The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 35–45

REJECTING OUTLIERS BY MAXIMUM NORMED RESIDUAL[1]

BY WILHELMINE STEFANSKY

University of California, Berkeley

1. Introduction. The maximum normed residual (MNR) has been proposed as a test statistic in connection with the problem of rejecting outliers. An outlying observation is one that does not fit in with the pattern of the remaining observations. Outliers may be mistakes, or else accurate but unexpected observations which could shed new light on the phenomenon under study. On the other hand, it is possible that an outlier is simply a manifestation of the inherent variability of the data. It is of interest, therefore, to test whether or not a given outlier comes from a population different from the one hypothesized.

A thorough discussion of the special case of detecting outliers in a single sample from a normal population is given by F. E. Grubbs in [5]. Several tests are discussed, among them tests based on the MNR, and tables of critical values are included. For unreplicated factorial designs C. Daniel [3] proposed a statistic equivalent to the MNR. T. S. Ferguson [4] has proved that the maximum Studentized residual possesses the optimum property of being admissible within all invariant procedures. For designs with residuals having a common variance, the MNR is equivalent to the maximum Studentized residual, hence also has this optimum property. Nevertheless, except for the case of a single sample from a normal population, critical values of the MNR are not available.

In 1936 E. S. Pearson and C. Chandra Sekar [7] noticed that for the case of a single sample from a normal distribution, critical values for statistics equivalent to the MNR can be calculated quite easily from tables of the t-distribution, provided the level α is not too large. In order to determine if α is not too large, it is necessary to know M_2, the largest value which the second largest among the absolute values of the normed residuals can take on.

In Section 4 the results of E. S. Pearson and C. Chandra Sekar [7] are extended to designs with the property that the residuals have a common variance. Applications are given in Section 5.

The main problem in extending the results of [7] is the calculation of M_2. In Section 3 a more general problem is dealt with, namely the calculation of M_k for arbitrary designs, where M_k denotes the largest value which the kth largest among the absolute values of the residuals can have. A method for calculating M_k is developed and this method yields an explicit expression for M_2 for designs having residuals with a common variance.

Designs with the property that the residuals have a common variance include all ordinary factorial designs, where the different levels of each factor are replicated

Received April 30, 1970.

[1] This paper is based on the author's doctoral dissertation and was prepared with the partial support of a National Science Foundation Grant GP 15283.

equally often, Latin squares and balanced incomplete blocks. An example of a design which does not have this property is a two-way layout with an unequal number of observations per cell.

2. Definition of the MNR **and equivalent statistics.** Let $\mathbf{y} = (y_1 \cdots y_n)'$ be an $n \times 1$ random vector with expected value $E\mathbf{y} = (Ey_1 \cdots Ey_n)'$ and covariance matrix $\sigma^2 I$, where I is the $n \times n$ identity matrix and σ^2 is a scalar, usually unknown. An observed value of $\mathbf{y}$ will also be denoted by $\mathbf{y}$. It will be assumed that $E\mathbf{y}$ is of the form $X'\boldsymbol{\beta}$ where X' is an $n \times q$ matrix of rank $p \leqq q$ and where $\boldsymbol{\beta}$ is a $q \times 1$ column vector. The vector $\hat{\mathbf{y}} = X'\hat{\boldsymbol{\beta}}$ is the vector of fitted values, where $\hat{\boldsymbol{\beta}}$ is a least squares estimate (LSE) of $\boldsymbol{\beta}$. Since $\hat{\mathbf{y}}$ is the projection of $\mathbf{y}$ onto the space spanned by the columns of X', $\hat{\mathbf{y}}$ is unique, even though $\hat{\boldsymbol{\beta}}$ is unique if and only if X' is of full rank. The difference between the observed and fitted values of the ith coordinate of $\mathbf{y}$ is called the ith residual. The $n \times 1$ vector of residuals is denoted by $\mathbf{e}$, so that

$$\mathbf{e} = (e_1 \cdots e_n)' = \mathbf{y} - \hat{\mathbf{y}}.$$

The quantity $z_i = e_i/\|\mathbf{e}\|$ is the ith normed residual, where $\|\mathbf{e}\|^2 = \sum_{i=1}^n e_i^2$ as usual, and the MNR $|z|^{(1)}$ is defined to be the largest among the absolute values of the normed residuals.

For detecting outliers in factorial designs C. Daniel [3] proposed the statistic

$$d_1 = \frac{|z|^{(1)}(n-p-1)^{\frac{1}{2}}}{[1-n(|z|^{(1)})^2/(n-p)]^{\frac{1}{2}}}$$

which is a strictly increasing function of $|z|^{(1)}$, hence a statistic equivalent to $|z|^{(1)}$.

A statistic closely related to d_1 is $F^{(1)}$, defined as follows. Let

$$F_i = \frac{(n-p-1)e_i^2/p_{ii}}{\|\mathbf{e}\|^2 - e_i^2/p_{ii}} = \frac{(n-p-1)z_i^2/p_{ii}}{1-z_i^2/p_{ii}},$$

where $p_{ii} = (1/\sigma^2)\operatorname{Var}(e_i)$. Then $F^{(1)} = \max(F_1, F_2, \cdots, F_n)$. If the residuals all have the same variance, $p_{ii} = (n-p)/n$, because $E[\sum_{i=1}^n e_i^2] = (n-p)\sigma^2$, as is well known. In this case, $F_i = n(n-p-1)z_i^2/(n-p-nz_i^2)$ so that $F^{(1)} = nd_1^2/(n-p)$.

T. S. Ferguson [4] considers the random variable (rv) $\max\{(n-p)^{\frac{1}{2}}|z_i|/p_{ii}^{\frac{1}{2}}(i = 1, \cdots, n)\}$ and refers to it as the maximum Studentized residual. If the residuals all have the same variance, then the maximum Studentized residual is equal to $n^{\frac{1}{2}}|z|^{(1)}$.

Lieblein [6] discusses the statistic L defined as follows. Let $\mathbf{y} = (y_1 y_2 y_3)'$ be a sample of size three from a distribution with mean μ and variance σ^2. Denote the rth largest among y_1, y_2, y_3 by $y^{(r)}(r = 1, 2, 3)$, so that $y^{(3)} \leqq y^{(2)} \leqq y^{(1)}$. Let y' and y'' be the closest two observations, with $y' \geqq y''$, and let y''' be the remaining observation. Then $L = (y'-y'')/(y^{(1)}-y^{(3)})$. The statistics L and $F^{(1)}$ are equivalent, because they are related by the identities

$$L = \frac{2}{(3F^{(1)})^{\frac{1}{2}}+1} \qquad F^{(1)} = \frac{(2-L)^2}{3L^2}.$$

Samples of size three are of interest, because in chemical laboratories determinations are often made in triplicate. In [6] Lieblein investigates the effects of the practice of keeping only the two closest observations y' and y'' and discarding the remaining observation y''' as less accurate. Rejecting the outlier y''' when L is too small is equivalent to rejecting y''' when the MNR is too large.

3. Calculation of M_k. In this section a general method is given for computing M_k, the largest value which the kth largest among $|z_1|, \cdots, |z_n|$ can take on. First some well-known results will be reviewed which are needed in the proof of Theorem 3.1. Let V denote the space spanned by the columns of X' and denote by $V^\perp$ the orthocomplement of V in R^n, where R^n is the space of $n \times 1$ column vectors with inner product $(\mathbf{v}, \mathbf{w}) = \mathbf{v}'\mathbf{w} = \sum_{i=1}^n v_i w_i$ and norm $||\mathbf{v}|| = (\mathbf{v}, \mathbf{v})^{\frac{1}{2}}$. The dimension of $V^\perp$ is $n-p$, because the rank of X' is p. Since $\hat{\mathbf{y}}$ is the projection of $\mathbf{y}$ onto V, the vector $\mathbf{e}$ of residuals is the projection of $\mathbf{y}$ onto $V^\perp$, i.e., $\mathbf{e} = P\mathbf{y}$, where P denotes the linear operator which projects onto $V^\perp$. Since P is a projection operator

(i) $PP = P$, i.e., P is idempotent.
(ii) $(\mathbf{v}, P\mathbf{w}) = (P\mathbf{v}, \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in R^n$, i.e., P is symmetric.

For $\mathbf{v} \in R^n$ let $|v|^{(k)}$ denote the kth largest among $|v_1|, \cdots, |v_n|$ $(k = 1, \cdots, n)$. Since the vector $\mathbf{z} = (e_1/||\mathbf{e}|| \cdots e_n/||\mathbf{e}||)'$ has norm one and belongs to $V^\perp$, the problem of calculating M_k has the following formulation. Subject to the conditions (i) $\mathbf{v} \in V^\perp$ and (ii) $||\mathbf{v}|| = 1$, calculate the maximum value which $|v|^{(k)}$ can have $(k = 1, \cdots, n)$.

Several definitions are needed. A *sign matrix* is a matrix whose off diagonal elements are all equal to zero and whose diagonal elements are all equal to plus or minus one. Sign matrices are of interest for this problem, because the vector of absolute values of the normed residuals $(|z_1| \cdots |z_n|)$ is equal to $S\mathbf{z}$ for some sign matrix S. For a given sign matrix S, denote SPS by Q. Denote the jth column of Q by Q_j. If $\mathbf{u}_j$ is the $n \times 1$ column vector with a one in the jth position and zeros elsewhere, then $Q_j = Q\mathbf{u}_j$. Observe that (i) $QQ = Q$ and (ii) $(\mathbf{v}, Q\mathbf{w}) = (Q\mathbf{v}, \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in R^n$.

Let $J = \{j_1, \cdots, j_k\}$ be a subset of $\{1, \cdots, n\}$ and let $\Xi(J) = \{\xi \in R^n: \xi_i \geqq 0\ (i = 1, \cdots, n);\ \xi_i = 0\ (i \notin J);\ \sum_{i=1}^n \xi_i = 1\}$. Let $H(J, S) = \{Q\xi: \xi \in \Xi(J)\}$, i.e., let $H(J, S)$ be the convex hull of $Q_j (j \in J)$.

The set $H(J, S)$ is easy to visualize when J contains two elements, say i and j. If Q_i and Q_j are linearly dependent, $H(J, S)$ is the line segment which joins them and is part of a line through the origin. $H(J, S)$ may or may not contain the origin. If Q_i and Q_j are linearly independent, $H(J, S)$ is again the line segment which joins them, but neither contains the origin nor is part of a line through the origin as shown in Figure 3.1.

If J contains three elements, say i, j and l, and exactly two of Q_i, Q_j and Q_l are linearly independent, then $H(J, S)$ is a triangle which is part of a plane through the origin and which may or may not contain the origin. The second case is shown in

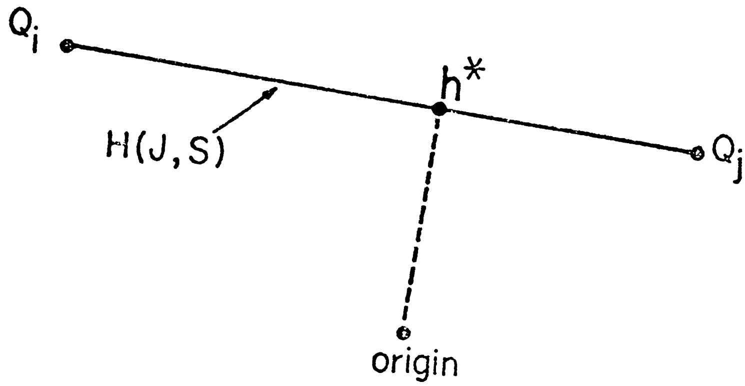

FIG. 3.1.

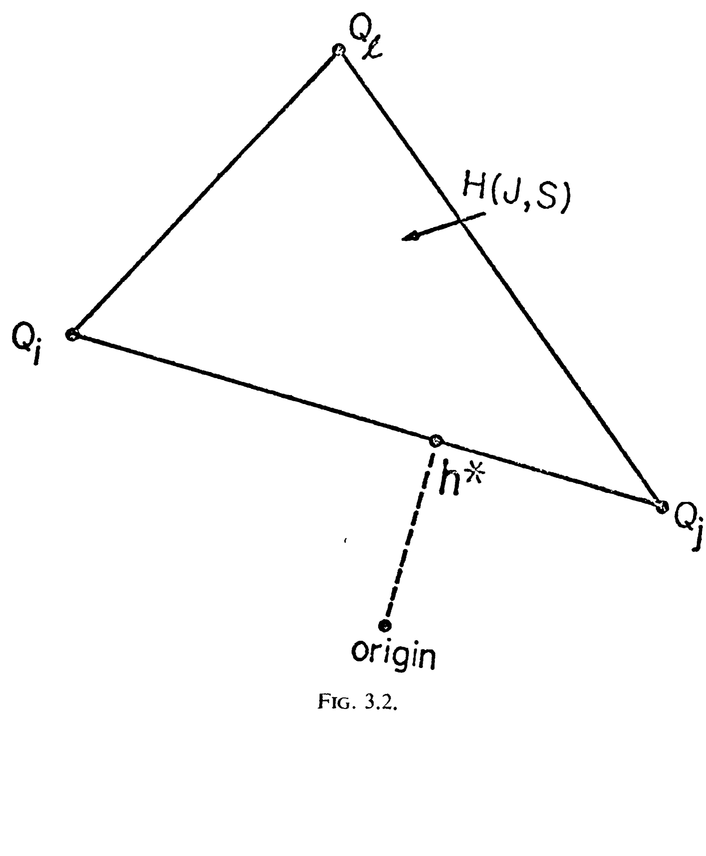

FIG. 3.2.

Figure 3.2. If all three of Q_i, Q_j and Q_l are linearly independent, $H(J, S)$ is a triangle which neither contains the origin nor is part of a plane through the origin.

The set $\Xi(J)$ is a closed, bounded convex set and consequently $H(J, S)$ is closed, bounded and convex also. Let $N(J, S) = \inf ||\mathbf{h}||$, where the infimum is taken over all $\mathbf{h} \in H(J, S)$. Since $H(J, S)$ is closed and bounded, there exists an $\mathbf{h}^* \in H(J, S)$ such that $||\mathbf{h}^*|| = N(J, S)$. If $H(J, S)$ contains the origin, $N(J, S) = 0$ and $\mathbf{h}^*$ is the origin. In Figures 3.1 and 3.2, $\mathbf{h}^*$ is the point in the intersection of L and the line which passes through the origin and is perpendicular to L, where L denotes the line segment joining Q_i and Q_j.

Observe that in Figures 3.1 and 3.2 the algebraic length of the projection of Q_i on $\mathbf{h}^*/||\mathbf{h}^*||$ is equal to the length of $\mathbf{h}^*$, i.e., $(Q_i, \mathbf{h}^*/||\mathbf{h}^*||) = ||\mathbf{h}^*||$, so that $h_i^* = (\mathbf{u}_i, \mathbf{h}^*) = (\mathbf{u}_i, Q\mathbf{h}^*) = (Q\mathbf{u}_i, \mathbf{h}^*) = (Q_i, \mathbf{h}^*) = ||\mathbf{h}^*||^2$. Similarly $h_j^* = ||\mathbf{h}^*||^2$. Since the algebraic length of the projection of Q_l on $\mathbf{h}^*/||\mathbf{h}^*||$ is larger than $||\mathbf{h}^*||$, $h_l^* > ||\mathbf{h}^*||^2$.

LEMMA 3.1. *For* $\mathbf{h}^* \in H(J, S)$, $||\mathbf{h}^*|| = N(J, S)$ *if and only if* $h_j^* \geqq ||\mathbf{h}^*||^2$ *for all* $j \in J$.

PROOF. Suppose that $h_j^* \geqq ||\mathbf{h}^*||^2$ for all $j \in J$. If $||\mathbf{h}^*|| = 0$, then $||\mathbf{h}^*|| = 0 \leqq ||\mathbf{h}||$ for all $\mathbf{h} \in H(J, S)$, so that $||\mathbf{h}^*|| = N(J, S)$. Therefore, suppose that $||\mathbf{h}^*|| > 0$. Let $\mathbf{h}$ be an arbitrary element of $H(J, S)$ and let $\boldsymbol{\eta} \in \Xi(J)$ be such that $\mathbf{h} = Q\boldsymbol{\eta}$. Let $\boldsymbol{\xi} \in \Xi(J)$ be such that $\mathbf{h}^* = Q\boldsymbol{\xi}$. By the Cauchy–Schwarz inequality, $(\mathbf{h}, \mathbf{h}^*) \leqq |(\mathbf{h}, \mathbf{h}^*)| \leqq ||\mathbf{h}||\,||\mathbf{h}^*||$. If $||\mathbf{h}^*||^2 \leqq h_j^*$ for all $j \in J$, then

$$||\mathbf{h}^*||^2 \leqq \textstyle\sum_{i=1}^n h_i^* \eta_i = (\boldsymbol{\eta}, \mathbf{h}^*) = (\boldsymbol{\eta}, Q\mathbf{h}^*) = (Q\boldsymbol{\eta}, \mathbf{h}^*) = (\mathbf{h}, \mathbf{h}^*).$$

Hence $||\mathbf{h}^*||^2 \leqq ||\mathbf{h}||\,||\mathbf{h}^*||$ for all $\mathbf{h} \in H(J, S)$. Since by assumption $||\mathbf{h}^*|| > 0$, it follows that $||\mathbf{h}^*|| \leqq ||\mathbf{h}||$ for all $\mathbf{h} \in H(J, S)$, so that $||\mathbf{h}^*|| = N(J, \mathrm{S})$.

Next the converse is proved. Suppose that $\mathbf{h}^* \in H(J, S)$ is such that $||\mathbf{h}^*|| \leqq ||\mathbf{h}||$ for all $\mathbf{h} \in H(J, S)$. If there is a $j_0 \in J$ such that $h_{j_0}^* < ||\mathbf{h}^*||^2$, choose $0 < \varepsilon < \min\{1, (||\mathbf{h}^*||^2 - h_{j_0}^*)/||\mathbf{h}^* - Q_{j_0}||^2\}$. Then $\mathbf{g} = (1-\varepsilon)\mathbf{h}^* + \varepsilon Q_{j_0}$ belongs to $H(J, S)$ and

$$\begin{aligned} ||\mathbf{g}||^2 &= ||\mathbf{h}^*||^2(1-\varepsilon)^2 + \varepsilon^2 ||Q_{j_0}||^2 + 2\varepsilon(1-\varepsilon)h_{j_0}^* \\ &= ||\mathbf{h}^*||^2 + \varepsilon^2 ||\mathbf{h}^* - Q_{j_0}||^2 + 2\varepsilon(h_{j_0}^* - ||\mathbf{h}^*||^2) \\ &< ||\mathbf{h}^*||^2 + \varepsilon(h_{j_0}^* - ||\mathbf{h}^*||^2) < ||\mathbf{h}^*||^2, \end{aligned}$$

which contradicts the assumption that $||\mathbf{h}^*|| \leqq ||\mathbf{h}||$ for all $\mathbf{h} \in H(J, S)$. Therefore $h_j^* \geqq ||\mathbf{h}^*||^2$ for all $j \in J$. □

If S is a sign matrix whose diagonal elements s_i for $i \notin J$ are equal to one and whose diagonal elements s_j for $j \in J$ are equal to plus or minus one, denote by $-S$ the sign matrix obtained from S by changing the sign of the diagonal elements s_j of S for all $j \in J$. Then obviously $N(J, S) = N(J, -S)$. Let $N(J) = \max N(J, S)$, where the maximum is taken over all sign matrices whose diagonal elements s_j are equal to plus or minus one for $j \in J$. If J has k elements, there are 2^k such matrices. However, since $N(J, S) = N(J, -S)$, at most 2^{k-1} of the numbers $N(J, S)$ need to be calculated in order to compute $N(J)$. Let $N_k = \max N(J)$, where the maximum is taken over all $\binom{n}{k}$ possible subsets J of $\{1, \cdots, n\}$ which have k elements.

THEOREM 3.1. *Subject to the conditions that*

$$\text{(i)} \qquad \mathbf{v} \in V^{\perp}$$

$$\text{(ii)} \qquad \|\mathbf{v}\| = 1,$$

the maximum value which $|v|^{(k)}$ *can have is equal to* N_k $(k = 1, \cdots, n)$.

PROOF. Suppose that $|v_{j_1}|, \cdots, |v_{j_k}|$ are the k largest among $|v_1|, \cdots, |v_n|$ and denote the set $\{j_1, \cdots, j_k\}$ by J. Let S be the sign matrix whose diagonal elements s_i are equal to one for $i \notin J$ and whose diagonal elements s_j for $j \in J$ are such that $s_j v_j = |v_j|$. Obviously $|v|^{(k)}$ is equal to or less than any weighted average of the $|v_{j_1}|, \cdots, |v_{j_k}|$. Formally, for any $\xi \in \Xi(J)$, $|v|^{(k)} \leqq (\xi, S\mathbf{v}) = (\xi, SP\mathbf{v}) = (SPS\xi, S\mathbf{v}) = (Q\xi, S\mathbf{v}) \leqq \|Q\xi\| \, \|S\mathbf{v}\| = \|Q\xi\| \, \|\mathbf{v}\| = \|Q\xi\|$. Hence $|v|^{(k)} \leqq \inf_{\xi \in \Xi(J)} \|Q\xi\| = \inf_{h \in H(J,S)} \|\mathbf{h}\| = N(J, S) \leqq N_k$. Since $\mathbf{v}$ was arbitrary, it follows that the maximum value which $|v|^{(k)}$ can have subject to conditions (i) and (ii) is equal to or less than N_k.

Next, let J_0 and S_0 be such that $N_k = N(J_0, S_0)$. Denote $S_0 P S_0$ by Q. There is an element $\mathbf{h}^*$ of $H(J_0, S_0)$ for which $\|\mathbf{h}^*\| = N(J_0, S_0) = N_k$, and there is a ξ in $\Xi(J_0)$ such that $\mathbf{h}^* = Q\xi$. The vector $\mathbf{v}^* = S_0 \mathbf{h}^*/\|\mathbf{h}^*\| = PS_0 \xi/\|Q\xi\|$ belongs to $V^{\perp}$ and has norm one. By Lemma 3.1, $\|\mathbf{h}^*\| = N(J_0, S_0)$ if and only if $h_j^* \geqq \|\mathbf{h}^*\|^2$ for all $j \in J_0$. Denote the jth diagonal element of the sign matrix S_0 by s_j. It follows that $s_j v_j^* = h_j^* \geqq \|\mathbf{h}^*\|^2$ for all $j \in J_0$. Since $N_k \geqq 0$, it must be the case that $s_j v_j^* = |v_j^*|$ for all $j \in J$. Since at least k of $|v_1^*|, \cdots, |v_n^*|$ are equal to or larger than N_k, $|v^*|^{(k)} \geqq N_k$. Hence the maximum value which $|v|^{(k)}$ can have subject to conditions (i) and (ii) is equal to or larger than N_k. □

If a method is found for calculating each value $N(J, S)$, the problem of calculating $M_k (k = 1, \cdots, n)$ is completely solved in view of Theorem 3.1 and the second paragraph of this section. The next theorem provides a method for calculating $N(J, S)$.

For a given subset K of $\{1, \cdots, n\}$, the matrix $Q^*(K)$ is the matrix obtained from Q by deleting the ith row and ith column from Q for all $i \notin K$.

THEOREM 3.2. *Let* $\mathbf{h}^* \in H(J, S)$ *be such that* $\|\mathbf{h}^*\| = N(J, S)$. *If* $N(J, S) > 0$, *then there is a subset* K *of* J *such that*

(i) $\mathbf{h}^* = Q\xi^*$ *for some* $\xi^* \in \Xi(K)$ *with* $\xi_i^* > 0$ *for all* $i \in K$.
(ii) $h_i^* = \|\mathbf{h}^*\|^2$ *for all* $i \in K$.
(iii) $Q^*(K)$ *is nonsingular.*

PROOF. Let $\Xi^* = \{\xi : \xi \in \Xi(J); \mathbf{h}^* = Q\xi\}$. Then Ξ^* is a closed and bounded convex set. Let ξ^* be an extreme point of Ξ^* and let K be the subset of J such that $\xi_i^* > 0$ if and only if $i \in K$. In view of Lemma 3.1, it must be the case that $h_i^* = \|\mathbf{h}^*\|^2$ for all $i \in K$, so that (i) and (ii) hold for this set K. In order to show that in addition (iii) holds, it is of course sufficient to show that the columns of $Q^*(K)$ are linearly independent. Suppose the columns of $Q^*(K)$ are linearly dependent. Then there is a $\mathbf{t} \in R^n$ with $t_i = 0$ for $i \notin K$, with $t_i \neq 0$ for at least one

$i \in K$ and such that the ith coordinate of $Q\mathbf{t}$ is zero for all $i \in K$. Then $0 = \mathbf{t}'Q'\xi^* = \mathbf{t}'Q\xi^* = \mathbf{t}'\mathbf{h}^* = ||\mathbf{h}^*||^2 \sum_{i \in K} t_i$. Since $||\mathbf{h}^*||^2 > 0$ by assumption, $\sum_{i \in K} t_i = 0$. It follows easily that $\xi_\varepsilon = \xi^* + \varepsilon\mathbf{t}$ belongs to Ξ^* for $|\varepsilon|$ sufficiently small. But then $\xi^* = \frac{1}{2}(\xi_\varepsilon + \xi_{-\varepsilon})$, contradicting the fact that ξ^* is an extreme point of Ξ^*. □

Suppose $K = \{i_1, \cdots, i_m\}$ with $i_1 < \cdots < i_m$. For the matrix $Q^*(K)$ in part (iii) of Theorem 3.2, denote the sum of the elements in the ith row of $[Q^*(K)]^{-1}$ by T_i and let $T = \sum_{i=1}^m T_i$. Since P is a positive semidefinite matrix, so are Q and $Q^*(K)$. Since $Q^*(K)$ is nonsingular, $Q^*(K)$ is in fact positive definite, and hence its inverse is positive definite also. It follows that $T > 0$.

COROLLARY 3.1. *The vector* ξ^* *in part* (i) *of Theorem* 3.2 *is unique and is given by*

$$\xi_{i_r}^* = T_r/T \qquad (r = 1, \cdots, m)$$

$$\xi_i^* = 0 \qquad (i \notin K).$$

PROOF. Let $\boldsymbol{\delta}$ be the $m \times 1$ column vector consisting entirely of ones. Denote $Q^*(K)$ simply by Q^*. Let ξ be the $m \times 1$ vector obtained from ξ^* by deleting all the zero coordinates of ξ^*. Then condition (ii) of Theorem 3.2 can be written as $Q^*\xi = ||\mathbf{h}^*||^2 \boldsymbol{\delta}$. Since Q^* is nonsingular, $\xi = ||\mathbf{h}^*||^2[Q^*]^{-1}\boldsymbol{\delta}$. Since $1 = \boldsymbol{\delta}'\xi = ||\mathbf{h}^*||^2 \boldsymbol{\delta}'[Q^*]^{-1}\boldsymbol{\delta} = ||\mathbf{h}^*||^2 T$, $||\mathbf{h}^*||^2 = 1/T$, so that $\xi_r = T_r/T$ $(r = 1, \cdots, m)$. □

Theorem 3.2 provides a method for calculating $N(J, S)$. For all subsets K of J examine $Q^*(K)$. If $Q^*(K)$ is nonsingular, calculate ξ^* according to Corollary 3.1. If $\xi_{i_r}^* > 0$ $(r = 1, \cdots, m)$, calculate $\mathbf{h}^* = Q\xi^*$. If $h_j^* \geqq ||\mathbf{h}^*||^2$ for all $j \in J$, then $N(J, S) = ||\mathbf{h}^*||$ by Lemma 3.1 and from the proof of Corollary 3.1 it follows easily that $||\mathbf{h}^*|| = (1/T)^{\frac{1}{2}}$, i.e., that $N(J, S)$ is equal to the square root of the reciprocal of the sum of the entries of $[Q^*(K)]^{-1}$. If it is not possible to find a subset K of J such that (a) $Q^*(K)$ is nonsingular, (b) the vector ξ^* calculated from $Q^*(K)$ according to Corollary 3.1 belongs to $\Xi(K)$ with $\xi_i^* > 0$ for all $i \in K$ and (c) $h_j^* \geqq ||\mathbf{h}^*||^2$ for all $j \in J$, where $\mathbf{h}^* = Q\xi^*$, then $N(J, S) = 0$ in view of Theorem 3.2, Lemma 3.1 and the fact that $N(J, S) \geqq 0$.

Let us illustrate the method for calculating $M_k = N_k$ by calculating M_2 for a design with the property that the residuals have a common variance. In this case the diagonal elements of P are all equal to $(n-p)/n$. Recall that ρ_{ij} denotes the correlation coefficient between the residuals e_i and e_j and that $R = \max |\rho_{ij}|$, where the maximum is taken over all pairs (i, j) with $i \neq j$ $(i, j = 1, \cdots, n)$. We shall show that $M_2 = N_2 = [(n-p)(1+R)/2n]^{\frac{1}{2}}$.

First let us find an explicit expression for $N(J, S)$, where S is an arbitrary sign matrix, and where J is an arbitrary subset of $\{1, \cdots, n\}$ of size 2. Let s_r denote the rth diagonal element of S as usual. Suppose $J = \{i, j\}$ and denote $s_i s_j$ by s. Then $Q^*(J)$ is equal to

$$\begin{pmatrix} (n-p)/n & s\rho_{ij}(n-p)/n \\ s\rho_{ij}(n-p)/n & (n-p)/n \end{pmatrix}.$$

Two cases can arise.

Case (i). $|\rho_{ij}| < 1$. In this case $Q^*(J)$ is nonsingular and

$$[Q^*(J)]^{-1} = \frac{n}{(n-p)(1-\rho_{ij}^2)}\begin{pmatrix} 1 & -s\rho_{ij} \\ -s\rho_{ij} & 1 \end{pmatrix}.$$

For the subset $K = \{i, j\} = J$, the vector ξ^* calculated according to Corollary 3.1 is given by $\xi_i^* = \frac{1}{2} = \xi_j^*$, $\xi_k^* = 0$ for $k \neq i, j$. Further, for $\mathbf{h}^* = Q\xi^*$, $h_i^* = (n-p)(1+s\rho_{ij})/2n = h_j^*$. Since $h_i^* = h_j^* = \|\mathbf{h}^*\|^2$, $N(J, S) = \|\mathbf{h}^*\|$ by Lemma 3.1 and consequently $N(J, S) = [(n-p)(1+s\rho_{ij})/2n]^{\frac{1}{2}}$.

Case (ii). $|\rho_{ij}| = 1$. In this case $Q^*(J)$ is singular. Let $K = \{i\} \subset J$. Then $[Q^*(K)]^{-1} = n/(n-p)$ and the vector ξ^* calculated according to Corollary 3.1 is given by $\xi_i^* = 1$, $\xi_k^* = 0$ for $k \neq i$. For $\mathbf{h}^* = Q\xi^*$, $h_i^* = (n-p)/n = \|\mathbf{h}^*\|^2$ and $h_j^* = s\rho_{ij}(n-p)/n$. If $s\rho_{ij} = 1$, $h_i^* = h_j^* = \|\mathbf{h}^*\|^2$ and $N(J, S) = \|\mathbf{h}^*\| = [(n-p)/n]^{\frac{1}{2}}$. However, if $s\rho_{ij} = -1$, $h_j^* < \|\mathbf{h}^*\|^2$. There is one other subset K of J for which $Q^*(K)$ is nonsingular, namely $K = \{j\}$. For this subset, the vector ξ^* is given by $\xi_j^* = 1$, $\xi_r^* = 0$ for $r \neq j$. For $\mathbf{h}^* = Q\xi^*$, $h_j^* = (n-p)/n = \|\mathbf{h}^*\|^2$ but $h_i^* < \|\mathbf{h}\|^2$. Since it is not possible to find a subset K of J for which (a) $Q^*(K)$ is nonsingular, (b) $\xi^* \in \Xi(K)$ with $\xi_k^* > 0$ for $k \in K$ and (c) $h_j^* \geqq \|\mathbf{h}^*\|^2$ for all $j \in J$, where $\mathbf{h}^* = Q\xi^*$, it must be the case that $N(J, S) = 0$ if $s\rho_{ij} = -1$. Note that whether $s\rho_{ij} = 1$ or $s\rho_{ij} = -1$, $N(J, S) = [(n-p)(1+s\rho_{ij})/2n]^{\frac{1}{2}}$, which is the same expression for $N(J, S)$ as the one found in Case (i).

The next step is to calculate $N(J) = \max N(J, S)$, where the maximum is taken over all possible sign matrices S for which the diagonal elements s_r are equal to one for all $r \notin J$. There are four such sign matrices S, but they give rise to only two different values $N(J, S)$. Either $s = s_i s_j = 1$ or $s = -1$, so that $N(J) = [(n-p)(1+|\rho_{ij}|)/2n]^{\frac{1}{2}}$.

Finally, $N_2 = \max N(J)$, where the maximum is taken over all possible subsets J of size 2. It is easily seen that $M_2 = N_2 = [(n-p)(1+R)/2n]^{\frac{1}{2}}$.

4. A method for calculating critical values of the MNR **in linear models with residuals having equal variances, provided α is sufficiently small.** The main result of this section is Theorem 4.1 in which the conditions are given under which the critical value of the MNR can be calculated from tables of the t-distribution. The $(1-\alpha)$ 100th percentiles of the distributions needed in this section will be denoted as follows.

F_α—F-distribution with 1 and $n-p-1$ df.

t_α—t-distribution with $n-p-1$ df.

$F_\alpha^{(1)}$—distribution of $F^{(1)}$.

D_α—distribution of the MNR.

LEMMA 4.1. *Let* $\mathbf{x} = (x_1 \cdots x_n)'$ *be a random vector whose coordinates have a common marginal distribution G. Let* $x^{(r)}$ *denote the* rth *largest coordinate* ($r = 1, 2$). *Then provided* $k >$ ess sup $x^{(2)}$, $P[x^{(1)} > k] = nP[X > k]$, *where X is a* rv *which has distribution G.*

PROOF. Let $A_i = [x_i > k]$. Then $[x^{(1)} > k] = \bigcup_{i=1}^{n} A_i$ and $P(A_i) = P[X > k]$, where X has distribution G. If $k > \text{ess sup } x^{(2)}$, then the events A_i are mutually exclusive, so that $P[x^{(1)} > k] = P[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} P(A_i) = nP[X > k]$. □

COROLLARY 4.1. *Let X' be an $n \times q$ matrix of rank $p \leqq q$. If* $\mathbf{y}$ *has the normal distribution with mean $X'\boldsymbol{\beta}$ and covariance matrix $\sigma^2 I$, then $F_\alpha^{(1)} = t^2_{\alpha/2n}$, provided $t^2_{\alpha/2n} >$* ess sup *$F^{(2)}$, where $F^{(2)}$ is the second largest among $F_1, \cdots, F_n$.*

PROOF. If $\mathbf{y}$ has the normal distribution with mean $X'\boldsymbol{\beta}$ and covariance matrix $\sigma^2 I$, then the $F_i (i = 1, \cdots, n)$ have the F-distribution with 1 and $n-p-1$ df as common marginal distribution. The corollary now follows easily from Lemma 4.1 and the fact that $nP[F > k] = 2nP[t > k^{\frac{1}{2}}]$, where F (t) is a rv having the F-distribution (t-distribution) with 1 and $n-p-1$ $(n-p-1)$ df. □

If the residuals have a common variance, the rv's $|z_i|$ and F_i are related by the equation $|z_i| = \{(n-p)F_i/[n(n-p-1+F_i)]\}^{\frac{1}{2}} = g(F_i)$. Recall that ρ_{ij} denotes the correlation coefficient between the residuals e_i and e_j and that $R = \max |\rho_{ij}|$, where the maximum is taken over all pairs (i, j) with $i \neq j$ $(i, j = 1, \cdots, n)$. In Section 3 it is shown that if the residuals have a common variance, $M_2 = [(n-p)(1+R)/2n]^{\frac{1}{2}}$. Since g is a strictly increasing function, Theorem 4.1 follows easily.

THEOREM 4.1. *Let X' be an $n \times q$ matrix of rank $p \leqq q$ and suppose that* $\mathbf{y}$ *has the normal distribution with mean $X'\boldsymbol{\beta}$ and covariance matrix $\sigma^2 I$. If the residuals have a common variance, $D_\alpha = g(t^2_{\alpha/2n})$ provided $g(t^2_{\alpha/2n}) > [(n-p)(1+R)/2n]^{\frac{1}{2}}$.*

5. Applications. The simplest example of a design with the property that all residuals have the same variance is a sample of size three. Borenius [2] noticed that in this case $|z|^{(1)}$ is always $\geqq M_2$. Therefore, if the sample is from a normal distribution, critical values of statistics equivalent to the MNR can be calculated from tables of the t-distribution for *all* levels α. This applies in particular to the statistic L defined in Section 2.

In [1] F. J. Anscombe provides a list of designs with residuals having a common variance. Where possible, the method of the previous section was used to calculate critical values of the MNR for these designs. It was assumed of course that the vector of observations $\mathbf{y}$ has a normal distribution. The results are summarized in Table 5.1.

The Plackett–Burman fraction is given in [8]. In the balanced incomplete blocks designs v is the number of varieties or treatment levels and k is the number of units in each block. It is supposed that interblock information is not recovered. Most of the other designs are factorial designs. Thus (6) is half of a factorial design with five factors each at two levels and $ABCDE$ is the generator of the alias subgroup. No interactions are estimated.

Acknowledgment. I wish to express my sincere appreciation to Professor E. L. Lehmann for his guidance and advice throughout the course of this research.

TABLE 5.1

Critical values of the maximum normed residual

Design	n	$n-p$	R	M_2	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$	$\alpha = .20$
(1) 2^3	8	4	1/2	.612	.700	.686	.673	.653
(2) 3^2	9	4	1/2	.577	.660	.648	.637	.620
(3) Plackett–Burman fraction of 2^5	12	6	2/3	.645	.675			
(4) Balanced incomplete blocks ($v = 4$, $k = 3$)	12	5	1/2	.559	.630	.611	.597	.576
(5) 2^4	16	11	3/11	.661	.697			
(6) $2^5/2$: *ABCDE*	16	10	1/5	.612	.682	.632		
(7) $2^5/2$: *ABCD*	16	10	2/5	.661	.682			
(8) $2^6/4$: *ABCD, CDEF*	16	9	5/9	.661	.665			
(9) $2^6/4$: *ABC, DEF*	16	9	1/3	.612	.665	.621		
(10) $2^7/8$: *ABC, CDE, EFG*	16	8	1/2	.612	.644			
(11) $2^8/16$: *ABC, CDE, EFG, AGH*	16	7	3/7	.559	.619	.588	.568	
(12) $2^9/32$: *ABC, CDE, EFG, AGH, BFI*	16	6	1/3	.500	.587	.565	.549	.528
(13) $2^{10}/64$: *ABE, ACF, ADG, BCH, BDI, CDJ*	16	5	3/5	.500	.548	.533	.523	.507
(14) Latin square	16	6	1/3	.500	.587	.565	.549	.528
(15) Balanced incomplete blocks ($v = 7$, $k = 3$)	21	8	1/2	.535	.567	.536		
(16) Latin square	25	12	1/4	.548	.577			
(17) Graeco–Latin square	25	8	3/8	.469	.522	.495	.479	
(18) $3^4/3$: *ABCD*	27	18	1/6	.624	.597			
(19) $3^9/729$: *ABD*, AB^2E, *ACF*, AC^2G, *BCH*, BC^2I	27	8	1/2	.471	.503	.478		

REFERENCES

[1] ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics* **2** 123--147.

[2] BORENIUS, G. (1958). On the distribution of the extreme values in a sample from a normal distribution. *Skand. Aktuarietidskr.* **41** 131–166.

[3] DANIEL, C. (1960). Locating outliers in factorial experiments. *Technometrics* **2** 149–156.

[4] FERGUSON, THOMAS S. (1961). On the rejection of outliers. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 253–287. Univ. of California.

[5] GRUBBS, FRANK E. (1969). Procedures for detecting outlying observations in samples. *Technometrics* **11** 1–21.

[6] LIEBLEIN, JULIUS (1952). Properties of certain statistics involving the closest pair in a sample of three observations. *J. Res. Nat. Bur. Standards* **48** 255–268.

[7] PEARSON, E. S. and SEKAR, C. CHANDRA (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika* **28** 308–320.

[8] PLACKETT, R. L. and BURMAN, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika* **33** 305–325.