Statistics & Probability Letters 8 (1989) 163–166
North-Holland
June 1989

EFFECT OF A MISSING OBSERVATION ON HOTELLING'S GENERALIZED T_0^2

Anant M. KSHIRSAGAR

Biostatistics Department, University of Michigan, Ann Arbor, MI 48109, USA

Received March 1988
Revised August 1988

Abstract: A generalization of the missing plot technique in univariate linear models to the multivariate case is given in this paper. The estimate of the missing observation vector produces the correct error matrix and the correct parameter estimates but the hypothesis matrix is biased. The adjustment required to the test, for achieving the desired level of significance, when the biased hypothesis matrix is used, is derived.

Keywords: missing observation, Hotelling's generalized T_0^2, linear combination of Wishart matrices, bias in hypothesis matrix, multivariate linear model.

1. Introduction

If an observation in a univariate model is missing, the missing plot technique consists in replacing the missing observation by an algebraic quantity that minimizes the error sum of squares (s.s.). This technique preserves the original design matrix and gives the correct error s.s. and the correct least squares estimates of parameters. In this paper, we show that, for a multivariate linear model, when one observation vector is missing, we can use this technique on each of the variables separately and then we get, not only the correct error s.s. for each variable, but also the correct error sum of products (s.p.) also, for every pair of variables.

However, as in univariate linear models, this technique does not give the correct hypothesis matrix for testing a linear hypothesis. Hotelling's generalized T_0^2, used for testing a linear hypothesis will then be biased.

Most practitioners will like to ignore this bias and proceed but then the desired level of significance is not achieved. The distribution of this T_0^2 and the modifications required to the test, to obtain the desired level of significance, are investigated.

2. Multivariate linear model

Consider the linear model,

$$Y = X\beta + \varepsilon, \tag{2.1}$$

where Y is $n \times p$, X is $n \times m$ of rank m, β is $m \times p$ and ε is $n \times p$. X is known, β is unknown and the rows of ε have independent p-variate normal distributions, with the same but unknown covariance matrix Σ. This is expressed as

$$V(\varepsilon) = \Sigma \otimes I_n, \tag{2.2}$$

where $\otimes$ denotes Kronecker product of two matrices, and V stands for the variance–covariance matrix. The least squares estimate of β is

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2.3}$$

and

$$V(\hat{\beta}) = \Sigma \otimes (X'X)^{-1}. \tag{2.4}$$

For testing the hypothesis

$$H_0: \beta = 0,$$

the hypothesis matrix is

$$H = \hat{\beta}'(X'X)\hat{\beta}. \tag{2.5}$$

which has the central Wishart distribution with m degrees of freedom (d.f.) and scale matrix Σ. We denote this by

$$H \sim W_p(H; \Sigma, m). \tag{2.6}$$

The error matrix is

$$E = Y'\left(I_n - X(X'X)^{-1}X'\right)Y \tag{2.7}$$

and has the $W_p(E; \Sigma; n-m)$ distribution, independent of H. The hypothesis H_0 is then tested by the Hotelling's generalized T_0^2 test criterion,

$$T_0^2 = \operatorname{tr}(E^{-1}H). \tag{2.8}$$

where tr stands for the trace of a matrix. The distribution of T_0^2 depends on $n-m$, m and p only. Let us denote the distribution of T_0^2 by $\mathrm{T}_0^2(n-m, m, p)$ and its $100(1-\alpha)\%$ point by $T_\alpha^2(n-m, m, p)$. These percentage points are available in Pillai and Young (1971). We reject H_0 if the observed T_0^2 exceeds $T_\alpha^2(n-m, m, p)$ and the level of significance is then α.

3. A missing observation in the multivariate model

Partition Y, X as

$$Y = \begin{bmatrix} Y_e \\ y_n' \end{bmatrix}, \qquad X = \begin{bmatrix} X_e \\ x_n' \end{bmatrix}, \tag{3.1}$$

where y_n' and x_n' are row vectors. Suppose the nth observation y_n' is missing. The correct least squares estimate of β is then

$$\hat{\beta} = (X_e'X_e)^{-1}X_e'Y_e. \tag{3.2}$$

where e stands for existing observations. The variance–covariance matrix of this $\hat{\beta}$ is

$$V(\hat{\beta}) = \Sigma \otimes (X_e'X_e)^{-1}, \tag{3.3}$$

and the correct error matrix is

$$Y_e'\left(I_{n-1} - X_e(X_e'X_e)^{-1}X_e'\right)Y_e, \tag{3.4}$$

with $n-1-m$ d.f.

However, usually the original design matrix X is well planned to yield an $(X'X)^{-1}$ with a good pattern but $(X_e'X_e)^{-1}$ is not so. The missing plot technique consists in augmenting the existing observation Y_e by a row vector l', forming the augmented matrix,

$$Y_a = \begin{bmatrix} Y_e \\ l' \end{bmatrix}. \tag{3.5}$$

The error matrix

$$E_a = Y_a'\left(I_n - X(X'X)^{-1}X'\right)Y_a \tag{3.6}$$

is then calculated and l is determined by minimizing the trace of E_a with respect to l. (This is the same as determining l_i, the ith component of l, by minimizing the ith diagonal element of E_a or in short using the missing plot technique separately for each of the p variables.) We find that

$$l' = x_n'(X'X)^{-1}X_e'Y_e. \tag{3.7}$$

If we substitute this value back in E_a, it is the same as the correct error matrix (3.4), and also that

$$\hat{\beta}_a = (X'X)^{-1}X'Y_a \tag{3.8}$$

is the correct least squares estimate $\hat{\beta}$ of (3.2). Thus not only the correct error sums of squares of the p variables are producted but also the correct error sum of products are also obtained. The degrees of freedom are however one less than the usual $n-m$.

But the matrix

$$H_a = \hat{\beta}_a(X'X)\hat{\beta}_a, \tag{3.9}$$

following (2.5) is *not the correct hypothesis matrix*, because the variance–covariance matrix of $\hat{\beta}_a$ is *not* (2.4), but

$$V(\hat{\beta}_a) = \Sigma \otimes (X_e'X_e)^{-1}. \tag{3.10}$$

Consequently, Hotelling's statistic

$$T_a^2 = \operatorname{tr}\left(E_a^{-1}H_a\right) \tag{3.11}$$

is not the correct one. The test that rejects H_0 when

$$T_a^2 \geqslant T_\alpha^2(n-1-m, m, p) \tag{3.12}$$

does not have α as the level of significance, as one might think. We need to derive the distribution of T_a^2 and seek modifications in the above test to achieve the desired level of significance α.

4. Distribution of T_a^2

The correct hypothesis matrix, when an observation is missing, is

$$H = \hat{\beta}_a'(X_e'X_e)\hat{\beta}_a \tag{4.1}$$

$$= \hat{\beta}_a'(X'X - x_n x_n')\hat{\beta}_a$$

$$= H_a - \hat{\beta}_a' x_n x_n' \hat{\beta}_a. \tag{4.2}$$

From (3.10),

$$V(\hat{\beta}_a' x_n) = x_n'(X_e'X_e)^{-1} x_n \cdot \Sigma$$

$$= x_n' \left[(X'X)^{-1} + \frac{(X'X)^{-1} x_n x_n' (X'X)^{-1}}{1 - x_n'(X'X)^{-1} x_n} \right] x_n \cdot \Sigma$$

$$= [b/(1-b)]\Sigma, \tag{4.3}$$

where

$$b = x_n'(X'X)^{-1} x_n. \tag{4.4}$$

Let

$$Z = (1-b)^{1/2} b^{-1/2} \hat{\beta}_a' x_n. \tag{4.5}$$

It is easy to see that, under H_0, Z has a p-variate normal distribution with zero means and covariance matrix Σ; that is

$$Z \sim N_p(0, \Sigma). \tag{4.6}$$

From (4.2),

$$H_a = H + b(1-b)^{-1} ZZ'$$

$$= (H - ZZ') + (1-b)^{-1} ZZ'$$

$$= H^* + (1-b)^{-1} ZZ', \tag{4.7}$$

where

$$H^* = H - ZZ'. \tag{4.8}$$

It can now be proved that H^* and ZZ' are independently distributed. A heuristic argument is that H is the hypothesis matrix for testing $\beta = 0$, while ZZ' is the hypothesis matrix for testing a linear function of β, and so ZZ' is a part of H. A rigorous proof will consist of expressing $H - ZZ'$ and ZZ' as $Y_e'PY_e$ and $Y_e'QY_e$ and showing that PQ is null and that P, Q are idempotent matrices.

Finally, therefore,

$$T_a^2 = \operatorname{tr}(E_a^{-1} H_a)$$

$$= \operatorname{tr} E_a^{-1}\left(H^* + (1-b)^{-1} ZZ'\right). \tag{4.9}$$

where

$$E_a \sim W_p(E_a;\ \Sigma;\ n-1-m), \tag{4.10}$$

$$H^* \sim W_p(H^*;\ \Sigma;\ m-1), \tag{4.11}$$

and

$$Z \sim N_p(0, \Sigma), \tag{4.12}$$

all the three being independent.

Since $H^* + (1-b)^{-1} ZZ'$ is a linear combination of independent Wishart matrices, and not the sum, the exact null distribution of T_a^2 is involved. But like the linear combination of independent χ^2 variables we shall approximate the distribution of

$$M = H^* + (1-b)^{-1} ZZ' \tag{4.13}$$

by that of a suitable multiple of a Wishart matrix. Tan and Gupta (1983) have studied such an approximation. In the next section, we derive this, by a slightly different simpler approach and then will return to the distribution of T_a^2.

5. Linear combination of Wishart matrices

Let us approximate the distribution of M defined in (4.13) by a Wishart distribution of scale matrix Γ and degrees of freedom f, both chosen in such a way that the first and second order moments of M agree with those of this Wishart distribution. The mean of the Wishart distribution if $f\Gamma$ and the covariance matrix of the p^2 elements of the Wishart matrix, written as a vector (notation vec) by stacking the columns of the matrix is (Muirhead, 1982)

$$f(I_{p^2} + K)(\Gamma \otimes \Gamma), \tag{5.1}$$

where the explicit of definition of K is not necessary here, though it is given in Muirhead (1982). Equating these to the mean

$$E(M) = \left[(m-1) + (1-b)^{-1}\right]\Sigma$$

$$= a_1 \Sigma, \quad \text{say}, \tag{5.2}$$

and to the covariance matrix

$$V(\text{vec } M) = a_2(I_{p^2} + K)(\Sigma \otimes \Sigma) \tag{5.3}$$

with

$$a_2 = (m-1) + (1-b)^{-2}, \tag{5.4}$$

we find

$$f = a_1^2/a_2, \qquad \Gamma = (a_2/a_1)\Sigma. \tag{5.5}$$

Thus M is approximately distributed as

$$\mathrm{W}_p(M;\ (a_2/a_1)\Sigma;\ a_1^2/a_2). \tag{5.6}$$

6. Approximate distribution of T_{a}^2

From (4.9),

$$\begin{aligned} T_{\mathrm{a}}^2 &= \operatorname{tr} E_{\mathrm{a}}^{-1} M \\ &= (a_2/a_1) \operatorname{tr}\left[((a_2/a_1)E_{\mathrm{a}})^{-1} M\right], \end{aligned} \tag{6.1}$$

where $(a_2/a_1)E_{\mathrm{a}}$ and M are independent Wishart matrices with the same scale matrix $(a_2/a_1)\Sigma$ and d.f. $n-1-m$ and f respectively. Hence

$$(a_1/a_2)T_{\mathrm{a}}^2 \sim \mathrm{T}_0^2(n-m-1,\ f,\ p). \tag{6.2}$$

Therefore, the hypothesis H_0 should be rejected when

$$T_a^2 > (a_2/a_1)T_\alpha^2(n-1-m,\ f,\ p), \tag{6.3}$$

to achieve the desired level of significance. If one uses the test (3.12), the level of significance will not be α but

$$\begin{aligned} \gamma &= \operatorname{Prob}\left(T_{\mathrm{a}}^2 \geqslant T_\alpha^2(n-1-m,\ m,\ p)\right) \\ &= \operatorname{Prob}\big((a_1/a_2)T_{\mathrm{a}}^2 \\ &\qquad \geqslant (a_1/a_2)T_\alpha^2(n-1-m,\ m,\ p)\big) \end{aligned} \tag{6.4}$$

and hence γ' will be given by

$$\begin{aligned} &(a_1/a_2)T_\alpha^2(n-1-m,\ m,\ p) \\ &\quad = T_\gamma^2(n-1-m,\ f,\ p). \end{aligned} \tag{6.5}$$

Even (6.3) and (6.5) are not exact as we approximated the distribution of M, but it is a better approximation as noticed by Tan and Gupta (1983).

7. Some comments

(a) It is possible to use the correct H given by (4.2) but in practice, it is more convenient to substitute the vector $\boldsymbol{l}$ of (3.7) for the missing observations and proceed in the "usual" way using the well patterned $X'X$ and $(X'X)^{-1}$.

(b) There are other multivariate criteria like Wilks' lambda namely

$$|E|/|E+H|$$

or Pillai's

$$\operatorname{tr}(E+H)^{-1}H$$

and the results in this paper can be used for these criteria too with some adjustments.

(c) We considered only one missing observation, but it is a straightforward generalization to extend it to several missing observations. If s observations are missing, E will have $n-s-m$ d.f.

(d) We considered the "full" hypothesis $\beta = 0$ in this paper, but a "sub" hypothesis or a general linear hypothesis can be dealt with very similarly.

(e) We assumed $X'X$ to be of full rank, but if that is not so, a generalized inverse should replace $(X'X)^{-1}$, with appropriate changes in the d.f. $n-m$ of E.

Acknowledgements

The author is indebted to Professors S. Kocherlakota and K. Kocherlakota of the University of Manitoba, for valuable discussions and facilities for carrying out this research in Winnipeg and to the NSERC grant of Canada.

References

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory* (Wiley, New York).

Pillai, K.C.S. and D.C. Young (1971), On the exact distribution of Hotelling's generalized T_0^2, *J. of Multivariate Anal.* **1**, 90–97.

Tan, W.Y. and R.P. Gupta (1983), On approximating a linear combination of central Wishart matrices with positive coefficients, *Comm. Statist.* **12**, 2589–2595.