Unbalanced multivariate repeated measurements with a Kronecker product structured covariance matrix

ISSN 1817 – 2695

Abdul Hussein Saber AL-Mouel and Rana Rony Faik

[1]*Department of Mathematics-College of Education*
[2]*Department of Mathematics-College of Science - University of Basrah*

((Received 2/6/2010 , Accepted 19/1/2011))

Abstract

In this article we consider a set of repeated measurements on p variables on each of the n individuals. Thus, data on each individual is unbalanced data. The n individuals themselves may be divided and randomly assigned to g groups. We test the usual hypotheses and studied the likehood ratio test of this test ,assuming that the data on an individual has a covariance matrix which is a Kronecker product of two positive definite matrices.

Key Words: Multivariate repeated **,** kronecker product , covariance matrix

(1-1)Introduction

Measurements on a variable (or a characteristic) made at several occasions or under different treatment conditions on the same experimental unit, lead to repeated measures (or longitudinal) data. Repeated measures for unbalanced data routinely occur in many diverse fields, such as medicine, psychology and education. The unbalanced data means that the occasions of measurement are not the same for all of the experimental units. Most previously published work regarding linear models is with the unbalanced data and correlated errors. Jennrich and Schluchter (1986)[4] , have used likelihood ratio test for unbalanced repeated measures models with structured covariance matrices .Also J. C. Keselman& H. J. Keselman(1990)[8] have used the unbalanced repeated measures designs .In our work we assume that the data on an individual has a covariance matrix which is a Kronecker product of two positive definite matrices.One of this matrices has an AR(1) structure . We used this assuming to test the usual hypotheses and studied the likehood ratio test of this test .

(2-1)Unbalanced data

In practice, the measurements on the individuals at all the t time points may not be available. This leads to unbalanced data. Suppose we have t_{ij} repeated measurements on p variables on the jth individual ($j = 1,...,n_i$)from the ith group ($i = 1, . . . , g$). Let $Y'_{ij} = \left(Y'_{ij1},..,Y'_{ijt_{ij}}\right)$ be the observational vector representing these measurements. Assume $\Omega = Cov(y_{ij}) = V_{ij} \otimes \Sigma$,where V_{ij} is a $t_{ij} \times t_{ij}$ positive definite matrix and, as before, Σ is a $p \times p$ positive definite matrix. The covariance matrix V_{ij} is the $t_{ij} \times t_{ij}$ submatrix of a $t \times t$ positive definite matrix V, t being the maximum of t_{ij} .

Consider the multivariate linear model $Y = XB + E$ where Y is the $n \times p$ observation matrix, obtained by stacking all $n = \sum_{i=1}^{g} \sum_{j=1}^{n_i} t_{ij}$, $p \times 1$ vectors,

Y_{ijk}, as rows and B is the m × p matrix of unknown parameters. Next, the design matrix X is obtained by stacking all the t_{ij} × m matrices X_{ij} one below the other, where X_{ij} is the design matrix corresponding to the jth individual in the ith group .

Let $\xi_{ij} = y_{ij} - E(y_{ij})$. Then the log-likelihood function, assuming that ξ_{ij}, i = 1, . . . , g; j = 1, . . . , n_i, forms a random sample from pt_{ij}-variate normal distribution with zero mean vector and covariance matrix $V_{ij} \otimes \Sigma$ and $n = \sum_{i=1}^{g}\sum_{j=1}^{n_i} t_{ij}$. Then the log-likelihood function, is proportional to

$$-\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i} \ln\left|V_{ij}\otimes\Sigma\right| - \frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\xi_{ij}\right)'(V_{ij}^{-1}\otimes\Sigma^{-1})\left(\xi_{ij}\right) \qquad \ldots(2-1-1)$$

Next, partition the $pt_{ij}\times 1$ vector ξ_{ij} into t_{ij} blocks of p × 1 vectors such that $\xi_{ij} = \left(\xi_{ij1},\cdots,\xi_{ijt_{ij}}\right)'$. Using this partition, rewrite (2-1-1) as

$$-\frac{p}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\ln\left|V_{ij}\right| - \frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i} t_{ij}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\xi_{ij}\right)'(V_{ij}^{-1}\otimes\Sigma^{-1})\left(\xi_{ij}\right) \quad \ldots(2-1-2)$$

Now the maximum likelihood estimate of Σ is given by

$$\hat{\Sigma} = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sum_{k=1}^{t_{ij}}\sum_{l=1}^{t_{ij}} \hat{v}_{kl}^{*}\,(\xi_{ijk})(\xi_{ijl})'}{\sum_{i=1}^{g}\sum_{j=1}^{n_i} t_{ij}} \qquad \ldots(2-1-3)$$

Where $\hat{v}_{kl}^{*}$ is the (k,l)th element of V_{ij}^{-1}.

Since V_{ij} has an AR(1) structure,

$$V_{ij} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{t_{ij}-1} \\ \rho & 1 & \rho & \cdots & \rho^{t_{ij}-2} \\ \rho^2 & \rho^2 & 1 & \cdots & \rho^{t_{ij}-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{t_{ij}-1} & \rho^{t_{ij}-2} & \rho^{t_{ij}-3} & \cdots & 1 \end{bmatrix}$$

It is well known that the determinant of this matrix is given by

$$\left|V_{ij}\right| = \left(1-\rho^2\right)^{t_{ij}-1}$$

and its inverse is

$$V_{ij}^{-1} = \frac{1}{(1-\rho^2)}\left[I_{t_{ij}} + \rho^2 C_{1ij} + \rho C_{2ij}\right]$$

(see, Roy, A. and Khattree, R.(2004))[5]).

Where C_{1ij} and C_{2ij} are $t_{ij} \times t_{ij}$ matrices such that C_1 = diag (0, 1, . . . , 1, 0) and C_2 is a tridiagonal matrix with 0 on the diagonal and 1 on the upper and lower diagonals.

Next, using

and V_{ij}^{-1}, $\left|V_{ij}\right| = \left(1-\rho^2\right)^{t_{ij}-1}$

rewrite (2-1-1) as

$$-\frac{p}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(t_{ij}-1\right)\ln\left(1-\rho^2\right)-\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i} t_{ij}\ln\left|\Sigma\right|-\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\xi_{ij}\right)'$$

$$\left(\frac{1}{(1-\rho^2)}\left[I_{t_{ij}}\otimes\Sigma^{-1}+\rho^2 C_{1ij}\otimes\Sigma^{-1}+\rho C_{2ij}\otimes\Sigma^{-1}\right]\right)\left(\xi_{ij}\right) \qquad \ldots(2-1-4)$$

Then ,

$$z=\frac{\partial \ln L}{\partial\rho}=0 \qquad \Rightarrow z=\frac{pa\rho}{(1-\rho^2)}-\frac{d\rho}{(1-\rho^2)^2}-\frac{1}{2}\frac{(-1-\rho^2)}{(1-\rho^2)^2}C-\frac{b\rho}{(1-\rho^2)}=0$$

$$2p\rho\left(1-\rho^2\right)a + C + C\rho^2 - 2b\rho - 2d\rho = 0$$

$$-2pa\rho^3 + C\rho^2 + \left[2pa - 2\left(d+b\right)\right]\rho + C = 0 \qquad \ldots(2-1-5)$$

Where ,

$$a=\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(t_{ij}-1\right)$$

$$b=\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\hat{\xi}_{ij}\right)'\left(C_{1ij}\otimes\Sigma^{-1}\right)\left(\hat{\xi}_{ij}\right)$$

$$c=\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\hat{\xi}_{ij}\right)'\left(C_{2ij}\otimes\Sigma^{-1}\right)\left(\hat{\xi}_{ij}\right)$$

$$d=\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\hat{\xi}_{ij}\right)'\left(I_{t_{ij}}\otimes\Sigma^{-1}\right)\left(\hat{\xi}_{ij}\right) \qquad \ldots(2-1-6)$$

However,

$$\hat{\Omega}=\hat{V}_{ij}\otimes\hat{\Sigma} \qquad \ldots(2-1-7)$$

where $\hat{V}_{ij}$ and $\hat{\Sigma}$ are the maximum likelihood estimates of V_{ij} and Σ respectively. It must be emphasized that V_{ij} , having the AR(1) structure is a function of ρ which must be estimated along with Σ .The maximum likelihood estimates $\hat{\Sigma}$ and $\hat{\rho}$ are obtained by simultaneously and iteratively solving the above equations (2-1-3) and (2-1-5). Thus, the maximum likelihood estimate $\hat{V}_{ij}$ is obtained from V_{ij} by replacing ρ by $\hat{\rho}$.

(2-2)The likelihood ratio test:

The likelihood ratio test compares the maximum value of the likelihood function L restricted to the region defined by the null hypothesis H_0, to the maximum of likelihood function L in the

$$\lambda = \frac{\max_{H_0} L}{\max_{H_a} L} \qquad ...(2-2-1)$$

unrestricted Region, H_a. Thus the ratio or a function of it, is used as the test statistic to test the null hypothesis H_o.
Where,
$H_o : \Omega = Cov(y_{ij}) = V_{ij} \otimes \Sigma$ vs. $H_a : \Omega =$ unstructured

It is well known that, for large samples and under normality assumption L = -2ln λ is approximately χ^2_α under H_o where the degrees of freedom (d.f.)α are equal to the number of parameters estimated under H_a minus the number of parameters estimated under H_o. To compute the test statistic -2ln λ, given by:

$$-2\ln\lambda = [-2\ln \max_{H_0} L(\mu, \Sigma, V_{ij}; Y)] - [-2\ln \max_{H_a} L(\mu, \Omega; Y)]$$

The likelihood functions under both structured H_0 ($\Omega = V_{ij} \otimes \Sigma$, V_{ij} AR (1)) and under H_a (Ω unstructured) are to be maximized separately. Now, the maximum likelihood estimates of V_{ij} and Σ is defined in equation (2-1-5) and (2-1-3) under the H_0.

Also, the maximum likelihood estimates of Ω under the H_a and H_0 are

$$\hat{\Omega}_a = \frac{1}{M}\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\xi_{ij})(\xi_{ij})' \qquad \text{where} \quad S = \sum_{i=1}^{g} n_i \qquad ...(2-2-2)$$

$$\hat{\Omega}_0 = \hat{V}_{ij} \otimes \hat{\Sigma}_0 \qquad ...(2-2-3)$$

Therefore, substituting the corresponding maximum likelihood estimates of the parameters into the likelihood function, the likelihood ratio is given by

$$\lambda = \frac{\left|\frac{B}{S}\right|^{\frac{M}{2}} EXP\left[-\frac{1}{2}\left(tr\,\hat{V}_{ij}^{-1}\right)\left(tr\,\hat{\Sigma}^{-1}\right)B\right]}{\left(\prod_{i=1}^{g}\prod_{j=1}^{n_i}\left|\hat{V}_{ij}^{-1}\right|^{\frac{p}{2}}\left|\hat{\Sigma}\right|^{\frac{t_{ij}}{2}}\right) EXP\left[-\frac{1}{2}Mpt_{ij}\right]} \qquad ...(2-2-4)$$

for i =1 , … , g , j = 1 , … , n_i

Where

$$B = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\xi_{ij})(\xi_{ij})' \qquad ...(2-2-5)$$

References

1- M.J.,Crowder and D.T. ,Hand" Analysis of Repeated Measures ", Chapman and Hall, London,(1990).
2- D.N. Geary " Modeling the Covariance Structure of Repeated Measuremements ", Biometrics, 45, 1153-1195, (1989).
3- D. F. , Morrison" Multivariate Statistical Methods", New York, McGraw Hill, (1976).
4- R.I.,Jennrich and M.D., Schluchter "Unbalanced Repeated Measures Models with Structured Covariance Matrices", Biometrics, 42, 805-820, (1986).
5- A. Roy and R. Khattree"On Implementation of a Test for Kronecker Product Covariance Structure for Multivariate Repeated Measures Data," Submitted, (2004).
6- N. H. Timm"Multivariate analysis of variance of repeated measurements". In: P. R. Krishnaiah(ed), Handbook of Statistics, vol. 1 (pp. 41- 87) (New York, North-Holland) , (1980).
7- E.F. Vonesh and V.M. Chinchilli" Linear and Nonlinear Models for the Analysis of Repeated Measurements ", Marcel Dakker, Inc., New York, (1997).
8- J. C. Keselman& H. J. Keselman, " Analysing unbalanced repeated measures designs", British Journal of Mathematical and Statistical Psychology, 43, pp. 265-282,(1990).
9- A.Roy," Some Contributions to Discrimination and Classification with Repeated Measures Data with Special Emphasis on Biomedical Applications", Ph.D. Dissertation, Department of Mathematics and Statistics, Oakland University, Rochester, MI, (2002).
10-A.H.S. AL Moul & R.R.Faik" Multivariate Repeated Measurements with a Kronecker Product Structured Covariance matrix"Iraq, Journal of Qudisiah of science of math. and computer ,Vol.1,No.1,(pp 241-247),(2009).

القياسات المتكررة المتعددة المتغيرات غير المتزنة مع ضرب كرونيكر لبنية مصفوفة التباين المشترك

عبد الحسين صبر الموويل[1] و رنا روني فائق[2]

[1] قسم الرياضيات / كلية التربية

[2] قسم الرياضيات / كلية العلوم

جامعة البصرة

المستخلص

تم في هذا البحث اعتماد مجموعة من القياسات المتكررة على p من المتغيرات لكل n من الوحدات التجريبية .لذا فــان البيانات في كل وحدة تجريبية هي بيانات غير متزنة . وتم تقسيم الوحدات التجريبية إلى g من المجموعات (groups) وقد تــم اختبار الفرضيات وايجاد المعيار لذلك الاختبار وافترضنا ان مصفوفة التباين المشترك لتلك البيانات تكون بشكل ضرب كرونيكر لاثنتين من المصفوفات .