USO DE FATORES MULTIPLICATIVOS PARA AJUSTAMENTO DE DADOS

ELIAS NUNES MARTINS[1], MARTINHO DE ALMEIDA E SILVA[1], MAURÍCIO MELLO DE ALENCAR[2], LUÍS OTÁVIO CAMPOS DA SILVA[2]

[1] Professores da UEM e da UFMG
[3] Pesquisadores da EMBRAPA

RESUMO: O uso do procedimento de ajustamento de dados por fatores multiplicativos conduz à estimação e predição viesadas. Para que estimativas e predições não-viesadas sejam obtidas após o ajustamento por fatores multiplicativos deve-se subtrair do vetor de observações ajustadas a média da subclasse escolhida como padrão e pré-multiplicar o resultado pela inversa da matriz de fatores de ajustamento.

PALAVRAS-CHAVE: estimação, modelo linear, predição, viesado

USING OF MULTIPLICATIVE FACTORS FOR DATA ADJUSTMENT

ABSTRACT: Using of multiplicative factors for data adjustment leads to biased estimations and predictions. For unbiased estimations and predictions based on adjusted data the average of the standard subclass must be subtracted from the vector of adjusted data, and the results should be multiplied by the inverse of the adjustment factor matrix.

KEYWORDS: biased, estimation, linear model, prediction

INTRODUÇÃO

Os fatores multiplicativos são calculados pela razão entre a média na subclasse escolhida como padrão e a média na subclasse a que pertence a observação que se deseja ajustar e o processo de ajustamento consiste em se multiplicar a observação pelo fator correspondente.

O uso de fatores multiplicativos para o ajustamento de dados tem sido uma prática comum na área de melhoramento animal (COOPER e HARGROVE, 1982; KEOWN e EVERETT, 1985; MARTINEZ et al., 1990; NORMAN et al., 1995; TORRES, 1999). A justificativa para esse procedimento é a redução do sistema de equações e a conseqüente diminuição das dificuldades computacionais e tempo de análise.

Entretanto, o uso desse procedimento causa alterações nos resultados de estimação e predição executados sobre os dados ajustados. O objetivo deste trabalho é mostrar que o procedimento de ajustamento de dados por fatores multiplicativos causa alterações nos resultados de estimação e predição executados sobre os dados ajustados e apresentar um procedimento que evita a obtenção de estimativas e predições viesadas, após ajustamento por fatores multiplicativos.

MATERIAL E MÉTODOS

Considere o modelo linear, na forma matricial

$$y = X\beta + \varepsilon$$

em que

y é o vetor de observações;

X é a matriz de incidência dos efeitos fixos ;

β é o vetor de efeitos fixos;

ε é o vetor de resíduos com estrutura de correlação entre seus elementos.

Define-se $\varepsilon = Za + e$, em que Z é a matriz de incidência dos efeitos aleatórios contidos no vetor a e e o novo vetor de resíduos sem estrutura de correlação entre seus elementos.

Sendo assim, a variância de y é dada por $Var(y) = V = ZGZ' + R$, em que

$G = A\sigma_a^2$ e $R = I\sigma_e^2$, sendo A a matriz de correlação entre os elementos de a, I uma matriz identidade e σ_a^2 e σ_e^2 as variâncias de a e e, respectivamente.

Admitindo-se no modelo dois fatores de efeito fixo A e B, respectivamente com p e q níveis, tem-se.

$$\beta' = \begin{bmatrix} \beta_a' & \beta_b' \end{bmatrix} = \begin{bmatrix} \beta_{a1} & \beta_{a2} & \cdots & \beta_{ap} & \beta_{b1} & \beta_{b2} & \cdots & \beta_{bq} \end{bmatrix}.$$

Impondo-se ao modelo as restrições $\mu = 0$ e $\sum_{i=1}^{q} \beta_{bi} = 0$, tem-se $X = \begin{bmatrix} X_a & X_b \end{bmatrix}$ e estimativas para β podem ser obtidas por $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$ e predições para a por $\hat{a} = GZ'V^{-1}(y - X\hat{\beta})$.

Fatores de correção multiplicativos para o fator de efeito fixo A, tomando-se como padrão o nível 1, seriam dados por

$$f' = \left[\frac{\hat{\beta}_{a1}}{\hat{\beta}_{a1}} \; \frac{\hat{\beta}_{a1}}{\hat{\beta}_{a2}} \cdots \frac{\hat{\beta}_{a1}}{\hat{\beta}_{ap}}\right]$$

e a o ajustamento dos dados e efetuado por meio de $y^* = Fy$, em que F é a matriz diagonal cujos elementos f_{ii} são os fatores de ajustamento correspondentes às observações contidas em y.

RESULTADOS E DISCUSSÃO

Esse processo de ajustamento conduz à obtenção de estimativas e predições viesadas, como se demonstra a seguir.

$$y^* = F(X\hat{\beta} + \hat{e}) = FX\hat{\beta} + F\hat{e} = F\begin{bmatrix} X_a & X_b \end{bmatrix}\begin{bmatrix} \hat{\beta}_a \\ \hat{\beta}_b \end{bmatrix} + F\hat{e} = FX_a\hat{\beta}_a + FX_b\hat{\beta}_b + F\hat{e}$$

$$= X_a\begin{bmatrix} \hat{\beta}_{a1} \\ \hat{\beta}_{a1} \\ \vdots \\ \hat{\beta}_{a1} \end{bmatrix} + FX_b\begin{bmatrix} \hat{\beta}_{b1} \\ \hat{\beta}_{b2} \\ \vdots \\ \hat{\beta}_{bq-1} \end{bmatrix} + F\hat{e} = \hat{\beta}_{a1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + FX_b\begin{bmatrix} \hat{\beta}_{b1} \\ \hat{\beta}_{b2} \\ \vdots \\ \hat{\beta}_{bq-1} \end{bmatrix} + F\hat{e} = \hat{\beta}_{a1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + F(X_b\hat{\beta}_b + \hat{e})$$

Desse resultado verifica-se que a estimação para os efeitos do fator B, após o ajustamento para o fator de efeito fixo A, produzirá estimativas viesadas pela matriz de ajustamento F. Por conseqüência as predições para a também estarão viesadas. As estimativas de componentes de variância obtidas a partir do vetor ajustado também serão viesadas, como se mostra a seguir.

Conforme LOPES et al. (1993), a estimação de componentes de variância a partir de formas quadráticas produzirá estimativas não-viesadas se a matriz núcleo da forma quadrática for ortogonal à matriz de incidência dos efeitos fixos de tal forma que o resultado do produto da matriz núcleo pelo vetor de observações seja o vetor de resíduos. A matriz que satisfaz esta condição é a matriz projetor ortogonal da parte aleatória das observações. Após o ajustamento, sob a restrição já imposta anteriormente para os efeitos do fator B, a matriz X^* de incidência dos efeitos fixos do novo modelo será a matriz de incidência do modelo inicial, adicionada de uma coluna referente à constante geral, mas sem as colunas referentes aos efeitos fixos do fator A. Isto posto, pode-se verificar se a condição de não-tendenciosidade é satisfeita.

Seja P a matriz projetor ortogonal no novo modelo após o ajustamento. Se o produto Py^* reproduzir o vetor $\hat{e}$ a condição de não-tendenciosidade estará satisfeita.

$$Py^* = P\{\hat{\beta}_{a1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + F(X_b\hat{\beta}_b + \hat{e})\}$$

Como o vetor coluna contendo elementos iguais ao número 1 é idêntico ao vetor coluna referente à constante no novo modelo o produto da matriz ortogonal por este vetor é nulo, tem-se

$$Py^* = PF(X_b\hat{\beta}_b + \hat{e}) = PFX_b\hat{\beta}_b + PF\hat{e}$$

Como $P = I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$, o produto $PF\hat{e} = [I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime}]F\hat{e}$ se torna $Py^* = PFX_b\hat{\beta}_b + F\hat{e}$

Verifica-se dessa forma que a condição de não-tendenciosidade não é satisfeita.

Para que estimativas e predições obtidas após ajustamento por fatores multiplicativos não sejam viesadas deve-se subtrair do vetor de observações ajustadas a média da subclasse escolhida como padrão e pré-multiplicar o resultado pela inversa da matriz de ajustamento, como mostrado a seguir.

$$F^{-1}(y^* - \hat{\beta}_{a1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}) = F^{-1}\{\hat{\beta}_{a1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + F(X_b\hat{\beta}_b + \hat{e}) - \hat{\beta}_{a1}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}\} = F^{-1}F(X_b\hat{\beta}_b + \hat{e}) = X_b\hat{\beta}_b + \hat{e}$$

Este procedimento é coerente com um modelo aditivo porque eqüivale a um processo de ajustamento por subtração, como segue.

$$y - X_a\hat{\beta}_a = X_a\hat{\beta}_a + X_b\hat{\beta}_b + \hat{e} - X_a\hat{\beta}_a$$

$$y - X_a\hat{\beta}_a = X_b\hat{\beta}_b + \hat{e}$$

Dessa forma o ajustamento não produzirá viés no processo de estimação e predição subseqüente.

Todavia, deve-se ter atenção ao fato de o ajustamento dos dados por subtração envolverá menor esforço computacional, visto que consiste apenas em subtrair de cada observação a média ou o efeito fixo correspondente da subclasse do fator que se deseja ajustar, não se justificando, portanto, o uso de fatores multiplicativos.

CONCLUSÃO

O uso do procedimento de ajustamento de dados por fatores multiplicativos causa alterações nos resultados de estimação e predição executados sobre os dados ajustados. Para que estimativas e predições não-viesadas sejam obtidas após o ajustamento por fatores multiplicativos deve-se subtrair do vetor de observações ajustadas a média da subclasse escolhida como padrão e pré-multiplicar o resultado pela inversa da matriz de fatores de ajustamento. No entanto, o ajustamento dos dados por subtração exigirá menor esforço computacional, não sendo recomendado o uso de ajustamento por fatores multiplicativos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. COOPER, J.B., HARGROVE, G.L. 1982. Age and months of calving adjustments of Holstein protein, milk and fat lactations yields. *J. Dairy Sci.* 65(8):1673-8.
2. KEOWN, J.F., EVERETT, R.W. 1985. Age-months adjustments factors for milk, fat and protein yields in Holstein cows. *J. Dairy Sci.* 68(10):2664-7.
3. LOPES, P.S.; MARTINS, E.N.; SILVA, M.A.; REGAZZI, A.J. 1993. *Estimação de componentes de variância*. Viçosa, Imprensa Universitária, UFV. 61p.

4. MARTINEZ, M.L., LEE, A.J., LIN,C.Y. 1990. Multiplicative age-season adjustments factors by maximun likelihood, gross comparisons and paired comparisons. *J. Dairy Sci.* 73(3):819-25.
5. NORMAN, H.D.; MEINERT, T.R.; SCHUTZ, M.M.; WRIGHT, J.R. 1995. Age and seasonal effects on Holstein yield for four regions of the United State over time. *J. Dairy Sci.* 78(8):1855-61.
6. TORRES, R.A. *Efeito da heterogeneidade de variância na avaliação gene´tica de bovinos da raça Holandesa no Brasil*. Belo Horizonte -MG: UFMG, 1998. 124p. Tese (Doutorado).