK}\sum_i T_i^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQT/GLT
Erro experimental (EE)	$(I - 1)(J - 1)$	$\frac{1}{K}\sum_{ij} Y_{ij.}^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2 - SQB - SQT$	$SQEe/GLEe$
Erro amostral (EA)	$IJ(K - 1)$	Diferença	$SQEa/GLEa$
Total (To)	$IJK - 1$	$\sum_{ijk} Y_{ij}^2 - \frac{1}{IJK}\left(\sum_{ijk} Y_{ijk}\right)^2$	

Tabela 10. Esperança matemática dos quadrados médios em blocos completos ao acaso, com k amostras por parcela. Modelos aleatório, fixo e misto.

Fonte de variação	Grau de liberdade	Valor esperado dos quadrados médios		
		Modelo aleatório	Modelo fixo	Modelo misto
Blocos (B)	$J - 1$	$\sigma^2 + K\delta^2 + IK\sigma_b^2$	$\sigma^2 + K\delta^2 + IK\Sigma B_j^2/(J - 1)$	$\sigma^2 + K\delta^2 + IK\sigma_b^2$
Tratamentos (T)	$I - 1$	$\sigma^2 + K\delta^2 + JK\sigma_t^2$	$\sigma^2 + K\delta^2 + JK\Sigma T_i^2/(I - 1)$	$\sigma^2 + K\delta^2 + JK\Sigma T_i^2/(I - 1)$
Erro experimental (EE)	$(I - 1)(J - 1)$	$\sigma^2 + K\delta^2$	$\sigma^2 + K\delta^2$	$\sigma^2 + K\delta^2$
Erro amostral (EA)	$IJ(K - 1)$	σ^2	σ^2	σ^2

A análise da Tabela 10 mostra que há a necessidade de primeiro efetuar um teste F para comprovar ou não a existência de um verdadeiro valor para δ^2. Caso o teste F seja significativo, torna-se evidente que as demais fontes de variação deverão ter suas hipóteses verificadas, com um teste F, comparadas ao erro experimental. Se o teste F para o erro experimental for não significativo, a melhor alternativa é a junção dos dois erros em um único, com graus de liberdade igual à soma dos graus de liberdade dos erros experimentais e amostral. Sua soma de quadrados seria também a soma das somas de quadrados desses mesmos erros, pois nesse caso se consideraria que o valor estimado de δ^2 seria igual a 0. Quando o erro experimental é significativo, ele então deve ser empregado em todos os testes relativos às médias de tratamentos.

Exemplo de aplicação

O pesquisador Paulo Hideo Nakano Rangel, nos ensaios de linhagens promissoras, em blocos ao acaso, mede a altura média dos perfilhos, além de fazer a contagem do número de perfilhos e de panículas em cada uma de cinco plantas tomadas ao acaso em cada parcela. Os dados na Tabela 13 são a altura em centímetros de apenas dez dos tratamentos, de um ensaio conduzido por ele, em 2001, na fazenda Palmital, da Embrapa Arroz e Feijão. A análise de variância encontra-se na Tabela 11.

Nesse caso, em particular, por não se verificar a significância do teste F para o erro experimental, é permitida, portanto, a união dos dois erros, de forma que a análise de variância fica sendo aquela vista na Tabela 12.

Tabela 11. Análise de variância de um ensaio em blocos completos ao acaso, com cinco amostras por parcela.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Blocos (B)	3	437,98	145,99	
Tratamentos (T)	9	713,90	79,32	3,28 (0,0080)
Erro experimental (EE)	27	652,48	24,17	1,52 (0,0604)
Erro amostral (EA)	160	2.548,40	15,93	
Total (To)	199	4.352,76		

Tabela 12. Análise de variância modificada por não haver sido encontrada significância no teste F para o erro experimental.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Blocos (B)	3	437,98	145,99	8,53(< ,0001)
Tratamentos (T)	9	713,90	79,32	4,63(< ,0001)
Erro combinado	187	3.200,88	17,12	
Total (To)	199	4.352,76		

As somas de quadrados foram calculadas segundo as expressões dadas na Tabela 9. Portanto:

$FC = 19.007^2/200 = 1.806.330,24$;

$SQTotal = (93^2 + 95^2 + ...+ 94^2) - FC = 4.352,76$;

$SQBlocos = (4.752^2 + 4.849^2 + 4.765^2 + 4.641^2)/50 - FC = 437,98$;

$SQTratamentos = (1.929^2 + 1.801^2 + \ldots + 1.940^2)/20 - FC = 713,90$;

$SQParcelas = (482^2 + 474^2 + \ldots + 477^2)/5 - FC = 1.804,36$;

$SQEe = 1.804,36 - 437,98 - 713,90 = 652,48$ e a $SQEa$ obtida por diferença.

Para a comparação de médias de tratamentos, utilizando-se o teste de Tukey a 5% de probabilidade o valor do teste é:

$$DHS = 4,53 * \sqrt{\frac{17,12}{20}} = 4,19$$

Se o erro experimental houvesse sido significativo, ele seria o erro a ser empregado nesse teste. Nesse caso, ter-se-ia então:

$$DHS = 4{,}86 * \sqrt{\frac{24{,}17}{20}} = 5{,}34$$

Portanto, o valor da diferença mínima entre duas médias seria 27,4% maior, uma vantagem importante. Deve ser ressaltado que, neste último caso, o uso dos valores médios das medidas de altura de plantas, teria resultado exatamente igual na significância do teste F para tratamentos e no valor do DHS, e praticamente nada se teria de ganho.

Uso do SAS: blocos completos casualizados (mais de uma observação por parcela)

Para o caso de mais de uma observação por parcela, ao usar-se o SAS, deve-se, em geral, proceder á análise em duas etapas. A primeira, para verificar se o erro experimental, aqui expresso como a interação entre BLOCOS e TRATAMENTOS, é significativa ou não. Caso o seja, a análise deve ser repetida agora com um comando de cálculo de médias de tratamentos e a respectiva comparação entre elas, com a especificação correta do erro a ser usado no teste, nesse caso, o erro experimental (BLOCO*TRATAMENTO). Se esse erro experimental não for significativamente diferente do erro amostral, uma nova análise deve ser realizada sem que se especifique no modelo essa interação entre blocos e tratamentos. Esse foi o procedimento adotado também no caso do delineamento inteiramente casualizado (Capítulo 2), o qual deveria ser adotado em todos os outros casos ou delineamentos.

Os programas para a primeira e segunda fases de análise são mostrados a seguir:

Primeira fase:

```
PROC ANOVA;CLASS BLOCO TRATAMENTO PLANTA;
MODEL ALT=BLOCO TRATAMENTO BLOCO*TRATAMENTO;
```

Segunda fase (erro experimental não significativo):

```
PROC ANOVA;CLASS BLOCO TRATAMENTO;
MODEL ALT=BLOCO TRATAMENTO;
MEANS TRATAMENTO/TUKEY;
```

Segunda fase (erro experimental significativo):

```
PROC ANOVA;CLASS BLOCO TRATAMENTO PLANTA;
MODEL ALT=BLOCO TRATAMENTO BLOCO*TRATAMENTO;
TEST H=TRATAMENTO E=BLOCO*TRATAMENTO;
MEANS TRATAMENTO/TUKEY E=BLOCO*TRATAMENTO;
```

Essa programação SAS foi aplicada aos dados da Tabela 13 e os resultados se veem nas Figuras 3 e 4, cujos comentários se seguem. A Figura 3 mostra o resultado da primeira fase da análise em que se verifica que o erro experimental (BLOCO*TRATAMENTO) é não significativo (probabilidade de F maior que 0,05), o que conduz à segunda fase com o modelo adequado para a situação, como se vê na Figura 4.

Tabela 13. Dados de altura de planta de um ensaio conduzido pelo pesquisador Paulo Hideo Nakano Rangel. Delineamento em blocos completos ao acaso, com cinco amostras por parcela.

Tratamento	Amostra	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Soma
BG 90 2	1	93	95	100	95	
BG 90 2	2	95	99	106	92	
BG 90 2	3	96	90	98	94	
BG 90 2	4	98	92	100	89	
BG 90 2	5	100	98	105	94	
Soma		482	474	509	464	1.929
CNAi 9920	1	95	94	98	93	
CNAi 9920	2	93	94	90	90	
CNAi 9920	3	89	90	85	90	
CNAi 9920	4	91	86	90	90	
CNAi 9920	5	93	90	82	78	
Soma		461	454	445	441	1.801
CNAi 9922	1	90	95	93	90	
CNAi 9922	2	101	96	95	90	
CNAi 9922	3	92	90	100	90	
CNAi 9922	4	102	96	95	90	
CNAi 9922	5	103	97	94	95	
Soma		488	474	477	455	1.894
CNAi 9924	1	90	96	99	93	
CNAi 9924	2	94	100	98	89	
CNAi 9924	3	99	100	99	95	
CNAi 9924	4	92	100	97	93	
CNAi 9924	5	95	103	94	105	
Soma		470	499	487	475	1.931

Continua...

Tabela 13. Continuação.

Tratamento	Amostra	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Soma
CNAi 9926	1	84	109	99	90	
CNAi 9926	2	93	99	95	99	
CNAi 9926	3	109	89	100	95	
CNAi 9926	4	95	95	95	90	
CNAi 9926	5	94	98	85	100	
Soma		475	490	474	474	1.913
CNAi 9928	1	94	100	97	97	
CNAi 9928	2	95	105	97	93	
CNAi 9928	3	93	97	95	90	
CNAi 9928	4	100	100	94	92	
CNAi 9928	5	93	100	92	85	
Soma		475	502	475	457	1.909
CNAi 9930	1	95	99	92	94	
CNAi 9930	2	90	100	93	85	
CNAi 9930	3	95	101	91	90	
CNAi 9930	4	92	102	91	102	
CNAi 9930	5	92	97	92	95	
Soma		464	499	459	466	1.888
CNAi 9932	1	90	95	98	93	
CNAi 9932	2	97	95	96	90	
CNAi 9932	3	93	95	97	94	
CNAi 9932	4	100	94	100	99	
CNAi 9932	5	102	95	97	99	
Soma		482	474	488	475	1.919
CNAi 9934	1	92	95	92	96	
CNAi 9934	2	95	101	93	90	
CNAi 9934	3	91	98	95	86	
CNAi 9934	4	95	94	98	95	
CNAi 9934	5	95	100	92	90	
Soma		468	488	470	457	1.883
CNAi 9936	1	102	95	94	95	
CNAi 9936	2	97	97	94	99	
CNAi 9936	3	93	98	99	96	
CNAi 9936	4	98	105	97	93	
CNAi 9936	5	97	100	97	94	
Soma		487	495	481	477	1.940
Total		4.752	4.849	4.765	4.641	19.007

```
                          The ANOVA Procedure
                        Class Level Information
Class         Levels  Values
BLOCO              4  1 2 3 4
TRATAMENTO        10  BG.90.2 CNAi.9920 CNAi.9922 CNAi.9924 CNAi.9926
                      CNAi.9928 CNAi.9930 CNAi.9932 CNAi.9934 CNAi.9936
PLANTA             5  1 2 3 4 5
                     Number of observations    200
                          The ANOVA Procedure
Dependent Variable: ALT
                                  Sum of
 Source                   DF       Squares    Mean Square   F Value   Pr > F
 Model                    39    1804.355000      46.265513      2.90   <.0001
 Error                   160    2548.400000      15.927500
 Corrected Total         199    4352.755000

               R-Square     Coeff Var      Root MSE      ALT Mean
               0.414532      4.199429      3.990927      95.03500

 Source                   DF      Anova SS    Mean Square   F Value   Pr > F
 BLOCO                     3    437.9750000    145.9916667      9.17   <.0001
 TRATAMENTO                9    713.9050000     79.3227778      4.98   <.0001
 BLOCO*TRATAMENTO         27    652.4750000     24.1657407      1.52   0.0604
```

Figura 3. Delineamento em blocos completos com mais de uma observação por parcela, Análise inicial usando o SAS.

```
                          The ANOVA Procedure
                        Class Level Information
Class         Levels  Values
BLOCO              4  1 2 3 4
TRATAMENTO        10  BG.90.2 CNAi.9920 CNAi.9922 CNAi.9924 CNAi.9926
                      CNAi.9928 CNAi.9930 CNAi.9932 CNAi.9934 CNAi.9936
                     Number of observations    200
                          The ANOVA Procedure
Dependent Variable: ALT
                                  Sum of
 Source                   DF       Squares    Mean Square   F Value   Pr > F
 Model                    12    1151.880000      95.990000      5.61   <.0001
 Error                   187    3200.875000      17.116979
 Corrected Total         199    4352.755000

               R-Square     Coeff Var      Root MSE      ALT Mean
               0.264632      4.353414      4.137267      95.03500

 Source                   DF      Anova SS    Mean Square   F Value   Pr > F
 BLOCO                     3    437.9750000    145.9916667      8.53   <.0001
 TRATAMENTO                9    713.9050000     79.3227778      4.63   <.0001
                          The ANOVA Procedure
                Tukey's Studentized Range (HSD) Test for ALT
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.

                Alpha                                    0.05
                Error Degrees of Freedom                  187
                Error Mean Square                    17.11698
                Critical Value of Studentized Range  4.52879
                Minimum Significant Difference        4.1897

        Means with the same letter are not significantly different.

 Tukey Grouping          Mean     N    TRATAMENTO
               A        97.000    20    CNAi.9936
               A        96.550    20    CNAi.9924
               A        96.450    20    BG.90.2
               A        95.950    20    CNAi.9932
               A        95.650    20    CNAi.9926
               A        95.450    20    CNAi.9928
               A        94.700    20    CNAi.9922
               A        94.400    20    CNAi.9930
          B    A        94.150    20    CNAi.9934
          B             90.050    20    CNAi.9920
```

Figura 4. Análise de variância com o uso de SAS para delineamento em blocos completos ao acaso, com mais de uma observação por parcela e erro experimental não significativo.

Teste de aditividade

Em 1949, para blocos completos ou ensaios de dupla classificação, com uma observação por unidade experimental, Tukey introduziu um teste para aditividade por meio da obtenção de uma soma de quadrados, com 1 grau de liberdade, que é devida à aditividade como segue:

a) Obtenção dos efeitos de tratamentos (t_i) e de blocos (b_j);

b) Cálculo dos Q_j dados por: $Q_j = \sum_i t_i * Y_{ij}$;

c) Obtenção do valor Q, dado por: $Q = \sum_j Q_j * b_j$;

d) $SQAdit = \frac{Q^2}{K}$, em que $K = \sum_i t_i^2 * \sum_j b_j^2$;

e) $SQResidual = SQErro - SQAdit$, com $(I - 1)(J - 1) - 1$ graus de liberdade.

Nas fórmulas apresentadas, os efeitos de tratamentos (t_i) e de blocos (b_j) são obtidos por meio da diferença entre as médias dos tratamentos e as médias de blocos e a média geral do ensaio, respectivamente, enquanto Y_{ij} representa o valor obtido na parcela correspondente ao tratamento i no bloco j.

O efeito da falta, ou não, de aditividade é verificado por meio de um teste F, dividindo a soma de quadrado de aditividade pelo quadrado médio residual. Significância nesse teste indica a falta de aditividade no modelo e, portanto, a necessidade de transformação nos dados.

Exemplo de aplicação

Utilizando-se os resultados obtidos nas Tabelas 4 e 6, com o experimento conduzido por Peloso e colaboradores, os efeitos de tratamentos e de blocos são aqueles mostrados na Tabela 14.

A aplicação das fórmulas dadas no item *Teste de aditividade* aos resultados obtidos nas Tabelas 5 e 14 resulta em:

Q_1 = (-642,32*1.112,5) + (-17,32*1.850,0 + ... + (551,48*2.112,5) = 2.161.020,83 e similarmente Q_2 = 2.334.731,77; Q_3 = 2.044.406,25 e Q_4 = 2.213.893,23.

Tabela 14. Obtenção dos efeitos de tratamentos e blocos para análise de aditividade.

	Efeito de tratamentos	Bloco 1	Bloco 2	Bloco 3	Bloco 4
CNFP 8015	-642,32	1.112,5	650,0	1.037,5	1.362,5
CNFP 8016	-17,32	1.850,0	1.637,5	1.787,5	1.387,5
CNFP 8017	-336,02	987,5	1.537,5	1.600,0	1.262,5
CNFP 8018	320,18	1.587,5	2.412,5	2.100,0	1.912,5
CNFP 8019	326,48	2.012,5	2.075,0	2.087,5	1.862,5
CNFP 8020	301,48	1.975,0	1.787,5	2.225,0	1.950,0
CNFP 8021	470,18	1.837,5	2.137,5	2.050,0	2.587,5
CNFP 8022	4,58	1.825,0	2.100,0	1.375,0	1.450,0
CNFP 8023	292,08	2.250,0	2.062,5	1.950,0	1.637,5
CNFP 8024	-564,12	562,5	1.437,5	1.550,0	925,0
CNFP 8025	-236,02	1.650,0	1.625,0	1.187,5	1.325,0
CNFP 8026	-579,82	1.362,5	862,5	1.225,0	962,5
Diamante negro	-79,82	1.600,0	1.725,0	1.750,0	1.337,5
FT Nobre	188,98	2.162,5	1.825,0	1.950,0	1.550,0
G. Brilhante	551,48	2.112,5	2.137,5	2.450,0	2.237,5
Efeito de blocos		-23,75	51,25	72,08	-99,59

$Q = (-23,75*2.161.020,83) + ... + (-99,59*2.213.893,23) = -4.768.491,753$; $K = [(-642,32)^2 + (-17,32)^2 + ... + (551,48)^2]*[(-23,75)^2 + ... + (-99,59)^2 = 40.057.387.285$.

Portanto, $SQAditividade = (-4.768.491,753)^2/40.057.387.285 = 567,65$ com 1 grau de liberdade; logo, a soma de quadrados residual passa a ser $3.002.010,42 - 567,61 = 3.001.442,77$ com 41 graus de liberdade.

O quociente entre seus quadrados médios resulta em: $F = 567,65/73205,92 = 0,008$.

Esse resultado é, evidentemente, não significativo, o que implica afirmar que a aditividade pode ser admitida sem qualquer problema. O SAS não apresenta opções para esse teste.

Usos diferenciados do delineamento em blocos completos

Existem casos em que pode haver a necessidade de alguma alteração de uso deste delineamento. Dois exemplos dessa situação são dados a seguir. Pimentel-Gomes (2000)

mostrou a possibilidade de que experimentos com baixo número de tratamentos, dois ou três, por exemplo, possam ser aplicados a mais de uma parcela em cada bloco. Outro caso é aquele que ocorre em experimentos com inseticidas para controle de insetos, em que o uso de testemunha sem inseticida pode ser repetida várias vezes em cada bloco. Este último caso se justifica pelo fato de, ao se ter a testemunha em uma única parcela em cada bloco, a área não tratada ser diminuta em relação às tratadas podendo apresentar número muito superior de insetos quando comparado à população normal de insetos, pois eles buscariam aí um refúgio contra os efeitos dos inseticidas. Por causa disso, a testemunha tem seu efeito exacerbado e mostra efeito relativo muito mais eficiente de controle dos produtos químicos do que o real. Outros exemplos podem ser facilmente encontrados em outras áreas de investigação.

Em ambas as situações, há um ganho em precisão, com maior número de graus de liberdade para o erro ou resíduo, enquanto no segundo ocorre adicionalmente uma diferenciação na precisão das medidas dos tratamentos, pois a testemunha terá muito mais repetições do que os demais tratamentos, mas esse inconveniente é normalmente compensado por uma maior precisão na avaliação do efeito real dos inseticidas.

Repetição de todos os tratamentos em cada bloco

Neste caso, o modelo matemático apresenta uma pequena variação, que é a adição de um indexador ($k = 1, 2, ..., K$) de identificação da parcela de cada tratamento em cada bloco. Dessa forma, o modelo apresenta a expressão

$$Y_{ijk} = m + b_j + t_i + e_{ijk} \quad [12]$$

e o quadro da análise de variância será construído mostrando essa alteração. É interessante notar que esses ensaios poderiam ser montados com o mesmo número de parcelas, em KJ blocos, mas, da maneira aqui mostrada, o erro nesse caso tem $J(K - 1)$ graus de liberdade a mais do que o experimento com mais blocos. Os testes de F são obtidos da maneira usual (Tabela 15).

Exemplo de aplicação e uso do SAS

Um dos exemplos clássicos desta situação é dado por Pimentel-Gomes (2000), tendo como base os dados de um experimento em que se testaram duas formas de aração do solo, realizado pelo engenheiro-agrônomo Duvilio Ometto, em blocos completos, com seis blocos e duas aplicações de cada forma de aração por bloco. Nessa obra, o leitor pode ver o exemplo calculado manualmente. O uso do SAS para esse tipo de ensaio é muito simples, já que o modelo é o mesmo adotado para os blocos completos tradicionais, ou seja:

Tabela 15. Análise de variância de um ensaio em blocos completos ao acaso, com todos os tratamentos repetidos k vezes em cada bloco.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Blocos (B)	$J - 1$	$\frac{1}{IK}\sum_j B_j^2 - \frac{1}{IJK}\left(\sum_{ij} Y_{ijk}\right)^2$	SQB/GLB
Tratamentos (T)	$I - 1$	$\frac{1}{JK}\sum_i T_i^2 - \frac{1}{IJK}\left(\sum_{ij} Y_{ijk}\right)^2$	SQT/GLT
Erro ou resíduo (E)	$IJK - J - I + 1$	$SQTo - SQB - SQT$	SQE/GLE
Total (To)	$IJK - 1$	$\frac{1}{JK}\sum_i T_i^2 - \frac{1}{IJK}\left(\sum_{ij} Y_{ijk}\right)^2$	

```
PROC ANOVA;CLASS BLOCO TRATAMENTO;
MODEL PROD=BLOCO TRATAMENTO;
MEANS TRAT;
```

Como esse experimento só apresenta dois tratamentos, não há a necessidade de comparação de médias, pois o teste F é suficiente para discriminar entre eles. O resultado de tal análise encontra-se na Figura 5 cujos resultados diferem daqueles de Pimentel Gomes por questão de aproximação.

Testemunha repetida mais de uma vez por bloco

O indexador k, também incluído no modelo neste caso, seria agora identificado como k_i, em que $K_i = 1, 2, \ldots, K$ se $i = 0$ (testemunha) e $k_i = 1$ nos demais casos. O quadro de análise de variância será agora calculado na forma mostrada na Tabela 16.

Como sempre os testes de F são obtidos por meio dos quocientes com o quadrado médio do erro. Nesse caso em particular, deve ser notado que a testemunha tem JK observações ao passo que os demais tratamentos apenas J observações, e isso implicará em modificações nas fórmulas dos testes para comparações de médias, se for desejada a comparação entre todos os pares de médias. Mas é evidente que, nesse caso, o teste mais

adequado seria o teste de Dunnett, específico para comparações das médias dos distintos tratamentos com a média da testemunha.

```
                          The ANOVA Procedure
                       Class Level Information
                Class           Levels    Values
                BLOCO                6    1 2 3 4 5 6
                TRATAMENTO           2    AN AP
                     Number of observations    24
                          The ANOVA Procedure
Dependent Variable: PROD
                                  Sum of
 Source                    DF       Squares    Mean Square   F Value   Pr > F
 Model                      6    15.63000000     2.60500000      3.67   0.0160
 Error                     17    12.06333333     0.70960784
 Corrected Total           23    27.69333333

              R-Square     Coeff Var      Root MSE     PROD Mean
              0.564396      13.55038      0.842382      6.216667

 Source                    DF       Anova SS    Mean Square   F Value   Pr > F
 BLOCO                      5    13.92333333     2.78466667      3.92   0.0151
 TRATAMENTO                 1     1.70666667     1.70666667      2.41   0.1394
                          The ANOVA Procedure
              Level of              ------------PROD-----------
              TRATAMENTO       N            Mean          Std Dev
              AN              12      5.95000000       1.09917324
              AP              12      6.48333333       1.07435675
```

Figura 5. Análise de variância de um bloco completo casualizado, dois tratamentos, seis blocos, com duas aparições de cada tratamento nos blocos com o uso do PROC ANOVA do SAS.

Tabela 16. Análise de variância de um experimento em blocos completos ao acaso, modificado, com o tratamento testemunha repetido k vezes em cada bloco.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Blocos (B)	J - 1	$\frac{1}{I+K-1}\sum_{j} B_j^2 - \frac{1}{J(I+K-1)}\left(\sum_{ijk_i} Y_{ijk_i}\right)^2$	*SQB*/*GLB*
Tratamentos (T)	I - 1	$\frac{1}{J}\sum_{i\neq 0} T_i^2 + \frac{T_0^2}{JK} - \frac{1}{J(I+K-1)}\left(\sum_{ijk_i} Y_{ijk_i}\right)^2$	*SQT*/*GLT*
Erro ou resíduo (E)	$(I - 1)(J - 1) + (K - 1)J$	*SQTo - SQB - SQT*	*SQE*/*GLE*
Total (To)	IJ - 1	$\sum_{ijk_i} Y_{ijk_i}^2 - \frac{1}{J(I+K-1)}\left(\sum_{ijk_i} Y_{ijk_i}\right)^2$	

Exemplo de aplicação com o uso do SAS

Vamos supor um experimento de aplicação de inseticidas agrícolas em arroz, em que se testaram quatro produtos novos e um tratamento testemunha sem aplicação de inseticidas. O delineamento é o de blocos completos casualizados, com quatro repetições, mas, nesse caso, em cada bloco, havia quatro parcelas com o tratamento testemunha. Dessa forma, os blocos tinham oito parcelas, das quais quatro com os inseticidas e quatro com a testemunha, todas distribuídas ao acaso. A Tabela 17 mostra os resultados obtidos em kg ha^{-1}.

Tabela 17. Dados de produção de arroz em kg ha^{-1}, de um experimento fictício com cinco tratamentos, nos quais a testemunha era repetida quatro vezes em cada bloco.

Tratamentos	Bloco 1	Bloco 2	Bloco 3	Bloco 4
Testemunha	2.354	2.145	2.000	2.198
Testemunha	2.123	2.234	1.980	2.001
Testemunha	2.200	2.147	2.076	2.190
Testemunha	2.148	2.341	2.075	2.165
Inseticida-1	2.256	2.248	2.165	2.346
Inseticida-2	2.378	2.400	2.376	2.303
Inseticida-3	2.590	2.820	2.700	2.487
Inseticida-4	3.054	2.943	2.965	2.689

A Figura 6 mostra o resultado da análise de variância feito com o SAS, utilizando-se o seguinte programa:

```
PROC ANOVA;CLASS BLOCO TRATAMENTO;
MODEL PROD=BLOCO TRATAMENTO;
MEANS TRATAMENTO;
MEANS TRATAMENTO/DUNNETT("TESTEMUNHA");
```

Como se pode ver, por duas vezes foi feita a especificação do comando MEANS. A primeira apenas para que se tivesse a impressão das médias dos tratamentos e a segunda para que se obtivesse a comparação dessas médias com o uso do teste de Dunnett, específico para a comparação de diversos tratamentos contra uma testemunha, como discutido no Capítulo 1. A análise de variância mostra um ganho de 12 graus de liberdade no erro, em razão do aumento do número de parcelas com o tratamento testemunha. O teste de Dunnett mostra que todos os tratamentos com inseticidas foram diferentes da testemunha. Se houvesse a necessidade de comparar todos os tratamentos entre si, outro teste

```
                          The ANOVA Procedure
                       Class Level Information
Class         Levels  Values
bloco              4  1 2 3 4
tratamento         5  INSETICIDA-1 INSETICIDA-2 INSETICIDA-3 INSETICIDA-4
                      TESTEMUNHA
                          Number of observations    32
                          The ANOVA Procedure
Dependent Variable: prod
                                 Sum of
 Source                   DF       Squares    Mean Square   F Value   Pr > F
 Model                     7    2400378.875     342911.268     33.51   <.0001
 Error                    24     245570.594      10232.108
 Corrected Total          31    2645949.469

               R-Square     Coeff Var      Root MSE     prod Mean
               0.907190      4.310324      101.1539      2346.781

 Source                   DF       Anova SS    Mean Square   F Value   Pr > F
 bloco                     3      88656.344      29552.115      2.89   0.0564
 tratamento                4    2311722.531     577930.633     56.48   <.0001
                          The ANOVA Procedure
            Level of                 ------------prod-----------
            tratamento        N             Mean          Std Dev

            INSETICIDA-1      4       2253.75000        73.992680
            INSETICIDA-2      4       2364.25000        42.256163
            INSETICIDA-3      4       2649.25000       143.255890
            INSETICIDA-4      4       2912.75000       156.695831
            TESTEMUNHA       16       2148.56250       108.694660
                          The ANOVA Procedure
                       Dunnett's t Tests for prod
NOTE: This test controls the Type I experimentwise error for comparisons of
all treatments against a control.

                    Alpha                          0.05
                    Error Degrees of Freedom         24
                    Error Mean Square           10232.11
                    Critical Value of Dunnett's t  2.67086
        Comparisons significant at the 0.05 level are indicated by ***.

                                  Difference
            tratamento              Between     Simultaneous 95%
            Comparison                Means   Confidence Limits

       INSETICIDA-4 - TESTEMUNHA      764.19      613.16   915.22  ***
       INSETICIDA-3 - TESTEMUNHA      500.69      349.66   651.72  ***
       INSETICIDA-2 - TESTEMUNHA      215.69       64.66   366.72  ***
       INSETICIDA-1 - TESTEMUNHA      105.19      -45.84   256.22
```

Figura 6. Análise de um experimento em blocos completos casualizados, em que o tratamento testemunha é repetido mais de uma vez por bloco, com o uso do SAS.

poderia ser aplicado, tal como Tukey, SNK, etc., bastando para isso especificar o novo teste num dos comandos MEANS. Nunca é demais recordar que os testes de médias só têm seu resultado validado se o teste F para tratamentos for significativo, como se vê na Figura 6, em que a probabilidade do teste F para tratamentos é muito inferior a 1%, mas apenas um dos tratamentos com inseticida difere da testemunha.

As análises de variância e as comparações de médias dos dois últimos exemplos desse capítulo, a serem feitas segundo as fórmulas dadas nas Tabelas 15 e 16, são deixadas aos leitores como exercícios.

Capítulo 4

Delineamento em quadrado latino

Neste delineamento, o princípio básico do controle local aparece de forma dupla, isto é, dois tipos de controle ou de blocos estão presentes. Na versão original, em agricultura, os blocos passaram a ser chamados de colunas e linhas e se apresentavam em forma perpendicular um ao outro. Isso se explica pelo fato de, na agricultura, tentar-se dessa forma controlar a variabilidade do solo em dois sentidos, e por isso o experimento em campo apresentava-se na forma de um quadrado. Essa dupla apresentação de blocos completos faz que o número de repetições seja igual ao de tratamentos; portanto, o desenho apresenta um número total de parcelas igual ao quadrado do número de tratamentos. A aparente vantagem de maior controle ambiental é bastante prejudicada, pois esse desenho leva a um aumento muito rápido de parcelas experimentais, o que pode inviabilizar sua execução e limitar o número de tratamentos que podem ser testados. Em geral, é recomendado que o número de tratamentos varie entre 3 e 8, podendo ser encontrados casos de até 12 tratamentos como número máximo. Outra desvantagem aparece com baixo número de tratamentos, pois aí o número de graus de liberdade do erro é também baixo. Como será visto mais adiante, os graus de liberdade do erro são dados por $(K - 1)(K - 2)$, em que K é o número de tratamentos. Portanto, verifica-se que é impossível trabalhar com apenas dois tratamentos, pois, nesse caso, o erro teria 0 (zero) grau de liberdade e não seria possível sua estimação.

Um exemplo de quadrado latino (com quatro tratamentos) é dado na Figura 1, na qual os tratamentos são identificados pelas letras latinas A, B, C e D. Nesse caso, vê-se

A	C	D	B
B	D	A	C
C	A	B	D
D	B	C	A

Figura 1. Esquema experimental de um quadrado latino, com quatro tratamentos, após casualização dos tratamentos.

claramente que, em cada linha e em cada coluna, todos os tratamentos aparecem uma única vez, formando assim blocos em dois sentidos.

Hoje em dia, esse duplo controle de variabilidade ambiental foi expandido e adaptado para os diversos tipos de problemas ou áreas de investigação, com a única condição de que esses fatores de bloqueio sejam independentes entre si e dos tratamentos, isto é, que não haja interação entre eles. Então, em experimentos agrícolas, é possível dispor todas as parcelas lateralmente. Dessa forma, cada posição dentro do bloco (linha) corresponde a uma coluna, muito útil quando ocorre um gradiente na variabilidade do solo, como se pode ver na Figura 2.

A C D B B D A C C A B D D B C A

Figura 2. Esquema experimental de um quadrado latino, com quatro tratamentos, após casualização dos tratamentos, modificado para medir gradiente de solo.

Outros exemplos de aplicações de quadrado latino em agricultura e pecuária são: pastos como linhas e animais como colunas em experimentos de adição suplementar de sais (tratamentos); tratores como linhas e as posições das rodas como colunas em testes de pneumáticos; safra utilizada como linha e provador como coluna em testes de tipos de processos de fabricação de queijos, sucos ou outros tipos de comidas e bebidas, e muitos outros.

A casualização de um quadrado latino deve ser sempre iniciada pela adoção de uma distribuição sistemática dos tratamentos em linhas e colunas, chamados de quadrados estandares em que a primeira coluna e a primeira linha possuem os tratamentos dispostos em ordem natural, como mostrados ao final deste capítulo. O passo seguinte consiste na casualização das linhas e depois das colunas, ou vice-versa. Para exemplificar essa aplicação, usar-se-á um experimento com quatro tratamentos. O primeiro passo é a adoção de uma forma de distribuição padrão, a primeira delas, por exemplo, é dada na Figura 3.

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

Figura 3. Esquema básico, padrão de um quadrado latino 4 x 4.

O passo 2 consiste na casualização das linhas (por exemplo: 2 4 1 3); e o passo 3, das colunas (por exemplo: 1 4 2 3). Em seguida, devem-se designar ao acaso as identificações (A, B, C, D) aos tratamentos. Os passos 2 e 3 apareceriam como se vê na Figura 4, de maneira que a última forma apresentada seria o desenho do experimento na sua implantação a campo.

B	A	D	C
D	C	A	B
A	B	C	D
C	D	B	A

B	C	A	D
D	B	C	A
A	D	B	C
C	A	D	B

Figura 4. Esquemas mostrando os resultados parcial e final de casualização de um quadrado latino 4 x 4.

Modelo matemático e análise de variância

A representação matemática mais comum para este delineamento experimental é

$$Y_{ijk} = m + c_i + l_j + t_k + e_{ijk} \quad [1]$$

em que i, j e $k = 1, 2, 3, ..., K$. Nesse caso, Y_{ijk} é a leitura do tratamento k na coluna i, linha j; m a média; c_i o efeito da coluna i; l_j o efeito da linha j; t_k o efeito do tratamento k e e_{ijk} o erro associado a essa medida, assumido ter distribuição normal, com média 0 e variância σ^2. Os erros são considerados independentes e de variância homogênea. Deve ser lembrado que os índices i, j e k não aparecem em todas as suas combinações possíveis, isto é, para cada combinação ij, existe um único valor de k, o do tratamento aplicado àquela combinação de linha e coluna.

A análise de variância é mostrada na Tabela 1.

Os testes de F, para tratamentos, linhas e colunas, independentemente do tipo do modelo, aleatório, fixo ou misto, são sempre obtidos pelos quocientes entre os quadrados médios respectivos e o quadrado médio do erro.

Tabela 1. Esquema de análise de variância de experimento em quadrado latino.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Coluna (C)	$K - 1$	$\frac{1}{K}\sum_i C_i^2 - \frac{1}{K^2}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQC/GLC
Linha (L)	$K - 1$	$\frac{1}{K}\sum_j L_j^2 - \frac{1}{K^2}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQL/GLL
Tratamento (T)	$K - 1$	$\frac{1}{K}\sum_k T_k^2 - \frac{1}{K^2}\left(\sum_{ijk} Y_{ijk}\right)^2$	SQT/GLT
Erro (E)	$(K - 1)(K - 2)$	$SQTo - SQC - SQL - SQT$	SQE/GLE
Total (To)	$K^2 - 1$	$\sum_{ijk} Y_{ijk}^2 - \frac{1}{K^2}\left(\sum_{ijk} Y_{ijk}\right)^2$	

Exemplo de aplicação

Os dados da Tabela 2 foram obtidos pelo pesquisador Evane Ferreira, em um experimento realizado em 1998, cujo objetivo era medir a resposta de um grupo de oito genótipos de arroz ao ataque inicial de pragas e se referem à produção de grãos em kg ha^{-1}. Os tratamentos eram: A: CNA 8540; B: CNA 8711; C: CNA 8172; D: CNA 8436; E: Primavera; F: Caiapó; G: Maravilha e H: Canastra.

Na Tabela 3, encontra-se a análise de variância, obtido com a aplicação das fórmulas antes descritas.

As somas de quadrados de colunas, linhas e tratamentos são dadas, respectivamente, por

$$\frac{1}{8}\sum_i (9.041{,}25^2 + \ldots + 9.878{,}75^2) - \frac{1}{8^2}(72.401{,}25)^2 = 721.891{,}58;$$

$$\frac{1}{8}(8.652{,}50^2 + \ldots + 9.590{,}00^2) - \frac{1}{8^2}(72.401{,}25)^2 = 1.710.324{,}78 \quad \text{e}$$

$$\frac{1}{8}(10.755{,}00^2 + \ldots + 11.353{,}75^2) - \frac{1}{8^2}(72.401{,}25)^2 = 6.269.123{,}60.$$

Tabela 2. Dados de produção de grãos de arroz, em kg ha^{-1}, em um experimento em quadrado latino 8 x 8.

	Coluna 1	Coluna 2	Coluna 3	Coluna 4	Coluna 5	Coluna 6	Coluna 7	Coluna 8	Total	Total de tratamentos
Linha 1	A 1.303,75	D 1.481,25	G 586,25	C 708,75	F 1.193,75	E 825,00	H 1.856,25	B 697,50	8.652,50	A 10.755,00
Linha 2	B 618,75	E 1.251,25	H 822,50	D 846,25	G 1.051,25	F 1.070,00	A 1.126,25	C 1.301,25	8.087,50	B 3.086,25
Linha 3	C 1.176,25	F 1.742,50	A 748,75	E 716,25	H 952,50	G 1.470,00	B 377,50	D 1.162,50	8.346,25	C 8.527,50
Linha 4	D 852,50	G 1.680,00	B 317,50	F 1.038,75	A 1.326,25	H 1.250,00	C 962,50	E 738,75	8.166,25	D 11.196,25
Linha 5	E 700,00	H 1.897,50	C 1.096,25	G 945,00	B 216,25	A 1.767,50	D 1.682,50	F 1.416,25	9.721,25	E 8.176,25
Linha 6	F 1.003,75	A 1.726,25	D 2.148,75	H 855,00	C 487,50	B 237,50	E 1.198,75	G 1.498,75	9.156,25	F 9.685,00
Linha 7	G 1.395,00	B 313,75	E 1.543,75	A 1.421,25	D 1.627,50	C 1.367,50	F 1.283,75	H 1.728,75	10.681,25	G 9.621,25
Linha 8	H 1.991,25	C 1.427,50	F 936,25	B 307,50	E 1.202,50	D 1.395,00	G 995,00	A 1.335,00	9.590,00	H 11.353,75
Total	9.041,25	11.520,00	8.200,00	6.838,75	8.057,50	9.382,50	9.482,50	9.878,75	72.401,25	

Tabela 3. Análise de variância para os dados da Tabela 2.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Linha	7	721.891,58	103.127,37	0,95 (0,48)
Coluna	7	1.710.324,78	244.332,11	2,26 (0,048)
Tratamento	7	6.269.123,61	895.589,09	8,27 (< ,0001)
Erro	42	4.548.852,20	108.306,00	
Total	63	13.250.192,16		

A soma de quadrados total é obtida por

$$\left(1.303,75^2 + \ldots + 1.355,00^2\right) - \frac{1}{8^2}\left(72.401,25\right)^2 = 13.250.192,16\,;$$

e a soma de quadrados do erro por diferença, como usual.

Dada a significância do teste F para tratamentos, caberia uma comparação de médias. Como entre os tratamentos estão incluídas quatro linhagens e quatro cultivares em uso corrente pelos agricultores, o teste de Dunnett poderia ser aplicado apenas no caso em que uma das cultivares fosse escolhida como testemunha. Fora isso, qualquer outro teste é perfeitamente cabível. No caso atual, cabe a aplicação do teste de Tukey.

A diferença honestamente significativa é igual a

$$DHS = 4,51 * \sqrt{\frac{108.306,00}{8}} = 524,65$$

e isso conduz ao seguinte agrupamento:

Canastra 1.419,2 a

CNA 8436 1.399,5 a

CNA 8540 1.344,4 a

Caiapó 1.210,6 a

Maravilha 1.202,7 a

CNA 8172 1.065,9 a

Primavera 1.022,0 a

CNA 8711 385,8 b

SAS e o delineamento quadrado latino

Assim como no caso do delineamento em blocos completos casualizados, o quadrado latino pode ser analisado com o uso do procedimento ANOVA, desde que não haja perda de dados. O programa, de maneira bem simples, é o seguinte:

```
PROC ANOVA;CLASS LINHA COLUNA TRATAMENTO;
MODEL PRODUCAO=LINHA COLUNA TRATAMENTO;
MEANS TRATAMENTO/TUKEY;
```

Mais uma vez, solicita-se a comparação de médias pelo teste de Tukey, mas uma série de outros testes poderia ser utilizada. O resultado de tal análise pode ser visualizado na Figura 5.

```
                    The ANOVA Procedure
                  Class Level Information
Class           Levels    Values
LINHA                8    1 2 3 4 5 6 7 8
COLUNA               8    1 2 3 4 5 6 7 8
TRATAMENTO           8    A B C D E F G H

                Number of observations    64
                    The ANOVA Procedure
Dependent Variable: PRODUCAO
                                  Sum of
Source                    DF       Squares    Mean Square   F Value   Pr > F
Model                     21     8701339.97      414349.52      3.83   0.0001
Error                     42     4548852.20      108306.00
Corrected Total           63    13250192.16

           R-Square     Coeff Var      Root MSE    PRODUCAO Mean
           0.656695      29.09110      329.0988         1131.270

Source                  DF       Anova SS    Mean Square   F Value   Pr > F
LINHA                    7     721891.577     103127.368      0.95   0.4779
COLUNA                   7    1710324.780     244332.111      2.26   0.0483
TRATAMENTO               7    6269123.608     895589.087      8.27   <.0001
                    The ANOVA Procedure
          Tukey's Studentized Range (HSD) Test for PRODUCAO
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.

                Alpha                                   0.05
                Error Degrees of Freedom                  42
                Error Mean Square                     108306
                Critical Value of Studentized Range  4.50911
                Minimum Significant Difference        524.65

        Means with the same letter are not significantly different.
      Tukey Grouping          Mean      N    TRATAMENTO
                  A        1419.2      8    H
                  A        1399.5      8    D
                  A        1344.4      8    A
                  A        1210.6      8    F
                  A        1202.7      8    G
                  A        1065.9      8    C
                  A        1022.0      8    E
                  B         385.8      8    B
```

Figura 5. Resultado da análise de variância de um ensaio em quadrado latino com o uso do procedimento ANOVA, do SAS. Dados na Tabela 2.

Dados perdidos em quadrado latino

Em qualquer experimento é factível a perda de dados, fato nunca desejado, mas que ocorre com certa frequência. Neste delineamento, pode-se perder uma, duas ou mais parcelas, ou perder-se totalmente uma linha, uma coluna ou um tratamento. Serão analisados, a seguir, os casos de uma ou duas parcelas perdidas e de linha, coluna ou tratamento perdido.

Uma parcela perdida

Neste caso, o seu valor pode ser estimado pela fórmula

$$X = \frac{K\,(L + C + T) - 2 * G}{(K - 1)\,(K - 2)} \qquad [2]$$

em que K representa o número de tratamentos; L, C e T os totais da linha, da coluna e do tratamento que apresentam a parcela perdida; e G o total disponível em todo o experimento.

Essa perda de informação, como se viu no caso de blocos completos, implica a diminuição de 1 grau de liberdade às somas de quadrados total e de erro.

No caso de comparação de médias, deve-se lembrar da variância dos contrastes que envolvem o tratamento com dado perdido, esse fato deve ser levado em consideração, ao utilizar-se para ele $K - 1$ repetições, como se vê no exemplo a seguir, e ter presente que o teste, nesse caso, é aproximado.

Exemplo

Suponha-se que o resultado referente ao tratamento E na linha 3 e coluna 4 tivesse sido perdido. Aos totais de linha (8.346,25), coluna (6.838,75) e tratamento (8.176,25) correspondentes deveria então ser subtraído o valor 716,25, e a aplicação da fórmula 2 daria a estimativa

$$X = \frac{8 * (7.630{,}0 + 6.122{,}5 + 7.460{,}0) - 2 * 71.685{,}05}{7 * 6} = 626.90\,,$$

que, como se vê, não é muito próxima ao valor real.

A análise de variância é agora aquela mostrada na Tabela 4.

Tabela 4. Análise de variância do ensaio em quadrado latino, quando se considerou o dado do tratamento E, linha 3, coluna 4, como perdido.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Linha	7	738.488,27	105.498,32	0,95 (0,48)
Coluna	7	1.760.595,254	251.513,61	2,27 (0,047)
Tratamento	7	6.289.517,676	898.502,53	8,11 (< 0,0001)
Erro	41	4.543.613,63	110.819,84	
Total	62	13.332.214,83		

A soma de quadrados de total foi obtida com o uso do valor 626,90 substituindo o valor perdido, enquanto para as somas de quadrados de coluna, linha e tratamento, foram utilizados os novos totais da linha 3 (7.630,00 + 626,90 = 8.256,90), da coluna 4 (6.122,50 + 626,90 = 6.749,40) e do tratamento E (7.460,00 + 626,90 = 8.086,90), com a aplicação das mesmas fórmulas utilizadas quando não há dado perdido. A perda da parcela experimental é considerada com a diminuição de 1 grau de liberdade no total e no erro.

No caso de comparação de médias, todos os tratamentos apresentam as mesmas médias anteriores, à exceção do tratamento E, cuja média é agora igual a 1.010,86, ligeiramente inferior à verdadeira média que era 1.022,0.

Uso do SAS no delineamento quadrado latino com um dado perdido

Assim como no caso do delineamento anterior, blocos completos casualizados, quando ocorre perda de dados no experimento, ao usar-se o SAS o procedimento adequado é o PROC GLM, que, como se vê, apresenta algumas discrepâncias nas somas de quadrados, nesse caso, linha, coluna e tratamentos, mas estima corretamente o erro e a(s) média(s) do(s) tratamento(s) com dado(s) perdido(s). O programa a ser empregado é então escrito:

```
PROC GLM;CLASS LINHA COLUNA TRATAMENTO;
MODEL PRODUCAO=LINHA COLUNA TRATAMENTO;
LSMEANS TRATAMENTO;
```

O resultado dessa análise pode ser conferido na Figura 6.

```
                    The GLM Procedure
                Class Level Information
Class           Levels    Values
LINHA                8    1 2 3 4 5 6 7 8
COLUNA               8    1 2 3 4 5 6 7 8
TRATAMENTO           8    A B C D E F G H

                Number of observations    64
NOTE: Due to missing values, only 63 observations can be used in this
analysis.
                    The GLM Procedure

Dependent Variable: PRODUCAO
                                  Sum of
Source                     DF       Squares    Mean Square   F Value   Pr > F
Model                      21    8531638.38     406268.49      3.67   0.0002
Error                      41    4543663.92     110821.07
Corrected Total            62   13075302.31

           R-Square     Coeff Var      Root MSE    PRODUCAO Mean
           0.652500      29.25655      332.8980          1137.858

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
LINHA                       7     669150.398      95592.914      0.86   0.5437
COLUNA                      7    1591441.503     227348.786      2.05   0.0714
TRATAMENTO                  7    6271046.482     895863.783      8.08   <.0001

Source                     DF    Type III SS    Mean Square   F Value   Pr > F
LINHA                       7     726011.685     103715.955      0.94   0.4897
COLUNA                      7    1653609.634     236229.948      2.13   0.0615
TRATAMENTO                  7    6271046.482     895863.783      8.08   <.0001
                    The GLM Procedure
                Least Squares Means

                                  PRODUCAO
TRATAMENTO                         LSMEAN
A                               1344.37500
B                                385.78125
C                               1065.93750
D                               1399.53125
E                               1010.86280
F                               1210.62500
G                               1202.65625
H                               1419.22500
```

Figura 6. Análise de variância de um quadrado latino, com perda de parcela, usando-se o procedimento GLM do SAS.

Duas parcelas perdidas

A perda de duas parcelas pode ocorrer de quatro formas distintas: numa mesma coluna, numa mesma linha, em linha e coluna diferentes para um mesmo tratamento ou para tratamentos diferentes. A cada caso corresponderão duas fórmulas de cálculo. A obtenção dessas fórmulas se faz da mesma forma empregada em blocos completos, e exemplifica-se a seguir para o caso mais complexo, de linha, coluna e tratamentos diferentes.

Fazendo Z igual à soma de quadrados de erro, com X e Y substituindo os valores perdidos e simplificando a expressão após eliminarem-se os termos independentes de X e Y, tem-se

$$Z = (X^2 + Y^2) - \frac{(C_x + X)^2 + (C_y + Y)^2}{K} - \frac{(L_x + X)^2 + (L_y + Y)^2}{K} - \frac{(T_x + X)^2 + (T_y + y)^2}{K} + \frac{2*(G + X + Y)^2}{K^2} \quad [3]$$

Igualando-se a 0 as derivadas parciais em relação a X e Y, e simplificando-se as expressões, obtém-se o seguinte sistema de equações:

$$(K - 1)(K - 2)X + 2Y = K(C_x + L_x + T_x) - 2G \quad [4]$$

$$2X + (K - 1)(K - 2)Y = K(C_y + L_y + T_y) - 2G \quad [5]$$

Resolvendo esse sistema, tem-se

$$X = \frac{(K - 1)(K - 2)(C_x + L_x + T_x) - 2K(C_y + L_y + T_y) + [4 - 2(K - 1)(K - 2)]G}{(K - 1)^2(K - 2)^2 - 4} \quad [6]$$

e

$$Y = \frac{(K - 1)(K - 2)(C_y + L_y + T_y) - 2K(C_x + L_x + T_x) + [4 - 2(K - 1)(K - 2)]G}{(K - 1)^2(K - 2)^2 - 4} \quad [7]$$

Procedimento igual permitirá obter as equações para os demais casos, com a devida substituição de variáveis. Nas fórmulas 3 a 7, C, L, e T representam os totais disponíveis nas linhas, colunas e tratamentos com parcela perdida, identificada pelo índice respectivo, e as demais variáveis têm o significado antes adotado. A análise de variância se procederá como no caso de uma parcela perdida, considerando que agora dois são os graus de liberdade a serem diminuídos do total e do erro.

Linha, coluna ou tratamento perdido

Estes casos, que não são incomuns, foram estudados por Yates (1936a) e também por Sanches (1978), e são relativamente fáceis de ser resolvidos, o que não ocorre quando se perdem duas ou mais linhas, colunas ou tratamentos, casos estudados por Yates e Hale (1939).

Quando se perde um tratamento, a análise de variância deve ser efetuada com o ajustamento da soma de quadrados de colunas ou de linhas, que é efetuado em dois passos:

a) Calcular os valores $Q_i = (K - 1)C_i - W_i$, em que C_i é o total da coluna (linha) i e W_i a soma das linhas (colunas) que não continham o tratamento perdido na coluna (linha) i.

b) A soma de quadrados ajustada para coluna ou linha é igual à soma dos quadrados de Q_i dividido por $K(K - 1)(K - 2)$.

As demais somas de quadrados são obtidas pelo método usual. As médias de tratamentos serão as médias aritméticas, e as comparações entre elas se darão pelos processos usuais, como se vê no exemplo mostrado na Tabela 5.

Para o ajuste da soma de quadrados de coluna, por exemplo, os valores Q_i devem ser calculados da seguinte forma:

$Q1 = 7*8.341,25 - (64.225,00 - 9.021,25) = 3.185,00$

$Q2 = 7*10.268,75 - (64.225,00 - 6.836,25) = 14.492,00$

$Q3 = 7*6.656,25 - (64.225,00 - 9.137,50) = -8.493,75$

$Q4 = 7*6.122,50 - (64.225,00 - 7.630,00) = -13.737,50$

$Q5 = 7*6.855,00 - (64.225,00 - 8.387,50) = -7.852,50$

$Q6 = 7*8.557,50 - (64.225,00 - 7.827,50) = 3.505,00$

$Q7 = 7*8.283,75 - (64.225,00 - 7.957,50) = 1.718,75$

$Q8 = 7*9.140,00 - (64.225,00 - 7.427,50) = 7.182,50$

E a soma de quadrados de coluna, ajustada para a perda de um tratamento é:

$SQC = (3.185,00^2 + ... + 7.182,502)/(8*7*6) = 1.814.073,41$

As demais somas de quadrados, total, linha, tratamento e erro se calculam da forma usual, e a análise de variância encontra-se na Tabela 6.

As médias de tratamentos poderão ser comparadas como sempre, lembrando que agora apenas sete tratamentos estão presentes, porém todos com oito repetições.

Por sua vez, a perda de uma linha ou de uma coluna faz que a soma de quadrados de tratamentos e suas respectivas médias sejam ajustadas para corrigir o efeito dessas perdas. O procedimento prevê a obtenção usual das somas de quadrados para total, linhas e colunas. Para tratamentos, deve-se primeiro obter os valores Q_k, dados por

$$Q_k = (K - 1)T_k - W_k \qquad [8]$$

Tabela 5. Dados de produção de grãos de arroz, em um experimento conduzido pelo pesquisador Evane Ferreira, ao considerar o tratamento E como perdido.

	Coluna 1	Coluna 2	Coluna 3	Coluna 4	Coluna 5	Coluna 6	Coluna 7	Coluna 8	Total	Total de tratamentos
Linha 1	A 1.303,75	D 1.481,25	G 586,25	C 708,75	F 1.193,75	E	H 1.856,25	B 697,50	7.827,50	A 10.755,00
Linha 2	B 618,75	E	H 822,50	D 846,25	G 1.051,25	F 1.070,00	A 1.126,25	C 1.301,25	6.836,25	B 3.086,25
Linha 3	C 1.176,25	F 1.742,50	A 748,75	E	H 952,50	G 1.470,00	B 377,50	D 1.162,50	7.630,00	C 8.527,50
Linha 4	D 852,50	G 1.680,00	B 317,50	F 1.038,75	A 1.326,25	H 1.250,00	C 962,50	E	7.427,50	D 11.196,25
Linha 5	E	H 1.897,50	C 1.096,25	G 945,00	B 216,25	A 1.767,50	D 1.682,50	F 1.416,25	9.021,25	E
Linha 6	F 1.003,75	A 1.726,25	D 2.148,75	H 855,00	C 487,50	B 237,50	E	G 1.498,75	7.957,50	F 9.685,00
Linha 7	G 1.395,00	B 313,75	E	A 1.421,25	D 1.627,50	C 1.367,50	F 1.283,75	H 1.728,75	9.137,50	G 9.621,25
Linha 8	H 1.991,25	C 1.427,50	F 936,25	B 307,50	E	D 1.395,00	G 995,00	A 1.335,00	8.387,50	H 11.353,75
Total	8.341,25	10.268,75	6.656,25	6.122,50	6.855,00	8.557,50	8.283,75	9.140,00	64.225,00	

em que T é o total do tratamento k e W_k é a soma de todas as linhas (colunas) onde aparece o tratamento k, considerando a perda de uma coluna (linha). A soma de quadrados de tratamentos ajustada é obtida da mesma forma que para o caso anterior, ou seja, pelo quociente entre a soma de quadrados dos Q_k e $K(K - 1)(K - 2)$. A média ajustada de cada tratamento é a soma entre o quociente de $(Q_k)/[K(K - 2)]$ e a média geral do experimento.

Tabela 6. Análise de variância do ensaio em quadrado latino 8 x 8, com um tratamento perdido.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Coluna	7	1.814.073,41	259.153,34	2,36 (0,0438)
Linha	7	618.745,54	88.392,22	0,80 (0,5889)
Tratamento	6	6.160.021,88	1.026.670,31	9,35 (< ,0001)
Erro	35	3.843.443,55	109.812,67	
Total	55	12.436.284,38		

Com o mesmo experimento com o qual se está trabalhando, suponha-se agora que a coluna 8, por qualquer motivo, haja sido perdida. O quadro de dados é agora o que está na Tabela 7.

Deve-se agora calcular os valores de Q_k, necessários para o cálculo da soma de quadrados de tratamentos ajustada para a perda de uma coluna:

$Q1 = 7*9.420,00 - (62.522,50 - 8.255,00) = 11.672,50$

$Q2 = 7*2.388,75 - (62.522,50 - 7.955,00) = -37.846,25$

$Q3 = 7*7.226,25 - (62.522,50 - 6.786,25) = -5.152,50$

$Q4 = 7*10.033,75 - (62.522,50 - 7.183,75) = 14.897,50$

$Q5 = 7*7.437,50 - (62.522,50 - 7.427,50) = -3.032,50$

$Q6 = 7*8.268,75 - (62.522,50 - 8.305,00) = 3.663,75$

$Q7 = 7*8.122,50 - (62.522,50 - 7.657,50) = 1.992,50$

$Q8 = 7*9.625,00 - (62.522,50 - 8.952,50) = 13.805,00$

Portanto, a soma de quadrados de tratamentos ajustada para a perda de uma coluna (8) é

$SQTr = (11.672,50^2 + ... + 13.805,00^2)/(8*7*6) = 6.054.276,36$

Tabela 7. Experimento de cultivares de arroz, em quadrado latino 8 x 8, com a coluna 8 eliminada.

	Coluna 1	Coluna 2	Coluna 3	Coluna 4	Coluna 5	Coluna 6	Coluna 7	Total de linhas	Total de tratamentos
Linha 1	A 1.303,75	D 1.481,25	G 586,25	C 708,75	F 1.193,75	E 825,00	H 1.856,25	7.955,00	A 9.420,00
Linha 2	B 618,75	E 1.251,25	H 822,50	D 846,25	G 1.051,25	F 1.070,00	A 1.126,25	6.786,25	B 2.388,75
Linha 3	C 1.176,25	F 1.742,50	A 748,75	E 716,25	H 952,50	G 1.470,00	B 377,50	7.183,75	C 7.226,25
Linha 4	D 852,50	G 1.680,00	B 317,50	F 1.038,75	A 1.326,25	H 1.250,00	C 962,50	7.427,50	D 10.033,75
Linha 5	E 700,00	H 1.897,50	C 1.096,25	G 945,00	B 216,25	A 1.767,50	D 1.682,50	8.305,00	E 7.437,50
Linha 6	F 1.003,75	A 1.726,25	D 2.148,75	H 855,00	C 487,50	B 237,50	E 1.198,75	7.657,50	F 8.268,75
Linha 7	G 1.395,00	B 313,75	E 1.543,75	A 1.421,25	D 1.627,50	C 1.367,50	F 1.283,75	8.952,50	G 8.122,50
Linha 8	H 1.991,3	C 1.427,50	F 936,25	B 307,50	E 1.202,50	D 1.395,00	G 995,00	8.255,00	H 9.625,00
Total	9.041,25	11.520,00	8.200,00	6.838,75	8.057,50	9.382,50	9.482,50	62.522,50	

com as demais somas de quadrados calculadas da forma usual, de forma que o quadro de análise de variância encontra-se na Tabela 8.

Tabela 8. Análise de variância do quadrado latino 8 x 8, com uma coluna perdida.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (probabilidade)
Coluna	6	1.612.243,69	268.707,28	2,29 (0,0569)
Linha	7	482.710,16	68.958,59	0,59 (0,7607)
Tratamento	7	6.054.276,37	864.896,62	7,38 (< ,0001)
Erro	35	4.101.860,74	117.196,02	
Total	55	12.251.090,96		

As médias de tratamentos ajustadas para a perda de uma coluna são:

CNA 8540 = 11.672,50/(8*6) + 1.116,47 = 1.359,65

CNA 8711 = -37.846,25/(8*6) + 1.116,47 = 328,01

CNA 8172 = -5.152,50/(8*6) + 1.116,47 = 1.009,13

CNA 8436 = 14.897,50/(8*6) + 1.116,47 = 1.426,83

Primavera = -3.032,50/(8*6) + 1.116,47 = 1.053,29

Caiapó = 3.663,75/(8*6) + 1.116,47 = 1.192,60

Maravilha = 1.992,50/(8*6) + 1.116,47 = 1.157,98

Canastra = 13.805,00/(8*6) + 1.116,47 = 1.404,07

Uso do SAS com perda total de tratamento, coluna ou linha

Para os casos de tratamento perdido ou linha ou coluna perdida, as Figuras 7 e 8 ilustram os resultados obtidos com o SAS.

No caso do tratamento perdido (Figura 7), o procedimento adotado foi o ANOVA, cujos resultados são idênticos àqueles obtidos anteriormente com o uso das fórmulas de ajustes. A programação também é a mesma usada no caso de presença de todos os dados, ou seja:

```
PROC GLM;CLASS LINHA COLUNA TRATAMENTO;
MODEL PRODUCAO=LINHA COLUNA TRATAMENTO;
LSMEANS TRATAMENTO;
```

O resultado se vê na Figura 7.

The ANOVA Procedure
Class Level Information

Class	Levels	Values
LINHA	8	1 2 3 4 5 6 7 8
COLUNA	8	1 2 3 4 5 6 7 8
TRATAMENTO	7	A B C D F G H

Number of observations 56

The ANOVA Procedure

Dependent Variable: PRODUCAO

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	8720191.07	436009.55	4.11	0.0001
Error	35	3716093.30	106174.09		
Corrected Total	55	12436284.38			

R-Square	Coeff Var	Root MSE	PRODUCAO Mean
0.701189	28.41144	325.8437	1146.875

Source	DF	Anova SS	Mean Square	F Value	Pr > F
LINHA	7	618745.536	88392.219	0.83	0.5677
COLUNA	7	1941423.661	277346.237	2.61	0.0280
TRATAMENTO	6	6160021.875	1026670.313	9.67	<.0001

Figura 7. Uso do SAS para um ensaio em quadrado latino com um tratamento perdido.

The GLM Procedure
Class Level Information

Class	Levels	Values
LINHA	8	1 2 3 4 5 6 7 8
COLUNA	7	1 2 3 4 5 6 7
TRATAMENTO	8	A B C D E F G H

Number of observations 56

The GLM Procedure

Dependent Variable: PRODUCAO

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	8149230.22	407461.51	3.48	0.0006
Error	35	4101860.74	117196.02		
Corrected Total	55	12251090.96			

R-Square	Coeff Var	Root MSE	PRODUCAO Mean
0.665184	30.66254	342.3390	1116.473

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LINHA	7	482710.156	68958.594	0.59	0.7607
COLUNA	6	1612243.694	268707.282	2.29	0.0569
TRATAMENTO	7	6054276.367	864896.624	7.38	<.0001

The GLM Procedure
Least Squares Means

TRATAMENTO	PRODUCAO LSMEAN
A	1359.65030
B	328.00967
C	1009.12946
D	1426.83780
E	1053.29613
F	1192.80134
G	1157.98363
H	1404.07738

Figura 8. Uso do SAS para a análise de variância de um quadrado latino com uma coluna perdida.

Mais de uma observação por parcela

Assim como nos delineamentos anteriores, é possível, muitas vezes, a obtenção de mais de uma observação por parcela ou unidade experimental. É fortemente recomendado que o número de observações por parcela seja absolutamente idêntico, pois qualquer diferença entre eles terá fortes implicações na análise de variância com a impossibilidade de se obter uma forma exata e simples para os testes F. O uso de mais de uma observação por parcela implica mudança no modelo matemático e consequente alteração do esquema de análise de variância, pela inclusão em ambos de um erro amostral e por considerar-se o erro anteriormente calculado como o erro experimental. O quadro da análise de variância e as fórmulas de cálculo são dados na Tabela 9, em que L é o número de observações por parcela e $Y_{ijk.}$ representa a soma dessas observações em cada parcela. Os demais símbolos têm o mesmo significado visto anteriormente.

A esperança matemática dos componentes da análise se dá na Tabela 10. Como se pode facilmente observar, o erro experimental é testado em relação ao erro amostral e as fontes coluna, linha e tratamento, com relação ao erro experimental. Também aqui, ao se verificar a não significância do erro experimental, é possível a aglutinação dos dois erros

Tabela 9. Análise de variância de um quadrado latino com l amostra por parcela.

Fonte de variação	Graus de liberdade (GL)	Soma de quadrados (SQ)
Coluna (C)	$K - 1$	$\frac{1}{LK}\sum_i C_i^2 - \frac{1}{LK^2}\left(\sum_{ijkl} Y_{ijkl}\right)^2$
Linha (L)	$K - 1$	$\frac{1}{LK}\sum_j L_j^2 - \frac{1}{LK^2}\left(\sum_{ijkl} Y_{ijkl}\right)^2$
Tratamento (T)	$K - 1$	$\frac{1}{LK}\sum_k T_k^2 - \frac{1}{LK^2}\left(\sum_{ijkl} Y_{ijkl}\right)^2$
Erro experimental (EE)	$(K - 1)(K - 2)$	$\frac{1}{L}\sum_{ijk} Y_{ijk.}^2 - \frac{1}{LK^2}\left(\sum_{ijkl} Y_{ijkl}\right)^2 - SQC - SQL - SQT$
Erro amostral (EA)	$(L - 1)K^2$	$SQTo - SQC - SQL - SQT - SQEe$
Total (To)	$K^2 - 1$	$\sum_{ijkl} Y_{ijkl}^2 - \frac{1}{LK^2}\left(\sum_{ijkl} Y_{ijkl}\right)^2$

Tabela 10. Esperança matemática dos quadrados médios na análise de variância de um quadrado latino com *l* amostra por parcela. Modelos aleatório, fixo e misto.

Fonte de variação	Esperança dos quadrados médios		
	Modelo aleatório	Modelo misto	Modelo fixo
Coluna	$\sigma^2 + L\sigma_e^2 + LK\sigma_c^2$	$\sigma^2 + L\sigma_e^2 + LK\sigma_c^2$	$\sigma^2 + L\sigma_e^2 + W_c^2$
Linha	$\sigma^2 + L\sigma_e^2 + LK\sigma_l^2$	$\sigma^2 + L\sigma_e^2 + LK\sigma_l^2$	$\sigma^2 + L\sigma_e^2 + W_l^2$
Tratamentos	$\sigma^2 + L\sigma_e^2 + LK\sigma_t^2$	$\sigma^2 + L\sigma_e^2 + W_t^2$	$\sigma^2 + L\sigma_e^2 + W_t^2$
Erro experimental	$\sigma^2 + L\sigma_e^2$	$\sigma^2 + L\sigma_e^2$	$\sigma^2 + L\sigma_e^2$
Erro amostral	σ^2	σ^2	σ^2

em um só erro, que terá como graus de liberdade a soma dos respectivos graus de liberdade e, como soma de quadrado, a soma das somas de quadrado dos erros experimental e amostral e com o qual se fará a aplicação do teste F para as demais fontes.

Exemplo de aplicação

Em 1985, o pesquisador Evane Ferreira conduziu um experimento com 11 tratamentos, em quadrado latino, em que foram tomadas quatro amostras em cada uma das parcelas, no que diz respeito ao número total de hastes e ao número de hastes mortas por cupim (*Synthermes* sp.) e lagarta-elasmo (*Elasmopalpus* sp.). Com base nesses números, calculou-se a proporção de hastes sobreviventes ao ataque de insetos, cujos resultados se mostram na Tabela 11. Os tratamentos testados foram:

Número do tratamento	Identificação do tratamento
1	IAC 47 - 50 sementes m^{-1}
2	1 + aldrin (sulco)
3	2 + carbofuran e monocrotofós
4	2 + carbofuran
5	2 + monocrotofós
6	IAC 47 - 30 sementes m^{-1}
7	6 + aldrin (sulco) + carbofuran + monocrotofós
8	IAC 165 - 70 sementes m^{-1}
9	8 + aldrin (sulco) + carbofuran + monocrotofós
10	IAC 165 + 40 sementes m^{-1}
11	10 + aldrin (sulco) + carbofuran + monocrotofós

Tabela 11. Dados referentes à proporção de hastes sobreviventes ao ataque de insetos. Experimento conduzido pelo pesquisador Evane Ferreira, em quadrado latino com quatro amostras por parcela.

Linhas	Colunas																							
	1		2		3		4		5		6		7		8		9		10		11			
1	2	0,9000	11	0,9130	10	0,8730	9	0,9394	5	0,8929	4	0,9149	6	0,8571	8	0,9783	7	0,8000	3	0,8387	1	0,7500
		0,8800		0,9024		1,0000		1,0000		0,7111		0,8649		0,8929		0,9524		0,7097		0,9063		0,7059
		0,9032		0,8333		0,7188		0,8983		0,8361		0,9365		0,2500		0,8158		0,9388		0,8571		0,6500
		0,8750		0,7778		0,8846		0,8966		0,8000		0,9667		0,6667		0,5714		0,7941		0,6071		0,3529
2	6	0,7368	3	0,9524	11	0,9821	5	0,9744	8	0,9375	2	0,9362	10	0,9556	9	0,9620	1	0,9219	4	0,9531	7	0,9310
		0,6757		0,9863		1,0000		0,8889		0,7105		0,9383		0,8571		0,9677		0,6364		0,9054		0,9130
		0,6400		0,9875		0,9702		0,9434		0,9180		0,8955		0,8421		0,9546		0,9091		0,9623		1,0000
		0,7895		0,9259		0,9167		0,8226		0,9298		0,8919		0,9057		1,0000		0,7344		0,9114		0,8491
3	3	0,9184	8	0,8750	1	0,8182	6	0,7273	9	0,9000	10	0,9250	11	0,9483	2	0,8929	5	0,8269	7	1,0000	4	0,9825
		0,9400		0,7593		0,8621		0,7381		0,9778		0,9394		0,9756		0,9778		0,8621		0,9630		0,9216
		0,9804		0,9444		0,8542		0,7759		0,9467		0,8462		0,9787		0,9677		0,8254		0,9556		0,9702
		0,9828		0,9767		0,9804		0,9245		1,0000		0,9706		0,9184		0,9589		0,8781		0,9524		0,9273
4	7	1,0000	9	1,0000	2	1,0000	3	0,9867	6	0,9556	11	0,9841	5	0,9390	4	1,0000	8	0,8824	1	0,9846	10	0,9057
		1,0000		1,0000		0,9756		1,0000		0,8824		0,9821		0,8986		0,9211		0,8793		0,9296		0,8718
		0,8788		0,9375		1,0000		0,9811		0,6154		0,9737		0,9750		0,9298		0,7647		0,8548		0,9231
		0,8750		0,7917		1,0000		0,9722		0,9778		0,9231		0,9200		0,9891		0,9048		0,9841		0,8889
5	4	0,8621	1	0,8525	6	0,8793	11	0,9524	10	0,9811	8	0,8971	2	0,8677	7	0,8571	3	1,0000	5	0,9039	9	0,9500
		0,9111		0,8889		0,9348		0,9487		0,8611		1,0000		0,9643		0,9492		0,9608		0,9787		0,9706
		0,8472		0,9057		0,8537		1,0000		0,9615		0,9138		0,9474		0,9583		0,9091		0,7755		0,9500
		0,8085		0,9231		0,9630		0,9592		1,0000		0,9286		0,9302		0,9846		0,9718		0,9737		0,9697
6	11	0,8800	7	1,0000	5	0,8298	4	0,9667	2	0,8421	1	0,8364	9	0,9048	6	0,8571	10	0,8781	8	0,9375	3	0,9857
		0,8824		1,0000		0,8853		0,9857		0,8281		0,8871		0,9444		0,7955		0,8182		1,0000		0,8853
		0,9118		0,9677		0,9149		0,9868		0,9403		0,8431		0,9859		0,9688		0,8611		0,8571		0,9286
		0,8409		0,9672		0,8571		0,9683		0,9136		0,9778		0,9355		0,9070		0,9460		0,9661		0,9575

Continua...

Tabela 11. Continuação.

Linhas	Colunas																							
	1		2		3		4		5		6		7		8		9		10		11			
7	10	0,7742	6	0,8525	9	0,9400	1	0,8293	11	0,9123	7	0,9189	8	0,9811	3	0,9867	4	0,9702	2	0,8478	5	0,8864
		0,6400		0,8333		0,8636		0,9697		0,9167		0,8667		0,9592		0,7353		0,9032		0,8462		0,9737
		0,8182		0,8235		0,9420		1,0000		0,9057		0,7872		0,9623		0,7857		0,9333		0,8966		0,9583
		0,7037		0,8000		0,9259		0,9565		0,9091		0,9531		0,9362		0,9677		0,7600		0,9200		0,9375
8	5	0,7750	4	0,9273	8	0,9167	2	0,9846	7	0,9583	3	0,9605	1	0,9643	10	0,9630	11	0,9130	9	0,9583	6	0,8571
		0,9091		0,9524		0,8393		0,9265		0,9508		0,9516		0,9130		0,8889		1,0000		0,9460		0,7391
		0,8333		0,9492		0,9722		0,9200		0,9348		0,8833		0,7931		1,0000		0,9744		1,0000		1,0000
		0,8571		0,9851		0,9460		0,9167		0,9167		0,9667		0,9348		0,9268		0,8857		0,9365		0,8462
9	1	0,9091	5	0,8000	7	1,0000	8	0,8421	4	0,8824	9	1,0000	3	0,9677	11	0,9722	6	0,7586	10	0,9231	2	0,9048
		1,0000		0,9556		0,9556		0,8846		0,8824		0,9730		1,0000		0,9032		0,8261		1,0000		0,9722
		0,9535		0,9333		0,9298		0,8947		0,9808		1,0000		0,9697		0,9815		0,8511		0,9706		0,8254
		0,9546		0,8710		0,8833		0,9474		0,9016		0,9825		0,8163		0,9020		0,9655		0,7200		0,9016
10	8	0,9474	2	0,8974	4	0,8615	10	0,7813	3	0,8983	5	0,9048	7	0,8793	1	0,7931	9	0,8889	6	0,8000	11	0,9130
		0,9524		0,8636		0,9107		0,8421		0,8500		0,9091		0,7692		0,8594		0,8409		0,8750		0,9231
		0,9787		0,8889		0,9756		0,9143		0,8667		0,7959		0,6818		0,6250		0,9250		0,8824		0,7308
		1,0000		0,9524		0,7761		0,8214		0,6613		0,8936		0,8298		0,9722		0,7674		0,8261		0,8929
11	9	0,9200	10	0,7647	3	1,0000	7	0,9318	1	0,9615	6	0,7059	4	0,9344	5	0,6889	2	0,8000	11	1,0000	8	1,0000
		1,0000		0,7500		0,8983		0,9841		0,7600		0,7222		1,0000		0,8000		1,0000		0,8387		0,9615
		0,7931		0,9333		0,9091		0,9388		0,9434		0,9575		0,9818		0,7679		0,8372		0,8085		0,9184
		1,0000		0,4375		0,9091		0,8039		0,8387		0,7167		0,8636		0,7174		0,7636		1,0000		0,8810

As somas de quadrados para linha, coluna e tratamentos são obtidas da forma usual tomando-se apenas o cuidado de lembrar que cada valor de total dessas linhas, colunas e tratamentos foi obtido de quatro medidas por parcela, o que quer dizer que, no presente caso, cada total é a soma de 44 valores. A soma de quadrados total não sofre qualquer modificação em sua maneira de cálculo, bem como o erro amostral (por diferença), enquanto para a obtenção do erro experimental devem ser somados os quatro valores de cada parcela. É necessário então montar um quadro auxiliar com esses totais, o qual se encontra na Tabela 12.

As somas de quadrados são então calculadas:

$$FC = \frac{(433,7341)^2}{484} = 388,6886$$

$$SQLin = \frac{36,0137^2 + 39,4220^2 + ... + 38,1425^2}{44} - FC = 0,5513$$

$$SQCol = \frac{38,6390^2 + 39,2393^2 + ... + 39,1634^2}{44} - FC = 0,0985$$

$$SQTrat = \frac{37,9744^2 + 40,0922^2 + ... + 40,6377^2}{44} - FC = 0,6045$$

$$SQtotal = 0,9000^2 + 0,9130^2 + ... + 0,8810^2 - FC = 4,2910$$

$$SQEe = \frac{3,5582^2 + 3,4265^2 + ... + 3,7609^2}{4} - FC - SQLin - SQCol - SQTrat = 0,9092$$

O erro amostral é calculado por diferença, restando todas as fontes de variação calculadas até agora da soma de quadrados total, ou seja,

$$SQEa = SQtotal - SQCol - SQLin - SQTrat - SQEe = 2,1275$$

Com esses dados, pode-se então montar o quadro de análise de variância (Tabela 13).

Como o erro experimental foi significativo ($F = 1,72$), os efeitos de tratamentos devem ser verificados utilizando esse erro como o denominador na obtenção do teste F para tratamentos, que, nesse caso, foi igual a 5,98, muitíssimo significativo, ou seja, significativo a um nível inferior a 1 por 10.000.

Tabela 12. Quadro auxiliar para análise de variância de um quadrado latino, com quatro amostras por parcela, mostrando os totais das parcelas, linhas, colunas e tratamentos.

	1	2	3	4	5	6	7	8	9	10	11	Total	Total de tratamentos
1	2	11	10	9	5	4	6	8	7	3	1		1
	3,5582	3,4265	3,4764	3,7343	3,2401	3,6830	2,6667	3,3179	3,2426	3,2092	2,4588	36,0137	37,9744
2	6	3	11	5	8	2	10	9	1	4	7		2
	2,8420	3,8521	3,8690	3,6293	3,4958	3,6619	3,5605	3,8843	3,2018	3,7322	3,6931	39,4220	40,0922
3	3	8	1	6	9	10	11	2	5	7	4		3
	3,8216	3,5554	3,5149	3,1658	3,8245	3,6812	3,8210	3,7973	3,3925	3,8710	3,8016	40,2468	40,4057
4	7	9	2	3	6	11	5	4	8	1	10		4
	3,7538	3,7292	3,9756	3,9400	3,4312	3,8630	3,7326	3,8400	3,4312	3,7531	3,5895	41,0392	40,7248
5	4	1	6	11	10	8	2	7	3	5	9		5
	3,4289	3,5702	3,6308	3,8603	3,8037	3,7395	3,7096	3,7492	3,8417	3,6318	3,8403	40,8060	38,2813
6	11	7	5	4	2	1	9	6	10	8	3		6
	3,5151	3,9349	3,4871	3,9075	3,5241	3,5444	3,7706	3,5284	3,5034	3,7607	3,7571	40,2333	35,9037
7	10	6	9	1	11	7	8	3	4	2	5		7
	2,9361	3,3093	3,6715	3,7555	3,6438	3,5259	3,8388	3,4754	3,5667	3,5106	3,7559	38,9895	40,1185
8	5	4	8	2	7	3	1	10	11	9	6		8
	3,3745	3,8140	3,6742	3,7478	3,7606	3,7621	3,6052	3,7787	3,7731	3,8408	3,4424	40,5734	40,0217
9	1	5	7	8	4	9	3	11	6	10	2		9
	3,8172	3,5599	3,7687	3,5688	3,6472	3,9555	3,7537	3,7589	3,4013	3,6137	3,604	40,4489	41,3863
10	8	2	4	10	3	5	7	1	9	6	11		10
	3,8785	3,6023	3,5239	3,3591	3,2763	3,5034	3,1601	3,2497	3,4222	3,3835	3,4598	37,8188	38,1878
11	9	10	3	7	1	6	4	5	2	11	8		11
	3,7131	2,8855	3,7165	3,6586	3,5036	3,1023	3,7798	2,9742	3,4008	3,6472	3,7609	38,1425	40,6377
Total	38,6390	39,2393	40,3086	40,3270	39,1509	40,0222	39,3986	39,3540	38,1773	39,9538	39,1634	433,7341	

Tabela 13. Análise de variância do quadrado latino com amostras por parcelas.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Colunas	10	0,0985	0,0098	
Linhas	10	0,5513	0,0551	
Tratamentos	10	0,6045	0,0604	5,98 (< 0,0001)
Erro experimental (EE)	90	0,9092	0,0101	1,72 (0,0003)
Erro amostral (EA)	363	2,1275	0,0059	
Total	483	4,2910		

Com esse resultado, verifica-se então que existe uma diferença acentuada entre os tratamentos na reação ao ataque de insetos. Para verificar quais tratamentos apresentam número maior de sobrevivência, deve-se aplicar um teste de médias. Como exemplo, aplica-se o teste de Scott-Knott, utilizando-se os procedimentos descritos no Capítulo 1. O resultado está representado na Tabela 14.

Tabela 14. Comparações de médias, utilizando o teste de Scott-Knott com 5% de probabilidade no nível de significância.

Identificação do tratamento	Média de sobrevivência	Grupo
[8] + aldrin (sulco) + carbofuran + monocrotofós	0,9406	A
[2] + carbofuran	0,9256	A
[10] + aldrin (sulco) + carbofuran + monocrotofós	0,9236	A
[2] + carbofuran e monocrotofós	0,9138	A
[6] + aldrin (sulco) + carbofuran + monocrotofós	0,9118	A
[1] + aldrin (sulco) [2]	0,9112	A
IAC 165 - 70 sementes m^{-1} [8]	0,9056	A
[2] + monocrotofós	0,8700	B
IAC 165 + 40 sementes m^{-1} [10]	0,8679	B
IAC 47 - 50 sementes m^{-1} [1]	0,8630	B
IAC 47 - 30 sementes m^{-1} [6]	0,8160	C

Exemplo recalculado com o uso do SAS

Para este caso, o procedimento ANOVA é suficiente por se tratar de conjunto de dados balanceados, lembrando que existe a necessidade de especificar o erro experimental. Além disso, os efeitos de tratamentos devem ser, em princípio, testados por esse erro.

Caso o erro experimental não seja significativo, uma segunda análise deve ser realizada sem a especificação do erro experimental. Como afirmado tantas vezes antes, o teste do erro experimental é feito contra o erro amostral, que o SAS identifica apenas como erro (ERROR). O erro experimental, nesse caso, é especificado como a interação entre os três fatores – linha, coluna e tratamento.

Assim o programa é estabelecido:

```
PROC ANOVA;CLASS LINHA COLUNA TRATAMENTO;
MODEL PHASTV=LINHA COLUNA TRATAMENTO LINHA*COLUNA*TRATAMENTO;
TEST H=TRATAMENTO E=LINHA*COLUNA*TRATAMENTO;
```

O resultado de tal processo se vê na Figura 9.

```
                          The ANOVA Procedure
                        Class Level Information
                Class           Levels    Values
                LINHA               11    1 2 3 4 5 6 7 8 9 10 11
                COLUNA              11    1 2 3 4 5 6 7 8 9 10 11
                TRATAMENTO          11    1 2 3 4 5 6 7 8 9 10 11
                          Number of observations    484
                          The ANOVA Procedure
Dependent Variable: PHASTV
                                     Sum of
 Source                      DF      Squares    Mean Square   F Value   Pr >
F
 Model                      120   2.16346884     0.01802891      3.08
<.0001
 Error                      363   2.12754106     0.00586099
 Corrected Total            483   4.29100990

              R-Square     Coeff Var      Root MSE    PHASTV Mean
              0.504186      8.542983      0.076557       0.896140

 Source                 DF       Anova SS    Mean Square   F Value   Pr > F
 LINHA                  10     0.55128846     0.05512885      9.41   <.0001
 COLUNA                 10     0.10429024     0.01042902      1.78   0.0628
 TRATAMENTO             10     0.60447757     0.06044776     10.31   <.0001
 LINHA*COLUNA*TRATAME   90     0.90341257     0.01003792      1.71   0.0003

                  Tests of Hypotheses Using the Anova MS for
                    LINHA*COLUNA*TRATAME as an Error Term
 Source                 DF       Anova SS    Mean Square   F Value   Pr > F
 TRATAMENTO             10     0.60447757     0.06044776      6.02   <.0001
```

Figura 9. Análise de variância de um experimento em quadrado latino, com amostras nas parcelas e uso do SAS.

Teste de aditividade

Para o delineamento em quadrado latino, o procedimento é dado a seguir.

Obter primeiro o valor e o erro esperados de cada unidade experimental, sendo o valor esperado dado por

$$\hat{Y}_{ijk} = \bar{Y}_{.j.} + \bar{Y}_{..k} + \bar{Y}_{i..} - 2 * \bar{Y} \quad [9]$$

ou seja, o valor esperado de cada unidade experimental é igual à soma das médias da linha, da coluna e do tratamento correspondentes menos duas vezes a média do experimento. O erro esperado é a diferença entre o valor real e o valor esperado, ou seja,

$$\hat{e}_{ijk} = Y_{ijk} - \hat{Y}_{ijk} \quad [10]$$

Segundo, após realizar uma nova análise de variância, empregando agora o quadrado do valor esperado como o valor de cada parcela, calcular a soma de quadrado de não aditividade por meio de

$$SQ(NAd) = \frac{(\sum_{ijk} \hat{Y}_{ijk} * \hat{e}_{ijk})^2}{SQ} \quad [11]$$

em que SQ é a soma de quadrado do erro obtido na análise anterior. O teste de F para não aditividade é igual a

$$F = \frac{SQ(NAd)}{QM\,\mathrm{Re}\,s} \quad [12]$$

em que

$$QM\,\mathrm{Re}\,s = \frac{SQErro - SQ(NAd)}{(K - 1)(K - 2) - 1} \quad [13]$$

e o $SQErro$ corresponde à soma de quadrado do erro da análise efetuada com os dados originais. O $QMRes$ tem agora $(K - 1)(K - 2) - 1$ graus de liberdade.

Exemplo de aplicação

Os dados da Tabela 15 referem-se ao número de perfilhos de arroz (Y_{ijh}), em um experimento conduzido pelo pesquisador Evane Ferreira, em Santo Antônio de Goiás, em

1998. Os valores de y_{ijk} são estimados segundo as fórmulas dadas anteriormente, e e_{ijk} são as diferenças entre os valores medidos e os estimados, enquanto C_h, L_j e T_i são os efeitos, respectivamente, de colunas, linhas e tratamentos.

Tabela 15. Dados de um experimento em quadrado latino com as estimativas de efeitos para o teste de aditividade.

Linhas		Colunas						Média	L_j
		1	2	3	4	5	6		
1	Tratamento	1	5	6	3	2	4	177,5	-13,7
	Y_{ijk}	155	191	203	187	162	167		
	y_{ijk}	155,4	194,4	180,6	190,9	170,6	173,2		
	e_{ijk}	-0,4	-3,4	22,4	-3,9	-8,6	-6,2		
2	Tratamento	5	3	4	1	6	2	190,5	-0,7
	Y_{ijk}	198	199	205	186	181	174		
	y_{ijk}	197,7	214,4	188,4	167,6	193,7	181,2		
	e_{ijk}	0,3	-15,4	16,6	18,4	-12,7	-7,2		
3	Tratamento	4	2	3	6	5	1	186,8	-4,4
	Y_{ijk}	187	213	195	177	188	161		
	y_{ijk}	194,2	198,9	191,6	198,6	184,7	153,1		
	e_{ijk}	-7,2	14,1	3,4	-21,6	3,3	7,9		
4	Tratamento	2	6	1	4	3	5	189,8	-1,4
	Y_{ijk}	204	225	135	190	204	181		
	y_{ijk}	192,2	212,1	158,2	196,4	194,7	185,4		
	e_{ijk}	11,8	12,9	-23,2	-6,4	9,3	-4,4		
5	Tratamento	6	4	5	2	1	3	202,8	11,6
	Y_{ijk}	209	218	196	209	175	210		
	y_{ijk}	215,4	219,9	200,6	204,4	171,4	205,4		
	e_{ijk}	-6,4	-1,9	-4,6	4,6	3,6	4,6		
6	Tratamento	3	1	2	5	4	6	199,8	8,6
	Y_{ijk}	216	181	178	215	203	206		
	y_{ijk}	214,1	187,4	192,7	206,2	197,9	200,7		
	e_{ijk}	1,9	-6,4	-14,7	8,8	5,1	5,3		
Média		194,8	204,5	185,3	194,0	185,5	183,2	191,2	
C_k		3,6	13,3	-5,9	2,8	-5,7	-8,1		

Tratamento	1	2	3	4	5	6
Média	165,5	190,0	201,8	195,0	194,8	200,2
T_i	-25,7	-1,2	10,6	3,8	3,6	8,9

A análise de variância executada com os valores estimados y_{ijk} conduz a uma estimativa de soma de quadrados de erro igual a 2.327.112,52, e a soma de quadrados do produto y_{ijk}*e_{ijk} dá um total igual a 100.739.342. Com esses valores, obtém-se a soma de quadrado para não aditividade igual a 43,29; e quociente entre 100.739.342 e 2.327.112,52. A análise de variância dos dados originais (Y_{ijh}) produz uma soma de quadrados de erro igual a 4.011,56. Desse valor, deve ser subtraído aquele correspondente à não aditividade. Portanto, tem-se que o novo quadrado médio residual com 19 GL é igual a

$$QM\,\mathrm{Re}\,s = \frac{4.011,56 - 43,29}{(6 - 1)(6 - 2) - 1} = 208,86 \qquad [14]$$

e o valor do teste F é igual a 0,207, resultado do quociente 43,29/208,86; portanto, não significativo.

Planos básicos de alguns quadrados latinos selecionados:

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4 (a)

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

4 x 4 (b)

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

4 x 4 (c)

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

4 x 4 (d)

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

5 x 5

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

6 x 6

A	B	C	D	E	F
B	F	D	C	A	E
C	D	E	F	B	A
D	A	F	E	C	B
E	C	A	B	F	D
F	E	B	A	D	C

7 x 7

A	B	C	D	E	F	G
B	C	D	E	F	G	A
C	D	E	F	G	A	B
D	E	F	G	A	B	C
E	F	G	A	B	C	D
F	G	A	B	C	D	E
G	A	B	C	D	E	F

8 x 8

A	B	C	D	E	F	G	H
B	C	D	E	F	G	H	A
C	D	E	F	G	H	A	B
D	E	F	G	H	A	B	C
E	F	G	H	A	B	C	D
F	G	H	A	B	C	D	E
G	H	A	B	C	D	E	F
H	A	B	C	D	E	F	G

9 x 9

A	B	C	D	E	F	G	H	I
B	C	D	E	F	G	H	I	A
C	D	E	F	G	H	I	A	B
D	E	F	G	H	I	A	B	C
E	F	G	H	I	A	B	C	D
F	G	H	I	A	B	C	D	E
G	H	I	A	B	C	D	E	F
H	I	A	B	C	D	E	F	G
I	A	B	C	D	E	F	G	H

10 x 10

A	B	C	D	E	F	G	H	I	J
B	C	D	E	F	G	H	I	J	A
C	D	E	F	G	H	I	J	A	B
D	E	F	G	H	I	J	A	B	C
E	F	G	H	I	J	A	B	C	D
F	G	H	I	J	A	B	C	D	E
G	H	I	J	A	B	C	D	E	F
H	I	J	A	B	C	D	E	F	G
I	J	A	B	C	D	E	F	G	H
J	A	B	C	D	E	F	G	H	I

11 x 11

A	B	C	D	E	F	G	H	I	J	K
B	C	D	E	F	G	H	I	J	K	A
C	D	E	F	G	H	I	J	K	A	B
D	E	F	G	H	I	J	K	A	B	C
E	F	G	H	I	J	K	A	B	C	D
F	G	H	I	J	K	A	B	C	D	E
G	H	I	J	K	A	B	C	D	E	F
H	I	J	K	A	B	C	D	E	F	G
I	J	K	A	B	C	D	E	F	G	H
J	K	A	B	C	D	E	F	G	H	I
K	A	B	C	D	E	F	G	H	I	J

12 x 12

A	B	C	D	E	F	G	H	I	J	K	L
B	C	D	E	F	G	H	I	J	K	L	A
C	D	E	F	G	H	I	J	K	L	A	B
D	E	F	G	H	I	J	K	L	A	B	C
E	F	G	H	I	J	K	L	A	B	C	D
F	G	H	I	J	K	L	A	B	C	D	E
G	H	I	J	K	L	A	B	C	D	E	F
H	I	J	K	L	A	B	C	D	E	F	G
I	J	K	L	A	B	C	D	E	F	G	H
J	K	L	A	B	C	D	E	F	G	H	I
K	L	A	B	C	D	E	F	G	H	I	J
L	A	B	C	D	E	F	G	H	I	J	K

Capítulo 5

Delineamentos em blocos incompletos

Existem muitas situações experimentais nas quais ocorre a impossibilidade de se obter um ambiente totalmente homogêneo, a fim de permitir o uso do delineamento completamente casualizado, ou mesmo restringir esse ambiente homogêneo ao tamanho adequado para a aplicação de todos os tratamentos, que é o caso dos delineamentos em blocos completos e em quadrado latino. Essa impossibilidade pode tanto ser devida ao ambiente quanto ao número de tratamentos, como, por exemplo: os ensaios de novas linhagens em melhoramento de plantas, em geral muito acima de 100 materiais distintos; os experimentos com animais, como no caso de alimentação de porcos em que o número de leitões por leitegada raramente ultrapassa dez animais; os testes de pneus em que cada veículo, em geral, só pode conter quatro unidades e os testes de palatabilidade ou de sabor, nos quais os provadores têm capacidade limitada de provas que possam executar, como nos testes de qualidade de bebidas (café, vinho, sucos, etc.). Em todas essas situações a solução a ser adotada é o uso de delineamentos em blocos incompletos, no qual cada bloco contém apenas uma parte dos tratamentos. Esses delineamentos foram introduzidos por Yates (1936b).

De acordo com o arranjo dos tratamentos nos blocos e a sua ocorrência em pares nesses blocos, os delineamentos em blocos incompletos podem ser classificados em balanceados (BIB) e parcialmente balanceados (BIPB). Alguns deles, segundo o número de tratamentos e sua disposição nos blocos, recebem denominações especiais, como reticulados quadrados, reticulados retangulares, reticulados cúbicos, quadrados reticulados e quadrados de Youden. Ressalte-se que os quadrados de Youden são sempre balanceados e os reticulados tanto podem ser balanceados quanto parcialmente balanceados. Além disso, os quadrados reticulados e os quadrados de Youden apresentam controle ambiental em duplo sentido.

Cochran e Cox (1983) classificam os BIB e BIPB em cinco tipos:

- Tipo I – os blocos podem ser agrupados em repetições.
- Tipo II – os blocos podem ser agrupados formando grupos de repetições.

- Tipo III – os blocos não podem ser agrupados formando repetições ou grupos de repetições.
- Tipo IV – o número de tratamentos é igual ao número de blocos que possuem um número de unidades experimentais menor que o número de tratamentos.
- Tipo V – experimentos pequenos.

Os experimentos de tipo IV e V são raríssimas vezes empregados; portanto, não serão detalhados aqui. Aos neles interessados, sugere-se a leitura da literatura referenciada.

Neste capítulo, em que se estudam os BIB e os BIPB de forma geral, e no Capítulo 6, no qual se estudam os reticulados, será adotado que t é o número de tratamentos, b o número de blocos, r o número de repetições, k o número de parcelas por bloco e λ o número de vezes em que cada par de tratamentos aparece junto em um bloco.

Um delineamento em blocos incompletos é dito balanceado quando todos os pares de tratamento aparecem juntos em um bloco o mesmo número de vezes, ou seja, λ é único. Quando λ apresenta valores distintos para distintos pares de tratamento, diz-se que o delineamento é parcialmente balanceado.

O valor de λ é dado pela expressão

$$\lambda = r(k - 1)/(t - 1) \qquad [1]$$

em que λ, r, t, k e b são inteiros e $rt = bk$.

Por exemplo, para $t = 9$ e $k = 5$, com $r = 10$ tem-se $\lambda = 5$ e $b = 18$. O valor de r (10) é o valor mínimo que se pode ter para um delineamento balanceado. As soluções de r menores que 10 levam a valores fracionados de λ se r for ímpar, e fracionados de b se r for par. A solução $r = 5$, com nove blocos, leva a um desenho parcialmente balanceado, com valores de λ iguais a 2 e 3.

Blocos incompletos balanceados (BIB)

Estes desenhos experimentais apresentam uma análise de variância relativamente mais difícil que aquelas discutidas até agora, além de uma certa redução nos graus de liberdade do erro. Esses aparentes problemas são compensados fortemente por uma redução na estimativa do erro experimental devida principalmente ao maior controle ambiental que introduzem pela utilização de blocos de tamanho reduzido.

Cochran e Cox (1983) e outros autores apresentam esquemas experimentais em BIB e outros tipos de blocos incompletos para os mais diversos números de tratamentos,

blocos, repetições e tamanho de blocos, dos quais um conjunto deles se apresenta ao final deste capítulo. A partir dessas definições, é possível determinar a eficiência relativa desses delineamentos quando comparados com seus equivalentes em blocos completos. Essa eficiência (E), dada a seguir, apresenta valor no intervalo $0 < E < 1$; portanto, os melhores esquemas serão os que apresentam valor de E próximo a 1.

$$E = \frac{\lambda t}{kr} \qquad [2]$$

Para os BIBs, dois tipos de análise de variância são possíveis: a análise intrablocos e a análise com recuperação da informação interblocos. No primeiro caso, os efeitos de tratamentos são ajustados para os efeitos de blocos em que se alocam, sendo sempre independentes. Na análise interblocos, realiza-se um ajuste nos efeitos de blocos com respeito aos tratamentos que contêm, e isso também é considerado no ajuste de médias dos tratamentos, que agora deixam de ser independentes; portanto, qualquer teste de médias será apenas aproximado (PIMENTEL-GOMES, 2000).

Casualização dos blocos e dos tratamentos

A aplicação dos tratamentos às parcelas e a ordem de instalação ou execução dos blocos são feitas de forma aleatória, posterior à definição da distribuição básica do desenho.

Suponha-se o esquema 5.7 com $t = 7, k = 3, b = 7, r = 3$ e $\lambda = 2$, tipo III, cuja distribuição básica é mostrada a seguir:

Bloco 1			Bloco 2			Bloco 3			Bloco 4			Bloco 5			Bloco 6			Bloco 7		
1	2	4	2	3	5	3	4	6	4	5	7	1	5	6	2	6	7	1	3	7

Após o sorteio dos blocos e dos tratamentos dentro de cada bloco, poder-se-ia ter a seguinte distribuição:

Bloco 4			Bloco 6			Bloco 1			Bloco 7			Bloco 2			Bloco 5			Bloco 3		
5	7	4	6	7	2	2	4	1	3	1	7	3	5	2	6	5	1	4	3	6

Essa casualização pode ser levada mais adiante na forma em que aos tratamentos se designem, de forma aleatória, os números de ordem que os identificarão. O exemplo

anterior mostra que a aleatorização de um experimento em blocos incompletos se faz em quatro etapas:

a) Escolha do desenho básico adequado às condições experimentais existentes e nível de eficiência do esquema.
b) Casualização da ordem de implantação dos blocos.
c) Casualização dos tratamentos em cada um dos blocos.
d) Designação aleatória aos tratamentos de seu número de ordem.

Modelos matemáticos e análise de variância intrablocos

Consideram-se aqui os três principais tipos de BIB: o tipo I, blocos formando repetições; o tipo II, blocos formando grupos de repetição; e o tipo III, em que é impossível a união de blocos formando repetição ou grupos de repetição. Apesar da sua aparente diversidade, os três tipos podem ser analisados como se fossem do tipo III, que pode ser considerado um modelo genérico para o delineamento em blocos incompletos.

Tipo I

O modelo matemático correspondente a este desenho experimental é

$$Y_{ijl(j)} = m + R_j + B_{l(j)} + T_i + e_{ijl(j)} \quad [3]$$

em que m é a média, R_j o efeito da repetição j ($j = 1, 2, ..., r$), $B_{l(j)}$ o efeito do bloco l ($l = 1, 2, ..., b$) dentro da repetição j, T_i o efeito do tratamento i ($i = 1, 2, ..., t$) e $e_{ijl(j)}$ o erro associado à observação $Y_{ijl(j)}$, com distribuição normal, de média 0 e variância σ^2. Os erros se supõem independentes e de variância homogênea.

Na Tabela 1, que representa este modelo, observa-se que as somas de quadrados para total, repetição e blocos dentro de repetições são calculadas da maneira usual, enquanto a soma de quadrados para tratamentos é ajustada para o efeito de blocos. Na fórmula dada, os valores Q_i são obtidos pela expressão $kT_i - A_i$, em que T_i é o total do tratamento i, e A_i é a soma de totais de cada bloco onde aparece o tratamento i ($Q_i = kT_i - A_i$).

Os valores ajustados das médias de tratamentos são calculados segundo a expressão

$$\bar{t}_i = m + \frac{Q_i}{\lambda t} \quad [4]$$

em que m representa a média geral do experimento e $Q_i/\lambda t$ é o efeito de tratamento ajustado para efeito de blocos.

Tabela 1. Análise de variância de um experimento em blocos incompletos, tipo I. Efeitos fixos.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Repetição (R)	r - 1	$\frac{1}{t}\sum R_j^2 - \frac{1}{rt}\left(\sum_{ijl(j)} Y_{ijl(j)}\right)^2$	*SQR*/*GLR*
Blocos dentro de repetição (B d. R)	$r(b$ - 1)	$\frac{1}{k}\sum B_{l(j)}^2 - \frac{1}{rt}\left(\sum_{ijl(j)} Y_{ijl(j)}\right)^2 - SQR$	*SQBd.R*/*GLBd.R*
Tratamentos ajustados (T)	t - 1	$\frac{\sum_i Q_i^2}{\lambda kt}$	*SQTaj.*/*GLTaj.*
Erro (E)	Diferença	Diferença	*SQE*/*GLE*
Total	rt - 1	$\sum Y_{ijl(j)}^2 - \frac{1}{rt}\left(\sum_{ijl(j)} Y_{ijl(j)}\right)^2$	

Para a comparação de médias, considerando o balanço do delineamento, pode ser adotado qualquer um dos testes apresentados anteriormente, segundo interesse ou necessidade do pesquisador. Nesse caso, a variância do contraste entre médias é expressa por

$$Var\,(Contraste) = \frac{k * QME * \sum_i c_i^2}{\lambda t} \qquad [5]$$

ou igual à variância de um contraste (se em blocos completos) multiplicado pelo inverso da eficiência (E) do delineamento. Portanto, isso implica que, quanto menos eficiente for o delineamento, maior será a variância dos contrastes entre médias.

Tipo II

No caso dos blocos incompletos de tipo II o modelo matemático e dado por

$$Y_{ijl(j)} = m + G_j + B_{l(j)} + T_i + e_{ijl(j)} \qquad [6]$$

em que G_j representa o efeito de grupo de repetição j (j = 1, 2, ..., J), e $B_{l(j)}$ o efeito de blocos l (l = 1, 2, ..., b) dentro do grupo de repetição j e os demais elementos têm o mesmo

significado anterior. O numero de repetições dentro de cada grupo de repetições é constante e é representado pela letra g.

A análise de variância apresenta, portanto, essa única modificação, como se vê na Tabela 2, permanecendo constantes os demais cálculos.

Para a comparação de médias, vale o mesmo registro feito para delineamentos do tipo I.

Tabela 2. Análise de variância de um experimento em blocos incompletos, tipo II. Efeitos fixos.

Fonte de variação	Graus de liberdade (GL)	Soma de quadrados (SQ)	Quadrados médios (QM)
Grupo de repetição (GR)	$J - 1$	$\frac{1}{gt}\sum G_j^2 - \frac{1}{rt}\left(\sum_{ijl(j)} Y_{ijl(j)}\right)^2$	$SQGR/GLGR$
Blocos dentro de repetição (B d. R)	$J(b - 1)$	$\frac{1}{k}\sum B_{l(j)}^2 - \frac{1}{rt}\left(\sum_{ijl(j)} Y_{ijl(j)}\right)^2 - SQGR$	$SQBd.R/GLBd.R$
Tratamentos ajustados (T)	$t - 1$	$\frac{\sum_i Q_i^2}{\lambda kt}$	SQT/GLT
Erro (E)	Diferença	Diferença	SQE/GLE
Total	$rt - 1$	$\sum Y_{ijl(j)}^2 - \frac{1}{rt}\left(\sum_{ijl(j)} Y_{ijl(j)}\right)^2$	

Tipo III

O modelo matemático correspondente a delineamentos do tipo III é idêntico ao de blocos completos, isto é,

$$Y_{ij} = m + B_j + T_i + e_{ij} \qquad [7]$$

com as mesmas interpretações anteriores, lembrando apenas que, nesse caso, cada um dos blocos contém apenas parte dos tratamentos. A Tabela 3 reflete esse modelo.

Para a comparação de médias, adota-se o mesmo procedimento usado para os tipos 1 e 2.

Tabela 3. Análise de variância de um experimento em blocos incompletos, tipo III. Efeitos fixos.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Blocos não ajustados (B)	b - 1	$\frac{1}{k}\sum B_j^2 - \frac{1}{rt}\left(\sum_{ij} Y_{ij}\right)^2$	*SQB*/*GLB*
Tratamentos ajustados (T)	t - 1	$\frac{\sum_i Q_i^2}{\lambda kt}$	*SQTaj.*/*GLTaj.*
Erro (E)	Diferença	Diferença	*SQE*/*GLE*
Total	rt - 1	$\sum Y_{ij}^2 - \frac{1}{rt}\left(\sum_{ij} Y_{ij}\right)^2$	

Análise com recuperação da informação interblocos

A forma de análise discutida até agora estima os efeitos de tratamentos tendo por base apenas sua própria informação, contida nos blocos onde aparece. Yates (1940), tendo como base o fato da não ortogonalidade dos delineamentos BIB, por não conterem os blocos todos os tratamentos, propôs uma nova forma de analisar esse tipo de experimento, de modo que os efeitos de tratamentos contidos nos blocos pudessem ser recuperados. Isso se chama análise com recuperação da informação interblocos. Essa recuperação de informação implica assumir que os blocos agora são aleatórios, com média 0 e variância igual a σ_b^2, e que existem modificações nos procedimentos estudados anteriormente. A primeira modificação é o ajuste na soma de quadrados de blocos para o efeito de tratamentos que contêm, enquanto as outras dizem respeito a novos ajustes nas médias de tratamentos e na soma de quadrado de tratamentos para contemplar essa recuperação de informação.

Segundo Kuehl (2001), se os blocos foram eficientes em reduzir o erro experimental, a recuperação interblocos contribui muito pouco na melhoria das estimações combinadas, e essa contribuição pode ser altamente interessante se a redução do erro for pequena.

A soma de quadrados de blocos ajustada é facilmente obtida por meio da diferença entre a soma das somas de quadrados de blocos não ajustada e a soma de tratamentos ajustada e a soma de quadrado de tratamentos não ajustada, isto é,

$$SQBaj = SQBnaj + SQTaj - SQTnaj \quad [8]$$

em que as somas de quadrados para blocos e tratamentos, não ajustadas, são obtidas pelo processo usual, e a soma de quadrados de tratamentos ajustada é aquela obtida na análise intrablocos.

A informação interblocos é definida pela variável R dada por

$$R = \frac{\sigma^2}{\sigma^2 + \sigma_b^2} \quad [9]$$

Como os valores das variâncias na fórmula 9 não são conhecidos, devem-se usar estimativas de seus valores, que serão baseadas nos quadrados médios do erro (QME) e de blocos ajustados ($QMBaj$), lembrando que, para o tipo I, teremos blocos dentro de repetições; e para o tipo II, blocos dentro de grupos de repetições. Essas estimativas serão obtidas a partir das relações a seguir, para os tipos I, II e III, respectivamente.

$$R = \frac{(r - 1)QME}{rQMBaj - QME} \quad [10]$$

$$R = \frac{[t(r - 1) - k(g - 1)]QME}{k(b - g)QMBaj - (t - k)QME} \quad [11]$$

$$R = \frac{QME}{QME - \frac{k(b - 1)}{bk - t}(QME - QMBaj)} \quad [12]$$

A partir dessas estimativas, devem ser obtidas, para os três casos, as estimativas

$$M_i = Q_i + RA_i - \frac{RkG}{t} \quad [13]$$

e os efeitos de tratamentos ajustados com recuperação da informação interblocos

$$t_i^+ = \frac{M_i}{\lambda t + (r - \lambda)R} \quad [14]$$

A nova soma de quadrados de tratamentos ajustada para essa recuperação de informação é dada por

$$SQT_{aj}^{+} = \frac{\sum M_i^2}{k[\lambda t + (r - \lambda)R]} \quad [15]$$

Para as fórmulas 13 a 15, a simbologia é aquela definida anteriormente, com a inclusão de G, que representa a soma de todos os dados disponíveis no ensaio.

Deve-se sempre ter em mente que, quando se trabalha com estimativas, é possível a obtenção de um valor negativo para σ_b^2. Nesse caso, deve-se assumir que seu valor será igual a 0 (zero), logo $R = 1$. Portanto, não haveria necessidade de análise para recuperação da informação interblocos por não haver efetividade dos blocos. Nesse caso um BIB de tipo I poderia ser reduzido a um bloco completo, e um BIB tipo III seria, na realidade, considerado um delineamento completamente casualizado.

Para a comparação de médias, qualquer um dos testes antes descrito pode ser empregado, sabendo-se que a variância de um contraste entre médias é dada pela expressão

$$V(Cont) = \frac{k * QME}{\lambda t + (r - \lambda)R} \sum_i C_i^2 \quad [16]$$

em que C_i são os coeficientes do contraste em estudo.

Blocos incompletos parcialmente balanceados (BIPB)

Bose e Nair (1939) introduziram esta classe de delineamentos que podem ocorrer pelos mais variados motivos, como número exageradamente grande de repetições para permitir o balanço, insuficiência de material experimental e até insuficiências de ordem financeira ou humana. Deve ser lembrado que o número mínimo de repetições para desenhos balanceados é $r = \lambda(t - 1)/(k - 1)$ com as mesmas restrições anteriores. Por exemplo, se $t = 11$ e $k = 4$, λ deveria ser igual a 6 ou múltiplo de 6 para um valor inteiro de r com todos os blocos de tamanho 4. Com $\lambda = 6$, r é igual a 20, o que implica 220 parcelas distribuídas em 55 blocos. Isso seria, na maioria das vezes, impraticável. Este delineamento tem eficiência (E) igual a 0,825.

Três propriedades básicas devem ser consideradas:

1) Todo bloco contem k parcelas, todo tratamento aparece r vezes e nenhum deles aparece mais de uma vez em um dado bloco.

2) Cada par de tratamentos ocorre junto em λ_1 e λ_2 blocos. Os que aparecem λ_1 vezes são chamados de primeiro associados, e os que aparecem λ_2 vezes chamam-se segundo associados.

3) Se dois tratamentos são i-ésimos associados, o número de tratamentos j-ésimos associados do primeiro deles, e o número de k-ésimos associados do segundo, é sempre igual, não importando com que par se comece. Cochran e Cox (1983) chamam esse número de $p_{jk}^{\ i}$.

Um desenho é dito BIPB com duas classes associadas desde que satisfaça a essas três propriedades descritas. A última propriedade, apesar de ser aparentemente a mais complicada, é a que torna factível a análise de variância. Além disso, ela define que apenas dois tipos de variâncias existem para os testes de médias. Ademais, essas variâncias são tão aproximadas que muito comumente pode ser utilizado um valor médio. Bem mais complicado é o caso de contrastes entre mais de duas médias. Se o número de médias envolvidas no contraste for 3, por exemplo, todas podem ter aparecido juntas em um bloco ou não. Se não, é possível que os três pares tenham o mesmo valor de λ ou tenham valores diferentes, e a cada caso haverá uma variância diferente. Com mais de três médias, as possíveis alternativas são em muito maior número.

Alguns esquemas básicos para blocos incompletos

Seguem-se alguns planos de delineamentos em blocos incompletos. Para cada um deles, tem-se: t = número de tratamentos, k = número de parcelas por bloco, r = número de repetições, b = número de blocos, λ = número de vezes que cada par de tratamentos aparece junto em um bloco e E = eficiência.

Esquema 5.1 – $t = 4, k = 2, r = 3, b = 6, \lambda = 1, E = 0{,}67$, Tipo V.

Bloco	Rep. 1		Rep. 2		Rep. 3	
(1)	1	2	(3)	1	3	(5)	1	4
(2)	3	4	(4)	2	4	(6)	2	3

Esquema 5.2 – $t = 5, k = 2, r = 4, b = 10, \lambda = 1, E = 0{,}62$, tipo V.

Bloco	Reps. 1 e 2		Bloco	Reps. 3 e 4	
(1)	1	2	(6)	1	4
(2)	3	4	(7)	2	3
(3)	2	5	(8)	3	5
(4)	1	3	(9)	1	5
(5)	4	5	(10)	2	4

Esquema 5.3 – $t = 6, k = 2, r = 5, b = 15, \lambda = 1, E = 0{,}60$, tipo I.

Bloco	Rep. 1		Bloco	Rep. 2		Bloco	Rep. 3		Bloco	Rep. 4		Bloco	Rep. 5	
(1)	1	2	(4)	1	3	(7)	1	4	(10)	1	5	(13)	1	6
(2)	3	4	(5)	2	5	(8)	2	6	(11)	2	4	(14)	2	3
(3)	5	6	(6)	4	6	(9)	3	5	(12)	3	6	(15)	4	5

Esquema 5.4 – $t = 6, k = 3, r = 5, b = 10, \lambda = 2, E = 0{,}80$, tipo III.

Bloco				Bloco			
(1)	1	2	5	(6)	2	3	4
(2)	1	2	6	(7)	2	3	5
(3)	1	3	4	(8)	2	4	6
(4)	1	3	6	(9)	3	5	6
(5)	1	4	5	(10)	4	5	6

Esquema 5.5 – $t = 6, k = 3, r = 10, b = 20, \lambda = 4, E = 0{,}80$, tipo I.

Bloco	Rep. 1			Rep. 2			Rep. 3			Rep. 4		
(1)	1	2	3	(3)	1	2	4	(5)	1	2	5	(7)	1	2	6
(2)	4	5	6	(4)	3	5	6	(6)	3	4	6	(8)	3	4	5

	Rep. 5			Rep. 6			Rep. 7			Rep. 8		
(9)	1	3	4	(11)	1	3	5	(13)	1	3	6	(15)	1	4	5
(10)	2	5	6	(12)	2	4	6	(14)	2	4	5	(16)	2	3	6

	Rep. 9			Rep. 10		
(17)	1	4	6	(19)	1	5	61
(18)	2	3	5	(20)	2	3	4

Esquema 5.6 – $t = 6, k = 4, r = 10, b = 15, \lambda = 6, E = 0{,}90$, tipo II.

Bloco	Reps. 1 e 2			
(1)	1	2	3	4
(2)	1	4	5	6
(3)	2	3	5	6

Bloco	Reps. 3 e 4			
(4)	1	2	3	5
(5)	1	2	4	6
(6)	3	4	5	6

Bloco	Reps. 5 e 6			
(7)	1	2	3	6
(8)	1	3	4	5
(9)	2	4	5	6

Bloco	Reps. 7 e 8			
(10)	1	2	4	5
(11)	1	3	5	6
(12)	2	3	4	6

Bloco	Reps. 9 e 10			
(13)	1	2	5	6
(14)	1	3	4	6
(15)	2	3	4	5

Esquema 5.7 – $t = 7, k = 3, r = 3, b = 7, \lambda = 1, E = 0{,}78$, tipo V.

Bloco			
(1)	1	2	4
(2)	2	3	5

(3)	3	4	6
(4)	4	5	7

(5)	5	6	1
(6)	6	7	2

(7)	7	1	3

Esquema 5.8 – $t = 7, k = 4, r = 4, b = 7, \lambda = 2, E = 0{,}88$, tipo V.

Bloco				
(1)	3	5	6	7
(2)	1	4	6	7
(3)	1	2	5	7

(4)	1	2	3	6
(5)	2	3	4	7
(6)	1	3	4	5

(7)	2	4	5	6

Esquema 5.9 – $t = 8, k = 2, r = 7, b = 28, \lambda = 1, E = 0{,}57$, tipo I.

Bloco	Rep. 1	
(1)	1	2
(2)	3	4
(3)	5	6
(4)	7	8

	Rep. 2	
(5)	1	3
(6)	2	8
(7)	4	5
(8)	6	7

	Rep. 3	
(9)	1	4
(10)	2	7
(11)	3	6
(12)	5	8

	Rep. 4	
(13)	1	5
(14)	2	3
(15)	4	7
(16)	6	8

	Rep. 5	
(17)	1	6
(18)	2	4
(19)	3	8
(20)	5	7

	Rep. 6	
(21)	1	7
(22)	2	6
(23)	3	5
(24)	4	8

	Rep. 7	
(25)	1	8
(26)	2	5
(27)	3	7
(28)	4	6

Esquema 5.10 – $t = 8$, $k = 4$, $r = 7$, $b = 14$, $\lambda = 3$, $E = 0{,}86$, tipo I.

Bloco	Rep. 1				Rep. 2				Rep. 3				Rep. 4			
(1)	1	2	3	4	(3)	1	2	7	8	(5)	1	3	6	8	(7)	1	4	6	7
(2)	5	6	7	8	(4)	3	4	5	6	(6)	2	4	5	7	(8)	2	3	5	8

	Rep. 5				Rep. 6				Rep. 7			
(9)	1	2	5	6	(11)	1	3	5	7	(13)	1	4	5	8
(10)	3	4	7	8	(12)	2	4	6	8	(14)	2	3	6	7

Esquema 5.11 – $t = 9$, $k = 4$, $r = 8$, $b = 18$, $\lambda = 3$, $E = 0{,}84$, tipo III.

Bloco																
(1)	1	2	3	4	(7)	1	4	8	9	(13)	2	5	6	8
(2)	1	2	5	6	(8)	1	5	7	9	(14)	3	5	8	9
(3)	1	2	7	8	(9)	2	3	8	9	(15)	4	6	7	9
(4)	1	3	5	7	(10)	2	4	5	9	(16)	3	4	5	6
(5)	1	4	6	8	(11)	2	6	7	9	(17)	3	6	7	8
(6)	1	3	6	9	(12)	2	3	4	7	(18)	4	5	7	8

Esquema 5.12 – $t = 9$, $k = 5$, $r = 10$, $b = 18$, $\lambda = 5$, $E = 0{,}90$, tipo III.

Bloco																
(1)	1	3	6	7	8	(7)	1	3	4	7	9	(13)	1	3	4	5	8
(2)	2	3	4	6	8	(8)	1	2	3	0	9	(14)	1	2	4	6	7
(3)	2	4	5	7	8	(9)	1	2	3	5	8	(15)	1	4	5	6	7
(4)	5	6	7	8	9	(10)	1	2	4	5	9	(16)	2	3	5	7	9
(5)	3	4	5	6	9	(11)	3	4	7	8	9	(17)	1	2	7	8	9
(6)	2	4	6	8	9	(12)	2	3	5	6	7	(18)	1	5	6	8	9

Esquema 5.13 – $t = 9, k = 6, r = 8, b = 12, \lambda = 5, E = 0{,}94$, tipo II.

Bloco	Reps. 1 e 2					
(1)	1	2	4	5	7	8
(2)	2	3	5	6	8	9
(3)	1	3	4	6	7	9

	Reps. 3 e 4					
(4)	1	2	5	6	7	9
(5)	1	3	4	5	8	9
(6)	2	3	4	6	7	8

	Reps. 5 e 6					
(7)	1	3	5	6	7	8
(8)	1	2	4	6	8	9
(9)	2	3	4	5	7	9

	Reps. 7 e 8					
(10)	4	5	6	7	8	9
(11)	1	2	3	4	5	6
(12)	1	2	3	7	8	9

Esquema 5.14 – $t = 10, k = 2, r = 9, b = 45, \lambda = 1, E = 0{,}56$, tipo I.

Bloco	Rep. 1	
(1)	1	2
(2)	3	4
(3)	5	6
(4)	7	8
(5)	9	10

	Rep. 2	
(6)	1	3
(7)	2	7
(8)	4	8
(9)	5	9
(10)	6	10

	Rep. 3	
(11)	1	4
(12)	2	10
(13)	3	7
(14)	5	8
(15)	6	9

	Rep. 4	
(16)	1	5
(17)	2	8
(18)	3	10
(19)	4	9
(20)	6	7

	Rep. 5	
(21)	1	6
(22)	2	9
(23)	3	8
(24)	4	10
(25)	5	7

	Rep. 6	
(26)	1	7
(27)	2	6
(28)	3	9
(29)	4	5
(30)	8	10

	Rep. 7	
(31)	1	8
(32)	2	3
(33)	4	6
(34)	5	10
(35)	7	9

	Rep. 8	
(36)	1	9
(37)	2	4
(38)	3	5
(39)	6	8
(40)	7	10

	Rep. 9	
(41)	1	10
(42)	2	5
(43)	3	6
(44)	4	7
(45)	8	9

Esquema 5.15 – $t = 10, k = 3, r = 9, b = 30, \lambda = 2, E = 0{,}74$, tipo III.

Bloco			
(1)	1	2	3
(2)	1	2	4
(3)	1	3	5
(4)	1	4	6
(5)	1	5	7
(6)	1	6	8

(7)	1	7	9
(8)	1	8	10
(9)	1	9	10
(10)	2	3	6
(11)	2	4	10
(12)	2	5	8

(13)	2	5	9
(14)	2	6	7
(15)	2	7	9
(16)	2	8	10
(17)	3	4	7
(18)	3	4	8

(19)	3	5	6
(20)	3	7	10
(21)	3	8	9
(22)	3	9	10
(23)	4	5	9
(24)	4	5	10

(25)	4	6	9
(26)	4	7	8
(27)	5	6	10
(28)	5	7	8
(29)	6	7	10
(30)	6	8	9

Esquema 5.16 – $t = 10, k = 4, r = 6, b = 15, \lambda = 2, E = 0{,}83$, tipo III.

Bloco				
(1)	1	2	3	4
(2)	1	2	5	6
(3)	1	3	7	8
(4)	1	4	9	10
(5)	1	5	7	9

(6)	1	6	8	10
(7)	2	3	6	9
(8)	2	4	7	10
(9)	2	5	8	10
(10)	2	7	8	9

(11)	3	5	9	10
(12)	3	6	7	10
(13)	3	4	5	8
(14)	4	5	6	7
(15)	4	6	8	9

Esquema 5.17 – $t = 10, k = 5, r = 9, b = 18, \lambda = 4, E = 0{,}89$, tipo III.

Bloco					
(1)	1	2	3	4	5
(2)	1	2	3	6	7
(3)	1	2	4	6	9
(4)	1	2	5	7	8
(5)	1	3	6	8	9
(6)	1	3	7	8	10

(7)	1	4	5	6	10
(8)	1	4	8	9	10
(9)	1	5	7	9	10
(10)	2	3	4	8	10
(11)	2	3	5	9	10
(12)	2	4	7	8	9

(13)	2	5	6	8	10
(14)	2	6	7	9	10
(15)	3	4	6	7	10
(16)	3	4	5	7	9
(17)	3	5	6	8	9
(18)	4	5	6	7	8

Esquema 5.18 – $t = 10, k = 6, r = 9, b = 15, \lambda = 5, E = 0{,}93$, tipo III.

Bloco						
(1)	1	2	4	5	8	9
(2)	5	6	7	8	9	10
(3)	2	4	5	6	9	10
(4)	1	2	4	6	7	8
(5)	3	4	7	8	9	10

(6)	2	3	4	6	8	10
(7)	1	2	6	7	9	10
(8)	1	3	5	6	8	9
(9)	1	2	3	8	9	10
(10)	2	3	4	5	7	9

(11)	1	4	5	7	8	10
(12)	1	2	3	5	7	10
(13)	2	3	5	6	7	8
(14)	1	3	4	5	6	10
(15)	1	3	4	6	7	9

Esquema 5.19 – $t = 11, k = 5, r = 5, b = 11, \lambda = 2, E = 0{,}88$, tipo IV.

Bloco					
(1)	1	2	3	5	8
(2)	2	3	4	6	9
(3)	3	4	5	7	10
(4)	4	5	6	8	11
(5)	5	6	7	9	1
(6)	6	7	8	10	2

(7)	7	8	9	11	3
(8)	8	9	10	1	4
(9)	9	10	11	2	5
(10)	10	11	1	3	6
(11)	11	1	2	4	7

Esquema 5.20 – $t = 11, k = 6, r = 6, b = 11, \lambda = 3, E = 0{,}92$, tipo IV.

Bloco						
(1)	4	6	7	9	10	11
(2)	1	5	7	8	10	11
(3)	1	2	6	8	9	11
(4)	1	2	3	7	9	10
(5)	2	3	4	8	10	11
(6)	1	3	4	5	9	11

(7)	1	2	4	5	6	10
(8)	2	3	5	6	7	11
(9)	1	3	4	6	7	8
(10)	2	4	5	7	8	9
(11)	3	5	6	8	9	10

Esquema 5.21 – $t = 13, k = 3, r = 6, b = 26, \lambda = 1, E = 0{,}72$, tipo III.

Bloco			
(1)	1	2	3
(2)	1	4	5
(3)	1	6	7
(4)	1	8	9
(5)	1	10	11
(6)	1	12	13
(7)	2	4	9
(8)	2	5	13
(9)	2	6	11

(10)	2	7	10
(11)	2	8	12
(12)	3	4	8
(13)	3	5	10
(14)	3	6	12
(15)	3	7	9
(16)	3	11	13
(17)	4	6	10
(18)	4	7	13

(19)	4	11	12
(21)	5	6	8
(21)	5	7	12
(22)	5	9	11
(23)	6	9	13
(24)	7	8	11
(25)	8	10	13
(26)	9	10	12

Esquema 5.22 – $t = 13, k = 4, r = 4, b = 13, \lambda = 1, E = 0{,}81$, tipo IV.

Bloco				
(1)	1	2	4	10
(2)	2	3	5	11
(3)	3	4	6	12
(4)	4	5	7	13
(5)	5	6	8	1

(6)	6	7	9	2
(7)	7	8	10	3
(8)	8	9	11	4
(9)	9	10	12	5
(10)	10	11	13	6

(11)	11	12	1	7
(12)	12	13	2	8
(13)	13	1	3	9

Esquema 5.23 – $t = 13, k = 9, r = 9, b = 13, \lambda = 6, E = 0{,}96$, tipo IV.

Bloco									
(1)	3	5	6	7	8	9	11	12	13
(2)	1	4	6	7	8	9	10	12	13
(3)	1	2	5	7	8	9	10	11	13
(4)	1	2	3	6	8	9	10	11	12
(5)	2	3	4	7	9	10	11	12	13
(6)	1	3	4	5	8	10	11	12	13
(7)	1	2	4	5	6	9	11	12	13

(8)	1	2	3	5	6	7	10	12	13
(9)	1	2	3	4	6	7	8	11	13
(10)	1	2	3	4	5	7	8	9	12
(11)	2	3	4	5	6	8	9	10	13
(12)	1	3	4	5	6	7	9	10	11
(13)	2	4	5	6	7	8	10	11	12

Esquema 5.24 – $t = 15, k = 3, r = 7, b = 35, \lambda = 1, E = 0{,}71$, tipo I.

Bloco	Rep. 1			Rep. 2			Rep. 3			Rep. 4		
(1)	1	2	3	(6)	1	4	5	(11)	1	6	7	(16)	1	8	9
(2)	4	8	12	(7)	2	8	10	(12)	2	9	11	(17)	2	13	15
(3)	5	10	15	(8)	3	13	14	(13)	3	12	15	(18)	3	4	7
(4)	6	11	13	(9)	6	9	15	(14)	4	10	14	(19)	5	11	14
(5)	7	9	14	(10)	7	11	12	(15)	5	8	13	(20)	6	10	12

	Rep. 5			Rep. 6			Rep. 7		
(21)	1	10	11	(26)	1	12	13	(31)	1	14	15
(22)	2	12	14	(27)	2	5	7	(32)	2	4	6
(23)	3	5	6	(28)	3	9	10	(33)	3	8	11
(24)	4	9	13	(29)	4	11	15	(34)	5	9	12
(25)	7	8	15	(30)	6	8	14	(35)	7	10	13

Esquema 5.25 – $t = 15, k = 7, r = 7, b = 15, \lambda = 3, E = 0{,}92$, tipo IV.

Bloco							
(1)	1	2	3	4	5	6	7
(2)	1	2	3	8	9	10	11
(3)	1	2	3	12	13	14	15
(4)	1	4	5	8	9	14	15
(5)	1	4	5	10	11	12	13
(6)	1	6	7	8	9	12	13
(7)	1	6	7	10	11	14	15
(8)	2	4	6	8	10	13	15
(9)	2	4	6	9	11	12	14
(10)	2	5	7	8	10	12	14
(11)	2	5	7	9	11	13	15
(12)	3	4	7	8	11	13	14
(13)	3	4	7	9	10	12	15
(14)	3	5	6	8	11	12	15
(15)	3	5	6	9	10	13	14

Esquema 5.26 – $t = 15, k = 8, r = 8, b = 15, \lambda = 4, E = 0{,}94$, tipo IV.

Bloco								
(1)	8	9	10	11	12	13	14	15
(2)	4	5	6	7	12	13	14	15
(3)	4	5	6	7	8	9	10	11
(4)	2	3	6	7	10	11	12	13
(5)	2	3	6	7	8	9	14	15
(6)	2	3	4	5	10	11	14	15
(7)	2	3	4	5	8	9	12	13
(8)	1	3	5	7	9	11	12	14
(9)	1	3	5	7	8	10	13	15
(10)	1	3	4	6	9	11	13	15
(11)	1	3	4	6	8	10	12	14
(12)	1	2	5	6	9	10	12	15
(13)	1	2	5	6	8	11	13	14
(14)	1	2	4	7	9	10	13	14
(15)	1	2	4	7	8	11	12	15

Esquema 5.27 – $t = 16, k = 6, r = 6, b = 16, \lambda = 2, E = 0{,}89$, tipo IV.

Bloco						
(1)	13	14	15	4	8	12
(2)	6	7	8	1	9	13
(3)	9	10	11	4	8	16
(4)	5	6	8	3	11	15
(5)	9	11	12	2	6	14
(6)	13	15	16	2	6	10
(7)	9	10	12	3	7	15
(8)	10	11	12	1	5	13
(9)	5	7	8	2	10	14
(10)	1	2	3	8	12	16
(11)	1	3	4	6	10	14
(12)	13	14	16	3	7	11
(13)	1	2	4	7	11	15
(14)	5	6	7	4	12	16
(15)	2	3	4	5	9	13
(16)	14	15	16	1	5	9

Esquema 5.28 – $t = 16, k = 6, r = 9, b = 24, \lambda = 3, E = 0{,}89$, tipo III.

Bloco						
(1)	1	2	5	6	11	12
(2)	1	2	7	8	13	14
(3)	1	2	9	10	15	16
(4)	1	3	5	7	10	12
(5)	1	3	6	8	13	15
(6)	1	3	9	11	14	16
(7)	1	4	5	8	10	11
(8)	1	4	6	7	13	16
(9)	1	4	9	12	14	15
(10)	2	3	5	8	14	15
(11)	2	3	6	7	9	12
(12)	2	3	10	11	13	16
(13)	2	4	5	7	14	16
(14)	2	4	6	8	9	11
(15)	2	4	10	12	13	15
(16)	3	4	5	6	15	16
(17)	3	4	7	8	9	10
(18)	3	4	11	12	13	14
(19)	5	6	9	10	13	14
(20)	5	7	9	11	13	15
(21)	5	8	9	12	13	16
(22)	6	7	10	11	14	15
(23)	6	8	10	12	14	16
(24)	7	8	11	12	15	16

Esquema 5.29 – $t = 16, k = 10, r = 10, b = 16, \lambda = 6, E = 0{,}96$, tipo III.

Bloco										
(1)	1	2	3	5	6	7	9	10	11	16
(2)	2	3	4	5	10	11	12	14	15	16
(3)	1	2	3	5	6	7	12	13	14	15
(4)	1	2	4	7	9	10	12	13	14	16
(5)	1	3	4	5	7	8	10	13	15	16
(6)	1	3	4	5	7	8	9	11	12	14
(7)	1	2	4	5	6	8	11	13	14	16
(8)	2	3	4	6	7	8	9	14	15	16
(9)	1	3	4	6	9	11	12	13	15	16
(10)	4	5	6	7	9	10	11	13	14	15
(11)	2	5	7	8	9	11	12	13	15	16
(12)	1	2	4	5	6	8	9	10	12	15
(13)	3	5	6	8	9	10	12	13	14	16
(14)	1	2	3	8	9	10	11	13	14	15
(15)	1	6	7	8	10	11	12	14	15	16
(16)	2	3	4	6	7	8	10	11	12	13

Esquema 5.30 – $t = 19$, $k = 3$, $r = 9$, $b = 57$, $\lambda = 1$, $E = 0{,}70$, tipo III.

Bloco			
(1)	1	2	19
(2)	1	3	6
(3)	1	4	18
(4)	1	5	16
(5)	1	7	14
(6)	1	8	13
(7)	1	9	12
(8)	1	10	11
(9)	1	15	17
(10)	2	3	17
(11)	2	4	16
(12)	2	5	15
(13)	2	6	14
(14)	2	7	13
(15)	2	8	12

(16)	2	9	11
(17)	2	10	18
(18)	3	4	15
(19)	3	5	14
(20)	3	7	12
(21)	3	8	11
(22)	3	10	13
(23)	3	9	16
(24)	3	18	19
(25)	4	5	19
(26)	4	6	12
(27)	4	7	11
(28)	4	8	10
(29)	4	9	17

(30)	4	13	14
(31)	5	6	11
(32)	5	7	10
(33)	5	8	9
(34)	5	12	17
(35)	5	13	18
(36)	6	7	8
(37)	6	9	10
(38)	6	13	15
(39)	6	16	19
(40)	6	17	18
(41)	7	9	15
(42)	7	16	18
(43)	7	17	19

(44)	8	14	19
(45)	8	15	18
(46)	8	16	17
(47)	9	13	19
(48)	9	14	18
(49)	10	12	19
(50)	10	14	17
(51)	10	15	16
(52)	11	12	18
(53)	11	13	17
(54)	11	14	16
(55)	11	15	19
(56)	12	13	16
(57)	12	14	15

Esquema 5.31 – $t = 19$, $k = 9$, $r = 9$, $b = 19$, $\lambda = 4$, $E = 0{,}94$, tipo IV.

Bloco									
(1)	1	2	3	4	5	6	7	8	9
(2)	1	2	4	5	10	13	14	17	19
(3)	1	2	7	8	11	12	14	16	17
(4)	1	2	6	9	12	13	15	16	19
(5)	1	3	4	6	11	12	14	18	19
(6)	1	3	7	9	10	13	14	16	18
(7)	1	3	5	8	10	11	15	16	19
(8)	1	4	8	9	11	13	15	17	18
(9)	1	5	6	7	10	12	15	17	18
(10)	2	3	4	7	10	11	12	13	15

(11)	2	3	5	6	11	13	16	17	18
(12)	2	3	8	9	10	12	17	18	19
(13)	2	4	6	8	10	14	15	16	18
(14)	2	5	7	9	11	14	15	18	19
(15)	3	4	5	9	12	14	15	16	17
(16)	3	6	7	8	13	14	15	17	19
(17)	4	5	7	8	12	13	16	18	19
(18)	4	6	7	9	10	11	16	17	19
(19)	5	6	8	9	10	11	12	13	14

Esquema 5.32 – $t = 19, k = 10, r = 10, b = 19, \lambda = 5, E = 0{,}95$, tipo IV.

Bloco										
(1)	10	11	12	13	14	15	16	17	18	19
(2)	3	6	7	8	9	11	12	15	16	18
(3)	3	4	5	6	9	10	13	15	18	19
(4)	3	4	5	7	8	10	11	14	17	18
(5)	2	5	7	8	9	10	13	15	16	17
(6)	2	4	5	G	8	11	12	15	17	19
(7)	2	4	6	7	9	12	13	14	17	18
(8)	2	3	5	6	7	10	12	14	16	19
(9)	2	3	4	8	9	11	13	14	16	19
(10)	1	5	6	8	9	14	16	17	18	19

(11)	1	4	7	8	9	10	12	14	15	19
(12)	1	4	5	6	7	11	13	14	15	16
(13)	1	3	5	7	9	11	12	13	17	19
(14)	1	3	4	6	8	10	12	13	16	17
(15)	1	2	6	7	8	10	11	13	18	19
(16)	1	2	4	5	9	10	11	12	16	18
(17)	1	2	3	6	9	10	11	14	15	17
(18)	1	2	3	5	8	12	13	14	15	18
(19)	1	2	3	4	7	15	16	17	18	19

Esquema 5.33 – $t = 21, k = 3, r = 10, b = 70, \lambda = 1, E = 0{,}70$, tipo I.

Bloco	Rep. 1			Rep. 2			Rep. 3			Rep. 4			Rep. 5		
(1)	1	2	3	(8)	1	4	15	(15)	1	5	17	(22)	1	6	9	(29)	1	7	21
(2)	4	5	6	(9)	2	5	11	(16)	2	4	14	(23)	2	7	16	(30)	2	13	17
(3)	7	8	9	(10)	3	9	16	(17)	3	7	11	(24)	3	8	21	(31)	3	10	18
(4)	10	11	12	(11)	6	17	20	(18)	6	10	19	(25)	4	17	19	(32)	4	8	11
(5)	13	14	15	(12)	7	12	19	(19)	8	16	20	(26)	5	10	13	(33)	5	16	19
(6)	16	17	18	(13)	8	13	18	(20)	9	15	18	(27)	11	15	20	(34)	6	12	15
(7)	19	20	21	(14)	10	14	21	(21)	12	13	21	(28)	12	14	18	(35)	9	14	20
	Rep. 6			Rep. 7			Rep. 8			Rep. 9			Rep. 10		
(36)	1	8	10	(43)	1	11	18	(50)	1	12	20	(57)	1	13	19	(64)	1	14	16
(37)	2	18	19	(44)	2	10	20	(51)	2	6	8	(58)	2	9	12	(65)	2	15	21
(38)	3	15	17	(45)	3	5	12	(52)	3	14	19	(59)	3	4	20	(66)	3	6	13
(39)	4	12	16	(46)	4	9	13	(53)	4	18	21	(60)	5	8	14	(67)	4	7	10
(40)	5	9	21	(47)	6	16	21	(54)	5	7	15	(61)	6	7	18	(68)	5	18	20
(41)	6	11	14	(48)	7	14	17	(55)	9	10	17	(62)	10	15	16	(69)	8	12	17
(42)	7	13	20	(49)	8	15	19	(56)	11	13	16	(63)	11	17	21	(70)	9	11	19

Esquema 5.34 – $t = 21, k = 5, r = 5, b = 21, \lambda = 1, E = 0{,}84$, tipo IV.

Bloco					
(1)	1	2	3	4	17
(2)	5	6	7	8	17
(3)	9	10	11	12	17
(4)	13	14	15	16	17
(5)	1	5	9	13	18
(6)	2	6	10	14	18
(7)	3	7	11	15	18

(8)	4	8	12	16	18
(9)	1	6	11	16	19
(10)	2	5	12	15	19
(11)	3	8	9	14	19
(12)	4	7	10	13	19
(13)	1	7	12	14	20
(14)	2	8	11	13	20

(15)	3	5	10	16	20
(16)	4	6	9	15	20
(17)	1	8	10	15	21
(18)	2	7	9	16	21
(19)	3	6	12	13	21
(20)	4	5	11	14	21
(21)	17	18	19	20	21

Esquema 5.35 – $t = 21, k = 7, r = 10, b = 30, \lambda = 3, E = 0{,}90$, tipo III.

Bloco							
(1)	2	5	10	11	17	19	20
(2)	3	6	11	12	18	20	21
(3)	4	7	12	13	19	21	15
(4)	5	1	13	14	20	15	16
(5)	6	2	14	8	21	16	17
(6)	7	3	8	9	15	17	18
(7)	1	4	9	10	16	18	19
(8)	3	4	8	13	17	19	20
(9)	4	5	9	14	18	20	21
(10)	5	6	10	8	19	21	15
(11)	6	7	11	9	20	15	16
(12)	7	1	12	10	21	16	17
(13)	1	2	13	11	15	17	18
(14)	2	3	14	12	16	18	19
(15)	1	6	9	12	17	19	20

(16)	2	7	10	13	18	20	21
(17)	3	1	11	14	19	21	15
(18)	4	2	12	8	20	15	16
(19)	5	3	13	9	21	16	17
(20)	6	4	14	10	15	17	18
(21)	7	5	8	11	16	18	19
(22)	1	2	4	8	9	11	21
(23)	2	3	5	9	10	12	15
(24)	3	4	6	10	11	13	16
(25)	4	5	7	11	12	14	17
(26)	5	6	1	12	13	8	18
(27)	6	7	2	13	14	9	19
(28)	7	1	3	14	8	10	20
(29)	1	2	3	4	5	6	7
(30)	8	9	10	11	12	13	14

Esquema 5.36 – $t = 25$, $k = 4$, $r = 8$, $b = 50$, $\lambda = 1$, $E = 0{,}78$, tipo III.

Bloco				
(1)	25	24	8	16
(2)	1	3	24	14
(3)	19	10	3	9
(4)	7	23	10	5
(5)	13	8	23	6
(6)	24	18	4	20
(7)	18	6	5	1
(8)	6	12	15	3
(9)	12	25	13	17
(10)	3	17	18	11
(11)	17	16	6	19
(12)	16	20	12	10
(13)	20	1	25	21
(14)	10	21	17	22
(15)	21	14	16	7
(16)	14	11	20	23
(17)	11	19	1	2
(18)	23	2	21	15
(19)	2	9	14	13
(20)	9	22	11	8
(21)	22	7	19	4
(22)	8	4	2	12
(23)	4	5	9	25
(24)	5	15	22	24
(25)	15	13	7	18
(26)	25	2	10	18
(27)	1	4	23	17
(28)	19	18	8	21
(29)	7	17	24	2
(30)	13	21	3	4
(31)	24	9	21	6
(32)	18	22	14	12
(33)	6	7	11	25
(34)	12	23	19	24
(35)	3	5	2	16
(36)	17	15	9	20
(37)	16	13	22	1
(38)	20	8	7	3
(39)	10	6	4	14
(40)	21	12	5	11
(41)	14	25	15	19
(42)	11	24	13	10
(43)	23	16	18	9
(44)	2	20	6	22
(45)	9	1	12	7
(46)	22	3	25	23
(47)	8	14	17	5
(48)	4	11	16	15
(49)	5	19	20	13
(50)	15	10	1	8

Esquema 5.37 – $t = 25$, $k = 9$, $r = 9$, $b = 25$, $\lambda = 3$, $E = 0{,}93$, tipo IV.

Bloco									
(1)	1	2	3	4	5	6	7	8	9
(2)	1	2	3	14	15	18	19	22	23
(3)	1	2	3	16	17	20	21	24	25
(4)	1	4	7	10	11	14	15	20	21
(5)	1	4	7	12	13	16	17	18	19
(6)	1	5	9	10	11	18	19	24	25
(7)	1	5	9	12	13	20	21	22	23
(8)	1	6	8	10	11	16	17	22	23
(9)	1	6	8	12	13	14	15	24	25
(10)	2	4	9	10	12	15	17	22	24
(11)	2	4	9	11	13	14	16	23	25
(12)	2	5	8	10	12	14	16	19	21
(13)	2	5	8	11	13	15	17	18	20
(14)	2	6	7	10	12	18	20	23	25
(15)	2	6	7	11	13	19	21	22	24
(16)	3	4	8	10	13	18	21	23	24
(17)	3	4	8	11	12	19	20	22	25
(18)	3	5	7	10	13	15	16	22	25
(19)	3	5	7	11	12	14	17	23	24
(20)	3	6	9	10	13	14	17	19	20
(21)	3	6	9	11	12	15	16	18	21
(22)	4	5	6	14	16	18	20	22	24
(23)	4	5	6	15	17	19	21	23	25
(24)	7	8	9	14	17	18	21	22	25
(25)	7	8	9	15	16	19	20	23	24

Esquema 5.38 – $t = 28$, $k = 4$, $r = 9$, $b = 63$, $\lambda = \lambda$, $E = 0{,}78$, tipo I.

Bloco	Rep. 1				Rep. 2				Rep. 3			
(1)	28	1	10	19	(8)	28	2	11	20	(15)	28	3	12	21
(2)	2	9	13	16	(9)	3	1	14	17	(16)	4	2	15	18
(3)	3	8	11	18	(10)	4	9	12	10	(17)	5	1	13	11
(4)	4	7	23	24	(11)	5	8	24	25	(18)	6	9	25	26
(5)	5	6	20	27	(12)	6	7	21	19	(19)	7	8	22	20
(6)	12	17	22	25	(13)	13	18	23	26	(20)	14	10	24	27
(7)	14	15	21	26	(14)	15	16	22	27	(21)	16	17	23	19
	Rep. 4				Rep. 5				Rep. 6			
(22)	28	4	13	22	(29)	28	5	14	23	(36)	28	6	15	24
(23)	5	3	16	10	(30)	6	4	17	11	(37)	7	5	18	12
(24)	6	2	14	12	(31)	7	3	15	13	(38)	8	4	16	14
(25)	7	1	26	27	(32)	8	2	27	19	(39)	9	3	19	20
(26)	8	9	23	21	(33)	9	1	24	22	(40)	1	2	25	23
(27)	15	11	25	19	(34)	16	12	26	20	(41)	17	13	27	21
(28)	17	18	24	20	(35)	18	10	25	21	(42)	10	11	26	22
	Rep. 7				Rep. 8				Rep. 9			
(43)	28	7	16	25	(50)	28	8	17	26	(57)	28	9	18	27
(44)	8	6	10	13	(51)	9	7	11	14	(58)	1	8	12	15
(45)	9	5	17	15	(52)	1	6	18	16	(59)	2	7	10	17
(46)	1	4	20	21	(53)	2	5	21	22	(60)	3	6	22	23
(47)	2	3	26	24	(54)	3	4	27	25	(61)	4	5 5	19	26
(48)	18	14	19	22	(55)	10	15	20	23	(62)	11	16	21	24
(49)	11	12	27	23	(56)	12	13	19	24	(63)	13	14	20	25

Capítulo 6

Delineamentos em reticulado

Os delineamentos em reticulado, também chamados de delineamentos em látice (do inglês *lattice* = reticulado), foram desenvolvidos por Yates na década de 1930. Basicamente existem quatro tipos de delineamentos em reticulado: o reticulado quadrado (YATES, 1936b), o reticulado retangular (HARSHBARGER, 1947, citado por COCHRAN; COX, 1983 e HARSHBARGER, 1949), o reticulado cúbico (YATES; HALE, 1939) e o quadrado de Youden (YOUDEN, 1940), a respeito do qual foi feita uma breve descrição no capítulo sobre quadrado latino.

Os reticulados, casos especiais de blocos incompletos, são sempre de tipo I, isto é, os blocos se juntam formando repetições e tanto podem ser balanceados quanto parcialmente balanceados e apresentam a análise de variância intrablocos, bem como a análise com recuperação interblocos. De uma forma geral, são, entre os BIB e BIPB, os desenhos mais utilizados na pesquisa agropecuária, apesar de algumas limitações que apresentam. Foram criados basicamente para resolver problemas de teste de linhagens desenvolvidas pelos melhoristas de plantas, que sempre trabalham com elevado número de linhagens.

A principal desvantagem desses desenhos é a limitação no número de materiais que podem ser testados. Os reticulados quadrados estão limitados a números de tratamentos iguais a um quadrado perfeito (daí seu nome), ou seja, o número de tratamentos é sempre igual a k^2 (9, 16, 25, 36, 49, 64, 81, 100, 121, ...), distribuídos em k blocos de k unidades cada um, em cada repetição. Os reticulados retangulares permitem o teste de $k(k+1)$ tratamentos distribuídos em $k+1$ blocos de k parcelas, por repetição, o que implica que o número de tratamentos que se podem testar nesse caso é 12 = 3*4, 20 = 4*5, 30 = 5*6, 42 = 6*7, 56 = 7*8, 72 = 8*9, 90 = 9*10, e assim por diante. Finalmente os reticulados cúbicos permitem o uso de k^3 tratamentos, isto é, 8 = 2^3, 27 = 3^3, 64 = 4^3, etc. Cada repetição possui k^2 blocos com k parcelas.

Se um pesquisador possuir 60 materiais para testar, por exemplo, ele terá de buscar outro tipo de delineamento, ou então adicionar outros quatro materiais (64 no total) para poder usar um reticulado quadrado (8^2) ou um reticulado cúbico (4^3), ou ainda eliminar quatro deles

(ficando com 56) para poder utilizar o reticulado retangular (7*8). A solução mais comum nesses casos é a inclusão de alguns materiais em uso corrente como testemunhas.

Reticulados quadrados

Estes delineamentos, conforme já foi mencionado, são do tipo I, blocos agrupados em repetição. Quando o valor básico k é primo ou potência de primo (2, 3, 4, 5, 7, 8, 9, ...), são conhecidas $k + 1$ repetições com arranjos dos tratamentos em forma diferente, e o uso de todas elas faz que o desenho seja balanceado. Na realidade, qualquer número de repetições pode ser adotado. Quando esse número for menor do que $k + 1$ usam-se alguns dos arranjos uma única vez ou alguns deles repetidos. Quando o número de repetições desejado é maior que $k + 1$, fato de ocorrência rara, usam-se alguns dos arranjos básicos repetidos. Quando k é um produto de primos (6, 10, 12, 15, ...), apenas três arranjos de repetições são conhecidos e o delineamento é então parcialmente balanceado. Para o uso de número de repetições diferente de três segue-se o mesmo critério anterior para k primo ou potência de primo. Para alguns valores de k, os arranjos básicos são dados ao final do capítulo.

Nos reticulados quadrados, os tratamentos que numa repetição aparecem juntos num bloco, nas demais repetições se distribuem entre os blocos, de forma que o valor de λ é sempre 1 se balanceado, e igual a 1 ou 0 se parcialmente balanceado, exceto quando se usam réplicas das repetições básicas. Nesse caso, o valor de λ será igual ao número de réplicas para os tratamentos que aparecem juntos em algum bloco e igual a 0 quando não aparecem.

Casualização do delineamento

A aleatorização para estes delineamentos segue um processo de quatro etapas:

1) Seleção dos arranjos básicos de repetição.
2) Casualização dos blocos em cada repetição.
3) Casualização dos tratamentos em cada bloco.
4) Designação aleatória dos números sequenciais aos tratamentos.

Para exemplificar o processo, adota-se o caso de $k = 3$, em que os quatro arranjos básicos (plano 7.3) são os seguintes:

1	2	3
4	5	6
7	8	9

1	4	7
2	5	8
3	6	9

1	5	9
2	6	7
3	4	8

1	6	8
2	4	9
3	5	7

Se o experimento for executado com quatro repetições, devem ser usados os quatro arranjos básicos, o que o faz um caso balanceado. Quando se pretende conduzi-lo com três repetições, podem ser escolhidos os três primeiros arranjos básicos, por exemplo. Se houver interesse no uso de seis repetições, podem ser escolhidos os mesmos três primeiros arranjos básicos que devem ser executados duas vezes cada um ou podem ser utilizados os quatro arranjos básicos e dois deles serem empregados duas vezes.

No caso de quatro repetições, com o uso das quatro formas, e admitindo que a ordem dos blocos em cada repetição, casualizada, seja 1, 3, 2; 2, 1, 3; 1, 2, 3 e 3, 1, 2, respectivamente, após a terceira etapa de casualização dos tratamentos nos blocos, ter-se-ia o seguinte desenho:

3	2	1
8	9	7
5	4	6

2	8	5
7	1	4
6	3	9

1	5	9
6	7	2
3	8	4

7	5	3
8	6	1
4	9	2

A última etapa seria identificar aleatoriamente cada tratamento com um dos números de 1 a 9.

Análise intrablocos de reticulados quadrados balanceados e parcialmente balanceados

O modelo matemático é aquele descrito para os BIB de tipo I.

O esquema de análise de variância é também idêntico ao já apresentado no caso de delineamentos balanceados. No entanto, para o caso de desenhos parcialmente balanceados a análise é ligeiramente modificada e se detalha a seguir para os casos de haver ou não replicação dos esquemas básicos.

Quando não ocorre réplica de esquemas básicos, os desenhos com duas repetições básicas são chamados de reticulados simples, e os com três repetições básicas de reticulados triplos. Os com mais repetições não recebem nomes especiais. Em geral, adotam-se as primeiras repetições básicas, dadas no item *Planos básicos de reticulados quadrados e reticulados retangulares* ao final do capítulo, na execução dos experimentos parcialmente balanceados.

Análise intrablocos de um reticulado parcialmente balanceado sem replicação de esquemas básicos

O modelo matemático é aquele descrito para os BIB de tipo I.

O esquema de análise de variância é também idêntico ao já apresentado no caso de desenhos em blocos incompletos balanceados. Entretanto, para o caso de desenhos reticulados, a análise no que se refere à soma de quadrados de tratamentos ajustada é ligeiramente modificada e é detalhada no próximo item.

Quando não ocorre repetição de planos básicos, os delineamentos com duas repetições básicas são chamados de reticulados simples, e os com três repetições básicas de reticulados triplos. Os com mais repetições não recebem nomes especiais. Em geral, adotam-se as primeiras repetições básicas, dadas ao final do capítulo, na execução dos experimentos parcialmente balanceados.

Exemplo de análise intrablocos de um reticulado quadrado parcialmente balanceado sem repetição de esquemas básicos

Os pesquisadores Orlando Peixoto Moraes, Emilio da Maia de Castro e Flávio Brezeghelo, da Embrapa Arroz e Feijão, realizaram na fazenda Capivara em Santo Antônio de Goiás, um ensaio de competição de cultivares de arroz em 1994–1995, em um delineamento em reticulado quadrado 5 x 5, com três repetições.

Os dados de produção, em kg ha^{-1}, encontram-se na Tabela 1, segundo o esquema utilizado em campo.

A análise de variância feita segundo as fórmulas apresentadas na Tabela 1 do Capítulo 5 encontra-se na Tabela 2 deste capítulo.

As somas de quadrados foram assim obtidas:

Fator de correção $(FC) = 172.886{,}25^2/75 = 398.528.739{,}19$

Soma de quadrados total:

$(SCTo) = (1.875{,}00^2 + 2.125{,}00^2 + ... + 1.687{,}50^2) - FC = 32.167.890{,}63$

Soma de quadrados de repetições:

$(SCR) = (1/25)*(49.281{,}25^2 + 62.906{,}25^2 + 60.698{,}75^2) - FC = 4.348.682{,}50$

Soma de quadrados de blocos (repetições):

$(SCBr) = (1/5)\ (8.656{,}25^2 + 11.500{,}00^2 + ... + 8.562{,}75^2) - FC - SCR = 10.544.375{,}00$

Tabela 1. Dados de produção de arroz de terras altas, em kg ha^{-1}, de um ensaio em reticulado quadrado 5^2, com três repetições.

Tratamento/produtividade					Total do bloco	Total da repetição
17	20	16	19	18		
1.875,00	2.125,00	2.250,00	1.562,50	843,75	8.656,25	
10	9	7	8	6		
937,50	2.093,75	3.000,00	2.843,75	2.625,00	11.500,00	
4	2	5	1	3		
1.531,25	1.531,25	2.031,25	1.843,75	1.562,50	8.500,00	
15	13	12	14	11		
2.750,00	2.187,50	2.187,50	2.312,50	1.937,50	11.375,00	
21	22	24	23	25		
1.031,25	1.968,75	2.500,00	1.875,00	1.875,00	9.250,00	49.281,25
24	9	4	19	14		
1.718,75	2.156,25	1.812,50	1.875,00	2.343,75	9.906,25	
5	15	25	20	10		
2.343,75	2.656,25	2.125,00	2.468,75	2.250,00	11.843,75	
1	16	6	11	21		
3.125,00	3.468,75	2.500,00	2.437,50	1.843,75	13.375,00	
12	2	7	22	17		
2.500,00	3.156,25	4.000,00	3.062,50	3.125,00	15.843,75	
13	18	23	3	8		
2.843,75	2.500,00	2.500,00	1.625,00	2.468,75	11.937,50	62.906,25
22	9	15	3	16		
1.625,00	1.562,50	2.843,75	1.875,00	2.812,50	10.718,75	
18	6	12	24	5		
1.593,75	2.875,00	2.937,50	3.750,00	3.218,75	14.375,00	
17	11	10	4	23		
3.312,50	3.031,25	2.500,00	1.906,25	2.531,25	13.281,25	
7	13	25	1	19		
3.750,00	2.593,75	2.500,00	2.812,50	2.375,00	14.031,25	
21	20	8	2	14		
1.375,00	1.718,75	2.531,25	1.250,00	1.687,50	8.562,50	60.968,75

Tabela 2. Análise de variância de um reticulado quadrado 5^2.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetições	2	4.346.562,50	2.173.281,25	14,90 < 0,0001
Blocos (Repetições)	12	10.544.375,00		
Tratamentos (Ajustados)	24	12.026.002,60	501.083,44	3,44 0,0004
Erro	36	5.250.950,52	145.859,74	
Total	74	32.167.890,63		

Para obter a soma de quadrados de tratamentos ajustada para efeitos de blocos, uma tabela adicional é muito importante, pois facilita em muito os cálculos (Tabela 3).

Nessa tabela, SB_i é igual à soma dos totais de blocos em que o tratamento está presente; Qi é igual a $k*T_i$ - SB_i, em que T_i é o total do tratamento i; SQ_{i1} é a soma dos Q_i dos tratamentos que estavam no mesmo bloco na repetição 1; SQ_{i2} e SQ_{i3} são os valores equivalentes para as repetições 2 e 3; t_i é o efeito ajustado para blocos do tratamento i e é obtido por meio da fórmula:

$$t_i = \frac{Q_i}{r * k} + \frac{SQ_{i1} + SQ_{i2} + SQ_{i3}}{r(r - 1)k^2} \quad [1]$$

Por último, a média ajustada dos tratamentos é a soma entre a média geral do ensaio e o valor t_i. Portanto, a soma de quadrados de tratamentos ajustada ($SQTaj.$) é dada por

$$SQTaj = \frac{\sum t_i Q_i}{k} \quad [2]$$

que, no exemplo que se está analisando, é:

$SQTaj$ = [3.000,00*229,79 + (-3.218,75)*(-233,75) + ... + (-2.625,00)*(-151,46)]/5

$SQTaj$ = 12.026.002,60

Assim, a soma de quadrados do engano é:

$SQE = SQTo - SQR - SQBr - SQTaj$ = 5.250.950,52

Tabela 3. Tabela para obtenção de médias de tratamentos ajustados em um reticulado quadrado 5^2.

Tratamento	Total de tratamentos	Sb_i	Q_i	SQ_{i1}	SQ_{i2}	SQ_{i3}	t_i	Média de tratamentos
1	7.781,25	35.906,25	3.000,00	-8.250,00	2.718,75	10.000,00	229,79	2.538,54
2	5.937,50	32.906,25	-3.218,75	-8.250,00	6.937,50	-1.562,50	-233,75	2.075,00
3	5.062,50	31.156,25	-5.843,75	-8.250,00	-8.062,50	5.781,25	-459,79	1.848,96
4	5.250,00	31.687,50	-5.437,50	-8.250,00	-3.843,75	-10.781,25	-515,00	1.793,75
5	7.593,75	34.718,75	3.250,00	-8.250,00	2.250,00	-3.437,50	153,75	2.462,50
6	8.000,00	39.250,00	750,00	9.093,75	2.718,75	-3.437,50	105,83	2.414,58
7	10.750,00	41.375,00	12.375,00	9.093,75	6.937,50	10.000,00	998,54	3.307,29
8	7.843,75	32.000,00	7.218,75	9.093,75	-8.062,50	-1.562,50	477,71	2.786,46
9	5.812,50	32.125,00	-3.062,50	9.093,75	-3.843,75	5.781,25	-130,63	2.178,13
10	5.687,50	36.625,00	-8.187,50	9.093,75	2.250,00	-10.781,25	-542,08	1.766,67
11	7.406,25	38.031,25	-1.000,00	5.500,00	2.718,75	-10.781,25	-83,75	2.225,00
12	7.625,00	41.593,75	-3.468,75	5.500,00	6.937,50	-3.437,50	-171,25	2.137,50
13	7.625,00	37.343,75	781,25	5.500,00	-8.062,50	10.000,00	101,67	2.410,42
14	6.343,75	29.843,75	1.875,00	5.500,00	-3.843,75	-1.562,50	125,63	2.434,38
15	8.250,00	33.937,50	7.312,50	5.500,00	2.250,00	5.781,25	577,71	2.886,46
16	8.531,25	32.750,00	9.906,25	2.375,00	2.718,75	5.781,25	732,92	3.041,67
17	8.312,50	37.781,25	3.781,25	2.375,00	6.937,50	-10.781,25	242,29	2.551,04
18	4.937,50	34.968,75	-10.281,25	2.375,00	-8.062,50	-3.437,50	-746,25	1.562,50
19	5.812,50	32.593,75	-3.531,25	2.375,00	-3.843,75	10.000,00	-178,54	2.130,21
20	6.312,50	29.062,50	2.500,00	2.375,00	2.250,00	-1.562,50	187,08	2.495,83
21	4.250,00	31.187,50	-9.937,50	-8.718,75	2.718,75	-1.562,50	-712,92	1.595,83
22	6.656,25	35.812,50	-2.531,25	-8.718,75	6.937,50	5.781,25	-142,08	2.166,67
23	6.906,25	34.468,75	62,50	-8.718,75	-8.062,50	-10.781,25	-179,58	2.129,17
24	7.968,75	33.531,25	6.312,50	-8.718,75	-3.843,75	-3.437,50	314,17	2.622,92
25	6.500,00	35.125,00	-2.625,00	-8.718,75	2.250,00	10.000,00	-151,46	2.157,29
Soma	173.156,25	865.781,25	0,00	0,00	0,00	0,00	0,00	57.718,75

Para os testes de médias, dois processos podem ser considerados: um mais exato, que considera se os tratamentos estavam ou não em um mesmo bloco, pois nesse caso os contrastes entre médias têm variâncias diferentes, ou outro que emprega um valor médio (PIMENTEL-GOMES, 2000). Esses três valores de variância das diferenças entre médias são dados a seguir:

- Tratamentos aparecem em um mesmo bloco:

$$V_{dif} = \frac{2 * QME}{r}\left(1 + \frac{1}{k}\right) \quad [3]$$

- Tratamentos não aparecem em um mesmo bloco:

$$V_{dif} = \frac{2 * QME}{r}\left(1 + \frac{r}{(r - 1)k}\right) \quad [4]$$

- Valor médio (pode ser usado nos casos a e b acima, mas é uma aproximação):

$$V_{dif} = \frac{2 * QME}{r}\left(1 + \frac{r}{(r - 1)(k + 1)}\right) \quad [5]$$

Uso do SAS com delineamentos em blocos incompletos

No caso dos delineamentos em blocos incompletos, reticulados quadrados, retangulares ou outros desse tipo, o SAS permite o uso do procedimento GLM, com o comando LSMEANS para a estimação das médias ajustadas e o uso do procedimento LATTICE para o caso dos reticulados. Para ilustrar o uso desses procedimentos, serão utilizados os dados do exemplo dado na seção intitulada *Exemplo de análise intrablocos de um reticulado quadrado parcialmente balanceado sem repetição de esquemas básicos*.

No caso do uso do GLM, o programa para a obtenção da análise de variância e das médias dos tratamentos ajustadas é o seguinte:

```
PROC GLM;CLASS REPETICAO BLOCO TRATAMENTO;
MODEL PRODUCAO=REPETICAO BLOCO(REPETICAO) TRATAMENTO/SS1;
LSMEANS TRATAMENTO;
```

O resultado desse procedimento se vê na Figura 1, cujos valores são exatamente aqueles obtidos na seção intitulada *Exemplo de análise intrablocos de um reticulado quadrado parcialmente balanceado sem repetição de esquemas básicos*. Se a opção SS1 fosse omitida, seria obtida também outra estimação de valores para a soma de quadrados de blocos dentro de repetição, ajustada para o fato de que os blocos não contêm todos os tratamentos, o que é mostrado pelo SAS como TYPE III SS, e que se ilustra na Figura 2.

```
                          The GLM Procedure
                       Class Level Information
Class         Levels  Values
REPETICAO          3  1 2 3
BLOCO              5  1 2 3 4 5
TRATAMENTO        25  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
                      23 24 25
                          Number of observations    75
                          The GLM Procedure
Dependent Variable: PRODUCAO
                                    Sum of
 Source                     DF       Squares    Mean Square   F Value   Pr > F
 Model                      38    26916940.10      708340.53      4.86   <.0001
 Error                      36     5250950.52      145859.74
 Corrected Total            74    32167890.63
            R-Square     Coeff Var      Root MSE    PRODUCAO Mean
            0.836764     16.54211      381.9159          2308.750

 Source                     DF      Type I SS    Mean Square   F Value   Pr > F
 REPETICAO                   2     4346562.50     2173281.25     14.90   <.0001
 BLOCO(REPETICAO)           12    10544375.00      878697.92      6.02   <.0001
 TRATAMENTO                 24    12026002.60      501083.44      3.44   0.0004
                          The GLM Procedure
                        Least Squares Means
                                      PRODUCAO
                       TRATAMENTO      LSMEAN
                       1            2538.54167
                       2            2075.00000
                       3            1848.95833
                       4            1793.75000
                       5            2462.50000
                       6            2414.58333
                       7            3307.29167
                       8            2786.45833
                       9            2178.12500
                       10           1766.66667
                       11           2225.00000
                       12           2137.50000
                       13           2410.41667
                       14           2434.37500
                       15           2886.45833
                       16           3041.66667
                       17           2551.04167
                       18           1562.50000
                       19           2130.20833
                       20           2495.83333
                       21           1595.83333
                       22           2166.66667
                       23           2129.16667
                       24           2622.91667
                       25           2157.29167
```

Figura 1. Uso do procedimento GLM do SAS para a análise de variância de um ensaio em reticulado quadrado 5 x 5, com três repetições.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REPETICAO	2	4346562.50	2173281.25	14.90	<.0001
BLOCO(REPETICAO)	12	6003398.44	500283.20	3.43	0.0020
TRATAMENTO	24	12026002.60	501083.44	3.44	0.0004

Figura 2. Análise de variância com a soma de quadrados de blocos dentro de repetição, ajustada para tratamentos, com o uso do SAS.

No caso de uso do procedimento LATTICE, as variáveis de classificação devem ter seu nome escrito da seguinte forma (em inglês): group, block, treatmnt[1].

O programa seria então:

PROC LATTICE;VAR PRODUCAO;

e o resultado se vê na Figura 3.

The Lattice Procedure
Analysis of Variance for PRODUCAO

Source	DF	Sum of Squares	Mean Square
Replications	2	4346563	2173281
Blocks within Replications (Adj.)	12	6003398	500283
Component B	12	6003398	500283
Treatments (Unadj.)	24	16566979	690291
Intra Block Error	36	5250951	145860
Randomized Complete Block Error	48	11254349	234466
Total	74	32167891	434701

Additional Statistics for PRODUCAO

Variance of Means in Same Block	111018
Variance of Means in Different Bloc	117907
Average of Variance	114462
LSD at .01 Level	920.06
LSD at .05 Level	686.15
Efficiency Relative to RCBD	136.56

Adjusted Treatment Means for PRODUCAO

Treatment	Mean
1	2554.64
2	2047.06
3	1801.88
4	1780.99
5	2482.54
6	2488.08
7	3387.77
8	2736.35
9	2107.97
10	1804.33
11	2296.07
12	2255.34
13	2448.68
14	2341.14
15	2846.67
16	2983.96
17	2615.12
18	1586.80
19	2074.02
20	2381.64
21	1543.60
22	2181.85
23	2179.58
24	2632.64
25	2160.02

Figura 3. Análise de variância de um reticulado quadrado com o uso do procedimento LATTICE do SAS.

[1] A escrita correta da palavra no inglês é *treatment*.

A análise de variância é feita com recuperação da informação interblocos. Note-se, no entanto, que o procedimento não mostra a soma de quadrados de tratamentos ajustada, mas as médias dos tratamentos estão ajustadas. Os valores dessas médias são distintos daqueles obtidos no primeiro caso e também diferentes das médias aritméticas dos valores obtidos no ensaio. O procedimento também faz um complemento de análise, estimando o valor do erro, considerando o experimento como um bloco completo ao acaso e também calculando o ganho em eficiência no uso do delineamento em reticulado, em substituição ao de blocos completos. No caso atual, esse ganho de eficiência foi estimado em 36,56%, o que mostra o acerto no uso do reticulado.

Deixa-se ao leitor a decisão sobre qual procedimento lhe parece o mais adequado.

Reticulados retangulares

Esta classe de delineamentos em blocos incompletos tem um número de tratamentos igual ao produto $k^*(k+1)$, que se distribui em $k+1$ blocos com k unidades experimentais em cada repetição. Esses delineamentos foram apresentados por Harshbarger em 1947 (citado por COCHRAN, COX, 1983 e HARSHBARGER, 1949). Os números permitidos de tratamentos são 12, 20, 30, 42, 56, 72, etc., de forma que os números de tratamentos que podem ser empregados complementam aqueles utilizados nos reticulados quadrados. São apresentadas três repetições para cada um dos números de tratamentos, ao final deste capítulo. Os livros que apresentam sua análise de variância, Cochran e Cox (1983) e Kempthorne (1975), por exemplo, são recomendados como referência para estudos adicionais.

O reticulado retangular tem algumas peculiaridades não encontradas em outros delineamentos em blocos incompletos. Para cada projeto devemos identificar as réplicas, e cada uma delas foi chamada de X, Y e Z. Um exame de qualquer bloco de uma réplica, quando comparado com os blocos das outras repetições, mostra que um e somente um dos blocos nas outras repetições não contém qualquer tratamento em comum com o bloco considerado. Esses dois blocos são chamados de sócios, e devem ser perfeitamente identificados, pois é muito importante poder distingui-los na análise de variação. Assim, se na primeira réplica um bloco se identifica como $X1$, seus sócios nas outras duas réplicas devem ser chamados de $Y1$ e $Z1$.

O modelo matemático é o mesmo que se adotou no reticulado quadrado, e a casualização do delineamento se faz da mesma forma vista nos reticulados quadrados.

Exemplo de aplicação

Um ensaio de competição de cultivares e linhagens de arroz irrigado foi conduzido pelo pesquisador Paulo Hideo Nakano Rangel em 2000–2001, na fazenda Palmital da Embrapa Arroz e Feijão, no delineamento reticulado retangular 5 x 6, com três repetições. Os dados estão na Tabela 4, enquanto as Tabelas 5 e 6 mostram alguns valores auxiliares para a análise de variância que se encontra na Tabela 7. A Tabela 5 mostra também as médias de tratamentos ajustadas para os efeitos de blocos.

Tabela 4. Dados de produção de grãos, em kg ha^{-1}, do ensaio em reticulado retangular, totais de blocos e de repetição, efeito de blocos (C_x, C_y e C_z) e fator de ajuste do bloco.

X3	15	13	12	11	14	Total de blocos	C_x	Fator de ajuste
	7.164,94	4.150,57	4.994,54	5.543,68	5.017,24	26.870,97	1.051,17	53,8219
X5	21	25	23	24	22			
	5.612,07	5.177,59	4.758,62	5.595,98	5.151,44	26.295,70	272,40	12,9118
X1	5	1	4	3	2			
	4.256,32	4.628,45	4.889,94	3.713,22	4.602,30	22.090,23	2.535,05	128,4428
X2	10	9	6	8	7			
	5.710,34	5.974,71	4.236,21	5.622,13	5.386,78	26.930,17	1.247,13	63,4259
X6	30	29	28	26	27			
	6.929,60	4.445,40	7.243,39	3.516,09	4.393,10	26.527,58	5.229,31	264,1405
X4	20	17	16	18	19			
	4.916,09	6.439,66	5.517,53	5.517,53	5.256,03	27.646,84	4.947,71	250,5652
						156.361,49	15.282,77	

Y6	10	25	15	5	20	Total de blocos	C_y	Fator de ajuste
	5.419,54	5.049,43	5.419,54	3.859,77	5.948,28	25.696,56	3.126,41	157,7337
Y4	24	8	3	29	13			
	6.275,86	5.710,34	5.287,36	5.909,77	6.344,83	29.528,16	-9.758,04	-493,5457
Y2	1	22	12	17	27			
	4.573,56	5.181,61	5.657,47	6.529,89	4.785,34	26.727,87	-1.616,36	-81,4667
Y5	14	30	19	4	9			
	6.066,67	7.086,49	6.397,70	3.739,37	5.419,54	28.709,77	-426,44	-22,4495
Y3	7	28	23	2	18			
	6.197,41	7.217,24	4.520,69	5.334,48	5.525,29	28.795,11	-1.512,93	-75,9216
Y1	21	11	26	6	16			
	5.181,61	4.968,39	4.445,40	5.445,98	5.962,07	26.003,45	-1.828,16	-92,3356
						165.460,92	-12.015,52	

Continua...

Tabela 4. Continuação.

$Z2$	29	14	16	23	5	Total de blocos	C_z	Fator de ajuste
	5.831,32	5.099,14	6.589,66	4.629,31	4.282,76	26.432,19	-2.550,30	-128,7241
$Z1$	7	27	13	25	19			
	5.282,18	5.177,59	4.445,40	6.249,71	6.537,36	27.692,24	-2.245,70	-113,4631
$Z6$	12	6	24	18	4			
	5.525,29	5.831,32	5.818,97	5.974,71	4.863,79	28.014,08	-4.149,99	-210,4521
$Z3$	1	21	30	8	20			
	6.223,56	5.726,72	6.979,31	6.118,97	5.700,57	30.749,13	-5.289,64	-267,0231
$Z4$	22	2	26	15	9			
	4.229,89	4.994,54	3.164,08	6.825,00	5.445,98	24.659,49	2.891,07	146,4991
$Z5$	3	11	17	10	28			
	3.556,32	4.441,38	5.075,86	4.236,21	7.687,93	24.997,70	8.077,31	407,8402
						162.544,83	-3.267,25	

Na Tabela 4, os valores C_x, C_y e C_z são obtidos pela diferença entre a soma dos totais de tratamentos que se encontram no bloco e o produto $r*B$, em que r é o número de repetições e B o total do bloco em questão. O fator de ajuste é igual a

$$\text{Fator de ajuste} = \lambda C_i - \mu S \qquad [6]$$

obtido a partir dos dados da Tabela 5.

Tabela 5. Efeitos de blocos (C_x, C_y e C_z) e seu total (S).

Número do bloco	C_x	C_y	C_z	S_m	λC_x	λC_y	λC_z	μS
1	2.535,05	-1.828,16	-2.245,7	-1.538,81	128,2735	-92,5049	-113,6324	-0,1693
2	1.247,13	-1.616,36	-2.550,3	-2.919,53	63,1048	-81,7878	-129,0452	-0,3211
3	1.051,17	-1.512,93	-5.289,64	-5.751,40	53,1892	-76,5543	-267,6558	-0,6327
4	4.947,71	-9.758,04	2.891,07	-1.919,26	250,3541	-493,7568	146,2880	-0,2111
5	272,4	-426,44	8.077,31	7.923,27	13,7834	-21,5779	408,7118	0,8716
6	5.229,31	3.126,41	-4.149,99	4.205,73	264,6031	158,1963	-209,9895	0,4626
Totais(R_j)	15.282,77	-12.015,52	-3.267,25	0,00	773,3081	-607,9854	-165,3227	0,0000

Tabela 6. Análise de variância de um reticulado retangular.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	$F\ (Pr > F)$
Repetição	2	1.439.298,90		
Blocos ajustados	15	8.493.032,31	566.202,15	
Tratamentos	29	54.575.223,16	1.881.904,25	6,41 (< 0,0001)
Erro	43	12.624.596,05	293.594,63	
Total	89	77.132.150,42		

Tabela 7. Valor total e média de produção de grãos, em kg ha^{-1}.

Tratamento	Total não ajustado	Total ajustado	Média ajustada
1	15.425,57	15.205,52	5.068,51
2	14.931,32	15.130,34	5.043,45
3	12.556,90	12.599,64	4.199,88
4	13.493,10	13.388,64	4.462,88
5	12.398,85	12.556,30	4.185,43
6	15.513,51	15.274,15	5.091,38
7	16.866,37	16.740,41	5.580,14
8	17.451,44	16.754,30	5.584,77
9	16.840,23	17.027,71	5.675,90
10	15.366,09	15.995,09	5.331,70
11	14.953,45	15.322,78	5.107,59
12	16.177,30	15.939,20	5.313,07
13	14.940,80	14.387,61	4.795,87
14	16.183,05	16.085,70	5.361,90
15	19.409,48	19.767,53	6.589,18
16	18.069,26	18.098,77	6.032,92
17	18.045,41	18.622,35	6.207,45
18	17.017,53	16.981,72	5.660,57
19	18.191,09	18.305,74	6.101,91
20	16.564,94	16.706,22	5.568,74
21	16.520,40	16.173,95	5.391,32
22	14.562,94	14.640,88	4.880,29
23	13.908,62	13.716,89	4.572,30
24	17.690,81	16.999,72	5.666,57
25	16.476,73	16.533,91	5.511,30
26	11.125,57	11.443,87	3.814,62
27	14.356,03	14.425,24	4.808,41
28	22.148,56	22.744,62	7.581,54
29	16.186,49	15.828,36	5.276,12
30	20.995,40	20.970,07	6.990,02

A partir desses valores, podem ser calculadas as somas de quadrados em que as referentes ao total (*SQTo*), repetição (*SQR*), tratamentos (*SQTr*) e erro (*SQE*) calculam-se da maneira usual, enquanto a soma de quadrados para blocos (*SQB*) é ajustada para efeito de tratamentos. Essas somas de quadrado foram obtidas usando as seguintes expressões:

$$FC = \frac{(484.367,24)^2}{90} = 2.606.795.813,17$$

$$SQTo = (7.164,94^2 + 4.150,57^2 + ... + 8.077,31^2) - FC = 77.132.150,42$$

$$SQR = \frac{(156.361,49^2 + 165.460,92^2 + 162.544,83^2}{30} - FC = 1.439.298,90$$

$$SQTr = \frac{(15.425,57^2 + 14.931,32^2 + ... + 20.995,40^2}{3} - FC = \frac{7.984.113.108,9316}{3} - FC = 54.575.223,16$$

A soma de quadrados de blocos, ajustada é dada pela fórmula

$$SQBaj = \frac{\sum C_{mj}^2}{r(rk - k - 1)} - \frac{\sum R_j^2}{r(k+1)(rk - k - 1)} - \frac{\sum S_m^2}{r(r - 1)(k+1)(rk - k - 1)}$$

em que C_{mj} são os efeitos de blocos em cada uma das repetições, R_j são os totais dos valores C de cada uma das repetições e S_m é a soma dos valores C dos blocos considerados sócios, como visto na Tabela 5. Nessas fórmulas, $m = k + 1$. Para esse exemplo, tem-se então:

$$SQBaj = \frac{2.535,05^2 + 1.247,13^2 + ... + (-4.149,99)^2}{3(3*5 - 5 - 1)} - \frac{15.287,77^2 + ... + (-3.267,25)^2}{3(5+1)(3*5 - 5 - 1)} -$$

$$\frac{(-1.538,81)^2 + ... + 4.205,73^2}{3(3 - 1)(5+1)(3*5 - 5 - 1)} = \frac{304.756.999,7026}{27} - \frac{388.610.702,3058}{162} - \frac{128.120.124,8704}{324} = 8.493.032,31$$

A tabela de análise de variância pode então ser construída como se mostra na Tabela 6.

Agora, chamando os quadrados médios de blocos ajustados de E_b e o de erro de E_e, podem-se calcular os valores de λ e μ, necessários para completar as Tabelas 5 e 7.

$$\lambda = \frac{r(E_b - E_e)}{r(2k - 1)E_b + (rk - 3k + r)E_e} \quad [7]$$

$$\mu = \frac{\lambda^3}{1 + 3\lambda} \quad [8]$$

Note-se que essas fórmulas são válidas apenas para reticulado retangular tríplice, isto é, com três repetições distintas. Para um reticulado retangular simples (duas repetições), as fórmulas são:

$$\lambda = \frac{r(E_b - E_e)}{r(k - 1)E_b + (rk - 2k + r)E_e} \quad [9]$$

$$\mu = \frac{\lambda^2}{1 + 2\lambda} \quad [10]$$

No caso atual, tem-se:

$$\lambda = \frac{3(566.202,15 - 293.594,63)}{3(2*5 - 1)566.202,15 + (3*5 - 3*5 + 3)293.594,63} = \frac{817.822,56}{16.168.241,94} = 0,0506$$

e

$$\mu = \frac{0,0506^3}{1 + 3*0,0506} = 0,00011$$

Com o uso dos valores de λ e μ, completa-se a Tabela 5, em que o fator de ajuste é obtido segundo a fórmula 6 deste capítulo. Como exemplo, para o bloco $X3$ tem-se que o fator de ajuste é igual a 53,1892 - (-0,6327) ou fator de ajuste = 53,8219, e assim para todos os demais. Finalmente a Tabela 7 pode ser completada. Nela o total de tratamentos ajustado é igual ao seu total não ajustado. Somam-se todos os fatores de ajustes dos blocos onde ele aparece. Dessa forma, para o tratamento 1, por exemplo, tem-se:

$$Total\ ajustado = 15.425,57 + (128,4428 - 81,4667 - 267,0231) = 15.205,5230.$$

Consequentemente, a média ajustada será esse valor dividido por 3, o número de repetições.

Para a comparação de médias, procedimentos semelhantes àqueles adotados para o reticulado quadrado podem ser aplicados aqui. No caso de haver réplicas das repetições básicas, a análise deve ser modificada, mas, por ser um procedimento não muito comum, não é apresentado neste livro. Sugere-se aos interessados a leitura de outros textos, como o de Cochran e Cox (1983).

Uso do SAS no caso dos reticulados retangulares

Assim como no caso do reticulado quadrado, para o reticulado retangular é possível usar os procedimentos GLM e LATTICE, com os resultados sendo interpretados da mesma maneira. As Figuras 4 e 5 ilustram o uso dos procedimentos para os mesmos dados apresentados no item anterior (*Exemplo de aplicação*), e a comparação dos resultados fica como um exercício para o leitor. A programação é idêntica à utilizada na seção *Uso do SAS com delineamentos em blocos incompletos*.

```
                          The GLM Procedure
                       Class Level Information
Class       Levels  Values
group            3  1 2 3
block            6  1 2 3 4 5 6
Treatmnt        30  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
                    24 25 26 27 28 29 30

                      Number of observations    90
                          The GLM Procedure
Dependent Variable: PROD
                                    Sum of
Source                      DF       Squares    Mean Square   F Value   Pr > F
Model                       46    64507554.20     1402338.13      4.78   <.0001
Error                       43    12624596.22      293595.26
Corrected Total             89    77132150.42

              R-Square     Coeff Var      Root MSE      PROD Mean
              0.836325      10.06798      541.8443       5381.858

Source                      DF      Type I SS    Mean Square   F Value   Pr > F
group                        2     1439297.16      719648.58      2.45   0.0982
block(group)                15    11719425.33      781295.02      2.66   0.0061
Treatmnt                    29    51348831.71     1770649.37      6.03   <.0001


Source                      DF    Type III SS    Mean Square   F Value   Pr > F
group                        2     1439297.16      719648.58      2.45   0.0982
block(group)                15     8493045.79      566203.05      1.93   0.0470
Treatmnt                    29    51348831.71     1770649.37      6.03   <.0001

                          The GLM Procedure
                        Least Squares Means
                     Treatmnt      PROD LSMEAN
                       1          5011.48184
                       2          5150.46562
                       3          4203.49173
                       4          4389.40861
                       5          4248.95966
                       6          4996.72467
                       7          5560.61977
                       8          5338.69909
                       9          5741.36335
                      10          5554.74936
                      11          5252.90993
                      12          5231.57638
                      13          4603.04119
                      14          5325.33880
                      15          6742.32475
                      16          6063.85164
                      17          6428.15560
                      18          5656.71385
                      19          6134.19629
                      20          5635.47077
                      21          5251.31350
                      22          4902.09132
                      23          4498.11445
                      24          5360.39982
                      25          5502.26515
                      26          3939.26316
                      27          4836.76298
                      28          7799.97339
                      29          5135.26589
                      30          6960.75458
```

Figura 4. Uso do procedimento GLM do SAS na análise de variância de um reticulado retangular 6 x 5.

The Lattice Procedure
Analysis of Variance for PROD

Source	DF	Sum of Squares	Mean Square
Replications	2	1439297	719649
Blocks within Replications (Adj.)	15	8493046	566203
Component B	15	8493046	566203
Treatments (Unadj.)	29	54575211	1881904
Intra Block Error	43	12624596	293595
Randomized Complete Block Error	58	21117642	364097
Total	89	77132150	866653

Additional Statistics for PROD

Variance of Means in Same Block	215096
Variance of Means in Different Bloc	224779
Average of Variance	220772
LSD at .01 Level	1266.33
LSD at .05 Level	947.57
Efficiency Relative to RCBD	109.95

Adjusted Treatment Means for PROD

Treatment	Mean
1	5075.72
2	5049.91
3	4196.73
4	4455.44
5	4185.59
6	5091.59
7	5587.34
8	5592.30
9	5673.71
10	5325.14
11	5107.10
12	5316.24
13	4802.42
14	5362.44
15	6591.58
16	6037.40
17	6205.21
18	5663.02
19	6098.76
20	5571.16
21	5390.91
22	4878.11
23	4572.85
24	5659.47
25	5503.84
26	3814.06
27	4808.58
28	7576.98
29	5276.61
30	6985.54

Figura 5. Uso do procedimento LATTICE do SAS na análise de um reticulado retangular 6 x 5.

Planos básicos de reticulados quadrados e reticulados retangulares

Reticulado quadrado 3 x 3 balanceado

Bloco	Rep. 1		
(1)	1	2	3
(2)	4	5	6
(3)	7	8	9

	Rep. 2		
(4)	1	4	7
(5)	2	5	8
(6)	3	6	9

	Rep. 3		
(7)	1	5	9
(8)	7	2	0
(9)	4	8	3

	Rep. 4		
(10)	1	8	6
(11)	4	2	9
(12)	7	5	3

Reticulado quadrado 4 x 4 balanceado

Bloco	Rep. 1			
(1)	1	2	3	4
(2)	5	6	7	8
(3)	9	10	11	12
(4)	13	14	15	16

	Rep. 2			
(5)	1	5	9	13
(6)	2	6	10	14
(7)	3	7	11	15
(8)	4	8	12	16

	Rep. 3			
(9)	1	6	11	16
(10)	5	2	15	12
(11)	9	14	3	8
(12)	13	10	7	4

	Rep. 4			
(13)	1	14	7	12
(14)	13	2	11	8
(15)	5	10	3	16
(16)	9	6	15	4

	Rep. 5			
(17)	1	10	15	8
(18)	9	2	7	16
(19)	13	6	3	12
(20)	5	14	11	4

Reticulado quadrado 5 x 5 balanceado

Bloco	Rep. 1				
(1)	1	2	3	4	5
(2)	6	7	8	9	10
(3)	11	12	13	14	15
(4)	16	17	18	19	20
(5)	21	22	23	24	25

	Rep. 2				
(6)	1	6	11	16	21
(7)	2	7	12	17	22
(8)	3	8	13	18	23
(9)	4	9	14	19	24
(10)	5	10	15	20	25

	Rep. 3				
(11)	1	7	13	19	25
(12)	21	2	8	14	20
(13)	16	22	3	9	15
(14)	11	17	23	4	10
(15)	6	12	18	24	5

	Rep. 4				
(16)	1	12	23	9	20
(17)	16	2	13	24	10
(18)	6	17	3	14	25
(19)	21	7	18	4	15
(20)	11	22	8	19	5

	Rep. 5				
(21)	1	17	8	24	15
(22)	11	2	18	9	25
(23)	21	12	3	19	10
(24)	6	22	13	4	20
(25)	16	7	23	14	5

	Rep. 6				
(26)	1	22	18	14	10
(27)	6	2	23	19	15
(28)	11	7	3	24	20
(29)	16	12	8	4	25
(30)	21	17	13	9	5

Reticulado quadrado balanceado 7 x 7

Rep. 1

Bloco							
(1)	1	2	3	4	5	6	7
(2)	8	9	10	11	12	13	14
(3)	15	16	17	18	19	20	21
(4)	22	23	24	25	26	27	28
(5)	29	30	31	32	33	34	35
(6)	36	37	38	39	40	41	42
(7)	43	44	45	46	47	48	49

Rep. 2

(8)	1	8	15	22	29	36	43
(9)	2	9	16	23	30	37	44
(10)	3	10	17	24	31	38	45
(11)	4	11	18	25	32	39	46
(12)	5	12	19	26	33	40	47
(13)	6	13	20	27	34	41	48
(14)	7	14	21	28	35	42	49

Rep. 3

(15)	1	9	17	25	33	41	49
(16)	43	2	10	18	26	34	42
(17)	36	44	3	11	19	27	35
(18)	29	37	45	4	12	20	28
(19)	22	30	38	46	5	13	21
(20)	15	23	31	39	47	6	14
(21)	8	16	24	32	40	48	7

Rep. 4

(22)	1	37	24	11	47	34	21
(23)	15	2	38	25	12	48	35
(24)	29	16	3	39	26	13	49
(25)	43	30	17	4	40	27	14
(26)	8	44	31	18	5	41	28
(27)	22	9	45	32	19	6	42
(28)	36	23	10	46	33	20	7

Rep. 5

(29)	1	30	10	39	19	48	28
(30)	22	2	31	11	40	20	49
(31)	43	23	3	32	12	41	21
(32)	15	44	24	4	33	13	42
(33)	36	16	45	25	5	34	14
(34)	8	37	17	46	26	6	35
(35)	29	9	38	18	47	27	7

Rep. 6

(36)	1	23	45	18	40	13	35
(37)	29	2	24	46	19	41	14
(38)	8	30	3	25	47	20	42
(39)	36	9	31	4	26	48	21
(40)	15	37	10	32	5	27	49
(41)	43	16	38	11	33	6	28
(42)	22	44	17	39	12	34	7

Rep. 7

(43)	1	16	31	46	12	27	42
(44)	36	2	17	32	47	13	28
(45)	22	37	3	18	33	48	14
(46)	8	23	38	4	19	34	49
(47)	43	9	24	39	5	20	35
(48)	29	44	10	25	40	6	21
(49)	15	30	45	11	26	41	7

Rep. 8

(50)	1	44	38	32	26	20	14
(51)	8	2	45	39	33	27	21
(52)	15	9	3	46	40	34	28
(53)	22	16	10	4	47	41	35
(54)	29	23	17	11	5	48	42
(55)	36	30	24	18	12	6	49
(56)	43	37	31	25	19	13	7

Reticulado quadrado balanceado 8 x 8

Rep. 1

Bloco								
(1)	1	2	3	4	5	6	7	8
(2)	9	10	11	12	13	14	15	16
(3)	17	18	19	20	21	22	23	24
(4)	25	26	27	28	29	30	31	32
(5)	33	34	35	36	37	38	39	40
(6)	41	42	43	44	45	46	47	48
(7)	49	50	51	52	53	54	55	56
(8)	57	58	59	60	61	62	63	64

Rep. 2

(9)	1	9	17	25	33	41	49	57
(10)	2	10	18	26	34	42	50	58
(11)	3	11	19	27	35	43	51	59
(12)	4	12	20	28	36	44	52	60
(13)	5	13	21	29	37	45	53	61
(14)	6	14	22	30	38	46	54	62
(15)	7	15	23	31	39	47	55	63
(16)	8	16	24	32	40	48	56	64

Rep. 3

(17)	1	10	19	28	37	46	55	64
(18)	9	2	51	44	61	30	23	40
(19)	17	50	3	36	29	62	15	48
(20)	25	42	35	4	21	14	63	56
(21)	33	58	27	20	5	54	47	16
(22)	41	26	59	12	53	6	39	24
(23)	49	18	11	60	45	38	7	32
(24)	57	34	43	52	13	22	31	8

Rep. 4

(25)	1	18	27	44	13	62	39	56
(26)	17	2	35	60	53	46	31	16
(27)	25	34	3	12	45	54	23	64
(28)	41	58	11	4	29	22	55	40
(29)	9	50	43	28	5	38	63	24
(30)	57	42	51	20	37	6	15	32
(31)	33	26	19	52	61	14	7	48
(32)	49	10	59	36	21	30	47	8

Rep. 5

(33)	1	26	43	60	21	54	15	40
(34)	25	2	11	52	37	62	47	24
(35)	41	10	3	20	61	38	31	56
(36)	57	50	19	4	45	30	39	16
(37)	17	34	59	44	5	14	55	32
(38)	49	58	35	28	13	6	23	48
(39)	9	42	27	36	53	22	7	64
(40)	33	18	51	12	29	46	63	8

Rep. 6

(41)	1	34	11	20	53	30	63	48
(42)	33	2	59	28	45	22	15	56
(43)	9	58	3	52	21	46	39	32
(44)	17	26	51	4	13	38	47	64
(45)	49	42	19	12	5	62	31	40
(46)	25	18	43	36	61	6	55	16
(47)	57	10	35	44	29	54	7	24
(48)	41	50	27	60	37	14	23	8

Rep. 7

(49)	1	42	59	52	29	38	23	16
(50)	41	2	19	36	13	54	63	32
(51)	57	18	3	28	53	14	47	40
(52)	49	34	27	4	61	46	15	24
(53)	25	10	51	60	5	22	39	48
(54)	33	50	11	44	21	6	31	64
(55)	17	58	43	12	37	30	7	56
(56)	9	26	35	20	45	62	55	8

Rep. 8

(57)	1	50	35	12	61	22	47	32
(58)	49	2	43	20	29	14	39	64
(59)	33	42	3	60	13	30	55	24
(60)	9	18	59	4	37	54	31	48
(61)	57	26	11	36	5	46	23	56
(62)	17	10	27	52	45	6	63	40
(63)	41	34	51	28	21	62	7	16
(64)	25	58	19	44	53	38	15	8

Rep. 9

(65)	1	58	51	36	45	14	31	24
(66)	57	2	27	12	21	38	55	48
(67)	49	26	3	44	37	22	63	16
(68)	33	10	43	4	53	62	23	32
(69)	41	18	35	52	5	30	15	64
(70)	9	34	19	60	29	6	47	56
(71)	25	50	59	20	13	46	7	40
(72)	17	42	11	28	61	54	39	8

Reticulado quadrado balanceado 9 x 9

Rep. 1

Bloco									
(1)	1	2	3	4	5	6	7	8	9
(2)	10	11	12	13	14	15	16	17	18
(3)	19	20	21	22	23	24	25	26	27
(4)	28	29	30	31	32	33	34	35	36
(5)	37	38	39	40	41	42	43	44	45
(6)	46	47	48	49	50	51	52	53	54
(7)	55	56	57	58	59	60	61	62	63
(8)	64	65	66	67	68	69	70	71	72
(9)	78	74	75	76	77	78	79	80	81

Rep. 2

(10)	1	10	19	28	37	46	55	64	73
(11)	2	11	20	29	38	47	56	65	74
(12)	3	12	21	30	39	48	57	66	75
(13)	4	13	22	31	40	49	58	67	76
(14)	5	14	23	32	41	50	59	68	77
(15)	6	15	24	33	42	51	60	69	78
(16)	7	16	25	34	43	52	61	70	79
(17)	8	17	26	35	44	53	62	71	80
(18)	9	18	27	36	45	54	63	72	81

Rep. 3

(19)	1	20	12	58	77	69	34	53	45
(20)	10	2	21	67	59	78	43	35	54
(21)	19	11	3	76	68	60	52	44	36
(22)	28	47	39	4	23	15	61	80	72
(23)	37	29	48	13	5	24	70	62	81
(24)	46	38	30	22	14	6	79	71	63
(25)	55	74	66	31	50	42	7	26	18
(26)	64	56	75	40	32	51	16	8	27
(27)	73	65	57	49	41	33	25	17	9

Rep. 4

(28)	1	11	21	31	41	51	61	71	81
(29)	19	2	12	49	32	42	79	62	72
(30)	10	20	3	40	50	33	70	80	63
(31)	55	65	75	4	14	24	34	44	54
(32)	73	56	66	22	5	15	52	35	45
(33)	64	74	57	13	23	6	43	53	36
(34)	28	38	48	58	68	78	7	17	27
(35)	46	29	39	76	59	69	25	8	18
(36)	37	47	30	67	77	60	16	26	9

Rep. 5

(37)	1	29	57	22	50	78	16	44	72
(38)	55	2	30	76	23	51	70	17	45
(39)	28	56	3	49	77	24	43	71	18
(40)	10	38	66	4	32	60	25	53	81
(41)	64	11	39	58	5	33	79	26	54
(42)	37	65	12	31	59	6	52	80	27
(43)	19	47	75	13	41	69	7	35	63
(44)	73	20	48	67	14	42	61	8	36
(45)	46	74	21	40	68	15	34	62	9

Rep. 6

(46)	1	56	30	13	68	42	25	80	54
(47)	28	2	57	40	14	69	52	26	81
(48)	55	29	3	67	41	15	79	53	27
(49)	19	74	48	4	59	33	16	71	45
(50)	46	20	75	31	5	60	43	17	72
(51)	73	47	21	58	32	6	70	44	18
(52)	10	65	39	22	77	51	7	62	36
(53)	37	11	66	49	23	78	34	8	63
(54)	64	38	12	76	50	24	61	35	9

Bloco	Rep. 7								
(55)	1	47	66	76	14	33	43	62	27
(56)	64	2	48	31	77	15	25	44	63
(57)	46	65	3	13	32	78	61	26	45
(58)	37	56	21	4	50	69	79	17	36
(59)	19	38	57	67	5	51	34	80	18
(60)	55	20	39	49	68	6	16	35	81
(61)	73	11	30	40	59	24	7	53	72
(62)	28	74	12	22	41	60	70	8	54
(63)	10	29	75	58	23	42	52	71	9

	Rep. 8								
(64)	1	74	39	67	32	24	52	17	63
(65)	37	2	75	22	68	33	61	53	18
(66)	73	38	3	31	23	69	16	62	54
(67)	46	11	57	4	77	42	70	35	27
(68)	55	47	12	40	5	78	25	71	36
(69)	10	56	48	76	41	6	34	26	72
(70)	64	29	21	49	14	60	7	80	45
(71)	19	65	30	58	50	15	43	8	81
(72)	28	20	66	13	59	51	79	44	9

	Rep. 9								
(73)	1	65	48	40	23	60	79	35	18
(74)	46	2	66	58	41	24	16	80	36
(75)	64	47	3	22	59	42	34	17	81
(76)	73	29	12	4	68	51	43	26	63
(77)	10	74	30	49	5	69	61	44	27
(78)	28	11	75	67	50	6	25	62	45
(79)	37	20	57	76	32	15	7	71	54
(80)	55	38	21	13	77	33	52	8	72
(81)	19	56	39	31	14	78	70	53	9

	Rep. 10								
(82)	1	38	75	49	59	15	70	26	36
(83)	73	2	39	13	50	60	34	71	27
(84)	37	74	3	58	14	51	25	35	72
(85)	64	20	30	4	41	78	52	62	18
(86)	28	65	21	76	5	42	16	53	63
(87)	19	29	66	40	77	6	61	17	54
(88)	46	56	12	67	23	33	7	44	81
(89)	10	47	57	31	68	24	79	8	45
(90)	55	11	48	22	32	69	43	80	9

Reticulado quadrado tríplice 6 x 6

Bloco	Rep. 1					
(1)	1	2	3	4	5	6
(2)	7	8	9	10	11	12
(3)	13	14	15	16	17	18
(4)	19	20	21	22	23	24
(5)	25	26	27	28	29	30
(6)	31	32	33	34	35	36

	Rep. 2					
(7)	1	7	13	19	25	31
(8)	2	8	14	20	26	32
(9)	3	9	15	21	27	33
(10)	4	10	16	22	28	34
(11)	5	11	17	23	29	35
(12)	6	12	18	24	30	36

	Rep. 3					
(13)	1	8	15	22	29	36
(14)	31	2	9	16	23	30
(15)	25	32	3	10	17	24
(16)	19	26	33	4	11	18
(17)	13	20	27	34	5	12
(18)	7	14	21	28	35	6

Reticulado quadrado tríplice 10 x 10

Rep. 1

Bloco										
(1)	1	2	3	4	5	6	7	8	9	10
(2)	11	12	13	14	15	16	17	18	19	20
(3)	21	22	23	24	25	26	27	28	29	30
(4)	31	32	33	34	35	36	37	38	39	40
(5)	41	42	43	44	45	46	47	48	49	50
(6)	51	52	53	54	55	56	57	58	59	60
(7)	61	62	63	64	65	66	67	68	69	70
(8)	71	72	73	74	75	76	77	78	79	80
(9)	81	82	83	84	85	86	87	88	89	90
(10)	91	92	93	94	95	96	97	98	99	100

Rep. 2

(11)	1	11	21	31	41	51	61	71	81	91
(12)	2	12	22	32	42	52	62	72	82	92
(13)	3	13	23	33	43	53	63	73	83	93
(14)	4	14	24	34	44	54	64	74	84	94
(15)	5	15	25	35	45	55	65	75	85	95
(16)	6	16	26	36	46	56	66	76	86	96
(17)	7	17	27	37	47	57	67	77	87	97
(18)	8	18	28	38	48	58	68	78	88	98
(19)	9	19	29	39	49	59	69	79	89	99
(20)	10	20	30	40	50	60	70	80	90	100

Rep. 3

(21)	1	12	23	34	45	56	67	78	89	100
(22)	91	2	13	24	35	46	57	68	79	90
(23)	81	92	3	14	25	36	47	58	69	80
(24)	71	82	93	4	15	26	37	48	59	70
(25)	61	72	83	94	5	16	27	38	49	60
(26)	51	62	73	84	95	6	17	28	39	50
(27)	41	52	63	74	85	96	7	18	29	40
(28)	31	42	53	64	75	86	97	8	19	30
(29)	21	32	43	54	65	76	87	98	9	20
(30)	11	22	33	44	55	66	77	88	99	10

Reticulado quadrado quádruplo 12 x 12

Bloco	Rep. 1											
(1)	1	2	3	4	5	6	7	8	9	10	11	12
(2)	13	14	15	16	17	18	19	20	21	22	23	24
(3)	25	26	27	28	29	30	31	32	33	34	35	36
(4)	37	38	39	40	41	42	43	44	45	46	47	48
(5)	49	50	51	52	53	54	55	56	57	58	59	60
(6)	61	62	63	64	65	66	67	68	69	70	71	72
(7)	73	74	75	76	77	78	79	80	81	82	83	84
(8)	85	86	87	88	89	90	91	92	93	94	95	96
(9)	97	98	99	100	101	102	103	104	105	106	107	108
(10)	109	110	111	112	113	114	115	116	117	118	119	120
(11)	121	122	123	124	125	126	127	128	129	130	131	132
(12)	133	134	135	136	137	138	139	140	141	142	143	144

	Rep. 2											
(13)	1	13	25	37	49	61	73	85	97	109	121	133
(14)	2	14	26	38	50	62	74	86	98	110	122	134
(15)	3	15	27	39	51	63	75	87	99	111	123	135
(16)	4	16	28	40	52	64	76	88	100	112	124	136
(17)	5	17	29	41	53	65	77	89	101	113	125	137
(18)	6	18	30	42	54	66	78	90	102	114	126	138
(19)	7	19	31	43	55	67	79	91	103	115	127	139
(20)	8	20	32	44	56	68	80	92	104	116	128	140
(21)	9	21	33	45	57	69	81	93	105	117	129	141
(22)	10	22	34	46	58	70	82	94	106	118	130	142
(23)	11	23	35	47	59	71	83	95	107	119	131	143
(24)	12	24	36	48	60	72	84	96	108	120	132	144

Rep. 3

(25)	1	14	27	40	57	70	83	96	101	114	127	140
(26)	2	13	28	39	58	69	84	95	102	113	128	139
(27)	3	16	25	38	59	72	81	94	103	116	125	138
(28)	4	15	26	37	60	71	82	93	104	115	126	137
(29)	5	18	31	44	49	62	75	88	105	118	131	144
(30)	6	17	32	43	50	61	76	87	106	117	132	143
(31)	7	20	29	42	51	64	73	86	107	120	129	142
(32)	8	19	30	41	52	63	74	85	108	119	130	141
(33)	9	22	35	48	53	66	79	92	97	110	123	136
(34)	10	21	36	47	54	65	80	91	98	109	124	135
(35)	11	24	33	46	55	68	77	90	99	112	121	134
(36)	12	23	34	45	56	67	78	89	100	111	122	133

Rep. 4

(37)	1	24	30	43	53	64	82	95	105	116	122	135
(38)	2	23	29	44	54	63	81	96	106	115	121	136
(39)	3	22	32	41	55	62	84	93	107	114	124	133
(40)	4	21	31	42	56	61	83	94	108	113	123	134
(41)	5	16	34	47	57	68	74	87	97	120	126	139
(42)	6	15	33	48	58	67	73	88	98	119	125	140
(43)	7	14	36	45	59	66	76	85	99	118	128	137
(44)	8	13	35	46	60	65	75	86	100	117	127	138
(45)	9	20	26	39	49	72	78	91	101	112	130	143
(46)	10	19	25	40	50	71	77	92	102	111	129	144
(47)	11	18	28	37	51	70	80	89	103	110	132	141
(48)	12	17	27	38	52	69	79	90	104	109	131	142

Reticulado retangular 3 x 4

Bloco	Rep. X		
X1	1	2	3
X2	4	5	6
X3	7	8	9
X4	10	11	12

	Rep. Y		
Y1	4	7	10
Y2	1	8	11
Y3	2	5	12
Y4	3	6	9

	Rep. Z		
Z1	6	8	12
Z2	2	9	10
Z3	3	4	11
Z4	1	5	7

Reticulado retangular 4 x 5

Bloco	Rep. X			
X1	1	2	3	4
X2	5	6	7	8
X3	9	10	11	12
X4	13	14	15	16
X5	17	18	19	20

	Rep. Y			
Y1	5	9	13	17
Y2	1	10	14	18
Y3	2	6	15	19
F4	3	7	11	20
F5	4	8	12	16

	Rep. Z			
Z1	8	11	15	18
Z2	2	9	16	20
Z3	4	7	14	17
Z4	1	5	12	19
Z5	3	6	10	13

Reticulado retangular 5 x 6

Bloco	Rep. X				
X1	1	2	3	4	5
X2	6	7	8	9	10
X3	11	12	13	14	15
X4	16	17	18	19	20
X5	21	22	23	24	25
X6	26	27	28	29	30

	Rep. Y				
Y1	6	11	16	21	26
Y2	1	12	17	22	27
Y3	2	7	18	23	28
Y4	3	8	13	24	29
Y5	4	9	14	19	30
Y6	5	10	15	20	25

	Rep. Z				
Z1	7	13	19	25	27
Z2	5	14	16	23	29
Z3	1	8	20	21	30
Z4	2	9	15	22	26
Z5	3	10	11	17	28
Z6	4	6	12	18	24

Reticulado retangular 6 x 7

Bloco	Rep. Z					
X1	1	2	3	4	5	6
X2	7	8	9	10	11	12
X3	13	14	15	16	17	18
X4	19	20	21	22	23	24
X5	25	26	27	28	29	30
X6	31	32	33	34	35	36
X7	37	38	39	40	41	42

	Rep. Y					
Y1	7	13	19	25	31	37
Y2	1	14	20	26	32	38
Y3	2	8	21	27	33	39
Y4	3	9	15	28	34	40
Y5	4	10	16	22	35	41
Y6	5	11	17	23	29	42
Y7	6	12	18	24	30	36

	Rep. Z					
Z1	12	17	22	28	33	38
Z2	2	13	24	29	35	40
Z3	4	9	20	25	36	42
Z4	6	11	16	27	32	37
Z5	1	7	18	23	34	39
Z6	3	8	14	19	30	41
Z7	5	10	15	21	26	31

Reticulado retangular 7 x 8

Rep. X

Bloco							
X1	1	2	3	4	5	6	7
X2	8	9	10	11	12	13	14
X3	15	16	17	18	19	20	21
X4	22	23	24	25	26	27	28
X5	29	30	31	32	33	34	35
X6	36	37	38	39	40	41	42
X7	43	44	45	46	47	48	49
X8	50	51	52	53	54	55	56

Rep. Y

Y1	8	15	22	29	36	43	50
Y2	1	16	23	30	37	44	51
Y3	2	9	24	31	38	45	52
Y4	3	10	17	32	39	46	53
Y5	4	11	18	25	40	47	54
Y6	5	12	19	26	33	48	55
Y7	6	13	20	27	34	41	56
Y8	7	14	21	28	35	42	49

Rep. Z

Z1	9	17	25	33	41	49	51
Z2	7	18	26	34	38	43	53
Z3	1	10	27	35	36	47	55
Z4	2	11	19	29	42	44	56
Z5	3	12	20	28	37	45	50
Z6	4	13	21	22	30	46	52
Z7	55	14	15	23	31	39	54
Z8	6	8	16	24	32	40	48

Reticulado retangular 8 x 9

Rep. X

Bloco								
X1	1	2	3	4	5	6	7	8
X2	9	10	11	12	13	14	15	16
X3	17	18	19	20	21	22	23	24
X4	25	26	27	28	29	30	31	32
X5	33	34	35	36	37	38	39	40
X6	41	42	43	44	45	46	47	48
X7	49	50	51	52	53	54	55	56
X8	57	58	59	60	61	62	63	64
X9	65	66	67	68	69	70	71	72

Rep. Y

Y1	9	17	25	33	41	49	57	65
Y2	1	18	26	34	42	50	58	66
Y3	2	10	27	35	43	51	59	67
Y4	3	11	19	36	44	52	60	68
Y5	4	12	20	28	45	53	61	69
Y6	5	13	21	29	37	54	62	70
Y7	6	14	22	30	38	46	63	71
Y8	7	15	23	31	39	47	55	72
Y9	8	16	24	32	40	48	56	64

Rep. Z

Z1	16	23	30	37	45	52	59	66
Z2	2	17	32	39	46	54	61	68
Z3	4	11	26	33	48	55	63	70
Z4	6	13	20	35	42	49	64	72
Z5	8	15	22	29	44	51	58	65
Z6	1	9	24	31	38	53	60	67
Z7	3	10	18	25	40	47	62	69
Z8	5	12	19	27	34	41	56	71
Z9	7	14	21	28	36	43	50	57

Reticulado retangular 9 x 10

Rep. X

Bloco									
X1	1	2	3	4	5	6	7	8	9
X2	10	11	12	13	14	15	16	17	18
X3	19	20	21	22	23	24	25	26	27
X4	28	29	30	31	32	33	34	35	36
X5	37	38	39	40	41	42	43	44	45
X6	46	47	48	49	50	51	52	53	54
X7	55	56	57	58	59	60	61	62	63
X8	64	65	66	67	68	69	70	71	72
X9	73	74	75	76	77	78	79	80	81
X10	82	83	84	85	86	87	88	89	90

Rep. Y

Y1	10	19	28	37	46	55	64	73	82
Y2	1	20	29	38	47	56	65	74	83
Y3	2	11	30	39	48	57	66	75	84
Y4	3	12	21	40	49	58	67	76	85
Y5	4	13	22	31	50	59	68	77	86
Y6	5	14	23	32	41	60	69	78	87
Y7	6	15	24	33	42	51	70	79	88
Y8	7	16	25	34	43	52	61	80	89
Y9	8	17	26	35	44	53	62	71	90
Y10	9	18	27	36	45	54	63	72	81

Rep. Z

Z1	11	21	31	41	51	61	71	81	83
Z2	9	22	32	42	52	62	64	76	84
Z3	1	12	33	43	53	63	68	73	87
Z4	2	13	23	44	54	55	65	79	89
Z5	3	14	24	34	46	56	72	75	90
Z6	4	15	25	35	45	57	67	74	82
Z7	5	16	26	36	37	47	66	77	85
Z8	6	17	27	28	38	48	58	78	86
Z9	7	18	19	29	39	49	59	69	88
Z10	8	10	20	30	40	50	60	70	80

Capítulo 7

Blocos aumentados de Federer (BAF)

Este delineamento foi desenvolvido por Federer (1956b), por ocasião de sua estada no Havaí. Foi preparado para ser aplicado na cultura da cana-de-açúcar. Mais tarde expandiu-se para outros cultivos de multiplicação assexuada e de plantas perenes e, posteriormente, para culturas anuais de multiplicação sexuada. O esquema consiste na formação de blocos completos com os materiais testemunhas e a inclusão nos blocos de linhagens novas a serem testadas, e cada linhagem aparece em um e somente um bloco, daí o seu nome. Esse esquema pode também ser aplicado nos delineamentos quadrado latino, inteiramente casualizado ou em qualquer outro delineamento (FEDERER, 1956a; MARTINEZ; GARZA, 1972). Os tratamentos testemunhas serão aqueles que permitirão a estimativa do erro.

Um exemplo pequeno com quatro blocos, quatro testemunhas (T1 a T4) e 12 linhagens (L1 a L12) encontra nas Tabelas 1 e 2. A casualização dos tratamentos aos blocos pode ser feita de duas maneiras. A primeira sugere o sorteio inicial das testemunhas em cada bloco e, em seguida, a casualização das linhagens nas parcelas remanescentes de cada bloco (Tabela 1). A segunda forma de se casualizarem os tratamentos consiste em fazer a casualização do ensaio como se fosse em blocos completos casualizados e, então, substituir os tratamentos não testemunhas pelas linhagens de forma ordenada (Tabela 2). Os tratamentos de 1 a 4 representam as testemunhas (T1 a T4).

Tabela 1. Esquema de casualização em campo de ensaio em blocos aumentados de Federer com quatro testemunhas e 12 linhagens.

T1	L12	T4	T3	L6	L2	T2
L4	T2	T3	L5	T1	L1	T4
T4	T3	L8	L3	T2	T1	L11
L10	T1	T2	L7	T4	T3	L9

Tabela 2. Esquema de casualização em campo de ensaio em blocos aumentados de Federer com quatro testemunhas e 12 linhagens, considerando-se o ensaio como blocos completos casualizados.

5(L1)	1	3	4	6(L2)	2	7(L3)
1	5(L4)	4	6(L5)	7(L6)	3	2
6(L8)	3	2	7(L9)	4	1	5(L7)
4	7(L12)	1	5(L10)	2	3	6(L11)

O BAF possui grande utilidade quando a disponibilidade de sementes é pequena, ou quando há problemas de área ou de mão de obra, já que economiza em ambas as situações. É de particular importância nas etapas iniciais de um processo de melhoramento de plantas quando o número de linhagens costuma ser muito alto.

Zimmermann (1997) mostrou que 200 linhagens avaliadas em BAF exigiriam 47,6% do número total de parcelas necessárias para o teste de 192 linhagens em reticulado quadrado 14 x 14, ou 44,4% para 206 linhagens em reticulado retangular 14 x 15, ou ainda 45,8% para se testarem as mesmas 200 linhagens em blocos completos casualizados com três repetições.

No entanto, este delineamento apresenta o seguinte inconveniente: cada parcela perdida com material em teste (não testemunha) implica a falta de informação sobre a linhagem. Esse defeito, entretanto, é minimizado pelo grande número de materiais em teste. Ademais, há ainda o fato de que, possivelmente, muitas linhagens no processo final de desenvolvimento sejam irmãs.

A comparação entre outros esquemas de delineamentos e o BAF mostra que as vantagens mais evidentes são as seguintes: os blocos têm seu tamanho mais limitado, portanto há maior possibilidade de haver homogeneidade de ambiente; o uso de mais de uma testemunha, alocadas nos blocos de forma aleatorizada, permite uma estimativa, não viciada, dos seus efeitos; o BAF tem um número adequado de graus de liberdade para a estimação do erro experimental.

A principal desvantagem deste delineamento é o fato de apresentar quatro variâncias diferentes para as diversas formas de comparação de médias, que são: entre testemunhas; entre linhagens no mesmo bloco; entre linhagens em blocos diferentes e entre testemunhas e linhagens. Como, no entanto, o melhorista usualmente não está interessado em comparar suas testemunhas entre si, e nem as linhagens entre si, mas sim em diferenciar as linhas que superaram ou que foram superadas pelas testemunhas, essa desvantagem é praticamente de efeito nulo para fins práticos de melhoramento. Pode ser adicionado ainda que, em muitos casos, neste estágio do programa – gerações iniciais – o melhorista

não está nem mesmo interessado em diferenças estatisticamente significativas, mas em ter boas estimativas da capacidade produtiva e de outras características da planta que o permitam selecionar com vantagem sobre as testemunhas.

Análise de variância e esperança dos quadrados médios

A análise de variância para o BAF apresenta a formação dada na Tabela 3 e reflete o modelo matemático associado a este delineamento.

Tabela 3. Análise de variância para o delineamento de blocos aumentados de Federer.

Fonte de variação	Graus de liberdade (GL)
Blocos (ignorando tratamentos)	B - 1
Tratamentos (ajustados)	$V + L$ - 1
Testemunhas	V - 1
Linhagens	L - 1
Testemunhas x linhagens	1
Erro	$(B$ - 1) $(V$ - 1)
Total	$BV + L$ - 1

As somas de quadrados total e de blocos (ignorando tratamentos) são calculadas da forma usual, isto é, são chamadas de Y_{ij} as observações feitas em cada parcela, e de $Y_{.j}$ os totais de blocos, em que *j* representa o número de blocos e *i* o número total de tratamentos (testemunhas mais linhagens). Dessa forma, tem-se:

$$SQTotal = \sum Y_{ij}^2 - FC \quad [1]$$

$$SQBlo\cos = \frac{1}{I}\sum Y_{.j}^2 - FC \quad [2]$$

$$FC = \frac{1}{IJ}\left(\sum Y_{ij}\right)^2 \quad [3]$$

Para o cálculo das somas de quadrados de tratamentos ajustados e suas partições, é necessário primeiro fazer o ajuste dos valores obtidos para as linhagens em teste. Isso é feito somando-se a cada valor de linhagem obtido o valor dado pela diferença MT_j - MT, em que MT_j representa a média das testemunhas no bloco j onde se encontra a linhagem, e MT a média geral das testemunhas em todo o ensaio. Sendo T_i o total da testemunha i e L_k o valor ajustado de cada linhagem, a soma de quadrados de tratamentos ajustada é dada por:

$$SQTrat.aj. = \frac{1}{J}\sum T_i^2 + \sum L_k^2 - FC \quad [4]$$

Para o caso dos blocos aumentados de Federer, as esperanças dos quadrados médios, assumindo-se que apenas os efeitos das linhagens são aleatórios, são mostradas na Tabela 4, em que k_1 é uma constante que depende do número de testemunhas e de linhagens em teste, além do número de blocos empregados.

Tabela 4. Esperança matemática dos quadrados médios, $E(QM)$, no delineamento de blocos aumentados de Federer.

Fonte de variação	Grau de liberdade (GL)	E(QM)
Blocos (ignorando tratamentos)	B - 1	
Tratamentos (ajustados)	$V + L$ -1	
Testemunhas	V - 1	
Linhagens	L - 1	$\sigma^2 + k_1\sigma^2_g$
Testemunhas x linhagens	1	
Erro	$(B$ - 1) $(V$ - 1)	σ^2
Total	$BV + L$ - 1	

Exemplo de aplicação

O pesquisador Joaquim Geraldo Cáprio da Costa, em seus trabalhos de melhoramento de feijão, ao trabalhar com linhagens do grupo preto, realizou um experimento em uma estação experimental em Ponta Grossa, PR, com plantio realizado em fevereiro de 2000. O experimento tinha um desenho de blocos aumentados de Federer, com 18 blocos, cada um com quatro testemunhas (tratamentos 1 a 4) e 12 linhagens, num total de 216 linhagens. A Tabela 5 reproduz os dados de apenas dez desses blocos com nove tratamentos em cada um – as quatro testemunhas e cinco linhagens. Os dados referem-se à produção de grãos em kg ha^{-1}.

Tabela 5. Dados de produção de grãos de feijão de um experimento em blocos aumentados de Federer.

											Total
Bloco 1	Tratamento	9	5	8	3	4	1	2	6	7	
	Produção	761	1.516	1.692	1.563	2.019	2.003	1.671	1.245	1.203	13.673
Bloco 2	Tratamento	21	20	2	3	4	1	17	18	19	
	Produção	1.536	1.888	1.838	1.818	2.210	1.866	1.786	1.756	2.037	16.735
Bloco 3	Tratamento	30	1	29	31	2	4	3	32	33	
	Produção	1.052	1.695	1.080	1.443	1.472	1.745	1.680	1.123	1.632	12.922
Bloco 4	Tratamento	2	41	42	45	43	44	3	1	4	
	Produção	1.113	1.357	1.219	1.135	1.076	1.384	1.516	1.773	1.881	12.454
Bloco 5	Tratamento	3	57	1	2	4	53	55	54	56	
	Produção	1.959	1.503	2.085	1.454	1.837	1.873	1.243	1.212	1.509	14.675
Bloco 6	Tratamento	67	65	4	68	1	69	3	2	66	
	Produção	1.414	1.279	2.077	1.586	1.846	1.956	1.775	1.728	1.546	15.207
Bloco 7	Tratamento	80	2	79	3	78	81	1	77	4	
	Produção	2.213	1.930	2.653	2.054	2.361	1.812	1.849	1.885	1.835	18.592
Bloco 8	Tratamento	90	91	89	2	3	92	1	4	93	
	Produção	1.145	1.194	1.920	1.415	1.589	1.138	1.840	2.062	1.418	13.721
Bloco 9	Tratamento	3	1	102	101	105	104	4	103	2	
	Produção	1.894	1.695	1.885	1.433	1.803	1.451	1.975	1.879	1.675	15.690
Bloco 10	Tratamento	115	114	4	117	2	113	1	3	116	
	Produção	2.356	1.450	2.191	2.351	1.687	2.180	1.757	1.879	1.605	17.456
											151.125

Como visto na seção anterior, o fator de correção, a soma de quadrados de blocos e a soma de quadrados total são obtidas da forma usual.

$$FC = \frac{(151.125)^2}{90} = 253.764.062,5$$

$$SQBlocos = \frac{1}{9}(13.673^2 + 16.375^2 + ... + 17.456^2) - FC = 4.071.067,39$$

$$SQBlocos = (761^2 + 1.516^2 + ... + 1.605^2) - FC = 11.147.428,50$$

Para a soma de quadrados de tratamentos ajustada, é necessário primeiro calcular a média geral de todas as testemunhas para o ensaio e para cada bloco, já que o fator de correção da produção das linhagens em teste é a diferença entre a média das testemunhas no ensaio e a média das testemunhas no bloco em que está a linhagem. A Tabela 6 mostra todos esses valores.

Tabela 6. Fatores de correção das linhagens segundo os efeitos das testemunhas em cada bloco.

	Média das testemunhas no ensaio	Média das testemunhas no bloco	Fator de correção das linhagens
Bloco 1	1.798,775	1.814,000	15,225
Bloco 2	1.798,775	1.933,000	134,225
Bloco 3	1.798,775	1.648,000	-150,775
Bloco 4	1.798,775	1.570,750	-228,025
Bloco 5	1.798,775	1.833,750	34,975
Bloco 6	1.798,775	1.856,500	57,725
Bloco 7	1.798,775	1.917,000	118,225
Bloco 8	1.798,775	1.726,500	-72,275
Bloco 9	1.798,775	1.809,750	10,975
Bloco 10	1.798,775	1.878,500	79,725

Assim, a linhagem de número 5, no bloco 1, tem como produção ajustada o valor dado por: 1.516 - 15,225 = 1.500,775; a linhagem 30, no bloco 3, tem a produção ajustada igual a 1.052 - (-150,775) = 1.202,775, e assim por diante. A Tabela 7 dá o total das testemunhas e o valor ajustado de cada linhagem.

Tabela 7. Valores de produção total das testemunhas e de produção ajustada das linhagens em um experimento em blocos aumentados de Federer.

Bloco 1	5	6	7	8	9
	1.500,775	1.229,775	1.187,775	1.676,775	745,775
Bloco 2	17	18	19	20	21
	1.651,775	1.621,775	1.902,775	1.753,775	1.401,775
Bloco 3	29	30	31	32	33
	1.230,775	1.202,775	1.593,775	1.273,775	1.782,775
Bloco 4	41	42	43	44	45
	1.585,025	1.447,025	1.304,025	1.612,025	1.363,025
Bloco 5	53	54	55	56	57
	1.838,025	1.177,025	1.208,025	1.474,025	1.468,025
Bloco 6	65	66	67	68	69
	1.221,275	1.488,275	1.356,275	1.528,275	1.898,275
Bloco 7	77	78	79	80	81
	1.766,775	2.242,775	2.534,775	2.094,775	1.693,775
Bloco 8	89	90	91	92	93
	1.992,275	1.217,275	1.266,275	1.210,275	1.490,275
Bloco 9	101	102	103	104	105
	1.422,025	1.874,025	1.868,025	1.440,025	1.792,025
Bloco 10	113	114	115	116	117
	2.100,275	1.370,275	2.276,275	1.525,275	2.271,275
Testemunhas	1	2	3	4	
	18.409	15.983	17.727	19.832	

Dessa forma, a soma de quadrados de tratamentos ajustada é dada por:

$$SQTrat.ajustada = \frac{1}{10}(18.409^2 + ... + 19.832^2) +$$
$$(1.500,775^2 + ... + 2.271,275^2) - FC = 6.403.550,136$$

O quadro de análise de variância encontra-se na Tabela 8.

Tabela 8. Análise de variância para um experimento em blocos aumentados de Federer, com quatro testemunhas e 50 linhagens de feijão. Dados de produção de grãos em kg ha^{-1}.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (PR > F)
Blocos (ignorando tratamentos)	9	4.071.067,389	452.340,821	18,15 (< ,0001)
Tratamentos (ajustados)	53	6.403.550,136	120.821,701	4,85 (< ,0001)
Testemunhas	3	766.572,275	255.524,092	10,25 (< 0,0001)
Linhagens	49	4.606.934,816	94.019,078	3,77 (0,0002)
Testamentos x linhagens	1	1.030.043,045	1.030.043,045	41,34 (< ,0001)
Erro	27	672.810,97	24.918,92	
Total	89	11.147.428,50		

Os desdobramentos da soma de quadrados de tratamentos ajustada se fazem da forma usual, com o uso das formas de soma de quadrados indicada para elas. Tem-se, então,

$$SQtestemunhas = \frac{1}{10}(\sum T^2) - \frac{1}{40}(\sum T)^2 =$$

$$\frac{1}{10}(18.409^2 + ... + 19.832^2) - \frac{1}{40}(18.409 + ... + 19.832)^2$$

em que T representa o total de cada uma das testemunhas, 10 o número de repetições para cada testemunha, e

$$SQ(TestxLinh) = \frac{1}{40}(\sum T)^2 + \frac{1}{50}(\sum L)^2 - \frac{1}{90}(\sum T + \sum L)^2 =$$

$$\frac{1}{40}(71.951)^2 + \frac{1}{50}(79.174)^2 - \frac{1}{90}(151.125)^2$$

em que T tem o significado dado na fórmula anterior e L representa o total ajustado de cada uma das linhagens. Finalmente, para linhagens a forma é:

$$SQLinhagens = (\sum L^2) - \frac{1}{50}(\sum L)^2 = (1.500,775^2 + ... + 1.525,275^2) - \frac{1}{50}(79.174)^2$$

É possível fazer a comparação de médias das linhagens com cada uma das testemunhas, com a melhor delas ou com a média delas. O problema maior dessas comparações é que as linhagens e as testemunhas ou sua média têm variância diferente, sendo sempre menor a variância das testemunhas. Esse problema, entretanto, é minimizado ao considerar que os pesquisadores, quando usam esse delineamento em seus trabalhos, estão em fases muito preliminares, com número elevado de novos materiais em teste, e o erro que se possa cometer em razão das variâncias diferentes não é significativo diante do grande número desses materiais. Em outras palavras, selecionar alguns materiais ruins ou perder alguns poucos bons não afeta o resultado final da pesquisa, que deverá ainda passar por muitas outras etapas.

Para um contraste de médias que inclua uma linhagem e a média das testemunhas, a estimativa da variância desse contraste é dada por

$$V(Contraste) = QME * \left(1 + \frac{1}{IJ}\right) \quad [5]$$

em que QME é o quadrado médio do erro, I o número de testemunhas e J o número de blocos.

No exemplo que se está apresentando, tem-se, então,

$$V(Contraste) = 24.918{,}92 * \left(1 + \frac{1}{4 * 10}\right) = 25.541{,}893$$

e a diferença mínima significativa (*DMS*) pelo teste t é igual a

$$DMS = 2{,}052 * (\sqrt{25.541{,}893}) = 327{,}95$$

Portanto, qualquer linhagem que difira da média das testemunhas por um valor maior que o valor do *DMS* será considerada diferente (melhor ou pior) que a média das testemunhas. Nesse caso em particular, as linhagens 78, 79, 115 e 117 são superiores às testemunhas, e as de número 6, 7, 9, 21, 29, 30, 32, 42, 43, 45, 54, 55, 57, 65, 67, 90, 91, 92, 101, 104 e 114 são inferiores à média das testemunhas. Verifica-se assim que somente quatro linhagens seriam selecionadas como mais produtivas.

Uso do SAS

Para se conseguir fazer no SAS a análise completa e exata de um ensaio em blocos aumentados de Federer, existe a necessidade de um pouco de programação extra. A programação SAS para este delineamento foi desenvolvida pelo pesquisador Valdenir Queiroz Ribeiro da Embrapa Meio Norte a quem se agradece a permissão de uso.

Como foi visto anteriormente, os tratamentos (aqui chamados de TRAT) dividem-se em dois grupos, um chamado de testemunhas ou tratamentos comuns (TC) e outro de tratamentos não comuns (TNC) ou linhagens. Deve-se então dividir esses tratamentos nos dois grupos, criando variáveis que permitam a obtenção das fontes de variação, como se observou no item *Exemplo de aplicação*. Logo, após a leitura dos dados, cujas variáveis seriam então BLOCO TRAT e PRODUCAO, segue-se a criação das novas variáveis de classificação, como se vê na primeira parte da programação SAS a seguir:

```
DATA BAF;INPUT BLOCO TRAT PRODUCAO;
IF TRAT LE 4 THEN TC_TNC=1;ELSE TC_TNC=0;
TNC=TRAT*(TC_TNC=0);TC=TRAT*(TC_TNC=1);
CARDS;
```

Após a leitura dos dados, a programação SAS, com o uso do procedimento GLM, tem a seguinte forma:

```
PROC GLM;CLASS BLOCO TC_TNC TNC TC;
MODEL ISD=BLOCO TC_TNC TNC(TC_TNC) TC(TC_TNC);
LSMEANS TC_TNC TNC(TC_TNC) TC(TC_TNC)/OUT=MED1;
```

Para obter a listagem das médias dos tratamentos em ordem sequencial, sem especificação de comuns e não comuns, a programação SAS tem a seguinte forma:

```
DATA MED1;SET MED1;
IF TC_TNC=0 AND TNC=. AND TC=. THEN delete;
IF TC_TNC=1 AND TNC=. AND TC=. THEN delete;
IF TC_TNC=1 AND TNC=0 AND TC=. THEN delete;
IF TC_TNC=0 AND TNC=. AND TC=0 THEN delete;
IF TC_TNC=1 AND TNC=. THEN TNC=TC;TRAT=TNC;
MEDIA_AJ=LSMEAN;KEEP VAR MEDIA_AJ;
PROC SORT;BY TRAT;PROC PRINT;
```

O resultado de tal procedimento aparece na Figura 1.

The GIM Procedure
Class Level Information

Class	Levels	Values
BLOCO	10	1 2 3 4 5 6 7 8 9 10
TC_TNC	2	0 1
TNC	51	0 5 6 7 8 9 17 18 19 20 21 29 30 31 32 33 41 42 43 44 45 53 54 55 56 57 65 66 67 68 69 77 78 79 80 81 89 90 91 92 93 101 102 103 104 105 113 114 115 116 117
TC	5	0 1 2 3 4

Number of observations 90

The GIM Procedure

Dependent Variable: PRODUCAO

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	62	10474617.53	168945.44	6.78	<.0001
Error	27	672810.97	24918.92		
Corrected Total	89	11147428.50			

R-Square	Coeff Var	Root MSE	PRODUCAO Mean
0.939644	9.400931	157.8573	1679.167

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCO	9	4071067.389	452340.821	18.15	<.0001
TC_TNC	1	1030043.045	1030043.045	41.34	<.0001
TNC(TC_TNC)	49	4606934.816	94019.078	3.77	0.0002
TC(TC_TNC)	3	766572.275	255524.092	10.25	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCO	9	492837.725	54759.747	2.20	0.0551
TC_TNC	1	1030043.045	1030043.045	41.34	<.0001
TNC(TC_TNC)	49	4606934.816	94019.078	3.77	0.0002
TC(TC_TNC)	3	766572.275	255524.092	10.25	0.0001

The GIM Procedure
Least Squares Means

TC_TNC	PRODUCAO LSMEAN
0	1583.48000
1	1798.77500

TNC	TC_TNC	PRODUCAO LSMEAN
5	0	1500.77500
6	0	1229.77500
7	0	1187.77500
8	0	1676.77500
9	0	745.77500
17	0	1651.77500
18	0	1621.77500
19	0	1902.77500
20	0	1753.77500
21	0	1401.77500
29	0	1230.77500
30	0	1202.77500
31	0	1593.77500

32	0	1273.77500
33	0	1782.77500
41	0	1585.02500
42	0	1447.02500
43	0	1304.02500
44	0	1612.02500
45	0	1363.02500
53	0	1838.02500
54	0	1177.02500
55	0	1208.02500
56	0	1474.02500
57	0	1468.02500
65	0	1221.27500
66	0	1488.27500
67	0	1356.27500
68	0	1528.27500
69	0	1898.27500
77	0	1766.77500
78	0	2242.77500
79	0	2534.77500
80	0	2094.77500
81	0	1693.77500
89	0	1992.27500
90	0	1217.27500
91	0	1266.27500
92	0	1210.27500
93	0	1490.27500
101	0	1422.02500
102	0	1874.02500
103	0	1868.02500
104	0	1440.02500
105	0	1792.02500
113	0	2100.27500
114	0	1370.27500
115	0	2276.27500
116	0	1525.27500
117	0	2271.27500
0	1	1798.77500

TC	TC_TNC	PRODUCAO LSMEAN
0	0	1583.48000
1	1	1840.90000
2	1	1598.30000
3	1	1772.70000
4	1	1983.20000

Obs	TRAT	MEDIA_AJ
1	1	1840.90
2	2	1598.30
3	3	1772.70
4	4	1983.20
5	5	1500.78
6	6	1229.78
7	7	1187.78
8	8	1676.78
9	9	745.78

10	17	1651.78
11	18	1621.78
12	19	1902.78
13	20	1753.78
14	21	1401.78
15	29	1230.78
16	30	1202.78
17	31	1593.78
18	32	1273.78
19	33	1782.78
20	41	1585.03
21	42	1447.03
22	43	1304.03
23	44	1612.03
24	45	1363.03
25	53	1838.03
26	54	1177.03
27	55	1208.03
28	56	1474.03
29	57	1468.03
30	65	1221.28
31	66	1488.28
32	67	1356.28
33	68	1528.28
34	69	1898.28
35	77	1766.78
36	78	2242.78
37	79	2534.78
38	80	2094.78
39	81	1693.78
40	89	1992.28
41	90	1217.28
42	91	1266.27
43	92	1210.27
44	93	1490.28
45	101	1422.03
46	102	1874.03
47	103	1868.03
48	104	1440.03
49	105	1792.03
50	113	2100.28
51	114	1370.28
52	115	2276.28
53	116	1525.28
54	117	2271.28

Figura 1. Análise de um experimento em blocos aumentados de Federer, usando procedimento GLM do SAS.

Para uniformizar a leitura dos resultados com o exemplo dado no item *Exemplo de aplicação*, TC_TNC corresponde ao contraste linhagens x testemunhas ou tratamentos não comuns x tratamentos comuns; TNC(TC_TNC), ao efeito das linhagens ou tratamentos não comuns; TC(TC_TNC), ao efeito das testemunhas ou tratamentos comuns; e BLOCO, ao

efeito de bloco não ajustado. Essa mesma interpretação ocorre nas médias chamadas pelo SAS de LSMEANS. A última lista de MEDIAS_AJ corresponde às médias dos 54 tratamentos testados, às quatro testemunhas (1 a 4) e às 50 linhagens (5 a 117). Para a comparação das médias, qualquer processo poderia ser usado, manualmente, pois, com o SAS, apenas o teste t seria aplicável com o uso da opção LSMEANS. Recorda-se uma vez mais que esses dados são apenas parte dos dados do experimento original.

Capítulo 8

Ensaios fatoriais

Nos primórdios da investigação agrícola, os estudos eram realizados para um fator de cada vez, mantendo-se constantes os demais. A partir dos resultados obtidos para esse fator, seu melhor resultado era incorporado como um dos fatores fixos, e outro fator era selecionado para estudo. Isso levava a ciclos de repetição e de ajustes aos melhores resultados de cada um dos fatores, até que as modificações encontradas fossem consideradas irrelevantes. Nas áreas de pesquisa de microbiologia, química e entre outras, em que cada ensaio pode durar algumas horas ou alguns dias, tal procedimento, apesar de oneroso e trabalhoso, permitia ao pesquisador determinar seus resultados em pouco tempo. Mas para o caso da agricultura e da pecuária e de outras áreas de conhecimento, em que cada ciclo de ensaio pode levar meses ou até anos, essa técnica é considerada completamente inviável.

Em um clássico trabalho sobre experimentos fatoriais, Yates (1937) os definiu como aqueles que incluem todas as combinações de vários fatores. No entanto, já em 1933, discutia os princípios de ortogonalidade e confundimento em experimentos repetidos. Os princípios de confundimento permitem que esses ensaios fatoriais sejam executados em blocos incompletos (Capítulo 10) e em réplicas fracionadas (Capítulo 11), em que só algumas das combinações são testadas.

De uma maneira geral, quando se fala de fatoriais não se está falando em delineamentos experimentais, e sim em delineamento de tratamentos. No entanto, como será visto no Capítulo 9, existem algumas modificações nos delineamentos básicos que só podem ser aplicadas aos ensaios fatoriais.

Chama-se de fator àquilo que se quer testar e de níveis às suas diferentes manifestações. Por exemplo, em estudos de adubação de plantas, três elementos – nitrogênio (N), fósforo (P) e potássio (K) – são considerados os macronutrientes. Assim, cada elemento é considerado um fator, e suas diferentes doses de aplicação os níveis. Se apenas um elemento for testado num ensaio, cada nível será chamado de tratamento. Por sua vez, se dois ou os três elementos forem testados, cada combinação entre os seus níveis é que será declarada como um tratamento.

Os fatores podem ser quantitativos, quando seus níveis se referem a doses, espaçamentos entre plantas, ou qualquer outra manifestação que possa ser mensurável, e quando, em geral, se está interessado em determinar um ponto ótimo e, portanto, um estudo de regressão associado. Quando os níveis de um fator não são mensuráveis (ex.: cultivares), o fator é dito qualitativo. Quando os níveis de um fator são sempre os mesmos para todos os níveis de outro(s) fator(es), o ensaio é chamado simplesmente de fatorial. Mas quando eles se manifestam de forma diferente para cada nível de outro fator, o ensaio é denominado de aninhado. Exemplos deste último caso são os testes de cultivares agrupadas por ciclo vegetativo, em que para cada ciclo se utilizam cultivares distintas. Outros exemplos seriam o estudo de fontes e procedências de fontes de adubos; cidades e supermercados de cada cidade; no caso de ensaios em pecuária, os genitores e seus descendentes, e vários outros.

Os fatoriais são mais comumente representados por uma potência (b^k) indicando que os k fatores possuem todos b níveis. Quando os diversos fatores apresentam números diferentes de níveis, isso se indica por meio de produtos de potências. Por exemplo, se o experimento conta com k fatores com b níveis e l fatores com c níveis, a representação é b^k x c^l. Os ensaios exploratórios, em geral, apresentam b com valor 2, enquanto os não exploratórios podem apresentar valores de b iguais a 3, 4, 5 ou maior.

Os fatores são usualmente representados por letras maiúsculas (A, B, C, ...), e cada um de seus níveis pelas correspondentes minúsculas acompanhadas de um índice indicador desse nível. Outro tipo de notação, útil para o caso de apenas dois níveis, é o uso apenas da letra minúscula, em que a presença da letra indica o nível mais alto do fator, e sua ausência, o valor mais baixo, enquanto a combinação de todos os fatores em nível mais baixo é representada por (1). Uma terceira forma de representação é a que utiliza apenas os números entre 0 e b, em que 0 significa ausência ou nível mais baixo, ou ainda a testemunha, e b o mais alto dos níveis.

Um experimento com três fatores, A, B e C, em 2 níveis cada um, por exemplo, teria seus oito tratamentos identificados por:

$$a_0b_0c_0,\ a_0b_1c_0,\ a_0b_0c_1,\ a_0b_1c_1,\ a_1b_0c_0,\ a_1b_1c_0,\ a_1b_0c_1,\ a_1b_1c_1$$

ou

$$(1),\ b,\ c,\ bc,\ a,\ ab,\ ac,\ abc$$

ou

$$000,\ 010,\ 001,\ 011,\ 100,\ 110,\ 101,\ 111.$$

Esta última notação é aquela que adotaremos neste livro. Note-se que a primeira e terceira formas de representação podem ser adotadas para qualquer número de fatores e de níveis, enquanto a segunda forma é válida apenas para fatoriais de base 2, independentemente do número de fatores.

Na análise de variância, a notação a ser empregada será a de representar os efeitos principais pela sua letra maiúscula, e os efeitos de interação pelas combinações entre as letras correspondentes unidas por um x. Nas fórmulas de soma de quadrados, as letras maiúsculas serão acompanhadas pelos índices adequados, mas para as interações desaparecerá o x.

Os ensaios fatoriais apresentam como grandes vantagens no que se refere às possibilidades de estudo dos efeitos dos fatores e das suas interações em um tempo muito reduzido, além de permitir o ajuste de superfícies de resposta no caso de fatores quantitativos. A grande e principal desvantagem é o crescimento muito rápido do número de tratamentos, quando se usam muitos fatores ou muitos níveis para os fatores, o que pode tornar impraticável a execução do ensaio. Essa desvantagem pode ser reduzida com o uso de confundimento ou de repetições fracionadas. Nesses casos, o custo a ser pago é a impossibilidade de estimação de alguns efeitos, geralmente as interações maiores, e a estimação simultânea de efeitos, respectivamente (Capítulos 10 e 11).

Um novo conceito, o de interação, é introduzido nos ensaios fatoriais. Define-se interação como a ação conjunta ou sinérgica de dois ou mais fatores. De forma geral, pode-se considerar uma interação como positiva, negativa ou nula. Ela é dita positiva, quando o efeito conjunto de dois fatores é maior que a soma de seus efeitos individuais, negativa quando for menor que essa soma e nula quando igualar a soma dos efeitos próprios. A Figura 1 ilustra esse conceito.

	B	0	1
A	0	10	15
	1	20	40

Interação positiva

0	1
10	15
20	20

Interação negativa

0	1
10	15
20	25

Interação nula

Figura 1. Exemplos numéricos de interação entre dois fatores.

Verifica-se aí que a mudança de nível 0 para 1 no fator A ocasiona um incremento de dez unidades, enquanto a mesma alteração em B ocasiona aumento de cinco unidades. A alteração simultânea de 0 para 1 em A e B leva ao aumento de 30 unidades no caso da interação positiva ($30 > 5 + 10$), de 10 unidades na negativa ($10 < 5 + 10$) e de 15 na interação

nula (15 = 5 + 10). A interação nula indica também não haver diferença na magnitude das respostas, enquanto as interações positiva e nula podem indicar alteração nas magnitudes das respostas ou até modificação do sentido das respostas.

Uma interação não significativa indica a independência entre os fatores, o que resulta em uma grande economia de tempo e esforço, além de levar usualmente a uma maior precisão na estimação dos efeitos dos fatores individuais pelo mais elevado número de repetições a que se submete cada um. A presença de interação significativa implica maior esforço por parte do pesquisador em detalhar mais a análise de seus dados.

Como afirmado anteriormente, fatoriais podem ser conduzidos em qualquer um dos delineamentos experimentais vistos até aqui, ou seja, completamente casualizado, blocos completos ao acaso, quadrado latino e mesmo em blocos incompletos ou reticulados. No entanto, no decorrer deste capítulo, adotar-se-á sempre o desenho em blocos completos por ser o desenho de maior ocorrência.

Análise de variância em ensaios fatoriais

Os modelos matemáticos para os ensaios fatoriais seguem os dos desenhos experimentais adotados, com a diferença que os efeitos t_i serão agora representados pelos efeitos principais (fatores) e por suas interações, e o índice i por tantos índices quantos forem os fatores em estudo. Por exemplo, no caso de três fatores, o índice i seria substituído por i, k e l (reservando-se o j para os blocos), e a letra t por A, B e C, de maneira que o modelo original

$$Y_{ij} = m + b_j + t_i + e_{ij} \quad [1]$$

é agora substituído por

$$Y_{ijkl} = m + b_j + A_i + B_k + C_l + AB_{ik} + AC_{il} + BC_{kl} + ABC_{ikl} + e_{ijkl} \quad [2]$$

De forma semelhante se procede para qualquer número de fatores em qualquer outro desenho experimental.

Percebe-se então que a análise de variância de um ensaio fatorial refere-se ao estudo detalhado dos efeitos de tratamentos, em que os graus de liberdade a eles associados são desdobrados em conjuntos menores de graus de liberdade ou até ao estudo de cada um deles individualmente. Deve ser lembrado que o teste F, feito para tratamentos e com mais

de um grau de liberdade, reflete o efeito médio de cada um dos graus de liberdade (GL) envolvidos e, portanto, das muitas comparações individuais envolvidas. Se apenas umas poucas dessas comparações for real, é possível que o teste F, para tratamentos, por causa das muitas não reais, não as detecte, por isso seu estudo individualizado deve sempre ser realizado.

De uma maneira geral, existem tantos efeitos principais quantos são os fatores envolvidos no estudo. As interações simples (entre dois fatores) são iguais às combinações entre eles tomados dois a dois; enquanto as interações duplas (entre três fatores) se igualam às combinações tomadas três a três, e assim por diante. O modelo dado na fórmula 2 demonstra isso de maneira muito clara: três fatores, logo três efeitos principais; três interações simples (combinações de três tomados dois a dois) e uma interação dupla (combinação de três, tomados três a três).

O efeito de um fator (principal) é medido pelo total de seus efeitos considerados sobre todos os outros fatores e repetições de forma que cada nível de um fator, A, por exemplo, no modelo dado na fórmula 2, é igual a

$$A_i = \sum_{jkl} Y_{ijkl} \quad \text{ou} \quad A_i = \sum_{kl} a_i b_k c_l \qquad [3]$$

As interações serão medidas pelos totais das combinações dos fatores em consideração sobre todos os outros fatores e repetições, isto é,

$$AB_{ik} = \sum_{jl} Y_{ijkl} \quad \text{ou} \quad AB_{ik} = \sum_{l} a_i b_k c_l \qquad [4]$$

e assim sucessivamente até a interação de todos os fatores em estudo. Nas segundas formas apresentadas, as somas se efetuam sobre todas as repetições.

As somas de quadrados para efeitos principais são obtidas da forma usual empregada para blocos e/ou tratamentos. Portanto, para o efeito principal A, no caso acima, sua soma de quadrados é dada por

$$SQA = \frac{1}{JKL} \sum_{i} A_i^2 \text{ - } \frac{1}{IJKL} (Y_{ijkl})^2 \qquad [5]$$

em que a parte subtraída é o fator de correção (FC).

No caso das interações, para a obtenção das somas de quadrados deve-se também subtrair todas as somas de quadrados dos efeitos maiores envolvidos. Para a soma de quadrados de uma interação simples (A x B), devem ser subtraídas as somas de quadrados dos fatores principais envolvidos (A e B); para uma interação dupla (A x B x C), devem ser subtraídas aquelas correspondentes aos fatores principais (A, B e C) e às interações simples (A x B, A x C e B x C), e assim sucessivamente, de maneira que a expressão para calcular a interação simples A x B, por exemplo, é dada por

$$SQAB = \frac{1}{JL}\sum_{ik} AB_{ik}^2 - FC - SQA - SQB \quad [6]$$

e para uma interação dupla (A x B x C) seria dada por

$$SQABC = \frac{1}{J}\sum_{ik} ABC_{ikl}^2 - FC - SQA - SQB - SQAB - SQAC - SQBC$$

As Tabelas 1 e 2 mostram as análises de variância para ensaios fatoriais com dois e três fatores, em blocos completos.

Tabela 1. Análise de variância de um fatorial com dois fatores, em blocos completos ao acaso, nos quais se omitiram os quadrados médios.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrado (SQ)
Blocos	$J - 1$	$SQR = \frac{1}{IK}\sum_{j} R_j^2 - \frac{1}{IJK}(Y_{ijk})^2$
Fator A	$I - 1$	$SQA = \frac{1}{JK}\sum_{i} A_i^2 - \frac{1}{IJK}(Y_{ijk})^2$
Fator B	$K - 1$	$SQB = \frac{1}{IJ}\sum_{k} B_k^2 - \frac{1}{IJK}(Y_{ijk})^2$
Interação A x B	$(I - 1)(K - 1)$	$SQAB = \frac{1}{J}\sum_{ik} AB_{ik}^2 - FC - SQA - SQB$
Erro	$(J - 1)(IK - 1)$	Por diferença
Total	$IJK - 1$	$SQTo = \sum_{ijk} Y_{ijk}^2 - \frac{1}{IJK}(Y_{ijk})^2$

Tabela 2. Análise de variância de um fatorial com três fatores, em blocos completos ao acaso, nos quais se omitiram os quadrados médios.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrado (SQ)
Blocos	$J - 1$	$SQR = \frac{1}{IKL}\sum_{j} R_j^2 - \frac{1}{IJKL}(Y_{ijkl})^2$
Fator A	$I - 1$	$SQA = \frac{1}{JKL}\sum_{i} A_i^2 - \frac{1}{IJKL}(Y_{ijkl})^2$
Fator B	$K - 1$	$SQB = \frac{1}{IJL}\sum_{k} B_k^2 - \frac{1}{IJKL}(Y_{ijkl})^2$
Interação A x B	$(I - 1)(K - 1)$	$SQAB = \frac{1}{JL}\sum_{ik} AB_{ik}^2 - FC - SQA - SQB$
Fator C	$L - 1$	$SQC = \frac{1}{IJK}\sum_{l} C_l^2 - \frac{1}{IJKL}(Y_{ijkl})^2$
Interação A x C	$(I - 1)(L - 1)$	$SQAC = \frac{1}{JK}\sum_{il} AC_{il}^2 - FC - SQA - SQC$
Interação B x C	$(IK - 1)(L - 1)$	$SQBC = \frac{1}{JI}\sum_{kl} BC_{kl}^2 - FC - SQB - SQC$
Interação A x B x C	$(I - 1)(K - 1)(L - 1)$	$SQABC = \frac{1}{J}\sum_{ikl} ABC_{ikl}^2 - FC - SQA - SQB - SQC - SQAB - SQAC - SQBC$
Erro	$(J - 1)(IKL - 1)$	Por diferença
Total	$IJKL - 1$	$SQTo = \sum_{ijkl} Y_{ijkl}^2 - \frac{1}{IJKL}(Y_{ijkl})^2$

Estudo dos graus de liberdade individuais

O estudo dos graus de liberdade individualizados para cada efeito principal ou interação se faz, usualmente, por meio de contrastes ortogonais que, no caso específico de apenas dois níveis, coincidem suas somas de quadrados com aquelas dadas na Tabela 2, como será visto em exemplos de aplicação mais adiante. Quanto se trabalha com fatores com mais de dois níveis, deve se considerar se os fatores são quantitativos ou qualitativos. No primeiro caso, o mais usual é o emprego dos contrastes para polinômios ortogonais para linha reta ou de

primeira ordem; quadrática ou de segunda ordem, etc. Quando se trata de fatores qualitativos, os contrastes a serem considerados são os ditos contrastes de interesse que devem ser definidos pelo pesquisador de acordo com as informações adicionais que dispõe sobre cada fator. Para as interações, os contrastes são obtidos a partir dos contrastes dos efeitos principais.

Fatores com níveis igualmente espaçados

Este caso é aquele no qual a diferença (q) entre duas doses ou níveis consecutivos é constante. Os níveis tanto podem ser representados por seus valores reais como codificados, que, nesse caso, referem-se ao quociente entre seu valor e a constante q. Assim, ao testar-se a aplicação de quatro doses de N, por exemplo, distanciadas em 50 kg ha^{-1} de princípio ativo, tanto se poderia representar as doses como 0, 50, 100 e 150 kg ha^{-1} de N, como por 0, 1, 2 e 3. Os contrastes a serem aplicados para o estudo dos graus de liberdade individualizados são aqueles referentes aos polinômios ortogonais que permitem a estimação de equações de regressão de primeiro grau (linear), segundo grau ou equação quadrática, de terceiro grau ou cúbica, etc.

A Tabela 3 dá os valores desses polinômios para o estudo de fatores com até seis níveis, que são os casos mais comuns, sendo extremamente raras as pesquisas que envolvam fatores com sete ou mais níveis. De qualquer forma, os valores dos coeficientes dos polinônmios são calculados a partir de fórmulas bem conhecidas. Após a tabela são mostradas essas fórmulas, até o quinto grau, e identificadas como P_{1i}, P_{2i}, P_{3i}, P_{4i} e P_{5i}, as quais podem ser usadas para qualquer número de níveis.

$$P_{1i} = x_i \; ; \quad P_{2i} = x_i^2 - \frac{n^2 - 1}{12} \; ; \quad P_{3i} = x_i^3 - \left(\frac{3n^2 - 7}{20} \right) x_i \qquad [7]$$

$$P_{4i} = x_i^4 - \left(\frac{3n^2 - 13}{14} \right) x_i^2 + \frac{3(n^2 - 1)(n^2 - 9)}{560} \qquad [8]$$

$$P_{5i} = x_i^5 - \left(\frac{5(n^2 - 7)}{18} \right) x_i^3 + \left(\frac{15n^4 - 230n^2 + 470}{1008} \right) x_i \qquad [9]$$

em que n representa o número de níveis e

$$x_i = \frac{X_i - \bar{X}}{q} \qquad [10]$$

Tabela 3. Coeficientes de polinômios ortogonais para fatores até seis níveis igualmente espaçados.

Número de níveis	Ordem do polinômio	Valor dos coeficientes dos polinômios ortogonais (P_i) para os níveis em teste						$\sum_i P_i^2$	M
		0	1	2	3	4	5		
2	Linear	-1	1					2	2
3	Linear	-1	0	1				2	1
	Quadrático	1	-2	1				4	3
4	Linear	-3	-1	1	3			20	2
	Quadrático	1	-1	-1	1			4	1
	Cúbico	-1	3	-3	1			20	10/3
5	Linear	-2	-1	0	1	2		10	1
	Quadrático	2	-1	-2	-1	2		14	1
	Cúbico	-1	2	0	-2	1		10	5/6
	4° Grau	1	-4	6	-4	1		70	35/12
6	Linear	-5	-3	-1	1	3	5	70	2
	Quadrático	5	-1	-4	-4	-1	5	84	3/2
	Cúbico	-5	7	4	-4	-7	5	180	5/3
	4° Grau	1	-3	2	2	-3	1	28	7/12
	5° Grau	-1	5	-10	10	-5	1	252	21/10

As somas de quadrados dos diferentes níveis de polinômios, cada um com um grau de liberdade, calculam-se como visto anteriormente para contrastes ortogonais (Capítulo 1), ou seja,

$$SQPolinômio = \frac{\sum_i (P_i T_i)^2}{r \sum_i P_i^2} \qquad [11]$$

e o valor do coeficiente de regressão associado é calculado por meio da fórmula

$$b_k = \frac{\sum_i (P_i T_i)^2}{r \sum_i P_i^2} \qquad [12]$$

A equação de regressão é então estimada como

$$Y_i - \bar{Y} = \sum_k b_k P_k M_k \quad [13]$$

em que o índice k representa o grau de polinômio e varia de 1 até o maior grau significativo; P_k representa a expressão polinomial dada ao final da tabela anterior e M_k corresponde às constantes também dadas na mesma tabela e $\bar{Y}$ o valor médio dos dados correspondentes.

Fatores com níveis não igualmente espaçados

Para o estudo dos graus de liberdade individuais quando os fatores quantitativos apresentam níveis não igualmente espaçados, os polinômios ortogonais anteriormente referidos não podem mais ser empregados. Existe então a necessidade de calcular os polinômios adequados para cada caso. Robson em 1959 apresentou uma fórmula recursiva [14] que permite o cálculo dos valores de cada polinômio.

$$f_h(X_i) = \frac{1}{c_h}\left[X_i^h - \sum_{j=0}^{h-1} f_j(X_i) \sum_{g=1}^{n} X_g^h \; f_j(X_g) \right] \quad [14]$$

em que c_h é dado pela expressão

$$c_h^2 = \sum_i \left[X_i^h - \sum_{j=0}^{h-1} f_j(X_i) \sum_{g=1}^{n} X_g^h \; f_j(X_g) \right]^2 \quad [15]$$

A aplicação dessas equações mostra que

$$c_0 = \sqrt{n} \quad [16]$$

e

$$f_0(X_1) = f_0(X_2) = \ldots f_0(X_n) = 1/\sqrt{n} \quad [17]$$

Além disso, mostra que o valor de c_1 é igual a

$$c_1 = \sqrt{\sum \left(X_i - \overline{X} \right)^2} \quad [18]$$

e

$$f_1(X_i) = \frac{1}{c_1}\left(X_i - \overline{X} \right) \quad [19]$$

e assim sucessivamente para c_2, f_2, etc.

Em estudos de adubação, é muito comum o uso de doses sucessivamente iguais ao dobro da dose anterior, por exemplo, 0, 50, 100, 200, 400, etc. A Tabela 4 contém os valores dos polinômios ortogonais para os casos de quatro e cinco níveis.

Tabela 4. Coeficientes dos polinômios ortogonais para tratamentos com níveis não igualmente espaçados (razão 2 entre doses).

Número de doses	Grau do polinômio	Valor das doses codificadas				
		0	1	2	4	8
4	Linear	-7	-3	1	9	
	Quadrático	7	-4	-8	5	
	Cúbico	-3	8	-6	1	
5	Linear	-3	-2	-1	1	5
	Quadrático	31	2	-19	-37	23
	Cúbico	-9	8	10	-11	2
	4° Grau	21	-64	56	-14	1

Fatores qualitativos

No caso de experimentos que envolvem fatores qualitativos, o estudo de graus de liberdade individuais se faz a partir dos chamados contrastes ortogonais de interesse. Nessa técnica, o pesquisador deve ter uma série de informações adicionais a cada um dos níveis para poder estruturar esses contrastes.

A melhor maneira de explicar essa técnica é por meio do uso de um exemplo. Suponha-se que um pesquisador está testando quatro cultivares de feijão de grão do tipo carioca, cujos nomes e informações adicionais constam da Tabela 5.

Tabela 5. Descrição de características de quatro cultivares de feijão do grupo comercial Carioca.

Cultivar	Cruzamento	Origem	Hábito de crescimento	Mancha-angular
Aporé (1)	Duplo	CNPAF	III	Resistente
IAPAR 81 (2)	Múltiplo	IAPAR	II	Susceptível
Pérola (3)	Duplo [Seleção d. (1)]	CNPAF	II/III	Intermediário
Rudá (4)	Simples	CIAT	II	Susceptível

Com quatro cultivares, existem logicamente três graus de liberdade que podem ser explorados individualmente, e isso se faz a partir de contrastes ortogonais. Ocorre que existe um grande número de conjuntos de três contrastes ortogonais que permitem essa exploração, e o problema de qual conjunto se deve usar depende das informações colaterais disponíveis. São essas informações que permitirão a construção dos contrastes de maior interesse. As Tabelas 6, 7 e 8 mostram bem o problema envolvido. As Tabelas 6 e 7 mostram dois conjuntos que poderiam ser de interesse para um melhorista de plantas, já que a primeira delas compara os diversos tipos de cruzamentos ou métodos de obtenção de novos materiais, enquanto a Tabela 7 explora as origens dos materiais.

Tabela 6. Contrastes ortogonais de interesse para estudo dos efeitos de métodos de cruzamento ou obtenção de novas linhagens/cultivares.

Contraste para tipos de cruzamento	(1)	(2)	(3)	(4)
1 - Simples x duplo	-1	0	-1	2
2 - (Simples e duplo) x múltiplo	-1	3	-1	-1
3 - (Seleção x original) dentro do duplo	-1	0	1	0

Tabela 7. Contrastes ortogonais de interesse para estudo dos efeitos de local de origem das linhagens/cultivares.

Contraste para origem	(1)	(2)	(3)	(4)
1 - Nacional x estrangeiro	-1	-1	-1	3
2 - CNPAF x Iapar	-1	2	-1	0
3 - Dentro do CNPAF	-1	0	1	0

Tabela 8. Contrastes ortogonais de interesse para estudo do efeito da reação das cultivares à mancha-angular.

Contraste para reação à mancha-angular	(1)	(2)	(3)	(4)
1 - Resistente x não resistentes	3	-1	-1	-1
2 - Susceptível x intermediário	0	-1	2	-1
3 - Entre susceptíveis	0	-1	0	1

A Tabela 8, que seria também de interesse para os fitopatologistas, explora os efeitos do tipo de reação de cada cultivar a uma doença comum em feijoeiro – a mancha-angular. Evidentemente que muitos outros contrastes de interesse poderiam ser obtidos, se novas informações colaterais fossem fornecidas. Nota-se que os contrastes da Tabela 8 seriam exatamente os mesmos para explorar os efeitos do hábito de crescimento das cultivares.

Os testes F na análise de variância

Uma dificuldade adicional nos ensaios fatoriais refere-se à aplicação do teste F, uma vez que seu uso correto dependerá das considerações sobre os fatores em estudo serem admitidos como fixos ou aleatórios.

No caso de todos os fatores serem assumidos como fixos, essa dificuldade desaparece e os testes de F serão sempre calculados em relação ao quadrado médio do erro. Quando todos, ou mesmo alguns, devem ser considerados aleatórios, a sua realização já é feita de forma diferente, pois o quadrado médio do erro pode agora não ser correto testador para os estimadores do teste F.

Um fator para ser considerado fixo ou aleatório, dependerá da forma de escolha dos seus níveis e basicamente da definição dada pelo pesquisador. Quando os níveis forem considerados uma amostra entre todos os possíveis valores de níveis para esse fator, então ele deverá ser definido como aleatório. No entanto, se os diferentes níveis forem claramente definidos pelo investigador de acordo com seus interesses, então ele deve ser considerado fixo. Exemplos disso podem ser dados: em competição de cultivares, se elas forem consideradas amostra representativa de um banco de germoplasma, o fator cultivar é aleatório, mas se elas forem selecionadas segundo algum critério (em uso na região, as melhores linhagens selecionadas no processo de melhoramento, etc.) então o fator cultivar deve ser definido como fixo. Em trabalhos que envolvem anos e/ou locais, estes devem usualmente ser considerados aleatórios. Fatores quantitativos, nos quais o investigador busca função de resposta em equações de regressão, definindo, portanto, os níveis a serem usados, devem ser considerados fixos.

As interações devem sempre ser consideradas aleatórias se pelo menos um de seus fatores componentes for aleatório.

As fontes de variação que forem consideradas aleatórias são, usualmente, consideradas como de distribuição normal com média 0 (zero) e variância σ^2, identificadas por um índice associado à fonte. Portanto, as esperanças dos seus quadrados médios mostrarão esse componente multiplicado pelo número de repetições com que cada um de seus níveis for testado. Quando as fontes de variação são consideradas fixas, a esperança dos quadrados mostrará isso pela inclusão de um componente quadrático dado pela soma dos quadrados dos efeitos dividida pelos graus de liberdade e multiplicada também pelo número de repetições de seus níveis. A seção *Obtenção das esperanças matemáticas dos quadrados médios* dá uma forma prática de obtenção desses componentes de variância.

Um teste F será considerado adequado quando os quadrados médios do numerador e do denominador tiverem valores esperados que difiram por apenas um componente, e os demais tenham todos coeficientes iguais. Quando isso não ocorre, é permitido fazer combinações entre fontes de variação (ZIMMERMANN, 1973).

A Tabela 9 mostra um exemplo de um ensaio fatorial, em blocos completos, com dois fatores (A e B), para as situações de fatores fixos, aleatórios e mistos (A – aleatório, B – fixo e interação A x B aleatória). Uma análise dessa tabela mostra que um teste F adequado para o caso de modelo fixo seria obtido para qualquer fonte, usando-se como comparador o quadrado médio do erro. Para o caso aleatório, a interação teria seu teste F feito com o quadrado médio do erro, enquanto ela, por sua vez, seria o testador adequado para os efeitos principais. No caso do modelo misto, o erro seria o testador adequado para o fator B e para a interação A x B, enquanto a interação seria o testador para o fator principal A.

Tabela 9. Valores esperados dos quadrados médios em um experimento fatorial, com dois fatores, em blocos completos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Valor esperado dos quadrados médios		
		Modelo fixo	Modelo aleatório	Modelo misto
Blocos	J - 1	$\sigma^2 + IK\sigma_b^2$	$\sigma^2 + IK\sigma_b^2$	$\sigma^2 + IK\sigma_b^2$
A	I - 1	$\sigma^2 + JK\Sigma a^2/(I - 1)$	$\sigma^2 + J\sigma_{AB}^2 + JK\sigma_A^2$	$\sigma^2 + J\sigma_{AB}^2 + JK\sigma_A^2$
B	K - 1	$\sigma^2 + IJ\Sigma b^2/(K - 1)$	$\sigma^2 + J\sigma_{AB}^2 + IJ\sigma_B^2$	$\sigma^2 + IJ\Sigma b^2/(K - 1)$
A x B	$(I - 1)(K - 1)$	$\sigma^2 + J\Sigma(ab)^2/(I - 1)(K - 1)$	$\sigma^2 + J\sigma_{AB}^2$	$\sigma^2 + J\sigma_{AB}^2$
Erro	$(J - 1)(IK - 1)$	σ^2	σ^2	σ^2

Fatorial 2 x 2 – exemplo de análise

Como exemplo, utiliza-se parte dos dados de um experimento conduzido em casa de vegetação pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão, realizado em 1999, cujos fatores eram a densidade de solo (1,12 g cm^{-3} e 1,39 g cm^{-3}) e doses de microelementos, 0 e 2 g por vaso de FTE-BR-12. A Tabela 10 mostra os resultados obtidos em três repetições, da matéria seca de plantas de arroz, transformados em logaritmo decimal enquanto a Tabela 11 mostra os valores totais dos fatores estudados e suas combinações, como uma tabela auxiliar da análise de variância. O esquema experimental adotado foi o de blocos completos ao acaso.

Tabela 10. Dados de matéria seca de feijão, transformados por logaritmo, em um ensaio fatorial 2 x 2, no delineamento de blocos completos ao acaso.[1]

Repetição	Densidade do solo (D)	Dose de microelementos (M)	Matéria seca (g) Log10(MS*1.000)	Total das repetições
1	0	0	2,9638	
1	0	2	3,4014	
1	2	0	3,6599	
1	2	2	3,6484	13,6735
2	0	0	3,3324	
2	0	2	2,9494	
2	2	0	3,8075	
2	2	2	3,5775	13,6668
3	0	0	3,3927	
3	0	2	2,9638	
3	2	0	3,9112	
3	2	2	3,6589	13,9266

[1] Dados cedidos pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão.

Tabela 11. Quadro dos valores totais dos fatores e suas combinações, necessários para o cálculo das somas de quadrados dos fatores e sua interação.[1]

	M = 0	M = 2	Soma
D = 0	9,6889	9,3146	19,0035
D = 2	11,3786	10,8848	22,2634
Soma	21,0675	20,1994	41,2669

[1] Dados cedidos pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão.

A análise de variância está apresentada na Tabela 12.

Tabela 12. Análise de variância para um ensaio fatorial 2 x 2 em blocos completos ao acaso.[1]

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	2	0,0110	0,0055	0,12 ns
Densidade (D)	1	0,8856	0,8856	20,08 (0,0042)
Microelemento (M)	1	0,0628	0,0628	1,42 ns
D x M	1	0,0012	0,0012	0,03 ns
Erro	6	0,2647	0,0441	
Total	11	1,2253		

[1] Dados cedidos pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão.

O coeficiente de variação foi igual a 6,11%.

As somas de quadrados, total e de repetições, foram obtidas da forma usual. Para densidade de solo, microelementos e sua interação, as somas de quadrados foram obtidas por:

$$SQDens = \frac{1}{6}(19{,}0035^2 + 22{,}2634^2) - \frac{1}{12}(41{,}2669)^2$$

$$SQMicro = \frac{1}{6}(21{,}0675^2 + 20{,}1994^2) - \frac{1}{12}(41{,}2669)^2$$

$$SQDensxMicro = \frac{1}{3}(9{,}6889^2 + 9{,}3146^2 + 11{,}3786^2 + 10{,}8848^2)$$
$$- \frac{1}{12}(41{,}2669)^2 - SQDens - SqMicro$$

Como se pode ver na Tabela 12, somente a densidade de solo tem efeito significativo sobre a produção de massa seca, com vantagem de melhor crescimento das plantas de arroz com maior densidade do solo.

Uso do SAS no fatorial 2 x 2 em blocos ao acaso

Este é um processo muito simples de ser realizado em SAS. O procedimento adequado é o ANOVA, e a especificação do modelo é direta. Por simplicidade de escrita, quando todos

os fatores e suas interações são de interesse, o SAS permite que se descreva o modelo apenas escrevendo o nome dos fatores unidos por uma barra vertical (|). Essa nomenclatura indica que todos os fatores simples e todas as interações entre eles são estimáveis. Assim, o programa é:

```
PROC ANOVA;CLASS REPETICAO DENSIDADE MICRO;
MODEL MAT_SEC=REPETICAO DENSIDADE|MICRO;
MEANS DENSIDADE|MICRO;
```

Não há necessidade de comparação de médias uma vez que os fatores se apresentam em apenas dois níveis, e o resultado do programa está detalhado na Figura 2.

```
                          The ANOVA Procedure
                       Class Level Information
                 Class          Levels    Values
                 REPETICAO           3    1 2 3
                 Densidade           2    0 2
                 MICRO               2    0 2

                   Number of observations    12

                          The ANOVA Procedure
Dependent Variable: MAT_SEC

                                    Sum of
 Source                 DF         Squares     Mean Square    F Value    Pr > F
 Model                   5      0.96053553      0.19210711       4.35    0.0508
 Error                   6      0.26467214      0.04411202
 Corrected Total        11      1.22520767

              R-Square     Coeff Var      Root MSE    MAT_SEC Mean
              0.783978      6.107421      0.210029        3.438908

 Source                 DF       Anova SS     Mean Square    F Value    Pr > F
 REPETICAO               2      0.01096671      0.00548336       0.12    0.8853
 Densidade               1      0.88557900      0.88557900      20.08    0.0042
 MICRO                   1      0.06279980      0.06279980       1.42    0.2778
 Densidade*MICRO         1      0.00119002      0.00119002       0.03    0.8749

                          The ANOVA Procedure
               Level of              -----------MAT_SEC-----------
               Densidade      N             Mean          Std Dev
               0              6       3.16725000       0.22942139
               2              6       3.71056667       0.12365904

                Level of              -----------MAT_SEC-----------
                MICRO          N             Mean          Std Dev
                0              6       3.51125000       0.35087428
                2              6       3.36656667       0.33070956

          Level of      Level of              -----------MAT_SEC-----------
          Densidade     MICRO          N             Mean          Std Dev
          0             0              3       3.22963333       0.23218429
          0             2              3       3.10486667       0.25690631
          2             0              3       3.79286667       0.12628746
          2             2              3       3.62826667       0.04427757
```

Figura 2. Resultado do programa SAS com ANOVA para um fatorial 2 x 2, em blocos completos casualizados.

Fatorial 3 x 3 – exemplo de análise

Para este caso, utiliza-se o mesmo experimento descrito no item *Fatorial 2 x 2 – exemplo de análise*, apenas que agora está sendo considerada a existência de três doses de densidade de solo (incluindo um valor intermediário de 1,26 g cm^{-3}) e de uma dose também intermediária de microelementos de 1 g por vaso de FTE-BR-12. Cada vaso continha três plantas de arroz e, adicionalmente aos microelementos, foi feita adubação com 4 g por vaso de formulado 5-30-15.

Para análise, considera-se a mesma variável do exemplo anterior, cujos valores estão mostrados na Tabela 13, enquanto na Tabela 14 são mostrados os valores auxiliares para estudo dos efeitos principais e interação entre fatores.

A análise de variância está reportada na Tabela 15, tendo sido utilizadas fórmulas correspondentes àquelas do exemplo anterior. Novamente se verifica que apenas densidade de solo teve efeito significativo, mas, como a interação com microelementos mostra um valor do teste de F pouco acima dos níveis de significância, é recomendado que essa interação seja estudada com um pouco mais de detalhe.

Como os microelementos não mostraram efeito significativo, deve-se então estudar seus efeitos dentro de cada uma das densidades de solo. Nesse caso, dada a característica de variável quantitativa e igualmente espaçada, o uso de polinômios ortogonais é recomendado. Para exemplificar seu uso, dentro da densidade mais baixa têm-se os contrastes mostrados a seguir. Nesse caso, os valores das matérias secas são oriundos da tabela de dupla entrada para densidade de solo e microelementos, e os coeficientes das tabelas de coeficientes de polinômios ortogonais oriundos da seção *Fatores com níveis igualmente espaçados*:

$$SQLinear(Md.D = 0) = \frac{(-1*9{,}6889 + 0*10{,}8926 + 1*9{,}3146)^2}{3*[(-1)^2 + (0)^2 + (1)^2]} = 0{,}02335568$$

e

$$SQQuadrático(Md.D = 0) = \frac{(1*9{,}6889 - 2*10{,}8926 + 1*9{,}3146)^2}{3*[(1)^2 + (-2)^2 + (1)^2]} = 0{,}42985555$$

De forma semelhante são obtidos os outros valores mostrados na Tabela 16, em que se verifica que existe um efeito quadrático para microelementos na densidade mais baixa e nenhum efeito nas demais densidades de solo.

A partir desse resultado, é possível então definir a equação de regressão de segundo grau para microelementos dentro da densidade de solo de 1,12 g cm^{-3}, como visto

Tabela 13. Dados de matéria seca de feijão em um ensaio fatorial 3 x 3, em blocos completos ao acaso, com três repetições.[1]

Repetição	Densidade	Microelemento	Logaritmo do peso da matéria seca
1	0	0	2,9638
2	0	0	3,3324
3	0	0	3,3927
1	0	1	3,5515
2	0	1	3,8698
3	0	1	3,4713
1	0	2	3,4014
2	0	2	2,9494
3	0	2	2,9638
1	1	0	3,6425
2	1	0	3,2810
3	1	0	3,7324
1	1	1	3,4886
2	1	1	3,4472
3	1	1	3,5575
1	1	2	3,2553
2	1	2	3,7753
3	1	2	3,8768
1	2	0	3,6599
2	2	0	3,8075
3	2	0	3,9112
1	2	1	3,7016
2	2	1	3,5611
3	2	1	3,1523
1	2	2	3,6484
2	2	2	3,5775
3	2	2	3,6590

[1] Dados cedidos pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão.

Tabela 14. Quadro dos valores totais dos fatores e suas combinações, necessário para o cálculo das somas de quadrados dos fatores e sua interação.

	Microelemento = 0	Microelemento = 1	Microelemento = 2	Soma
Densidade = 0	9,6889	10,8926	9,3146	29,8961
Densidade = 1	10,6559	10,4932	10,9073	32,0564
Densidade = 2	11,3786	10,4150	10,8848	32,6784
Soma	31,7234	31,8007	31,1067	94,6309

Tabela 15. Análise de variância para um ensaio fatorial 3 x 3 em blocos completos ao acaso.[1]

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	2	0,00962540	0,00481270	0,09 (0,9148)
Densidade (D)	2	0,47390018	0,23695009	4,41 (0,0299)
Microelemento (M)	2	0,03214758	0,01607379	0,30 (0,7457)
D x M	4	0,60488310	0,15122078	2,81 (0,0608)
Erro	16	0,86043704	0,05377732	
Total	26	1,98099331		

[1] Dados cedidos pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão.

Tabela 16. Análise de variância para o estudo da interação entre os dois fatores no ensaio fatorial 3 x 3 em blocos completos ao acaso.[1]

Fonte de variação	Grau de liberdade (GL)	Quadrado médio (QM)	F (Pr > F)
Micro lin. d. D = 0	1	0,02335568	0,43 (0,5193)
Micro qua. d. D = 0	1	0,42985555	7,99 (0,0121)
Micro lin. d. D = 1	1	0,01053551	0,20 (0,6640)
Micro qua. d. D = 1	1	0,01848154	0,34 (0,5659)
Micro lin. d. D = 2	1	0,04063845	0,76 (0,3975)
Micro qua. d. D = 2	1	0,11416395	2,12 (0,1645)

Lin = linear; qua = quadrático; d. = para dentro; D = densidade.

[1] Dados cedidos pelo pesquisador Pedro Marques da Silveira, da Embrapa Arroz e Feijão.

anteriormente no item *Fatores com níveis igualmente espaçados*, em que a equação é dada por

$$Y_i - \overline{Y} = \sum_k b_k P_k M_k \qquad [20]$$

de forma que se tem

$$MS = 29,8961/9 \ \frac{(-1*9,6889 + 0*10,8926 + 1*9,3146)}{3*[(-1)^2 + (0)^2 + (1)^2]} \ * x \ *1 +$$

$$\frac{(1*9,6889 - 2*10,8926 + 1*9,3146)}{3*[(1)^2 + (-2)^2 + (1)^2]} \ * (x^2 - 8/13)*3$$

Logo,

$$MS = 3,6308 - 0,0624(X - 1) - 0,4635(X - 1)^2$$

$$MS = 3,2297 + 0,8646X - 0,4635X^2$$

em que X é a dose de microelementos aplicada por vaso.

A equação revela então que a resposta possui um ponto de máximo igual a 3,6329, que se verifica para uma dose de 0,93 g de FTE-BR-12 por vaso, o que corresponde a 4,29 g de matéria seca por três plantas por vaso.

Fatorial 3 x 3 com o uso do SAS

Para a análise inicial do experimento com um fatorial 3 x 3, o programa SAS é descrito de forma idêntica àquele apresentado na seção *Uso do SAS no fatorial 2 x 2 em blocos ao acaso*, ou seja:

```
PROC ANOVA;CLASS REPETICAO DENSIDADE MICRO;
MODEL MAT_SEC=REPETICAO DENSIDADE|MICRO;
MEANS DENSIDADE|MICRO;
```

No entanto, considerando-se as discussões na seção *Fatorial 3 x 3 – exemplo de análise*, o estudo detalhado da interação necessita ser feito, com o uso de contrastes. Nesse caso, o procedimento a ser usado é o GLM, com a opção CONTRAST. Da mesma forma como anteriormente, usam-se os contrastes ortogonais, para três níveis, com igual espaçamento entre eles.

O programa SAS então passa a ter a seguinte configuração:

```
PROC GLM;CLASS REPETICAO DENSIDADE MICRO;
MODEL  MAT_SEC=REPETICAO DENSIDADE|MICRO;
CONTRAST "MICRO LIN d. DENS=0" MICRO -1 0 1 DENSIDADE*MICRO -1 0 1 0 0 0 0 0 0;
CONTRAST "MICRO QUA d. DENS=0" MICRO 1 -2 1 DENSIDADE*MICRO 1 -2 1 0 0 0 0 0 0;
CONTRAST "MICRO LIN d. DENS=1" MICRO -1 0 1 DENSIDADE*MICRO 0 0 0 -1 0 1 0 0 0;
CONTRAST "MICRO QUA d. DENS=1" MICRO 1 -2 1 DENSIDADE*MICRO 0 0 0 1 -2 1 0 0 0;
CONTRAST "MICRO LIN d. DENS=2" MICRO -1 0 1 DENSIDADE*MICRO 0 0 0 0 0 0 -1 0 1;
CONTRAST "MICRO QUA d. DENS=2" MICRO 1 -2 1 DENSIDADE*MICRO 0 0 0 0 0 0 1 -2 1;
MEANS DENSIDADE|MICRO;
```

O resultado desse programa aparece na Figura 3, e maiores detalhes sobre a construção de contrastes podem ser encontrados no Capítulo 18 e nos manuais do SAS.

Para a obtenção da equação de regressão de segundo grau para microelementos dentro da densidade 0 (zero), existe a necessidade de um pequeno artifício: calcular as médias para os níveis de micro elementos dentro da dose zero de densidade e depois usar o PROC GLM para ajustar a equação de regressão. O programa para isso é:

```
PROC SORT DATA=A;BY DENSIDADE MICRO;
DATA B;SET A;IF DENSIDADE=0;
PROC SORT;BY MICRO;
PROC MEANS MEAN NOPRINT;BY MICRO;VAR MAT_SEC;OUTPUT OUT=MEDIA
MEAN=MAT_SEC;
PROC GLM;MODEL MAT_SEC=MICRO MICRO*MICRO;
```

O resultado está na Figura 4, da qual se usa apenas o resultado final da estimação dos parâmetros da equação de regressão de segundo grau.

Como foram usadas apenas as três médias correspondentes às doses de microelementos na densidade 0 (zero), não existem graus de liberdade para o teste da análise de variância, a qual já foi realizada antes com todos os dados do experimento, e de lá vem a significância da regressão. Por isso, dessa análise se usaria apenas a estimação dos parâmetros da equação de regressão.

```
                    The GLM Procedure
                Class Level Information
Class          Levels    Values
REPETICAO           3    1 2 3
DENSIDADE           3    0 1 2
MICRO               3    0 1 2

  Number of observations    27
                    The GLM Procedure
Dependent Variable: MAT_SEC
                                Sum of
Source                  DF      Squares    Mean Square   F Value   Pr > F
Model                   10   1.12058928     0.11205893      2.08   0.0919
Error                   16   0.86044547     0.05377784
Corrected Total         26   1.98103475

             R-Square     Coeff Var      Root MSE    MAT_SEC Mean
             0.565659      6.616542      0.231900        3.504859

Source                  DF    Type I SS    Mean Square   F Value   Pr > F
REPETICAO                2   0.00961796     0.00480898      0.09   0.9149
DENSIDADE                2   0.47393591     0.23696796      4.41   0.0299
MICRO                    2   0.03213757     0.01606879      0.30   0.7458
DENSIDADE*MICRO          4   0.60489783     0.15122446      2.81   0.0608

Source                  DF   Type III SS    Mean Square   F Value   Pr > F
REPETICAO                2   0.00961796     0.00480898      0.09   0.9149
DENSIDADE                2   0.47393591     0.23696796      4.41   0.0299
MICRO                    2   0.03213757     0.01606879      0.30   0.7458
DENSIDADE*MICRO          4   0.60489783     0.15122446      2.81   0.0608

Contrast                DF  Contrast SS    Mean Square   F Value   Pr > F
MICRO LIN d. DENS=0      1   0.02335008     0.02335008      0.43   0.5193
MICRO QUA d. DENS=0      1   0.42988083     0.42988083      7.99   0.0121
MICRO LIN d. DENS=1      1   0.01054204     0.01054204      0.20   0.6639
MICRO QUA d. DENS=1      1   0.01847683     0.01847683      0.34   0.5659
MICRO LIN d. DENS=2      1   0.04062328     0.04062328      0.76   0.3976
MICRO QUA d. DENS=2      1   0.11416235     0.11416235      2.12   0.1645

                    The GLM Procedure
Level of                ----------MAT_SEC----------
DENSIDADE          N            Mean          Std Dev
0                  9      3.32178889       0.31263026
1                  9      3.56184444       0.21555883
2                  9      3.63094444       0.21020007

Level of                ----------MAT_SEC----------
MICRO              N            Mean          Std Dev
0                  9      3.52482222       0.30270756
1                  9      3.53343333       0.19418516
2                  9      3.45632222       0.33804202

Level of   Level of     ----------MAT_SEC----------
DENSIDADE  MICRO    N            Mean          Std Dev
0          0        3      3.22963333       0.23218429
0          1        3      3.63086667       0.21077206
0          2        3      3.10486667       0.25690631
1          0        3      3.55196667       0.23893033
1          1        3      3.49776667       0.05571843
1          2        3      3.63580000       0.33340778
2          0        3      3.79286667       0.12628746
2          1        3      3.47166667       0.28536181
2          2        3      3.62830000       0.04431219
```

Figura 3. Uso de contrastes para estimação de efeito linear e quadrático em um fatorial 3 x 3 com o SAS.

```
                          The GLM Procedure
                       Number of observations    3
                          The GLM Procedure

Dependent Variable: MAT_SEC
                                  Sum of
Source                     DF       Squares    Mean Square   F Value   Pr > F
Model                       2    0.15107697     0.07553848         .       .
Error                       0    0.00000000      .
Corrected Total             2    0.15107697

                R-Square     Coeff Var      Root MSE    MAT_SEC Mean
                1.000000             .             .        3.321789

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
MICRO                       1    0.00778336     0.00778336         .       .
MICRO*MICRO                 1    0.14329361     0.14329361         .       .

Source                     DF     Type III SS    Mean Square   F Value   Pr > F
MICRO                       1    0.11507162     0.11507162         .       .
MICRO*MICRO                 1    0.14329361     0.14329361         .       .

                                                  Standard
          Parameter                Estimate          Error    t Value    Pr > |t|
          Intercept             3.229633333              .          .           .
          MICRO                 0.864850000              .          .           .
          MICRO*MICRO          -0.463616667              .          .           .
```

Figura 4. Uso do SAS com o procedimento GLM para obtenção de uma equação de regressão no caso de uma fatorial 3 x 3.

Fatorial 3 x 3 x 3 – exemplo de aplicação

Os dados aqui utilizados pertencem ao pesquisador Alberto Baeta dos Santos, da Embrapa Arroz e Feijão, que, no ano agrícola 1997–1998, em um ensaio realizado na Fazenda Palmital daquela unidade de pesquisa, investigava o efeito do N, do K e das épocas de aplicação desses elementos na produtividade da soca do arroz irrigado. Ressalte-se que originalmente os dois elementos foram testados com quatro níveis, mas, para esse exemplo, a dose mais alta foi eliminada.

O delineamento experimental adotado foi o de blocos completos ao acaso, com três repetições. As Tabelas 17 e 18 mostram os dados obtidos referentes à produção de grãos, em kg ha^{-1}. A Tabela 17 contém todos os valores, e a Tabela 18 contém os quadros auxiliares que se constroem para facilitar a obtenção das somas de quadrados e das médias de cada um dos fatores e de suas interações simples.

A análise de variância é a que está sumarizada na Tabela 19.

A soma de quadrados para repetição se obteve da forma

$$SQRep = \frac{1}{27}(28.508,22^2 + 27.801,10^2 + 27.587,13^2) - \frac{1}{81}(83.896,45)^2$$

Tabela 17. Dados de produção de grãos de arroz em um ensaio fatorial 3^3, em blocos completos ao acaso.[1]

Nitrogênio	Potássio	Época	Repetição 1	Repetição 2	Repetição 3
1	1	1	1.066,31	886,36	1.125,69
1	1	2	1.022,19	841,58	1.509,76
1	1	3	1.137,93	1.393,92	1.259,35
1	2	1	922,84	943,84	1.454,8
1	2	2	1.056,45	506,26	983,85
1	2	3	747,22	697,66	783,39
1	3	1	914,94	858,15	719,37
1	3	2	718,15	860,4	603,87
1	3	3	671,60	1.050,53	784,58
2	1	1	1.090,63	882,44	1.067,93
2	1	2	1.508,95	1.352,49	901,28
2	1	3	1.003,57	877,54	742,91
2	2	1	1.108,57	1.759,42	660,58
2	2	2	974,27	1.242,17	716,86
2	2	3	1.518,25	1.331,46	1.048,17
2	3	1	1.494,00	1.001,68	1.941,66
2	3	2	875,19	411,64	825,1
2	3	3	1.387,65	1368,2	810,17
3	1	1	826,87	1.958,82	986,09
3	1	2	1.394,67	964,25	1.126,04
3	1	3	871,14	1.031,83	1.472,08
3	2	1	788,42	530,53	916,45
3	2	2	1.074,87	554,86	754,31
3	2	3	910,33	936,04	760,7
3	3	1	913,07	1.318,21	998,53
3	3	2	1.057,20	786,19	978,79
3	3	3	1.452,94	1.454,63	1.654,82
	Soma		28.508,22	27.801,10	27.587,13

[1] Dados cedidos pelo pesquisador Alberto Baeta dos Santos, da Embrapa Arroz e Feijão.

Tabela 18. Quadros auxiliares para a análise de variância, mostrando os totais de cada um dos fatores e suas combinações, em pares.[1]

	K1	K2	K3	Soma
N1	10.243,09	8.096,31	7.181,59	25.520,99
N2	9.427,74	10.359,75	10.115,29	29.902,78
N3	10.631,79	7.226,51	10.614,38	28.472,68
Soma	30.302,62	25.682,57	27.911,26	83.896,45

	EP1	EP2	EP3	Soma
N1	8.892,3	8.102,51	8.526,18	25.520,99
N2	11.006,91	8.807,95	10.087,92	29.902,78
N3	9.236,99	8.691,18	10.544,51	28.472,68
Soma	29.136,2	25.601,64	29.158,61	83.896,45

	EP1	EP2	EP3	Soma
K1	9.891,14	10.621,21	9.790,27	30.302,62
K2	9.085,45	7.863,9	8.733,22	25.682,57
K3	10.159,61	7.116,53	10.635,12	27.911,26
Soma	29.136,2	25.601,64	29.158,61	83.896,45

[1] Dados cedidos pelo pesquisador Alberto Baeta dos Santos, da Embrapa Arroz e Feijão.

Tabela 19. Análise de variância do experimento fatorial 3^3.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
REP	2	17.212,452	8.606,226	0,11 (0,8971)
N	2	369.848,685	184.924,342	2,34 (0,1066)
K	2	395.438,565	197.719,283	2,50 (0,0920)
N x K	4	1.059.864,611	264.966,153	3,35 (0,0163)
EP	2	310.440,150	155.220,075	1,96 (0,1508)
N x EP	4	196.891,274	49.222,819	0,62 (0,6487)
K x EP	4	632.986,780	158.246,695	2,00 (0,1081)
N x K x EP	8	1.081.209,619	135.151,202	1,71 (0,1186)
Erro	52	4.113.885,580	79.113,184	
Total	80	8.177.777,716		

e fórmulas semelhantes se adotaram para as somas de quadrados dos fatores principais e suas interações, das quais se mostram aquelas para K, épocas de aplicação e sua interação.

$$SQK = \frac{1}{27}(30.306,62^2 + 25.682,57^2 + 27.911,26^2) - \frac{1}{81}(83.896,45)^2$$

$$SQEP = \frac{1}{27}(29.136,20^2 + 25.601,64^2 + 29.158,61^2) - \frac{1}{81}(83.896,45)^2$$

$$SQKxEP = \frac{1}{9}(9.891,14^2 + 10.621,21^2 + ... + 7.116,53^2 + 10.635,12^2) -$$

$$\frac{1}{81}(83.896,45)^2 - SQK - SQEP$$

Como se vê na Tabela 19, a interação entre N e K é significativa, o que implica que estudos adicionais devem ser feitos para explicar essa interação. Como ambos os efeitos principais não foram significativos, é possível estudar essa interação tanto do ponto de vista do N dentro de cada um dos níveis de K, quanto ao contrário. Como exemplo de aplicação, a Tabela 20 mostra o estudo dos efeitos lineares e quadráticos de N dentro de cada um dos três níveis de K. Esse estudo se faz da mesma forma mostrada no exemplo anterior de um fatorial 3 x 3.

Tabela 20. Estudo adicional da interação N x K, onde se estudam os efeitos linear e quadrático do nitrogênio dentro de cada nível de potássio.

Fonte	Grau de liberdade (GL)	Quadrado médio (QM)	F (Pr >F)
N linear d. K = 0	1	8.393,7606	0,11 (0,7459)
N quadrático d. K = 0	1	75.518,0807	0,95 (0,3331)
N linear d. K = 1	1	42.030,6689	0,53 (0,4693)
N quadrático d. K = 1	1	539.336,2041	6,82 (0,0118)
N linear d. K = 2	1	654.669,2880	8,28 (0,0058)
N quadrático d. K = 2	1	109.765,2936	1,39 (0,2442)

N = nitrogênio; K = potássio; d. = para dentro.

Os efeitos significativos, quadrático para N dentro de K = 1 e linear para N dentro de K = 2, associados à não significância de N na dose 0 (zero) para K, mostram claramente o porquê da significância da interação simples entre esses dois elementos nesse ensaio. Da mesma forma que anteriormente poder-se-ia obter ambas as equações de regressão para N.

Obtenção das fórmulas de somas de quadrados e dos valores esperados de quadrados médios

Para os estudos que envolvem ensaios fatoriais ou ensaios aninhados ou ainda uma combinação de ambas as formas com um número razoável de fatores, com fatores sendo definidos como fixos e/ou aleatórios, pode ser difícil para um pesquisador encontrar os modelos já previamente desenvolvidos.

Montgomery (1991) apresentou um conjunto de regras práticas para definir as formas das somas de quadrados e os respectivos valores esperados, que independem de o pesquisador ter conhecimento da teoria matemática envolvida e que são muito fáceis de ser aplicadas.

Forma prática de obtenção das fórmulas de somas de quadrados

As regras, relativamente simples, para definir as somas de quadrados são:

- **Regra 1:** o erro do modelo ($\varepsilon_{ij...m}$) é sempre escrito como $\varepsilon_{(ij...)m}$, em que o subscrito m indica o subscrito de repetição.
- **Regra 2:** além da média geral e do erro, o modelo contém todos os efeitos principais e qualquer interação que o pesquisador considera existir. Se existem k fatores num estudo, e todas as interações estão presentes, então existem k efeitos principais, $k(k - 1)/2$ efeitos de interações entre dois fatores, $k(k - 1)(k - 2)/6$ efeitos de interações entre três fatores, um efeito de interação entre k fatores. Quando um fator no modelo é indicado entre parêntesis, não existem interações entre aquele fator e os demais naquele termo.
- **Regra 3:** para cada termo no modelo, divida os subscritos em três tipos: a) vivos – presentes e não em parêntesis; b) mortos – presentes entre parêntesis; c) ausentes – subscritos não aparecem no termo. Por exemplo: $(\tau\beta)_{ij}$, i e j estão vivos e *k* está ausente, e em $\varepsilon_{(ij)k}$, k está vivo e i e j estão mortos.
- **Regra 4:** o número de graus de liberdade para qualquer termo no modelo é igual ao produto do número de níveis associado a cada subscrito morto e o número de níveis menos um de cada subscrito vivo.
- **Regra 5:** Para obter a fórmula de cálculo de qualquer soma de quadrado, primeiro faça a expansão do número de graus de liberdade a ela associada. Por exemplo, para o fator β_j os graus de liberdade expandidos são $b - 1$. Cada elemento nessa quantidade é uma representação simbólica de uma soma de quadrados não

corrigida, e o valor 1 representa o fator de correção. Usando a letra y para representar o dado lido e os subscritos, i, j, k, ..., m, para cada um dos fatores, ou a sua substituição por um ponto para representar o resultado da soma em relação àquele subscrito, ter-se-ia que o fator de correção seria escrito como:

$$1 = \frac{1}{ab...r}\sum_{i}^{a}\sum_{j}^{b}...\sum_{m}^{r} y_{ij...m}^{2} = \frac{1}{ab...r} y_{.....}^{2} \quad [21]$$

E a soma de quadrados correspondente ao b da expansão seria escrita:

$$b = \frac{1}{ac...r}\sum_{j}^{b} y_{.j....}^{2}$$

Dessa forma, a soma de quadrados corrigida para o fator β seria dada por

$$SS\beta = \frac{1}{ac...r}\sum_{j}^{b} y_{.j...}^{2} - \frac{1}{ab...r} y_{....}^{2} \quad [22]$$

De forma semelhante, seriam definidas as fórmulas de cálculo para todas as fontes de variação.

Obtenção das esperanças matemáticas dos quadrados médios

Ainda segundo Montgomery (1991) as seguintes regras devem ser aplicadas para a obtenção das esperanças matemáticas, de forma prática:

- Regra 1 – cada efeito tem ou um componente de variância (efeito aleatório) ou a fator fixo (efeito fixo) associado a ele. Se existe interação, e ela contém pelo menos um efeito aleatório, é considerada aleatória. Um componente de variância tem seu subscrito identificado por uma letra grega para identificar um efeito aleatório. Um efeito fixo é sempre identificado pela soma de quadrados dos componentes do modelo associado com o fator, dividido pelos seus graus de liberdade. Num modelo misto, com dois fatores, A fixo e B aleatório, o componente de variância de B é dado por σ_{β}^{2} e da interação AB por $\sigma_{\alpha\beta}^{2}$ enquanto para o fator A, a expressão é: $\frac{\sum_{i}^{a}\alpha_{i}^{2}}{a-1}$.

- Regra 2 – para obter a esperança dos quadrados médios, prepare uma tabela da seguinte forma: existe uma linha para cada componente do modelo, uma coluna para cada subscrito e uma coluna para a esperança do quadrado médio. Sobre cada subscrito escreva o número de níveis do fator associado a ele e a identificação do fator como fixo (F) ou aleatório (A). Repetições devem sempre ser consideradas como aleatórias.

  a) Em cada linha, escreva 1 na coluna em que houver subscrito morto coincidente com o subscrito da coluna.

  b) Em cada linha, escreva 0 se a coincidência de subscritos vivos for verificada e o fator é fixo, ou 1 se o fator for aleatório.

  c) Nos espaços vazios, escreva o número de níveis que encabeça cada coluna.

  d) Para obter a esperança do quadrado médio de um componente qualquer do modelo, primeiro cubra todas as colunas iniciadas por subscrito vivo do componente. Então, em cada linha que contenha pelo menos os mesmos subscritos do componente, tome o produto dos números visíveis e multiplique-o pelo componente fixo ou aleatório apropriado dado na regra 1.

A Tabela 21 mostra a aplicação dessas regras para o caso de um modelo misto, com dois fatores, no delineamento inteiramente ao acaso, cujos componentes de variância foram mostrados na regra 1.

Tabela 21. Estimação da esperança dos quadrados médios para os fatores principais, sua interação e o erro em um ensaio fatorial em que se assume que o fator A é fixo e o fator B e a interação A x B são aleatórios.

Fator	F A I	A b j	A R K	Quadrado médio esperado
α_I	0	B	R	$\sigma^2 + \frac{bn\sum \alpha_i^2}{a - 1}$
β_J	A	1	R	$\sigma^2 + r\sigma^2_{\alpha\beta} + ar\sigma^2_{\beta}$
$\alpha\beta_{IJ}$	0	1	1	$\sigma^2 + r\sigma^2_{\alpha\beta}$
$\varepsilon_{(IJ)K}$	1	1	1	σ^2

Ensaios fatoriais com tratamentos adicionais

Muitas vezes o pesquisador, ao executar um experimento num esquema fatorial, necessita de algumas informações adicionais o que o leva a acrescentar um ou vários tratamentos adicionais ao desenho. As situações mais comuns ocorrem nos ensaios em que se testam fontes e doses (adubos, inseticidas, fungicidas, etc.), nos quais uma testemunha sem aplicação do produto é acrescentada, ou em ensaios de adubação para estudos de macroelementos (N, P e K), em que se adicionam tratamentos com microelementos ou com aplicação de calcário.

As análises de variância devem então ser modificadas de acordo com essas adições, e o uso de contrastes é a solução mais adequada para esses casos. Quando o tratamento adicional é a testemunha, pode também ser empregado o teste de Dunnett. Como é quase impossível fazer uma generalização de esquema de análise, esses casos serão discutidos com a apresentação de um exemplo.

Testemunha e tratamento adicionados

A Tabela 22 mostra o número de sementes infectadas com *Fusarium* (um total de 40 sementes), bem como sua transformação por arco seno da raiz quadrada da proporção de sementes infectadas, $[arcsen(\sqrt{(número/40)})]$, e refere-se a um experimento no delineamento completamente ao acaso, num esquema fatorial 3 x 3 + 2. O primeiro fator era a aplicação de fungicidas às sementes, e o segundo fator o uso associado ou não de um polímero. Os dois tratamentos adicionais eram a aplicação isolada do polímero e de uma testemunha sem qualquer tratamento.

A identificação dos tratamentos que compunham o esquema fatorial é a seguinte: F1 = benomyl, F2 = ethanethiol e F3 = carbendazim; puro = aplicação apenas do fungicida, antes = aplicação de um polímero após a aplicação do fungicida e mistura = aplicação do fungicida em mistura com o polímero.

Esse experimento foi conduzido no laboratório de fitopatologia da Embrapa Arroz e Feijão, no ano de 1999, e é parte da tese de mestrado de Larissa Leandro Pires, sob orientação conjunta de Jefferson Luiz da Silva Costa e Cláudio Bragantini.

A análise de variância (Tabela 23) foi executada com os dados transformados em arco seno da raiz quadrada da proporção de sementes infectadas, $arcsen(\sqrt{(número/40)})$.

A soma de quadrados total foi obtida da forma usual. Para a obtenção das somas de quadrados do fatorial, deve-se estruturar um quadro auxiliar (Tabela 24), como mostrado anteriormente. As somas de quadrados dos efeitos de fatorial x adicionais e polímero x testemunha são obtidas com o uso de contrastes ortogonais, que também poderiam ser utilizados para as somas de quadrados do fatorial.

Tabela 22. Número de sementes infectadas com *Fusarium* em amostras de 40 sementes, e sua transformação [$arcsen(\sqrt{(número/40)})$] em um ensaio em delineamento completamente ao acaso, num esquema fatorial 3 x 3 + 2.

Tratamento	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Repetição 5	Soma
F1 Puro	5 (0,36137)	0 (0,00000)	0 (0,00000)	0 (0,00000)	0 (0,00000)	5 (0,36137)
F1 Antes	1 (0,15878)	1 (0,15878)	0 (0,00000)	0 (0,00000)	0 (0,00000)	2 (0,31756)
F1 Mistura	3 (0,27741)	0 (0,00000)	2 (0,22551)	0 (0,00000)	5 (0,36137)	10 (0,86429)
F2 Puro	7 (0,43161)	8 (0,46365)	5 (0,36137)	13 (0,60661)	4 (0,32175)	37 (2,18499)
F2 Antes	3 (0,27741)	0 (0,00000)	0 (0,00000)	5 (0,36137)	0 (0,00000)	8 (0,63878)
F2 Mistura	6 (0,39770)	5 (0,36137)	6 (0,39770)	6 (0,39770)	4 (0,32175)	27 (1,87622)
F3 Puro	0 (0,00000)	0 (0,00000)	0 (0,00000)	2 (0,22551)	0 (0,00000)	2 (0,22551)
F3 Antes	0 (0,00000)	0 (0,00000)	0 (0,00000)	1 (0,15878)	0 (0,00000)	1 (0,15878)
F3 Mistura	0 (0,00000)	0 (0,00000)	0 (0,00000)	0 (0,00000)	0 (0,00000)	0 (0,00000)
Polímero	6 (0,39770)	7 (0,43161)	5 (0,36137)	9 (0,49422)	11 (0,55202)	38 (2,23692)
Testemunha	14 (0,63305)	8 (0,46365)	7 (0,43161)	10 (0,52360)	6 (0,39770)	45 (2,44961)

Tabela 23. Análise de variância de um ensaio fatorial 3 x 3 + 2, em delineamento completamente ao acaso com cinco repetições. Dados de porcentagem de sementes atacadas transformados por $arcsen(\sqrt{(\%)})$.

Fonte	Grau de liberdade (GL)	Som de quadrado (SQ)	Quadrado médio (QM)	F (Pr > F)
Fungicidas	2	0,6652	0,3326	27,05 (< 0,0001)
Formas	2	0,1197	0,0598	4,87 (0,0123)
Fungicidas x formas	4	0,1904	0,0476	3,87 (0,026)
Fatorial x adicionais	1	0,8450	0,8450	68,72 (< 0,0001)
Polímero x testemunha	1	0,0045	0,0045	0,37 (0,5473)
Erro	44	0,5411	0,0123	
Total	54	2,2839		

Tabela 24. Quadro auxiliar para obtenção das somas de quadrados dos fatores principais e sua interação.

	Puro	Antes	Misturado	Soma
F1	0,36137	0,31756	0,86429	1,54322
F2	2,18499	0,63878	1,87622	4,69999
F3	0,22551	0,15878	0,00000	0,38429
Soma	2,77187	1,11512	2,74051	6,62750

De forma que se tem então:

$$SQFung = \frac{1}{15}(1{,}54322^2 + 4{,}69999^2 + 0{,}38429^2) - \frac{1}{45}(6{,}62750)^2$$

$$SQFormas = \frac{1}{15}(2{,}77187^2 + 1{,}11512^2 + 2{,}74051^2) - \frac{1}{45}(6{,}62750)^2$$

$$SQFungxFormas = \frac{1}{5}(0{,}36137^2 + 0{,}31756^2 + ... + 0{,}00000^2) -$$

$$\frac{1}{45}(6{,}62750)^2 - SQFung - SQFormas$$

A Tabela 25 mostra os contrastes necessários para as outras duas somas de quadrados, fatorial x adicionais e polímero x testemunha, que são obtidos pela aplicação das formas:

$$SQContraste = \frac{(\sum c_i T_i)^2}{r\sum c_i^2} \quad [23]$$

Então a soma de quadrados de polímero x testemunha, por exemplo, é dada por:

$$SQ(PolimeroxTestemunha) = \frac{(1*2{,}23692 + (-1)*2{,}44961)^2}{5*[1^2 + (-1)^2]}$$

Estudos adicionais poderiam ter sido feitos, como, por exemplo, explorar a significância da interação entre fungicidas e formas de aplicação, que deixamos para a curiosidade do leitor em explorar.

Tabela 25. Contrastes ortogonais para estimação de efeitos de tratamentos adicionais em um experimento fatorial.

F1 puro	F1 antes	F1 mistura	F2 puro	F2 antes	F2 mistura	F3 puro	F3 antes	F3 mistura	Polímero	Testemunha
0,36137	0,31756	0,86429	2,18499	0,63878	1,87622	0,22551	0,15878	0,00000	2,23692	2,44961
2	2	2	2	2	2	2	2	2	-9	-9
0	0	0	0	0	0	0	0	0	1	-1

Uso do SAS para experimentos fatoriais com tratamentos adicionais.

No caso do uso do SAS para a análise de ensaios fatoriais com tratamentos adicionais, existe a necessidade de uso do comando CONTRAST para obtenção das somas de quadrados tanto do fatorial como das comparações adicionais como se mostrou no item *Testemunha e tratamento adicionados*. O programa SAS para essa análise seria então dada por:

```
PROC GLM;CLASS TRAT;
MODEL PSI=TRAT;
CONTRAST "FUNGI" TRAT -1 -1 -1 0 0 0 1 1 1 0 0,
        TRAT 1 1 1 -2 -2 -2 1 1 1 0 0;
CONTRAST "FORMA" TRAT -1 0 1 -1 0 1 -1 0 1 0 0,
        TRAT 1 -2 1 1 -2 1 1 -2 1 0 0;
CONTRAST "FUNG X FORM" TRAT 1 0 -1 0 0 0 -1 0 1 0 0,
            TRAT -1 2 -1 0 0 0 1 -2 1 0 0,
            TRAT -1 0 1 2 0 -2 -1 0 1 0 0,
            TRAT 1 -2 1 -2 4 -2 1 -2 1 0 0;
CONTRAST "FATORIAL X ADIC" TRAT 2 2 2 2 2 2 2 2 2 -9 -9;
CONTRAST "POLIM X TEST" TRAT 0 0 0 0 0 0 0 0 0 1 -1;
MEANS TRAT;
```

Os resultados obtidos se encontram na Figura 5.

Ensaios fatoriais com uma repetição

Em muitas situações é necessário trabalhar com apenas uma repetição dos tratamentos, seja por falta de material experimental, por deficiência de mão de obra ou de área experimental, seja mesmo por questão de custo. Nesse caso, a análise ainda é possível desde que uma suposição adicional seja feita ao modelo experimental, qual seja, a de que a

```
                          The GLM Procedure
                       Class Level Information
Class       Levels  Values
TRAT            11  F1Antes F1Mistura F1Puro F2Antes F2Mistura F2Puro
                    F3Antes F3Mistura F3Puro Polimero Testemunha

                        Number of observations    55

                          The GLM Procedure

Dependent Variable: PSI

                                  Sum of
Source                     DF       Squares     Mean Square    F Value    Pr > F
Model                      10    1.82485872      0.18248587      14.84    <.0001
Error                      44    0.54106258      0.01229688
Corrected Total            54    2.36592131

                R-Square     Coeff Var      Root MSE      PSI Mean
                0.771310      53.90670      0.110891      0.205710

Source                     DF     Type I SS     Mean Square    F Value    Pr > F
TRAT                       10    1.82485872      0.18248587      14.84    <.0001

Source                     DF    Type III SS     Mean Square    F Value    Pr > F
TRAT                       10    1.82485872      0.18248587      14.84    <.0001

Contrast                   DF    Contrast SS     Mean Square    F Value    Pr > F
FUNGI                       2    0.66519071      0.33259536      27.05    <.0001
FORMA                       2    0.11972659      0.05986330       4.87    0.0123
FUNG X FORM                 4    0.19038289      0.04759572       3.87    0.0089
FATORIAL X ADIC             1    0.84503482      0.84503482      68.72    <.0001
POLIM X TEST                1    0.00452370      0.00452370       0.37    0.5473

                          The GLM Procedure
                Level of                      ------------PSI------------
                TRAT              N                 Mean          Std Dev
                F1Antes           5           0.06351200       0.08696739
                F1Mistura         5           0.17285800       0.16507567
                F1Puro            5           0.07227400       0.16160958
                F2Antes           5           0.12775600       0.17743774
                F2Mistura         5           0.37524400       0.03378946
                F2Puro            5           0.43699800       0.11012412
                F3Antes           5           0.03175600       0.07100857
                F3Mistura         5           0.00000000       0.00000000
                F3Puro            5           0.04510200       0.10085114
                Polimero          5           0.44738400       0.07625292
                Testemunha        5           0.48992200       0.09248681
```

Figura 5. Análise de um experimento fatorial 3 x 3 + 2, usando o procedimento GLM do SAS e o comando CONTRAST.

interação de ordem maior é de pouca importância ou não existente e a de que ela pode ser então utilizada como um estimador da variância experimental. Essa situação pode também ocorrer em ensaios fatoriais em que se proceda ao confundimento de parte de interações de ordem elevada como será visto no Capítulo 10.

Exemplo de aplicação

A partir do mesmo experimento sobre os efeitos de adubação nitrogenada e potássica e suas épocas de aplicação, realizado por Alberto Baeta dos Santos, mostra-se na Tabela 26 a altura das plantas no cultivo da soca, para a repetição 1.

Tabela 26. Dados de altura de plantas, em cm, em um cultivo de soca de arroz irrigado, obtidos em um experimento fatorial 3 x 3 x 3. Uma repetição.

Nitrogênio	Potássio	Época	Altura
1	1	1	56,9
1	1	2	55,9
1	1	3	53,3
1	2	1	56,7
1	2	2	57,4
1	2	3	52,1
1	3	1	51,5
1	3	2	55,5
1	3	3	49,0
2	1	1	57,6
2	1	2	55,4
2	1	3	56,7
2	2	1	55,1
2	2	2	53,2
2	2	3	60,3
2	3	1	52,7
2	3	2	57,0
2	3	3	59,2
3	1	1	57,5
3	1	2	54,7
3	1	3	55,8
3	2	1	55,4
3	2	2	57,2
3	2	3	55,6
3	3	1	56,1
3	3	2	61,0
3	3	3	57,3

A partir desses dados, podem-se construir as tabelas auxiliares para a análise de variância, para estimar os efeitos principais e suas interações simples (Tabela 27).

A soma de quadrados total, que nesse caso é a soma de quadrados de tratamentos, é obtida por

$$SQTotal = (56,9^2 + 55,9^2 + ... + 57,3^2) - \frac{1}{27}(1.506,1)^2$$

Tabela 27. Quadros auxiliares para obtenção das somas de quadrados dos fatores principais e suas interações simples.

	K1	K2	K3	Soma
N1	166,1	166,2	156,0	488,3
N2	169,7	168,6	168,9	507,2
N3	168,0	168,2	174,4	510,6
Soma	503,8	503,0	499,3	1.506,1
	EP1	EP2	EP3	Soma
N1	165,1	168,8	154,4	488,3
N2	165,4	165,6	176,2	507,2
N3	169,0	172,9	168,7	510,6
Soma	499,5	507,3	499,3	1.506,1
	EP1	EP2	EP3	Soma
K1	172,0	166,0	165,8	503,8
K2	167,2	167,8	168,0	503,0
K3	160,3	173,5	165,5	499,3
Soma	499,5	507,3	499,3	1.506,1

e da mesma forma que nos exemplos anteriores de análise de ensaios fatoriais se obtêm as somas de quadrados para os efeitos principais e suas interações simples, como, por exemplo,

$$SQN = \frac{1}{9}(488{,}3^2 + 507{,}2^2 + 510{,}6^2) - \frac{1}{27}(1.506{,}1)^2$$

$$SQK = \frac{1}{9}(503{,}8^2 + 503{,}0^2 + 499{,}3^2) - \frac{1}{27}(1.506{,}1)^2$$

e

$$SQNxK = \frac{1}{3}(166{,}1^2 + 166{,}2^2 + ... + 174{,}4^2) - \frac{1}{27}(1.506{,}1)^2 - SQN - SQK$$

Resumindo, a análise de variância é o que se mostra na Tabela 28, na qual se vê que, além de N, suas interações com K e épocas de aplicação e a interação de K com épocas de aplicação também foram significativas. Evidentemente que essas interações e efeitos principais, a exemplo de outros casos anteriores, poderiam ser mais bem estudadas,

aplicando-se contrastes ortogonais, tanto no que se refere aos polinômios ortogonais, no caso de fatores quantitativos, quanto aos contrastes de interesse, no caso de fatores qualitativos.

Tabela 28. Análise de variância do experimento fatorial 3 x 3 x 3 em um delineamento completamente ao acaso.

Fonte	Graus de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Nitrogênio	2	32,0763	16,0382	8,20 (0,0116)
Potássio	2	1,2807	0,6404	0,33 (0,7301)
N x K	4	30,6570	7,6643	3,92 (0,0476)
Épocas	2	4,6252	2,3126	1,18 (0,3551)
N x épocas	4	61,7659	15,4415	7,89 (0,0070)
K x épocas	4	33,2415	8,3104	4,25 (0,0391)
Erro (N x K x épocas)	8	15,6541	1,9568	
Total (Tratamentos)	26	179,3007		

N = nitrogênio; K = potássio.

Uso do SAS em fatoriais 3 x 3 x 3 com uma repetição

No caso de fatoriais com apenas uma repetição, o procedimento adotado é o mesmo utilizado nos casos anteriores (ANOVA). No entanto, deve-se ter o cuidado de, na descrição do modelo matemático, dentro do comando MODEL, não especificar as interações que serão utilizadas como estimadores do erro. Para os dados do exemplo dado no item anterior (*Exemplo de aplicação*), a programação SAS então tem a seguinte forma:

```
PROC ANOVA;CLASS N K EPOCA;
MODEL ALT= N K N*K EPOCAS N*EPOCAS K*EPOCAS;
```

O resultado se vê na Figura 6.

```
                          The ANOVA Procedure
                       Class Level Information
                    Class         Levels    Values
                    N                  3    1 2 3
                    K                  3    1 2 3
                    EPOCA              3    1 2 3
                     Number of observations    27
                          The ANOVA Procedure
Dependent Variable: ALT
                                   Sum of
 Source                  DF        Squares    Mean Square   F Value   Pr > F
 Model                   18    163.6466667      9.0914815      4.65   0.0162
 Error                    8     15.6540741      1.9567593
 Corrected Total                26    179.3007407
                R-Square     Coeff Var      Root MSE      ALT Mean
                0.912694      2.507718      1.398842      55.78148
 Source                  DF       Anova SS    Mean Square   F Value   Pr > F
 N                        2    32.07629630    16.03814815      8.20   0.0116
 K                        2     1.28074074     0.64037037      0.33   0.7301
 N*K                      4    30.65703704     7.66425926      3.92   0.0476
 EPOCA                    2     4.62518519     2.31259259      1.18   0.3551
 N*EPOCA                  4    61.76592593    15.44148148      7.89   0.0070
 K*EPOCA                  4    33.24148148     8.31037037      4.25   0.0391
```

Figura 6. Uso do SAS para a análise de um fatorial 3 x 3 x 3 com uma repetição. Interação dos três fatores usada como erro.

Capítulo 9

Experimentos em parcelas divididas e suas variações

Como foi visto no capítulo anterior, os ensaios fatoriais podem ser executados em qualquer um dos delineamentos estatísticos básicos. No entanto, dada a sua estruturação por meio de combinações dos níveis de cada um dos fatores, surgem algumas novas possibilidades na sua instalação, que podem ser motivadas pelos seguintes fatores: pela necessidade de um maior tamanho de parcela para um ou mais fatores; por trazer uma maior simplicidade de aplicação; por alguma necessidade técnica ou mesmo por causa da forma como se coletam as informações. Esses fatos levam a condições particulares que permitem um novo esquema para efetuar a análise de variância.

A mais conhecida dessas variações é aquela denominada de parcelas divididas que, na realidade, implica que alguns fatores tenham parcelas de tamanho maior e outros de menor área. Estes últimos ficam então contidos nas parcelas de cada nível dos fatores anteriores. Outra forma também muito utilizada é aquela que leva aos experimentos em faixa, na qual os fatores se colocam de forma perpendicular uns sobre os outros. Essas duas formas podem também ser combinadas levando ao que se denomina de ensaios em faixa com parcelas divididas. O fato mais significativo desses arranjos diferenciados diz respeito à forma de se fazer a casualização dos fatores, já que ocorrem algumas restrições na alocação dos tratamentos ou necessidade de fazer uma sequência de sorteio.

Segundo o número de fatores e as restrições possíveis nos sorteios dos níveis dos fatores, existe uma infinidade de maneiras de se conduzirem os experimentos com parcelas divididas e suas variações e, em continuação, serão vistas algumas delas.

Parcelas divididas

Os experimentos em parcelas divididas podem ter origem no planejamento ou na forma de condução. No planejamento, os distintos fatores em estudo são divididos segundo

algum critério, uns para constituírem as parcelas de tamanho maior, e outros as de tamanho menor, que estarão então contidas naquelas. O processo de divisão de parcelas pode ser realizado em etapas sucessivas dando origem a diversas variações nesse desenho. O fator alocado às parcelas maiores, que chamamos de principal, é colocado segundo o delineamento experimental escolhido, enquanto o outro fator é colocado em cada uma das parcelas principais, com sorteios independentes. Para tornar mais claro esse princípio, imagine-se que dois fatores, chamados de A e B, serão testados, de modo que o fator A fique nas parcelas principais, segundo o delineamento em blocos casualizados, e o fator B nas parcelas secundárias. O sorteio da colocação dos diferentes níveis dos dois fatores se dá em sequência, isto é, primeiro define-se a colocação dos níveis do fator A nas parcelas em cada um dos blocos; depois, para cada uma delas, define-se a ordem de locação dos níveis do fator B. Supondo que cada fator tenha três níveis, e que se usem três blocos, têm-se as seguintes fases: fase 1 – sorteio do fator A nas parcelas (Figura 1); fase 2 – divisão das parcelas em três subparcelas e sorteio do fator B, independentemente para cada parcela (Figura 2).

Bloco 1	Bloco 2	Bloco 3
a1	a2	a3
a3	a1	a2
a2	a3	a1

Figura 1. Casualização do fator A às parcelas experimentais nos três blocos.

	Bloco 1				Bloco 2				Bloco 3			
a1	b1	b3	b2	a2	b2	b3	b1	a3	b2	b3	b1
a3	b2	b1	b3	a1	b3	b2	b1	a1	b2	b1	b3
a2	b3	b1	b2	a3	b1	b2	b3	a2	b1	b3	b2

Figura 2. Casualização do fator B às subparcelas experimentais em cada uma das parcelas, nos três blocos.

A Figura 2 mostra que na, primeira linha de parcelas do bloco 1, por exemplo, as três subparcelas conteriam as combinações a1b1, a1b3 e a1b2, nessa ordem. De igual modo, ocorre nas demais parcelas, de maneira que a última linha de subparcelas no bloco 3 teria os tratamentos a2b1, a2b3 e a2b2. Vê-se, portanto, que, para o fator B, os níveis de A têm a função de blocos.

Se três fatores estivessem disponíveis, teria sido possível alocar as combinações de dois deles às parcelas principais e o terceiro às subparcelas; ou, então, um às parcelas e as

combinações dos outros dois às subparcelas. Com mais fatores, as possibilidades se tornam maiores; portanto, muitas soluções são possíveis de ser obtidas. Imaginando a primeira situação apontada para três fatores, em blocos ao acaso com três repetições, fatores A e C com três níveis e o fator B com apenas dois níveis, um possível desenho esquemático do ensaio poderia ser o que está demonstrado na Figura 3.

	Bloco 1				Bloco 2				Bloco 3		
a1b2	c1	C3	c2	a3b2	c2	c3	c1	a2b2	c1	c2	c3
a3b1	c2	C3	c1	a2b1	c3	c1	c2	a1b1	c2	c1	c3
a1b1	c3	C1	c2	a1b2	c2	c3	c1	a2b1	c3	c1	c2
a2b2	c2	C3	c1	a2b2	c1	c2	c3	a3b2	c2	c3	c1
a3b2	c1	C2	c3	a3b1	c3	c1	c2	a1b2	c1	c3	c2
a2b1	c2	C1	c3	a1b1	c2	c3	c1	a3b1	c2	c1	c3

Figura 3. Resultado do sorteio de um experimento em parcelas divididas com dois fatores (A e B) nas parcelas e um terceiro (C) nas subparcelas.

Uma das consequências de tal distribuição dos tratamentos às parcelas experimentais é o aparecimento de dois erros na análise de variância, um chamado de erro a, que é utilizado para testar os efeitos dos tratamentos alocados às parcelas principais, e o outro chamado de erro b, que é usado no teste das hipóteses concernentes aos fatores nas subparcelas e às interações entre os fatores de parcelas e subparcelas, ao assumir-se um modelo fixo. De maneira geral o erro a tem maior valor absoluto que o erro b, que adicionalmente apresenta maior número de graus de liberdade. Esse fato determina que às subparcelas sejam alocados os fatores que necessitam de maior precisão de teste, além das considerações feitas anteriormente sobre as causas de definição de quais fatores seriam usados nas parcelas maiores e quais nas parcelas menores.

O modelo matemático segue o delineamento básico empregado e incorpora essas modificações a respeito do erro. Para o primeiro exemplo, mostrado anteriormente, o modelo matemático adequado seria, então,

$$Y_{ijk} = m + b_j + A_i + ea_{ij} + B_k + AB_{ik} + eb_{ijk} \quad [1]$$

Para o segundo exemplo seria

$$Y_{ijk} = m + b_j + A_i + B_k + AB_{ik} + ea_{ijk} + C_l + AC_{il} + BC_{kl} + ABC_{ikl} + eb_{ijkl} \quad [2]$$

Outra consequência desse tipo de delineamento é a dificuldade que surge para a comparação de médias ou estudos de graus de liberdade individuais. Procurando tornar mais claras essas consequências, examina-se o primeiro exemplo (dois fatores em blocos ao acaso). Cinco possibilidades existem nesse caso: comparar médias dos níveis do fator A ($\overline{A}_i$ - $\overline{A}_{i'}$), médias dos níveis do fator B ($\overline{B}_k$ - $\overline{B}_{k'}$), médias dos níveis do fator B dentro de cada nível de A ($\overline{A}_i\overline{B}_k$ - $\overline{A}_i\overline{B}_{k'}$), médias de A em cada um dos níveis do fator B ($\overline{A}_i\overline{B}_k$ - $\overline{A}_{i'}\overline{B}_k$) e finalmente qualquer combinação entre níveis de dois fatores ($\overline{A}_i\overline{B}_k$ - $\overline{A}_{i'}\overline{B}_{k'}$).

Para as três primeiras, os testes são diretos, utilizando-se o erro A no primeiro deles e o erro b nos outros dois. No entanto, as duas últimas possibilidades, levam ao envolvimento dos dois erros experimentais. A Tabela 1 mostra as fórmulas para estimativa das variâncias da diferença entre duas médias para todos esses casos.

Tabela 1. Estimativa da variância da diferença de duas médias no caso de um ensaio em parcelas divididas com dois fatores.

Comparação de médias	Variância da diferença entre duas médias
$(\overline{A}_i$ - $\overline{A}_{i'})$	$(2*QMEa)/(r*K)$
$(\overline{B}_k$ - $\overline{B}_{k'})$	$(2*QMEb)/(r*I)$
$(\overline{A}_i\overline{B}_k$ - $\overline{A}_i\overline{B}_{k'})$	$(2*QMEb)/r$
$(\overline{A}_i\overline{B}_k$ - $\overline{A}_{i'}\overline{B}_k)$	$[2*(K - 1)*QMEB + QMEa]/(r*K)$
$(\overline{A}_i\overline{B}_k$ - $\overline{A}_{i'}\overline{B}_{k'})$	$[2*(K - 1)*QMEB + QMEa]/(r*K)$

Para as duas últimas comparações de médias, há a necessidade de se calcular o número de graus de liberdade, pois a variância indicada é uma média ponderada dos dois erros da análise de variância. Satterthwaite (1946) propôs a aproximação desses graus de liberdade por meio da fórmula 3.

$$GL = \frac{[(K - 1)\,QMEb + QMEa]^2}{\frac{[(K - 1)\,QMEb]^2}{GLEb} + \frac{(QMEa)^2}{GLEa}} \qquad [3]$$

Na Tabela 1 e na fórmula 3, K é o número de níveis do fator B, I o número de níveis do fator A, r o número de repetições, $QMEa$, $QMEb$, $GLEa$ e $GLEb$ são, respectivamente, os quadrados médios e os graus de liberdade dos erros a e b.

Como já discutido nos capítulos sobre delineamentos básicos e ensaios fatoriais, o quadro de análise de variância é construído segundo o modelo matemático adotado. A Tabela 2 contém o modelo correspondente à fórmula 1 (dois fatores em blocos ao acaso), com o fator A nas parcelas e o fator B nas subparcelas e ambos os fatores são considerados como fixos.

Na Tabela 2, o fator de correção (FC) é calculado da forma usual.

Tabela 2. Análise de variância de um experimento em parcelas divididas em blocos completos ao acaso, com dois fatores.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F
Blocos (R)	$J - 1$	$\frac{1}{IK}\sum_{j} Y_{.j.}^2 - FC$	SQR/GLR	$QMR/QMEa$
Fator A (A)	$I - 1$	$\frac{1}{JK}\sum_{i} Y_{i..}^2 - FC$	SQA/GLA	$QMA/QMEa$
Erro a (Ea)	$(J - 1)(I - 1)$	$\frac{1}{K}\sum_{ij} Y_{ij.}^2 - FC - SQR - SQA$	$SQEa/GLEa$	
Fator B (B)	$K - 1$	$\frac{1}{IJ}\sum_{k} Y_{..k}^2 - FC$	SQB/GLB	$QMB/QMEb$
A*B (AB)	$(I - 1)(K - 1)$	$\frac{1}{J}\sum_{ik} Y_{i.k}^2 - FC - SQA - SQB$	$SQAB/GLAB$	$QMAB/QMEb$
Erro b (Eb)	$(J - 1)(K - 1)I$	Por diferença	$SQEb/GLEb$	
Total	$IJK - 1$	$\sum_{ijk} Y_{ijk}^2 - FC$		

Para o outro caso discutido, três fatores em blocos ao acaso, sendo os fatores A e B alocados às parcelas e o fator C às subparcelas, a análise de variância, também considerando os três fatores como fixos, é o que se mostra na Tabela 3.

Ao se considerarem os fatores A e B como aleatórios, ou um aleatório e o outro fixo, para o primeiro caso em discussão, a realização dos testes F deve ser modificada segundo os componentes de variância esperados. As Tabelas 4 e 5 mostram esses valores para todos os quatro casos.

Tabela 3. Análise de variância de um experimento em blocos ao acaso com parcelas divididas, possuindo três fatores – dois deles alocados às parcelas principais e um deles à subparcela.

Fonte de variação	Graus de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F
Blocos (R)	$J - 1$	$\frac{1}{IKL}\sum_{j} Y_{.j..}^2 - FC$	SQR/GLR	$QMR/QMEa$
Fator A (A)	$I - 1$	$\frac{1}{JKL}\sum_{i} Y_{i...}^2 - FC$	SQA/GLA	$QMA/QMEa$
Fator B (B)	$K - 1$	$\frac{1}{IJL}\sum_{k} Y_{..k.}^2 - FC$	SQB/GLB	$QMB/QMEa$
A*B (AB)	$(I - 1)(K - 1)$	$\frac{1}{JL}\sum_{ik} Y_{i.k.}^2 - FC - SQA - SQB$	$SQAB/GLAB$	$QMAB/QMEa$
Erro a (Ea)	$(J - 1)(I - 1)(K - 1)$	$\frac{1}{L}\sum_{ijk} Y_{ijk.}^2 - FC - SQR - SQA - SQB - SQAB$	$SQEa/GLEa$	
Fator C (C)	$L - 1$	$\frac{1}{IJK}\sum_{l} Y_{...l}^2 - FC$	SQC/GLC	$QMC/QMEb$
A*C	$(I - 1)(L - 1)$	$\frac{1}{JK}\sum_{il} Y_{i..l}^2 - FC - SQA - SQC$	$SQAC/GLAC$	$QMAC/QMEb$
B*C	$(K - 1)(L - 1)$	$\frac{1}{IJ}\sum_{kl} Y_{..kl}^2 - FC - SQB - SQC$	$SQBC/GLBC$	$QMBC/QMEb$
A*B*C	$(I - 1)(K - 1)(L - 1)$	$\frac{1}{J}\sum_{i.kl} Y_{i.kl}^2 - FC - SQA - SQB - SQC - SQAB - SQAC - SQBC$	$SQABC/GLABC$	$QMABC/QMEb$
Erro b (Eb)	$(J - 1)(L - 1)IK$	Por diferença	$SQEb/GLEb$	
Total	$IJKL - 1$	$\sum_{ijkl} Y_{ijkl}^2 - FC$		

Tabela 4. Esperança dos quadrados médios em um experimento em blocos ao acaso, com parcelas divididas, considerando primeiro os fatores A e B como fixos e depois ambos como aleatórios.

Fontes de variação	Graus de liberdade (GL)	Esperança dos quadrados médios	
		A e B fixos	A e B aleatórios
Blocos	$J - 1$	$\sigma^2 + K\sigma^2_{\alpha} + IK\sigma^2_r$	$\sigma^2 + K\sigma^2_{\alpha} + IK\sigma^2_r$
A	$I - 1$	$\sigma^2 + K\sigma^2_{\alpha} + JK\frac{\sum_i a_i^2}{I - 1}$	$\sigma^2 + K\sigma^2_{\alpha} + J\sigma^2_{ab} + JK\sigma^2_a$
Erro a	$(I - 1)(J - 1)$	$\sigma^2 + K\sigma^2_{\alpha}$	$\sigma^2 + K\sigma^2_{\alpha}$
B	$K - 1$	$\sigma^2 + IJ\frac{\sum_k b_k^2}{(K - 1)}$	$\sigma^2 + J\sigma^2_{ab} + IJ\sigma^2_b$
A*B	$(I - 1)(K - 1)$	$\sigma^2 + J\frac{\sum_{ik}(ab_{ik})^2}{(I - 1)(K - 1)}$	$\sigma^2 + J\sigma^2_{ab}$
Erro b	$(J - 1)(K - 1)I$	σ^2	σ^2

Tabela 5. Esperança dos quadrados médios em um experimento em blocos ao acaso, com parcelas divididas, considerando primeiro o fator A como aleatório e B como fixo, e depois A como fixo e B como aleatório.

Fontes de variação	Graus de liberdade (GL)	Esperança dos quadrados médios	
		A aleatório e B fixo	A fixo e B aleatório
Blocos	$J - 1$	$\sigma^2 + K\sigma^2_{\alpha} + IK\sigma^2_r$	$\sigma^2 + K\sigma^2_{\alpha} + IK\sigma^2_r$
A	$I - 1$	$\sigma^2 + K\sigma^2_{\alpha} + JK\sigma^2_a$	$\sigma^2 + K\sigma^2_{\alpha} + J\sigma^2_{ab} + JK\frac{\sum_i a_i^2}{I - 1}$
Erro a	$(I - 1)(J - 1)$	$\sigma^2 + K\sigma^2_{\alpha}$	$\sigma^2 + K\sigma^2_{\alpha}$
B	$K - 1$	$\sigma^2 + J\sigma^2_{ab} + IJ\frac{\sum_k b_k^2}{(K - 1)}$	$\sigma^2 + J\sigma^2_{ab} + IJ\sigma^2_b$
A*B	$(I - 1)(K - 1)$	$\sigma^2 + J\sigma^2_{ab}$	$\sigma^2 + J\sigma^2_{ab}$
Erro b	$(J - 1)(K - 1)I$	σ^2	σ^2

A partir dos resultados obtidos com a aplicação das regras de estimação dos valores esperados dos quadrados médios estabelecidas no Capítulo 8, fica evidente que, no caso em que ambos os fatores são aleatórios, o teste de F adequado para o fator B será obtido dividindo-se seu quadrado médio pelo quadrado médio da interação A*B. Por sua vez, para o fator A, o valor correto do teste F exige o quociente entre duas funções de quadrados médios, isto é, o numerador será a soma entre os quadrados médios do fator A e do erro b, enquanto o denominador é a soma dos quadrados médios do erro a e da interação A*B. Esses procedimentos são adotados sempre que necessários e segundo as especificações de efeitos aleatórios e fixos.

Exemplo de aplicação – dois fatores

No ano agrícola 1998–1999, os pesquisadores Anne Sitarama Prabhu e Gisele Barata da Silva realizaram um experimento na fazenda Capivara, da Embrapa Arroz e Feijão, onde testaram os efeitos de três doses de fongorene aplicado às sementes nas doses de 0, 400 g e 800 g de produto por 100 kg de sementes, em seis cultivares de arroz de sequeiro: Maravilha, Canastra, Confiança, Caiapó, Primavera e Rio Paranaíba. O delineamento adotado foi o de parcelas divididas em blocos ao acaso. As cultivares foram alocadas às parcelas e os tratamentos de semente às subparcelas. Mediu-se a percentagem de área foliar lesionada por brusone em cinco datas, e depois se estabeleceu a área sob a curva do progresso da doença (ASCPD). A Tabela 6 mostra esses dados, transformados em logaritmos, para três dessas cultivares.

Na Tabela 7, apresenta-se o quadro auxiliar que mostra os totais de cada combinação entre os fatores em estudo, bem como os totais gerais de cada nível dos dois fatores.

Tabela 6. Dados de área sob a curva do progresso da doença.[1]

Bloco	Cultivar	Dose de Fongorene (g)			Total das parcelas	Total de blocos
		0	400	800		
1	Canastra	6,29754	3,38405	3,71637	13,39796	
	Maravilha	4,92718	2,68410	2,33117	9,94245	
	R. Paranaíba	6,38072	4,22442	2,26124	12,86638	36,20679
2	Canastra	6,03998	3,43865	2,74181	12,22044	
	Maravilha	6,03634	3,27450	2,66966	11,98050	
	R. Paranaíba	6,12295	3,00642	2,73404	11,86341	36,06435
3	Canastra	6,13508	3,02796	2,77633	11,93937	
	Maravilha	5,44038	3,30615	3,40768	12,15421	
	R. Paranaíba	6,75723	4,33899	3,58060	14,67682	38,7704

[1] Dados cedidos pelo pesquisador Anne Sitarama Prabhu, da Embrapa Arroz e Feijão, e por Gisele Barata da Silva, professora da Universidade Rural da Amazônia.

Tabela 7. Quadro auxiliar para a obtenção das somas de quadrados dos fatores principais e sua interação.

Cultivar	Dose de Fongorene (g)			Total
	0	400	800	
Canastra	18,47260	9,85066	9,23451	37,55777
Maravilha	16,40390	9,26475	8,40851	34,07716
R. Paranaíba	19,26090	11,56983	8,57588	39,40661
Total	54,13740	30,68524	26,21890	111,04154

As somas de quadrados são obtidas aplicando-se as fórmulas dadas anteriormente:

$$SQCultivares = \frac{1}{9}(37{,}55777^2 + 34{,}07716^2 + 39{,}40661^2) - \frac{(111{,}04154)^2}{27} = 1{,}627261$$

$$SQFongorene = \frac{1}{9}(54{,}13740^2 + 30{,}68524^2 + 26{,}21890^2) - \frac{(111{,}04154)^2}{27} = 49{,}977598$$

$$SQBlocos = \frac{1}{9}(36{,}20679^2 + 36{,}06435^2 + 38{,}77040^2) - \frac{(111{,}04154)^2}{27} = 0{,}515372$$

$$SQErro(a) = \frac{1}{3}(13{,}39796^2 + 9{,}94345^2 + \ldots + 14{,}67682^2) - \frac{(111{,}04154)^2}{27} - SQBlocos - SQCultivares = 2{,}247697$$

$$SQCult * Trat = \frac{1}{3}(18{,}47260^2 + 9{,}85066^2 + \ldots + 8{,}57588^2) - \frac{(111{,}04154)^2}{27} - SQCultivares - SQFongorene = 0{,}908262$$

$$SQTotal = (6{,}29754^2 + 3{,}38405^2 + \ldots + 3{,}58060^2) - \frac{(111{,}04154)^2}{27} = 56{,}913816$$

e a *SQErro b*, obtida por diferença.

A análise de variância que resume todas essas informações é apresentada na Tabela 8.

Tabela 8. Análise de variância para os dados de área sob a curva do progresso da doença.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	2	0,515372	0,257686	1,89 (0,1937)
Cultivares	2	1,627261	0,813630	1,45 (0,3365)
Erro a	4	2,247697	0,561924	
Tratamento (fongorene)	2	49,977598	24,988799	183,11 (< ,0001)
Cultivar*tratamento	4	0,908262	0,227065	1,66 (0,2225)
Erro b	12	1,637626	0,136469	
Total	26	56,913816		

Lembrando que os tratamentos com o fungicida fongorene são do tipo quantitativo, igualmente espaçados, e altamente significativos em seu efeito, cabe agora um estudo mais detalhado desse efeito, por meio do estudo de regressão. Nesse caso, é possível estimar os efeitos linear e quadrático, pela aplicação dos polinômios ortogonais. Tem-se, então,

$$SQFong.Linear = \frac{(-1*54{,}13740 + 0*30{,}68524 + 1*26{,}21890)^2}{9*2} = \frac{(-27{,}91850)^2}{18} = 43{,}302396$$

e

$$SQFong.Quadrático = \frac{(1*54{,}13740 + (-2)*30{,}68524 + 1*26{,}21890)^2}{9*6} = \frac{(18{,}98582)^2}{54} = 6{,}675202$$

Ambos os efeitos são significativos ao aplicar o teste de F. A equação de resposta obtida é igual a $LSACPD = B_1M_1P_1 + B_2M_2P_2$ ou

$$LASCPD = \frac{-27{,}91850}{18} * 1 * \left(\frac{X - 400}{400}\right) + \frac{18{,}98582}{54} * 3 * \left[\left(\frac{X - 400}{400}\right)^2 - \frac{9 - 1}{12}\right]$$

$$LASCPD = 6{,}01527 - 0{,}00915X + 0{,}0000066X^2$$

cujo ponto de mínimo se verifica a uma dose de fongorene (X) igual a 693 g por 100 kg de sementes, e valor de $LASCPD = 2{,}84396$.

Uso do SAS em parcelas divididas

Para fazer a análise de variância de um delineamento com parcelas divididas, dois procedimentos podem ser usados: ANOVA e GLM. O ANOVA deve ser usado quando os fatores em estudo forem qualitativos e o interesse maior for a comparação de médias. Quando pelo menos um dos fatores for quantitativo, o procedimento mais adequado é o GLM, pela necessidade de estabelecer contrastes para estudos de regressão no fator quantitativo.

No presente caso, o segundo fator, dose de fungicida, é quantitativo, logo os efeitos linear e quadrático para esse fator podem ser estudados. Caso a interação seja significativa, pode-se também estudar esses efeitos dentro de cada um dos níveis do(s) fator(es) qualitativo(s). No exemplo que se apresenta no item *Exemplo de aplicação – dois fatores*, essa interação não foi significativa; portanto, o estudo de regressão se fará apenas para o fator.

O programa SAS para esse caso é:

```
PROC GLM;CLASS BLOCO CULTIVAR FONGOR;
MODEL PAF_L=BLOCO CULTIVAR BLOCO*CULTIVAR FONGOR CULTIVAR*FONGOR;
TEST H=CULTIVAR E=BLOCO*CULTIVAR;
CONTRAST "FONGOR_LIN" FONGOR -1 0 1;
CONTRAST "FONGOR_QUA" FONGOR 1 -2 1;
MEANS CULTIVAR/TUKEY E=BLOCO*CULTIVAR;
```

Como visto no item *Exemplo de aplicação – dois fatores*, não há efeito significativo das cultivares, mas, se houvesse, o teste de média deveria ser aplicado utilizando o erro adequado (nesse caso, erro a). Por isso, no programa exemplo, faz-se essa especificação apenas como ilustração do processo.

Os resultados se veem na Figura 4.

Caso a interação entre cultivares e fungicidas tivesse sido significativa, o estudo de regressão deveria ser feito para cada uma das cultivares, e o programa SAS então deveria incluir as seguintes seis linhas de programação:

```
CONTRAST "LIN D. CANASTRA" FONGOR -1 0 1 CULTIVAR*FONGOR -1 0 1 0 0 0 0 0 0;
CONTRAST "QUA D. CANASTRA" FONGOR 1 -2 1 CULTIVAR*FONGOR 1 -2 1 0 0 0 0 0 0;
CONTRAST "LIN D. MARAVILHA" FONGOR -1 0 1 CULTIVAR*FONGOR 0 0 0 -1 0 1 0 0 0;
CONTRAST "QUA D. MARAVILHA" FONGOR 1 -2 1 CULTIVAR*FONGOR 0 0 0 1 -2 1 0 0 0;
CONTRAST "LIN D. R.PARANAI" FONGOR -1 0 1 CULTIVAR*FONGOR 0 0 0 0 0 0 -1 0 1;
CONTRAST "QUA D. R.PARANAI" FONGOR 1 -2 1 CULTIVAR*FONGOR 0 0 0 0 0 0 1 -2 1;
```

```
                    The GLM Procedure
                  Class Level Information
          Class          Levels    Values
          BLOCO               3    1 2 3
          CULTIVAR            3    CANASTRA MARAVILHA R.PARANAIBA
          FONGOR              3    0 400 800
                  Number of observations    27
                    The GLM Procedure

Dependent Variable: PAF_L
                                Sum of
 Source                  DF       Squares    Mean Square   F Value   Pr > F
 Model                   14    55.27617485     3.94829820     28.93   <.0001
 Error                   12     1.63762689     0.13646891
 Corrected Total         26    56.91380174
               R-Square     Coeff Var      Root MSE    PAF_L Mean
               0.971226      8.982457      0.369417      4.112650

 Source                  DF     Type I SS    Mean Square   F Value   Pr > F
 BLOCO                    2     0.51537378     0.25768689      1.89   0.1937
 CULTIVAR                 2     1.62725528     0.81362764      5.96   0.0159
 BLOCO*CULTIVAR           4     2.24769971     0.56192493      4.12   0.0251
 FONGOR                   2    49.97757940    24.98878970    183.11   <.0001
 CULTIVAR*FONGOR          4     0.90826667     0.22706667      1.66   0.2225

 Source                  DF    Type III SS    Mean Square   F Value   Pr > F
 BLOCO                    2     0.51537378     0.25768689      1.89   0.1937
 CULTIVAR                 2     1.62725528     0.81362764      5.96   0.0159
 BLOCO*CULTIVAR           4     2.24769971     0.56192493      4.12   0.0251
 FONGOR                   2    49.97757940    24.98878970    183.11   <.0001
 CULTIVAR*FONGOR          4     0.90826667     0.22706667      1.66   0.2225

 Contrast                DF    Contrast SS    Mean Square   F Value   Pr > F
 FONGOR_LIN               1    43.30236901    43.30236901    317.31   <.0001
 FONGOR_QUA               1     6.67521039     6.67521039     48.91   <.0001

 Tests of Hypotheses Using the Type III MS for BLOCO*CULTIVAR as an Error
Term
 Source                  DF    Type III SS    Mean Square   F Value   Pr > F
 CULTIVAR                 2     1.62725528     0.81362764      1.45   0.3365

              Tukey's Studentized Range (HSD) Test for PAF_L
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.
                Alpha                                      0.05
                Error Degrees of Freedom                      4
                Error Mean Square                      0.561925
                Critical Value of Studentized Range    5.04024
                Minimum Significant Difference          1.2594
        Means with the same letter are not significantly different.
      Tukey Grouping          Mean      N    CULTIVAR
                  A        4.3785      9    R.PARANAIBA
                  A        4.1731      9    CANASTRA
                  A        3.7864      9    MARAVILHA
```

Figura 4. Exemplo de uso do SAS para ensaios com parcelas divididas.

Parcelas subdivididas

Como foi afirmado anteriormente, o processo de subdivisão de parcelas pode ser estendido segundo a necessidade do pesquisador e do aumento no número de fatores sendo estudados. Deve-se tomar cuidado nessas subdivisões, pois, apesar de uma aparente maior simplicidade na execução do experimento a campo, esse fato acarreta maiores complicações na análise dos dados, com o surgimento de novos erros experimentais (um para cada divisão); portanto, há uma maior diversidade de estimadores de variância para

as comparações de médias, como já foi visto no caso de parcelas com apenas uma divisão, e que aumentam de forma extremamente rápida à medida que se aumentam os níveis de subdivisões. Nesse caso de parcelas subdivididas, pelo menos três fatores estão envolvidos: um para parcelas, outro para subparcelas e um terceiro para as sub-subparcelas. Deve ser lembrado que também aqui mais de um fator pode ser alocado a cada nível de divisão de parcela.

No caso de apenas três fatores, em blocos completos casualizados, o modelo matemático adotado é:

$$Y_{ijkl} = m + b_j + A_i + ea_{ij} + B_k + AB_{ik} + eb_{ijk} + C_l + AC_{il} + BC_{kl} + ABC_{ikl} + ec_{ijkl} \quad [4]$$

Existem nesse caso três erros experimentais: o erro a (ea) para o teste de efeito do fator A, aplicado nas parcelas; o erro b (eb) para o teste do fator B, aplicado nas subparcelas, e de sua interação com o fator A; e o erro c (ec) para testar os efeitos do fator C, aplicado às sub-subparcelas, e de suas interações simples com os fatores A e B e da interação dupla entre os três fatores, desde que assumido que todos os fatores têm efeito fixo e não aleatório. No caso em que todos ou apenas um ou dois fatores sejam aleatórios, há a necessidade de estimação dos valores esperados dos quadrados médios para a determinação da forma correta dos testes F.

Para as comparações de médias dos efeitos principais utilizam-se os erros experimentais correspondentes, como visto no parágrafo anterior. Para as comparações de médias que envolvem as combinações entre dois fatores apenas, podem ser aplicadas as considerações e combinações adequadas entre os erros experimentais, como visto para parcelas divididas. No que se refere às comparações que envolvam combinações entre os três fatores, dependendo da combinação de interesse, os estimadores das variâncias dos contrastes podem envolver um, dois ou até mesmo os três erros experimentais.

Exemplo de aplicação

Em um ensaio realizado em 1996-1997 na fazenda Capivara, o engenheiro-agrônomo Evane Ferreira, da Embrapa Arroz e Feijão, testou em um delineamento em blocos completos, com parcelas subdivididas, três fatores: épocas de plantio (25/11, 12/12 e 26/12/1996), aplicação de inseticidas (sem e com aplicação de fipronil) no sulco de plantio e cultivares (num total de 21 cultivares e linhagens).

As Tabelas 9 e 10 mostram os resultados obtidos com a variável peso em gramas de 100 espiguetas, para a primeira e a terceira época de plantio (EPC), as duas aplicações de

Tabela 9. Dados de peso de 100 espiguetas (g) de arroz. Delineamento em blocos completos, com parcelas subdivididas.

EPC	INS	GEN	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
1	1	CNA 8441	2,36	2,31	2,35	2,29	9,31
1	1	CNA 8394	3,07	2,9	2,88	2,74	11,59
1	1	Canastra	2,60	2,58	2,55	2,63	10,36
Total de subparcelas			8,03	7,79	7,78	7,66	
1	2	CNA 8441	2,49	2,39	2,26	2,21	9,35
1	2	CNA 8394	2,86	2,74	2,73	2,85	11,18
1	2	Canastra	2,64	2,55	2,77	2,72	10,68
Total de subparcelas			7,99	7,68	7,76	7,78	
Total das parcelas			16,02	15,47	15,54	15,44	
3	1	CNA 8441	1,95	1,88	1,93	1,92	7,68
3	1	CNA 8394	2,77	2,88	2,34	2,78	10,77
3	1	Canastra	2,44	2,41	2,47	2,52	9,84
Total de subparcelas			7,16	7,17	6,74	7,22	
3	2	CNA 8441	2,39	2,32	2,42	2,48	9,61
3	2	CNA 8394	2,75	2,91	3,02	2,96	11,64
3	2	Canastra	2,63	2,63	2,64	2,62	10,52
Total de subparcelas			7,77	7,86	8,08	8,06	
Total das parcelas			14,93	15,03	14,82	15,28	
Total dos blocos			30,95	30,50	30,36	30,72	122,53

EPC = época de plantio; INS = aplicações de fipronil; GEN = três cultivares.

fipronil (INS) e três cultivares (GEN). As épocas de plantio foram implantadas nas parcelas principais, o inseticida nas subparcelas e os genótipos nas sub-subparcelas.

A análise de variância encontra-se na Tabela 11.

As somas de quadrados foram obtidas a partir das fórmulas dadas para o caso de dois fatores e expandidas para três fatores:

$$SQEPC = \frac{1}{24}(62,47^2 + 60,06^2) - \frac{(122,53)^2}{48} = 0,1210$$

$$SQREP = \frac{1}{12}(30,95^2 + 30,50^2 + 30,36^2 + 30,72^2) - \frac{(122,53)^2}{48} = 0,0167$$

$$SQErro(a) = \frac{1}{6}(16{,}02^2 + 15{,}47^2 + ... + 15{,}28^2) - \frac{(122{,}53)^2}{48} - SQREP - SQEPC = 0{,}0395$$

$$SQINS = \frac{1}{24}(59{,}55^2 + 62{,}98^2) - \frac{(122{,}53)^2}{48} = 0{,}2451$$

$$SQEPC * INS = \frac{1}{12}(31{,}21^2 + 31{,}26^2 + 28{,}29^2 + 31{,}77^2) - \frac{(122{,}53)^2}{48} - SQEPC - SQINS = 0{,}2596$$

$$SQErro(b) = \frac{1}{3}(8{,}03^2 + 7{,}79^2 + ... + 8{,}06^2) - \frac{(122{,}53)^2}{48} - SQREP - SQEPC - SQErro(a) - SQINS - SQEPC * INS = 0{,}0583$$

$$SQGEN = \frac{1}{16}(35{,}95^2 + 45{,}18^2 + 41{,}40^2) - \frac{(122{,}53)^2}{48} = 2{,}6913$$

$$SQEPC * GEN = \frac{1}{8}(18{,}66^2 + 22{,}77^2 + ... + 20{,}36^2) - \frac{(122{,}53)^2}{48} - SQEPC - SQGEN = 0{,}0333$$

$$SQINS * GEN = \frac{1}{8}(16{,}99^2 + 22{,}36^2 + ... + 21{,}20^2) - \frac{(122{,}53)^2}{48} - SQINS - SQGEN = 0{,}0333$$

$$SQEPC * INS * GEN = \frac{1}{4}(9{,}31^2 + 11{,}59^2 + ... + 10{,}52^2) - \frac{(122{,}53)^2}{48} - SQEPC - SQINS - SQGEN - SQEPC * INS$$
$$SQEPC * GEN - SQINS * GEN = 0\ 0333$$

$$SQTotal = (2{,}36^2 + 2{,}31^2 + ... + 2{,}62^2) - \frac{(122{,}53)^2}{48} = 3{,}8858$$

A $SQErro(c)$ foi obtida por diferença entre a soma de quadrados total e a soma de todas as demais somas de quadrado.

A partir da análise de variância, conclui-se facilmente que os três efeitos principais foram significativos, isto é, existe efeito de época de plantio de aplicação de inseticidas e de

Tabela 10. Quadros auxiliares para obtenção das somas de quadrados dos fatores principais e suas interações simples. Delineamento em blocos ao acaso com parcelas subdivididas.

INS			
EPC	1	2	Total
1	31,26	31,21	62,47
3	28,29	31,77	60,06
Total	59,55	62,98	122,53

GEN				
EPC	CNA 8441	CNA 8394	Canastra	Total
1	18,66	22,77	21,04	62,47
3	17,29	22,41	20,36	60,06
Total	35,95	45,18	41,40	122,53

GEN				
INS	CNA 8441	CNA 8394	Canastra	Total
1	16,99	22,36	20,20	59,55
2	18,96	22,82	21,20	62,98
Total	35,95	45,18	41,40	122,53

EPC = época de plantio; INS = aplicações de fipronil; GEN = três cultivares.

Tabela 11. Análise de variância para o experimento em parcelas subdivididas em blocos ao acaso.

Fonte	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
REP	3	0,0167	0,0056	
EPC	1	0,1210	0,1210	9,17 (0,0564)
Erro a	3	0,0395	0,0132	
INS	1	0,2451	0,2451	25,27 (0,0024)
EPC*INS	1	0,2596	0,2596	26,76 (0,0021)
Erro b	6	0,0583	0,0097	
GEN	2	2,6913	1,3457	118,04 (< ,0001)
EPC*GEN	2	0,0333	0,0167	1,46 (0,2520)
GEN*INS	2	0,0732	0,0366	3,21 (0,0582)
EPC*GEN*INS	2	0,0742	0,0371	3,25 (0,0563)
Erro c	24	0,2736	0,0114	
Total	47	3,8858		

EPC = época de plantio; INS = aplicações de fipronil; GEN = três cultivares.

genótipos, e apenas para este último caberia um teste de médias. Das interações, apenas a interação entre época de plantio e inseticidas foi significativa. Um estudo detalhado dessa interação, deixada a cargo do leitor, mostraria que não há efeito de inseticida na primeira época de plantio, mas um efeito significativo na segunda.

Uso do SAS em parcelas subdivididas em blocos completos ao acaso

Para o exemplo em estudo, o procedimento ANOVA é suficiente para que se faça a análise de variância e para que se conheçam as médias dos tratamentos. Nesse caso em particular, apenas as médias de genótipos poderiam sofrer um teste de médias do tipo SNK, Tukey e, até mesmo, DUNNETT, ao se considerar a cultivar Canastra como a testemunha perante as duas linhagens em estudo. O programa a seguir mostra como se obtém essa análise, com as indicações de erros adequados para o teste F para fatores nas parcelas principais e subparcelas. Os comandos para teste de médias são deixados ao leitor como exercício.

```
PROC ANOVA;CLASS BLOCO EPC INS GEN;
MODEL P100E=BLOCO EPC BLOCO*EPC INS EPC*INS BLOCO*EPC*INS GEN EPC*GEN
INS*GEN EPC*INS*GEN;
TEST H=EPC E=BLOCO*EPC;
TEST H=INS EPC*INS E=BLOCO*EPC*INS;
```

Os resultados dessa análise aparecem na Figura 5.

Experimentos em faixas

Em algumas situações experimentais, pode ocorrer que mais de um fator em estudo requeira parcelas de tamanho muito grande. Nesse caso, uma das soluções é a utilização de uma modificação do esquema de parcelas divididas, chamada de experimento em faixa. Para tornar claro o procedimento, imagine-se um ensaio em que dois fatores são o interesse maior do pesquisador. Um deles é a forma de preparo do solo (fator A): aração normal (A1), plantio direto (A2) e aração com subsolagem (A3), por exemplo. O outro é o uso de fertilizantes fosfatados para melhoria do solo (fator B), com aplicação de grandes quantidades para incorporação ao solo e cujas fontes sejam: superfosfato simples (B1), superfosfato triplo (B2) e fosfato natural (B3).

```
                    The ANOVA Procedure
                  Class Level Information
               Class         Levels    Values
               BLOCO              4    1 2 3 4
               EPC                2    1 3
               INS                2    1 2
               GEN                3    CANASTRA CNA.8394 CNA.8441
                             Number of observations    48
                                The ANOVA Procedure
Dependent Variable: P100E
                                Sum of
Source                 DF      Squares     Mean Square    F Value    Pr > F
Model                  23   3.61211458      0.15704846      13.77    <.0001
Error                  24   0.27363333      0.01140139
Corrected Total        47   3.88574792

                  R-Square     Coeff Var      Root MSE    P100E Mean
                  0.929580      4.182902      0.106777      2.552708

Source                 DF        Anova SS     Mean Square    F Value    Pr > F
BLOCO                   3   0.01668958      0.00556319       0.49    0.6939
EPC                     1   0.12100208      0.12100208      10.61    0.0033
BLOCO*EPC               3   0.03947292      0.01315764       1.15    0.3477
INS                     1   0.24510208      0.24510208      21.50    0.0001
EPC*INS                 1   0.25960208      0.25960208      22.77    <.0001
BLOCO*EPC*INS           6   0.05827917      0.00971319       0.85    0.5433
GEN                     2   2.69132917      1.34566458     118.03    <.0001
EPC*GEN                 2   0.03330417      0.01665208       1.46    0.2520
INS*GEN                 2   0.07317917      0.03658958       3.21    0.0582
EPC*INS*GEN             2   0.07415417      0.03707708       3.25    0.0563

Tests of Hypotheses Using the Anova MS for BLOCO*EPC as an Error Term
Source                 DF        Anova SS     Mean Square    F Value    Pr > F
EPC                     1   0.12100208      0.12100208       9.20    0.0562

Tests of Hypotheses Using the Anova MS for BLOCO*EPC*INS as an Error Term
Source                 DF        Anova SS     Mean Square    F Value    Pr > F
INS                     1   0.24510208      0.24510208      25.23    0.0024
EPC*INS                 1   0.25960208      0.25960208      26.73    0.0021
```

Figura 5. Análise de variância de um ensaio em parcela subdividida com o SAS.

Adotando-se o delineamento em blocos completos, com quatro repetições, a distribuição dos tratamentos às parcelas se faz de forma sequencial, sorteando-se primeiro os níveis do fator A para as parcelas constituídas no sentido horizontal, por exemplo; e, em seguida, os níveis do fator B às parcelas formadas num sentido vertical. Dessa forma, têm-se as seguintes fases: fase 1 – casualização do fator A em cada bloco (Figura 6); fase 2 – casualização do fator B, de forma perpendicular ao fator A em cada bloco (Figura 7).

Os efeitos dos tratamentos serão então medidos em cada uma das pequenas parcelas formadas pela intersecção das parcelas maiores. Na Figura 7, a unidade experimental X, por exemplo, conteria os efeitos de preparo de solo A2 = plantio direto e uso de superfosfato simples (B1); a unidade Y conteria a combinação A3B3, ou seja, aração com subsolagem

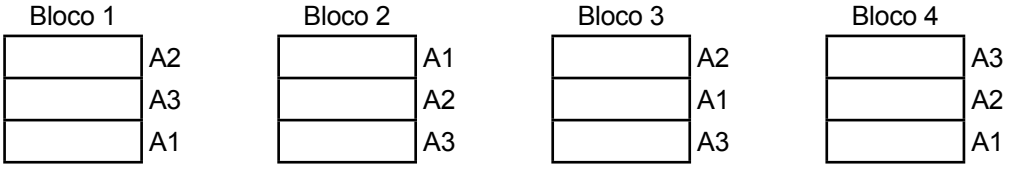


Figura 6. Primeira fase de casualização de tratamentos em um experimento em faixa em blocos completos ao acaso.

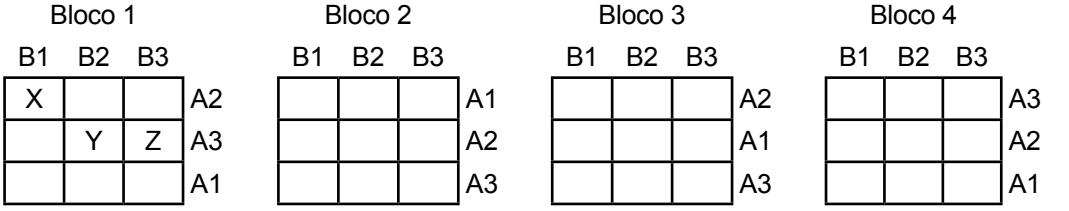


Figura 7. Segunda fase de casualização de tratamentos em um experimento em faixa em blocos completos ao acaso.

mais o uso de fosfato natural, enquanto a unidade Z teria a combinação A3B2, isto é, aração com subsolagem mais superfosfato triplo.

O modelo matemático deve ser então modificado a partir daquele empregado no caso das parcelas divididas, a fim de refletir essa nova restrição de casualização, que implica agora na existência de três erros experimentais: um para o fator aplicado na faixa horizontal, outro para aquele nas faixas verticais e um terceiro para o teste das interações entre os dois fatores. Dessa forma, tem-se então:

$$Y_{ijk} = m + b_j + A_i + ea_{ij} + B_k + eb_{jk} + AB_{ik} + ec_{ijk} \quad [5]$$

A análise de variância apresenta o esquema mostrado na Tabela 12 quando os fatores são considerados fixos.

Quando algum dos fatores for considerado aleatório, apenas a coluna referente ao teste F deve ser modificada segundo a esperança do quadrado médio respectivo.

Para a comparação de médias, devem ser considerados os mesmos casos discutidos em parcelas divididas. Para as comparações entre as médias do fator A, usa-se o quadrado médio do erro a; para as médias do fator B, usa-se o erro b; para as comparações de médias de A dentro de cada nível de B, é necessária uma combinação entre os erros a e c; e para as comparações entre médias de B dentro de cada nível de A, usa-se uma combinação entre

Tabela 12. Modelo de análise de variância de um experimento em faixa no delineamento em blocos completos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F
Blocos (R)	$J - 1$	$\frac{1}{IK}\sum_{j} Y_{.j.}^2 - FC$	SQR/GLR	$QMR/QMEa$
Fator A (A)	$I - 1$	$\frac{1}{JK}\sum_{i} Y_{i..}^2 - FC$	SQA/GLA	$QMA/QMEa$
Erro a (Ea)	$(J - 1)(I - 1)$	$\frac{1}{K}\sum_{ij} Y_{ij.}^2 - FC - SQR - SQA$	$SQEa/GLEa$	
Fator B (B)	$K - 1$	$\frac{1}{IJ}\sum_{k} Y_{..k}^2 - FC$	SQB/GLB	$QMB/QMEb$
Erro b (Eb)	$(J - 1)(K - 1)$	$\frac{1}{I}\sum_{i} Y_{i..}^2 - FC - SQR - SQB$	$SQEb/GLEb$	
A*B (AB)	$(I - 1)(K - 1)$	$\frac{1}{J}\sum_{ik} Y_{i.k}^2 - FC - SQA - SQB$	$SQAB/GLAB$	$QMAB/QMEc$
Erro c (Ec)	$(I - 1)(J - 1)(K - 1)$	Por diferença	$SQEc/GLEc$	
Total	$IJK - 1$	$\sum_{ijk} Y_{ijk}^2 - FC$		

os erros b e c. Para qualquer outra combinação de médias, é necessária uma combinação entre os três erros. Essas formas são mostradas na Tabela 13 (ZIMMERMANN, 1973).

Assim como no caso das parcelas divididas, mais fatores podem ser testados e o esquema de análise de variância é modificado para representar a existência desses novos elementos. Por exemplo, um terceiro fator poderia ser considerado e ser aplicado também às parcelas verticais, juntamente com o fator B, de forma que as parcelas naquele sentido contivessem, cada uma delas, uma das combinações entre os fatores B e C. O modelo matemático apresentaria agora a forma a seguir, com o quadro de análise de variância modificado segundo ele.

$$Y_{ijk} = m + b_j + A_i + ea_{ij} + B_k + C_l + BC_{jkl} + AB_{ik} + AC_{il} + ABC_{ikl} + ec_{ijk} \quad [6]$$

Para a obtenção das somas de quadrados e de suas respectivas esperanças, podem ser aplicadas as regras práticas mostradas no Capítulo 8.

Tabela 13. Estimativa da variância da diferença de duas médias para um experimento em faixa no delineamento em blocos completos ao acaso.

Comparação de médias	Variância da diferença entre duas médias
$(\overline{A}_i - \overline{A}_{i'})$	$(2*QMEa)/(J*K)$
$(\overline{B}_k - \overline{B}_{k'})$	$(2*QMEb)/(J*I)$
$(\overline{A}_i\overline{B}_k - \overline{A}_i\overline{B}_{k'})$	$2[QMEb + (I - 1)QMEc]/IJ$
$(\overline{A}_i\overline{B}_k - \overline{A}_{i'}\overline{B}_k)$	$2[QMEa + (\mathrm{K} - 1)QMEc]/JK$
$(\overline{A}_i\overline{B}_k - \overline{A}_{i'}\overline{B}_{k'})$	$2[IQMEa + KQMEb + (IK - I - K)QMEc]/IJK$

Exemplo de aplicação

No ano agrícola de 1998–1999, o pesquisador Cleber Guimarães conduziu dois experimentos para testar o efeito da aplicação de adubos em arroz após pastagem e empregou nesses ensaios três cultivares. Um experimento foi conduzido com preparo de solo com arado escarificador, e o outro com grade aradora. As aplicações de adubo foram feitas nas doses descritas na Tabela 14.

Tabela 14. Formulação de adubação empregada no ensaio de cultivo de arroz após pasto.[1]

	N	P_2O_5	K_2O	Micros	Zn	Quantidade da mistura aplicada
1	6	30	27	45	2	175 kg ha^{-1}
2	12	60	54	90	4	350 kg ha^{-1}
3	18	90	81	135	6	525 kg ha^{-1}

[1] Dados cedidos pelo pesquisador Cleber Guimarães, da Embrapa Arroz e Feijão.

As cultivares empregadas foram: Caiapó, Maravilha e Primavera. Para esse exemplo, estão sendo usadas as duas últimas. Os dados de produção de grãos em kg ha^{-1} são apresentados na Tabela 15.

A Tabela 16 contém os totais dos fatores e de todas as combinações entre eles.

A análise de variância se mostra completa na Tabela 17, na qual as somas de quadrados foram obtidas pela aplicação das fórmulas dadas anteriormente, algumas das quais se mostram como exemplo de uso.

Tabela 15. Dados de produção de grãos de arroz, em ensaio em faixas em blocos ao acaso.

Bloco 1			
	Maravilha	Primavera	Total
Adubação 1	1.752,76	2.983,33	4.736,09
Adubação 2	1.913,57	3.232,41	5.145,98
Adubação 3	1.852,62	2.894,26	4.746,88
Total	5.518,95	9.110,00	14.628,95
Bloco 2			
	Maravilha	Primavera	Total
Adubação 1	1.686,29	2.954,67	4.640,96
Adubação 2	1.814,98	2.931,08	4.746,06
Adubação 3	1.738,03	2.717,51	4.455,54
Total	5.239,30	8.603,26	13.842,56
Bloco 3			
	Maravilha	Primavera	Total
Adubação 1	1.631,19	2.628,98	4.260,17
Adubação 2	1.811,67	2.915,48	4.727,15
Adubação 3	1.678,96	2.688,05	4.367,01
Total	5.121,82	8.232,51	13.354,33
Bloco 4			
	Maravilha	Primavera	Total
Adubação 1	1.535,76	2.627,62	4.163,38
Adubação 2	1.782,22	2.863,81	4.646,03
Adubação 3	1.634,93	2.651,23	4.286,16
Total	4.952,91	8.142,66	13.095,57

Tabela 16. Quadro auxiliar para obtenção das somas de quadrados dos fatores e sua interação.

	Maravilha	Primavera	Total
Adubação 1	6.606,00	11.194,60	17.800,60
Adubação 2	7.322,44	11.942,78	19.265,22
Adubação 3	6.904,54	10.951,05	17.855,59
Total	20.832,98	34.088,43	54.921,41

Tabela 17. Análise de variância para o ensaio em faixa no delineamento em blocos ao acaso. Dados de produção de arroz.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	3	227.401,62	75.800,54	
Adubação	2	172.299,68	86.149,84	19,36 (0,0024)
Erro a	6	26.692,62	4.448,76	
Cultivar	1	7.321.123,11	7.321.123,11	968,79 (< 0,0001)
Erro b	3	22.670,94	7.556,98	
Adubação*Cultivar	2	26.006,24	13.003,12	3,90 (0,0822)
Erro c	6	20.003,89	3.333,98	
Total	23	7.816.198,11		

$$SQBlocos = \frac{1}{6}(14.628,95^2 + 13.842,56^2 + 13.354,33^2 + 13.095,57^2) - \frac{(54.921,41)^2}{24} = 227.401,62$$

$$SQAdub = \frac{1}{8}(17.800,60^2 + 19.265,22^2 + 17.855,59^2) - \frac{(54.921,41)^2}{24} = 172.299,68$$

$$SQErro(a) = \frac{1}{2}(4.736,09^2 + 5.145,98^2 + ... + 4.286,16^2) - \frac{(54.921,41)^2}{24} - SQBlocos - SQAdub = 26.692,62$$

Considerando-se o efeito significativo de adubação, pode-se fazer um estudo adicional de seu efeito e, nesse caso, medir os componentes linear e quadrático de uma possível regressão, como se vê na Tabela 18.

Como se pode ver, todo efeito de adubação se deve ao efeito quadrático e, além disso, poder-se-ia obter a equação de regressão. As somas de quadrados linear e quadrática

Tabela 18. Análise de variância detalhada para o fator adubação, medindo os efeitos de possíveis modelos de regressão linear e quadrática.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Adubação	2	172.299,68	86.149,84	19,36 (0,0024)
Linear	1	188,99	188,99	0,04 (0,8435)
Quadrático	1	172.110,69	172.110,69	38,69 (0,0008)
Erro a	6	26.692,62	4.448,76	

foram obtidas usando os coeficientes dos polinômios ortogonais de primeiro e segundo grau, para doses igualmente espaçadas como mostrado na Tabela 1 do Capítulo 8.

SAS e os experimentos em faixa

Para este exemplo, um dos fatores – nível de adubação – é quantitativo; portanto, é possível fazer também um estudo da resposta desse fator no que se refere à equação de regressão. Logo, o programa SAS deve utilizar o procedimento GLM, como se vê a seguir:

```
PROC GLM;CLASS BLOCO ADUBACAO CULTIVAR;
MODEL PRODUCAO=BLOCO ADUBACAO BLOCO*ADUBACAO CULTIVAR
BLOCO*CULTIVAR ADUBACAO*CULTIVAR;
TEST H=ADUBACAO E=BLOCO*ADUBACAO;
TEST H=CULTIVAR E=BLOCO*CULTIVAR;
CONTRAST "ADUB LIN" ADUBACAO -1 0 1/E=BLOCO*ADUBACAO;
CONTRAST "ADUB QUA" ADUBACAO 1 -2 1/E=BLOCO*ADUBACAO;
```

O resultado se vê na Figura 8.

Experimentos em faixas com parcelas divididas

Em muitos ensaios, é possível que dois fatores necessitem de uma parcela maior e que um terceiro, por exemplo, requeira área menor. Nesse caso, uma das soluções é o uso de experimentos em faixa para os dois fatores que requeiram a área maior e a subdivisão das parcelas para acomodar o terceiro fator em estudo. A Figura 9 mostra como ficaria o experimento para uma repetição.

```
                    The GLM Procedure
                  Class Level Information
                  Class         Levels    Values
                  BLOCO              4    1 2 3 4
                  ADUBACAO           3    1 2 3
                  CULTIVAR           2    MARAVILHA PRIMAVERA
                               Number of observations    24
                    The GLM Procedure
Dependent Variable: PRODUCAO
                              Sum of
Source                 DF        Squares     Mean Square    F Value    Pr > F
Model                  17     7796194.220      458599.660     137.55    <.0001
Error                   6       20003.887        3333.981
Corrected Total        23     7816198.107

                  R-Square     Coeff Var      Root MSE    PRODUCAO Mean
                  0.997441      2.523197      57.74064          2288.392

Source                 DF       Type I SS     Mean Square    F Value    Pr > F
BLOCO                   3      227401.622       75800.541      22.74    0.0011
ADUBACAO                2      172299.683       86149.841      25.84    0.0011
BLOCO*ADUBACAO          6       26692.617        4448.769       1.33    0.3675
CULTIVAR                1     7321123.113     7321123.113    2195.91    <.0001
BLOCO*CULTIVAR          3       22670.941        7556.980       2.27    0.1810
ADUBACAO*CULTIVAR       2       26006.244       13003.122       3.90    0.0822

Source                 DF      Type III SS     Mean Square    F Value    Pr > F
BLOCO                   3      227401.622       75800.541      22.74    0.0011
ADUBACAO                2      172299.683       86149.841      25.84    0.0011
BLOCO*ADUBACAO          6       26692.617        4448.769       1.33    0.3675
CULTIVAR                1     7321123.113     7321123.113    2195.91    <.0001
BLOCO*CULTIVAR          3       22670.941        7556.980       2.27    0.1810
ADUBACAO*CULTIVAR       2       26006.244       13003.122       3.90    0.0822

Tests of Hypotheses Using the Type III MS for BLOCO*ADUBACAO as an Error
Term
Source                 DF      Type III SS     Mean Square    F Value    Pr > F
ADUBACAO                2     172299.6826      86149.8413      19.36    0.0024

Tests of Hypotheses Using the Type III MS for BLOCO*CULTIVAR as an Error
Term
Source                 DF      Type III SS     Mean Square    F Value    Pr > F
CULTIVAR                1     7321123.113     7321123.113     968.79    <.0001

Tests of Hypotheses Using the Type III MS for BLOCO*ADUBACAO as an Error
Term
Contrast               DF      Contrast SS     Mean Square    F Value    Pr > F
ADUB LIN                1        188.9938        188.9938       0.04    0.8435
ADUB QUA                1     172110.6888     172110.6888      38.69    0.0008
```

Figura 8. Análise de variância de um experimento em faixa com o uso do procedimento GLM do SAS.

Bloco 1

B1		B3		B2		
C1	C2	C2	C1	C1	C2	A1
C2	C1	C1	C2	C1	C2	A3
C2	C1	C2	C1	C2	C1	A2

Figura 9. Esquema de um bloco em um experimento em faixas com parcelas divididas. Fator A compondo as faixas horizontais, fator B as faixas verticais e fator C nas subparcelas.

O modelo matemático, nesse caso, teria a seguinte expressão:

$$Y_{ijk} = m + b_j + A_i + ea_{ij} + B_k + eb_{jk} + AB_{ik} + ec_{ikj} + C_l + AC_{il} + BC_{kl} + ABC_{ikl} + ed_{ijkl} \quad [7]$$

A análise de variância, nesse caso, tem a forma mostrada na Tabela 19.

Os testes F para comprovação dos efeitos das distintas causas de variação devem ser determinados segundo as pressuposições feitas ao modelo (fixo, aleatório ou misto). No caso de ser considerado fixo, os testes F se fazem sempre com o erro especificado imediatamente após os efeitos respectivos. Assim, tratamentos A serão testados com o erro a, tratamentos B com o erro b, etc. Caso o modelo seja considerado aleatório ou misto, os testes F devem ser determinados segundo os componentes de variância de cada uma das fontes de variação.

Para as comparações de médias, os mesmos critérios devem ser seguidos, lembrando que, nesse caso, existe uma gama enorme de possibilidades.

Esse procedimento pode ser adotado para outros tipos de delineamentos, como o caso de um quadrado latino, devendo-se apenas lembrar que o modelo matemático deve ser modificado para acomodar as características do delineamento que se adote.

Exemplo de aplicação

Os dados mostrados na Tabela 20 referem-se à massa de 100 grãos de feijão obtidos em um experimento realizado por Luis Fernando Stone e um de seus estudantes Regis Vilela Bagatini, em 1998, no delineamento em blocos completos ao acaso, com duas repetições. Nesse experimento, foram testadas nas faixas horizontais as lâminas de água aplicada em irrigação por aspersão, nas faixas verticais as formas de preparo de solo e nas subparcelas doses de N. O experimento foi realizado na fazenda Capivara, sede da Embrapa Arroz e Feijão.

Para a realização da análise de variância, constroem-se alguns quadros auxiliares, que se mostram na Tabela 21, pois eles facilitam em muito a realização desta análise e dão ao investigador um melhor conhecimento dos resultados que está obtendo.

A análise de variância, executada segundo as fórmulas dadas na Tabela 19, está retratada na Tabela 22. Como exemplo de aplicação, tem-se que:

Tabela 19. Modelo de análise de variância de um experimento em faixas com parcelas divididas em blocos completos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
Blocos (R)	$J - 1$	$\frac{1}{IKL}\sum_{j} Y_{.j..}^2 - FC$	SQR/GLR
Fator A (A)	$I - 1$	$\frac{1}{JKL}\sum_{i} Y_{i...}^2 - FC$	SQA/GLA
Erro a (Ea)	$(J - 1)(I - 1)$	$\frac{1}{KL}\sum_{ij} Y_{ij..}^2 - FC - SQR - SQA$	$SQEa/GLEa$
Fator B (B)	$K - 1$	$\frac{1}{IJL}\sum_{k} Y_{..k.}^2 - FC$	SQB/GLB
Erro b (Eb)	$(J - 1)(K - 1)$	$\frac{1}{IL}\sum_{jk} Y_{.jk.}^2 - FC - SQR - SQB$	$SQEb/GLEb$
A*B (AB)	$(I - 1)(K - 1)$	$\frac{1}{JL}\sum_{ik} Y_{i.k.}^2 - FC - SQA - SQB$	$SQAB/GLAB$
Erro c (Ec)	$(I - 1)(J - 1)(K - 1)$	$\frac{1}{L}\sum_{ijk} Y_{ijk.}^2 - FC - SQA - SQB - SQR - SQEa - SQEb - SQAB$	$SQEc/GLEc$

Continua...

Tabela 19. Continuação.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)
C	$L - 1$	$\frac{1}{IJK}\sum_{l} Y_{...l}^2 - FC$	SQC/GLC
A*C (AC)	$(I - 1)(L - 1)$	$\frac{1}{JK}\sum_{il} Y_{i..l}^2 - FC - SQA - SQC$	$SQAC/GLAC$
B*C (BC)	$(K - 1)(L - 1)$	$\frac{1}{JI}\sum_{kl} Y_{..kl}^2 - FC - SQB - SQC$	$SQBC/GLBC$
A*B*C (ABC)	$(I - 1)(K - 1)(L - 1)$	$\frac{1}{L}\sum_{ijk} Y_{ijk.}^2 - FC - SQA - SQB - SQC - SQAB - SQAC - SQBC$	$SQABC/GLABC$
Erro d (Ed)	$(J - 1)(L - 1)IK$	Por diferença	$SQEd/GLEd$
Total	$IJK - 1$	$\sum_{ijk} Y_{ijk}^2 - FC$	

Tabela 20. Dados de massa de 100 grãos de feijão em um experimento em faixa com parcelas divididas em blocos ao acaso.

Repetição 1													
	Psolo 1				Psolo 2				Psolo 3				
	Nit 1	Nit 2	Nit 3	Soma	Nit 1	Nit 2	Nit 3	Soma	Nit 1	Nit 2	Nit 3	Soma	Total
Lam 1	26,33	27,85	27,13	81,31	25,10	27,67	24,93	77,70	25,00	28,03	29,65	82,68	241,69
Lam 2	24,04	25,22	28,32	77,58	25,19	27,77	27,28	80,24	25,89	24,27	25,83	75,99	233,81
Lam 3	25,85	25,70	26,97	78,52	25,63	27,11	25,62	78,36	26,16	24,86	25,51	76,53	233,41
Lam 4	23,20	20,32	23,94	67,46	29,28	26,03	28,60	83,91	26,23	25,49	24,65	76,37	227,74
Total	99,42	99,09	106,36	304,87	105,20	108,58	106,43	320,21	103,28	102,65	105,64	311,57	936,65
Repetição 2													
	Psolo 1				Psolo 2				Psolo 3				
	Nit 1	Nit 2	Nit 3	Soma	Nit 1	Nit 2	Nit 3	Soma	Nit 1	Nit 2	Nit 3	Soma	Total
Lam 1	25,87	28,64	29,31	83,82	27,80	27,25	25,56	80,61	28,53	26,38	32,45	87,36	251,79
Lam 2	27,16	26,49	25,99	79,64	24,63	26,91	28,47	80,01	26,68	27,64	24,80	79,12	238,77
Lam 3	27,11	24,44	28,06	79,61	25,77	27,46	26,20	79,43	26,83	27,55	27,19	81,57	240,61
Lam 4	23,00	23,43	23,42	69,85	28,71	26,45	26,25	81,41	26,64	26,82	26,88	80,34	231,60
Total	103,14	103,00	106,78	312,92	106,91	108,07	106,48	321,46	108,68	108,39	111,32	328,39	962,77

Psolo = formas de preparo do solo; Nit = nitrogênio; Lam = lâminas de água aplicada em irrigação por aspersão.

Tabela 21. Quadros auxiliares para obtenção das somas de quadrados dos fatores principais e suas interações no experimento em faixa com parcelas divididas em blocos ao acaso.

	Psolo 1	Psolo 2	Psolo 3	Total
Lam 1	165,13	158,31	170,04	493,48
Lam 2	157,22	160,25	155,11	472,58
Lam 3	158,13	157,79	158,10	474,02
Lam 4	137,31	165,32	156,71	459,34
Total	617,79	641,67	639,96	1.899,42

	Nit 1	Nit 2	Nit 3	Total
Psolo 1	202,56	202,09	213,14	617,79
Psolo 2	212,11	216,65	212,91	641,67
Psolo 3	211,96	211,04	216,96	639,96
Total	626,63	629,78	643,01	1.899,42

	Nit 1	Nit 2	Nit 3	Total
Lam 1	158,63	165,82	169,03	493,48
Lam 2	153,59	158,30	160,69	472,58
Lam 3	157,35	157,12	159,55	474,02
Lam 4	157,06	148,54	153,74	459,34
Total	626,63	629,78	643,01	1.899,42

	Psolo 1			Psolo 2			Psolo 3		
	Nit 1	Nit 2	Nit 3	Nit 1	Nit 2	Nit 3	Nit 1	Nit 2	Nit 3
Lam 1	52,20	56,49	56,44	52,90	54,92	50,49	53,53	54,41	62,10
Lam 2	51,20	51,71	54,31	49,82	54,68	55,75	52,57	51,91	50,63
Lam 3	52,96	50,14	55,03	51,40	54,57	51,82	52,99	52,41	52,70
Lam 4	46,20	43,75	47,36	57,99	52,48	54,85	52,87	52,31	51,53

Psolo = formas de preparo do solo; Nit = nitrogênio; Lam = lâminas de água aplicada em irrigação por aspersão.

$$SQLAM = \frac{1}{18}(493{,}48^2 + 472{,}58^2 + 474{,}02^2 + 459{,}34^2) - \frac{1.899{,}42^2}{72} = 32{,}9710$$

$$SQREP = \frac{1}{26}(936{,}72^2 + 962{,}77^2) - \frac{1.899{,}42^2}{72} = 9{,}4758$$

e

$$SQErroa = \frac{1}{9}(241{,}69^2 + 233{,}81^2 + \ldots + 231{,}60^2) - \frac{1.899{,}42^2}{72} - 32{,}9710 - 9{,}4758 = 1{,}2660$$

Tabela 22. Análise de variância para o experimento em faixa com parcelas divididas em blocos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Rep	1	9,4758	9,4758	
Lam	3	32,9710	10,9903	26,04 (0,0119)
Erro a	3	1,2660	0,4220	
Psolo	2	14,7873	7,3937	2,91 (0,2556)
Erro b	2	5,0775	2,5387	
Lam*Psolo	6	67,6311	11,2718	35,89 (0,0002)
Erro c	6	1,8844	0,3141	
Nit	2	6,2953	3,1476	2,11 (0,1432)
Lam*Nit	6	14,2557	2,3759	1,59 (0,1926)
Psolo*Nit	4	7,4711	1,8678	1,25 (0,3161)
Lam*Psolo*Nit	12	39,4927	3,2911	2,21 (0,0479)
Erro d	24	35,8102	1,4921	
Total	71	236,4180		

Rep = repetição; Psolo = formas de preparo do solo; Nit = nitrogênio; Lam = lâminas de água aplicada em irrigação por aspersão.

As demais somas de quadrados são deixadas para o leitor como um exercício.

Esses resultados mostram que existem efeitos significativos da lâmina de água, de sua interação com as formas de preparo de solo e da interação dos três fatores em estudo. Para um conhecimento melhor do porquê da significância das interações, podem ser usados dois procedimentos: a comparação de médias de um fator dentro de cada nível do outro fator e o desdobramento das fontes de variação no quadro da análise de variância. Como exemplo, será feita a comparação das médias de lâminas de água dentro de cada um dos preparos de solo, dado que este último fator mostrou-se não significativo na análise mostrada na Tabela 22. Para isso, pode ser utilizado o teste de Tukey, por exemplo. Nesse caso, a variância a ser usada no teste será dada por uma ponderação entre o erro a e o erro c, como mostrado no caso do delineamento em faixas. Dessa forma, tem-se

$$(0,4220 + 2*0,3141)/6 = 0,1750$$

com sete graus de liberdade, determinados pela aplicação da fórmula de Satterthwaite (1946) (fórmula 1).

O valor de DHS é igual a

$$DHS = 4{,}68(\sqrt{0{,}1750}) = 1{,}96$$

o que indica que qualquer diferença entre duas médias superior ao valor de 1,96 será considerada significativa a 5% de probabilidade.

A partir do quadro auxiliar (Tabela 21), que mostra os totais das combinações entre lâmina de água e formas de preparo de solo, obtêm-se os resultados apresentados na Tabela 23.

Tabela 23. Comparações de médias de massa de 100 grãos como efeito de lâminas de água dentro de cada uma das formas de preparo de solo pelo teste de Tukey a 5% de probabilidade.

	Preparo de solo 1	Preparo de solo 2	Preparo de solo 3
Lâmina 1	27,52 a	26,35 a	28,34 a
Lâmina 2	26,20 a	26,71 a	25,85 b
Lâmina 3	25,85 a	26,30 a	26,35 b
Lâmina 4	22,88 b	27,55 a	26,12 b

A não existência de efeitos significativos para lâminas de água na forma de preparo de solo 2 e a sua existência em formas distintas nas outras duas formas de preparo de solo mostram o porquê da interação significativa entre forma de preparo de solo e lâminas de água.

Uso do SAS e os experimentos em faixa com parcela dividida

Novamente aqui os procedimentos adotados podem ser o ANOVA, caso não haja necessidade de estudos adicionais por meio de contrastes, ou o GLM, caso isso seja interessante para o investigador (estudo por contrastes). Considerando-se os resultados mostrados na seção *Exemplo de aplicação*, e sem a informação complementar dos valores dos níveis de lâmina de água e de N, será usado o procedimento ANOVA, como segue:

```
PROC ANOVA;CLASS BLO LAM PSOL NIT;
MODEL M100=BLO LAM BLO*LAM PSOL BLO*PSOL LAM*PSOL BLO*LAM*PSOL NIT
LAM*NIT PSOL*NIT LAM*PSOL*NIT;
TEST H=LAM E=BLO*LAM;
TEST H=PSOL E=BLO*PSOL;
TEST H=LAM*PSOL E=BLO*LAM*PSOL;
```

O resultado dessa programação aparece na Figura 10.

```
                    The ANOVA Procedure
                  Class Level Information
            Class         Levels    Values
            BLO                2    1 2
            LAM                4    1 2 3 4
            PSOL               3    1 2 3
            NIT                3    1 2 3
             Number of observations    72

                    The ANOVA Procedure

Dependent Variable: M100
                               Sum of
Source                 DF         Squares     Mean Square    F Value    Pr > F
Model                  47     200.6077500       4.2682500       2.86    0.0035
Error                  24      35.8102000       1.4920917
Corrected Total        71     236.4179500

                R-Square     Coeff Var      Root MSE     M100 Mean
                0.848530      4.630301      1.221512      26.38083

Source                 DF        Anova SS     Mean Square    F Value    Pr > F
BLO                     1      9.47575556      9.47575556       6.35    0.0188
LAM                     3     32.97103889     10.99034630       7.37    0.0012
BLO*LAM                 3      1.26597778      0.42199259       0.28    0.8373
PSOL                    2     14.78732500      7.39366250       4.96    0.0158
BLO*PSOL                2      5.07746944      2.53873472       1.70    0.2037
LAM*PSOL                6     67.63105278     11.27184213       7.55    0.0001
BLO*LAM*PSOL            6      1.88439722      0.31406620       0.21    0.9700
NIT                     2      6.29527500      3.14763750       2.11    0.1432
LAM*NIT                 6     14.25566944      2.37594491       1.59    0.1926
PSOL*NIT                4      7.47105000      1.86776250       1.25    0.3161
LAM*PSOL*NIT           12     39.49273889      3.29106157       2.21    0.0479

Tests of Hypotheses Using the Anova MS for BLO*LAM as an Error Term
Source                 DF        Anova SS     Mean Square    F Value    Pr > F
LAM                     3     32.97103889     10.99034630      26.04    0.0119
Tests of Hypotheses Using the Anova MS for BLO*PSOL as an Error Term
Source                 DF        Anova SS     Mean Square    F Value    Pr > F
PSOL                    2     14.78732500      7.39366250       2.91    0.2556
Tests of Hypotheses Using the Anova MS for BLO*LAM*PSOL as an Error Term
Source                 DF        Anova SS     Mean Square    F Value    Pr > F
LAM*PSOL                6     67.63105278     11.27184213      35.89    0.0002
```

Figura 10. Análise de variância de um ensaio em faixa com parcela dividida usando o procedimento ANOVA do SAS.

Capítulo 10

Confundimento em ensaios fatoriais

Como foi visto no Capítulo 8, uma das grandes desvantagens dos ensaios fatoriais é o crescimento muito rápido do número de tratamentos a serem testados, tanto pelo aumento do número de fatores quanto pelo número de níveis de cada fator, o que torna muitas vezes impraticável a sua execução, pois os blocos passam a ter uma área tão extensa, ou requerem uma quantidade tão grande de material experimental homogêneo, que não se consegue a devida uniformidade de ambiente.

Para resolver esse problema, Yates (1937) introduziu o princípio de confundimento, que consiste em dividir os tratamentos em blocos incompletos, de tal forma que certos efeitos (principais ou de interação) tenham como estimadores os mesmos contrastes ou combinações de tratamentos que estimam os efeitos desses blocos. De maneira geral, recomenda-se que o confundimento seja realizado com as interações de ordem maior e, portanto, de explicação extremamente difícil ou de pouca importância, ou então com interações ou efeitos principais os quais o pesquisador não tem interesse em conhecer ou medir pelo fato de já possuir um conhecimento suficiente a respeito delas. O uso do confundimento acarreta, em geral, também uma estimação mais precisa do erro experimental e dos fatores ou efeitos não confundidos.

No caso de experimentos cujos fatores têm apenas dois níveis, o procedimento é relativamente simples e há uma perfeita coincidência nos efeitos confundidos. Quando o número de níveis é maior que três, o procedimento é ligeiramente mais complicado, e não ocorre essa coincidência entre efeitos de blocos e efeitos de interação ou efeitos principais, o que será discutido mais adiante por meio dos exemplos de aplicação.

O confundimento pode ser total ou parcial. É total quando em todas as repetições, para a divisão do bloco em blocos incompletos, usa-se o mesmo contraste de definição, e é parcial quando se empregam contrastes de definição diferentes nas distintas repetições.

Para o confundimento, a divisão dos blocos completos em blocos incompletos se faz sempre segundo a base de definição do fatorial, isto é, fatoriais de base 2 podem ter blocos incompletos com a metade, um quarto, um oitavo, etc. dos tratamentos. Por sua vez, os de base 3 teriam um terço, um nono, um vinte sete avos, e assim por diante. Quando a divisão dos blocos é feita em valores superiores ao valor da base, há a necessidade do uso de mais de um contraste de definição.

O uso de confundimento em fatoriais mistos, que apresentam fatores com número diferentes de níveis, é muito mais complicado do que aqueles de base constante e não serão discutidos aqui.

Confundimento em fatoriais de base 2

Recordemos que, nos fatoriais 2^k, os contrastes para qualquer efeito principal ou interação são constituídos apenas por -1 e +1. Assim, a divisão dos tratamentos em dois blocos implica que um deles deveria conter todos os tratamentos cujos coeficientes no contraste de definição têm valor -1; enquanto o outro, aqueles com valor +1.

A Tabela 1 mostra os contrastes para um fatorial 2^3 para os efeitos principais e suas interações.

Tabela 1. Coeficientes dos contrastes que permitem medir os efeitos de cada efeito principal, as interações simples e dupla, para um fatorial 2^3.

Efeito	Tratamento							
	(1)	a	B	ab	c	ac	bc	abc
A	-1	+1	-1	+1	-1	+1	-1	+1
B	-1	-1	+1	+1	-1	-1	+1	+1
A*B	+1	-1	-1	+1	+1	-1	-1	+1
C	-1	-1	-1	-1	+1	+1	+1	+1
A*C	+1	-1	+1	-1	-1	+1	-1	+1
B*C	+1	+1	-1	-1	-1	-1	+1	+1
A*B*C	-1	+1	+1	-1	+1	-1	-1	+1

Para a construção de dois blocos incompletos, que contenham cada um a metade dos tratamentos, deve-se primeiro definir o contraste de definição. Como afirmado anteriormente, em geral utiliza-se a interação de maior ordem como contraste de definição, de forma que um bloco contenha os tratamentos com coeficientes -1 e o outro, os de

coeficiente +1. Assim, um bloco conteria os tratamentos (1), ab, ac e bc, enquanto o outro os tratamentos a, b, c e abc.

Em um experimento com quatro repetições, por exemplo, se todas elas utilizassem o mesmo contraste de definição, haveria um confundimento total do efeito da interação de segunda ordem, em geral de pouca importância. No entanto, se houvesse o interesse de estimá-la, poder-se-ia utilizar um procedimento de confundimento parcial adotando-se as interações simples e a interação dupla como contrastes de definição para cada uma das quatro repetições. As Figuras 1 e 2 mostram a constituição dos blocos em cada uma das repetições, nos dois casos, sem a devida casualização dos blocos dentro de cada repetição e também sem a casualização de tratamentos dentro de cada bloco.

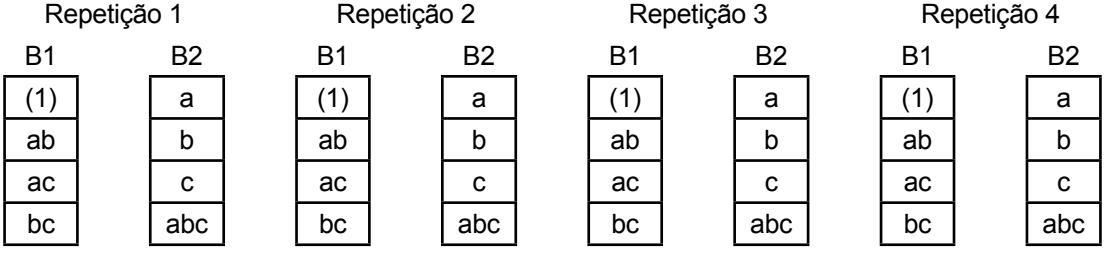


Figura 1. Distribuição dos tratamentos em blocos dentro das repetições com confundimento total da interação dupla A*B*C.

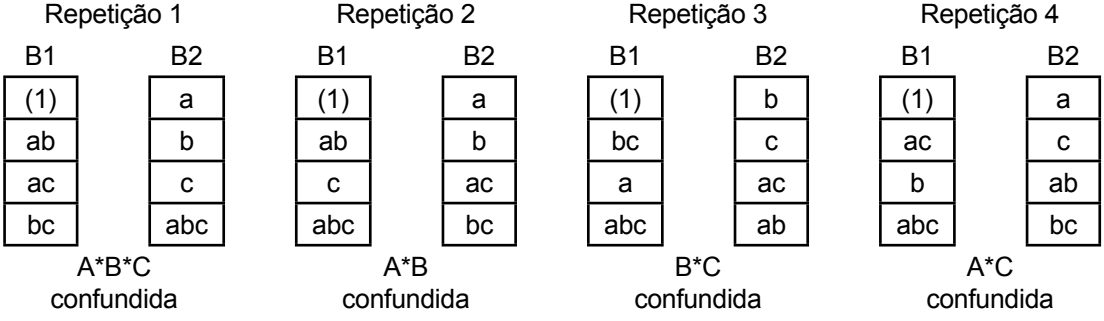


Figura 2. Distribuição dos tratamentos em blocos dentro das repetições com confundimento parcial das interações simples e dupla.

Análise de um fatorial 2^3 com confundimento total da interação A*B*C

O pesquisador Alberto Baeta dos Santos conduziu na cidade de Nobres, MT, experimento visando ao conhecimento do efeito de algumas práticas culturais (densidade de

plantio, espaçamento de plantio e adubação nitrogenada em cobertura) sobre a produtividade e incidência de brusone em arroz de terras altas. Os dados mostrados na Tabela 2 referem-se à produtividade de grãos (kg ha^{-1}) no ano agrícola de 1981–1982. O ensaio foi conduzido como um fatorial em blocos ao acaso com quatro repetições. As densidades de plantio são de 50 e 90 sementes por metro, os espaçamentos entre linhas de 35 cm e 50 cm e o terceiro fator presença ou ausência de adubação nitrogenada em cobertura. A Tabela 2 mostra esses dados para o caso do confundimento total da interação dupla. Mostram-se aí as quatro repetições com dois blocos cada uma, as identificações dos tratamentos e os respectivos dados, além dos totais de blocos e repetições. Convém recordar que as identificações (1) e 000 representam o mesmo tratamento, assim como a e 100, b e 010, como explicado no Capítulo 8.

Tabela 2. Dados de produtividade em kg ha^{-1} com confundimento total da interação dupla no ensaio de práticas culturais em arroz de terras altas em Nobres MT, 1981–1982.

Repetição 1		Repetição 2		Repetição 3		Repetição 4	
Bloco 1	Bloco 2	Bloco 1	Bloco 2	Bloco 1	Bloco 2	Bloco 1	Bloco 2
101	001	000	001	011	111	101	111
2.820	2.783	2.732	3.499	2.927	2.232	3.400	3.527
011	111	110	111	000	100	000	010
3.053	3.187	2.416	3.493	2.702	1.908	2.526	2.644
000	100	101	100	101	001	011	001
2.168	1.985	2.514	2.556	2.438	2.616	2.856	3.077
110	010	011	010	110	010	110	100
1.951	3.052	3.089	3.841	2.288	2.648	2.635	2.048
9.992	11.007	10.751	13.389	10.355	9.404	11.417	11.296
20.999		24.140		19.759		22.713	

Para a análise de variância, deve ser montado um quadro similar ao da Tabela 1, que contém os totais de cada tratamento multiplicados pelos coeficientes dos contrastes ali contidos. O total do tratamento 000 foi obtido somando-se todos os dados para esse tratamento nas quatro repetições e assim para os demais tratamentos. Dessa forma, estrutura-se a Tabela 3.

Na Tabela 3, os valores dos contrastes correspondem à soma dos totais de tratamentos em cada contraste, enquanto a soma de quadrado dos contrastes foi obtida pelo uso da mesma fórmula mostrada no Capítulo 1. Como exemplo, para o fator A, tem-se

Tabela 3. Quadro para obtenção das somas de quadrados dos fatores não confundidos em um fatorial 2^3.

Efeito	Tratamento								Valor do contraste	Soma de quadrado
	000	001	010	011	100	101	110	111		
A	-10.128	-11.975	-12.185	-11.925	8.497	11.172	9.290	12.439	-4.815	724.507,0313
B	-10.128	-11.975	12.185	11.925	-8.497	-11.172	9.290	12.439	4.067	516.890,2813
C	-10.128	11.975	-12.185	11.925	-8.497	11.172	-9.290	12.439	7.411	171.6341,281
A*B	10.128	11.975	-12.185	-11.925	-8.497	-11.172	9.290	12.439	53	87,78125
A*C	10.128	-11.975	12.185	-11.925	-8.497	11.172	-9.290	12.439	4.237	561.005,2813
B*C	10.128	-11.975	-12.185	11.925	8.497	-11.172	-9.290	12.439	-1.633	83.334,03125

$$SQA = \frac{(\sum c_i T_i)^2}{r * \sum c_i^2} = \frac{[(-10.128) + (-11.975) + ... + (12.439)]^2}{4 * 8} =$$

$$\frac{(-4.815)^2}{32} = 724.507,0313$$

e assim, sucessivamente, para todos os fatores principais e interações simples.

Para a montagem do quadro da análise de variância, devem-se ainda calcular as somas de quadrados: total, repetições, blocos dentro de repetições e erro, sendo esta última obtida por diferença. Assim:

$$SQTotal = 2.820^2 + 2.783^2 + ... + 2.048^2 - FC = 7.622.960,22$$

$$SQRep = \frac{20.999^2 + 24.140^2 + ... + 22.713^2}{8} - FC = 1.384.277,59$$

$$SQBlocos(Rep) = \frac{9.992^2 + 11.007^2 + ... + 11.296^2}{4} - FC - SQRep = 1.113.538,88$$

A análise de variância está apresentada na Tabela 4.

Tabela 4. Análise de variância de um fatorial 2^3 com confundimento total da interação dupla com quatro repetições.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	3	1.384.277,59		
Bloco(Rep)	4	1.113.538,88		
A	1	724.507,03	724.507,03	8,56 (0,009)
B	1	516.890,28	516.890,28	6,11 (0,024)
C	1	1.716.341,28	1.716.341,28	20,29 (0,001)
A*B	1	87,78	87,78	0,00 (0,975)
A*C	1	561.005,28	561.005,28	6,63 (0,019)
B*C	1	83.334,03	83.334,03	0,98 (0,334)
Erro	18	1.522.978,06	84.609,89	
Total	31	762.2960,22		

Uso do SAS para um fatorial 2^3 com confundimento total de uma interação

Para esta análise, o procedimento ANOVA é suficiente para que se obtenham todas as fontes de variação, suas somas de quadrados, seus quadrados médios e testes F. O programa é o seguinte:

```
PROC ANOVA;CLASS REP BLOCO FAT_A FAT_B FAT_C;
MODEL PROD=REP BLOCO(REP) FAT_A  FAT_B FAT_C FAT_A*FAT_B FAT_A*FAT_C
FAT_B*FAT_C;
```

O resultado está apresentado na Figura 3.

```
                    The ANOVA Procedure
                   Class Level Information
Class          Levels    Values
REP                 4    1 2 3 4
BLOCO               2    1 2
FAT_A               2    0 1
FAT_B               2    0 1
FAT_C               2    0 1

Number of observations    32

                    The ANOVA Procedure
Dependent Variable: PROD

                              Sum of
Source                DF      Squares    Mean Square   F Value   Pr > F
Model                 13   6099982.156     469229.397      5.55   0.0005
Error                 18   1522978.063      84609.892
Corrected Total       31   7622960.219

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.800212      10.62434      290.8778      2737.844

Source                DF       Anova SS    Mean Square   F Value   Pr > F
REP                    3   1384277.594     461425.865      5.45   0.0076
BLOCO(REP)             4   1113538.875     278384.719      3.29   0.0344
FAT_A                  1    724507.031     724507.031      8.56   0.0090
FAT_B                  1    516890.281     516890.281      6.11   0.0237
FAT_C                  1   1716341.281    1716341.281     20.29   0.0003
FAT_A*FAT_B            1        87.781         87.781      0.00   0.9747
FAT_A*FAT_C            1    561005.281     561005.281      6.63   0.0191
FAT_B*FAT_C            1     83334.031      83334.031      0.98   0.3341
```

Figura 3. Resultado da análise de variância de um fatorial 2^3 com confundimento total da interação dupla com o uso do SAS.

Análise de um fatorial 2^3 com confundimento parcial das interações simples e da interação dupla

Considere-se agora o mesmo ensaio, mas assumindo que as interações simples e a interação dupla foram confundidas uma em cada repetição. Os dados são apresentados na Tabela 5.

Tabela 5. Dados de produtividade do ensaio fatorial 2^3, assumindo as interações simples A*B, B*C e A*C nas repetições 2, 3 e 4, respectivamente, e a interação dupla A*B*C na repetição 1.

Repetição 1		Repetição 2		Repetição 3		Repetição 4	
Bloco 1	Bloco 2	Bloco 1	Bloco 2	Bloco 1	Bloco 2	Bloco 1	Bloco 2
101	001	000	101	011	101	101	011
2.820	2.783	2.732	2.514	2.927	2.438	3.400	2.856
011	111	110	011	000	110	000	110
3.053	3.187	2.416	3.089	2.702	2.288	2.526	2.635
000	100	001	100	111	001	111	001
2.168	1.985	3.499	2.556	2.232	2.616	3.527	3.077
110	010	111	010	100	010	010	100
1.951	3.052	3.493	3.841	1.908	2.648	2.644	2.048
9.992	11.007	12.140	12.000	9.769	9.990	12.097	10.616
20.999		24.140		19.759		22.713	

Para a obtenção das somas de quadrados dos fatores principais e interações entre eles, deve-se montar um quadro de contrastes, lembrando agora que, para as interações, as somas para os totais de tratamentos são parciais, isto é, para cada uma delas somam-se os valores correspondentes das repetições em que a interação não está confundida. Assim, para a interação simples A*B somam-se os valores dos tratamentos nas repetições 1, 3 e 4, para a interação A*C, os totais de tratamentos são obtidos somando-se os valores das repetições 1 a 3, e assim sucessivamente. Esses valores, já multiplicados pelos coeficientes dos contrastes, encontram-se na Tabela 6.

As somas de quadrados para total, repetições, blocos dentro de repetições e erro são obtidas como como calculado no item *Análise de um fatorial 23, com confundimento total da interação A*B*C*, de forma que a análise de variância agora é aquela mostrada na Tabela 7.

Tabela 6. Contraste de estimação dos efeitos principais e das interações, valor dos contrastes e suas somas de quadrados.

Fator e interação	Tratamento								Valor do contraste	Soma de quadrado do contraste
	000	001	010	011	100	101	110	111		
A	-10.128	-11.975	-12.185	-11.925	8.497	11.172	9.290	12.439	-4.815	724.507,03
B	-10.128	-11.975	12.185	11.925	-8.497	-11.172	9.290	12.439	4.067	516.890,28
C	-10.128	11.975	-12.185	11.925	-8.497	11.172	-9.290	12.439	7.411	1.716.341,28
A*B	7.396	8.476	-8.344	-8.836	-5.941	-8.658	6.874	8.946	-87	315,38
A*C	7.602	-8.898	9.541	-9.069	-6.449	7.772	-6.655	8.912	2.756	316.480,67
B*C	7.426	-9.359	-9.537	8.998	6.589	-8.734	-7.002	10.207	-1.412	83.072,67
A*B*C	-7.960	9.192	9.133	-8.872	6.512	-8.352	-7.339	9.252	1.566	102.181,50

Tabela 7. Análise de variância de um fatorial 2^3 com confundimento parcial das interações simples e da interação dupla.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	3	1.384.277,59		
Bloco(Rep)	4	411.503,38		
A	1	724.507,03	724.507,03	5,20 (0,036)
B	1	516.890,28	516.890,28	3,71 (0,071)
C	1	1.716.341,28	1.716.341,28	12,32 (0,003)
A*B	1	315,38	315,38	0,00 (0,963)
A*C	1	316.480,67	316.480,67	2,27 (0,150)
B*C	1	83.072,67	83.072,67	0,60 (0,451)
A*B*C	1	102.181,50	102.181,50	0,73 (0,404)
Erro	17	2.367.390,44	139.528,26	
Total	31	7.622.960,22		

Como pode ser observado, as somas de quadrados dos efeitos principais não sofrem qualquer alteração na sua estimação nesses dois casos, bem como no exemplo que se segue, pois eles nunca estiveram em confundimento. Apenas os efeitos de interação mostram essa alteração ou não são estimáveis por estarem parcial ou totalmente confundidos com os blocos nas repetições.

Uso do SAS em experimentos fatoriais com confundimento parcial de interações simples e dupla

Para fazer esta análise, com confundimento parcial das interações simples e dupla, o procedimento a ser empregado é o GLM, com a opção de soma de quadrados do tipo 1 (para maiores detalhes ver os manuais do SAS). Dessa forma, o programa tem o seguinte aspecto:

```
PROC GLM;CLASS REP BLOCO FAT_A FAT_B FAT_C;
MODEL PROD=REP BLOCO(REP) FAT_A|FAT_B|FAT_C/SS1;
```

Os resultados que se encontram na Figura 4 são absolutamente iguais àqueles mostrados na Tabela 7.

```
                    The GLM Procedure
                  Class Level Information
Class                 Levels    Values
REP                        4    1 2 3 4
BLOCO                      2    1 2
FAT_A                      2    0 1
FAT_B                      2    0 1
FAT_C                      2    0 1

                Number of observations    32
                    The GLM Procedure

Dependent Variable: PROD

                                 Sum of
Source                    DF       Squares    Mean Square   F Value   Pr > F
Model                     14    5255569.771     375397.841      2.70   0.0274
Error                     17    2367390.448     139258.262
Corrected Total           31    7622960.219

                  R-Square     Coeff Var      Root MSE     PROD Mean
                  0.689439      13.63019      373.1732      2737.844

Source                    DF     Type I SS    Mean Square   F Value   Pr > F
REP                        3    1384277.594     461425.865      3.31   0.0451
BLOCO(REP)                 4     411503.375     102875.844      0.74   0.5784
FAT_A                      1     724507.031     724507.031      5.20   0.0357
FAT_B                      1     516890.281     516890.281      3.71   0.0709
FAT_A*FAT_B                1        315.375        315.375      0.00   0.9626
FAT_C                      1    1716341.281    1716341.281     12.32   0.0027
FAT_A*FAT_C                1     316480.667     316480.667      2.27   0.1500
FAT_B*FAT_C                1      83072.667      83072.667      0.60   0.4505
FAT_A*FAT_B*FAT_C          1     102181.500     102181.500      0.73   0.4036
```

Figura 4. Análise de um fatorial 2^3, com confundimento parcial das interações simples e dupla com o uso do SAS.

Fatorial de base 2 com blocos que contêm menos que a metade dos tratamentos.

Ocorrem certas situações em que a simples divisão do bloco completo em dois blocos incompletos não é suficiente. Em geral, isso acontece quando o número de fatores é muito elevado ou por condições experimentais em que o bloco deve ter um número muito baixo de parcelas. A solução então é dividir o bloco em quatro, oito, ou dezesseis blocos menores, por exemplo, o que se consegue utilizando mais de um contraste de definição de confundimento.

Para dividir um bloco em quatro menores, são necessários dois contrastes de definição, mas isso implica que um terceiro contraste está sendo empregado, já que a divisão em quatro blocos acarreta o uso de três graus de liberdade para blocos em cada repetição. Esse terceiro contraste corresponde ao produto dos dois escolhidos pelo pesquisador, no qual se eliminam todas as identificações de fatores que aparecem com expoente 2.

Por exemplo, se os dois fatores escolhidos para definição da partição do bloco forem as interações AB e AC, cujo produto seria A^2BC, tem-se que a interação BC também estará confundida ao eliminar-se do resultado o fator A^2; se a escolha para partição dos blocos tivesse recaído nas interações ABC e AB, o fator C é que estaria confundido, pois $ABC*AB = A^2B^2C = C$. Verifica-se, portanto, o cuidado que se deve ter na escolha dos efeitos para que não se provoque o confundimento de fatores principais ou de interações nas quais o pesquisador possa ter interesse em conhecer.

Para a definição dos tratamentos que serão implantados em cada um dos blocos, deve-se recorrer aos contrastes de definição dos efeitos. Em um fatorial 2^3, se houvesse a necessidade do uso de quatro blocos com duas parcelas por bloco em cada repetição, e assumindo o uso das interações AB e BC, que implicaria também no confundimento da interação AC, cada bloco conteria uma das quatro possíveis combinações de sinais + e - dos dois contrastes mostrados na Tabela 8, ou seja, um bloco conteria os tratamentos (1) e abc (+1 nos dois contrastes), um segundo os tratamentos b e ac (-1 nos dois contrastes), um terceiro os tratamentos a e bc (-1 no contraste A*B e +1 para B*C) e o quarto os tratamentos ab e c (+1 em A*B e -1 em B*C).

Tabela 8. Contrastes para definição de confundimento total das interações A*B e B*C gerando blocos com um quarto dos tratamentos.

	Tratamentos							
Efeitos	(1)	a	b	ab	c	ac	bc	abc
A*B	+1	-1	-1	+1	+1	-1	-1	+1
B*C	+1	+1	-1	-1	-1	-1	+1	+1

Exemplo de aplicação

Retornando ao exemplo anterior, e assumindo agora um confundimento de duas das interações simples, como mostrado na Figura 3, obtém-se a distribuição dos tratamentos apresentada na Tabela 9.

Com o confundimento de duas das interações simples, A*B e B*C, automaticamente se fez também o confundimento da terceira interação simples A*C, facilmente verificável pelo simples produto das interações no módulo 2.

A análise de variância, nesse caso, permite apenas a estimação dos efeitos dos fatores principais e da interação dupla. O quadro da análise de variância se monta então seguindo essas limitações e existe também a necessidade de um quadro auxiliar com os contrastes para estimação dos efeitos não confundidos. A Tabela 10 contém esses contrastes e a Tabela 11 a análise de variância para esse experimento.

Tabela 9. Distribuição dos tratamentos em blocos de duas unidades para confundimento das interações simples e respectivas produções.

Repetição 1				Repetição 2				Repetição 3				Repetição 4			
Bloco 1	Bloco 2	Bloco 3	Bloco 4	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Bloco 1	Bloco 2	Bloco 3	Bloco 4
111	011	101	001	111	011	101	001	111	011	101	001	111	011	101	001
3.187	3.053	2.820	2.783	3.493	3.089	2.514	3.499	2.232	2.927	2.438	2.616	3.527	2.856	3.400	3077
000	100	010	110	000	100	010	110	000	100	010	110	000	100	010	110
2.168	1.985	3.052	1.951	2.732	2.556	3.841	2.416	2.702	1.908	2.648	2.288	2.526	2.048	2.644	2635
5.355	5.038	5.872	4.734	6.225	5.645	6.355	5.915	4.934	4.835	5.086	4.904	6.053	4.904	6.044	5.712
20.999				24.140				19.759				22.713			

Tabela 10. Contrastes para estimação das somas de quadrados dos efeitos não confundidos num fatorial 2^3, com blocos que contêm um quarto dos tratamentos.

Fator	Tratamento								Efeito do contraste	Soma de quadrado do contraste
	000	001	010	011	100	101	110	111		
A	-10.128	-11975	-12.185	-11.925	8.497	11.172	9.290	12.439	-4.815	724.507,0313
B	-10.128	-11975	12.185	11.925	-8.497	-11.172	9.290	12.439	4.067	516.890,2813
C	-10.128	11.975	-12.185	11.925	-8.497	11.172	-9.290	12.439	7.411	1.716.341,2810
A*B*C	-10.128	11.975	12.185	-11.925	8.497	-11.172	-9.290	12.439	2.581	208.173,7813

Tabela 11. Análise de variância de um fatorial 2^3 com as interações simples confundidas com blocos.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	3	1.384.277,59		
Bloco(Rep)	12	961.297,12		
A	1	724.507,03	724.507,03	4,12 (0,065)
B	1	516.890,28	516.890,28	2,94 (0,112)
C	1	1.716.341,28	1.716.341,28	9,75 (0,009)
A*B*C	1	208.173,78	208.173,78	1,18 (0,298)
Erro	12	2.111.473,12	175.956,094	
Total	31	7.622.960,22		

Uso do SAS em fatoriais de base 2 com confundimento das interações simples

Novamente aqui o procedimento adequado é o GLM, com a opção de soma de quadrado tipo 1. No comando MODEL, não devem ser especificadas as interações simples, pois elas estão em confundimento com os blocos. No entanto, se forem explicitadas, o resultado mostrará para elas uma soma de quadrados e graus de liberdade iguais a 0 (zero), uma vez que estão totalmente confundidas. Assim, o programa apresenta o seguinte formato:

```
PROC GLM;CLASS REP BLOCO FAT_A FAT_B FAT_C;
MODEL PROD=REP BLOCO(REP) FAT_A FAT_B FAT_C FAT_A*FAT_B*FAT_C/SS1;
```

O resultado se vê na Figura 5.

Procedimento geral para confundimento em fatoriais

Dada a maior complexidade dos contrastes de definição de efeitos nos fatoriais de base 3 ou maior, o procedimento adotado no caso dos fatoriais de base 2 não pode ser adotado. Um procedimento mais geral é então necessário e está baseado no uso da função residual de módulo m ou resíduo $mod\ m$. Para um inteiro qualquer n, o resíduo $mod\ m$ é o

```
                          The GLM Procedure
                      Class Level Information
                Class         Levels    Values
                REP                4    1 2 3 4
                BLOCO              4    1 2 3 4
                FAT_A              2    0 1
                FAT_B              2    0 1
                FAT_C              2    0 1

                 Number of observations    32
                      The GLM Procedure
Dependent Variable: PROD
                                Sum of
 Source                 DF      Squares    Mean Square   F Value   Pr > F

 Error                  12   2111473.125     175956.094
 Corrected Total        31   7622960.219

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.723011      15.32123      419.4712      2737.844

 Source                 DF     Type I SS    Mean Square   F Value   Pr > F
 REP                     3   1384277.594     461425.865      2.62   0.0986
 BLOCO(REP)             12    961297.125      80108.094      0.46   0.9063
 FAT_A                   1    724507.031     724507.031      4.12   0.0652
 FAT_B                   1    516890.281     516890.281      2.94   0.1122
 FAT_C                   1   1716341.281    1716341.281      9.75   0.0088
 FAT_A*FAT_B*FAT_C       1    208173.781     208173.781      1.18   0.2981
```

Figura 5. Análise de variância de um fatorial 2^3 com confundimento total das interações simples com o uso do SAS.

resto da divisão de n por m, ou seja, r, que se escreve como $n = r\ (mod\ m)$ de forma que r pode assumir os valores 0, 1, 2, ..., m - 1.

Nos experimentos fatoriais o valor de m é a base desses fatoriais de forma que, para fatoriais de base 2, r assume apenas os valores 0 e 1; para base 3, r pode ter os valores 0, 1 e 2, e assim sucessivamente.

O uso de resíduo $mod\ m$ se faz associado às equações de definição de contraste, em que cada fator é representado por uma variável, usualmente $x_1, x_2, ..., x_k$, em que k é o número de fatores envolvidos no ensaio fatorial e os níveis de cada fator são representados por 0, 1, ..., m - 1. As variáveis podem ter coeficientes iguais a 0, 1, ..., m - 1, e as equações envolvem tantas variáveis quantos efeitos envolvidos nos contrastes de definição de confundimento. Por exemplo, se num fatorial 2^3 se pretende confundir o efeito principal A, a equação correspondente é $x_1 + 0*x_2 + 0*x_3 = r\ (mod\ 2)$ ou $x_1 = r\ (mod\ m)$, e cada valor possível de r define a constituição de cada um dos blocos incompletos; para confundir-se a interação AB, a equação é $x_1 + 0*x_2 + 0*x_3 = r\ (mod\ 2)$ ou $x_1 + x_2 = r\ (mod\ m)$.

No caso de um fatorial 2^3, os tratamentos são mostrados na Tabela 12, na qual se faz a associação dos dois tipos de identificação, em que 0 corresponde à ausência (ou nível mais baixo) do fator e 1 à presença (ou nível mais alto).

Tabela 12. Associação entre tipos de identificação de tratamentos num fatorial 2^3.

Tratamento							
(1)	a	b	ab	c	ac	bc	abc
000	100	010	110	001	101	011	111

Para fazer o confundimento da interação ABC, a equação de definição é dada por $x_1 + x_2 + x_3 = r\ (mod\ 2)$, de forma que r assume apenas os valores 0 e 1, que definem, então, a constituição dos dois blocos. Desse modo, tem-se:

$$000 \rightarrow 0 + 0 + 0 = 0$$
$$100 \rightarrow 1 + 0 + 0 = 1$$
$$010 \rightarrow 0 + 1 + 0 = 1$$
$$110 \rightarrow 1 + 1 + 0 = 2 = 0\ (mod\ 2)$$
$$001 \rightarrow 0 + 0 + 1 = 1$$
$$101 \rightarrow 1 + 0 + 1 = 2 = 0\ (mod\ 2)$$
$$011 \rightarrow 0 + 1 + 1 = 2 = 0\ (mod\ 2)$$
$$111 \rightarrow 1 + 1 + 1 = 3 = 1\ (mod\ 2)$$

Portanto, um bloco teria os tratamentos 000, 110, 101 e 011, em que a equação possui resultado igual a 0, e o outro, os tratamentos 100, 010, 001 e 111, com resultado igual a 1, grupos idênticos aos obtidos pelo uso dos coeficientes -1 e +1 do procedimento anterior.

No caso de confundimento da interação AB, a equação de definição é dada por $X_1 + X_2 + 0{*}X_3 = r\ (mod\ 2)$, de forma que r assume apenas os valores 0 e 1, que darão então a constituição dos dois blocos:

$$000 \rightarrow 0 + 0 + 0{*}0 = 0$$
$$100 \rightarrow 1 + 0 + 0{*}0 = 1$$
$$010 \rightarrow 0 + 1 + 0{*}0 = 1$$
$$110 \rightarrow 1 + 1 + 0{*}0 = 2 = 0\ (mod\ 2)$$
$$001 \rightarrow 0 + 0 + 0{*}1 = 0$$
$$101 \rightarrow 1 + 0 + 0{*}1 = 1$$
$$011 \rightarrow 0 + 1 + 0{*}1 = 1$$
$$111 \rightarrow 1 + 1 + 0{*}1 = 2 = 0\ (mod\ 2)$$

Dessa forma, um bloco teria os tratamentos 000, 110, 001 e 111, em que a equação possui resultado igual a 0, e o outro, os tratamentos 100, 010, 101 e 011, para resultado igual a 1, que também aqui correspondem aos coeficientes -1 e +1 usados anteriormente.

Para a divisão dos blocos em conjuntos de tratamentos menores que a sua metade, usa-se mais de um contraste de definição de confundimento, como anteriormente, e cada bloco conterá as possíveis combinações de resultados 0 e 1 de cada uma das equações.

Suponha-se que para o fatorial 2^3 haja a necessidade de uso de blocos de tamanho 2, ou seja, cada repetição deverá ser composta por quatro blocos. Como antes, as interações AB e BC, usadas como contrastes de definição fornecem as equações $x_1 + x_2 + 0*x_3 = r\ (mod\ 2)$ e $0*x_1 + x_2 + x_3 = r\ (mod\ 2)$, cujos resultados são:

$000 \rightarrow 0 + 0 + 0*0 = 0$ $\qquad$ $000 \rightarrow 0*0 + 0 + 0 = 0$
$100 \rightarrow 1 + 0 + 0*0 = 1$ $\qquad$ $100 \rightarrow 0*1 + 0 + 0 = 0$
$010 \rightarrow 0 + 1 + 0*0 = 1$ $\qquad$ $010 \rightarrow 0*0 + 1 + 0 = 1$
$110 \rightarrow 1 + 1 + 0*0 = 2 = 0 \pmod 2$ $\qquad$ $110 \rightarrow 0*1 + 1 + 0 = 1$
$001 \rightarrow 0 + 0 + 0*1 = 0$ $\qquad$ $001 \rightarrow 0*0 + 0 + 1 = 1$
$101 \rightarrow 1 + 0 + 0*1 = 1$ $\qquad$ $101 \rightarrow 0*1 + 0 + 1 = 1$
$011 \rightarrow 0 + 1 + 0*1 = 1$ $\qquad$ $011 \rightarrow 0*0 + 1 + 1 = 2 = 0 \pmod 2$
$111 \rightarrow 1 + 1 + 0*1 = 2 = 0 \pmod 2$ $\qquad$ $111 \rightarrow 0*1 + 1 + 1 = 2 = 0 \pmod 2$

Dessa forma, o bloco 1 teria as soluções 0 e 0, logo 000 e 111; o bloco 2 as soluções 0 e 1, portanto 110 e 001; o bloco 3 os resultados 1 e 0, cujos tratamentos são 100 e 011, e finalmente o bloco 4 os tratamentos 010 e 101, cujas soluções são 1 e 1 para as duas equações. Recorde-se que nesse caso a interação AC também está confundida com blocos.

Confundimento de fatoriais 3^2 em blocos com um terço dos tratamentos

Com fatoriais de base 3, existe uma única forma de se confundir um efeito principal, mas existem duas ou mais formas de se confundir cada uma das interações. Em um fatorial 3^2 com dois fatores em três níveis cada um, os tratamentos são 00, 01, 02, 10, 11, 12, 20, 21 e 22, e para confundir o efeito de A, que possui dois graus de liberdade, cada bloco deve ser dividido em três, sendo a equação de definição de contraste dada por $x_1 + 0*x_2 = r(mod\ 3)$, em que r pode, portanto assumir os valores 0, 1 e 2. Nesse caso, tem-se:

$00 \rightarrow 0 + 0*0 = 0$
$01 \rightarrow 0 + 0*1 = 0$
$02 \rightarrow 0 + 0*2 = 0$

$10 \rightarrow 1 + 0*0 = 1$
$11 \rightarrow 1 + 0*1 = 1$
$12 \rightarrow 1 + 0*2 = 1$
$20 \rightarrow 2 + 0*0 = 2$
$21 \rightarrow 2 + 0*1 = 2$
$22 \rightarrow 2 + 0*2 = 2$

O primeiro bloco conterá os tratamentos 00, 01 e 02; o segundo, 10, 11 e 12 e o terceiro, 10, 11 e 12.

Para a interação AB, que tem quatro graus de liberdade, podem ser confundidos apenas dois graus de liberdade, os quais serão definidos pelas equações $x_1 + x_2 = r\ (mod\ 3)$ e $x_1 + 2*x_2 = r\ (mod\ 3)$. Vale lembrar que essa interação é usualmente dividida em quatro componentes dados pelas combinações entre os efeitos lineares e quadráticos de A e B, mas que não têm qualquer correspondência com os graus de liberdade que se confundem pelo uso das equações anteriores. A primeira equação de definição, $x_1 + x_2 = r\ (mod\ 3)$, leva a:

$00 \rightarrow 0 + 0 = 0$
$01 \rightarrow 0 + 1 = 1$
$02 \rightarrow 0 + 2 = 2$
$10 \rightarrow 1 + 0 = 1$
$11 \rightarrow 1 + 1 = 2$
$12 \rightarrow 1 + 2 = 3 = 0\ (\text{mod } 3)$
$20 \rightarrow 2 + 0 = 2$
$21 \rightarrow 2 + 1 = 3 = 0\ (\text{mod } 3)$
$22 \rightarrow 2 + 2 = 4 = 1\ (\text{mod } 3)$

O bloco 1 conteria 00, 12 e 21; o bloco 2 conteria 01, 10 e 22, enquanto o bloco 3 teria os tratamentos 02, 11 e 20. A outra equação, $x_1 + 2*x_2 = r\ (mod\ 3)$, tem como resultado:

$00 \rightarrow 0 + 2*0 = 0$
$01 \rightarrow 0 + 2*1 = 2$
$02 \rightarrow 0 + 2*2 = 4 = 1\ (\text{mod } 3)$
$10 \rightarrow 1 + 2*0 = 1$
$11 \rightarrow 1 + 2*1 = 3 = 0\ (\text{mod } 3)$
$12 \rightarrow 1 + 2*2 = 5 = 2\ (\text{mod } 3)$
$20 \rightarrow 2 + 2*0 = 2$
$21 \rightarrow 2 + 2*1 = 4 = 1\ (\text{mod } 3)$
$22 \rightarrow 2 + 2*2 = 6 = 0\ (\text{mod } 3)$

Dessa forma, um bloco conteria agora os tratamentos 00, 11 e 22; enquanto um segundo bloco teria os tratamentos 02, 10 e 21; e o terceiro bloco, os tratamentos 01, 12 e 20.

Exemplo de aplicação em fatorial 3^2

No ano agrícola 1999–2000, no Estado do Tocantins, o pesquisador Homero Aidar conduziu experimentos com a cultura do feijoeiro, testando espaçamento entre linhas de plantio (0,45 cm, 0,6 cm e 0,75 cm) e densidade de plantio (7, 10 e 13 sementes por metro). Os dados a seguir referem-se à cultivar Jalo. Para aproveitar as condições da fazenda onde se instalou o experimento adotou-se o delineamento completamente ao acaso, mas aqui se considera o de blocos ao acaso com confundimento parcial de dois graus de liberdade da interação densidade (A) x espaçamento (B). A Tabela 13 contém as identificações dos tratamentos, distribuídos nos três blocos em cada uma das repetições, segundo o esquema mostrado no item anterior e o resultado da medida da altura da planta, média de dez plantas.

Tabela 13. Dados de altura de planta, em centímetros, medidos no experimento fatorial 3^2, conduzido em Tocantins.[1]

Repetição 1			Repetição 2			Repetição 3			Repetição 4		
Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3
00	01	02	00	01	02	00	01	02	00	01	02
118	124	120	116	128	124	121	120	124	114	119	117
11	12	10	11	12	10	11	12	10	11	12	10
110	121	115	133	130	119	120	115	124	117	112	111
22	20	21	22	20	21	22	20	21	22	20	21
103	112	124	120	117	123	113	112	119	117	104	118
331	357	359	369	375	366	354	347	367	348	335	346
	1.047			1.110			1.068			1.029	

1 Dados cedidos pelo pesquisador Homero Aidar, da Embrapa Pesca e Aquicultura.

A Tabela 14 contém os totais de todas as combinações de tratamentos, bem como os totais gerais de cada nível de cada um dos fatores em estudo.

Tabela 14. Valores totais dos diversos níveis de cada um dos fatores e suas combinações. Ensaio fatorial 3^2, com confundimento parcial da interação.

A (Densidade)	B (Espaçamento)			Soma
	0 (0,45 cm)	1 (0,60 cm)	2 (0,75 cm)	
0 (7 semente m^{-1})	469	491	485	1.445
1 (10 semente m^{-1})	469	480	478	1.427
2 (13 semente m^{-1})	445	484	453	1.382
Soma	1.383	1.455	1.416	4.254

A análise de variância se encontra detalhada na Tabela 15, e os valores das somas de quadrados foram obtidos da forma usual, como feito em todos os exemplos anteriores, ou seja:

$$SQTotal = (118^2 + 110^2 + ... + 111^2 + 118^2) - \frac{(4.254)^2}{36} = 1.419{,}00$$

$$SQRep = \frac{1.047^2 + 1.110^2 + ... + 1.029^2}{9} - \frac{4.254^2}{36} = 405{,}00$$

$$SQBl(Rep) = \frac{331^2 + 357^2 + ... + 346^2}{3} - \frac{4.254^2}{36} - SQRep = 278{,}00$$

$$SQDens = \frac{1.445^2 + 1.427^2 + 1.382^2}{12} - \frac{4.254^2}{36} = 175{,}50$$

$$SQEsp = \frac{1.383^2 + 1.455^2 + 1.416^2}{12} - \frac{4.254^2}{36} = 216{,}50$$

Para obter a soma de quadrados da parte não confundida da interação densidade x espaçamento, deve-se primeiro obter a soma de quadrados de blocos, depois a soma de quadrados (completa) para a interação a partir dos dados da Tabela 14 e, em seguida, a diferença entre essas somas de quadrados que representa a parte não confundida.

Para a soma de quadrados de blocos, primeiro é necessário primeiramente somar os valores de cada um dos blocos, segundo sua identificação em todas as repetições. Assim: bloco 1 = 331 + 369 + 354 + 348 = 1.402; bloco 2 = 357 + 375 + 347 + 335 = 1.414 e bloco 3 = 359 + 366 + 367 + 346 = 1.438.

$$SQBlocos = \frac{1.402^2 + 1.414^2 + 1.438^2}{12} - \frac{4.254^2}{36} = 56{,}00$$

$$SQ(DensxEsp) = \frac{469^2 + 491^2 + ... + 453^2}{4} - \frac{4.254^2}{36} -$$

$$SQDens - SQEsp = 77{,}5$$

Então, segue-se que a soma de quadrados da parte não confundida é igual a $SQ(Dens \text{ x } Esp)nc = 77{,}5 - 56{,}00 = 21{,}5$.

A soma de quadrados do erro é obtida por diferença.

Tabela 15. Análise de variância de um fatorial 3^2, com confundimento parcial da interação entre fatores.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetições	3	405,00	135,00	
Blocos (Rep)	8	278,00	34,75	
Densidade	2	175,50	87,75	4,90 (0,020)
Espaçamento	2	216,50	108,25	6,04 (0,01)
Densidade x espaçamento (nc)	2	21,50	10,75	0,60 (0,559)
Erro	18	322,50	17,92	
Total	35	1.419,00		

Uso do SAS em fatoriais de base 3, 2 fatores e confundimento parcial da interação simples

Novamente aqui o procedimento adequado é o GLM, com a especificação de soma de quadrados do tipo 1. O programa é o seguinte:

```
PROC GLM;CLASS REP BLOCO FAT_A FAT_B;
MODEL ALTURA=REP BLOCO(REP) FAT_A|FAT_B/SS1;
```

Como se pode ver o modelo especifica os fatores principais e sua interação, como nos casos de fatoriais completos, mas o procedimento faz o ajuste do confundimento, estimando corretamente a soma de quadrados da interação e considera apenas os dois graus de liberdade não confundidos. O resultado se vê na Figura 6.

```
                          The GLM Procedure
                      Class Level Information
                Class         Levels    Values
                REP                4    1 2 3 4
                BLOCO              3    1 2 3
                FAT_A              3    0 1 2
                FAT_B              3    0 1 2

                  Number of observations    36
                          The GLM Procedure
Dependent Variable: ALTURA
                                 Sum of
 Source                DF       Squares    Mean Square   F Value   Pr > F
 Model                 17    1096.500000     64.500000      3.60   0.0050
 Error                 18     322.500000     17.916667
 Corrected Total       35    1419.000000

             R-Square     Coeff Var      Root MSE    ALTURA Mean
             0.772727      3.582066      4.232808       118.1667

 Source                DF     Type I SS    Mean Square   F Value   Pr > F
 REP                    3    405.0000000    135.0000000      7.53   0.0018
 BLOCO(REP)             8    278.0000000     34.7500000      1.94   0.1159
 FAT_A                  2    175.5000000     87.7500000      4.90   0.0200
 FAT_B                  2    216.5000000    108.2500000      6.04   0.0098
 FAT_A*FAT_B            2     21.5000000     10.7500000      0.60   0.5594
```

Figura 6. Análise de uma fatorial 3^2, com confundimento parcial da interação e uso do procedimento GLM do SAS.

Confundimento em fatorial 3^3

Os fatoriais com três fatores em três níveis, que são aplicados de forma intensa por pesquisadores da área agrícola, principalmente, foram estudados primeiramente por Yates (1937), que mostrou que cada dois graus de liberdade da interação dupla A*B*C podem ser utilizados para a constituição de blocos com nove tratamentos cada um. Yates chamou esses quatro conjuntos de grupos W, X, Y e Z.

As equações de definição de confundimento são:

Grupo W: $x_1 + 2*x_2 + 2*x_3 = r\ (mod\ 3)$;
Grupo X: $x_1 + 2*x_2 + x_3 = r\ (mod\ 3)$;
Grupo Y: $x_1 + x_2 + 2*x_3 = r\ (mod\ 3)$ e
Grupo Z: $x_1 + x_2 + x_3 = r\ (mod\ 3)$.

O uso da quarta equação (grupo Z) faz que os três blocos incompletos, cada um com nove tratamentos, tenham as seguintes composições:

Bloco 1 ($r = 0$) → 000, 012, 021, 102, 111, 120, 201, 210 e 222;
Bloco 2 ($r = 1$) → 001, 010, 022, 100, 112, 121, 202, 211 e 220 e
Bloco 3 ($r = 2$) → 002, 011, 020, 101, 110, 122, 200, 212 e 221.

De forma similar, podem-se obter as composições dos blocos para os demais grupos. Também aqui, como mostrado no caso dos fatoriais de base 2, um único contraste de definição pode ser adotado em todas as repetições (confundimento total) ou podem ser adotados contrastes distintos para cada uma das réplicas (confundimento parcial).

Exemplo de aplicação

No ano agrícola de 1978–1979, o pesquisador Nand Kumar Fagéria fez um estudo sobre a absorção e a translocação de zinco (Zn) em arroz de terras altas sob a influência de P e calcário, num delineamento em blocos ao acaso com três repetições. O experimento foi realizado na fazenda Capivara da Embrapa Arroz e Feijão. Os fatores em estudo tinham três níveis e eram, respectivamente, zinco (A) nas doses 0, 5 e 10 kg ha^{-1}, fósforo (B) nas doses de 0, 50 e 100 kg ha^{-1} de P_2O_5 e calcário (C) em 0, 1,5 e 3 t ha^{-1}.

Os dados de produtividade de grãos em kg ha^{-1}, com confundimento parcial de dois graus de liberdade da interação dupla, em que se utilizou o grupo Z de Yates, encontram-se na Tabela 16.

Tabela 16. Dados de produtividade de grãos em kg ha^{-1}, no ensaio fatorial 3^3, com confundimento de dois graus de liberdade da interação dupla.

Repetição 1			Repetição 2			Repetição 3		
Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3
000	001	002	000	001	002	000	001	002
1.200	1.600	1.250	950	1.650	2.150	1.300	1.800	2.350
012	010	011	012	010	011	012	010	011
3.900	3.500	4.100	4.200	4.300	3.850	4.550	3.400	4.550
021	022	020	021	022	020	021	022	020
4.200	4.150	4.300	3.950	4.950	4.900	4.500	4.550	4.750
102	100	101	102	100	101	102	100	101
1.800	1.150	1.450	2.600	2.200	1.650	3.150	4.300	2.350
111	112	110	111	112	110	111	112	110
3.350	4.200	3.850	4.500	3.750	4.200	4.250	4.450	4.250
120	121	122	120	121	122	120	121	122
4.700	4.200	4.450	5.600	5.200	4.640	4.850	4.600	5.000
201	202	200	201	202	200	201	202	200
1.100	1.850	1.100	2.000	2.350	1.900	2.250	1.800	1.500
210	211	212	210	211	212	210	211	212
2.600	3.500	3.500	3.750	3.900	4.450	4.500	2.150	3.800
222	220	221	222	220	221	222	220	221
4.000	4.450	4.850	3.650	5.350	3.750	4.700	4.950	3.350
26.850	28.600	28.850	31.200	33.650	31.490	34.050	32.000	31.900
	84.300			96.340			97.950	

As Tabelas 17, 18 e 19 contêm os dados dos três fatores e suas combinações que são necessários para a obtenção das somas de quadrados respectivas.

Tabela 17. Dados de produtividade para os fatores zinco e fósforo e suas combinações.

A (Zinco)	B (Fósforo)			Soma
	0	1	2	
0	14.250	36.350	40.250	90.850
1	20.650	36.800	43.240	100.690
2	15.850	32.150	39.050	87.050
Soma	50.750	105.300	122.540	278.590

Tabela 18. Dados de produtividade para os fatores zinco e calcário e suas combinações.

A (Zinco)	C (Calcário)			Soma
	0	1	2	
0	28.600	30.200	32.050	90.850
1	35.100	31.550	34.040	100.690
2	30.100	26.850	30.100	87.050
Soma	93.800	88.600	96.190	278.590

Tabela 19. Dados de produtividade para os fatores fósforo e calcário e suas combinações.

C (Calcário)	B (Fósforo)			Soma
	0	1	2	
0	15.600	34.350	43.850	93.800
1	15.850	34.150	38.600	88.600
2	19.300	36.800	40.090	96.190
Soma	50.750	105.300	122.540	278.590

A análise de variância é obtida da maneira usual, que já foi descrita nos Capítulos 8 e 9 e nos exemplos anteriores deste capítulo. A diferença, neste caso, reside, como no item *Exemplo de aplicação em fatorial* 3^2, na soma de quadrado da parte não confundida da interação dupla. Aqui a soma de quadrados da interação dupla, total, deve ser calculada como mostrado no Capítulo 8 e depois subtraída a soma de quadrados de blocos, como mostrado também em *Exemplo de aplicação em fatorial* 3^2.

Para exemplificar, calculam-se, a seguir, algumas dessas somas de quadrados.

$$SQRep = \frac{84.300^2 + 96.340^2 + 97.950^2}{27} - \frac{278.590^2}{81} = 4.121.928,40$$

$$SQZinco = \frac{90.850^2 + 100.690^2 + 87.050^2}{27} - \frac{278.590^2}{81} = 3.670.558,00$$

$$SQCalcario = \frac{93.800^2 + 88.600^2 + 96.190^2}{27} - \frac{278.590^2}{81} = 1.115.558,00$$

$$SQ(ZincoxCalcário) = \frac{28.600^2 + 30.200^2 + ... + 30.100^2}{9} - \frac{278.590^2}{81}$$
$$- SQZinco - SQCalcário$$

$$SQ(ZincoxCalcário) = 1.067.264,20$$

Para a soma de quadrados da parte não confundida da interação dupla, deve-se começar pela obtenção da soma de quadrados da interação dupla, como se não tivesse havido o confundimento, e depois subtrair a soma de quadrados para blocos, como se mostra a seguir. A soma de quadrados para tratamentos é obtida da forma usual com os dados mostrados na Tabela 20.

$$SQTratamentos = \frac{3.450^2 + 5.050^2 + ... + 12.350^2}{3} - \frac{278.590^2}{81} = 114.693.402,50$$

Tabela 20. Totais de tratamentos para o ensaio fatorial 3^3, com confundimento parcial da interação dupla.

A (Zinco)	B (Fósforo)	C (Calcário)		
		0	1	2
0	0	3.450	5.050	5.750
0	1	11.200	12.500	12.650
0	2	13.950	12.650	13.650
1	0	7.650	5.450	7.550
1	1	12.300	12.100	12.400
1	2	15.150	14.000	14.090
2	0	4.500	5.350	6.000
2	1	10.850	9.550	11.750
2	2	14.750	11.950	12.350

Para calcular a soma de quadrados de blocos, deve-se obter o total de cada um dos blocos, o que pode ser facilmente feito a partir da Tabela 16, ou seja:

Bloco 1 = 26.850 + 31.200 + 34.050 = 92.100,

Bloco 2 = 28.600 + 33.650 + 32.000 = 94.250 e

Bloco 3 = 28.850 + 31.490 + 31.900 = 92.240.

$$SQBlocos = \frac{92.100^2 + 94.250^2 + 92.240^2}{27} - \frac{278.590^2}{81} = 107.187,65$$

A soma de quadrados da interação não confundida é igual à diferença entre a soma de quadrados de tratamentos e a soma das somas de quadrados dos fatores principais, das interações simples e de blocos, de forma que se tem: $SQ(Zn*P*Cal) = 1.463.044,40$.

Pode-se agora montar a análise de variância, como se vê na Tabela 21.

Tabela 21. Análise de variância para o fatorial 3^3, com confundimento parcial da interação dupla. Ensaio com três repetições.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	2	4.121.928,40	2.060.964,20	
Bloco (Rep)	6	989.340,70	164.890,10	
Zinco (Zn)	2	3.670.558,00	1.835.279,00	6,52 (0,003)
Fósforo (P)	2	104.033.632,10	52.016.816,00	184,78 (< 0,001)
Calcário (Cal)	2	1.115.558,00	557.779,00	1,98 (0,149)
Zn*P	4	1.290.967,90	322.742,00	1,15 (0,346)
Zn*Cal	4	1.067.264,20	266.816,00	0,95 (0,445)
P*Cal	4	1.945.190,10	486.297,50	1,73 (0,159)
Zn*P*Cal	6	1.463.044,40	243.840,70	0,87 (0,526)
Erro	48	13.511.985,20	281.499,70	
Total	80	133.209.469,10		

Uso do SAS em fatoriais 3 x 3 x 3 com confundimento parcial da interação dupla

O exemplo dos casos anteriores de fatoriais (com confundimento) também se aplica aqui, com a especificação do modelo completo e o uso da opção SS1.

```
PROC GLM;CLASS REP BLOCO FAT_A FAT_B FAT_C;
MODEL PROD=REP BLOCO(REP) FAT_A|FAT_B|FAT_C/SS1;
```

O resultado se vê na Figura 7.

```
                          The GLM Procedure
                     Class Level Information
Class                 Levels    Values
REP                        3    1 2 3
BLOCO                      3    1 2 3
FAT_A                      3    0 1 2
FAT_B                      3    0 1 2
FAT_C                      3    0 1 2

                 Number of observations    81
                          The GLM Procedure

Dependent Variable: PROD

                                  Sum of
Source                     DF       Squares    Mean Square   F Value   Pr > F
Model                      32    119697484.0      3740546.4     13.29   <.0001
Error                      48     13511985.2       281499.7
Corrected Total            80    133209469.1

                  R-Square     Coeff Var      Root MSE     PROD Mean
                  0.898566      15.42618      530.5654      3439.383

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
REP                         2      4121928.4      2060964.2      7.32   0.0017
BLOCO(REP)                  6       989340.7       164890.1      0.59   0.7399
FAT_A                       2      3670558.0      1835279.0      6.52   0.0031
FAT_B                       2    104033632.1     52016816.0    184.78   <.0001
FAT_A*FAT_B                 4      1290967.9       322742.0      1.15   0.3462
FAT_C                       2      1115558.0       557779.0      1.98   0.1490
FAT_A*FAT_C                 4      1067264.2       266816.0      0.95   0.4447
FAT_B*FAT_C                 4      1945190.1       486297.5      1.73   0.1594
FAT_A*FAT_B*FAT_C           6      1463044.4       243840.7      0.87   0.5265
```

Figura 7. Análise de um fatorial 3^3 com confundimento de dois graus de liberdade da interação entre os três fatores, com o uso do SAS.

Análise de ensaios fatoriais com confundimento e uma repetição

Muitas vezes, por necessidades financeiras, ou de material, decide-se executar um experimento com apenas uma repetição. Esse procedimento é perfeitamente possível nos ensaios fatoriais, quando, ao assumir a inexistência de efeitos de interações maiores, estas podem ser usadas como um estimador do erro experimental. Isso permite a realização dos testes para os demais efeitos de interesse, condicionado ao número de graus de liberdade existentes, pois, quando estes são muito pequenos, somente grandes diferenças consequem ser detectadas.

Em fatoriais de base 2, quando o número de fatores é elevado, podem ser considerados de interesse os efeitos principais e as interações simples e duplas, e todas as demais serem combinadas para a constituição de um estimador da variância experimental. Com fatoriais de base 3 ou maior, muitas vezes as interações simples e duplas podem ser desdobradas em seus componentes lineares e quadráticos, e os de ordem maior podem ser incorporados ao estimador do erro experimental.

É também possível o uso de confundimento nesses casos, e extremamente útil quando o número de tratamentos é elevado.

Para um melhor entendimento dessas possíveis variações, o uso de exemplos de aplicação é fundamental.

Fatorial 2^5 com interações de ordem 3 ou maior como estimadores da variância experimental

Neste caso, não raro na agricultura em virtude dos múltiplos fatores que afetam uma cultura, o número de tratamentos é muito grande, o que torna impraticável a execução de experimentos em blocos completos e mesmo com o uso de repetições. Além disso, as interações de ordem superior à dupla, ou mesmo simples, são de difícil interpretação. Isso leva à necessidade de se dividirem os blocos em duas ou mais partes (confundimento), bem como à aplicação de apenas uma repetição.

Os dados utilizados neste exemplo foram obtidos no chamado experimento central, que procurava resumir as principais técnicas em investigação para arroz irrigado no final da década de 1970. O pesquisador Luis Fernando Stone era o responsável pelo projeto "Efeito do conjunto de tecnologia aplicado ao sistema de produção de arroz irrigado", que estudava os cinco principais problemas da cultura: controle de invasoras (0 = Stam F-34 e 1 = Ronstar), adubação (0 = 30 kg ha^{-1} de N e 1 = 60 kg ha^{-1} de N), irrigação (0 = permanente

e 1 = intermitente), controle de doenças (0 = sem controle e 1 = aplicação de Manzate) e cultivar (0 = IAC 435 e 1 = IR 841-63-5-I-9-33). Do experimento original, utiliza-se apenas a primeira repetição, enquanto a interação de quarta ordem (cinco fatores) foi confundida com blocos. A Tabela 22 resume essa informação, mostrando a produtividade obtida nas parcelas em kg ha^{-1}, em que M = manejo da irrigação, A = adubação, C = cultivar, D = controle de doenças e I = controle de invasoras.

Tabela 22. Dados de produtividade de arroz irrigado, em ensaio fatorial 2^5, uma repetição com confundimento da interação de quarta ordem MACDI.

Bloco 1								
Tratamento	00000	00011	00101	00110	01001	01010	01100	01111
MACDI	2.445,0	2.157,5	2.945,0	2.545,0	2.920,0	3.820,0	2.407,5	4.345,0
Tratamento	10001	10010	10100	10111	11000	11011	11101	11110
MACDI	1.545,0	1.645,0	1.545,0	2.445,0	2.982,5	2.795,0	3.320,0	1670,0
Bloco 2								
Tratamento	00001	00010	00100	00111	01000	01011	01101	01110
MACDI	3.220,0	3.695,0	2.857,5	3.470,0	4.020,0	4.432,5	3.682,5	3.382,5
Tratamento	10000	10011	10101	10110	11001	11010	11100	11111
MACDI	1.595,0	2.082,5	2.682,5	2.582,5	3.270,0	3.070,0	2.607,5	3.870,0

Para efetuar a análise de variância, é interessante montar quadros com os contrastes de estimação dos efeitos principais e interações simples (Tabelas 23 e 24), lembrando que, no caso de fatoriais de base 2, os coeficientes são sempre -1 e +1, em que -1 está associado ao nível mais baixo do fator e +1 ao mais alto. Para as interações, os coeficientes são obtidos por meio de produto dos coeficientes.

Tendo-se nas Tabelas 23 e 24 a estimação por meio de contrastes dos efeitos principais e das interações, e lembrando que suas somas de quadrados são calculadas facilmente por serem contrastes, falta calcular as somas de quadrados total, de blocos e do erro, que se calculam da forma usual, como já demonstrado tantas vezes.

Por exemplo:

$$SQTotal = (2.445^2 + 3.220^2 + ... + 3.870^2) -$$

$$\frac{(2.445 + 3.220 + ... + 3.870)^2}{32} = 20.428.432,62$$

Tabela 23. Estimação dos efeitos principais no caso do fatorial 2^5, com confundimento da interação de ordem maior e com uma repetição.

Tratamento	Produção	Irrigação (M)	Adubação (A)	Cultivar (C)	Doença (D)	Invasora (I)
00000	2.445,0	-1	-1	-1	-1	-1
00001	3.220,0	-1	-1	-1	-1	1
00010	3.695,0	-1	-1	-1	1	-1
00011	2.157,5	-1	-1	-1	1	1
00100	2.857,5	-1	-1	1	-1	-1
00101	2.945,0	-1	-1	1	-1	1
00110	2.545,0	-1	-1	1	1	-1
00111	3.470,0	-1	-1	1	1	1
01000	4.020,0	-1	1	-1	-1	-1
01001	2.920,0	-1	1	-1	-1	1
01010	3.820,0	-1	1	-1	1	-1
01011	4.432,5	-1	1	-1	1	1
01100	2.407,5	-1	1	1	-1	-1
01101	3.682,5	-1	1	1	-1	1
01110	3.382,5	-1	1	1	1	-1
01111	4.345,0	-1	1	1	1	1
10000	2.082,5	1	-1	-1	-1	-1
10001	1.595,0	1	-1	-1	-1	1
10010	1.545,0	1	-1	-1	1	-1
10011	1.645,0	1	-1	-1	1	1
10100	1.545,0	1	-1	1	-1	-1
10101	2.682,5	1	-1	1	-1	1
10110	2.582,5	1	-1	1	1	-1
10111	2.445,0	1	-1	1	1	1
11000	2.982,5	1	1	-1	-1	-1
11001	3.270,0	1	1	-1	-1	1
11010	3.070,0	1	1	-1	1	-1
11011	2.795,0	1	1	-1	1	1
11100	2.607,5	1	1	1	-1	-1
11101	3.320,0	1	1	1	-1	1
11110	1.670,0	1	1	1	1	-1
11111	3.870,0	1	1	1	1	1
Efeito		-12.637,5	13.137,5	662,5	2.887,5	5.537,5

Tabela 24. Estimação dos efeitos das interações simples para o fatorial 2^5, com confundimento da interação de ordem maior e com uma repetição.

Tratamento	Produção	M*A	M*C	M*D	M*I	A*C	A*D	A*I	C*D	C*I	D*I
00000	2.445,0	1	1	1	1	1	1	1	1	1	1
00001	3.220,0	1	1	1	-1	1	1	-1	1	-1	-1
00010	3.695,0	1	1	-1	1	1	-1	1	-1	1	0
00011	2.157,5	1	1	-1	-1	1	-1	-1	-1	-1	-2
00100	2.857,5	1	-1	1	1	-1	1	1	-1	-1	-1
00101	2.945,0	1	-1	1	-1	-1	1	-1	-1	1	0
00110	2.545,0	1	-1	-1	1	-1	-1	1	1	-1	0
00111	3.470,0	1	-1	-1	-1	-1	-1	-1	1	1	0
01000	4.020,0	-1	1	1	1	-1	-1	-1	1	1	0
01001	2.920,0	-1	1	1	-1	-1	-1	1	1	-1	0
01010	3.820,0	-1	1	-1	1	-1	1	-1	-1	1	0
01011	4.432,5	-1	1	-1	-1	-1	1	1	-1	-1	0
01100	2.407,5	-1	-1	1	1	1	-1	-1	-1	-1	-1
01101	3.682,5	-1	-1	1	-1	1	-1	1	-1	1	1
01110	3.382,5	-1	-1	-1	1	1	1	-1	1	-1	0
01111	4.345,0	-1	-1	-1	-1	1	1	1	1	1	2
10000	2.082,5	-1	-1	-1	-1	1	1	1	1	1	2
10001	1.595,0	-1	-1	-1	1	1	1	-1	1	-1	0

Continua...

Tabela 24. Continuação.

Tratamento	Produção	M*A	M*C	M*D	M*I	A*C	A*D	A*I	C*D	C*I	D*I
10010	1.545,0	-1	-1	1	-1	1	-1	1	-1	1	1
10011	1.645,0	-1	-1	1	1	1	-1	-1	-1	-1	-1
10100	1.545,0	-1	1	-1	-1	-1	1	1	-1	-1	0
10101	2.682,5	-1	1	-1	1	-1	1	-1	-1	1	0
10110	2.582,5	-1	1	1	-1	-1	-1	1	1	-1	0
10111	2.445,0	-1	1	1	1	-1	-1	-1	1	1	0
11000	2.982,5	1	-1	-1	-1	-1	-1	-1	1	1	0
11001	3.270,0	1	-1	-1	1	-1	-1	1	1	-1	0
11010	3.070,0	1	-1	1	-1	-1	1	-1	-1	1	0
11011	2.795,0	1	-1	1	1	-1	1	1	-1	-1	-1
11100	2.607,5	1	1	-1	-1	1	-1	-1	-1	-1	-2
11101	3.320,0	1	1	-1	1	1	-1	1	-1	1	0
11110	1.670,0	1	1	1	-1	1	1	-1	1	-1	-1
11111	3.870,0	1	1	1	1	1	1	1	1	1	1
Efeito		1.787,5	2.812,5	-3.812,5	1.537,5	-4.712,5	1.462,5	3.812,5	1.637,5	8.787,5	4.783,3

$$SQBlocos = \frac{(2.445 + ... + 3.320)^2 + (3.220 + ... + 3.870)^2}{16} -$$
$$\frac{(2.445 + 3.220 + ... + 3.870)^2}{32} = 2.524.223,63$$

$$SQIrrigação = \frac{(-12.637,5)^2}{32} = 4.990.825,20$$

$$SQ(MxC) = \frac{(2.812,5)^2}{32} = 247.192,38$$

Assim, para todos os demais fatores principais e interações simples, lembrando que o erro, como sempre, se obtém por diferença, mas que, neste caso em particular, representa a soma das somas de quadrados das interações duplas, triplas e quádruplas, de pouco interesse ao pesquisador ou de interpretação extremamente complicada. Dessa forma, a análise de variância se completa como se vê na Tabela 25.

Tabela 25. Análise de variância de um fatorial 2^5, com uma repetição e com a interação de ordem maior confundida com blocos, e as interações duplas, triplas e quádruplas combinadas como estimador do erro.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	1	2.524.223,63	2.524.223,63	18,76 (> 0,001)
Irrigação (M)	1	4.990.825,20	4.990.825,20	37,09 (< 0,001)
Adubação (A)	1	5.393.559,57	5.393.559,57	40,08 (< 0,001)
Cultivar (C)	1	13.715,82	13.715,82	0,10 (0,754)
Doenças (D)	1	490.668,94	490.668,94	3,65 (0,076)
Invasoras (I)	1	1.245.239,26	1.245.239,26	9,25 (0,008)
M*A	1	99.848,63	99.848,63	0,74 (0,403)
M*C	1	247.192,38	247.192,38	1,84 (0,195)
M*D	1	234.184,57	234.184,57	1,74 (0,207)
M*I	1	167.114,26	167.114,26	1,24 (0,283)
A*C	1	693.989,26	693.989,26	5,16 (0,038)
A*D	1	4.692,38	4.692,38	0,03 (0,854)
A*I	1	288.325,20	288.325,20	2,14 (0,164)
C*D	1	9.887,70	9.887,70	0,07 (0,790)
C*I	1	2.006.254,88	2.006.254,88	14,91 (0,002)
D*I	1	122,07	122,07	0,00 (0,976)
Erro	15	2.018.588,87	134.572,59	
Total	31	20.428.432,62		

Uso do SAS em fatoriais de base 2 com uma repetição: estimação de efeitos principais e interações simples

O SAS para este tipo de ensaio pode ser realizado com o procedimento ANOVA, sem maiores problemas. Para isso, usa-se o programa:

```
PROC ANOVA;CLASS BLOCO M A C D I;
MODEL PRODUCAO=BLOCO M A C D I M*A M*C M*D M*I A*C A*D A*I C*D C*I D*I;
```

A Figura 8 mostra o resultado desse processamento.

```
                    The ANOVA Procedure
                  Class Level Information
Class          Levels    Values
BLOCO               2    1 2
M                   2    0 1
A                   2    0 1
C                   2    0 1
D                   2    0 1
I                   2    0 1

                Number of observations    32
                    The ANOVA Procedure
Dependent Variable: PRODUCAO
                              Sum of
Source                DF      Squares    Mean Square   F Value   Pr > F
Model                 16  18409843.75     1150615.23      8.55   <.0001
Error                 15   2018588.87      134572.59
Corrected Total       31  20428432.62

           R-Square     Coeff Var      Root MSE    PRODUCAO Mean
           0.901187      12.75242      366.8414         2876.641

Source                DF       Anova SS    Mean Square   F Value   Pr > F
BLOCO                  1   2524223.633   2524223.633     18.76   0.0006
M                      1   4990825.195   4990825.195     37.09   <.0001
A                      1   5393559.570   5393559.570     40.08   <.0001
C                      1     13715.820     13715.820      0.10   0.7539
D                      1    490668.945    490668.945      3.65   0.0755
I                      1   1245239.258   1245239.258      9.25   0.0082
M*A                    1     99848.633     99848.633      0.74   0.4026
M*C                    1    247192.383    247192.383      1.84   0.1954
M*D                    1    234184.570    234184.570      1.74   0.2069
M*I                    1    167114.258    167114.258      1.24   0.2827
A*C                    1    693989.258    693989.258      5.16   0.0383
A*D                    1      4692.383      4692.383      0.03   0.8544
A*I                    1    288325.195    288325.195      2.14   0.1639
C*D                    1      9887.695      9887.695      0.07   0.7900
C*I                    1   2006254.883   2006254.883     14.91   0.0015
D*I                    1       122.070       122.070      0.00   0.9764
```

Figura 8. Resultado usado com o procedimento ANOVA, do SAS, na análise de um fatorial 2^5, com uma repetição e com a interação de ordem maior confundida com blocos, e as interações duplas, triplas e quádruplas combinadas como estimador do erro.

Fatorial 3^3 com confundimento parcial da interação dupla

Como exemplo de aplicação, considera-se o uso dos valores médios dos tratamentos do ensaio utilizado na Tabela 16 como o resultado obtido de um ensaio com apenas uma repetição. Evidentemente que o grupo de confundimento é o mesmo anterior. Os dados estão contidos na Tabela 26.

Assim como no caso do ensaio com três repetições, devem-se montar os três quadros auxiliares para cálculo das somas de quadrados dos fatores e de suas interações simples (Tabelas 27, 28 e 29). A partir da Tabela 26, podem-se calcular as somas de quadrados total e de blocos. A soma dos quadrados do erro se obtém por diferença.

Tabela 26. Valores médios (arredondados para o inteiro mais próximo) dos dados do ensaio na Tabela 16.

Bloco 1	Bloco 2	Bloco 3
000	001	002
1.150	1.683	1.917
012	010	011
4.217	3.733	4.167
021	022	020
4.217	4.550	4.650
102	100	101
2.517	2.550	1.817
111	112	110
4.033	4.133	4.100
120	121	122
5.050	4.667	4.697
201	202	200
1.783	2.000	1.500
210	211	212
3.617	3.183	3.917
222	220	221
4.117	4.917	3.983
30.701	31.416	30.748
	92.864	

Tabela 27. Valores totais dos fatores zinco e fósforo e de suas combinações.

A (Zinco)	B (Fósforo)			Soma
	0	1	2	
0	4.750	12.117	13.417	30.284
1	6.884	12.266	14.414	33.564
2	5.283	10.717	13.017	29.017
Soma	16.917	35.100	40.848	92.865

Tabela 28. Valores totais dos fatores zinco e calcário e de suas combinações.

A (Zinco)	C (Calcário)			Soma
	0	1	2	
0	9.533	10.067	10.684	30.284
1	11.700	10.517	11.347	33.564
2	10.034	8.949	10.034	29.017
Soma	31.267	29.533	32.065	92.865

Tabela 29. Valores totais dos fatores fósforo e calcário e de suas combinações.

C (Calcário)	B (Fósforo)			Soma
	0	1	2	
0	5.200	11.450	14.617	31.267
1	5.283	11.383	12.867	29.533
2	6.434	12.267	13.364	32.065
Soma	16.917	35.100	40.848	92.865

As somas de quadrados são obtidas da forma usual, algumas das quais são mostradas a seguir.

$$SQBlocos = \frac{30.701^2 + 31.416^2 + 30.748^2}{9} - \frac{92.865^2}{27} = 35.542{,}89$$

$$SQZinco = \frac{30.284^2 + 33.564^2 + 29.017^2}{9} - \frac{92.865^2}{27} = 1.223.662{,}89$$

$$SQCalcario = \frac{31.267^2 + 29.533^2 + 32.065^2}{9} - \frac{92.865^2}{27} = 372.392,00$$

$$SQ(ZincoxCalcário) = \frac{9.533^2 + 10.067^2 + ... + 10.034^2}{9} - \frac{92.865^2}{27}$$

$$SQZinco - SQCalcário$$

$$SQ(ZincoxCalcário) = 356.285,11$$

Da mesma forma, são calculadas as demais somas de quadrados para P, calcário, e para as interações Zn x P e P x calcário.

A soma de quadrados total é dada por:

$$SQTotal = 1.150^2 + 4.217^2 + ... + 3.983^2 - \frac{92.865^2}{27} = 38.234.502,67$$

Finalmente pode-se montar a análise de variância (Tabela 30).

Tabela 30. Análise de variância de um fatorial, com uma repetição e confundimento parcial da interação dupla. Parte não confundida da interação dupla utilizada como erro.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	2	35.542,89	17.771,44	
Zinco (Zn)	2	1.223.662,89	611.831,44	7,53 (0,023)
Fósforo (P)	2	34.679.768,67	17.339.884,33	213,28 (< 0,001)
Calcário (Cal)	2	372.392,00	186.196,00	2,29 (0.182)
Zn*P	4	430.611,11	107.652,78	1,32 (0,361)
Zn*Cal	4	356.285,11	89.071,28	1,10 (0,437)
P*Cal	4	648.430,00	162.107,50	1,99 (0,215)
Erro	6	487.810,00	81.301,67	
Total	26	38.234.502,67		

Uso do SAS em fatoriais 3 x 3 x 3 – uma repetição e confundimento parcial da interação dupla

Assim como no exemplo dado na Figura 8, o programa SAS é muito simples, conforme se vê a seguir.

```
PROC ANOVA;CLASS BLOCO FAT_A FAT_B FAT_C;
MODEL PRODUCAO=BLOCO FAT_A FAT_B FAT_C FAT_A*FAT_B FAT_A*FAT_C
FAT_B*FAT_C;
```

O resultado está apresentado na Figura 9.

```
                    The ANOVA Procedure
                     Class Level Information
                  Class         Levels    Values
                  BLOCO              3    1 2 3
                  FAT_A              3    0 1 2
                  FAT_B              3    0 1 2
                  FAT_C              3    0 1 2
                   Number of observations    27
                      The ANOVA Procedure
Dependent Variable: PRODUCAO
                                  Sum of
Source                  DF       Squares    Mean Square   F Value   Pr > F
Model                   20    37746692.67     1887334.63     23.21   0.0004
Error                    6      487810.00       81301.67
Corrected Total         26    38234502.67

              R-Square     Coeff Var      Root MSE    PRODUCAO Mean
              0.987242      8.290132      285.1345          3439.444

Source                  DF       Anova SS    Mean Square   F Value   Pr > F
BLOCO                    2       35542.89       17771.44      0.22   0.8098
FAT_A                    2     1223662.89      611831.44      7.53   0.0232
FAT_B                    2    34679768.67    17339884.33    213.28   <.0001
FAT_C                    2      372392.00      186196.00      2.29   0.1824
FAT_A*FAT_B              4      430611.11      107652.78      1.32   0.3606
FAT_A*FAT_C              4      356285.11       89071.28      1.10   0.4374
FAT_B*FAT_C              4      648430.00      162107.50      1.99   0.2146
```

Figura 9. Análise de um ensaio fatorial 3^3, com uma repetição e confundimento parcial da interação dupla, com o uso do SAS.

Fatorial 3^3 com partes das interações simples adicionadas ao estimador de erro

Como dito no início desta secção, muitas vezes o uso das interações maiores pode não garantir um número adequado de graus de liberdade para a execução do teste F, e somente diferenças muito grandes são detectadas como significativas. Lembrando que, ao fazer as

análises de variância com fatores quantitativos, estes podem ter seus efeitos estudados em maior detalhe por meio do estudo de polinômios ortogonais, cada um deles com 1 grau de liberdade, bem como as interações entre dois fatores podem ser também estudadas em detalhes e desdobradas nas interações polinomiais linear x linear, linear x quadrática, quadrática x linear e quadrática x quadrática, como, por exemplo, no caso de fatoriais em base 3. Pode ser que essas interações não sejam todas de interesse do pesquisador e parte delas seja adicionada ao estimador do erro. Supondo ser esse o caso para o exemplo anterior e que então seja de interesse estudar cada um dos graus de liberdade individualmente para os fatores principais e interação linear x linear para as interações simples, segue que nove graus de liberdade sejam adicionados ao erro, que passará, portanto, a contar com um total de 15 graus de liberdade.

O processo de estimação desses componentes polinomiais é extremamente simples e já foi discutido em detalhes anteriormente. Na Tabela 31, apresenta-se exemplo de um dos fatores principais (P) e da interação Zn x P.

Tabela 31. Valores totais para os fatores fósforo e zinco e suas combinações e coeficientes dos polinômios linear e quadrático no caso de três níveis igualmente espaçados.

A (Zinco)	B (Fósforo)			Soma	Contraste linear	Contraste quadrático
	0	1	2			
0	4.750 (1)	12.117 (0)	13.417 (-1)	30.284	-1	1
1	6.884 (0)	12.266 (0)	14.414 (0)	33.564	0	-2
2	5.283 (-1)	10.717 (0)	13.017 (1)	29.017	1	1
Soma	16.917	35.100	40.848	92.865		
Contraste linear	-1	0	1			
Contraste quadrático	1	-2	1			

Para a estimação dos efeitos linear e quadrático para P, aplicam-se as fórmulas de soma de quadrados de contrastes, neste caso multiplicando o total de cada um dos níveis de P pelo correspondente coeficiente, ou seja:

$$SQFosforo(linear) = \frac{[(-1)*16.917 + 0*35.100 + 1*40.848]^2}{9*2} =$$

$$\frac{(23.931)^2}{18} = 31.816.264,50$$

$$SQFosforo(quadrático) = \frac{[1*16.917 + (-2)*35.100 + 1*40.848]^2}{9*6} = \frac{(-12.435)^2}{54} = 2.863.504,17$$

Nessas duas fórmulas, nos denominadores, o número 9 representa a quantidade de unidades experimentais que deram origem ao total correspondente a cada nível de P, enquanto 2 e 6 representam as somas dos quadrados dos coeficientes dos contrastes linear e quadrático, respectivamente.

Para a estimação do efeito da interação linear x linear, componente da interação simples Zn x P, o valor correspondente a cada combinação entre os níveis dos dois fatores é multiplicado pelo produto dos coeficientes dos contrastes linear para Zn e para P que, no caso atual, estão representados na Tabela 31, entre parêntesis, em cada uma das células, e o resultado se processa como no caso dos contrastes para um fator simples. Isto é:

$$SQ(ZnLinxPLin) = \frac{[1*4.750 + (-1)*5.283 + (-1)*13.417 + 1*13.017]^2}{3*4} = \frac{(-933)^2}{12} = 72.540,75$$

Isso significa que três graus de liberdade serão transferidos para o erro juntamente com a diferença na soma de quadrados que é igual a 358.070,36. O mesmo procedimento se aplica às outras interações simples, de forma que o quadro de análise de variância apresenta-se na Tabela 32.

Como se pode ver, para os fatores principais Zn e P há um efeito quadrático significativo, enquanto para calcário não foi observada qualquer significância das regressões polinomiais. Das interações simples, apenas foi significativa a combinação linear para a interação P x calcário. Pouca alteração se nota na estimação do erro, que é agora estimado com mais precisão, ou seja, com um maior número de graus de liberdade.

Uso do SAS em experimento fatorial 3^3, com partes das interações simples adicionadas ao estimador de erro

Para se obter esta análise, o procedimento adequado é o GLM, pois só ele permite a criação de contrastes. Deve-se usar uma opção do PROC GLM chamada de OUTSTAT=nome

Tabela 32. Análise de variância para fatorial 3^3, com desdobramento dos efeitos principais e interações simples, com confundimento parcial da interação dupla e uma repetição.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	2	35.542,89	17.771,44	
Zinco linear (Zn lin)	1	89.182,72	89.182,72	1,09 (0,313)
Zinco quadrático	1	1.134.480,17	1.134.480,17	13,89 (0,002)
Fósforo linear (P lin)	1	31.816.264,50	31.816.264,50	389,66 (< 0,001)
Fósforo quadrático	1	2.863.504,17	2.863.504,17	35,07 (< 0,001)
Calcário linear (Cal lin)	1	35.378,00	35.378,00	0,43 (0,522)
Calcário quadrático	1	337.014,00	337.014,00	4,13 (0,060)
Zn lin x P lin	1	72.540,75	72.540,75	0,89 (0,360)
Zn lin x Cal lin	1	110.400,08	110.400,08	1,35 (0,263)
P lin x Cal lin	1	515.430,75	515.430,75	6,31 (0,024)
Erro	15	1.224.764,64	81.650,98	
Total	26	38.234.502,67		

do conjunto de dados. Essa opção cria um conjunto de dados que conterá uma série de variáveis, entre elas, as fontes, os graus de liberdade, as somas de quadrados, os testes F e suas probabilidades. Esse conjunto de dados deve ser posteriormente trabalhado para que se obtenha a análise de variância que aparece na Tabela 32. O programa inicial é:

```
PROC GLM OUTSTAT=B NOPRINT;CLASS BLOCO FAT_A FAT_B FAT_C;
MODEL PRODUCAO=BLOCO FAT_A FAT_B FAT_C FAT_A*FAT_B FAT_A*FAT_C
FAT_B*FAT_C/SS1;
CONTRAST "A LIN" FAT_A -1 0 1;
CONTRAST "A QUA" FAT_A 1 -2 1;
CONTRAST "B LIN" FAT_B -1 0 1;
CONTRAST "B QUA" FAT_B 1 -2 1;
CONTRAST "C LIN" FAT_C -1 0 1;
CONTRAST "C QUA" FAT_C 1 -2 1;
CONTRAST "ALIN X BLIN" FAT_A*FAT_B 1 0 -1 0 0 0 -1 0 1;
CONTRAST "ALIN X BQUA" FAT_A*FAT_B -1 2 -1 0 0 0 1 -2 1;
CONTRAST "AQUA X BLIN" FAT_A*FAT_B -1 0 1 2 0 -2 -1 0 1;
CONTRAST "AQUA X BQUA" FAT_A*FAT_B 1 -2 1 -2 4 -2 1 -2 1;
CONTRAST "ALIN X CLIN" FAT_A*FAT_C 1 0 -1 0 0 0 -1 0 1;
CONTRAST "ALIN X CQUA" FAT_A*FAT_C -1 2 -1 0 0 0 1 -2 1;
```

```
CONTRAST "AQUA X CLIN" FAT_A*FAT_C -1 0 1 2 0 -2 -1 0 1;
CONTRAST "AQUA X CQUA" FAT_A*FAT_C 1 -2 1 -2 4 -2 1 -2 1;
CONTRAST "BLIN X CLIN" FAT_B*FAT_C 1 0 -1 0 0 0 -1 0 1;
CONTRAST "BLIN X CQUA" FAT_B*FAT_C -1 2 -1 0 0 0 1 -2 1;
CONTRAST "BQUA X CLIN" FAT_B*FAT_C -1 0 1 2 0 -2 -1 0 1;
CONTRAST "BQUA X CQUA" FAT_B*FAT_C 1 -2 1 -2 4 -2 1 -2 1;
DATA C;SET B;FONTE=_SOURCE_;DROP _NAME_ _SOURCE_ _TYPE_ F PROB;
IF FONTE="BLOCO" OR FONTE="A LIN" OR FONTE="A QUA" OR FONTE="B LIN" OR
FONTE="B QUA" OR
FONTE="C LIN" OR FONTE="C QUA" OR FONTE="ALIN X BLIN" OR FONTE="ALIN X
CLIN" OR FONTE="BLIN X CLIN";MS=SS/DF;
DATA D;SET B;FONTE=_SOURCE_;DROP _SOURCE_ _NAME_ _TYPE_ F PROB;
IF FONTE="ERROR" OR FONTE="ALIN X BQUA" OR FONTE="AQUA X BLIN" OR
FONTE="AQUA X BQUA" OR FONTE="ALIN X CQUA"
OR FONTE="AQUA X CLIN" OR FONTE="AQUA X CQUA" OR FONTE="BLIN X CQUA" OR
FONTE="BQUA X CLIN" OR FONTE="BQUA X CQUA";
PROC MEANS SUM NOPRINT;VAR DF;OUTPUT OUT=E SUM=DF;
PROC MEANS DATA=D SUM NOPRINT;VAR SS;OUTPUT OUT=F SUM=SS;
DATA EF;MERGE E F;FONTE="ERROR";MS=SS/DF;KEEP FONTE DF SS MS;
DATA ANA;SET C EF;PROC PRINT;VAR FONTE DF SS MS;
```

O resultado final desta programação encontra-se na Figura 10. Verifica-se ali a necessidade de calcular ainda os testes F e verificar suas probabilidades.

Obs	FONTE	DF	SS	MS
1	BLOCO	2	35542.89	17771.44
2	A LIN	1	89182.72	89182.72
3	A QUA	1	1134480.17	1134480.17
4	B LIN	1	31816264.50	31816264.50
5	B QUA	1	2863504.17	2863504.17
6	C LIN	1	35378.00	35378.00
7	C QUA	1	337014.00	337014.00
8	ALIN X BLIN	1	72540.75	72540.75
9	ALIN X CLIN	1	110400.08	110400.08
10	BLIN X CLIN	1	515430.75	515430.75
11	ERROR	15	1224764.64	81650.98

Figura 10. Análise de variância de um ensaio 3^3, com confundimento parcial da interação dupla, e o uso da opção OUTSTAT=data_name do PROC GLM, no SAS.

Capítulo 11

Análise de grupo de experimentos

É comum na pesquisa agrícola, bem como em outros ramos da ciência, a realização de experimentos repetidos em diferentes locais, anos ou condições experimentais e, por meio da análise desses experimentos como um grupo homogêneo de ensaios, verificar o efeito dessas condições sobre os tratamentos. Esse efeito conjunto ou interação, identificado como locais x tratamentos, ou anos x tratamentos, ou ainda ensaios x tratamentos, é de especial interesse em muitos ramos de conhecimento e fornece aos pesquisadores uma gama de informações extremamente úteis sobre os tratamentos.

De maneira geral, procura-se que todos os ensaios tenham uma variância homogênea, ou seja, que a estimação do erro de cada um deles seja igual ou pelo menos não seja muito diferente. Além disso, devem, preferencialmente, ter todos eles os mesmos tratamentos, ou pelo menos um tratamento em comum, o qual é necessário para garantir o princípio da conectividade, como foi visto no caso dos ensaios em blocos incompletos ou nos de blocos aumentados. Quando apenas um ou poucos tratamentos são comuns, Pavate (1961) apresenta a solução para blocos incompletos, e Pimentel-Gomes e Guimarães (1968) e Pimentel-Gomes (1970) apresentam solução para blocos completos. Em qualquer uma das situações, é necessária a realização das análises individuais dos experimentos e, posteriormente, de uma análise conjunta.

Neste capítulo, vai-se estudar a análise de grupos de experimentos com todos os tratamentos em comum para os delineamentos completamente ao acaso e blocos completos ao acaso, com os tratamentos em estrutura fatorial ou não, além do caso de alguns tratamentos comuns em blocos completos ao acaso e quadrado latino.

Análise de experimentos com todos os tratamentos em comum

As análises individuais se executam como visto nos Capítulos 2, 3 e 4 e, para a análise conjunta, o modelo matemático é modificado para inclusão do efeito de ensaios (locais,

anos, condições e suas combinações) e da interação ensaios x tratamentos. No caso de blocos completos, a fonte de variação repetição ou blocos é estudada dentro de ensaios e o mesmo se dá em quadrado latino com linhas e colunas. A interação ensaios x tratamentos pode ser desmembrada em componentes menores, caso os tratamentos venham de uma estrutura fatorial ou os ensaios componham, por sua vez, um tipo de estrutura fatorial.

Para o caso mais simples, do delineamento completamente ao acaso, o modelo matemático é

$$Y_{ijk} = m + E_k + t_i + (E^*t)_{ik} + e_{ij(k)} \quad [1]$$

em que E_k indica o efeito de ensaios, t_i o efeito de tratamentos, $(E^*t)_{ik}$ o efeito da interação ensaios x tratamentos, $e_{ij(k)}$ o erro associado a cada repetição de tratamento em cada um dos ensaios. Caso os ensaios sejam uma combinação de experimentos conduzidos em vários locais em mais de um ano, por exemplo, o efeito E_k pode ser desdobrado nos efeitos individuais de locais, anos e da interação local x ano, com equivalente desdobramento da interação ensaio x tratamento (E^*t). Dessa forma, o modelo se modifica para:

$$Y_{ijkl} = m + L_k + A_l + (L^*A)_{kl} + t_i + (L^*t)_{ik} + (A^*t)_{il} + (L^*A^*t)_{ikl} + e_{ij(kl)} \quad [2]$$

No caso de blocos completos o modelo apresenta a seguinte forma:

$$Y_{ijk} = m + E_k + b(E)_{j(k)} + t_i + (E^*t)_{ik} + e_{ij(k)} \quad [3]$$

Por sua vez, para quadrado latino o modelo é:

$$Y_{ijk} = m + E_k + c(E)_{j(k)} + l(E)_{m(k)} + t_i + (E^*t)_{ik} + e_{ijm(k)} \quad [4]$$

Estes dois últimos modelos podem ser alterados da mesma forma que foi feita para o modelo de delineamento completamente ao acaso, caso os ensaios tenham uma estrutura fatorial.

Para simplificar a análise de variância, nos exemplos que se seguem, assume-se que todos os efeitos são fixos. Caso algum deles ou todos sejam aleatórios na análise de variância, deve-se modificar a forma de execução dos testes F como se viu anteriormente.

Não havendo perda de dados em qualquer ensaio, a análise conjunta se faz de uma maneira muito simples. A soma de quadrados total resulta da soma das somas de

quadrados total de cada ensaio mais a soma dos respectivos fatores de correção menos um fator de correção comum, enquanto os graus de liberdade total são iguais ao total de parcelas experimentais em todos os ensaios menos 1. Com exceção da soma de quadrados de tratamentos, as demais somas de quadrado e respectivos graus de liberdade são obtidos por meio da soma dos graus de liberdade e das somas de quadrados de cada ensaio individual, de tal maneira que a soma de quadrados de ensaios é igual à soma dos fatores de correção individuais menos o fator de correção comum. Para tratamentos, há a necessidade de obtenção de sua soma de quadrados, utilizando-se os totais de cada tratamento para todos os ensaios, enquanto a soma de quadrados para a interação ensaios x tratamentos se obtém por diferença. Isso ficará muito claro nos exemplos a seguir. A comparação de médias se faz da maneira usual.

Exemplo de aplicação: ensaios no delineamento completamente ao acaso

Com a cultura de arroz irrigado, foram conduzidos ensaios com diferentes manejos de água, dos quais se escolheram dois como exemplo: um com inundação contínua durante todo o ciclo da cultura e outro com inundação até a diferenciação do primórdio floral e drenado após essa fase. Os dados das Tabelas 1 e 2 referem-se à produtividade de grãos.

As análises individuais destes dois ensaios (Tabelas 3 e 4) foram feitas segundo o procedimento apresentado no Capítulo 2 e, portanto, não são detalhadas aqui.

Tabela 1. Dados de produtividade de grãos de arroz irrigado em lavoura conduzida com inundação contínua durante todo o ciclo.[1]

Tratamentos	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Repetição 5	Repetição 6	Total
CNA 7830	7.012	7.348	6.894	7.108	6.833	6.587	41.782
CNA 8502	6.957	7.865	6.767	7.164	6.064	6.742	41.559
CNA 8540	5.738	5.554	5.848	5.370	5.811	5.554	33.875
Epagri 108	7.026	6.783	7.807	6.466	6.691	6.966	41.739
Formoso	7.471	6.892	7.517	7.773	7.544	7.280	44.477
Metica 1	8.120	7.894	7.181	7.052	7.654	7.814	45.715
Supremo 1	6.489	6.188	6.148	6.285	6.528	6.217	37.855
Total							287.002

Dados cedidos pelo pesquisador Alberto Baeta dos Santos, da Embrapa Arroz e Feijão.

Tabela 2. Dados de produtividade de grãos de arroz irrigado em lavoura conduzida com inundação contínua até a fase de diferenciação do primórdio floral e drenada após essa fase.[1]

Tratamentos	Repetição 1	Repetição 2	Repetição 3	Repetição 4	Repetição 5	Repetição 6	Total
CNA 7830	7.660	6.205	6.173	7.395	6.504	6.770	40.707
CNA 8502	7.084	7.334	7.227	8.120	7.857	6.592	44.214
CNA 8540	5.988	6.616	6.357	5.138	5.581	4.953	34.633
Epagri 108	6.520	6.162	6.535	7.697	6.617	6.838	40.369
Formoso	7.233	6.922	6.886	6.482	6.563	8.033	42.119
Metica 1	6.832	7.268	7.177	5.971	7.475	6.830	41.553
Supremo 1	5.920	6.132	6.868	6.422	6.039	5.724	37.105
Total							280.700

Dados cedidos pelo pesquisador Alberto Baeta dos Santos, da Embrapa Arroz e Feijão.

Tabela 3. Análise de variância para o ensaio 1. Inundação durante todo o ciclo da cultura.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr >F)
Tratamentos	6	16.074.588,24	2.679.098,04	19,64 < ,0001
Erro	35	4.774.483,67	136.413,82	
Total	41	20.849.071,90		

Tabela 4. Análise de variância para o ensaio 2. Inundação durante todo o ciclo até a fase de diferenciação floral.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr >F)
Tratamentos	6	10.401.915,00	1.733.652,50	5,59 0,0004
Erro	35	10.849.364,33	309.981,84	
Total	41	21.251.279,33		

Para os ensaios 1 e 2 os fatores de correção são:

$$FC1 = \frac{287.002^2}{42} = 1.961.194.000,10 \quad \text{e} \quad FC2 = \frac{280.700^2}{42} = 1.876.011.666,67$$

Esses valores serão utilizados a seguir para a análise conjunta desses ensaios, para o cálculo da soma de quadrados de ensaios e da soma de quadrados total, como foi dito anteriormente.

Para a realização da análise de variância conjunta, monta-se a Tabela 5, que contém os totais de tratamentos para cada um dos ensaios e para o conjunto deles. Na realidade, apenas o total geral de cada tratamento é necessário para a realização dos cálculos como será visto a seguir.

Tabela 5. Análise de variância com os dados totais de produtividade de grãos para os dois ensaios e o total geral.

Tratamento	Ensaio 1	Ensaio 2	Total
CNA 7830	41.782	40.707	82.489
CNA 8502	41.559	44.214	85.773
CNA 8540	33.875	34.633	68.508
Epagri 108	41.739	40.369	82.108
Formoso	44.477	42.119	86.596
Metica 1	45.715	41.553	87.268
Supremo 1	37.855	37.105	74.960
Total	287.002	280.700	567.702

O fator de correção geral para a análise conjunta é igual a

$$FCg = \frac{567.702^2}{84} = 3.836.732.866,71$$

Dessa forma, inicia-se então a análise conjunta.

$$SQTotal = SQTotal_1 + SQTotal_2 + FC1 + FC2 - FCg$$

$$SQTotal = 20.849.071,90 + 21.251.279,33 + 1.961.194.000,10 + 1.876.011.666,67 - 3.836.732.866,71 = 42.573.151,29$$

$$SQEnsaios = FC1 + FC2 - FCg$$

$$SQEnsaios = 1.961.194.000,10 + 1.876.011.666,67 - 3.836.732.866,71 = 472.800,06$$

$$SQErro = SQErro_1 + SQErro_2$$

$$SQErro = 4.774.483,67 + 10.849.364,33 = 15.623.848,00$$

$$SQTrat = \frac{\sum T_i^2}{JK} - FCg$$

$$SQTrat = \frac{82.849^2 + 85.773^2 + ... + 74.960^2}{2 * 6} - 3.836.732.866,71 = 24.107.551,45$$

e finalmente

$$SQ(Ens \text{ x } Trat) = SQTrat1 + SQTrat2 - SQTrat$$

$$SQ(Ens \text{ x } Trat) = 16.074.588,24 + 10.401.915,00 - 24.107.551,45 = 2.368.951,79$$

Para a comparação de médias, pode ser empregado qualquer um dos testes discutidos no Capítulo 1. Como se pode ver na Tabela 6, não houve interação entre ensaios e tratamentos, o que é uma indicação de que as cultivares responderam de maneira semelhante nos dois experimentos, que, como se sabe, diferiram apenas no manejo de água. O fato de o efeito de ensaios também não ter sido significativo indica que, para essas cultivares, nas condições desses experimentos, o fato de manter a cultura drenada após a diferenciação do primórdio floral não afetou significativamente a produtividade da cultura.

Tabela 6. Análise de variância conjunta quando os dois ensaios foram conduzidos no delineamento completamente ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr >F)
Ensaios	1	472.800,06	472.800,06	2,12 (0,1500)
Tratamentos	6	24.107.551,45	4.017.925,24	18,00 (0 < ,0001)
Ensaios x tratamentos	6	2.368.951,79	394.825,30	1,77 (0,1181)
Erro	70	15.623.848,00	223.197,83	
Total	83	42.573.151,29		

Uso do SAS na análise conjunta de experimentos inteiramente casualizados

As análises individuais são realizadas conforme mostrado no Capítulo 2. Para a análise conjunta, o programa SAS tem a seguinte forma:

```
PROC ANOVA;CLASS ENSAIO TRAT;
MODEL PROD=ENSAIO TRAT ENSAIO*TRAT;
```

O resultado da análise está na Figura 1.

```
                          The ANOVA Procedure
                       Class Level Information
Class       Levels  Values
ENSAIO           2  1 2
TRAT             7  CNA 7830 CNA 8502 CNA 8540 EPAGRI 108 FORMOSO
                    METICA 1 SUPREMO 1

                        Number of observations    84
                          The ANOVA Procedure

Dependent Variable: PROD
                                  Sum of
 Source                DF       Squares    Mean Square   F Value   Pr > F
 Model                 13    26949303.29     2073023.33      9.29   <.0001
 Error                 70    15623848.00      223197.83
 Corrected Total       83    42573151.29

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.633012      6.990429      472.4382      6758.357

 Source                DF      Anova SS    Mean Square   F Value   Pr > F
 ENSAIO                 1     472800.05     472800.05      2.12   0.1500
 TRAT                   6   24107551.45    4017925.24     18.00   <.0001
 ENSAIO*TRAT            6    2368951.79     394825.30      1.77   0.1181
```

Figura 1. Análise conjunta de ensaios em desenho inteiramente ao acaso, com todos os tratamentos comuns, usando o SAS.

Exemplo de aplicação – ensaios em blocos completos ao acaso

Para exemplificar o procedimento de análise conjunta de experimentos em blocos completos ao acaso, muito similar ao que se demonstrou no exemplo anterior, utilizam-se dois experimentos conduzidos pelo pesquisador Anne Sitarama Prabhu e pela estudante Gisele Barata da Silva, no ano de 1999, em que se avaliaram as cultivares para área foliar atacada por brusone (*Pyricularia Orizae* L.) em diferentes datas, e nos quais se calculou a área sob a curva do progresso da doença. O experimento 1 foi semeado na densidade de 80

sementes por metro e o experimento 2 na densidade de 200 sementes por metro. Não foi, nesse caso, feita qualquer proteção das sementes ou das plantas ao fungo. A análise que se apresenta nas Tabelas 7 a 10 foi efetuada com os dados transformados por logaritmo natural, em que se procurou maior homogeneização das variâncias.

Tabela 7. Dados de área sob a curva do progresso da doença, transformados por logaritmo natural, para o experimento semeado com 80 sementes por metro.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Total
Caiapó	5,74616	6,05715	6,49288	18,29619
Canastra	6,29754	6,03998	6,13508	18,47260
Confiança	6,35712	5,83999	5,75745	17,95456
Maravilha	4,92718	6,03634	5,44038	16,40390
Primavera	6,26963	5,91024	6,47247	18,65234
R. Paranaíba	6,38072	6,12295	6,75723	19,26090
Total	35,97835	36,00665	37,05549	109,04049

Tabela 8. Dados de área sob a curva do progresso da doença, transformados por logaritmo natural, para o experimento semeado com 200 sementes por metro.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Total
Caiapó	4,82607	5,40380	4,92129	15,15116
Canastra	4,99789	5,23228	4,62698	14,85715
Confiança	5,67312	5,30094	5,51192	16,48598
Maravilha	4,37519	5,04110	4,05820	13,47449
Primavera	6,08142	5,96179	4,87302	16,91623
R. Paranaíba	5,94118	5,86669	5,08187	16,88974
Total	31,89487	32,80660	29,07328	93,77475

Tabela 9. Análise de variância para o ensaio 1. Densidade de semeadura de 80 sementes por metro.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr >F)
Blocos	2	0,1256	0,0628	
Tratamentos	5	1,5652	0,3130	2,26 (0,1272)
Erro	10	1,3829	0,1383	
Total	17	3,0737		

Tabela 10. Análise de variância para o ensaio 2. Densidade de semeadura de 200 sementes por metro.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr >F)
Blocos	2	1,2628	0,6314	
Tratamentos	5	3,1489	0,6298	6,12 (0,0075)
Erro	10	1,0284	0,1028	
Total	17	5,4401		

A Tabela 11 é necessária para calcular a soma de quadrados de tratamentos na análise conjunta e, se for o caso, a significância da interação ensaios x tratamentos. Pode ser adaptada para conter as médias de cada ensaio e a média geral dos tratamentos.

Tabela 11. Tabela auxiliar para análise de variância conjunta, com os totais dos tratamentos em cada um dos ensaios e o total geral.

Tratamento	Ensaio 1	Ensaio 2	Total
Caiapó	18,29619	15,15116	33,44735
Canastra	18,47260	14,85715	33,32975
Confiança	17,95456	16,48598	34,44054
Maravilha	16,40390	13,47449	29,87839
Primavera	18,65234	16,91623	35,56857
R. Paranaíba	19,26090	16,88974	36,15064
Total	109,04049	93,77475	202,81524

A única diferença na realização da análise conjunta (Tabela 12) é que agora aparece uma nova fonte de variação, a de blocos dentro dos ensaios, representada por Rep(ensaios), que é obtida pela simples soma das somas de quadrados de repetições (blocos) de cada um dos ensaios individuais.

Ficam a cargo do leitor a interpretação dos resultados, a comparação de médias e a realização dos cálculos necessários para o preenchimento da Tabela 12.

Tabela 12. Análise de variância conjunta. Experimentos conduzidos no delineamento em blocos completos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr >F)
Ensaios	1	6,4734	6,4734	53,69 < ,0001
Rep(ensaios)	4	1,3884	0,3471	
Tratamentos	5	4,1314	0,8262	6,85 0,0007
Ensaios x culturas	5	0,5828	0,1165	0,97 0,4614
Erro	20	2,4112	0,1206	
Total	35	14,9872		

Uso do SAS na análise conjunta de experimentos em blocos completos ao acaso

Novamente aqui, a análise individual é realizada como mostrado no Capítulo 3, e a análise conjunta tem a seguinte programação SAS:

```
PROC ANOVA;CLASS ENSAIO BLOCO TRAT;
MODEL AREA_L=ENSAIO BLOCO(ENSAIO) TRAT ENSAIO*TRAT;
```

O resultado se vê na Figura 2.

Grupo de experimentos com um ou alguns tratamentos comuns

Quando nem todos os tratamentos aparecem em todos os ensaios, dois casos devem ser considerados: somente um deles é comum ou mais de um é comum. Apesar de os modelos matemáticos serem diferentes nesses dois casos, é comum a ambos os processos a necessidade de ajuste das médias dos tratamentos não comuns de acordo com a resposta do(s) tratamento(s) comum(ns), da mesma forma como se viu no Capítulo 7, com relação ao desenho em blocos aumentados de Federer. Quando apenas um tratamento é comum aos experimentos, não aparece no modelo qualquer referência a esse fato. No entanto, se houver dois ou mais tratamentos em comum, inclui-se no modelo o efeito da interação tratamentos comuns x ensaios.

```
                          The ANOVA Procedure
                        Class Level Information
Class        Levels   Values
ENSAIO            2   1 2
BLOCO             3   1 2 3
TRAT              6   Caiapo Canastra Confianca Maravilha Primavera
                      R.Paranaiba

                      Number of observations    36
                          The ANOVA Procedure
Dependent Variable: AREA_L
                                 Sum of
 Source                DF       Squares    Mean Square   F Value   Pr > F
 Model                 15    12.57598816     0.83839921      6.95   <.0001
 Error                 20     2.41123349     0.12056167
 Corrected Total       35    14.98722164

              R-Square     Coeff Var      Root MSE    AREA_L Mean
              0.839114      6.163204      0.347220       5.633757

 Source                DF      Anova SS    Mean Square   F Value   Pr > F
 ENSAIO                 1    6.47341160     6.47341160     53.69   <.0001
 BLOCO(ENSAIO)          4    1.38841093     0.34710273      2.88   0.0493
 TRAT                   5    4.13135385     0.82627077      6.85   0.0007
 ENSAIO*TRAT            5    0.58281177     0.11656235      0.97   0.4614
```

Figura 2. Análise conjunta de ensaios em blocos completos casualizados com todos os tratamentos comuns, usando o SAS.

Quando existe apenas um tratamento comum entre os experimentos, a soma de quadrados das fontes se fará como no caso de todos os tratamentos comuns, exceto para tratamentos que se obtêm também por soma das somas de quadrados nos experimentos, como se verá nos exemplos dados nos itens *Exemplo de aplicação: blocos completos ao acaso com um tratamento comum* e *Exemplo de aplicação: quadrado latino com um tratamento comum*. Se houver mais de um tratamento em comum, há necessidade da construção de um quadro auxiliar de dupla entrada. Uma delas corresponde aos ensaios e a outra aos tratamentos comuns. A partir desse quadro, obtém-se a soma de quadrados para a interação tratamentos comuns x ensaios. A diferença entre a soma das somas de quadrado de tratamentos de cada ensaio e a soma de quadrados da interação tratamentos comuns x ensaios representa a soma de quadrados de tratamentos ajustada, como se verá no item *Exemplo de aplicação: blocos completos ao acaso com mais de um tratamento comum*.

Uma complicação aparece no que se refere às comparações de médias, pois aí há que se considerar o fato de as comparações se darem entre tratamentos comuns, entre tratamentos não comuns de um mesmo grupo ou de grupos distintos ou ainda entre tratamentos comuns e não comuns, pois os diferentes números de repetições e as diferentes ocorrências dentro de cada bloco afetam a estimação da variância dos contrastes respectivos.

Os tratamentos comuns têm como média a média aritmética dos dados respectivos em todos os experimentos. Por sua vez, os tratamentos não comuns necessitam de uma

correção que será subtraída de sua média aritmética, e esse ajuste é obtido pela diferença entre a média dos tratamentos comuns no seu experimento e a média geral dos tratamentos comuns para todos os experimentos. Deve ser lembrado que a soma das correções das médias dos tratamentos não comuns é igual a zero.

As comparações de tratamentos, ou contrastes entre tratamentos, apresentam as seguintes formas de estimação de sua variância:

- Dois tratamentos não comuns, mas de um mesmo ensaio

$$\hat{V}(t_i - t_{i'}) = \frac{2 * QME}{J} \quad [5]$$

em que J é o número de repetições e QME o quadrado médio do erro da análise conjunta.

- Dois tratamentos não comuns e de ensaios diferentes

$$\hat{V}(t_i - t_{i'}) = \frac{2 * QME}{J}\left(1 + \frac{1}{c}\right) \quad [6]$$

em que c representa o número de tratamentos comuns entre os ensaios.

- Tratamento não comum e tratamento comum

$$\hat{V}(t_i - t_{i'}) = \frac{2 * QME}{J}\left(1 + \frac{1}{c} + \frac{1}{K} + \frac{1}{cK}\right) \quad [7]$$

em que K é o número de ensaios.

- Entre dois tratamentos comuns

$$\hat{V}(t_i - t_{i'}) = \frac{2 * QME}{JK} \quad [8]$$

Cabe recordar que este último caso só ocorre quando existe mais de um tratamento comum. Nas comparações entre tratamentos não comuns de ensaios diferentes e entre um tratamento não comum e um tratamento comum o valor de c é igual a 1.

Exemplo de aplicação: blocos completos ao acaso com um tratamento comum

O exemplo que será discutido a seguir é composto de dois experimentos em blocos completos ao acaso, nos quais foi estudada a patogenicidade de fungos ao percevejo do grão de arroz, em 1988, sob responsabilidade das pesquisadoras Eliana Dias Quintela e Maria Goretti Lima. Nesses experimentos, foram estudados onze fungos no que se refere à sua patogenicidade. Cada experimento continha cinco repetições (blocos). A testemunha sem infecção (1) era o tratamento comum entre os experimentos. Os dados se referem à proporção de percevejos infectados e transformados por arco seno da raiz quadrada da proporção (Tabelas 13 e 14).

Tabela 13. Dados de proporção de insetos infectados, transformados por arco seno da raiz quadrada – ensaio 1.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Bloco 5	Total
1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	1,0799	1,0799	0,7854	0,3063	0,9553	4,2068
3	0,4909	0,4909	0,7297	0,4637	0,3614	2,5365
4	0,6847	0,6847	0,4909	0,5236	0,6155	2,9994
5	0,6155	0,6155	0,5495	1,1072	0,7137	3,6013
6	0,7854	0,7854	0,6847	0,9553	0,3876	3,5984
7	1,0799	1,0799	0,8691	0,8411	0,5796	4,4497
Total	4,7363	4,7363	4,1093	4,1971	3,6131	21,3921

Tabela 14. Dados de proporção de insetos infectados, transformados por arco seno da raiz quadrada – ensaio 2.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Bloco 5	Total
1	0,0000	0,0000	0,3218	0,0000	0,0000	0,3218
9	0,8411	0,9117	0,5796	1,5708	0,8571	4,7603
10	0,4909	0,7297	0,5236	0,5236	0,5236	2,7914
11	0,6155	0,5796	0,7854	0,9117	0,8571	3,7493
12	0,5796	1,1072	0,7854	0,5236	0,4909	3,4867
13	0,7297	0,6155	0,9553	0,7297	0,9117	3,9420
Total	3,2568	3,9437	3,9511	4,2595	3,6404	19,0515

Assim como foi feito na seção *Análise de experimentos com todos os tratamentos em comum*, as análises individuais (Tabelas 15 e 16) se fizeram segundo os processos estabelecidos no Capítulo 2, por isso não se repetem aqui. A análise de variância conjunta para esses dois ensaios encontra-se na Tabela 17, na qual as somas de quadrados para ensaios e total foram obtidas como mostrado no caso das análises conjuntas com todos os tratamentos comuns aos ensaios. A soma de quadrados para tratamentos nesse caso, por se ter apenas um tratamento em comum, também é obtida efetuando-se a soma das somas de quadrados de tratamentos dos ensaios individuais.

Tabela 15. Análise de variância para os dados do ensaio 1, sobre patogenicidade de fungos ao percevejo do grão de arroz.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrado (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	4	0,1081	0,0270	
Tratamentos	6	2,4158	0,4026	9,12 (> 0,0001)
Erro	24	1,0591	0,0441	
Total	34	3,5830		

Tabela 16. Análise de variância para os dados do ensaio 2, sobre patogenicidade de fungos ao percevejo do grão de arroz.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrado (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	4	0,0958	0,0239	
Tratamentos	5	2,3633	0,4726	9,48 (< 0,0001)
Erro	20	0,9973	0,0499	
Total	29	3,4564		

A grande diferença desta análise com relação àquelas em que todos os tratamentos são comuns aos ensaios é, como já afirmado anteriormente, a necessidade de que as médias dos tratamentos não comuns sejam ajustadas para correção do efeito da sua ausência dos outros ensaios. Esse ajuste se faz a partir das médias do tratamento comum que, por estar em todos os ensaios, reflete em sua média geral esse fato. Assim, o fator de correção para cada ensaio, para os tratamentos não comuns, é a diferença entre a média do tratamento

comum naquele ensaio e sua média em todos os ensaios. É evidente que a soma de todos os fatores de correção será sempre igual a 0 (zero). A Tabela 18 mostra como se procede a esse ajuste de médias.

Tabela 17. Análise de variância conjunta para os dois ensaios sobre patogenicidade de fungos ao percevejo do grão de arroz.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrado (SQ)	Quadrado médio (QM)	F (Pr > F)
Ensaios	1	0,0136	0,0136	
Blocos (ensaios)	8	0,2039	0,0255	
Tratamentos	11	4,7790	0,4345	9,30 < ,0001
Erro	44	2,0565	0,0467	
Total	64	7,0530		

Tabela 18. Tabela de ajuste de médias para os tratamentos não comuns.

Tratamento	Média no ensaio 1 (M_{1i})	Média no ensaio 2 (M_{2i})	Fator de correção (FC)	Média ajustada (M_{ki} - FC)
1	0,0000	0,0644	Média aritmética	0,0322
2	0,8026		0,0000 - 0,0322 = -0,0322	0,8348
3	0,6105		0,0000 - 0,0322 = -0,0322	0,6427
4	0,5860		0,0000 - 0,0322 = -0,0322	0,6182
5	0,7131		0,0000 - 0,0322 = -0,0322	0,7453
6	0,6785		0,0000 - 0,0322 = -0,0322	0,7107
7	0,8512		0,0000 - 0,0322 = -0,0322	0,8834
9		0,9521	0,0644 - 0,0322 = 0,0322	0,9199
10		0,5583	0,0644 - 0,0322 = 0,0322	0,5261
11		0,7499	0,0644 - 0,0322 = 0,0322	0,7177
12		0,6973	0,0644 - 0,0322 = 0,0322	0,6651
13		0,7884	0,0644 - 0,0322 = 0,0322	0,7562

A comparação das médias se faz com o auxílio das fórmulas dadas anteriormente para obtenção das variâncias dos contrastes entre duas médias. No exemplo atual, tem-se:

a) Duas médias de tratamentos não comuns de um mesmo ensaio:

$$\hat{V}(t_{\hat{i}} - t_{i'}) = \frac{2 * QME}{J} = \frac{2 * 0{,}0467}{5} = 0{,}01868$$

b) Duas médias de tratamentos não comuns de ensaios distintos:

$$\hat{V}(t_{\hat{i}} - t_{i'}) = \frac{2 * QME}{J}\left(1 + \frac{1}{c}\right) = \frac{2 * 0{,}0467}{5}\left(1 + \frac{1}{1}\right) = 0{,}03736$$

c) Tratamentos não comuns contra o tratamento comum:

$$\hat{V}(t_{\hat{i}} - t_{i'}) = \frac{2 * QME}{J}\left(1 + \frac{1}{c} + \frac{1}{K} + \frac{1}{cK}\right) = \frac{2 * 0{,}0467}{5}\left(1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{1 * 2}\right) = 0{,}05604$$

Utilizando-se o teste t para essas comparações para o caso *a*, tem-se que

$$DMS = t_{(44;0,05)} * \sqrt{(0{,}01868)} = 2{,}02 * 0{,}1367 = 0{,}2761$$

Portanto quaisquer duas médias não comuns de cada ensaio que difira por um valor maior que esse devem ser consideradas diferentes. No ensaio 1, não se verificam diferenças significativas entre os tratamentos 2 ao 7, e no ensaio 2 o tratamento 10 será considerado diferente do tratamento 9. Fica a cargo do leitor verificar se todos os tratamentos não comuns devem ser considerados diferentes do tratamento comum.

Uso do SAS na análise conjunta de blocos completos ao acaso com um tratamento em comum

Neste caso, como existe a necessidade de ajuste das médias dos tratamentos não comuns, o procedimento SAS a ser usado é o GLM, com o comando LSMEANS. Dessa forma, o programa é o seguinte:

```
PROC GLM;CLASS ENSAIO BLOCO TRAT;
MODEL PERC=ENSAIO BLOCO(ENSAIO) TRAT/SS1;
LSMEANS TRAT;
```

O resultado se vê na Figura 3.

```
                          The GLM Procedure
                       Class Level Information
                Class         Levels    Values
                ENSAIO             2    1 2
                BLOCO              5    1 2 3 4 5
                TRAT              12    1 2 3 4 5 6 7 9 10 11 12 13

                      Number of observations    65
                          The GLM Procedure
Dependent Variable: PERC
                                 Sum of
 Source                  DF      Squares    Mean Square   F Value   Pr > F
 Model                   20   4.99656011     0.24982801      5.35   <.0001
 Error                   44   2.05650132     0.04673867
 Corrected Total         64   7.05306143

               R-Square     Coeff Var      Root MSE     PERC Mean
                0.708424      34.90308      0.216191      0.619405

 Source                  DF     Type I SS    Mean Square   F Value   Pr > F
 ENSAIO                   1   0.01363182     0.01363182      0.29   0.5919
 BLOCO(ENSAIO)            8   0.20387198     0.02548400      0.55   0.8159
 TRAT                    11   4.77905631     0.43445966      9.30   <.0001

                          The GLM Procedure
                        Least Squares Means
                        TRAT      PERC LSMEAN
                        1          0.03218000
                        2          0.83476000
                        3          0.64272000
                        4          0.61822000
                        5          0.74526000
                        6          0.71068000
                        7          0.88342000
                        9          0.91988000
                        10         0.52610000
                        11         0.71768000
                        12         0.66516000
                        13         0.75620000
```

Figura 3. Análise de variância de experimentos em blocos completos casualizados com um tratamento comum e uso do SAS.

Exemplo de aplicação: blocos completos ao acaso, com mais de um tratamento comum

Para exemplificar a análise conjunta de experimentos em blocos ao acaso com alguns tratamentos comuns, utilizaram-se dois experimentos de competição de cultivares de feijão do grupo preto, realizados em 1999–2000, em dois municípios do Estado de Goiás: um em Senador Canedo e outro em Santo Antônio de Goiás. Em ambos os locais se mantiveram as três cultivares testemunhas como tratamentos comuns. Das 12 linhagens testadas, as seis primeiras foram eliminadas do ensaio de Santo Antônio de Goiás e as seis últimas do ensaio de Senador Canedo. As Tabelas 19 e 20 mostram os dados de produtividade de grãos, em kg ha^{-1}, desses dois experimentos com os tratamentos selecionados em cada um. Cada experimento foi conduzido no período de inverno, sob irrigação por aspersão.

Tabela 19. Dados de produtividade de grãos, em kg ha^{-1}, para o ensaio de Senador Canedo, GO.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
CNFP 8015	3.700,00	3.950,00	4.050,00	2.950,00	14.650,00
CNFP 8016	3.600,00	3.750,00	3.550,00	3.000,00	13.900,00
CNFP 8017	3.925,00	2.800,00	2.825,00	3.600,00	13.150,00
CNFP 8018	3.875,00	3.825,00	4.000,00	3.825,00	15.525,00
CNFP 8019	4.200,00	3.250,00	3.950,00	3.375,00	14.775,00
CNFP 8020	4.850,00	3.200,00	3.450,00	4.200,00	15.700,00
Diamante negro	3.300,00	2.500,00	4.025,00	3.950,00	13.775,00
FT Nobre	4.750,00	4.950,00	4.700,00	3.575,00	17.975,00
G. Brilhante	4.025,00	3.575,00	3.125,00	3.825,00	14.550,00
Total	36.225,00	31.800,00	33.675,00	32.300,00	134.000,00

Tabela 20. Dados de produtividade de grãos, em kg ha^{-1}, para o ensaio de Santo Antônio de Goiás, GO.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Bloco 4	Total
CNFP 8021	4.075,00	2.252,78	3.388,89	2.663,89	12.380,56
CNFP 8022	2.544,44	3.133,33	2.866,67	2.533,33	11.077,77
CNFP 8023	2.763,89	2.250,00	1.941,67	2.791,67	9.747,23
CNFP 8024	3.036,11	2.369,44	3.033,33	2.922,22	11.361,10
CNFP 8025	2.591,67	1.466,67	2.813,89	2.750,00	9.622,23
CNFP 8026	2.652,78	2.927,78	2.838,89	2.305,56	10.725,01
Diamante negro	1.913,89	1.850,00	1.988,89	2.080,56	7.833,34
FT Nobre	3.713,89	3.158,33	3.250,00	2.877,78	13.000,00
G. Brilhante	2.400,00	1.827,78	3.316,67	2.669,44	10.213,89
Total	25.691,67	21.236,11	25.438,90	23.594,45	95.961,13

Também aqui as análises individuais (Tabelas 21 e 22) se processaram como mostrado no Capítulo 3.

Tabela 21. Análise de variância para os dados de produtividade de grãos, em kg ha^{-1}, no ensaio de Senador Canedo, GO.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	3	1.309.583,33	436.527,78	
Ntrat	8	4.003.472,22	500.434,03	1,95 (0,0980)
Erro	24	6.151.666,67	256.319,44	
Total	35	11.464.722,22		

Tabela 22. Análise de variância para os dados de produtividade de grãos, em kg ha^{-1}, no ensaio de Santo Antônio de Goiás, GO.

Fonte de variaço	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	3	1.415.040,18	471.680,06	
Ntrat	8	4.801.328,23	600.166,03	3,16 (0,0137)
Erro	24	4.559.091,00	189.962,13	
Total	35	10.775.459,41		

Na análise conjunta apresentada na Tabela 23, as somas de quadrados de ensaios, blocos (ensaios), erro e total foram obtidas como estudado nos exemplos anteriores. As diferenças aqui surgem na forma de obtenção das somas de quadrados de tratamentos e da interação ensaios x tratamentos comuns. Para o cálculo desta última soma de quadrados, é necessário construir-se o quadro auxiliar com os totais dos tratamentos comuns em cada um dos ensaios e o total geral desses tratamentos em todos os ensaios, além do total para ensaios medido apenas pelos tratamentos comuns, como se vê na Tabela 24.

Tabela 23. Análise de variância conjunta para os ensaios de Senador Canedo e Santo Antônio de Goiás.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Ensaios	1	20.096.605,98	20.096.605,98	
Blocos(ensaios)	6	2.724.623,51	454.103,92	
Ntrat	14	8.641.450,34	617.246,45	2,77 (0,0044)
Ensaios x Ntrat comuns	2	163.350,12	81.675,06	0,37 (0,6954)
Erro	48	10.710.757,67	223.140,78	
Total	71	42.336.787,62		

Tabela 24. Quadro auxiliar com os valores totais dos tratamentos comuns nos dois ensaios.

Tratamento comum	Ensaio 1	Ensaio 2	Soma
Diamante negro	13.775,00	7.833,34	21.608,34
FT Nobre	17.975,00	13.000,00	30.975,00
G. Brilhante	14.550,00	10.213,89	24.763,89
Soma	46.300,00	31.047,23	77.347,23

A fórmula para a interação ensaios x tratamentos comuns é dada por

$$SQ(ExTc) = \frac{\sum Tc_{ik}^2}{J} - \frac{\sum Tc_i^2}{KJ} - \frac{\sum Ec_k^2}{cJ} + \frac{G^2}{cJK} \quad [9]$$

em que J é o número de repetições, K o número de ensaios, c o número de tratamentos comuns, Tc_{ik} é o total do tratamento comum i no ensaio k, Tc_i é o total de tratamento comum i em todos os ensaios, Ec_k é o total dos tratamentos comuns no ensaio k e G é o total geral do quadro auxiliar. No exemplo atual, esta soma de quadrados é:

$$SQ(ExTc) = \frac{13.775^2 + 7.833{,}34^2 + ... + 10.213{,}23^2}{4} -$$

$$\frac{21.608{,}34^2 + 30.975^2 + 24.763{,}89^2}{2*4} - \frac{46.300^2 + 31.047{,}23^2}{3*4} +$$

$$\frac{77.347{,}23^2}{3*2*4} = 163.351{,}34$$

A soma de quadrados de tratamentos ajustada para ensaios é igual à diferença entre a soma de quadrados de tratamentos obtida em cada ensaio e a soma de quadrados da interação calculada anteriormente. Assim,

$$SQTrat = 4.003.472{,}22 + 4.801.346{,}88 - 163.351{,}34 = 8.641.467{,}76$$

Finalmente, o ajuste das médias de tratamentos não comuns se faz outra vez restando a cada uma um fator de correção típico de cada ensaio, dado pela diferença entre as médias dos tratamentos comuns no ensaio e a média dos tratamentos comuns em todos os ensaios como se vê na Tabela 25.

Para a comparação de médias de tratamentos, empregam-se as mesmas fórmulas mostradas ao início deste capítulo.

Tabela 25. Tabela para obtenção das médias de tratamentos não comuns ajustadas.

Tratamento	Média no ensaio 1 (M)	Média no ensaio 2 (M)	Fator de correção (FC)	Média ajustada (M – FC)
CNFP 8015	3.662,50		3.858,33 - 3.222,80 = 635,53	3.026,97
CNFP 8016	3.475,00		3.858,33 - 3.222,80 = 635,53	2.839,47
CNFP 8017	3.287,50		3.858,33 - 3.222,80 = 635,53	2.651,97
CNFP 8018	3.881,25		3.858,33 - 3.222,80 = 635,53	3.245,77
CNFP 8019	3.693,75		3.858,33 - 3.222,80 = 635,53	3.058,27
CNFP 8020	3.925,00		3.858,33 - 3.222,80 = 635,53	3.289,47
CNFP 8021		3.095,14	2.587,27 - 3.222,80 = -635,53	3.730,67
CNFP 8022		2.769,44	2.587,27 - 3.222,80 = -635,53	3.404,97
CNFP 8023		2.436,81	2.587,27 - 3.222,80 = -635,53	3.072,34
CNFP 8024		2.840,28	2.587,27 - 3.222,80 = -635,53	3.475,81
CNFP 8025		2.405,56	2.587,27 - 3.222,80 = -635,53	3.041,09
CNFP 8026		2.681,25	2.587,27 - 3.222,80 = -635,53	3.316,78
Diamante negro	3.443,75	1.958,33		2.701,04
FT Nobre	4.493,75	3.250,00		3.871,88
G. Brilhante	3.637,50	2.553,47		3.095,48
Média das comuns	3.858,33	2.587,27		3.222,80

Uso do SAS na análise conjunta de blocos completos ao acaso com mais de um tratamento em comum

Para o caso de mais de um tratamento comum entre os ensaios, o SAS requer o uso do procedimento GLM com o a opção SS1 do comando MODEL sem o comando para a estimação de médias, como segue:

```
PROC GLM;CLASS ENSAIO BLOCO TRAT;
MODEL PROD=ENSAIO BLOCO(ENSAIO) TRAT ENSAIO*TRAT/SS1;
```

Para a obtenção das médias ajustadas dos tratamentos não comuns, existe a necessidade de uma segunda execução do SAS, já que o programa anterior com o uso do comando LSMEANS forneceria apenas as médias aritméticas dos tratamentos comuns e informaria que as médias dos tratamentos não comuns são não estimáveis. A solução para que se obtenham as médias ajustadas é a mesma mostrada para o delineamento de blocos aumentados de Federer.

O programa completo então é o seguinte:

a) Para a entrada de dados:

```
INPUT ENSAIO NTRAT $ 5-15 BLOCO PROD TRAT;
IF TRAT GE 13 THEN TC_TNC=1;ELSE TC_TNC=0;
TNC=TRAT*(TC_TNC=0);TC=TRAT*(TC_TNC=1);
```

NTRAT e TRAT são respectivamente os valores dos tratamentos em formato alfanumérico e numérico, ENSAIO e BLOCO são as identificações numéricas dos ensaios e dos blocos e PROD é a variável cujos valores correspondem à produção de grãos em kg ha^{-1}. Os tratamentos comuns têm numeração 13 ou maior para esse caso.

b) Para a análise de variância

A análise de variância cujo resultado se vê na Figura 4, que foi obtida com o seguinte programa:

```
PROC GLM;CLASS ENSAIO BLOCO NTRAT;
MODEL PROD=ENSAIO BLOCO(ENSAIO) NTRAT ENSAIO*NTRAT/SS1;
PROC GLM NOPRINT;CLASS ENSAIO BLOCO TC_TNC TNC TC;
MODEL PROD=ENSAIO BLOCO(ENSAIO)TC_TNC TNC(TC_TNC) TC(TC_TNC);
LSMEANS TC_TNC TNC(TC_TNC) TC(TC_TNC)/NOPRINT OUT=MED1;
DATA MED1;SET MED1;
IF TC_TNC=0 AND TNC=. AND TC=. THEN delete;
IF TC_TNC=1 AND TNC=. AND TC=. THEN delete;
IF TC_TNC=1 AND TNC=0 AND TC=. THEN delete;
IF TC_TNC=0 AND TNC=. AND TC=0 THEN delete;
IF TC_TNC=1 AND TNC=. THEN TNC=TC;TRAT=TNC;
MEDIA_AJ=LSMEAN; KEEP TRAT MEDIA_AJ;PROC SORT;BY TRAT;
DATA B;SET A;IF ENSAIO=1 AND BLOCO=1 AND TRAT LE 6;KEEP TRAT NTRAT;
DATA C;SET A;IF ENSAIO=2 AND BLOCO=1;KEEP TRAT NTRAT;
DATA BC;SET B C;PROC SORT;BY TRAT;
DATA FIM;MERGE BC MED1;BY TRAT;
PROC PRINT;VAR NTRAT MEDIA_AJ;
```

O primeiro procedimento GLM fornece a análise de variância, e o segundo, com todos os demais passos de programação, serve apenas para a estimação das médias ajustadas.

```
                          The GLM Procedure
                       Class Level Information
Class       Levels  Values
ENSAIO           2  1 2
BLOCO            4  1 2 3 4
NTRAT           15  CNFP 8015 CNFP 8016 CNFP 8017 CNFP 8018 CNFP 8019
                    CNFP 8020 CNFP 8021 CNFP 8022 CNFP 8023 CNFP 8024
                    CNFP 8025 CNFP 8026 D.NEGRO FT.NOBRE G.BRILHANTE
                         Number of observations    72
                          The GLM Procedure
Dependent Variable: PROD
                                      Sum of
 Source            DF       Squares    Mean Square   F Value   Pr > F
 Model             23    31626029.95     1375044.78      6.16   <.0001
 Error             48    10710757.67      223140.78
 Corrected Total   71    42336787.62
               R-Square     Coeff Var      Root MSE     PROD Mean
               0.747011      14.78998      472.3778      3193.905

 Source            DF     Type I SS    Mean Square   F Value   Pr > F
 ENSAIO             1    20096605.98    20096605.98     90.06   <.0001
 BLOCO(ENSAIO)      6     2724623.51      454103.92      2.04   0.0790
 NTRAT             14     8641450.34      617246.45      2.77   0.0044
 ENSAIO*NTRAT       2      163350.12       81675.06      0.37   0.6954
```

Obs	NTRAT	MEDIA_AJ
1	CNFP 8015	3026.97
2	CNFP 8016	2839.47
3	CNFP 8017	2651.97
4	CNFP 8018	3245.72
5	CNFP 8019	3058.22
6	CNFP 8020	3289.47
7	CNFP 8021	3730.67
8	CNFP 8022	3404.97
9	CNFP 8023	3072.34
10	CNFP 8024	3475.81
11	CNFP 8025	3041.09
12	CNFP 8026	3316.78
13	D.NEGRO	2701.04
14	FT.NOBRE	3871.88
15	G.BRILHANTE	3095.49

Figura 4. Análise de variância conjunta de experimentos em blocos completos ao acaso, com alguns tratamentos comuns usando o SAS.

Exemplo de aplicação: quadrado latino com um tratamento comum

No ano agrícola, o pesquisador Evane Ferreira, da Embrapa Arroz e Feijão, testou 26 linhagens/cultivares para resistência a insetos e os avaliou em cinco quadrados latinos 6 x 6, em que uma dos cultivares (26 = Bonança) era comum a todos os ensaios. Desses experimentos, apresentam-se as Tabelas 26 e 27 que contêm os dados de produtividade de espiguetas em gramas, para dois deles. Os demais materiais eram: 1 = Canastra, 2 =

Tabela 26. Dados de produtividade de espiguetas, em gramas, para o ensaio 1. Para cada coluna, o primeiro número indica o tratamento e o segundo a produtividade de espiguetas.

Linha	Coluna1		Coluna 2		Coluna 3		Coluna 4		Coluna 5		Coluna 6		Total	Tratamento	Total de tratamentos
1	1	55	3	33	26	77	2	20	4	64	5	52	301	1	348
2	26	52	2	19	5	37	1	64	3	44	4	39	255	2	211
3	4	21	26	55	3	49	5	53	1	48	2	25	251	3	283
4	5	30	1	40	4	21	26	55	2	34	3	50	230	4	201
5	3	77	5	21	2	75	4	12	26	62	1	71	318	5	256
6	2	38	4	44	1	70	3	30	5	63	26	17	262	26	318
Total		273		212		329		234		315		254	1.617		1.617

Tabela 27. Dados de produtividade de espiguetas, em gramas, para o ensaio 2. Para cada coluna o primeiro número indica o tratamento e o segundo a produtividade de espiguetas.

Linha	Coluna 1		Coluna 2		Coluna 3		Coluna 4		Coluna 5		Coluna 6		Total	Tratamento	Total de tratamentos
1	26	79	10	44	8	67	9	20	7	83	6	30	323	6	183
2	7	53	6	71	10	67	26	98	9	19	8	50	358	7	329
3	6	27	26	88	9	39	10	55	8	66	7	67	342	8	319
4	10	30	9	66	7	58	8	49	6	12	26	36	251	9	218
5	9	31	8	48	6	10	7	10	26	62	10	52	213	10	327
6	8	39	7	58	26	90	6	33	10	79	9	43	342	26	453
Total		259		375		331		265		321		278	1.829		1.829

CNAs8818, 3 = CNA 8540, 4 = CNAs8989, 5 = CNAs8957, 6 = CNAs8960, 7 = CNAs8952, 8 = CNAs8817, 9 = CNA 8711 e 10 = CNAs8938. Os tratamentos de 1 a 5 e o tratamento comum compunham o ensaio 1, enquanto os de número 6 a 10, mais o tratamento comum, o ensaio 2.

A análise de variância de cada um desses ensaios foi realizada como mostrado no Capítulo 4 e se encontram nas Tabelas 28 e 29. Os fatores de correção dos dois ensaios foram respectivamente iguais a 72.630,25 e 92.923,36.

Tabela 28. Análise de variância para o ensaio 1, dados mostrados na Tabela 26.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Coluna	5	1.738,25	347,65	
Linha	5	918,92	183,78	
Tratamento	5	2.832,25	566,45	1,67 (0,18)
Erro	20	6.799,33	339,97	
Total	35	12.288,75		

Tabela 29. Análise de variância para o ensaio 2, dados mostrados na Tabela 27.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Coluna	5	1.712,81	342,56	
Linha	5	2.875,14	575,03	
Tratamento	5	7.602,14	1.520,43	4,31 (0,0080)
Erro	20	7.047,56	352,38	
Total	35	19.237,65		

A análise de variância conjunta que se mostra na Tabela 30 foi realizada segundo as indicações dadas ao início deste capítulo.

Tabela 30. Análise de variância conjunta para os dois ensaios sobre resistência de arroz ao ataque de insetos.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Ensaio	1	624,22	624,22	
Coluna(Ensaio)	10	3.451,06	345,11	
Linha(Ensaio)	10	3.794,06	379,41	
Tratamento	10	10.434,39	1.043,44	3,01 (0,0063)
Erro	40	13.846,89	346,17	
Total	71	32.150,62		

$$FCc = \frac{(1.617 + 1.829)^2}{36 + 36} = 164.929,39$$

$$SQTotal = SQTotal_1 + SQTotal_2 + FC1 + FC2 - FCc = 12.288,75 + 19.237,65 + 72.630,25 + 92.923,36 - 164.929,39 = 32.150,62$$

$$SQEnsaio = FC1 + FC2 - FCc = 72.630,25 + 92.923,36 - 164.929,39 = 624,22$$

$$SQCol(Ensaio) = SQCol_1 + SQCol_2 = 1.738,25 + 1.712,81 = 3.451,06$$

$$SQLin(Ensaio) = SQLin_1 + SQLin_2 = 918,92 + 2.875,14 = 3.794,06$$

$$SQErro = SQErro_1 + SQErro_2 = 6.799,33 + 7.047,56 = 13.846,89$$

A soma de quadrados de tratamentos ajustada se obtém, como afirmado anteriormente, por diferença.

$$SQTrat = SQTotal - SQEnsaio - SQCol(Ensaio) - SQLin(Ensaio) - SQErro = 10.434,39$$

ou

$$SQTrat = SQTrat_1 + SQTrat_2 = 2.832,25 + 7.602,14 = 10.434,39$$

As médias de tratamentos se obtêm subtraindo de cada média não comum um fator de correção, calculado da mesma forma que se calculou o caso em blocos completos, enquanto a média do tratamento comum é a média aritmética dos dois ensaios.

A Tabela 31 resume tudo isso.

Tabela 31. Ajuste de médias para os dados dos dois ensaios de arroz.

Tratamento	Média no ensaio 1 (M)	Média no ensaio 2 (M)	Fator de correção (FC)	Média ajustada (M - FC)
1	58,00		53,00 - 64,25 = -11,25	69,25
2	35,17		53,00 - 64,25 = -11,25	46,42
3	47,17		53,00 - 64,25 = -11,25	58,47
4	33,50		53,00 - 64,25 = -11,25	44,75
5	42,67		53,00 - 64,25 = -11,25	53,92
6		30,50	75,50 - 64,25 = 11,25	19,25
7		54,83	75,50 - 64,25 = 11,25	43,58
8		53,17	75,50 - 64,25 = 11,25	41,92
9		36,33	75,50 - 64,25 = 11,25	25,08
10		54,50	75,50 - 64,25 = 11,25	43,25
26	53,00	75,50	Média aritmética	64,25

Uso do SAS na análise conjunta de ensaio em quadrado latino com um tratamento em comum

A exemplo do que foi feito para os delineamentos inteiramente casualizados e blocos completos ao acaso, o programa SAS, usando o procedimento GLM, com a opção SS1 no comando MODEL é o seguinte:

```
PROC GLM;CLASS ENSAIO COLUNA LINHA TRAT;
MODEL PROD=ENSAIO COLUNA(ENSAIO) LINHA(ENSAIO) TRAT/SS1;
LSMEANS TRAT;
```

O resultado apresenta-se na Figura 5.

```
                    The GLM Procedure
                  Class Level Information
Class         Levels    Values
ENSAIO             2    1 2
COLUNA             6    1 2 3 4 5 6
LINHA              6    1 2 3 4 5 6
TRAT              11    1 2 3 4 5 6 7 8 9 10 26

                Number of observations    72
                    The GLM Procedure

Dependent Variable: PROD

                              Sum of
Source                DF      Squares     Mean Square    F Value    Pr > F
Model                 31  18303.72222       590.44265       1.71    0.0560
Error                 40  13846.88889       346.17222
Corrected Total       71  32150.61111

                R-Square     Coeff Var      Root MSE     PROD Mean
                0.569312      38.87437      18.60570      47.86111

Source                DF      Type I SS    Mean Square    F Value    Pr > F
ENSAIO                 1     624.22222      624.22222       1.80    0.1869
COLUNA(ENSAIO)        10    3451.05556      345.10556       1.00    0.4623
LINHA(ENSAIO)         10    3794.05556      379.40556       1.10    0.3884
TRAT                  10   10434.38889     1043.43889       3.01    0.0063
                    The GLM Procedure
                  Least Squares Means
                  TRAT      PROD LSMEAN
                  1          69.2500000
                  2          46.4166667
                  3          58.4166667
                  4          44.7500000
                  5          53.9166667
                  6          19.2500000
                  7          43.5833333
                  8          41.9166667
                  9          25.0833333
                  10         43.2500000
                  26         64.2500000
```

Figura 5. Análise conjunta de ensaios em delineamento quadrado latino, com um tratamento comum, usando o SAS.

Análise conjunta de experimentos, com informação parcial ou com delineamentos diferentes

Como se viu neste capítulo, a análise conjunta de experimentos, quando se trata de grupos com o mesmo delineamento, com todos os tratamentos em comum ou só um ou uns poucos em comum e sem dados perdidos, não apresenta qualquer dificuldade de execução. Os problemas começam a surgir quando aparecem dados perdidos, quando os diversos experimentos têm delineamentos diferentes ou o número de tratamentos comuns não é constante em todos os ensaios. Uma solução rápida é trabalhar com as médias dos

tratamentos de cada um dos ensaios e, com essas médias, calcular os efeitos devidos a ensaios, tratamentos e a interação ensaios x tratamentos comuns e utilizar uma média ponderada dos erros dos diversos experimentos como uma estimativa do erro da análise conjunta. Evidentemente que, quanto menor for a divergência de informação entre os distintos ensaios, mais precisa será essa análise. Um exemplo é apresentado a seguir para mostrar o procedimento.

Exemplo de aplicação – dados completos dos ensaios não estão disponíveis

Alfredo do Nascimento Júnior e Valter Martins de Almeida (1996), pesquisadores da Empaer-MT, apresentaram na *8ª Reunião da comissão técnica regional de feijão – Região II*, em 1994, os resultados dos ensaios de competição de cultivares e linhagens de feijão no Estado de Mato Grosso para quatro grupos comerciais dessa cultura, em cultivos irrigado e de sequeiro. No relatório da comissão, aparecem apenas as médias dos cultivares, o coeficiente de variação das análises conjuntas por cultivo. Informa-se também que o delineamento experimental adotado foi o de blocos completos com quatro repetições e que os ensaios foram conduzidos por um período de 3 anos (1992, 1993 e 1994), num total de 11 experimentos nas seguintes situações: Nova Mutum (irrigado durante os 3 anos), Campo Novo dos Parecis (irrigado e sequeiro em 1993 e 1994), São José dos Quatro Marcos (sequeiro em 1993 e 1994) e Tangará da Serra (sequeiro em 1993 e irrigado em 1994). Para o grupo preto, os resultados se mostram na Tabela 32. (NASCIMENTO JÚNIOR; ALMEIDA, 1996)

Conforme se viu, foram realizados seis ensaios irrigados e cinco de sequeiro, e as médias gerais desses ensaios, calculadas a partir da Tabela 32, foram iguais a 2.419,00 e 849,44, respectivamente. Utilizando-se a fórmula do coeficiente de variação, é possível calcular os quadrados médios respectivos, que são

$$QME = \left(\frac{m * CV}{100} \right)^2 = \left(\frac{2.419,00 * 8,07}{100} \right)^2 = 38.108,2325$$

e

$$QME = \left(\frac{m * CV}{100} \right)^2 = \left(\frac{849,44 * 17,46}{100} \right)^2 = 21.996,5158$$

Tabela 32. Produtividades médias, em kg ha^{-1}, coeficientes de variação dos grupos de ensaios irrigados e de sequeiro e média geral.

Genótipo	Média nos ensaios irrigados	Média nos ensaios de sequeiro	Média geral
AN 911120	2.645,67	1.064,00	1.911,88
Diamante Negro	2.619,00	1.031,33	1.875,21
AN 910390	2.583,33	982,67	1.854,54
AN 910375	2.583,33	980,00	1.840,45
AN 910960	2.463,33	949,00	1.734,54
AN 910021	2.462,67	860,00	1.729,33
AN 910970	2.369,33	849,33	1.646,15
AN 910902	2.358,33	778,33	1.638,94
AN 910961	2.337,33	775,67	1.612,18
AN 910891	2.317,67	742,00	1.587,37
Rio Tibagi	2.250,33	711,00	1.541,09
FE 821681	2.240,00	690,00	1.507,88
FE 732007	2.216,67	629,33	1.595,20
CV (%)	8,07	17,46	

Esses quadrados médios não são muito diferentes; portanto, permitem fazer uma análise de variância conjunta. Lembrando que cada ensaio mostrava 36 graus de liberdade para o erro, tem-se que os ensaios irrigados possuíam um total de 216 GL e os de sequeiro 180 GL para os erros nessas análises, logo a análise conjunta tem um total de 396 GL para o erro. Dessa forma, o erro da análise conjunta é igual a:

$$SQErro = 38.108,2325*216 + 21.996,5158*180 = 1.219.0751,0640$$

Nos ensaios irrigados, cada genótipo tinha um total de 24 repetições e nos de sequeiro 20 repetições. Multiplicando cada média por seu número de repetições, monta-se a Tabela 33.

Para montar a análise conjunta desses dois grupos de ensaios, aplicam-se as seguintes fórmulas:

$$SQGrupos = \frac{(754.727,76)^2}{312} + \frac{(220.853,20)^2}{260} - \frac{[975.580,96]^2}{572}$$

$$SQGrupos = 349.373.440,0690$$

Tabela 33. Totais de genótipos para os dois grupos de ensaios e o total geral.

Genótipo	Total nos ensaios irrigados	Total nos ensaios de sequeiro	Total geral
AN 911120	63.496,08	21.280,00	84.776,08
Diamante Negro	62.856,00	20.626,60	83.482,60
AN 910390	61.999,92	19.653,40	81.653,32
AN 910375	61.999,92	19.600,00	81.599,92
AN 910960	59.119,92	18.980,00	78.099,92
AN 910021	59.104,08	17.200,00	76.304,08
AN 910970	56.863,92	16.986,60	73.850,52
AN 910902	56.599,92	15.566,60	72.166,52
AN 910961	56.095,92	15.513,40	71.609,32
AN 910891	55.624,08	14.840,00	70.464,08
Rio Tibagi	54.007,92	14.220,00	68.227,92
FE 821681	53.760,00	13.800,00	67.560,00
FE 732007	53.200,08	12.586,60	65.786,68
Total	754.727,76	220.853,20	975.580,96

Dessa mesma forma, podem-se obter as somas de quadrados de genótipos e da interação cultivos x genótipos, como se vê a seguir.

$$SQGenótipos = \frac{\sum T_i^2}{r} - FC = \frac{84.776{,}08^2 + ...65.786{,}68^2}{44} - 1.663.912.953{,}69672$$

$$SQGenótipos = 11.230.844{,}2543$$

A soma de quadrados para a interação grupos x genótipos é dada por:

$$SQGrxGe = \frac{\sum T_{ij}^2}{J_i} - FC - SQGrupos - SQGenótipos$$

$$SQGrxGe = 2.024.637.307{,}5776 - 1.663.912.953{,}6967 - 349.373.440{,}0690 - 11.230.844{,}2543$$

$$SQGrxGe = 120.069{,}5576$$

Finalmente, a análise de variância é apresentada na Tabela 34.

Tabela 34. Análise de variância conjunta de dois grupos de ensaios de feijão.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Grupos	1	349.373.440,0690	349.373.440,0690	11.348,92 (> 0,0001)
Genótipos	12	11.230.844,2543	935.903,6878	30,40 (> 0,0001)
Grupos x genótipos	12	120.069,5576	10.005,7965	0,32 (0,9857)
Erro	396	12.190.751,0640	30.784,7249	

Como se viu, contando apenas com as médias de tratamentos de dois grupos de ensaios e dos coeficientes de variação, e sabendo o delineamento adotado, é perfeitamente possível fazer-se a análise conjunta desses grupos de ensaios. Procedimentos semelhantes podem ser aplicados em outros casos. A soma de quadrados dos erros foi calculada a partir dos valores do coeficiente de variação de cada grupo, que pode muitas vezes não estar disponível. No entanto, se é conhecido o valor de algum teste de médias, tipo Tukey, Duncan, ou mesmo a diferença mínima significativa pelo teste t, estes podem ser usados da mesma forma para cálculo da estimação.

Capítulo 12

Regressão e correlação

Em trabalhos experimentais que envolvam coleta de dados em populações, além da aplicação de estatísticas de posição e dispersão, discutidas de forma resumida no Capítulo 1, pode haver o interesse do pesquisador em medir ou definir a existência de possíveis associações entre as variáveis medidas.

Essas associações, que são definidas pelo uso de modelos matemáticos que estabeleçam a forma da ligação entre as variáveis, podem envolver duas ou mais variáveis. No caso de apenas duas variáveis, é muito comum o uso de modelos polinomiais de forma linear, quadrática, cúbica, ou de ordem maior, ou modelos mais complicados do tipo exponencial, logístico, etc. Quando mais de duas variáveis estão envolvidas, os modelos lineares múltiplos ou de superfície de resposta de ordem 2 ou superior são os indicados.

Em capítulos anteriores, discutiu-se o uso de modelos polinomiais em associação com dados experimentais, ou seja, quando se produziu o aparecimento da informação pertinente. Da mesma forma, os estudos de superfície de resposta com dados provocados serão discutidos no Capítulo 14. O estudo simultâneo de análise de experimentos e regressão para ajuste de dados será feito no Capítulo 13. Neste capítulo, serão estudados apenas os casos de dados coletados de forma amostral, apesar de as técnicas aqui empregadas poderem ser perfeitamente aplicadas a qualquer forma de obtenção de dados.

Com o uso de dados de ensaio planejados, com todas as repetições, em que a variável independente é constituída pelos níveis dos tratamentos, o erro é estimado incluindo nele os efeitos dessas repetições mais o efeito dos graus de liberdade de tratamentos não incluídos no estudo de regressão. Essa parte da soma de quadrados de tratamentos, que não é considerada no estudo de regressão, constitui o que se chama de falta de ajuste. Os valores de cada repetição constituem pontos múltiplos da variável dependente para cada um dos pontos da variável independente, e essa variação nunca poderá ser estimada pelo modelo de regressão. Por exemplo, suponha-se que são estudados cinco níveis de adubação nitrogenada, em blocos completos, com cinco repetições. A análise de variância mostrará que os 24 graus de liberdade se dividem em 4 graus para blocos, 4 para tratamentos e 16 para o erro experimental. O estudo de regressão para modelo quadrático pode ser feito com

o uso de contrastes, como foi visto em capítulos anteriores, e terá apenas dois graus de liberdade. Ao estudar o modelo quadrático em processos de regressão, sem considerar o delineamento experimental, o erro terá 22 graus de liberdade, os 16 do erro experimental, os 4 de blocos e os 2 da falta de ajuste. Alguns procedimentos do SAS permitem isolar essa falta de ajuste do erro, mas nunca separa a parte devida aos blocos. Se forem usadas apenas as médias dos tratamentos, o que se chama de falta de ajuste é que constituirá o erro, nesse exemplo com dois graus de liberdade.

É comum desdobrar o estudo de associação de variáveis em dois componentes: o estudo da correlação, em geral linear e simples, e o de regressão, que procura definir a forma ou modelo da associação. No caso de regressão, discutem-se os modelos de regressão linear simples, de regressão linear múltipla, e alguns outros modelos mais complexos. Adverte-se o leitor que neste livro se fará um estudo inicial de regressão e correlação. Aos interessados em maiores detalhes, sugerem-se os livros *Applied linear regression* (WEISBERG, 1985), *Applied Regression Analysis* (DRAPER; SMITH, 1998) e *Introdução aos modelos de regressão linear e não linear* (SOUZA, 1998), entre tantos outros disponíveis.

Apenas como ilustração, mostram-se a seguir algumas figuras geradas segundo modelos matemáticos adequados para possíveis estudos de regressão. Entre os modelos, estão aqueles ditos polinomiais simples, polinomiais múltiplos, lineares por transformação e não lineares.

Alguns modelos de regressão

Os modelos de regressão aqui apresentados podem ser divididos em três categorias: modelos lineares, modelos lineares por transformação e modelos não lineares. Estes últimos são analisados por meio de processos matemáticos iterativos, como o NLIN, disponível no SAS. Para todos os três casos, pode-se ter uma ou mais de uma variável independente.

Modelos lineares – uma variável independente

Incluem-se aqui os modelos chamados de polinomiais, que podem ser de primeiro grau, segundo grau, etc.

O linear ou primeiro grau (Figura 1) é dado por

$$Y_i = b_0 + b_1 X_i \qquad [1]$$

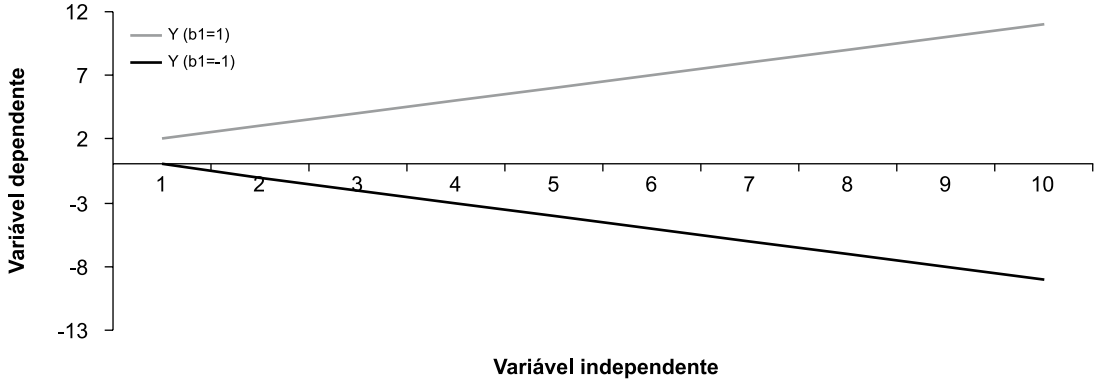


Figura 1. Exemplos de modelos de regressão polinomial de primeiro grau.

o quadrático ou de segundo grau (Figura 2) com a forma

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 \qquad [2]$$

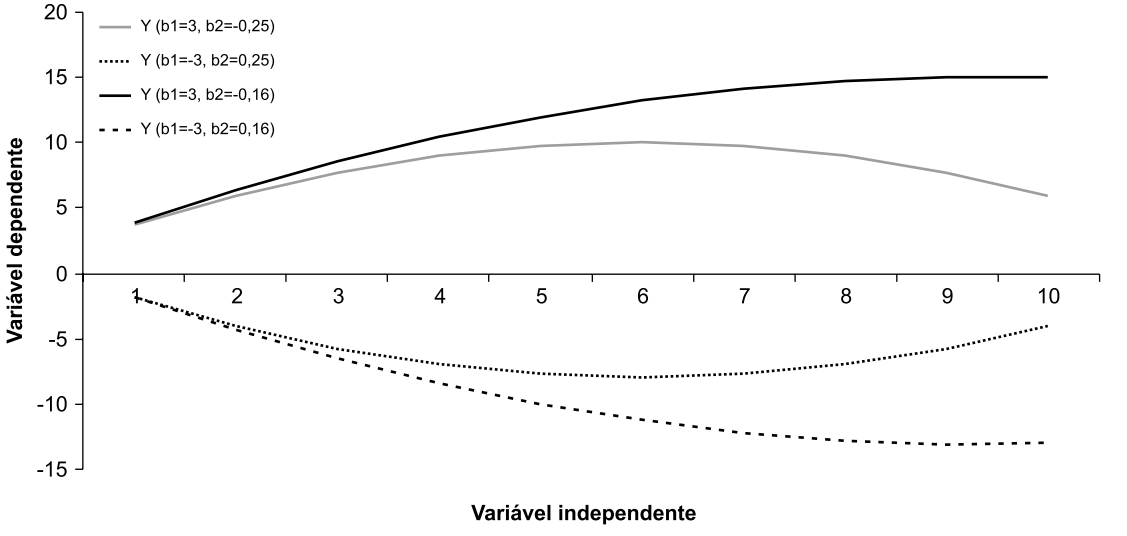


Figura 2. Exemplos de modelos de regressão polinomial de segundo grau.

o cúbico ou de terceiro grau (Figura 3) é representado por

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3 \qquad [3]$$

ou de ordem maior.

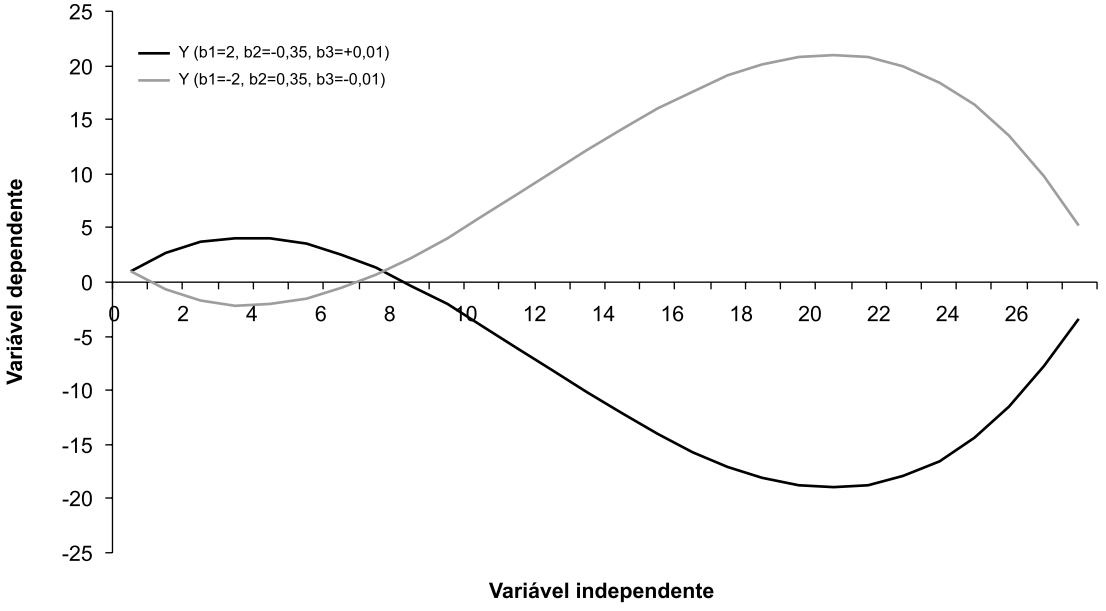


Figura 3. Exemplos de modelos de regressão polinomial de terceiro grau.

Modelos lineares – duas ou mais variáveis independentes

Os modelos representados a seguir são descritos para duas variáveis, mas muito facilmente podem ser estendidos para três ou mais. O modelo de primeiro grau

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} \quad [4]$$

representa um plano no espaço tridimensional. Na Figura 4, o plano foi desenhado com os três parâmetros b_0, b_1 e b_2 iguais a 1. O modelo de segundo grau

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X^2_{1i} + b_4X^2_{2i} + b_5X_{1i}X_{2i} \quad [5]$$

representa uma cúpula ou concha em três dimensões. Para a obtenção da Figura 5, os parâmetros foram definidos como: b_0 = 1.027,45; b_1 = 287,57; b_2 = 308,76; b_{11} = -41,33; b_{22} = -53,71 e b_{12} = -32,43 para as variáveis X_1 e X_2, chamadas de N e P, que variaram entre 0 e 4.

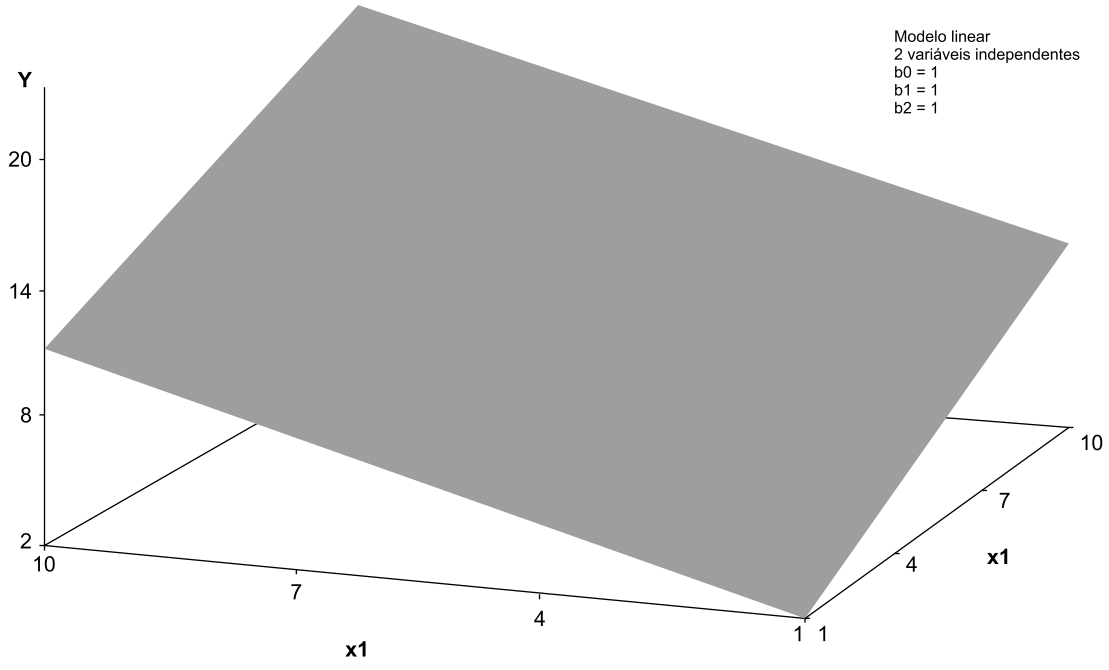


Figura 4. Modelo de regressão de múltipla de primeiro grau ou superfície de resposta de primeira ordem.

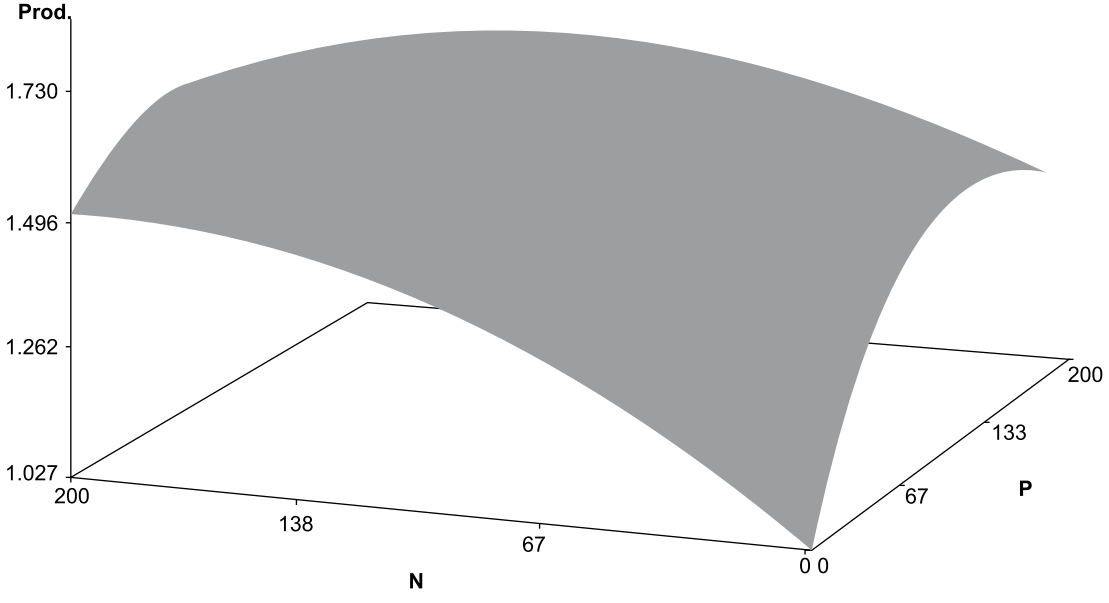


Figura 5. Modelo de regressão múltipla de segundo grau ou superfície de resposta de segundo grau.

Modelos lineares por transformação – uma variável independente

Para muitos modelos de regressão cuja aparência não é aquela de um modelo linear, uma transformação da variável independente, ou da variável resposta ou ambas, permite que o modelo seja expresso como um modelo linear, o que leva ao uso dos processos estudados nas seções anteriores.

a) Modelo exponencial:

$$Y_i = a * e^{b*X_i} \quad [6]$$

em que $e = 2{,}7182$, e o modelo pode ser visto de diversas maneiras, dependendo dos valores de a e b, como na Figura 6.

Para a estimativa dos parâmetros, a e b, deve-se aplicar a transformação por logaritmo natural. Dessa forma, tem-se

$$ln(Y_i) = ln(a) + bX_i * ln(e)$$

$$ln(Y_i) = ln(a) + bX_i$$

uma vez que $ln(e) = 1$. Fazendo $ln(Y_i) = Wi$ e $ln(a) = C$, o modelo terá a seguinte forma:

$$W_i = C + bX_i$$

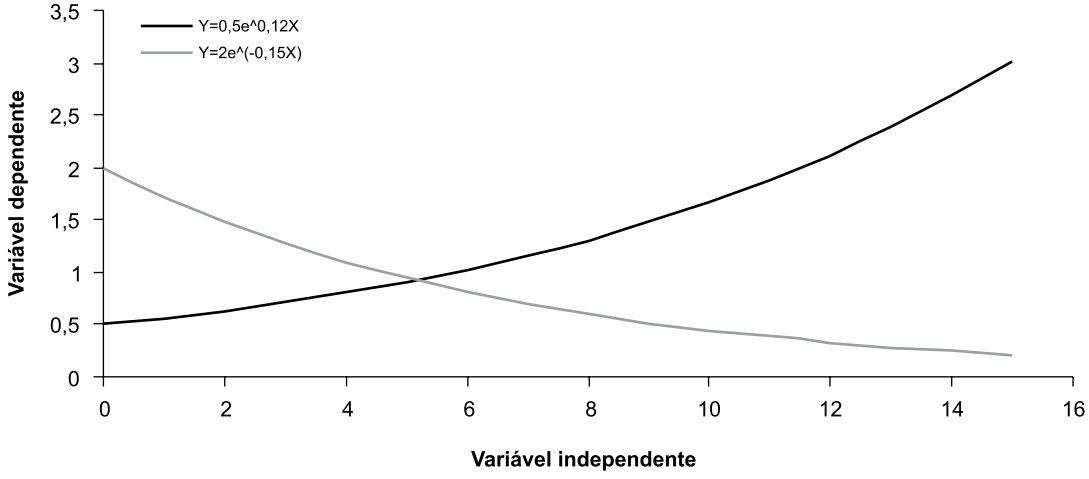


Figura 6. Modelo de regressão exponencial.

b) Modelo potência

$$Y_i = a * b^{X_i} \text{ ou } Y_i = a * X_i^b \quad [7]$$

Também aqui o modelo, dependendo dos valores dos parâmetros, a e b, apresenta diversas formas. Na Figura 7 mostram-se os dois tipos do modelo.

Além disso, deve-se aplicar a transformação de logaritmo (natural ou base 10) para a estimativa dos parâmetros $log(Y_i) = log(a) + log(b)X_i$. Uma vez que $log(Y_i) = W_i$, $log(a) = C$ e $log(b) = D$, segue $Wi = C + DX_i$.

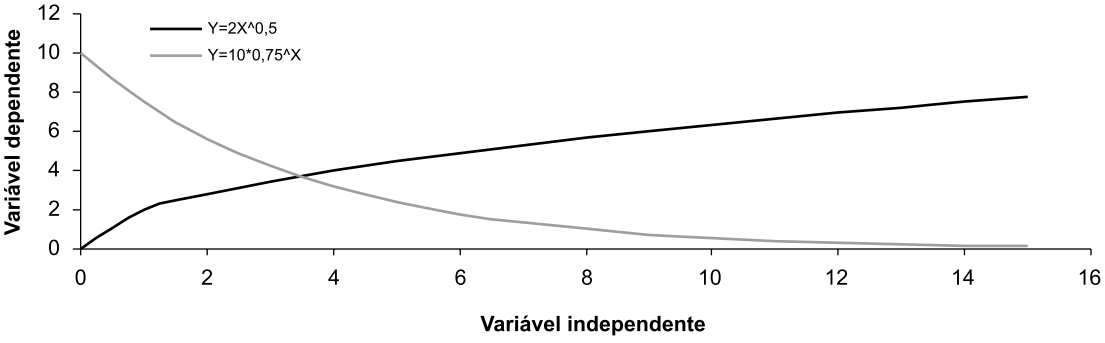


Figura 7. Modelo de regressão do tipo potência.

c) Modelo logístico

$$Y_i = \frac{K}{1 + e^{a+bX_i}} \quad [8]$$

Este modelo é usado em processos de crescimento de populações, animais ou vegetais, em estado natural, sem a intervenção do homem. Aqui o parâmetro K é chamado, em ecologia, de capacidade de suporte ambiental, ou seja, o número máximo de indivíduos que o meio ambiente suporta, enquanto a e b são os parâmetros de população inicial e taxa de crescimento. O desenvolvimento de duas populações, das quais uma serve de alimento para a outra, presa e predador, é explicado pelo uso simultâneo de duas dessas curvas. A Figura 8 mostra um exemplo deste modelo, em que K é considerado igual a 1, muito utilizado em estudos de crescimento de área foliar afetada por uma enfermidade.

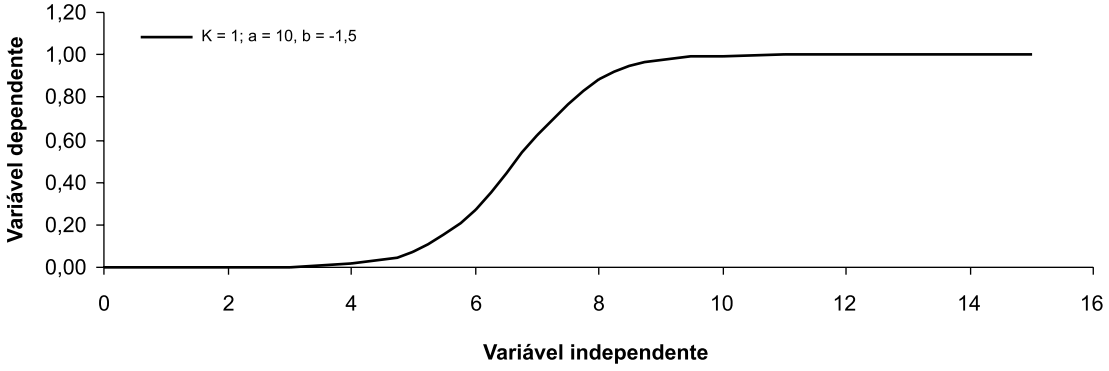


Figura 8. Modelo de regressão logística.

Se neste modelo o valor de K não é conhecido, apenas processos não lineares podem ser aplicados para estimativa dos parâmetros K, a e b. Se K é conhecido então é possível fazer transformação no modelo para torná-lo linear.

$$Y_i = \frac{K}{1 + e^{a+bX_i}}$$

Primeiramente, deve-se procurar a forma inversa, depois multiplicar ambos os termos do modelo por k para obtê-la na forma:

$$\frac{K}{Y_i} = 1 + e^{a+bX_i}$$

Agora, deve-se subtrair 1 em cada lado do modelo e finalmente aplicar logaritmo natural para obter o modelo final, isto é,

$$ln\left(\frac{K}{Y_i} - 1\right) = a + bX_i$$

ou

$$ln\left(\frac{K - Y_i}{Y_i}\right) = a + bX_i$$

Nesse caso, evidentemente K deve ser maior ou igual ao máximo valor de Y, e o valor mínimo de Y deve ser maior que 0 (zero).

d) Modelo de Mitscherlich

O modelo de Mitscherlich, usado em estudos de fertilidade de solo, tem a seguinte expressão:

$$y_i = A[1 - 10^{-c(x_i + b)}] \quad [9]$$

em que A corresponde à produção máxima teórica possível, c é o coeficiente de eficácia do nutriente e b o teor do nutriente encontrado no solo. Por ser um modelo que inclui somas e produtos entre seus parâmetros, ele é evidentemente não linear e a estimação de seus parâmetros se faz por processos iterativos. O modelo pode ser modificado para mais de um nutriente, e para ordens superiores ao primeiro grau. A Figura 9 mostra o aspecto desse modelo para um nutriente e corresponde ao ajuste que se demonstra na seção *Modelos de regressão não lineares*.

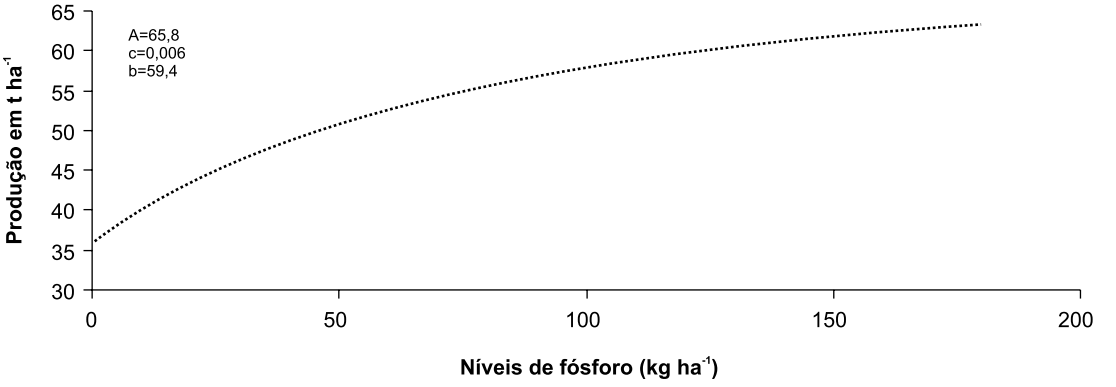


Figura 9. Modelo de regressão de Mitscherlich.

Estudos de correlação

Uma medida de correlação é uma variável aleatória que é usada em situações nas quais os dados consistem em pares de números, como em uma amostra aleatória bivariada de tamanho n, que é representada por (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n), em que cada par tem distribuição bivariada idêntica.

Exemplos de variáveis aleatórias bivariadas incluem: X_i que representa altura da *i*-ésima planta e Y_i que representa a produtividade de grãos, ou X_i que representa a nota de uma avaliação para resistência a doença e Y_i que indica o seu grau de resistência a uma praga.

Por tradição, uma medida de correlação entre X e Y deveria satisfazer os seguintes requerimentos para ser aceitável.

- A medida de correlação deveria assumir valores somente entre -1 e +1.
- Se os valores maiores de X tendem a ser pareados com os valores maiores de Y e, portanto, os valores menores de X e Y tendem a ser pareados também, então a medida de correlação deveria ser positiva, e próxima a $+1$, se a tendência é forte. De forma que seria possível falar de uma correlação positiva entre X e Y.
- Se os valores maiores de X tendem a ser pareados com os valores menores de Y, e vice-versa, então a medida de correlação deveria ser negativa e próxima de -1, se a tendência é forte. Assim, diz-se que X e Y são negativamente correlacionados.
- Se os valores de X parecem ser pareados aleatoriamente com os valores de Y, a medida de correlação deveria ser próxima a zero. Isso seria o caso quando X e Y são independentes, e possivelmente em alguns casos em que X e Y não são independentes. Então se diz que X e Y são não correlacionados ou que têm correlação zero.

A medida de correlação mais comumente usada é o coeficiente de correlação de produtos de momentos de Pearson, chamado de *r* e definido como:

$$r = \frac{cov(x, y)}{\sqrt{var(y) * var(x)}} \qquad [10]$$

Essa medida de correlação pode ser usada com quaisquer dados de uma natureza numérica sem qualquer requerimento concernente à escala de medida ou ao tipo da distribuição básica, embora seja difícil de interpretar, a menos que a escala de medida seja pelo menos de intervalo. Ela tem os requerimentos necessários de uma medida de correlação aceitável. Entretanto, *r* é uma variável aleatória e, como tal, tem uma distribuição de função que depende da distribuição de função bivariada de (*X*, *Y*).

Fisher (1921, citado por DRAPER; SMITH, 1981), apresentou a transformação *z* de *r* dada por

$$z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right) \qquad [11]$$

em que z tem uma distribuição aproximadamente normal, com média e variância dadas por

$$\mu = \frac{1}{2} ln\left(\frac{1+\rho}{1-\rho}\right) \text{ e } \sigma^2 = \frac{1}{n-3} \qquad [12] \text{ e } [13]$$

e isso permite que se obtenha um intervalo de confiança para o valor de ρ.

Para testar uma hipótese nula (H_0) de que $\rho = 0$, é indicada a aplicação de um teste t com $n - 2$ graus de liberdade, como se mostra a seguir.

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}} \qquad [14]$$

Os valores necessários para o cálculo de r seguem os padrões até aqui mostrados no decorrer deste livro. Assim:

$$Cov(X,Y) = \left[\sum X_i Y_i - \frac{1}{n}(\sum x_i)(\sum Y_i)\right]/(n-1) \qquad [15]$$

$$Var(X) = \left[\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right]/(n-1) \qquad [16]$$

e

$$Var(Y) = \left[\sum Y_i^2 - \frac{1}{n}(\sum Y_i)^2\right]/(n-1) \qquad [17]$$

Na prática, a divisão por $n - 1$ em todas as três formas apresentadas não se realiza, pois elas se anulam na fórmula de cálculo de r, e os valores assim calculados recebem os nomes de soma de produtos [$SP(X,Y)$] e soma de quadrados [$SQ(X)$ e $SQ(Y)$].

Exemplo de aplicação

Em um plantio de feijão com a cultivar Jalo Precoce, foram coletadas plantas, em dias sequenciais (idades), e medido o índice de área foliar (IAF) em uma área experimental sob responsabilidade do pesquisador Pedro Marques da Silveira, em 2001, na fazenda Capivara, sede da Embrapa Arroz e Feijão. Os dados obtidos se encontram na Tabela 1.

Tabela 1. Dados de índice de área foliar e idade das plantas na amostragem e valores auxiliares para cálculo do coeficiente de correlação.

Leitura	Idade (X)	Índice de área foliar (IAF) (Y)	X^2	Y^2	X*Y
1	16	0,16115	256	0,025969323	2,5784
2	23	0,38815	529	0,150660423	8,92745
3	30	0,71196	900	0,506887042	21,3588
4	36	1,16696	1.296	1,361795642	42,01056
5	43	2,06033	1.849	4,244959709	88,59419
6	57	2,20084	3.249	4,843696706	125,44788
7	64	1,96282	4.096	3,852662352	125,62048
8	71	1,87116	5.041	3,501239746	132,85236
Total	340	10,52337	17.216	18,48787094	547,39012

Com o uso das fórmulas 15, 16 e 17, sem a divisão por n -1, tem-se:

$$SP(X,Y) = (16*0,16115 + 23*0,38815 + ... + 71*1,87116) - (1/8)(340*10,52337)$$

$$SP(X,Y) = 547,39012 - 447,24322 = 100,14690$$

$$SQ(X) = (256 + 529 + ... + 5.041) - (1/8)(340)^2 = 17.216 - 14.450 = 2.766$$

$$SQ(Y) = (0,025969323 + ... + 3,501239746) - (1/8)(10,52337)^2$$

$$SQ(Y) = 18,48787094 - 13,84266452 = 4,64520642$$

Tem-se, então, que o valor do coeficiente de correlação é igual a:

$$r = \frac{100,14690}{\sqrt{2766 * 4,64520642}} = \frac{100,14690}{113,35184} = 0,8835$$

A aplicação do teste t leva ao seguinte resultado:

$$t = \frac{0,8835}{\sqrt{(1 - 0,8835^2)/(8 - 2)}} = \frac{0,8835}{0,1656} = 5,34$$

Esse valor tem seis graus de liberdade. Um exame do Anexo 2 mostra que a probabilidade associada a esse valor está entre 0,005 e 0,001. Trata-se, portanto, de um valor altamente significativo, diferente de 0 (zero). Logo, pode-se inferir que existe uma grande associação entre a idade da planta e o índice de área foliar.

É possível também calcular um intervalo de confiança para esse valor utilizando a transformação *z*, dada antes, de forma que

$$z = \frac{1}{2}\ln\left(\frac{1 + 0,8835}{1 - 0,8835}\right) = \frac{1}{2}\ln\left(\frac{1,8835}{0,1165}\right) = 1,3915$$

e a variância é igual a $1/(8 - 3)$ ou $0,2$. Dessa maneira, o *IC* a 95% de probabilidade é

$$1,3915 \pm 1,96 * \sqrt{0,2}$$

ou

$$1,3915 \pm 0,8765$$

ou, ainda,

$$0,515 - 2,268$$

Ao se aplicarem esses valores na fórmula inversa de *z*, para obtenção de valores de *r*, em que

$$r = \frac{e^{2z} - 1}{1 + e^{2z}} \qquad [18]$$

tem-se que o intervalo para ρ está dado por

$0,473 \leq \rho \leq 0,978$

Análise de regressão linear

Este assunto é estreitamente relacionado ao de correlação de dados, em que se está examinando uma amostra aleatória (X_1, Y_1), ..., (X_n, Y_n) da variável aleatória bivariada (X, Y). Métodos de correlação enfatizam a estimação do grau de dependência entre X e Y. Métodos de regressão são usados para inspecionar a relação entre X e Y mais estreitamente. Um importante objetivo dos métodos de regressão é predizer um valor de Y no par (X, Y) em que somente o valor de X é conhecido, com base na informação que se pode obter das observações prévias (X_1, Y_1) até (X_n, Y_n). Por exemplo, se X representa a nota de avaliação de resistência a uma doença de feijão e Y representa a produtividade da planta com aquela doença, observações de muitas outras plantas podem ajudar a predizer qual será a produtividade de uma planta de feijão com um determinado nível de incidência da doença. Por certo, Y é ainda uma variável aleatória, de maneira que não se pode esperar determinar Y somente por conhecer o valor associado de X, mas conhecer X ajudaria a fazer uma melhor previsão de Y.

A regressão de Y em X é $E(Y|X = x)$. A equação de regressão é $y = E(Y|X = x)$. Se a equação de regressão é conhecida, pode-se representar a equação num gráfico utilizando y como ordenada e x como abcissa. Mas a equação de regressão é raramente conhecida. Para que a regressão de Y sobre X possa ser estimada, várias observações são necessárias e deve ser esperado que uma grande amplitude de valores para X e Y seja observada, pois o ajuste do modelo e as predições que se façam a partir desse modelo só serão válidos para a amplitude estudada nos dados coletados. Como será visto mais adiante, o intervalo de confiança para uma predição é tanto maior quanto mais o valor de X difere do valor médio, podendo chegar, em alguns casos, a valores de intervalos muito maiores que a própria predição.

A regressão de Y sobre X é dita ser linear se o gráfico da equação de regressão é uma linha reta e, nesse caso, a equação é da forma

$$E(Y|X = x) = A + Bx \qquad [19]$$

para os parâmetros A, chamado de ponto de interseção, e B, chamado de inclinação ou coeficiente angular.

Usualmente as constantes A e B são desconhecidas e devem ser estimadas a partir dos dados. Se todas as observações de Y e X são usadas para estimar A e B, é feito o máximo uso dos dados, e uma boa estimativa de $E(Y|X = x)$ para cada x pode ser esperada. Um método comumente aceito para estimar A e B é o chamado método dos quadrados mínimos.

O método dos quadrados mínimos para estimar a e b de A e B, na equação de regressão $y = A + Bx$, é o método que minimiza a soma dos quadrados dos desvios

$$SQDesvios = \sum_{i=1}^{n} [\, Y_i - (A + BX_i)\,]^2 \qquad [20]$$

das observações (X_1, Y_1), ... (X_n, Y_n).

A ideia atrás do método de quadrados mínimos é que uma estimativa da linha de regressão deveria ser próxima aos valores observados de X e Y, porque a linha de regressão verdadeira é provavelmente próxima às observações. Portanto, a estimativa é selecionada de maneira que a distância vertical e_i entre Y_i e a linha de regressão estimada, que é igual a $a + bX_i$ diretamente acima ou abaixo de Y_i seja pequena quando todos os pontos são considerados. Não se pode simplesmente fazer a soma dos e's pequena porque a soma deles seria igual a zero, mesmo quando a linha de regressão estimada não é próxima às observações. Isto é, os valores absolutos das distâncias ε podem ser grandes, mas os valores positivos cancelariam os negativos, dando soma zero. Para evitar isso, minimiza-se a soma dos quadrados dos valores de ε_i, em que $\varepsilon_i = Y_i - (a + bX_i)$.

O método de quadrados mínimos fornece as estimativas $y = a + bx$, da linha de regressão verdadeira $y = A + Bx$, em que a e b são calculados a partir de

$$b = \frac{\sum XY - \frac{1}{n}(\sum X)(\sum |Y)}{\sum X^2 - \frac{1}{n}(\sum X)^2} \qquad [21]$$

e

$$a = \overline{Y} - b\overline{X} \qquad [22]$$

em que $\bar{Y}$ e $\bar{X}$ são as médias aritméticas amostrais das variáveis X e Y.

Como se pode ver, a fórmula de cálculo de b coincide com o numerador daquela empregada para obtenção de r, exceto pela divisão por uma constante $(n - 1)$. Isso implica que o mesmo sinal de r será encontrado para b e que há uma estreita relação entre essas

estatísticas. Caso os valores de X sejam igualmente espaçados, o uso de contrastes, como feito no Capítulo 8, também pode ser adotado.

No caso de regressão, os testes de hipóteses podem ser aplicados tanto para *a* quanto para *b*; no entanto, este último é muito mais importante para o pesquisador. O valor *a* indica o ponto de corte da linha de regressão no eixo das abcissas, em que o valor 0 (zero) indica que a linha passa pelo ponto de origem. O valor de *b* mostra a inclinação da linha em relação ao eixo das ordenadas. Um valor igual a 0 (zero) seria um indicador de que a linha é paralela a esse eixo; portanto, nenhuma associação existe entre as variáveis X e Y, pois, para qualquer valor de X, o valor de Y é constante e igual ao ponto de intersecção a.

Os valores esperados de a e b são A e B, respectivamente, e suas variâncias são dadas por

$$Var(a) = \frac{\sum X_i^2}{n\left(\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right)} * s^2 \qquad [23]$$

e

$$Var(b) = \frac{s^2}{\left(\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right)} \qquad [24]$$

em que s^2 é o estimador do erro ou o quadrado médio do erro na análise de variância, como se verá mais adiante.

Ao assumir que os erros (ε_i) têm distribuição normal, com média zero e variância σ^2, intervalos de confiança ou provas de hipóteses acerca desses parâmetros podem ser obtidos.

Para a análise de variância, pode ser construído um quadro com as características mostradas na Tabela 2.

No caso de predição de valores, os valores preditos médios ($\hat{Y}_0$) para um determinado valor de X_0 apresentam uma estimativa de variância dada por

$$Var(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - \frac{1}{x}(\sum X)^2}\right] * s^2 \qquad [25]$$

Tabela 2. Análise de variância para um estudo de regressão linear.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	Teste de F
Regressão (b)	1	$\frac{\left[\sum XY - \frac{1}{n}(\sum X)(\sum Y)\right]^2}{\sum X^2 - \frac{1}{n}(\sum X)^2}$	*SQR*/*GLR*	*QMR*/*QME*
Erro	$n - 2$	Diferença	*SQE*/*GLE*	
Total	$n - 1$	$\sum Y_i^2 - 1/n(\sum Y_i)^2$		

em que s^2 é o quadrado médio do erro na análise de variância, dada na Tabela 2. Logo, é possível determinar um intervalo de confiança para essa estimativa dado por:

$$\hat{Y}_0 \pm t_{\alpha,GLE}\sqrt{Var(\hat{Y}_0)} \qquad [26]$$

Exemplo de aplicação

Para este exemplo, utilizam-se os mesmos dados de índice de área foliar e idade das plantas (Tabela 1) do exemplo anterior. Dessa forma, os valores dos parâmetros são então calculados:

$$b = \frac{(16*0{,}16115 + ...71*1{,}87116 - \frac{1}{8}(340)(10{,}52337)}{(16^2 + ... + 71^2) - \frac{1}{8}(340)^2} = \frac{100{,}14690}{2766} = 0{,}0362$$

e

$$a = \frac{10{,}52337}{8} - 0{,}0362 * \frac{340}{8} = -0{,}223$$

O quadro da análise de variância (Tabela 3) é facilmente montado a partir dos dados dos parâmetros acima calculados e dos valores disponibilizados na Tabela 1.

Tabela 3. Análise de variância para os dados da Tabela 1. Regressão linear de índice de área foliar como uma função da idade das plantas.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Regressão (b)	1	3,6260	3,6260	21,34 (0,003)
Erro	6	1,0192	0,1699	
Total	7	4,6452		

Na Tabela 3, a soma de quadrados devida à regressão foi obtida utilizando-se a fórmula dada anteriormente

$$SQReg = \frac{[100,14690]^2}{2766} = 3,6260$$

enquanto a soma de quadrados total é igual a

$$SQTotal = (0,025969323 + ... + 3,501239746) - (1/8)(10,52337)^2$$

$$SQTotal = 18,48787094 - 13,84266452 = 4,6452$$

A equação de regressão então é estimada por

$$\hat{Y}_i = -0,223 + 0,0363X_i$$

o que implica que, a cada dia de idade, o IAF cresce 0,0362 unidade (m^2 de área foliar por m^2 de solo). As estimativas das variâncias de *a* e *b* são dadas por:

$$Var(a) = \frac{\sum X_i^2}{n\left(\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right)} * s^2 = \frac{17.216 * 0,1699}{8 * 2.766} = 0,1322$$

e

$$Var(b) = \frac{s^2}{\left(\sum X_i^2 - \frac{1}{n}(\sum X_i)^2\right)} = \frac{0,1699}{2766} = 0,0000614$$

A partir da equação ajustada, é possível estimar valores de IAF para plantas com idades dentro da faixa de idades amostradas. Além disso, pode-se estimar a variância dessa estimativa, para a qual se pode, portanto, calcular um intervalo de confiança. Para uma idade igual a 30 dias, tem-se:

$$\hat{Y}_0 = -0{,}223 + 0{,}0363X_0 = -0{,}223 + 0{,}0363*30 = 0{,}866$$

Sua variância é igual a

$$Var(\hat{Y}_0) = \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum X^2 - \frac{1}{x}(\sum X)^2}\right] * s^2 = \left[\frac{1}{8} + \frac{(30 - 42{,}5)^2}{2766}\right] * 0{,}1699 = 0{,}0308$$

e o intervalo de confiança dessa estimativa é dado por

$$\hat{Y}_0 \pm t_{\alpha,GLE}\sqrt{Var(\hat{Y}_0)} = 0{,}866 \pm 2{,}976\sqrt{0{,}0308} = 0{,}866 \pm 0{,}521$$

Isso implica que o verdadeiro valor de IAF para 30 dias está no intervalo (0,345; 1,387).

Modelos polinomiais de ordem 2 ou superior

Muitas vezes o interesse do pesquisador pode ser o de ajustar uma equação diversa daquela de uma linha reta, como, por exemplo, uma equação de segundo grau, de terceiro ou de grau superior.

Uma equação de segundo grau tem a forma

$$E(Y|X = x) = B_0 + B_1X + B_2X^2 \qquad [27]$$

e uma de terceiro grau, a forma

$$E(Y|X = x) = B_0 + B_1X + B_2X_2 + B_3X^3 \qquad [28]$$

e assim sucessivamente para graus maiores.

Nos estudos no campo biológico, a equação de segundo grau é uma das mais interessantes, por isso ela será estudada a seguir com mais detalhes.

Para a estimação dos parâmetros B_0, B_1 e B_2, adota-se o mesmo procedimento visto para a equação da linha reta, ou seja, emprega-se o método dos mínimos quadrados.

Assim, tem-se que:

$$SQDesvios = \sum_{i=1}^{n} [Y_i - (B_0 + B_1 X_i + B_2 X_i^2)]^2 \quad [29]$$

A partir das derivadas parciais da expressão com relação aos parâmetros B_i, encontra-se um sistema de três equações com três variáveis (parâmetros), que leva à seguinte solução, em que X_2 é a media dos valores X_i^2:

$$b_0 = \overline{Y} - b_1 \overline{X} - b_2 X_2 \quad [30]$$

Para b_1 e b_2, deve-se resolver o seguinte sistema de equações:

$$b_1 \left[\sum X_i^2 - \frac{1}{n} (\sum X_i)^2 \right] + b_2 \left[\sum X_i^3 - \frac{1}{n} \sum X_i^2 \sum X_i \right] = \sum X_i Y_i - \frac{1}{n} \sum X_i \sum Y_i \quad [31]$$

$$b_1 \left[\sum X_i^3 - \frac{1}{n} \sum X_i^2 \sum X_i \right] + b_2 \left[\sum X_i^4 - \frac{1}{n} (\sum X_i^2)^2 \right] = \sum X_i^2 Y_i - \frac{1}{n} \sum X_i^2 \sum Y_i \quad [32]$$

Para a análise de variância desse tipo de modelo, cabe também a construção de um quadro conforme apresentado na Tabela 4.

Caso haja interesse em conhecer a contribuição à melhoria de ajuste da linha de regressão dada a inclusão do componente quadrático, deve-se primeiro fazer a análise do modelo linear (primeiro grau), e a diferença entre as duas somas de quadrados mediria essa contribuição. Esse procedimento será mostrado no exemplo a seguir. Da mesma forma que no caso da regressão linear, é possível estimar as variâncias das estimativas dos parâmetros b_i, valores de $\hat{Y}_0$ e sua variância e, portanto, o intervalo de confiança (DRAPER; SMITH, 1998).

Exemplo de aplicação

Voltando ao exemplo referente à Tabela 1, considera-se aqui a aplicação de um modelo polinomial de segundo grau aos dados de IAF como função da idade das plantas. A partir da Tabela 1, desenvolve-se a Tabela 5, como auxiliar dos cálculos necessários.

Tabela 4. Análise de variância para um estudo de regressão polinomial de segundo grau.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	Teste de F
Regressão (b_1 e b_2)	2	$b_1\sum(X-\bar{X})(Y-\bar{Y})+b_2\sum(X^2-X2)(Y-\bar{Y})$	SQR/GLR	QMR/QME
Erro	$n-3$	Diferença	SQE/GLE	
Total	$n-1$	$\sum Y_i^2 - \frac{1}{n}(\sum Y_i)^2$		

Tabela 5. Dados auxiliares para a análise de uma regressão polinomial de segundo grau para IAF como função da idade das plantas.

Leitura	Y	X	X^2	X^3	X^4	X*Y	X^2Y	Y^2
1	0,16115	16	256	4.096	65.536	2,5784	41,2544	0,025969323
2	0,38815	23	529	12.167	279.841	8,92745	205,33135	0,150660423
3	0,71196	30	900	27.000	810.000	21,3588	640,764	0,506887042
4	1,16696	36	1.296	46.656	1.679.616	42,01056	1.512,38016	1,361795642
5	2,06033	43	1.849	79.507	3.418.801	88,59419	3.809,55017	4,244959709
6	2,20084	57	3.249	185.193	10.556.001	125,44788	7.150,52916	4,843696706
7	1,96282	64	4.096	262.144	16.777.216	125,62048	8.039,71072	3,852662352
8	1,87116	71	5.041	357.911	25.411.681	132,85236	9.432,51756	3,501239746
Total	10,52337	340	17.216	974.674	58.998.692	547,39012	30.832,03752	18,48787094

Primeiro, devem-se estimar b_1 e b_2, utilizando o sistema de equações dado acima. Dessa forma, tem-se:

$$b_1\left[17.216 - \frac{1}{8}(340)^2\right] + b_2\left[974.674 - \frac{1}{8}(17.216)(340)\right] =$$
$$547,39012 - \frac{1}{8}(340)(10,52337)$$

$$b_1\left[974.674 - \frac{1}{8}(17.216)(340)\right] + b_2\left[58.998.692 - \frac{1}{8}(17.216)^2\right] =$$
$$30.832,03752 - \frac{1}{8}(17.216)(10,52337)$$

ou

$$b_1(2.766) + b_2(242.994) = 100,14690$$

$$b_1(242.994) + b_2(21.949.860) = 8.185,74528,$$

de onde se obtém então que $b_1 = 0,12543$ e $b_2 = -0,00101565$.

Portanto,

$$b_0 = 10,52337/8 - 0,12543*(340/8) - (-0,00101565)*(17.216/8)$$

$$b_0 = -1,82967$$

Dessa forma, a equação de regressão de segundo grau é

$$\hat{Y}_i = -1,82967 + 0,12543b_1 - 0,00102b_2.$$

A análise de variância é mostrada na Tabela 6.

Tabela 6. Análise de variância para um estudo de regressão polinomial de segundo grau para IAF como função da idade da planta.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Regressão (b_1 e b_2)	2	4,24757	2,12388	26,70 (< 0,001)
Erro	5	0,39765	0,07953	
Total	7	4,64522		

A soma de quadrados em razão da regressão foi obtida aplicando-se a fórmula dada anteriormente, onde X_2 é a média aritmética dos quadrados de X_1:

$$SQReg = b_1 \sum (X - \overline{X})(Y - \overline{Y}) + b_2 \sum (X^2 - X2)(Y - \overline{Y})$$

$$SQReg = 0{,}12543 * 100{,}1469 - 0{,}00102 * 8.185{,}74528 = 4{,}24757$$

A soma de quadrados total é a mesma obtida na análise de regressão linear e do erro por diferença. Caso se decida verificar a importância da inclusão do componente de segundo grau no modelo, a soma de quadrados de regressão pode ser dividida, como afirmado anteriormente, e o quadro de análise de variância para esse fato está na Tabela 7.

A soma de quadrados devida a b_1 é aquela obtida na análise da regressão linear, e a soma de quadrados devida a b_2 dado que $b_1(b_2/b_1)$ foi obtida por diferença entre as somas de quadrados de regressão nas duas análises.

Tabela 7. Análise de variância para um estudo de regressão polinomial de segundo grau para IAF como função da idade da planta.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Regressão (b_1)	1	3,62600	3,62600	45,59 (< 0,001)
Regressão (b_2/b_1)	1	0,62157	0,62157	7,82 (0,05)
Erro	5	0,39765	0,07953	
Total	7	4,64522		

O uso de matrizes para a estimação de parâmetros e análise de variância de modelos de regressão lineares

Uma leitura cuidadosa das seções anteriores mostrará ao leitor uma certa sequência lógica na estimação de parâmetros e na análise de variância adotada. O uso de matrizes para a representação dos dados e dos modelos permite uma generalização rápida dos procedimentos adotados. Retornando ao modelo linear simples ou polinomial de ordem 1, isto é,

$$Y_i = A + BX_i + \varepsilon_i \quad [33]$$

em que i varia de 1 até n, pode-se representar o conjunto de observações em forma matricial e então as observações Y são representadas pelo vetor Y de dimensões n x 1, as observações X por uma matriz X de ordem n x 2, em que a primeira coluna será composta apenas por 1s e a segunda coluna serão os valores de X_i, os parâmetros serão contidos num vetor 2 x 1 chamado de β, e os erros de medida, num vetor ε de dimensão n x 1. O modelo é escrito, portanto, na forma

$$Y = X\beta + \varepsilon \quad [34]$$

Percebe-se claramente que pequenos ajustes nas descrições da matriz X e no vetor β fazem que esse modelo matricial seja único na representação dos modelos lineares.

A soma de quadrados de erros que deve ser minimizada, segundo os procedimentos adotados nos itens *Estudos de correlação* e *Análise de regressão linear*, é dada por

$$\varepsilon'\varepsilon = (Y - X\beta)'(Y - X_\beta) \quad [35]$$

que leva a que a estimação dos parâmetros seja igual a

$$\hat{\beta} = (X'X)^{-1}X'Y \quad [36]$$

Da mesma forma, pode-se representar a soma de quadrados total por

$$SQTotal = Y'Y - Y'11'Y/n \quad [37]$$

em que 1 é um vetor 1 x n de $1s$ e n o número de observações. A soma de quadrados de regressão é dada por

$$SQb_1/b_0 = \beta' X'Y - \hat{Y}'11'Y/n \quad [38]$$

e a soma de quadrados de erro, ou desvio de regressão, é obtida por diferença entre a $SQTotal$ e SQb_1/b_0.

A variância dos parâmetros (β) é dada pela expressão

$$V(\beta) = (X'X)^{-1}\sigma^2 \quad [39]$$

em que σ^2 é dado pelo erro puro, se houver, ou pelo erro estimado por falta de ajuste.

Para modelos mais completos, polinômios de ordem 2 ou superior ou de regressão múltipla, o vetor β e a matriz X são modificadas segundo o modelo adotado, mas as soluções apresentadas são as mesmas.

Exemplo de aplicação de matrizes

Como exemplo de aplicação, vai-se adotar o mesmo caso do exemplo dado na Tabela 1, na qual houve um ajuste do modelo polinomial de segundo grau. Neste caso, os vetores Y e β e a matriz X são iguais a:

$$Y = \begin{bmatrix} 0,16115 \\ 0,38815 \\ 0,71196 \\ 1,16696 \\ 2,06033 \\ 2,20084 \\ 1,96282 \\ 1,87116 \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & 16 & 256 \\ 1 & 23 & 529 \\ 1 & 30 & 900 \\ 1 & 36 & 1.296 \\ 1 & 43 & 1.849 \\ 1 & 57 & 3.249 \\ 1 & 64 & 4.096 \\ 1 & 71 & 5.041 \end{bmatrix} \quad \text{e} \quad \beta = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

Para a estimação de β, faz-se a aplicação da fórmula [28], de modo que se tem então:

$$\beta = \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 16 & 23 & 30 & 36 & 43 & 57 & 64 & 71 \\ 16^2 & 23^2 & 30^2 & 36^2 & 43^2 & 57^2 & 64^2 & 71^2 \end{bmatrix} \begin{bmatrix} 1 & 16 & 16^2 \\ 1 & 23 & 23^2 \\ 1 & 30 & 30^2 \\ 1 & 36 & 36^2 \\ 1 & 43 & 43^2 \\ 1 & 57 & 57^2 \\ 1 & 64 & 64^2 \\ 1 & 71 & 71^2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 16 & 23 & 30 & 36 & 43 & 57 & 64 & 71 \\ 16^2 & 23^2 & 30^2 & 36^2 & 43^2 & 57^2 & 64^2 & 71^2 \end{bmatrix} \begin{bmatrix} 0{,}16115 \\ 0{,}38815 \\ 0{,}71196 \\ 1{,}16696 \\ 2{,}06033 \\ 2{,}20084 \\ 1{,}92282 \\ 1{,}87116 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 8 & 340 & 17.216 \\ 340 & 17.216 & 974.674 \\ 17.216 & 974.674 & 58.998.692 \end{bmatrix}^{-1} \begin{bmatrix} 10{,}52337 \\ 547{,}39012 \\ 30.832{,}03752 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 4{,}92824 & -0{,}24588 & 0{,}00262 \\ -0{,}24588 & 0{,}01317 & -0{,}00014 \\ 0{,}00262 & -0{,}00014 & 0{,}0000016 \end{bmatrix} \begin{bmatrix} 10{,}52337 \\ 547{,}39012 \\ 30.832{,}03752 \end{bmatrix}$$

estimativa dos parâmetros é igual a

$$B = \begin{bmatrix} -1{,}82697 \\ 0{,}12543 \\ -0{,}00102 \end{bmatrix}$$

Utilizando as demais equações, poder-se-ia realizar a análise completa como anteriormente, mas esse resultado se deixa ao leitor como exercício. Pequenas diferenças podem surgir, dada a precisão com que se realizam os cálculos.

Uso do SAS para estudos de regressão e correlação

O SAS possui uma série de procedimentos para estudos de correlação e regressão, mas evidentemente não será possível discutir todos eles aqui. Entre as múltiplas opções do SAS, os procedimentos CORR, GLM, NLIN, REG e RSREG são os mais utilizados. O NLIN foi desenvolvido especificamente para estudos de modelos não lineares e não linearizáveis,

enquanto o RSREG é feito para estudos de superfície de resposta e será olhado com mais detalhes no Capítulo 14.

Uso do SAS para correlação

Para este caso específico, o procedimento mais simples é o CORR, cuja sintaxe é: PROC CORR;VAR "nomes das variáveis". Para os dados da Tabela 1, o programa seria:

PROC CORR;VAR IDADE IAF;

O resultado é apresentado na Figura 10.

The CORR Procedure
2 Variables: IDADE IAF

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
IDADE	8	42.50000	19.87820	340.00000	16.00000	71.00000
IAF	8	1.31542	0.81462	10.52337	0.16115	2.20084

Pearson Correlation Coefficients, N = 8
Prob > |r| under H0: Rho=0

	IDADE	IAF
IDADE	1.00000	0.88350 0.0036
IAF	0.88350 0.0036	1.00000

Figura 10. Determinação do coeficiente de correlação linear com o uso do SAS (PROC CORR).

O procedimento mostra as estatísticas básicas das variáveis em estudo e a matriz de coeficientes de correlação de Pearson e suas probabilidades. A diagonal da matriz é sempre composta de 1, pois refere-se à correlação da variável com ela mesma.

Uso do SAS para estudos de modelos de regressão lineares ou linearizáveis

Neste caso, usando os mesmos dados da Tabela 1, podem ser usados os procedimentos GLM e REG, cujos resultados são apresentados nas Figuras 10 e 11.

Usando o procedimento GLM, para o modelo linear de primeiro grau o programa é:

PROC GLM;MODEL IAF=IDADE/SS1;OUTPUT OUT=B P=IAF_EST RESIDUAL=DESVIOS;
PROC PRINT DATA=B;

O resultado é visto na Figura 11.

```
                          The GLM Procedure
                       Number of observations    8
                          The GLM Procedure

Dependent Variable: IAF

                                  Sum of
Source                     DF       Squares    Mean Square   F Value   Pr > F
Model                       1    3.62595827     3.62595827     21.34   0.0036
Error                       6    1.01924815     0.16987469
Corrected Total             7    4.64520642

                  R-Square     Coeff Var      Root MSE      IAF Mean
                  0.780581      31.33282      0.412159      1.315421

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
IDADE                       1    3.62595827     3.62595827     21.34   0.0036

                                                Standard
          Parameter              Estimate           Error    t Value    Pr > |t|
          Intercept          -.2233506363      0.36354601      -0.61      0.5615
          IDADE             0.0362063973      0.00783679       4.62      0.0036

                    Obs    IDADE      IAF      IAF_EST     DESVIOS

                      1       16    0.16115    0.35595    -0.19480
                      2       23    0.38815    0.60940    -0.22125
                      3       30    0.71196    0.86284    -0.15088
                      4       36    1.16696    1.08008     0.08688
                      5       43    2.06033    1.33352     0.72681
                      6       57    2.20084    1.84041     0.36043
                      7       64    1.96282    2.09386    -0.13104
                      8       71    1.87116    2.34730    -0.47614
```

Figura 11. Resultado do uso do GLM do SAS para estudo de uma regressão linear de primeiro grau.

Para o caso de se desejar usar o procedimento REG o programa seria:

```
PROC REG;MODEL IAF=IDADE/P R;
```

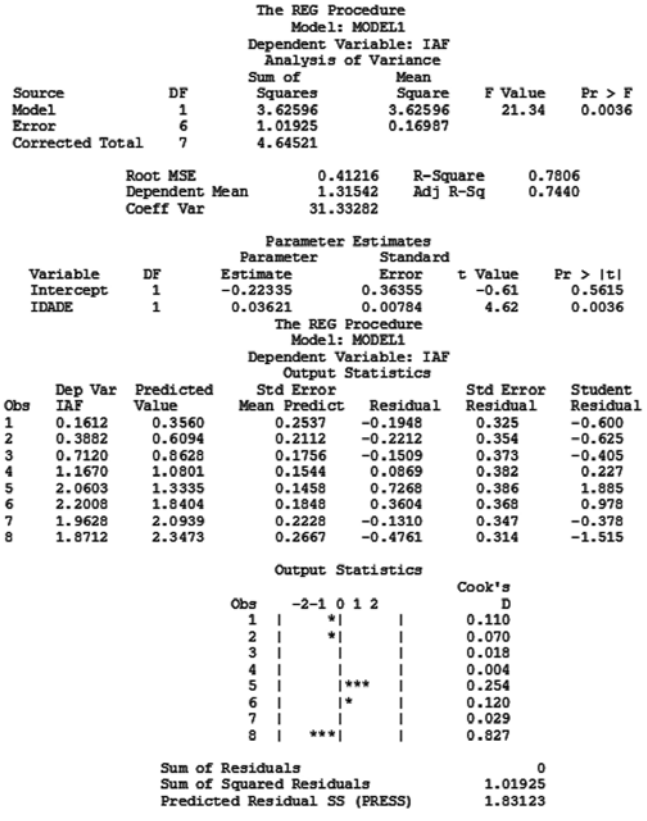
O resultado aparece na Figura 12.

Para o caso de um ajuste de modelo de segunda ordem ou polinomial quadrático, o programa para obtenção da análise com o uso do GLM é:

```
PROC GLM;MODEL IAF=IDADE IDADE*IDADE/SS1;OUTPUT OUT=C P=IAF_EST
RESIDUAL=DESVIOS;PROC PRINT DATA=C;
```

O resultado é visto na Figura 13.

O uso do procedimento REG para este caso, ou do GLM e do REG para modelos de ordem maior, fica por conta do leitor, lembrando apenas que, no caso dos modelos não lineares, mas linearizáveis, a transformação necessária para a(s) variável(eis) deve ser feita já na entrada dos dados.



The REG Procedure
Model: MODEL1
Dependent Variable: IAF
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.62596	3.62596	21.34	0.0036
Error	6	1.01925	0.16987		
Corrected Total	7	4.64521			

Root MSE	0.41216	R-Square	0.7806
Dependent Mean	1.31542	Adj R-Sq	0.7440
Coeff Var	31.33282		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.22335	0.36355	-0.61	0.5615
IDADE	1	0.03621	0.00784	4.62	0.0036

The REG Procedure
Model: MODEL1
Dependent Variable: IAF
Output Statistics

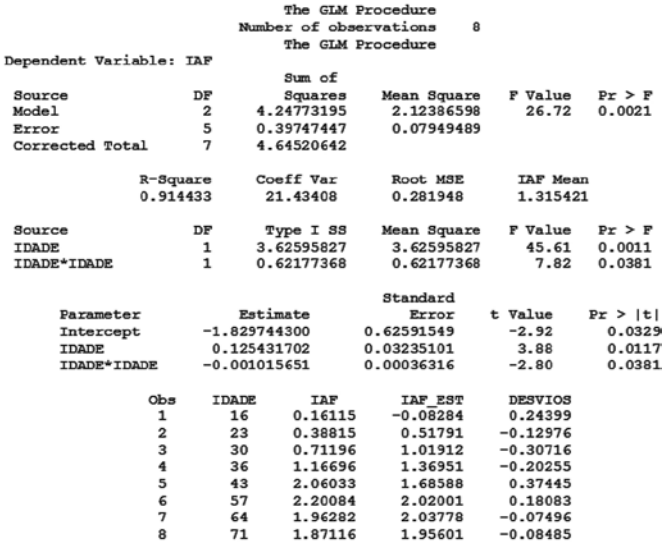
Obs	Dep Var IAF	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual
1	0.1612	0.3560	0.2537	-0.1948	0.325	-0.600
2	0.3882	0.6094	0.2112	-0.2212	0.354	-0.625
3	0.7120	0.8628	0.1756	-0.1509	0.373	-0.405
4	1.1670	1.0801	0.1544	0.0869	0.382	0.227
5	2.0603	1.3335	0.1458	0.7268	0.386	1.885
6	2.2008	1.8404	0.1848	0.3604	0.368	0.978
7	1.9628	2.0939	0.2228	-0.1310	0.347	-0.378
8	1.8712	2.3473	0.2667	-0.4761	0.314	-1.515

Output Statistics

Obs	-2-1 0 1 2	Cook's D
1	*| |	0.110
2	*| |	0.070
3	| |	0.018
4	| |	0.004
5	|*** |	0.254
6	|* |	0.120
7	| |	0.029
8	***| |	0.827

Sum of Residuals	0
Sum of Squared Residuals	1.01925
Predicted Residual SS (PRESS)	1.83123

Figura 12. Resultado do uso do procedimento REG do SAS para um modelo linear de primeira ordem.



The GLM Procedure
Number of observations 8
The GLM Procedure
Dependent Variable: IAF

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4.24773195	2.12386598	26.72	0.0021
Error	5	0.39747447	0.07949489		
Corrected Total	7	4.64520642			

R-Square	Coeff Var	Root MSE	IAF Mean
0.914433	21.43408	0.281948	1.315421

Source	DF	Type I SS	Mean Square	F Value	Pr > F
IDADE	1	3.62595827	3.62595827	45.61	0.0011
IDADE*IDADE	1	0.62177368	0.62177368	7.82	0.0381

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-1.829744300	0.62591549	-2.92	0.0329
IDADE	0.125431702	0.03235101	3.88	0.0117
IDADE*IDADE	-0.001015651	0.00036316	-2.80	0.0381

Obs	IDADE	IAF	IAF_EST	DESVIOS
1	16	0.16115	-0.08284	0.24399
2	23	0.38815	0.51791	-0.12976
3	30	0.71196	1.01912	-0.30716
4	36	1.16696	1.36951	-0.20255
5	43	2.06033	1.68588	0.37445
6	57	2.20084	2.02001	0.18083
7	64	1.96282	2.03778	-0.07496
8	71	1.87116	1.95601	-0.08485

Figura 13. Resultado do uso do GLM, do SAS, para um modelo de regressão de segunda ordem ou quadrático.

Modelos lineares múltiplos e o SAS

Em muitos casos, a associação entre variáveis envolve mais de uma variável independente (seção *Modelos lineares – duas ou mais variáveis independentes*), isto é, X_1, X_2, ..., X_k. Nesse caso, fala-se de um modelo de regressão linear múltipla. Os procedimentos a serem adotados são os mesmos mostrados para a regressão polinomial e, nesta seção, será analisado o caso de primeiro grau. Caso no modelo se incluam as variáveis independentes na segunda potência e seus produtos lineares, tem-se um modelo de superfície de resposta de segundo grau, que será estudado no Capítulo 14.

Usando dados fornecidos pelo pesquisador Luís Fernando Stone, da Embrapa Arroz e Feijão, obtidos em um experimento sobre o manejo da cultura do arroz irrigado, foi ajustado um modelo de regressão linear múltipla de primeiro grau, no qual a variável dependente era a produção de grãos em kg ha^{-1}, com a altura da planta medida em centímetros e o número de perfilhos por metro linear de plantio. Com o emprego do procedimento REG do SAS, obteve-se o resultado mostrado na Figura 14. O programa usado para isso foi:

```
PROG REG;MODEL PROD=NPER ALT;
```

The REG Procedure
Model: MODEL1
Dependent Variable: PROD

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	22593835	11296918	18.07	<.0001
Error	125	78162209	625298		
Corrected Total	127	100756044			

Root MSE	790.75765	R-Square	0.2242
Dependent Mean	2787.48047	Adj R-Sq	0.2118
Coeff Var	28.36819		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-241.75559	509.39031	-0.47	0.6359
NPER	1	27.11334	5.26357	5.15	<.0001
ALT	1	15.30053	3.27862	4.67	<.0001

Figura 14. Resultado da análise de uma regressão linear múltipla de primeiro grau, usando-se o procedimento REG do SAS.

Com base nesses resultados, vrifica-se que a equação de regressão ajustada é:

PROD = -241,75 + 27,11*NPER + 15,3*ALT.

Isso indica uma associação positiva entre as variáveis com cada perfilho por metro linear de plantio e contribui com um crescimento na produção de 27,11 kg ha^{-1}. Cada centímetro da altura da planta tem contribuição positiva de 15,3 kg ha^{-1}. O processo de ajuste

da equação é aquele adotado na seção *Modelos polinomiais de ordem 2 ou superior*, em que a segunda variável é usada no lugar do quadrado da primeira variável.

No caso de existirem muitas variáveis independentes, se o investigador quiser saber quais estão associadas à variável dependente, pode também usar o procedimento REG e algum de seus métodos de seleção de variáveis, para estudar essas associações e escolher as variáveis de importância. Exemplo dessas variáveis seria a produção de grãos (variável dependente), que pode ser influenciada por uma série grande de variáveis, em geral os componentes de produção (variáveis independentes), tais como: no caso da cultura do arroz, número de perfilhos, percentagem de perfilhos férteis, peso de 100 grãos, altura da planta, incidência de doenças e pragas, entre outras; no caso do feijão e de outras leguminosas, número de plantas, número de vagens, número de grãos por vagem, peso de 100 grãos, altura da inserção da primeira vagem, altura da planta, e assim por diante para cada cultura de interesse.

Para exemplificar esse procedimento, serão usados os dados de um experimento do pesquisador Alberto Baeta dos Santos, no qual foram estudados os efeitos de espaçamento entre linhas e a densidade de semeadura na linha. Além da produção, no experimento foram avaliadas as variáveis, a altura de planta, o acamamento de plantas, o número de perfilhos e de panículas. Esses dados serão usados com o procedimento REG, que permite a seleção de variáveis, com opção de nove processos de seleção. Serão mostrados a seguir dois desses processos, primeiro o que obriga a inclusão de todas as variáveis e depois um para seleção das variáveis significativas, segundo critérios de probabilidade definidos pelo pesquisador.

Para o primeiro caso, o programa é:

```
PROC REG;MODEL PROD=PERF PAN ALT ACA/SELECTION=NONE;
```

A opção SELECTION=NONE, que pode ser omitida, implica que todas as variáveis independentes devem ser incluídas no modelo, independentemente de serem ou não significativas ao influenciar a variável dependente.

Para o segundo caso, o programa é:

```
PROC REG;MODEL PROD=PERF PAN ALT ACA/SELECTION=FORWARD sle=0.1 sls=0.05;
```

Nesse caso, três opções foram incluídas no comando MODEL: SELECTION, SLE e SLS. A opção SELECTION indica a escolha de variáveis a serem incluídas no modelo. Nesse caso, FORWARD indica ao programa que escolha a mais significativa delas, depois a segunda, a terceira, e assim por diante. SLE indica o critério de entrada de uma variável no modelo dado em porcentagem de significância, o mesmo para SLS, que é o critério para permanência. No exemplo anterior, a probabilidade de entrada era de 10% e a de permanência 5%. Para maiores detalhes, recomenda-se a leitura dos manuais do SAS. Os resultados obtidos pelos dois programas anteriores são apresentados nas Figuras 15 e 16, respectivamente.

The REG Procedure
Model: MODEL1
Dependent Variable: prod
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1856565	464141	38.96	0.0018
Error	4	47659	11915		
Corrected Total	8	1904223			

Root MSE	109.15426	R-Square	0.9750
Dependent Mean	2750.38889	Adj R-Sq	0.9499
Coeff Var	3.96868		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	74.11835	1629.86411	0.05	0.9659
PERF	1	-13.70934	4.37376	-3.13	0.0350
PAN	1	17.04710	6.57243	2.59	0.0604
ALT	1	33.05098	11.75534	2.81	0.0482
ACA	1	-242.15811	33.70085	-7.19	0.0020

Figura 15. Uso de procedimento REG com a opção SELECTION=NONE.

Modelos de regressão não lineares e o SAS

Evidentemente existem muitos modelos de regressão não lineares além do modelo logístico mostrado ao início deste capítulo. Um deles, que foi empregado com certa frequência nos estudos de fertilidade de solo, e discutido com detalhes por Pimentel-Gomes (2000), é o modelo de Mitscherlich, que tem a seguinte forma:

$$y = A[1 - 10^{-c(x+b)}] \qquad [40]$$

em que A é a produção máxima teórica, c é o coeficiente de eficácia, b é o teor do elemento contido no solo e x é a dose do elemento aplicada no ensaio.

```
                          The REG Procedure
                              Model: MODEL1
                         Dependent Variable: PROD
                         Forward Selection: Step 1
          Variable ACA Entered: R-Square = 0.8003 and C(p) = 26.9242
                              Analysis of Variance
                              Sum of           Mean
Source                 DF     Squares         Square    F Value    Pr > F
Model                   1     1523857        1523857      28.04    0.0011
Error                   7      380366          54338
Corrected Total         8     1904223

                    Parameter     Standard
      Variable       Estimate        Error   Type II SS  F Value  Pr > F
      Intercept   3871.09125    225.44036     16021656   294.85  <.0001
      ACA         -237.32521     44.81503      1523857    28.04  0.0011
                       Bounds on condition number: 1, 1
------------------------
                         Forward Selection: Step 2
          Variable ALT Entered: R-Square = 0.9129 and C(p) = 10.9195
                              Analysis of Variance
                              Sum of           Mean
Source                 DF     Squares         Square    F Value    Pr > F
Model                   2     1738378         869189      31.45    0.0007
Error                   6      165845          27641
Corrected Total         8     1904223

                    Parameter     Standard
      Variable       Estimate        Error   Type II SS  F Value  Pr > F
      Intercept  -1221.91325   1835.22306        12253     0.44  0.5303
      ALT           42.85359     15.38256       214521     7.76  0.0318
      ACA         -231.15455     32.03969      1438735    52.05  0.0004
                  Bounds on condition number: 1.0048, 4.0192
-------------------------
No other variable met the 0.1000 significance level for entry into the
model.
                          The REG Procedure
                              Model: MODEL1
                         Dependent Variable: PROD
                       Summary of Forward Selection
      Variable  Number     Partial     Model
Step  Entered   Vars In    R-Square    R-Square     C(p)    F Value   Pr > F
  1    ACA         1       0.8003      0.8003    26.9242      28.04   0.0011
  2    ALT         2       0.1127      0.9129    10.9195       7.76   0.0318
```

Figura 16. Uso do procedimento REG, com a opção SELECTION=FORWARD, para escolha de variáveis independentes significantes.

O ajuste dessa equação com o SAS é feito com o procedimento NLIN. Para exemplificar seu uso, serão usados os dados médios dos tratamentos citados por Pimentel-Gomes (2000), na seção 7.7 de seu livro. O programa é o seguinte:

```
PROC NLIN METHOD=MARQUARDT BEST=5;
MODEL PROD=A*(1-10**((-C)*(X+B)));
PARMS A=67 C=0.006 B=64.5;
DER.A=1-10**((-C)*(X+B));
DER.C=A*Log(10)*10**((-C)*(X+B))*(X+B);
DER.B=A*Log(10)*C*10**((-C)*(X+B));
```

O resultado aparece na Figura 17.

EQUACAO DE MITSCHERLIC

The NLIN Procedure

Dependent Variable PROD

Grid Search

A	C	B	Sum of Squares
67.0000	0.00600	64.5000	1468.1

The NLIN Procedure

Dependent Variable PROD

Method: Marquardt

Iterative Phase

Iter	A	C	B	Sum of Squares
0	67.0000	0.00600	64.5000	1468.1
1	65.7745	0.00592	59.3630	1319.7
2	65.7765	0.00592	59.3761	1319.7

NOTE: Convergence criterion met.

Estimation Summary

Method	Marquardt
Iterations	2
R	1.48E-6
PPC(C)	4.973E-7
RPC(C)	0.000813
Object	8.885E-6
Objective	1319.657
Observations Read	27
Observations Used	27
Observations Missing	0

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	3	69626.6	23208.9	23.88	<.0001
Residual	24	1319.7	54.9857		
Uncorrected Total	27	70946.3			
Corrected Total	26	3945.4			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
A	65.7765	10.6705	43.7539	87.7992
C	0.00592	0.00448	-0.00333	0.0152
B	59.3761	33.1508	-9.0433	127.8

Approximate Correlation Matrix

	A	C	B
A	1.0000000	-0.9670857	0.8910239
C	-0.9670857	1.0000000	-0.9689791
B	0.8910239	-0.9689791	1.0000000

Figura 17. Resultado do uso do procedimento NLIN, do SAS, para estimação não linear dos parâmetros de uma equação de Mitscherlich.

É necessário ressaltar que esse procedimento, em geral, apresenta um problema um pouco sério, do qual depende muito a convergência do processo. Trata-se da definição dos valores iniciais dos parâmetros a serem estimados. Valores muito discrepantes dos reais podem levar a não convergência do procedimento.

Outro problema reclamado por alguns usuários, não estatísticos, é o da obtenção das primeiras derivadas do modelo em relação aos parâmetros. No caso presente, houve uma convergência rápida do processo, mas deve ser ressaltado que dos três parâmetros apenas um deles, *A*, foi significativo, como se pode ver na Figura 14. As discrepâncias encontradas

nas estimativas dos parâmetros, quando comparadas com as de Pimentel-Gomes (2000), podem ser atribuídas ao fato de que o autor usou as médias das três doses de fósforo para a estimação de A, b e c, enquanto aqui foram usadas as médias de todos os 27 tratamentos, correspondentes às combinações dos três fatores em estudo. A equação aqui estimada é:

$$y = 65{,}7765*[1 - 10^{-0{,}00592+(x + 59{,}3761)}] \quad [41]$$

Capítulo 13

Análise de covariância

É comum em pesquisa a avaliação de diversas variáveis em cada unidade experimental, e ocorre frequentemente a associação entre essas variáveis, o que pode afetar a resposta dos tratamentos ou influenciar a medida do erro experimental, por não serem as variações encontradas devidas aos fatores em estudo, mas a essa correlação entre as variáveis. Para resolver esse problema de influências estranhas aos fatores em estudo, foi desenvolvida a chamada análise de covariância, que associa o estudo de regressão de uma ou mais de uma variável sobre aquela que é objeto da análise de variância.

Como exemplo desse problema, pode-se pensar no caso de produção de grãos e do número de plantas por unidade de área num ensaio de competição de cultivares. Sabe-se que o número de plantas é importante na definição da produtividade de uma cultura. Se esse número é muito variável num ensaio, e se essa variação não é uma resposta aos tratamentos aplicados, as diferenças (ou igualdades) em produção de grãos podem não ser devidas a eles, mas a essa variação no estande.

A correção desse efeito indesejado deve ser efetuada, e em geral se processa por meio do uso de regressão linear. Mas o modelo de regressão pode ser modificado para um modelo quadrático, exponencial, ou por outro que se julgue o mais adequado para o problema em questão. Vários outros usos podem ser encontrados para análise de covariância, mas aqui serão estudados apenas os casos de ajuste de tratamentos e a estimativa de valores perdidos em experimentos. Será dada ênfase ao delineamento de blocos completos por ser ainda o de maior uso na pesquisa e na experimentação em geral, mas o procedimento pode ser utilizado em qualquer desenho experimental. Como exemplo, na última seção será feito um estudo de estimativa de dois valores perdidos em quadrado latino.

O modelo matemático para a análise de covariância

Para a realização da análise de covariância o modelo matemático a ser empregado deve associar o modelo de uma regressão ao modelo do delineamento empregado no

ensaio. Como se viu no capítulo anterior, o modelo matemático para uma regressão linear é dado por $Y_i = a + bX_i$ e, para uma regressão de segundo grau, por $Y_i = b_0 + b_1X_i + b_2X_i^2$, por exemplo. No Capítulo 2 e nos seguintes, foram vistos os modelos para diversos delineamentos, como o completamente ao acaso ($Y_{ij} = m + t_i + e_{ij}$), blocos completos ($Y_{ij} = m + r_j + t_i + e_{ij}$), quadrado latino ($Y_{ijk} = m + c_i + l_j + t_k + e_{ijk}$) e tantos outros.

Numa análise de covariância linear, no delineamento completamente ao acaso, o modelo completo será dado por

$$Y_{ij} = m + t_i + b\,(X_{ij} - \bar{x}) + e_{ij} \qquad [1]$$

em que Y_{ij} representa a medida da variável de interesse no tratamento i na repetição j; m a média geral do ensaio; t_i o efeito do tratamento i; b o coeficiente de regressão de X sobre Y; X_{ij} representa o valor da variável secundária na parcela do tratamento i na repetição j; $\bar{x}$ a média da variável secundária em todo o ensaio; e e_{ij} o erro associado a Y_{ij}. Se o delineamento fosse o de blocos completos, haveria a inclusão do componente r_j, que representaria o efeito do bloco j; ou, no caso de um quadrado latino, haveria a inclusão de dois componentes referentes ao efeito de linha e de coluna. Caso o processo fosse o de covariância de segundo grau, e o delineamento o de um bloco completo, o modelo seria:

$$Y_{ij} = m + r_j + t_i + b1(X_{ij} - \bar{x}) + b_2(X\text{ij} - \bar{x})^2 + e_{ij} \qquad [2]$$

Os procedimentos da análise de covariância serão vistos a seguir para casos particulares que, no entanto, podem facilmente ser generalizados para qualquer delineamento.

Ajuste de tratamentos no delineamento em blocos completos

O quadro de análise de covariância é semelhante ao de uma análise de variância com todas as fontes de variação devidas ao modelo do delineamento adotado; no entanto, a coluna de soma de quadrados é referida como de soma de produtos, pois incluirá, além da variável principal, também a variável secundária e o produto entre elas. Além disso, deverá incluir a análise da regressão (efeito de b) e os correspondentes ajustes devido a essa regressão. Esses ajustes correspondem à perda de um grau de liberdade para o erro e aos valores das somas de quadrados de tratamentos e erro devidos a essa regressão. Isso se explica facilmente pelo fato de o parâmetro b de regressão tomar um grau de liberdade para si, o qual vem do erro, que, portanto, terá sua soma de quadrados reduzida por esse mesmo

efeito, e do ajuste na soma de quadrados de tratamentos, já que as médias de tratamentos serão ajustadas pelo efeito da variável secundária caso a regressão seja significativa.

Um comentário deve ser feito neste ponto: para muitos autores, a análise de variância para a variável secundária deve ser realizada primeiro e, no caso de o efeito de tratamentos ser significativo, a análise de covariância não deveria ser levada à frente, pois um dos possíveis efeitos de tratamentos seria refletido nesta variável secundária. Outros autores, entre eles Steel e Torrie (1981), sugerem que, mesmo nesse caso, a análise de covariância seja feita dependendo das variáveis em estudo e de como esse efeito pode ser interpretado.

A Tabela 1 mostra, esquematicamente, o quadro auxiliar para a análise de covariância para um delineamento em blocos completos.

Tabela 1. Esquema inicial de análise de covariância, no delineamento em blocos completos ao acaso.

Fonte de variação	Grau de liberdade	Soma de produtos			Grau de liberdade ajustados	Soma de quadrados ajustada
		XX	XY	YY		
Blocos	J - 1	R_{xx}	R_{xy}	R_{yy}		
Tratamentos	T - 1	T_{xx}	T_{xy}	T_{yy}	T - 1	$T_{yy}(aj) = S_{yy}(aj)$ - $E_{yy}(aj)$
Erro	$(J$ - 1$)(T$ - 1$)$	E_{xx}	E_{xy}	E_{yy}	$(J$ - 1$)(T$ - 1$)$ - 1	$E_{yy}(aj) = E_{yy}$ - Sb
Regressão	-	-	-	-	1	$Sb = (E_{xy})^2/E_{xx}$
Tratamentos + erro	$J(T$ - 1$)$	S_{xx}	S_{xy}	S_{yy}	$J(T$ - 1$)$ - 1	$S_{yy}(aj) = S_{yy}$ - $(S_{xy})^2/S_{xx}$
Total	JT - 1	t_{xx}	t_{xy}	t_{yy}		

As somas de quadrados (XX e YY) para total (t_{xx} e t_{yy}), blocos (R_{xx} e R_{yy}), tratamentos (T_{xx} e T_{yy}) e erro (E_{xx} e E_{yy}) são calculadas segundo as fórmulas descritas no Capítulo 3. As somas de produtos para essas mesmas fontes se calculam segundo as fórmulas a seguir:

$$SPtotal = t_{xy} = \sum X_{ij} Y_{ij} - \frac{(\sum X_{ij})(\sum Y_{ij})}{IJ} \quad [3]$$

$$SPBlo\cos = R_{xy} = \frac{\sum Bx_j By_j}{I} - \frac{(\sum Bx_j)(By_j)}{IJ} \quad [4]$$

e

$$SPTratamentos = T_{xy} = \frac{\sum Tx_j Ty_j}{I} - \frac{(\sum Tx_j)(Ty_j)}{IJ} \quad [5]$$

Como sempre, a soma de produtos para erro se calcula por diferença:

$$SPErro = E_{xy} = t_{xy} - R_{xy} - T_{xy} \quad [6]$$

O coeficiente de regressão linear b se calcula da seguinte forma:

$$b = E_{xy} / E_{xx} \quad [7]$$

Como consequência dos ajustes na análise, o quadro final da análise de covariância é aquele mostrado na Tabela 2.

Tabela 2. Esquema final de análise de covariância, no delineamento em blocos completos ao acaso.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	$J - 1$	R_{yy}		
Tratamentos (ajustados)	$T - 1$	$T_{yy}(aj)$	$SQTr(aj)/GLTr$	$QMTr(aj)/QME(aj)$
Regressão	1	Sb	Sb	$Sb/QME(aj)$
Erro (ajustado)	$(J - 1)(T - 1) - 1$	$E_{yy}(aj)$	$SQE(aj)/GLE$	
Total	$JT - 1$	t_{yy}		

Finalmente, há a necessidade de ajuste das médias dos tratamentos, caso haja significância da regressão. Esse ajuste se processa da forma

$$\overline{T}_i(aj) = \overline{T}_i - b * (\overline{X}_i - \overline{X}) \quad [8]$$

em que $\bar{T}_i$ é a média aritmética do tratamento i; b o coeficiente de regressão; $\bar{X}_i$ a média aritmética da variável secundária no tratamento i; e $\bar{X}$ a média aritmética da variável secundária no experimento.

Para uma melhor compreensão do processo, um exemplo dessa análise é apresentado a seguir, utilizando-se o mesmo ensaio empregado no Capítulo 3, mas com a restrição no número de tratamentos para apenas sete, para maior facilidade de cálculos. As informações pertinentes são apresentadas na Tabela 3.

Para a análise de covariância, a variável produtividade será considerada a variável principal (Y); e o estande, a variável secundária (X), associada à produtividade sobre a qual se medirá o efeito de regressão linear.

Para a montagem do quadro inicial de uma análise de covariância (Tabela 4), as somas de produtos e quadrados para as fontes definidas no modelo matemático do delineamento experimental para as duas variáveis se calculam da forma usual, como foi visto em cada um dos capítulos referentes a esses desenhos de ensaios e no estudo de regressão e correlação. Para maior claridade de entendimento, seguem alguns desses cálculos.

$$SPtotal = t_{xy} = (1.587,5 * 95 + 2.412,5 * 103 + ... + 1.637,5 * 92 - \frac{56.187,5 * 2.820}{28} = 26.253,57$$

$$SPBlocos = R_{xy} = \frac{13.600,0 * 719 + ... 13.637,5 * 658}{7} - \frac{56.187,5 * 2.820}{28} = 7.096,43$$

$$SPTratamentos = T_{xy} = \frac{8.012,5 * 379 + ... + 8.937,5 * 389}{4} - \frac{56.187,5 * 2.820}{28} = -4.205,80$$

A partir desses dados, pode-se calcular então o valor do coeficiente de regressão que é dado por:

$$b = \frac{(23.362,95)}{1.567,00} = 14,90$$

Tabela 3. Dados de produtividade em kg ha^{-1} (PROD) e o estande em plantas por área útil para sete tratamentos de um ensaio de competição de cultivares em blocos completos ao acaso, com quatro repetições.[1]

Tratamento	Bloco 1		Bloco 2		Bloco 3		Bloco 4		Soma	
	Produtiv.	Estande	Produtiv.	Estande	Produtiv.	Estande	Produtiv.	Estande	Produtiv.	Estande
CNFP 8018	1.587,5	95	2.412,5	103	2.100,0	91	1.912,5	90	8.012,5	379
CNFP 8019	2.012,5	124	2.075,0	114	2.087,5	106	1.862,5	88	8.037,5	432
CNFP 8020	1.975,0	106	1.787,5	116	2.225,0	111	1.950,0	101	7.937,5	434
CNFP 8021	1.837,5	92	2.137,5	101	2.050,0	90	2.587,5	118	8.612,5	401
CNFP 8022	1.825,0	101	2.100,0	115	1.375,0	98	1.450,0	92	6.750,0	406
CNFP 8023	2.250,0	98	2.062,5	112	1.950,0	92	1.637,5	77	7.900,0	379
G. Brilhante	2.112,5	103	2.137,5	108	2.450,0	86	2.237,5	92	8.937,5	389
Soma	13.600,0	719	14.712,5	769	14.237,5	674	13.637,5	658	56.187,5	2.820

[1] Dados cedidos pela pesquisadora Maria José del Peloso, da Embrapa Arroz e Feijão.

Tabela 4. Análise inicial de covariância, no delineamento em blocos completos ao acaso, para o ensaio de competição de cultivares de feijão.[1]

Fonte de variação	Grau de liberdade	Soma de produtos			Grau de liberdade ajustados	Soma de quadrados ajustada
		X	XY	Y		
Blocos	3	1.066,00	7.096,43	120.954,24		
Tratamentos	6	790,71	-4.205,80	706.752,23	6	899.420,78
Erro	18	1.567,00	23.362,95	1.277.756,70	17	929.430,32
Regressão					1	348.326,38
Tratamentos + erro	24	2.357,71	19.157,14	1.984.508,93	23	1.828.851,10
Total	27	3.423,71	26.253,57	215.463,17		

[1] Dados cedidos pela pesquisadora Maria José del Peloso, da Embrapa Arroz e Feijão.

A soma de quadrados de regressão é, então,

$$SQ\,Regressão = \frac{(23.362,95)^2}{1.567,00} = 348.326,38$$

o que leva a uma soma de quadrados de erro, ajustada para o efeito de regressão de

$$E_{yy}(aj) = 1.277.756,70 - 348.326,38 = 929.430,32$$

Por sua vez, para que se obtenha a soma de quadrados de tratamentos ajustada, primeiro se deve calcular a soma de quadrados de tratamento + erro, ajustada, segundo a fórmula dada na Tabela 1, que é:

$$S_{yy}(aj) = 1.984.508,93 - (19.157,14)2/2.357,71 = 1.828.851,10$$

Daí segue que a soma de quadrados de tratamentos ajustada é então igual a:

$$T_{yy}(aj) = 1.828.851,10 - 929.430,32 = 899.420,78$$

A tabela final da análise de covariância está dada na Tabela 5.

Tabela 5. Esquema final da análise de covariância para o ensaio de competição de cultivares de feijão.[1]

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	3	120.954,24		
Tratamentos (aj)	6	899.420,78	149.903,46	2,74 (0,047)
Regressão	1	348.326,38	348.326,38	6,37 (0,022)
Erro (aj)	17	929.430,32	54.672,37	
Total	27	215.463,17		

[1] Dados cedidos pela pesquisadora Maria José del Peloso, da Embrapa Arroz e Feijão.

Para o cálculo das médias de tratamentos ajustadas, a montagem de um quadro é interessante e facilita muito o pesquisador. Para o experimento que está sendo analisado, o valor médio de estande para todo o ensaio ($\bar{X}$) é igual a 100,71, e o valor de b é igual a 14,90. O resultado final encontra-se na Tabela 6.

Tabela 6. Ajuste das médias de tratamentos como função da análise de covariância. Média da variável secundária igual a 100,71 e coeficiente de regressão (b) igual a 14,90.

Tratamento	Estande médio ($\bar{X}_i$)	Produtividade média ($\bar{T}_i$)	($\bar{X}_i$ - $\bar{X}$)	Produtividade média ajustada $\bar{T}_i(aj) = \bar{T}_i - b * (\bar{X}_i - \bar{X})$
CNFP 8018	94,75	2.003,13	-5,96	2.091,99
CNFP 8019	108,00	2.009,38	7,29	1.900,82
CNFP 8020	108,50	1.984,38	7,79	1.868,37
CNFP 8021	100,25	2.153,13	-0,46	2.160,04
CNFP 8022	101,50	1.687,50	0,79	1.675,79
CNFP 8023	94,75	1.975,00	-5,96	2.063,87
G. Brilhante	97,25	2.234,38	-3,46	2.285,99

Uso do SAS

Para fazer a análise de covariância com o SAS, o procedimento adequado é o GLM, com uso da opção LMEANS, que resultará nas médias dos tratamentos ajustados.

O programa para essa análise é:

```
PROC GLM;CLASS BLOCO TRAT;
MODEL PROD=BLOCO TRAT ST;
LSMEANS TRAT;
```

Para exemplificar o seu uso, empregam-se os mesmos dados da Tabela 3, onde a variável corresponde a Estande. O resultado se vê na Figura 1.

```
                          The GLM Procedure
                        Class Level Information
Class       Levels  Values
BLOCO            4  1 2 3 4
TRAT             7  CNFP.8018 CNFP.8019 CNFP.8020 CNFP.8021 CNFP.8022
                    CNFP.8023 G.BRILHANTE

                       Number of observations    28
                          The GLM Procedure
Dependent Variable: PROD
                                Sum of
 Source                DF       Squares    Mean Square   F Value   Pr > F
 Model                 10    1176032.744     117603.274      2.15   0.0791
 Error                 17     929430.426      54672.378
 Corrected Total       27    2105463.170

               R-Square     Coeff Var      Root MSE     PROD Mean
               0.558562      11.65205      233.8213      2006.696

 Source             DF     Type I SS    Mean Square   F Value   Pr > F
 BLOCO               3    120954.2411     40318.0804      0.74   0.5441
 TRAT                6    706752.2321    117792.0387      2.15   0.0997
 ST                  1    348326.2705    348326.2705      6.37   0.0218

 Source             DF    Type III SS    Mean Square   F Value   Pr > F
 BLOCO               3    116916.8483     38972.2828      0.71   0.5577
 TRAT                6    899420.9131    149903.4855      2.74   0.0474
 ST                  1    348326.2705    348326.2705      6.37   0.0218

                          The GLM Procedure
                          Least Squares Means
                          TRAT            PROD LSMEAN
                          CNFP.8018        2092.04860
                          CNFP.8019        1900.74976
                          CNFP.8020        1868.29509
                          CNFP.8021        2160.04720
                          CNFP.8022        1675.78551
                          CNFP.8023        2063.92360
                          G.BRILHANTE      2286.02524
```

Figura 1. Análise de covariância em um delineamento em blocos completos ao acaso com o uso do SAS.

Quando se comparam os resultados representados na Tabela 5 e na Figura 1, verifica-se que a soma de quadrados dos blocos vem do SS1 do SAS, e a soma de quadrados dos tratamentos ajustados da SS3 do SAS. As pequenas diferenças nas médias ajustadas vêm do grau de precisão matemática das estimativas das médias aritméticas da produção e do estande, e principalmente do coeficiente de regressão. Caso fosse necessário conhecer o valor do coeficiente de regressão, seria preciso explicitar ao modelo a opção SOLUTION. Nesse caso, o programa seria então:

```
PROC GLM;CLASS BLOCO TRAT;
MODEL PROD=BLOCO TRAT ST/SOLUTION;
LSMEANS TRAT;
```

Essa opção leva o SAS a imprimir a matriz de informação (Figura 2), na qual se vê que a estimativa obtida pelo SAS do coeficiente de regressão é igual a 14,9093468, ligeiramente maior que o valor de 14,90, anteriormente calculado:

Parameter		Estimate		Error	t Value	Pr > |t|
Intercept		826.0644955	B	552.7277241	1.49	0.1534
BLOCO	1	-135.2814506	B	135.1671968	-1.00	0.3309
BLOCO	2	-82.8482135	B	156.1848068	-0.53	0.6027
BLOCO	3	51.6357788	B	125.7098287	0.41	0.6864
BLOCO	4	0.0000000	B	.	.	.
TRAT	CNFP.8018	-193.9766330	B	165.9947303	-1.17	0.2587
TRAT	CNFP.8019	-385.2754781	B	177.1105475	-2.18	0.0440
TRAT	CNFP.8020	-417.7301514	B	178.1907244	-2.34	0.0315
TRAT	CNFP.8021	-125.9780404	B	166.2834855	-0.76	0.4591
TRAT	CNFP.8022	-610.2397239	B	167.2315363	-3.65	0.0020
TRAT	CNFP.8023	-222.1016330	B	165.9947303	-1.34	0.1985
TRAT	G.BRILHANTE	0.0000000	B	.	.	.
ST		14.9093468		5.9067622	2.52	0.0218

Figura 2. Matriz de informação acerca dos parâmetros estimáveis numa análise de covariância obtida com o SAS.

Estimativa de um valor perdido

Outra aplicação da análise de covariância diz respeito ao cálculo de valores perdidos num experimento. Para exemplificar o uso do procedimento, será considerado o caso de um dado perdido num ensaio em quadrado latino. Nesse caso, a variável secundária não é medida no ensaio, é uma variável dicotômica, pois apresenta apenas dois valores: zero (0) nas parcelas em que os dados foram colhidos e um (1) na parcela cujo dado não está disponível. O coeficiente de regressão (*b*) estimado, em valor absoluto, é a estimativa do dado não disponível. Para o caso do delineamento em quadrado latino, o modelo matemático é dado por:

$$Y_{ijk} = m + c_j + l_k + t_i + bX_{ijk} + e_{ijk} \quad [9]$$

Se o delineamento fosse o de blocos completos, dever-se-ia apenas incluir o componente bX_{ij} ao modelo usual daquele desenho. O mesmo procedimento pode ser empregado em outros delineamentos.

As médias dos tratamentos serão as médias aritméticas dos dados disponíveis, com exceção do tratamento com valor perdido, pois, aos dados disponíveis, deve-se primeiro adicionar o valor calculado pela análise de covariância.

O quadro de análise de covariância é similar àquele apresentado na seção anterior para o caso de uma variável secundária real e deve ser estruturado segundo o modelo do delineamento experimental adotado.

O esquema mostrado na Tabela 7 considera um delineamento em quadrado latino.

A análise de variância, portanto, será igual àquela representada na Tabela 8.

Tabela 7. Quadro inicial de análise de covariância, no delineamento em quadrado latino.

Fonte de variação	Grau de liberdade	Soma de produtos			Grau de liberdade ajustados	Soma de quadrados ajustada
		XX	XY	YY		
Linhas	K - 1	L_{xx}	L_{xy}	L_{yy}		
Colunas	K - 1	C_{xx}	C_{xy}	C_{yy}		
Tratamentos	K - 1	T_{xx}	T_{xy}	T_{yy}	K - 1	$T_{yy}(aj) = S_{yy}(aj) - E_{yy}(aj)$
Erro	$(K - 1)(K - 2)$	E_{xx}	E_{xy}	E_{yy}	$(K - 1)(K - 2) - 1$	$E_{yy}(aj) = E_{yy} - Sb$
Regressão					1	$Sb = (E_{xy})^2/E_{xx}$
Tratamentos + erro	$(K - 1)^2$	S_{xx}	S_{xy}	S_{yy}	$(K - 1)^2 - 1$	$S_{yy}(aj) = S_{yy} - (S_{xy})^2/S_{xx}$
Total	K^2 - 1	t_{xx}	t_{xy}	t_{yy}		

Tabela 8. Esquema final de análise de covariância no delineamento quadrado latino.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Linhas	K - 1	L_{yy}		
Colunas	K - 1	C_{yy}		
Tratamentos (ajustados)	K - 1	$T_{yy}(aj)$	$SQTr(aj)/GLTr$	$QMTr(aj)/QME(aj)$
Regressão	1	Sb	Sb	$Sb/QME(aj)$
Erro (ajustado)	$(K - 1)(K - 2) - 1$	$E_{yy}(aj)$	$SQE(aj)/GLE$	
Total	K^2 - 1	t_{yy}		

No caso de mais de uma parcela perdida, devem ser criadas tantas variáveis secundárias quantos são os valores não disponíveis. O procedimento de ajuste nesse caso é um pouco mais complicado (STEEL; TORRIE, 1981).

Exemplo de aplicação de estimação de valor perdido em quadrado latino

Como exemplo, será utilizado um ensaio de controle de insetos-praga de arroz conduzido por Evane Ferreira, pesquisador da Embrapa Arroz e Feijão, já mostrado como exemplo para o estudo de aditividade em quadrados latinos, mas agora com a variável produção de grãos por área útil da parcela. Assume-se como perdida a parcela da linha 6, coluna 2, correspondente ao tratamento 1. Na Tabela 9, mostram-se os dados colhidos no experimento (o valor "perdido" está substituído por 0), os valores da variável auxiliar (X) que só contêm zeros e um valor 1 na correspondência ao dado perdido.

Para o preenchimento do quadro inicial da análise de covariância, é necessário calcular as somas de quadrados das duas variáveis e sua soma de produtos. Nesse caso, simplificam-se bastante os cálculos para a variável auxiliar e para o produto entre elas, pois a variável auxiliar apresenta apenas um valor diferente de zero que é aquele referente ao dado perdido.

Para a variável auxiliar:

$SQLinha_{xx} = SQColuna_{xx} = SQTratamentos_{xx} = 1/6 - 1/36 = 0,1389$

$SQTotal_{xx} = 1 - 1/36 = 0,9722$

$SQErro_{xx} = SQTotal_{xx} - SQLinha_{xx} - SQColuna_{xx} - SQTratamentos_{xx}$

$SQErro_{xx} = 0,9722 - 3 * 0,1389 = 0,5555$

Para a soma de produtos entre as variáveis:

$SPLinha_{xy} = (3.114,5 * 1)/6 - (21.157,6 * 1)/36 = -68,6278$

$SPColuna_{xy} = (2.523,3 * 1)/6 - (21.157,6 * 1)/36 = -167,1611$

$SPTratamento_{xy} = (3.057,2 * 1)/6 - (21.157,6 * 1)/36 = -78,1778$

$SPTotal_{xy} = -(21.157,6 * 1)/36 = -587,7111$

$SPErro_{xx} = -587,7111 - (-68,6278 - 167,1611 - 78,1778) =$

$SPErro_{xy} = -273,7444$

Tabela 9. Dados de um experimento em quadrado latino com um dado perdido (0) e a variável auxiliar X.[1]

Linha		Coluna						Soma	Total dos tratamentos
		1	2	3	4	5	6		
	Tratamento	5	2	3	4	1	6		1
1	Y_{ijk}	570,3	533,7	651,9	610,9	720,8	722,1	3.809,7	3.057,2
	X	0	0	0	0	0	0	0	1
	Tratamento	4	6	2	1	3	5		2
2	Y_{ijk}	467,6	549,1	594,6	585,2	654,9	688,9	3.540,3	3.601,3
	X	0	0	0	0	0	0	0	0
	Tratamento	3	4	5	2	6	1		3
3	Y_{ijk}	687,1	524,5	580,3	586,2	663,6	781,2	3.822,9	3.839,6
	X	0	0	0	0	0	0	0	0
	Tratamento	1	5	6	3	2	4		4
4	Y_{ijk}	507,5	480,2	465,3	646,8	719,5	778,2	3.597,5	3.575,7
	X	0	0	0	0	0	0	0	0
	Tratamento	6	3	1	5	4	2		5
5	Y_{ijk}	385,4	435,8	462,5	664,0	628,8	696,2	3.272,7	3.629,7
	X	0	0	0	0	0	0	0	0
	Tratamento	2	1	4	6	5	3		6
6	Y_{ijk}	471,1	0,0	565,7	668,6	646,0	763,1	3.114,5	3.454,1
	X	0	1	0	0	0	0	1	0
Soma	Y_{ijk}	3.089,0	2.523,3	3.320,3	3.761,7	4.033,6	4.429,7	21.157,6	
	X	0	1	0	0	0	0	1	

[1] Dados cedidos pelo pesquisador Evane Ferreira, da Embrapa Arroz e Feijão.

Finalmente, as somas de quadrados para a variável principal calculam-se da forma usual, como visto no Capítulo 4.

$$SQLinha_{yy} = \frac{3.809,7^2 + ... + 3.114,5^2}{6} - \frac{(21.157,6)^2}{36} = 67.907,46$$

$$SQColuna_{yy} = \frac{3.089,0^2 + ... + 4.429,7^2}{6} - \frac{(21.157,6)^2}{36} = 394.762,38$$

$$SQTratamento_{yy} = \frac{3.057,2^2 + ... + 3.454,1^2}{6} - \frac{(21.157,6)^2}{36} = 57.030,24$$

$$SQTotal_{yy} = 570,3^2 + ... + 763,1^2 - \frac{(21.157,6)^2}{36} = 720.778,54$$

$$SQErro_{yy} = 720.778,54 - 67.907,46 - 394.762,38 - 57.030,24 = 201.078,46$$

Dessa forma, pode-se montar a Tabela 10.

Tabela 10. Análise de covariância no delineamento em quadrado latino.

Fonte de variação	Grau de liberdade	Soma de produtos			Grau de liberdade ajustados	Soma de quadrados ajustada
		XX	XY	YY		
Linhas	5	0,1389	-68,6278	67.907,46		
Colunas	5	0,1389	-167,1611	394.762,38		
Tratamentos	5	0,1389	-78,1778	57.030,24	5	13.572,16
Erro	20	0,5555	-273,7444	201.078,45	19	66.193,61
Regressão					1	134.884,8
Tratamentos + erro	25	0,6944	-351,9222	258.108,69	24	79.765,77
Total	35	0,9722	-587,7111	720.778,54		

O valor do coeficiente de regressão, que nesse caso, em valor absoluto, estima o valor perdido, é igual ao quociente entre a soma de produto do erro (XY) pela soma de quadrados do erro (XX) ou:

$$b = \frac{-273,7444}{0,5555} = -492,74$$

A análise de covariância é então vista na Tabela 11.

Como afirmado anteriormente, as médias de tratamentos são as médias aritméticas obtidas no experimento; no entanto, no caso do tratamento com valor perdido, deve-se ao total obtido somar-se o valor do coeficiente de regressão, em valor absoluto, antes de se efetuar o cálculo da média. A Tabela 12 mostra esses valores.

Deve ser destacado que o valor estimado por covariância é o mesmo que teria sido estimado se fosse empregada a fórmula para cálculo de um valor perdido em quadrado latino dada no Capítulo 4. A maior diferença é que, nesse caso, a soma de quadrados ajustada para tratamentos já está corrigida para possíveis vícios, como assinalam Steel e Torrie (1981).

Tabela 11. Quadro final de análise de covariância no delineamento quadrado latino.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Linhas	5	67.907,46		
Colunas	5	394.762,38		
Tratamentos (ajustados)	5	13.572,16	2.714,43	0,78 (0,576)
Regressão	1	134.884,80	134.884,80	38,72 (< 0,001)
Erro (ajustado)	19	66.193,61	3.483,87	
Total	35	720.778,54		

Tabela 12. Cálculo da média ajustada para o tratamento com dado perdido e das médias aritméticas dos demais tratamentos.

Tratamento	Total do ensaio	Valor perdido estimado	Total geral	Média total geral/6
1	3.057,2	492,74	3.549,94	591,7
2	3.601,3	0,00	3.601,3	600,2
3	3.839,6	0,00	3.839,6	639,9
4	3.575,7	0,00	3.575,7	596,0
5	3.629,7	0,00	3.629,7	605,0
6	3.454,1	0,00	3.454,1	575,7

Uso do SAS na análise de covariância com estimação do valor da parcela perdida

Para este caso de covariância, o SAS deve ser usado também com o procedimento GLM e com a opção SOLUTION no comando MODEL. Porém o comando para as médias é o MEANS, para posterior ajuste da média do tratamento com dado perdido, manualmente, pois o comando LSMEANS produz médias ajustadas diferentes daquelas obtidas no exemplo. O programa, portanto, é:

```
PROC GLM;CLASS LINHA COLUNA TRAT;
MODEL PROD=LINHA COLUNA TRAT VAR_AUX/SOLUTION;
MEANS TRAT;
```

A variável VAR_AUX no programa corresponde ao nome dado à variável auxilia x, que contém somente 0 e 1 e o resultado aparece na Figura 3.

The GIM Procedure
Class Level Information

Class	Levels	Values
LINHA	6	1 2 3 4 5 6
COLUNA	6	1 2 3 4 5 6
TRAT	6	1 2 3 4 5 6

Number of observations 36

The GIM Procedure

Dependent Variable: PROD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	654584.9209	40911.5576	11.74	<.0001
Error	19	66193.6147	3483.8745		
Corrected Total	35	720778.5356			

R-Square	Coeff Var	Root MSE	PROD Mean
0.908164	10.04309	59.02435	587.7111

Source	DF	Type I SS	Mean Square	F Value	Pr > F
LINHA	5	67907.4589	13581.4918	3.90	0.0134
COLUNA	5	394762.3822	78952.4764	22.66	<.0001
TRAT	5	57030.2422	11406.0484	3.27	0.0267
VAR_AUX	1	134884.8376	134884.8376	38.72	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
LINHA	5	33997.3129	6799.4626	1.95	0.1327
COLUNA	5	249714.5189	49942.9038	14.34	<.0001
TRAT	5	13572.1591	2714.4318	0.78	0.5769
VAR_AUX	1	134884.8376	134884.8376	38.72	<.0001

Parameter		Estimate	Standard Error	t Value	Pr > |t|
Intercept		712.3766667 B	40.32131049	17.67	<.0001
LINHA	1	33.7433333 B	36.54429105	0.92	0.3674
LINHA	2	-11.1566667 B	36.54429105	-0.31	0.7635
LINHA	3	35.9433333 B	36.54429105	0.98	0.3377
LINHA	4	-1.6233333 B	36.54429105	-0.04	0.9650
LINHA	5	-55.7566667 B	36.54429105	-1.53	0.1436
LINHA	6	0.0000000 B	.	.	.
COLUNA	1	-223.4500000 B	34.07772712	-6.56	<.0001
COLUNA	2	-235.6100000 B	36.54429105	-6.45	<.0001
COLUNA	3	-184.9000000 B	34.07772712	-5.43	<.0001
COLUNA	4	-111.3333333 B	34.07772712	-3.27	0.0041
COLUNA	5	-66.0166667 B	34.07772712	-1.94	0.0677
COLUNA	6	0.0000000 B	.	.	.
TRAT	1	15.9733333 B	36.54429105	0.44	0.6670
TRAT	2	24.5333333 B	34.07772712	0.72	0.4803
TRAT	3	64.2500000 B	34.07772712	1.89	0.0748
TRAT	4	20.2666667 B	34.07772712	0.59	0.5590
TRAT	5	29.2666667 B	34.07772712	0.86	0.4011
TRAT	6	0.0000000 B	.	.	.
VAR_AUX		-492.7400000	79.18948176	-6.22	<.0001

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The GIM Procedure

Level of		PROD		VAR_AUX	
TRAT	N	Mean	Std Dev	Mean	Std Dev
1	6	509.533333	277.849712	0.16666667	0.40824829
2	6	600.216667	94.620197	0.00000000	0.00000000
3	6	639.933333	109.068321	0.00000000	0.00000000
4	6	595.950000	106.781623	0.00000000	0.00000000
5	6	604.950000	76.929078	0.00000000	0.00000000
6	6	575.683333	131.832916	0.00000000	0.00000000

Figura 3. Análise de covariância em um delineamento em quadrado latino com uma parcela perdida e uso de variável auxiliar.

Verifica-se pelos resultados que as somas de quadrados para linha e coluna são obtidas na soma de quadrados tipo SS1, e a soma de quadrados de tratamentos ajustada para a perda de parcela por variável auxiliar está na soma de quadrados do tipo SS3. As médias dos tratamentos sem dado perdido são as médias aritméticas, enquanto a média do tratamento 1, ajustada para o dado perdido, é igual a sua média aritmética mais a média aritmética da variável auxiliar, multiplicada pelo valor absoluto da estimativa da variável auxiliar obtida com a opção SOLUTION. Nesse caso, a média do tratamento 1 é

$$509{,}533333 + 0{,}6666667 * 492{,}74 = 591{,}6567$$

que é o valor obtido na Tabela 12.

Capítulo 14

Experimentos fatoriais fracionados

Análise de variância e uso para estudo de superfícies de resposta

Quando o número de fatores ou o número de níveis de cada fator é elevado, o total de tratamentos a ser ensaiado é extremamente grande, o que torna, em geral, impraticável a sua aplicação completa. Nesses casos, a melhor solução é usar apenas parte dos tratamentos, eliminando muitos deles do experimento. Além disso, muitas vezes o interesse maior do pesquisador ao montar um experimento está no ajuste de equações de regressão, e não numa possível comparação entre tratamentos, seja somente para medir a associação entre variáveis seja para definir pontos de solução ótimos, sejam eles de máximo ou de mínimo.

Evidentemente que os tratamentos que serão utilizados nos ensaios não podem ser definidos de qualquer forma, há a necessidade de critérios de escolha para descarte ou manutenção de tratamentos. Em primeiro lugar, esses critérios devem permitir que o tratamento testemunha esteja incluído, pois qualquer medida de progresso só pode ser efetuada a partir da uma dada origem de conhecimento; e em segundo, devem proporcionar que os tratamentos mantidos estimem os efeitos principais e as interações de primeira ordem, pelo menos. No caso das superfícies de resposta, como os modelos mais empregados são os de segunda ordem, isto é, os modelos contêm apenas os efeitos lineares e quadráticos de cada fator, e as combinações lineares entre pares dos fatores em estudo, isso também é possível quando os tratamentos estimam aqueles efeitos.

Existem vários tipos de delineamentos experimentais que permitem o estudo de superfícies de resposta com um número reduzido de tratamentos, mesmo quando o número de fatores não é baixo ou com muitos níveis para cada um deles. Conagin e Jorge (1976),

Malheiros e Perecin (1983) e Zimmermann e Conagin (1986) fizeram vários estudos com experimentos fatoriais do tipo $(1/5)5^3$ procurando determinar os melhores planejamentos de níveis para ajuste de superfície de resposta de segundo grau. Os autores verificaram que o uso de apenas um experimento desse tipo, em estudos de adubação de plantas, leva a uma grande porcentagem de pontos de sela. Isso pode ser corrigido pela utilização de grupos de experimentos, pela escolha dos níveis de cada elemento, e pela melhoria na precisão do experimento indicada pelo coeficiente de variação.

Construção e análise de delineamentos fatoriais fracionados

O procedimento mais indicado na construção de um ensaio fatorial fracionado é o uso dos contrastes de definição de confundimento, como explicado em capítulo anterior, associado à escolha de uma das soluções de resíduo $mod\ m$, de forma que uma fração deste experimento é empregada, e essa fração está associada à base do fatorial. Por exemplo, em fatoriais 3^k, pode ser utilizado um terço, um nono ou qualquer fração de uma potência de 3; em fatoriais 5^k, poderiam ser usados um quinto ou um vinte e cinco avos, por exemplo, dos tratamentos totais possíveis. Essas frações, representadas por 3^{k-j} ou 5^{k-j}, podem ou não ser repetidas, dependendo do número de tratamentos a serem mantidos e do tipo de informação que o pesquisador está buscando.

Numa análise de variância comum, na qual o interesse maior seria o de conhecer os possíveis efeitos de cada fator e de suas interações, esse processo apresenta o inconveniente da existência de aliases, isto é, um contraste de tratamentos pode estimar mais de um efeito, o que implica que os contrastes de definição de confundimento devem ser bem escolhidos para que não ocorra que dois efeitos principais sejam aliases, ou que efeitos principais não sejam aliases de interações de primeira ordem, ou ainda que efeitos de interesse de estudo não sejam aliases. Recorde-se que, quando um contraste estima dois efeitos simultaneamente, cabe ao pesquisador definir qual deles deve ser considerado estimado e qual não deve ser considerado. De uma maneira geral efeitos que representem um número menor de fatores são os mais indicados para essa escolha, pois deles, em geral, espera-se um efeito muito mais acentuado. Para uma melhor compreensão desses fatos, alguns exemplos serão apresentados a seguir.

Metade de um fatorial 2^k

A metade dos tratamentos de um fatorial na base 2 é a maior fração que se pode trabalhar. Para isso, é necessário definir apenas um contraste de confundimento. Supondo

que k seja igual a 4, o ensaio conterá 16 tratamentos, portanto cada metade terá apenas 8 tratamentos. Um ensaio deste tipo possibilita a estimação de 15 efeitos, quatro principais (A, B, C e D), seis interações simples (AB, AC, AD, BC, BD e CD), quatro interações duplas (ABC, ABD, ACD e BCD) e uma tripla (ABCD), e qualquer um deles permite um procedimento de confundimento.

O uso da interação de maior ordem (ABCD) como contraste de definição de confundimento, cuja equação associada é $x_1 + x_2 + x_3 + x_4 = r\ (mod\ 2)$, com apenas duas soluções, $r = 0$ e $r = 1$, é a que permite o estudo de todos os efeitos principais e de metade das interações simples. Como o valor de k é par (4), a solução $r = 0$ contém tratamentos cujo produto dos coeficientes é +1 e a solução $r = 1$ aqueles cujo produto de coeficientes é -1. Dessa forma, para $r = 0$ os tratamentos serão: (1), ab, ac, ad, bc, bd, cd e abcd. Por sua vez, para $r = 1$ tem-se: a, b, c, d, abc, abd, acd e bcd. Qualquer uma dessas duas combinações de tratamentos pode ser empregada como solução de meio fatorial 2^4, mas, como a solução $r = 0$ contém a testemunha, ela deveria ser preferencialmente a metade utilizada.

Para a determinação dos aliases, deve-se proceder ao produto dos demais efeitos pelo efeito de definição de confundimento e aplicar aos expoentes dos produtos obtidos o residual $mod\ m$. Nesse caso, tem-se:

$$A \rightarrow A*ABCD = A^2BCD = BCD$$

$$B \rightarrow B*ABCD = AB^2CD = ACD$$

$$C \rightarrow C*ABCD = ABC^2D = ABD$$

$$D \rightarrow D*ABCD = ABCD^2 = ABC$$

$$AB \rightarrow AB*ABCD = A^2B^2CD = CD$$

$$AC \rightarrow AC*ABCD = A^2BC^2D = BD$$

$$AD \rightarrow AD*ABCD = A^2BCD^2 = BC$$

Os resultados mostram que cada efeito principal tem como álias uma interação de segunda ordem, e as interações de primeira ordem têm como aliases outra interação de primeira ordem. Numa análise de variância, poder-se-ia, portanto, estimar os efeitos principais e algumas das interações de primeira ordem, mas cabe ao investigador decidir entre os aliasses quais são os efeitos de maior interesse.

Esse tipo de experimento, no entanto, não permite o estudo de superfícies de resposta por estarem todos os fatores presentes em apenas dois níveis. Para uma análise de variância realmente informativa, que permita a estimação das interações simples além dos

efeitos principais, o exemplo anterior mostra que, pelo menos, cinco fatores devem estar envolvidos, pois, nesse caso, os efeitos principais seriam aliases das interações tríplices, e as interações simples aliases das interações duplas, ao usar a interação de ordem maior como contraste de definição. Para fazer uma análise de variância, é necessária a existência de repetições para que seja possível a estimação do erro ou variância, o que não ocorria em muitos casos discutidos em capítulos anteriores, com ou sem confundimento, em que experimentos com uma réplica eram analisados e as interações de ordem maior eram empregadas como estimadores do erro. A vantagem desse tipo de ensaio que se apresenta agora é que, se houver perda de uma ou de poucas parcelas, a análise continua sendo possível, na maioria dos casos, se o interesse é o estudo de superfícies de resposta. Nesse caso, muitas vezes uma única aplicação de cada tratamento é suficiente.

Análise de variância de um fatorial 2^{4-1}

Para efetuar a análise de variância de um fatorial fracionado, adotam-se os mesmos procedimentos já explicados para um fatorial completo, apenas com o cuidado de explicitar apenas aquelas fontes de variação que são realmente estimáveis. A definição de quais são as interações estimadas deve ser feita pelo pesquisador, em razão de seu conhecimento a respeito dos fatores em estudo ou do interesse que ele possa ter em algumas das interações estimáveis. Esse conhecimento ou interesse deriva de experimentos ou estudos anteriores.

No exemplo que se vai discutir a seguir, são estimáveis os efeitos principais, e metade das interações de primeira ordem como visto no item *Metade de um fatorial* 2^k. O experimento foi realizado em 1981, na fazenda sede da Embrapa Arroz e Feijão, com quatro fatores (cultivares, calcário, gesso e P) no delineamento em blocos completos com três repetições. Para os fatores abióticos, o nível zero correspondeu à não aplicação do insumo; e para cultivares, correspondeu à mais utilizada pelos agricultores da época (IAC 47). O nível 1 de cada um dos fatores era: IAC 165 (cultivar), 5.000 kg ha^{-1} de calcário, 1.500 kg ha^{-1} de gesso e 50 kg ha^{-1} de P_2O_5. Os tratamentos, em número de oito, foram: 0000, 1001, 1010, 1100, 0011, 0101, 0110 e 1111. A variável analisada é o número médio de perfilhos por planta. A Tabela 1 contém os dados obtidos.

A Tabela 2 mostra os contrastes necessários para estimação dos efeitos principais e das interações simples, consideradas de interesse pelo pesquisador, que são aquelas referentes às interações de cultivares com os fatores abióticos. A tabela contém também os efeitos de cada uma das fontes, obtidos pela soma dos produtos entre os coeficientes dos contrastes e os totais de tratamentos.

Para a análise de variância, devem-se primeiro calcular as somas de quadrados de blocos, tratamentos, total e erro, da forma tradicional, e depois as somas de quadrados dos

Tabela 1. Número médio de perfilhos por planta em um ensaio fatorial fracionado (2^{4-1}) 2 x 2 x 2 x 2, em blocos ao acaso, com três repetições.

Tratamento	Bloco 1	Bloco 2	Bloco 3	Total
0000	5,67	5,72	5,32	16,71
0011	5,25	6,10	6,95	18,30
0101	5,80	4,30	5,45	15,55
0110	5,65	4,20	3,95	13,80
1001	4,02	4,37	4,10	12,49
1010	3,85	4,20	3,00	11,05
1100	5,35	4,35	4,80	14,50
1111	4,45	4,75	5,10	14,30
Total	40,04	37,99	38,67	116,70

Tabela 2. Contrastes para estimação dos efeitos principais e das interações de interesse no caso de um fatorial fracionado (2^{4-1}).

	0000	0011	0101	0110	1001	1010	1100	1111	Efeitos
Cultivar (Cult)	-1	-1	-1	-1	1	1	1	1	-12,02
Calcário (Cal)	-1	-1	1	1	-1	-1	1	1	-0,40
Gesso (G)	-1	1	-1	1	-1	1	-1	1	-1,80
Fósforo (P)	-1	1	1	-1	1	-1	-1	1	4,58
Cult*Cal	1	1	-1	-1	-1	-1	1	1	10,92
Cult*G	1	-1	1	-1	-1	1	-1	1	-1,48
Cult*P	1	-1	-1	1	1	-1	-1	1	-2,10
Total dos tratamentos	16,71	18,30	15,55	13,80	12,49	11,05	14,50	14,30	

efeitos principais e das interações de interesse. Conforme apresentado no Capítulo 3, em blocos ao acaso, tem-se:

$$SQTotal = (5{,}67^2 + 5{,}72^2 + ... + 4{,}74^2 + 5{,}10^2) - \frac{116{,}70^2}{24} = 18{,}2783$$

$$SQBlocos = \frac{40{,}04^2 + 37{,}99^2 + 38{,}67^2}{8} - \frac{116{,}70^2}{24} = 0{,}2726$$

$$SQTratamentos = \frac{16,71^2 + 18,30^2 + ... + 14,30^2}{3} - \frac{116,70^2}{24} = 12,2793$$

Portanto, com a soma de quadrados de erro dada pela diferença entre o total e tratamentos e blocos, tem-se:

$$SQErro = 18,2783 - 0,2726 - 12,2793 = 5,7264$$

Para as somas de quadrados dos efeitos principais e interações, deve-se lembrar de que elas são dadas pela fórmula

$$SQContraste = \frac{(\sum c_i T_i)^2}{r * \sum c_i^2}$$

em que o valor entre parênteses no numerador representa o efeito medido do contraste, e o denominador apresenta o produto entre o número de repetições e a soma de quadrados dos coeficientes do contraste. No caso atual, essa soma é sempre igual a oito, portanto o denominador para qualquer um dos contrastes é igual a 24. Dessa forma, pode-se então calcular as somas de quadrados faltantes, isto é:

$$SQCultivar = \frac{[(-1*16,71) + (-1*18,30) + ... + (1*14,30)]^2}{3*8} = \frac{(-12,02)^2}{24} = 6,0200$$

$$SQCalcario = \frac{[(-1*16,71) + (-1*18,30) + ... + (1*14,30)]^2}{3*8} = \frac{(-0,40)^2}{24} = 0,0067$$

e assim sucessivamente até

$$SQ(CultxP) = \frac{[(1*16,71) + (-1*18,30) + ... + (1*14,30)]^2}{3*8} = \frac{(-2,10)^2}{24} = 0,1838$$

A análise de variância encontra-se na Tabela 3.

Por interesse do pesquisador responsável, as interações simples estimadas foram apenas aquelas que envolviam cultivares, ou seja, o interesse era verificar como a troca de cultivares (tradicional por melhorada) seria afetada pelo uso de corretivos do solo e de um adubo químico (P). Nesse caso, os aliases são: gesso x P, calcário x P e calcário x gesso, respectivamente para as três interações mostradas na Tabela 3. Isso implica que a interação

Tabela 3. Análise de variância de um fatorial fracionado (2^{4-1}), em blocos ao acaso, com três repetições. Efeitos principais e interações simples de interesse.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Blocos	2	0,2726	0,1363	0,33 (0,722)
Cultivares (Cult)	1	6,0200	6,0200	14,72 (0,002)
Calcário (Cal)	1	0,0067	0,0067	0,02 (0,900)
Gesso (G)	1	0,1350	0,1350	0,33 (0,575)
Fósforo (P)	1	0,8740	0,8740	2,14 (0,166)
Cult*Cal	1	4,9686	4,9686	12,15 (0,004)
Cult*G	1	0,0913	0,0913	0,22 (0,644)
Cult*P	1	0,1838	0,1838	0,45 (0,514)
Erro	14	5,7264	0,4090	
Total	23	18,2783		

significativa de cultivar x calcário também poderia ser interpretada como um efeito significativo da interação gesso x P. Para saber qual dessas duas interações está sendo estimada, existe a necessidade de realizar um experimento que meça ambas as interações.

Uso do SAS na análise de variância de um fatorial 2^{4-1}

Para obter essa análise com o SAS, pode ser usado o procedimento ANOVA. No comando MODEL, devem-se especificar os fatores principais e as interações que são do interesse do pesquisador estimar. Para o caso dos dados mostrados na Tabela 1, o programa é:

```
PROC ANOVA;CLASS CULT CA G P REP; MODEL PERF=REP CULT CA G P
CULT*CA CULT*G CULT*P;
MEANS CULT CA G P CULT*CA CULT*G CULT*P;
```

O resultado se vê na Figura 1.

Um terço de um fatorial na base 3

Como foi visto anteriormente, ao se estudar o confundimento em fatoriais de base 3, cada contraste de definição apresenta três soluções possíveis, quais sejam, as de resultados iguais a 0, 1 ou 2.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
CULT	2	0 1
CA	2	0 1
G	2	0 1
P	2	0 1
REP	3	1 2 3

Number of observations 24

The ANOVA Procedure

Dependent Variable: PERF

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	12.55189167	1.39465463	3.41	0.0198
Error	14	5.72635833	0.40902560		
Corrected Total	23	18.27825000			

R-Square	Coeff Var	Root MSE	PERF Mean
0.686712	13.15272	0.639551	4.862500

Source	DF	Anova SS	Mean Square	F Value	Pr > F
REP	2	0.27257500	0.13628750	0.33	0.7222
CULT	1	6.02001667	6.02001667	14.72	0.0018
CA	1	0.00666667	0.00666667	0.02	0.9002
G	1	0.13500000	0.13500000	0.33	0.5747
P	1	0.87401667	0.87401667	2.14	0.1659
CULT*CA	1	4.96860000	4.96860000	12.15	0.0036
CULT*G	1	0.09126667	0.09126667	0.22	0.6439
CULT*P	1	0.18375000	0.18375000	0.45	0.5136

The ANOVA Procedure

Level of CULT	N	PERF Mean	PERF Std Dev
0	12	5.36333333	0.85589223
1	12	4.36166667	0.61792664

Level of CA	N	PERF Mean	PERF Std Dev
0	12	4.87916667	1.13214566
1	12	4.84583333	0.61587275

Level of G	N	PERF Mean	PERF Std Dev
0	12	4.93750000	0.68141731
1	12	4.78750000	1.08860315

Level of P	N	PERF Mean	PERF Std Dev
0	12	4.67166667	0.87845767
1	12	5.05333333	0.90028615

Level of CULT	Level of CA	N	PERF Mean	PERF Std Dev
0	0	6	5.83500000	0.62631462
0	1	6	4.89166667	0.82789895
1	0	6	3.92333333	0.48467171
1	1	6	4.80000000	0.37947332

Level of CULT	Level of G	N	PERF Mean	PERF Std Dev
0	0	6	5.37666667	0.55694404
0	1	6	5.35000000	1.14061387
1	0	6	4.49833333	0.49845428
1	1	6	4.22500000	0.73942545

Level of CULT	Level of P	N	PERF Mean	PERF Std Dev
0	0	6	5.08500000	0.79891802
0	1	6	5.64166667	0.88736501
1	0	6	4.25833333	0.80586393
1	1	6	4.46500000	0.40618961

Figura 1. Análise de variância de um fatorial 2^{4-1}, usando o procedimento ANOVA do SAS.

No caso de um fatorial 3^3, viu-se que é possível confundir dois graus de liberdade da interação dupla, e que isso pode ser feito de quatro formas diferentes, as quais constituem os grupos W, X, Y e Z. Associando-se as equações de definição de confundimento com a representação de efeitos, o grupo W seria definido por AB^2C^2, o X por AB^2C, o Y por ABC^2 e o grupo Z por ABC.

Para o grupo W, os três grupos de nove tratamentos definidos pelas soluções $r = 0$, $r = 1$ ou $r = 2$ são: {000, 012, 021, 101, 110, 122, 202, 211 e 220}, {002, 011, 020, 100, 112, 121, 201, 210 e 222} e {001, 010, 022, 102, 111, 120, 200, 212 e 221}. O uso de qualquer um desses terços mostra que os aliases seriam:

$$A \rightarrow \quad A*AB^2C^2 = A^2B^2C^2 = ABC$$

$$A \rightarrow \quad A*(AB^2C^2)^2 = A^3B^4C^4 = BC$$

$$B \rightarrow \quad B*AB^2C^2 = AB^3C^2 = AC^2$$

$$B \rightarrow \quad B*(AB^2C^2)^2 = A^2B^5C^4 \ = ABC^2$$

$$C \rightarrow \quad C*AB^2C^2 = AB^2C^3 = AB^2$$

$$C \rightarrow \quad C*(AB^2C2)^2 = A^2B^4C^5 = AB2C$$

$$AB \rightarrow AB*AB^2C^2 = A^2B^3C^2 = AC$$

$$AB \rightarrow AB*(AB^2C^2)^2 = A^3B^5C^4 = BC^2$$

Portanto, os oito graus de liberdade seriam estimados pelos seguintes quatro grupos de aliases: A, BC e ABC; B, AC^2 e ABC^2; C, AB^2 e AB^2C e AB, AC e BC^2. Vê-se claramente que os efeitos principais e parte de uma interação simples podem ser estimados.

Processo semelhante se aplica aos casos de maior número de fatores envolvidos que adicionalmente apresentam maiores possibilidades de estimação e de ajuste de superfícies de resposta. Para três fatores, como foi visto anteriormente, apenas nove tratamentos – portanto oito graus de liberdade – são testados, e uma superfície de resposta, nesse caso, que teria em seu modelo a necessidade de estimação de dez parâmetros, é evidentemente impossível. Como no caso de fatoriais de base 2, para que se faça uma simples análise de variância, também é necessário que o experimento apresente repetições.

Análise de variância de um fatorial 3^{3-1}

Para exemplificar o uso de um terço de um fatorial 3^3, serão usados os blocos que continham o tratamento 000, no exemplo apresentado na Tabela 16 do Capítulo 10 e que se reproduz na Tabela 4.

Tabela 4. Dados de produção em kg ha^{-1} obtidos em um ensaio com 1/3 de um fatorial 3 x 3 x 3.[1]

Tratamento	Repetição 1	Repetição 2	Repetição 3	Totais
0	1200	950	1300	3450
12	3900	4200	4550	12650
21	4200	3950	4500	12650
102	1800	2600	3150	7550
111	3350	4500	4250	12100
120	4700	5600	4850	15150
201	1100	2000	2250	5350
210	2600	3750	4500	10850
222	4000	3650	4700	12350
Totais	26850	31200	34050	92100

[1] Dados cedidos pelo pesquisador Nand Kumar Fagéria, da Embrapa Arroz e Feijão.

Para a análise de variância desse tipo de experimento, tem-se como fontes de variação: repetição, fator A (Zn), fator B (P), fator C (C), parte da interação A*B (Zn*P), erro e total. As somas de quadrado são obtidas da forma tradicional, exceto para a parte da interação A*B, não confundida, que se obtém por diferença entre as somas de quadrado de tratamentos e a soma das somas de quadrados dos fatores principais (A, B e C). Para o exemplo que se está apresentando, as somas de quadrados então são dadas por:

$$SQTotal = (1.200^2 + 3.900^2 + ... + 4.500^2 + 4.700^2) - \frac{(92.100)^2}{27} = 360.265.000 - 314.163.333,33$$

$$SQTotal = 46.101.666,67$$

$$SQ\,Rep = \frac{26.850^2 + 31.200^2 + 34.050^2}{9} - 314.163.333,33 = 2.921.666,67$$

$$SQTrat = \frac{3.450^2 + 12.650^2 + ... + 12.350^2}{3} - 314.163.333,33 = 40.420.000,00$$

$$SQErro = 46.101.666,67 - 2.921.666,67 - 40.420.000,00 = 2.760.000,00$$

A soma de quadrados de tratamentos deve ser estudada segundo seus componentes. Dessa forma, tem-se, então,

$$SQZn = \frac{28.750^2 + 34.800^2 + 28.550^2}{9} - 314.163.333,33 = 2.803.888,89$$

$$SQP = \frac{16.350^2 + 35.600^2 + 40.150^2}{9} - 314.163.333,33 = 35.470.555,56$$

$$SQZn = \frac{29.450^2 + 30.100^2 + 32.550^2}{9} - 314.163.333,33 = 593.888,89$$

e finalmente

$$SQZn * P(não\,conf) = 40.420.000 - 2.803.888,89 - 35.470.555,56 - 593.888,89 = 1.551.666,66$$

o que permite agora elaborar a Tabela 5, com a análise de variância desse fatorial fracionado.

Tabela 5. Análise de variância de um fatorial fracionado (1/3 de 3 x 3 x 3).

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > F)
Repetição	2	2.921.666,67		
Zn	2	2.803.888,89	1.401.944,44	8,13 (0,004)
P	2	35.470.555,56	17.735.277,78	102,81 (< 0,001)
C	2	593.888,89	296.944,44	1,72 (0,210)
Zn*P (não confundida)	2	1.551.666,66	775.833,33	4,50 (0,028)
Erro	16	2.760.000,00	172.500,00	
Total	26	46.101.666,67		

Os dois graus de liberdade de cada um dos fatores principais poderiam ser estudados independentemente pelo ajuste de contrastes de efeitos linear e quadrático, mas isso é impossível para a parte não confundida da interação simples, pelo simples fato de que

ela não corresponde a nenhuma das quatro possíveis divisões, que, conforme foi visto no Capítulo 10, seriam: (linear x linear), (linear x quadrática), (quadrática x linear) e (quadrática x quadrática).

Uso do SAS na análise de variância de um fatorial 3^{3-1}

Para o caso de confundimento parcial de interação, o procedimento adequado é o GLM com a opção SS1, como se pode ver no programa a seguir. Seus resultados encontram-se na Figura 2.

```
PROC GLM;CLASS REP ZN P C;
MODEL PROD=REP ZN P C ZN*P/SS1;
MEANS ZN P C;
```

```
                    The GLM Procedure
                  Class Level Information
Class         Levels    Values
REP                3    1 2 3
Zn                 3    0 1 2
P                  3    0 1 2
C                  3    0 1 2
 Number of observations    27
                    The GLM Procedure
Dependent Variable: PROD
                             Sum of
Source                DF     Squares    Mean Square   F Value   Pr > F
Model                 10  43341666.67     4334166.67     25.13   <.0001
Error                 16   2760000.00      172500.00
Corrected Total       26  46101666.67

              R-Square     Coeff Var      Root MSE     PROD Mean
              0.940132      12.17583      415.3312      3411.111

Source                DF      Type I SS    Mean Square   F Value   Pr > F
REP                    2    2921666.67     1460833.33      8.47   0.0031
Zn                     2    2803888.89     1401944.44      8.13   0.0037
P                      2   35470555.56    17735277.78    102.81   <.0001
C                      2     593888.89      296944.44      1.72   0.2103
Zn*P                   2    1551666.67      775833.33      4.50   0.0282
                    The GLM Procedure
 Level of                 ------------PROD------------
 Zn              N              Mean           Std Dev
 0               9        3194.44444       1550.69429
 1               9        3866.66667       1218.09277
 2               9        3172.22222       1234.43082
 Level of                 ------------PROD------------
 P               N              Mean           Std Dev
 0               9        1816.66667        752.080448
 1               9        3955.55556        646.357314
 2               9        4461.11111        587.780666
 Level of                 ------------PROD------------
 C               N              Mean           Std Dev
 0               9        3272.22222       1792.71287
 1               9        3344.44444       1256.59372
 2               9        3616.66667        946.37466
```

Figura 2. Análise de variância de um terço de um fatorial 3^3, com três repetições, confundimento parcial de interação (grupo Z), usando o procedimento GLM do SAS.

Fatoriais fracionados de base 4 ou maior

Com quatro ou mais níveis para cada fator, já é possível a realização de análise para ajuste de superfície de resposta de segundo grau. No entanto, para uma simples análise de variância na qual o interesse maior está centrado no estudo dos efeitos principais e de todas as interações simples, continua sendo necessário o uso de pelo menos cinco fatores, ou seja, com quatro ou menos fatores, somente parte das interações simples pode ser estimada; e, dependendo do contraste de definição, talvez parte das interações de ordem maior possa também ser estimada.

Como exemplo de construção de um fatorial fracionado com um quinto dos tratamentos de um fatorial 5 x 5 x 5, usa-se o modelo desenvolvido por Conagin e Jorge (1976), e também por Jorge e Conagin (1976), em que se demonstram as possíveis maneiras de distribuição dos 25 tratamentos em cinco blocos, uma das quais é a que se vê na Tabela 6.

Tabela 6. Distribuição dos tratamentos que compõe um quinto de fatorial 5 x 5 x 5.

Bloco	Tratamento				
1	000	234	413	142	321
2	111	340	024	203	432
3	222	401	130	314	043
4	333	012	241	420	104
5	444	123	302	031	210

Fonte: Conagin e Jorge (1976).

Evidentemente é possível distribuir esses 25 tratamentos em um único bloco, e repetir esse bloco várias vezes, de modo que se componha um experimento em blocos completos, com apenas essa fração de tratamentos. O empecilho seria o tamanho do bloco, definido pelo tamanho das parcelas a serem usadas pelo pesquisador. Essa opção é necessária se o investigador deseja fazer uma análise de variância completa. A solução de Conagin e Jorge (1976) é útil para o caso de ajustes de superfícies de resposta, tópico que será discutido a seguir.

Uso de fatoriais fracionados e composto central rotatório no ajuste de superfícies de resposta

Como discutido até agora para o ajuste de superfícies de resposta de segundo grau, existe a necessidade de que pelo menos três níveis de cada fator sejam testados no

experimento, e que o número total de tratamentos seja maior que o número de parâmetros do modelo de regressão a ser estimado, de forma que se permita também a estimação do erro, isso quando o pesquisador vai usar apenas uma repetição ou uma fração dos tratamentos de um fatorial. No caso de dois fatores, o modelo de superfície de resposta de segundo grau possui seis parâmetros; com três fatores, esse número já sobe para 10 e, com quatro fatores, já serão 15, incluídos aí o ponto de intersecção (b_0) como se vê na Tabela 7. Percebe-se por esses números que o menor experimento fatorial que pode ser usado é o 3 x 3, completo (nove tratamentos), para dois fatores, um quarto de 4 x 4 x 4 (16 tratamentos) para 3 fatores e um terço de um 3 x 3 x 3 x 3 (27 tratamentos) para o caso de quatro fatores. O delineamento fracionado proposto por Conagin e Jorge (1976), visto no item *Fatoriais fracionados de base 4 ou maior*, para três fatores contém 25 tratamentos, mais de 50% maior que um terço de um 4 x 4 x 4 e 25% maior que o composto central rotatório que será visto a seguir.

Tabela 7. Modelos completos de superfície de resposta de segundo grau para dois, três e quatro fatores.

Número de fatores	Modelo de superfície de resposta de segundo grau
2	$Y_{ij} = b_0 + b_1A + b_2B + b_{11}A^2 + b_{22}B^2 + b_{12}AB$
3	$Y_{ij} = b_0 + b_1A + b_2B + b_3C + b_{11}A^2 + b_{22}B^2 + b_{33}C^2 + b_{12}AB + b_{13}AC + b_{23}BC$
4	$Y_{ij} = b_0 + b_1A + b_2B + b_3C + b_4D + b_{11}A^2 + b_{22}B^2 + b_{33}C^2 + b_{44}D^2 + b_{12}AB + b_{13}AC + b_{14}AD + b_{23}BC + b_{24}BD + b_{34}CD$

Outra opção para a escolha dos tratamentos que permitam a estimação de uma superfície de resposta é a adoção de delineamentos do tipo central composto. Esse delineamento apresenta N unidades experimentais, e é descrito da forma

$$N = 2^n + 2n + m$$

em que n é o numero de fatores em estudo e m o número de parcelas que recebem o tratamento central (000...); 2^n representa um fatorial completo a dois níveis, $2n$ os pontos axiais em que uma das ordenadas é igual $a \pm \alpha$ e as demais iguais a 0. Para que um desenho desses seja também rotatório, é necessário que se tenha o valor de $\alpha = (2^n)^{1/4}$. Para dois

fatores o valor de α é igual à raiz quadrada de 2 (1,414), para três fatores $\alpha = (8)^{1/4}$ ou 1,682 e assim sucessivamente.

Verifica-se por esses números que o delineamento central composto requer sempre cinco níveis de cada um dos fatores ($-\alpha$, -1, 0, 1, α) e que, comparados com os fatoriais base 3, os desenhos centrais compostos rotatórios têm sempre um número menor de tratamentos, exceto no caso de dois fatores. A Tabela 8 mostra o número de unidades experimentais para os delineamentos centrais compostos rotatórios de dois até cinco fatores.

Tabela 8. Número de pontos fatoriais ou axiais, número de unidades com ponto central e número total de parcelas para estudos de superfície de resposta de segundo grau, no delineamento composto central rotatório com dois a seis fatores.

Número de fatores (n)	Valor α	Ponto de fatorial 2^n	Ponto axial	Unidade com o ponto central	Total de unidades experimentais
2	1,414	4	4	5	13
3	1,682	8	6	6	20
4	2,000	16	8	7	31
5	2,378	32	10	10	52
6	2,828	64	12	15	91

Para dois e três fatores, na Tabela 9 descrevem-se todos os pontos a serem utilizados para o estudo de superfícies de resposta nesse delineamento.

Tabela 9. Identificação das combinações de fatores a serem usadas em compostos centrais rotatórios com dois e três fatores. Para dois fatores α = 1,414 e para três fatores α = 1,682.

	2 fatores	3 fatores
Pontos do fatorial 2^n	(-1,-1), (1,-1), (-1,1), (1,1)	(-1,-1,-1), (-1,-1,1), (-1,1,-1), (1,-1,-1) (-1,1,1), (1,-1,1), (1,1,-1) (1,1,1)
Pontos axiais	$(-\alpha,0)$, $(\alpha,0)$, $(0,-\alpha)$, $(0,\alpha)$	$(-\alpha,0,0)$, $(0,-\alpha,0)$, $(0,0,-\alpha)$, $(\alpha,0,0)$, $(0,\alpha,0)$, $(0,0,\alpha)$
Ponto central	(0,0)[1]	(0,0,0)[2]

[1] Repetido cinco vezes.

[2] Repetido seis vezes.

Para a análise de variância e estimação dos parâmetros de uma superfície de resposta, utilizam-se os mesmos princípios discutidos no Capítulo 12. Deve ser destacado que, nesse caso, a determinação de pontos de máximo e mínimo implica o estudo dos valores e vetores próprios.

Exemplo de uso de fatorial fracionado em estudos de superfície de resposta

Como exemplo de aplicação, serão usados os dados gerados por Zimmermann e Conagin (1986), para um ensaio com uma só repetição de um fatorial fracionado correspondente a 1/5 de um 5 x 5 x 5, cujos dados são apresentados na Tabela 10. O modelo é aquele descrito nas seções anteriores para uma superfície de segundo grau com três fatores, e contém, portanto, dez parâmetros, o que leva a uma análise de variância cujo erro possui 15 graus de liberdade.

Tabela 10. Dados de produção de grãos de terras altas.

N	P	K	Produção (kg ha^{-1})
1	1	1	1.605,77
2	2	2	1.917,00
3	3	3	2.025,72
4	4	4	1.959,55
5	5	5	1.730,59
2	3	4	2.060,93
3	4	5	1.782,17
4	5	1	2.044,60
5	1	2	1.932,94
1	2	3	1.888,83
3	5	2	1.837,20
4	1	3	1.964,88
5	2	4	1.851,14
1	3	5	2.087,41
2	4	1	1.961,96
4	2	5	2.026,54
5	3	1	2.035,03
1	4	2	1.959,44
2	5	3	2.069,28
3	1	4	1.982,09
5	4	3	1.979,45
1	5	4	1.964,41
2	1	5	1.716,11
3	2	1	1.981,46
4	3	2	2.202,02
	Total		48.566,52

Fonte: dados simulados por Zimmermann e Conagin (1986).

Na Tabela 11 está descrito o esquema da análise de variância em que se vê que os dez parâmetros dão origem aos nove graus de liberdade correspondentes aos efeitos lineares (3 GL), quadráticos (3 GL) e interação linear entre os fatores (3 GL).

Tabela 11. Análise de variância para uma superfície de resposta de segundo grau, três fatores.

Fonte de variação	Grau de liberdade (GL)	Soma de quadrados (SQ)	Quadrado médio (QM)	F (Pr > f)
N	1	5.384,94	5.384,94	0,56 (ns)
P	1	15.005,36	15.005,36	1,57 (ns)
K	1	7.259,64	7.259,64	0,76 (ns)
N*N	1	16.359,69	16.359,69	1,72 (ns)
P*P	1	84.635,05	84.635,05	8,88 (0,01)
K*K	1	35.652,02	35.652,02	3,74 (ns)
N*P	1	51.107,59	51.107,59	5,36 (0,03)
N*K	1	53.581,23	53.581,23	5,62 (0,03)
P*K	1	4.162,89	4.162,89	0,44 (ns)
Erro	15	142.969,57	9.531,30	
Total	24	416.117,98		

A soma de quadrado total é obtida da forma tradicional, com o uso da fórmula

$$SQTotal = \sum Y_{ijk}^2 - \frac{(\sum Y_{ijk})^2}{n}$$

que para os dados tem o seguinte valor:

$$SQTotal = (1.605,77^2 + 1.917,00^2 + ... + 2.202,02^2) - \frac{(48.566,52)^2}{25} = 416.117,98$$

Neste caso, o modelo de regressão é aquele mostrado na Tabela 7 para três fatores, qual seja:

$$Y_{ijk} = b_0 + b_1 N_i + b_2 P_j + b_3 K_k + b_{11} N_i^2 + b_{22} P_j^2 + b_{33} K_k^2 + b_{12} NP_{ij} + b_{13} NK_{ik} + b_{23} PK_{jk} + \varepsilon_{ijk}$$

No sistema de equações para a solução dessa superfície de resposta, em forma matricial, a matriz X possui dimensões 25 x 10, de forma que a inversa, de $X'X$, tem dimensão 10 x 10. Na matriz X, cada coluna corresponde a um dos parâmetros da equação e é dada por:

$$X = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 4 & 4 \\
1 & 3 & 3 & 3 & 9 & 9 & 9 & 9 & 9 & 9 \\
1 & 4 & 4 & 4 & 16 & 16 & 16 & 16 & 16 & 16 \\
1 & 5 & 5 & 5 & 25 & 25 & 25 & 25 & 25 & 25 \\
1 & 2 & 3 & 4 & 4 & 9 & 16 & 6 & 8 & 12 \\
1 & 3 & 4 & 5 & 9 & 16 & 25 & 12 & 15 & 20 \\
1 & 4 & 5 & 1 & 16 & 25 & 1 & 20 & 4 & 5 \\
1 & 5 & 1 & 2 & 25 & 1 & 4 & 5 & 10 & 2 \\
1 & 1 & 2 & 3 & 1 & 4 & 9 & 2 & 3 & 6 \\
1 & 3 & 5 & 2 & 9 & 25 & 4 & 15 & 6 & 10 \\
1 & 4 & 1 & 3 & 16 & 1 & 9 & 4 & 12 & 3 \\
1 & 5 & 2 & 4 & 25 & 4 & 16 & 10 & 20 & 8 \\
1 & 1 & 3 & 5 & 1 & 9 & 25 & 3 & 5 & 15 \\
1 & 2 & 4 & 1 & 4 & 16 & 1 & 8 & 2 & 4 \\
1 & 4 & 2 & 5 & 16 & 4 & 25 & 8 & 20 & 10 \\
1 & 5 & 3 & 1 & 25 & 9 & 1 & 15 & 5 & 3 \\
1 & 1 & 4 & 2 & 1 & 16 & 4 & 4 & 2 & 8 \\
1 & 2 & 5 & 3 & 4 & 25 & 9 & 10 & 6 & 15 \\
1 & 3 & 1 & 4 & 9 & 1 & 16 & 3 & 12 & 4 \\
1 & 5 & 4 & 3 & 25 & 16 & 9 & 20 & 15 & 12 \\
1 & 1 & 5 & 4 & 1 & 25 & 16 & 5 & 4 & 20 \\
1 & 2 & 1 & 5 & 4 & 1 & 25 & 2 & 10 & 5 \\
1 & 3 & 2 & 1 & 9 & 4 & 1 & 6 & 3 & 2 \\
1 & 4 & 3 & 2 & 16 & 9 & 4 & 12 & 8 & 6
\end{bmatrix}$$

A matriz Y, com dimensões 25 x 1, corresponde à ultima coluna da Tabela 10, o sistema completo para estimação do vetor β, de parâmetros da equação, que é dado por $(X'X)^{-1}X'Y$, fornece as seguintes estimativas de parâmetros:

$$\beta = \begin{bmatrix} 1.070,02 \\ 197,60 \\ 275,59 \\ 172,35 \\ -12,16 \\ -31,57 \\ -15,86 \\ -15,65 \\ -22,44 \\ -7,30 \end{bmatrix}, \text{ em que } X'Y = \begin{bmatrix} 48.566,52 \\ 146.218,45 \\ 146.565,74 \\ 145.097,08 \\ 536.274,93 \\ 536.994,78 \\ 529.037,08 \\ 438.466,48 \\ 433.599,23 \\ 435.964,73 \end{bmatrix} \text{ e}$$

$$(X'Y)^{-1} = \begin{bmatrix} 25 & 75 & 75 & 75 & 275 & 275 & 275 & 225 & 225 & 225 \\ 75 & 275 & 225 & 225 & 1.125 & 825 & 825 & 825 & 825 & 675 \\ 75 & 225 & 275 & 225 & 825 & 1.125 & 825 & 825 & 675 & 825 \\ 75 & 225 & 225 & 275 & 825 & 825 & 1.125 & 675 & 825 & 825 \\ 275 & 1.125 & 825 & 825 & 4.895 & 3.025 & 3.025 & 3.375 & 3.375 & 2.505 \\ 275 & 825 & 1.125 & 825 & 3.025 & 4.895 & 3.025 & 3.375 & 2.485 & 3.375 \\ 275 & 825 & 825 & 1.125 & 3.025 & 3.025 & 4.895 & 2.505 & 3.375 & 3.375 \\ 225 & 825 & 825 & 675 & 3.375 & 3.375 & 2.505 & 3.025 & 2.505 & 2.485 \\ 225 & 825 & 675 & 825 & 3.375 & 2.485 & 3.375 & 2.505 & 3.025 & 2.505 \\ 225 & 675 & 825 & 825 & 2.505 & 3.375 & 3.375 & 2.485 & 2.505 & 3.025 \end{bmatrix}^{-1}$$

A soma de quadrados dos parâmetros, com nove graus de liberdade, é dada por $\beta' X'Y - \hat{Y}'11'Y/n$; a soma de quadrados total, com 24 graus de liberdade, é obtida por meio da operação matricial $Y'Y - Y'11'Y/n$; e a soma de quadrados de erro por diferença entre esses valores. Para a obtenção da soma de quadrados dos efeitos lineares e quadráticos de cada um dos fatores da equação, o uso de contrastes ortogonais é admitido. No entanto, essa técnica não é exata para os efeitos das interações entre os fatores. Nesse caso, o fatorial em estudo é incompleto, portanto os contrastes ortogonais para as interações não são perfeitamente estimáveis, como se discutiu no Capítulo 12.

Após a obtenção das estimativas do modelo de superfície de resposta, deve-se determinar o tipo de resposta obtido, máximo, mínimo ou ponto de sela, e onde ocorre esse ponto de solução ótima. O processo consiste na obtenção das derivadas parciais

em relação às variáveis do modelo, e a resolução do sistema de equações assim obtidas. As segundas derivadas permitem definir se a solução obtida tem um máximo (todas são negativas), mínimo (todas são positivas) ou ponto de sela (valores positivos e negativos). A obtenção das raízes do sistema de equações define onde ocorreu o ponto de solução definido pela segunda derivada. A partir da construção das matrizes dos coeficientes desse sistema de equações, podem ser obtidos os vetores e os valores próprios que indicarão os mesmos resultados anteriores.

A análise de variância desse modelo é dada na Tabela 11.

Com os valores dados na matriz β, constrói-se o modelo de superfície de resposta:

$$Prod_{ijk} = 1.070,02 + 197,6N_i + 275,59P_j + 172,35K_k - 12,16N^2_i - 31,57P^2_{ik} - 15,86K^2_{jk} - 15,65NP_{ij} - 22,44NK_{ik} - 7,30PK_{jk}$$

As derivadas parciais com relação às variáveis N, P e K, resultam no seguinte sistema de equações:

$$197,60 - 24,32N - 15,65P - 22,44K = 0$$

$$275,59 - 63,14P - 15,65N - 7,30K = 0$$

$$172,35 - 31,72K - 22,44N - 7,30P = 0$$

ou

$$24,32N + 15,65P + 22,44K = 197,60$$

$$15,65N + 63,14P + 7,30K = 275,59$$

$$22,44N + 7,30P + 31,72K = 172,35$$

Os valores de N, P e K, que dão origem ao ponto estacionário, são obtidos ao ser solucionado esse sistema de equações formado com as primeiras derivadas. Adotando-se o processo de substituição, tem-se que: $N = 5,3286$; $P = 2,9298$ e $K = 0,9887$. Para definir se o ponto estacionário é um ponto de máximo, de mínimo ou um ponto de sela, deve-se construir uma matriz, M, que contenha na sua diagonal os valores de $b11$, $b22$ e $b33$ e fora das diagonais a metade dos valores de $b12$, $b13$ e $b23$. As raízes características dessa matriz, se negativas, indicam um ponto de máximo; se positivas, um ponto de mínimo; e, se houver valores positivos e negativos, um ponto de sela. Para o exemplo que se está estudando, a matriz M é igual a

$$M = \begin{bmatrix} -12{,}16 & -7{,}82 & -11{,}22 \\ -7{,}82 & -31{,}56 & -3{,}65 \\ -11{,}22 & -3{,}65 & -15{,}86 \end{bmatrix}$$

cujas raízes características são, respectivamente, -8,5988; -81,0564 e -148,6716, todas negativas, o que indica que o ponto estacionário é um ponto de máximo.

Uso do SAS para análise de superfícies de resposta

No caso de superfícies de resposta, o SAS possui um procedimento adequado para esse tipo de ajuste, chamado RSREG, o qual, para o exemplo trabalhado no item *Exemplo de uso de fatorial fracionado em estudos de superfície de resposta*, é

```
PROC RSREG;MODEL PROD=N P K;
```

cujo resultado está representado na Figura 3.

Para que se obtenha uma análise de variância idêntica àquela mostrada na Tabela 11, é necessário o uso procedimento GLM, cuja programação seria a apresentada a seguir e cujo resultado se vê na Figura 4.

```
PROC GLM;
MODEL PROD=N P K N*N P*P K*K N*P N*K P*K/SS1 SOLUTION;
```

The RSREG Procedure

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
N	3.000000	2.000000
P	3.000000	2.000000
K	3.000000	2.000000

Response Surface for Variable PROD

Response Mean	1942.660800
Root MSE	97.628404
R-Square	0.6564
Coefficient of Variation	5.0255

Regression	DF	Type I Sum of Squares	R-Square	F Value	Pr > F
Linear	3	27650	0.0664	0.97	0.4340
Quadratic	3	136647	0.3284	4.78	0.0157
Crossproduct	3	108852	0.2616	3.81	0.0328
Total Model	9	273148	0.6564	3.18	0.0232

Residual	DF	Sum of Squares	Mean Square
Total Error	15	142970	9531.305360

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Parameter Estimate from Coded Data
Intercept	1	1070.022920	186.558367	5.74	<.0001	2061.824682
N	1	197.604821	89.433166	2.21	0.0431	20.755600
P	1	275.591856	87.938136	3.13	0.0068	34.647200
K	1	172.352698	89.433166	1.93	0.0731	-24.099200
N*N	1	-12.156697	12.593845	-0.97	0.3497	-48.626787
P*N	1	-15.652837	11.054047	-1.42	0.1772	-62.611350
P*P	1	-31.565603	11.772497	-2.68	0.0171	-126.262414
K*N	1	-22.442776	10.912322	-2.06	0.0575	-89.771104
K*P	1	-7.305374	11.054047	-0.66	0.5187	-29.221498
K*K	1	-15.859641	12.593845	-1.26	0.2272	-63.438564

The RSREG Procedure

Factor	DF	Sum of Squares	Mean Square	F Value	Pr > F
N	4	93538	23385	2.45	0.0910
P	4	102953	25738	2.70	0.0708
K	4	74067	18517	1.94	0.1555

The RSREG Procedure

Canonical Analysis of Response Surface Based on Coded Data

Factor	Critical Value Coded	Uncoded
N	1.164278	5.328557
P	-0.035100	2.929800
K	-1.005635	0.988730

Predicted value at stationary point: 2085.416767

Eigenvalues	Eigenvectors N	P	K
-8.598687	0.784049	-0.133329	-0.606209
-81.056145	0.454015	-0.542778	0.706585
-148.672933	0.423246	0.829226	0.365031

Stationary point is a maximum.

Figura 3. Análise de uma superfície de resposta de segundo grau, usando o procedimento RSREG do SAS.

The GLM Procedure
Number of observations 25
The GLM Procedure

Dependent Variable: PROD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	273148.4024	30349.8225	3.18	0.0232
Error	15	142969.5804	9531.3054		
Corrected Total	24	416117.9828			

R-Square	Coeff Var	Root MSE	PROD Mean
0.656421	5.025499	97.62840	1942.661

Source	DF	Type I SS	Mean Square	F Value	Pr > F
N	1	5384.93664	5384.93664	0.56	0.4639
P	1	15005.35585	15005.35585	1.57	0.2288
K	1	7259.64301	7259.64301	0.76	0.3966
N*N	1	16359.68881	16359.68881	1.72	0.2099
P*P	1	84635.04801	84635.04801	8.88	0.0093
K*K	1	35652.02368	35652.02368	3.74	0.0722
N*P	1	51107.59141	51107.59141	5.36	0.0351
N*K	1	53581.22666	53581.22666	5.62	0.0316
P*K	1	4162.88833	4162.88833	0.44	0.5187

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1070.022920	186.5583671	5.74	<.0001
N	197.604821	89.4331661	2.21	0.0431
P	275.591856	87.9381361	3.13	0.0068
K	172.352698	89.4331661	1.93	0.0731
N*N	-12.156697	12.5938447	-0.97	0.3497
P*P	-31.565603	11.7724969	-2.68	0.0171
K*K	-15.859641	12.5938447	-1.26	0.2272
N*P	-15.652837	11.0540466	-1.42	0.1772
N*K	-22.442776	10.9123222	-2.06	0.0575
P*K	-7.305374	11.0540466	-0.66	0.5187

Figura 4. Análise de uma superfície de resposta de segundo grau, usando o procedimento GLM do SAS.

Capítulo 15

Estatística não paramétrica

Este capítulo se refere às aplicações de testes não paramétricos ou de distribuição livre aos estudos de regressão e correlação, bem como à maioria dos delineamentos experimentais vistos em capítulos anteriores do ponto de vista paramétrico. Esses procedimentos apresentam a desvantagem de não serem gerais, isto é, não existe um teste único que possa ser aplicado na análise de variância de qualquer delineamento similar ao teste F; para cada delineamento foi desenvolvido um teste específico. Ademais, faz-se um estudo introdutório à teoria necessária aos testes não paramétricos, de interesse daqueles que desejarem se aprofundar no assunto.

Estatísticas de distribuição livre, definições e teoremas

Nas áreas de teste de hipótese e intervalo de confiança, o campo da estatística não paramétrica baseia-se pesadamente sobre um conceito conhecido como propriedade de distribuição livre. Esse conceito estabelece que um teste de hipótese de distribuição livre é aquele no qual o nível de significância permanece constante sob uma classe de pressuposições distribucionais fundamentais. A grande maioria dos testes de distribuição livre se baseia na transformação dos dados em postos ou na ordem relativa dos valores das variáveis medidas. Para um melhor entendimento sobre esse tipo de teste, uma série de definições e teoremas é apresenta a seguir.

Definição 1

A estatística T é dita de distribuição livre no campo real A se a distribuição de T é a mesma para toda distribuição conjunta em A.

Definição 2

A estatística T é dita ser não paramétrica de distribuição livre, se a classe A de distribuições conjuntas para a qual T tem a mesma distribuição inclui mais de uma forma distribucional.

Definição 3

A observação Z_i é dita ter posto R_i entre Z_1, ..., Z_N desde que a R_i-ésima estatística de ordem seja unicamente definida.

Teorema 1 (RANDLES; WOLFE, 1979)

Seja Z_1, ..., Z_N uma amostra aleatória de uma distribuição contínua, e seja R^* igual a (R_1..., R_N) o vetor de postos, em que R_i é a posição de Z_i na amostra. Se $R = \{r: r$ é uma permutação dos inteiros 1, ..., $N\}$, então R^* é uniformemente distribuído em R.

Corolário 1

Denote Z_1,..., Z_N uma amostra aleatória de uma distribuição contínua, e seja R^* o correspondente vetor de postos, em que R_i é o posto de Z_i entre as N variáveis aleatórias. Então:

$P[R_i = r] = 1/N$, para $r = 1, ..., N$,

$= 0$, nos outros casos,

e se *i* for diferente de *j*,

$P[R_i = r, R_j = s] = 1/N(N - 1)$, se r e s são inteiros, satisfazendo a condição de que $r^1s = 1, ..., N$, e

$= 0$, nos outros casos.

Corolário 2

Seja R^*, o vetor de postos, correspondente à amostra aleatória de tamanho N de uma distribuição contínua. Então:

$$E[R_i] = (N + 1)/2,$$
$$Var[R_i] = (N + 1)\ (N - 1)/12$$

e para i diferente de j,

$$\text{Cov}\ [R_i, R_j] = -(N + 1)/12.$$

Definição 4

Uma estatística definida como $V(R^*)$, que é uma função de $Z_1, ..., Z_N$, somente por meio do vetor de postos R^* é chamada de uma estatística de ordem.

Definição 5

O índice E_n, definido como a eficiência de um teste não paramétrico, é igual à razão entre o tamanho da amostra para um teste paramétrico e o tamanho da amostra para um teste não paramétrico, para um mesmo poder.

Correlação

Do ponto de vista paramétrico, os estudos a respeito de correlação encontram-se no Capítulo 12. Aqui se verificará que, no caso não paramétrico, os procedimentos são bem diversos, apesar de um dos testes existentes (ρ de Spearman) utilizar a mesma formula básica apresentada anteriormente para o coeficiente de correlação de Pearson, que é a medida de correlação mais comumente usada, chamada de r e definida como:

$$r = \frac{cov(x, y)}{\sqrt{var(y) * var(x)}} \qquad [1]$$

Como se viu no Capítulo 12, essa medida de correlação pode ser usada com quaisquer dados de uma natureza numérica, sem qualquer requerimento concernente à escala de medida ou ao tipo da distribuição básica, embora seja difícil de interpretar, a menos que a escala de medida seja pelo menos de intervalo. Infelizmente, a função de distribuição de r depende da distribuição de função bivariada de (X, Y); portanto, r não tem valor como um teste estatístico não paramétrico.

As medidas de correlação apresentadas a seguir são funções somente dos postos assinalados às observações. Elas possuem funções de distribuição que são independentes da função de distribuição bivariada de (X, Y) se X e Y são independentes e contínuos. Elas podem mesmo ser usadas como medidas de correlação em certos tipos de dados não numéricos.

Rho (ρ) de Spearman

Para a aplicação desse teste, requer-se que os dados consistam de uma amostra aleatória bivariada de tamanho n, (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n), aos quais se aplica a transformação de posto. Isto é, definindo-se $R(X_i)$ como o posto de X_i, quando comparado com os outros valores de X, para $i = 1, 2,..., n$; tem-se que $R(X_i) = 1$ se X_i é o menor de todos, $R(X_i) = 2$ é o segundo menor, e assim por diante, e o posto n é dado ao maior de todos os valores de X_i. Similarmente, $R(Y_i)$ será igual a 1, 2, ..., ou n, dependendo da magnitude relativa de Y_i quando comparado com todos os valores de Y.

Os dados podem também consistir de observações não numéricas ocorrendo em n pares se as observações são tais que possam ser transformadas em posto da maneira acima descrita. O posicionamento pode ser baseado na qualidade da observação (da pior para a melhor) ou de acordo com o grau de preferência amarrado à observação, e assim por diante. No caso de empates, assinala-se a cada valor empatado a média dos postos que eles teriam se não tivesse havido empate.

A medida de correlação dada por Spearman (1904) é usualmente designada por ρ (rho) e, se não houver empates, é definida como:

$$\rho = \frac{\sum_{i=1}^{n}\left[R(X_i) - \frac{(n+1)}{2}\right] * \left[R(Y_i) - \frac{(n+1)}{2}\right]}{\frac{n(n^2-1)}{12}} \qquad [2]$$

Uma forma equivalente, mas computacionalmente mais fácil é

$$\rho = 1 - \frac{6T}{n(n^2 - 1)} \quad [3]$$

em que

$$T = \sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2 \quad [4]$$

Se houver muitos empates, computa-se a medida de correlação de Spearman por meio da fórmula do coeficiente de correlação de Pearson com os valores das observações transformados em posto. Essa substituição pode ser feita mesmo quando não há empates, pois pode ser mostrado que as duas fórmulas se equivalem.

O coeficiente de correlação de posto de Spearman é frequentemente utilizado como um teste estatístico para a independência entre duas variáveis aleatórias. Na realidade, o rho (ρ) de Spearman é insensível a alguns tipos de dependência, então é melhor ser específico no que ser refere aos tipos de dependência que podem ser detectados. Portanto, as hipóteses tomam a seguinte forma.

a) Teste bilateral

H_0: X_i e Y_i são mutuamente independentes.

H_1: ou há uma tendência de os valores maiores de X parearem com os valores maiores de Y, ou de os valores menores de X parearem com os valores maiores de Y.

b) Teste unilateral para correlação positiva

H_0: X_i e Y_i são mutuamente independentes.

H_1: há uma tendência de os valores maiores de X e Y parearem.

c) Teste unilateral para correlação negativa

H_0: X_i e Y_i são mutuamente independentes

H_1: há uma tendência de os valores menores de X parearem com os maiores de Y, e vice-versa.

As hipóteses alternativas dadas estabelecem a existência de correlação entre X e Y, de forma que a hipótese nula de "não correlação entre X e Y" seria mais precisa que a declaração de independência dada nas hipóteses nulas. Entretanto, usar-se-á essa hipótese, porque ela é de uso mais amplo e mais fácil de interpretar.

O rho (ρ) de Spearman pode ser usado como um teste estatístico para as hipóteses precedentes. O Anexo 10 dá os quantis de rho (ρ) sob a pressuposição de independência, a hipótese nula. Então, H_0 em B é rejeitada se rho (ρ) é muito grande (ao nível α se rho (ρ) excede o quantil 1 - α); enquanto H_0 em C é rejeitada se rho (ρ) é muito pequeno, e o teste bilateral envolve a rejeição de H_0 se rho (ρ) excede o quantil 1 - $\alpha/2$ ou se rho (ρ) é menor que o quantil $\alpha/2$.

Ao invés do rho (ρ) de Spearman, é usualmente mais conveniente usar diretamente a estatística T, dada na fórmula 4.

$$T = \sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2$$

Note-se que, no caso de muitos empates, rho (ρ) deveria sempre ser usado. Entretanto, se o número de empates é moderado, o uso de T elimina alguma da aritmética envolvida no cálculo de rho (ρ). O teste nessa forma é chamado de teste de Hotelling-Pabst (HOTELLING; PABST, 1936), que enfatizou a natureza não paramétrica do teste de Spearman. Os quantis de T são dados no Anexo 11. Note-se, entretanto, que T é grande quando rho (ρ) é pequeno, e vice-versa. Portanto, a H_0 em B é rejeitada ao nível α se T é menor que o seu quantil α. Também H_0 em X é rejeitada se T excede seu quantil 1 - α.

Tau (τ) de Kendall

Também aqui se presume que os dados consistem de uma amostra aleatória bivariada de tamanho n, (X_i, Y_i) para $i = 1, 2, ..., n$; mas, para a sua aplicação, há a necessidade de se verificar se ocorre concordância ou não entre as observações. Segundo Kendall (1938), duas observações são ditas concordantes se ambos os membros de uma observação são maiores que os respectivos membros da outra observação, e ele chamou de N_c o número de pares de observação concordantes entre todos os pares possíveis obtidos a partir das combinações de n observações tomadas duas a duas. Por sua vez, um par de observações é chamado de discordante quando os dois números de uma observação diferem em direções opostas aos respectivos membros da outra observação, e aqui se chamará de N_d o número total de pares de observação discordantes. Pares com empates entre os membros

respectivos não são nem concordantes nem discordantes. Porque as n observações podem ser pareadas em

$$C_n^2 = \frac{n(n-1)}{2} \quad [5]$$

maneiras diferentes, o número de pares concordantes N_c mais o número de pares discordantes N_d mais o número de pares com empates têm de totalizar $n(n - 1)/2$.

Os dados podem também consistir de observações não numéricas ocorrendo em n pares se as observações são tais que N_c e N_d podem ser computados.

A medida de correlação proposta por Kendall (1938) é

$$\tau = \frac{N_c - N_d}{\frac{n(n-1)}{2}} \quad [6]$$

Se todos os pares são concordantes, o tau (τ) de Kendall é igual a 1. Se todos os pares são discordantes, o valor é igual a - 1; portanto, o tau (τ) de Kendall satisfaz os requerimentos antes estabelecidos.

O cálculo de tau (τ) é simplificado se as observações (X_i, Y_i) são arranjadas em uma coluna de acordo com os valores crescentes de X. Assim, cada Y pode ser comparado somente com aqueles abaixo dele, e o número de comparações concordantes e discordantes é facilmente determinado. Então, cada par de observações é considerado somente uma vez.

O tau (τ) de Kendall pode ser usado como um teste estatístico para testar hipótese nula de independência entre X e Y, com alternativas possíveis de duas ou de uma cauda como descrito no rho (ρ) de Spearman. Alguma aritmética pode ser economizada, entretanto, utilizando-se a diferença $N_c - N_d$ como um teste estatístico. Assim, usar-se-á como teste estatístico de Kendall o valor definido como

$$T = N_c - N_d \quad [7]$$

Quantis de T são dados em Conover (1980) (Anexo12).

O rho (ρ) de Spearman tende a ser maior que o tau (τ) de Kendall, em valor absoluto. Entretanto, como um teste de significância, não existe razão para preferir um em relação ao outro, porque ambos produzirão resultados praticamente idênticos, na maioria dos casos.

Em princípio, as distribuições exatas de rho (ρ) e tau (τ) são muito simples de obter, embora na prática o procedimento seja tedioso mesmo para valores moderados de n. As distribuições exatas são encontradas sob a pressuposição de que X_i e Y_i são independentes e identicamente distribuídos. Então, cada um dos n arranjos dos postos dos X_i pareados com os postos dos Y_i é igualmente provável. As funções de distribuição são obtidas simplesmente pela contagem de número de arranjos, que dão um valor particular de rho (ρ) ou tau (τ), e pela divisão daquele número por n para encontrar a probabilidade daquele valor de rho (ρ) ou tau (τ). Uma forma do teorema de limite central é aplicada para que se obtenham distribuições aproximadas para grandes amostras, porque ambos, rho (ρ) e tau (τ), são baseados na soma de variáveis aleatórias. Rho (ρ) e tau (τ) têm distribuição de probabilidade simétrica em torno de zero, de forma que zero é a média de ambos. As variâncias são mais difíceis de ser obtidas. Divisão de rho (ρ) e tau (τ) por suas respectivas variâncias resultará então numa variável aleatória que é aproximadamente distribuída como uma variável aleatória normal padrão para grandes valores de n. A aproximação é bastante boa quando usada para encontrar os quantis de tau (τ) para n maior ou igual a 8, mas não tão boa quando usada para encontrar os quantis de rho (ρ).

Se (X_i, Y_i), $i = 1, 2, ..., n$, são variáveis aleatórias normais bivariadas independentes e identicamente distribuídas, ambos, rho (ρ) e tau (τ), tem uma eficiência relativa assintótica de $9/\pi^2 = 0{,}912$, relativo ao teste paramétrico que usa a correlação de Pearson como um teste estatístico (STUART, 1954).

O tau (τ) de Kendall é usualmente considerado mais difícil de calcular que o rho (ρ) de Spearman. A vantagem maior do tau (τ) de Kendall é que sua distribuição se aproxima da distribuição normal mais rapidamente, de forma que a aproximação normal é melhor para o tau (τ) do que para o rho (ρ), quando a hipótese nula de independência entre X e Y é verdadeira. Outra vantagem do tau (τ) é sua interpretação simples e direta no que se refere à probabilidade de observarem-se pares concordantes e discordantes.

Exemplo de aplicação

Como exemplo de uso dos coeficientes de correlação de Spearman (1904) e de Kendall (1938), serão usados dados de análise de solos aluviais em dois estados da região Nordeste do Brasil (Piauí e Rio Grande do Norte), os quais apresentaram pH de solo abaixo de 7. Será analisada a associação existente entre teor de matéria orgânica no solo (em porcentagem) e disponibilidade de P ($mg*kg^{-1}$), extraídos de Fagéria et al. (1994). A tendência

natural dos pesquisadores nesses casos é o uso do coeficiente de regressão de Pearson (paramétrico), o qual supõe que a amostra provenha de uma distribuição normal bivariada, o que em geral não se verifica, exceto por aproximação com grandes amostras. Esse não é o caso presente, pois a amostra tem apenas sete observações. Além disso, uma das variáveis é medida em porcentagem, o que dificulta a aceitação de distribuição normal.

Os dados encontram-se na Tabela 1.

Tabela 1. Valores de matéria orgânica e fósforo nas amostras de sete solos aluviais do Nordeste brasileiro. N_c indica número de pares concordantes e N_d o número de pares discordantes.

Origem	Teor de matéria orgânica (%)	Conteúdo de fósforo (mg kg^{-1})	Posto do teor de matéria orgânica	Posto do conteúdo de fósforo	N_c	N_d
Picos, PI	0,6	2,0	1	1	6	0
Luzilândia, PI	0,9	16,6	2	4	3	2
Parnaíba, PI	1,1	3,4	3	2	4	0
Piripiri, PI	1,2	7,2	4	3	3	0
Ipanguaçu, RN	1,3	73,3	5	7	0	2
Caicó, RN	1,4	30,2	6	5	1	0
Teresina, PI	1,5	38,9	7	6		

Para o cálculo do ρ de Spearman, dado na fórmula a seguir, deve-se calcular primeiro o valor de T, dado por:

$$T = \sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2 = (1 - 1)^2 + (2 - 4)^2 + (3 - 2)^2 +$$
$$(4 - 3)^2 + (5 - 7)^2 + (6 - 5)^2 + (7 - 6)^2 = 12$$

Portanto, como $n = 7$, tem-se que

$$\rho = 1 - \frac{6T}{n(n^2 - 1)} = 1 - \frac{6 * 12}{7 * 48} = 1 - 0{,}2143 = 0{,}7857$$

e, segundo o Anexo 14, a probabilidade de encontrar esse valor está entre 0,025 e 0,01, o que indica a existência, significativa, de uma correlação entre essas variáveis.

Da mesma forma, seria possível calcular o coeficiente τ de Kendall, que, como se viu, é calculado por

$$\tau = \frac{N_c - N_d}{\frac{n(n-1)}{2}}$$

N_c e N_d indicam o número de pares concordantes e discordantes, respectivamente. Recorde-se que o número de pares concordantes é obtido após a ordenação dos postos da variável independente (X), no caso o teor de matéria orgânica, e que se considera concordante qualquer par de postos em que os dois valores crescem ou diminuem simultaneamente. Por exemplo, o par (1 - 1) é concordante com o par (2 - 3), pois o posto de X no segundo par é maior que no primeiro e o mesmo ocorre com o posto de Y. A partir da Tabela 1, tem-se que $N_c = 17$ e $N_d = 4$, logo

$$\tau = \frac{17 - 4}{\frac{7(7-1)}{2}} = \frac{13}{21} = 0{,}6190$$

valor um pouco menor que aquele obtido pelo coeficiente de Spearman, que, de acordo com o Anexo 16, tem uma probabilidade de ocorrência de 0,025 e, da mesma forma, é possível assumir a existência de correlação entre esses elementos no solo.

Apenas como curiosidade, o valor da correlação paramétrica é igual a 0,5944, não significativo.

Uso do SAS para estimar os coeficientes de correlação de Spearman e Kendall

Para a obtenção dos coeficientes de correlação não paramétricos, o procedimento SAS a ser usado é o mesmo do caso paramétrico, o PROC CORR, apenas com a necessidade de especificar o teste adequado. Para os dados da Tabela 1, será usado o programa abaixo, empregando os dois testes como exemplo.

```
PROC CORR SPEARMAN KENDALL;VAR MO P;
```

O resultado se vê na Figura 1, na qual se verifica que os valores calculados dos coeficientes de correlação de Spearman e de Kendall são iguais aos mostrados na secção

anterior. No entanto, há discrepância no valor da probabilidade do teste, e isso se deve ao uso da distribuição normal padrão como aproximação do teste, como discutido na seção *Tau (τ) de Kendall*, em lugar da distribuição exata dada nos Anexos 14 e 16.

The CORR Procedure
2 Variables: MO P

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
MO	7	1.14286	0.31015	1.20000	0.60000	1.50000
P	7	24.51429	25.60745	16.60000	2.00000	73.30000

Spearman Correlation Coefficients, N = 7
Prob > |r| under H0: Rho=0

	MO	P
MO	1.00000	0.78571 0.0362
P	0.78571 0.0362	1.00000

Kendall Tau b Correlation Coefficients, N = 7
Prob > |r| under H0: Rho=0

	MO	P
MO	1.00000	0.61905 0.0509
P	0.61905 0.0509	1.00000

Figura 1. Estudo de correlação não paramétrica com o SAS.

Análise de regressão linear

Aqui também se está examinando uma amostra aleatória (X_1, Y_1), ..., (X_n, Y_n) da variável aleatória bivariada (X, Y), mas a grande diferença é que os métodos de correlação enfatizam a estimação do grau de dependência entre X e Y, enquanto os métodos de regressão são usados para inspecionar a relação entre X e Y mais estreitamente, como já afirmado no Capítulo 12.

Lembre-se que a regressão de Y sobre X é dita ser linear se a equação de regressão é da forma $E(Y|X = x) = a + bx$, para as constantes a, chamada ponto de intersecção, e b, chamada de inclinação ou coeficiente angular.

Um método comumente aceito para estimar a e b é o chamado método dos quadrados mínimos, que minimiza a soma dos quadrados dos desvios das observações (X_1, Y_1), ..., (X_n, Y_n).

$$SQ = \sum_{i=1}^{n} [Y_i - (A - BX_i)]^2 \quad [8]$$

As estimativas dos parâmetros a e b são aquelas já mostradas no Capítulo 12 e dadas por

$$\hat{b} = \frac{\sum Y_i X_i - \frac{1}{n} \sum Y_i \sum X_i}{\sum X_i^2 - \frac{1}{n} (\sum X_i)^2} \quad [9]$$

e

$$\hat{a} = \overline{Y} - \hat{b}\overline{X} \quad [10]$$

Para testar a hipótese a respeito da inclinação da reta (b), é utilizado o rho (ρ) de Spearman, que pode ser adaptado para testar as seguintes hipóteses a respeito do coeficiente de inclinação. Represente b_0 algum número especificado.

a) Teste bilateral

$H_0: b = b_0$

$H_1: b^1 \neq b_0$

b) Teste unilateral

$H_0: b = b_0$

$H_1: b > b_0$

c) Teste unilateral

$H_0: b = b_0$

$H_1: b < b_0$

Para cada par (X_i, Y_i), computa-se $Y_i - b_0 * X_i = U_i$. Encontra-se, então, o coeficiente de correlação de postos de Spearman sobre os pares (X_i, U_i), $i = 1, ..., n$, como descrito anteriormente. A significância de rho (ρ) é feita como anteriormente.

Para a obtenção de um intervalo de confiança do coeficiente angular, o procedimento prevê que, para cada par de pontos (X_i, Y_i) e (X_j, Y_j), tal que $i < j$ e $X_i^1 X_j$, deve ser computada inclinação de dois pontos, $S_{ij} = (Y_i - Y_j)/(X_i - X_j)$.

Os valores de S_{ij} de todas as possíveis inclinações (N) são ordenados, obtendo-se então a sequência $S^1 < S^2 < ... < S^n$, em que S^i são as inclinações ordenadas.

Para um intervalo de confiança $1 - \alpha$, encontra-se $w_{1-\alpha/2}$ a partir do Anexo 12.

Sejam r e s dados por

$$r = (N - w1 - ./2))/2 \qquad [11]$$

e

$$s = (N + w_{1-\alpha/2})/2 + 1 \qquad [12]$$

Arredonda-se r para baixo e s para cima ao inteiro mais próximo, se eles já não forem inteiros. Então, o intervalo de confiança para b é dado pelo intervalo (S^r, S^s).

Cabem aqui alguns comentários. O primeiro deles diz respeito ao intervalo de confiança para b, baseado no tau (τ) de Kendall, um conceito completamente diferente daquele de quadrados mínimos; portanto, é possível, embora improvável, o estimador de quadrados mínimos para b estar fora do intervalo de confiança de b. Isso poderia acontecer, por exemplo, quando um valor de Y muito maior ou muito menor do que se poderia esperar estivesse entre os dados. Tal observação poderia puxar a linha de quadrados mínimos para ajustá-la à custa das outras observações. O segundo deles refere-se ao fato de que nenhuma pressuposição acerca da distribuição de (X, Y) foi feita, de forma que o método de quadrados mínimos é distribuição livre. A única pressuposição a ser feita é a de que os residuais

$$Y_i - E(Y_i|X_i) = Y_i - (a + bX_i) \qquad [13]$$

são independentes de X_i, de forma que as pressuposições que dizem respeito ao rho (ρ) de Spearman são obedecidas. Note-se que os postos de $(Y_i - a - bX_i)$ são os mesmos postos de $U_i = (Y_i - b_0 X_i)$ de forma que se pode testar H_0: $b = b_0$ sem conhecer a. Assim como o

rho (ρ) de Spearman é simplesmente o r de Pearson computado sobre os postos, esse teste é o análogo de postos ao computar-se r sobre os pares (X_i, U_i), e esse é o procedimento paramétrico usual para testar as mesmas hipóteses nulas, válidas com a pressuposição adicional de que (X, Y) tem a distribuição normal bivariada. Sob tais condições e a condição de que as observações em X são igualmente espaçadas, a Eficiência Relativa Assintótica (ERA) desse procedimento é $(3/\pi)^{1/3} = 0,98$ de acordo com Stuart (1954, 1956); para outras distribuições o ERA é sempre maior ou igual a 0,95 (LEHMANN, 1975). Para maiores detalhes sobre esse procedimento, recomenda-se a leitura de Conover (1980).

Exemplo de aplicação

Como exemplo de aplicação, será estudado o mesmo conjunto de dados utilizados no estudo de correlação.

A partir dos dados da Tabela 1, pode-se obter uma estimativa do parâmetro B da equação de regressão, empregando-se a fórmula 9, que é:

$$\hat{b} = \frac{\sum Y_i X_i - \frac{1}{n}\sum Y_i \sum X_i}{\sum X_i^2 - \frac{1}{n}(\sum X_i)^2}$$

$$\hat{b} = \frac{(0,6*2 + 0,9*16,6 + 1.1*3,4 + ... + 1,5*38,9) - \frac{1}{7}(0,6 + 0,9 + ... + 1,5)(2 + 16,6 + .. + 38,9)}{(0,6^2 + 0,9^2 + ... + 1,5^2) - \frac{1}{7}(0,6 + 0,9 + ... + 1,5)^2}$$

$$\hat{b} = \frac{224,44 - \frac{1}{7}(8)(171,6)}{9,72 - \frac{1}{7}(8)^2} = \frac{28,32571}{0,57714} = 49,0792$$

Com esse valor de b, pode-se então fazer a aplicação de algum teste que se considere adequado, como, por exemplo, H_0: $b = 50$ contra H_1: $b \neq 50$. Com esse hipotético valor de b_0, pode-se calcular a variável U_i, e calcular o coeficiente de correlação de Spearman

para comprovar ou não a validade da hipótese formulada. Esses valores de U_i e os postos correspondentes estão também descritos na Tabela 2.

Tabela 2. Valores de matéria orgânica, fósforo, U_i e os respectivos postos nas amostras de sete solos aluviais do Nordeste brasileiro.

Origem	Teor de matéria orgânica (%) X_i	Conteúdo de fósforo (mg kg^{-1}) Y_i	$U_i = Y_i - b_0 * X_i$	Posto de X_i	Posto de U_i
Picos, PI	0,6	2,0	-28	1	6
Luzilândia, PI	0,9	16,6	-28,4	2	5
Parnaíba, PI	1,1	3,4	-51,6	3	2
Piripiri, PI	1,2	7,2	-52,8	4	1
Ipanguaçu, RN	1,3	73,3	8,3	5	7
Caicó, RN	1,4	30,2	-39,8	6	3
Teresina, PI	1,5	38,9	-36,1	7	4

Pode-se agora aplicar o teste de Spearman, com os postos de X_i e U_i, de forma que se tem

$$T = \sum_{i=1}^{n} [R(X_i) - R(Y_i)]^2 = (1 - 6)^2 + (2 - 5)^2 + (3 - 2)^2 +$$
$$(4 - 1)^2 + (5 - 7)^2 + (6 - 3)^2 + (7 - 4)^2 = 66$$

e portanto

$$\rho = 1 - \frac{6T}{n(n^2 - 1)} = 1 - \frac{6 * 66}{7 * 48} = 1 - 1{,}1786 = -0{,}1786$$

Pelo Anexo 14, verifica-se que, com probabilidade de 5%, o valor é de -0,6786, muito menor que o obtido; portanto, não se pode rejeita a hipótese nula.

Para a determinação do intervalo confiança (IC) de b, é necessário primeiramente calcular todas as inclinações de dois pontos, ordená-las e depois, com o auxílio do Anexo 14, definir os limites superior e inferior do IC. Como o número de observações (n) é

igual a 7, e não existem empates entre os valores de matéria orgânica, tem-se um total de 21 inclinações para calcular. A primeira delas, por exemplo, é:

$$S_1 = (2 - 16,6)/(0,6 - 0,9) = 48,6667$$

Da mesma forma, calculam-se todas as demais 20 inclinações, que depois são ordenadas. O uso do Anexo 14 mostra que, para um intervalo de confiança 1 - α, o valor de $w_{1-\alpha/2}$ com $\alpha = 0,05$ é igual a 13. Agora podem-se obter os valores de r e s e as ordenadas para os limites inferior e superior, respectivamente. Então:

$$r = (N - w_{1-\alpha/2})/2 = (21 - 13)/2 = 4$$

e

$$s = (N + w_{1-\alpha/2})/2 + 1 = (21 + 13)/2 + 1 = 18$$

O quarto e décimo oitavo valores ordenados das inclinações de dois pontos consecutivos são iguais a -31,3333 e 115. Pode-se então concluir que o intervalo de confiança para b, com 95% de confiança, é: (-31,33 ≤ b ≤ 115).

Duas amostras independentes

A comparação de duas amostras independentes é feita por um teste conhecido como teste de Mann-Whitney (MANN; WHITNEY, 1947) e também como teste de Wilcoxon (1949). Formas equivalentes desse teste apareceram na literatura sob diversos nomes, provavelmente por causa do apelo intuitivo do procedimento. Embora primariamente um teste para duas amostras, o teste de Mann-Whitney pode ser aplicado em muitas situações diferentes desta.

A situação ocorre quando o experimentador que obteve as duas amostras, por uma série de razões se convence de que não deve utilizar o teste t.

Uma aproximação intuitiva ao problema das duas amostras é combinar ambas numa única amostra ordenada e então assinalar postos aos valores amostrais, do menor para o maior, sem considerar a que população cada valor pertence. Então o teste estatístico poderia ser a soma dos postos dados aos valores de uma população. Se a soma é muito pequena (ou muito grande), existe alguma indicação de que os valores daquela população tendem a ser menores (ou maiores, se for o caso), que os valores da outra população. Portanto, a

hipótese nula de não diferença entre populações pode ser rejeitada se os postos associados com uma amostra tendem a ser maiores do que aqueles da outra.

Entre as possíveis razões que justifiquem o uso de um teste não paramétrico estão as que indicam que os postos podem ser considerados preferíveis aos dados atuais pelos seguintes motivos: a) os números assinalados às observações não têm sentido por si mesmos, mas somente numa comparação ordinal com as outras observações; o número não contém mais informação do que os postos, tal é a natureza dos dados ordinais; b) os números têm significado, mas a função de distribuição não é normal, a teoria de probabilidade está usualmente fora de alcance quando o teste estatístico é baseado nos dados atuais; c) o tamanho da amostra ou a natureza dos dados não dão suporte ao uso da distribuição normal como aproximação.

A teoria de probabilidade de estatísticas baseadas em postos é relativamente simples e não depende da distribuição na maioria dos casos. Deve ser ressaltado que a ERA do teste de Mann-Whitney nunca é tão má quando comparada com o teste t para duas amostras, a contrapartida paramétrica usual, e que o contrário não é verdadeiro; a ERA do teste t comparada ao teste de Mann-Whitney pode ser tão pequena quanto zero, ou "infinitamente" má.

O teste de Mann-Whitney

Seja $X_1, X_2, ..., X_n$ a amostra aleatória de tamanho n da população 1 e $Y_1, Y_2, ..., Y_m$ a amostra aleatória de tamanho m da população 2. O procedimento consiste em agrupar as duas amostras, ordenar o conjunto do menor para o maior e assinalar os postos de 1 até $n + m$, e chamar de $R(X_i)$ e $R(Y_j)$ os postos assinalados a X_i e Y_j para todo i e j. Por conveniência, $N = n + m$. Quando valores amostrais forem exatamente iguais uns aos outros (empate), deve-se assinalar a cada um a média dos postos que eles teriam se não estivessem empatados.

As hipóteses a serem testadas podem se referir à forma da distribuição de probabilidade ou ao valor da probabilidade de que $X < Y$, ou ainda à média das populações. Para a correta aplicação do teste de Mann-Whitney, se as hipóteses se referem às distribuições de probabilidade ou ao valor das probabilidades, apenas as seguintes três pressuposições são necessárias: a) ambas as amostras são amostras aleatórias das suas respectivas populações; b) existe independência dentro de cada amostra, bem como independência mútua entre as duas amostras; c) a escala de medida é pelo menos ordinal. Para o teste de médias, deve ser adicionada que existe uma diferença entre as funções de distribuição, isto é, $F(X) \neq G(Y)$ ou $F(X) = G(Y + c)$, em que c é uma constante qualquer. O teste pode apresentar as formas bilateral ou unilaterais, de forma que se tem para as diferentes situações:

- Teste bilateral

  Sejam $F(x)$ e $G(x)$ as funções de distribuição correspondentes às populações 1 e 2, respectivamente, de X e Y. Então:

  H_0: $F(x) = G(x)$ para todo x

  H_1: $F(x) \neq G(x)$ para algum x

  Em muitas situações reais, qualquer diferença entre distribuições implica que $P(X < Y)$ não mais é igual a 1/2. Portanto, será possível adotar:

  H_0: $P(X < Y) = 1/2$

  H_1: $P(X < Y) \neq 1/2$

- Teste unilateral

  H_0: $P(X < Y) = 1/2$

  H_1: $P(X < Y) > 1/2$

- Teste unilateral

  H_0: $P(X < Y) = 1/2$

  H_1: $P(X < Y) < 1/2$

Quando as hipóteses são estabelecidas em termos das médias de X e Y, se elas existem, tem-se que:

- Teste bilateral

  H_0: $E(X) = E(Y)$

  H_1: $E(X) \neq E(Y)$

- Teste unilateral

  H_0: $E(X) = E(Y)$

  H_1: $E(X) < E(Y)$

- Teste unilateral

  H_0: $E(X) = E(Y)$

  H_1: $E(X) > E(Y)$

O teste de Mann-Whitney é não viciado e consistente quando se testam as hipóteses precedentes.

Se não ocorrem empates, ou se há somente alguns, a soma dos postos assinalados à população 1 pode ser usada como um teste estatístico.

$$T = \sum_{i=1}^{n} R(X_i) \qquad [14]$$

Se existem muitos empates, subtraia a média de T e divida pelo desvio padrão para obter

$$T_1 = \frac{T - \frac{n(N+1)}{2}}{\sqrt{\frac{nm}{N(N-1)} \sum_{i=1}^{N} R_i^2 - \frac{nm(N+1)^2}{4(N-1)}}} \qquad [15]$$

em que $\sum_{i=1}^{N} R_i^2$ refere-se à soma dos quadrados de todos os N postos ou postos médios realmente utilizados em ambas as amostras.

Os quantis w_p de T são dados no Anexo 7 para p = 0,001, 0,005, 0,01, 0,025, 0,05 e 0,1. Os quantis superiores não são dados, mas podem ser calculados por subtração de $n(N+1)$. Isto é, $w_{1-p} = n(N+1) - w_p$.

Como uma alternativa para usar quantis superiores, a estatística T', definida como

$$T' = n(N+1) - T \qquad [16]$$

pode ser usada com os quantis inferiores sempre que um teste de quantil superior é desejado.

Os quantis aproximados de T_1 são dados no Anexo 1, na distribuição normal.

Nos seguintes conjuntos de regras, substitua T_1 por T se T_1 está realmente sendo utilizado.

a) Teste bilateral – rejeite A_0 ao nível de significância α se T é menor que o quantil $w_{\alpha/2}$, $1 - \alpha/2$ se T é maior que o quantil $w_{1-\alpha/2}$. Aceite H_0 se T está entre ou é igual aos dois quantis.

b) Teste unilateral – valores pequenos de T indicam que H_1 é verdadeiro. Portanto, rejeite H_0 ao nível de significância α se T é menor do que o α-ésimo quantil w_α. Aceite H_0 se T é maior ou igual a w_α.

c) Teste unilateral – valores grandes de T, ou valores pequenos de T', indicam que H_1 é verdadeiro. Portanto, rejeite H_0 ao nível de significância α se T é maior do que $w_{1-\alpha}$ ou (equivalentemente) se T' é menor que w_α. Aceite H_0 se T é maior ou igual a $w_{1-\alpha}$.

A distribuição nula de T é encontrada quando se pressupõe que X_i e Y_j são identicamente distribuídos. Isso é estritamente verdadeiro somente quando H_0 é verdadeiro no teste bilateral. Entretanto, nos testes unilaterais, α é encontrado maximizando a probabilidade de T cair na região de rejeição apropriada, e aquela probabilidade é um máximo, sob H_0, quando as duas populações têm distribuições idênticas. Portanto, a distribuição de T é encontrada quando se assume que X_i e Y_j são identicamente distribuídos, não importando qual das três formas de H_0 está sendo testada.

Se X_i e Y_j são independentes e identicamente distribuídos, todo arranjo dos X e dos Y na amostra combinada ordenada é igualmente provável. Esse é o princípio básico atrás de muitos testes de postos. Se X_i e Y_j são independentes e identicamente distribuídos, o posto assinalado aos X_i na amostra combinada deveria parecer uma seleção casual de n inteiros extraídos de 1 até $n + m$. Isto é, não há razão para que qualquer posto particular tenha uma chance melhor que qualquer outro de ser assinalado a um valor de X_i. Pelo fato de cada número de 1 até $n + m$ ter igual probabilidade de ser assinalado a X_i como seu posto, e pelo fato de n números diferentes serem selecionados como postos dos X, a distribuição de probabilidade de T e a soma dos postos podem ser obtidas considerando-se a distribuição de probabilidade da soma de n inteiros selecionados ao acaso, sem substituição, entre os inteiros de 1 até $n + m$.

O número de maneiras de selecionar n inteiros de um número total de $n + m$ inteiros é igual a $(m + n)!/m!n!$, e cada maneira tem probabilidade igual de acordo com a premissa básica anteriormente estabelecida. Então, a probabilidade de $T = k$ pode ser encontrada contando-se um número de conjuntos diferentes de n inteiros de 1 até $n + m$ que somam k e, então, dividindo-se aquele número pelo valor dado anteriormente.

Visto que T é a soma dos postos dos nX para grandes valores de n e m, o teorema do limite central pode ser aplicado para obter uma distribuição aproximada para T, sabendo-se que a média é dada por $n(N + 1)/2$ e a variância por $nm(N + 1)/12$.

Exemplo de aplicação

Como exemplo de aplicação de teste de Mann-Whitney, serão usados os teores em matéria orgânica dos solos aluviais e gley húmicos obtidos por Fagéria et al. (1994). Como foi visto anteriormente, o teste prevê a designação de postos a cada uma das observações, considerando-se as duas amostras como uma só amostra. Os resultados dessa designação de postos e os dados estão na Tabela 3.

Tabela 3. Dados dos teores em matéria orgânica em amostras de dois tipos de solo de várzea brasileiros, e os postos correspondentes após a ordenação dos dados.

Matéria orgânica em sete solos aluviais (A)	Matéria orgânica em sete solos gley húmicos (B)	Posto de A (considerando amostra única)	Posto de B (considerando amostra única)
0,6		1	
	0,7		2
0,9		3,5	
	0,9		3,5
1,1		5	
1,2		6	
1,3		7	
1,4		8	
1,5		9,5	
	1,5		9,5
	2,2		11
	3,7		12
	10,5		13
	16,8		14

O valor do teste de Mann-Whitney pode agora ser obtido, e como ocorreram apenas dois empates, é possível utilizar a forma mais simples desse teste de modo que o valor de T será dado pela soma dos postos assinalados aos dados da amostra A (solos aluviais), e o valor deverá ser comparado com os dados do Anexo 12 com $n = m = 7$.

$$T = 1 + 3{,}5 + 5 + 6 + 7 + 8 + 9{,}5 = 40$$

Pelo Anexo 12, é possível ver que esse valor está associado a um nível de probabilidade de 5%. Logo, podemos concluir que os teores de matéria orgânica desses dois tipos de solos são diferentes, e que os solos aluviais possuem um menor teor dessa matéria.

Uso do SAS para teste de duas amostras independentes pelo teste de Wilcoxon

Para o problema de testar duas amostras independentes, o SAS inclui o procedimento NPAR1WAY, com a opção WILCOXON. A entrada de dados para esse problema se faz com um mínimo de duas variáveis, uma de classificação, correspondente às amostras, e outra de resposta, correspondente à variável de interesse na análise. Para os dados da Tabela 3, as variáveis são respectivamente SOLO e MO. O programa então tem a seguinte conformação:

```
PROC  NPAR1WAY WILCOXON;CLASS SOLO;VAR MO;
```

O resultado aparece na Figura 2.

```
                    The NPAR1WAY Procedure

             Wilcoxon Scores (Rank Sums) for Variable mo
                    Classified by Variable solo
                  Sum of      Expected       Std Dev          Mean
solo       N       Scores      Under H0      Under H0         Score
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
gley       8          80.0          64.0      8.625543     10.000000
aluv       7          40.0          56.0      8.625543      5.714286
                 Average scores were used for ties.

                    Wilcoxon Two-Sample Test
                   Statistic             40.0000
                   Normal Approximation
                   Z                     -1.7970
                   One-Sided Pr <  Z      0.0362
                   Two-Sided Pr > |Z|     0.0723

                   t Approximation
                   One-Sided Pr <  Z      0.0470
                   Two-Sided Pr > |Z|     0.0939

              Z includes a continuity correction of 0.5.
                        Kruskal-Wallis Test
                   Chi-Square            3.4409
                   DF                         1
                   Pr > Chi-Square       0.0636
```

Figura 2. Teste de Mann-Whitney ou de Wilcoxon, com o uso do SAS.

Novamente aqui se vê uma discrepância na significância do teste, em razão do uso das aproximações normal ou pelo teste de t, o que comprova mais uma vez a necessidade de uso das tabelas dos testes não paramétricos com o uso de amostras pequenas. Deve-se ressaltar que, nesse caso, não há diferença entre as conclusões dos testes, sendo significativa a diferença entre as amostras em qualquer situação. Outro fato interessante é o uso do teste de Kruskal-Wallis (KRUSKAL; WALLIS, 1952), equivalente ao de Wilcoxon com duas amostras, mas que aqui usa a aproximação do teste de qui-quadrado, não significativo.

Delineamento inteiramente ao acaso

A situação experimental a ser estudada é aquela em que k amostras aleatórias têm sido obtidas ou induzidas por meio de ensaio conduzido neste delineamento (Capítulo 2). Nesse caso, o objetivo é testar a hipótese nula de que todas as populações são idênticas contra a alternativa de que algumas das populações tendem a fornecer valores observados maiores do que outras. O termo "maior" aplica-se a observações de variáveis aleatórias, mas atualmente qualquer observação que possa ser arranjada em ordem crescente, de acordo com alguma propriedade, pode ser analisada usando o teste de Kruskal-Wallis (KRUSKAL; WALLIS, 1952) em uma maneira semelhante à análise de dados não numéricos usando o teste de Mann-Whitney.

O teste de Kruskal-Wallis usa mais informação contida nas observações do que outros testes não paramétricos para a mesma situação. Isto é, o teste estatístico de Kruskal-Wallis é uma função dos postos das observações na amostra combinada, como foi feito no teste de Mann-Whitney, enquanto outros testes estatísticos (teste de mediana, por exemplo) dependem somente do fato de serem conhecidas as observações que estão abaixo ou acima de uma mediana.

Chamando a i-ésima amostra aleatória (tratamento) de tamanho n_i, por $X_{i1}, X_{i2}, ..., X_{ini}$ e fazendo N o número total de observações

$$N = \sum_{i=1}^{k} n_i \qquad [17]$$

deve-se assinalar o posto 1 à menor das N observações, posto 2 à segunda menor, e assim por diante, até a maior das N observações, a qual recebe posto N. O passo seguinte é a obtenção de R_i, a soma dos postos assinalados à i-ésima amostra ou tratamento, dado por

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}) \qquad [18]$$

e computado para cada amostra.

Se os postos podem ser assinalados de diferentes maneiras pelo fato de algumas observações serem iguais umas às outras, deve ser dado o posto médio a cada uma das observações empatadas.

As pressuposições feitas para a aplicação desse teste são: a) todas as amostras são amostras aleatórias de suas populações respectivas; b) além da independência dentro de cada amostra, existe independência mútua entre as várias amostras; c) a medida de escala é pelo menos ordinal; d) ou as k funções de distribuição populacionais são idênticas, ou algumas das populações tendem a produzir valores maiores do que outras.

Essas pressuposições levam ao teste das seguintes hipóteses:

H_0: Todas as k funções de distribuição populacionais são idênticas.

H_1: Pelo menos uma das populações tende a produzir observações maiores do que pelo menos uma das outras.

A hipótese alternativa H_1 é algumas vezes substituída por:

H_1: As k populações não têm médias idênticas

O teste estatístico T é definido como

$$T = \frac{\sum_{i=1}^{k} \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}}{S^2} \qquad [19]$$

em que N e R_i são iguais aos valores definidos anteriormente e

$$S^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} R(X_{ij})^2 - \frac{N(N+1)^2}{4}}{N-1} \qquad [20]$$

Se não existem empates S^2, simplifica-se para $N(N+1)/12$, e o teste estatístico se reduz a

$$T = \frac{12 \sum_{i=1}^{k} \frac{R_i^2}{n_i}}{N(N+1)} - 3(N+1) \qquad [21]$$

Se o número de empates é moderado, haverá pouca diferença entre os valores calculados pelas duas equações de T.

Se $k = 3$, e todos os tamanhos de amostras são 5 ou menor, e não há empates, o quantil exato pode ser obtido do Anexo 8. Tabelas mais completas são dadas por Iman et al. (1975). Quando existem empates, ou quando tabelas exatas não estão disponíveis, os quantis aproximados podem ser obtidos do Anexo 2, a distribuição de qui-quadrado com k - 1 graus de liberdade. Rejeite-se H_0 ao nível α, se T excede o quantil $1 - \alpha$ obtido.

Se, e somente se, a hipótese nula é rejeitada, devem-se determinar quais pares de populações tendem a diferir. Pode-se dizer que as populações i e j parecem ser diferentes se a inigualdade a seguir definida é satisfeita:

$$\left[\frac{R_i}{n_i} - \frac{R_j}{n_j}\right] > t_{(1-\frac{\alpha}{2})} \sqrt{S^2 \frac{N - 1 - T}{N - k}} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \qquad [22]$$

em que R_i e R_j são as somas dos postos das duas amostras, $t_{(1-\alpha/2)}$ é o quantil $(1 - \alpha/2)$ da distribuição de t com $N - k$ graus de liberdade, S^2 vem da equação anterior e T é o teste estatístico. Esse procedimento é repetido para todos os pares de populações. O nível α usado nessa comparação deve ser o mesmo do teste de Kruskal-Wallis.

A distribuição exata de T é encontrada sob a pressuposição de que todas as observações foram obtidas de uma ou de diversas populações idênticas. O método é o da casualização, o qual foi também usado para encontrar a distribuição do teste estatístico de Mann-Whitney. Isto é, sob a pressuposição antecedente, cada arranjo dos postos 1 a N em grupos de tamanhos $n_1, n_2, ..., n_k$, respectivamente, é igualmente provável, e ocorre com probabilidade $n_1!n_2! \ldots n_k!/N!$, que é o recíproco do número de maneiras em que N postos podem ser divididos em grupos de tamanhos $n_1, n_2, ..., n_k$. O valor de T é calculado para cada arranjo. As probabilidades associadas a iguais valores de T são então somadas para dar a distribuição de probabilidade de T.

A aproximação para grandes amostras para distribuição de T foi definida por Kruskal (1952), enquanto Gabriel e Lachembruch (1969) mostraram que a aproximação de qui-quadrado é boa mesmo quando o tamanho de amostra é pequeno.

Para duas amostras, o teste de Kruskal-Wallis é equivalente ao teste de Mann-Whitney.

O procedimento de comparações múltiplas é simplesmente o procedimento paramétrico usual, o DMS ou LSD de Fisher, calculado com os postos ao invés dos dados, como descrito por Conover e Iman (1979).

A violação da pressuposição de normalidade usualmente tem pouco efeito sobre a distribuição da estatística F quando a hipótese nula é verdadeira. Entretanto, o poder do teste F pode ser consideravelmente menor do que o teste de Kruskal-Wallis para certos tipos de não normalidade quando H_0 é falsa. Dados que contenham valores muito discrepantes são mais bem adaptados ao teste de Kruskal-Wallis. A ERA do teste de Kruskal-Wallis relativa ao teste F nunca é menor do que 0,864, mas pode ser tão alto quanto infinito se a função de distribuição tem formas idênticas, mas diferem somente nas suas médias. Se as populações são normais, a ERA é igual a 0,955. Para distribuições uniformes, a ERA é igual a 1,0; para distribuições duplo-exponenciais, ela é de 1,5.

Exemplo de aplicação

Um experimento inteiramente ao acaso foi conduzido para que fossem testadas várias cultivares de arroz, em casa de vegetação, das quais três delas (A, B e C) estão sendo utilizadas como exemplo de uso do teste de Kruskal-Wallis, em que, em cada parcela, contou-se o número de hastes mortas por *Elasmopalpus lignosellus*. Cada cultivar foi repetida três vezes (dados cedidos pelo pesquisador da Embrapa Arroz e Feijão, Evane Ferreira).

Os dados medidos estão na Tabela 4.

Tabela 4. Número de hastes mortas por *E. lignosellus* L., e respectivos postos, num experimento inteiramente casualizado.

Número de hastes mortas por cultivar			Posto do número de hastes mortas		
A	B	C	A	B	C
12	11	16	4	3	8
13	9	17	5	2	9
15	7	14	7	1	6
Soma dos postos por cultivar			16	6	23

Usando-se a fórmula simplificada dada anteriormente, já que não houve empates, tem-se:

$$T = \frac{12[\frac{16^2}{3} + \frac{6^2}{3} + \frac{23^2}{3}]}{9\,(10)} - 3(10)$$

$$T = 6{,}69^{**}$$

Verificando-se o Anexo 8, encontra-se que o teste de Kruskal-Wallis pode ser considerado significativo a uma probabilidade de 0,01; portanto, a hipótese de igualdade de resposta das cultivares pode ser rechaçada.

Para comparações múltiplas, usando-se a fórmula dada anteriormente, tem-se:

$R_1 = 16, R_2 = 6, R_3 = 23$

$t_{0,995} = 3,707$ (6 GL)

$S^2 = (N + 1)N/12 = 90/12$ e

$T = 6,49$

Logo,

$$|R_i - R_j| > 3,707 \sqrt{\frac{90}{12} * \frac{8 - 6,49}{6}} \sqrt{\frac{2}{3}}$$

$|R_i - R_j| > 3,707 * 1,3739 * 0,8165$

$|R_i - R_j| > 4,15$

Portanto, verifica-se que

$|R_1 - R_2| = 10$ ** $|R_1 - R_3| = 7$ ** $|R_2 - R_3| = 17$ **

o que confirma que as três cultivares apresentam reação distinta ao ataque desse inseto.

Uso do SAS para análise não paramétrica do delineamento inteiramente ao acaso (mais de duas amostras)

Também aqui o procedimento NPAR1WAY deve ser usado conforme se vê no programa a seguir:

```
PROC NPAR1WAY WILCOXON;CLASS TRAT;VAR HASTES;
```

A opção continua sendo WILCOXON, mas, com mais de duas amostras ou níveis da variável de classificação, ele corresponde ao teste de Kruskal-Wallis. O resultado se vê na Figura 3.

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable HASTES
Classified by Variable TRAT

TRAT	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
A	3	16.0	15.0	3.872983	5.333333
B	3	6.0	15.0	3.872983	2.000000
C	3	23.0	15.0	3.872983	7.666667

Kruskal-Wallis Test

Chi-Square	6.4889
DF	2
Pr > Chi-Square	0.0390

Figura 3. Resultado da análise de variância de um delineamento inteiramente casualizado, com o uso do procedimento NPAR1WAY do SAS.

Novamente aqui, foi usada a aproximação pelo teste de qui-quadrado, que mostra uma diferença significativa entre os tratamentos com 3,9%, enquanto o teste exato, aplicado na seção *Exemplo de aplicação*, mostra significância de 1%.

Delineamento em blocos completos casualizados

Neste caso, considera-se o problema de analisar diversas amostras relacionadas, o que é uma extensão de problema examinado no item anterior e que no caso paramétrico foi estudado no Capítulo 3.

Dois testes serão apresentados, o primeiro deles é um teste que é uma extensão daquele de Wilcoxon para o caso de diversas amostras relacionadas (QUADE, 1979). O segundo é o teste de Friedman (FRIEDMAN, 1937), que é uma extensão do teste de sinal e requer menos pressuposições do que o de Quade. O teste de Friedman é o mais conhecido para esta situação experimental, mas tem menos poder em algumas situações.

As $k (k \geq 2)$ amostras são arranjadas em um experimento que é delineado no esquema de blocos completos, que possuem cada um k unidades experimentais como visto no Capítulo 3. Nesse caso, cada tratamento aplicado uma única vez dentro de cada bloco é considerada.

Teste de Quade

O procedimento consiste em:

1) Assinalar postos a cada observação de um dado bloco, e então chamar de $R(X_{ij})$ o posto, de 1 a k, assinalado a Y_{ij} dentro do bloco i; isto é, para o bloco i as variáveis

aleatórias Y_{i1}, Y_{i2}, ..., Y_{ik}, são comparadas umas às outras e o posto 1 é assinalado ao menor dos valores observados, o posto 2 ao segundo menor, e assim por diante, até o posto k, o qual é assinalado à maior observação nesse bloco. Postos são assinalados em todos os b blocos. Usam-se postos médios no caso de empates.

2) O próximo passo usa novamente as observações originais Y_{ij}. Postos são assinalados aos blocos de acordo com o tamanho do intervalo amostral de cada um deles, que consiste na diferença entre a maior e a menor observação dentro daquele bloco. Assinale-se posto 1 ao bloco com o menor intervalo, posto 2 ao bloco com o segundo menor intervalo, e assim por diante, até o bloco com o maior intervalo, que recebe posto b. Usam-se postos médios em caso de empates. Seja Q_1, Q_2, ..., Q_b os postos assinalados aos blocos $1, 2, ..., b$, respectivamente.

3) Finalmente, o posto de bloco Q_i é multiplicado pela diferença entre o posto dentro do bloco i, $R(X_{ij})$, e o posto médio dentro de blocos, $(k + 1)/2$, para obter o produto S_{ij}

$$S_{ij} = Q_i \, [R(X_{ij}) - (k + 1)/2]$$

que é a estatística que representa o tamanho relativo de cada observação dentro do bloco, ajustado para refletir a significância relativa do bloco no qual ela aparece.

Denote-se por S_j a soma para cada tratamento

$$S_j = \sum_{i=1}^{b} S_{ij} \qquad [23]$$

para $j = 1, 2, ..., k$.

As pressuposições necessárias para a aplicação desse teste são:

1) As b variáveis aleatórias k-variadas são mutuamente independentes (os resultados dentro de um bloco não influenciam os resultados dentro dos outros blocos).

2) Dentro de cada bloco as observações podem ser ordenadas de acordo com algum critério de interesse.

3) O intervalo amostral pode ser determinado dentro de cada bloco, de forma que os blocos possam ser ordenados.

4) Deve ser notado que esse teste é válido mesmo quando existem muitos empates nos postos.

As hipóteses a serem testadas são:

H_0: cada ordenação das variáveis aleatórias dentro de um bloco é igualmente provável (isto é, os tratamentos têm efeitos idênticos).

H_1: pelo menos um dos tratamentos tende a produzir valores observados maiores do que pelo menos um dos outros tratamentos.

A aplicação desse teste implica primeiramente o cálculo da soma de quadrados total dada por

$$A_1 = \sum_{i=1}^{b} \sum_{j=1}^{k} S_{ij}^2 \quad [24]$$

em que S_{ij} foi definido anteriormente. Note-se que se não existem empates, A_1 simplifica-se para

$$A_1 = b(b + 1)(2b + 1)k(k + 1)(k - 1)/72 \quad [25]$$

Depois, calcula-se o termo

$$B_1 = \frac{1}{b} \sum_{j=1}^{k} S_j^2 \quad [26]$$

em que S_j, também definido anteriormente, corresponde à soma de quadrados de tratamentos. O teste estatístico é dado por

$$T_1 = \frac{(b - 1)B_1}{A_1 - B_1} \quad [27]$$

Se $A_1 = B_1$, considere que aquele ponto esteja na região crítica e calcule o nível crítico como $\alpha = (1/k!)^{b-1}$.

Rejeita-se a hipótese nula ao nível α se T_1 excede o quantil $1 - \alpha$ da distribuição F, como dado nos Anexos 3 e 4 com $k1 = k - 1$ e $k2 = (b - 1)(k - 1)$ graus de liberdade. Na realidade, a distribuição de F somente aproxima a distribuição exata de T_1, pois tabelas exatas ainda não são disponíveis. Quanto maior se torna b, mais a distribuição de F se aproxima da distribuição exata.

As comparações múltiplas podem ser feitas somente se o procedimento anterior resultar na rejeição da hipótese nula. Tratamentos i e j são considerados diferentes se a inigualdade

$$| S_i - S_j | > t_{(1-\frac{\alpha}{2})} \sqrt{\frac{2b(A_1 - B_1)}{(b-1)(k-1)}} \quad [28]$$

é satisfeita, em que S_i, S_j, A_1 e B_1 foram dados anteriormente, e $t_{1-\alpha/2}$ é obtido do Anexo 2 para o teste de t com $(b - 1)(k - 1)$ graus de liberdade. Essa comparação é feita para todos os pares de tratamentos, usando o mesmo α usado no teste de Quade.

Teste de Friedman

Prefere-se o teste de Friedman quando comparações entre os diferentes blocos não são possíveis. O teste é mais fácil de ser executado do que o teste de Quade; portanto, é mais simples de entender. Se o número de tratamentos for cinco ou mais, esse teste parece ser mais poderoso do que o teste de Quade.

O procedimento consiste em encontrar os postos dentro dos blocos ($R(X_{ij})$), como descrito no teste de Quade, e então somar os postos de cada tratamento para obter R_j, em que:

$$R_j = \sum_{i=1}^{b} R(X_{ij}) \quad [29]$$

para $j = 1, 2, ..., k$. Note-se que nenhuma comparação entre os diferentes blocos é feita nesse teste.

Apenas as duas primeiras pressuposições feitas no teste de Quade são necessárias para esse teste, enquanto as hipóteses são exatamente as mesmas.

Para sua aplicação, começa-se por calcular o termo A_2, dado por

$$A_2 = \sum_{i=1}^{b} \sum_{j=1}^{k} [R(X_{ij})]^2 \quad [30]$$

Se não existem empates, A_2 simplifica-se para

$$A_2 = \frac{bk(k+1)\,(2k+1)}{6} \quad [31]$$

Em seguida, calcula-se o termo

$$B_2 = \frac{1}{b}\sum_{j=1}^{k} R_j^2 \quad [32]$$

em que R_j foi definido anteriormente. O teste estatístico é

$$T_2 = \frac{(b-1)\left[B_2 - \frac{bk(k+1)^2}{4}\right]}{A_2 - B_2} \quad [33]$$

Se $A_2 = B_2$, considere que aquele ponto esteja na região crítica e compute o nível crítico como $\alpha = (1/k)^{b-1}$.

Rejeita-se a hipótese nula ao nível α se T_2 excede o quantil $1 - \alpha$ na distribuição de F, como dado nos Anexos 3 e 4, em que $k1 = k - 1$ e $k2 = (b - 1)(k - 1)$. Também aqui, a distribuição de F somente aproxima a distribuição exata de T_2, mas a aproximação é bastante boa e melhora quanto maior é o valor de b.

A forma usual do teste de Friedman compara a estatística

$$T_3 = \frac{(k-1)\left[bB_2 - \frac{b^2 k(k+1)^2}{4}\right]}{A_2 - \frac{bk(k+1)^2}{4}} \quad [34]$$

com quantis da distribuição de qui-quadrado, com $k - 1$ graus de liberdade. Mas estudos de Iman e Davenport (1980) mostram que a aproximação F para T_2 é claramente superior à aproximação anterior de qui-quadrado.

Para comparar as médias de tratamentos, pode ser usada a estatística a seguir, se, e somente se, o teste de Friedman resulta em rejeição da hipótese nula, de forma que os tratamentos *i* e *j* são considerados diferentes se a inigualdade é satisfeita.

$$|R_i - R_j| > t_{(1-\frac{\alpha}{2})}\sqrt{\frac{2b(A_2 - B_2)}{(b-1)(k-1)}} \quad [35]$$

em que R_i, R_j, A_2 e B_2 foram dados anteriormente e $t_{(1-\alpha/2)}$ é o quantil $1 - \alpha/2$ da distribuição de t com $(b - 1)(k - 1)$ graus de liberdade. O valor de α é o mesmo usado no teste de Friedman.

Sobre as distribuições de probabilidade e outros comentários

As distribuições exatas de T_1, T_2, e T_3 são encontradas sob a pressuposição de que cada ordenamento dentro de um bloco é igualmente provável, e isso corresponde à hipótese nula. Existem $k!$ possíveis arranjos de postos $R(X_{1j})$ dentro de um bloco; portanto, há $(k!)b$ possíveis arranjos de postos no conjunto total de b blocos. A declaração precedente implica que cada um desses $(k!)b$ arranjos é igualmente provável sob a hipótese nula. Portanto, as distribuições de probabilidade de T_1, T_2, e T_3 podem ser encontradas para um determinado número de k amostras e b blocos, meramente por listarem-se todos os possíveis arranjos de postos e, então, calculando-se T_1, T_2, e T_3 para cada arranjo.

As aproximações das distribuições de T_1, T_2, e T_3, que usam as distribuições de F ou qui-quadrado, são justificadas usando o teorema do limite central. Detalhes do desenvolvimento das distribuições assintóticas podem ser encontrados nos seguintes trabalhos: Quade (1972, 1979), Lawler (1978) para T_1, Iman e Davenport (1980) para T_2 e Friedman (1937) para T_3.

Se o teste F é calculado usando-se os postos em vez dos dados, a estatística é igual a T_2. Se o teste F é computado sob os postos ponderados S_{ij}, o resultado é a estatística T_1, e o ajuste para empates é automaticamente incorporado no teste F.

Existe uma relação direta entre T_2 e T_3:

$$T_2 = \frac{(b-1)T_3}{b(k-1) - T_3} \quad [36]$$

O método de comparações múltiplas descritas anteriormente é meramente o procedimento paramétrico, conhecido como diferença mínima significativa de Fisher (FISHER; YATES, 1957), mas computado sobre os números S_{ij} ou sobre os postos $R(X_{ij})$.

Para k amostras, a ERA do teste de Friedman relativo ao teste paramétrico usual F é dependente do número de amostras k e igual a $(0{,}955)k/(k + 1)$ se as populações são normais, $k/(k + 1)$ se as populações são uniformes, e $3k/2(k + 1)$ se as populações são

duplo-exponenciais. De fato, a ERA do teste de Friedman em relação ao teste F nunca cai abaixo de $(0{,}864)k/(k+1)$ sob hipóteses alternativas de translação (NOETHER, 1967a).

Para $k = 2$, a ERA do teste de Quade em relação ao teste t é igual ao do teste de postos assinalados de Wilcoxon (0,955), quando as distribuições são normais. Da mesma forma que ocorre com o teste de Wilcoxon, a ERA relativa ao teste t nunca cai abaixo de 0,864, mas pode ser tão grande quanto infinito. A ERA desse teste para mais do que duas amostras nunca foi encontrada.

Exemplo de aplicação

Como exemplo de uso desses dois testes não paramétricos no delineamento em blocos completos casualizados, serão usadas 8 das 28 cultivares de arroz irrigado, testadas em um experimento conduzido pelo pesquisador Paulo Hideo Nakano Rangel (comunicação pessoal), na fazenda Palmital, da Embrapa Arroz e Feijão, no ano agrícola 2001–2002. Os dados referem-se ao número de dias para floração.

Considerando-se que esses dados aparecem sempre em uma amplitude que raramente ultrapassa os 40 dias, que o número de medidas por tratamento é, em geral reduzido (três ou quatro) e que são de natureza discreta, o uso de um teste não paramétrico é plenamente justificado. A Tabela 5 apresenta os resultados retirados desse ensaio e os respectivos postos por cultivar em cada bloco.

Teste de Friedman

Com os dados da Tabela 5, referentes aos postos por blocos, em que se observa a não ocorrência de empates, a aplicação do teste de Friedman pode ser feita com o uso da forma mais simples. Os cálculos necessários são aqueles requeridos pelas fórmulas 31 e 32.

$$A_2 = \frac{bk(k+1)(2k+1)}{6}, \quad B_2 = \frac{1}{b}\sum_{j=1}^{k} R_j^2$$

e

$$T_2 = \frac{(b-1)\left[B_2 - \frac{bk(k+1)^2}{4}\right]}{A_2 - B_2}$$

Tabela 5. Número de dias para floração, postos e identificação dos tratamentos (cultivares) em um experimento em blocos ao acaso.[1]

Cultivar	Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3	Total
	Número de dias para floração			Posto por bloco			
CNAi 8859	95	91	95	5	4	4	13
CNAi 8864	93	95	93	3	5	3	11
CNAi 8876	94	89	96	4	3	5	12
Jaburu	109	107	109	7	7	7	21
Bigua	104	103	103	6	6	6	18
CNA 8569	119	113	119	8	8	8	24
CNAi 9052	85	85	84	1	1	1	3
CNAi 9053	90	88	88	2	2	2	6
Total	789	771	787				

[1] Dados cedidos pelo pesquisador Paulo H. N. Rangel, da Embrapa Arroz e Feijão.

em que b é o número de blocos, k o número de tratamentos e R_j é o total dos postos de cada tratamento. No exemplo, $b = 3$, $k = 8$, e os valores R_j estão dados na Tabela 5, na última coluna, de forma que se tem, então,

$$A_2 = \frac{bk(k+1)(2k+1)}{6} = \frac{3*8(8+1)(2*8+1)}{6} = \frac{3.672}{6} = 612$$

$$B_2 = \frac{1}{b}\sum_{j=1}^{k} R_j^2 = \frac{1}{3}(13^2 + 11^2 + 12^2 + 21^2 + 18^2 + 24^2 + 3^2 + 6^2) = \frac{1.820}{3} = 606{,}67$$

e finalmente

$$T_2 = \frac{(b-1)\left[B_2 - \frac{bk(k+1)^2}{4}\right]}{A_2 - B_2} = \frac{2*\left[606{,}67 - \frac{3*8(8+1)^2}{4}\right]}{612 - 606{,}67} = \frac{241{,}34}{5{,}33} = 45{,}28$$

Nos Anexos 3 e 4, com 7 e 14 graus de liberdade encontram-se os valores 2,76 e 4,28 para 5% e 1% de probabilidade, respectivamente. Como o valor calculado de T_2 supera em muito esses valores, pode-se rejeitar a hipótese nula de igualdade de tratamentos e aceitar

a alternativa de diferença entre os cultivares no período de floração. A definição de onde se encontram essas diferenças será mostrada ao final do exemplo.

Teste de Quade

Neste caso, pode-se também aplicar o teste de Quade. Para isso, monta-se a Tabela 6, que contém, além dos postos dos tratamentos em cada bloco, os postos dos próprios blocos, definidos de acordo com o intervalo de valores observados em cada um deles e dos valores de postos ponderados, conforme prevê o teste. Os blocos devem ser também classificados (postos) de acordo com o valor da amplitude das medidas feitas em cada um deles, de forma que o bloco com menor amplitude de variação de dados receba o posto 1, o segundo maior o posto 2 e assim sucessivamente. Examinando os dados coletados na Tabela 5, verifica-se que as amplitudes dos três blocos são, respectivamente, 34 (119-85), 28 (113 - 85) e 35 (119 - 84), logo os postos dos blocos são 2, 1 e 3, como se vê na Tabela 6. Os valores ponderados para cada tratamento (S_{ij}) em cada um dos blocos é obtido pelo produto entre o posto do bloco e a diferença entre o posto do tratamento e a média dos postos, o que é dado pela fórmula

$$S_{ij} = Q_i\,[R(X_{ij}) - (k + 1)/2] \qquad [37]$$

em que Q_i é o posto do bloco, $R(X_{ij})$ é o posto do tratamento i no bloco j e $(k + 1)/2$ é a média dos postos em cada bloco. Para o tratamento 1 no bloco 1, por exemplo, tem-se

$$S_{11} = Q_1\,[R(X_{11}) - (k + 1)/2] = 2*[5 - (8 +1)/2] = 2*(5 - 4{,}5) = 1$$

Por sua vez, para o tratamento 2, no mesmo bloco, tem-se $S_{21} = 2*(3 - 4{,}5) = -3$, e assim sucessivamente.

Com os escores S_{ij}, procede-se o cálculo do teste de Quade, que, como foi visto anteriormente, necessita da soma de quadrados total (A_1) e da soma de quadrados de tratamento (B_1), para finalmente obter-se T_1, o teste de Quade, que tem distribuição F, com k - 1 e $(b - 1)(k - 1)$ graus de liberdade.

$$A_1 = \sum_{i=1}^{b}\sum_{j=1}^{k} S_{ij}^2 \;,\; B_1 = \frac{1}{b}\sum_{j=1}^{k} S_j^2 \quad \text{e} \quad T_l = \frac{(b-1)B_1}{A_1 - B_1}$$

Tabela 6. Postos e escores ponderados pelo posto dos blocos com o respectivo total e a identificação dos tratamentos (cultivares) em um experimento em blocos ao acaso.

Cultivar	Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3	Total
	Posto por bloco			Valor ponderados			
CNAi 8859	5	4	4	1,5	-0,5	-1,0	0
CNAi 8864	3	5	3	-4,5	0,5	-3,0	-7
CNAi 8876	4	3	5	-1,5	-1,5	1,0	-2
Jaburu	7	7	7	7,5	2,5	5,0	15
Bigua	6	6	6	4,5	1,5	3,0	9
CNA 8569	8	8	8	10,5	3,5	7,0	21
CNAi 9052	1	1	1	-10,5	-3,5	-7,0	-21
CNAi 9053	2	2	2	-7,5	-2,5	-5,0	-15
Posto do bloco	3	1	2				

Para os valores deste ensaio, tem-se que

$$A_1 = [1,5^2+(-4,5)^2+(-1,5)^2+...+(-5,0)^2] = 544$$

e

$$B_1 = (1/3)[(0)^2+(-7)^2+(-2)^2+...(-15)^2] = 488,67$$

logo,

$$T_1 = \frac{(b - 1)B_1}{A_1 - B_1} = \frac{2 * 488,67}{544 - 488,67} = 9,84$$

Como o teste tem $7\ (8 - 1)$ e $14\ (3 - 1)*(8 - 1)$ graus de liberdade, nos Anexos 3 e 4, os valores tabelados para esses graus de liberdade são, respectivamente, 2,76 e 4,28, com 5% e 1% de probabilidade. Como o valor calculado é muito maior que esses valores, assume-se que se pode rejeitar a hipótese nula e aceitar a alternativa de que existem diferenças entre os tratamentos, resultado concordante com o obtido pelo teste de Friedman.

Comparação de médias

Para a comparação de médias, sendo os testes significativos, pode-se usar a soma dos postos (Friedman) ou a soma dos escores de postos ponderados (Quade), cujas fórmulas de comparação são dadas a seguir.

$$T_1 = \frac{(b-1)B_1}{A_1 - B_1} = \frac{2 * 488{,}67}{544 - 488{,}67} = 9{,}84$$

e

$$|S_i - S_j| > t_{(1-\frac{\alpha}{2})}\sqrt{\frac{2b(A_1 - B_1)}{(b-1)(k-1)}}$$

Para esse exemplo, serão usadas as duas fórmulas, lembrando que, no uso real desses testes, apenas uma das formas deveria ser empregada, a que se segue ao teste empregado na análise dos dados. Em ambos os casos o valor do teste t, com 14 graus de liberdade e 5% de probabilidade é igual a 2,51; portanto, para o teste de Friedman tem-se:

$$|R_i - R_j| > t_{(1-\frac{\alpha}{2})}\sqrt{\frac{2b(A_1 - B_1)}{(b-1)(k-1)}} = 2{,}51\sqrt{\frac{2 * 3(612 - 606{,}67)}{2 * 7}} = 2{,}19$$

Logo, qualquer diferença entre as somas de postos de dois tratamentos que ultrapasse esse valor será considerada significativa, e os tratamentos diferentes entre si. A Tabela 7 mostra os resultados obtidos.

Para o teste de Quade, tem-se

$$|S_i - S_j| > t_{(1-\frac{\alpha}{2})}\sqrt{\frac{2b(A_1 - B_1)}{(b-1)(k-1)}} = 2{,}51\sqrt{\frac{2 * 3(588 - 488)}{2 * 7}} = 16{,}43$$

com a mesma interpretação do teste anterior, cujos resultados também estão na Tabela 7.

Tabela 7. Identificação dos tratamentos (cultivares), postos, escores ponderados e significância das diferenças entre pares de tratamentos pelos testes de Friedman e de Quade.

Cultivar	Total dos postos (Friedman)	Comparação dos postos	Total dos escores ponderados (Quade)	Comparação dos escores ponderados
CNA 8569	24	A	21	A
Jaburu	21	B	15	AB
Bigua	18	C	9	ABC
CNAi 8859	13	D	0	BCD
CNAi 8876	12	D	-2	CD
CNAi 8864	11	D	-7	CDE
CNAi 9053	6	E	-15	DE
CNAi 9052	3	F	-21	E
Diferença mínima		2,19		16,43

Uso do SAS

Para o teste de Friedman, usam-se dois procedimentos do SAS: primeiro o procedimento RANK para substituir os valores da(s) variável(is) por seus postos dentro de cada bloco e, em seguida, o procedimento ANOVA para a obtenção do teste F para tratamentos, que equivale ao teste de Friedman. A programação SAS para isso é a seguinte:

```
PROC SORT;BY BLOCO;
PROC RANK OUT=NP1;BY BLOCO;VAR variável;
PROC ANOVA DATA=NP1;CLASS BLOCO TRAT;
MODEL variável=BLOCO TRAT;MEANS TRAT;
```

O resultado dessa análise encontra-se na Figura 4.

Como seria de se esperar, a soma de quadrados de blocos é igual a 0 (zero), uma vez que os dados foram transformados em seus postos, dentro de cada bloco; portanto, todos os blocos têm valor total igual. Note-se que a soma de quadrado de erro corresponde à diferença entre A_2 e B_2, igual a 5,33, calculada no teste de Friedman. A pequena diferença encontrada entre os valores do teste F e do teste de Friedman deve-se à aproximação de cálculos. A comparação de médias, nesse caso, não leva aos mesmos resultados que a comparação de totais, como foi mostrado na seção *Teste de Friedman*.

```
                          The ANOVA Procedure
                       Class Level Information
Class       Levels  Values
BLOCO            3  1 2 3
TRAT             8  BIGUA CNA.8569 CNAi.8859 CNAi.8864 CNAi.8876
                    CNAi.9052 CNAi.9053 JABURU
                          Number of observations    24

                          The ANOVA Procedure
Dependent Variable: DIAS   Values of DIAS Were Replaced by Ranks
                                Sum of
 Source                DF       Squares    Mean Square   F Value   Pr > F
 Model                  9    120.6666667     13.4074074     35.19   <.0001
 Error                 14      5.3333333      0.3809524
 Corrected Total       23    126.0000000

               R-Square     Coeff Var      Root MSE     DIAS Mean
               0.957672      13.71585      0.617213      4.500000

 Source                DF      Anova SS    Mean Square   F Value   Pr > F
 BLOCO                  2     0.0000000      0.0000000      0.00   1.0000
 TRAT                   7    120.6666667     17.2380952     45.25   <.0001
                          The ANOVA Procedure
               Level of              -------------DIAS------------
               TRAT          N           Mean          Std Dev
               BIGUA         3      6.00000000       0.00000000
               CNA.8569      3      8.00000000       0.00000000
               CNAi.8859     3      4.33333333       0.57735027
               CNAi.8864     3      3.66666667       1.15470054
               CNAi.8876     3      4.00000000       1.00000000
               CNAi.9052     3      1.00000000       0.00000000
               CNAi.9053     3      2.00000000       0.00000000
               JABURU        3      7.00000000       0.00000000
```

Figura 4. Análise de variância de um experimento em blocos completos ao acaso, com o uso de postos para o teste de Friedman.

Para o teste de Quade, o programa SAS é um pouco mais elaborado, pois existe a necessidade de obter os postos dos blocos e depois calcular os escores típicos do teste de Quade, e esse programa se encontra a seguir. O valor 4,5, subtraído do posto do tratamento em cada bloco, corresponde ao valor médio dos postos dentro dos blocos.

```
PROC SORT;BY BLOCO;
PROC MEANS SUM NOPRINT;BY BLOCO;VAR DIAS;
OUTPUT OUT=NP0 SUM=Q;
PROC RANK OUT=NP01;VAR Q;
DATA NP1;MERGE NP NP01;BY BLOCO;
PROC RANK OUT=NP1;BY BLOCO;VAR DIAS;
DATA NP1;SET NP1;SCORE=Q*(DIAS-4.5);
PROC ANOVA DATA=NP1;CLASS BLOCO TRAT;
MODEL SCORE=BLOCO TRAT;MEANS TRAT;
```

O resultado desse processamento está representado na Figura 5.

```
                          The ANOVA Procedure
                         Class Level Information
Class       Levels  Values
BLOCO            3  1 2 3
TRAT             8  BIGUA CNA.8569 CNAi.8859 CNAi.8864 CNAi.8876
                    CNAi.9052 CNAi.9053 JABURU
                         Number of observations    24
                          The ANOVA Procedure
Dependent Variable: SCORE
                                 Sum of
 Source                 DF       Squares    Mean Square   F Value   Pr > F
 Model                   9    488.6666667     54.2962963      7.65   0.0005
 Error                  14     99.3333333      7.0952381
 Corrected Total        23    588.0000000

              R-Square     Coeff Var      Root MSE    SCORE Mean
              0.831066       Infty        2.663689             0

 Source                 DF      Anova SS    Mean Square   F Value   Pr > F
 BLOCO                   2     0.0000000      0.0000000      0.00   1.0000
 TRAT                    7    488.6666667     69.8095238      9.84   0.0002

                          The ANOVA Procedure
              Level of              ------------SCORE------------
              TRAT          N             Mean          Std Dev
              BIGUA         3       3.00000000       1.50000000
              CNA.8569      3       7.00000000       3.50000000
              CNAi.8859     3       0.00000000       1.32287566
              CNAi.8864     3      -2.33333333       2.56580072
              CNAi.8876     3      -0.66666667       1.44337567
              CNAi.9052     3      -7.00000000       3.50000000
              CNAi.9053     3      -5.00000000       2.50000000
              JABURU        3       5.00000000       2.50000000
```

Figura 5. Análise de variância de um experimento em blocos completos ao acaso usando os escores de Quade.

Os valores aqui encontrados correspondem àqueles obtidos anteriormente, e as observações referentes às comparações de médias são as mesmas feitas para o teste de Friedman.

Delineamento em blocos incompletos balanceados

No delineamento em blocos incompletos casualizados (Capítulo 5), apenas parte dos tratamentos são aplicados a cada um dos blocos. Além disso, para que o delineamento seja balanceado, exige-se que:

1) Cada bloco contenha k unidades experimentais.
2) Cada tratamento apareça em r blocos.
3) Cada tratamento apareça com cada outro tratamento em um número igual de vezes.

Durbin (1951) apresentou um teste de postos que pode ser usado para testar a hipótese nula de não haver diferenças entre tratamentos nesses casos. Deve-se preferir o teste de Durbin aos testes paramétricos se as pressuposições de normalidade não são obedecidas, se um método mais fácil é desejado, ou se as observações consistem meramente de postos. O teste de Durbin reduz-se ao teste de Friedman se o número de tratamentos é igual ao número de unidades experimentais por bloco. Se a terceira condição dada não é completamente satisfeita, o teste de Durbin é ainda válido na maioria das situações.

Teste de Durbin

Será usada aqui a mesma notação usada no Capítulo 5, isto é:

t = número de tratamentos a serem examinados.

k = número de unidades experimentais por bloco ($k < t$).

b = número total de blocos.

r = número de vezes que cada tratamento aparece ($r < b$).

λ = número de blocos nos quais os tratamentos i e j aparecem juntos (λ é o mesmo para todos os pares de tratamentos).

Representando por X_{ij} os resultados do tratamento j no bloco i se o tratamento j aparece no bloco i, deve-se primeiro ordenar os X_{ij} dentro de cada bloco, assinalando o posto 1 à menor observação no bloco, posto 2 à segunda menor, e assim por diante, até o posto k, que é assinalado à maior de todas as observações no bloco i, já que existem apenas k observações dentro de cada bloco. Denote-se por $R(X_{ij})$ o posto de X_{ij}, em que X_{ij} existe. O passo seguinte é computar a soma dos postos assinalados aos r valores observados para o j-ésimo tratamento e chamar essa soma de R_j. Então, R_j pode ser escrito como

$$R_j = \sum_{i=1}^{b} R(X_{ij}) \qquad [38]$$

em que somente r valores de $R(X_{ij})$ existem para cada tratamento j.

Se as observações são "não numéricas", mas passíveis de ser ordenadas dentro dos blocos de acordo com algum critério de interesse, o posto de cada observação é anotado e os valores de R_j são calculados como antes.

Se houver empate, recomenda-se assinalar o posto médio às observações empatadas. Esse procedimento muda a distribuição nula do teste estatístico, mas o efeito é desprezível se o número de empates não for excessivo.

As pressuposições feitas são:

1) Os blocos são mutuamente independentes um do outro.
2) Dentro de cada bloco, as observações podem ser ordenadas em ordem crescente, de acordo com algum critério de interesse (um número moderado de empates pode ser tolerado).

As hipóteses são as mesmas feitas para os delineamentos anteriores.

O teste estatístico de Durbin é definido como

$$T = \frac{12(t - 1)}{rt(k - 1)(k + 1)} \sum_{j=1}^{t} [R_j - \frac{r(k + 1)}{2}]^2 \quad [39]$$

e pode ser escrito numa forma equivalente:

$$T = \frac{12(t - 1)}{rt(k - 1)(k + 1)} \sum_{j=1}^{t} R_j^2 - \frac{3r(t - 1)(k + 1)}{k - 1} \quad [40]$$

Rejeita-se a hipótese nula a um nível aproximado de significância se o teste estatístico T de Durbin exceder o quantil $1 - \alpha$ de uma variável aleatória qui-quadrado com $t - 1$ graus de liberdade.

Distribuição de probabilidade

O desenvolvimento teórico do teste de Durbin (DURBIN, 1951) é muito parecido com o do teste de Friedman. Isto é, a distribuição exata do teste estatístico de Durbin é encontrada sob a pressuposição de que cada arranjo dos k postos dentro de um bloco é igualmente provável visto que não há diferenças entre tratamentos. Existem $k!$ maneiras igualmente prováveis de arranjar os postos dentro de cada bloco, e existem b blocos. Portanto, cada arranjo de postos sobre o conjunto total de b blocos é igualmente provável e tem probabilidade $1/(k!)^b$ associado com ele, porque existem $(k!)^b$ diferentes arranjos possíveis. O teste estatístico de Durbin é calculado para cada conjunto e então a função de distribuição é determinada, da mesma forma como ocorreu com o teste estatístico de Friedman. A distribuição exata não é fácil de ser encontrada na maioria dos casos, de forma que a distribuição do teste estatístico de Durbin é aproximada pela distribuição de qui-quadrado com $t - 1$ graus de liberdade, se o número de repetições r de cada tratamento é grande. Na prática, a aproximação é usada mesmo se r for tão pequeno quanto 2 ou 3, por força da necessidade. O nível α declarado é provavelmente não muito preciso nessas circunstâncias.

Noether (1967b) discute o teste de Durbin e suas generalizações e mostra que a ERA do teste de Durbin, relativamente ao seu equivalente teste paramétrico, é a mesmo do teste de Friedman.

No caso de significância do teste de Durbin, dois tratamentos i e j podem ser considerados diferentes se as suas somas de postos satisfazem a inigualdade

$$|R_j - R_i| > t_{(1-\frac{\alpha}{2})} \sqrt{\frac{r(k+1)(k-1)[bk(t-1)-tT]}{6(t-1)(bk-t-b+1)}}$$

em que $t_{(1-\alpha/2)}$ é o quantil $1 - \alpha/2$ da distribuição de t com $bk - t - b + 1$ graus de liberdade

Exemplo de aplicação

Do mesmo ensaio de competição de cultivares realizado em 1994–1995 pelos pesquisadores Orlando Peixoto Moraes, Emílio da Maia de Castro e Flávio Brezeghelo, da Embrapa Arroz e Feijão, em um delineamento em reticulado quadrado 5 x 5, com três repetições, mostra-se a Tabela 8 com os dados referentes ao número de dias para floração e os postos equivalentes, segundo a metodologia definida no teste de Durbin.

Tabela 8. Dados de floração e respectivos postos em um ensaio em reticulado quadrado.

Repetição 1	Bloco 1	Tratamento	1	2	3	4	5
		Floração (posto)	95 (4,5)	84 (3)	95 (4,5)	81 (1,5)	81 (1,5)
	Bloco 2	Tratamento	6	7	8	9	10
		Floração (posto)	102 (5)	89 (3)	101 (4)	77 (1)	79 (2)
	Bloco 3	Tratamento	11	12	13	14	15
		Floração (posto)	112 (5)	103 (3)	102 (2)	108 (4)	100 (1)
	Bloco 4	Tratamento	16	17	18	19	20
		Floração (posto)	102 (4)	110 (5)	95 (3)	83 (2)	81 (1)
	Bloco 5	Tratamento	21	22	23	24	25
		Floração (posto)	82 (2)	104 (5)	102 (4)	79 (1)	87 (3)

Continua...

Tabela 8. Continuação.

Repetição 2	Bloco 1	Tratamento	1	6	11	16	21
		Floração (posto)	78 (1)	103 (4)	107 (5)	99 (3)	79 (2)
	Bloco 2	Tratamento	2	7	12	17	22
		Floração (posto)	83 (1,5)	83 (1,5)	103 (4)	104 (5)	100 (3)
	Bloco 3	Tratamento	3	8	13	18	23
		Floração (posto)	94 (2)	102 (5)	101 (4)	83 (1)	99 (3)
	Bloco 4	Tratamento	4	9	14	19	24
		Floração (posto)	81 (3)	74 (1)	107 (5)	80 (2)	82 (4)
	Bloco 5	Tratamento	5	10	15	20	25
		Floração (posto)	77 (2,5)	75 (1)	101 (5)	77 (2,5)	84 (4)
Repetição 3	Bloco 1	Tratamento	2	8	14	20	21
		Floração (posto)	94 (3)	100 (4)	110 (5)	79 (1)	81 (2)
	Bloco 2	Tratamento	3	9	15	16	22
		Floração (posto)	91 (2)	74 (1)	101 (4)	98 (3)	107 (5)
	Bloco 3	Tratamento	4	10	11	17	23
		Floração (posto)	79 (2)	75 (1)	105 (5)	104 (4)	100 (3)
	Bloco 4	Tratamento	5	6	12	18	24
		Floração (posto)	75 (1)	102 (4)	103 (5)	84 (3)	76 (2)
	Bloco 5	Tratamento	1	7	13	19	25
		Floração (posto)	81 (2)	86 (4)	101 (5)	78 (1)	82 (3)

Dada a utilização de posto, por bloco, como resposta da variável de interesse, assim como nos casos dos testes de Friedman e de Quade, os efeitos de blocos e de repetições são exatamente iguais, portanto assumidos zero (0), de forma que apenas as somas de quadrados, total e de tratamentos devem ser calculadas. Os totais de tratamentos são mostrados na Tabela 9.

Tabela 9. Identificação e total dos postos por tratamento de um ensaio em blocos incompletos.

Tratamentos	1	2	3	4	5
Total	7,5	7,5	8,5	6,5	5
Tratamentos	6	7	8	9	10
Total	13	8,5	13	3	4
Tratamentos	11	12	13	14	15
Total	15	12	11	14	10
Tratamentos	16	17	18	19	20
Total	10	14	7	5	4,5
Tratamentos	21	22	23	24	25
Total	6	13	10	7	10

Considerando-se desprezível o número de empates verificado neste experimento (apenas quatro), o teste de Durbin pode ser aplicado sem grandes problemas na sua distribuição de probabilidade. Como se viu anteriormente, a forma do teste é

$$T = \frac{12(t - 1)}{rt(k - 1)(k + 1)} \sum_{j=1}^{t} [R_j - \frac{r(k + 1)}{2}]^2$$

em que, nesse caso, $t = 25$, $r = 3$, $k = 5$ e os R_j são os totais de tratamentos dados no quadro acima. Portanto, tem-se que

$$\frac{12(t - 1)}{rt(k - 1)(k + 1)} = \frac{12(25 - 1)}{3 * 25(5 - 1)(5 + 1)} = 0{,}16$$

e

$$\frac{r(k + 1)}{2} = \frac{3(5 + 1)}{2} = 9$$

logo

$$T = 0{,}16*[(7{,}5 - 9)^2 + (7{,}5 - 9)^2 + (8{,}5 - 9)^2 + ... + (10 - 9)^2] = 0{,}16+292{,}50 = 46{,}50$$

Como se sabe, T tem distribuição de qui-quadrado (χ^2), com, neste caso, 24 graus de liberdade. O Anexo 1 dá como valores críticos para 5% e 1%, respectivamente, 36,42 e

42,98. Como o valor calculado supera esses limites, pode-se assumir que a igualdade entre tratamentos para número de dias para floração pode ser rejeitada. Para se compararem os tratamentos, pode-se utilizar a forma do teste da diferença mínima significativa dada em *Delineamento em blocos incompletos balanceados*. Tem-se que quaisquer dois tratamentos cujos totais de postos difiram por um valor maior que

$$t_{(1-\frac{\alpha}{2})}\sqrt{\frac{r(k+1)(k-1)[bk(t-1)-tT]}{6(t-1)(bk-t-b+1)}}$$

em que t tem 36 graus de liberdade $(bk - t - b + 1)$, serão considerados diferentes.

Para esse ensaio, o valor é igual a 6,92 (ver cálculo a seguir):

$$2{,}34 * \sqrt{\frac{3(5+1)(5-1)[15*5(25-1)-25*46{,}80]}{6(25-1)(15*5-25-15+1)}} = 6{,}92$$

Uso do SAS

Para calcular o teste de Durbin com o auxílio do SAS, existe a necessidade de um processamento próprio, pois os procedimentos de análise de variância do software não apresentam este teste. Esse programa é dado a seguir para o exemplo que se está discutindo, no qual a base é cálculo dos totais dos tratamentos com base nos postos e necessita de pequenas alterações para outros delineamentos, no que se refere aos valores 9 e 0,16, que aparecem na fórmula do teste e que são função dos números de tratamentos e de repetições e do tamanho dos blocos.

```
PROC SORT;BY REP BLOCO;
PROC RANK OUT=NP1;BY REP BLOCO;VAR FLOR;
PROC SORT;BY TRAT;
PROC MEANS SUM NOPRINT;BY TRAT;VAR FLOR;
OUTPUT OUT=POSTO SUM=FLOR;
DATA POSTO;SET POSTO;VAL=(FLOR-9)**2;
PROC MEANS SUM NOPRINT;VAR VAL;
OUTPUT OUT=DURB SUM=D;
DATA DURB;SET DURB;TESTE=D*0.16;
PROC PRINT;VAR TESTE;
TITLE "TESTE DE DURBIN PARA BLOCOS INCOMPLETOS";
```

O resultado é a seguinte informação, provida pelo SAS, na qual fica faltando apenas a significância do teste.

```
TESTE DE DURBIN PARA BLOCOS INCOMPLETOS
                    Obs    TESTE
                     1      46.8
```

Delineamento quadrado latino

Como visto no Capítulo 4, no delineamento em quadrado latino, três fontes de variação são investigadas simultaneamente, desde que não haja interação entre os três fatores e todos eles tenham o mesmo número de níveis.

Usualmente, na pesquisa agrícola um fator é chamado "linha", o outro "coluna", e o terceiro são os tratamentos em teste.

Neste delineamento, existem duas restrições de casualização a fim de que cada tratamento apareça em cada linha e em cada coluna, uma única vez.

Dois testes serão estudados neste item: o de Bennett (1966) e o de Zimmermann (1983). O primeiro é uma extensão de outros testes já vistos anteriormente, enquanto o segundo introduz a noção da soma de postos como variável substituta na análise.

Teste de Bennett

Também aqui se considera X_{ijk} o valor da variável associada à linha i, à coluna j e ao tratamento k. Além disso, assume-se que $R(X_{ijk})$ é o seu posto, de 1 a N, em que $N = n^2$. Isso quer dizer que todas as N observações são ordenadas, dando-se posto 1 à menor delas, posto 2 à segunda menor, e assim por diante, até posto N, ao qual é dado a observação de maior valor. No caso de empate entre valores, deve ser usado o posto médio.

Finalmente, calcula-se $R_{..k}$, a soma dos postos assinalados ao tratamento k, isto é,

$$R_{..k} = \sum_{ij} X_{ijk}$$

para $k = 1, 2, ..., n$.

As pressuposições necessárias são as mesmas duas feitas para outros delineamentos, isto é:

1) As n variáveis aleatórias n-variadas são mutuamente independentes,

2) As observações podem ser ordenadas, de acordo com algum critério de interesse.

Da mesma forma, as hipóteses nula e alternativa são:

H_0: cada ordenação das variáveis aleatórias é igualmente provável, isto é, os tratamentos têm efeitos idênticos.

H_1: pelo menos um tratamento tende a produzir valores maiores do que pelo menos um dos outros tratamentos.

O teste estatístico é dado por

$$B = \frac{12}{N(N+1)}\left(\frac{1}{n}\sum_{k} R_{..k}^{2}\right) - 3(N+1)$$

A hipótese nula é rejeitada ao nível α se B exceder o quantil $1 - \alpha$ da distribuição de qui-quadrado com $n - 1$ graus de liberdade.

A distribuição de qui-quadrado é a distribuição assintótica de B, e Zimmermann (1983) mostrou que B tem a mesma distribuição do teste de Kruskal-Wallis.

Teste de Zimmermann

Como no teste anterior, os dados consistem de n n-variadas variáveis aleatórias, mutuamente independentes. Deve-se encontrar o posto de cada observação primeiro dentro de cada linha e depois dentro de cada coluna e, então, somar os dois postos para obter uma nova variável, chamada Z_{ijk}.

Com essa variável, obtêm-se as somas de quadrados total e de tratamentos, como na análise paramétrica usual.

1) As n variáveis aleatórias n-variadas são mutuamente independentes.

2) As observações podem ser ordenadas dentro das linhas e dentro das colunas por algum critério de interesse.

As mesmas pressuposições feitas para o teste de Bennett podem ser feitas para o teste de Zimmermann.

As hipóteses nula e alternativa agora apresentam a seguinte expressão:

H_0: cada ordenação das variáveis aleatórias nas linhas e/ou nas colunas é igualmente provável, isto é, os tratamentos têm efeitos idênticos.

H_1: pelo menos um tratamento tende a produzir valores maiores do que pelo menos um dos outros tratamentos.

O teste estatístico é dado por

$$Z = [nT/(1 - T)]^{(n-1)/(n+1)}$$

em que

$$T = SQtrat/SQtotal$$

Zimmermann (1983) também mostrou que

$$Z' = nT/(1 - T)$$

tem assintoticamente uma distribuição de F com $n - 1$ e $n(n - 1)$ graus de liberdade, além de ter mostrado que é necessário o uso do expoente $(n - 1)/(n + 1)$ para uma melhor aproximação à distribuição de F, em razão do tamanho das amostras, normalmente trabalhadas nesse delineamento.

Se $T = 1$, considere-se Z estar na região crítica com um nível de significância dado por $(1/n!)^{n-1}$.

Exemplo de aplicação

Do mesmo ensaio utilizado como exemplo de aplicação de aditividade em quadrados latinos (Tabela 15 do Capítulo 4), conduzido por Evane Ferreira, apresentam-se a seguir os dados de colmos atacados por *E. lignosellus* L.

Teste de Bennett

Na Tabela 10, além das identificações de linhas, colunas e tratamentos, mostram-se os números de colmos atacados e o posto equivalente segundo a metodologia prescrita pelo teste de Bennett, na qual empates recebem o posto médio das observações empatadas.

Tabela 10. Número de colmos atacados por *E. lignosellus* L. e posto (entre parêntesis), segundo metodologia de Bennett (1966) em um quadrado latino 6 x 6.

Linha		Colunas						Total de linhas	Total de tratamentos
		1	2	3	4	5	6		
1	Tratamento	5	2	3	4	1	6		1
	CAE	20 (34)	5 (7)	8 (15)	6 (9,5)	8 (15)	5 (7)	87,5	124,5
2	Tratamento	4	6	2	1	3	5		2
	CAE	17 (32,5)	4 (4)	9 (19,5)	7 (11,5)	8 (15)	16 (30,5)	113	102,5
3	Tratamento	3	4	5	2	6	1		3
	CAE	16 (30,5)	8 (15)	9 (19,5)	4 (4)	10 (23)	9 (19,5)	111,5	122
4	Tratamento	1	5	6	3	2	4		4
	CAE	8 (15)	14 (29)	5 (7)	3 (1,5)	11 (25,5)	10 (23)	101	112,5
5	Tratamento	6	3	1	5	4	2		5
	CAE	4 (4)	17 (32,5)	12 (27,5)	9 (19,5)	6 (9,5)	7 (11,5)	104,5	158
6	Tratamento	2	1	4	6	5	3		6
	CAE	22 (35)	30 (36)	10 (23)	3 (1,5)	11 (25,5)	12 (27,5)	148,5	46,5
Total de colunas		151	124	111,5	47,5	113,5	119	666	

CAE = colmos atacados por *E. lignosellus* L.

O teste apresenta a forma descrita no item *Teste de Bennett*, isto é,

$$B = \frac{12}{N(N+1)}\left(\frac{1}{n}\sum_{k} R_{..k}^{2}\right) - 3(N+1)$$

em que $N = 36$, $n = 6$ (o número básico deste quadrado latino) e $R_{..k}$ é a soma dos postos do tratamento k.

Nesse caso, tem-se o seguinte resultado:

$$B = \frac{12}{36(36+1)}\left(\frac{1}{6}(124{,}5^2 + 102{,}5^2 + 122^2 + 112{,}5^2 + 158^2 + 46{,}5^2)\right) -$$
$$3(36+1) = 10{,}13$$

Como a distribuição do teste de Bennett é a de qui-quadrado com cinco graus de liberdade, verifica-se no Anexo 1 que os níveis críticos para 5% e 1% são, respectivamente, 11,07 e 15,09. Como o valor do teste é inferior a esses valores, não se pode rechaçar a hipótese nula de igualdade entre os tratamentos.

Teste de Zimmermann

Para o teste de Zimmermann (ZIMMERMANN, 1983; ZIMMERMANN; ROBERTSON, 1990), existe a necessidade de obtenção de escores dados pela soma dos postos de cada tratamento, ordenados tanto por linha quanto por coluna. Os resultados da aplicação desses postos e sua soma encontram-se detalhados na Tabela 11. Os números entre parêntesis indicam, o primeiro, o posto na linha e, o segundo, o posto na coluna. O valor fora dos parêntesis é a soma desses postos e são os valores com os quais se deve proceder à análise. Os dados originais estão detalhados na Tabela 10.

Para a aplicação do teste de Zimmermann, deve-se obter primeiro, com o uso das fórmulas já tão conhecidas, as somas de quadrados de tratamentos e total, para depois trabalhar a forma do teste que tem uma distribuição de F com n - 1 e $n(n$ - 1) graus de liberdade. No exemplo que se está discutindo, têm-se, portanto, 5 e 30 graus de liberdade para o teste.

A soma de quadrados total é igual a

$$SQTo = \sum Z_{ijk}^2 - \frac{1}{36}\left(\sum Z_{ijk}\right)^2 = (11^2 + 3{,}5^2 + 6{,}5^2 + \ldots + 9^2) -$$
$$\frac{1}{36}252^2 = 2.102{,}50 - 1.764{,}00 = 338{,}50$$

enquanto a soma de quadrados de tratamentos é:

$$SQTrat = \frac{1}{6}\sum Z_{..k}^2 - \frac{1}{36}\left(\sum Z_{ijk}\right)^2 = \frac{1}{6}(48{,}5^2 + 41{,}5^2 + 43{,}5^2 + 40^2 + 57{,}5^2 + 21^2) -$$
$$1.764{,}50 = 121{,}67$$

Tabela 11. Escores de Zimmermann e postos por linha e por coluna (parcelas da soma entre parêntesis), segundo metodologia de Zimmermann (ZIMMERMANN, 1983; ZIMMERMANN; ROBERTSON, 1990), em um quadrado latino 6 x 6, para o número de colmos atacados por *E. lignosellus* L. mostrados na Tabela 10.

Linha		Coluna						Total de linhas	Total de tratamentos
		1	2	3	4	5	6		
1	Tratamento	5	2	3	4	1	6		1
	CAE	11 (6+5)	3,5 (1,5+2)	6,5 (4,5+2)	7 (3+4)	7 (4,5+2,5)	2,5 (1,5+1)	37,5	48,5
2	Tratamento	4	6	2	1	3	5		2
	CAE	10 (6+4)	2 (1+1)	7,5 (4+3,5)	7 (2+5)	5,5 (3+2,5)	11 (5+6)	43,0	41,5
3	Tratamento	3	4	5	2	6	1		3
	CAE	9 (6+3)	5 (2+3)	7 (3,5+3,5)	4 (1+3)	9 (5+4)	6,5 (3,5+3)	40,5	43,5
4	Tratamento	1	5	6	3	2	4		4
	CAE	5 (3+2)	10 (6+4)	3 (2+1)	2,5 (1+1,5)	10,5 (5+5,5)	8 (4+4)	39,0	40,0
5	Tratamento	6	3	1	5	4	2		5
	CAE	2 (1+1)	11 (6+5)	11 (5+6)	10 (4+6)	3 (2+1)	5 (3+2)	42,0	57,5
6	Tratamento	2	1	4	6	5	3		6
	CAE	11 (5+6)	12 (6+6)	7 (2+5)	2,5 (1+1,5)	8,5 (3+5,5)	9 (4+5)	50,00	21,0
Total de colunas		48,0	43,5	42,0	33,0	43,5	42,0	252,0	

CAE = colmos atacados por *E. lignosellus* L.

Esses valores obtidos podem ser agora aplicados para calcular o teste de Zimmermann, dado pela fórmula

$$Z = [nT/(1 - T)]^{(n-1)/(n+1)}$$

em que $T = SQ\ trat/SQ\ total$. Portanto, têm-se:

$$T = 121{,}67/338{,}50 = 0{,}3594$$

e

$$Z = [(6*0{,}3594)/(1 - 0{,}3594)]^{5/7} = (3{,}3666)^{0{,}7142} = 2{,}38$$

Pelo Anexo 3, para 5 e 30 graus de liberdade, os valores críticos do teste F a 5% e 1% de probabilidade são, respectivamente, 2,53 e 3,70. Aqui, novamente não é possível rejeitar a hipótese nula, ou seja, a de igualdade entre os tratamentos.

Uso do SAS

Para este delineamento, o teste de Bennett é de fácil obtenção, uma vez que foi demonstrado que ele corresponde ao teste de Kruskal-Wallis; portanto, o procedimento SAS adequado é o NPAR1WAY. O programa SAS é:

```
PROC NPAR1WAY WILCOXON;CLASS TRAT;VAR COLMOS;
```

O resultado do processamento se vê na Figura 6, na qual a pequena diferença no valor do teste de qui-quadrado se deve a aproximações de cálculo.

Para a obtenção do teste de Zimmermann, uma programação mais extensa é necessária e, para o caso presente, ela é dada a seguir, lembrando que o valor de n para o exemplo é igual a 6 (dimensão do quadrado latino), e que esse valor aparece no cálculo de Z, primeiro multiplicando o valor de T e depois como expoente da função, que é $[(n - 1)/(n + 1)]$.

```
PROC SORT DATA=QLNP;BY LINHA;
PROC RANK OUT=LINHAR;BY LINHA;VAR COLMOS;
DATA LINHAR;SET LINHAR;LR=COLMOS;KEEP LINHA COLUNA TRAT LR;
PROC SORT;BY LINHA COLUNA TRAT;
PROC SORT DATA=QLNP;BY COLUNA;
PROC RANK OUT=COLUNAR;BY COLUNA;VAR COLMOS;
DATA COLUNAR;SET COLUNAR;CR=COLMOS;KEEP LINHA COLUNA TRAT CR;
PROC SORT;BY LINHA COLUNA TRAT;
DATA POSTO;MERGE LINHAR COLUNAR;BY LINHA COLUNA TRAT;
ESCORE=LR+CR;*PROC PRINT;
PROC ANOVA OUTSTAT=D NOPRINT;CLASS LINHA COLUNA TRAT;
MODEL ESCORE=LINHA COLUNA TRAT;
DATA D1;SET D;
IF _SOURCE_ EQ "TRAT";FONTE1=_SOURCE_;KEEP FONTE1 SS;
PROC MEANS DATA=D SUM NOPRINT;VAR SS;
OUTPUT OUT=D2 SUM=SS1;
DATA D2;SET D2;FONTE2="TOTAL";
DATA FIM;MERGE D1 D2;T=SS/SS1;Z=((6*T)/(1-T))**(5/7);
PROC PRINT;VAR FONTE1 SS FONTE2 SS1 T Z;
```

O resultado de tal programação para os dois testes, está na Figura 6, na qual falta apenas a significância do teste estatístico de Zimmermann.

TESTE DE BENNETT PARA O QUADRADO LATINO

The NPAR1WAY Procedure
Wilcoxon Scores (Rank Sums) for Variable COLMOS
Classified by Variable TRAT

TRAT	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
5	6	158.00	111.0	23.484037	26.333333
2	6	102.50	111.0	23.484037	17.083333
3	6	122.00	111.0	23.484037	20.333333
4	6	112.50	111.0	23.484037	18.750000
1	6	124.50	111.0	23.484037	20.750000
6	6	46.50	111.0	23.484037	7.750000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	10.1949
DF	5
Pr > Chi-Square	0.0699

TESTE DE ZIMMERMANN PARA O QUADRADO LATINO

Obs	FONTE1	SS	FONTE2	SS1	T	Z
1	TRAT	121.667	TOTAL	338.5	0.35943	2.37996

Figura 6. Uso do procedimento NPAR1WAY para a análise não paramétrica de um quadrado latino com o teste de Bennett.

Generalização da transformação de posto

A maior parte dos procedimentos não paramétricos apresentados são exemplos de procedimentos que apareceram por aplicar-se à transformação de posto aos dados (isto é, substituir os dados por seus postos) e então pelo uso do processo paramétrico. O uso mais óbvio dessa transformação foi visto no estudo de correlação, quando o processo usual de obtenção do coeficiente de correlação conhecido como r de Pearson foi aplicado aos postos e chamado de rho (r) de Spearman. Também a forma usual de obtenção de linhas de regressão foi utilizada sobre os postos.

Outros procedimentos não paramétricos, tais como o teste de Wilcoxon, o teste de Mann-Whitney e o teste de Kruskal-Wallis, são aplicações menos óbvias da transformação de posto. É fácil mostrar que o teste de Wilcoxon dos postos assinalados é o equivalente ao teste t para uma amostra utilizando-se os dados transformados em postos e que o teste de

Mann-Whitney é o equivalente ao teste t para duas amostras depois da transformação de posto.

Pode ser mostrado também que a estatística F computada sobre os postos das observações é uma função do teste T de Kruskal-Wallis na seguinte forma

$$FR = \frac{T/(k\text{-}1)}{(N - 1 - T)/(N - k)}$$

e porque FR aumenta ou diminui, acompanhando os aumentos ou diminuições de T, o procedimento de transformação de posto é equivalente ao teste de Kruskal-Wallis.

Esses exemplos e outros neste trabalho, tais como o teste de inclinação de linha, o teste de Friedman, o teste de Durbin e o teste de Zimmermann, ilustram a ideia de aplicar os testes estatísticos paramétricos usuais, ou seus equivalentes, sobre os postos das observações para obter um procedimento não paramétrico com alta eficiência na maioria das situações. Por certo, o truque em cada caso é ordenar as observações, de tal maneira que todas as possíveis ordenações sejam igualmente prováveis sob a hipótese nula. Essa técnica foi utilizada com sucesso por Worsley (1977) na análise de conglomerados, e por Shirley (1977) para contrastar doses crescentes de um tratamento.

Em situações em que tais métodos de ordenação não sejam possíveis, ou sejam difíceis, o princípio da transformação de posto pode ainda ser utilizado com sucesso. Duas áreas da estatística na qual a transformação de posto é útil, mesmo sabendo que ela não resulta em procedimentos não paramétricos, são as áreas de delineamentos experimentais e de regressão múltipla.

Para analisar um delineamento experimental utilizando a transformação de posto, primeiramente é necessário ordenar todas as observações da menor até a maior e então aplicar a análise de variância normal aos postos. O resultado é um procedimento que é apenas condicionalmente de distribuição livre. Entretanto, ele é "robusto", o que significa que o verdadeiro nível de significância é bastante próximo ao nível de significância utilizado no teste, não importando qual distribuição de probabilidade tem a população. O procedimento resultante usualmente tem boa eficiência (CONOVER; IMAN, 1976; IMAN, 1974).

O procedimento recomendado nos delineamentos experimentais para os quais não existem testes não paramétricos é utilizar a análise de variância comum sobre os dados e de novo sobre os dados transformados em postos. Se os dois procedimentos derem resultados praticamente idênticos, as pressuposições que fundamentam a análise de variância provavelmente serão razoáveis e a análise paramétrica é válida. Quando os dois procedimentos dão resultados substancialmente diferentes, a análise com os postos é provavelmente mais

precisa do que aquela sobre os dados originais e deveria ser preferida. Essa transformação de posto foi usada em delineamentos experimentais por Crouse (1967, 1968), Lemmer e Stoker (1967), MacDonald (1971), Hamilton (1976) e Scheirer et al. (1976).

Em regressão múltipla, cada variável é ordenada separadamente, como na regressão bivariada. Os métodos de regressão usuais são aplicados sobre os postos. O resultado é um método de regressão robusto, que não é sensível a valores estranhos ou distribuições não normais, como o são os métodos aplicados sobre os dados. Como antes, o recomendado é analisar os dados, analisar os postos, e interpretar os resultados à luz de ambas as análises.

Pode ser considerada procedimento não paramétrico a análise feita sob dados transformados ou por desvios normais aleatórios ou por notas normais esperadas. No primeiro caso (desvios normais aleatórios), substitui-se a amostra aleatória a ser analisada com números que pareçam ter vindo de uma distribuição normal, isto é, a observação original de posto k é substituída por um número que tem posto k numa amostra pseudonormal, que pode ser obtida de tabelas, tais como as do livro *A Million Random Digits with 100,000 Normal Deviates* da Rand Corporation (1955) ou por programas especiais de computador. Esse princípio é explicado com mais detalhes por Bell e Doksum (1965).

O outro processo (notas normais esperadas) foi sugerido por Fisher e Yates (1957).

Capítulo 16

Análise exploratória de dados

Introdução

Após a coleta dos dados, seja por amostragem seja por experimentos planejados, os investigadores necessitam conhecer a forma da sua distribuição, seus parâmetros (estimativas) e, principalmente, verificar se entre eles não existem valores errados ou extremamente discrepantes, que prejudicariam a melhor estimativa, os testes de hipóteses propostos a priori e a interpretação dos resultados.

Em muitos casos, e principalmente na amostragem, o volume de dados disponível é de uma magnitude tal que pode se tornar impossível tirar informações úteis a partir de uma simples inspeção desses dados. Ocorre então a necessidade de utilização de ferramentas que permitam fazer um resumo de toda a informação contida nas coleções de dados, permitindo assim uma melhor visualização de seu conteúdo. Essas ferramentas podem ser divididas em dois grupos: estudos numéricos ou estatísticos e elaboração de gráficos. Nestes últimos, há possibilidade de união com as informações estatísticas.

Análise numérica

Por análise numérica entende-se o cálculo ou a estimativa dos parâmetros de uma população seja por base o censo seja a amostragem e a sua interpretação. Esses parâmetros foram definidos no Capítulo 1 e aqui se fará apenas um exemplo de aplicação.

Exemplo de aplicação de análise numérica

Os dados apresentados na Tabela 1 correspondem aos dados do rendimento de feijão em grão (kg ha^{-1}) de 88 municípios amostrados entre os 293 que possui Santa Catarina, em 2009. Esses dados estão disponíveis no site do Instituto Brasileiro de Geografia e Estatística

(IBGE). Uma informação interessante é o fato de 20 dos municípios daquele estado não apresentarem dados de produção de feijão.

Esses mesmos dados serão também empregados nas seções *Gráficos e diagramas* e *Uso do SAS*.

Medidas de posição

Usam-se as definições e as fórmulas dadas no Capítulo 1 e os dados da Tabela 1:

Média: $\overline{X} = m = \dfrac{\sum_{i=1}^{n} X_i}{n} = \dfrac{675 + 980 + ... + 879}{88} = 1.188,3$

Moda: o valor mais frequente é 1.200, que aparece seis vezes entre os 88 municípios amostrados, e nenhum outro valor apresenta essa frequência ou frequência maior.

Mediana: $\tilde{x} = \begin{Bmatrix} x_{([n+1]/2)}, & \text{ímpar} \\ \dfrac{x_{(n/2)} + x_{([n/2]+1)}}{2}, & \text{par} \end{Bmatrix}$.

Como a amostra tem um número par de elementos (88), o valor da mediana é dado por $\tilde{x} = \left\{ \dfrac{x_{44} + x_{45}}{2} \right\} = \dfrac{1.050 + 1.060}{2} = 1.055$.

Quartil 1 (25%): o valor deste quartil é 900, pois os valores X_{22} e X_{23}, na amostra ordenada, têm este valor.

Quartil 3 (75%): o valor do quartil 3 é igual a 1.459, e a média entre os valores 1.430 e 1.488 corresponde, respectivamente, aos valores de número 66 e 67 da amostra ordenada.

Mínimo e máximo: na amostra ordenada, esses valores correspondem a 400 e 2.400.

Tabela 1. Amostra aleatória simples de 88 municípios de Santa Catarina, para os dados de rendimento de feijão em grão, em kg ha^{-1}, no ano de 2009.

Município	Rendimento	Município	Rendimento
Vitor Meireles	675	Nova Veneza	960
Planalto Alegre	980	Canelinha	1.030
Guarujá do Sul	1.500	Governador Celso Ramos	1.000
Nova Itaberaba	918	Serra Alta	1.157
Rio Fortuna	1.107	Ermo	1.000
Rancho Queimado	1.200	Modelo	846
Paraíso	1.488	Aurora	950
Santa Terezinha	1.800	Gravatal	1.037
Florianópolis	500	Frei Rogério	2.100
Santa Cecília	2.000	Abelardo Luz	1.770
Monte Carlo	1.500	Caçador	1.800
Antônio Carlos	500	Santiago do Sul	1.430
São Joaquim	900	São João do Sul	912
Palhoça	1.300	Dona Emma	1.020
Seara	1.200	Cordilheira Alta	765
São Domingos	1.867	Alfredo Wagner	1.050
Iporã do Oeste	1.756	São João do Oeste	1.623
Agronômica	1.409	Chapecó	900
Petrolândia	1.200	São Miguel da Boa Vista	1.062
Treze Tílias	2.400	Bandeirante	1.088
Campo Belo do Sul	1.700	Rio do Oeste	723
Cunhataí	960	Campo Erê	1.396
Catanduvas	1.800	Orleans	889
Chapadão do Lageado	813	Mafra	1.908
Ibicaré	1.391	Doutor Pedrinho	900
Fraiburgo	1.900	Anita Garibaldi	1.000
Lauro Muller	723	Lontras	1.200
Imaruí	900	Saudades	1.160
Braço do Trombudo	950	Quilombo	733
Pinheiro Preto	1.800	Benedito Novo	600
Romelândia	880	Tubarão	1.200
Mondaí	1.414	Içara	908
Araranguá	993	São João do Itaperiú	791

Continua...

Tabela 1. Continuação.

Município	Rendimento	Município	Rendimento
Timbó Grande	1.400	Vargeão	1.200
Iomerê	1.900	Arroio Trinta	1.633
Ponte Serrada	1.900	Rio dos Cedros	1.000
Santo Amaro da Imperatriz	400	Forquilhinha	905
Witmarsum	1.350	Cocal do Sul	892
Pouso Redondo	1.428	Meleiro	800
Balneário Barra do Sul	500	Arabutã	866
Porto União	1.228	Novo Horizonte	1.033
Três Barras	1.060	Rio Rufino	1.500
Ponte Alta	1.500	Caxambu do Sul	911
Iraceminha	950	São Martinho	879

Fonte: dados do IBGE (2011).

Medidas de dispersão

Novamente usam-se as fórmulas e as definições do Capítulo 1:

Variância:

$$s^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2}{n - 1} =$$

$$\frac{(675^2 + 980^2 + ... + 879^2) - \frac{1}{88}(675 + 980 + ... + 879)^2}{87} = 172.674,52$$

Desvio padrão:

$$dp = s = \sqrt{s^2} = \sqrt{172.674,52} = 451,54$$

Coeficiente de variação:

$$CV = \frac{s}{m} * 100 = 415,54/1.188,3 * 100 = 34,97\%$$

Assimetria (usando a fórmula 10 do Capítulo 1, segue que o valor de CA é igual a 0,622):

$$CA = \frac{n}{(n - 1)*(n - 2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Curtose (usando a fórmula 11 do Capítulo 1, segue que o valor da curtose é igual a -0,155):

$$Curtose = \left\{ \frac{n(n+1)}{(n - 1)(n - 2)(n - 3)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{n(n - 1)^2}{(n - 2)(n - 3)}$$

Os parâmetros estimados mostram que existe uma variabilidade um pouco acentuada nos dados de rendimento de feijão em grão, dada por um valor de CV de aproximadamente 35%, e o valor da assimetria, positivo, mostra uma concentração de valores à esquerda da média, portanto com uma cauda no sentido dos valores de alto rendimento. O pequeno valor (e negativo) da curtose mostra que existe pequena alteração no desenho da curva dos dados, no sentido de um pequeno achatamento. Os valores diferentes da média e da mediana, com esta última de menor valor, confirmam a assimetria da curva dos dados com certa predominância dos valores de menor rendimento.

Gráficos e diagramas

A representação dos dados de uma amostra em gráficos ou diagramas permite uma visualização rápida e eficiente da informação neles contida. Três tipos de gráficos ou diagramas devem ser considerados: histograma, ramos e folhas e caixa e bigodes. O primeiro e o segundo tipo são usados para mostrar as frequências de valores, que, no entanto, apresentam regras distintas de construção; enquanto o terceiro, caixa e bigodes, feito em geral em associação com o diagrama de ramos e folhas, requer informação de algumas medidas de posição e dispersão para a sua construção, como será visto a seguir. No caso de experimentos planejados, o gráfico de pontos ou de dispersão é, em geral, suficiente para verificar a consistência dos valores obtidos para os tratamentos em teste. No caso de experimentos com controle local, também é possível obter informação de parte dos efeitos de blocos.

Histograma

O histograma é um gráfico de barras, horizontais ou verticais, que representa a frequência, frequência acumulada, frequência relativa ou frequência relativa acumulada, correspondentes aos valores existentes em uma amostra.

Se os dados forem do tipo discreto, contagens ou notas de valor inteiro, também chamados de valores indicativos, ou se se tratam de variáveis qualitativas, as barras são definidas pelos valores, se não são em número muito grande. Se o número de classes é muito alto, que dê origem a um gráfico com número excessivo de barras, os valores podem ser agrupados em classes de valores. Para esse tipo de dados, o número mínimo de barras pode ser tão pequeno quanto 2 (dois), quando se trata de variáveis do tipo Bernoulli, de sucesso ou fracasso, prejudicado ou não prejudicado, por exemplo.

Se os dados forem de natureza contínua, como peso, altura, área, volume, diâmetro ou outros desse tipo, cada barra será associada a um intervalo de valores. O número de barras é definido pela raiz quadrada do número total de valores ($\sqrt{n}$). Como o valor do número de classes assim definido é, muito usualmente, um número não inteiro, deve-se tomar o inteiro superior. É necessário também cuidar para que o número de barras não seja nem muito pequeno nem muito grande.

A amplitude de valores em cada uma das classes é obtida pela divisão da amplitude dos dados pelo número de classes definidas. O valor superior da primeira classe é obtido somando-se ao valor mínimo dos dados a amplitude de classe. Para as demais classes, o valor superior é obtido pela soma da amplitude de classe ao limite superior da classe anterior.

Exemplo de aplicação

Os dados da Tabela 1 serão empregados para demonstrar a construção de um histograma. Neste caso, os dados são de natureza contínua; portanto, a construção do diagrama será feita com o uso de intervalos de valores. O número de barras ou colunas é determinado pelo número inteiro imediatamente superior à raiz quadrada do número de elementos na amostra, e o intervalo de classe é definido apropriadamente. Cada coluna ou barra tem a altura equivalente à frequência de valores no intervalo. A classe 1, por exemplo, mostraria a frequência de valores entre o valor mínimo e o limite definido para ela, a classe 2 mostra a frequência entre o limite da classe 1 e o limite da classe 2, e assim por diante. A Tabela 2 resume esses cálculos e a Figura 1, o histograma obtido (construído com o Excel da Microsoft).

Mais uma vez, observa-se a concentração dos dados à esquerda e uma cauda acentuada à direita, o que confirma a assimetria calculada e o alto valor da variância.

Tabela 2. Determinação do número e dos limites superiores das colunas e frequência observada para a construção do histograma.

Número de classes (NC) =	$\cong \sqrt{n}$	$\sqrt{88} = 9{,}38$	10	Frequências
Amplitude dos dados (AD) =	Máximo - mínimo	2.400 - 400	2.000	
Amplitude das classes (AC)	AD/NC	2.000/10	200	
Classe 1	Mínimo + AC	400 + 200	600	5
Classe 2	Classe 1 + AC	600 + 200	800	7
Classe 3	Classe 2 + AC	800 + 200	1.000	27
Classe 4	…	1.000 + 200	1.200	17
Classe 5	…	1.200 + 200	1.400	6
Classe 6	…	1.400 + 200	1.600	9
Classe 7	…	1.600 + 200	1.800	9
Classe 8	…	1.800 + 200	2.000	6
Classe 9	Classe 8 * AC	2.000 + 200	2.200	1
Classe 10	Classe 9 + AC	2.200 + 200	2.400	1

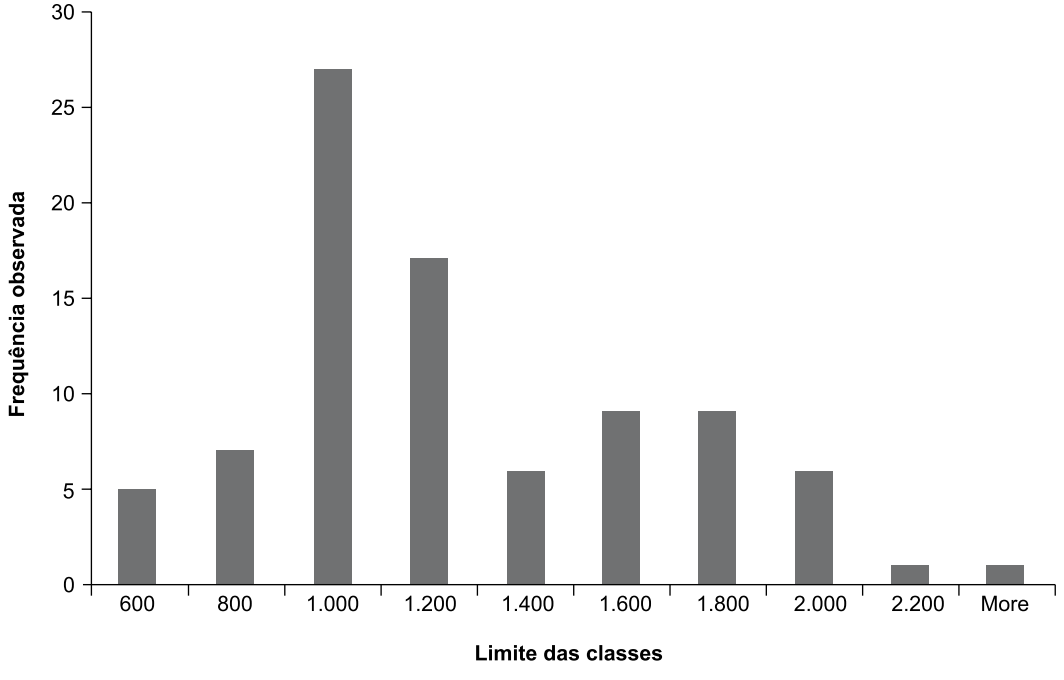


Figura 1. Histograma dos dados de rendimento de feijão em grão, em kg ha^{-1}, em Santa Catarina, ano de 2009. Amostra de 88 municípios.

Fonte: dados do IBGE (2011).

Ramos e folhas

O diagrama de ramos e folhas é outra forma de representação visual de uma série de valores, conformados pelo menos de dois dígitos. Para sua construção os números se dividem em duas partes. Uma parte se chama ramo, formada por um ou mais dígitos principais, e a outra folha, que contém o resto dos dígitos. Para permitir uma representação interessante dos dados, o número de ramos é relativamente pequeno quando comparado com o número total de observações. Esse número, em geral, está entre 5 e 20.

O diagrama apresenta, portanto, duas colunas de informação: uma, à esquerda, que corresponde aos ramos; e a outra, à direita, que representa as folhas. É também muito comum que se adicione uma terceira coluna, mais à direita, que contém o valor das frequências em cada ramo. É recomendável indicar em alguma parte do diagrama a ordem de grandeza dos ramos e folhas.

Algumas vezes é recomendável aumentar o número de ramos. Para duplicá-los, por exemplo, ao fazer a divisão de cada ramo, em um deles ficam as folhas cujos valores tenham como primeiro dígito valor 4 ou menor, e no outro, valor 5 ou maior.

Exemplo de aplicação

Ainda com os dados do rendimento de feijão em grão em Santa Catarina, em 2009, como apresentado na Tabela 1, verificou-se que, se forem usados os valores das centenas desse rendimento como ramos e os valores das dezenas como folhas, o número de ramos é igual a 25, um pouco acima do recomendado, mas que será usado de qualquer forma como exemplo. O resultado é visto na Figura 2, também construída com o auxílio do Excel da Microsoft e com o uso de um complemento elaborado por Berk e Carey (2001).

A Figura 2 mostra mais uma vez a assimetria, a concentração de dados na faixa de baixos rendimentos e uma cauda acentuada para os valores de alto rendimento.

Caixa e bigodes

O diagrama de caixa e bigodes, além de proporcionar uma visualização dos dados da amostra, representa também algumas das estatísticas de posição e dispersão e permite identificar alguns valores como valores atípicos e/ou valores atípicos extremos. A caixa ou retângulo central tem como lados extremos os valores dos quartis 3 e 1 e uma linha central para o quartil 2 (mediana). Dentro desse retângulo, o símbolo + ou uma linha tracejada de cor diferente representa a posição da média. Os bigodes, desenhados como linhas perpendiculares ao retângulo tem um comprimento igual a 1,5 vez a distância interquartílica.

Ramo x 100	Rendimento	Frequência
4	0	1
5	000	3
6	07	2
7	22369	5
8	01467889	8
9	0000001115556689	16
10	000023335668	12
11	056	3
12	0000002	7
13	0599	4
14	001238	6
15	0000	4
16	23	2
17	057	3
18	00006	5
19	0000	4
20	0	1
21	0	1
22		0
23		0
24	0	1

Figura 2. Diagrama de ramos e folhas para os dados de rendimento de feijão em grão, em Santa Catarina, em 2009. Amostra de 88 municípios.
Fonte: dados do IBGE (2011).

Os valores situados entre o final dos pontos extremos do bigode até uma distância igual a 1,5 vez a distância interquartílica são considerados valores atípicos, e aqueles, além desses pontos, são chamados de atípicos extremos.

Exemplo de aplicação

Com os mesmos dados da Tabela 1, e usando os resultados das estimativas dos parâmetros obtidas no item *Medidas de posição*, é possível elaborar uma tabela auxiliar (Tabela 3) para a construção do diagrama de caixa e bigodes. O resultado é visto na Figura 3, também construída com o auxílio do Excel da Microsoft e com o uso de complemento elaborado por Berk e Carey (2001).

Tabela 3. Valores necessários para a construção do diagrama de caixa e bigodes.

Elemento de construção	Valor
Média	1.188,3
Mediana (quartil 2)	1.055
Quartil 1	900
Quartil 3	1.459
Distância interquartílica	559
Comprimento dos bigodes	838,5
Limite para bigode superior	2.297,5
Limite para bigode inferior	61,5
Valor máximo para valores atípicos superiores	3.136
Valor mínimo para valores atípicos inferiores	-777 (deve ser considerado como zero, pois os valores da amostra são todos positivos)

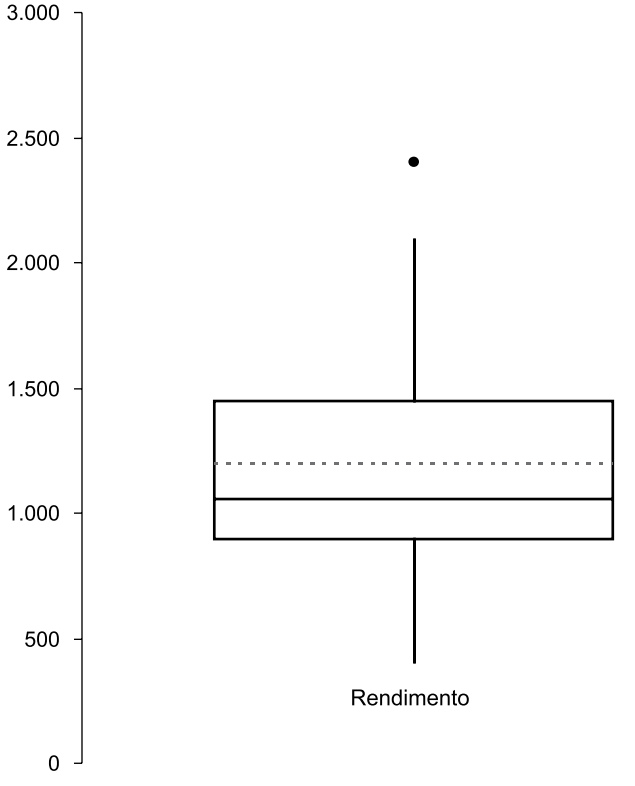


Figura 3. Diagrama de caixa e bigodes para os dados de rendimento de feijão em grão em Santa Catarina, em 2009. Amostra de 88 municípios.
Fonte: dados do IBGE (2011).

Como visto nas Figuras 1 e 2, o diagrama de caixa e bigodes mostra uma assimetria dos dados, claramente visível pela distinção entre as localizações da média (linha tracejada vermelha) e a mediana (linha central da caixa). Ademais, o valor identificado como um ponto preto no alto da figura, que representa o valor máximo (2.400), é apontado como um valor atípico.

Gráfico de pontos

O gráfico de pontos é o mais simples de todos os gráficos, pois se trata apenas da representação dos dados segundo sua ordem de coleta ou, algumas vezes, segundo o valor de alguma outra variável de interesse do pesquisador. Essa variável pode ser o tempo ou a data de leitura, o tratamento experimental ou qualquer outra variável que se possa assumir que esteja associada ao dado que se está analisando.

Esses gráficos são muito empregados em estudos de correlação e regressão (veja Capítulo 12), mas também podem ser utilizados em análise exploratória de dados. Na seção a seguir serão apresentados dois exemplos de aplicação – um com os dados de rendimento de feijão mostrados na Tabela 1 e outro para dados de um experimento em blocos ao acaso, utilizado no Capítulo 3. Por simplicidade de uso, os gráficos de pontos desta seção serão obtidos com o auxílio do Excel da Microsoft.

Exemplo com dados de amostragem

No caso de amostras, um gráfico como o mostrado na Figura 4 é o suficiente para que o investigador tenha uma boa ideia dos dados obtidos. Uma amostra adequada mostraria todos os dados dentro de uma faixa de valores sem grandes discrepâncias entre eles. No caso atual, verifica-se que apenas um dado se destaca, aquele próximo à linha dos 2.500 kg ha^{-1}, como já seria de se prever a partir do diagrama de caixa de bigodes (Figura 3).

Exemplo com dados de um ensaio em blocos completos

Para exemplificar o uso do gráfico de pontos em um ensaio, utilizou-se o experimento mostrado no Capítulo 3, Tabela 4, referente a um ensaio regional de feijão-preto, em 1999, em Santo Antônio de Goiás. O ensaio continha 15 variedades de feijão, com quatro repetições no delineamento em blocos completos casualizados. Para efeitos do gráfico, mostrado na Figura 5, obtido por meio do software Excel da Microsoft, os tratamentos em ordem alfabética foram numerados de 1 a 15. Cada ponto de cada tratamento refere-se ao rendimento em uma das quatro repetições. Quanto mais distantes estiverem os pontos

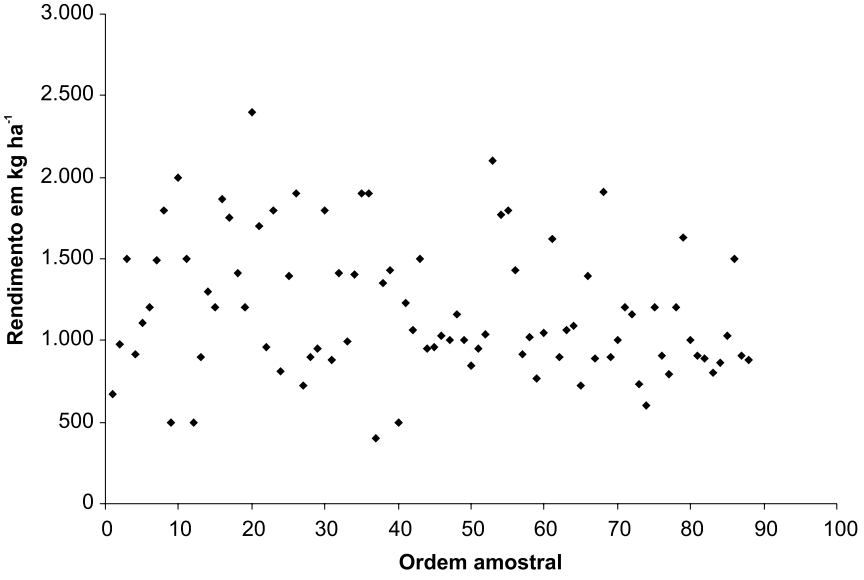


Figura 4. Gráfico de pontos dos rendimentos de feijão, em kg ha^{-1}, obtidos na amostragem dos municípios de Santa Catarina, em 2009.
Fonte: dados do IBGE (2011).

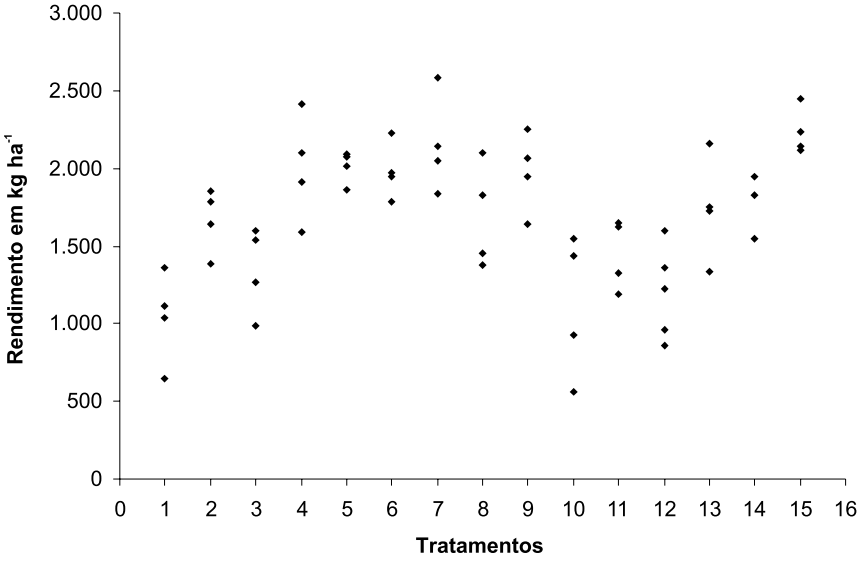


Figura 5. Gráfico de pontos dos rendimentos de feijão, em kg ha^{-1}, obtidos no ensaio regional de feijão-preto em 1999, no Município de Santo Antônio de Goiás (ver Tabela 4 do Capítulo 3).

para um mesmo tratamento, maior deve ser a variância desse tratamento, lembrando que o efeito de bloco deve ser ainda levado em conta. De uma maneira geral, pode-se inferir que o tratamento 10 apresenta a maior variabilidade, enquanto os tratamentos 5 e 15 as menores. Na seção seguinte, com o uso do SAS tem-se uma melhor apreciação desses fatos.

Uso do SAS na análise exploratória de dados

O SAS é um software de uso mundial, bastante potente e flexível em suas opções de análises. Entre os muitos procedimentos que possui para esse tipo de análise, serão utilizados apenas alguns.

Análises numéricas

Para as análises numéricas, definidas no Capítulo 1, o procedimento adequado é:

```
PROC UNIVARIATE; VAR REND;
```

com resultado e interpretação idêntica às precedentes (Figura 6).

Análises gráficas

Para a realização das análises gráficas, a obtenção dos diagramas de caixa e bigodes (*box plot*) e ramos e folhas (*stem and leafs*), basta acrescentar a opção *plot* ao procedimento. Dessa forma, serão obtidas ambas as análises, numéricas e gráficas, além de um gráfico de distribuição normal:

```
PROC UNIVARIATE PLOT;VAR REND;
```

O resultado da opção *plot* encontra-se na Figura 7.

Para produzir um histograma, o procedimento adotado foi o PROC CHART com a seguinte sintaxe:

```
PROC CHART;VBAR=REND/LEVELS=11;
```

O resultado pode ser visto na Figura 8.

The UNIVARIATE Procedure

Variable: rend

Moments

N	88	Sum Weights	88
Mean	1188.26136	Sum Observations	104567
Std Deviation	415.541234	Variance	172674.517
Skewness	0.62247704	Kurtosis	-0.1552547
Uncorrected SS	139275609	Corrected SS	15022683
Coeff Variation	34.9705247	Std Error Mean	44.2968444

Basic Statistical Measures

Location		Variability	
Mean	1188.261	Std Deviation	415.54123
Median	1055.000	Variance	172675
Mode	1200.000	Range	2000
		Interquartile Range	559.00000

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	26.82497	Pr > |t|	<.0001
Sign	M	44	Pr >= |M|	<.0001
Signed Rank	S	1958	Pr >= |S|	<.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	2400
99%	2400
95%	1900
90%	1800
75% Q3	1459
50% Median	1055
25% Q1	900
10%	733
5%	600
1%	400
0% Min	400

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
400	37	1900	36
500	40	1908	68
500	12	2000	10
500	9	2100	53
600	74	2400	20

Figura 6. Resultado da análise numérica com o uso do procedimento UNIVARIATE do SAS.

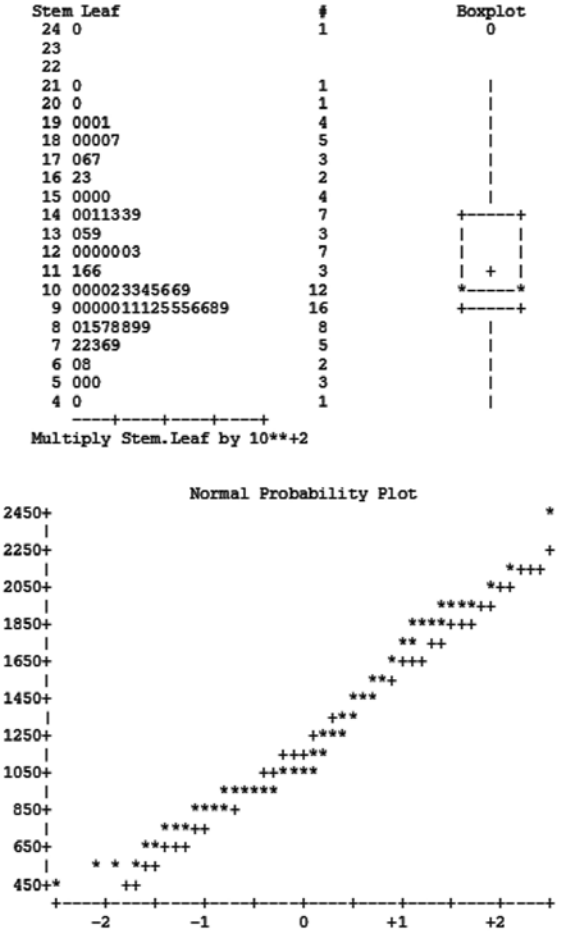


Figura 7. Diagrama de caixa e bigodes e gráfico de distribuição normal, para os dados de rendimento de feijão em grão (em kg ha^{-1}), em Santa Catarina, no ano de 2009.

Fonte: dados do IBGE (2011).

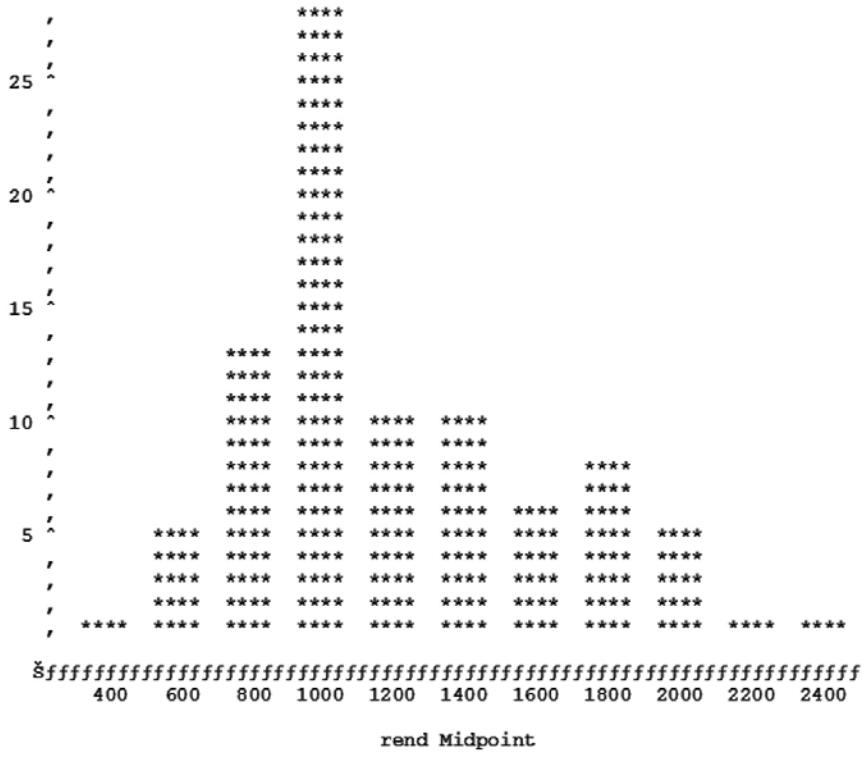


Figura 8. Histograma para os dados de rendimento de feijão em grão (em kg ha^{-1}), em Santa Catarina, no ano de 2009, obtido com o uso do SAS.
Fonte: dados do IBGE (2011).

Análise gráfica para um delineamento em blocos completos casualizados

Usar o SAS para uma análise gráfica, principalmente no caso de um gráfico de pontos, é muito simples e, ao mesmo tempo, muito revelador nos seus resultados. Para o caso particular de blocos completos casualizados, o interesse maior é a verificação dos resultados obtidos por tratamento em função dos blocos. Se a suposição inicial, de homogeneidade de blocos, se verificar, os dados obtidos por blocos deveriam aparecer no gráfico sempre na mesma ordem: os mais férteis ou de maior rendimento, no exemplo em estudo, na parte superior de cada tratamento, e o mais pobre na parte mais baixa. Inversões nessa ordem indicam certa heterogeneidade de blocos. Para obter esse tipo de gráfico, o procedimento SAS e sua sintaxe seriam:

```
PROC PLOT; PLOT REND*TRAT=REP;
```

em que REND corresponde ao rendimento obtido em kg ha^{-1} de feijão em grão, TRAT o nome do tratamento e REP a identificação numérica de cada bloco.

O resultado obtido com esse procedimento aparece na Figura 9.

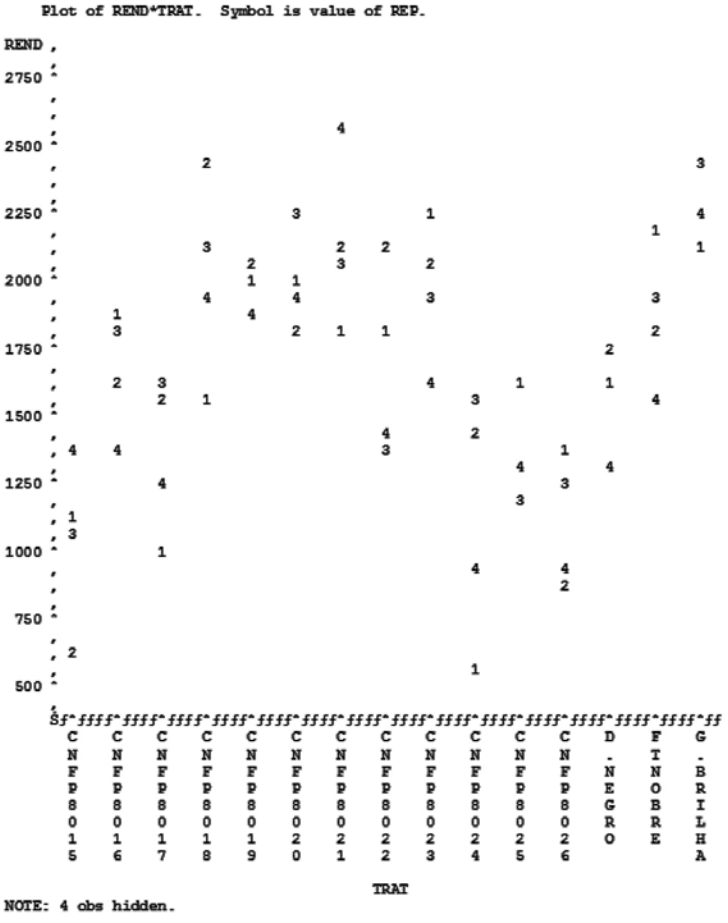


Figura 9. Gráfico de pontos para os dados do ensaio regional de feijão-preto em Santo Antônio de Goiás, em 1999. Exemplo de uso do SAS.

Na Figura 9 não existe a constância de aparecimento dos pontos correspondentes aos blocos em cada tratamento, como desejado, o que indica certa heterogeneidade. Quando se verifica a ANOVA, apresentada no Capítulo 3, observa-se que o CV foi de aproximadamente 16%, o que é considerado médio e condiz com o observado no gráfico.

A nota que aparece ao fim da figura (*Note: 4 obs hidden*) indica que quatro dos pontos não aparecem no gráfico por serem coincidentes ou muito próximos do rendimento do bloco imediatamente anterior, ou seja, se o número 2, por exemplo, não aparece na coluna de um tratamento, o rendimento desse bloco 2 foi muito parecido ou igual ao do bloco 1, e assim por diante. A variabilidade dentro de cada tratamento tem a mesma interpretação dada na seção *Exemplo com dados de um ensaio em blocos completos*.

Capítulo 17

Princípios e técnicas de amostragem

Introdução

Como foi visto no Capítulo 1, alguns procedimentos são adotados para a obtenção de dados necessários para a verificação de certas hipóteses, seja para a criação de novas tecnologias seja para a adaptação dessas tecnologias a novos ambientes. Mas muitas vezes existe a necessidade de conhecer a realidade de certas populações existentes na natureza. Aqui também as metodologias e as técnicas estatísticas e de probabilidade são aplicadas a esses dados que são simplesmente medidos na forma em que estão disponíveis na natureza. A forma de escolhê-los e de analisá-los se estuda na área chamada de amostragem.

No entanto, muitas vezes a amostragem também é aplicada em experimentos planejados, e isso ocorre com certa frequência, pela necessidade de medir alguns parâmetros por planta ou de tomar essas medidas em diferentes datas. Mas isso deve ser feito com cuidado, pois essa amostragem é diversa daquela feita sobre indivíduos cuja existência não foi planejada pelo pesquisador. A amostragem feita no experimento é baseada em indivíduos cujas condições de existência foram definidas pelo pesquisador. Portanto, isso implica que todos os indivíduos da "população" a ser amostrada estão em condições idênticas, e suas diferenças são basicamente devidas ao acaso, enquanto a amostragem é feita em populações naturais, e assim logicamente em condições bem diversas umas das outras. É para este último caso que este capítulo é dedicado.

Com uma população sendo definida como o conjunto de todos os indivíduos que compartilham uma ou mais de uma característica, é fácil perceber que uma população pode conter um número finito ou infinito, ou ainda finito incontável, de indivíduos, dependendo da definição e do número de elementos que possuem as características de definição.

O estudo estatístico pode ser feito sobre todos os indivíduos que constituem a população, e nesse caso tem-se um censo, ou sobre uma parte desses elementos, que é a amostragem.

A realização dos censos é, em geral, custosa, pois exige muito tempo de preparo, coleta de dados e análise, além de requerer equipes de pessoas muito bem treinadas; e, por isso, é muito pouco aplicada. Exemplos de censos são aqueles realizados pelos governos dos países, periodicamente, para que se possa conhecer sua população humana, sua agricultura, indústria, etc. Por isso é muito comum a realização de amostragens nas quais somente uma parte da população é avaliada. Diferentes técnicas ou procedimentos existem para isso. Este capítulo tem como foco a amostragem aleatória simples e a amostragem aleatória estratificada, assim como o caso específico de amostragem de variáveis para o estudo de proporções. Outros detalhes podem ser encontrados no livro de Cochran (1977).

Da mesma forma que os ensaios ou experimentos, um trabalho de amostragem (ou censo) requer a elaboração de um projeto, e alguns passos devem ser seguidos.

- Objetivos da amostragem: definição exata do que se procura conhecer da população.
- População: definir com precisão a população objeto da amostragem, bem como da população a ser amostrada. Conforme foi visto anteriormente, muitas vezes as duas não são coincidentes. Nesse caso, as informações serão válidas apenas para a última. Por exemplo, ao definir população de plantas de arroz, incluem-se nela todas as plantas em cultivo num determinado país ou região num certo sistema de cultivo. Mas, se a amostragem se faz apenas em fazendas que distem no máximo a 30 km de rodovias pavimentadas, a população objeto da amostragem é distinta daquela definida.
- Unidades de amostragem: os elementos a serem medidos devem ser bem definidos antes de iniciar o trabalho (planta, panícula, vagem, fruto, área de cultivo, etc.). Deve-se também definir as condições em que os elementos se desenvolvem para que possam ser incluídos na amostragem (por exemplo, se estão em regime de competição idêntico ou não, o estágio da planta, etc.).
- Precisão e erro aceitáveis: o erro aceitável refere-se ao nível de significância das provas estatísticas que se pensa em aplicar às estimativas dos parâmetros populacionais. Precisão diz respeito ao tamanho do erro na estimativa que é aceitável para aquele nível de significância.
- Dados a serem coletados: devem ser definidas as variáveis que vão ser coletadas nos indivíduos da amostra, de maneira que se possa alcançar o objetivo determinado, sem deixar qualquer variável omitida ou incluir variáveis não pertinentes

ao objetivo. Evidentemente, variáveis de identificação do indivíduo devem ser, sempre que possível, incluídas.

- Métodos de medição: devem-se definir os instrumentos de medição, seu grau de precisão e a forma de obtê-los, por entrevista ou medida direta, exame (físico, químico, radiológico, ou outros disponíveis), correio, correio eletrônico ou combinações entre eles. Também se refere à forma de registrar ou de codificar os dados.
- Seleção da amostra: existem várias maneiras de selecionar a amostra, mas, para cada uma delas, estimativas de tamanho de amostra podem ser feitas de acordo com o grau de precisão desejado. Além disso, os custos relativos e o tempo necessário para realizar o trabalho devem ser considerados.
- Elaboração do questionário e o teste preliminar: o questionário deve ser elaborado de forma cuidadosa a fim de que forneça toda informação requerida para obtenção dos objetivos. Mas, ao mesmo tempo, deve ser elaborado para que não seja extremamente grande e/ou tedioso para que aqueles que devem respondê-lo ou aplicá-lo não o façam de maneira descuidada, desinteressada, ou mesmo se recusem a fazê-lo. Antes de sua aplicação final deve ser feito um teste preliminar com as pessoas que vão aplicá-lo, e com alguns possíveis membros da população, para verificar se tudo está claro, bem entendido, a fim de que não haja dúvidas durante sua aplicação final.

Estatísticas de interesse

Ao realizar uma amostragem, em geral, o que se busca é conhecer ou estimar certos parâmetros da população objetivo. Entre eles, podem ser destacados os seguintes: média, moda, mediana, variância, desvio padrão, erro da média, total, variância do total, proporção e variância da proporção. Além disso, há também interesse em provar certas hipóteses a respeito desses parâmetros.

Parâmetros e suas estimativas

Quando se trabalha com censos, as estatísticas obtidas são os verdadeiros valores dos parâmetros da população. No entanto, quando se trabalha com amostras, as estatísticas são as estimativas dos parâmetros dessa população. Neste capítulo, Y_i será chamado de valor do indivíduo i da variável Y na população, e y_i de valor do indivíduo i para a variável Y na amostra. Ou seja, quando se refere à população, serão empregadas letras maiúsculas (ou gregas); enquanto para as amostras, serão usadas letras minúsculas. Os cálculos necessários

para os parâmetros de interesse foram quase todos definidos no Capítulo 1, de forma que aqui serão definidos apenas os demais.

Estimativa por intervalo da média

Para certas situações, a estimativa pontual, como feita anteriormente no item *Medidas de tendência central ou de posição* (Capítulo 1), não é informação suficiente e segura, motivo pelo qual é desejável conhecer uns limites inferior e superior dessa estimativa a qual se vincula uma probabilidade, α, de forma que o intervalo assim definido apresente uma confiança igual a $1 - \alpha$ de conter o verdadeiro parâmetro da população. Em outras palavras, deseja-se conhecer os limites I e S tal que:

$$P\,(I = \mu = S) = 1 - \alpha \qquad [1]$$

Sendo conhecida a variância da população, a estimativa por intervalos da média populacional é dada por:

$$\bar{y} - z_{1-\alpha/2} * S/\sqrt{n} \;\leq\; \mu \;\leq\; \bar{y} + z_{1-\alpha/2} * S/\sqrt{n} \qquad [2]$$

Para o caso mais comum em que a variância da população é desconhecida, é possível usar a estimativa amostral (s^2), como está na fórmula 2 (Capítulo 1). Além disso, associando-se a hipótese usual de que a população segue a distribuição normal, a distribuição t é então a distribuição adequada para encontrar esse intervalo de confiança:

$$\bar{y} - t_{\alpha/2;(n-1)} * s/\sqrt{n} \;\leq\; \mu \;\leq\; \bar{y} + t_{\alpha/2;(n-1)} * s/\sqrt{n} \qquad [3]$$

Para os dados da amostragem mostrada no Capítulo 16, tem-se que a estimativa da média foi igual a 1.188,3 e a variância 172.674,52 para um tamanho de amostra (n) de 88.

$$1.188{,}3 - 1{,}9876 * \sqrt{172.674{,}52/88} \;\leq\; \mu \;\leq\; 1.188{,}3 + 1{,}9876 * \sqrt{172.674{,}52/88}$$

$$1.100{,}22 \;\leq\; \mu \;\leq\; 1.276{,}3$$

Esse resultado implica que o valor real da média está situado entre 1.100,22 kg ha^{-1} e 1.276,30 kg ha^{-1}. Considerando-se, no entanto, que a relação n/N ($N = 293$) não é desprezível, o valor da média do erro padrão dada na fórmula 3 deveria ser substituído pela raiz quadrada da variância da média dada na fórmula 5.

Nesse caso, o intervalo de confiança da média será igual a

$1.126{,}67 \leq \mu \leq 1.249{,}85$

Variância da média

A variância da média de uma amostra é dada por

$$s^2_{\bar{y}} = \frac{s^2}{n} \quad [4]$$

enquanto a variância dessa média para uma população finita é igual a

$$s^2_{\bar{Y}} = \frac{s^2}{n} * \frac{N - n}{N} \quad [5]$$

O quociente $\frac{N - n}{N}$ é chamado de fator de correção para população finita. No entanto, pode ser eliminado do cálculo sempre que a relação n/N for muito pequena.

Total

No caso de populações finitas, é possível calcular o total de uma variável E, que é obtida por meio de

$$\hat{Y} = N * \bar{y} \quad [6]$$

quando se trabalha com amostras, ou por

$$Y = N * \overline{Y} = \sum_{i=1}^{N} Y_i \quad [7]$$

no caso do censo.

Variância do total

A variância do total de uma população finita é dada por:

$$s^2_{\hat{Y}} = \frac{N^2 s^2}{n} * \frac{N - n}{N} \quad [8]$$

Proporção

Proporção é definida como a frequência relativa de ocorrência de uma variável no total de indivíduos da população ou da amostra. Em estudos com arroz, por exemplo, existe o interesse de se conhecer a proporção de perfilhos férteis (com panícula). A variável Y_i pode ter apenas dois valores, um se ocorrer o fenômeno de interesse (perfilho fértil) e zero se não ocorrer (perfilho sem panícula). Na prática, contam-se o número de perfilhos férteis e o número total de perfilhos amostrados (n). O quociente entre essas duas quantidades resulta na proporção de perfilhos férteis por planta. Dessa forma, tem-se

$$\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n} \quad [9]$$

e sua variância é

$$s^2_p = \frac{\hat{p}(1 - \hat{p})}{n} \quad [10]$$

Provas de hipótese

Este assunto já foi discutido com certo detalhe no Capítulo 1. Aqui se fará apenas a discussão de tópicos específicos de amostragem ou censo, como é o caso de prova de uma média, prova de duas médias, prova de média de dados emparelhados e prova de duas variâncias.

Provas de hipótese para uma média

A hipótese nula para a média é representada por $H_0 : \mu = \mu_0$ para uma hipótese alternativa que pode ser bilateral $H_1 : \mu \neq \mu_0$ ou unilateral $H_1 : \mu < \mu_0$ ou $H_1 : \mu > \mu_0$.

A prova t para esses casos tem n - 1 graus de liberdade e a forma

$$t = \frac{\bar{y} - \mu_0}{\sqrt{\frac{s^2}{n}\left(\frac{N - n}{N}\right)}} \quad [11]$$

quando se trabalha com amostras de populações finitas, ou a forma

$$t = \frac{\bar{y} - \mu_0}{\sqrt{\frac{s^2}{n}}} \quad [12]$$

quando são usadas estimativas para populações infinitas ou finitas, desde que o quociente n/N possa ser considerado desprezível.

Quando a variância da população é conhecida, a prova de hipótese para uma média segue então a distribuição normal padrão e tem as formas anteriores modificadas para

$$z = \frac{\bar{y} - \mu_0}{\sqrt{\frac{\sigma^2}{n}\left(\frac{N - n}{N}\right)}} \quad [13]$$

e

$$z = \frac{\bar{y} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \quad [14]$$

Para os mesmos dados amostrais do Capítulo 16, seria possível formular hipótese de que a média de rendimento em Santa Catarina não é diferente da média nacional. Pelos dados do IGBE (2011), em 2009 o rendimento médio do feijão em grão no Brasil foi de 802,57 kg ha^{-1}. Dessa forma, as hipóteses seriam

$$H_0 : \mu = 802,57$$

e

$$H_1 : \mu \neq 802,57$$

usando os mesmos dados anteriores: estimativa da média igual a 1.188,3, variância de 172.674,52 e tamanho de amostra (*n*) de 88. Com o auxílio da fórmula 11, tem-se que:

$$t = \frac{1.188,3 - 802,57}{\sqrt{\frac{172.674,52}{88}\left(\frac{293 - 88}{293}\right)}}$$

$$t = 10,41$$

Esse valor de *t*, com 87 graus de liberdade, tem uma probabilidade de $5,9*10^{-17}$, o que permite recusar a hipótese nula e aceitar a alternativa de que o rendimento de feijão em grão em Santa Catarina é diferente daquele obtido no País como um todo.

Prova de hipótese para a proporção

A hipótese nula para a proporção é representada por $H_0 : p = p_0$ para uma hipótese alternativa que pode ser bilateral $H_1 : p \neq p_0$ ou unilateral $H_1 : p < p_0$ ou $H_1 : p > p_0$.

A prova *t* também tem $n - 1$ graus de liberdade e as mesmas formas da média, apenas substitui-se no numerador a estimativa da média pela estimativa da proporção e no denominador a fórmula para estimativa da variância da proporção, como visto no item *Proporção*.

Provas de hipótese para duas médias

Neste caso, considerando-se apenas os casos em que as variâncias populacionais são desconhecidas, a prova *t* é aplicável, mas ter-se-á de considerar primeiro se as variâncias de cada média podem ser consideradas iguais ou não, e isso pode ser comprovado como está descrito no item *Prova de hipótese para variâncias*, que será apresentado ainda neste capítulo.

A hipótese nula, H_0: $\mu_1 = \mu_2$, pode ser contrastada com a hipótese alternativa bilateral, H_1: $\mu_1 \neq \mu_2$, ou com as hipóteses alternativas unilaterais, H_1: $\mu_1 < \mu_2$ ou H_1: $\mu_1 > \mu_2$.

A forma da prova t para esses dois casos é:

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{V(dif)}} \qquad [15]$$

O valor da variância da diferença entre duas médias depende dos valores das variâncias das duas amostras. Por serem amostras de populações independentes, seus valores podem ser considerados iguais ou não. Para isso, deve ser aplicada inicialmente uma prova de igualdade de variâncias, como será mostrado no item *Prova de hipótese para variâncias*. O resultado dessa prova determina então a aplicação de cálculo da variância da diferença dado nos itens *Variâncias iguais* ou *Variâncias diferentes*, com seus respectivos graus de liberdade.

Variâncias iguais

Neste caso, a prova *t* tem $(n_1 + n_2 - 2)$ graus de liberdade, e a variância da diferença entre as duas médias, $V(dif)$, é obtida por

$$V(dif) = \left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right] * \left[\frac{1}{n_1} + \frac{1}{n_2}\right] \qquad [16]$$

em que s_1^2 e s_2^2 são as estimativas das variâncias das duas amostras, e n_1 e n_2 são os respectivos tamanhos de amostras.

Variâncias diferentes

Quando as variâncias das amostras devem ser consideradas diferentes, tem-se

$$V(dif) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \qquad [17]$$

de modo que os graus de liberdade são estimados por (Satterthwaite, 1946)

$$GL = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2 / n_1\right)^2}{n_1 + 1} + \frac{\left(s_2^2 / n_2\right)^2}{n_2 + 1}} - 2 \qquad [18]$$

em que s_1^2, s_2^2, n_1 e n_2 têm o mesmo significado do item *Variâncias iguais*.

Dados pareados

Em algumas situações, uma variável é medida em duas situações distintas, ou em dois tempos consecutivos, como, por exemplo, antes e depois de algum evento, sobre os mesmos elementos da amostra, o que não deveria ser considerado como a existência de duas amostras, mas de dados pareados. Nessa situação, não se está interessado nas estimativas dos parâmetros em cada uma das situações, mas nas diferenças existentes entre as medidas tomadas, ou seja, $D_i = Y_{2i} - Y_{1i}$. Assim as estimativas da média e da variância, bem como a prova de hipótese a respeito da média, são feitas com as diferenças D_i, e o processo é aquele discutido no item *Provas de hipótese para uma média*.

Prova de hipótese para variâncias

A maneira mais simples de verificar a hipótese nula H_0: $\sigma_1^2 = \sigma_2^2$, com respeito à alternativa H_1: $\sigma_1^2 \neq \sigma_2^2$, é a prova F, que corresponde ao quociente entre as duas estimativas de variâncias:

$$F = \frac{s_1^2}{s_2^2} \qquad [19]$$

A prova tem os graus de liberdade correspondentes aos tamanhos das amostras $n_1 - 1$ e $n_2 - 1$.

Uso do SAS para uma amostra

Como foi visto no Capítulo 16, as estatísticas básicas são obtidas com o uso do procedimento UNIVARIATE. Da mesma forma, esse procedimento pode ser usado para testes

com a média e para o cálculo do intervalo de confiança (IC) tanto da média como da variância e do desvio padrão. No Capítulo 16, o teste t para a média é feito assumindo-se que μ_0 é igual a zero, mas é possível usar uma opção do procedimento para teste para outros valores. O programa a seguir repete as estatísticas obtidas para o capítulo anterior e, além disso, testa a média para um valor de μ_0 igual à média do País (MU = 802,57) e obtém os ICs (CIBASIC) para os três casos citados anteriormente.

```
PROC UNIVARIATE MU0=802.57 CIBASIC;VAR REND;
```

Na Figura 1, são mostradas apenas as estatísticas concernentes a este capítulo. Verifica-se diferença no valor do teste t para a média, pois o SAS não faz a correção para população finita.

```
    Basic Confidence Limits Assuming Normality
Parameter          Estimate     95% Confidence Limits
Mean                   1188          1100        1276
Std Deviation     415.54123     361.91401   487.97292
Variance             172675        130982      238118

          Tests for Location: Mu0=802.57
   Test           -Statistic-     -----p Value------
   Student's t    t  8.706972     Pr > |t|    <.0001
   Sign           M        32     Pr >= |M|   <.0001
   Signed Rank    S      1673     Pr >= |S|   <.0001
```

Figura 1. Uso do procedimento UNIVARIATE para testes para uma amostra.

Uso do SAS para duas amostras

No caso de duas amostras independentes, o procedimento SAS adequado é o TTEST. Como exemplo de aplicação, foi retirada uma amostra de 83 municípios de Santa Catarina, no ano de 2008, também para estudo da produtividade do feijão, e essa amostra foi comparada com a amostra de 2009, dada no Capítulo 16. O programa SAS para esse caso é o seguinte:

```
PROC TTEST;CLASS ANO; VAR PROD;
```

O resultado obtido está na Figura 2, na qual é possível ver que o sistema calcula as estatísticas básicas, os intervalos de confiança para as médias e para os desvios padrões, para cada uma das classes, bem como para a diferença (Diff) entre as classes. O teste t é

feito sob as duas suposições, igualdade e diferença de variâncias, mas o teste F para variâncias mostra que, nesse caso, o adequado é aquele com igualdade de variâncias. Os dados ali mostrados indicam que a hipótese de nulidade ou de igualdade das médias deve ser rejeitada.

The TTEST Procedure

Statistics

Variable	ano	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev
prod	08	83	1231.1	1324.4	1417.8	370.94	427.55
prod	09	88	1100.2	1188.3	1276.3	361.91	415.54
prod	Diff (1-2)		8.871	136.16	263.45	380.87	421.41

Variable	ano	Upper CL Std Dev	Std Err	Minimum	Maximum
prod	08	504.71	46.93	780	2416
prod	09	487.97	44.297	400	2400
prod	Diff (1-2)	471.68	64.48		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
prod	Pooled	Equal	169	2.11	0.0362
prod	Satterthwaite	Unequal	168	2.11	0.0363

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
prod	Folded F	82	87	1.06	0.7924

Figura 2. Uso do procedimento TTEST para a comparação de médias de duas amostras independentes.

Uso do SAS para amostras pareadas

Como exemplo de aplicação para esta situação, foi retirada do site do IBGE (2011) nova amostra de produtividade de feijão em municípios de Santa Catarina, no ano de 2008, para os mesmos que foram amostrados em 2009. Essas amostras podem ser tratadas como pareadas, se for considerado que os agricultores que produziram essa leguminosa nos dois anos devem ser os mesmos e não existe nada que leve a supor o contrário. Para tal situação, devem-se calcular as diferenças em produtividade nos dois anos e, então, aplicar o procedimento MEANS, conforme o programa abaixo, uma vez que a hipótese nula seria de que não há diferenças entre as médias.

```
PROC MEANS N MEAN STDERR T PRT;VAR DIFF;
```

O resultado pode ser visto na Figura 3. Também ali se verifica a existência de diferença significativa na produtividade do feijão nos dois anos em estudo. Como a diferença média é negativa, isso indica que o ano de 2008 apresentou maior produtividade que 2009.

```
                The MEANS Procedure

            Analysis Variable: PROD

 N            Mean       Std Error     t Value     Pr > |t|
-----------------------------------------------------------
88      -95.4545455      34.6705042       -2.75       0.0072
-----------------------------------------------------------
```

Figura 3. Uso do procedimento MEANS, do SAS, para o teste de médias em amostras pareadas.

Nesse caso, seria possível também adotar o procedimento UNIVARIATE, como visto na secção *Uso do SAS para uma amostra*, apenas retirando a opção MU0. Dessa forma, a programação seria:

```
PROC  UNIVARIATE CIBASIC;VAR PROD;
```

Na Figura 4, mostra-se apenas parte do resultado do processamento, que contém as partes de interesse para o teste de média.

```
Basic Confidence Limits Assuming Normality
   Parameter          Estimate     95% Confidence Limits
   Mean              -95.45455    -164.36593   -26.54316
   Std Deviation     325.23816     283.26490   381.92940
   Variance             105780         80239      145870
                Tests for Location: Mu0=0
      Test           -Statistic-     -----p Value------

      Student's t    t  -2.75319    Pr > |t|    0.0072
      Sign           M     -11.5    Pr >= |M|   0.0086
      Signed Rank    S      -485    Pr >= |S|   0.0047
```

Figura 4. Uso do procedimento UNIVARIATE, do SAS, para o teste de média de dados pareados.

Técnicas de amostragem

Conforme foi discutido no início deste capítulo, censo é a observação ou coleta de dados sobre todos os indivíduos de uma população; portanto, essa população deve ser finita. Trata-se de um processo caro, que demanda muito tempo e pessoal treinado para sua realização. Para populações infinitas, finitas incontáveis ou mesmo finitas compostas por grande número de indivíduos, a amostragem é o procedimento indicado, considerando-se

amostragem como a observação ou estudo de apenas alguns dos componentes da população. Diversas técnicas existem para isso, mas, neste capítulo, serão discutidas as duas mais comuns: amostragem aleatória simples e amostragem aleatória estratificada. A amostragem para o estudo de proporções, que pode ser feita tanto com uma técnica quanto com a outra, será apresentada ao final por mostrar uma característica especial na definição do número de amostras necessárias.

Amostragem aleatória simples

Considera-se aqui que os indivíduos que constituem a população não possuem características especiais que os diferencie um do outro além de seus caracteres próprios que os fazem conformar uma população uniforme. As populações podem ser finitas ou infinitas. As amostras são retiradas da população de forma aleatória de maneira que cada conjunto de n elementos retirados da população tenha a mesma probabilidade de ser o escolhido. Todos os possíveis conjuntos deveriam ser descritos, mas a tarefa é, em geral, larga e tediosa, já que mesmo para populações não muito grandes o número de possíveis conjuntos pode chegar a milhares ou até milhões de possíveis amostras.

Algumas alternativas são: restringir a amostra à parte da população que é de fácil acesso; selecionar a amostra ao azar ou utilizar voluntários para constituir a amostra. Para escolher a amostra ao azar, podem-se eleger os membros da amostra completamente ao azar ou selecionar a amostra por períodos, ou seja, a cada k elementos 1 é selecionado, mas apenas 1 indivíduo do primeiro conjunto de k elementos é tomado ao azar e outros são escolhidos sistematicamente. Nesse caso, o número de elementos na amostra é igual a N/k. K é definido de forma que a amostra tenha o tamanho n adequado, como definido no item *Tamanho de amostra*. No caso de populações infinitas, são amostrados indivíduos a cada k elementos encontrados, que pertençam à população até que se tenha avaliado n elementos.

As estimativas e as provas de hipótese a respeito dos parâmetros se fazem como especificado no item *Provas de hipótese* e no Capítulo 1. O principal problema se resume a determinar o tamanho de amostra adequado e a forma de determiná-lo se dá a seguir.

Tamanho de amostra

Duas condições são determinantes na definição do tamanho da amostra: o erro aceitável e a precisão desejada. Por erro aceitável se está referindo ao nível de significância do intervalo de confiança para a média, e por precisão desejada ao tamanho desse intervalo. A significância (α) para a distribuição de t é, em geral, definida como igual a 5% ou 0,05.

A precisão (*d*) é uma certa percentagem da média que se aceita como máximo erro na estimativa, como, por exemplo, 5% ou 10% da média (0,05*média ou 0,10*média), ou qualquer outra percentagem que se considere adequada.

Para populações infinitas, o tamanho da amostra, n, é calculado pela equação 20:

$$n = \frac{(t * s)^2}{d^2} \qquad [20]$$

em que t é o valor de tabela da distribuição t, com $nf1$ graus de liberdade e probabilidade igual a $\alpha/2$; s é o desvio padrão (estimado) da população e p é a precisão. É claro que o processo acima é um processo iterativo, já que o valor de t é dependente de n, que está sendo estimado. O desvio padrão, os graus de liberdade e a média são valores obtidos em trabalhos prévios ou de literatura. Quando esses valores não estão disponíveis, é comum a realização de uma pré-amostragem de tamanho não muito pequeno nem muito grande que os permita estimar.

Quando se trabalha com populações finitas, o valor n é denominado de n_0, e sobre isso é necessário fazer a correção para população finita, de modo que:

$$n = \frac{n_0}{1 + n_0 / N} \qquad [21]$$

Se o valor da variância disponível for proveniente de censos anteriores ou de amostras grandes, a fórmula 20 pode ser modificada, substituindo-se o valor de t da distribuição de t, por z da distribuição normal padronizada. Nesse caso, a fórmula é aplicada apenas uma vez. A fórmula 21 não tem qualquer modificação nesse caso.

Amostragem aleatória estratificada

Muitas vezes a população de interesse apresenta subpopulações, com características próprias, inclusive com variâncias distintas, de maneira que é necessário, além de conhecer os parâmetros da população, conhecer também os parâmetros desses grupos menores de indivíduos. Esse caso é o que se denomina de amostragem estratificada. As subpopulações podem ser formadas por variados processos ou características, como estrato social, nível de educação, bairros, cidades, províncias, países, raças, variedades, formas de cultivo, origem, etc.

É evidente que as estimativas de parâmetros discutidas anteriormente (na seção *Estatísticas de interesse*) são válidas somente para cada estrato. As estimativas para a população como um todo são específicas, assim como a definição do tamanho da amostra e sua distribuição entre os distintos estratos.

Nesse caso, a população tem h estratos, ($h = 1, 2, \ldots, H$), cada um com tamanho N_h, e a população tamanho N, que é igual à soma de todos os NH. Cada estrato tem seu fator de ponderação $W_h = N_h/N$. Da mesma forma, o tamanho total da amostra é n com n_h mostra por estrato.

As estimativas de média e variância de cada estrato são iguais a:

$$\bar{y}_h = \frac{y_{h1} + y_{h2} + \ldots + y_{hn}}{n_h} = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \quad [22]$$

e

$$s_h^2 = \frac{\sum_{i=1}^{n_h} y_{hi}^2 - \frac{\left(\sum_{i=1}^{n_h} y_{hi}\right)^2}{n_h}}{n_h - 1} \quad [23]$$

Estimativas dos parâmetros da população

Para obter a estimativa da média da população, chamada agora de média estratificada, usa-se a fórmula:

$$\bar{y}_{est} = \sum_{h=1}^{H} W_h * \bar{y}_h \quad [24]$$

A variância dessa média estratificada obtém-se por

$$s^2_{(\bar{y}_{est})} = \frac{1}{N^2} \sum_{h=1}^{H} g_h s_h^2 \quad [25]$$

em que

$$g_h = \frac{N_h(N_h - n_h)}{n_h} \quad [26]$$

O número efetivo de graus de liberdade dessa variância é calculado conforme a fórmula de Satterthwaite (1946):

$$GL = \frac{\left(\sum g_h s_h^2\right)^2}{\sum \frac{g_h^2 s_h^4}{n_h - 1}} \quad [27]$$

Definição do tamanho da amostra

O tamanho da amostra para um valor determinado da variância de média estratificada

$$\left(V = s^2_{(\bar{y}_{est})} = \frac{1}{N^2}\sum_{h=1}^{H} g_h s_h^2\right) \quad [28]$$

é definido por

$$n = \frac{\left(\sum W_h s_h\right)^2}{V + \sum W_h s_h^2 / N} \quad [29]$$

A partição dessa amostragem entre os estratos, procurando-se minimizar a variância da média estratificada, é dada por:

$$n_h = n * \frac{N_h s_h}{\sum N_h s_h} \quad [30]$$

Seleção de amostras

A seleção de indivíduos em cada um dos estratos faz-se de maneira aleatória e, para cada um deles, deve-se seguir os procedimentos indicados no caso de amostra aleatória simples.

Amostragem para proporções

Em amostragem para proporções, considere-se somente a(s) variável(is) que apresentam apenas dois valores possíveis, 0 (zero) e 1 (um), em que 0 significa que o indivíduo na amostra não possui a característica, e 1 indica que ele a tem. Exemplos desse tipo de amostragem são as pesquisa a respeito das preferências eleitorais, de incidência de enfermidades ou vacinações, de existência de defeitos em peças ou produtos, etc.

A estimativa de p (proporção ou percentagem de ocorrência da característica) e sua variância foram detalhadas no item *Proporção*.

Tamanho da amostra

Uma análise simples da fórmula para a variância de p, para um valor n, constante, mostra que a máxima variância ocorre quando $p = 0{,}5$ e que ela diminui quando p aumenta ou diminui.

Admitindo-se certa margem de erro (d) na estimativa com uma significância igual a α, ou seja, $Prob(|p < P| = d) = \alpha$, e associando-se essa probabilidade com o intervalo de confiança para p, segue que:

$$n_0 = \frac{z^2 * (p * q)}{d^2} \quad [31]$$

Com o ajuste para população finita segue que:

$$n = \frac{n_0}{1 + n_0 / N} \quad [32]$$

Nas fórmulas 31 e 32, z é a abcissa da curva normal que define uma área α da cauda, q é igual a 1 - p.

O produto $p*q$ corresponde à variância e é fácil verificar que ela é máxima quando $p = q = 0,5$; diminuindo à medida que p cresce ou diminui. Dessa forma, pode-se determinar o tamanho máximo de amostragem para determinados valores de α e d.

Exemplo de definição de tamanho de amostra

Admitindo-se que o valor da precisão desejada (d) seja igual a 0,05, ou seja, um intervalo de confiança de 5% para mais ou para menos, e sabendo-se que, para um erro (α) de 5%, o valor de z é aproximadamente igual a 1,96, o tamanho máximo de amostra é de 385 como se vê a seguir, para uma população infinita. Caso a população seja finita, basta fazer o ajuste como descrito na fórmula 32.

$$n_0 = \frac{1,96^2 * (0,5 * 0,5)}{0,05^2} = \frac{3,8416 * 0,25}{0,0025} = 384,16 \approx 385$$

Dessa forma, é possível para esse tipo de amostragem a criação de tabelas com o número adequado de amostras para diversos tamanhos de população, erro e precisão desejados (Tabela 1).

Tabela 1. Número de amostras necessárias em amostragem para proporções com vários tamanhos de população (N), erro (α) e precisão (d), para diversos valores de p. Para valores de p maiores que 0,50, deve-se usar o valor correspondente a 1fp.

		N = infinito			N = 10.000			N = 5.000			N = 1.000		
	p	d = 0,1	d = 0,05	d = 0,025	d = 0,1	d = 0,05	d = 0,025	d = 0,1	d = 0,05	d = 0,025	d = 0,1	d = 0,05	d = 0,025
α = 0,05	0,01	4	16	61	4	16	61	4	16	61	4	16	58
	0,05	19	73	292	19	73	284	19	72	276	19	69	227
	0,10	35	139	554	35	138	525	35	136	499	34	123	357
	0,15	49	196	784	49	193	728	49	189	678	47	164	440
	0,20	62	246	984	62	241	896	62	235	823	59	198	496
	0,25	73	289	1.153	73	281	1.034	72	274	937	69	225	536
	0,30	81	323	1.291	81	313	1.144	80	304	1.027	75	245	564
	0,35	88	350	1.399	88	339	1.228	87	328	1.094	81	260	584
	0,40	93	369	1.476	93	356	1.287	92	344	1.140	86	270	597
	0,45	96	381	1.522	96	368	1.321	95	355	1.167	88	276	604
	0,50	97	385	1.537	97	371	1.333	96	358	1.176	89	278	606
α = 0,025	0,01	5	20	80	5	20	80	5	20	79	5	20	75
	0,05	24	96	382	24	96	368	24	95	355	24	88	277
	0,10	46	181	723	46	178	675	46	175	632	44	154	420
	0,15	64	256	1.024	64	250	929	64	244	850	61	204	506
	0,20	81	322	1.285	81	312	1.139	80	303	1.023	75	244	563
	0,25	95	377	1.506	95	364	1.309	94	351	1.158	87	274	601
	0,30	106	422	1.686	105	405	1.443	104	390	1.261	96	297	628
	0,35	115	457	1.827	114	438	1.545	113	419	1.339	104	314	647
	0,40	121	482	1.927	120	460	1.616	119	440	1.391	108	326	659
	0,45	125	497	1.987	124	474	1.658	122	453	1.422	112	332	666
	0,50	126	502	2.008	125	479	1.673	123	457	1.433	112	335	668

Continua...

Tabela 1. Continuação.

		N = infinito			*N* = 10.000			*N* = 5.000			*N* = 1.000		
	p	*d* = 0,1	*d* = 0,05	*d* = 0,025	*d* = 0,1	*d* = 0,05	*d* = 0,025	*d* = 0,1	*d* = 0,05	*d* = 0,025	*d* = 0,1	*d* = 0,05	*d* = 0,025
α = 0,01	0,01	11	43	172	11	43	170	11	43	167	11	42	147
	0,05	52	206	823	52	202	761	52	198	707	50	171	452
	0,10	98	390	1.559	98	376	1.349	97	362	1.189	90	281	610
	0,15	139	553	2.209	138	525	1.810	136	498	1.533	123	357	689
	0,20	174	693	2.771	172	649	2.170	169	609	1.783	149	410	735
	0,25	203	812	3.248	199	752	2.452	196	699	1.969	169	449	765
	0,30	228	910	3.637	223	835	2.668	219	770	2.106	186	477	785
	0,35	247	985	3.940	242	897	2.827	236	823	2.204	199	497	798
	0,40	260	1.040	4.157	254	943	2.937	248	861	2.270	207	510	807
	0,45	268	1.072	4.287	262	969	3.001	255	883	2.309	212	518	811
	0,50	271	1.083	4.330	264	978	3.022	258	891	2.321	214	520	813

Capítulo 18

Usando o SAS na análise dos dados

Introdução

Conforme foi discutido no Capítulo 1, para a realização de uma investigação científica cinco etapas devem ser seguidas, das quais duas delas são o objeto principal deste livro: o planejamento do experimento ou amostragem e a tabulação e análise dos dados. Ambos os aspectos foram discutidos nos Capítulos 1 a 17. Nesta segunda edição, revista e ampliada, foi introduzido o uso de um software para a análise dos dados, o Statistical Analysis System (SAS), usado pela Embrapa desde seu início e por centenas de instituições no mundo. Outros softwares podem ser usados, evidentemente, mas o amplo conhecimento do SAS, sua potência e facilidade de uso o tornaram indicado para este livro.

Evidentemente, não se pretende neste capítulo fazer uma descrição exaustiva do software, tarefa impossível, pois ele possui uma multitude de processos e opções que não pode ser alcançada em um livro texto de estatística. O que se pretende aqui é descrever a estrutura básica de um programa SAS e mostrar quais são os procedimentos mais utilizados em estatística experimental e estatística amostral. Ao leitor interessado em um conhecimento maior do SAS, recomenda-se a realização de treinamentos específicos e o uso de seus manuais (SAS, 2011).

De uma forma geral, um programa SAS deveria conter três partes: uma descritiva do projeto e dos dados obtidos, uma entrada de dados, criação e transformação de variáveis e finalmente a análise dos dados. A primeira parte é totalmente facultativa, uma vez que não tem qualquer influência na entrada e na análise dos dados, mas é recomendada pelo fato de todos os detalhes da investigação ficarem aí armazenados.

A segunda parte, entrada dos dados, contém três comandos essenciais: DATA, INPUT e CARDS.

A terceira parte, análise dos dados, aparece após a leitura dos dados e será discutida mais detalhadamente em seções específicas deste capítulo.

Ao ser utilizado, o SAS abre uma janela no computador composta de três seções: uma para digitação dos dados e do programa, chamada de EDITOR; outra para a saída do programa, comentários e indicação de erros, tempo de processamento, etc., chamada de LOG; e uma terceira, chamada de OUTPUT, na qual aparece o resultado do programa feito, caso não haja erros de programação ou de leitura dos dados.

Informações gerais e entrada de dados

As informações gerais que compõem a primeira parte de um programa SAS são, em geral, as seguintes:

- Identificação do projeto e dos pesquisadores responsáveis.
- Datas relativas a sua execução, por exemplo, data de plantio, data de emergência, data de floração, data de frutificação, data de colheita, em se tratando de plantas; data de cobertura, data do parto, data das pesagens, etc., em se tratando de animais, e assim por diante.
- Descrição das variáveis de classificação e das variáveis resposta.
- Delineamento experimental adotado, número de repetições, tamanho da parcela, precisão estatística adotada, testes a serem empregados, etc.

Deve ser lembrado que algumas dessas informações podem também ser consideradas variáveis resposta, principalmente quando se trata de função dos tratamentos.

Em geral, esta primeira parte do programa consiste de alguns comandos SAS, não obrigatórios, mas de uso recomendado. São eles:

```
OPTIONS PS= xx LS=XX;
COMMENT
```

No comando OPTIONS, PS = xx e LS = xx indicam ao sistema, em caso de necessidade de impressão do resultado, o tamanho da página referindo-se ao número de linhas (PS) e ao número de caracteres por linha (LS). Para o papel formato A4, o número de linhas é de 60 e o de caracteres por linha é de 80.

O comando COMMENT indica ao sistema que o que se segue são apenas informações gerais que não devem ser processadas. Todo comando ou procedimento SAS termina com um ponto e vírgula (;).

Um exemplo desses comentários pode ser retirado de um arquivo para processamento de dados de um projeto do pesquisador Luís Fernando Stone, da Embrapa Arroz e Feijão, no qual as datas de realização de operações foram omitidas, como segue.

```
OPTIONS PS=60 LS=80;
COMMENT


PROJETO: EFEITO DO CONJUNTO DE TECNOLOGIAS APLICADAS AO
   SISTEMA DE PRODUÇÃO DE ARROZ IRRIGADO
PESQUISADOR RESPONSÁVEL: LUÍS FERNANDO STONE
LINHA DE PESQUISA: MANEJO E TRATOS CULTURAIS
DELINEAMENTO EXPERIMENTAL: BLOCOS COMPLETOS CASUALIZADOS,
              COM 4 REPETIÇÕES. FATORIAL 2⁵ EM PARCELAS
              SUBDIVIDIDAS. FATORES "A" E "M" NAS PARCELAS,
              OS DEMAIS NAS SUBPARCELAS.
VARIÁVEIS DE CLASSIFICAÇÃO:
           REP=REPETIÇÃO(1,2,3,4)
             I=CONTROLE DE INVASORAS (1=STAM F-34
             2=RONSTAR)
             A=ADUBAÇÃO NPK(1=30-90-30 2=60-90-30)
             M=IRRIGAÇÃO(1=PERMANENTE 2=INTERMITENTE)
             C=CULTIVAR (1=IAC 435 2=IR 841-63-5-L-9-33)
             D=CONTROLE DE DOENÇAS (1=SEM CONTROLE
                     2=CONTROLE COM MANZATE 50PM)
VARIÁVEIS RESPOSTA:
           GCH=NÚMERO DE GRÃOS CHEIOS MÉDIA DE 10
               PANÍCULAS
           GVA=NÚMERO DE GRÃOS VAZIOS MEDIA DE 10
               PANÍCULAS
         P100=PESO DE 100 GRÃOS
         ALT=ALTURA
         NPAN=NÚMERO DE PANÍCULAS / M
         NPER=NÚMERO DE PERFILHOS / M
         PROD=PRODUÇÃO
       PPFERT=PORCENTAGEM DE PERFILHOS FÉRTEIS
       PGVAZ=PORCENTAGEM DE GRÃOS VAZIOS
AS VARIÁVEIS EM PORCENTAGEM SERÃO TRANSFORMADAS PELA
FUNÇÃO ARSIN (ARCO SENO) E AS EM NÚMERO PELA FUNÇÃO SQRT
(RAIZ QUADRADA);
```

Na segunda parte, o primeiro comando, DATA, nomeia o conjunto de dados a serem introduzidos com a seguinte sintaxe:

DATA "nome"; em que "nome" é a denominação com que se quer vincular os dados. Caso não seja usado o SAS, deve ser nomeado como DATA A. Se houver mais de um conjunto

de dados, cada um deverá ter seu nome. Caso fique a cargo do sistema a necessidade de nomeá-los, serão chamados sucessivamente de A, B, C, etc.

O segundo comando, INPUT, identifica e qualifica as variáveis, com a sintaxe:

INPUT var1 var2 var3, … ; em que var1, var2, var3, …, são os nomes que se quer dar às variáveis, sejam elas de classificação ou de resposta. Podem ou não ser seguidas de números que indiquem sua posição de leitura nas linhas de dados.

Finalmente, o terceiro, CARDS, indica o início da leitura dos dados.

A necessidade de criação ou transformação de variáveis deve ser suprida por operações matemáticas entre os comandos INPUT e CARDS.

A seguir, encontram-se dois exemplos a respeito disso. O primeiro deles foi retirado do mesmo processamento anterior, dos quais se mostrou parte das informações gerais.

```
DATA fjpz;
INPUT REP 1 M 2 A 3 C 4 D 5 I 6 GCH 7-10 1 GVA 11-14 1 P100 15-18 2
ALT 19-21 NPAN 22-23 NPER 24-26 PROD 27-32 1;
PGVAZ=(GVA*100)/(GCH+GVA);
PGVAZT=ARSIN(SQRT(PGVAZ/100));
PPFERT=(NPAN*100)/NPER;
PPFERTT=ARSIN(PPFERT/100);
GCHT=SQRT(GCH);
GVAT=SQRT(GVA+0.5);
NPANT=SQRT(NPAN);
NPERT=SQRT(NPER);
CARDS;
111111 666 4627311048 5324450;
```

No exemplo citado, o conjunto de dados se chamará *fjpz*, as variáveis são todas elas numéricas e lidas em sequência. A variável REP está na primeira coluna de dados (REP 1); a variável M, na segunda (M 2); a variável GCH está nas colunas 7 a 10, com 1 decimal (GCH 7-10 1), e assim por diante. Após o comando INPUT aparecem as funções de criação ou de transformação de variáveis, antes do comando CARDS. Por exemplo, a variável PGVAZ foi criada como porcentagem em função da leitura dos números de grãos cheios e vazios por panícula (PGVAZ=(GVA*100)/(GCH+GVA)). Logo após sua criação, já vem a função matemática para sua transformação em arco seno (PGVAZT=ARSIN(SQRT(PGVAZ/100)). Após CARDS, aparece a primeira linha de informações.

O outro exemplo encontra-se no Capítulo 3, no qual uma das variáveis de classificação era de carácter alfanumérico. Nesse caso, no comando INPUT a variável é seguida pelo símbolo $, que indica variável é alfanumérica, e depois pelo número das colunas, inicial ou

final, onde aparece, o que define o número máximo de caracteres que ela possui. Nesse exemplo, o tamanho máximo da variável é de 11 caracteres localizados entre as colunas 1 e 11, inclusive. As duas outras variáveis são numéricas e, como estão sempre separadas por espaços em branco, não existe a necessidade de identificação de suas posições ou colunas. Segue-se também uma linha de dados

```
DATA A;
INPUT TRATAMENTO $ 1-11 BLOCO PROD;
CARDS;
CNFP8015    1     1112.5;
```

Caso exista um arquivo tipo texto (nome.txt), ou planilha do Excel ou Lotus, é possível fazer a leitura desses dados usando o procedimento PROC IMPORT. Suponha que se tenha um arquivo chamado baf.txt, que possui três variáveis, e que, nesse arquivo, a primeira linha contenha os nomes das variáveis e a segunda linha e seguintes contenham os dados dessas variáveis, e que os dados e os nomes sejam separados por espaço. Para ler esses dados e criar um arquivo SAS, o programa é o seguinte:

```
PROC IMPORT DATAFILE="C:\DOCUMENTS AND SETTINGS\FRANCISCO ZIMMERMAN\
MY DOCUMENTS\LIVRO\ARQUIVOS_TRABALHO\BAF.TXT" OUT=MY2;
GETNAMES=YES;
```

O comando DATAFILE indica o caminho para a localização do arquivo baf.txt no disco C. O comando OUT= fornece o nome do arquivo SAS de trabalho, e o comando GETNAMES= indica que a primeira linha do arquivo contém os nomes das variáveis.

Programa semelhante pode ser feito para outros tipos de arquivos. Para tal recomenda-se a leitura dos manuais SAS.

Análise dos dados

No que diz respeito à análise de dados, procura-se agrupar os procedimentos e comandos segundo o tipo de análise a ser empregada no estudo dos dados. Alguns procedimentos podem ser usados em mais de um tipo de análise, ou podem ser usados de forma diferenciada conforme a informação desejada pelo pesquisador. Por causa disso, esta seção será subdividida da seguinte forma:

- Procedimentos gerais.

- Procedimentos para análise exploratória de dados.
- Procedimentos para análise de dados de amostras.
- Procedimentos para análise de variância.
- Procedimentos para estudos de regressão.
- Procedimentos para estatística não paramétrica.

Procedimentos gerais

Incluem-se aqui alguns procedimentos que podem ser usados pelo usuário independentemente do tipo de análise em que se está interessado após a leitura dos dados. Outros detalhes ou procedimentos podem ser encontrados no manual do SAS, chamado de SAS BASICS ou Base SAS 9.2 Procedure Guide.

- DATA nome; SET nome;

Usado para criar um subconjunto de dados, ou a criação de novas variáveis, a partir do conjunto identificado pelo comando SET. Exemplos disso se seguem:

DATA B; SET A; IF TRAT LE 10;

O conjunto de dados B terá todos os dados do conjunto geral A, para os tratamentos identificados por números entre 1 e 10, inclusive.

DATA A; SET A; LPROD=LOG(PROD);

O conjunto de dados A terá uma nova variável, chamada de LPROD, cujo valor é igual ao logaritmo natural de PROD.

DATA C;SET A B;

O conjunto de dados C será composto por todos os dados dos conjuntos A e B, colocados um acima do outro. As variáveis deveriam ser todas comuns aos dois conjuntos A e B. Caso isso não ocorra, as variáveis inexistentes em uma base de dados aparecem como tendo os valores perdidos no conjunto final nas observações que pertencem ao outro conjunto de dados.

- DATA nome; MERGE nome1 nome2 nome3 ….;BY var1 var2 …;

Aqui o que se deseja é colocar dois conjuntos de dados, lado a lado, que possuam variáveis comuns, identificadas pelo comando BY, e outras não comuns, não identificadas

em BY, de modo que o conjunto final contenha as variáveis comuns e não comuns. Esse comando deve vir antecedido pelos comandos a seguir.

PROC SORT DATA=nome; BY var1 var2 …; (um para cada conjunto citado no comando MERGE)

Por exemplo, suponha-se que existam dois conjuntos de dados, A e B, em que em A apareçam as variáveis BLOCO TRAT GERM NP ALT, e em B apareçam BLOCO TRAT PROD P1OO NVAG. Nesse caso, tem-se então:

```
PROC SORT DATA=A;BY BLOCO TRAT;
PROC SORT DATA=B;BY BLOCO TRAT;
DATA C;MERGE A B;BY BLOCO TRAT;
```

O conjunto C tem então as variáveis BLOCO TRAT GERM NP ALT PROD P100 NVAG; nessa ordem.

- PROC PRINT:VAR var1 var2 var3 …;

O procedimento PRINT faz a listagem virtual do conjunto de dados das variáveis citadas no comando VAR. Esse comando, quando não usado, faz que todas as variáveis do conjunto de dados sejam listadas. É bom lembrar aqui que a impressão em papel de um processamento SAS se faz pela impressão do arquivo de saída criado pelo sistema na janela OUTPUT, o qual deve ter um nome antes de ser gravado. Se não houver interesse em gravar o resultado do processamento, o conteúdo da janela OUTPUT pode ser impresso diretamente. Para o exemplo dado em b, caso se deseje apenas imprimir as variáveis de classificação e a produção obtida no ensaio, o comando seria:

```
PROC PRINT;VAR BLOCO TRAT PROD;
```

Para imprimir todo o arquivo, basta omitir o comando VAR.

- PROC MEANS opções;BY var1 var2 …;VAR var1 var2 …;

O procedimento MEANS permite calcular as estatísticas básicas das variáveis listadas no comando VAR para todos os níveis das variáveis incluídas no comando BY. O comando BY só deve ser incluído caso seja necessário calcular as estatísticas para as combinações de níveis das variáveis citadas no comando. Permite também a criação de conjunto de

dados com essas estatísticas. As variáveis listadas no comando BY não podem aparecer no comando VAR. Exemplos desses usos encontram-se a seguir:

```
PROC MEANS MEAN VAR STDERR;BY ANO;VAR X1 X2 X2;
```

Aqui se está pedindo o cálculo da média, da variância e do desvio padrão das variáveis X1, X2 e X3, para cada um dos anos incluídos no conjunto de dados. Isso só é válido se, para cada ano, mais de uma observação existe.

```
PROC SORT;BY ANO;
PROC MEANS MEAN NOPRINT;BY ANO;VAR X1 X2 X3;
OUTPUT OUT DATA=A MEAN=X1 X2 X3;
PROC PRINT DATA=A;
```

No exemplo citado, é criado um conjunto de dados, chamado de A, que contém quatro variáveis, ANO, X1, X2 e X3, além de duas outras criadas pelo sistema, chamadas de _TYPE_ e _FREQ_. A variável _TYPE_ contém informação sobre o tipo da variável e _FREQ_ contém a frequência de observações para cada uma das variáveis.

Procedimentos para análise exploratória de dados

Como visto no Capítulo 16, alguns poucos procedimentos são necessários para esse tipo de análise. Os procedimentos mais comuns são então:

- PROC UNIVARIATE opções; VAR var1 var2 var3 …;

Este procedimento, quando usado sem opções, produz uma saída que contém, para cada uma das variáveis citadas no comando VAR, as estatísticas numéricas desejadas, quais sejam: os momentos – N, média, variância, desvio padrão, assimetria, curtose, pesos, soma total, somas de quadrados (corrigida e não corrigida), erro padrão da média e coeficiente de variação; as estatísticas básicas de posição e de dispersão, teste para a média ($H0 : \mu = 0$), quantis e observações extremas.

O comando para análise da variável PRODUCAO seria então:

```
PROC UNIVARIATE;VAR PRODUCAO;
```

Entre as opções que podem ser utilizadas, a de maior interesse para a análise exploratória é a opção PLOT, que produz a saída adicional do gráfico de ramos e folhas, do gráfico de caixa e bigodes e o gráfico de distribuição normal. O programa então é:

```
PROC UNIVARIATE PLOT;VAR PRODUCAO;
```

- PROC CHART;VAR var1 var2 var3 …/opções;

Este procedimento é utilizado para gerar um histograma para cada uma das variáveis listadas no comando VAR. Entre as opções, a mais interessante é a que define o número de colunas no histograma, chamada LEVELS. O programa para gerar um histograma com dez classes ou dez colunas é:

PROC CHART;VAR PRODUCAO/LEVELS=10;

- PROC PLOT;PLOT var1*var2=var3;

O procedimento leva à construção de um gráfico de pontos, no qual os eixos são definidos pelas variáveis var1 (abcissa) e var2 (ordenada), e a variável var3 identifica os pontos no gráfico. Esta última variável não necessita ser usada e pode ser substituída por um símbolo qualquer, entre aspas, por exemplo, "*" ou "+". Nesses casos, os pontos serão identificados por um asterisco ou pelo sinal mais. Na forma mais simples que pode ser usada no comando PLOT, aparecem apenas as variáveis var1 e var2 (PROC PLOT; PLOT var1*var2;). Nesse caso os pontos no gráfico serão identificados pelos algarismos 1, 2, 3, …, em que 1 indica a existência de um ponto, 2 a ocorrência de dois pontos coincidentes, e assim por diante.

As figuras 6 a 9 (Capítulo 16) contêm exemplos de aplicação desses procedimentos.

Procedimentos para análise de dados de amostras

Para este tipo de dado, além dos processos mostrados no item *Procedimentos para análise exploratória de dados*, os testes de médias e de variâncias, além das estimativas das médias por intervalo, fazem parte da análise de dados amostrais, os quais foram mostrados no Capítulo 17, com resultados nas Figuras 1 a 4, e são apresentados aqui com um pouco mais de detalhes.

- Procedimentos para teste de uma média e intervalos de confiança – o procedimento aqui utilizado é o UNIVARIATE, com a seguinte sintaxe:

  PROC UNIVARIATE MU0=valor CIBASIC;VAR var1 var2 …;

A linha de programação citada, além das estatísticas básicas, como discutido no item *Procedimentos para análise exploratória de dados*, leva o sistema a um teste t da média, com hipótese nula (H0 : μ = valor), e ao cálculo dos intervalos de confiança para a média, para a variância e para o desvio padrão.

O "valor" pode ser qualquer número com o qual se quer testar a média da amostra.

- Procedimentos para teste de duas médias

Para realizar o teste t de comparação de duas médias oriundas de duas amostras independentes, o procedimento SAS adequado é o TTEST, com a seguinte sintaxe:

PROC TTEST;CLASS var; VAR var1 var2 …;

A variável de classificação (CLASS var) é aquela que identifica as duas amostras, e o comando VAR identifica quais variáveis medidas nas amostras devem ser analisadas. O resultado mostrará as estatísticas básicas para cada um dos dois níveis de classificação (média, desvio padrão) e seus intervalos de confiança, além do erro padrão da média e dos valores máximo e mínimo. Também apresenta a diferença entre as duas médias, e seu desvio padrão além dos intervalos de confiança. Depois faz o teste t entre as duas médias, tanto sob a hipótese de igualdade de variâncias quanto de diferença entre elas, e por último o teste f para as variâncias. Resultado desse processamento pode ser visto na Figura 2 do Capítulo 17.

- Procedimentos para teste de médias de dados pareados

Este é o caso de amostras pareadas, situação discutida no Capítulo 17, na seção *Provas de hipótese para duas médias*. Nesse caso, já ao se fazer a entrada dos dados das amostras coletadas nas duas situações distintas, deve-se calcular a diferença entre cada um dos pares de dados, e depois aplicar o procedimento MEANS para verificação da hipótese de nulidade na diferença entre as situações. A sintaxe de programação é a seguinte:

PROC MEANS N MEAN STDERR T PRT;VAR diferença;

O significado dessa programação é que se está solicitando ao sistema o número de observações (N), a média das diferenças (MEAN), o seu desvio padrão (STDERR), o teste t para média (H0 : $\mu = 0$) e a probabilidade desse teste. A variável a ser testada é aquela identificada pelo comando VAR (diferença).

Seria possível também usar o procedimento UNIVARIATE, cujo resultado é mais completo, e com a necessidade de usar somente a opção CIBASIC. Exemplo dessa situação também aparece no Capítulo 17, na secção *Uso do SAS para amostras pareadas*.

Procedimentos para análise de variância

No caso dos delineamentos de experimentos, dois procedimentos SAS permitem a realização de praticamente todas as análises de variância. São eles o ANOVA e o GLM.

O procedimento ANOVA é o adequado para os experimentos balanceados e completos, isto é, sem falta de dados. Incluem-se nesse caso os delineamentos inteiramente casualizado, os blocos completos casualizados e o quadrado latino, com um ou mais de um fator e em suas variações de parcelas divididas, em faixa, etc. Sua vantagem sobre o GLM para essas situações é a economia de memória e a rapidez de processamento. Por este último comentário, fica fácil deduzir que o outro procedimento, GLM, pode ser aplicado também a essas situações, bem como àquelas com dados perdidos, não balanceados, em blocos incompletos, em análise de covariância e naquelas em que há a necessidade de uso de contrastes, seja para comparação de tratamentos seja para estudos de regressão, além de ser também usado em análise de regressão, como será visto na seção *Procedimentos para análise de regressão*.

Outro procedimento que pode ser usado é o LATTICE, específico para os reticulados quadrados balanceados e parcialmente balanceados, e para alguns reticulados retangulares.

O ANOVA tem uma sintaxe básica composta pelos seguintes comandos:

```
PROC ANOVA opções; BY variáveis; CLASS variáveis;
MODEL respostas = efeitos;
TEST H=efeitos E=erro;
MEANS efeitos/opções;
```

Os comandos PROC ANOVA, CLASS e MODEL são obrigatórios e devem aparecer nessa ordem. Os demais aparecem após esses comandos em qualquer ordem, além de outros que podem ser consultados nos manuais do SAS[1].

Existem apenas quatro opções do PROC ANOVA. A mais usada é a opção DATA=nome do conjunto de dados, que especifica qual deve ser usado no processamento se houver mais de um conjunto.

O comando BY especifica a repetição do processamento de análise de variância para cada nível da variável ou combinações de níveis das variáveis especificadas. Por sua vez, o comando CLASS determina qual ou quais variáveis são usadas como variáveis de classificação ou efeitos no modelo matemático.

[1] SAS/STAT User´s Guide. v. 1 (SAS, 2011).

Para o comando MODEL, "respostas" corresponde às variáveis de resposta para as quais se deseja a análise de variância, "efeitos" corresponde às variáveis de classificação e devem ser apresentadas segundo o modelo matemático do delineamento experimental.

O comando TEST é usado quando algum ou alguns dos efeitos devem ser testados por outro efeito que é considerado erro, além do erro experimental normalmente calculado. Nesse caso, H especifica qual ou quais efeitos devem ser testados, e E especifica qual efeito é considerado o testador. O comando TEST é normalmente usado em experimentos com parcelas divididas e em suas variações, em experimentos com múltiplas observações por parcela ou nos casos em que algum fator é considerado aleatório.

Finalmente o comando MEANS especifica sobre que "efeitos" devem ser calculadas e testadas as médias: os testes para comparação de médias, seu nível de probabilidade (ALPHA=porcentagem), o erro a ser usado (E = efeito, como em TEST), e a forma de apresentação (LINES), definidos em "opções". Entre os testes disponíveis no SAS, destacam-se: Bonferroni (BON), Duncan (DUNCAN), Dunnett (DUNNETT, DUNNETTL e DUNNETTU), Ryan-Einot-Gabriel-Welsch (REGWF e REGWQ), Scheffé (SCHEFFE), Student-Neuman-Keuls (SNK), t (T e LSD) e Tukey (TUKEY). A opção LINES determina que a igualdade entre duas médias seja representada por linha contínua em vez de letras que é a forma padrão.

Nos Capítulos 3 em diante, podem ser encontradas diferentes formas de uso do procedimento ANOVA. Algumas delas encontram-se nos exemplos a seguir:

- Delineamento inteiramente casualizado

```
PROC ANOVA;CLASS TRAT;
MODEL RESPOSTA=TRAT;
MEANS TRAT/TUKEY ALPHA=0.01;
```

- Delineamento em blocos completos ao acaso, com mais de uma medida por parcela

```
PROC ANOVA;CLASS BLOCO TRAT;
MODEL RESPOSTA=BLOCO TRAT BLOCO*TRAT;
TEST H=TRAT E=BLOCO*TRAT;
MEANS TRAT/SNK E=BLOCO*TRAT;
```

- Experimento fatorial, com três fatores (A, B e C) em blocos completos.

```
PROC ANOVA;CLASS BLOCO A B C;
MODEL RESPOSTA=BLOCO A|B|C;
MEANS A|B|C/LSD;
```

No exemplo *a*, a opção ALPHA=0.01 determina que o teste de Tukey seja realizado no nível de probabilidade de 1%. No exemplo *b*, a interação BLOCO*TRAT corresponde ao erro experimental, o qual deve ser usado como o termo de erro ao ser realizada a comparação de médias pelo teste de SNK, indicado pela opção E=BLOCO*TRAT, enquanto o erro estimado na análise (error) corresponde ao erro amostral. Já no exemplo *c*, a forma de escrita A|B|C indica ao sistema que devem ser considerados na análise os efeitos principais (A, B e C), as interações simples (A*B, A*C, B*C) e a interação dupla (A*B*C), o que simplifica muito o programa.

O procedimento GLM tem uma sintaxe básica muito parecida com a de ANOVA e com a de outros comandos próprios:

```
PROC GLM opções; BY variáveis; CLASS variáveis;
MODEL respostas = efeitos/opções;
TEST H=efeitos E=erro;
MEANS efeitos/opções;
LSMEANS efeitos/opções;
CONTRAST "nome" efeitos "valores para os efeitos"/opções;
```

Entre as opções do PROC GLM, merecem destaque DATA=(nome do conjunto de dados) e OUTSTAT=(nome do conjunto de dados). Para a opção DATA, valem os comentários feitos no PROC ANOVA. Para a opção OUTSTAT, recomenda-se a leitura do exemplo no Capítulo 10 (seção *Fatorial 3^3 com partes das interações simples adicionadas ao estimador de erro*, subitem *Uso do SAS*), em que essa opção cria um novo conjunto de dados, que contém as variáveis _TYPE_, _SOURCE_, _NAME_, SS, DF, F e PROB. A variável alfanumérica _SOURCE_ contém o nome do efeito no modelo ou do contraste, para o qual as estatísticas correspondentes são geradas. _NAME_ contém o nome da variável dependente no modelo. SS, DF, F e PROB contêm respectivamente a soma de quadrados, os graus de liberdade, o valor do teste F e a sua probabilidade, correspondente a cada uma das fontes em _SOURCE_.

Os comandos BY, CLASS, MODEL, TEST e MEANS têm a mesma função que no PROC ANOVA.

Em MODEL, algumas opções devem ser realçadas. SOLUTION imprime a solução para as equações normais, útil no caso da análise de covariância, pois ali se tem a estimativa do coeficiente de regressão para ajuste das médias. SS1, SS2, SS3 e SS4 imprimem as somas de quadrados segundo o tipo de funções estimáveis para cada efeito. Seus usos foram estudados em capítulos anteriores.

LSMEANS efeitos/opções deve ser usado no caso dos delineamentos em blocos incompletos, em análise de covariância e nos casos de experimentos com dados perdidos,

em que as médias devem ser ajustadas por algum procedimento adequado. Veja exemplos no Capítulo 4 e nos capítulos subsequentes. Diferentemente do comando MEANS, as comparações de médias só são realizadas pelo teste t, com as opções PDIFF e TDIFF. PDIIFF imprime as probabilidades para a hipótese de igualdade entre todas as combinações entre duas médias, enquanto TDIFF imprime também os valores dos testes t.

CONTRAST="nome" efeitos "valores para efeitos"/opções é usado nos casos de necessidade de estudos de regressão para os tratamentos do tipo quantitativo (contrastes polinomiais) ou comparações especiais de médias de tratamentos qualitativos (contrastes de interesse). Para a obtenção correta das somas de quadrados para os contrastes, é necessário lembrar que a ordem correta dos "valores para efeitos" depende da ordem em que os efeitos são explicitados no comando CLASS, principalmente no caso de ensaios fatoriais. Essa ordem estabelece a sequência dos efeitos na construção da matriz X'X.

Por exemplo, num fatorial 3 x 3 com quatro repetições, se em CLASS designamos a ordem BL A B, a matriz, na parte que interessa aos contrastes, terá suas colunas identificadas como: A1 A2 A3 B1 B2 B3 A1B1 A1B2 A1B3 A2B1 A2B2 A2B3 A3B1 A3B2 A3B3.

Os contrastes para os efeitos linear e quadrático de A e B serão então digitados como:

```
CONTRAST "LIN DE A" A -1 0 1;
CONTRAST "QUA DE A" A 1 -2 1;
CONTRAST "LIN DE B" B -1 0 1;
CONTRAST "QUA DE B" B 1 -2 1;
```

A interação de A e B, que possui quatro graus de liberdade, terá quatro linhas de informações, correspondentes às combinações das interações linear de A com linear de B, linear de A com quadrática de B, quadrática de A com linear de B e quadrática de A com quadrática de B. Podem ser escritos os quatro contrastes, um para grau de liberdade, ou apenas um contraste, que terá então os 4 GL da interação. As duas formas encontram-se a seguir, devendo ser notado que os vetores de contrastes para o segundo caso terminam com vírgula (,), exceto o último deles, que tem a conclusão tradicional do SAS de ponto e vírgula (;).

```
CONTRAST "LIN A X LIN B" A*B 1 0 -1 0 0 0 -1 0 1;
CONTRAST "LIN A X QUA B" A*B -1 2 -1 0 0 0 1 -2 1;
CONTRAST "QUA A X LIN B" A*B -1 0 1 2 0 -2 -1 0 1;
CONTRAST "QUA A X QUA B" A*B 1 -2 1 -2 4 -2 1 -2 1;
```

ou

```
CONTRAST "INTER A x B" A*B 1 0 -1 0 0 0 -1 0 1,
                           A*B -1 2 -1 0 0 0 1 -2 1,
                           A*B -1 0 1 2 0 -2 -1 0 1,
                           A*B 1 -2 1 -2 4 -2 1 -2 1;
```

Se o interesse do pesquisador está em conhecer o efeito de um fator dentro de cada um dos níveis do outro fator, por exemplo, linear e quadrático de A dentro do nível 1 de B, os dois contrastes serão:

```
CONTRAST "LIN A d. B=1" A -1 0 1 A*B -1 0 0 0 0 0 1 0 0;
CONTRAST "QUA A d. B=1" A 1 -2 1 A*B 1 0 0 -2 0 0 1 0 0;
```

No caso de CLASS ser escrito como BL B A, a matriz então terá as seguintes identificações: B1 B2 B3 A1 A2 A3 B1A1 B1A2 B1A3 B2A1 B2A2 B2A3 B3A1 B3A2 B3A3. Os contrastes então deverão ser reescritos segundo essa ordem. Para exemplificar, os dois últimos contrastes serão então escritos como:

```
CONTRAST "LIN A d. B=1" A -1 0 1 A*B -1 0 1 0 0 0 0 0 0;
CONTRAST "QUA A d. B=1" A 1 -2 1 A*B 1 -2 1 0 0 0 0 0 0;
```

No caso de variáveis qualitativas, cujos níveis sejam alfanuméricos, a ordem seguida na matriz para cada fator segue a ordem ascendente alfabética dos nomes.

Vários casos de uso do GLM são dados na maioria dos capítulos deste livro, mas alguns exemplos são:

- Blocos completos ao acaso

```
PROC GLM;CLASS BLOCO TRAT;
MODEL RESPOSTA=BLOCO TRAT;
MEANS TRAT/DUNNETT("testemunha");
```

- Blocos completos com perda de parcela

```
PROC GLM;CLASS BLOCO TRAT;
MODEL RESPOSTA=BLOCO TRAT;
LSMEANS TRAT/PDIFF;
```

- Quadrado latino com análise de covariância

```
PROC GLM;CLASS LINHA COLUNA TRAT;
MODEL RESPOSTA=LINHA COLUNA TRAT X/SOLUTION;
LSMEANS TRAT/PDIFF;
```

- Reticulado quadrado com análise intrablocos

```
PROC GLM;CLASS REP BLOCO TRAT;
MODEL RESPOSTA=REP BLOCO(REP) TRAT;
LSMEANS TRAT;
```

O procedimento LATTICE, de uso mais restrito, quando comparado com o procedimento GLM, apresenta algumas diferenças no que diz respeito ao quadro de análise de variância e à estimação das médias ajustadas. O LATTICE realiza a análise e o ajuste de médias com a recuperação da informação interblocos, enquanto o GLM faz o que se chama de análise intrablocos. Um exemplo dessa diferença foi mostrado no Capítulo 6. O procedimento LATTICE requer uma padronização dos nomes das variáveis referentes à repetição, bloco e tratamento, que devem sempre ser GROUP, BLOCK e TREATMNT. A programação para o LATTICE, nesse caso, é simplesmente:

- Reticulado quadrado com análise com recuperação da informação interblocos

```
PROC LATTICE;VAR RESPOSTA;
```

Nos exemplos de uso do GLM, a opção DUNNETT("testemunha") indica que o teste de Dunnett é o indicado, nesse caso, para testar a hipótese nula $\mathrm{H0}\colon\ t_i \neq t_0$, em que t_0 corresponde à testemunha ou tratamento padrão. A expressão "testemunha" indica qual tratamento deve ser considerado o padrão para comparação. No exemplo c, a variável X que aparece no modelo é a covariável pela qual se deseja ajustar os valores do experimento e deve constar na lista de variáveis lidas no programa. A opção SOLUTION pede a impressão das soluções do sistema de matrizes para que se possa conhecer o valor do coeficiente de regressão de X. Nos exemplos b e c, o comando LSMEANS solicita as médias de tratamentos por quadrados mínimos, que correspondem à média aritmética no caso de não haver perda de dados ou de covariável para ajuste dessas médias. Já no exemplo d, essa opção serve para ajustar as médias pelos efeitos de blocos, por se tratar de delineamento em blocos incompletos.

Procedimentos para análise de regressão

Incluem-se aqui apenas quatro dos procedimentos SAS para regressão para dados obtidos em amostragem, pois o estudo de regressão com dados de ensaios planejados deve ser feito por meio de contrastes, como visto na seção *Procedimentos para análise de variância*.

Os procedimentos são: GLM, REG, RSREG e NLIN. Os dois primeiros são utilizados no caso de modelos matemáticos lineares ou lineares por transformação, o terceiro deles é específico para modelos de superfície de resposta de segundo grau para duas ou mais variáveis independentes, e o quarto apenas para os modelos matemáticos não lineares.

O procedimento GLM, para estudos de regressão, requer basicamente os seguintes dois comandos:

PROG GLM opções; MODEL resposta=modelo/opções;

A opção mais usada em PROC GLM é DATA=nome do conjunto de dados, usada quando se tem mais de um conjunto de dados criado no processamento; portanto, a opção determina qual deles deve ser empregado na análise.

No comando MODEL, "resposta" corresponde à variável dependente, enquanto "modelo" corresponde às variáveis ou à variável independente na ou nas potências requeridas no modelo. Seguem-se alguns exemplos de modelos nos quais Y é variável resposta e $X1$, $X2$, … são as variáveis independentes.

- Linear simples ($Y = b_0 + b_1 * X1$)

  $Y = X1$;

- Linear múltiplo ($Y = b_0 + b_1 * X1 + b_2 * X2$)

  $Y = X1\ X2$;

- Linear de segundo grau ($Y = b_0 + b_1 * X1 + b_2 * X_{12}$)

  Y = $X1\ X1 * X1$;

- Exponencial ($Y = b_0 * e^{b}{}_1 * 1$)

  $LY = X1$;

Neste último caso, a variável Y deve ser antecipadamente transformada em seu logaritmo natural, ou seja, $LY = log(Y)$.

Entre as opções que podem ser usadas no estudo de regressão, estão as seguintes:

- ALPHA=porcentagem (os valores possíveis de porcentagem são os de 10%, 5% e 1%).
- CLM (limites de confiança para o valor predito médio de cada observação).
- XPX (imprime a matriz X'X).
- I (imprime a matriz inversa de X'X).
- P (imprime os valores observados, preditos e resíduos de cada observação que não tenha valor perdido).
- NOINT (omite o parâmetro do ponto de intersecção do modelo).

O procedimento REG é específico para estudos de regressão para modelos lineares aditivos diretos ou por transformação; permite a inclusão ou a seleção de variáveis independentes ou modelos por uma série de processos, além da elaboração de gráficos; tem uma sintaxe complexa, mas que pode ser simplificada em sua parte principal de elaboração:

PROG REG opções; BY variáveis;
MODEL dependentes=independentes/opções;

Entre as opções para PROG REG, devem ser destacadas: DATA="nome do conjunto de dados", que define qual dos conjuntos de dados deve ser usado no estudo de regressão; CORR, que define a impressão da matriz de correlações para todas as variáveis listadas em MODEL; OUTEST=""nome do conjunto de dados", que cria um conjunto de dados com as estimativas dos parâmetros e estatísticas opcionais. BY "variáveis" indica que, para cada nível ou categoria das variáveis listadas, deve ser realizada uma análise de regressão segundo o modelo em MODEL.

O comando MODEL permite o uso de uma ou mais variáveis dependentes para as quais se deseja o ajuste do modelo representado pelas variáveis independentes. Uma das restrições que o procedimento apresenta é que no caso de inclusão de variáveis independentes em qualquer potência diferente de 1, ou a inclusão de produtos de variáveis, essas "novas variáveis" devem ser criadas antecipadamente em algum comando DATA, seja na leitura e criação do conjunto de dados inicial seja durante a execução do programa. Isso quer dizer que o uso de X1*X1, por exemplo, tão familiar no GLM, não é permitido no REG, e deveria ser criado antes, ou seja, deve-se criar uma nova variável, por exemplo, X1Q=X1*X1. As opções que o comando MODEL possui são numerosas e, para um amplo conhecimento

delas, é recomendado a leitura do manual do SAS (SAS/STAT User's Guide). Aqui serão apresentadas apenas algumas delas como segue:

- SELECTION=nome

  Especifica o método a ser usado na seleção de variáveis a serem incluídas no modelo. Os métodos disponíveis são: FORWARD (ou F), BACKWARD (ou B), STEPWISE, MAXR, MINR, RSQUARE, ADJRSQ, CP e NONE (que indica o uso de todas as variáveis). O padrão de uso é NONE, e somente um dos métodos pode ser especificado. Para usar mais de um método, devem ser usados tantos comandos MODEL quantos forem os métodos que se deseja usar.

- SLE=valor

  Indica o nível de significância para inclusão no modelo usado nos métodos FORWARD e STEPWISE.

- SLS=valor

  Indica o nível de significância para permanência no modelo usado nos métodos BACKWARD e STEPWISE.

- XPX

  Imprime a matriz X'X para o modelo. A matriz é ladeada pelas matrizes X'Y e Y'Y.

- I

  Imprime a matriz inversa $(X'X)^{-1}$. A matriz inversa é impressa ladeada pelas matrizes de estimativa de parâmetros e da soma de quadrados de erro.

- INCLUDE=n

  Obriga a inclusão das primeiras n variáveis no modelo, segundo o método de seleção de variáveis.

- NOINT

  Elimina o ponto de intersecção no modelo, ou seja, $b_0 = 0$.

Alguns exemplos de uso do PROC REG seguem:

- Modelo linear simples ($y = b_0 + b_1X$)

  PROC REG; MODEL Y=X;

- Modelo linear múltiplo com três variáveis ($y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$)

  PROC REG; MODEL Y = X1 X2 X3;

- Modelo linear múltiplo com seis variáveis e processo de seleção STEPWISE

  PROC REG SELECTION=STEPWISE;
  MODEL Y = X1 X2 X3 X4 X5 X6;

- Modelo linear múltiplo com seis variáveis e processo de seleção STEPWISE e inclusão obrigatória de $X1$

  PROC REG SELECTION=STEPWISE INCLUDE=1;
  MODEL Y = X1 X2 X3 X4 X5 X6;

- Modelo linear múltiplo com seis variáveis e processo de seleção STEPWISE e critérios de seleção de inclusão e exclusão por significância estatística

  PROC REG SELECTION=STEPWISE SLE=0.1 SLS=0.05;
  MODEL Y = X1 X2 X3 X4 X5 X6;

- Modelo linear simples de segundo grau ($y = b_0 + b_1X + b_2X_2$)

  Neste caso, deve-se primeiro criar a variável para o quadrado de X, previamente ao estudo de regressão. Por exemplo:

DATA A; SET A; XQ = X * X;
PROC REG; MODEL Y = X XQ;

O procedimento RSREG é específico para estudos de superfície de resposta de segundo grau, para duas ou mais variáveis independentes. O resultado desse procedimento, no entanto, é muito completo, pois, além do ajuste da equação e de sua análise de variância, ele também fornece a análise canônica da superfície de resposta, com seus valores críticos, valor predito no ponto estacionário, valores e vetores próprios e a definição do tipo de ponto estacionário (máximo, mínimo ou sela). A sintaxe do procedimento é:

```
PROC RSREG opções ;BY variáveis;
MODEL resposta=modelo/opções;
```

As opções no comando PROC são apenas três: DATA=nome do conjunto de dados; NOPRINT e OUT=nome do conjunto de dados.

DATA indica em qual conjunto de dados se encontram os valores das variáveis especificadas em BY e em MODEL; NOPRINT suprime a impressão dos resultados para o caso de se desejar apenas a opção OUT, que vai especificar onde as estatísticas devem ser guardadas. BY indica que o modelo deve ser ajustado para cada nível da variável ou das variáveis ali mencionadas e MODEL "resposta" indica o nome da variável dependente e o "modelo" contém a relação das variáveis independentes. As opções usadas em MODEL são variadas, e as mais importantes entre elas são: a) LACKFIT, para o caso de as variáveis independentes possuírem valores repetidos (nesse caso, deve ser usado previamente um procedimento de ordenação das variáveis independentes, PROC SORT, para agrupar esses valores repetidos); b) PREDICT, que solicita a impressão dos valores preditos pelo modelo. Outras opções disponíveis podem ser consultadas nos manuais do SAS.

Suponha que um ensaio de fertilização do solo foi realizado com os três elementos principais, e a resposta medida foi a produtividade em grãos. Se apenas uma repetição foi feita, com os elementos em diversos níveis, e se os dados armazenados num conjunto de dados forem chamados de A, o programa para a análise de variância seria:

```
PROC RSREG DATA=A;MODEL PROD=N P K/PREDICT;
```

Caso houvesse repetição dos tratamentos, o programa seria modificado para:

```
PROC SORT;BY N P K;
PROC RSREG; MODEL PROD=N P K/LACKFIT PREDICT;
```

Exemplo de uso desse procedimento pode ser visto no Capítulo 14.

O procedimento NLIN é específico para análise de modelos de regressão não lineares, de forma que os procedimentos anteriormente descritos, GLM, REG e RSREG, não são aplicáveis. A sintaxe simplificada desse procedimento é:

```
PROC NLIN opções;
MODEL resposta=expressão do modelo;
PARAMETERS parâmetro= valor, …, parâmetro=valor;
DER.parâmetro =expressão;
DER.parâmetro.parâmetro =expressão;
```

Entre as opções do comando PROC, devem ser destacadas: DATA=nome do conjunto de dados, para indicar o conjunto de dados a serem usados; MAXITER=número, que especifica o número máximo de iterações a serem usadas pelo sistema; e METHOD=método de ajuste, no qual se indica o processo iterativo a ser usado. Os métodos disponíveis são: GAUSS, MARQUARDT, NEWTON, GRADIENT e DUD. Se METHOD não é especificado, mas as declarações DER existem, o método usado é o GAUSS. Quando as declarações DER também não são declaradas, então o sistema usa o método DUD. Quando a opção MAXITER=número não é usada, o sistema realiza 50 iterações e "número" deve ser um valor positivo e inteiro.

No comando MODEL, "resposta" indica o nome da variável dependente, e "expressão do modelo" é a escrita do modelo matemático completo.

A declaração PARAMETERS apresenta o nome dos parâmetros do modelo, "parâmetro", seguida do valor, valores, ou sequência e intervalos de ajuste dos parâmetros. As declarações DER especificam as derivadas parciais de primeira (DER.parâmetro) ou segunda ordem (DER.parâmetro.parâmetro) de cada um dos parâmetros do modelo.

No Capítulo 12, encontra-se um exemplo de uso com o modelo de MITSCHERLICH.

Procedimentos para estatística não paramétrica

No caso de estatística não paramétrica, o SAS não é muito pródigo em opções, uma vez que o único procedimento específico é o PROC NPAR1WAY. No entanto, outros procedimentos permitem também análises não paramétricas diretamente, como o PROC CORR, ou em associação com o PROC RANK, como o PROC ANOVA ou PROC GLM ou ainda o PROC MEANS.

- Correlação e regressão

  No caso da correlação, a sintaxe para a obtenção dos testes de Spearman ou de Kendall foi vista no Capítulo 15. A sintaxe básica é:

  PROC CORR SPEARMAN KENDALL;VAR variáveis;

  em que "variáveis" corresponde aos nomes das variáveis das quais se deseja conhecer a associação em pares.

Já para o estudo de regressão, deve-se lembrar de que o processo de estimação dos parâmetros pode ser considerado um processo não paramétrico; portanto, ambos os procedimentos GLM e REG podem ser usados para essa estimação. Infelizmente os testes estatísticos e os intervalos de confiança para as estimativas devem ser feitos como indicados no Capítulo 15.

- Duas amostras independentes e o delineamento inteiramente ao acaso

  Para duas amostras independentes ou para o delineamento inteiramente casualizado, o procedimento recomendado é o NPAR1WAY, com a sintaxe:

  PROC NPAR1WAY WILCOXON;
  CLASS variável; VAR variáveis:

  A declaração CLASS "variável" indica qual é a variável que identifica as duas amostras ou os tratamentos, e a declaração VAR "variáveis" identifica a variável ou as variáveis a serem analisadas. Exemplos podem ser encontrados nos Capítulos 15 e 17.

- Blocos completos ao acaso

  Para este delineamento, existe a necessidade de uso de dois procedimentos: RANK e ANOVA ou RANK e GLM. No caso do teste de Friedman, o programa é o seguinte:

  PROC SORT;BY BLOCO;
  PROC RANK OUT=NP1;BY BLOCO;VAR variável;
  PROC ANOVA DATA=NP1;CLASS BLOCO TRAT;
  MODEL variável=BLOCO TRAT;MEANS TRAT;

  em que BLOCO e TRAT são os nomes no programa SAS para identificar os blocos e os tratamentos, e "variável" é o nome da variável resposta que se quer analisar

Quando se trata do teste de QUADE, o programa já é um pouco mais elaborado, mas também aqui se usam as variáveis BLOCO e TRAT com o mesmo sentido usado no teste de Friedman. A variável PM, que aparece na linha de comando para modificação do conjunto de dados NP1, corresponde ao valor médio dos postos por bloco e é igual a $(1 + t)/2$, em que t é o número de tratamentos. O conjunto de dados NP, citado na criação do conjunto de dados NP1, é o conjunto de dados originais lido no início do programa SAS.

Exemplos de aplicação desses programas podem ser vistos no Capítulo 15 (seção *Delineamento em blocos completos casualizados*, subitem *Uso do SAS)*.

PROC SORT;BY BLOCO;
PROC MEANS SUM NOPRINT;BY BLOCO;VAR DIAS;
OUTPUT OUT=NP0 SUM=Q;

```
PROC RANK OUT=NP01;VAR Q;
DATA NP1;MERGE NP NP01;BY BLOCO;
PROC RANK OUT=NP1;BY BLOCO;VAR DIAS;
DATA NP1;SET NP1;SCORE=Q*(DIAS- PM);
PROC ANOVA DATA=NP1;CLASS BLOCO TRAT;
MODEL SCORE=BLOCO TRAT;MEANS TRAT;
```

- Blocos incompletos

  Para a obtenção do teste de Durbin, uma programação mais extensa é necessária, e se mostra a seguir, enquanto um exemplo de aplicação pode ser visto no Capítulo 15, na seção *Delineamento em blocos incompletos balanceados*. As variáveis $V1$ e $V2$ devem ser calculadas a priori e seus valores substituídos neste programa. As fórmulas são dadas também na mesma secção, e correspondem a

$$V1 = \frac{r(k+1)}{2}$$

e

$$V2 = \frac{12(t-1)}{rt(k-1)(k+1)}$$

em que t é o número de tratamentos, r o número de repetições e k o tamanho do bloco.

```
PROC SORT;BY REP BLOCO;
PROC RANK OUT=NP1;BY REP BLOCO;VAR FLOR;
PROC SORT;BY TRAT;
PROC MEANS SUM NOPRINT;BY TRAT;VAR FLOR;
OUTPUT OUT=POSTO SUM=FLOR;
DATA POSTO;SET POSTO;VAL=(FLOR-V1)**2;
PROC MEANS SUM NOPRINT;VAR VAL;
OUTPUT OUT=DURB SUM=D;
DATA DURB;SET DURB;TESTE=D*V2;
PROC PRINT;VAR TESTE;
TITLE "TESTE DE DURBIN PARA BLOCOS INCOMPLETOS";
```

- Quadrado latino

No caso do delineamento em quadrado latino, no qual existem dois testes disponíveis, um deles, o de Bennett, é de aplicação muito fácil pelo fato de usar o procedimento NPAR1WAY da mesma forma que se usou para duas amostras e para o delineamento inteiramente ao acaso, daí sua restrição de uso, pois desconsidera o desenho original do experimento.

No programa a seguir, TRAT é o nome dado a tratamentos e "variável" é o nome da variável que se quer analisar:

```
PROC NPAR1WAY WILCOXON;CLASS TRAT;
VAR "variável";
```

Já para o teste de Zimmermann, não há uma solução tão fácil e o programa para sua obtenção é o seguinte:

```
PROC SORT DATA=QLNP;BY LINHA;
PROC RANK OUT=LINHAR;BY LINHA;VAR variável;
DATA LINHAR;SET LINHAR;LR=variável;KEEP LINHA COLUNA TRAT LR;
PROC SORT;BY LINHA COLUNA TRAT;
PROC SORT DATA=QLNP;BY COLUNA;
PROC RANK OUT=COLUNAR;BY COLUNA;VAR variável;
DATA COLUNAR;SET COLUNAR;CR=variável;KEEP LINHA COLUNA TRAT CR;
PROC SORT;BY LINHA COLUNA TRAT;
DATA POSTO;MERGE LINHAR COLUNAR;BY LINHA COLUNA TRAT;
ESCORE=LR+CR;*PROC PRINT;
PROC ANOVA OUTSTAT=D NOPRINT;CLASS LINHA COLUNA TRAT;
MODEL ESCORE=LINHA COLUNA TRAT;
DATA D1;SET D;
IF _SOURCE_ EQ "TRAT";FONTE1=_SOURCE_;KEEP FONTE1 SS;
PROC MEANS DATA=D SUM NOPRINT;VAR SS;
OUTPUT OUT=D2 SUM=SS1;
DATA D2;SET D2;FONTE2="TOTAL";
DATA FIM;MERGE D1 D2;
T=SS/SS1;Z=((n*T)/(1-T))**((n-1)/(n+1));
PROC PRINT;VAR FONTE1 SS FONTE2 SS1 T Z;
```

Nesse programa, QLNP refere-se ao conjunto de dados originais inserido no programa SAS; "variável" corresponde ao nome da variável resposta que se quer analisar; LINHA, COLUNA e TRAT correspondem aos fatores de classificação de um quadrado latino; "n" é o

número de tratamentos ou o tamanho do quadrado do latino e seu valor numérico deve ser aí introduzido.

Exemplos dos dois testes encontram-se no Capítulo 15, na seção *Delineamento quadrado latino*.

Referências

AMARAL, A. M.; MUNIZ, J. A.; SOUZA, M. Avaliação do coeficiente de variação como medida de precisão na experimentação com citros. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 32, n. 12, p. 1221-1225, 1997.

BELL, C. B.; DOKSUM, K. A. Some new distribution-free statistics. **Annals of Mathematical Statistical**, Ann Arbor, v. 36, n. 203-214, 1965.

BENNETT, B. M. Use and estimating efficiency of rank-order tests in randomized block, Latin square, and split-plot design. **Biometrie Praximetrie**, Gembloux, v. 7, p. 79-87, 1966.

BERK, K. N.; CAREY, P. **Análisis de datos con Microsoft® Excel**. México, DF: Thomson Learning, 2001.

BOSE, R. C.; NAIR, K. R. Partially balanced incomplete block designs. **Sankhyã**, Calcutta, v. 4, p. 337-372, 1939.

CAMPOS, H. **Estatística experimental não-paramétrica**. Piracicaba: Esalq, 1983.

COCHRAN, W. G. **Sampling Techniques**. New York: J. Willey, 1977.

COCHRAN, W. G.; COX, G. M. **Diseños experimentales**. México: Trillas, 1983. 661 p.

CONAGIN, A.; JORGE, J. de P. N. Delineamento (1/5) (5 x 5 x 5) em blocos. **Bragantia**, Campinas, v. 36, p. 23-58, 1976.

CONOVER, W. J. **Practical nonparametric statistics**. New York: J. Wiley, 1980.

CONOVER, W. J. Rank tests for one sample, two samples and k samples without the assumption of a continuous distribution function. **The Annals of Statistics**, Hayward, v. 1, n. 6, p. 1105-1125, 1973.

CONOVER, W. J.; IMAN, R. L. **On multiple comparison procedures**. Los Alamos: Los Alamos Scientific Laboratory, 1979. (Technical Report, LA-7677-MS).

CONOVER, W. J.; IMAN, R. L. On some alternative procedures using ranks for the analysis of experimental designs. **Communications in Statistics**, Parte A, Theory and Methods, New York, v. 5, p. 1349-1368, 1976.

COSTA, N. H. A. D.; SERAPHIN, J. C.; ZIMMERMANN, F. J. P. Novo método de classificação de coeficientes de variação para a cultura do arroz de terras altas. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 37, n. 3, p. 243-249, 2002.

CROUSE, C. F. A distribution-free method of analyzing a 2m factorial experiment. **South African Statistical Journal**, Cape Town, v. 2, p. 101-108, 1968.

CROUSE, C. F. A class of distribution-free analysis of variance tests. **South African Statistical Journal**, Cape Town, v. 1, p. 75-80, 1967.

DE-LA-LOMA, J. L. **Experimentación agrícola**. México, DF: Uteha, 1966. 493 p.

DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: J. Wiley, 1998. 706 p.

DUNCAN, D. B. Multiple range and multiple F tests. **Biometrics**, Washington, DC, v. 11, n. 1, p. 1-42, 1955.

DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. **Journal of the American Statistical Association**, New York, v. 50, n. 272, p. 1096-1121, 1955. Reimpresso.

DUNNETT, C. W. New tables for multiple comparison with a control. **Biometrics**, Washington, DC, v. 20, n. 3, p. 482-491, 1964

DURBIN, J. Incomplete blocks in ranking experiments. **British Journal of Statistical Psychology**, London, UK, v. 4, n. 2, p. 85-90, 1951.

ESTEFANEL, V.; PIGNATARO, I. A. B.; STORCK, L. Avaliação do coeficiente de variação de experimentos com algumas culturas agrícolas. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRÍCOLA, 2., 1987, Londrina. **Anais...** Londrina: UEL: RBRAS. 1987. p. 115-131.

FAGÉRIA, N. K.; BARBOSA FILHO, M. P.; ZIMMERMANN, F. J. P. Caracterização química e granulométrica de solos de várzea de alguns estados brasileiros. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 29, n. 2, p. 267-274, 1994.

FEDERER, W. T. A method for evaluating genetic progress in a sugar cane breeding program. **Hawaiian Planters Record**. Honolulu, v. 55, n. 2, p. 177-189, 1956a.

FEDERER, W. T. Augmented (or Hoonuiaku) designs. **Hawaiian Planters Record**, Honolulu, v. 55, n. 2, p. 191-208, 1956b.

FEDERER, W. T. **Experimental design**: theory and application. New York: Macmillan, 1955. 584 p.

FISHER, R. A.; YATES, F. **Statistical tables for biological, agricultural, and medical research**. 5. ed. Edinburgh: Oliver â Boyd, 1957.

FRIEDMAN, W. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. **Journal of the American Statistical Association**, New York, v. 32, n. 200, p. 675-701, 1937.

GABRIEL, K. R.; LACHENBRUCH, P. A. Non-parametric ANOVA in small samples: A Monte Carlo study of the adequacy of the asymptotic approximation. **Biometrics**, Washington, DC, v. 25, n. 3, p. 593-596, 1969.

GARCIA, C. H. **Tabelas para classificação do coeficiente de variação**. Piracicaba: IPEF, 1989. 12 p. (Circular técnica, 171).

GOMEZ, K. A.; GOMEZ, A. A. **Statistical procedures for agricultural research**. New York: J. Wiley, 1976. 680 p.

HAMILTON, B. L. A Monte Carlo test of the robustness of parametric and nonparametric analysis of covariance against unequal regression slopes. **Journal of the American Statistical Association**, New York, v. 71, n. 356, p. 864-869, 1976.

HARRIS, M.; HORVITZ, D. G.; MOOD, A. M. On the determination of sample sizes in designing experiments. **Journal of the American Statistical Association**, New York, v. 43, n. 243, p. 391-402, 1948.

HARSHBARGER, B. Triple rectangular lattices. **Biometrics**, Washington, DC, v. 5, n. 1, p. 1-13, 1949.

HARTER, H. L. Error rates and sample sizes for range tests in multiple comparisons. **Biometrics**, Washington, DC, v. 13, n. 4, p. 511-536, 1957.

HOTELLING, H.; PABST, M. R. Rank correlation and tests of significance involving no assumption of normality. **Annals of Mathematical Statatistics**, Ann Arbor, v. 7, n. 1, p. 29-43, 1936.

IBGE. Banco de Dados Agregados. **Produção agrícola municipal**: feijão: Santa Catarina: 2008 e 2009. Disponível em: <http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=1002&z=p&o=26>. Acesso em: 21 jun. 2011.

IMAN, R. L. A power study of a rank transform for the two-way classification model when interaction may be present. **Canadian Journal of Statistics**, Montreal, 2, n. 1/2, p. 227-39, 1974.

IMAN, R. L.; DAVENPORT, J. J. Approximations of the critical region of the Friedman statistic. **Communications in Statistics**, Parte A, Theory and Methods, New York, v. 9, n. 6, 571-595, 1980.

IMAN, R. L.; QUADE, D.; ALEXANDER, D. A. Exact probability levels for the Kruskal-Wallis test. In: HARTER, H. L.; OWEN, D. B. (Ed.). **Selected Tables in Mathematical Statistics**. Providence: American Mathematical Society, 1975. v. 3, p. 329-384.

KALCKMANN, R. E. **Técnica de experimentação agrícola**. Rio de Janeiro: Ministério da Agricultura, 1958. 166 p.

KEMPTHORNE, O. **The design and analysis of experiments**. Huntington: Robert E. Krieger, 1975. 631 p.

KENDALL, M. G. A new measure of rank correlation. **Biometrika**, London, UK, v. 30, n. 1/2, p. 81-93, 1938.

KRUSKAL, W. H. A nonparametric test for the several sample problem. **Annals of Mathematical Statistics**, Ann Arbor, v. 23, n. 4, p. 525-540, 1952.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks on one-criterion variance analysis. **Journal of the American Statistical Association**, New York, v. 47, n. 260, p. 583-621, 1952. Correções publicadas no v. 48, p. 907-911, 1953.

KUEHL, R. O. **Diseño de experimentos**: principios estadísticos para el diseño y análisis de investigaciones. 2nd. ed. México, DF: Thomson, 2001. 666 p.

LAWLER, K. R. **A distribution-free generalization of the Wilcoxon signed-ranks tests for the two-way layout**. 1978. 29 p. Thesis (Master of Science) – Texas Tech University, Lubbock.

LEHMANN, E. L. **Nonparametrics**: statistical methods based on ranks. San Francisco: Holden-Day, 1975.

LEMMER, H. H.; STOKER, D. J. A distribution-free analysis of variance for the two-way classification. **South African Statistical Journal**, Cape Town, v. 1, p. 67-74, 1967.

MACDONALD, P. The analysis of a 2n experiment by means of ranks. **Journal of the Royal Statistical Society**, Series C, Applied Statistics, London, UK, v. 20, n. 3, p. 259-75, 1971.

MALHEIROS, E. B.; PERECIN, D. Posição das doses de nutrientes na análise de uma superfície de resposta à adubação. **Revista de Matemática e Estatística**, Marília, v. 1, p. 69-78, 1983.

MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables in stochastically larger than the other. **Annals of Mathematical Statistics**, Ann Arbor, v. 18, n. 1, p. 50-60, 1947.

MARTINEZ GARZA, A. **Diseño y analisis de experimentos com caña de azucar**. México, DF: Talleres Gráficos de la nación, 1972. 204 p.

MONTGOMERY, D. C. **Design and analysis of experiments**. New York: J. Wiley, 1991. 418 p.

NASCIMENTO JÚNIOR, A. do; ALMEIDA, V. M. Desempenho dos ensaios regionais de feijão em nível de região II do Brasil no estado do Mato Grosso. In: REUNIÃO DA COMISSÃO TÉCNICA REGIONAL DE FEIJÃO - REGIÃO II, 8., 1994, Goiânia. **Relatório técnico**: resultados do período 1992-94. Goiânia: Embrapa-CNPAF, 1996. 134 p. (Embrapa-CNPAF. Documentos, 72).

NOETHER, G. E. **Elements of nonparametric statistics**. New York: J. Wiley, 1967a.

NOETHER, G. E. Wilcoxon confidence intervals for location parameters in the discrete case. **Journal of the American Statistical Association**, New York, v. 62, n. 317, p. 84-108, 1967b.

PAVATE, M. V. Combined analysis of balanced incomplete block designs with some common treatments. **Biometrics**, Washington, DC, v. 17, n. 1, p. 111-119, 1961.

PEARSON, E. S.; *HARTLEY,* H. O. **Biometrika tables for statisticians**. 3rd. ed. Cambridge: University Press, 1966. v. 1.

PIMENTEL GOMES, F. An extension of the method of joint analysis of experiments in complete randomized blocks. **Biometrics**, Washington, v. 26, n. 2, p. 332-336, 1970.

PIMENTEL GOMES, F.: GUIMARÃES, R. F. Joint analysis of experiments in complete randomised blocks with some common treatments. **Biometrics**, Washington, DC, v. 14, n. 4, p. 521-526, 1958.

PIMENTEL-GOMES, F. **Curso de Estatística experimental**. 14. ed. Piracicaba: Edição do autor, 2000. 477 p.

PITMAN, E. J. G. **Notes on non-parametric statistical inference**. New York: Columbia University, 1949. (Mimeografado).

QUADE, D. **Analyzing randomized blocks by weighted ranking**. Amsterdam, NL: Stichting Mathematisch Centrum, 1972. (Technical Report, SW, 18/72).

QUADE, D. Using weighted rankings in the analysis of complete blocks with additive block effects. **Journal of the American Statistical Association**, New York, v. 74, n. 367, p. 680-683, 1979.

RAND CORPORATION. **A million random digits with 100,000 normal deviates**. Glencoe: Free Press, 1955.

RANDLES, R. H.; WOLFE, D. A. **Introduction do the theory of non-parametric statistics**. New York: J. Wiley, 1979.

ROBSON, D. S. A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. **Biometrics**, Washington, DC, v. 15, n. 2, p. 187-191, 1959.

SACHS, L. **Applied statistics**: a handbook of techniques. New York: Springer-Verlag, 1982.

SANCHES, S. F. **Análise estatística de um experimento em quadrado latino com perda de parcelas, linha, coluna ou tratamento**. 1978. 100 f. Dissertação (Mestrado) –Universidade de São Paulo, Piracicaba.

SAS. **SAS Products**. Disponível em: <http://support.sas.com/software/products/stat/index.html#documentation>. Acesso em: 21 jun. 2011.

SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. **Biometrics Bulletin**, Washington, DC, v. 2, n. 6, p. 110-114, 1946.

SCAPIM, C. A.; CARVALHO, C. G. P.; CRUZ, C. D. Uma proposta de classificação dos coeficientes de variação para a cultura do milho. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 30, n. 5, p. 683-686, 1995.

SCHEFFÉ, H. **The analysis of variance**. New York: J. Wiley, 1959. 461 p.

SCHEIRER, C. J.; RAY, W. S.; HARE, N. The analysis of ranked data derived from completely randomized factorial designs. **Biometrics**, Washington, DC, v. 32, n. 2, p. 429-434, 1976.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, Washington, DC, v. 30, n. 3, p. 507-512, 1974

SHIRLEY, E. A non-parametric equivalent of William's test for contrasting increasing dose levels of a treatment. **Biometrics**, Washington, DC, v. 33, n. 2, p. 386-389, 1977.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não-linear**. Brasília, DF: Embrapa, SEA, *1998.* 489 p.

SPEARMAN, C. The proof and measurement of association between two things. **American Journal Psychology**, Champaign, 15, n. 1, p. 72-101, 1904.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics, with special reference to the biological sciences**. New York: Mc-Graw Hill, 1981.

STEVENS, S. S. On the theory of scales of measurement. **Science**, Washington, DC, v. 103, n. 2684, p. 677-680, 1946.

STUART, A. Asymptotic relative efficiency of tests and the derivatives of their power functions. **Skandinavisk Aktaurietidskrift**, Stockholm, SE, n. 3-4, p. 163-169, 1954.

STUART, A. The efficiencies of test of randomness against normal regression. **Journal of the American Statistical Association**, New York, v. 51, n. 274, p. 285-287, 1956.

TUKEY, J. W. Comparing individual means in the analysis of variance. **Biometrics**, Washington, v. 5, n. 2, p. 99-114, 1949a.

TUKEY, J. W. One degree of freedom for non-additivity. **Biometrics**, Washington, DC, v. 5, n. 3, p. 232-242, 1949b.

TUKEY, J. W. Reply to Query 113. **Biometrics**, Washington, DC, v. 11, p. 111-113, 1955.

WEISBERG, S. **Applied linear regression**. New York : J. Wiley, *1985*.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics**, Washington, DC, v. 1, n. 6, p. 80-83, 1945.

WILCOXON, F. **Some rapid approximate statistical procedures**. Stamford: American Cyanamid: Stamford Research Laboratories, 1949.

WORSLEY, K. J. A nonparametric extension of a cluster analysis method by Scott and Knott. **Biometrics**, Washington, DC, v. 33, n. 3, p. 532-535, 1977.

YATES, F. A new method of arranging variety trials involving a large number of varieties. **Journal of Agricultural Science**, Cambridge, v. 26, n. 3, p. 424-455. 1936b.

YATES, F. **Design and analysis of factorial experiments**. London: Imperial Bureau of Soil Sciences, 1937. (Imperial Bureau of Soil Sciences. Technical Communication, 35).

YATES, F. Incomplete Latin squares. **Journal of Agricultural Science**, Cambridge, v. 26, n. 2, p. 301-315, 1936a.

YATES, F. The recovery of interblock information in balanced incomplete block designs. **Annals of Eugenics**, London, v. 10, p. 317-325, 1940.

YATES, F.; HALE, R. W. The analysis of Latin squares when two or more rows, columns or treatments are missing. **Journal of the Royal Statistical Society**, London, v. 6, p. 67-79, 1939.

YOUDEN, W. J. Experimental designs to increase accuracy of greenhouse studies. **Contributions from the Boyce Thompson Institute**, New York, v. 11, p. 219-228, 1940.

ZIMMERMANN, F. J. P. **Análise de um experimento de adubação e calagem em milho, executado num delineamento em faixa durante oito anos no mesmo local**. 1973. Dissertação (Mestrado) – Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, SP.

ZIMMERMANN, F. J. P. **Distribution-free tests in the Latin square design**. 1983. Tese (Doutorado) – University of California, Riverside.

ZIMMERMANN, F. J. P. Efeito da heterogeneidade de variância e distribuição dos dados sobre o poder e tamanho do teste F. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 22, n. 11/12, p. 1209-1213. 1987.

ZIMMERMANN, F. J. P. Estadística aplicada a la selección recurrente. In: GUIMARÃES, E. P. (Ed). **Selección recurrente en arroz**. Cali: CIAT, 1997. 240 p.

ZIMMERMANN, F. J. P.; ROBERTSON, C. A. A new distribution-free test for the Latin square design. **Revista de Matemática e Estatística**, São Paulo, v. 8, p. 57-67, 1990.

ZIMMERMANN, F. J. Z.; CONAGIN, A. Ajuste de modelos polinomiais de 2º grau em pesquisas com fertilizantes. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 21, n. 9, p. 971-978, 1986.

Anexo 1. Teste de χ^2.

Graus de liberdade	Níveis de probabilidade									
	0,50	0,25	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	0,45	1,32	1,64	2,71	3,84	5,02	6,63	7,88	10,83	12,12
2	1,39	2,77	3,22	4,61	5,99	7,38	9,21	10,60	13,82	15,20
3	2,37	4,11	4,64	6,25	7,81	9,35	11,34	12,84	16,27	17,73
4	3,36	5,39	5,99	7,78	9,49	11,14	13,28	14,86	18,47	20,00
5	4,35	6,63	7,29	9,24	11,07	12,83	15,09	16,75	20,51	22,11
6	5,35	7,84	8,56	10,64	12,59	14,45	16,81	18,55	22,46	24,10
7	6,35	9,04	9,80	12,02	14,07	16,01	18,48	20,28	24,32	26,02
8	7,34	10,22	11,03	13,36	15,51	17,53	20,09	21,95	26,12	27,87
9	8,34	11,39	12,24	14,68	16,92	19,02	21,67	23,59	27,88	29,67
10	9,34	12,55	13,44	15,99	18,31	20,48	23,21	25,19	29,59	31,42
11	10,34	13,70	14,63	17,28	19,68	21,92	24,73	26,76	31,26	33,14
12	11,34	14,85	15,81	18,55	21,03	23,34	26,22	28,30	32,91	34,82
13	12,34	15,98	16,98	19,81	22,36	24,74	27,69	29,82	34,53	36,48
14	13,34	17,12	18,15	21,06	23,68	26,12	29,14	31,32	36,12	38,11
15	14,34	18,25	19,31	22,31	25,00	27,49	30,58	32,80	37,70	39,72
16	15,34	19,37	20,47	23,54	26,30	28,85	32,00	34,27	39,25	41,31
17	16,34	20,49	21,61	24,77	27,59	30,19	33,41	35,72	40,79	42,88
18	17,34	21,60	22,76	25,99	28,87	31,53	34,81	37,16	42,31	44,43
19	18,34	22,72	23,90	27,20	30,14	32,85	36,19	38,58	43,82	45,97
20	19,34	23,83	25,04	28,41	31,41	34,17	37,57	40,00	45,31	47,50
21	20,34	24,93	26,17	29,62	32,67	35,48	38,93	41,40	46,80	49,01
22	21,34	26,04	27,30	30,81	33,92	36,78	40,29	42,80	48,27	50,51
23	22,34	27,14	28,43	32,01	35,17	38,08	41,64	44,18	49,73	52,00
24	23,34	28,24	29,55	33,20	36,42	39,36	42,98	45,56	51,18	53,48
25	24,34	29,34	30,68	34,38	37,65	40,65	44,31	46,93	52,62	54,95
26	25,34	30,43	31,79	35,56	38,89	41,92	45,64	48,29	54,05	56,41
27	26,34	31,53	32,91	36,74	40,11	43,19	46,96	49,65	55,48	57,86
28	27,34	32,62	34,03	37,92	41,34	44,46	48,28	50,99	56,89	59,30
29	28,34	33,71	35,14	39,09	42,56	45,72	49,59	52,34	58,30	60,73
30	29,34	34,80	36,25	40,26	43,77	46,98	50,89	53,67	59,70	62,16
32	31,34	36,97	38,47	42,58	46,19	49,48	53,49	56,33	62,49	64,99
34	33,34	39,14	40,68	44,90	48,60	51,97	56,06	58,96	65,25	67,80
36	35,34	41,30	42,88	47,21	51,00	54,44	58,62	61,58	67,98	70,59
38	37,34	43,46	45,08	49,51	53,38	56,90	61,16	64,18	70,70	73,35

Continua...

Anexo 1. Continuação.

Graus de liberdade	Níveis de probabilidade									
	0,50	0,25	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
40	39,34	45,62	47,27	51,81	55,76	59,34	63,69	66,77	73,40	76,10
44	43,34	49,91	51,64	56,37	60,48	64,20	68,71	71,89	78,75	81,53
48	47,34	54,20	55,99	60,91	65,17	69,02	73,68	76,97	84,04	86,90
50	49,33	56,33	58,16	63,17	67,50	71,42	76,15	79,49	86,66	89,56
55	54,33	61,67	63,58	68,80	73,31	77,38	82,29	85,75	93,17	96,16
60	59,33	66,98	68,97	74,40	79,08	83,30	88,38	91,95	99,61	102,70
65	64,33	72,28	74,35	79,97	84,82	89,18	94,42	98,10	105,99	109,16
70	69,33	77,58	79,71	85,53	90,53	95,02	100,43	104,21	112,32	115,58
80	79,33	88,13	90,41	96,58	101,88	106,63	112,33	116,32	124,84	128,26
90	89,33	98,65	101,05	107,57	113,15	118,14	124,12	128,30	137,21	140,78
100	99,33	109,14	111,67	118,50	124,34	129,56	135,81	140,17	149,45	153,16
200	199,33	213,10	216,61	226,02	233,99	241,06	249,45	255,26	267,54	272,42
300	299,33	316,14	320,40	331,79	341,40	349,87	359,91	366,84	381,42	387,20
400	399,33	418,70	423,59	436,65	447,63	457,31	468,72	476,61	493,13	499,67
500	499,33	520,95	526,40	540,93	553,13	563,85	576,49	585,21	603,45	610,64

Nota: valores calculados com o uso da fórmula INVC de Excel© de Microsoft©.

Anexo 2. Teste t de Student.

Graus de liberdade	Níveis de probabilidade									
	0,50	0,25	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	1,00	2,41	3,08	6,31	12,71	25,45	63,66	127,32	636,58	1.273,16
2	0,82	1,60	1,89	2,92	4,30	6,21	9,92	14,09	31,60	44,70
3	0,76	1,42	1,64	2,35	3,18	4,18	5,84	7,45	12,92	16,33
4	0,74	1,34	1,53	2,13	2,78	3,50	4,60	5,60	8,61	10,31
5	0,73	1,30	1,48	2,02	2,57	3,16	4,03	4,77	6,87	7,98
6	0,72	1,27	1,44	1,94	2,45	2,97	3,71	4,32	5,96	6,79
7	0,71	1,25	1,41	1,89	2,36	2,84	3,50	4,03	5,41	6,08
8	0,71	1,24	1,40	1,86	2,31	2,75	3,36	3,83	5,04	5,62
9	0,70	1,23	1,38	1,83	2,26	2,69	3,25	3,69	4,78	5,29
10	0,70	1,22	1,37	1,81	2,23	2,63	3,17	3,58	4,59	5,05
11	0,70	1,21	1,36	1,80	2,20	2,59	3,11	3,50	4,44	4,86
12	0,70	1,21	1,36	1,78	2,18	2,56	3,05	3,43	4,32	4,72
13	0,69	1,20	1,35	1,77	2,16	2,53	3,01	3,37	4,22	4,60
14	0,69	1,20	1,35	1,76	2,14	2,51	2,98	3,33	4,14	4,50
15	0,69	1,20	1,34	1,75	2,13	2,49	2,95	3,29	4,07	4,42
16	0,69	1,19	1,34	1,75	2,12	2,47	2,92	3,25	4,01	4,35
17	0,69	1,19	1,33	1,74	2,11	2,46	2,90	3,22	3,97	4,29
18	0,69	1,19	1,33	1,73	2,10	2,45	2,88	3,20	3,92	4,23
19	0,69	1,19	1,33	1,73	2,09	2,43	2,86	3,17	3,88	4,19
20	0,69	1,18	1,33	1,72	2,09	2,42	2,85	3,15	3,85	4,15
21	0,69	1,18	1,32	1,72	2,08	2,41	2,83	3,14	3,82	4,11
22	0,69	1,18	1,32	1,72	2,07	2,41	2,82	3,12	3,79	4,08
23	0,69	1,18	1,32	1,71	2,07	2,40	2,81	3,10	3,77	4,05
24	0,68	1,18	1,32	1,71	2,06	2,39	2,80	3,09	3,75	4,02
25	0,68	1,18	1,32	1,71	2,06	2,38	2,79	3,08	3,73	4,00
26	0,68	1,18	1,31	1,71	2,06	2,38	2,78	3,07	3,71	3,97
27	0,68	1,18	1,31	1,70	2,05	2,37	2,77	3,06	3,69	3,95
28	0,68	1,17	1,31	1,70	2,05	2,37	2,76	3,05	3,67	3,93
29	0,68	1,17	1,31	1,70	2,05	2,36	2,76	3,04	3,66	3,92
30	0,68	1,17	1,31	1,70	2,04	2,36	2,75	3,03	3,65	3,90
32	0,68	1,17	1,31	1,69	2,04	2,35	2,74	3,01	3,62	3,87
34	0,68	1,17	1,31	1,69	2,03	2,35	2,73	3,00	3,60	3,85
36	0,68	1,17	1,31	1,69	2,03	2,34	2,72	2,99	3,58	3,83
38	0,68	1,17	1,30	1,69	2,02	2,33	2,71	2,98	3,57	3,81

Continua...

Anexo 2. Continuação.

Graus de liberdade	Níveis de probabilidade									
	0,50	0,25	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
40	0,68	1,17	1,30	1,68	2,02	2,33	2,70	2,97	3,55	3,79
44	0,68	1,17	1,30	1,68	2,02	2,32	2,69	2,96	3,53	3,76
48	0,68	1,16	1,30	1,68	2,01	2,31	2,68	2,94	3,50	3,73
50	0,68	1,16	1,30	1,68	2,01	2,31	2,68	2,94	3,50	3,72
55	0,68	1,16	1,30	1,67	2,00	2,30	2,67	2,92	3,48	3,70
60	0,68	1,16	1,30	1,67	2,00	2,30	2,66	2,91	3,46	3,68
65	0,68	1,16	1,29	1,67	2,00	2,29	2,65	2,91	3,45	3,66
70	0,68	1,16	1,29	1,67	1,99	2,29	2,65	2,90	3,43	3,65
80	0,68	1,16	1,29	1,66	1,99	2,28	2,64	2,89	3,42	3,63
90	0,68	1,16	1,29	1,66	1,99	2,28	2,63	2,88	3,40	3,61
100	0,68	1,16	1,29	1,66	1,98	2,28	2,63	2,87	3,39	3,60
200	0,68	1,15	1,29	1,65	1,97	2,26	2,60	2,84	3,34	3,54
300	0,68	1,15	1,28	1,65	1,97	2,25	2,59	2,83	3,32	3,52
400	0,68	1,15	1,28	1,65	1,97	2,25	2,59	2,82	3,32	3,51
500	0,67	1,15	1,28	1,65	1,96	2,25	2,59	2,82	3,31	3,50

Nota: valores calculados com o uso da fórmula INVC de Excel© de Microsoft©.

Anexo 3. Teste F. Nível de probabilidade = 0,05.

Graus de liberdade do denominador	Níveis de probabilidade																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246	246	247	247	248	248
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43	19,43	19,44	19,44	19,44	19,45
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65
12	3,26	3,49	3,49	3,49	3,49	3,49	3,49	3,89	3,89	3,89	3,89	3,89	3,89	3,89	3,89	3,89	3,89	3,89	3,89	3,89
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15	2,13	2,11	2,10	2,08	2,07
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11	2,09	2,07	2,05	2,04	2,03

Continua...

Anexo 3. Continuação.

Graus de liberdade do denominador	Níveis de probabilidade																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07	2,05	2,03	2,02	2,00	1,99
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04	2,02	2,00	1,99	1,97	1,96
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01	1,99	1,98	1,96	1,95	1,93
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97	1,95	1,93	1,92	1,90	1,89
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,99	1,96	1,94	1,92	1,90	1,88	1,87	1,85
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,89	1,87	1,85	1,84
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,95	1,92	1,90	1,88	1,86	1,84	1,83	1,81
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96	1,93	1,90	1,88	1,86	1,84	1,82	1,81	1,79
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87	1,85	1,83	1,81	1,80	1,78
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93	1,90	1,88	1,85	1,83	1,81	1,79	1,78	1,76
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,82	1,80	1,78	1,76	1,75
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81	1,79	1,77	1,75	1,74	1,72
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79	1,77	1,75	1,73	1,72	1,70
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77	1,75	1,73	1,71	1,69	1,68
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,72	1,69	1,67	1,66	1,64	1,62
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77	1,74	1,71	1,69	1,66	1,64	1,62	1,61	1,59

Nota: valores calculados com o uso da fórmula INVC de Excel© de Microsoft©.

Anexo 4. Teste de F. Nível de probabilidade = 0,01.

Graus de liberdade do denominador	Graus de liberdade do numerador																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107	6126	6143	6157	6170	6181	6191	6201	6209
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,41	99,42	99,42	99,43	99,43	99,44	99,44	99,44	99,45	99,45
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05	26,98	26,92	26,87	26,83	26,79	26,75	26,72	26,69
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37	14,31	14,25	14,20	14,15	14,11	14,08	14,05	14,02
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07	3,02	2,98	2,94	2,91	2,88	2,85	2,83
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89	2,85	2,82	2,79	2,76	2,74
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90	2,86	2,81	2,78	2,75	2,72	2,69	2,66

Continua...

Anexo 4. Continuação.

	Graus de liberdade do numerador																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84	2,79	2,75	2,72	2,68	2,65	2,63	2,60
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79	2,74	2,70	2,66	2,63	2,60	2,57	2,55
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,80	2,74	2,69	2,64	2,60	2,56	2,53	2,50	2,47	2,44
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52	2,48	2,45	2,42	2,39	2,37
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62	2,56	2,52	2,47	2,44	2,40	2,37	2,35	2,32
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,71	2,64	2,58	2,53	2,48	2,44	2,40	2,37	2,33	2,31	2,28
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56	2,51	2,46	2,42	2,38	2,35	2,32	2,29	2,27
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53	2,47	2,42	2,38	2,34	2,31	2,28	2,25	2,23
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44	2,39	2,35	2,31	2,28	2,25	2,22	2,20
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40	2,35	2,31	2,27	2,23	2,20	2,18	2,15
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36	2,31	2,27	2,23	2,20	2,17	2,14	2,12
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31	2,27	2,22	2,19	2,15	2,12	2,09	2,07
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31	2,25	2,20	2,16	2,12	2,09	2,06	2,03	2,00
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22	2,17	2,13	2,09	2,06	2,03	2,00	1,97
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22	2,17	2,12	2,07	2,04	2,00	1,97	1,94	1,92

Nota: valores calculados com o uso da fórmula INVC de Excel© de Microsoft©.

Anexo 5. Comparações múltiplas com uma testemunha (teste de Dunnett) – α = 0,05 (teste unilateral).

Graus de liberdade	Número de tratamentos a serem comparados excluindo a testemunha												
	2	3	4	5	6	7	8	9	10	11	12	15	20
5	2,44	2,68	2,85	2,98	3,08	3,16	3,24	3,30	3,36	3,41	3,45	3,57	3,72
6	2,34	2,56	2,71	2,83	2,92	3,00	3,07	3,12	3,17	3,22	3,26	3,37	3,50
7	2,27	2,48	2,62	2,73	2,82	2,89	2,95	3,01	3,05	3,10	3,13	3,23	3,36
8	2,22	2,42	2,55	2,66	2,74	2,81	2,87	2,92	2,96	3,01	3,04	3,14	3,25
9	2,1,8	2,37	2,50	2,60	2,68	2,75	2,81	2,86	2,90	2,94	2,97	3,06	3,18
10	2,15	2,34	2,47	2,56	2,64	2,70	2,76	2,81	2,85	2,89	2,92	3,01	3,12
11	2,13	2,31	2,44	2,53	2,60	2,67	2,72	2,77	2,81	2,85	2,88	2,96	3,07
12	2,11	2,29	2,41	2,50	2,58	2,64	2,69	2,74	2,78	2,81	2,84	2,93	3,03
13	2,09	2,27	2,39	2,48	2,55	2,61	2,66	2,71	2,75	2,78	2,82	2,90	3,00
14	2,08	2,25	2,37	2,46	2,53	2,59	2,64	2,69	2,72	2,76	2,79	2,87	2,97
15	2,07	2,24	2,36	2,44	2,51	2,57	2,62	2,67	2,70	2,74	2,77	2,85	2,95
16	2,06	2,23	2,34	2,43	2,50	2,56	2,61	2,65	2,69	2,72	2,75	2,83	2,93
17	2,05	2,22	2,33	2,42	2,49	2,54	2,59	2,64	2,67	2,71	2,74	2,81	2,91
18	2,04	2,21	2,32	2,41	2,48	2,53	2,58	2,62	2,66	2,69	2,72	2,80	2,89
19	2,03	2,20	2,31	2,40	2,47	2,52	2,57	2,61	2,65	2,68	2,71	2,79	2,88
20	2,03	2,19	2,30	2,39	2,46	2,51	2,56	2,60	2,64	2,67	2,70	2,77	2,87
24	2,01	2,17	2,28	2,36	2,43	2,48	2,53	2,57	2,60	2,64	2,66	2,74	2,83
30	1,99	2,15	2,25	2,33	2,40	2,45	2,50	2,54	2,57	2,60	2,63	2,70	2,79
40	1,97	2,13	2,23	2,31	2,37	2,42	2,47	2,51	2,54	2,57	2,60	2,67	2,75
60	1,95	2,10	2,21	2,28	2,35	2,39	2,44	2,48	2,51	2,54	2,56	2,63	2,72
120	1,93	2,08	2,18	2,26	2,32	2,37	2,41	2,45	2,48	2,51	2,53	2,60	2,68
∞	1,92	2,06	2,16	2,23	2,29	2,34	2,38	2,42	2,45	2,48	2,50	2,56	2,64

Fonte: C. W. Dunnett (1955, 1964) e tabelas adicionais produzidas por C. W. Dunnett (1980).

Anexo 6. Comparações múltiplas com uma testemunha (teste de Dunnett) − α = 0,01 (unilateral).

Graus de liberdade	Número de tratamentos excluindo a testemunha												
	2	3	4	5	6	7	8	9	10	11	12	15	20
5	3,90	4,21	4,43	4,60	4,73	4,85	4,94	5,03	5,11	5,17	5,24	5,39	5,59
6	3,61	3,88	4,07	4,21	4,33	4,43	4,51	4,59	4,64	4,70	4,76	4,89	5,06
7	3,42	3,66	3,83	3,96	4,07	4,15	4,23	4,30	4,35	4,40	4,45	4,57	4,72
8	3,29	3,51	3,67	3,79	3,88	3,96	4,03	4,09	4,14	4,19	4,23	4,34	4,48
9	3,19	3,40	3,55	3,66	3,75	3,82	3,89	3,94	3,99	4,04	4,08	4,18	4,31
10	3,11	3,31	3,45	3,56	3,64	3,71	3,78	3,83	3,88	3,92	3,96	4,06	4,18
11	3,06	3,25	3,38	3,48	3,56	3,63	3,69	3,74	3,79	3,83	3,86	3,96	4,08
12	3,01	3,19	3,32	3,42	3,50	3,56	3,62	3,67	3,71	3,75	3,79	3,88	3,99
13	2,97	3,15	3,27	3,37	3,44	3,51	3,56	3,61	3,65	3,69	3,73	3,81	3,92
14	2,94	3,11	3,23	3,32	3,40	3,46	3,51	3,56	3,60	3,64	3,67	3,76	3,87
15	2,91	3,08	3,20	3,29	3,36	3,42	3,47	3,52	3,56	3,60	3,63	3,71	3,82
16	2,88	3,05	3,17	3,26	3,33	3,39	3,44	3,48	3,52	3,56	3,59	3,67	3,78
17	2,86	3,03	3,14	3,23	3,30	3,36	3,41	3,45	3,49	3,53	3,56	3,64	3,74
18	2,84	3,01	3,12	3,21	3,27	3,33	3,38	3,42	3,46	3,50	3,53	3,61	3,71
19	2,83	2,99	3,10	3,18	3,25	3,31	3,36	3,40	3,44	3,47	3,50	3,58	3,68
20	2,81	2,97	3,08	3,17	3,23	3,29	3,34	3,38	3,42	3,45	3,48	3,56	3,65
24	2,77	2,92	3,03	3,11	3,17	3,22	3,27	3,31	3,35	3,38	3,41	3,48	3,57
30	2,72	2,87	2,97	3,05	3,11	3,16	3,21	3,24	3,28	3,31	3,34	3,41	3,50
40	2,68	2,82	2,92	2,99	3,05	3,10	3,14	3,18	3,21	3,24	3,27	3,34	3,42
60	2,64	2,78	2,87	2,94	3,00	3,04	3,08	3,12	3,15	3,18	3,20	3,27	3,35
120	2,60	2,73	2,82	2,89	2,94	2,99	3,03	3,06	3,09	3,12	3,14	3,20	3,28
∞	2,56	2,68	2,77	2,84	2,89	2,93	2,97	3,00	3,03	3,06	3,08	3,14	3,21

Fonte: C. W. Dunnett (1955, 1964) e tabelas adicionais produzidas por C. W. Dunnett (1980).

Anexo 7. Comparações múltiplas com uma testemunha (teste de Dunnett) − α = 0,05 (bilateral).

Graus de liberdade	Número de tratamentos excluindo a testemunha												
	2	3	4	5	6	7	8	9	10	11	12	15	20
5	3,03	3,29	3,48	3,62	3,73	3,82	3,90	3,97	4,03	4,09	4,14	4,26	4,42
6	2,86	3,10	3,26	3,39	3,49	3,57	3,64	3,71	3,76	3,81	3,86	3,97	4,11
7	2,75	2,97	3,12	3,24	3,33	3,41	3,47	3,53	3,58	3,63	3,67	3,78	3,91
8	2,67	2,88	3,02	3,13	3,22	3,29	3,35	3,41	3,46	3,50	3,54	3,64	3,76
9	2,61	2,81	2,95	3,05	3,14	3,20	3,26	3,32	3,36	3,40	3,44	3,53	3,65
10	2,57	2,76	2,89	2,99	3,07	3,14	3,19	3,24	3,29	3,33	3,36	3,45	3,57
11	2,53	2,72	2,84	2,94	3,02	3,08	3,14	3,19	3,23	3,27	3,30	3,39	3,50
12	2,50	2,68	2,81	2,90	2,98	3,04	3,09	3,14	3,18	3,22	3,25	3,34	3,45
13	2,48	2,65	2,78	2,87	2,94	3,00	3,06	3,10	3,14	3,18	3,21	3,29	3,40
14	2,46	2,63	2,75	2,84	2,91	2,97	3,02	3,07	3,11	3,14	3,18	3,26	3,36
15	2,44	2,61	2,73	2,82	2,89	2,95	3,00	3,04	3,08	3,12	3,15	3,23	3,33
16	2,42	2,59	2,71	2,80	2,87	2,92	2,97	3,02	3,06	3,09	3,12	3,20	3,30
17	2,41	2,58	2,69	2,78	2,85	2,90	2,95	3,00	3,03	3,07	3,10	3,18	3,27
18	2,40	2,56	2,68	2,76	2,83	2,89	2,94	2,98	3,01	3,05	3,08	3,16	3,25
19	2,39	2,55	2,66	2,75	2,81	2,87	2,92	2,96	3,00	3,03	3,06	3,14	3,23
20	2,38	2,54	2,65	2,73	2,80	2,86	2,90	2,95	2,98	3,02	3,05	3,12	3,22
24	2,35	2,51	2,61	2,70	2,76	2,81	2,86	2,90	2,94	2,97	3,00	3,07	3,16
30	2,32	2,47	2,58	2,66	2,72	2,77	2,82	2,86	2,89	2,92	2,95	3,02	3,11
40	2,29	2,44	2,54	2,62	2,68	2,73	2,77	2,81	2,85	2,87	2,90	2,97	3,06
60	2,27	2,41	2,51	2,58	2,64	2,69	2,73	2,77	2,80	2,83	2,86	2,92	3,00
120	2,24	2,38	2,47	2,55	2,60	2,65	2,69	2,73	2,76	2,79	2,81	2,87	2,95
∞	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69	2,72	2,74	2,77	2,83	2,91

Fonte: C. W. Dunnett (1955, 1964) e tabelas adicionais produzidas por C. W. Dunnett (1980).

Anexo 8. Comparações múltiplas com uma testemunha (teste de Dunnett) − α = 0,01 (bilateral).

Graus de liberdade	Número de tratamentos excluindo a testemunha												
	2	3	4	5	6	7	8	9	10	11	12	15	20
5	4,63	4,98	5,22	5,41	5,56	5,69	5,80	5,89	5,98	6,05	6,12	6,30	6,52
6	4,21	4,51	4,71	4,87	5,00	5,10	5,20	5,28	5,35	5,41	5,47	5,62	5,81
7	3,95	4,21	4,39	4,53	4,64	4,74	4,82	4,89	4,95	5,01	5,06	5,19	5,36
8	3,77	4,00	4,17	4,29	4,40	4,48	4,56	4,62	4,68	4,73	4,78	4,90	5,05
9	3,63	3,85	4,01	4,12	4,22	4,30	4,37	4,43	4,48	4,53	4,57	4,68	4,82
10	3,53	3,74	3,88	3,99	4,08	4,16	4,22	4,28	4,33	4,37	4,42	4,52	4,65
11	3,45	3,65	3,79	3,89	3,98	4,05	4,11	4,16	4,21	4,25	4,29	4,39	4,52
12	3,39	3,58	3,71	3,81	3,89	3,96	4,02	4,07	4,12	4,16	4,19	4,29	4,41
13	3,33	3,52	3,65	3,74	3,82	3,89	3,94	3,99	4,04	4,08	4,11	4,20	4,32
14	3,29	3,47	3,59	3,69	3,76	3,83	3,88	3,93	3,97	4,01	4,05	4,13	4,24
15	3,25	3,43	3,55	3,64	3,71	3,78	3,83	3,88	3,92	3,95	3,99	4,07	4,18
16	3,22	3,39	3,51	3,60	3,67	3,73	3,78	3,83	3,87	3,91	3,94	4,02	4,13
17	3,19	3,36	3,47	3,56	3,63	3,69	3,74	3,79	3,83	3,86	3,90	3,98	4,08
18	3,17	3,33	3,44	3,53	3,60	3,66	3,71	3,75	3,79	3,83	3,86	3,94	4,04
19	3,15	3,31	3,42	3,50	3,57	3,63	3,68	3,72	3,76	3,79	3,83	3,90	4,00
20	3,13	3,29	3,40	3,48	3,55	3,60	3,65	3,69	3,73	3,77	3,80	3,87	3,97
24	3,07	3,22	3,32	3,40	3,47	3,52	3,57	3,61	3,64	3,68	3,70	3,78	3,87
30	3,01	3,15	3,25	3,33	3,39	3,44	3,49	3,52	3,56	3,59	3,62	3,69	3,78
40	2,95	3,09	3,19	3,26	3,32	3,37	3,41	3,44	3,48	3,51	3,53	3,60	3,68
60	2,90	3,03	3,12	3,19	3,25	3,29	3,33	3,37	3,40	3,42	3,45	3,51	3,59
120	2,85	2,97	3,06	3,12	3,18	3,22	3,26	3,29	3,32	3,35	3,37	3,43	3,51
∞	2,79	2,92	3,00	3,06	3,11	3,15	3,19	3,22	3,25	3,27	3,29	3,35	3,42

Fonte: C. W. Dunnett (1955, 1964) e tabelas adicionais produzidas por C. W. Dunnett (1980).

Anexo 9. Valores de q para o teste de Tukey ($q_{\alpha,n,gl}$). Probabilidade = 0,05.

Graus de liberdade do erro	Número de médias a comparar (n)																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18,1	26,7	32,8	37,2	40,5	43,1	45,4	47,3	49,1	50,6	51,9	53,2	54,3	55,4	56,3	57,2	58,0	58,8	59,6
2	6,09	8,28	9,80	10,89	11,73	12,43	13,03	13,54	13,99	14,39	14,75	15,08	15,38	15,65	15,91	16,14	16,36	16,57	16,77
3	4,50	5,88	6,83	7,51	8,04	8,47	8,85	9,18	9,46	9,72	9,95	10,16	10,35	10,52	10,69	10,84	10,98	11,12	11,24
4	3,93	5,00	5,76	6,31	6,73	7,06	7,35	7,60	7,83	8,03	8,21	8,37	8,52	8,67	8,80	8,92	9,03	9,14	9,24
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17	7,32	7,47	7,60	7,72	7,83	7,93	8,03	8,12	8,21
6	3,46	4,34	4,90	5,31	5,63	5,89	6,12	6,32	6,49	6,65	6,79	6,92	7,04	7,14	7,24	7,34	7,43	7,51	7,59
7	3,34	4,16	4,68	5,06	5,35	5,59	5,80	5,99	6,15	6,29	6,42	6,54	6,65	6,75	6,84	6,93	7,01	7,08	7,16
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18	6,29	6,39	6,48	6,57	6,65	6,73	6,80	6,87
9	3,20	3,95	4,42	4,76	5,02	5,24	5,43	5,60	5,74	5,87	5,98	6,09	6,19	6,28	6,36	6,44	6,51	6,58	6,65
10	3,15	3,88	4,33	4,66	4,91	5,12	5,30	5,46	5,60	5,72	5,83	5,93	6,03	6,12	6,20	6,27	6,34	6,41	6,47
11	3,10	3,82	4,26	4,58	5,82	5,03	5,20	5,35	5,49	5,61	5,71	5,81	5,90	5,98	6,06	6,14	6,20	6,27	6,33
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,40	5,51	5,61	5,71	5,80	5,88	5,95	6,02	6,09	6,15	6,21
13	3,06	3,73	4,15	4,46	4,69	4,88	5,05	5,19	5,32	5,43	5,53	5,63	5,71	5,79	5,86	5,93	6,00	6,06	6,11
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46	5,56	5,64	5,72	5,79	5,86	5,92	5,98	6,03
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40	5,49	5,57	5,65	5,72	5,79	5,85	5,91	5,96
16	3,00	3,65	4,05	4,34	4,56	4,74	4,90	5,03	5,15	5,26	5,35	5,44	5,52	5,59	5,66	5,73	5,79	5,84	5,90
17	2,98	3,62	4,02	4,31	4,52	4,70	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,55	5,61	5,68	5,74	5,79	5,84
18	2,97	3,61	4,00	4,28	4,49	4,67	4,83	4,96	5,07	5,17	5,27	5,35	5,43	5,50	5,57	5,63	5,69	5,74	5,79
19	2,96	3,59	3,98	4,26	4,47	4,64	4,79	4,92	5,04	5,14	5,23	5,32	5,39	5,46	5,53	5,59	5,65	5,70	5,75
20	2,95	3,58	3,96	4,24	4,45	4,62	4,77	4,90	5,01	5,11	5,20	5,28	5,36	5,43	5,50	5,56	5,61	5,66	5,71
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10	5,18	5,25	5,32	5,38	5,44	5,50	5,55	5,59
30	2,89	3,48	3,84	4,11	4,30	4,46	4,60	4,72	4,83	4,92	5,00	5,08	5,15	5,21	5,27	5,33	5,38	5,43	5,48
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,74	4,82	4,90	4,98	5,05	5,11	5,17	5,22	5,27	5,32	5,36
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,00	5,06	5,11	5,15	5,20	5,24
120	2,80	3,36	3,69	3,92	4,10	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,90	4,95	5,00	5,04	5,09	5,13
∞	2,77	3,32	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,68	4,74	4,80	4,84	4,89	4,93	4,97	5,01

Fonte: adaptado de Pearson e Hartley (1966).

Anexo 10. Valores de q para o teste de Tukey ($q_{\alpha,n,gl}$). Probabilidade = 0,01.

Graus de liberdade do erro	Número de médias a comparar (n)																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90	135	164	86	202	216	227	237	246	253	260	266	272	272	282	286	290	294	298
2	14,0	19,0	22,3	24,7	26,6	28,2	29,5	30,7	31,7	32,6	33,4	34,1	34,8	35,4	36,0	36,5	37,0	37,5	37,9
3	8,26	10,6	12,2	13,3	14,2	15,0	15,6	16,2	16,7	17,1	17,5	17,9	18,2	18,5	18,8	19,1	19,3	19,5	19,8
4	6,51	8,12	9,17	9,96	10,6	11,1	11,5	11,9	12,3	12,6	12,8	13,1	13,3	13,5	13,7	13,9	14,1	14,2	14,4
5	5,70	6,97	7,80	8,42	8,91	9,32	9,67	9,97	10,24	10,48	10,70	10,89	11,08	11,24	11,40	11,55	11,68	11,81	11,93
6	5,24	6,33	7,03	7,56	7,97	8,32	8,61	8,87	9,10	9,30	9,49	9,65	9,81	9,95	10,08	10,21	10,32	10,43	10,54
7	4,95	5,92	6,54	7,01	7,37	7,68	7,94	8,17	8,37	8,55	8,71	8,86	9,00	9,12	9,24	9,35	9,46	9,55	9,65
8	4,74	5,63	6,20	6,63	6,96	7,24	7,47	7,68	7,87	8,03	818	8,31	8,44	8,55	8,66	8,76	8,85	8,94	9,03
9	4,60	5,43	5,96	6,35	6,66	6,91	7,13	7,32	7,49	7,65	7,78	7,91	8,03	8,13	8,23	8,32	8,41	8,49	8,57
10	4,48	5,27	5,77	6,14	6,43	6,67	6,87	7,05	7,21	7,36	7,48	7,60	7,71	7,81	7,91	7,99	8,07	8,15	8,22
11	4,39	5,14	5,62	5,97	6,25	6,48	6,67	6,84	6,99	7,13	7,25	7,36	7,46	7,56	7,65	7,73	7,81	7,88	7,95
12	4,32	5,04	5,50	5,84	6,10	6,32	6,51	6,67	6,81	6,94	7,06	7,17	7,26	7,36	7,44	7,52	7,59	7,66	7,73
13	4,26	4,96	5,40	5,73	5,98	6,19	6,37	6,53	6,67	6,79	6,90	7,01	7,10	7,19	7,27	7,34	7,42	7,48	7,55
14	4,21	4,89	5,32	5,63	5,88	6,08	6,26	6,41	6,54	6,66	6,77	6,87	6,96	7,05	7,12	7,20	7,27	7,33	7,39
15	4,17	4,83	5,25	5,56	5,80	5,99	6,16	6,31	6,44	6,55	6,66	6,76	6,84	6,93	7,00	7,07	7,14	7,20	7,26
16	4,13	4,78	5,19	5,49	5,72	5,92	6,08	6,22	6,35	6,46	6,56	6,66	6,74	6,82	6,90	6,97	7,03	7,09	7,15
17	4,10	4,74	5,14	5,43	5,66	5,85	6,01	6,15	6,27	6,38	6,48	6,57	6,66	6,73	6,80	6,87	6,94	7,00	7,05
18	4,07	4,70	5,09	5,38	5,60	5,79	5,94	6,08	6,20	6,31	6,41	6,50	6,58	6,65	6,72	6,79	6,85	6,91	6,96
19	4,05	4,67	5,05	5,33	5,55	5,73	5,89	6,02	6,14	6,25	6,34	6,43	6,51	6,58	6,65	6,72	6,78	6,84	6,89
20	4,02	4,64	5,02	5,29	5,51	5,69	5,84	5,97	6,09	6,19	6,29	6,37	6,45	6,52	6,59	6,65	6,71	6,76	6,82
24	3,96	4,54	4,91	5,17	5,37	5,54	5,69	5,81	5,92	6,02	6,11	6,19	6,26	6,33	6,39	6,45	6,51	6,56	6,61
30	3,89	4,45	4,80	5,05	5,24	5,40	5,54	5,65	5,76	5,85	5,93	6,01	6,08	6,14	6,20	6,26	6,31	6,36	6,41
40	3,82	4,37	4,70	4,93	5,11	5,27	5,39	5,50	5,60	5,69	5,77	5,84	5,90	5,96	6,02	6,07	6,12	6,17	6,21
60	3,76	4,28	4,60	4,82	4,99	5,13	5,25	5,36	5,45	5,53	5,60	5,67	5,73	5,79	5,84	5,89	5,93	5,98	6,02
120	3,70	4,20	4,50	4,71	4,87	5,01	5,12	5,21	5,30	5,38	5,44	5,51	5,56	5,61	5,66	5,71	5,75	5,79	5,83
∞	3,64	4,12	4,40	4,60	4,76	4,88	4,99	5,08	5,16	5,23	6,01	5,35	5,40	5,45	5,49	5,54	5,57	5,61	5,65

Anexo 11. Valores da amplitude “studentizada” para 5% Y 1% (teste de Duncan).

Graus de liberdade do erro	Nível de proteção (α)	Número de médias a comparar (n)													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	,05	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0
	,01	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0
2	,05	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09
	,01	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0
3	,05	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50
	,01	8,26	8,5	8,6	8,7	8,8	8,9	8,9	9,0	9,0	9,0	9,1	9,2	9,3	9,3
4	,05	3,93	4,01	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02
	,01	6,51	6,8	6,9	7,0	7,1	7,1	7,2	7,2	7,3	7,3	7,4	7,4	7,5	7,5
5	,05	3,64	3,74	3,79	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83
	,01	5,70	5,96	6,11	6,18	6,26	6,33	6,40	6,44	6,5	6,6	6,6	6,7	6,7	6,8
6	,05	3,46	3,58	3,64	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68
	,01	5.24	5,51	5,65	5,73	5,81	5,88	5,95	6,00	6,0	6,1	6,2	6,2	6,3	6,3
7	,05	3,35	3,47	3,54	3,58	3,60	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61	3,61
	,01	4,95	5,22	5,37	5,45	5,53	5,61	5,69	5,73	5,8	5,8	5,9	5,9	6,0	6,0
8	,05	3,26	3,39	3,47	3,52	3,55	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56	3,56
	,01	4,74	5,00	5,14	5,23	5,32	5,40	5,47	5,51	5,5	5,6	5,7	5,7	5,8	5,8
9	,05	3,20	3,34	3,41	3,47	3,50	3,52	3,52	3,52	3,52	5,4	3,52	3,52	3,52	3,52
	,01	4,60	4,86	4,99	5,08	5,17	5,25	5,32	5,36	3,52	5,5	5,5	5,6	5,7	5,7
10	,05	3,15	3,30	3,37	3,43	3,46	3,47	3,47	3,47	3,47	3,47	3,47	3,47	3,47	3,48
	,01	4,48	4,73	4,88	4,96	5,06	5,13	5,20	5,24	5,28	5,36	5,42	5,48	5,54	5,55
11	,05	3,11	3,27	3,35	3,39	3,43	3,44	3,45	3,46	4,46	3,46	3,46	3,46	3,47	3,48
	,01	4,39	4,6	4,77	4,86	4,94	5,01	5,06	5,12	5,15	5,24	5,28	5,34	5,38	5,39
12	,05	3,08	3,23	3,33	3,36	3,40	3,42	3,44	3,44	3,46	3,46	3,46	3,46	3,47	3,48
	,01	4,32	4,55	4,68	4,76	4,81	4,92	4,96	5,02	5,07	5,13	5,17	5,22	5,24	5,26
13	,05	3,06	3,21	3,30	3,35	3,38	3,41	3,42	3,44	3,45	3,45	3,46	3,46	3,47	3,47
	,01	4,26	4,48	4,62	4,69	4,74	4,84	4,88	4,94	4,98	5,04	5,08	5,13	5,14	5,15

Continua...

Anexo 11. Continuação.

Graus de liberdade do erro	Nível de proteção (α)	Número de médias a comparar (n)													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
14	,05	3,03	3,18	3,27	3,33	3,37	3,39	3,41	3,42	3,44	3,45	3,46	3,46	3,47	3,47
	,01	4,21	4,42	4,55	4,63	4,70	4,78	4,83	4,87	4,91	4,96	5,00	5,04	5,06	5,07
15	,05	3,01	3,16	3,25	3,31	3,36	3,38	3,40	3,42	3,43	3,44	3,45	3,46	3,47	3,47
	,01	4,17	4,37	4,50	4,58	4,64	4,72	4,77	4,81	4,84	4,90	4,94	4,97	4,99	5,00
16	,05	3,00	3,15	3,23	3,30	3,34	3,37	3,39	3,41	3,43	3,44	3,45	3,46	3,47	3,47
	,01	4,13	4,34	4,45	4,54	4,60	4,67	4,72	4,76	4,79	4,84	4,88	4,91	4,93	4,94
17	,05	2,98	3,13	3,22	328	3,33	3,36	3,38	3,40	3,42	3,44	3,45	3,46	3,47	3,47
	,01	4,10	4,30	4,41	4,50	4,56	4,63	4,68	4,72	4,75	4,80	4,83	4,86	4,88	4,89
18	,05	2,97	3,12	3,21	3,27	3,32	3,35	3,37	3,39	3,41	3,43	3,45	3,46	3,47	3,47
	,01	4,07	4,27	4,38	4,46	4,53	4,59	4,64	4,68	4,71	4,76	4,79	4,82	4,84	4,85
19	,05	2,96	3,11	3,19	3,26	3,31	3,35	3,37	3,39	3,41	3,43	3,44	3,46	3,47	3,47
	,01	4,05	4,24	4,35	4,43	4,50	4,56	4,61	4,64	4,67	4,72	4,76	4,79	4,81	4,82
20	,05	2,95	3,10	3,18	3,25	3,30	3,34	3,36	3,38	3,40	3,43	3,44	3,46	3,46	3,47
	,01	4,02	4,22	4,33	4,40	4,47	4,53	4,58	4,61	4,65	4,69	4,73	4,76	4,78	4,79
22	,05	2,93	3,08	3,17	3,24	3,29	3,32	3,35	3,37	3,39	3,42	3,44	3,45	3,46	3,47
	,01	3,99	4,17	4,28	4,36	4,42	4,48	4,53	4,57	4,60	4,65	4,68	4,71	4,74	4,75
24	,05	2,92	3,07	3,15	3,22	3,28	3,31	3,34	3,37	3,38	3,41	3,44	3,45	3,46	3,47
	,01	3,96	4,14	4,24	4,33	4,39	4,44	4,49	4,53	4,57	4,62	4,64	4,67	4,70	4,72
26	,05	2,91	3,06	3,14	3,21	3,27	3,30	3,34	3,36	3,38	3,41	3,43	3,45	3,46	3,47
	,01	3,93	4,11	4,21	4,30	4,36	4,41	4,46	4,50	4,53	4,58	4,62	4,65	4,67	4,69
28	,05	2,90	3,04	3,13	3,20	3,26	3,30	3,33	3,35	3,37	3,40	3,43	3,45	3,46	3,47
	,01	3,91	4,08	4,18	4,28	4,34	4,39	4,43	4,47	4,51	4,56	4,60	4,62	4,65	4,67
30	,05	2,89	3,04	3,12	3,20	3,25	3,29	3,32	3,35	3,37	3,40	3,43	3,44	3,46	3,47
	,01	3,89	4,06	4,16	4,22	4,32	4,36	4,41	4,45	4,48	4,54	4,58	4,61	4,63	4,65
40	,05	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	3,35	3,39	3,42	3,44	3,46	3,47
	,01	3,82	3,99	4,10	4,17	4,24	4,30	4,34	4,37	4,41	4,46	4,51	4,54	4,57	4,59

Continua...

Anexo 11. Continuação.

Graus de liberdade do erro	Nível de proteção (α)	Número de médias a comparar (n)													
		2	3	4	5	6	7	8	9	10	12	14	16	18	20
60	,05	2,83	2,98	3,08	3,14	3,20	3,24	3,28	3,31	3,33	3,37	3,40	3,43	3,45	3,47
	,01	3,76	3,92	4,03	4,12	4,17	4,23	4,27	4,31	4,34	4,39	4,44	4,47	4,50	4,53
100	,05	2,80	2,95	3,05	3,12	3,18	3,22	3,26	3,29	3,32	3,36	3,40	3,42	3,45	3,47
	,01	3,71	3,86	3,98	4,06	4,11	4,17	4,21	4,25	4,29	4,35	4,38	4,42	4,45	4,48
00	,05	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26	3,29	3,34	3,38	3,41	3,44	3,47
	,01	3,64	3,80	3,90	3,98	4,04	4,09	4,14	4,17	4,20	4,26	4,31	4,34	4,38	4,41

Fonte: D. B. Duncan (1955).

Anexo 12. Quantis do teste de Mann-Whitney.

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	,001	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	,005	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4
	,01	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5
	,025	3	3	3	3	3	3	4	4	4	5	5	5	5	5	5	6	6	6	6
	,05	3	3	3	4	4	4	5	5	5	5	6	6	7	7	7	7	8	8	8
	,10	3	4	4	5	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11
3	,001	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7
	,005	6	6	6	6	6	6	6	7	7	7	8	8	8	9	9	9	9	10	10
	,01	6	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12
	,025	6	6	6	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15
	,05	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	18
	,10	7	8	8	9	10	11	12	12	13	14	15	16	17	17	18	19	20	21	22
4	,001	10	10	10	10	10	10	10	10	11	11	11	12	12	12	13	13	14	14	14
	,005	10	10	10	10	11	11	12	12	13	13	14	14	15	16	16	17	17	18	19
	,01	10	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21
	,025	10	10	11	12	13	14	15	15	16	17	18	19	20	21	22	22	23	24	25
	,05	10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
	,10	11	12	14	15	16	17	18	20	21	22	23	24	26	27	28	29	31	32	33
5	,001	15	15	15	15	15	15	16	17	17	18	18	19	19	20	21	21	22	23	23
	,005	15	15	15	16	17	17	18	19	20	21	22	23	23	24	25	26	27	28	29
	,01	15	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	,025	15	16	17	18	19	21	22	23	24	25	27	28	29	30	31	33	34	35	36
	,05	16	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
	,10	17	18	20	21	23	24	26	28	29	31	33	34	36	38	39	41	43	44	46

Continua...

Anexo 12. Continuação.

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	,001	21	21	21	21	21	21	23	24	25	26	26	27	28	29	30	31	32	33	34
	,005	21	21	22	23	24	25	26	27	28	29	31	32	33	34	35	37	38	39	40
	,01	21	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44
	,025	21	23	24	25	27	28	30	32	33	35	36	38	39	41	43	44	46	47	49
	,05	22	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
	,10	23	25	27	29	31	33	35	37	39	41	43	45	47	49	51	53	56	58	60
7	,001	28	28	28	28	29	30	31	32	34	35	36	37	38	39	40	42	43	44	45
	,005	28	28	29	30	32	33	35	36	38	39	41	42	44	45	47	48	50	51	53
	,01	28	29	30	32	33	35	36	34	40	41	43	45	46	48	50	52	53	58	57
	,025	28	30	32	34	35	37	39	41	43	45	47	49	51	53	55	57	59	61	63
	,05	29	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
	,10	30	33	35	37	40	42	45	47	50	52	55	57	60	62	65	67	70	72	75
8	,001	36	36	36	37	38	39	41	42	43	45	46	48	49	51	52	54	55	57	58
	,005	36	36	38	39	41	43	44	46	48	50	52	54	55	57	59	61	63	65	67
	,01	36	37	39	41	43	44	46	48	50	52	54	56	59	61	63	65	67	69	71
	,025	37	39	41	43	45	47	50	52	54	56	59	61	63	66	68	71	73	75	78
	,05	38	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
	,10	39	42	44	47	50	53	56	59	61	64	67	70	73	76	79	82	85	88	91
9	,001	45	45	45	47	48	49	51	53	54	56	58	60	61	63	65	67	69	71	72
	,005	45	46	47	49	51	53	55	57	59	62	64	66	68	70	73	75	77	79	82
	,01	45	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86
	,025	46	48	50	53	56	58	61	63	66	69	72	74	77	80	83	85	88	91	94
	,05	47	50	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
	,10	48	51	55	58	61	64	68	71	74	77	81	84	87	91	94	98	101	104	108

Continua...

Anexo 12. Continuação.

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	,001	55	55	56	57	59	61	62	64	66	68	70	73	75	77	79	81	83	85	88
	,005	55	56	58	60	62	65	67	69	72	74	77	80	82	85	87	90	93	95	98
	,01	55	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103
	,025	56	59	61	64	67	70	73	76	79	82	85	89	92	95	98	101	104	108	111
	,05	57	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
	,10	59	62	66	69	73	77	80	84	88	92	95	99	103	107	110	114	118	122	126
11	,001	66	66	67	69	71	73	75	77	79	82	84	87	89	91	94	96	99	101	104
	,005	66	67	69	72	74	77	80	83	85	88	91	94	97	100	103	106	109	112	115
	,01	66	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120
	,025	67	70	73	76	80	83	86	90	93	97	100	104	107	111	114	118	122	125	129
	,05	68	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
	,10	70	74	78	82	86	90	94	98	103	107	111	115	119	124	128	132	136	140	145
12	,001	78	78	79	81	83	86	88	91	93	96	98	102	104	106	110	113	116	118	121
	,005	78	80	82	85	88	91	94	97	100	103	106	110	113	116	120	123	126	130	133
	,01	78	81	84	87	90	93	96	100	103	107	111	114	117	121	125	128	132	135	139
	,025	80	83	86	90	93	97	101	105	108	112	116	120	124	128	132	136	140	144	148
	,05	81	84	88	92	96	100	105	109	111	117	121	126	130	134	139	143	147	151	156
	,10	83	87	91	96	100	105	109	114	118	123	128	132	137	142	146	151	156	160	165
13	,001	91	91	93	95	97	100	103	106	109	112	115	118	121	124	127	130	134	137	140
	,005	91	93	95	99	102	105	109	112	116	119	123	126	130	134	137	141	145	149	152
	,01	92	94	97	101	104	108	112	115	119*	123	127	131	135	139	143	147	151	155	159
	,025	93	96	100	104	108	112	116	120	125	129	133	137	142	146	151	155	159	164	168
	,05	94	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
	,10	96	101	105	110	115	120	125	130	135	140	145	150	155	160	166	171	176	181	186

Continua...

Anexo 12. Continuação.

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
14	,001	105	105	107	109	112	115	118	121	125	128	131	135	138	142	145	149	152	156	160
	,005	105	107	110	113	117	121	124	128	132	136	140	144	148	152	156	160	164	169	173
	,01	106	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179
	,025	107	111	115	119	123	128	132	137	142	146	151	156	161	165	170	175	180	184	189
	,05	109	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
	,10	110	116	121	126	131	137	142	147	153	158	164	169	175	180	186	191	197	203	208
15	,001	120	120	122	125	128	133	135	138	142	145	149	153	157	161	164	168	172	176	180
	,005	120	123	126	129	133	137	141	145	150	154	158	163	167	172	176	181	185	190	194
	,01	121	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201
	,025	122	126	131	135	140	145	150	155	160	165	170	175	180	185	191	196	201	206	211
	,05	124	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
	,10	126	131	137	143	148	154	160	166	172	178	184	189	195	201	207	213	219	225	231
16	,001	136	136	139	142	145	148	152	156	160	164	168	172	176	180	185	189	193	197	202
	,005	136	139	142	146	150	155	159	164	168	173	178	182	187	192	197	202	207	211	216
	,01	137	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	2163	219	224
	,025	138	143	148	152	158	163	168	174	179	184	190	196	201	207	212	218	223	229	235
	,05	140	145	151	156	162	167	173	179	185	191	197	202	2508	214	220	226	232	238	244
	,10	142	148	154	160	166	173	179	185	191	198	204	211	217	223	230	236	243	249	256
17	,001	153	154	156	159	163	167	171	175	179	1836	188	192	197	201	206	211	215	220	224
	,005	153	156	160	164	169	173	178	183	188	193	198	203	208	214	219	224	229	235	240
	,01	154	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247
	,025	156	160	165	171	176	182	188	193	199	205	211	217	223	229	235	241	247	253	259
	,05	157	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
	,10	160	166	172	179	185	192	199	206	212	219	226	233	239	246	253	260	267	274	281

Continua...

Anexo 12. Continuação.

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
18	,001	171	172	175	178	182	186	190	195	199	204	209	214	218	223	228	233	238	243	248
	,005	171	174	178	183	188	193	198	203	209	214	219	225	230	236	242	247	253	259	264
	,01	172	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272
	,025	174	179	184	190	196	202	208	214	220	227	233	239	246	252	258	265	271	278	284
	,05	176	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
	,10	178	185	192	199	206	213	220	227	234	241	249	256	263	270	278	285	292	300	307
19	,001	190	191	194	198	202	206	211	216	220	225	231	236	241	246	251	257	262	268	273
	,005	191	194	198	203	208	213	219	224	230	236	242	248	254	260	265	272	278	284	290
	,01	192	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298
	,025	193	198	204	210	216	223	229	236	243	249	256	263	269	276	283	290	297	304	310
	,05	195	201	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	320
	,10	198	205	212	219	227	234	242	249	257	264	272	280	288	295	303	311	319	326	334
20	,001	210	211	214	218	223	227	232	237	243	248	253	259	265	270	276	281	287	293	299
	,005	211	214	219	224	229	235	241	247	253	259	265	271	278	284	290	297	303	310	316
	,01	212	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325
	,025	213	219	225	231	238	245	251	259	266	273	280	287	294	301	309	316	323	330	338
	,05	215	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349
	,10	218	226	233	241	249	257	265	273	281	289	297	305	313	321	330	338	346	354	362

Anexo 13. Quantis do teste de Kruskal-Wallis para pequenas amostras.

Sample sizes	$w_{0,90}$	$w_{0,95}$	$W_{0,99}$
2, 2, 2	3,7143	4,5714	4,5714
3, 2, 1	3,8571	4,2857	4,2857
3, 2, 2	4,4643	4,5000	5,3571
3, 3, 1	4,0000	4,5714	5,1429
3, 3, 2	4,2500	5,1389	6,2500
3, 3, 3	4,6202	5,6000	6,4889
4, 2, 1	4,0179	4,8214	4,8214
4, 2, 2	4,1667	5,1250	6,0000
4, 3, 1	3,8889	5,0000	5,8333
4, 3, 2	4,444	5,7273	6,7091
4, 4, 1	4,0667	4,8667	6,1667
4, 4, 2	4,4455	5,2364	6,8727
4, 4, 3	4,773	5,5758	7,1364
4, 4, 4	4,5000	5,6538	7,5385
5, 2, 1	4,0500	4,4500	5,2500
5, 2, 2	4,2933	5,0400	6,1333
5, 3, 1	3,8400	4,8711	6,4000
5, 3, 2	4,4946	5,1055	6,8218
5, 3, 3	4,4121	5,5152	6,9818
5, 4, 1	3,9600	4,8600	6,8400
5, 4, 2	4,5182	5,2682	7,1182
5, 4, 3	4,5231	5,6308	7,3949
5, 4, 4	4,6187	5,6176	7,7440
5, 5, 1	4,0364	4,9091	6,8364
5, 5, 2	4,5077	5,2462	7,2692
5, 5, 3	4,2363	5,6264	7,5429
5, 5, 4	4,5200	5,6429	7,7914
5, 5, 5	4,5000	5,6600	7,9800

Fonte: Conover (1980).

Anexo 14. Quantis do teste de Spearman

n	p = ,900	,950	,975	,990	,995	,999
4	,8000	,8000				
5	,7000	,8000	,9000	,9000		
6	,6000	,7714	,8286	,8857	,9429	
7	,5357	,6786	,7450	,8571	,8929	,9643
8	,5000	,6190	,7143	,8095	,8571	,9286
9	,4667	,5833	,6833	,7667	,8167	,9000
10	,4424	,5515	,6364	,7333	,7818	,8667
11	,4182	,5273	,6091	,7000	,7455	,8364
12	,3986	,4965	,5804	,6713	,7273	,8182
13	,3791	,4780	,5549	,6429	,6978	,7912
14	,3626	,4593	,5341	,6220	,6747	,7670
15	,3500	,4429	,5179	,6000	,6536	,7464
16	,3381	,4265	,5000	,5824	,6324	,7265
17	,3260	,4118	,4853	,5637	,6152	,7083
18	,3148	,3994	,4716	,5480	,5975	,6904
19	,3070	,3895	,4579	,5333	,5825	,6737
20	,2977	,3789	,4451	,5203	,5684	,6586
21	,2909	,3688	,4351	,5078	,5545	,6455
22	,2829	,3597	,4241	,4963	,5426	,6318
23	,2767	,3518	,4150	,4852	,5306	,6186
24	,2704	,3435	,4061	,4748	,5200	,6070
25	,2646	,3362	,3977	,4654	,5100	,5962
.26	,2588	,3299	,3894	,4564	,5002	,5856
27	,2540	,3236	,3822	,4481	,4915	,5757
28	,2490	,3175	,3749	,4401	,4828	,5660
29	,2443	,3113	,3685	,4320	,4744	,5567
30	,2400	,3059	,3620	,4251	,4665	,5479

Fonte: Conover (1980).

Anexo 15. Quantis do teste de Hotelling-Pabst

N	p=,001	,005	,010	,025	,050	,100	(1/3) $n(n^{2}-1)$
4					2	2	20
5			2	2	4	6	40
6		2	4	6	8	14	70
7	1	6	8	14	18	26	112
8	6	12	16	24	32	42	168
9	12	22	28	38	50	64	240
10	22	36	44	60	74	92	330
11	36	56	66	86	104	128	440
12	52	78	94	120	144	172	572
13	76	110	130	162	190	226	728
14	106	148	172	212	246	290	910
15	142	194	224	270	312	364	1.120
16	186	250	284	340	390	450	1.360
17	238	314	356	420	480	550	1.632
18	300	390	438	512	582	664	1.938
19	372	476	532	618	696	790	2.280
20	454	574	638	738	826	934	2.660
21	546	686	758	870	972	1.092	3.080
22	652	810	892	1.020	1.134	1.270	3.542
23	772	950	1.042	1.184	1.312	1.464	4.048
24	904	1.104	1.208	1.366	1.510	1.678	4.600
25	1.050	1.274	1.390	1.566	1.726	1.912	5.200
26	1.212	1.462	1.590	1.786	1.960	2.168	5.850
27	1.390	1.666	1.808	2.024	2.216	2.444	6.552
28	1.586	1.890	2.046	2.284	2.494	2.744	7.308
29	1.800	2.134	2.306	2.564	2.796	3.068	8.120
30	2.032	2.398	2.584	2.868	3.120	3.416	8.990

Fonte: Conover (1980).

Anexo 16. Quantis do Teste de Kendall.

n	P = ,900	,950	,975	,990	,995
4	4	4	6	6	6
5	6	6	8	8	10
6	7	9	11	11	13
7	9	11	13	15	17
8	10	14	16	18	20
9	12	16	18	22	24
10	15	19	21	25	27
11	17	21	25	29	31
12	18	24	28	34	36
13	22	26	32	38	42
14	23	31	35	41	45
15	27	33	39	47	51
16	28	36	44	50	56
17	32	40	48	56	62
18	35	43	51	61	67
19	37	47	55	65	73
20	40	50	60	70	78
21	42	54	64	76	84
22	45	59	69	81	89
23	49	63	73	87	97
24	52	66	78	92	102
25	56	70	84	98	108
26	59	75	89	105	115
27	61	79	93	111	123
28	66	84	98	116	128
29	68	88	104	124	136
30	73	93	109	129	143
31	75	97	115	135	149
32	80	102	120	142	158
33	84	106	126	150	164
34	87	111	131	155	173
35	91	115	137	163	179
36	94	120	144	170	188
37	98	126	150	176	196
38	103	131	155	183	203
39	104	137	161	191	211
40	110	142	168	198	220
41	114	146	174	206	228
42	119	151	181	213	235
43	123	157	187	221	245
44	128	162	194	228	252
45	132	168	200	236	262

Continua...

Anexo 16. Continuação.

n	P = ,900	,950	,975	,990	,995
46	135	173	207	245	271
47	141	179	213	253	279
48	144	186	220	260	288
49	150	191	228	268	296
50	153	197	233	277	305
51	159	203	241	285	315
52	162	208	248	294	324
53	168	214	256	302	334
54	173	221	263	311	343
55	177	227	269	319	353
56	182	232	276	328	362
57	186	240	284	336	372
58	191	245	291	345	381
59	197	251	299	355	391
60	202	258	306	364	402

Fonte: Conover (1980).

Anexo 17. Valores da distribuição normal padronizada (=0 e =1)

Probabilidade	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,50	0,0000	0,0025	0,0050	0,0075	0,0100	0,0125	0,0150	0,0175	0,0201	0,0226
0,51	0,0251	0,0276	0,0301	0,0326	0,0351	0,0376	0,0401	0,0426	0,0451	0,0476
0,52	0,0502	0,0527	0,0552	0,0577	0,0602	0,0627	0,0652	0,0677	0,0702	0,0728
0,53	0,0753	0,0778	0,0803	0,0828	0,0853	0,0878	0,0904	0,0929	0,0954	0,0979
0,54	0,1004	0,1030	0,1055	0,1080	0,1105	0,1130	0,1156	0,1181	0,1206	0,1231
0,55	0,1257	0,1282	0,1307	0,1332	0,1358	0,1383	0,1408	0,1434	0,1459	0,1484
0,56	0,1510	0,1535	0,1560	0,1586	0,1611	0,1637	0,1662	0,1687	0,1713	0,1738
0,57	0,1764	0,1789	0,1815	0,1840	0,1866	0,1891	0,1917	0,1942	0,1968	0,1993
0,58	0,2019	0,2045	0,2070	0,2096	0,2121	0,2147	0,2173	0,2198	0,2224	0,2250
0,59	0,2275	0,2301	0,2327	0,2353	0,2378	0,2404	0,2430	0,2456	0,2482	0,2508
0,60	0,2533	0,2559	0,2585	0,2611	0,2637	0,2663	0,2689	0,2715	0,2741	0,2767
0,61	0,2793	0,2819	0,2845	0,2871	0,2898	0,2924	0,2950	0,2976	0,3002	0,3029
0,62	0,3055	0,3081	0,3107	0,3134	0,3160	0,3186	0,3213	0,3239	0,3266	0,3292
0,63	0,3319	0,3345	0,3372	0,3398	0,3425	0,3451	0,3478	0,3505	0,3531	0,3558
0,64	0,3585	0,3611	0,3638	0,3665	0,3692	0,3719	0,3745	0,3772	0,3799	0,3826
0,65	0,3853	0,3880	0,3907	0,3934	0,3961	0,3989	0,4016	0,4043	0,4070	0,4097
0,66	0,4125	0,4152	0,4179	0,4207	0,4234	0,4261	0,4289	0,4316	0,4344	0,4372
0,67	0,4399	0,4427	0,4454	0,4482	0,4510	0,4538	0,4565	0,4593	0,4621	0,4649
0,68	0,4677	0,4705	0,4733	0,4761	0,4789	0,4817	0,4845	0,4874	0,4902	0,4930
0,69	0,4958	0,4987	0,5015	0,5044	0,5072	0,5101	0,5129	0,5158	0,5187	0,5215

Continua...

Anexo 17. Continuação.

Probabilidade	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,70	0,5244	0,5273	0,5302	0,5330	0,5359	0,5388	0,5417	0,5446	0,5476	0,5505
0,71	0,5534	0,5563	0,5592	0,5622	0,5651	0,5681	0,5710	0,5740	0,5769	0,5799
0,72	0,5828	0,5858	0,5888	0,5918	0,5948	0,5978	0,6008	0,6038	0,6068	0,6098
0,73	0,6128	0,6158	0,6189	0,6219	0,6250	0,6280	0,6311	0,6341	0,6372	0,6403
0,74	0,6433	0,6464	0,6495	0,6526	0,6557	0,6588	0,6620	0,6651	0,6682	0,6713
0,75	0,6745	0,6776	0,6808	0,6840	0,6871	0,6903	0,6935	0,6967	0,6999	0,7031
0,76	0,7063	0,7095	0,7128	0,7160	0,7192	0,7225	0,7257	0,7290	0,7323	0,7356
0,77	0,7388	0,7421	0,7454	0,7488	0,7521	0,7554	0,7588	0,7621	0,7655	0,7688
0,78	0,7722	0,7756	0,7790	0,7824	0,7858	0,7892	0,7926	0,7961	0,7995	0,8030
0,79	0,8064	0,8099	0,8134	0,8169	0,8204	0,8239	0,8274	0,8310	0,8345	0,8381
0,80	0,8416	0,8452	0,8488	0,8524	0,8560	0,8596	0,8632	0,8669	0,8706	0,8742
0,81	0,8779	0,8816	0,8853	0,8890	0,8927	0,8965	0,9002	0,9040	0,9078	0,9116
0,82	0,9154	0,9192	0,9230	0,9269	0,9307	0,9346	0,9385	0,9424	0,9463	0,9502
0,83	0,9542	0,9581	0,9621	0,9661	0,9701	0,9741	0,9782	0,9822	0,9863	0,9904
0,84	0,9945	0,9986	1,0027	1,0069	1,0110	1,0152	1,0194	1,0237	1,0279	1,0322
0,85	1,0364	1,0407	1,0451	1,0494	1,0537	1,0581	1,0625	1,0669	1,0714	1,0758
0,86	1,0803	1,0848	1,0893	1,0939	1,0985	1,1031	1,1077	1,1123	1,1170	1,1217
0,87	1,1264	1,1311	1,1359	1,1407	1,1455	1,1503	1,1552	1,1601	1,1650	1,1700
0,88	1,1750	1,1800	1,1850	1,1901	1,1952	1,2004	1,2055	1,2107	1,2160	1,2212
0,89	1,2265	1,2319	1,2372	1,2426	1,2481	1,2536	1,2591	1,2646	1,2702	1,2759

Continua...

Anexo 17. Continuação.

Probabilidade	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,90	1,2816	1,2873	1,2930	1,2988	1,3047	1,3106	1,3165	1,3225	1,3285	1,3346
0,91	1,3408	1,3469	1,3532	1,3595	1,3658	1,3722	1,3787	1,3852	1,3917	1,3984
0,92	1,4051	1,4118	1,4187	1,4255	1,4325	1,4395	1,4466	1,4538	1,4611	1,4684
0,93	1,4758	1,4833	1,4909	1,4985	1,5063	1,5141	1,5220	1,5301	1,5382	1,5464
0,94	1,5548	1,5632	1,5718	1,5805	1,5893	1,5982	1,6072	1,6164	1,6258	1,6352
0,95	1,6449	1,6546	1,6646	1,6747	1,6849	1,6954	1,7060	1,7169	1,7279	1,7392
0,96	1,7507	1,7624	1,7744	1,7866	1,7991	1,8119	1,8250	1,8384	1,8522	1,8663
0,97	1,8808	1,8957	1,9110	1,9268	1,9431	1,9600	1,9774	1,9954	2,0141	2,0335
0,98	2,0537	2,0748	2,0969	2,1201	2,1444	2,1701	2,1973	2,2262	2,2571	2,2904
0,99	2,3263	2,3656	2,4089	2,4573	2,5121	2,5758	2,6521	2,7478	2,8782	3,0902

Nota: valores calculados com o uso da fórmula INV.NORMP(probabilidade) de Excel© da Microsoft©.

Livraria Embrapa

Na Livraria Embrapa, você encontra
livros, DVDs e CD-ROMs sobre
agricultura, pecuária, negócio agrícola, etc.

Para fazer seu pedido, acesse:
www.embrapa.br/livraria

ou entre em contato conosco
Fone: (61) 3448-4236
Fax: (61) 3448-2494
livraria@embrapa.br

Você pode também nos encontrar nas redes sociais:

 facebook.com/livrariaembrapa

 twitter.com/livrariaembrapa

Impressão e acabamento
Embrapa Informação Tecnológica

O papel utilizado nesta publicação foi produzido conforme a certificação do Bureau Veritas Quality International (BVQI) de Manejo Florestal.

Embrapa
Arroz e Feijão

A segunda edição do livro *Estatística aplicada à pesquisa agrícola* foi ampliada com novos procedimentos estatísticos e informações para atender à demanda de pesquisadores agrícolas. Esta segunda edição apresenta o programa SAS (*Statistical Analises System*) de análise de delineamento e procedimentos.

A obra oferece os conhecimentos necessários para que pesquisadores e estudantes planejem seus experimentos e analisem seus dados da forma mais adequada para cada caso. É, assim, um importante aliado na pesquisa, em que o leitor poderá tirar o máximo proveito de dados experimentais, testando hipóteses, comparando tratamentos e validando conclusões.

Ministério da
Agricultura, Pecuária e Abastecimento

GOVERNO FEDERAL
BRASIL
PAÍS RICO É PAÍS SEM POBREZA

ISBN 978-85-7035-344-3
9 788570 353443
CGPE 11226