

# Automação da Representação de Resultados de Classificação Linear para Características de Tipo

Leonardo Carvalho Nápolis Costa<sup>1,\*</sup>, Fernanda Nascimento Almeida<sup>1,\*</sup>,  
Marcos Vinícius G. Barbosa da Silva<sup>1</sup>, Rui da Silva Verneque<sup>1</sup>, Wagner Arbex<sup>1,†</sup>

<sup>1</sup> Empresa Brasileira de Pesquisa Agropecuária – Embrapa  
R. Eugênio do Nascimento, 610 – 36033-380 – Juiz de Fora, MG – Brasil

{leonardocnc, almeida.fn}@gmail.com,  
{marcos.vb.silva, rui.verneque, wagner.arbex}@embrapa.br

**Abstract.** This paper describes the results from an automated graphics generator used to represent data evaluation of linear type traits. The aim was to generate boxplot graphs to assist in the readability of the results for animal breeding programs of Embrapa Dairy Cattle that includes the breeds Dairy Gyr and Gyr-Holstein. This procedure was developed using the R mathematical package, which enables iterative process automation and optimization of different types of tasks. Thus, it avoids manual intervention directly on the data for the creation of graphics which are published in their respective “Sire Summary”.

**Resumo.** Este artigo descreve os resultados obtidos a partir da geração automatizada de gráficos para representação de avaliação de características lineares de tipo. O intuito foi a geração de gráficos do tipo boxplot para auxiliar a legibilidade dos resultados dos programas de melhoramento genético da Embrapa Gado de Leite, das raças Gir Leiteiro e Girolando. Este procedimento foi desenvolvido com o pacote matemático R, permitindo a automação de processos iterativos e a otimização de diferentes tipos de tarefas. Assim, evita-se a intervenção manual direta sobre os dados para a criação dos gráficos que são publicados nos respectivos “Sumários de Touros”.

## 1. Introdução

Os programas de melhoramento genético da Embrapa Gado de Leite fazem a avaliação de animais para a produção de leite e de outras características de importância econômica. Estes programas buscam, a cada ano, incorporar novos e importantes aspectos de avaliações, resultados, tecnologias e informações que beneficiem os criadores [de Freitas et al. 2013]. Para tanto, os programas de melhoramento genético animal necessitam de meios que auxiliem no processo de representação de seus resultados. Neste sentido, o uso do software R tem se apresentado a opção mais adequada permitindo, por exemplo, a automação dos processos iterativos e a otimização de outras tarefas.

Neste sentido, foi possível implementar, com o uso do R, um procedimento padrão para a estruturação de gráficos advindos dos resultados dos testes dos animais analisados e posterior publicação no chamado “Sumário de Touro”, que traz os resultados da avaliação

\*Bolsista da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG.

†Autor correspondente.

SP 6174  
8/203

genética de diferentes raças bovinas. Cada programa de melhoramento genético, em geral, atende à uma raça em específico, desenvolvendo suas ações de pesquisa e seu sumário, sendo este individualizado para cada raça.

Como muitos desses resultados são apresentados graficamente, é imprescindível aprimorar os processos de geração e apresentação de seus gráficos. Com isso, além de melhorar a estrutura e o modelo dos gráficos, fornecendo uma visualização mais precisa da informação representada também, apresenta uma alternativa que diminua significativamente o tempo de realização dessa tarefa.

Esse artigo tem como objetivo discutir a utilização do software R para a geração automática de gráficos, evitando-se a intervenção manual direta sobre os dados para a criação desses gráficos, para suas publicações nos respectivos “Sumários de Touros” das raças Gir Leiteiro e Girolando e, ainda, a importância da legibilidade de tais gráficos para a interpretação dos resultados.

## 2. Materiais e Métodos

As características atualmente avaliadas nos programas de melhoramento genético das raças Gir Leiteiro e Girolando são: altura da garupa, perímetro torácico, comprimento do corpo, comprimento da garupa, largura entre os ísquios, largura entre os ílios e ângulo da garupa no grupo das características corporais. Além dessas, no grupo de pernas/pés avaliam-se o ângulo de cascos, posição das pernas, no sistema mamário, o úbere anterior, úbere posterior, profundidade de tetas, comprimento de tetas, diâmetro de tetas, facilidade de ordenha, temperamento e comprimento do umbigo. A classificação linear é a metodologia aplicada para a avaliação das medidas de conformação dos animais utilizada pelas associações de criadores de gado leiteiro, que agrupa as características como corporais, pernas/pés e sistema mamário. Neste sentido, a análise exploratória de dados envolve o uso de técnicas de estatística para identificar padrões que podem estar escondidos em um grupo de números.

Para a representação gráfica dos resultados da avaliação das características lineares dos bovinos animais, utilizaram-se, nesta abordagem, o software R, o gráfico do modelo *boxplot* e os próprios dados da avaliação de características lineares em bovinos. Com isso, o R é um software “livre” (*free software*), inicialmente desenvolvido para uso em trabalhos estatísticos e matemáticos mas, extrapolou essa expectativa e hoje é multiplataforma [R Core Team 2013]. Portanto, possui versões para diversos sistemas operacionais, tais como, GNU/Linux, Microsoft Windows e OS X.

O software R integra rotinas prontas como também, permite que novas rotinas sejam programadas para manipulação de dados, possibilita que sejam feitos inúmeros cálculos matemáticos e estatístico, dos mais diversos graus de complexidade, e a geração de gráficos, tanto por meio dos modelos já disponíveis, pela adaptação desses mesmos modelos ou, ainda, pelo desenvolvimento de novos modelos. Além disso, o uso do R permite: a manipulação e armazenamento efetivo dos dados; o uso de operadores para cálculo sobre variáveis indexadas e cálculo matricial; a análise de dados com o uso de uma integrada coleção de ferramentas; a análise exploratória de dados pela visualização em tela ou por meio de impressão, com o uso de recursos gráficos; e, o desenvolvimento de rotinas com o uso de uma linguagem de programação robusta, simples e eficiente,

que inclui estruturas condicionais, estruturas cíclicas, funções recursivas, e capacidade de entrada e saída de dados etc..

O R tem seu próprio ambiente de processamento e, muitas vezes, recebe essa designação caracterizando-o como um sistema completamente planeado e coerente, e não apenas como um conjunto ampliado de ferramentas específicas e rígidas, como é o caso de outros softwares de análise de dados [R Core Team 2013]. Possui procedimentos estatísticos em inúmeros de pacotes livremente disponíveis na Internet e que podem ser carregados de acordo com a necessidade.

A análise exploratória de dados envolve o uso de técnicas de estatística para identificar padrões que podem estar escondidas em um grupo de números. Os gráficos são frequentemente usados como ilustrações de dados e não como ferramentas de raciocínio para aprender algo novo na esfera do contexto, adquirir informação nova ou aprender a partir dos dados [Wild and Pfannkuch 1999] [Konold and Pollatsek 2002].

Um importante diagrama comumente usado em análises estatísticas para comparação de dois ou mais conjuntos de dados correspondentes às categorias de uma variável qualitativa é o *boxplot* (gráfico de caixa) [Tukey 1977]. Este diagrama é usado para avaliar a simetria dos dados, sua dispersão e a existência ou não de valores atípicos – ou *outliers* – no conjunto de dados. No *boxplot* os resultados são apresentados por meio de um diagrama de caixa, que capta importantes aspectos de um conjunto de dados por meio do resumo de cinco valores estatísticos: o valor mínimo, primeiro quartil, mediana ou segundo quartil, terceiro quartil e o valor máximo. Pode ainda ser utilizado para uma comparação visual entre dois ou mais grupos. Por exemplo, duas ou mais caixas são colocadas lado a lado e se compara a variabilidade entre elas, a mediana e assim por diante. O uso do *boxplot* têm sido considerado uma poderosa ferramenta que pode auxiliar melhorar o raciocínio sobre um informação quantitativa.

Neste trabalho, foi utilizado uma variante do *boxplot* onde os quartis não foram marcados. Assim, desprezou-se o primeiro e o terceiro quartis e a avaliação das características lineares foi representada pelos valores mínimo, médio e máximo dos valores típicos.

Estudos mostram que há influência dos fatores de meio ambiente como rebanho, ano e estação de classificação, estágio da lactação, idade ou ordem do parto, classificador, entre outros, sobre as características lineares de tipo. Assim, identificar e quantificar os fatores de meio ambiente que influenciam as características consideradas na classificação linear para tipo podem, por meio da seleção de animais funcionais e produtivos, auxiliar no aumento da produção leiteira [Esteves et al. 2004].

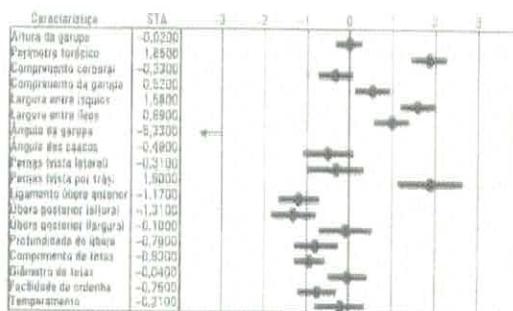
### **3. Resultados e Discussão**

A representação de qualquer conteúdo por meio de um gráfico requer que dados sejam transformados em pontos que, por sua vez, são distribuídos entre os eixos da dimensão em que se quer trabalhar. Sob esse aspecto, em geral, esses pontos são posicionados no plano cartesiano e assim permitem a interpretação de alguma informação, inferida a partir da leitura do gráfico. No entanto, sua interpretação será tão útil e atrativa quanto forem precisas as marcações dos pontos na estrutura do gráfico e, além disso, deve-se primar pela simplicidade no modelo do gráfico, pois, uma imagem com abundância de informações gráficas, pode prejudicar o seu entendimento.

Além disso, existem vários tipos, estruturas e modelos de gráficos, que se diferem por sua adequação ao representarem diferentes resultados de diferentes estudos e experimentos. Também, sua construção deve ser factível e rápida, dentro da necessidade da tarefa a ser realizada.

Os primeiros dados relativos às características lineares para avaliação de tipo, nos programas de melhoramento genético das raças Gir Leiteiro e Girolando, começaram a ser mensurados e armazenados, respectivamente, em meados dos anos 90 e na segunda metade da última década. A partir desses, as avaliações para essas características começaram a ser publicadas, o que exigiu uma representação gráfica, e uma representação adotada internacionalmente é o uso de gráficos, como o *boxplot*.

A construção do *boxplot* exige diversos critérios e, portanto, faz com que sua construção se torne um pouco trabalhosa. Inicialmente, os gráficos eram gerados com o uso de planilhas eletrônicas e com muita intervenção manual e, portanto, o resultado trazia pouca precisão, o que poderia prejudicar a sua leitura e interpretação. As Figuras 1 e 2 exemplificam tais gráficos, onde percebem-se problemas de legibilidade e precisão.



**Figura 1.** Gráfico *boxplot* para o resultado das avaliações de classificação linear para tipo utilizado no programa de melhoramento genético da raça Gir Leiteiro, em meados da última década [da Silva Verneque et al. 2003].

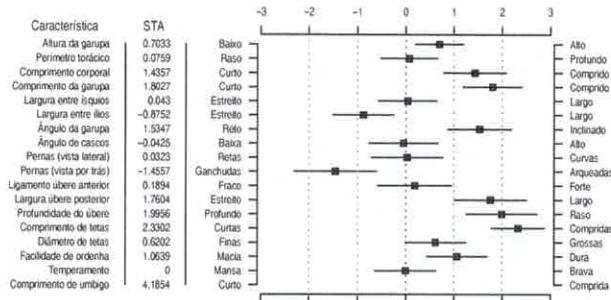


**Figura 2.** Gráfico *boxplot* para o resultado das avaliações de classificação linear para tipo utilizado no programa de melhoramento genético da raça Girolando, em 2011 [da Silva et al. 2011].

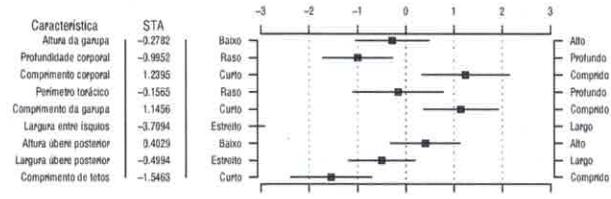
Com a utilização do software R foi possível alcançar um novo padrão gráfico e tipográfico e, notadamente, foi percebido diferenças na estrutura e modelos dos gráficos (Figuras 3 e 4), tornando-os mais precisos e, como consequência, favorecendo a leitura e interpretação dos mesmos. O procedimento para criação automatizada desses gráficos não trouxe benefícios somente para àqueles que lêem o conteúdo dos sumários e fazem uso de suas informações, mas, também trouxe melhorias sob a ótica de operacionalização do seu processo de construção.

Com a implantação desse procedimento, evita-se algumas etapas em que eram necessárias intervenções manuais diretas, tanto no preparo dos dados como na confecção

do próprio gráfico, permitindo otimizar etapas do processamento dos dados, reduzir o tempo despendido nesse processo e, ainda, maior controle sobre a estrutura e modelo do gráfico.



**Figura 3.** Gráfico *boxplot* para o resultado das avaliações de classificação linear para tipo utilizado no programa de melhoramento genético da raça Gir Leiteiro, em 2012 [da Silva Verneque et al. 2012].



**Figura 4.** Gráfico *boxplot* para o resultado das avaliações de classificação linear para tipo utilizado no programa de melhoramento genético da raça Girolando, em 2012 [da Silva et al. 2012].

#### 4. Conclusões

O procedimento ora apresentado, com a utilização de recursos computacionais de software adequados e gratuitos, possibilita uma eficiente maneira de construção de gráficos, agregada às metas de sejam obtidos bons resultados, em curto prazo, na geração dos gráficos *boxplot* para a leitura e interpretação dos resultados das avaliações para classificação linear de tipo dos programas de melhoramento genético de bovinos da Embrapa Gado de Leite

#### Agradecimentos

Os autores agradecem os revisores pelos pertinentes comentários e observações, à Fundação de Apoio à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro na concessão de bolsas e à Empresa Brasileira de Pesquisa Agropecuária (Embrapa) onde esse trabalho foi realizado.

#### Referências

- da Silva, M. V. G. B., de Carvalho Paiva, L., de Aguiar Rodrigues Cembranelli, M., Martins, M. F., de Freitas, A. F., Costa, C. N., Rodrigues, W. B. R., Arbex, W., Caetano, A. R., dos Santos, G. G., Bruneli, F. A. T., do Carmo Panetto, J. C., Viana, B. A.,

**SBIAGRO** 2013  
Anais do IX Congresso Brasileiro de  
Agroinformática  
Agronomáfrica: Inovação para a Sustentabilidade do  
Agronegócio Brasileiro

http://www.129.241.80/shia.org/analyse/fortiace-chimique

卷之三

Instituição: 1- Universidade Federal de Mato Grosso (UFMT); 2- Universidade de São Paulo (USP); São Carlos-SP	STI/Protecionaria de Tecnologia da Informação
Realização	Aponto
Promoção	Tutoria
Sensibilização	FAPEMAT
Sistêmico	Embrapa
Institucional	Mato Grosso
Bases e Trabalhos	Pronex
Tópico: Competição do Índice de Vegetação da Div. Agrícola	Instituto Federal de Mato Grosso (IFMT) / Selo Vida
Autores: Marcos A. Fausto   Nájel G. Machado   J. Ruybal   Marília dos Santos   Guilherme Agripêcius	Cuiabá-MT
Instituição: 1- Instituto Federal de Mato Grosso (IFMT); 2- Instituto Federal de Mato Grosso (IFMT) / Selo Vida	Cuiabá-MT
Tópico: Game Theory Concepts and Land-Use Dynamics	Unicamp
Autores: Fernando L. Graça   Mairon O. Souza   Bruno F. Costa   Fabiana S. Stohman   Antonio M. Buarque   Cícero E. Coelho	Campinas-SP
Instituição: 1- Secretaria de Gestão Estratégica, Entrada-Brasilia-CF	Unicamp
Bases e Trabalhos	Unicamp
Tópico: Implementação de Redes de Serviços Sem Fio e Sistemas de Informação para Agroindústria através de Serviços	Universidade Estadual do ABC, Santo André-SP
Autores: Willian F. Costa   Fabiana S. Stohman   Antonio M. Buarque   Cícero E. Coelho	Santo André-SP
Instituição: 1- Escola Politécnica, Universidade de São Paulo, São Paulo-SP; 2- Centro de Matemática, Computação e Ciência, Universidade Federal do ABC, Santo André-SP	Santo André-SP
Bases e Trabalhos	Unicamp
Tópico: Análise de Modelos de Simulação para Sistemas Agropecuários	Unicamp
Autores: Alain Sastre-Jiménez   Marlon Salete Marques Neto   Bruno Afonso   Hugo Leonardo Pelle Silveira   Thais Manuela Gonçalves	Caxias do Sul-RS
Instituição: 1- Departamento de Informática, Universidade Estadual de Ponta Grossa (UEPG)	Ponta Grossa-PR
Bases e Trabalhos	Unicamp
Tópico: Modelagem para Fertilização e Calagem na Cultura de Cana-de-Açúcar	Unicamp
Autores: Franklin José F. Fábio   Silvana Z.3. Felipe   Silvana Z.3. Felipe   Hugo Alvarino Venegasi   Alexandre de Castro	Caxias do Sul-RS
Instituição: 1- Fundação Joaquim Nabuco   Universidade Federal Rural de Pernambuco, Departamento de Agropecuária, Área das Sobras, CCA-UFRPE	Recife-PE
Bases e Trabalhos	Unicamp
Tópico: Aplicação do Processamento Digital de Imagens na Cultura da Banana	Unicamp
Autores: Silvia H. Medeiros-Góis   Silvia Domingues Lima   Denilson Eduardo Rozante   Wilson da Silva Moreira   Eduardo Nardini Gomes   Pedro Henrique Manzani	Ribeirão Preto-SP
Instituição: 1- Universidade Federal Rural do Rio Grande do Sul   2- Universidade Estadual de Roraima	Porto Alegre-RS
Bases e Trabalhos	Unicamp
Tópico: Pesquisa Operacional aplicada a decisões de Planejamento Agrícola em caso de programação dinâmica	Unicamp
Autores: Verena O. Monteiro   Thiago G. Reis   Adriano Leite   Silvia Vicente Carvalho Filho	São Paulo-SP
Instituição: 1- Universidade de Flórida, Department of Agricultural and Biological Engineering, 2- Universidade Superior de Agricultura Luiz de Queiroz	São Paulo-SP
Bases e Trabalhos	Unicamp
Tópico: Compartilhamento entre uso de modelo estatístico e modelagem fuzzy para criação do desempenho de frangos de corte	Unicamp
Autores: Guilherme F. Tavares   Leonardo Schiassi   Isac P. de Souza   Mariana M. Melo	São Paulo-SP
Instituição: 1- Universidade de São Paulo, São Paulo-SP; 2- Instituto de Ciências Agropecuárias e Ambientais da Universidade Federal de Mato Grosso (UFMT)	São Paulo-SP
Bases e Trabalhos	Unicamp
Tópico: Utilização de modelagem fuzzy na predição do ganho de peso de frango de corte	Unicamp
Autores: Isac P. de Souza   Leonardo Schiassi   Isac P. de Souza   Mariana M. Melo   Guilherme F. Tavares	São Paulo-SP
Instituição: 1- Universidade de São Paulo, São Paulo-SP; 2- Instituto de Ciências Agropecuárias e Ambientais da Universidade Federal de Mato Grosso (UFMT)	São Paulo-SP
Bases e Trabalhos	Unicamp
Tópico: Integração de Imagens R em Sistemas de Monitoramento Agropecuário visando a sensibilização	Unicamp
Autores: Jocelino Jr.   Oliveira Jr.   Mário B. Rosa   Rodrigo E. B. Sales   Cátia L. Zuganella   Antônio S. Melo   Guilherme G. Maihara   Celso Alberto Alves Meira	São Paulo-SP
Instituição: 1- Universidade Federal de Mato Grosso (UFMT); 2- Fundação ABC Pasqua   3- Desenvolvimento Agropecuário, Castro-PR	Castro-PR
Bases e Trabalhos	Unicamp
Tópico: Integração de Imagens R em Sistemas de Monitoramento Agropecuário visando a sensibilização	Unicamp
Autores: Jocelino Jr.   Oliveira Jr.   Mário B. Rosa   Rodrigo E. B. Sales   Cátia L. Zuganella   Antônio S. Melo   Guilherme G. Maihara   Celso Alberto Alves Meira	São Paulo-SP
Instituição: 1- Universidade Federal de Mato Grosso (UFMT); 2- Fundação ABC Pasqua   3- Desenvolvimento Agropecuário, Castro-PR	Castro-PR
Bases e Trabalhos	Unicamp

