

Erkki P. Liiski*

INFLUENTIAL OBSERVATIONS IN THE GENERALIZED ANALYSIS
OF VARIANCE MODEL

1. A GENERALIZED MULTIVARIATE ANALYSIS
OF VARIANCE MODEL (GMANOVA)

First we describe in this section a generalization of the standard MANOVA (multivariate analysis of variance) model suggested in P o t t h o f f and R o y [10]. This model is also known as the growth curves model, although it is a very general model applying to a variety of multivariate situations. Pott-hoff and Roy's modification to the MANOVA model is the addition of a within-subject design matrix T describing the structure of an individual curve. The model becomes

$$E(Y) = X B T' \quad (1.1)$$

$n \times q \quad n \times m \quad m \times p \quad p \times q$

where each row of Y has a multivariate normal distribution with mean vector μ_i and variance-covariance matrix Σ , and X and T are known $n \times m$ and $q \times p$ matrices of rank m and p , $m < n$ and $p \leq q$. B is an unknown $m \times p$ parameter matrix. Denote as $Y_{(i)}$ the rows of Y , so that $Y' = (Y_{(1)}, Y_{(2)}, \dots, Y_{(n)})$.

The matrix X in the model (1.1) is the usual design matrix, consisting of indicator variables specifying treatment group and possible covariates. For example, in the m -group case with n_i subjects per group (1-way analysis of variance)

* Lecturer at the University of Tampere and the Academy of Finland, Tampere Finland.

$$X'_{m \times n} = \begin{bmatrix} 1'_{n_1} & 0' & \dots & 0' \\ 0' & 1'_{n_2} & \dots & 0' \\ \vdots & \vdots & \ddots & \vdots \\ 0' & 0' & \dots & 1'_{n_m} \end{bmatrix}$$

where $1'_{n_i}$ denotes an $n_i \times 1$ vector of unities in the design matrix. Very often in growth curve applications T is a matrix whose rows are powers of the time t_i : $t_{ij} = t_i^{j-1}$

$$T_{q \times p} = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{p-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_q & t_q^2 & \dots & t_q^{p-1} \end{bmatrix}$$

For the j th subject we could have the univariate linear model

$$y_{(j)} = T \beta + \varepsilon_{(j)}; \quad j = 1, 2, \dots, n, \quad (1.2)$$

where $\text{var}(\varepsilon_{(j)}) = \Sigma$ for every j and $\beta' = (\beta_0, \beta_1, \dots, \beta_{p-1})$. The growth curve associated with the j th individual is of the form

$$E(y) = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_{p-1} t^{p-1},$$

and the generalized least squares estimator of β is

$$\beta = (T' \Sigma^{-1} T)^{-1} T' \Sigma^{-1} y_{(j)}.$$

We suppose that $\beta = (\beta_{10}, \beta_{11}, \dots, \beta_{1,p-1})'$ when the subject falls in the i th group. In this 1-way MANOVA situation the parameter matrix

$$B_{m \times p} = \begin{bmatrix} \beta_{10} & \beta_{11} & \dots & \beta_{1,p-1} \\ \beta_{20} & \beta_{21} & \dots & \beta_{2,p-1} \\ \vdots & \vdots & & \vdots \\ \beta_{m0} & \beta_{m1} & & \beta_{m,p-1} \end{bmatrix} = \begin{bmatrix} \beta'(1) \\ \beta'(2) \\ \vdots \\ \beta'(m) \end{bmatrix}$$

We rewrite the growth curves model (1.1) by putting it into the vector form. Let us define the matrix operation vec , which rearranges the columns of a matrix underneath each other. Thus, for example, $\text{vec}Y' = y'(1), y'(2), \dots, y'(n)'$ is an $nq \times 1$ vector and $\text{vec}B' = (\beta'(1), \beta'(2), \dots, \beta'(m))'$ is a $pm \times 1$ vector, where $y_{(i)}$ is the i th row of Y and $\beta_{(j)}$ is the j th row of B . Since we want to put the rows of Y underneath each other, we write first $E(Y') = TB'X'$ and note that $\text{vec}(TB'X') = (X \otimes T) \text{vec}B'$, where $X \otimes T$ is the Kronecker product of X and T . Therefore we have the generalized linear model defined by the following equations

$$E(\text{vec}Y') = (X \otimes T) \text{vec}B' \quad (1.3)$$

and

$$\text{cov}(\text{vec}Y') = I \otimes \Sigma,$$

where $\Sigma = \text{cov} y_{(i)}$ for every $i = 1, 2, \dots, n$. Let us suppose, first, that Σ were known. Then the BLUE (best linear unbiased estimate) for $\text{vec}B'$ is

$$\begin{aligned} \text{vec}\hat{B}' &= [(X \otimes T)' (I \otimes \Sigma)^{-1} (X \otimes T)]^{-1} (X \otimes T)' (I \otimes \Sigma)^{-1} \text{vec}Y' \\ &= [(X'X)^{-1} X' \otimes (T'\Sigma^{-1}T)^{-1} T'\Sigma^{-1}] \text{vec}Y' \end{aligned} \quad (1.4)$$

which is equivalent to

$$\hat{B} = (X'X)^{-1} X' Y \Sigma^{-1} T (T'\Sigma^{-1}T)^{-1} \quad (1.5)$$

The covariance matrix of the elements of \hat{B} is given as

$$\text{cov}(\text{vec}\hat{B}) = (X'X)^{-1} \otimes (T'\Sigma^{-1}T)^{-1} \quad (1.6)$$

Let us consider the estimation of estimable linear functions of the form CBD, where C and D are known matrices of order $g \times m$ and $p \times v$ respectively. This can be written in vector form as $\text{vec}(D'B'C') = (C\otimes D')\text{vec}B'$. The BLUE for the estimable function $(C\otimes D')\text{vec}B'$ is

$$(C\otimes D')[(X'X)^{-1}X'Y(T'\Sigma^{-1}T)^{-1}T'\Sigma^{-1}] \text{vec}Y' \quad (1.7)$$

which means that

$$\hat{C}BD = C(X'X)^{-1}X'Y\Sigma^{-1}T(T'\Sigma^{-1}T)^{-1}D \quad (1.8)$$

is the BLUE of CBD. The covariance matrix of $(C\otimes D')\text{vec}\hat{B}'$ is directly computed to be

$$\begin{aligned} & (C\otimes D')[(X'X)^{-1} \otimes (T'\Sigma^{-1}T)^{-1}](X'\otimes D) \\ & = C(X'X)^{-1}C'\otimes D'(T'\Sigma^{-1}T)^{-1}D \end{aligned} \quad (1.9)$$

In practice Σ is usually unknown and should be estimated. Let us reparametrize the model (1.1) for a while so that $\Gamma = BT'$. Then we have an ordinary multivariate linear model $E(Y) = X\Gamma$, where Γ' is restricted to the space spanned by the columns of T. The general theory of multivariate linear models shows (see e.g. Arnold [2], p. 350) that

$$\hat{\Sigma} = \frac{1}{n-m} (Y - X\hat{\Gamma})'(Y - X\hat{\Gamma}) \quad (1.10)$$

is an unbiased estimate of Σ and is independent of $\hat{\Gamma} = (X'X)^{-1}X'Y$. In fact $\hat{\Sigma}$ follows the Wishart distribution $W_q(n-m, \frac{1}{n-m}\Sigma)$ with $n-m$ degrees of freedom. If we replace Σ by $\hat{\Sigma}$ in (1.5) and (1.8), we obtain an empirical estimate for B:

$$\tilde{B} = (X'X)^{-1}X'Y\hat{\Sigma}T(T'\hat{\Sigma}T)^{-1} \quad (1.11a)$$

and for CBD:

$$\tilde{C}\tilde{B}D = C(X'X)^{-1}X'Y\hat{\Sigma}^{-1}T(T'\hat{\Sigma}^{-1}T)^{-1}D \quad (1.11b)$$

In a similar way we obtain empirical estimates for the covariance matrices of these estimators by replacing Σ by $\hat{\Sigma}$ in the expressions (1.6) and (1.9). K h a t r i [6] showed that \hat{B} is the maximum-likelihood (ML) estimate of B with the property that

$$|(Y - XBT')'(Y - XBT')| \text{ is minimum at } B = \hat{B} \quad 1.12$$

However, \hat{B} is no longer BLUE. Note that $\hat{\Sigma}$ is not the ML estimate of Σ ; the ML estimate under the normality assumption (K h a t r i [6]) is

$$\hat{\Sigma}_{ML} = \frac{1}{n}S + \frac{1}{n}(\hat{\Gamma} - \hat{B}T')'X'X(\hat{\Gamma} - \hat{B}T')$$

where $S = Y'[I - X(X'X)^{-1}X']Y$ and $\hat{\Gamma} = (X'X)^{-1}X'Y$. Generally \hat{B} is an unbiased estimator of B for all symmetric distributions of Y . Although \hat{B} is not BLUE, $\hat{\Sigma}$ is a consistent estimator of Σ and if n is large, $\hat{\Sigma}$ would be near Σ , and \hat{B} would be near the BLUE.

2. INFLUENCE OF A PART OF DATA

We are interested in the effect of deleting a part of measurements from data. Let \hat{B} be the estimate of B based on full data and let \hat{B}_A be an alternative estimator based on a subset of data. The subset of data can be obtained by deleting observations (some rows of Y), by deleting measurements at a given time-point (certain columns of the observation matrix Y and corresponding rows of T) or by deleting any other subset of measurements from the data. In this application the influence of deleting measurements at given time-points is of primary interest, but we will consider the problem of assessing the influence more generally.

Let I be an s vector of indices that specify the incomplete observations. $Y_{(I)}$ denote the observations, where no measurements are deleted. $Y_{(I)}$ can also be empty. Let Y_I denote that set of observations from which some measurements are deleted. Further, we partition Y_I such that Y_{IJ} contains the deleted measurements and $Y_{I(J)}$ the rest of the data contained in Y . For

example, data on bulls born in 1966 contain 208 bulls measured at 12 time-points. If we delete for 10 bulls measurements at the ages of 90 and 120 days, Y_{IJ} will be a 10×2 matrix and $Y_{I(J)}$ a 10×10 matrix. The growth curve model (1.1) can be expressed as a set of models as follows

$$\begin{aligned} E(Y_{(I)}) &= X_{(I)}BT' \\ E(Y_{I(J)}) &= X_I BT'K \end{aligned} \quad (2.1)$$

and

$$E(Y_{IJ}) = X_I BT' \bar{K},$$

where $X_{(I)}$ is the $n_1 \times m$ known across-individuals design matrix for complete observations and X_I is the $n_2 \times m$ design matrix for incomplete observations. The matrices B and T are the same as in model (1.1). K is a $q \times q_1$ incidence matrix of zeroes and ones which indicate the times with measurements for cases indexed by I . Correspondingly \bar{K} is a $q \times (q - q_1)$ incidence matrix indicating the times with missing measurements. When measurements at the ages of 90 and 120 days are deleted, \bar{K} is as follows

$$\bar{K} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and K is a 13×11 matrix of zeroes and ones. Note that $KK' + \bar{K}\bar{K}' = I_{13 \times 13}$. Denoting $\text{vec}Y'_{(I)} = \underline{y}_1$, $\text{vec}Y'_{I(J)} = \underline{y}_2$ and $\text{vec}Y'_{IJ} = \underline{y}_3$ we may write the model as follows

$$\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \end{pmatrix} = \begin{pmatrix} X_{(I)} \otimes T \\ X_I \otimes K'T \\ X_I \otimes \bar{K}'T \end{pmatrix} \text{vec}B' + \text{vec}E' \quad (2.2)$$

Let us further denote $K'T = T_{(J)}$, $\bar{K}'T = T_J$, and $K'EK = \Sigma_{(J)}$ where " (J) " denotes the indices of cases deleted from the data. If \underline{y}_3 is deleted from data, one obtains

$$\begin{aligned}
\text{vec} \tilde{B}'_{(IJ)} &= [X'_{(I)} X_{(I)} \otimes T' \Sigma^{-1} T + \\
&+ X'_I X_I \otimes T'_{(J)} \Sigma^{-1}_{(J)} T_{(J)}]^{-1} [(X_{(I)} \otimes \Sigma^{-1} T)' y_1 + \\
&+ (X_I \otimes \Sigma^{-1}_{(J)} T_{(J)})' y_2] \quad (2.3)
\end{aligned}$$

Substituting S into (2.3) in place of Σ yields the estimate $\tilde{B}_{(IJ)}$ for B , when y_{IJ} is deleted from the data. The empirical influence function is now

$$IF(IJ) = \tilde{B} - \tilde{B}_{(IJ)} \quad (2.4)$$

When $\Sigma = I$, we can derive convenient matrix formulas for the difference $\hat{B} - \hat{B}_{(IJ)}$. However, in this connection we do not consider this an important special case.

2.1. Measuring Influence at the Design Stage

In order to obtain a measure of information we compare traces of covariance matrices. Although measurements indexed by IJ are unavailable, the proportion of change in the trace of covariance matrix of the estimator of $(C \otimes D)' \text{vec} B'$ can be defined as

$$I_{(IJ)}(CBD) = \frac{\text{tr} V(\hat{CB}_{(IJ)}^D) - \text{tr} V(\hat{CBD})}{\text{tr} V(\hat{CBD})} \quad (2.5)$$

Note that $CBD = \{c'_{(i)} B d_j\}$, where $C = (c_{(1)}, c_{(2)}, \dots, c_{(g)})$ and $D = (d_1, d_2, \dots, d_v)$. However, we adopt a slightly different approach, which is more flexible and more simple. We calculate information on every $c'_{(i)} B d_j$ according to the formula (2.5), and denote it as

$$I_{(IJ)}(c'_{(i)} B d_j) = \frac{[c_{(i)} \otimes d_j]' [V(\hat{B}_{(IJ)}) - V(\hat{B})] [c_{(i)} \otimes d_j]}{[c_{(i)} \otimes d_j]' V(B) [c_{(i)} \otimes d_j]} \quad (2.6)$$

As information measure for CBD could be defined as a weighted sum of the information measures (2.6). If all elements of CBD

are of equal interest, then the arithmetic mean of $I_{(IJ)}(c'_{(i)}Bd_j)$ would be appropriate:

$$I_{(IJ)}^{(CBD)} = \sum_{i=1}^g \sum_{j=1}^v I_{(IJ)}(c'_{(i)}Bd_j) / (gv) \quad (2.7)$$

Therefore with respect to the parameter matrix B , the information contained in the measurements indexed by IJ is simply the average of information figures of the elements of B . We suggest calculating the information matrix

$$IM_{(IJ)}(B) = \{I_{(IJ)}(\beta_{ij})\} \quad (2.8)$$

which proves a useful statistic. Looking at this matrix we can see the information contained in the observations indexed by IJ with respect to every element of B . We can easily set that

$$I_{(IJ)}(c'_{(i)}Bd_j) = c'_{(i)}IM_{(IJ)}(B)d_j \quad (2.9)$$

G h o s h [5] suggested a kind of measure similar to that in (2.6) in the context of ordinary linear models.

2.2. Influence at the Inference Stage

Perhaps the most popular influence measure at the inference stage in the context of regression models is the distance measure proposed by C o o k [4]. No similar measure can be used straightforwardly in a growth curves model. In order to derive a measure for influence suitable in multivariate situations, we consider first testing the hypothesis

$$H_0 : CBD = M \quad (2.10)$$

in the model

$$E(\text{vec} Y') = (X \otimes T) \text{vec} B' \quad (2.11)$$

The hypothesis (2.10) can be expressed equivalently by

$$(C \otimes D') \text{vec} B' = \text{vec} M' \quad (2.12)$$

In order to simplify notation we choose $M = 0$ in the sequel. If necessary $(C\hat{D}')\text{vec}B'$ could easily be replaced by $(C\hat{D}')\text{vec}B' - \text{vec}M'$ in the following formulas. If Σ were known, we could use the ordinary χ^2 statistic known from the theory of linear models. If we denote $C\hat{D}' = K'$, we have

$$\begin{aligned} \chi^2 &= (K'\text{vec}\hat{B}')' \{K' (X\hat{\omega}T)' (I\hat{\omega}\Sigma)^{-1} (X\hat{\omega}T)^{-1} K\}^{-1} (K'\text{vec}\hat{B}') = \\ &= (K'\text{vec}\hat{B}')' \{[C(X'X)^{-1}C']^{-1} \hat{\omega}[D'T'\Sigma^{-1}T)^{-1}D]\} (K'\text{vec}\hat{B}') = \\ &= \text{tr} \{S_H [D'(T'\Sigma^{-1}T)^{-1}D]^{-1} \} \end{aligned} \quad (2.13)$$

where

$$S_H = (C\hat{B}D)' [C(X'X)^{-1}C']^{-1} (C\hat{B}D) \quad (2.14)$$

It is well-known that under $H_0 : CBD = 0$ the statistic (2.13) follows the central χ^2 distribution with s and $n - mp$ degrees of freedom, where $s = gv$ is the number of rows in $C\hat{D}'$ being full row rank. In certain cases Σ might be known from other experiments, and this "estimate" could be used in place of Σ . Perhaps in some applications there are good reasons for the use of $\sigma^2 I$ in place of Σ . Usually, however, Σ is unknown.

Now we consider testing the hypothesis (2.12), when Σ is unknown. As noted in the preceding section, $\hat{\Sigma}$ defined in (1.10) is an unbiased estimator of Σ in the restricted linear model

$$E(Y) = X\Gamma \quad 2.15$$

where $\Gamma = BT'$. Otherwise the assumptions are the same as in the model (1.1). It is known (e.g. Muirhead [9], p. 430) that the maximum likelihood estimates of Γ and Σ in the model are $\hat{\Gamma} = (X'X)^{-1}X'Y$ and $(1/n)S = (1/n)Y'[I - X(X'X)^{-1}X']Y$. Moreover $(\hat{\Gamma}, S)$ is sufficient for (Γ, Σ) . The maximum likelihood estimates $\hat{\Gamma}$ and S are independently distributed and $S \sim W_q(n-m, \Sigma)$. Since $(1/n)S$ is the ML estimate of Σ in the model (2.15), $(1/n)S$ is a consistent estimator of Σ , or in other words $(1/n)S$ converges in probability to Σ as n increases without limit. We write $(1/n)S^P \rightarrow \Sigma$ as $n \rightarrow \infty$. This result can also be proved using the weak law of large members

(Arnold [2], p. 365) and the fact that $S \sim W_q(n-m, \Sigma)$. Naturally the unbiased estimator $\hat{\Sigma}$ is also consistent.

If we substitute $\hat{\Sigma}_{ML}$ for Σ in the expression of the statistic (2.13), we no longer know the distribution of the resulting statistic. However, the asymptotic distribution of this statistic can be obtained.

After this substitution we have

$$(K' \text{vec} \tilde{B})' \{K' [(X\tilde{\Sigma}T)' (I \otimes \hat{\Sigma}_{ML})^{-1} (X\tilde{\Sigma}T)]^{-1} K\}^{-1} (K' \text{vec} \tilde{B}) \quad (2.16)$$

which is the Wald Statistic for testing $H_0: K' \text{vec} \tilde{B}' = 0$ (see e.g. Silvey [15], p. 116). Wald [17] showed that under the null hypothesis (2.16) is asymptotically distributed as $\chi^2(s)$, where s is the number of rows in K' . We noted above that $(1/n)S$ is a consistent estimator of Σ (and also $\hat{\Sigma} = [1/(n-m)]S$). Since $\hat{\Sigma}_{ML}$ is a consistent estimator of Σ , the estimators Σ_{ML} and $\hat{\Sigma}_{ML}$ are asymptotically the same. We substitute $\hat{\Sigma}$ for $\hat{\Sigma}_{ML}$ in (2.16), which yields the statistic

$$Q_n = (n-m) \text{tr}(S_H S_E^{-1}) \quad (2.17)$$

where S_H is as in (2.13) and

$$S_E = D'(T'S^{-1}T)^{-1}D \quad (2.18)$$

Not it is easy to see that also

$$Q_n \sim \chi^2(s) \quad (2.19)$$

asymptotically, when H_0 is true. We call W_n the Wald Statistic. Kleibaum [7] proposed this approach for testing linear hypotheses in this generalized growth curve model.

To determine the degree of influence the measurements indexed by IJ have in the estimate \tilde{B} , we suggest the measure defined by

$$D_{(IJ)}(\text{CBD}) = \text{tr}(S_H S_E^{-1}) \quad (2.20)$$

where

$$S_H = [c(\tilde{B} - \tilde{B}_{(IJ)})D]'[c(X'X)^{-1}c']^{-1}[c(\tilde{B} - \tilde{B}_{(IJ)})D] \quad (2.21)$$

We again adopt the same approach as in introducing the design stage measure and calculate first the influence of every $c'_{(i)}Bd_j$. Thus we have

$$D_{(IJ)}(c'_{(i)}Bd_j) = \frac{[c'_{(i)}(\tilde{B} - \tilde{B}_{(IJ)})d_j]^2}{c'_{(i)}(X'X)^{-1}c_{(i)} \cdot d_j'(T'S_{(IJ)}^{-1}T)d_j} \quad (2.22)$$

The following relation holds between these measures:

$$D_{(IJ)}(c'_{(i)}Bd_j) = I_{(IJ)}(c'_{(i)}Bd_j) \cdot \frac{[c'_{(i)}(\tilde{B} - \tilde{B}_{(IJ)})d_j]^2}{V(c'_{(i)}\tilde{B}_{(IJ)}d_j) - V(c'_{(i)}\tilde{B}d_j)} \quad (2.23)$$

Suppose now for a moment that $\Sigma = I$. Then we can write

$$\begin{aligned} [c'_{(i)}(\hat{B} - \hat{B}_{(IJ)})d_j]^2 &= \\ &= (c'_{(i)}\otimes d'_j)(X'X\otimes T'T)^{-1}(X_I\otimes T_J)'AA'(X_I\otimes T_J)(X'X\otimes T'T)^{-1}(c_{(i)}\otimes d_j), \end{aligned} \quad (2.24)$$

where

$$A = (I - H_{IJ})^{-1}[\text{vec}Y'_{(IJ)} - (X_I\otimes T_J)\text{vec}B'] \quad (2.25)$$

It can easily be shown that $E(AA') = (I - H_{IJ})^{-1}$. Therefore (2.6) and (2.24) yield the result

$$E[D_{(IJ)}(c'_{(i)}Bd_j)] = I_{(IJ)}(c'_{(i)}Bd_j) \quad (2.26)$$

When $\Sigma = I$, the maximum value of $D_{(IJ)}(c'_{(i)}Bd_j)$ is $A'(X_I\otimes T_J)(X'X\otimes T'T)^{-1}(X_I\otimes T_J)'A$, which is the largest eigenvalue of the matrix $(X_I\otimes T_J)'AA'(X_I\otimes T_J)(X'X\otimes T'T)^{-1}$. Further, it can easily be discerned (see e.g. Ghosh [5]) that

$$E[\max_{\underline{c}, \underline{d}} D_{(IJ)}(\underline{c}' B \underline{d})] = \max_{\underline{c}, \underline{d}} I(\underline{c}' B \underline{d}) \quad (2.27)$$

It should be emphasised that the preceding identities do not hold in the case where \hat{B} is not equal to $(X'X)^{-1}X'YT(T'T)^{-1}$.

3. AN EXAMPLE

Now we investigate the influence of deleting measurements at different time-points, when data on 208 bulls born in 1966 are under consideration. A polynomial of third degree was fitted them. In Figure 1 the values of the inference stage influence measure (2.20) for different ages are given.

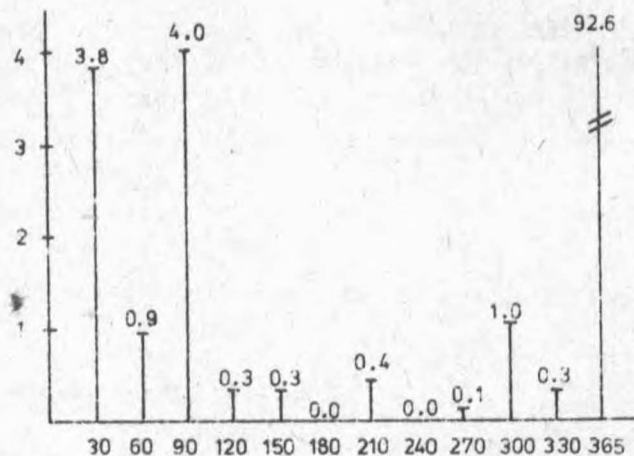


Fig. 1. The values of the influence measure $D_{(J)}^{(8)}$ for different time-points when the measurements at the corresponding point are deleted. Data concern bulls born in 1966

When the measurements are equally spaced, it is expected that the observations at the beginning and at the end of the sample period are most influential. This follows from the fact that deleting observations at the ends of the sample period reduces most the variance of \hat{B} (see the identities (2.5) and (2.26)). Note that the statistic (2.17) is the Lawley-Hotelling trace statistic. This is used as a basis for our influence measure, sin-

ce the statistic can be easily interpreted as a distance measure as can be seen from (2.16) and from (2.21). The magnitude of $(n - m)D_{(IJ)}(\text{CBD})$ may be assessed by comparing it to the probability points of the corresponding Lawley-Hotelling statistic under H_0 : $\text{CBD} = 0$.

The degree of influence is greatest at the points, 30, 90 and 365 days of ages. For economic and other practical reasons, measurements at the ages 30, 90, 120 and 150 days were not taken after the year 1970. If we delete these ages from the data on bulls born in 1966, the degree of influence $D_{(J)}(B) = 302.7$, where $J = (1, 3, 4, 5)$. Deleting the first three ages 30, 60 and 90 days yields the value 3593.4 for $D_{(J)}(B)$, but dropping out the four ages 60, 120, 150 and 240 gives the value 0.5. The 95% significance point of the corresponding Lawley-Hotelling statistic is 0.06. Therefore 0.5 is highly significant. If 4 weighing times must be deleted, one natural approach is to find such points which have least influence on the estimates. Finding the minimum is not straightforward since some time-points may be jointly influential but individually uninfluential, and conversely: some time-points may be individually influential but jointly uninfluential.

However, there might also be some practical side-conditions for selecting weighing times. On the other hand, it may be important to attain a good fit to data in individual bulls specially at a given age interval. Of course, the influence of a given set of measurements also depends on the mathematical form of a chosen growth curve. Therefore this influence analysis serves as a means for comparing the robustness of various models to missing measurements and to different study designs (i.e. sets of target ages).

REFERENCES

- [1] Anderson T. W. (1958), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.
- [2] Arnold S. F. (1981), *The Theory of Linear Models and Multivariate Analysis*, Wiley, New York.

- [3] B e l s l e y D. A., K u h, W e l c h R. E. El. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, J. Wiley and Sons, New York.
- [4] C o o k R. D. (1977), *Detection of Influential Observations in Linear Regression*, "Technometrics", 19, 15-18.
- [5] C o o k R. D., W e i s b e r g S. (1980), *Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression*, "Technometrics", 22, 495-508.
- [6] C o o k R. D., W e i s b e r g S. (1982), *Residuals and Influence in Regression*, Chapman and Hall, New York.
- [7] G h o s h S. (1983), *Influential Observations in View of Design and Inference*, Comm. in Statist. Theory and Meth., 12(14), 1675-1683.
- [8] K h a t r i C. G. (1966), *A Note on a MANOVA Model Applied to Problems in Growth Curves*, Ann. Inst. Statist. Math., 18, 75-86.
- [9] K l e i n b a u m D. G. (1973), *A Generalization of the Growth Curve Model which Allows Missing Data*, J. Multiv. Anal., 3, 117-124.
- [10] M o r r i s o n D. F. (1976), *Multivariate Statistical Methods* (2nd Ed.), McGraw-Hill Kogakusha, LTD, Tokyo.
- [11] M u i r h e a d R. J. (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.
- [12] P o t t h o f f R. F., R o y S. N. (1964), *A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems*, "Biometrika", 51, 313-326.
- [13] R a o C. R. (1965), *The Theory of Least Squares when the Parameters Are Stochastic and Its Application to the Analysis of Growth Curves*, "Biometrika", 52, 447-58.
- [14] R a o C. R. (1966), *Convariance Adjustment and Related Problems* [in: *Multivariate Analysis*, Academic Press, New York, 87-103.
- [15] R a o C. R. (1967), *Least Squares Theory Using an Estimated Dispersion Matrix and Its Application to Measurement of Signals*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 355-372.
- [16] R o h a t g i V. K. (1976), *An Introduction to Probability Theory and Mathematical Statistics*, Wiley, New York.
- [17] S i l v e y S. D. (1975), *Statistical Inference*, Chapman and Hall, London.
- [18] T i m m N. H. (1975), *Multivariate Analysis with Applications in Education and Psychology*, Wadsworth Publishing Co., Belmont.

- [19] W a l d A. (1943), *Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations Is Large*, Transac. Americ. Math. Soc., 54, 426-482.

Erkki P. Liski

OBSERWACJE WPŁYWOWE W UOGÓLNIONEJ ANALIZIE
MODELU WARIANCJI

Podano opis modelu GMANOVA wielowymiarowej analizy wariancji (zwanego czasem modelem krzywych wzrostu).

Dyskutowano problemy analizy skutków występowania wpływowych wyników obserwacji na własności estymatorów. Okazało się, że skutki te są różne w zależności od kształtu estymowanej funkcji parametrycznej.

Proponuje się pomiar tych skutków w fazie planowania eksperymentów oraz w fazie analizy danych eksperymentalnych. Wyniki analizy zilustrowano rezultatami badań eksperymentalnych z zakresu hodowli zwierząt.