ROBUST TEST OF LINEAR HYPOTHESIS FOR THE LINEAR MODEL

*by Ann Inez N. Gironella**
*and George A. Milliken***

INTRODUCTION

Consider the general linear model

$$y = X\beta + \epsilon \tag{1.1}$$

where y is an n x 1 vector of observations, X is an n x p matrix of known fixed quantities of rank $p(p \leqslant n)$, β is a p x 1 vector of unknown parameters, and ϵ is an n x 1 vector of random errors.

It is desired to test the hypothesis

$$H_o . H\beta = h \tag{1.2}$$

where H is a d x p matrix of full row rank $(d \leqslant p)$ and h is a known d x 1 vector. The unconstrained model (1.1) is called the full model and the model subject to the constraint (1.2) is called the reduced model.

If the errors, ϵ_i, are assumed to be i.i.d $N(0, \sigma^2)$, the least squares estimate of β, β_{LS}, is obtained by minimizing the sum of squares $SS = \sum_{i=1}^{n} (y_i - x_i'\beta)^2$ where x_i' is the ith row of X. The solution is $\beta_{LS} = (X'X)^{-1} X'Y$ and the classical F-statistic used to test H_o is

*Associate Professor, University of the Philippines at Los Baños, College, Laguna, Philippines.
**Associate Professor, Kansas State University, Manhattan, Kansas.

$$F = \frac{[SSR\ (Reduced) - SSR\ (Full)]/d}{\hat{\sigma}^2}$$

$$= \frac{n-p}{d}\left[\frac{(H\hat{\beta}_{LS} - h)' [H(X'X)^{-1} H']^{-1} (H\hat{\beta}_{LS} - h)}{y'[I - X(X'X)^{-1} X']y}\right]$$

where SSR (Reduced) is the residual sum of squares in the reduced model and SSR (Full) is the residual sum of squares in the full model.

If the errors are nonnormal or if the distribution of the ϵ'_is is long-tailed, the least squares procedure yields poor estimates of β, particularly in the single location case (Andrews, et. al., 1972). Huber (1973) proposed a robust method for "deflating the influence of gross errors of any kind" by minimizing some expression which is less sensitive to extreme values of the residuals. He defined an M-estimate of β as one that minimizes the dispersion of residuals.

$$D_M = \Sigma_{i=1}^{n} \rho\, (y_i - x_i'\beta) \tag{1.3}$$

or, equivalently, as the solution to the system of equations

$$\sum_{i=1}^{n} \psi\, (y_i - x_i'\beta)x_{ij} = 0, j = 1, 2, \ldots, p \tag{1.4}$$

where $\psi\ (t)$ is the first derivative of $\rho(t)$ with respect to t, and ρ is a convex function.

Hettmansperger and McKean (1977) have developed a robust alternative to least squares for testing hypotheses about the parameter β. Their approach is based on the ranks of the residuals. They defined a statistic F_R and showed that, when properly normalized, its sampling distribution is asympotic $\underline{\chi}^2$. Schrader (1976) also developed a robust alternative to least squares based on Huber's M-estimates. He defined a statistic F_M and showed that this statistic, properly normalized, has asymptotic $\underline{\chi}^2$ – distribution. Schrader and McKean (1977) showed that these two robust test statistics can be applied to most analysis of variance problems.

In this paper, a test of linear hypothesis based on the asymptotic normality of Huber's M-estimates will be considered. The statistics proposed by Schrader (1976) requires lengthly iterative computational procedures. A simpler method is proposed that would considerably reduce calculations. The procedure basically applies weighted least squares techniques to Huber's M-estimates.

BACKGROUND THEOREMS

Theorem 1. Under the regularity conditions given by Huber (1973),

$$\hat{\beta}_M - \beta \xrightarrow{D} MVN(O, K(\psi,F)\ (X'X)^{-1})$$

where $\hat{\beta}_M$ is the Huber M-estimate of β and $K(\psi, F) = E(\psi^2)/\,E(\psi')^2$.

The symbol $\xrightarrow{D}$ denotes convergence in distribution.

Proof: Huber (1973).

Theorem 2. Let D_M (Full) = minimum value of $\Sigma_{i=1}^{n}\ \rho\ (y_i - x_i'\beta)$ under the full model (1.1), and let D_M (Red) = minimum value of $\Sigma_i^n\ \rho\ (y_i - x'\beta)$ under the model subject to constraint (1.2), then

$$\frac{2E(\psi')}{E(\psi^2)}\ [D_M\ (\text{Red}) - D_M\ (Full)] \xrightarrow{D} x^2\ (d)$$

where $\chi^2\ (d)$ is a central χ^2 – random variable with d degrees of freedom.

Proof: Schrader (1976).

ONE-STEP PROCEDURES FOR ESTIMATING β

Huber's class of M-estimates, defined as the solution to the system of equations (1.4), can be calculated only iteratively. Bickel (1975) proposed one-step procedures for estimating β in the general linear model using Huber's ρ-function

$$\rho(t) - \begin{cases} \frac{1}{2} t^2, & |t| \leqslant c \\ c|t| - \frac{1}{2} c^2, & |t| > c \end{cases} \tag{3.1}$$

with $c = k\hat{\sigma}$ for scale invariance and k is some positive constant. Computationally, one-step estimates are obtained as follows:

let $s^+ = \left[i / \ (y_i - x_i'\beta^*) > k\sigma \right]$, $S^- = \left[i / \ (y_i - x_i'\beta^*) < -k\hat{\sigma} \right]$

$S^o = \left[i | \ |y_i - x_i'\beta^*| \leqslant k\hat{\sigma} \right]$ where $\hat{\beta}^*$ is an initial shift invariant estimate of β,

$$\hat{\sigma} = \frac{med\ |y_i - x_i'\beta^*|}{\phi^{-1}\left(\frac{3}{4}\right)}$$

a robust estimate of σ, and ϕ^{-1} is the inverse of the standard normal cdf. Replace any residual

$$y_i - \Sigma_{j=1}^{p} \ x_{ij}\beta_j^*$$

by $k\hat{\sigma}$ if $i \epsilon S^+$ and by $-k\hat{\sigma}$ if $i \epsilon S^-$. If $i \notin S^o$, replace x_{ij} by 0 for $j = 1, 2, \ldots, p$. If the resulting vector of modified residuals is denoted by R^* and the resulting matrix of modified x_{ij} by X^*, then the one step estimator of β is

$$\hat{\beta}_M = \beta^* + (X^{*\prime}X^*)^{-1} X'R^*$$

One-step procedures can readily be applied in the single location case, the p-sample case, regression through the origin, and most simple analysis of variance situations where a robust estimate of β, usually the median, can be used as an initial estimate of β.

In more complicated situations, robust initial estimates of β are not readily available and one may have to use a least squares estimator of β as starting point – a poor estimate of β in problems that are dealt with robust procedures. Andrews (1974) has proposed an iterative procedure for estimating β when one-step procedures are felt inadequate.

GENERAL PROCEDURE

Theorem 2 makes it possible to test hypothesis about the parameter of the general linear model with Huber's M-estimates in a manner very much like that of least squares theory. Schrader (1976) defined an F-like, F_M, statistics as

$$F_M = \frac{[D_M\ (Red) - D_M\ (Full)]/d}{\hat{\lambda}_M}$$

where

$$\hat{\lambda}_M = \frac{\frac{1}{n-p}\ \Sigma_{i=1}^{n}\ \psi^2\ (e_i)}{2\frac{1}{n}\ \Sigma_{i=1}^{n}\ \psi'\ (e_i)}$$

is a consistent estimator of $\lambda_M = E(\psi^2)/2E(\psi')$. His asymptotic theory shows that F_M can be compared with the F-distribution with d and $(n-p)$ degrees of freedom. This test statistic requires calculating the dispersion

$$D_M = \Sigma_{i=1}^{n}\ \rho\ (y_i - x_i\beta)$$

for both the full and the reduced models.

For example, in the analysis of covariance model with n observations per cell, the response is described as

$$y_{ij} = \alpha_i + \beta_i x_{ij} + \epsilon_{ij};\ i= \quad i = 1, 2, \ldots, t; \quad j = 1, 2, \ldots, n$$

where the ϵ_{ij} are independent with distribution function F assumed to be symmetric about 0. Suppose it is of interest to test the hypothesis, $H_o: \beta_1 = \beta_2 = \ldots = \beta_t = \beta, \beta$ unspecified. Under the full model, $2t$ parameters have to be estimated from whence D_M (Full) is obtained. Then the model is restricted by the conditions of the null hypothesis to yield

$$y_{ij} = \alpha_i + \beta x_{ij} + \epsilon_{ij};\ i = 1, 2, \ldots, t;, j = 1, 2, \ldots, n.$$

Under this reduced model, $(t + 1)$ parameters have to be estimated to obtain D_M (Red).

A method for testing $H_o: H\beta = h(H$ of rank $d \leqslant p)$ is proposed using the asymptotic normality of Huber's M-estimates.

Define the $\hat{\beta}_M$ --model as

$$\hat{\beta}_M = I\beta + \epsilon^* \tag{4.1}$$

where ϵ^* is asymptotically distributed as a MNV $(O, K[\psi, F]\ (X'X)^{-1})$ and

$$K(\psi, F) = \frac{E\psi 2(\epsilon_i)}{[EX'(\epsilon_i)]^2},\ \epsilon_i = y_i - x'\beta.$$

Note that $K(\psi, F)$ parallels $\sigma 2$ in the usual least squares procedure. A natural estimate of $K(\psi, F)$, suggested by Huber (1973), is

$$K(\psi, F) = \frac{\frac{1}{n-p}\ \Sigma_{i=1}\ \psi^2\ (y_i - x_1'\ \hat{\beta}_M)}{[\frac{1}{n}\ \Sigma_{i=1}^n\ \psi'\ (y_i - x_i'\hat{\beta}_M)]^2}. \tag{4.2}$$

From the $\hat{\beta}_M$ – model, test of linear hypothesis about β can be carried out. The general approach is to constrain the model (4.1) subject to H_o. This yields.

$$\hat{\beta}_M - H^-h = (I - H^-H)\beta + \epsilon^* \tag{4.3}$$

where H^- denotes the Moore-Penrose generalized inverse. Weighted least squares techniques can be applied to the constrained model (4.3) to obtain the sum of squares due to testing H_o, denoted by SSH_o i.e.,

$$SSH_o = (\hat{\beta}_M - H^- h)' \beta (\hat{\beta}_M - H^- h)$$

where
$B = X'X - X'X (I - H^- H) [(I - H^- H) X'X (I - H^- H)]^- (I - H^- H) X'X$

An alternative form of SSH_o is given as

$$SSH_o = (H\hat{\beta}_M - h)' [H(X'X)^{-1} H']^{-1} (H\hat{\beta}_M - h)$$

since $H \hat{\beta}_M - h$ is asymptotically distributed as a

$$MVN[(H\hat{\beta}_M - h),\ K(\psi, F)H(X'X)^{-1}H'].$$

Now, since SSH_o is a quadratic form in $\hat{\beta}_M$, $\hat{\beta}_M$ is asymptotically multi-variate normal with mean β and covariance matrix $V = K(\psi, F)\ (X'X)^{-1}$, and $\frac{1}{K(\psi, F)} B\ V$ is idempotent of rank d, then

$$\frac{SSH_o}{K(\psi, F)} \xrightarrow{D} \chi^2 \left[d, \frac{1}{2K(\psi, F)} (\beta - H^- h)' B (\beta - H^- h)\right]$$

This result can be summarized in the following theorem.
Theorem. In the $\hat{\beta}_M$ – model, $\hat{\beta}_M = I\beta + \epsilon^*$ is asymptotically

$$MVN\ (O,\ K(\psi, F)\,(X'X)^{-1}),$$

$$\frac{SSH_O}{K(\psi, F)} \xrightarrow{D} X^2 \left[d, \frac{1}{2K(\psi, F)} (\beta - H^- h)' B (\beta - H^- h) \right]$$

where $SSH_o = (\hat{\beta}_M - H^- h)' B(\hat{\beta}_M - H^- h)$ is the sum of squares for testing the hypothesis $H_o : H\beta = h$ vs. $H_a : H\beta \neq h$.

Corollary. Under $H_o : H\beta = h$,

$$\frac{SSH_o}{K(\psi, F)} \xrightarrow{D} \chi^2 (d).$$

If a consistent estimator of $K(\psi, F)$ is used, it is proposed to use the statistic

$$F_c = \frac{SSH_o / d}{K(\psi, F)} \tag{4.4}$$

under H_o, and compare it with the F-distribution with d and $(n-p)$ degrees of freedom. Small sample properties of this statistic are examined via Monte Carlo methods.

To illustrate the general procedure, consider a two-way design model,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

$$i = 1, \ldots, b;\; j = 1, \ldots, t;\; k = 1, \ldots, r.$$

This can be equivalently expressed as

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \quad \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}.$$

M-estimates of μ_{ij}, denoted by $\mu_{M\,ij}$, can be obtained by applying Bickel's one-step procedure. Then an estimate of $K(\psi, F)$ is computed as

$$K(\psi, F) = \frac{\frac{1}{bt(r-1)} \Sigma_{i=1}^{b} \Sigma_{j=1}^{t} \Sigma_{j=1}^{t} \Sigma_{k=1}^{r} \psi^2 (y_{ijk} - \hat{\mu}_{M\,ij})}{[\frac{1}{btr} \Sigma_{i=1}^{b} \Sigma_{j=1}^{t} \sum_{k=1}^{r} \psi(y_{ijk} - \hat{\mu}_{m\,ij})]^2} \tag{4.5}$$

The $\hat{\beta}_M$ – model is thus,

$$\hat{\mu}_M = \mu + \epsilon^* \tag{4.6}$$

where ϵ^* is asymptotically $MVN\ (O, \frac{K(\psi, F)}{r} I_{bt})$, since $(X'X)^{-1} = (1/r)\, I_{bt}$.

Suppose it is desired to test the hypothesis of no interaction, H_o: $\mu_{ij} - \mu_{ik} + \mu_{pk} - \mu_{pj} = 0$ for every pair (i, j) and (p, k), $i \neq p$, $j \neq k$. The hypothesis $H_o: H\beta = h$ then appears as $H_o: H\mu = O$, where (Milliken, 1971)

$$H = B_t \otimes B_b$$

$$\underset{(t-1)xt}{B_t} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 1 & -2 & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & & . \\ . & . & . & & . \\ 1 & 1 & 1 & \dots & -(t-1) \end{bmatrix}$$

$$\underset{(b-1)xb}{B_b} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 1 & -2 & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & & . \\ . & . & . & & . \\ 1 & 1 & 1 & \dots & -(b-1) \end{bmatrix}$$

The constrained model is

$$\hat{\mu}_M = (I - H^- H)\mu + \epsilon^*$$

$$= [I_{bt} - (I - \frac{1}{t} J_t) \otimes (I_b - \frac{1}{t} J_b)]\mu + \epsilon^*,$$

since $H^- H = (B_t \otimes B_b)^- (B_t \times B_b) = (I_t - \frac{1}{t} J_T) \otimes (I_b - \frac{1}{b} J_b)$.

So that

$$SSH_o = \hat{\mu}_M' B \ \hat{\mu}_M = r\hat{\mu}_M' [(I_t - \frac{1}{t} J_t) x (I_b - \frac{1}{b} J_b)] \hat{\mu}_M$$

Since

$$B = X'X - X'X (I - H^- H) [(I_t - H^- H) X'X (I - H^- H)]^- (I - H^- H) X'X$$

$$= rI - rI (I - H^- H) [(I - H^- H)^- (I - H^- H)$$

$$= r(I - I + H^- H) = rH^- H.$$

The F_c – statistic can then be calculated from equation (4.4) with an estimate of $K(\psi, F)$ given by equation (4.5)

Another approach to testing linear hypothesis about β is by reparametrizing the $\hat{\beta}_M$ – model to a design structure as

$$\hat{\beta}_M = A \tau + \epsilon^* \tag{4.7}$$

where ϵ^* is asymptotically MVN $(O, K(\psi, F) (X'X)^{-1})$. A is a pxs design matrix, and τ is an sxl vector of unknown parameters. Suppose it is of interest to test the hypothesis H_o: $H \tau = O$, where H is a dxs matrix of rank $d \leqslant p$ and $H \tau$ is a set of linearly estimable functions. The model (4.7) is then constrained by the conditions of H_o from which the sum of squares for testing H_o is obtained.

For example, the $\hat{\beta}_M$ – model for the two-way design model previously discussed (equation 4.6) can be reparametrized as follows:

$$\hat{\mu}_M = \begin{bmatrix} \hat{\mu}_{M11} \\ \hat{\mu}_{M12} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mu}_{Mbt} \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \cdot \\ \cdot \\ \cdot \\ \mu_{bt} \end{bmatrix} + \epsilon^* = \begin{bmatrix} \mu + \alpha_1 + \beta_1 + \gamma_{11} \\ \mu + \alpha_2 + \beta_2 + \gamma_{12} \\ \cdot \\ \cdot \\ \cdot \\ \mu + \alpha_b + \beta_t + \gamma_{bt} \end{bmatrix}$$

$$= A \tau + \epsilon^*$$

where

$$A = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ . & . & . & & . & . & . & & . & . & . & & . \\ . & . & . & & . & . & . & & . & . & . & & . \\ . & . & . & & . & . & . & & . & . & . & & . \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$= (j_t \otimes j_b, \; j_t \otimes I_b, I_t \otimes j_b, I_{bt})$ and

$\tau = (\mu, \alpha_1 \alpha_2, \dots, \alpha_b, \beta_1, \beta_2, \dots \beta_t, \gamma_{11}, \gamma_{12}, \dots, \gamma_{BT})$.

Suppose it is desired to test the hypothesis of no interaction, $H_o: \gamma_{11} = \gamma_{12} = . \; . \; . = \gamma_{bt} = 0$, the model constrained by H_o is

$$\hat{\mu}_M = C\tau^* + \epsilon^* \qquad (4.8)$$

where $C = (j_t \otimes j_b, j_t \otimes I_b, I_t \otimes j_b)$

and $\tau^* = (\mu, \alpha_1, \alpha_2, \dots, \alpha_b, \beta_1, \beta_2, \dots, \beta_t)$.

The sum of squares for testing H_o is the sum of squares residual of

The sum of squares for testing H_o is the sum of squares residual of
the constrained model (4.8), i.e.,

$$SSH_o = r\hat{\mu}'_M (I - CC^-)\hat{\mu}_M = r\mu'_M [I_t - \frac{1}{t} J_t) \otimes (I_b - \frac{1}{b} J_b)] \hat{\mu}_M .$$

SIMULATION STUDY FOR THE STATISTIC F_c

The distribution of the test statistic F_c was studied via Monte Carlo methods. The simulation study was conducted by using the one-way analysis of variance model,

$$y_{ij} = \mu_i + \epsilon_{ij},$$

where $i = 1, 2, \ldots, t$ population, $j = 1, 2, \ldots, r$ observations per population. The y_{ij}'s were generated from the $N(10, 1)$ with probability $(1 - p)$ and from the $N(10, 16)$ with probability p using the normal random number generator Super Duper (Marsaglia et. al., 1976). For each population, Huber's ρ-function (given in equation 3.1) was used with $c = 1.5\sigma$, together with Bickel's one-step procedure, to estimate the μ_i's. Starting values of μ_i, μ_i^*, were taken as the median of the r observations from each population, that is,

$$\mu_i^* = \underset{j}{med}\left\{y_{ij}\right\}$$

and a corresponding robust estimate of σ was computed as

$$\hat{\sigma} = \frac{\underset{j}{med} \mid y_{ij} - \mu_i^* \mid}{\Phi^{-1}(\frac{3}{3})}, j = 1, 2, \ldots, r.$$

An estimate of $K(\psi, F)$ was computed as follows:

$$K(\psi, F) = \frac{\frac{1}{t(r-1)} \Sigma_{i=1}^{t} \Sigma_{j=1}^{r} \psi^2 (y_{ij} - \hat{\mu}_{Mi})}{[\frac{1}{rt} \Sigma_{i=1}^{t} \Sigma_{j=1}^{r} \psi' (y_{ij} - \hat{\mu}_{Mi})]^2}$$

The $\hat{\beta}_M$ – model is then $\hat{\mu}_M = \mu + \epsilon^*$, where ϵ^* is asymptotically MVN $(O, [K(\psi, F)/r] I_t)$. Under the hypothesis of no population mean difference, $H_o: \mu_1 = \mu_2 = \ldots = \mu_t = \mu$, μ unspecified, the statistic F_c reduces to

$$F_c = \frac{SST/(t-1)}{K(\psi, F)/r}$$

where

$$SST = \Sigma_{i=1}^{t} \hat{\mu}_{Mi}^2 - [\Sigma_{i=1}^{t} \hat{\mu}_{Mi}]^2/r$$

and

$$\hat{\mu}_{Mi} = \text{one-step M-estimate of } \mu_i.$$

A simulation of size 500 was run for each test case. The frequency distribution of F_c was compared to the distribution of the theoretical F with $(t-1)$ and $t(r-1)$ degrees of freedom for each case. The cases studied were:

Case	*t*	*r*	*p*
1	3	5	.00
2	3	5	.10
3	3	5	.20
4	3	11	.00
5	3	11	.10
6	3	11	.20
7	5	9	.00
8	5	9	.10
9	5	9	.20
10	5	25	.00
11	5	25	.10
12	5	25	.20

The cases where $p = .00$ are the noncontaminated cases and were included in the study to determine how the statistic F_c compares with the theoretical F-distribution having $(t - 1)$ and $t(r - 1)$ degrees of freedom under normal conditions. The results are shown graphically in Figures 1a, 1b, and 1c.

The simulation results indicate that when the data are not contaminated, the distribution of F_c reasonably fits the theoretical F-distribution even with moderate sample sizes, the exception being Case 1. However, with 10% contamination, a very good fit is observed for all cases studied. In cases where 20% of the data are contaminated, the fit of F_c is not quite so good as when 10% are contaminated, but it may be observed that even with small samples sizes the distribution of F_c is close to that of the theoretical F-distribution.

A simulation study of the power of the test was also run for all twelve cases. The simulated power of F_c was compared to the respective noncentral F-distribution.

The noncentrality parameter, λ, was computed for the F-distribution as

$$\lambda = \frac{r}{2\sigma_c^2} \Sigma_{i=1}^{t} (\mu_i - \bar{\mu})^2,$$

where

$$\sigma_c^2 = p(\sigma')^2 + (1-p)\sigma^2;$$

$(\sigma')^2 = 16$, the variance of the contaminating normal distribution;

$\sigma^2 = 1$, the variance of the uncontaminated normal distribution,; and

$$\bar{\mu} = \Sigma_{i=1}^{t} \mu_i / t$$

λ was then transformed to $\phi = \sqrt{2\lambda/t}$. The results of the simulation study are shown in Table I.

The power study shows that F_c is a good statistic for testing linear hypotheses about the parameters in the $\hat{\beta}_M$ – model when the data are contaminated. Its power is greater than than of the usual F test under 10% and 20% contamination. Yet with no contamination. the power of F_c is less than the usual F, but the loss in power is not more than 3%.

In conclusion, for the cases studied, the sampling distribution of F_c is adequately described by the F-distribution with $(t - 1)$ and $t(r - 1)$ degrees of freedom when data are not contaminated. When the data are contaminated, the distribution of F_c closely fits the usual F-distribution and its power is greater than of the usual F.

EXAMPLE

The following example was discussed by Box and Cox (1964). The data are records of survival times of animals in a 3 x 4 factorial experiment with 4 observations per cell. The model assumed is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where $i = 1, 2, 3, j = 1, 2, 3, 4, k = 1, 2, 3, 4$. The model can be equivalently written as

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \quad \mu_{ij} = \mu + \alpha_i + \beta_j + \alpha_{ij}$$

Least squares estimates of μ_{ij} are given in Table II. The usual analysis of variance, shown in Table III, indicated no evidence of a poison x treatment (Pxt) interaction at the 5% level. This was the conclusion reached by Box and Cox. Schrader's (1977) robust analysis of the data indicated otherwise.

Bickel's one-step procedure was applied to each individual cell with $k = 1.5$ and σ estimated for each cell. Initial estimates of the μ_{ij}'s were taken to be the median of the y_{ijk}'s in the (i, j)th cell. Robust estimates of μ_{ij}, $\hat{\mu}_{M\,ij}$, are shown in Table IV. After the $\hat{\mu}_{M\,ij}$'s had been obtained, weighted least squares techniques were applied to test for main effects and interactions The results are given in Table V. No error sum of squares is entered in the table, unlike the conventional analysis of variance table.

The procedure was repeated with $k = 1.0$ in order to compare results with Schrader. The results are shown in Table VI and VII. In his analysis, Schrader used Huber's ρ-function with $k = 1.0$ and an estimate of scale as the eightieth percentile point of the error distribution. Comparison of sum of squares column from Table VII and the dipersion column of Table 4.4 (Schrader, 1976) indicates a close agreement in the numerical values. The main difference is in the MSE and mean dispersion for error entry in the respective tables. The statistic F_c failed to detect the presence of interaction effects for both $k = 1.0$ and $k = 1.5$ at the 5% level – the same conclusion reached by Box and Cox.

TABLE I. POWER STUDY

Means	*Size of the test*	*Power of the usual F-test* *p*=.00	*p*=.10	*p*=.20	Power of F_c *p*=.00	*p*=.10	*p*=.20
			t=3 r=5				
		ϕ=3.2	ϕ=2.0	ϕ=1.6			
10, 13, 13	.01	.936	.492	.288	.918	.772	.544
	.05	.995	.783	.580	.990	.922	.698
	.10	1.000	.882	.722	.998	.966	.880
			t=3 r=11				
		ϕ=2.7	ϕ=4.4	ϕ=1.4			
10, 11, 12	.01	.916	.448	.276	.904	.770	.598
	.05	.981	.706	.530	.978	.966	.818
	.10	.993	.812	.662	.990	.958	.894
			t=5 r=9				
		ϕ=2.7	ϕ=1.7	ϕ=1.3			
10, 10, 11	.01	.983	.609	.346	.980	.876	.696
12, 12	.05	.998	.827	.603	.998	.964	.864
	.10	1.000	.901	.726	1.000	.974	.936
			t=5 r=25				
		ϕ=1.9	ϕ=1.2	ϕ=0.9			
10, 10.5,	.01	.806	2.94	.147	.782	.580	.376
10.5, 11, 11	.05	.930	.580	.329	.920	.800	.654
	.10	.964	.661	.450	.954	.872	.772

TABLE II. LEAST SQUARES ESTIMATES OF μ_{ij}

	Treatment			
Poison	*A*	*B*	*C*	*D*
I	0.41	0.88	0.57	0.61
II	0.32	0.82	0.38	0.68
III	0.21	0.34	0.24	0.33

TABLE III. ANOVA FOR POISON X TREATMENT DATA

Source	*df*	*SS*	*MS*	*F*	$F_{.05}$
Poison	2	1.0330	0.5164	23.3	3.26
Treatment	3	0.9212	0.3071	13.8	2.86
P x T	6	0.2501	0.0417	1.9	2.35
Error	36	0.8807	0.2220		

Table IV. ONE-STEP M-ESTIMATES OF μ_{ij} (k = 1.5)

	Treatment			
Poison	*A*	*B*	*C*	*D*
I	0.44	0.87	0.57	0.63
II	0.32	0.82	0.38	0.67
III	0.21	0.34	0.24	0.33

TABLE V. ROBUST ANOVA FOR POISON x TREATMENT DATA
(k = 1.5)

Source	*df*	*SS*	*MS*	F_c
–				–
Poison	2	0.26629	0.13314	20.64
Treatment	3	0.21960	0.07320	11.35
P x T	6	0.05829	0.00972	1.51
Error	36	–	0.00645	

TABLE VI. ONE-STEP M-ESTIMATES OF μ_{ij} ($k = 1.0$)

	Treatment			
Poison	*A*	*B*	*C*	*D*
I	0.44	0.85	0.55	0.64
II	0.32	0.78	0.38	0.64
III	0.22	0.34	0.24	0.32

TABLE VII. ROBUST ANOVA FOR POISON x TREATMENT DATA ($k = 1.0$)

Source	*df*	*SS*	*MS*	$F_c c$
Poison	2	0.25379	0.12689	17.38
Treatment	3	0.19685	0.06562	8.98
P x T	6	0.04519	0.00753	1.03
Error	36	–	0.00730	

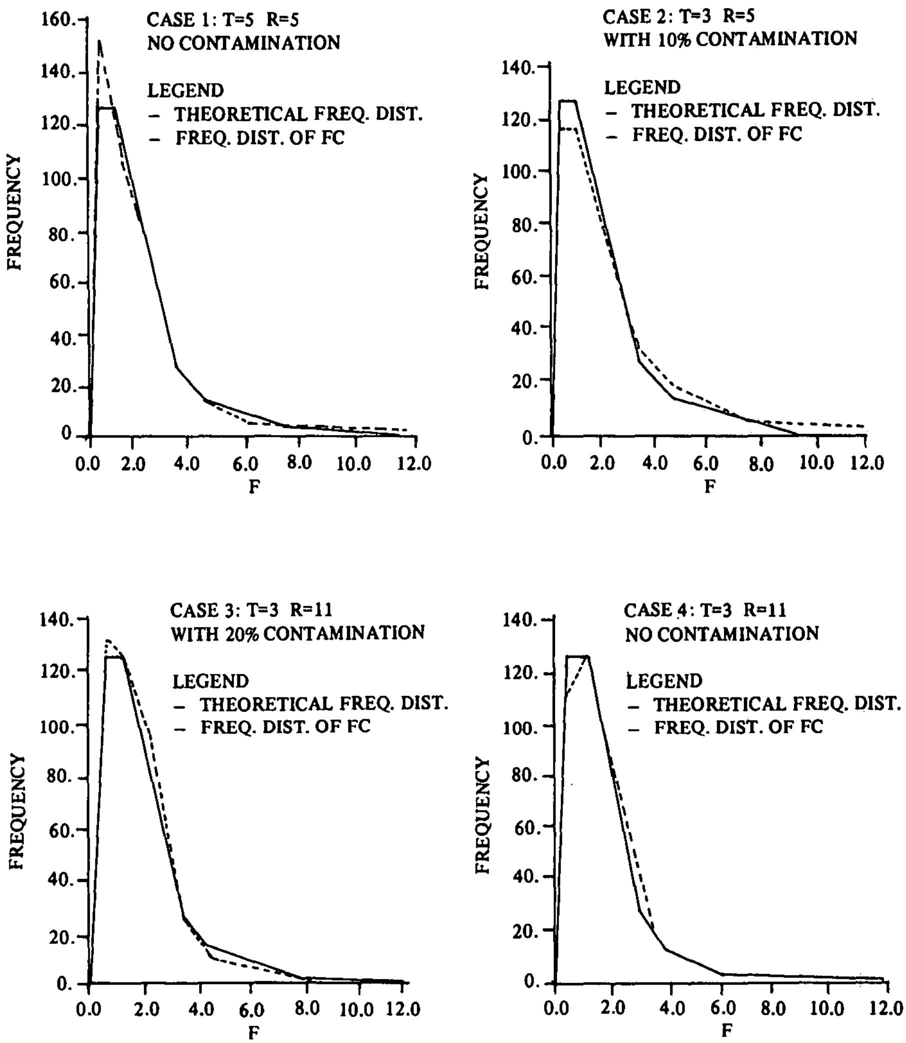

Figure 1a Frequency distribution of F and FC

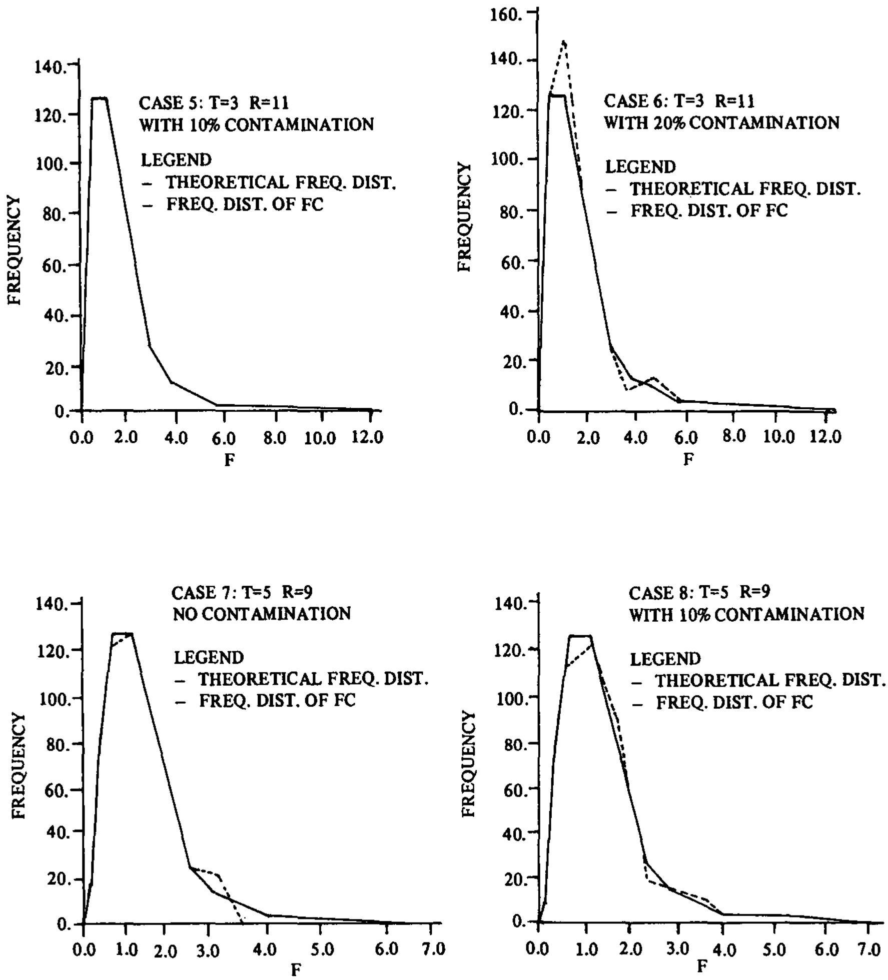

Figure 1b Frequency distribution of F and FC

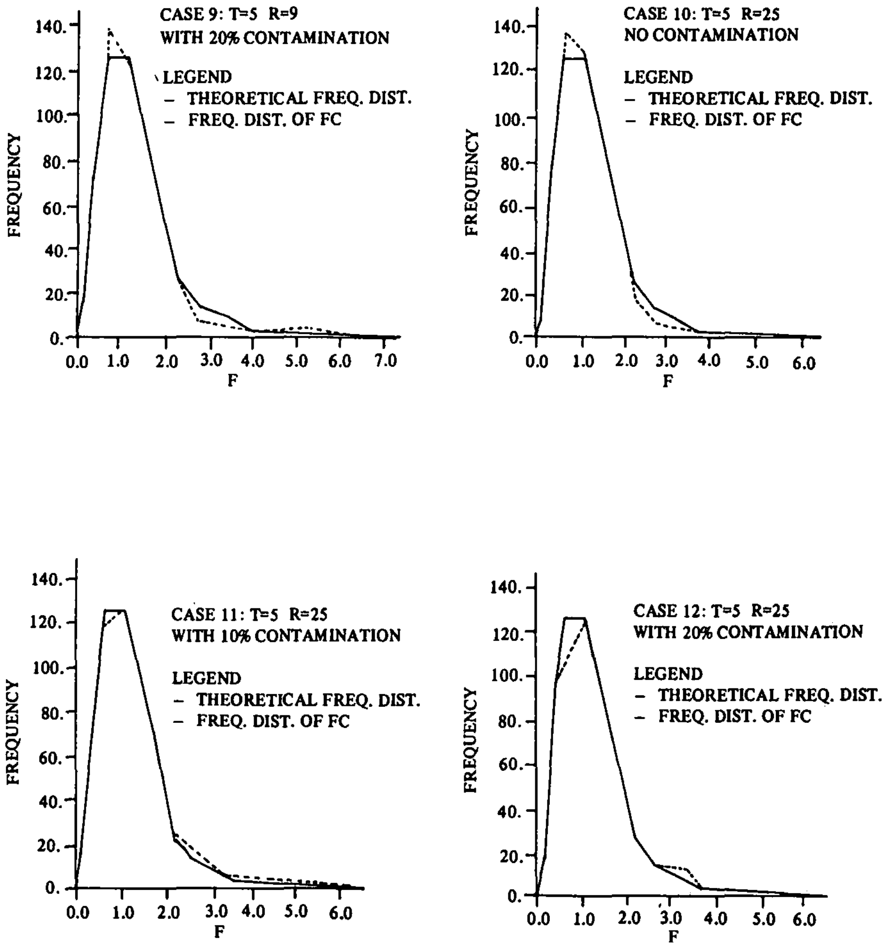

Figure 1c Frequency distribution of F and FC

LITERATURE CITED

1. Andrews, D. F. (1974). "A robust method for multiple linear regression". *Technometrics.* 16, 523-31.
2. Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W., (1972). *Robust Estimation of Location: Survey and Advances.* New Jersey: Princeton University Press.
3. Bickel, P. J. (1975). "One-step Huber estimates in the linear model". *JASA.* 70, 428-34.
4. Box, G. E. P. and Cox, D. R. (1964). *"An Analysis of transformations". J. Roy. Statist. Soc. B.* 26, 211-52.
5. Hettmansperger, T. P. and McKean, J. W. (1977). "A robust alternative based on ranks to least squares in analyzing linear models". *Technometrics.* 19, 275-84.
6. Huber, P. J. (1973). "Robust regression: Asymptotics, conjectures and Monte Carlo". *Ann. Statist.* 1, 799-821.
7. Marsaglia, G. and Ananthanarayanan, K. and Paul, N. J. (1976). "Improvements on fast methods for generating normal random variables". Information processing letters. 5, 27-30.
8. Schrader, R. M. (1976). *Robust Inferences Based on M-estiamtes in the Linear Model.* Ph.D Dissertation. Pennsylvania State University.
9. Schrader, R. M. and McKean, J. W. (1977). "Robust analysis of variance". *Commun. Statist.* A. 6, 879-94.