

# Transformação de Dados e Estimativas e Herdabilidade em Bovinos

ALFREDO RIBEIRO DE FREITAS<sup>1</sup>

## Resumo

Dados de pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos doze ( $y_3$ ) e aos dezoito meses ( $y_4$ ) de bovinos Canchim contendo efeitos de sexo (S), ano de nascimento (ANO), ordem de parto (OP) e efeito aleatório de touros (T) foram estudados por meio de estatísticas descritivas. Rejeitou-se ( $P < 0,01$ ) a hipótese de que os dados seguem uma distribuição normal, com a curva de distribuição apresentando-se assimétrica e platicúrtica; verificou-se ainda heterogeneidade de variâncias entre os efeitos de T. Utilizando-se estes dados, estimativas de herdabilidade de análises univariadas foram obtidas por Máxima Verossimilhança Restrita utilizando-se um modelo que incluiu os efeitos de S, ANO, OP e T, e com as variáveis  $y_i$  ( $i=1, \dots, 4$ ) analisadas em cinco escalas: a) ajustadas linearmente para as idades padrão, b)  $\log(y_i)$ , c) padronizada  $z_i = (y_i - \bar{y}) / s$ , d) aproximação de BOX e COX e e) escores normais de BLOM. A transformação dos dados pelo método dos escores normais de BLOM aproximou os dados a uma distribuição normal, indicando ser mais apropriada do que as outras.

Palavras-chave: bovinos, características de crescimento, transformação de dados, estimativas de herdabilidade

## DATA TRANSFORMATION AND HERITABILITY ESTIMATES IN CATTLE

## Abstract

Data on birth ( $y_1$ ), weaning ( $y_2$ ) 12 months ( $y_3$ ) and 18 months ( $y_4$ ) weights of Canchim beef cattle that included the sex (S), birth order (OP) and years (ANO) as fixed effects and random effects of sire (T) were studied by descriptive statistics. Rejected ( $P < 0.01$ ) the hypothesis that the data follow the normal distribution, showing skewed and platykurtic; also was verified variance heterogeneity across sires. Using this data set, heritability estimates were obtained for all variables using the univariate restricted maximum likelihood (REML) method in a sire model that included, in addition to the mean, the effects of S, OP, ANO and T. The variables were submitted to five

transformations: a) ajust linear for the standard ages, b)  $\log(y_i)$ , c)  $z_i = (y_i - \bar{y}) / s$ , d) BOX & COX approximation and e) normal scores of BLOM. The BLOM normal scores appears to fit better than the others transformations.

Key words: cattle, growth traits, data transformation, heritability estimates.

## Introdução

A metodologia dos modelos mistos (MM) com soluções BLUP ("Best Linear Unbiased Predictor") é a ideal para estimar parâmetros genéticos no melhoramento animal. No entanto, para se obterem estimadores confiáveis é necessário preocupar com a qualidade dos dados. Por exemplo, na construção dos MM deve-se usar o vetor de dados na forma normal multivariada (4). Segundo (7) os coeficientes de assimetria e de curtose são indicadores mais importantes de como o afastamento da normalidade influencia a análise de variância.

Vários pesquisadores tem preocupado com a qualidade dos dados nas mais diversas áreas: transformação e homogeneidade de variância no melhoramento animal (1), transformação de dados para aplicação em distância euclidiana e análise de conglomerados "cluster" (8); transformação em dados altamente assimétricos e com presença de 'outliers' (3). No entanto, com base na literatura nenhum procedimento individual soluciona os problemas de transformação dos dados.

O objetivo deste trabalho foi comparar cinco tipos de transformação em dados de pesos de bovinos quanto à estimativas de coeficientes de assimetria, curtose e herdabilidade.

## Material e Métodos

Foram utilizados pesos ao nascimento ( $y_1$ ), à desmama ( $y_2$ ), aos doze ( $y_3$ ) e aos dezoito meses ( $y_4$ ) de 3649 bovinos canchim nascidos de 1969 a 1994 na Fazenda Canchim, São Carlos, SP. Estatísticas descritivas desses dados foram obtidas por meio do Proc Univariate do SAS (6).

Estimativas de herdabilidade foram estimadas por meio do modelo misto  $y = Xb + Zu + \varepsilon$ , considerando-se os efeitos fixos de sexo (S), ano de nascimento (ANO), ordem de parto (OP) e efeito aleatório de touros. As estimativas de variâncias e covariâncias foram obtidas por meio da Máxima Verossimilhança Restrita (REML), utilizando-se o Proc Varcomp do SAS (6). Os erros padrão das estimativas de herdabilidade foram obtidos segundo a expansão pela série de de TAYLOR (5).

As estimativas acima foram obtidas com variáveis  $y_i$  ( $i=1, \dots, 4$ ) analisadas em cinco escalas:

- a)  $y_i$  ajustado linearmente para as idades padrão: (205 ( $y_2$ ), 365 ( $y_3$ ) e 540 ( $y_4$ ) dias;
- b)  $\log(y_i)$
- c) padronizada  $z_i = (y_i - \bar{y}) / s$
- d) BOX e COX (3)
- e) Escores normais de BLOM (6)

<sup>1</sup> EMBRAPA/CPPSE Caixa Postal 339, 13 560-970 São Carlos - SP.

## Resultados e Discussão

As estimativas de assimetria, curtose e de herdabilidade obtidas com os cinco tipos de transformação dos dados estão na Tabela 1.

Analisando os coeficientes de assimetria e de curtose, verifica-se que a transformação dos escores normais de BLOM aproximou os dados da distribuição normal, uma vez que estes coeficientes podem ser considerados nulos, destacando-se a seguir a transformação de BOX e COX, que proporcionou resultados satisfatórios para a assimetria.

As estimativas de herdabilidade de  $y_2$  e  $y_4$  obtidas pelo método de BLOM foram superiores às demais; porém, observou-se similaridade nos erros padrão obtidos de todas as transformações. Verifica-se, portanto, que associando-se os resultados obtidos pelo método dos escores normais de BLOM quanto à assimetria e curtose e estimativas de herdabilidade, esta transformação se comportou como a mais adequada para esta natureza de dados.

## Conclusões

A transformação dos dados de crescimento de bovinos pelo método dos escores normais de BLOM aproximou os dados a uma distribuição normal, comportando-se como a mais adequada para esta natureza de dados.

## Referências Bibliográficas

- 1 - BOLDMAN, K.G., FREEMAN, A.E. Adjustment for heterogeneity of variance by herd production level in dairy cow and sire evaluation. *Journal of Dairy Science*, Champaign, v.73, n.2, p.503-512, Feb.1990
- 2 - CARVALHO, J.F. Tópicos em planejamento de experimentos: desenvolvimento de planos experimentais assistido por computador. *Concórdia: EMBRAPA-CNPSA*, 1996. 126p. (Notas de aula).
- 3 - COOK, R.D.; WEISBERG, S. Transforming a response variable for linearity *Biometrika*, London, v.81, n.4, Dec., p.731-737, 1994
- 4 - HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, Washington, v. 72, n., p.320-340, 1977
- 5 - KENDALL, M.G. & STUART, A. The advanced theory of statistics. London, Charles Griffing, 1969. V.1.
- 6 - SAS INSTITUTE. In: SAS/STAT User's guide: statistics. versão 6, vol 2, 4. ed. Cary, 1993
- 7 - SCHEFFÉ, A. The analysis of variances. John Wiley, 1959. 477p.
- 8 - STODDARD, A.M. Standardization of measures prior to cluster analysis. *Biometrics*, Oxford, v.35, n.4, december, p.766-773, 1979

TABELA 1 - Estimativas de assimetria (A), curtose (C) e de herdabilidade obtidas dos pesos ao nascimento ( $y_1$ ), desmama ( $y_2$ ), doze ( $y_3$ ) e 18 meses ( $y_4$ ) de Canchim obtidas sob cinco transformação de dados.

| Característica                      | Transformação*      |        |                     |        |                     |        |                     |        |                     |         |
|-------------------------------------|---------------------|--------|---------------------|--------|---------------------|--------|---------------------|--------|---------------------|---------|
|                                     | T1                  |        | T2                  |        | T3                  |        | T4                  |        | T5                  |         |
|                                     | A                   | C      | A                   | C      | A                   | C      | A                   | C      | A                   | C       |
| $y_1$                               | 0,30 01             | 0,1200 | 0,3021              | 0,1101 | 0,3021              | 0,1101 | -0,2057             | 0,2326 | 0,0157              | -0,0395 |
| $y_2$                               | -0,0943             | 1,1920 | 0,5768              | 4,8000 | 0,5768              | 4,8000 | -0,3546             | 0,9660 | 0,0000              | -0,0227 |
| $y_3$                               | -0,16 11            | 0,4020 | 0,4965              | 0,9762 | 0,4965              | 0,9762 | -0,2213             | 0,4406 | 0,0000              | -0,0227 |
| $y_4$                               | 0,3821              | 1,4103 | 0,5503              | 1,9511 | 0,5503              | 1,9511 | -0,2665             | 0,7657 | 0,0000              | -0,0227 |
| Herdabilidade: $h^2 \pm s (h^2) \#$ |                     |        |                     |        |                     |        |                     |        |                     |         |
| $y_1$                               | 0,3089 $\pm$ 0,0000 |        | 0,2996 $\pm$ 0,0678 |        | 0,3089 $\pm$ 0,0686 |        | 0,3089 $\pm$ 0,0686 |        | 0,3007 $\pm$ 0,0679 |         |
| $y_2$                               | 0,5787 $\pm$ 0,0991 |        | 0,5715 $\pm$ 0,0986 |        | 0,5787 $\pm$ 0,0991 |        | 0,5836 $\pm$ 0,0995 |        | 0,6425 $\pm$ 0,1047 |         |
| $y_3$                               | 0,4548 $\pm$ 0,1016 |        | 0,4814 $\pm$ 0,1058 |        | 0,4545 $\pm$ 0,1016 |        | 0,4810 $\pm$ 0,1057 |        | 0,5026 $\pm$ 0,1081 |         |
| $y_4$                               | 0,4483 $\pm$ 0,0912 |        | 0,4419 $\pm$ 0,0887 |        | 0,4484 $\pm$ 0,0912 |        | 0,4520 $\pm$ 0,0914 |        | 0,4897 $\pm$ 0,0944 |         |

T1: ( $y_i$ )      T2:  $\log(y_i)$       T3:  $(y_i - \bar{y}) / s$

T4: BOX e COX      T5: BLOM

# erros padrão obtidos pela aproximação de Taylor (5)