

## ESTUDO DE DADOS DE PESAGENS DE BOVINOS

### USANDO-SE A TRANSFORMAÇÃO BOX-COX<sup>1</sup>

Alfredo Ribeiro de Freitas<sup>2</sup>, Luiz Otávio Campos da Silva<sup>3</sup>, Kepler Euclides Filho<sup>3</sup>,  
José Eduardo dos Reis<sup>4</sup>, Cláudio Massaki Kakuda<sup>5</sup>,  
Fabio H. Milanetto Ferreira<sup>5</sup>

<sup>1</sup>Financiado pelo CNPq

<sup>2</sup> Embrapa Pecuária Sudeste, CP 339, 13560-970 São Carlos-SP;

<sup>3</sup> Embrapa Gado de Corte, CP 154, 79002-970 Campo Grande- MS;

<sup>4,5</sup> Pós-graduação em São Carlos, SP: <sup>4</sup>ICMC-USP, <sup>5</sup>Instituto de Física-USP.

#### Introdução

No estudo de desenvolvimento ponderal de bovinos de corte, os pesos dos animais obtidos ao longo do tempo, normalmente, apresentam as seguintes características: a) correlação entre os erros, cuja magnitude depende do intervalo entre as medidas; b) vícios ou tendências nas pesagens, acarretando dados bimodais ou multimodais, presença de *outliers*, entre outras; c) não-normalidade, assimetria e curtose; d) heterogeneidade de variâncias. Estas ocorrências implicam em predição viciada do valor genético, redução do progresso genético e seleção desproporcional de animais oriundos de ambientes com variâncias diferentes, entre outros problemas. Para minimizá-los (PELTIER et al, 1998) e obter parâmetros genéticos mais confiáveis, recomenda-se transformar a variável resposta. Uma das transformações conhecidas e amplamente usada é a família de transformação *Box-Cox*. Por essa transformação, o parâmetro  $\lambda$ , tal que:  $y_i^\lambda = (y_i^\lambda - 1) / \lambda$ , para  $\lambda \neq 0$  ou  $y_i^\lambda = \log y_i$ , para  $\lambda=0$ , apresenta distribuição normal. O objetivo deste trabalho foi estimar  $\lambda$ , por máxima verossimilhança, por meio de uma macro BOXGLM.SAS do SAS, considerando-se dados de pesagens, do nascimento até os dois anos de idade, de animais da raça Gir.

#### Material e Métodos

Foram utilizados dados de nove pesagens: uma ao nascimento (PN) e oito (P1 a P8), realizadas em intervalos trimestrais, até os dois anos de idade, de 35.715 de bovinos da raça Gir, machos e fêmeas, nascidos de 1976 a 1992, cujas informações se encontram no Arquivo Zootécnico Nacional das Raças Zebuínas, mantido pela ABCZ/Embrapa Gado de Corte. A estimativa de  $\lambda$ , obtida por máxima verossimilhança, tal que  $y_i^\lambda = (y_i^\lambda - 1) / \lambda$ , para  $\lambda \neq 0$ , apresenta distribuição aproximadamente normal e foi estimada, considerando-se todas as observações, por meio da aplicação da macro BOXGLM.SAS do SAS (<http://www.math.yorku.ca/SCS/sasmac/boxglm.html>, acesso em 24/04/2001). Para visualizar características associadas as distribuições, como assimetria, curtose e *outliers*, foram utilizados diagramas de caixa (*Box-Plot*) em cada pesagem, antes e após o uso da transformação Box-Cox, por meio do SAS (SAS, 2000).

#### Resultados e Discussão

Os gráficos de diagramas de caixa (Figuras 1A, 1B), são efetivos para comparar distribuições de dados contínuos e revelam características importantes, tais como a evolução do crescimento de animais Gir segundo a idade, existência de heterogeneidade de variâncias, com variabilidade crescente na resposta dos animais com a idade, fato associado

ao fenômeno da “inflação de variância” (DAVIDIAN e GILTINAN, 1996). No diagrama de caixa, a linha horizontal no meio da caixa representa a mediana; a parte inferior e superior da caixa indica, respectivamente, o primeiro ( $Q_1$ ) e o terceiro ( $Q_3$ ) quartil, correspondendo aos elementos de posição 25<sup>o</sup> e 75<sup>o</sup> da amostra. As caixas estreitas (“whiskers”), acima e abaixo da caixa central, possuem distância não superior a 1,5 vezes a distância interquartílica ( $Q_3 - Q_1$ ). Os valores mais extremos isolados e que recebem marcações individuais são considerados dados discrepantes da amostra (*outliers*). Observando-se os dados analisados na escala original (Figura 1A), verifica-se assimetria positiva em todas as pesagens, presença de *outliers* e, ainda, variância crescente na resposta dos animais com a idade. As medidas de assimetria para as variáveis originais ( $y_i$ ) PN, P<sub>1</sub> a P<sub>8</sub> e transformadas ( $y_i^\lambda$ , transformação de Box-Cox), foram, pela ordem, (0,21 - 0,18; 1,21 - 0,68; 0,73 - 0,39; 0,53 - 0,20; 0,74 - 0,42; 0,78 - 0,46; 0,83 - 0,51; 0,83 - 0,52; 0,90 - 0,62). Os valores para a curtose, para as variáveis  $y_i$  e  $y_i^\lambda$ , para PN, P<sub>1</sub> a P<sub>8</sub>, foram, respectivamente: 2,02 - 1,76; 2,85 - 0,89; 0,82 - 0,16; 0,68 - 0,22; 0,97 - 0,37; 0,88 - 0,34; 0,88 - 0,30; 0,90 - 0,30 e 0,66 - 0,20. Observa-se que a transformação de Box-Cox foi efetiva na redução dos valores da assimetria, curtose e heterogeneidade de variâncias, para todas as pesagens. Este fato, também, pode ser comprovado por meio da análise da Figura 1B que revela não haver assimetria viesada à direita, que era nítida na Figura 1A. A solução de  $\lambda$ , por máxima verossimilhança, ( $\lambda = 0,6$ ) e log da função de verossimilhança (*likelihood*) igual a -52085,26, coincidem com o valor máximo para a estatística F (Figura 1C) e valor mínimo para o QMR. (Figura 1D). Conclui-se que a transformação pela família de Box-Cox foi efetiva para reduzir os coeficientes de simetria e de curtose. Segundo MEUWISSEN et al., (1996), apenas com a correção da heterogeneidade de variâncias dentro de rebanhos em dados de produção de leite, reduziu-se o viés dos valores genéticos em cerca de 19%.

### Conclusões

A transformação pela família de Box-Cox foi efetiva para reduzir a simetria, a curtose e a heterogeneidade de variâncias em dados de desenvolvimento ponderal de bovinos.

### Referências Bibliográficas

DAVIDIAN, M., GILTINAN, D.M. 1996. Nonlinear models for repeated measurement data: monographs on statistics and applied probability 62, Chapman & Hall, London,

9p.

FRIENDLY, M. Power transformations by Box-Cox method with graphic display of maximum likelihood solution, t-values for model effects, and influence of observations on choice of power. Disponível em: <http://www.math.yorku.ca/SCS/sasmac/boxglm.html>. Acesso em 24/04/2001.

MEUWISSEN, T.H.E.; JONG, G.; ENGEL, B. Joint estimation of breeding values and heterogeneous variances of large data sets. *Journal of Dairy Science*, Champaign, v.79, n.2, p.310-316, Feb. 1996

PELTIER, M R; WILCOX, C J; SHARP, D C . Technical note: Application of the Box-Cox data transformation to animal science experiments *Jour. Anim. Sci*, v.76, n.3, p.847-849, 1998.

SAS Institute 2000. *SAS/INSIGHT User's Guide*. versão 8.2, versão para Windows Cary, NC, USA.

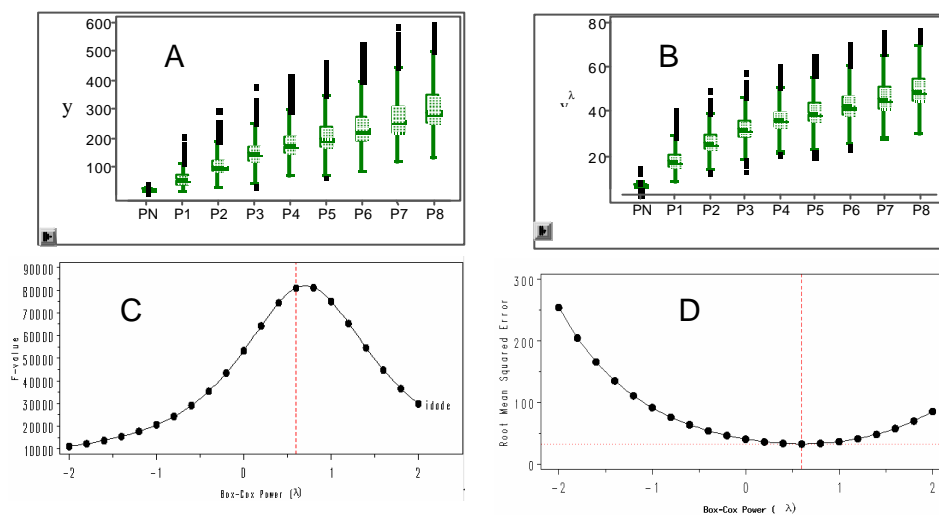


Figura 1. *Box-plots* para dados de pesagens na escala original (A), transformados com  $\lambda = 0,6$  (B) e solução de  $\lambda$  por máxima verossimilhança (C, D).