



Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA
Núcleo Tecnológico para Informática Agropecuária - NTIA
Rodovia SP 340, Km 105,4 - Caixa Postal 5010
13031 Campinas SP - Fone (0192) 42-9753

COMUNICADO TÉCNICO

CT-NTIA-8804
Setembro 1988

Absorção de Efeitos Principais em Modelos Lineares (*)

Maria Fernanda Moura
Bolsista PIEP, NTIA/EMBRAPA

I. Introdução

Em muitas áreas da pesquisa agropecuária, especialmente em pesquisa animal, os modelos utilizados possuem fatores com grande número de níveis. Nem sempre há a necessidade, ou o interesse, de se testar a significância dos efeitos de todos os fatores, mas apenas de um subconjunto deles, ajustado para os demais.

Absorção é a técnica que possibilita reduzir o modelo e testar apenas os efeitos do subconjunto de fatores de interesse.

A proposta para solucionar o problema, desenvolvida neste trabalho, aplica-se à absorção de efeitos principais. A técnica utilizada baseia-se em um processo de redução das equações normais do modelo, obtido diretamente de sua matriz de especificação, tendo como único pré-requisito, a nível computacional, a ordenação dos dados pelos efeitos a serem absorvidos.

Este trabalho é parte do projeto "Desenvolvimento de Módulos no SOC (Software Científico - EMBRAPA) para Análise de Dados Gerados pela Pesquisa Animal".

(*) Trabalho apresentado no VIII SINAPE, julho/88, IMPA/CNPq, RJ

II. Absorção de um Fator :

Suponha o seguinte modelo aditivo com dois fatores A e B:

$$y_{ijk} = u + a_i + b_j + e_{ijk}$$

onde :

$$i = 1, 2, \dots, p$$

$$j = 1, 2, \dots, q$$

$$k = 1, 2, \dots, n$$

ij

$$e_{ijk} \sim \text{i.i.d.}(0, \sigma^2)$$

e com as restrições usuais:

$$\begin{array}{c} \begin{array}{c} p \\ \hline \backslash \\ \quad a_i = 0 \\ / \\ \hline i=1 \end{array} \end{array} \quad \begin{array}{c} \begin{array}{c} q \\ \hline \backslash \\ \quad b_j = 0 \\ / \\ \hline j=1 \end{array} \end{array}$$

Dadas :

X : matriz de especificação (delineamento)
Y : vetor das variáveis observadas

as equações normais do modelo são :

$$X'X B = X'Y$$

Para resolver as equações, pode-se utilizar a seguinte estrutura :

$$(1) \begin{array}{c} \left| \begin{array}{cc} X'X & X'Y \\ \hline Y'X & Y'Y \end{array} \right| \end{array} \quad \begin{array}{c} \left| \right| \\ \text{de dimensão} \\ \left| \right| \end{array} (p+q) \times (p+q)$$

e aplicar o operador "SWEEP" (Goodnight, 1979) à estrutura, que a cada passo fornece somas de quadrados para cada efeito e ao final do processo uma estimativa do vetor B.

O processo de absorção, desenvolvido neste trabalho, foi adaptado a esse método de resolução.

Absorver $u + a_i$ equações, do modelo proposto, é construir a seguinte estrutura :

$$(2) \quad \begin{bmatrix} C & | & S \\ \hline s' & | & Q \end{bmatrix} \quad \text{de dimensão } qxq$$

A estrutura (2) corresponde à estrutura (1), após "p" iterações do SWEEP e removidas as partes que correspondem aos efeitos de u e A. Ou seja, se R de ordem (p+q) é a estrutura (1) após "p" iterações do SWEEP, e

$$T = \begin{bmatrix} 0 & I_q \\ q \times (p+q) & q \times p \end{bmatrix}$$

então : $\text{estrutura}(2) = T \cdot R \cdot T'$

Essa estrutura (2), pode ser obtida diretamente da matriz de especificação e do vetor Y.

Sejam :

$$X = \begin{bmatrix} u & a1 & a2 & \dots & a(p-1) & b1 & b2 & \dots & b(q-1) \\ 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & -1 & -1 & \dots & -1 & 0 & 0 & \dots & 1 \\ 1 & -1 & -1 & \dots & -1 & -1 & -1 & \dots & -1 \end{bmatrix}$$

e,

$$Y = \begin{bmatrix} y111 \\ y112 \\ y121 \\ y212 \\ \cdot \\ \cdot \\ ypqn \end{bmatrix}$$

ou ainda,

$X = [u \ a1 \ a2 \ \dots \ a(p-1) \ b1 \ b2 \ \dots \ b(q-1)]$
cujos elementos são vetores colunas definidos acima.

Então,

$$C = \begin{bmatrix} b_1' b_1 - \frac{\sum_{i=1}^p (u_i' b_1)^2}{n_1} & \dots & \dots \\ b_2' b_1 - \frac{\sum_{i=1}^p (u_i' b_2)(u_i' b_1)}{n_1} & b_2' b_2 - \frac{\sum_{i=1}^p (u_i' b_2)^2}{n_1} & \dots \\ \vdots & \vdots & \vdots \\ b' (q-1) b_1 - \frac{\sum_{i=1}^p (u_i' b(q-1))(u_i' b_1)}{n_1} & \dots & \dots \end{bmatrix}$$

$$S = \begin{bmatrix} Y' b_1 - \frac{\sum_{i=1}^p (u_i' y)(u_i' b_1)}{n_1} \\ \vdots \\ Y' b(q-1) - \frac{\sum_{i=1}^p (u_i' y)(u_i' b(q-1))}{n_1} \end{bmatrix}$$

e,

$$Q = \begin{bmatrix} Y' Y - \frac{\sum_{i=1}^p (u_i' y)^2}{n_1} \end{bmatrix}$$

$$C = \begin{bmatrix} \frac{p}{--} \frac{q}{--} & & & \\ & \backslash & \backslash & (u' c_1)^2 \\ & / & / & i_j i_j \\ & & & \frac{ni_j}{-----} \\ & i=1 & j=1 & \\ \dots & & & \dots \\ & \frac{p}{--} \frac{q}{--} & & \\ & \backslash & \backslash & (u' c(r-1)) (u' c_1) \\ & / & / & i_j i_j i_j i_j \\ & & & \frac{ni_j}{-----} \\ & i=1 & j=1 & \end{bmatrix}$$

$$S = \begin{bmatrix} \frac{p}{--} \frac{q}{--} & & & \\ & \backslash & \backslash & (u' y) (u' c_1) \\ & / & / & i_j i_j i_j i_j \\ & & & \frac{ni_j}{-----} \\ & i=1 & j=1 & \\ \dots & & & \\ & \frac{p}{--} \frac{q}{--} & & \\ & \backslash & \backslash & (u' y) (u' c(r-1)) \\ & / & / & i_j i_j i_j i_j \\ & & & \frac{ni_j}{-----} \\ & i=1 & j=1 & \end{bmatrix}$$

e,

$$Q = \begin{bmatrix} \frac{p}{--} \frac{q}{--} & & & \\ & \backslash & \backslash & (u' y)^2 \\ & / & / & i_j i_j \\ & & & \frac{ni_j}{-----} \\ & i=1 & j=1 & \end{bmatrix}$$

onde,

- u_i : partição do vetor u no nível i de A
 b_i : partição do vetor b no nível i de A
 y_i : partição do vetor Y no nível i de A
 n_i : frequência do nível i de A

Aplicando "SWEEP" à estrutura (2) obtem-se $SQ(b/u, a)$ e $SQRes$. Para obter $SQ(a/u)$ e SQT (corrigida pela media), pode-se utilizar alguns resultados que são calculados durante o processo de absorção (montagem da estrutura (2)), pois :

$$SQ(a/u) = \sum_{i=1}^p \frac{(u_i' y_i)^2}{n_i} - n \bar{y}^2 \quad e,$$

$$SQT(cor.) = Y'Y - n \bar{y}^2$$

Note que, a implementação computacional desses resultados é elementar, desde que os dados estejam ordenados por A .

Note ainda, que o resultado obtido para a absorção de u a1 equações é genérico, seja qual for o número de fatores não absorvidos no modelo (aditivo ou não).

III. Absorção de mais de um Fator :

Para absorver mais de um fator, usando o método proposto, os fatores devem seguir uma classificação hierárquica.

Por exemplo, dado o modelo :

$$y_{ijkl} = u + a_i + b_{j(i)} + c_k + e_{ijkl}$$

absorvendo $u + a_i + b_{j(i)}$, tem-se :

onde,

ck_{ij} : partição do vetor ck no nível j de B dentro do nível i de A

u_{ij} : partição do vetor u no nível j de B dentro do nível i de A

y_{ij} : partição do vetor y no nível j de B dentro do nível i de A

n_{ij} : frequência do nível j de B dentro do nível i de A

Assim, se os dados forem ordenados por A e B dentro de A , a implementação computacional é elementar.

Este resultado também generaliza-se para qualquer número de fatores a serem absorvidos, desde que obedecem a uma classificação hierárquica, e qualquer número de fatores não absorvidos.

IV. Bibliografia :

- ."A Tutorial on the SWEEP Operator "
James H. Goodnight
The American Statistician, August 1979, vol.33, n.3
- ."Least_Squares Analysis of Data with Unequal Subclass Numbers"
Walter R. Harvey - July 1968
Agricultural Research Service - U.S.Dept. Agriculture