The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 347-355

Influence Analysis on a Test Statistic in Canonical Correlation Analysis[1)]

Kang-Mo Jung[2)]

Abstract

We propose a method for detecting influential observations that have a large influence on the likelihood ratio test statistic for the two sets of variables are uncorrelated with one another. For this purpose we derive a local influence measure for the likelihood ratio test statistic under certain perturbation scheme. An illustrative example is given to show the effectiveness of the proposed method on the identification of influential observations.

Keywords : Canonical correlation analysis, influential observations, local influence.

1. Introduction

The detection of outliers and influential observations has a long history. However, many diagnostic measures have been proposed for influence analyses in the context of estimation. A few works that treat detection of influential observations for test statistics in multivariate analysis are found. Among others, Jung and Kim (1999) investigated the influence of observations on the likelihood ratio test (LRT) statistics in the multivariate Behrens-Fisher problem based on influence curves. Influence analysis in testing problems is very important because in extreme situations, few observations can dominate our conclusion about the hypothesis as can be seen in Section 5.

The LRT statistic in the canonical correlation analysis contains the canonical correlation coefficients which can be obtained by the eigenvalues of the matrix composed of partitioned covariance matrices. It is well known that the covariance matrix is very sensitive to influential observations, and so is the LRT statistic. Case deletion diagnostics are widely used in many statistical analyses (Cook and Weisberg, 1982). However, case deletion diagnostics require

1) This work was supported by Kunsan National University Grant (KSNU-2000).
2) Assistant Professor, Department of Informatics and Statistics, Kunsan National University, Kunsan, 573-701, Korea.
E-mail : kmjung@kunsan.ac.kr

amount of computation time. The local influence method was inspired by Cook (1986) as a general method of assessing the influence of perturbations of the model, and was adapted to canonical correlation analysis by Jung (2000).

In this work the method of local influence is adapted to the LRT statistics for testing the hypothesis that the two sets of variables are uncorrelated with one another, or the hypothesis that some canonical correlation coefficients are not significant. Jung (2000) considered the perturbation scheme which is free of distribution, because usually canonical correlation analysis does not assume that the data is normally distributed. However, the LRT statistic for the testing problems needs the normality. Therefore we considered the perturbation based on the normal population. It is well known that the sample covariance matrix is more sensitive to influential observations than the sample mean vector. Thus, for getting a local influence measure the perturbation is chosen in which a weight is put on the covariance matrix for an observation. Under this perturbation scheme, the parameters of interest are estimated. Then the perturbation vector and the perturbed estimator form a surface in a certain Euclidean space. We use the slope maximizing the slope vector of the surface for investigating the influence of observations on the LRT statistics.

Section 2 discusses the preliminaries for canonical correlation analysis. We described the local influence procedure and derivation on the LRT statistics in Sections 3 and 4, respectively. In Section 5, a numerical example is given and it will show that the local influence method is effective.

2. LRT Statistic in Canonical Correlation Analysis

Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ are p and q dimensional random vectors, respectively. Suppose further that $\boldsymbol{x}$ and $\boldsymbol{y}$ have means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively, and that

$$E[(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)^T]=\boldsymbol{\Sigma}_{11}, \quad E[(\boldsymbol{y}-\boldsymbol{\mu}_2)(\boldsymbol{y}-\boldsymbol{\mu}_2)^T]=\boldsymbol{\Sigma}_{22},$$
$$E[(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{y}-\boldsymbol{\mu}_2)^T]=\boldsymbol{\Sigma}_{12}=\boldsymbol{\Sigma}_{21}^T .$$

Then the random vector $\boldsymbol{z}=(\boldsymbol{x}^T, \boldsymbol{y}^T)^T$ has the covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is partitioned such that

$$\boldsymbol{\Sigma}=\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

The squared canonical correlation coefficient ρ_i^2 for $i=1,\cdots,r$, and $r=\min(p,q)$ are the positive roots of the generalized eigenvalue problem

$$|\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}-\rho_i^2\boldsymbol{\Sigma}_{11}|=0 ,$$

with $\rho_1^2 > \cdots > \rho_r^2$ and $\rho_j^2 = 0$ if $j > r$. For $i = 1, \cdots, r$ the canonical vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ satisfy the following equations

$$(\boldsymbol{\Sigma}_{12} \ \boldsymbol{\Sigma}_{22}^{-1} \ \boldsymbol{\Sigma}_{21} - \rho_i^2 \boldsymbol{\Sigma}_{11}) \boldsymbol{\alpha}_i = \mathbf{0} ,$$
$$(\boldsymbol{\Sigma}_{21} \ \boldsymbol{\Sigma}_{11}^{-1} \ \boldsymbol{\Sigma}_{12} - \rho_i^2 \boldsymbol{\Sigma}_{22}) \boldsymbol{\beta}_i = \mathbf{0} ,$$

with the normalized constraints

$$\boldsymbol{\alpha}_i^T \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha}_j = \boldsymbol{\beta}_i^T \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_j = \delta_{ij} ,$$

where δ_{ij} is the Kronecker delta.

Consider the hypothesis

$$H_0 : \boldsymbol{\Sigma}_{12} = \mathbf{0} , \tag{1}$$

which means the two sets of variables are uncorrelated with one another. Under the normality, the LRT statistic for testing H_0 is given by

$$T = -\{n - (p+q+3)/2\} \log \prod_{i=1}^{r} (1 - \widehat{\rho_i^2}), \tag{2}$$

where n is the number of observations, it can be obtained by using Bartlett's approximation, and it is approximately distributed as a chi-squared distribution with pq degrees of freedom (Mardia, et al., 1979, p. 288).

Note that Bartlett proposed a similar statistic to test the hypothesis that only m of the population canonical correlation coefficients are non-zero. This test is based on the statistic

$$-\{n - (p+q+3)/2\} \log \prod_{i=m+1}^{r} (1 - \widehat{\rho_i^2})$$

and it is asymptotically chi-squared distributed with $(p-m)(q-m)$ degrees of freedom. (Mardia, et al., 1979, p. 289).

3. Local Influence

Let τ be a statistic and denote by $\tau(\boldsymbol{w})$ the perturbed statistic under a perturbation scheme which can be characterized by a perturbation vector $\boldsymbol{w} = (w_1, \cdots, w_n)^T$. The perturbation vector is expressed as $w_u = 1 + a l_u$ for $u = 1, \cdots, n$. The scalar a represents the magnitude of the perturbation and the vector $\boldsymbol{l} = (l_1, \cdots, l_n)^T$ of unit length its direction. The perturbation scheme is chosen such that $\tau = \tau(\boldsymbol{w}_0)$, where $\boldsymbol{w}_0$ is called the null point.

The $(n+1)$ by 1 vector $(\boldsymbol{w}^T, \tau(\boldsymbol{w}))^T$ forms a surface in the $(n+1)$-dimensional

Euclidean space as $\boldsymbol{w}$ varies over a certain space. The direction vector $\boldsymbol{l}_{\max}$ maximizes the slope of the surface at $\boldsymbol{w}=\boldsymbol{w}_0$ and gives influence information on the statistic (Lawrance, 1988). The maximum vector can be obtained by

$$\boldsymbol{l}_{\max}=\dot{\tau}/\sqrt{\dot{\tau}^T\dot{\tau}}\ ,$$

where $\dot{\tau}=\left(\left.\frac{\partial\tau(\boldsymbol{w})}{\partial w_1}\right|_{\boldsymbol{w}=\boldsymbol{w}_0},\cdots,\left.\frac{\partial\tau(\boldsymbol{w})}{\partial w_n}\right|_{\boldsymbol{w}=\boldsymbol{w}_0}\right)^T$. Unlike Cook's likelihood displacement, in this case $\boldsymbol{l}_{\max}$ does not vanish, so it provides valuable information about the local behaviour of $\tau(\boldsymbol{w})$.

To get a local influence measure on the LRT (2), we will derive the first order partial derivatives of $T(\boldsymbol{w})$ evaluated at the null point under a perturbation scheme, which is given by

$$T(\boldsymbol{w})=-\{n-(p+q+3)/2\}\log\prod_{i=1}^{r}(1-\hat{\rho}_i^2(\boldsymbol{w})) \tag{3}$$

where $\hat{\rho}_i(\boldsymbol{w})$ is the perturbed canonical correlation coefficient. For investigating the influence of observations on the LRT (2), it is enough to get the first order partial derivatives of $\hat{\rho}_i(\boldsymbol{w})$.

4. Derivation

We consider a perturbation which the uth observation $\boldsymbol{z}_u$ is perturbed according to

$$\boldsymbol{z}_u\sim N(\boldsymbol{\mu},\ \boldsymbol{\Sigma}/w_u)\ , \tag{4}$$

for $u=1,\cdots,n$. When $w_u=1$ for all u, the perturbed model (4) reduces to the unperturbed model, that is the null point $\boldsymbol{w}_0$ becomes 1_n, where $1_n=(1,\cdots,1)^T$ of order n. The maximum likelihood estimators for the mean vector and covariance matrix under the perturbed model are given by

$$\begin{aligned}\hat{\boldsymbol{\mu}}(\boldsymbol{w}) &= \sum_{u=1}^{n} w_u\,\boldsymbol{z}_u/\sum_{v=1}^{n} w_v,\\ S(\boldsymbol{w}) &= \sum_{u=1}^{n} w_u(\boldsymbol{z}_u-\hat{\boldsymbol{\mu}}(\boldsymbol{w}))(\boldsymbol{z}_u-\hat{\boldsymbol{\mu}}(\boldsymbol{w}))^T/n.\end{aligned} \tag{5}$$

Such perturbation was termed the case-weights perturbation scheme by Cook (1986).

Under the perturbation in equation (4), the estimators $\hat{\rho}_i(\boldsymbol{w})$, $\hat{\boldsymbol{\alpha}}_i(\boldsymbol{w})$ and $\hat{\boldsymbol{\beta}}_i(\boldsymbol{w})$ satisfy

$$|\,Q(\boldsymbol{w})-\hat{\rho}_i^{\,2}(\boldsymbol{w})\,S_{11}(\boldsymbol{w})|=0 \tag{6}$$

where $Q(\boldsymbol{w})=S_{12}(\boldsymbol{w})\,S_{22}^{-1}(\boldsymbol{w})\,S_{21}(\boldsymbol{w})$ and $S_{ij}(\boldsymbol{w})$ is the estimator of $\boldsymbol{\Sigma}_{ij}$ under the perturbation in equation (4) for $i,j=1,2$. Here, the ordering of $\hat{\rho}_i^2(\boldsymbol{w})$ is determined by $\hat{\rho}_i^2(\mathbf{1}_n)=\hat{\rho}_i^2$.

To get the first order partial derivatives of $\hat{\rho}_i^2(\boldsymbol{w})$, from equation (6) we should get the partial derivatives of generalized eigenvalues. We considered a generalized eigenvalue problem such as

$$\boldsymbol{A}\,\boldsymbol{\gamma}_i=\lambda_i\,\boldsymbol{B}\,\boldsymbol{\gamma}_i,\qquad i=1,\dots,k$$

where $\boldsymbol{A}$ is a $k\times k$ symmetric matrix, $\boldsymbol{B}$ is a $k\times k$ positive-definite symmetric matrix and $\boldsymbol{\gamma}_i$ is the eigenvector associated with the ith eigenvalue λ_i, normalized to satisfy $\boldsymbol{\gamma}_i^T\,\boldsymbol{B}\,\boldsymbol{\gamma}_i=\delta_{ij}$. Assume that the eigenvalue are all distinct with $\lambda_1>\cdots>\lambda_k$. Let $\boldsymbol{A}(\boldsymbol{w})$ and $\boldsymbol{B}(\boldsymbol{w})$ be the perturbed matrices by the perturbation vector $\boldsymbol{w}$. Also $\boldsymbol{A}(\boldsymbol{w}_0)=\boldsymbol{A}$ and $\boldsymbol{B}(\boldsymbol{w}_0)=\boldsymbol{B}$. Then a generalized eigenvalue problem for the perturbed matrices $\boldsymbol{A}(\boldsymbol{w})$ and $\boldsymbol{B}(\boldsymbol{w})$ can be written as

$$\boldsymbol{A}(\boldsymbol{w})\,\boldsymbol{\gamma}_i(\boldsymbol{w})=\lambda_i(\boldsymbol{w})\,\boldsymbol{B}(\boldsymbol{w})\,\boldsymbol{\gamma}_i(\boldsymbol{w}),\qquad i=1,\dots,k \tag{7}$$

and the normalized constraint is inherited as $\boldsymbol{\gamma}_i^T(\boldsymbol{w})\,\boldsymbol{B}(\boldsymbol{w})\,\boldsymbol{\gamma}_i(\boldsymbol{w})=\delta_{ij}$. Assume that $\boldsymbol{A}(\boldsymbol{w})$, $\boldsymbol{B}(\boldsymbol{w})$, $\lambda_i(\boldsymbol{w})$ and $\boldsymbol{\gamma}_i(\boldsymbol{w})$ are differentiable at $\boldsymbol{w}=\boldsymbol{w}_0$. Differentiating equation (7) with respect to w_u evaluated at the null point yields

$$\boldsymbol{A}_u\,\boldsymbol{\gamma}_i+\boldsymbol{A}\,\boldsymbol{\gamma}_{i,u}=\lambda_{i,u}\boldsymbol{B}\,\boldsymbol{\gamma}_i+\lambda_i\,\boldsymbol{B}_u\,\boldsymbol{\gamma}_i+\lambda_i\boldsymbol{B}\,\boldsymbol{\gamma}_{i,u}\ ,$$

where the subscript u denotes the partial differentiation with respect to w_u evaluated at the null point. Then by left-multiplying $\boldsymbol{\gamma}_i^T$ to above equation gives

$$\lambda_{i,u}=\boldsymbol{\gamma}_i^T(\boldsymbol{A}_u-\lambda_i\,\boldsymbol{B}_u)\,\boldsymbol{\gamma}_i\ .$$

See Jung (2000) for details. From (6) we have the partial derivatives of $\hat{\rho}_i^2(\boldsymbol{w})$ such as

$$\hat{\rho}_{i,u}^2\equiv\left.\frac{\partial\,\hat{\rho}_i^2(\boldsymbol{w})}{\partial w_u}\right|_{\boldsymbol{w}=\boldsymbol{w}_0}=\hat{\boldsymbol{a}}_i^T(\boldsymbol{Q}_{1,u}-\hat{\rho}_i^2\,\boldsymbol{S}_{11,u})\,\hat{\boldsymbol{a}}_i\ , \tag{8}$$

since the generalized eigenvector of (6) is $\hat{\boldsymbol{a}}_i$.

Let $\boldsymbol{K}=\boldsymbol{S}_{11}^{-1/2}\,\boldsymbol{S}_{12}\,\boldsymbol{S}_{22}^{-1/2}$. The singular value decomposition gives $\boldsymbol{K}=(\boldsymbol{a}_1,\dots,\boldsymbol{a}_r)\,\boldsymbol{D}(\boldsymbol{b}_1,\dots\boldsymbol{b}_r)^T$, where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are the standardized eigenvectors of $\boldsymbol{K}\,\boldsymbol{K}^T$ and $\boldsymbol{K}^T\boldsymbol{K}$, respectively, and $\boldsymbol{D}=\operatorname{diag}(d_1^2,\dots,d_r^2)$ for d_i^2 is the eigenvalues of $\boldsymbol{K}\,\boldsymbol{K}^T$ and $\boldsymbol{K}^T\boldsymbol{K}$. Thus we have $\boldsymbol{a}_i=\boldsymbol{S}_{11}^{1/2}\,\hat{\boldsymbol{a}}_i$, $\boldsymbol{b}_i=\boldsymbol{S}_{22}^{1/2}\,\hat{\boldsymbol{\beta}}_i$ and $d_i^2=\hat{\rho}_i^2$.

Some algebra yields

$$S_{22}^{-1} S_{21} \hat{\alpha}_i = \hat{\rho}_i \hat{\beta}_i \quad \text{for} \quad i=1,\ldots,r. \tag{9}$$

Differentiating (5) with respect to w_u gives $S_u=[(z_u-\bar{z})(z_u-\bar{z})^T]/n$. Thus we obtain $S_{11,u}=(x_u-\bar{x})(x_u-\bar{x})^T/n$ and other $S_{ij,u}$ are similarly defined according to i,j for $i,j=1,2$. From the identity $S_{22}^{-1}(w)\, S_{22}(w)=I$ we have $S_{22,u}^{-1} = - S_{22}^{-1} S_{22,u} S_{22}^{-1}$. Thus we obtain

$$Q_u= S_{12,u} S_{22}^{-1} S_{21}- S_{12} S_{22}^{-1} S_{22,u} S_{22}^{-1} S_{21}+ S_{12} S_{22}^{-1} S_{21,u}.$$

By putting Q_u and $S_{11,u}$ into (8) with (9) we obtain

$$\hat{\rho}_{i,u}^{\,2}=\frac{1}{n}[2\hat{\rho}_i e_{iu} f_{iu}- \hat{\rho}_i^2 e_{iu}^2- \hat{\rho}_i^2 f_{iu}^2],$$

where $e_{iu}= \hat{\alpha}_i^T(x_u-\bar{x})$, $f_{iu}= \hat{\beta}_i^T(y_u-\bar{y})$. Here e_{iu} and f_{iu} denote the ith pair scores of x_u and y_u, respectively.

From (4), we obtain the partial derivatives $\partial T(w)/\partial w_u$ evaluated at $w= 1_n$ as

$$\begin{aligned}\left.\frac{\partial T(w)}{\partial w_u}\right|_{w=1_n} &= c\sum_{i=1}^{r}\frac{1}{1-\hat{\rho}_i}[2\hat{\rho}_i e_{iu} f_{iu}- \hat{\rho}_i^2 e_{iu}^2- \hat{\rho}_i^2 f_{iu}^2] \\ &= c\sum_{i=1}^{r}\frac{\hat{\rho}_i}{1-\hat{\rho}_i}[2(x_u-\bar{x})^T \hat{\alpha}_i \hat{\beta}_i^T(y_u-\bar{y}) \\ &\quad - \hat{\rho}_i(x_u-\bar{x})^T \hat{\alpha}_i \hat{\alpha}_i^T(x_u-\bar{x}) \\ &\quad - \hat{\rho}_i(y_u-\bar{y})^T \hat{\beta}_i \hat{\beta}_i^T(y_u-\bar{y})],\end{aligned}$$

where $c=-\{n-(p+q+3)/2\}$. Let $P = \sum_{i=1}^{r}\frac{\hat{\rho}_i}{1-\hat{\rho}_i}\begin{bmatrix} -\hat{\rho}_i\, \alpha_i\, \alpha_i^T & \alpha_i\, \beta_i^T \\ \beta_i\, \alpha_i^T & -\hat{\rho}_i\, \beta_i\, \beta_i^T\end{bmatrix}$. Then we have

$$\dot{T}= \left.\frac{\partial T(w)}{\partial w}\right|_{w=1_n} = \text{diag}(\tilde{Z} P \tilde{Z}^T), \tag{10}$$

where $\tilde{Z}= Z- 1_n \bar{z}^T$, because the constant c can be cancelled out by normalization.

For testing the hypothesis that only m of the population canonical correlation coefficients are non-zero, we may investigate the influence of observations on the test statistic based on the local influence measure given by the $n\times 1$ vector

$$\text{diag}(\tilde{Z} P_m \tilde{Z}^T),$$

where $P_m = \sum_{i=m+1}^{r}\frac{\hat{\rho}_i}{1-\hat{\rho}_i}\begin{bmatrix} -\hat{\rho}_i\, \alpha_i\, \alpha_i^T & \alpha_i\, \beta_i^T \\ \beta_i\, \alpha_i^T & -\hat{\rho}_i\, \beta_i\, \beta_i^T\end{bmatrix}$.

5. Numerical Example

The local influence method is applied to the head-length data (Mardia, et al., 1979, p. 121, Table 5.1.1) previously analyzed by Romanazzi (1992) and Jung (2000). For this data set, $n=25$, $p=2$, $q=2$. The two sample canonical correlation coefficients are $\hat{\rho}_1^{\,2}=0.6217$, $\hat{\rho}_2^{\,2}=0.0029$. Our analysis will be confined to the LRT statistic for the hypothesis (1). The LRT statistic based on the full data set is $T=20.96$, and therefore we conclude that the null hypothesis is strongly rejected by comparing $\chi^2_4(0.05)=9.49$, where $\chi^2_k(\alpha)$ is the upper αth percentile of the χ^2 distribution with k degrees of freedom.

We obtain information about influential observations for the LRT statistic T using the local influence method with the index plot of $l_{\max}$ by (10). The index plot is presented in Fig. 1. From this plot we may conclude that observations 16, 20 and 24 are candidate for influential observations. And also we can observe that observation 16 is most influential.

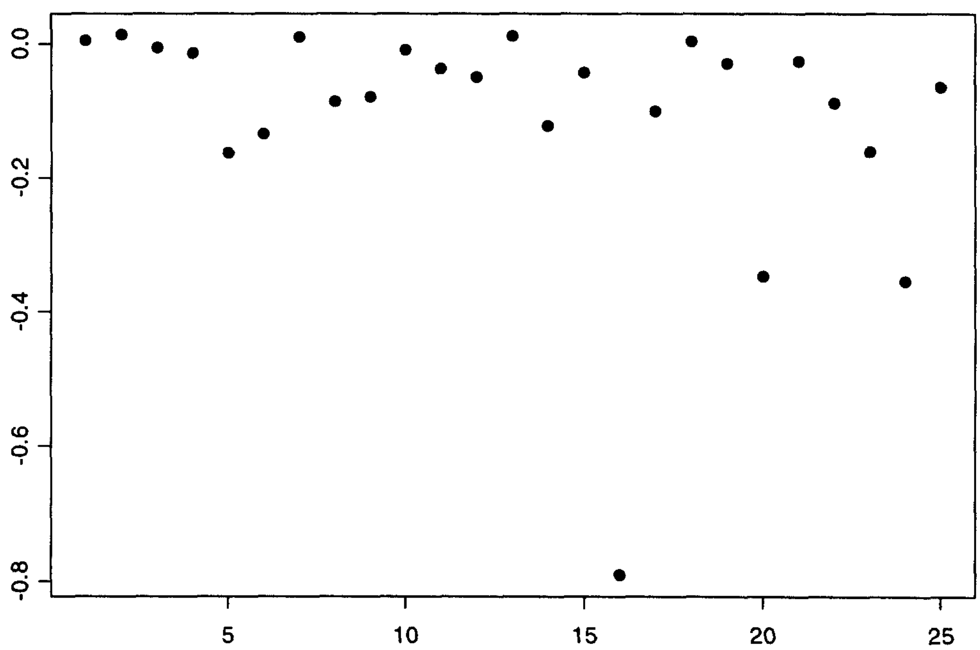

Fig. 1 : The index plot for the LRT statistic T using the local influence method

For a confirmatory analysis, we conduct the single and double case deletions, and the results are summarized in Table 1. Numbers in Table 1 are arranged in decreasing order of $|T-T_{(J)}|$, where $T_{(J)}$ denotes the test statistic after deletion of the corresponding index set J. The case deletion results show that observations 16, 20 and 24 are individually influential and observations 20 and 24 with observation 16 are jointly influential, which are reflected in Fig. 1. Furthermore, $T_{(16,20,24)}=8.71$ gives that the null hypothesis is not rejected. It implies that opposite conclusions are made by removing observations 16, 20, 24 or not. We conclude that observations 16, 20 and 24 are influential observations on the LRT statistic, and observation 16 is most influential.

This example shows that the local influence method for detecting influential observations in testing the uncorrelatedness between groups of variables, based on the canonical correlation coefficients, is very informative, in that it gives much information about individually and jointly influential observations. Confirmatory analysis using the case deletion method could be a giant time-consuming job. Therefore, the local influence method is efficient in detecting influential observations.

Table 1 : Single and double case deletions results for the LRT statistic T

J	$T_{(J)}$	$T-T_{(J)}$	J	$T_{(J)}$	$T-T_{(J)}$
16	14.99	5.98	16, 20	11.61	9.35
24	18.04	2.93	16, 24	11.97	8.99
20	18.27	2.69	5, 16	12.81	8.16

References

[1] Cook, R. D. (1986). Assessment of local influence (with discussion), *Journal of the Royal Statistical Society, Series B,* Vol. 48, pp. 133-169.

[2] Cook, R. D. and Weisberg, S. (1982). *Residuals and influence in regression*, Chapman and Hall, New York.

[3] Jung, K.-M. (2000). Local influence assessment in canonical correlation analysis, *Journal of Applied Statistics*, Vol. 27, pp. 293-301.

[4] Jung, K.-M. and Kim, M. G. (1999). Influence analysis of the likelihood ratio test in

multivariate Behrens-Fisher problem, *The Korean Communications in Statistics*, Vol. 6, pp. 939-946.

[5] Lawrance , A. J. (1988). Regression transformation diagnostics using local influence, *Journal of the American Statistical Association*, Vol. 83, 1067-1072.

[6] Mardia, K. V., Kent. J. T. and Bibby, J. M. (1979). *Multivariate analysis*, Academic Press, New York.

[7] Romanazzi, M. (1992). Influence in canonical correlation analysis, *Psychometrika*, Vol. 57, pp. 237-259.