Communications for Statistical Applications and Methods
2013, Vol. 20, No. 3, 217–223

DOI: http://dx.doi.org/10.5351/CSAM.2013.20.3.217

Influence Analysis of the Common Mean Problem

Myung Geun Kim[1,a]

[a]Department of Mathematics Education, Seowon University

Abstract

Two influence diagnostic methods for the common mean model are proposed. First, an investigation of the influence of observations according to minor perturbations of the common mean model is made by adapting the local influence method which is based on the likelihood displacement. It is well known that the maximum likelihood estimates are in general sensitive to influential observations. Case-deletions can be a candidate for detecting influential observations. However, the maximum likelihood estimators are iteratively computed and therefore case-deletions involve an enormous amount of computations. An approximation by Newton's method to the maximum likelihood estimator obtained after a single observation was deleted can reduce much of computational burden, which will be treated in this work. A numerical example is given for illustration and it shows that the proposed diagnostic methods can be useful tools.

Keywords: Case deletions, common mean, local influence, Newton method.

1. Introduction

The common mean problem has a long history of research since the pioneering work by Graybill and Deal (1959). It has real-life applications in wide areas, for example in clinical trials (Kelleher, 1996), in soil engineering (Zacks, 1966, 1970), in a balanced incomplete block design (Montgomery, 1997; Pal *et al.*, 2007), and in meta analysis (Chang and Pal, 2008). Some detailed references about point estimation of the common mean can be found in Kubokawa (1991) and Ma *et al.* (2011). Recently Pal *et al.* (2007) studied the maximum likelihood estimation of the parameters of the common mean model. For estimating the common mean with unknown variances, they showed that the maximum likelihood estimator has better overall performance than the Graybill-Deal estimator.

In this work, an investigation of the influence of observations according to minor perturbations of the common mean model is considered. In Section 2, it is performed by adapting the local influence method introduced by Cook (1986) which is based on the likelihood displacement. It is well known that the maximum likelihood estimates are in general sensitive to influential observations. Thus we need some diagnostic methods of detecting observations that have large influence on the estimates. To this end case-deletions can be a candidate. As will be seen later, the maximum likelihood estimators are iteratively computed and therefore case-deletions involve an enormous amount of computations. An approximation by Newton's method to the maximum likelihood estimator obtained after a single observation was deleted can reduce much of computational burden, which will be treated in Section 3. A numerical example is given for illustration in Section 4.

[1] Professor, Department of Mathematics Education, Seowon University, Cheongju 361-742, Korea.
E-mail: mgkim@seowon.ac.kr

2. Local Influence

2.1. Maximum likelihood estimation

We consider the common mean model characterized by two normal distributions with common mean, $N(\mu, \sigma_1^2)$ and $N(\mu, \sigma_2^2)$. A random sample $\{x_{i1}, \ldots, x_{in_i}\}$ is drawn from $N(\mu, \sigma_i^2)$ for each $i = 1, 2$ and both samples are assumed to be independent. The joint log-likelihood function of μ, σ_1^2, σ_2^2 based on both samples can be written as

$$L\left(\mu, \sigma_1^2, \sigma_2^2\right) = -\sum_{i=1}^{2} \frac{n_i}{2} \left[\log(2\pi) + \log\left(\sigma_i^2\right) + \frac{s_i^2}{\sigma_i^2} + \frac{1}{\sigma_i^2} (\bar{x}_i - \mu)^2\right],$$

where

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad \text{and} \quad s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2.$$

Let $\hat{\mu}$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ be the maximum likelihood estimators(MLEs) of μ, σ_1^2, σ_2^2, respectively and then they are found by iteratively solving the following likelihood equations

$$\hat{\mu} = \sum_{i=1}^{2} \frac{n_i}{\hat{\sigma}_i^2} \bar{x}_i \Big/ \sum_{i=1}^{2} \frac{n_i}{\hat{\sigma}_i^2},$$

$$\hat{\sigma}_i^2 = s_i^2 + (\bar{x}_i - \hat{\mu})^2, \quad (i = 1, 2).$$

Since $\bar{x}_i - \hat{\mu} = n_2\hat{\sigma}_1^2(\bar{x}_1 - \bar{x}_2)/(n_2\hat{\sigma}_1^2 + n_1\hat{\sigma}_2^2)$, an alternative form for $\hat{\sigma}_1^2$ becomes

$$\hat{\sigma}_1^2 = s_1^2 + \left[\frac{n_2\hat{\sigma}_1^2}{n_2\hat{\sigma}_1^2 + n_1\hat{\sigma}_2^2}\right]^2 (\bar{x}_1 - \bar{x}_2)^2.$$

Exchanging the role of the subscripts, we can have an equivalent form for $\hat{\sigma}_2^2$. More details about the maximum likelihood estimation can be found in Pal *et al.* (2007).

For later use, let $\ddot{L} = \partial^2 L(\mu, \sigma_1^2, \sigma_2^2)/\partial\theta\partial\theta^T$ evaluated at $\theta = \hat{\theta}$, where $\theta = (\mu, \sigma_1^2, \sigma_2^2)^T$. Then $\ddot{L}$ is a 3×3 matrix whose components are computed as

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2\right)}{\partial\mu^2} = -\frac{n_1}{\hat{\sigma}_1^2} - \frac{n_2}{\hat{\sigma}_2^2},$$

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2\right)}{\partial\mu\partial\left(\sigma_1^2\right)} = -\frac{n_1}{\hat{\sigma}_1^4} (\bar{x}_1 - \hat{\mu}),$$

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2\right)}{\partial\mu\partial\left(\sigma_2^2\right)} = -\frac{n_2}{\hat{\sigma}_2^4} (\bar{x}_2 - \hat{\mu}),$$

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2\right)}{\partial\left(\sigma_i^2\right)^2} = -\frac{n_i}{2}\left[-\frac{1}{\hat{\sigma}_i^4} + \frac{2s_i^2}{\hat{\sigma}_i^6} + \frac{2(\bar{x}_i - \hat{\mu})^2}{\hat{\sigma}_i^6}\right] = -\frac{n_i}{2\hat{\sigma}_i^4}, \quad (i = 1, 2),$$

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2\right)}{\partial\left(\sigma_1^2\right)\partial\left(\sigma_2^2\right)} = 0,$$

where all of the above partial derivatives are evaluated at $\theta = \hat{\theta}$.

2.2. Perturbed model

Let $n = n_1 + n_2$. A perturbation vector $w = (w_1, \ldots, w_n)^T$ in the n-dimensional Euclidean space characterizes the perturbed model, and it can be expressed as $w = 1_n + \alpha d$, where 1_n is an $n \times 1$ vector whose elements are all equal to one, the vector d of unit length represents the direction towards which the perturbation w moves, and the scalar α denotes the magnitude of the perturbation w along the direction d. We consider a perturbed model in which

$$\begin{cases} x_{1j} \sim N\left(\mu, \dfrac{\sigma_1^2}{w_j}\right), & (j = 1, \ldots, n_1), \\ x_{2k} \sim N\left(\mu, \dfrac{\sigma_2^2}{w_{n_1+k}}\right), & (k = 1, \ldots, n_2). \end{cases}$$

Under this perturbation scheme, the perturbed joint log-likelihood function of μ, σ_1^2, σ_2^2 based on both samples becomes

$$L\left(\mu, \sigma_1^2, \sigma_2^2 \,|\, w\right) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{n}\log(w_i) - \frac{n_1}{2}\log\left(\sigma_1^2\right) - \frac{1}{2}\sum_{j=1}^{n_1}\frac{w_j\left(x_{1j}-\mu\right)^2}{\sigma_1^2} - \frac{n_2}{2}\log(\sigma_2^2) - \frac{1}{2}\sum_{k=1}^{n_2}\frac{w_{n_1+k}(x_{2k}-\mu)^2}{\sigma_2^2}.$$

Let $\hat{\mu}(w)$, $\hat{\sigma}_1^2(w)$, $\hat{\sigma}_2^2(w)$ be the MLEs of μ, σ_1^2, σ_2^2 under the above perturbed model, respectively, and then it is easily seen that they are also iteratively computed by the following equations

$$\begin{aligned} \hat{\mu}(w) &= \frac{\sum_{j=1}^{n_1} w_j x_{1j}/\hat{\sigma}_1^2(w) + \sum_{k=1}^{n_2} w_{n_1+k} x_{2k}/\hat{\sigma}_2^2(w)}{\sum_{j=1}^{n_1} w_j/\hat{\sigma}_1^2(w) + \sum_{k=1}^{n_2} w_{n_1+k}/\hat{\sigma}_2^2(w)}, \\ \hat{\sigma}_1^2(w) &= \frac{1}{n_1}\sum_{j=1}^{n_1} w_j\left(x_{1j} - \hat{\mu}(w)\right)^2, \\ \hat{\sigma}_2^2(w) &= \frac{1}{n_2}\sum_{k=1}^{n_2} w_{n_1+k}\left(x_{2k} - \hat{\mu}(w)\right)^2. \end{aligned}$$

Since $\hat{\sigma}_i^2 = s_i^2 + (\bar{x}_i - \hat{\mu})^2 = (1/n_i)\sum_{j=1}^{n_i}(x_{ij} - \hat{\mu})^2$ in the unperturbed model, the form of the $\hat{\sigma}_i^2(w)$ in the perturbed model can easily be guessed just as given in the above.

Let $\Delta = \partial^2 L(\mu, \sigma_1^2, \sigma_2^2 \,|\, w)/\partial\theta\partial w^T$ evaluated at $\theta = \hat{\theta}$ and $w = 1_n$. A closed form of Δ is not available but its components are easily computed as

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2 \,|\, w\right)}{\partial\mu\partial w_j} = \begin{cases} \dfrac{x_{1j} - \hat{\mu}}{\hat{\sigma}_1^2}, & (1 \le j \le n_1), \\ \dfrac{x_{2k} - \hat{\mu}}{\hat{\sigma}_2^2}, & (j = n_1 + k,\ 1 \le k \le n_2), \end{cases}$$

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2 \,|\, w\right)}{\partial\left(\sigma_1^2\right)\partial w_j} = \begin{cases} \dfrac{1}{2}\dfrac{(x_{1j} - \hat{\mu})^2}{\hat{\sigma}_1^4}, & (1 \le j \le n_1), \\ 0, & (n_1 + 1 \le j \le n), \end{cases}$$

$$\frac{\partial^2 L\left(\mu, \sigma_1^2, \sigma_2^2 \,|\, w\right)}{\partial\left(\sigma_2^2\right)\partial w_j} = \begin{cases} 0, & (1 \leq j \leq n_1), \\ \dfrac{1}{2}\dfrac{(x_{2k} - \hat{\mu})^2}{\hat{\sigma}_2^4}, & (j = n_1 + k,\ 1 \leq k \leq n_2), \end{cases}$$

where all of the above derivatives are evaluated at $\theta = \hat{\theta}$ and $w = 1_n$.

2.3. Local influence procedure

In this subsection, we will briefly describe the local influence introduced by Cook (1986) which can be adapted to the common mean model with computational results obtained in the previous subsections. Let $\ddot{F} = \Delta^T \ddot{L}^{-}1\Delta$. Then we can compute $\ddot{F}$, numerically. In the case of the common mean model, the likelihood displacement can be defined by

$$LD(w) = 2\left[L\left(\hat{\mu}, \hat{\sigma}_1^2, \hat{\sigma}_2^2\right) - L\left(\hat{\mu}(w), \hat{\sigma}_1^2(w), \hat{\sigma}_2^2(w)\right)\right].$$

The surface of interest is formed by the vector $(w, LD(w))$ as w varies over a certain space. The eigenvector l_{max} corresponding to the largest absolute eigenvalue of $2\ddot{F}$ provides information about influential observations. Observations associated with relatively large absolute component of l_{max} would be influential. The largest absolute eigenvalue of $2\ddot{F}$ is the maximum curvature of the curve which is the portion of the surface cut out by the plane spanned by the vectors $1_{(n+1)}$ and $(l_{max}, 0)$, where $1_{(n+1)}$ is the $(n+1) \times 1$ vector with its $(n+1)$st element equal to one and the others being zero. Refer to O'Neill (2006) for more details about differential geometry.

3. Case Deletion Diagnostics

In what follows, a quantity with a subscript $(-r)$ implies that the quantity has been obtained after the r^{th} observation was deleted. Let $L_{(-r)} = L_{(-r)}(\mu, \sigma_1^2, \sigma_2^2)$ be the joint log-likelihood based on the reduced sample with the r^{th} observation of the first sample deleted. Similarly, the case of deleting the r^{th} observation of the second sample can be treated, which is therefore not reported here. Let $\dot{L}_{(-r)} = \partial L_{(-r)}/\partial\theta$ evaluated at $\theta = \hat{\theta}$ and $\ddot{L}_{(-r)} = \partial^2 L_{(-r)}/\partial\theta\partial\theta^T$ evaluated at $\theta = \hat{\theta}$.

The influence of deleting the r^{th} observation on the MLE $\hat{\theta}$ can be investigated by the quantity $\hat{\theta} - \hat{\theta}_{(-r)}$. If $\hat{\theta}_{(-r)}$ resides far away from $\hat{\theta}$, then the r^{th} observation is regarded as influential one. As in $\hat{\theta}$, an explicit form of $\hat{\theta}_{(-r)}$ is not available and for each r we should take steps of iteratively estimating the parameters, which incurrs a heavy computational burden. To avoid this computational difficulty, we can use an approximation to $\hat{\theta}_{(-r)}$ which can be obtained by the Newton's method as

$$\hat{\theta}_{(-r)} = \hat{\theta} - \ddot{L}_{(-r)}^{-1}\dot{L}_{(-r)}.$$

For more details about the Newton's method, refer to Chap. 5 of Cook and Weisberg (1982) or Chap. 10 of Kennedy and Gentle (1980). An approximation to $\hat{\theta}_{(-r)}$ given in the above depends on $\hat{\theta}$ for all r. Hence there is no need to do an estimation process again and this approximation reduces computational burden. An application of Newton method to case deletion diagnostics in other fields of statistics, for example in nonlinear structural equation model can be found in Lee and Lu (2003).

Based on the reduced sample with the r^{th} observation deleted from the i^{th} sample, the sample mean and the sample variance are

$$\bar{x}_{i(-r)} = \frac{1}{n_i - 1}\sum_{j \neq r} x_{ij} = \frac{n_i}{n_i - 1}\bar{x}_i - \frac{1}{n_i - 1}x_{ir},$$

$$s_{i(-r)}^2 = \frac{1}{n_i - 1}\sum_{j \neq r}(x_{ij} - \bar{x}_{i(-r)})^2 = \frac{n_i}{n_i - 1}s_i^2 - \frac{n_i}{(n_i - 1)^2}(x_{ir} - \bar{x}_i)^2.$$

The joint log-likelihood function of μ, σ_1^2 , σ_2^2 based on the reduced sample is

$$\begin{aligned} L_{(-r)} = &-\frac{n_1-1}{2}\log(2\pi) - \frac{n_1-1}{2}\log\left(\sigma_1^2\right) - \frac{n_1-1}{2}\frac{s_{1(-r)}^2}{\sigma_1^2} - \frac{n_1-1}{2}\frac{(\bar{x}_{1(-r)}-\mu)^2}{\sigma_1^2} \\ &- \frac{n_2}{2}\log(2\pi) - \frac{n_2}{2}\log\left(\sigma_2^2\right) - \frac{n_2}{2}\frac{s_2^2}{\sigma_2^2} - \frac{n_2}{2}\frac{(\bar{x}_2-\mu)^2}{\sigma_2^2}. \end{aligned}$$

For computing $\dot{L}_{(-r)}$, we need the following computational results:

$$\begin{aligned} \frac{\partial L_{(-r)}}{\partial \mu} &= \frac{n_1-1}{\hat{\sigma}_1^2}\left(\bar{x}_{1(-r)}-\hat{\mu}\right) + \frac{n_2}{\hat{\sigma}_2^2}\left(\bar{x}_2-\hat{\mu}\right), \\ \frac{\partial L_{(-r)}}{\partial\left(\sigma_1^2\right)} &= -\frac{n_1-1}{2}\left[\frac{1}{\hat{\sigma}_1^2} - \frac{s_{1(-r)}^2}{\hat{\sigma}_1^4} - \frac{(\bar{x}_{1(-r)}-\hat{\mu})^2}{\hat{\sigma}_1^4}\right], \\ \frac{\partial L_{(-r)}}{\partial\left(\sigma_2^2\right)} &= -\frac{n_2}{2}\left[\frac{1}{\hat{\sigma}_2^2} - \frac{s_2^2}{\hat{\sigma}_2^4} - \frac{(\bar{x}_2-\hat{\mu})^2}{\hat{\sigma}_2^4}\right] = 0. \end{aligned}$$

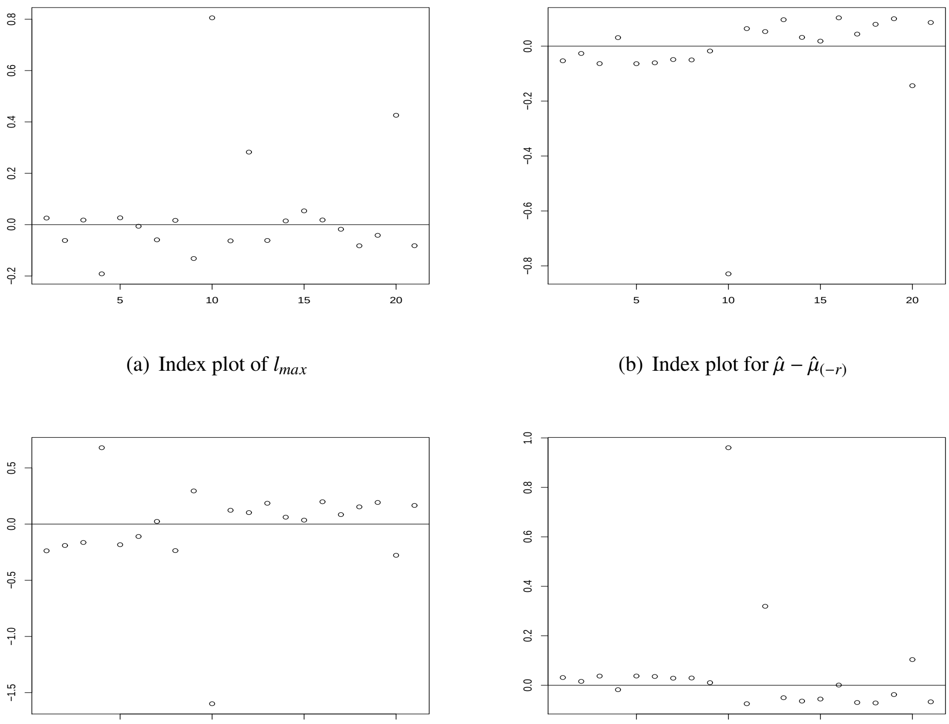
Also, the elements of $\ddot{L}_{(-r)}$ are computed as

$$\begin{aligned} \frac{\partial^2 L_{(-r)}}{\partial \mu^2} &= -\frac{n_1-1}{\hat{\sigma}_1^2} - \frac{n_2}{\hat{\sigma}_2^2}, \\ \frac{\partial^2 L_{(-r)}}{\partial\mu\partial\left(\sigma_1^2\right)} &= -\frac{n_1-1}{\hat{\sigma}_1^4}\left(\bar{x}_{1(-r)}-\hat{\mu}\right), \\ \frac{\partial^2 L_{(-r)}}{\partial\mu\partial\left(\sigma_2^2\right)} &= -\frac{n_2}{\hat{\sigma}_2^4}(\bar{x}_2-\hat{\mu}), \\ \frac{\partial^2 L_{(-r)}}{\partial\left(\sigma_1^2\right)^2} &= -\frac{n_1-1}{2}\left[-\frac{1}{\hat{\sigma}_1^4} + \frac{2s_{1(-r)}^2}{\hat{\sigma}_1^6} + \frac{2\left(\bar{x}_{1(-r)}-\hat{\mu}\right)^2}{\hat{\sigma}_1^6}\right], \\ \frac{\partial^2 L_{(-r)}}{\partial\left(\sigma_2^2\right)^2} &= -\frac{n_2}{2}\left[-\frac{1}{\hat{\sigma}_2^4} + \frac{2s_2^2}{\hat{\sigma}_2^6} + \frac{2\left(\bar{x}_2-\hat{\mu}\right)^2}{\hat{\sigma}_2^6}\right] = -\frac{n_2}{2\hat{\sigma}_2^4}, \\ \frac{\partial^2 L_{(-r)}}{\partial\left(\sigma_1^2\right)\partial\left(\sigma_2^2\right)} &= 0. \end{aligned}$$

Note that all of the above partial derivatives are evaluated at $\theta = \hat{\theta}$.

4. A Numerical Example

In this section we will illustrate the local influence method and an approximation to case-deletions with a data set taken from Problem 14.10 of Neter *et al.* (1990) in which data on productivity improvements for three levels of firms were collected. Here we will investigate the first two levels: Low level for the first group and Moderate level for the second group. The observations for the first group are labeled as 1 to 9 in order and those for the second group as 10 to 21.



(a) Index plot of l_{max}

(b) Index plot for $\hat{\mu} - \hat{\mu}_{(-r)}$

(c) Index plot for $\hat{\sigma}_1^2 - \hat{\sigma}_{1(-r)}^2$

(d) Index plot for $\hat{\sigma}_2^2 - \hat{\sigma}_{2(-r)}^2$

Figure 1: *Influence analysis*

The maximum likelihood estimates based on the full sample are $\hat{\mu} = 7.84$, $\hat{\sigma}_1^2 = 1.52$ and $\hat{\sigma}_2^2 = 0.61$. The largest absolute eigenvalue of $2\ddot{F}$ is 2.46. An index plot of l_{max} is included in Figure 1(a). It shows that observations 10, 20, 12 are potentially influential and observation 4 is a little bit influential.

Single case deletions are performed. Index plots for the values $\hat{\mu} - \hat{\mu}_{(-r)}$, $\hat{\sigma}_1^2 - \hat{\sigma}_{1(-r)}^2$, $\hat{\sigma}_2^2 - \hat{\sigma}_{2(-r)}^2$ are displayed in Figures 1(b) to (d), respectively. We can see that observation 10 has the largest influence on all of the estimates $\hat{\mu}$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$. Observation 4 also influences the estimation of σ_1^2. We can see that the role of observation 10 in estimating σ_1^2 is opposite to that of observation 4 because deletion of observation 10 increases the estimate $\hat{\sigma}_1^2$ while deletion of observation 4 decreases it. The influence of observation 12 is remarkable in estimating σ_2^2 compared with the other parameters. Observations 10 and 12 have a similar role in estimating σ_2^2 in that deleting each of them reduces the estimate $\hat{\sigma}_2^2$. Observation 20 has a little influence on $\hat{\mu}$. This example shows that case deletions yield some useful information about the behavior of observations in estimating the parameters.

References

Chang, C. H. and Pal, N. (2008). Testing on the common mean of several normal distributions, *Computational Statistics and Data Analysis*, **53**, 321–333.

Cook, R. D. (1986). Assessment of local influence (with discussions), *Journal of the Royal Statistical Society (B)*, **48**, 133–169.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators, *Biometrics*, **15**, 543–550.

Kelleher, T. (1996). A Bayes method for estimating the common mean of two normal populations, *Communications in Statistics: Theory and Methods*, **25**, 2141–2157.

Kennedy, W. and Gentle, J. (1980). *Statistical Computing*, Dekker, New York.

Kubokawa, T. (1991). Estimation of the common mean and its applications, *Sugaku Expositions*, **4**, 97–110 (Published by American Mathematical Society).

Lee, S. Y. and Lu, B. (2003). Case-Deletion diagnostics for nonlinear structural equation models, *Multivariate Behavioral Research*, **38**, 375–400.

Ma, T., Ye, R. and Jia, L. (2011). Finite-sample properties of the Graybill-Deal estimator, *Journal of Statistical Planning and Inference*, **141**, 3675–3680.

Montgomery, D. C. (1997). *Design and Analysis of Experiments*, 4th ed., John Wiley & Sons, New York.

Neter, J., Wasserman, W. and Kutner, M. (1990). *Applied Linear Statistical Models*, 3rd ed., Irwin.

O'Neill, Barrett (2006). *Elementary Differential Geometry*, 2nd ed., Academic Press.

Pal, N., Lin, J.-J., Chang, C.-H. and Kumar, S. (2007). A revisit to the common mean problem: Comparing the maximum likelihood estimator with the Graybill-deal estimator, *Computational Statistics and Data Analysis*, **51**, 5673–5681.

Zacks, S. (1966). Unbiased estimation of the common mean of two normal distributions based on small samples of equal size, *Journal of the American Statistical Association*, **61**, 467–476.

Zacks, S. (1970). Bayes and fiducial equivariant estimators of the common mean of two normal distributions, *Annals of Mathematical Statistics*, **41**, 59–69.

Received April 2, 2013; Revised April 25, 2013; Accepted April 25, 2013