PUBVET, Publicações em Medicina Veterinária e Zootecnia.

Conceitos estatísticos aplicados à experimentação zootécnica

Ana Karina Dias Salman[1] e Poliana Fernanda Giachetto[2]

[1]Zootecnista, Doutora, Pesquisadora da Embrapa, Rondônia.

[2]Zootecnista, Doutora, Pesquisadora da Embrapa Informática Agropecuária.

Resumo

Diante da importância dos produtos de origem animal para o cenário econômico brasileiro, é constante a busca por inovações científicas e tecnológicas com resultados econômicos e práticos para serem aplicados em sistemas de produção animal ou em algum outro setor da cadeia produtiva. Através da experimentação científica, pesquisadores utilizam diversos recursos metodológicos para alcance de melhores índices de produtividade, com base em resultados confiáveis e consolidados. Dentre esses recursos, os métodos estatísticos auxiliam, dentro dos preceitos da metodologia científica, a testar hipóteses levantadas e interpretar os resultados observados em um determinado estudo. Esse trabalho objetivou sumarizar os princípios básicos de estatística descritiva e de experimentação que fundamentam e auxiliam o planejamento de estudos científicos, dando enfoque à produção animal. Dessa forma, as informações aqui contidas podem servir como um guia básico, prático e em linguagem acessível para estudantes, técnicos e pesquisadores interessados em experimentação zootécnica.

Palavras-chave: zootecnia, metodologia científica, experimentação zootécnica

Statistical concepts applied to animal science experimentation

Abstract

Facing the importance of animal products to the Brazilian economic scenery, is always actual and constant to seek for scientific and technological innovations with economic and practical results that should be applied in animal production system or other animal production chain. Through the scientific experimentation, researchers can use many methodological tools for reaching better yield indices based on reliable and solid results. Among these tools, statistical methods aid to test hypothesis and to interpret observed results in a specific study, following the scientific principles. This document aimed to summarize the basic principles of scientific experimentation and descriptive statistical analysis which underline and help to plan scientific studies, focusing on animal science. Thus, this material contains information that should be useful for students, technical staffs and researchers with interest in a basic and practical guide for animal science experimentation.

Keywords: animal science, scientific methodology, animal science experimentation

INTRODUÇÃO

Na pesquisa agropecuária, os dados obtidos nos experimentos, quando atendem às pressuposições do modelo matemático, são passíveis de serem submetidos a algum tipo de análise estatística, sendo a mais frequente a análise da variância, na qual é importante ter conhecimento da natureza dos tratamentos avaliados, pois quando estes são de efeito fixo, a análise aplicada visa estimar os efeitos individuais de cada um e compará-los entre si (DAL'COL LÚCIO *et al.,* 2003). Por outro lado, quando os tratamentos são de efeito aleatório, a análise visa à estimação dos componentes da variância, que possuem grande importância no melhoramento genético, animal ou vegetal (BARBIN, 1993). Para Volpato (2000), a estatística é uma importante ferramenta auxiliar no estabelecimento do discurso científico sobre os fenômenos observados, mas nem sempre imprescindível e sugere que o bom senso também seja considerado como recurso de análise a ser considerado em um estudo cientifico.

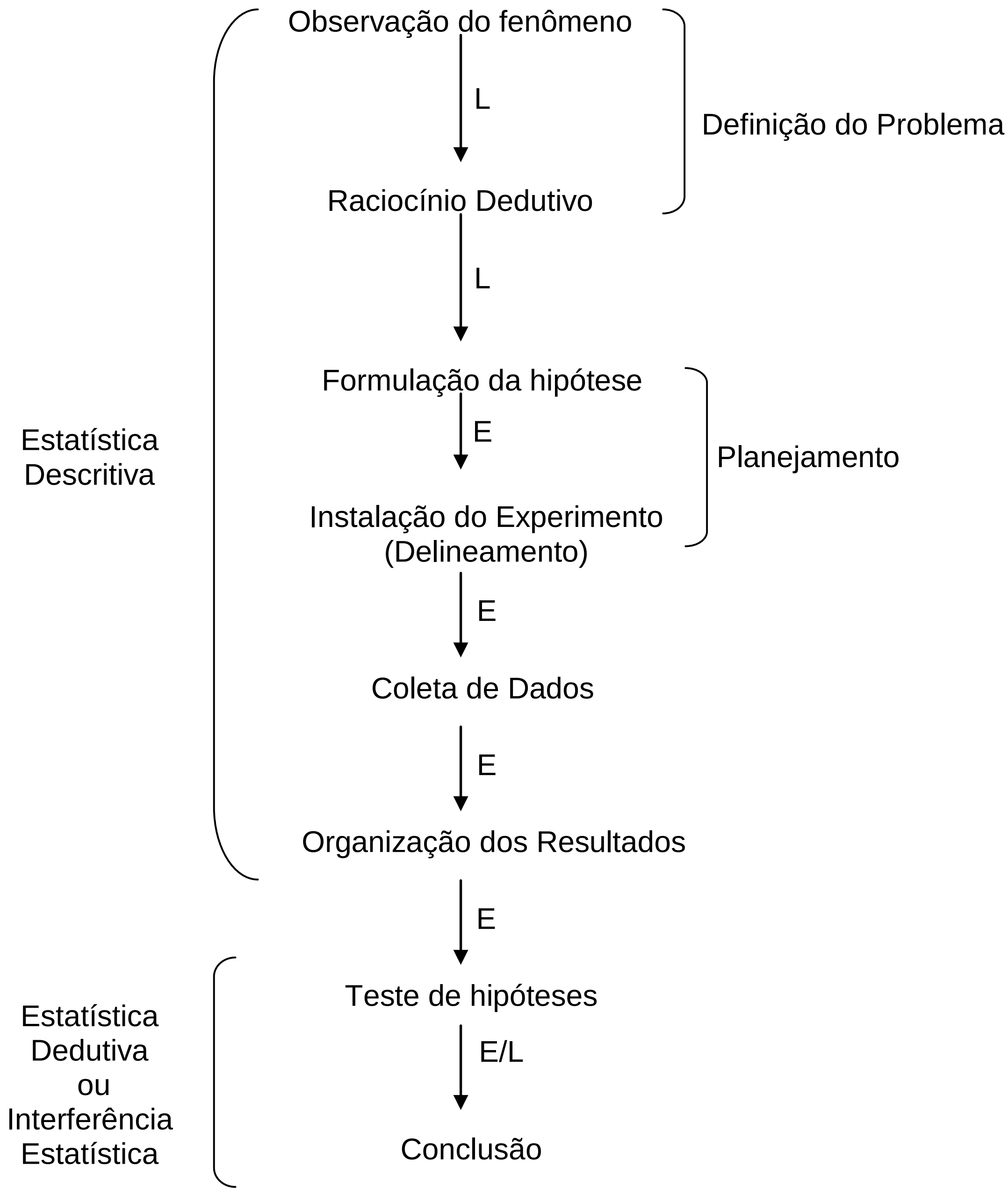
Sendo assim, esta revisão tem por objetivo sumarizar os princípios básicos de estatística descritiva e de experimentação que fundamentam e auxiliam o planejamento de estudos científicos, dando enfoque à produção animal.

CONCEITOS DE ESTATÍSTICA DESCRITIVA

Segundo Sampaio (2007), a estatística descritiva auxilia:

1) Na amostragem de dados ou no delineamento de um experimento;
2) Na apresentação dos resultados (tabelas e gráficos) e no estudo descritivo dos dados,
3) Na análise e interpretação dos resultados.

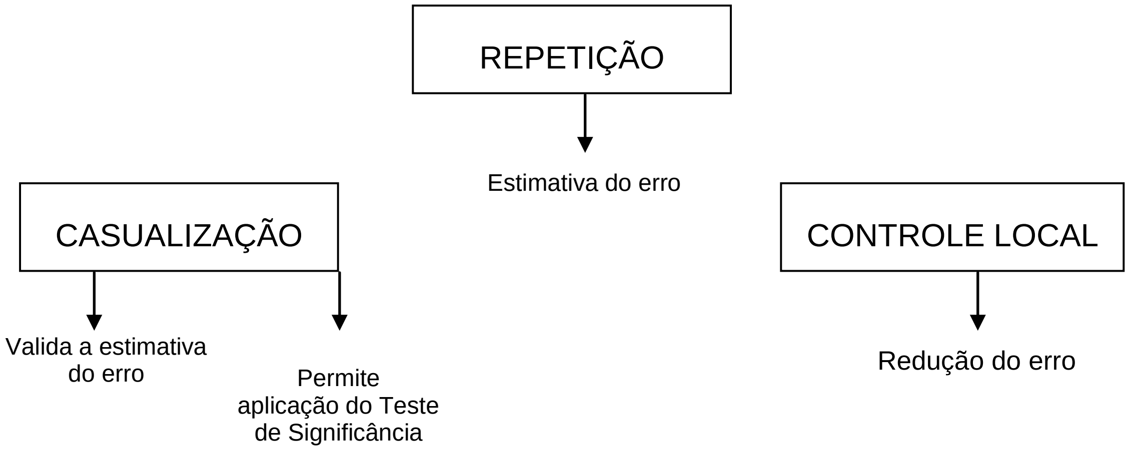
Na metodologia científica, a lógica e o julgamento crítico, além de um considerável domínio sobre o tema abordado, são necessários para o levantamento do problema e formulação das hipóteses. Estas, no entanto, só podem ser testadas com o auxílio da estatística. A seguir, o fluxograma ilustra os passos da metodologia e os pontos em que são necessárias as intervenções da lógica (L) ou da estatística (E):



É importante destacar, como se observa no fluxograma acima, que a estatística deve ser utilizada desde a fase de planejamento do experimento, seja ele de qualquer natureza. A correta aplicação dos modelos estatísticos e, consequentemente, a adequada análise e interpretação dos resultados só é possível quando o experimento obedece aos princípios básicos da experimentação.

PRINCÍPIOS BÁSICOS DE EXPERIMENTAÇÃO

Os princípios básicos da experimentação, que são a **casualização**, **repetição** e o **controle local**, visam garantir a estimativa da variação individual, desconsiderando os demais fatores que podem afetar as repostas dos animais durante a condução do experimento:



Repetição das Unidades Experimentais

A aplicação de um tratamento sobre um grupo de indivíduos provoca várias respostas, para as quais precisa-se estimar uma média. Porém, dentro de cada grupo experimental existe uma variação individual entre as respostas observadas, que também precisa ser estimada. A confiabilidade dessas estimativas depende do número de vezes que a aplicação do tratamento é repetida e o número de repetições adequado varia de acordo com a variável em questão.

Em geral, quanto maior o número de repetições, maior é a exatidão das estimativas e o número de repetições pode ser aumentado com a diminuição do número de tratamentos.

Ex. Em um experimento para avaliação de 3 rações (A,B e C), a probabilidade de qualquer uma das três rações ser a melhor é igual a 1/3 (33,33%). Se houver duas repetições, essa probabilidade (1/3*1/3=1/9) é reduzida para 11,11%. E, no caso de 3 repetições, essa probabilidade é mais reduzida ainda (1/3*1/3*1/3=1/27=3,7%).

Casualização das unidades experimentais

Este procedimento visa garantir que um tratamento não será continuamente favorecido ou desfavorecido, nas sucessivas repetições, devido a fatores de variação de origem conhecida ou não. Neste caso, os tratamentos devem ser distribuídos aleatoriamente entre as unidades experimentais.

As variações que contribuem para o surgimento do erro experimental são convertidas em *variáveis aleatórias.*

As respostas biológicas são variáveis cuja magnitude depende dos indivíduos onde foram observadas. Logo, cada grupo experimental deve ter a mesma chance de receber indivíduos com variações semelhantes.

Princípio do controle local

Objetiva eliminar o erro experimental entre os grupos experimentais através da uniformização do grupo de animais das parcelas, da aplicação dos tratamentos e do local onde o experimento será conduzido.

Os grupos de animais utilizados para testar os tratamentos devem ser homogêneos quanto à idade, categoria, peso, sexo, raça ou grau de sangue, ou ainda outras características importantes, que devem ser consideradas para cada estudo em particular.

Na Figura 1 está um exemplo de experimento instalado com um grupo de animais homogêneos recebendo os tratamentos sob as mesmas condições de meio.

Figura 1. Experimento instalado com um grupo de animais homogêneos recebendo os tratamentos sob as mesmas condições de meio. (Foto: Marcos Santos/USP imagens)

Os tratamentos devem ser aplicados a cada grupo experimental de maneira semelhante para garantir as mesmas condições do agente causador de resposta sobre cada animal.

Ex_1. Se o tratamento é imposto por injeções, o grupo controle deve receber o mesmo volume de material inerte para que todos os animais de todos os grupos experimentais fiquem sujeito ao mesmo fator de estresse, o qual pode interferir na resposta individual.

No trabalho de Cunha *et al.* (2005), em que o objetivo foi avaliar em ratos os efeitos da nandrolona (esteróides anabólicos androgênicos, EAA) e do exercício aeróbio sobre o peso corporal, triglicerídeos, glicose e reservas de glicogênio, os animais foram aleatoriamente divididos em quatro grupos: 1) sedentário + veículo (SV); 2) treinado + veículo (TV); 3) sedentário + EAA (SEAA) e 4) treinado + EAA (TEAA). Nesse caso, os grupos 1 e 2 foram os grupos controles que receberam propilenoglicol (veículo) na mesma quantidade que os animais dos grupos 3 e 4 tratados com esteróides anabólicos androgênicos.

Ex_2. Se o objetivo do experimento é testar o efeito de determinada substância via oral em animais, as doses aplicadas devem ser calculadas considerando o peso vivo individual para evitar que as diferenças de peso afetem as respostas.

No trabalho de Rangel *et al.* (2005) cujo objetivo foi verificar o grau de sensibilidade dos helmintos gastrintestinais de bovinos a produtos a base de lactonas macrocíclicas, foram realizados dois testes. No primeiro, foram utilizados três grupos com 13 animais cada, utilizando diferentes produtos: T2, T3, T4 e um grupo-controle T1 com 11 animais. No segundo, foram utilizados um grupo controle T5, com 11 animais, e outros dois grupos utilizando diferentes produtos: T6 e T7, com 15 animais cada. No primeiro dia de cada teste os animais de cada grupo controle, T1 e T5, receberam tratamento placebo (solução salina) e serviram de controle negativo. No mesmo dia os animais dos outros grupos receberam os tratamentos com os produtos administrados em doses de acordo com o peso vivo.

Com relação à uniformidade do meio, muitas vezes durante a instalação de um ensaio ocorre falta de espaço para se realizarem todas as repetições necessárias dentro de um mesmo local ou os indivíduos não estão completamente disponíveis ao mesmo tempo. Nesse caso, a escolha do delineamento adequado é imprescindível para que os efeitos de local e tempo sejam estimados e as variações individuais sejam calculadas de maneira mais precisa possível.

CONCEITOS DE ESTATÍSTICA

a) População e Amostra

Amostra = número representativo de observações dentro de um universo muito maior (População).

Escolha ao acaso: todos os indivíduos da população têm a mesma probabilidade de fazer parte da *amostra populacional*

As amostras aleatórias devem:

- Apresentar médias e variações semelhantes àquelas da população
- Ser representativas para que as conclusões possam ser extrapoladas para a população

Na Figura 2 tem-se um esquema que exemplifica como uma amostra pode representar uma população.

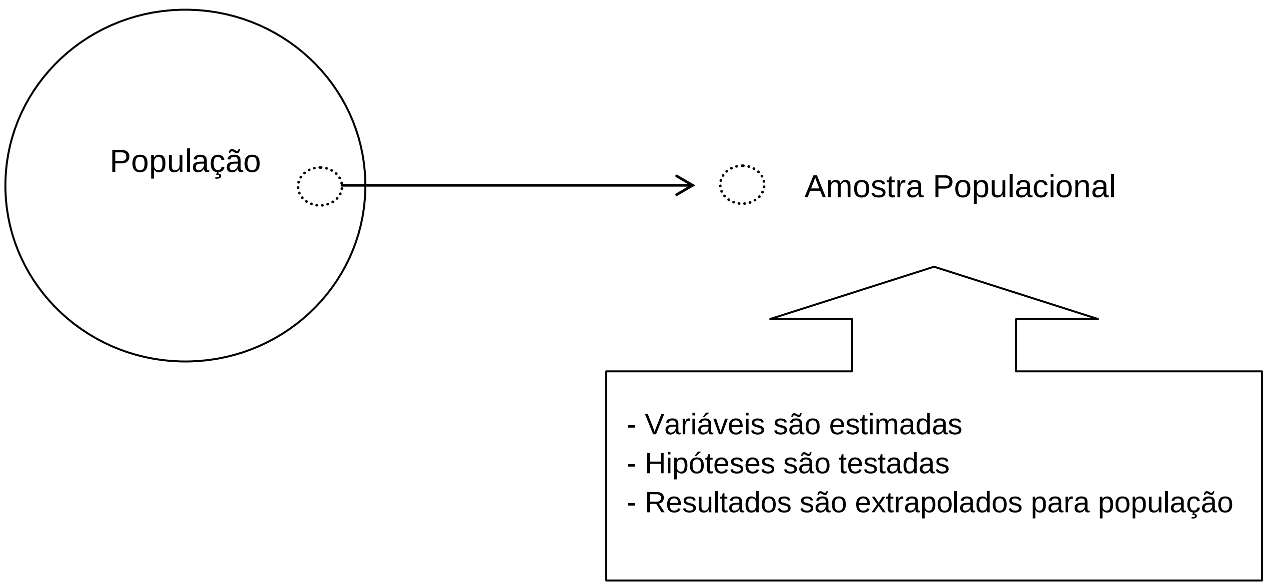


Figura 2. Esquema de uma amostra representando um universo maior (população).

b) Variável

Toda investigação ou experimento tem sempre um fato a ser avaliado, que é o objeto de estudo, o qual pode ser representado por X, Y, Z, etc.. Os valores atribuídos às variáveis são denominados de *observações* (x_1, x_2, ..., x_n).

Ex. Em um estudo sobre o comportamento de 1000 famílias de papagaios, pretende-se avaliar duas variáveis por ocasião do primeiro acasalamento: a idade do macho (X) e a atitude do macho (Y).

Variável X : 1000 observações - x_1, x_2, ..., x_{1000}

Variável Y : 1000 observações - y_1, y_2, ..., y_{1000}

Quando podem ser medidas ou estimadas, as variáveis são denominadas de *quantitativas.* No caso das variáveis que não são passiveis de mensuração e que são apenas observadas são classificadas como *qualitativas*. No exemplo acima, a variável X é quantitativa e a Y é qualitativa.

As variáveis também podem ser classificadas em *discreta* (obtida através de contagem, ex. nº de aves por galpão, nº de grãos por espiga, etc.) e *contínua* (obtida por medições ou estimativas, ex. altura, peso, comprimento, etc.).

c) Média, Desvio Padrão e Coeficiente de Variação

Na tentativa de deixar mais claro os conceitos de média, desvio padrão e coeficiente de variação, Sampaio (2007) propôs a seguinte simulação:

Supondo a existência de 3 variáveis (X, Y e Z) e três repetições (animal 1, animal 2, animal 3), de acordo com o quadro abaixo:

Variáveis	**X**	**Y**	**Z**
Animal 1	5	2	7
Animal 2	7	8	7
Animal 3	9	11	7
Total	**21**	**21**	**21**
Média	**7**	**7**	**7**

A média é equivalente para as três variáveis, porém, a variável Z apresentou respostas mais homogêneas entre os três indivíduos, sem variação alguma. Logo, a média dessa variável é a que melhor traduz a capacidade de resposta individual.

As variáveis X e Y apresentaram respostas mais instáveis e a média das observações fica mais distante daquela da população. Dessa maneira, fica clara a necessidade de se medir essa instabilidade de resposta.

A maneira mais simples de medir essa instabilidade é através da avaliação das diferenças entre a média e cada valor observado (*Desvio da Média*).

Desvios em relação à média 7

	X - 7	Y - 7	Z - 7
	5 - 7 = -2	2 - 7 = -5	0
	7 - 7 = 0	8 - 7 = 1	0
	9 - 7 = 2	11 - 7 = 4	0
Total	0	0	0

Não há desvios em Z. Pode-se observar que os desvios em X são menores que em Y. O cálculo dos desvios da média não diferencia os três grupos pelo fato do total ser o mesmo (zero). Uma alternativa para quantificar essa instabilidade é calcular a média dos quadrados dos desvios.

Média dos quadrados dos desvios

	$(X - 7)^2$	$(Y - 7)^2$	$(Z - 7)^2$
	$(-2)^2=4$	$(-5)^2 = 25$	0
	$(0)^2 = 0$	$(1)^2 = 1$	0
	$(2)^2 = 4$	$(4)^2 = 16$	0
Total	8	42	0
$(\text{desvio})^2$ médio	8/3	42/3	0/3
Desvio Médio	$\sqrt{8/3}$	$\sqrt{42/3}$	0

Este procedimento permite caracterizar grupos com diferentes instabilidades pela avaliação do desvio médio obtido pela raiz quadrada como reversão da operação anterior de elevação à potência 2, ou seja, a instabilidade média seria estimada pela equação:

$$\sqrt{\sum_{i=1}^{3} \frac{(X_i - 7)^2}{3}}$$

Através do exemplo exposto, pudemos vislumbrar a importância da estatística descritiva na organização e análises de dados observados em um experimento com três variáveis e três repetições. Agora, podemos definir média, desvio padrão e coeficiente de variação.

Média ($\overline{X}$) = valor mais provável da variável X

$$\overline{X} = \sum_{i=1}^{n} \frac{\mathrm{Xi}}{n}$$

onde *n* é o número de observações.

Pode-se notar que para se estimar o valor de X basta fazer *n* observações. Quanto maior o número de observações mais o valor de $\overline{X}$ pode ficar próximo do valor médio da população (μ).

Ex.

Espessura em micra do epitélio da mucosa vaginal de porcas no diestro:

43, 58, 17, 39, 62, 38, 23, 31, 45, 49

n = 10; $\sum X$ = 405; $\overline{X} = \frac{405}{10} = 40,5$

Desvio Padrão (s ou σ): calcula-se a partir do valor estimado da média ($\overline{X}$), obtida de uma amostra restrita. Porém, quando a amostra é abrangente e engloba todo o universo possível de respostas e, portanto, o valor da média real (μ), persiste a dedução efetuada para a avaliação da instabilidade de uma variável (σ):

$$(1) \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

Como tal situação não ocorre em situação experimental, precisamos estimar a média a partir de uma amostra restrita e, nesse caso, o valor do σ fica subestimado.

Nos estudos com amostras restritas, cada valor estimado e utilizado para outras estimativas diminui em um grau de liberdade o tamanho da amostra utilizada. Assim, a equação para estimar o desvio padrão a partir da média de uma amostra restrita fica da seguinte maneira:

$$s = \sqrt{\sum_{i=1}^{n} \frac{(X \quad - \overline{X})^2}{n-1}}$$

Como: $\overline{X} = \sum \frac{X}{n}$

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

Então: $s = \pm\sqrt{\frac{18227 - \frac{(405)^2}{10}}{9}} = 14{,}24$

Coeficiente de Variação (CV): avalia a instabilidade da variável em relação à média da população. É expresso em percentagem e pode variar de 0 (quando não houver variação nas observações) até 100% (quando *s* for tão alto quanto a média).

Os valores da média e do desvio padrão dão uma idéia da instabilidade da variável, mas não permitem a comparação de instabilidades de variáveis com unidades diferentes ou que tenham sido observadas em ensaios diferentes.

Ex.

Peso ao nascer de bezerros: $\overline{X}$ = 21 kg e s = 3 kg

Digestibilidade capim: $\overline{X}$ = 72% e s = 4%

Embora os *s* das duas variáveis pareçam próximos, o peso ao nascer é mais instável que a digestibilidade e isso pode ser comprovado calculando-se o CV:

$$CV\ (\%) = s/\overline{X}$$

$CV_{peso} = 3/21 = 14,3\%$

$CV_{digestibilidade} = 4/72 = 5,6\%$

O CV varia de acordo com o tipo de variável. De modo geral, os CVs de respostas animais oscilam entre 20 e 30%. Respostas de digestibilidade *in situ* de capim apresentam CV baixos. Porém, respostas hormonais ou imunológicas apresentam coeficientes de variação muito altos, acima de 30%.

APRESENTAÇÃO DOS DADOS

Em qualquer estudo, as observações (valores, dados) devem ser organizadas para melhor visualização da variável em estudo. Essa organização é denominada de série estatística e pode ser apresentada em tabelas e gráficos.

Os dados sobre os quais se tiram conclusões podem ser de dois tipos: de enumeração ou de medição. A coleta dos mesmos deve ser a mais rigorosa possível, visto que a coleta mal feita de um ou mais dados pode levar a erros de interpretação dos resultados e a falsas conclusões.

Considerando as etapas de um ensaio:

Coleta de dados (+ rigorosa possível)
↓
Ordená-los de acordo com critérios fundamentados → Distribuição dos dados
↓
Testes das diferenças
↓
Conclusões

Tipos de distribuição

Ex.

Lote de 300 suínos com peso de abate variando de 63 a 117 kg; $\overline{X}$ = 90 kg e s = 12 kg

Nesse intervalo, qual a frequência de ocorrência dos demais valores contidos dentro dele?

Estudo de distribuição de frequências:

N^{o} classes = 2,5 $\sqrt[4]{n}$ = 2,5 $\sqrt[4]{300}$ = 11 classes

Intervalo de Classes (IC) = amplitude dos valores observados / n^{o}. ideal de classes

IC = (117 – 63) / 11 = 54/11 = 4,909 = 5

$$\left.\begin{array}{l}\text{Amplitude do gráfico} = 5 \times 11 = 55 \\ \text{Amplitude observada} = 117 - 63 = 54\end{array}\right\} \text{Diferença de 1 unidade}$$

Essa diferença deve ser partilhada entre os valores mínimos e máximos:

½ = 0,5

$V_{mín.} = 63 - 0,5 = 62,5$ e $V_{max.} = 117 + 0,5 = 117,5$

Dessa forma, a frequência observada das diferentes classes de peso ao abate de suínos pode ser organizada como na Tabela 1 e na Figura 3:

Tabela 1. Frequência observada (%) das diferentes classes de peso ao abate (Kg) de suínos.

Classes de peso (kg)	Frequência Observada (%)
62,5 a 67,5	06
67,5 a 72,5	12
72,5 a 77,5	25
77,5 a 82,5	38
82,5 a 87,5	44
87,5 a 92,5	52
92,5 a 97,5	47
97,5 a 102,5	34
102,5 a 107,5	23
107,5 a 112,5	14
112,5 a 117,5	05

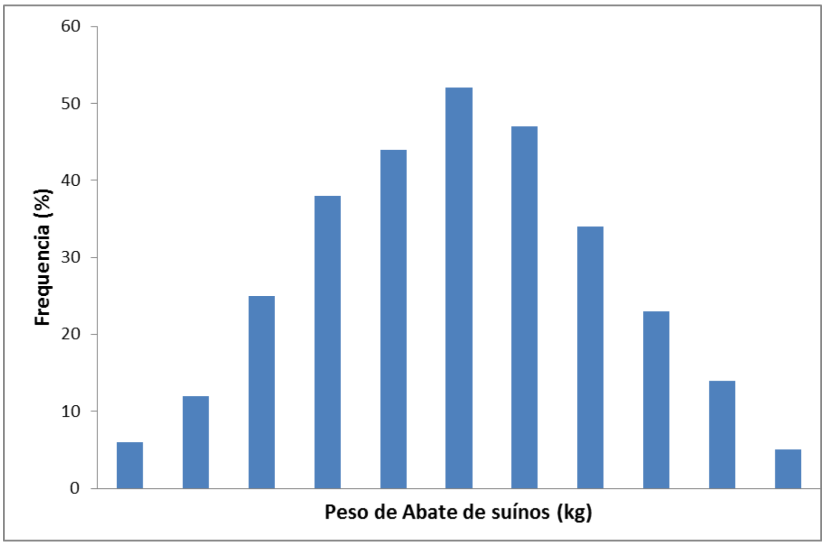


Figura 3.Frequência observada segundo a classe de peso ao abate de suínos (*Fonte: Sampaio, 2007*).

Diante de uma amostra infinitamente grande, esses dados podem ser ajustados de acordo com a função de distribuição normal (Gauss) (Figura 4):

$$Y_i = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x_i - \overline{x})^2}{2s^2}}$$

Onde:

Y_i = ordenada vertical de X_i (peso), que depende do seu valor médio μ

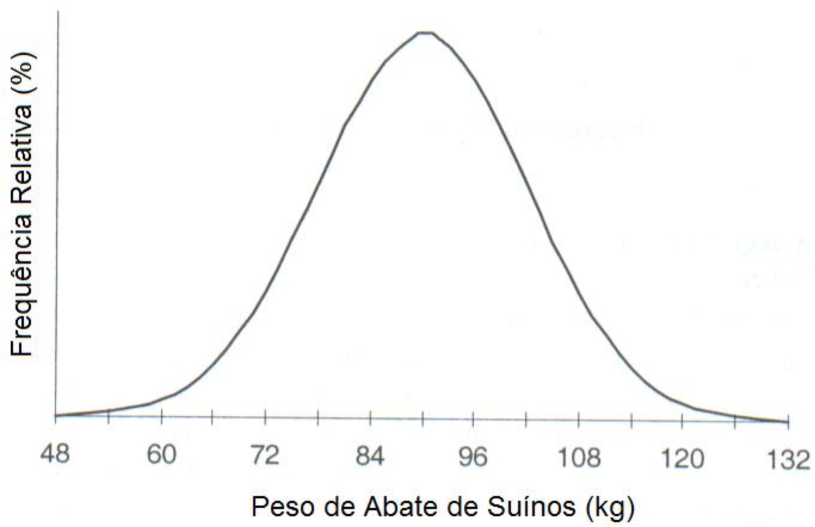


Figura 4. Distribuição de frequência da variável peso ao abate de suínos, com média de 90 kg, desvio padrão de 12 kg e n infinito (*Fonte: Sampaio, 2007*)

A importância dessa distribuição de dados de acordo com a equação da curva normal consiste no fato da mesma permitir que se identifique junto a uma população infinita um intervalo de valores que seja de nosso interesse.

a. Medidas de Posição

A partir das medidas de posição pode-se calcular um valor central onde se acumulam a maioria dos dados observados.

Moda

- Valor que ocorre com maior frequência

- Serve para se determinar o valor correto depois de uma série de repetições

- Pode ser calculado para qualquer tipo de variável (quantitativa e qualitativa)

Mediana

- Valor que não excede e não é excedido por mais da metade das observações
- Indicada quando existem dados discrepantes na amostra
- Para calculá-la deve colocar os valores em ordem crescente e, em seguida, escolher o valor central. Em caso de número par de observações, deve-se achar a média dos dois valores centrais.

  Ex.

  5 alunos – 5 notas (9, 8, 0, 9, 10) - $\overline{X}$ = 36/5 = 7,2

  Em ordem crescente: 0 – 8 – $\underline{9}$ – 9 – 10

  Mediana = 9

Média

- É a mais simples das representações quantitativas
- Desvantagem: é muito influenciada pelos valores mínimos e máximos
- Representada por $\overline{X}$
- Quando os valores são observados com frequências diferentes, a equação para calculá-la é a seguinte:

$$\overline{X} = \sum_{i=1}^{n} f_i X_i / \sum_{i=1}^{n} f_i$$

Ex.

Valor	Frequência
5	3
8	2
6	4
2	1

$$\overline{X} = \frac{3x5 + 2x8 + 4x6 + 1x2}{3 + 2 + 4 + 1} = 5{,}7$$

b. Medidas de dispersão

Análise Completa dos dados →

- Apresentação
- Medidas de posição

} - Inadequado
- Dados diferem entre si (variabilidade)

- Medidas de dispersão

↓

Informa o grau de variação dos dados

Amplitude total = diferença entre os valores máximo e mínimo

- Rápida e de fácil mensuração
- Pouco informativa
- Aumenta com o maior número de observações (n)

Ex. Peso ao nascer de leitões

Leitegada 1	Leitegada 2
0,9	1,1
1,2	1,1
1,2	1,2
1,3	1,3
1,4	1,4
1,4	1,5
1,6	1,5
X= 1,3	X= 1,3

Leitegada 1: Amplitude variação 1,6 – 0,9 = 0,7

Leitegada 2: Amplitude variação 1,5 – 1,1 = 0,4

Variância (s^2): é a soma dos quadrados dos desvios em relação à média que mede a dispersão dos dados em torno da média

Como na prática a μ é uma estimativa ($\overline{X}$), o número de observações perde 1 grau de liberdade,

então: $s^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 / n-1$

Quando as observações ocorrem com diferentes frequências ($f_{i,}$):

então: $s^2 = \sum_{i=1}^{n} f_i (X_i - \overline{X})^2 / \sum_{i=1}^{n} f_i - 1$

$$\text{Variância} \begin{cases} \text{População: } \sigma^2 \\ \text{Amostra: } s^2 \end{cases}$$

A expressão $\sum_{i=1}^{n}(x_i - \overline{x})^2$ pode ser transformada:

Soma dos dados observados

$$\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}$$

Soma dos quadrados individuais

Correção (C)

Desvio Padrão (s) = raiz quadrada da variância que, ao contrário dessa, é expressa na mesma unidade das observações

A medida que o número de observações (r) diminui, o desvio padrão (s) aumenta (Figura 5) :

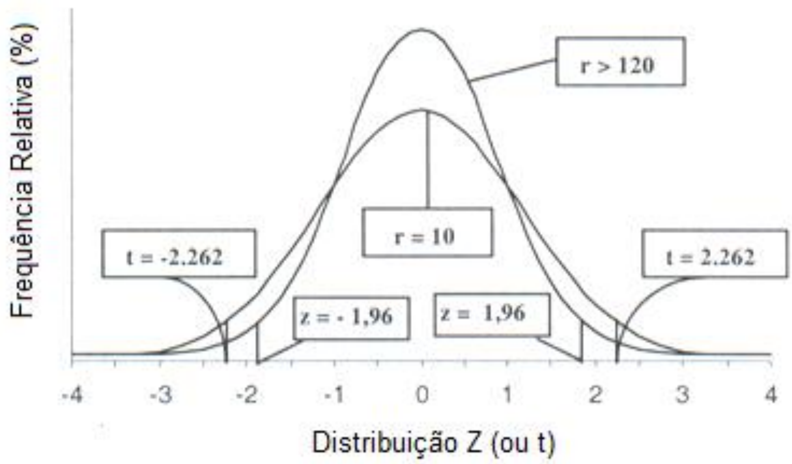


Figura 5. Relação entre o número de observações e o desvio padrão (*extraído de Sampaio, 2*)

Covariância (Cov) = mede o quanto que duas características variam juntas. Pode uma medida positiva ou negativa.

$$\text{Cov (x,y)} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

c. Associação de Variáveis Quantitativas

Quando várias respostas são tomadas simultaneamente de um mesmo indivíduo, em alguns casos, pode-se observar através de análise estatística uma associação entre algumas variáveis.

Correlação

Quando a associação entre duas variáveis se manifesta sem que seja possível estabelecer efeito causativo de uma delas sobre a outra, não existe explicação biológica para a associação. A associação entre altura da cernelha com o comprimento do pescoço em equinos aparentemente não mantém uma relação previsível.

O estudo de correlação entre duas variáveis é útil quando se deseja saber se a variação de uma delas acompanha proporcionalmente ou inversamente à variação da outra. Um exemplo seria a prova de ganho de peso que é realizada para testar reprodutores bovinos através de ensaios de desempenho utilizando os filhos dos reprodutores testados em confinamento durante um dado período. Para verificar se os novilhos apresentariam mesmo desempenho em sistema de pastagem, foi realizado um estudo de correlação em que pares de irmãos, filhos dos touros testados, foram divididos entre duas provas, uma realizada em confinamento e a outra em pastagem. Como o desempenho do novilho em confinamento não depende do desempenho de seu irmão no campo (e vise-versa), se os resultados mostrarem-se correlacionados, significa que os dois tipos de prova são adequados para avaliar o desempenho de filhos de reprodutores.

O grau de correlação poderia ser medido através do conceito de covariância, cuja equação reduzida é:

$$\mathrm{Cov}_{(x,y)} = \sum_{i=1}^{n} XY - \frac{\sum_{i=1}^{n} X - \sum_{i=1}^{n} Y}{n}$$

A desvantagem dessa equação é que ela é dada nas mesmas unidades de X e Y. Como a medida da correlação precisa ser adimencional, esta pode ser estimada através do *coeficiente de correlação (r),* que é o quociente entre a Cov $_{(x,\ y)}$ e o produto dos desvios padrão de x e y:

$$r_{x,\ y} = \frac{\mathrm{Cov}_{(x,y)}}{\sqrt{s_x^2 \cdot s_y^2}}$$

Sabendo-se que:

$$s^2 = \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}$$

Com a substituição dos valores correspondentes:

$$r_{x,\,y} = \frac{\sum XY - \sum X \cdot \sum Y n}{\sqrt{[\sum X^2 - (\sum X)^2 n] \cdot [\sum Y^2 - (\sum Y)^2 n]}}$$

Como o coeficiente de correlação (r) varia de +1 a -1, essa medida é mais adequada para avaliar o grau de relacionamento linear entre duas variáveis quantitativas do que a covariância. Portanto, a correlação pode ser positiva, negativa ou nula. A correlação é positiva quando o aumento de uma variável corresponde ao aumento de outra (ex. produção de leite x produção de gordura); é negativa quando o aumento de uma variável corresponde à diminuição da outra (ex. produção de leite x % de gordura); e é nula quando as características não se associam. O r expressa o grau de aproximação dos pontos à reta no diagrama de dispersão. Logo, define o grau de fidelidade com que a reta descreve a relação entre duas variáveis.

Se duas variáveis X e Y não estão correlacionadas (r = 0), os valores de X variam independentes de Y e, vice-versa. Graficamente, se cada eixo representar uma variável, os pontos se encontrarão dispersos por todo o quadrante (Figura 6).

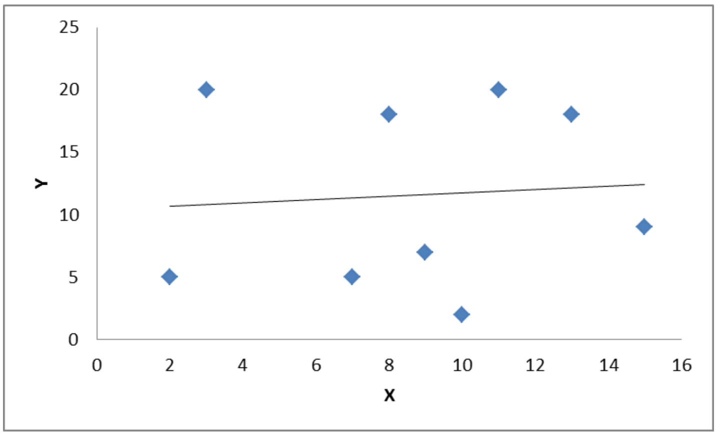


Figura 6. Representação gráfica de duas variáveis (X e Y) que não estão correlacionadas (r =0).

Uma correlação alta negativa (r = -0,95) mostrará uma dispersão inversa entre X e Y com os maiores valores de X correspondendo aos menores de Y, concentrados em torno de uma reta descendente na diagonal (Figura 7).

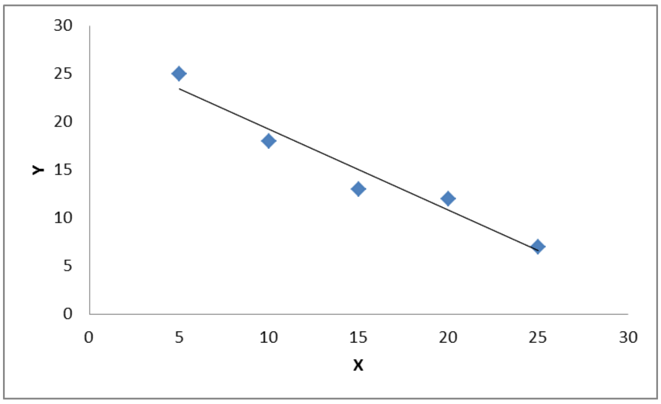


Figura 7. Representação gráfica de duas variáveis (X e Y) que estão correlacionadas negativamente (r = - 0,95).

Se X e Y estiverem completamente e positivamente correlacionados (r = 1), os pontos estarão alinhados na diagonal sob uma reta imaginária ascendente (Figura 8).

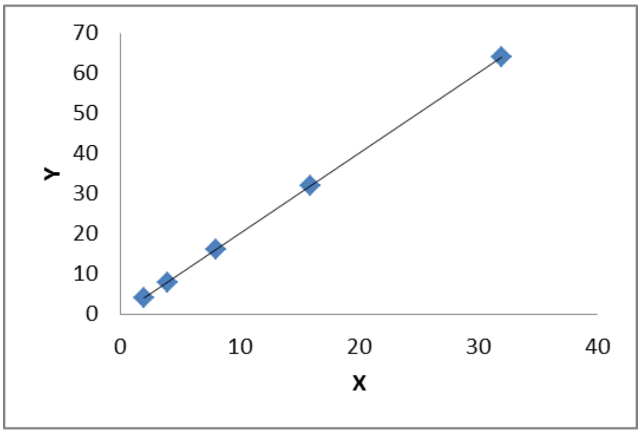


Figura 8. Representação gráfica de duas variáveis (X e Y) que estão correlacionadas positivamente (r = 1).

No caso de X e Y estarem levemente correlacionados (r = 0,30), a dispersão dos pontos não é tão dispersa e nem tão próxima da reta imaginária (Figura 9).

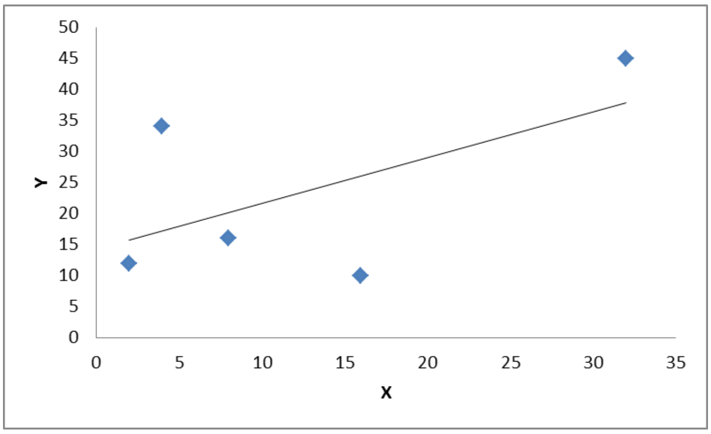


Figura 9. Representação gráfica de duas variáveis (X e Y) que estão levemente correlacionadas (r = 0,30).

A percepção e a interpretação da correlação entre duas variáveis são importantes ferramentas no planejamento de um ensaio balanceado. Neste caso, os fatores testados não devem se associar (r = 0) para que os efeitos dos mesmos não sejam confundidos. Por exemplo, em um delineamento inteiramente casualizado com quatro tratamentos em fatorial 2x2 e 5 repetições, o planejamento poderia ser feito da seguinte maneira:

Tratamento	Proteína (%)	Fibra (%)
1	16	9
2	16	12
3	18	9
4	18	12

Para cada um dos valores de proteína (P) e fibra (F) dos quatro tratamentos estarão repetidos cada um cinco vezes (n=20). A correlação entre P e F seria:

$$r = \frac{3570 - 340x210/20}{\sqrt{(5800 - 340^2/20)(2250 - 210^2/20)}}$$

Já que:

$$\sum X = 340$$

$$\sum X^2 = 5800$$

$$\sum XY = 3570$$

$$\sum Y = 210$$

$$\sum Y^2 = 2250$$

Regressão Linear

Quando a associação entre duas variáveis se manifesta por dependência clara de uma variável em relação à outra. Nesse caso, o efeito do agente causativo (variável independente) é estudado para verificar a variação linear na resposta observada (variável dependente). Por exemplo, o peso abate de bovinos depende do nível protéico da dieta, a qualidade do ovo decresce com o tempo de estocagem, a produção diária de leite aumenta com a idade da vaca até 6 – 7 anos e depois decresce. Um exemplo em genética quantitativa é a regressão do valor gênico do indivíduo em relação ao seu fenótipo (herdabilidade).

Todos esses exemplos têm seus próprios padrões de variação expressos por representações gráficas e equações de retas ascendentes ou descendentes ou parabólicas.

A análise de variância, em geral, pressupõe a independência entre os diversos tratamentos testados. Porém, quando esta hipótese não se verifica (tratamentos não são independentes entre si), a análise de variância não é válida. Existem casos em que os diferentes tratamentos variam apenas no nível do fator que se deseja testar. Por exemplo, em um ensaio que se deseja avaliar o desempenho de bovinos alimentados com a mesma ração mas com diferentes níveis de cana de açúcar picada (25, 50 e 75%) em substituição a silagem de milho espera-se observar a diminuição do desempenho (peso ao abate, em kg) com o aumento da cana na dieta com, concomitante, diminuição nos custos com alimentação. Esta dependência entre desempenho e nível de cana na dieta pode ser matematicamente definida pela equação:

$$\hat{y} = a + \mathrm{b}x$$

Onde:

$\hat{y}$ = estimativa do desempenho (peso ao abate) dos bovinos alimentados com ração contendo diferentes níveis de cana picada

a = coeficiente linear de regressão (valor de y quando x =0)

b = coeficiente de regressão da % de cana sobre a resposta y (peso ao abate)

O estudo deste modelo permite:

a) Conhecer a relação matemática entre as variáveis. Dessa maneira, é possível estimar valores da variável dependente em níveis não estudados da variável independente. Por exemplo, o peso ao abate poderá ser estimado para níveis de 25 e 75% de cana na dieta.
b) Avaliar como a variável independente (% de cana) influencia a resposta medida (peso ao abate) através do coeficiente de regressão (*b*) que define a variação em *y* para cada unidade de *x.*
c) Verificar se a resposta observada é realmente uma função linear da variável independente ou se é uma função mais complexa.

Interpretação dos parâmetros da equação de regressão

Se no exemplo citado o modelo matemático de $\hat{y}$, que é a estimativa do peso ao abate (kg) do bovino alimentado com ração contendo *x*% de cana, tivesse sido o seguinte:

$$\hat{y} = 550 - 2X$$

Então, poderíamos interpretar os parâmetros da seguinte maneira:

a = 550 -> para o nível zero de cana o peso ao abate seria de 550 kg. O valor de *a* neste caso não permite uma explicação biológica, ele apenas corresponde ao ponto de interseção que a reta $\hat{y}=a+bx$ apresenta com o eixo vertical *y.*

b = - 2 -> para cada 1% de cana na ração o desempenho potencial do animal diminui (o valor é negativo) 2 kg no peso final. Se o valor de *b* tivesse sido - 4, para cada 1% de aumento no teor de cana haveria redução de 4 kg no peso de abate. Isto significa que quanto maior o valor de *b* maior é a influência de *x* sobre *y* e maior é a inclinação da reta traçada no gráfico.

Quando b =0 e $\hat{y} =$ a, nenhuma inclinação é observada, definindo uma reta paralela ao eixo horizontal *x,* o que significação que não há associação entre as duas variáveis e *x* é independente de *y*.

Estimativas de desempenho de bovinos poderão ser feitas com segurança utilizando níveis de cana dentro daquele intervalo estudado ou até utilizando níveis próximos do intervalo, como por exemplo:

x= 20%, $\hat{y} =$ 550 – 2 (20) = 510 kg

x = 80%, $\hat{y} =$ 550 – 2 (80) = 390 kg

x = 100%, $\hat{y} =$ 550 – 2 (100) = 390 kg

Os parâmetros *a* e *b* podem ser tanto positivos quanto negativos. Os valores de *b* positivos traduzem uma resposta crescente, enquanto os negativos refletem resultados decrescentes de *y* em relação à *x*.

Estimativa dos parâmetros da equação de regressão linear

Considerando os resultados do estudo da associação do peso a desmama em relação ao número de leitões por leitegada:

No. Leitegada	X	Y	X^2	Y^2	XY
1	5	60	25	3600	300
2	6	72	36	5184	432
3	4	50	16	2500	200
4	8	84	64	7056	672
5	10	108	100	11664	1080
6	7	74	49	5476	518
7	6	62	36	3844	372
8	7	70	49	4900	490
Total	53	580	375	44224	4064
	$\sum X$	$\sum Y$	$\sum X^2$	$\sum Y^2$	$\sum XY$

Pergunta-se:

a) Qual é a variável independente? e a dependente?
A variável independente (x) é o número de leitões desmamados e a variável dependente (y) é o peso da leitegada.

b) Qual é a equação que permite estimar o peso da leitegada a desmama a partir do número de leitões desmamados?

- Cálculo do coeficiente de regressão:

$$b_{x,\,y} = \frac{\sum XY - (\sum X \cdot \sum Y)/n}{\sum X^2 - (\sum X)^2/n} = \frac{4064 - (53x580)/8}{375 - (53)^2/8} = \frac{4064 - 3842{,}5}{375 - 351{,}1} = +9{,}277$$

- Cálculo do coeficiente de linearidade (*a*):

$$a = \bar{y} - b\bar{x} \Rightarrow a = \frac{580}{8} - 9{,}277 \times \left(\frac{53}{8}\right) = 11{,}037$$

A equação de regressão linear, bem como o coeficiente de determinação da curva (R^2) desses dados são mostrados na Figura 10.

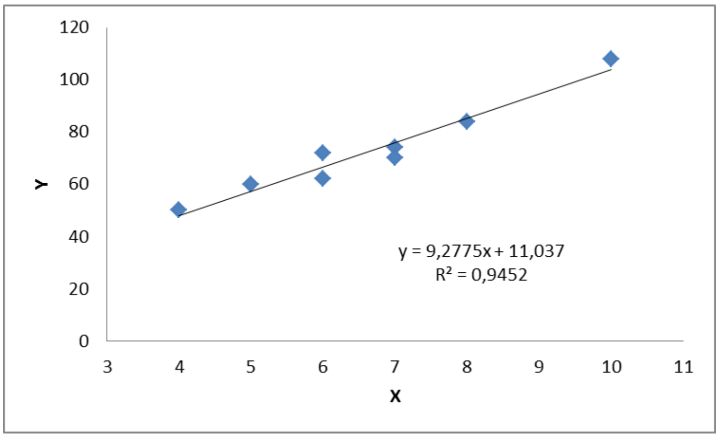


Figura 10. Representação gráfica da associação do peso a desmama em relação ao número de leitões por leitegada.

d. Análise de Variância

O estudo da variância (s^2) entre indivíduos é fundamental em qualquer investigação científica. No entanto, as situações experimentais envolvem muitos fatores como, por exemplo, diferenças de idade e sexo entre os animais da amostra; diferenças de instalações e de período de avaliação, entre outros. Estes fatores quando não identificados e não controlados pelo uso de um delineamento experimental adequado se incorporam na estimativa da variância (s^2) individual.

O objetivo da análise de variância é controlar os efeitos dessas fontes de variação para assegurar que o valor estimado da variância entre os indivíduos corresponda àquele valor sem a interferência de fatores externos que podem levar a superestimativa. No entanto, para que análise de variância seja adequada, dois requisitos básicos precisam ser atendidos:

a) A resposta que está sendo analisada deve ser uma *variável com distribuição normal.*
b) Os tratamentos avaliados devem apresentar variâncias iguais. Este princípio reconhece que a instabilidade de uma variável não depende do grupo experimental onde ela está sendo medida.

Ex.

Cinco famílias de reprodutores White Leghorn escolhidos ao acaso foram avaliadas através dos pesos das progênies à 8ª. semana de idade (em g):

Reprodutores				
A	B	C	D	E
687	618	618	600	617
691	680	687	657	658
793	592	763	669	674
675	683	747	606	611
700	631	678	718	678
753	691	738	693	788
704	694	731	669	650
717	732	603	648	690
5720	5321	5565	5260	5366

1º.) Determinar a soma dos quadrados total (SQT) para medir a variação total dos dados:

$$\text{SQT} = \sum x^2 - \frac{(\sum x)^2}{n}$$

Fator de Correção (FC):

$$\text{FC} = \frac{(\sum x)^2}{n} = (5720 + 5321 + 5565 + 5260 + 5366)^2/40 = 18.539.545$$

$$\sum x^2 = 687^2 + 691^2 + ... + 690^2 = 18.641.548$$

SQT = 18.539.545 – 18.641.548 = 102.003

No. total de observações = 5 x 8 = 40

No. graus de liberdade da SQT = 40 -1 = 39

2º.) Determinação da soma de quadrados entre reprodutores (SQE), que é reflexo da variação de cada reprodutor em relação à média geral

$$SQE = (\frac{5720^2}{8} + ... + \frac{5366^2}{8}) - FC$$

SQE = 18.557.777,75 – 18.539.545 = 18.232,75

No. g. l. = 5-1 = 4

3º.) Determinação da soma de quadrados dentro de reprodutores (SQD)

SQD = SQT – SQE = 102.003 – 18.232,75 = 83.770,245

No. g.l. resíduo = 39-4 = 35

4º.) A divisão dessas somas de quadrados pelos seus respectivos graus de liberdade leva às variâncias correspondentes. Estas são conhecidas por quadrados médios.

$$QME = \frac{SQE}{gl} = \frac{18.232,75}{4} = 4.558,19$$

$$QMD = \frac{SQD}{gl} = \frac{81.687}{35} = 2.393,43$$

5º.) Comparar as variâncias pelo teste de Fisher (F)

$$F = \frac{QME}{QMD} = \frac{4.558,19}{2.393,43} = 1,9$$

Ftab (g.l. 35 x 4) = 2,65 (nível de 5% de significância)

Fcalc (1,9) < Ftab (2,65): Não há diferença estatisticamente significativa, ao nível de 5% de probabilidade, entre as médias de peso medido na progênie com 8 semanas de idade de 5 reprodutores escolhidos ao acaso.

Tabela de Análise de Variância

Fatores de Variação	G.L.	SQ	QM	F
Entre reprodutores	4	18.232,75	4.558,19	1.9^{ns}
Dentro de reprodutores	35	83.770,245	2.393,43	
Total	39	10.2003		

CONSIDERAÇÕES FINAIS

Tão importante quanto o planejamento de um estudo envolvendo a experimentação com animais, é entender os fundamentos das ferramentas estatísticas para auxiliar não só na análise, mas também na interpretação dos dados observados. Apesar de existirem inúmeros softwares disponíveis para realização de análises estatísticas, os princípios e fundamentos são basicamente os mesmos e conhece-los é de suma importância para os estudos científicos. No caso da experimentação zootécnica, os cuidados a serem tomados podem variar de acordo com a espécie animal com que se trabalha, principalmente para atender os princípios do controle local. De qualquer forma, mesmo após um cuidadoso planejamento e execução de um estudo, as análises estatísticas e a forma de apresentação dos dados irão auxiliar na publicação dos mesmos em revistas de impacto.

LITERATURA CITADA E CONSULTADA

BARBIN, D. **Componentes de Variância: Teoria e Aplicações**. 2.ed. Piracicaba: FEALQ, 1993. 120p.

CUNHA, T.S.; TANNO, A.P.; MOURA, M.J.C.S; MARCONDES, F.K. Relação entre a administração de esteróide anabólico androgênico, treinamento físico aeróbio e supercompensação do glicogênio. **Revista Brasileira de Medicina do Esporte.** v. 11, n.3, p.187-192, 2005

DAL'COL LÚCIO, A.; LOPES, S. J.; STORCK, L.; CARPES, R. H.; LIEBERKNECHT, D.; NICOLA, M. C. Características Experimentais das Publicações da Ciência Rural de 1971 a 2000. **Ciência Rural**, v.33, n.1, p. 161-4, 2003.

PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 15ed. Piracicaba:FEALQ, 2009.451p.

RANGEL, V.B.; LEITE, R.C.; OLIVEIRA, P.R.; SANTOS JR., E.J. Resistência de *Cooperia* spp. e *Haemonchus* spp. às avermectinas em bovinos de corte. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.57, n.2, p.186-190, 2005

SAMPAIO, I. B. M. **Estatística Aplicada Á Experimentação Animal.** 3ed. Belo Horizonte:FEP MVZ Editora, 2007. 265p.

VOLPATO, G. L. **Ciência: da Filosofia à Publicação**. 2ed. Jaboticabal: FUNEP, 2000. 234p.