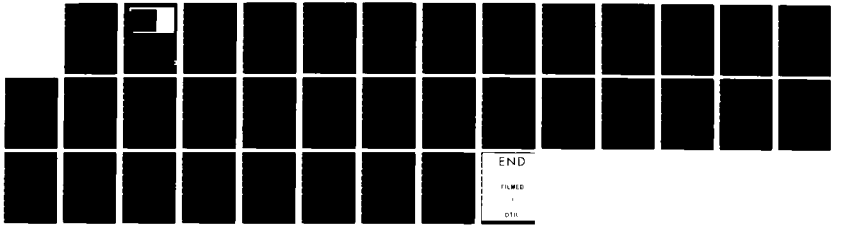
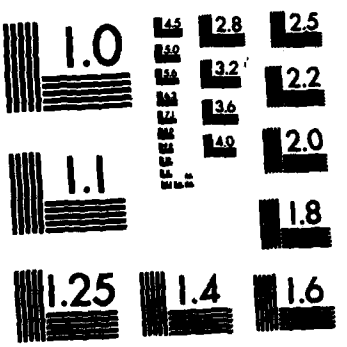
AD-A124 369 PERCENTAGE POINTS AND POWER CALCULATIONS FOR OUTLIER TESTS IN A REGRESSIO..(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER C FUCHS ET AL. OCT 82 1/1

UNCLASSIFIED MRC-TSR-2436 DAAG29-80-C-0041 . F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A124369

MRC Technical Summary Report #2436

PERCENTAGE POINTS AND POWER CALCULATIONS FOR OUTLIER TESTS IN A REGRESSION SITUATION

Camil Fuchs and Norman R. Draper

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

October 1982

(Received September 14, 1982)

DTIC FILE COPY

Approved for public release
Distribution unlimited

DTIC
ELECTE
FEB 15 1983
D
A

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

83 02 014 127

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

PERCENTAGE POINTS AND POWER CALCULATIONS FOR OUTLIER TESTS IN A REGRESSION SITUATION

Camil Fuchs* and Norman R. Draper**

Technical Summary Report #2436
October 1982

ABSTRACT

The usefulness of an extra sum of squares statistic Q_K for detecting K outliers has been discussed previously in the context of two-way tables. (See Gentleman and Wilk, 1975a, 1975b; John and Draper, 1978; and Draper and John, 1980.) That work is extended here to straight line regression situations arising from and motivated by a specific set of research data. Percentage points for the appropriate test statistics are obtained by simulation, approximations for these percentage points are suggested, and power calculations are made for various designs and outlier situations. Correct determination of K and position(s) of the outlier(s) appear to be important in influencing power.

AMS(MOS) Subject Classification: 62J05, 62J99

Key Words: Detection of outliers; Q_K statistic; Residuals;

Work Unit Number 4 - Statistics and Probability

* Statistics Departments, Tel-Aviv University, Tel-Aviv, and Haifa University, Haifa, Israel
** Statistics Department, University of Wisconsin, 1210 West Dayton St., Madison, WI 53706

Sponsored by the U. S. Army under Contracts Nos. DAAG29-80-C-0041 and DAAG29-80-C-0113.

SIGNIFICANCE AND EXPLANATION

Previous work in two-way tables on the use of an extra sum of squares statistic Q_K to detect outliers is extended to the straight line regression situation, motivated by some specific research data. Percentage points for the appropriate test statistics are obtained via simulation, and approximations for these percentage points are suggested. Power calculations, made for various derived designs and outliers situations, show the effects of the choice of K, the number of potential outliers assumed, and the positions of the outliers in the predictor variable space.

DTIC
COPY
INSPECTED
3

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

PERCENTAGE POINTS AND POWER CALCULATIONS FOR OUTLIER TESTS IN A REGRESSION SITUATION

Camil Fuchs* and Norman R. Draper**

1. INTRODUCTION AND NOTATION

In prior work by Gentleman and Wilk (1975a, 1975b), by John and Draper (1978), and by Draper and John (1980), the utility of the Q_K statistic for checking outliers in data used to fit a linear model was explored. If one or more observations (K in general) are suspect, extra dummy variables $\underset{\sim}{\theta}$ can be inserted in the model to represent the discrepancies, and the Q_K statistic is simply the extra sum of squares due to the estimates of the parameters associated with the dummy variables. Specifically, if our original model is $\underset{\sim}{y} = \underset{\sim}{\underset{\sim}{X}}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, we can write

$$\underset{\sim}{y} = \begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}_1 & \underset{\sim}{0} \\ \underset{\sim}{X}_2 & \underset{\sim}{I} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\theta} \end{pmatrix} + \underset{\sim}{\varepsilon} \qquad (1.1)$$

where $\underset{\sim}{y} = (\underset{\sim}{y}_1', \underset{\sim}{y}_2')'$ is an $n \times 1$ vector of response observations, $\underset{\sim}{X} = (\underset{\sim}{X}_1', \underset{\sim}{X}_2')'$ is an $n \times p$ matrix of predictor variable values, $\underset{\sim}{\beta}$ is a $p \times 1$ vector of model parameters, $\underset{\sim}{\theta}$ is a $K \times 1$ vector of additional parameters and $\underset{\sim}{\varepsilon}$ is an $n \times 1$ vector of random errors distributed $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \underset{\sim}{I}\sigma^2)$. The positioning of the K potential outliers as the last elements of $\underset{\sim}{y}$ is for convention only.

---

* Statistics Departments, Tel-Aviv University, Tel-Aviv, and Haifa University, Haifa, Israel
** Statistics Department, University of Wisconsin, 1210 West Dayton St., Madison, WI 53706

Sponsored by the U. S. Army under Contracts Nos. DAAG29-80-C-0041 and DAAG29-80-C-0113.

If, a priori, the number K and the locations implied by choice of $\underset{\sim}{X}_2$ are specified, and if s^2 is an unbiased estimate of σ^2 independent of Q_K, then

$$F_K = \{Q_K/K\}/s^2 \tag{1.2}$$

has a non-central F distribution with noncentrality parameter

$$\lambda_K = \underset{\sim}{\theta}'(\underset{\sim}{I}-\underset{\sim}{H}_{22})\underset{\sim}{\theta}/2\sigma^2, \tag{1.3}$$

where $\underset{\sim}{H}_{22}$ is the $K \times K$ lower right part of $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$ when $\underset{\sim}{H}$ is partitioned as

$$\underset{\sim}{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} X_1(X'X)^{-1}X_1' & X_1(X'X)^{-1}X_2' \\ X_2(X'X)^{-1}X_1' & X_2(X'X)^{-1}X_2' \end{bmatrix}. \tag{1.4}$$

(See Cook, 1979; Ellenberg, 1976.) $\underset{\sim}{H} = ((h_{ij}))$ is sometimes called the hat matrix, because $\hat{\underset{\sim}{y}} = \underset{\sim}{H}\underset{\sim}{y}$ so the h values give relative contributions ("leverage") of the corresponding observations to the fitted value at each point. Note that, in general, trace $\underset{\sim}{H} = p$, the number of parameters in the model. For $K = 1$, the parameter of non-centrality is $\lambda_1 = \theta_n^2(1-h_{nn})/2\sigma^2$.

In most practical situations, the F distribution is not appropriate because the number and the locations of the outliers have to be elucidated from the data. In that case, the quadratic form Q_K has to be calculated for all the permutations of K omissions. The largest Q_K over all the permutations of K omissions is denoted by Q_{Km}.

Because

$$Q_K = \underset{\sim}{e}_2'(\underset{\sim}{I}-\underset{\sim}{H}_{22})^{-1}\underset{\sim}{e}_2 \tag{1.5}$$

where

$$\underset{\sim}{e} = \begin{pmatrix} \underset{\sim}{e}_1 \\ \underset{\sim}{e}_2 \end{pmatrix} = (\underset{\sim}{I}-\underset{\sim}{H}) \begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \end{pmatrix} \tag{1.6}$$

with the obvious split of the residual vector $\underset{\sim}{e}$ into the first (n-K) and last K elements, it is clear that Q_{1m} corresponds to the square of the largest standardized residual in modulus; in fact the ith (i=1,2,...,n) Q_1 value is

$$Q_{1i} = e_i^2/(1-h_{ii}). \tag{1.7}$$

When K = 2, an interesting subtlety arises. We can always write

$$Q_{2m} = \frac{r_1^2}{1-h_{11}} + \frac{\{r_2^{(1)}\}^2}{1-h_{22}^{(1)}} \quad . \tag{1.8}$$

In this expression, r_1 denotes <u>one</u> of the residuals e_i whose location provides the largest Q_2 and $r_2^{(1)}$ denotes the residual that would occur in the second location that provides the largest Q_2, if the observation in the first location were dropped from the data. (Note that we can also write

$$Q_{2m} = \frac{\{r_1^{(2)}\}^2}{1-h_{11}^{(2)}} + \frac{r_2^2}{1-h_{22}} \tag{1.9}$$

since the relationship is perfectly symmetrical.) Now, if $r_1/(1-h_{11})^{1/2}$ <u>is</u> the largest in modulus standardized residual, then $r_2^{(1)}/(1-h_{22}^{(1)})^{1/2}$ is the largest in modulus (adjusted) standardized residual obtained after removal of the observation corresponding to r_1. (Similar remarks apply with subscripts 1 and 2 reversed.) However, neither $r_1/(1-h_{11})^{1/2}$ nor $r_2/(1-h_{22})^{1/2}$ need be the largest in modulus standardized residual! In our subsequent simulations, we nevertheless assume that one of them is, and so compute Q_{2m} in a "stepwise" fashion. This greatly simplifies the simulation procedure for K = 2, and is an adequate approximation in view of the fact that the correct and approximate simulation results appear to be identical well over 99% of the time, in practice. (To test this, we performed 3000 simulations calculating Q_{2m} both directly and in stepwise fashion. The results were identical in all except 8 cases.) Similar considerations would apply for $K \geq 3$.

The Q_{Km} and the corresponding residual mean square s_{Km}^2 provide the test statistic

$$F_{Km} = \{Q_{Km}/K\}/s^2_{Km}. \qquad (1.10)$$

The distribution of F_{Km} is non-standard and unknown. In previous work, percentage points have been generated for various specific cases of two-way tables. See John and Draper (1978) and Draper and John (1980) both for details, and for useful approximations to those percentage points for K = 1,2,3.

The present study focuses on the case of straight line regression and extends previous investigations. It examines critical values for the outlier test based on F_{Km}, approximations to those critical values, and also the power of these tests. The effects of various design configurations and of mis-specification of K on the power calculations is also examined. The study was triggered by analysis of data from an experiment on the relationship between the number of viable cells injected into the intestine of host rats and the number of γ-glutamyl transpeptidase colonies [GT+] formed in the liver lobes of those animals. The injected cell suspensions were prepared from donor rats livers subjected to a standard carginogen diet of 2-acetylaminofluorene (AAF) concurrent with a two third hepatectomy (PH). This AAF/PH regiment has been used extensively in recent years and is clearly toxic to the health of the test animals. The host animals were also subjected to an AAF/PH regiment and were sacrificed 10 days post the PH and the injection of the cells from the donor animals. This experiment is from an ongoing research study of the mechanism by which carcinogens induce liver cancer in experimental animals. For further details see Laishes and Rolfe (1980).

The data, shown in Table 1, were obtained on three different days denoted as A, B and C, with 13, 8, and 5 survivor host rats (not cannibalized)

Table 1. Average Counts of GT(+) Colonies in Two Standard Liver Sections and the Number of Viable Cells Injected ($\times 10^{-5}$)

	Case No.	y=GT(+)	Cells injected ($\times 10^{-5}$)
Experiment A	1	83.0	5
	2	59.5	5
	3	100.0	7
	4	92.5	7
	5	71.0	10
	6	112.5	10
	7	141.0	15
	8	56.0	3
	9	11.5	1
	10	9.5	1
	11	1.0	0
	12	0.0	0
	13	54.5	5
Experiment B	14	48.0	5
	15	131.0	10
	16	52.0	3
	17	32.5	3
	18	14.0	1
	19	6.5	1
	20	0.5	0
	21	64.0	5
Experiment C	22	136.0	10
	23	63.0	5
	24	9.5	1
	25	24.5	1
	26	0.0	0

respectively. Laishes and Rolfe (1980) found that a straight line regression model represents adequately the association between x = the number of viable liver cells injected into host rats and y = the number of GT(+) colonies observed on day ten post infection (R^2=0.90). When the test animals were autopsied, it was observed that two animals from Experiment A were obviously afflicted with a severe cholestatic disorder (exhibiting a yellowish, jaundiced liver)and one animal from Experiment C had received, through technical error, an incomplete PH (revealed by the presence of a portion of the median liver lobe). Thus, despite efforts to control animal health, a small percent were overly diseased at the time of the sampling. The question aroused in this experiment is whether regression diagnostics are able to detect the diseased animals as outliers. The concern is that, in other experiments, disease states that could influence liver colony developments might be overlooked. Additionally, it is also desirable to develop tools for detecting animals afflicted by subclinical disease states, which can only be revealed by complete pathological and microbiological work-ups.

The design from the above mentioned experiment served as an initial design in Monte-Carlo simulations intended to investigate the statistics F_{1m} and F_{2m} under various conditions in the linear regression model. (A brief analysis of the original data is given in Section 5.)

2. DESIGNS CHOSEN FOR SIMULATIONS

Initially we looked at 15 different designs comprising various numbers of points at the x-locations of the original data. From these 15 we selected the four shown in Table 2 for our investigation. In this table, the letters x, f denote the sites and frequency of points at that site, respectively. The h values are the corresponding values of the diagonal of the $\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$ matrix. We also record the Σh_{ii}^2 values at the foot of each column. Smaller values of Σh_{ii}^2 denote designs "more robust" to outliers, as described by Box and Draper (1975). The table also shows the locations at which one outlier will be added, at which pairs of outliers will be added and the corresponding values of $\underset{\sim}{H}_{22}$ and $|\underset{\sim}{I}-\underset{\sim}{H}_{22}|$. $\underset{\sim}{H}_{22}$ is the part of the $\underset{\sim}{H}$ matrix that corresponds to the two sites and $|\underset{\sim}{I}-\underset{\sim}{H}_{22}|$ is a spatial measure of the positions of these sites, lower values indicating more "remoteness" from the rest of the data. (See Draper and John, 1981.)

(Note: the words "cases" and "locations" have different meanings. "Cases" refer to specific animals in Table 1. "Locations" refer to x-sites counted off in the various designs. For example, location 15 is at x = 5 in Design 1, x = 7 in Design 2, x = 15 in Design 3, and x = 0 in Design 4.)

The four designs used were selected to achieve a representative range of the characteristics that occurred, within limits. For example, the fifteen Σh_{ii}^2 values ranged from 0.1537 to 0.4755 (apart from two designs with 25 points on one end of the x-range and one point at the other, for which

Table 2. Designs Used for Simulation Studies and Locations of Added Outliers.

	Design Number							
	1		2		3		4	
x	f	h	f	h	f	h	f	h
0	4	0.0858	8	0.0809	13	0.0769	20	0.0457
1	6	0.0667	2	0.0691	0	-	1	0.0394
3	3	0.0432	2	0.0511	0	-	1	0.0443
5	6	0.0394	2	0.0410	0	-	1	0.0725
7	2	0.0553	2	0.0386	0	-	1	0.1239
10	4	0.1161	2	0.0496	0	-	1	0.2445
15	1	0.3159	8	0.1067	13	0.0769	1	0.5617
Σh_i^2		0.2308		0.1695		0.1537		0.4412
One outlier at locations		5,25,26		14,15		1,14		21,25
where x =		1,10,15		5,7		0,15		1,10
Pair of outliers at locations		(25,25)		(14,15)		(1,2)		(25,25)
with H_{22} =		$\begin{bmatrix} .116 & .185 \\ .185 & .316 \end{bmatrix}$		$\begin{bmatrix} .041 & .038 \\ .038 & .039 \end{bmatrix}$		$\begin{bmatrix} .077 & .077 \\ .077 & .077 \end{bmatrix}$		$\begin{bmatrix} .039 & .024 \\ .024 & .244 \end{bmatrix}$
and $\lvert I-H_{22}\rvert$ =		.570		.921		.846		.725
Pairs of outliers at locations		(5,25)				(1,14)		
with = H_2		$\begin{bmatrix} .067 & -.008 \\ -.008 & .116 \end{bmatrix}$				$\begin{bmatrix} .077 & 0 \\ 0 & .077 \end{bmatrix}$		
and $\lvert I-H_{22}\rvert$ =		.825				.852		

Σh_{ii}^2 took the uncharacteristic value 1.04). The aim was to enable the assessment of the effects of several factors on the empirical power of the tests, specifically these:

(a) the use of tests based on F_{1m} and on F_{2m} both in the presence of a single outlier and when two outliers are present.

(b) the h-value (and the appropriate λ_1 value) corresponding to each outlier site.

(c) the values of the elements in the $\underset{\sim}{H}_{22}$ matrix corresponding to the two outliers sites.

(d) the $|I-\underset{\sim}{H}_{22}|$ corresponding to $\underset{\sim}{H}_{22}$ from (c).

(e) the λ_2 value (in the case of two outliers).

Obviously, by definition, the values of the λ's are functions of the appropriate h and θ values; see Eq. (1.3). The λ's are used here as general measures of departure from the null hypothesis and not in connection with any (inappropriate in the present context) non-central F-distribution.

3. CRITICAL VALUES FOR F_{Km} IN STRAIGHT LINE REGRESSION MODELS.

John and Draper (1978) and Draper and John (1980) suggest the following sequential strategy for the detection of the number of outliers and of their location in a two-way ANOVA table with one observation per cell:

(a) Determine K, the maximum reasonable number of outliers in the data.

(b) Test H_0: no outliers versus H_1: there are 1 or 2, ... or K outliers, by comparing F_{Km} with the appropriate critical value of the null distribution of F_{Km}.

(c) If H_0 is rejected, the y value associated with the standardized residual with maximum modulus value is declared an outlier and deleted, and we return to (b) with K replaced by K - 1. The algorithm stops when H_0 is not rejected. For K = 1, John and Draper (1978) suggest the use of the conservative critical value derived from the Bonferonni inequality $F_{1,n-p-K}(\alpha/n)$, where $F_{f_1,f_2}(c) = P(F_{f_1,f_2} > c)$ is the upper tail of a central F-variate with (f_1,f_2) degrees of freedom. For $K \geq 2$ (actually, for K = 2,3) Draper and John (1980) found that for testing H_0 at a specified level α, good approximations to the critical values are obtained by setting:

$$m = \frac{3}{4}n \left(1 - \frac{(K+1)n}{1600}\right) \tag{3.1}$$

in Andrews' (1971) formula

$$\alpha = \frac{1}{K!} \prod_{i=0}^{K-1} (m-1) F_{K,n-p-K}(F_{Km}) - \frac{m-K}{(K-1)!} \prod_{i=0}^{K-2} (m-i) F_{K-1,n-p-K+1}\left[\frac{K}{K-1} \frac{n-p-K+1}{n-p-K} F_{Km}\right] \tag{3.2}$$

which, in its original context, provided a bound on the probability of obtaining K extreme residuals.

We now proceed to assess empirically the critical values for the tests based on F_{1m} and F_{2m} in a straight line regression setting. The critical values may depend on the design configuration, namely on both the sample size and on the x-values. We first evaluate for fixed sample size the effect of the design on the critical values. Obviously it would be highly desirable if the critical values are relatively constant and thus do not have to be regenerated individually for each design. The effect of the design on the null distribution of F_{1m} and F_{2m} was evaluated for the four designs described in Section 2. The empirical null distribution of F_{1m} and F_{2m} was generated using 3000 samples for each design. The upper 10%, 5%, 2.5% and 1% of the cumulative distributions of F_{1m} and F_{2m} can be found in Table 3. The table also records the values $F^{-1}_{K,n-p-K}\,[\alpha/\binom{n}{K}]$, K = 1,2, derived using the Bonferonni inequality.

First note that the four designs yielded very similar empirical critical values for both F_{1m} and F_{2m}.

The empirical percentiles of F_{1m} are very similar to the $100(\alpha/n)$ percentiles of the $F_{1,n-p-1}$ distribution. In most cases, the empirical percentiles slightly exceeded the theoretical upper bounds $F^{-1}_{1,n-p-1}(\alpha/n)$ based on the Bonferonni inequality. The Bonferonni critical values for F_{1m} are clearly very tight upper bounds and their use is recommended for all regression designs. This concurs with Draper and John's procedure for the detection of a single outlier in a two-way ANOVA design.

Table 3. Empirical Percentiles of F_{1m} and F_{2m} for the Four Designs Presented in Table 2. The Values Listed in Parentheses are the Relative Frequencies by which F_{Km} Exceeded $F_{K,n-p-K}\{\alpha/\binom{n}{K}\}$, K = 1,2.

K = 1

α	DESIGN 1	DESIGN 2	DESIGN 3	DESIGN 4	$F_{K,n-p-K}\{\alpha/\binom{n}{K}\}$
.10	10.33(.100)	10.42(.104)	10.35(.101)	10.59(.110)	10.334
.05	12.48(.053)	12.47(.052)	12.59(.056)	12.42(.051)	12.257
.025	14.59(.027)	14.59(.027)	14.59(.027)	14.25(.025)	14.315
.01	16.97(.009)	17.32(.011)	17.76(.012)	18.06(.012)	17.246

K = 2

.10	9.90(.043)	10.05(.049)	10.17(.043)	10.16(.050)	11.943
.05	11.51(.023)	11.86(.025)	11.66(.022)	11.96(.026)	13.435
.025	13.19(.012)	13.45(.012)	13.08(.012)	13.46(.013)	15.025
.01	15.33(.005)	15.57(.006)	15.48(.006)	16.11(.006)	17.285

The comparison of the percentiles of the F_{2m} distribution for various designs is more relevant. The 95th percentile of the null empirical distribution of F_{2m} in the original design (11.507) corresponds to the 95.7, 95.4 and 95.8-th percentiles of the empirical distributions of F_{2m} generated under the designs 2, 3 and 4. The differences are small at other percentiles as well. Thus it appears that the critical values generated under the original design (either by simulation or computed from an approximation formula) can be safely used with other designs, a reassuring result. In the power study described in the next section, we use the critical values obtained from the null distribution of F_{2m} under the original design. Note that for α = 0.1, 0.05, 0.025 and 0.01, $F^{-1}_{2,n-p-2}\,[\alpha/\binom{n}{2}]$ corresponds to approximately half of the nominal α.

Using the x-values of the original design (Table 1), 3000 samples were now generated to provide the null distribution of F_{2m} for 1, 2, 3, 4, 5, and 6 y-values at each of the 26 x-values. This enabled us to list the empirical percentiles of F_{2m} for various n's and α's and to develop an approximating formula for the critical values. Following John and Draper (1978), we do that by estimating Andrews' parameter m in (3.2) with the formula set at prespecified probability values. Specifically for each value of $n = 26j$, $j = 1,\ldots,6$, let $F^{(t)}_{2m}$ be the upper 100t-th percentile of the empirical cumulative distribution of F_{2m} and let $\alpha_t = 1-t/3000$, $t = 1,\ldots,3000$. Denote the resulting solution of (3.2) by m_t. We thus obtained 3000 × 6 triplets (m_t,α_t,n). The extreme 1%, in both tails of the six empirical

distributions of F_{2m} were deleted due to their higher instability. The equation $m/n = a + bn$ was fitted to the remaining 17,640 points by ordinary least squares giving $m/n = 0.59 - 0.0015n$ with $R^2 = 0.26$. The fit is obviously not satisfactory. The addition of an α-term to the regression equation yielded $m/n = 0.778 - 0.378n - 0.0015n$ with $R^2 = 0.94$, a definite improvement. (Obviously, since the vector of α's is orthogonal to the vector of n's the coefficent of n remains unchanged when the α-term is added to the equation.) For $\alpha = 0.05$ the relationship is $m = 0.76n(1-0.0015n)$ or roughly $m = \frac{3}{4}n(1 - \frac{3n}{2000})$.

We thus conclude that, in a straight line regression, the critical values for F_{2m} at $\alpha = 0.05$ can be obtained by substituting in (3.2) $m = \frac{3}{4}n(1 - \frac{3n}{2000})$. Note the great similarity to the John and Draper (1978) approximation for the two-way table investigation $m = \frac{3}{4}n(1 - \frac{3n}{1600})$. However, if the critical values are sought at other α's the effect of α on the value of m cannot be ignored and the critical values should be obtained by substituting in (3.2) the value $m = n(0.778-0.378\alpha-0.0015n)$. Based on the results from the comparisons of the four designs and on the similarity of the approximating formula with the one obtained for two-way designs, we speculate that these approximations are valid over a wide range of designs.

4. POWER OF THE TESTS BASED ON THE STATISTICS F_{1m} AND F_{2m}.

We now turn to investigate the behaviour of the tests based on F_{1m} and F_{2m} in the presence of one and two outliers. Specifically, for each of the four designs and the locations of the assumed outliers given in Table 2, we generated 3000 samples according to model (1.1). For convenience we refer to a specific design together with the sites at which the outliers are located as a configuration. In all, we have evaluated 15 configurations. The outliers were of size $\pm 3\sigma$ and $\pm 5\sigma$. The empirical power is defined as the percent out of 3000 samples that the statistic F_{Km} exceeds its $\alpha = 0.05$ critical value. Note that the observation(s) thus identified as outlier(s) may or may not be the actual outlier(s) (although typically they would be).

4.1 Simulation results, K = 1.

In the case of a single outlier the simulation study addresses the following issues:

(a) How the power of the test based on F_{1m} varies with the leverage at the outlier's site and with the overall measure of robustness Σh_{ii}^2.

(b) How the power of the test based on F_{2m} behaves when only a single outlier is present.

The upper panel of Table 4 compares two configurations with almost equal h-values at the outlier's site but with very different Σh_{ii}^2's. We observe that, for equal outlier's size, there is little variation in power between the two configurations.

Table 4. Empirical Power Results (×100) with K = 1.

	DESIGN 2; Outlier at site 14			DESIGN 4; Outlier at site 21		
θ	$\lambda\sigma^2$	F_{1m}	F_{2m}	$\lambda\sigma^2$	F_{1m}	F_{2m}
3σ	4.31	32.7	30.5	4.32	34.5	31.0
5σ	11.99	90.1	86.6	12.01	89.8	85.8
-3σ	4.31	34.2	32.4	4.32	32.5	28.6
-5σ	11.99	90.1	86.4	12.01	90.7	86.6
h - value at outlier's site		.041			.039	
Σh_{ii}^2		.169			.441	

	DESIGN 1; Outlier at site 26			DESIGN 4; Outlier at site 25		
θ	$\lambda\sigma^2$	F_{1m}	F_{2m}	$\lambda\sigma^2$	F_{1m}	F_{2m}
3σ	3.08	23.7	21.8	3.40	24.9	22.6
5σ	8.55	73.4	68.3	9.45	78.9	73.5
-3σ	3.08	22.3	19.6	3.40	24.7	22.6
-5σ	8.55	73.9	69.3	9.40	78.3	73.6
h - value at outlier's site		.316			.244	
Σh_{ii}^2		.231			.441	

The lower panel of Table 4 compares two configurations with different h-values at the outlier's site and different (and reversed in direction) Σh_{ii}^2's. Powers are lower than above (because of higher h-values) and increase as expected with decreasing h values, and this effect is not reversed by a lower Σh_{ii}^2 value in the first of the two configurations.

The implication from the whole of Table 4 is that power does not change with Σh_{ii}^2, but increases as the h-value of the outlier site decreases.

In all four configurations in Table 4, the use of the test based on F_{2m} (instead of F_{1m}) resulted in a decrease of power.

4.2 Simulation results, K = 2.

In configurations with two outliers, let θ_1 and θ_2 be their respective sizes, so that $\underset{\sim}{\theta} = (\theta_1,\theta_2)'$. The following issues are addressed in the simulation study with K = 2:

(a) How the power varies with the $(\underset{\sim}{H}_{22},|\underset{\sim}{I}-\underset{\sim}{H}_{22}|,\lambda_2)$ which are related to the outliers' locations and with the overall measure of robustness Σh_{ii}^2.

(b) How the power varies with the relative position of the outliers.

(c) How the power of the test based on F_{1m} is affected by the presence and the position of the second outlier. We note that the performance of F_{1m} when one than a single outlier is present may be of interest in the cases when one does not recognize the presence of a second outlier and/or when the test is performed in a "stepwise" fashion (see, e.g., Anscombe, 1960).

In Table 5 we compare the empirical power of the tests in two configurations. The first configuration has both a small $|\underset{\sim}{I}-\underset{\sim}{H}_{22}|$ value at the outliers' locations and a smaller Σh_{ii}^2. From Table 5 we observe:

(a) In general, the power of the tests increases monotonically with λ_2. (Note that, unlike λ_1, λ_2 is not a monotonic function of $|\underset{\sim}{I}-\underset{\sim}{H}_{22}|$.)

(b) Neither the relative position of the two outliers (measured by h_{12}) nor the value of Σh_{ii}^2 appear to affect the power of the test based on F_{2m}.

(c) When the two outliers have equal signs, the test based on F_{1m} has a smaller power than the one obtained by F_{2m}.

(d) The magnitude of the loss of power due to the use of F_{1m} (instead of F_{2m}) depends on h_{12}. For a large h_{12} value and sign (θ_1) = sign (θ_2), the decrease in power may be considerable and may reverse the monotonicity with λ_2. Table 6 presents several additional configurations from all four designs to further illustrate this point.

Tables 7 and 8 present some comparisons of power achieved by F_{1m} when, in the same design (a) K = 1, versus (b) K = 2 with $|\theta_1| = |\theta_2|$. The assumed outlier in (a) is included in the pair of outliers in (b). The configurations in Table 7 have similar and relatively small h_{12}-values. We observe that, when the h-value of the second outlier is small (Design 2, site 15) and sign (θ_1) = sign (θ_2), the power obtained when K = 2 is smaller than when K = 1. When the two outliers have opposite signs, the two configurations yield similar power. When the h-value of the second outlier is large (Design 4, site 25) the power when K = 2 is in general smaller than when K = 1. Again, the power decreases when the two outliers have the same signs.

In Table 8 we investigate the effect of the relative position of the outliers. When the two outliers are at the extremes of the x-range (h_{12}=0) the power of the F_{1m} test is generally smaller than when K = 1 and is not

Table 5. Empirical Power Results (×100) when K = 2.

	DESIGN 1; Outliers at sites (25,26)			DESIGN 4; Outliers at sites (21, 25)		
(θ_1,θ_2)	$\lambda_2\sigma^2$	F_{1m}	F_{2m}	$\lambda_2\sigma^2$	F_{1m}	F_{2m}
$(3\sigma,3\sigma)$	5.389	11.8	20.3	7.503	27.5	34.5
$(3\sigma,5\sigma)$	9.751	34.9	50.5	13.401	63.2	74.6
$(3\sigma,-3\sigma)$	8.723	52.8	50.8	7.941	32.0	39.5
$(3\sigma,-5\sigma)$	15.306	86.5	84.9	14.130	67.1	75.4
$(5\sigma,3\sigma)$	11.350	59.3	64.4	15.042	80.9	83.3
$(5\sigma,5\sigma)$	14.971	30.9	76.6	20.843	74.2	93.3
$(5\sigma,-3\sigma)$	16.906	92.1	90.5	15.771	84.6	86.9
$(5\sigma,-5\sigma)$	24.230	97.8	98.2	22.058	85.0	94.7
$\underset{\sim}{H}_{22}$ at outliers' sites		$\begin{bmatrix} .116 & .185 \\ .185 & .316 \end{bmatrix}$			$\begin{bmatrix} .039 & .024 \\ .024 & .244 \end{bmatrix}$	
$\lvert \underset{\sim}{I} - \underset{\sim}{H}_{22} \rvert$		.570			.725	
Σh_{ii}^2		.231			.441	

Table 6. Empirical Power (×100) Using F_{1m}, as a Function of λ_2 and h_{12}.

Design	Outliers' sites	h_{12}	(θ_1,θ_2)	$\lambda_2\sigma^2$	Power
1	(25,26)	0.185	$(5\sigma,-3\sigma)$	16.906	92.1
			$(5\sigma,5\sigma)$	14.971	30.9
1	(5,25)	-0.008	$(5\sigma,-3\sigma)$	15.519	77.9
			$(5\sigma,5\sigma)$	22.921	83.8
3	(1,2)	0.077	$(5\sigma,-3\sigma)$	16.845	86.1
			$(5\sigma,5\sigma)$	21.154	66.2
3	(1,14)	0	$(5\sigma,-3\sigma)$	15.692	78.8
			$(5\sigma,5\sigma)$	23.076	82.3
4	(21,25)	0.024	$(5\sigma,-3\sigma)$	15.771	84.6
			$(5\sigma,5\sigma)$	20.843	74.2

Table 7. Comparison of Power (×100) of the Test Based on F_{1m} When K = 1 (A Single Outlier of Size θ) Versus K = 2 (Outliers of Sizes θ_1, θ_2).

		DESIGN 2		DESIGN 4	
θ	θ_1, θ_2	K = 1 at site 14	K = 2 at sites (14,15)	K = 1 at site 21	K = 2 at sites (21,25)
3σ	3σ,3σ	32.7	27.6	34.5	27.5
	3σ,-3σ		36.7		32.0
5σ	5σ,5σ	90.1	76.2	89.8	74.2
	5σ,-5σ		89.6		85.0
-3σ	-3σ,-3σ	34.2	27.7	32.5	28.4
	-3σ,3σ		36.9		32.9
-5σ	-5σ,-5σ	90.1	77.4	90.7	76.8
	-5σ,5σ		89.2		84.4
H_{22} at outliers' sites		0.041	$\begin{bmatrix} 0.041 & 0.038 \\ 0.038 & 0.039 \end{bmatrix}$	0.039	$\begin{bmatrix} 0.039 & 0.024 \\ 0.024 & 0.244 \end{bmatrix}$
Σh_{ii}^2		0.169		0.441	

Table 8. A Further Comparison of Power (×100) of the Test Based on F_{1m} When K = 1 Versus K = 2.

		DESIGN 3		
θ	θ_1, θ_2	K = 1 at site 1	K = 2 at sites (1,2)	K = 2 at sites (1,14)
3σ	$3\sigma, 3\sigma$	31.3	22.4	32.3
	$3\sigma, -3\sigma$		42.2	32.5
5σ	$5\sigma, 5\sigma$	88.7	66.2	82.3
	$5\sigma, -5\sigma$		92.8	82.6
-3σ	$-3\sigma, -3\sigma$	30.8	24.6	31.8
	$-3\sigma, 3\sigma$		43.3	31.8
-5σ	$-5\sigma, -5\sigma$	89.5	65.6	83.4
	$-5\sigma, 5\sigma$		93.4	81.1
H_{22}	at outliers' sites	0.077	$\begin{bmatrix} 0.077 & 0.077 \\ 0.077 & 0.077 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0 \\ 0 & 0.077 \end{bmatrix}$
Σh_{ii}^2			0.154	

affected by the signs of the θ's. However, when the two outliers are clustered (h_{12}= 0.077) and sign (θ_1) $\neq$ sign (θ_2), the power is larger when K = 2 than when K = 1. The opposite is true when sign (θ_1) = sign (θ_2).

5. AN EXAMPLE

A brief diagnostic analysis based on the statistics F_{Km} is now performed on the data from Table 1. No attempt is made to carry out a complete analysis of this set of data here but only to present an application of the use of the F_{Km} statistics. For further analyses see Laishes and Rolfe (1980) and Fuchs (1980). The fitted equation was $\hat{y} = 6.38 + 10.59x$ with $R^2 = 0.89$. The plot of y versus x indicates no obvious deviation from the fitted model. When we attempt a stepwise deletion of outliers (or, equivalently, assume at first that no more than one outlier is present), case 12 is detected as an outlier. After the deletion of case 12, case 13 is detected next. No further outliers were detected. We note that cases 12 and 13 correspond to the two jaundiced animals.

When a simultaneous detection of a pair of outliers was attempted, Q_{2m} selected the cases 12 and 13 as potential outliers and, in the subsequent testing procedure, both were labelled as outliers. When K = 3 was postulated, Q_{3m} selected cases 6, 12 and 13 as potential outliers but subsequent analyses identified only cases 12 and 13 as outliers. Thus all tests detected both the two jaundiced animals, and only these.

Next we performed the diagnostic analysis on the data from each of the three days (A, B and C) separately. No outliers were detected, not even for experiment A which included the two jaundiced animals. The reason for this is that the design in Experiment A is very "non-robust". The two jaundiced cases happen to have extreme x-values (x = 10 and x = 15, respectively). This alters considerably the ability to detect them as outliers in data set

A alone. When all data are combined however, three more observations with $x = 10$ are recorded and the fact that case 12 is an outlier becomes obvious, which then leads to the detection of the second jaundiced animal.

6. CONCLUSIONS

The value of Q_{Km} as an "outliers" statistic has been established for several years. In this article, we examine the distributional properties of the related F_{Km} statistic for the case of a straight line regression model and for K = 1,2. We have also extended previous investigations of F_{Km} by carrying out power calculations for various designs.

We found that a correct determination of the K in advance has considerable influence on the power of the test. When two outliers were present, the test based on F_{1m} performed more poorly than that based on F_{2m}. The loss in power is especially large when the two outliers are close to each other with the same sign. A decrease in power also results from the use of F_{2m} when only one outlier is present.

The recommendation mentioned by both Box and Draper (1975) and Draper and John (1980) that it would seem desirable to choose the experimental design so that Σh_{ii}^2 is as small as possible is valid when one expects <u>random</u> outliers at unknown positions. Here, however, the experimental design appears to affect the power through the leverage at the outliers' sites. The practical implication is that, if the experimenter has some prior knowledge about the experimental sites which are prone to outliers (as is the case in carcinogenic studies at high dosages) it may be wise to decrease the h-values at those sites even at the expense of overall robustness.

Our final comment concerns the formulas found for approximating the generated percentage points of F_{2m} for the straight line situation. Previously, Draper and John (1980) remarked that, for two way tables, the

approximating formula $m = \frac{3}{4}n\{1-(K+1)n/1600\}$ worked well for both $K = 2$ and 3. Here, for the case $K = 2$ and $\alpha = 0.05$, a very similar formula emerged, namely $m = \frac{3}{4}n\{1-(K+1)n/2000\}$. This leads us to speculate that, at least for $\alpha = 0.05$, either of these formulas (which differ very little for moderate n) would provide an adequate method for obtaining critical test values in a wide variety of design circumstances. For other α values, a more general formula is offered.

ACKNOWLEDGEMENTS

N. R. Draper was partially sponsored by the United States Army under Contracts Nos DAAG29-80-C-0041 and DAAG29-80-C-0113.

We are grateful to Sarah Hellman for related computations and helpful discussions.

REFERENCES

ANDREWS, D. F. (1971), "Significance Tests Based on Residuals", Biometrika, 58, 139-148.

ANSCOMBE, F. J. (1960), "Rejection of Outliers", Technometrics, 2, 123-147.

BOX, G. E. P. and DRAPER, N. R. (1975), "Robust Designs", Biometrika, 62, 347-352.

COOK, D. R. (1979), "Influential Observations in Linear Regression", Journal of the American Statistical Association, 74, 169-174.

DRAPER, N. R. and JOHN, J. A. (1980), "Testing for Three or Fewer Outliers in Two-Way Tables", Technometrics, 22, 9-15.

DRAPER, N. R. and JOHN, J. A. (1981), "Influential Observations and Outliers in Regression", Technometrics, 23, 21-26.

DRAPER, N. R. and SMITH, H. (1981), Applied Regression Analysis, New York: Wiley.

ELLENBERG, J. H. (1976), "Testing for a Single Outlier from a General Linear Regression", Biometrics, 32, 637-645.

FUCHS, C. (1980), "Detection of Damaged Animals in Carcinogeneity Studies with Laboratory Animals", University of Wisconsin Statistical Laboratory Report 80/3.

GENTLEMAN, J. F. and WILK, M. B. (1975a), "Detecting Outliers in a Two-Way Table: I. Statistical Behavior of Residuals", Technometrics, 17, 1-14.

GENTLEMAN, J. F. and WILK, M. B. (1975b), "Detecting Outliers: II. Supplementing the Direct Analysis of Residuals", Biometrics, 31, 387-410.

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

LAISHES, B.A. and ROLFE, P. B. (1980), "Quantitative Assessment of Liver Colony Formation and Hepatocellular Carcinoma Incidence in Rat Livers Receiving Intravenous Injections of Isogenic Liver Cells Isolated During Hepatocarcinogenesis", Cancer Research, 40, 4133-4143.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 2436	2. GOVT ACCESSION NO.: AD-A124 369	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): PERCENTAGE POINTS AND POWER CALCULATIONS FOR OUTLIER TESTS IN A REGRESSION SITUATION		5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Camil Fuchs and Norman R. Draper		8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0041 & DAAG29-80-C-0113
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709		12. REPORT DATE: October 1982
		13. NUMBER OF PAGES: 29
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report): UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Detection of outliers; Q_K statistics; Residuals.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The usefulness of an extra sum of squares statistic Q_K for detecting K outliers has been discussed previously in the context of two-way tables. (See Gentleman and Wilk, 1975a, 1975b; John and Draper, 1978; and Draper and John, 1980.) That work is extended here to straight line regression

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (cont.)

situations arising from and motivated by a specific set of research data. Percentage points for the appropriate test statistics are obtained by simulation, approximations for these percentage points are suggested, and power calculations are made for various designs and outlier situations. Correct determination of K and position(s) of the outlier(s) appear to be important in influencing power.