

## **Transformação de Box-Cox Em Dados de Medidas Repetidas de Bovinos de Corte**

Cátia Valéria Presotti, Pós-graduanda do Curso de Estatística da UFSCar

[catia\\_v\\_p@hotmail.com](mailto:catia_v_p@hotmail.com)

Alfredo Ribeiro de Freitas, Pesquisador da Embrapa Pecuária Sudeste. Bolsista do CNPq.

[ribeiro@cnpse.embrapa.br](mailto:ribeiro@cnpse.embrapa.br)

Waldomiro Barioni Júnior, Pesquisador da Embrapa Pecuária Sudeste.

[barioni@cnpse.embrapa.br](mailto:barioni@cnpse.embrapa.br)

**RESUMO:** O objetivo foi analisar dados de pesagens ( $y$ ) de bovinos. De cada animal foram feitas nove pesagens do nascimento aos dois anos de idade. A variável  $y$  foi transformada por  $y^\lambda = (y^\lambda - 1)/\lambda$  ( $\lambda \neq 0$ ), a qual foi eficiente na redução dos coeficientes de assimetria e da heterogeneidade de variância dos dados.

**Palavras chave:** Heterogeneidade de variâncias, Transformação de Box-Cox e Medidas repetidas.

### **1. INTRODUÇÃO**

No estudo do desenvolvimento ponderal de bovinos de corte, a análise das pesagens realizadas no indivíduo, dentro de certo período, pode compreender três aplicações:

1) Produzir testes estatísticos para efeitos fixos em regressão, análise de variância (ANOVA): univariada, multivariada e estimar variâncias de efeitos aleatórios; o procedimento tradicionalmente adotado é o GLM do SAS (SAS, 1999-2001), que utiliza o método de quadrados mínimos (Littel et al., 1998; SAS, 1999-2001).

2) Estimar parâmetros genéticos de dados de pesos e de ganhos de peso padronizados à determinada idade, e mais recentemente, utilizando-se todos os pesos em análise de regressão aleatória (Cyrillo, 2003, Dias, 2004). Nesse caso, para a estimação de componentes de variâncias, geralmente utilizam-se de programas em linguagem FORTRAN e método da Máxima Verossimilhança Restrita (Boldman et al. 1995) e Inferência Bayesiana (Van Tassel & Van Vleck, 1995).

Nas aplicações 1 e 2 em dados de desenvolvimento ponderal de bovinos, deve-se considerar algumas características inerentes aos dados e que interferem nas análises: a) são irregulares no tempo; b) possuem estrutura incompleta; c) as avaliações adjacentes são mais estreitamente correlacionadas que as demais; d) a variabilidade da resposta dos indivíduos em função do tempo é crescente.

Para minimizar os problemas de a) a d) recomenda-se transformar a variável resposta  $y$  por meio da família de transformação de Box-Cox (Peltier et al, 1998), que utiliza o parâmetro de transformação  $\lambda$ , tal que  $y_i^\lambda = (y_i^\lambda - 1)/\lambda$  ( $\lambda \neq 0$ ) ou  $y_i^\lambda = \log y_i$ , ( $\lambda=0$ ). Segundo Freitas et al. (2002) esta transformação foi eficiente para reduzir os coeficientes de assimetria e de curtose de dados de nove pesagens, do nascimento até os dois anos de idade, de bovinos da raça Gir, machos e fêmeas, oriundos do Arquivo da ABCZ.

O objetivo deste estudo foi propor análise para as aplicações 1 e 2 acima em dados de pesagens de bovinos, estimando-se  $\lambda$  por máxima verossimilhança tal que  $y^\lambda = (y^\lambda - 1)/\lambda$ , para  $\lambda \neq 0$ , seja aproximadamente normal; avaliar a eficiência desta transformação na estabilização das variâncias e na redução dos coeficientes de assimetria e de curtose.

## 2. METODOLOGIA

Foram analisados dados de nove pesagens: ao nascimento ( $P_0$ ) e oito ( $P_1$  a  $P_8$ ), realizadas em intervalos trimestrais, até os dois anos de idade, de bovinos Nelore, Guzerá, Gir e Indubrasil, machos e fêmeas, oriundos da Associação Brasileira de Criadores de Zebu – ABCZ.

A variável resposta  $y_i$  foi analisada na escala  $y_i^\lambda$ , em que  $\lambda$  é obtido da família de transformação de Box-Cox, tal que  $y_i^\lambda = (y_i^\lambda - 1)/\lambda$ , ( $\lambda \neq 0$ ). Para cada raça foi obtida uma estimativa de  $\lambda$  para os dados de pesagens do animal, independente de medidas repetidas. Para a estimativa de  $\lambda$ , utilizou-se uma adaptação da macro BOXGLM.SAS (Friendly, 2000), disponível no SAS (<http://www.math.yorku.ca/SCS/sasmac/boxglm.html>. Acesso em 20/03/2004).

Foi adotado o modelo misto usual para análises de medidas repetidas (Littell et al., 1998)  $y_{ijk} = \mu + \alpha_i + d_{ij} + t_k + (\alpha\alpha)_{ik} + \varepsilon_{ijk}$ , em que  $y_{ijk}$  é a resposta da avaliação  $k$  no indivíduo  $j$  no sexo  $i$ ;  $\mu$  é a média global,  $\alpha_i$  é o efeito fixo de sexo  $i$ ,  $d_{ij}$  é o efeito aleatório do animal  $j$  no sexo  $i$ ,  $t_k$  é o efeito fixo do tempo  $k$ ,  $(\alpha\alpha)_{ik}$  é o efeito de interação e  $\varepsilon_{ijk}$  é o erro aleatório. Os efeitos  $d_{ij}$  são considerados independentes e com variância  $\sigma_d^2$ .

## 3. RESULTADOS E DISCUSSÃO

A Tabela 1 mostra a média e o número de observações correspondentes às nove pesagens, do nascimento até dois anos de idade, das raças Nelore, Guzerá, Gir e Indubrasil, valores que foram utilizados para a confecção da Figura 1. O número de observações ilustra o comportamento típico de pesagem em bovinos, ou seja, a grande redução de animais em pesagens subsequentes, implicando em estrutura incompleta dos dados.

**Tabela 1** – Média e número de observações correspondentes a cada peso das raças Nelore, Guzerá, Gir e Indubrasil.

Peso	Nelore		Guzerá		Gir		Indubrasil	
	Média	N	Média	N	Média	N	Média	N
P <sub>0</sub>	29,4	49984	31,9	41514	25,0	35731	32,4	15172
P <sub>1</sub>	72,0	47476	67,8	40936	59,4	35504	75,0	14833
P <sub>2</sub>	130,7	45644	121,8	38918	109,0	31720	132,8	14027
P <sub>3</sub>	179,6	45046	166,5	38523	151,2	31960	184,1	13907
P <sub>4</sub>	211,3	37856	195,7	33235	182,1	27210	225,0	12210
P <sub>5</sub>	242,0	31508	223,8	27884	210,8	22277	262,9	10276
P <sub>6</sub>	276,9	26345	255,6	23034	239,3	18196	297,3	8387
P <sub>7</sub>	306,1	17799	286,8	15521	263,5	13312	325,6	6065
P <sub>8</sub>	342,1	3916	312,7	3962	291,6	3055	358,5	1710

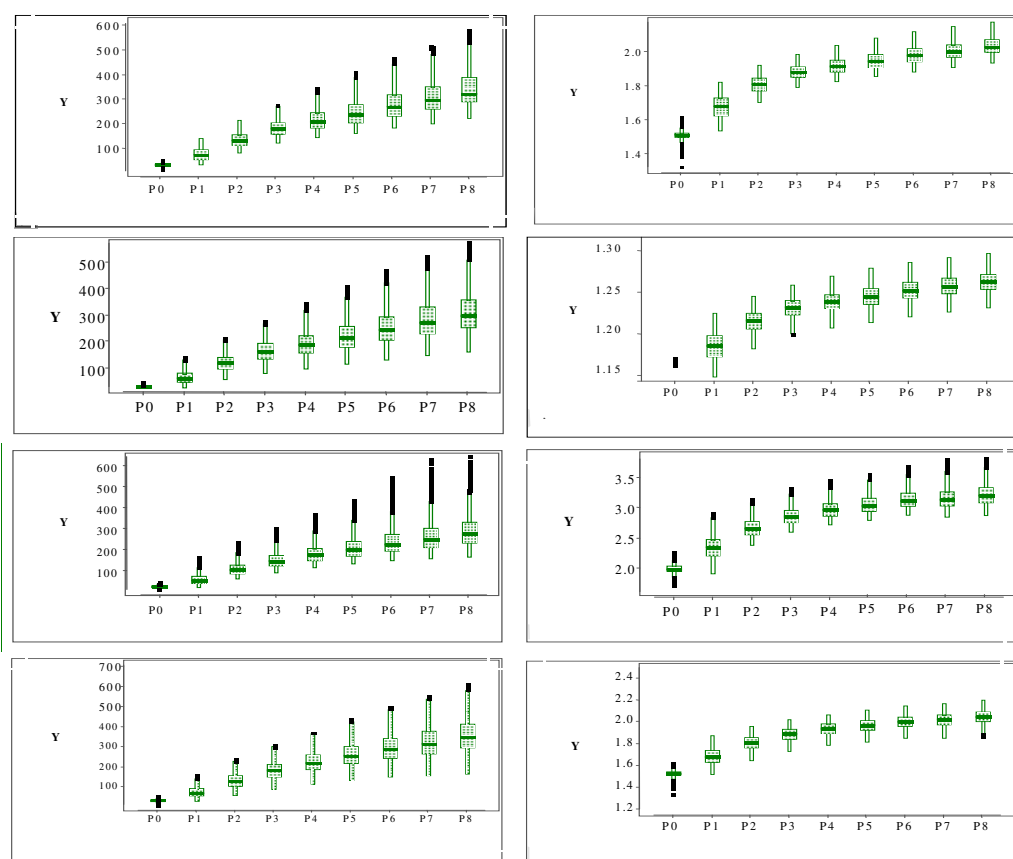
A Tabela 2 apresenta os coeficientes de assimetria e de curtose; à esquerda, os valores obtidos dos pesos observados  $y$  e, à direita, os valores obtidos dos pesos transformados por meio da família de transformação de Box-Cox. As estimativas de  $\lambda$ , obtida por máxima verossimilhança, foram: 0,12302 (Nelore), 0,15201 (Guzerá), 0,04082 (Gir) e 0,12245 (Indubrasil), indicando que a transformação logarítmica, normalmente usada para dados de pesos não é adequada, pois, nesse caso ter-se-ia  $\lambda = 0$ . Observa-se que a transformação  $y^\lambda$ , foi efetiva na redução dos coeficientes de assimetria para todas as pesagens e raças.

**Tabela 2** – Valores de assimetria (A) e curtose (C), obtidos dos pesos  $y$  (esquerda) e  $y^\lambda$  (direita).  $P_i$  ( $i = 0, \dots, 8$ ), refere-se aos pesos ao nascimento ( $P_0$ ) e até 24 meses de idade.

Peso	Nelore		Guzerá		Gir		Indubrasil	
	A $y$   $y^\lambda$	C $y$   $y^\lambda$	A $y$   $y^\lambda$	C $y$   $y^\lambda$	A $y$   $y^\lambda$	C $y$   $y^\lambda$	A $y$   $y^\lambda$	C $y$   $y^\lambda$
P <sub>0</sub>	0,53 -0,04	2,27 2,39	0,63 1,63	2,65 2,63	0,39 -0,12	1,10 0,84	0,38 -0,19	0,87 1,59
P <sub>1</sub>	0,43 -0,09	-0,48 -0,74	0,61 -0,09	-0,01 -0,58	0,94 0,00	1,08 -0,35	0,71 0,05	0,22 -0,58
P <sub>2</sub>	0,39 0,02	-0,54 -0,74	0,43 -0,13	-0,27 -0,48	0,84 0,22	0,64 -0,42	0,42 -0,11	-0,31 -0,48
P <sub>3</sub>	0,40 0,05	-0,30 -0,54	0,31 -0,22	-0,28 -0,35	0,71 0,16	0,49 -0,42	0,30 -0,19	-0,35 -0,39
P <sub>4</sub>	0,62 0,24	-0,01 -0,45	0,53 -0,06	0,06 -0,29	0,88 0,33	0,79 -0,28	0,33 -0,14	-0,34 -0,38
P <sub>5</sub>	0,79 0,39	0,17 -0,42	0,69 0,09	0,19 -0,37	0,97 0,42	0,93 -0,26	0,39 -0,08	-0,31 -0,43
P <sub>6</sub>	0,79 0,40	0,09 -0,49	0,72 0,14	0,18 -0,40	1,08 0,49	1,25 0,18	0,42 -0,06	-0,30 -0,46
P <sub>7</sub>	0,87 0,46	0,25 -0,37	0,70 0,12	0,17 -0,42	1,10 0,47	1,41 -0,11	0,45 -0,05	-0,22 -0,38
P <sub>8</sub>	0,93 0,55	0,30 -0,32	0,76 0,15	0,39 -0,32	1,01 0,34	0,99 -0,21	0,54 -0,03	-0,08 -0,22

Informações valiosas sobre os dados originais e transformados podem ser obtidas também por meio dos gráficos *Box-Plot* (Figura 1), como por exemplo, a evolução do crescimento de animais segundo a idade e a existência de heterogeneidade de variâncias, denominada de “inflação de variância” (Davidian e Giltinan, 1996). No gráfico de caixa (*Box-Plot*), a linha horizontal no meio da caixa representa a mediana, enquanto que a parte inferior e superior, indicam, respectivamente, o primeiro ( $Q_1$ ) e o terceiro ( $Q_3$ ) quartil. As caixas estreitas (“whiskers”) acima e abaixo da caixa central possuem distância não superior a 1,5 vezes à distância interquartilica ( $Q_3 - Q_1$ ), sendo que os valores mais extremos isolados são considerados *outliers*.

Observando-se os dados analisados na escala original (Figura 1, esquerda), observa-se para todas as pesagens das quatro raças bovinas, que a caixa estreita é mais alongada na parte acima da caixa central e com comprimento superior a  $1,5(Q_3 - Q_1)$ . Este aspecto da Figura 1 mostra que os dados de pesagens possuem assimetria positiva e que a variância é crescente na resposta dos animais com a idade; verificam-se também alguns valores extremos e isolados nos dados de pesos da raça Gir, indicando presença de *outliers*. Usando-se raciocínio análogo com os dados transformados por meio dos valores de  $\lambda$  (Figura 1, direita), observa-se redução acentuada da assimetria e com homogeneidade de variâncias entre as pesagens.



**Figura 1-** Gráfico de Caixa para dados de pesos de bovinos na escala original  $y$  (esquerda) e transformados  $y^\lambda$  (direita). Os gráficos de cima para baixo estão associados às raças Nelore, Guzera, Gir e Indubrasil, respectivamente.  $P_i$  referem-se aos pesos ao nascimento ( $P_0$ ) e até dois anos de idade.

Analisando-se a Tabela 1 e a Figura 1, verifica-se que a utilização dos valores de  $\lambda$  para analisar dados em uma nova escala, foi efetiva para a melhoria da qualidade dos dados de pesagens de bovinos, com reflexo positivo na ANOVA. Na análise de variância, há evidências de que os valores da assimetria, da curtose, assim como os desvios de normalidade dos dados e a heterogeneidade da variância, influenciam as inferências obtidas (Brownie et al., 1990). Desta forma, será possível a

obtenção de inferências mais precisas de efeitos fixos em análises de variâncias e estimação de parâmetros genéticos mais eficientes de efeitos aleatórios. Do ponto de vista teórico, esta melhoria na qualidade dos dados de pesagens após o uso desta transformação já era esperada, pois em uma ANOVA com os dados na escala  $y^\lambda$ , o valor de  $\lambda$  obtido por máxima verossimilhança representa o valor máximo para a estatística  $F$  e o valor mínimo para o Quadrado Mínimo Residual.

#### 4. CONCLUSÕES

A transformação dos dados de pesos de bovinos Nelore, Guzerá, Gir e Indubrasil por meio da família de transformação de Box-Cox, foi efetiva na redução dos coeficientes de assimetria e da heterogeneidade de variância para todas as pesagens e raças.

A análise dos dados de pesos de bovinos por meio do modelo misto usual para análises de medidas repetidas foi mais eficiente do que o modelo simples nulo ajustado pelo procedimento GLM do SAS ( $R = I\sigma^2_e$ ), indicando, para estes conjuntos de dados, que a análise univariada normalmente feita por meio do procedimento GLM, considerando-se medidas repetidas como subparcela, não é apropriada.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Boldman, K.G. e Kriese, L.A. e Van Vleck, L.D. et al. (1995). *A manual for use of MTDFREML. A set of programs to obtain estimates of variance and covariance (DRAFT)*, Lincoln: Department of Agriculture, Agriculture Research Service, 120p.
- Brownie, C. e Boos, D.D. e Oliver, J.H. (1990). *Modifying the t and ANOVA F tests when treatment is expected to increase variability relative to controls*. Biometrics, 46, n.1, p. 259-266.
- Cyrillo, J.N. dos S.G. (2003). *Estimativas de funções de covariância para crescimento de animais machos Nelore utilizando modelos de regressão aleatória*. Jaboticabal: Universidade do Estado de São Paulo, 75p. Tese (Doutorado em Zootecnia) – Faculdade de Ciências Agrárias e Veterinárias – UNESP
- Davidian, M. e Giltinan, D. M. (1996). *Nonlinear models for repeated measurement data*. 2.ed. London: Chapman Hall. 359p.
- Dias, L.T. (2004). *Modelos de regressão aleatória para características de crescimento de bovinos da raça Tabapuã*. Jaboticabal: Universidade do Estado de São Paulo, 59p. Tese (Doutorado em Zootecnia) – Faculdade de Ciências Agrárias e Veterinárias – UNESP.
- Freitas, A. R. e Silva, L.O.C. e Euclides F. K e Reis J.E. e Kakuda, C. M. e Ferreira, F.H.M. (2002). *Estudo de dados de pesagens de bovinos usando a transformação de box-cox*. In: *sociedade brasileira de melhoramento animal*. IV, Campo Grande, MS. Anais. Campo Grande: SBMA. p.212-214
- Littell, R.C. e Henry, P.R. e Ammerman, C.B. (1998). *Statistical analysis of repeated measures data using SAS procedures*. Journal of Animal Science. 76:1216-1231.
- Peltier, M R. e Wilcox, C J e Sharp, D C. (1998). *Technical note: Application of the Box-Cox data transformation to animal science experiments*. Journal of Animal Science, v.76, n.3, p.847-849.
- Van Tassel, C.P. e Van Vleck, L.D. (1995). *A manual for use of MTGSAM. A set of fortran programs to apply gibbs sampling to animal models for variance component estimation*. (DRAFT). Lincoln: Department of Agriculture/ Agriculture Research Service, 86p.