ADDITIONAL VARIABLES AND ADJUSTED ESTIMATES WITH ARBITRARY KNOWN VARIANCE-COVARIANCE STRUCTURE

Robert Schall
Messerschmitt-Bolkow-Blohm Gmbtt
D-8000 Munchen, Germany
and University of Cape Town

Timothy T. Dunne
University of Cape Town
Rondebosch
RSA

Technical Report No. 474
June, 1986

RESIDUAL RANKS TESTING FOR OUTLIERS IN 2^n DESIGNS

Robert Schall
Messerschmitt-Bolkow-Blohm Gmbtt
D-8000 Munchen, Germany
and University of Cape Town

Timothy T. Dunne*
University of Cape Town
Republic of South Africa

Key words and phrases: outliers; slippage problems; nonparametric tests; experimental designs

ABSTRACT

For a variety of experimental designs with 2^n observations, inter alia 2^n factorial designs, simple data sets with 2^n observations and two-sample problems where both sample sizes are respectively 2^{n-1}, we present a method to apply well known nonparametric tests such as those of Mann-Wilcoxon, Wald-Wolfowitz and Fisher-Yates to residuals for the detection of outliers.

* Partial support from University of Minnesota & by sabbatical leave grant from FRD, CSIR, South Africa.

1. INTRODUCTION

We consider the linear model (LM)$(Y, X\beta, \sigma^2 V)$ or equivalently

$$y = X\beta + e; \quad \text{cov}(e) = \sigma^2 V \tag{1.1}$$

where V is an arbitrary known non-negative definite and symmetric matrix. It is well known (see e.g. Rao, 1976) that the best linear unbiased estimator (BLUE) $\hat{X\beta}$ for $X\beta$ is given by $\hat{X\beta} = P_{x|vz}y$, where $P_{x|vz}$ denotes the projection operator onto the column space C(X) along the space C(VZ), Z being a matrix of maximum rank such that $Z'X = 0$. As Rao (1976) points out, $P_{x|vz}$ need not be unique, but $\hat{X\beta}$ or $P_{x|vz}y$ is unique since $y \in C([X:V]) = C([X:VZ])$, with probability 1 (w.p.1), if the model (1.1) is consistent with the data. One possible choice for $P_{x|vz}$ is $P_{x|vz} = X(X'V^*X)^-X'V^*$, where V^* is a g-inverse of V in the manner of Rao (1971), i.e. $V^* = (V + XUX')^-$, for U such that $C([X:V]) = C(V + XUX')$ and $C(V) \cap C(XUX') = \{0\}$. One possible choice for the projection operator $P_{vz|x}$ is then $P_{vz|x} = (I - X(X'V^*X)^-X'V^*)$. The main computational task involved in the analysis of the LM(1.1) is the computation of $P_{x|vz}$, whether we wish to estimate $X\beta$ or to test linear hypotheses in the model.

If the model (1.1) is augmented by the new variables A (say), we obtain the model

$$y = [X:A]\begin{bmatrix}\beta \\ \lambda\end{bmatrix} + e; \qquad \text{cov}(e) = \sigma^2 V \tag{1.2}$$

Seemingly, in order to compute the BLUE $\hat{X\beta} + \hat{A\lambda}$ for $X\beta + A\lambda$ under model (1.2) we have to compute a new a projection operator $P_{\dot{x}|v\dot{z}}$, where $\dot{X} = [X:A]$ and $\dot{Z}$ is a matrix of maximum rank such that $\dot{Z}'\dot{X} = \{0\}$. But for the case $V = I$ it is well known (see e.g. Searle, 1971; similar developments are given in Rao, 1973

nonparametric tests on the transformed data. Thus the method is easily implemented and derives its efficiency from that nonparametric test which is selected.

1. APPLICABLE DESIGNS

Consider the design matrix T of a 2^n factorial design. If all main effects and interactions are fitted (the "full model"), T is of the order $(2^n \times 2^n)$ and it can be written as

$$T = D \otimes D \ \ldots \otimes D = (\otimes D)^n, \text{ where } D = \begin{bmatrix} 1 & -1 \\ 1 & +1 \end{bmatrix}$$

and $\otimes$ denotes the Kronecker product of matrices.

The columns of T are mutually orthogonal and $2^{-n/2}T$ is an orthogonal matrix.

Now let

$$S = S^{(n)} = \{\text{matrix X of order } (2^n \times p) \mid \text{the p columns of X span the same space as some } q \leq p \text{ columns of T}\}$$

be a class of design matrices.

The method for the nonparametric testing for outliers outlined in the next section applies to all designs where the design matrix X is from S, or equivalently to all linear models $(Y, X\beta, \sigma^2 I)$ where $X \in S$.

This class includes a wide variety of experimental designs as illustrated by the following:

Example 1.1 A simple sample with 2^n observations can be modelled as a $LM(Y, X\mu, \sigma^2 I)$ where $X = [1,\ldots 1]'$ of dimension 2^n. Clearly $[1,\ldots,1]'$ is the first column of T and thus $X \in S$.

Example 1.2 A two-sample problem with sample sizes 2^{n-1} each can be modelled

(a) $\hat{y}^{(2)} = \hat{y}^{(3)}$

(b) $X\hat{\beta}^{(1)} = X\hat{\beta}^{(3)}$

(c) $P_{vz|x}A\hat{\lambda}^{(2)} = P_{vz|x}A\hat{\lambda}^{(3)} = P_{vz|x}A\hat{\lambda}^{(4)} = P_{vz|x}A\hat{\lambda}$

For the second equality we require $y \in C([P_{vz|x}A:V])$, that is to say, model (4) is consistent with the data. Otherwise, and in any case, y can be replaced by $\hat{e}^{(1)} = y - X\hat{\beta}^{(1)}$.

(d) $X\hat{\beta}^{(2)} = X\hat{\beta}^{(1)} - P_{x|vz}A\hat{\lambda}$.

This equation holds if $X\beta$ is estimable under model (2), [equivalently, $C(X) \cap C(A) = \{0\}$]. Otherwise (d) yields the BLUE $l'\hat{\beta}$ for any estimable linear function $l'\beta$ of β under (2).

(e) $P_{vz|x}A\hat{\lambda}$ is uncorrelated with $X\hat{\beta}^{(3)} = X\hat{\beta}^{(1)}$.

(f) The additional sum of squares due to fitting A in model (2) is

$$SSA = (P_{vz|x}A\hat{\lambda})'V^{-}(P_{vz|x}A\hat{\lambda})$$

(g) The total sum of squares in (2) (and (3)) can be decomposed into uncorrelated sums of squares as

$$\begin{aligned} SS &= SSR^{(2)} + SSE^{(2)} \\ &= SSR^{(1)} + SSA + SSE^{(2)} \\ &= SSR^{(1)} + SSA + (SSE^{(1)} - SSA), \end{aligned}$$

where SSR and SSE denote respectively the sum of squares for regression and the sum of squares for error in the model in question.

(h) The F-statistics associated with the hypothesis H_0: $A\lambda = 0$ in (2) (and (3)) is given by

$$F = \frac{SSA}{SSE^{(2)}} \cdot \frac{s-a}{a} = \frac{SSA}{SSE^{(1)} - SSA} \cdot \frac{s-a}{a},$$

which under normality follows an $F_{a,s}$-distribution, where a and s

Table 1.4: List of suitable designs.

(2^n observations are required in each case)

1	Simple sample
2	Two-sample problems with sample sizes 2^{n-1} each
3	2^k-sample problems ($k \in N$) with sample sizes $2^n i$, $i=1,\ldots k$
4	2^n factorial designs
5	2^n factorial designs in 2^r replicates
6	2^n factorial designs in 2^b blocks
7	Fractional 2^n factional designs
8	Two-way layouts with 2^t treatments in 2^{n-t} blocks
9	Complete block designs with 2^b blocks and 2^t treatments
10	Latin squares, Graeco-Latin squares

We see that the method can be applied in almost any ANOVA-type situation whenever the number of observations is 2^n. This condition can be taken into consideration even at the design stage of an experiment. But if this condition is not met and we have a simple sample (say) of $2^n < N < 2^{n+1}$ observations, we can choose some 2^n observations (including the possible outliers) at random from the sample and carry out the test for outliers in the reduced sample at the cost of some loss of power.

2. NONPARAMETRIC TESTING FOR A SINGLE OUTLIER

Let be $(Y, X\beta, \sigma^2 I)$ a linear model (LM) where $X \in S$. Without loss of generality we wish to test whether the last observation is an outlier. We write m for 2^n.

The adjusted model for an outlier, following John and Draper (1978) is

(e) $\operatorname{cov}(P_{vz|x}\, A\lambda, X\hat{\beta}^{(3)}) = \operatorname{cov}(P_{vz|x}\hat{A\lambda}, X\hat{\beta}^{(1)})$, from (b)

$= \operatorname{cov}(P_{vz|x}\hat{A\lambda}, P_{x|vz}\, y)$

$= 0.$

(f) $SSR^{(2)} = \hat{y}^{(2)\prime} V^* \hat{y}^{(2)}$

$= (X\hat{\beta}^{(2)} + \hat{A\lambda})' V^* (X\hat{\beta}^{(2)} + \hat{A\lambda})$

$= (X\hat{\beta}^{(1)} + P_{vz|x}\hat{A\lambda})' V^* (X\hat{\beta}^{(1)} + P_{vz|x}\hat{A\lambda})$, from (b)

$= (X\hat{\beta}^{(1)})' V^* X\hat{\beta}^{(1)} + (P_{vz|x}\hat{A\lambda})' V^* (P_{vz|x}\hat{A\lambda})$, from (e)

$= SSR^{(1)} + SSA.$

Then (g) and (h) are direct consequences of (f).

The lemma allows a complete treatment of the augmented model (2), without explicitly fitting this model. The projection operators $P_{x|vz}$ and $P_{vz|x}$ are known from and computed during the analysis of the original model (1). The BLUE $P_{vz|x}\hat{A\lambda}$ for $P_{vz|x}A\lambda$ can be computed using the reduced model (4) in an economical manner. Thereafter, adjusted parameter estimates and the extra sum of squares due to fitting A can be obtained using the results (d) and (f) of the lemma. Results (g) and (h) give the corresponding ANOVA and allow us to test the hypothesis H_0: $A\lambda = 0$.

For the formulation of the lemma we required that the additional variables A satisfy the condition $C(A) \subset C([X:V])$. If this is not the case, quantities like $P_{x|vz}A$ and $P_{vz|x}A$ are not invariant over the special choice of the projection operator in question, and the models (3) and (4) are not well-defined. However, since the estimated residual vector $\hat{e}$ in a linear model is from the space C(V) w.p.1 (Zyskind and Martin, 1969), we can write, with respect to model (2),

$\tilde{Y}_{q+1},\ldots\tilde{Y}_m$ in (5) and (6) are uncorrelated and identically distributed.

But more important is the following property: while ranking $\tilde{Y}_{q+1},\ldots,\tilde{Y}_m$ any perturbation of the data is equiprobable under H_o: $\gamma=0$ if any perturbation of the error terms $e_2,\ldots,e_m$ of the original data $Y_2,\ldots Y_m$ is equiprobable (even if the terms are not independently distributed).

Usually it is assumed that $e_1,\ldots e_m$ are independent and identically distributed, and thus we can apply such nonparametric tests as Mann-Wilcoxon, Fisher-Yates and Wald-Wolfowitz to the two-sample problem (6) in a routine way. (For an overview on nonparametric test procedures see e.g. Kendall and Stuart, 1973).

3. MORE THAN ONE POSSIBLE OUTLIER

In the case of 2 possible outliers (without loss of generality being the last 2 observations) the adjusted model similar to (3) can be written as

$$Y = \begin{bmatrix} X_{(=)} & : & 0 & 0 \\ x_{m-1} & : & 0 & 1 \\ x_m & : & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \\ \delta \end{bmatrix} + e \quad .$$

After transformation by $2^{-n/2}T$ and removal of a q terms we arrive at

$$\begin{bmatrix} \tilde{Y}_{q+1} \\ \tilde{Y}_m \end{bmatrix} = \begin{bmatrix} c_{q+1} & : & d_{q+1} \\ c_m & : & d_m \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} + \tilde{e}_{(-t)} \qquad (7)$$

similar to (6), where $\underline{d} = [d_1,\ldots,d_m]'$ is the second last column of T. Similarly, this can be generalized for $\ell > 2$ possible outliers, if 2^ℓ is not too large compared with 2^n. Save for the constant $2^{-n/2}$, the vectors $\underline{c}$, $\underline{d}$,... contain only components +1 and -1 and simultaneous nonparametric test procedures such as the Kruskal-Wallis k-sample test can be applied, with k =

in the case of A as in (3.1), and

$$y_1 - V_{12}V_{22}^{-1}\, y_2 = (X_1 - V_{12}V_{22}^{-1}X_2)\beta + (e_1 - V_{12}V_{22}^{-1}e_2),$$

$$\operatorname{cov}(e_1 - V_{12}V_{22}^{-1}e_2) = \sigma^2(V_{11} - V_{12}V_{22}^{-1}V_{21}) \qquad (3.4)$$

in the case of A as in (3.2), where V_{22} is taken to be nonsingular (see Schall and Dunne, 1985).

Here $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ and $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ are partitioned comformably with the partitioning of V in (3.2). The respective reductions of the LM(1.1) to the models (3.3) and (3.4) can be seen as two different methods to downdate a linear model. Model (3.3) is obtained by removing the mean $X_2\beta$ of y_2 from the model (1.1), whereas (3.4) is obtained by removing the error term e_2 from the model (1.1). The latter case is illustrated by writing e_1 as

$$\begin{aligned} e_1 &= e_1 - V_{12}V_{22}^{-1}e_2 + V_{12}V_{22}^{-1}e_2 \\ &= \tilde{e}_1 + V_{12}V_{22}^{-1}e_2. \end{aligned} \qquad (3.5)$$

Clearly, $\tilde{e}_1$ is uncorrelated with e_2, and thus $\tilde{e}_1$ is the error for e_1 adjusted for the covariate e_2. Note that during the reduction of model (1.1) to the model (3.3) this adjustment is not made, and hence the error term e_2 is still in the model (3.3), through correlation, as can be seen from (3.5). It is in fact only the mean $X_2\beta$ of y_2 which is removed from (1.1) to obtain (3.3). Of course, the two methods are identical in the case $V = I$, or more generally when V is a diagonal matrix, and in this case $X_2\beta$ and e_2, and thus the whole observation y_2, are removed from the model (1.1) simultaneously, by fitting the dummy variables (3.1) or (3.2) (see John and Draper, 1978).

Let N and M respectively denote the matrices

for a given α. But with the method of the foregoing sections the suggested test can be applied if the total sample size is $m=2^n$.

If the first sample consists only of one observation we may treat this observation as an outlier and test for significance as outlined in Section 2.

In the case of sample size 2 for the first sample we have the model

$$\begin{bmatrix} Y_1 \\ \\ Y_{m-2} \\ Y_{m-1} \\ Y_m \end{bmatrix} = \begin{bmatrix} 1 & : & 0 \\ \\ 1 & : & 0 \\ 1 & : & 1 \\ 1 & : & 1 \end{bmatrix} \begin{bmatrix} \mu_2 \\ \mu_1 \end{bmatrix} + e$$

After transformation by $2^{-n/2}T$ and possibly some rearrangement of the transformed data we arrive at

$$\begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \\ \\ \\ \\ \tilde{Y}_m \end{bmatrix} = 2^{-n/2} \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \\ 1 \\ 0 \\ -2 \\ \\ -2 \end{bmatrix} \mu_2 + \tilde{e}$$

We can now apply a 3-sample rank test (preferably based on the H-statistic).

5. ON THE EFFICIENCY OF THE OUTLIER-TESTS

The non-null distribution of the test-statistics based on the ranks of the transformed observations $\tilde{Y}_{q+1},\ldots\tilde{Y}_m$ is certainly intractable in the general case, but the efficiency of the tests can be judged by a comparison with the normal alternative. If the observations Y in the original model (3) follow a

We require that $V_{22}\lambda$, and hence λ, is estimable in (3.3) and (3.4), and that $C(A) \subset C([X:V])$. If $A=V_2$, the latter condition is trivially satisfied. When λ is not estimable in (3.3) or (3.4), then the true inverses M_{22}^{-1} and N_{22}^{-1} of M_{22} and N_{22} must be replaced by g-inverses M_{22}^{-} and N_{22}^{-} respectively, and the formulae (ii) and (iii) yield the BLUE's for any estimable linear function of the quantities in question. Downdating formulae are especially useful while treating the problem of outliers and influential observations in a linear regression model $(Y, X\beta, \sigma^2 I)$ (see Cook and Weisberg, 1982). Consequently, the above formulae will apply in treating the same problem in the general linear model $(Y, X\beta, \sigma^2 V)$ (see Schall and Dunne, 1985 and 1986).

REFERENCES

John, J.A. and Draper, N.R. (1978). On Testing for Two Outliers or One Outlier in Two-Way Tables, Technometrics, 20, 69-78.

Kendall, M.G. and Stuart, A. (1973). The Advanced Theory of Statistics, Vol. 2, Griffin, 3rd Edition.