

RT-MAE-9203

**PAINEL MULTIVARIADO - ANÁLISE ATRAVÉS
DO MODELO DE COMPONENTES DE VARIÂNCIA**

Wilton de Oliveira Bussab

Lúcia Pereira Barroso

Palavras Chaves: Variance Components Models, Multivariate
(Key words) Panels, Finite Populations.

Classificação AMS: 62K99
(AMS Classification)

Painel Multivariado - Análise Através do Modelo de Componentes de Variância

Wilton de Oliveira Bussab

Lúcia Pereira Barroso

Departamento de Estatística

Universidade de São Paulo

Caixa Postal 20570, Ag. Iguatemi

04198 - São Paulo, SP, Brasil

Abstract

In this paper we consider Variance Components Models for Multivariate Panels. We develop a rule for converting covariance matrix under infinite populations models into covariance matrix under finite populations models.

Key words: Variance Components Models, Multivariate Panels, Finite Populations.

1 Introdução

Em modelos de componentes de variância é usual supor que os efeitos aleatórios são extraídos de uma população infinita. Searle e Fawcett (1970) desenvolveram uma regra para converter esperanças dos quadrados médios para o modelo com população infinita em esperanças para o modelo com população finita e apresentaram estimadores da componente de variância entre classes para os dois modelos, no caso univariado.

O caso multivariado, para o modelo infinito, foi estudado por autores entre os quais Reinsel (1982, 1984, 1985) e Rao (1987).

Neste relatório apresentamos uma extensão da regra de conversão de Searle/Fawcett para o caso multidimensional, considerando o seu uso em painéis. Na seção 2, definimos o modelo adotado para medidas repetidas com as suposições para populações infinitas e finitas. Na seção 3, apresentamos a regra de conversão.

2 O painel multivariado

Painel é uma amostra selecionada num primeiro instante, sendo os dados coletados em várias ocasiões, sempre nessa mesma amostra. Quando, em cada instante, coletamos os dados em mais do que uma variável, a amostra em questão passa a formar um painel multivariado.

Estamos medindo p características, que indicaremos por $Y_1, Y_2, \dots, Y_p$, de n indivíduos em t ocasiões. Assim Y_{ijk} indicará a i -ésima característica correspondente ao j -ésimo indivíduo na k -ésima ocasião. Os dados podem ser representados como na tabela 1.

Tabela 1: DADOS

indivíduo	instantes				
	1	...	k	...	t
1	$\begin{pmatrix} Y_{111} \\ Y_{211} \\ \vdots \\ Y_{p11} \end{pmatrix}$	...	$\begin{pmatrix} Y_{11k} \\ Y_{21k} \\ \vdots \\ Y_{p1k} \end{pmatrix}$	...	$\begin{pmatrix} Y_{11t} \\ Y_{21t} \\ \vdots \\ Y_{p1t} \end{pmatrix}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\begin{pmatrix} Y_{1jt} \\ Y_{2jt} \\ \vdots \\ Y_{pjt} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1jk} \\ Y_{2jk} \\ \vdots \\ Y_{pjk} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1jt} \\ Y_{2jt} \\ \vdots \\ Y_{pjt} \end{pmatrix}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	$\begin{pmatrix} Y_{1jt} \\ Y_{2jt} \\ \vdots \\ Y_{pjt} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1jk} \\ Y_{2jk} \\ \vdots \\ Y_{pjk} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1jt} \\ Y_{2jt} \\ \vdots \\ Y_{pjt} \end{pmatrix}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\begin{pmatrix} Y_{1nt} \\ Y_{2nt} \\ \vdots \\ Y_{pnt} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1nk} \\ Y_{2nk} \\ \vdots \\ Y_{pnk} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1nt} \\ Y_{2nt} \\ \vdots \\ Y_{pnt} \end{pmatrix}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	$\begin{pmatrix} Y_{1nt} \\ Y_{2nt} \\ \vdots \\ Y_{pnt} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1nk} \\ Y_{2nk} \\ \vdots \\ Y_{pnk} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1nt} \\ Y_{2nt} \\ \vdots \\ Y_{pnt} \end{pmatrix}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$\begin{pmatrix} Y_{1nt} \\ Y_{2nt} \\ \vdots \\ Y_{pnt} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1nk} \\ Y_{2nk} \\ \vdots \\ Y_{pnk} \end{pmatrix}$	...	$\begin{pmatrix} Y_{1nt} \\ Y_{2nt} \\ \vdots \\ Y_{pnt} \end{pmatrix}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Adotamos o modelo usual para medidas repetidas, no qual o tempo é um fator fixo e o indivíduo um fator aleatório, ou seja, os dados seguem o seguinte modelo linear:

$$Y_{jk} = A_k + \alpha_j + e_{jk} ,$$

onde

$Y'_{jk} = (Y_{1jk}, Y_{2jk}, \dots, Y_{pjk})$,	vetor observado
$A'_k = (A_{1k}, A_{2k}, \dots, A_{pk})$,	vetor fixo medindo o efeito do tempo
$\alpha'_j = (\alpha_{1j}, \alpha_{2j}, \dots, \alpha_{pj})$,	fator aleatório medindo o efeito do indivíduo
$e'_{jk} = (e_{1jk}, e_{2jk}, \dots, e_{pjk})$,	vetor dos erros

As suposições sob os dois modelos, população infinita e finita, serão dadas adiante.

Seja Y o vetor de todas as observações, dado por:

$$Y' = (Y_{111}, Y_{211}, \dots, Y_{p11}, \dots, Y_{112}, Y_{212}, \dots, Y_{p12}, \dots, Y_{1n1}, Y_{2n1}, \dots, Y_{p n1}, \dots, Y_{1nt}, Y_{2nt}, \dots, Y_{pnt})$$

$$\text{com } E(Y) = \mu \quad Var(Y) = V.$$

A matriz de dispersão V exerce papel importante em problemas de testes de hipóteses ou de estimação, desenvolvidos quase sempre no caso de populações infinitas. Daí o interesse em estabelecer uma regra para converter essa matriz sob o modelo de população infinita (V_∞) para a matriz sob o modelo de população finita (V_F), para que possam ser adaptados modelos desenvolvidos de um caso para outro.

2.1 Suposições para o modelo de população infinita

Para quaisquer j, k , façamos

$$\alpha_j \sim N_p(0, \Sigma_\alpha)$$

$$e_{jk} \sim N_p(0, \Sigma_e)$$

com as suposições usuais:

α_j 's independentes

e_{jk} 's independentes

α_j 's e e_{jk} 's independentes

$$\begin{aligned} \Sigma_{\alpha} &= \begin{pmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1p} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{p1} & \tau_{p2} & \cdots & \tau_{pp} \end{pmatrix} & \Sigma_{\epsilon} &= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \\ (p \times p) & & (p \times p) & \end{aligned}$$

Assim,

$$\mathbf{V}_{\infty} = \begin{pmatrix} \mathbf{W}_{\infty} & 0 & \cdots & 0 \\ 0 & \mathbf{W}_{\infty} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{W}_{\infty} \end{pmatrix}$$

$(ntp \times ntp)$

com

$$\mathbf{W}_{\infty} = \begin{pmatrix} \Sigma_{\alpha} + \Sigma_{\epsilon} & \Sigma_{\alpha} & \cdots & \Sigma_{\alpha} \\ \Sigma_{\alpha} & \Sigma_{\alpha} + \Sigma_{\epsilon} & \cdots & \Sigma_{\alpha} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{\alpha} & \Sigma_{\alpha} & \cdots & \Sigma_{\alpha} + \Sigma_{\epsilon} \end{pmatrix}$$

$(tp \times tp)$

Então, $\mathbf{V}_{\infty}$ pode ser escrito como:

$$\mathbf{V}_{\infty} = \mathbf{I}_n \odot \mathbf{1}_t \mathbf{1}_t' \odot \Sigma_{\alpha} + \mathbf{I}_n \odot \mathbf{I}_t \odot \Sigma_{\epsilon} \quad (1)$$

onde

$\mathbf{I}_t$: matriz identidade de dimensão t

$\mathbf{1}_t$: vetor coluna de 1's de dimensão t

$\odot$: produto de Kronecker

2.2 Suposições para o modelo de população finita

Neste caso, supõe-se que n elementos são sorteados, sem reposição, de uma população de tamanho N e que, para qualquer i ,

$$\begin{aligned} \sum_{j=1}^N \alpha_{ij} &= 0 \\ \mathbf{S}_{ii'} &= \frac{1}{N-1} \sum_{j=1}^N \alpha_{ij} \alpha_{i'j} \end{aligned}$$

$$\underset{(p \times p)}{S_\alpha} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{pmatrix}$$

Vamos provar que:

$$\underset{(ntp \times ntp)}{V_F} = \begin{pmatrix} \mathbf{W}_F & \mathbf{U}_F & \cdots & \mathbf{U}_F \\ \mathbf{U}_F & \mathbf{W}_F & \cdots & \mathbf{U}_F \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{U}_F & \mathbf{U}_F & \cdots & \mathbf{W}_F \end{pmatrix}$$

onde

$$\underset{(tp \times tp)}{\mathbf{U}_F} = -\frac{1}{N} \begin{pmatrix} \mathbf{S}_\alpha & \mathbf{S}_\alpha & \cdots & \mathbf{S}_\alpha \\ \mathbf{S}_\alpha & \mathbf{S}_\alpha & \cdots & \mathbf{S}_\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_\alpha & \mathbf{S}_\alpha & \cdots & \mathbf{S}_\alpha \end{pmatrix}$$

e

$$\underset{(tp \times tp)}{\mathbf{W}_F} = (1 - N^{-1}) \begin{pmatrix} \mathbf{S}_\alpha & \mathbf{S}_\alpha & \cdots & \mathbf{S}_\alpha \\ \mathbf{S}_\alpha & \mathbf{S}_\alpha & \cdots & \mathbf{S}_\alpha \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_\alpha & \mathbf{S}_\alpha & \cdots & \mathbf{S}_\alpha \end{pmatrix} + \begin{pmatrix} \Sigma_\epsilon & 0 & \cdots & 0 \\ 0 & \Sigma_\epsilon & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_\epsilon \end{pmatrix}$$

ou seja,

$$\mathbf{V}_F = \mathbf{I}_n \otimes \mathbf{1}_t \mathbf{1}_t' \otimes \mathbf{S}_\alpha - \frac{1}{N} \mathbf{1}_n \mathbf{1}_n' \otimes \mathbf{1}_t \mathbf{1}_t' \otimes \mathbf{S}_\alpha + \mathbf{I}_n \otimes \mathbf{I}_t \otimes \Sigma_\epsilon. \quad (2)$$

Observando que:

i)

$$0 = \left(\sum_{j=1}^N \alpha_{ij} \right) \left(\sum_{j'=1}^N \alpha_{i'j'} \right) = \sum_{j=1}^N \alpha_{ij} \alpha_{i'j} + \sum_{j=1}^N \sum_{\substack{j'=1 \\ j' \neq j}}^N \alpha_{ij} \alpha_{i'j'} = (N-1)S_{ii'} + \sum_{j=1}^N \sum_{\substack{j'=1 \\ j' \neq j}}^N \alpha_{ij} \alpha_{i'j'}$$

$$\Rightarrow \sum_{j=1}^N \sum_{\substack{j'=1 \\ j' \neq j}}^N \alpha_{ij} \alpha_{i'j'} = -(N-1)S_{ii'}$$

ii)

$$E(\alpha_{ij}) = \frac{1}{N} \sum_{j=1}^N \alpha_{ij} = 0$$

$$Cov(\alpha_{ij}, \alpha_{i'j}) = E(\alpha_{ij}\alpha_{i'j}) = \frac{1}{N} \sum_{j=1}^N \alpha_{ij}\alpha_{i'j} = (1 - N^{-1})S_{ii'}$$

$$Cov(\alpha_{ij}, \alpha_{i'j'}) = E(\alpha_{ij}\alpha_{i'j'}) = \frac{1}{N(N-1)} \sum_{j=1}^N \sum_{\substack{j'=1 \\ j' \neq j}}^N \alpha_{ij}\alpha_{i'j'} = -\frac{S_{ii'}}{N}$$

podemos calcular as covariâncias entre os elementos do vetor $\mathbf{Y}$. Assim:

dentro do indivíduo

no mesmo tempo

$$Cov(Y_{ijk}, Y_{i'jk}) = Cov(A_{ik} + \alpha_{ij} + e_{ijk}, A_{i'k} + \alpha_{i'j} + e_{i'jk}) = (1 - N^{-1})S_{ii'} + \sigma_{ii'}$$

entre tempos

$$Cov(Y_{ijk}, Y_{i'j'k'}) = Cov(A_{ik} + \alpha_{ij} + e_{ijk}, A_{i'k'} + \alpha_{i'j'} + e_{i'j'k'}) = (1 - N^{-1})S_{ii'}$$

entre indivíduos

$$Cov(Y_{ijk}, Y_{i'j'k'}) = Cov(A_{ik} + \alpha_{ij} + e_{ijk}, A_{i'k'} + \alpha_{i'j'} + e_{i'j'k'}) = -\frac{S_{ii'}}{N}$$

A matriz $\mathbf{W}_F$ reflete as covariâncias "dentro do indivíduo", enquanto que $\mathbf{U}_F$ reflete as covariâncias "entre indivíduos", devidas ao sorteio sem reposição.

Para se chegar à expressão (2) basta escrever a matriz $\mathbf{V}_F$ detalhadamente e identificar a igualdade.

A comparação entre as expressões (1) e (2) leva à mesma idéia que aparece em amostragem, quando da multiplicação pelo fator de correção para populações finitas no cálculo das variâncias dos estimadores (Cochran (1977)).

3 A regra de conversão

Para obtermos a regra de conversão da matriz $\mathbf{V}_\infty$ para a matriz $\mathbf{V}_F$ basta verificar que os elementos que as compõem são:

Fonte	Símbolo	em V_{∞}	em V_F
Dentro do indivíduo			
no mesmo tempo	$Cov(Y_{ijk}, Y_{ijk})$	$\tau_{ii'} + \sigma_{ii'}$	$(1 - N^{-1})S_{ii'} + \sigma_{ii'}$
entre tempos	$Cov(Y_{ijk}, Y_{ijk'})$	$\tau_{ii'}$	$(1 - N^{-1})S_{ii'}$
Entre indivíduos	$Cov(Y_{ijk}, Y_{ijk'})$	0	$-N^{-1}S_{ii'}$

Assim, a regra de conversão consiste em substituir $\tau_{ii'}$ por $S_{ii'}$ e deduzir de cada elemento a parcela $\frac{S_{ii'}}{N}$.

Referências

- [1] Cochran, W.G. (1977). *Sampling Techniques*. New York, John Wiley, 428 p.
- [2] Rao, C. R. (1987). Prediction of future observations in growth curve models. *Statistical Science*, vol 2, 4, 434-471.
- [3] Reinsel, G. C. (1982). Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *JASA*, **377**, 190-195.
- [4] Reinsel, G. C. (1984). Estimation and prediction in a multivariate random-effects generalized linear model. *JASA*, **386**, 406-414.
- [5] Reinsel, G. C. (1985). Mean squared error properties of empirical bayes estimators in a multivariate random effects general linear model. *JASA*, **391**, 642-650.
- [6] Searle, S. R. e Fawcett, R. F. (1970). Expected mean squares in variance components models having finite populations. *Biometrics*, **26**, 243-254.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

1991

- 9101 - RAVISHANKAR, K.; Convergence of a random motion to a point, 11p.
- 9102 - FERRARI, P.A. & KAVISHANKAR, R.; Shocks in asymmetric exclusion automata, 17p.
- 9103 - ANDRE, C.D.S. & PERES, C.A.; An Algorithm for the Msaee Estimation of the Multistage Dose-Response model, 14p.
- 9104 - FERRARI, P.A.; Microscopic shocks in one dimensional driven systems, 21p.
- 9105 - BOLFARINE, H., LEITE, J.G. & RODRIGUES, J.; On the Estimation of the Size of a Finite and Closed Population, 17p.
- 9106 - IGLESIAS, P., SANDOVAL, M.C. & PEREIRA, C.A.B.; Predictive Likelihood in Finite Populations, 17p.
- 9107 - CORDEIRO, G.M., PAULA, G.A. & BOTTER, D.A.; Improved Likelihood Ratio Tests for Dispersion Models, 32p.
- 9108 - WERKEMA, H.C.C. & TANAKA, N.I.; Alguns Aspectos sobre Filtragens em Filas, $M/M^{1/2}/N$ com "Feedback", 35p.
- 9109 - WECHSLER, S.; Bayesian Note on Game Theory, 3p.
- 9110 - WECHSLER, S.; Rationality in Rubin's Sense, 16p.
- 9111 - BOLFARINE, H., ELIAN, S.N., RODRIGUES, J. & ZACKS, S.; Optimal Prediction of the Finite Population Regression Coefficient in Finite Populations, 13p.
- 9112 - PAULA, G.A.; Assessing Local Influence in Restricted Regression Models, 20p.
- 9113 - RODRIGUES, F.W. & WECHSLER, S.; A Discrete Bayes Explanation of a Failure Rate Paradox, 03p.
- 9114 - FERRARI, P.A., MARTINEZ, S. & PICCO, P. Existence of Quasi Stationary Distributions. A Renewal Dynamical Approach, 13p.
- 9115 - FERRARI, P.A., MARTINEZ, S. & PICCO, P. Existence of Non Trivial Quasi Stationary Distributions in the Birth and Death Chain, 20p.
- 9116 - FERRARI, P.A.; Shcock in the Burgers Equation and the Asymmetric Simple Exclusion Process, 40p.

9117 - BOLFARINE, H., NASCIMENTO, J.A. & RODRIGUES, J. On the Estimation of the True x Values in a Simple Regression Model When Both Variables Have Error, 15p.

9118 - BUSSAB, W.O. & ARTES, R. Recuperação de Salários Brutos Através de Salários Líquidos, 12p.

9119 - PEREIRA, C.A.B. & ROGATKO, A. Maternal Glycemia and Cardiotocography: A Bayesian Analysis, 11p.

9120 - LEITE, J.G., PEREIRA, C.A.B. & RODRIGUES, F.U. Waiting Time to Exhaust Lottery Numbers, 8p.

9121 - INOUE, L.Y.T. Inferência Comparada em Ensaios Clínicos: A Violação do Princípio da Verossimilhança, 28p.

1992

9201 - BOLFARINE, H., NASCIMENTO, J.A. & RODRIGUES, J. Comparing Several Regression Models with Measurement Errors. A Bayesian Approach, 16p.

9202 - BOLFARINE, H. & SANDOVAL, M.C. Empirical Bayesian Prediction in the Location Error in Variables Superpopulation Model, 26p.

The complete list of Relatórios do Departamento de Estatística, INE-USP, will be sent upon request.

- Departamento de Estatística
INE-USP
Caixa Postal 20.570
01498 - São Paulo, Brasil