ROBUST METHODS FOR FACTORIAL EXPERIMENTS WITH OUTLIERS

Raymond J. Carroll*

University of North Carolina at Chapel Hill

Abstract

Two factorial experiments with possible outliers (John (1978)) are reanalyzed by means of robust regression techniques.

Key Words: Factorial experiment, outliers, robustness, M-estimates

---

*This research was supported by the Air Force Office of Scientific Research under Grant AFOSR-75-2796.

1. Introduction

In a recent paper, John (1978) discussed the effects and detection of outliers in factorial experiments. His methods were based on least squares methodology and can be quickly summarized as follows. Consider the usual linear model based on n observations:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad (\text{rank } (X) = p) \ .$$

If one suspects that m outliers are present, then the model could be written as

$$\underline{Y} = X\underline{\beta} + \sum_{i=1}^{m} \theta_i \underline{d}_i + \varepsilon \ ,$$

where $\underline{d}_i$ is a vector with 1 in the row corresponding to the i-th suspected outlier and 0 elsewhere (i = 1,2,...,m). The presence of outliers would then be tested by H_0: $\underline{\theta} = 0$. However, the percentage point F_α which one would compare with the F-statistic F* should not be the usual upper α percentage point of the F distribution since the observations being tested have the most extreme residuals. John however discusses a simple modification when there is only one suspected outlier (m = 1 -- use the upper α/n percentage point of the F-distribution). For two suspected outliers (m = 2), a deeper analysis is required.

As discussed in the book by Barnett and Lewis (1978), a second method for dealing with outliers in dependent variables involves the techniques of robustness pioneered by Huber (1964, 1973, 1977). His class of M-estimates has been specifically designed to be insensitive to outliers and to retain high efficiency when the errors are heavier-tailed than the normal, two properties not possessed by least squares. The basics of M-estimates are reviewed in Section 2. In Section 3 we apply these methods to John's first example, while in Section 4 we discuss John's second example. We find in both cases that one pass through a robust regression program based on M-estimates yields results similar to those obtained by John.

Daniel and Wood (1971) and Andrews (1974) analyze a data set in a regression context. The former use least squares and decide (after detailed analysis) that there are three outliers. Andrews uses robust techniques similar to those presented here and shows that the decision favoring three outliers can be reached in an easier and somewhat more routine fashion. Thus, the advantages of robust techniques are not limited to the factorial experiments analyzed herein.

The analyses presented suggest that robust regression can be a valuable tool for statisticians concerned with the possibility of outliers or data which are not normally distributed. Depending on one's computing capabilities, this tool can be more or less routinely used in statistical analyses, although the advisability of routine use by even the most naive users is a matter of some debate (see the discussion of Bickel's (1976) paper).

2. M-estimates of Regression

Least squares estimates minimize

(2.1) $$\sum_{i=1}^{n} \rho((y_i - \underline{x}_i \beta)/\sigma) \; ,$$

where $\rho(x) = \frac{1}{2}x^2$. The quadratic form of ρ is what makes least squares sensitive to outliers. This can also be seen if one defines $\psi = \rho'$, for then one solves

$$(2.2) \qquad \sum_{i=1}^{n} \psi((y_i - \underline{x}_i\underline{\beta})/\sigma)\underline{x}_i = 0 \ ,$$

where for least squares $\psi(z) = z$. In order to achieve robustness against outliers and high efficiency for distributions heavier-tailed than the normal, Huber (1964), Andrews et al. (1972) and Hampel (1974) suggest that ψ be a bounded function, and that scale be estimated in one of two ways:

(I) (Commonly called Huber's Proposal 2) Simultaneously solve (2.2) and

$$(2.3) \qquad (n-p)^{-1}\sum\psi^2((y_i - x_i\beta)/\sigma) = E_\Phi\psi^2(Z) \ ,$$

the expectation taken under the standard normal distribution.

(II) (Commonly called MAD) Solve (2.2) with

$$(2.4) \qquad \hat{\sigma} = (\text{median absolute residual from median})/.6745 \ .$$

This is asymptotically equal to one for the normal model.

In both cases, the solution is found iteratively. One chooses a starting value for σ, solves (2.2), then updates σ by (2.3) or (2.4), etc., continuing until convergence. Algorithms are available in Huber (1973, 1977) and Dutter (1976); the author has adapted these algorithms for use in the SAS computer programs (a card deck is available upon request). In neither case is the computation burdensome.

Some typical choices of ψ (see also Gross (1976)) are

<u>Huber's</u> $\psi(x) = -\psi(-x) = \max(-k,\min(x,k))$, with k often taken as 1.5 or 2.0.

<u>Hampel's</u>

$$\psi(x) = -\psi(-x) = \begin{cases} x & 0 \le x \le a \\ a & a \le x \le b \\ a\frac{(c-x)}{(c-b)} & b \le x \le c \\ 0 & x > c \, . \end{cases}$$

<u>Andrews'</u>

$$\psi(x) = -\psi(-x) = \begin{cases} \text{sine}(x/c) & 0 \le x \le c\pi \\ 0 & x \ge c\pi \, . \end{cases}$$

The constant c is sometimes taken as 2.1.

The fact that Hampel's ψ and Andrews' ψ both redescend to zero suggests (Hampel (1974)) that they give no weight to gross outliers, while Huber's ψ will give some weight to these outliers but not nearly so much as least squares. The redescending ψ functions (unlike Huber's ψ) can have problems with convergence; for this reason we adopt the convention of first estimating $\underline{\beta}$ by Huber's method and then using at most two iterations of the algorithm for Hampel's and Andrews' methods. In all these cases, under proper conditions, the robust regression estimate $\hat{\beta}_R$ of $\underline{\beta}$ is asymptotically normally distributed with mean $\underline{\beta}$ and covariance matrix which can be estimated by

$$(2.5) \qquad ((n-p)^{-1}\textstyle\sum\psi^2(r_i))(n^{-1}\textstyle\sum\psi'(r_i))^{-2}\,\hat{\sigma}^2(X'X)^{-1} \, ,$$

where the standardized residuals are

$$r_i = (y_i - \underline{x}_i\hat{\underline{\beta}})/\hat{\sigma} \ .$$

Huber (1973) and Andrews *et al*. (1972) show in simulation experiments that the estimates $\hat{\beta}_R$ are generally more efficient than the least squares estimates; they are only slightly more variable than least squares for the normal model but are considerably less variable for heavier-tailed models.

Inference about the parameters can take at least two forms, both based on the approximation (2.4). Schrader and Hettmansperger (1979) suggest an analysis of the usual drop in sum of squares statistic using $\sum\rho(r_i)$. Bickel (1976, discussion section) suggests the approach used here. Let

$$\lambda = n^{-1} \sum_{i=1}^{n} \psi'(r_i)$$

$$\eta = 1 + (p/n)(1-\lambda)/\lambda \ .$$

The term η is suggested by Huber (1973, equation 7.20) as a variance inflation factor for (2.4) if p/n is not small. Define *pseudo-values*

$$\tilde{Y}_i = \underline{x}_i\hat{\beta}_R + \eta\hat{\sigma}\psi(r_i)/\lambda \ .$$

Then, the least squares estimates for the model $\underline{\tilde{Y}} = X\underline{\beta} + \underline{\varepsilon}$ are $\hat{\beta}_R$ (this follows from (2.2)). Bickel suggests that asymptotically correct tests can be obtained by defining the pseudo-values and using them in conventional least squares packages. Schrader and Hettmansperger (1979) note that this approach may have some difficulties if the error distributions are very heavy tailed.

In the examples below, we used $\hat{\sigma}$ given by (2.3); for Huber's ψ we took $k = 1.5$, for Hampel's $a = 1.5$, $b = 3.5$, $c = 8.0$, while for Andrews', $c = 2.1$.

3. First Example

John's first example is a 3^{4-1} fractional replication of a 3^4 experiment. The effects of each factor are split into linear and quadratic components (AL, AQ, BL, BQ, CL, CQ, DL, DQ) and three interactions are formed by multiplication (ALBL, ALCL, BLCL). Observation 11 is a suspected outlier; the predicted values and residuals for the four methods are given in Table 1. The obvious conclusion from Table 1 is that the Hampel and Andrews methods are particularly robust in that their predicted value for observation 11

(i) is close to John's refitted value;

(ii) hardly changes when the original observation $y = 14$ is replaced by the refitted value $y = 62.33$.

In Table 2 we present significance levels for the effects using the original observation and then using the refitted observation 11. In Table 3 we present parameter estimates and standard errors for least squares (original and modified data) and Andrews. The standard errors for Andrews estimate using the modified data are all within 2% of those given from an analysis of the original data, while the parameter estimates themselves do not change significantly. Thus

(i) the Hampel and Andrews methods applied to the original data give analyses similar to John's.

(ii) The Hampel and Andrews methods do not change to any large extent after observation 11 is modified.

We conclude that for the 3^{4-1} example, the robust regression methods compare favorably with John's method. As designed, the Hampel and Andrews estimates are particularly insensitive to the single gross outlier. The Huber estimates, while preferable to the ordinary least squares estimates, still show some sensitivity to the outlier.

4. Second Example

The second example John uses to illustrate his method is a confounded 2^5 experiment, the block effects confounding the highest order interaction. After analysis he concludes that two suspected outliers are not really outliers and should not be refitted. In Table 3 we present significance levels for the four tests.

The major difference between least squares and the robust estimates exhibited in Table 3 is that the latter show a main effect in B significant at the .05 level, while for least squares the significance level is approximately .13. Although John's analysis suggests that there may well be no *gross* outliers, we see that the treatment combinations ad and d are sufficiently discrepant from the others so as to inflate the least squares mean square error and thus obscure what appears to be a significant main effect. The Hampel and Andrews methods also find the BE interaction to be (moderately) significant, although the CE effect is insignificant by the Andrews method.

In the previous section we found that a gross outlier can radically affect a least squares analysis, while having a much smaller effect on the robust methods. In this example we have seen that slightly discrepant observations (perhaps due to a distribution heavier-tailed than the normal) can inflate the least squares mean square error, causing a potential loss of efficiency.

References

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*.

Andrews, D.F. (1974). A robust method for multiple linear regression. *Technometrics, 16*, 523-531.

Barnett, V. and Lewis, T. (1978). *Outliers in Statistical Data*. Wiley: New York.

Bickel, P.J. (1976). Another look at robustness. *Scand. J. Statist. 3*, 145-168.

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics, 19*, 15-18.

Daniel, C. and Wood, F.S. (1971). *Fitting Equations to Data*. Wiley: New York.

Dutter, R. (1976). LINWDR: Computer linear robust curve fitting program, Res. Rep. #10, Fachgruppe für Stat., Eidenössische Technische Hochschule, Zurich.

Gross, A.M. (1976). Confidence interval robustness for long-tailed symmetric distributions. *J. Am. Statist. Assoc. 71*, 409-416.

Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Am. Statist. Assoc. 69*, 383-393.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist. 35*, 73-101.

Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist. 1*, 799-821.

Huber, P.J. (1977). *Robust Statistical Procedures*, SIAM, Philadelphia, PA.

John, J.A. (1978). Outliers in factorial experiments. *Appl. Statist.* *27*, 111-119.

Schrader, R.M. and Hettmansperger, T.P. (1979). Robust analysis of variance based upon a likelihood ratio criterion. To appear in *Biometrika*.

Table 1

Predicted values and residuals for observation 11 in the 3^{4-1} experiment, based on the original observations. John's refitted value is 62.33.

	Least Squares	Huber	Hampel	Andrews
Observed value	14	14	14	14
Predicted value	46.2	55.7	57.7	59.4
Residual	-32.2	-41.7	-43.7	-45.4
Refitted value	62.33	62.33	62.33	62.33
Predicted value after refitting	62.33	61.2	61.2	61.6
Residual after refitting	0	1.1	1.1	0.7

Table 2

Significance levels for the 3^{4-1} experiment using the four methods.

Blank values indicate a significance level greater than 0.10.

Effects	Original Observations (Y_{11} = 14)				Modified Observations (Y_{11} = 62.33)			
	Methods				Methods			
	Least Squares	Huber	Hampel	Andrews	Least Squares	Huber	Hampel	Andrews
AL	.04	.00	.00	.00	.01	.00	.00	.00
AQ		.02	.01	.01	.01	.00	.01	.01
BL			.10	.08	.03	.06	.07	.07
BQ			.06	.09	.07	.10	.10	(.11)
CL								
CQ			.05	.06	.03	.02	.03	.04
DL								
DQ								
ALBL								
ALCL								
BLCL		.00	.00	.01	.01	.00	.00	.01
Predicted Value	46.22	55.70	57.72	59.35	62.33	61.20	61.17	61.60
Standard Error	7.3	4.1	3.7	4.7	4.4	3.9	4.0	4.6

Table 3

Estimates and standard errors for the 3^{4-1} experiment.

Effects	Least Squares (Original Data) Estimate	Least Squares (Original Data) S.E.	Andrews (Original Data) Estimate	Andrews (Original Data) S.E.	Least Squares (Modified Data) Estimate	Least Squares (Modified Data) S.E.
AL	6.83	2.99	6.57	1.91	6.83	1.78
AQ	-1.54	1.72	-3.07	1.10	-3.33	1.03
BL	-1.50	2.99	-3.53	1.91	-4.19	1.78
BQ	1.13	1.72	1.99	1.10	2.02	1.03
CL	-2.94	2.99	-2.58	1.91	-2.94	1.78
CQ	- .65	1.72	-2.22	1.10	-2.44	1.03
DL	1.33	2.99	- .96	1.91	-1.35	1.78
DQ	- .59	1.72	.41	1.10	.30	1.03
ALBL	.92	3.66	.19	2.32	.92	2.18
ALCL	-2.08	3.66	-2.62	2.32	-2.08	2.18
BLCL	6.00	3.66	6.59	2.32	6.00	2.18

Table 4

Significance levels for the confounded 2^5 experiment.

Blanks indicate a level greater than 0.10.

Source	Least Squares	Huber	Hampel	Andrews
A	.01	.00	.00	.00
B		.03	.02	.02
C	.03	.01	.01	.03
D	.00	.00	.00	.00
E			(.11)	.08
AB				
AC	.04	.04	.04	.07
AD				
AE	.04	.00	.00	.00
BC				
BD				
BE			.09	.06
CE	(.11)	.06	.07	
DE				
Mean Square Error	2.60	1.48	1.35	1.65

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) Robust Methods for Factorial Experiments with Outliers		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series #1244
7. AUTHOR(s) Raymond J. Carroll		8. CONTRACT OR GRANT NUMBER(s) AFOSR 75 2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina at Chapel Hill		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, DC		12. REPORT DATE July 1979
		13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS (*if different from Controlling Office*)		15. SECURITY CLASS. (*of this report*)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Distribution unlimited - approved for public release

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*)

Factorial experiment, outliers, robustness, M-estimates

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

Two factorial experiments with possible outliers (John (1978)) are reanalyzed by means of robust regression techniques.

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)