ISSN 1684-8403
J. Stat.
Vol. 10, No.1 (2003) PP 49-65

OUTLIERS IN DESIGNED EXPERIMENTS
I - CLASSICAL ROBUST & RESISTANT METHODS

AHMED F. SIDDIQI

June 2, 2004, Revised, June 15, 2005

Outliers in Designed Experiments;

I - Classical Robust & Resistant Methods

ABSTRACTS: The robust regression analysis and the designed experiments are two of the fastest growing fields in contemporary Statistics. There has been very little overlap between these fields. In designed experiments, designs were contrived for the efficient use of the least square estimates to maximize response, while in robust regression analysis, robust alternatives to the non-robust conventional least square estimates were developed. This paper, the first in the series of three, is an attempt to bridge the gap. It discusses classical robust, like M estimators using a ψ function developed by Huber, Hampel, & Tukey, and resistant estimators of these for-regression-analysis techniques to deal with outliers in the domain of designed experiments. Further, a Monte-Carlo simulation is carried out to appraise the efficiency of these methods in designed experiments and then a factorial experiment, with possible outliers, is reanalyzed. It is revealed that these techniques, with some precautions and modifications, work excellent in designed experiments.

Keywords: Outliers, M-Estimates, Huber Estimates, Hampel Estimates, Tukeys' bi weight method, Resistant Methods, Least Trimmed Median Squares methods.

1. INTRODUCTION

In the introduction to The Design of Experiment, Fisher [12] states statistical procedure and experimental designs are only two different aspects of the same whole, and that whole comprises all the logical requirements of the complete process of adding to natural knowledge of experimentation. Investigators in virtually all fields of inquiry, usually to

---

[1] School of Business & Economics University of Management & Technology

Lahore, Pakistan

discover something about a particular process or system and to improve the process functioning, perform experiments. Observations coming out from these experiment are an explicit manifestation of different facets of the process, or system, and exhibiting a peculiar harmony, balance and alliance in their magnitudes. Many a times, this is not the case and some surprising observation(s) stands apart from the bulk of data and the should-be harmony, balance and alliance shatters. These surprising observations are called outliers, in statistical literature, and the focal point of this study.

Experience shows that in a typical industrial setting 1 to 10% of all measurements performed for the purpose of the planned experiments results in grossly surprised data and should be treated as outlier (see, for related comments, Daniel [8], Anscombe [2], Morgenthaler & Schumacher [23]). Whole complexion of the data, and of its interpretation, changes because of these outliers. The presence of outliers in a set of data, quoting Hawkins [15] and Barnett & Lewis [3], may by tackled by (a) their outright rejection from the data by considering it an error, (b) further analyzes for suspected outliers for ascertaining their validity, (c) their incorporation by using special modifications either in the model or in the design, or (d) re-experimentation, to be sure what is happening. It has been a common practice, among the experimenters, to throw such observations out, just by labeling them as erroneous. But, these cannot be erroneous all the time and in that case the experimenter may run the chance of loosing an important, may be the most important, aspect of the experiment. On the other hand, once the measurements have been taken, it is in most situations impossible to check the validity of single measurements by repeating selected experiments. It is, therefore, important to search for specialized methods that may used to adjust the analysis, and interpretations, for possible surprising observations without kicking them out.

The literature on outliers in designed experiments seems to be divided into two broad categories; (a) using for-regression techniques and (b) specialized design based techniques. The for-regression techniques are pioneered by Huber, and will be discussed in section 0. While the specialized design based techniques are discussed by Box & Draper [4], & [5]), Herzberg & Andrews [16], Andrews & Herzberg [1] and Draper & Herzberg [10] primarily for developing robust design from central composite designs. Then in 1980s Taguchi and his school of thought give new edge to these robust designs (see, for example, Taguchi & Wu [29], Taguchi [30] & [31]). His parameter designs are a sort of industrial adaptation of statistical experimental designs, and are still controversial among statisticians. As a result, the experimental designs robust to outliers provide more rigorous statistical approach to analysis by considering noise variables as outliers.

Here, I am studying these surprising observations with emphasis on for-regression techniques to anticipate them, instead of throwing them out, in the domain of experimental designs. This first paper, in a series of three, will discuss the classical robust, like M estimators, Huber estimates, Tukey's bi-weight method, and resistant methods least trimmed median squares, since these considered to be the most robust family of estimators. While the remaining two would discuss the non-parametric and neural network methods respectively. Section 0 is giving a brief introduction of these methods, while the 3rd and 4th sections deal with their applications. A Monte-Carlo simulation is carried out, in the 3rd section, to compare the performance and efficiency of these methods. While in the 4th section, an experiment is reanalyzed to see how well these estimators perform on a real data.

2. ROBUST & RESISTANT METHODS

Robust and resistant methods are developed primarily to reduce the malignant effect of potential delinquencies in the application of statistical theory. These procedures can both minimize the effect of such delinquencies as well as help to identify them. In the domain of regression analysis, robust techniques are pioneered by Huber ([19], [20], and [21]) primarily to address the problem of outliers. Since then, several robust estimates for regression with high breakdown point have been proposed. Starting with M-estimates, the statistical literature, now, has a robust method named almost after each alphabet addressing not only the problem of outliers but almost all sorts of delinquencies. The resistant methods are, simply, another class of robust methods where a regression analysis is carried out by using only good points in the data set, thereby achieving a regression estimator with a high breakdown point. These include the least median of squares estimates and the least trimmed squares estimates proposed by Rosseeuw [26], the scale estimates proposed by Rousseeuw & Yohi [27], the MM estimates proposed by Yohai [33], the τ estimates proposed by Yohai & Zamar [34], etc. These estimates have a very high computational complexity (see Pena & Yohai [24], and Faraway [11]) and thus usual algorithms compute only approximate solutions.

The majority of approaches in robust regression assume that the independent variables, predictors, are random and outliers can appear also in these independent variables. For designed experiments, these assumptions make no sense. The analysis, presented in this paper, suggests that robust, and resistant, regression may be a valuable tool for experimentalists concerned with the possibility of outliers or of data which are not normally distributed. Let us discuss some classical robust and resistant methods. Consider a linear regression model of the form

$$\{1\} \qquad Y = X\theta + \varepsilon$$

Where Y is a (n x 1) vector of response variables, X is a (n x p) known design matrix of rank p, θ is a (p x 1) vector of unknown parameters, and ε is a (n x 1) vector of errors. The statistical theory necessitates a number of precincts in the application of this model including the normality of errors with mean zero and a constant variance. In presence of outlier(s), this normality precinct is no more be abided by, that results in misinterpretation of the data especially when methods are used which relied heavily on the normal distribution, like ordinary least square estimation. Here is another approach to this age-old method of least square estimation to adjust it for outliers. The least squares estimation attempts to minimize

$$\{2\} \qquad \sum_{i}^{n} \xi\left(y_i - \hat{\theta}\right)$$

where θ is a location parameter and ξ is of continuous type. Suppose that this minimization can be achieved by differentiating and solving the function in Eq.{2}; i.e., finding the appropriate $\hat{\theta}$ that satisfies

$$\{3\} \qquad \sum_{i}^{n} \psi\left(y_i - \hat{\theta}\right) = 0$$

where ψ(x) is the first differential of ξ(x). The solution of this equation (Eq.{3}) that minimizes the sum in Eq.{2} is called M-estimator of θ, the parameter in concern. Care should be taken to select ξ (.) so that the resultant estimator protect us against, at least a small percentage (around 10%) of outliers and, in addition, efficient enough (around 95%) in case the data actually enjoys the normality precincts. There are several solutions available for the equation in Eq.{3}. Rey [25] gives a comprehensive discussion on a number of such solutions. The function sign*(x)* in the

Table 1 is equal to -1, 0, or 1 according to whether *x* is less than, equal to, or greater than zero. The constants *a*, *b*, and *c* may be chosen so that the resultant estimator has desirable properties. The fact that Hampel's, Andrews' and Tukey's *ψ(x)* descend to zero suggests that they give no weight to unusually large outliers (see Hampel [14]). But, this is not the case with Huber's that does give some weight (no doubt much smaller than the least squares) to these unusually large outliers. But, such a descending function may have problems with convergence. As Venables & Ripley [32] observes that Huber's *y(x)* corresponds to a convex optimization problem and gives a unique solution, while the other may have multiple local minima and a good starting point, for iterations, is desirable.

The least trimmed mean squares methods, on the other hand, attempts to minimize a randomly selected subset of errors, $\sum_i^q \hat{\varepsilon}_i^2$ where *q* is some number less than *n* and *i*) indicates sorting; that is, they are defined by

$$\{4\} \qquad \hat{\theta} = \arg\left(\min S(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)\right)$$

where $\hat{\varepsilon}_i = y_i - x_i\hat{\theta}$, $i = 1, 2, \ldots, n$ and S is some appropriate scale of residuals. The usual approximate solution to the estimates defined by Eq.{4} are of the form

$$\{5\} \qquad \hat{\theta} = \arg\left(\min_{\theta \in \mathrm{A}} S(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)\right)$$

where $\mathrm{A} = \left(\theta^{(1)}, \ldots, \theta^{(n)}\right)$ is a sorted finite set of parameter estimates. Selection of the element of this finite set is quite scheming. Rosseeuw [26] proposed a scheme of choosing random n sub-samples of p different design points defined by

$$\{6\} \qquad \min_{\theta} \sum \left| y_i - \hat{\theta} \right|_{(i)}^2$$

Table 1: Some Popular Solutions for ξ(x) and ψ(x)

Criterion	ξ (x)	ψ(x)	Range
Huber	$\frac{1}{2}x^2$	x	$\lvert x\rvert \le c$
	$\left(\lvert x\rvert c - \frac{1}{2}c^2\right)$	c.sign(x)	$\lvert x\rvert > c$
Andrews	$a\left(1-\cos\frac{x}{a}\right)$	$\sin\frac{x}{a}$	$\lvert x\rvert \le a\pi$
	2a	0	$\lvert x\rvert > a\pi$
Hampel	$\frac{1}{2}x^2$	x	$\lvert x\rvert > a$
	$\left(\lvert x\rvert a - \frac{1}{2}b^2\right)$	a.sign(x)	$a < \lvert x\rvert \le b$
	$\frac{a\left(c\lvert x\rvert - 0.5x^2\right)}{c-b} - \frac{7}{6}a^2$	$\frac{a.\mathrm{sign}(x)\left(c-\lvert x\rvert\right)}{c-b}$	$b < \lvert x\rvert \le c$
	$a(b+c-a)$	0	$c < \lvert x\rvert$
Tukey	Bi-weight	$x\left(1-\frac{x^2}{a}\right)^3$	$a \ge \lvert x\rvert$
		0	$a < \lvert x\rvert$
	Bi-square	$\frac{c^2}{3}\left(1-\left(\frac{x}{a}\right)^2\right)^3$	$c \ge \lvert x\rvert$
		$\frac{c^2}{3}$	$c < \lvert x\rvert$

This fit is very resistant, and needs no scale estimate. Unlike the robust regression, it can reject values that fit badly because their x_i are outlier. However, this merit is of no use in case of experimental data where the x's are fixed. Hettmansperger & Sheather [17] and Davies [9] remark that it displays marked sensitivity to central data values. There are also differing opinions over the selection of q. Rosseeuw suggests that the sum should be taken over the smallest

{7} $$q = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+1}{2} \right\rfloor$$

squared errors, while Marazzi ([22], p. 200) opines for Rosseeuw & Leroy [28] show, for a small fraction p/2, the probability of getting a clean subset is given by $(1-(1-(1-\eta)^p)^n)$ where η is the fraction of outliers. They also compute the number of sub-samples required to make this probability (1-α) is given by

{8} $$n_{\eta,\alpha,p} \cong \frac{-\log \alpha}{(1-\eta)^p}$$

Incidentally, this number increases exponentially with an increase in p. Resultantly, methods based on random sub-sampling can be appropriate only when p is not very large. Venables & Ripley [32], however, is using an iterative procedure with different functions of the sorted squared errors, in writing algorithms (named lts, lms, lqs) for S-Plus and R, including quantile squared errors, sum of the quantile smallest squared errors with different default values. To control the complexity and time intensity of the procedures, he uses three levels; (1) 5p, (2) exhaustive enumeration up to 5000 samples, and (3) complete enumeration. The complete enumeration may take several hours to have a solution.

3. MONTE CARLO SIMULATION FOR COMPARISON

It is very difficult to discriminate these methods for the best, or even for an optimal. Monte Carlo Simulation, however, has the advantage that it is a brutal force technique that solve many problems for which no other solution exist. Here is a Monte Carlo simulation for discriminating among these methods. Consider a 2^4 full factorial experimental design for a second order model with all the two factor interaction terms. Translating, the model in Eq.{1} for the given situation results in

{9} $$y_i = \theta_0 + \sum_i^{16} \theta_i X_i + \sum_i^{16} \sum_j^{16} \theta_{ij} X_i X_j + \varepsilon_i$$

Suppose the true values for the model parameters are known and are given by Table 2.

Table 2: True Values for the Model Parameters

θ_0	θ_1	θ_2	θ_3	θ_4	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}
5	3	6	8	4	5	1	7	0.4	3	6

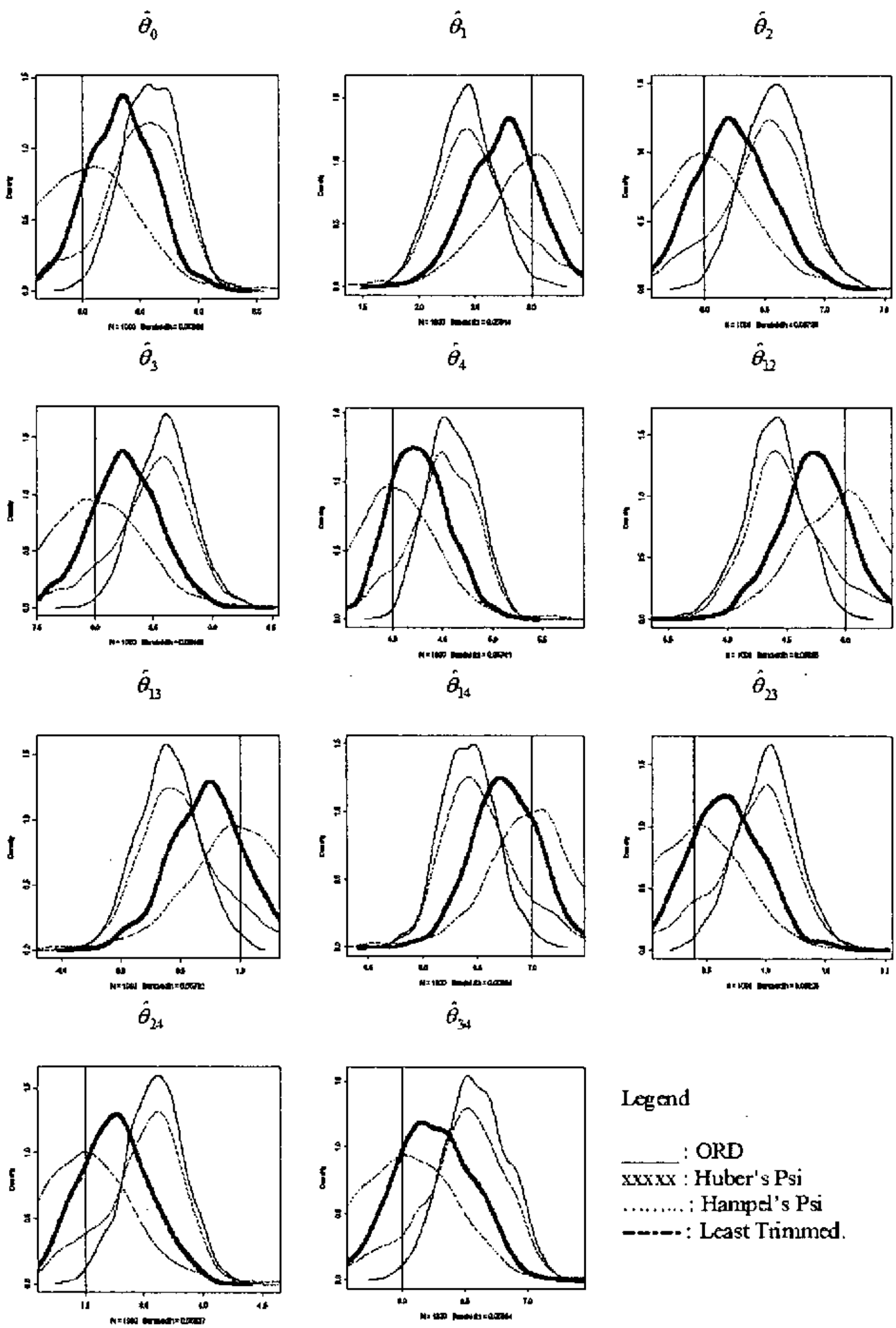
One thousand samples are taken from a normal distribution to estimate these model parameters employing different methods of estimation, described in the previous section. An outlier is introduced by making use of mean shift model approach (see Cook & Weisberg, [7], pp. 20) assuming that vector d (a 16 x 1 vector with zeros everywhere except at the point where the outlier is to be introduced) is added to error vector. Results are shown in Figure 1, which is showing the empirical distributions for all the eleven model parameter estimates under different methods.

The fit is considered to be as good as it is close to the true value. The solid black line is used for the ordinary least square fit, the thick line is used for a M estimator under Hubers' ψ function with tuning constant equal to unity, the dotted line is used for a M estimator under the Hampel's ψ function with initial set of parameters estimates taken from the ordinary least squares fit, while the spotted line is used for a least trimmed mean square fit. A computer code, written in R, is available in appendix for this simulation, which takes approximately 3 hours on a PIV 2.4 GHz, 512 MB RAM machine.

The vertical line in each panel is showing the actual value of the parameter. Evidently, the least square fit is pathetically poor even in the presence of a single moderate outlier. No doubt, it is more efficient but it stands a way apart from the true value. While the robust estimators seem to be less efficient but they are more close to the true values. Among the Robust methods, the Hampel's M estimates are behaving just like that of ordinary least squares'. Hubers's estimates, however, are doing slightly better. But, the least trimmed mean square fit is the best. It manages to estimate all the model parameters with approximately zero standard error.

Let us take a real data example, from the literature, to see the behavior of these estimates.

Figure 1: Empirical Distributions for Parameter Estimation under Different Methods



An Example

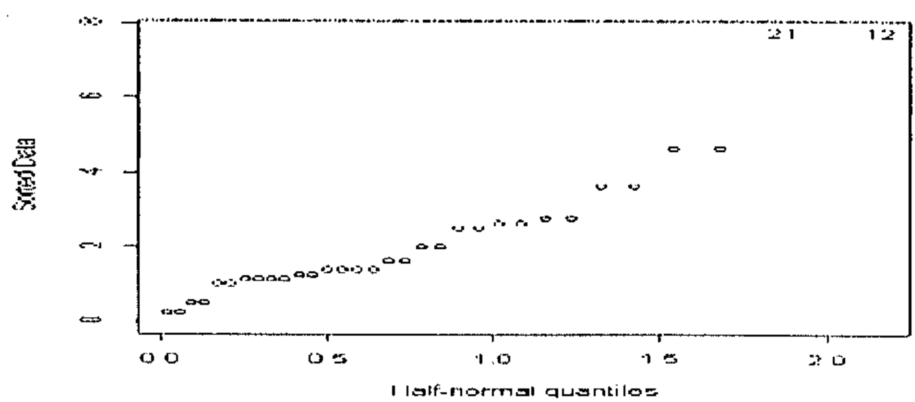
I am taking an already analyzed data from Goldsmith & Boddy ([13], second example). The previous analysis would be used to compare the efficiency of the new methods. The example relates to a 2^{6-1} fractional factorial design used to look into the fibrillation occurring in a polyester tape when two contra-rotating air jets twist it. The analysis was done with SAS. The factors explored are

A :	Tape Width	B :	Tape Thickness
C :	Type of Jet	D :	Tape Speed
E :	Air Pressure	F :	Tape Tension

with the defining contrast F = ABCDE. The formal analysis of the, by Goldsmith & Boddy [13], using all main effects and only some two-factor interactions, and making use of their own special method based on the residual sums of squares, shows that the data may have two outliers [observation #12 and #16]. I am using here, however, all the two factor interactions. The half-normal plot, in
Figure 2, produced by R, for the least square residuals.

Figure 2: Half Normal Plot for Residuals

Goldsmith & Boddy [13] Data



There are two observations (#12 and #21) stand apart from the data, may be outliers, and the normality assumption also seems to be in danger. Resultantly, the least square estimation can not be relied upon. Let us apply the robust procedure to extract *M*-estimates for the model parameters. Error! Not a valid bookmark self-reference. gives these parameter estimates, while the Table 4 gives the weight for design points, for different selection of $\psi(x)$, constants and initial values for iterations involved, explained in Table 1.

The weight allocation renders the Huber fit, with a tuning constant c=1.8, closest to the ordinary least square solution. A lower tuning constant gives a less than unity weight even to contiguous-to-suspected-outliers, but smaller tuning constant increases the residual standard error. The parameter estimates, and their significance status, for the intercept and all the included interaction effects are the same for different values of c.

The two suspected outliers (design point #12 and #21) revealed by the half-normal plot, in Figure 2, are allocated the least weights. The Hampel's psi function, with initial values of parameter's estimates given either by ordinary least squares (*ls*) or by least trimmed squares with 200 samples (*lts*), tells a different story.

The weight allocation, in this case, is quite different from that of Huber's. There are some design points that are given zero weights. The function with least trimmed mean square estimators are initial values renders the same design points as outliers, while the other initial values renders a different set of suspected design points. The 8th and 9th columns give the *M* estimates by using Tukeys' biweight psi function. Quite interestingly, this gives another set of suspected design points, different from the previous. It also renders a few effects as significant.

Table 3: Parameter Estimators under Different ψ(x)

Linear Regression		Huber			Hampel		Tukey		Resistant Estimates
		C = 0.2	c =1.0	c = 1.8	ls	lts	ls	lts	
Intercept	9.125	9.125	9.125	9.125	8.625	9.125	11.625	8.563	8.613
A	1	0.750	0813	0625	-0.250	0.292	-1.750	1	1.625
B	2.875	2.75	2.688	2.5	1.5	2.167	2	4	3.375
C	-2.813	-2.193	-2.165	-2.438	-0.5	-2.105	-3	-2.5	-2.125
D	1.5	0.443	0.415	1.125	1.75	0.792	-0.25	0.75	0.125
E	-2.25	-2.114	-2.125	-1.875	-0.75	-1.542	1.25	-2.5	-2.125
F	0.813	0.136	0.125	0.438	-1.25	0.105	0.25	0.75	0.125
AB	1.375	1.375	1.375	1.375	0.75	1.375	1.875	0.813	0.94
AC	-1.063	-1.063	-1.063	-1.063	0.25	-1.063	-1	-1.313	-1.9
AD	1.25	1.25	1.25	1.25	0.375	1.249	3.25	1.938	2.69
AE	-0.625	-0.625	-0.625	-0.625	1.375	-0.625	-2.875	-0.813	-0.6
AF	0.813	0.813	0.813	0.813	-0.625	0.813	0.375	1.438	1.3
BC	-0.688	-0.688	-0.688	-0.688	-0.875	-0.688	-2	-1.563	-1.6
BD	0.875	0.875	0.875	0.875	-1.5	0.875	2.5	2.188	2.3
BE	-0.875	-0.875	-0.875	-0.875	-0.25	-0.875	-1.625	-0.563	-0.9
BF	1.063	1.063	1.063	1.063	1	1.062	2.125	1.188	1.69
CD	-1.063	-1.063	-1.063	-1.063	0.25	-1.063	-1.625	-2.188	-1.8
CE	1.313	1.313	1.313	1.313	-0.25	1.312	1.25	1.563	1.4
CF	-1.125	-1.125	-1.125	-1.125	-1.75	-1.125	-2	-2.688	-2.4
DE	-1.75	-1.75	-1.75	-1.75	-1.375	-1.749	-4.5	-3.188	-3.2
DF	0.563	0.563	0.563	0.563	-2.125	-0.563	0.75	-0.188	-0.3
EF	-0.563	-0.563	-0.563	-0.563	0.625	-0.562	-0.625	-0.688	-1.3
$\hat{\sigma}_\varepsilon$	5.121	0.5726	0.4892	2.224	1.668	-	-	-	-

The main difference between the Huber method and the Tukey's biweight method is the sharper solution with Tukey's. Both techniques minimize some criterion function computed from the residuals, but the Huber criterion is a convex function of the unknown regression parameters while the Tukey's may have more than one local minima, as Hogg [18] points out. This is particularly true with the data with more than one fit, which is the case here. Collins [6] also warns the use of these sensitive-to-multi-minima estimators in case of acute asymmetric data. The very last column gives the estimates for the trimmed median squares. These estimates are quite different in magnitude, significance (calculated through the Bootstrap method) and the sign. Most of the interactions effects have different signs, while only two of them are significant. The Monte Carlo simulations renders these estimators are the best.

4. DISCUSSION & CONCLUSION

The classical least square approach for estimating the model parameters is founded on stringent precincts. But, the real word does not behave as nicely as described by these precincts. The performance and the valid application of the procedure requires strict adherence to the assumptions. Consequently, semi parametric (as discussed in this paper) and non parametric approaches are sometimes the only possible solutions. In many practical situations, the experimenter does have an idea about the experimental error and thus the fault, if exist, can easily be sensed. This paper addresses what to do next. There exists a bunch of robust and resistant methods and I have tried to compare some of them with particular reference to experimental designs. The Monte-Carlo simulation and the analysis of the 2^{6-1} factorial design data led us to the following conclusions;

- In a designed experiment, these techniques are applied only on response data and not on X data.
- A very unusual observation should be taken away from the data either for a separate peeking or dustbin. All the robust and resistant methods are sensitive to grossly unusual data points(s).

Table 4: Weights for Design Points under Different $\psi(x)$

Obs.	Linear Regression	Huber			Hampel		Tukey	
		c=0.2	c=1.0	c=1.8	ls	lts	ls	lts
1	1	0.0899	0.3914	1	0	1	1	0
2	1	1	1	1	0	1	0	1
3	1	1	1	1	0.9884	1	1	1
4	1	0.0219	0.0932	1	1	1	1	1
5	1	1	1	1	1	1	1	0
6	1	0.1738	1	1	1	1	1	1
7	1	0.7181	1	1	1	1	0	1
8	1	0.0278	0.1173	1	1	1	0	1
9	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	0
11	1	0.1868	1	1	1	1	1	0
12	1	0.0107	0.0459	0.4004	0	0.0561	1	1
13	1	0.0663	0.2795	1	0	1	1	1
14	1	1	1	1	0	1	1	1
15	1	1	1	1	1	1	1	1
16	1	0.0899	0.3913	1	1	1	1	1

17	1	0.0899	03913	1	0	1	0	1
18	1	1	1	1	1	1	0	0
19	1	1	1	1	1	1	0	1
20	1	0.0663	0.2795	1	1	1	1	1
21	1	0.108	0.04585	0.4004	1	00561	0	0
22	1	0.1868	1	1	1	1	0	1
23	1	1	1	1	1	1	0	1
24	1	1	1	1	0	1	1	1
25	1	0.0278	0.1173	1	0	1	1	0
26	1	0.7181	1	1	1	1	1	1
27	1	0.1738	1	1	1	1	0	0
28	1	1	1	1	1	1	1	1
29	1	0.0219	0.0932	1	1	1	1	0
30	1	1	1	1	0	1	1	0
31	1	1	1	1	1	1	1	1
32	1	0.0899	0.3914	1	1	1	1	1

- The robust estimators are biased and are less efficient than that of ordinary least squares, but they estimate the parameters with more accuracy.
- The robust procedures are two prong attack; they identify and locate the outliers and secondly adjust the analysis for their presence.
- The experimenter does not have to worry about the magnitude and number of outliers, like the design based procedures.

APPENDIX

```
## Monte Carlo Simulation
## Code written in R
library(Mass)
library(lqs)
bols<-matrix(0,1000,11)
bhub<-matrix(0,1000,11)
bham<-matrix(0,1000,11)
bres<-matrix(0,1000,11)
## Construction of Design Matrix
a<-c(rep(c(-1,1),c(8,8)))
b<-rep(c(rep(c(-1,1),c(4,4))),2)
c<-rep(c(rep(c(-1,1),c(2,2))),4)
d<-rep(c(-1,1),8)
x<-cbind(1,a,b,c,d,a*b,a*c,a*d,b*c,b*d,c*d)
## Supposed Values of Parameter Estimates
bta<-c(5,3,6,8,4,5,1,7,0.4,3,6)
## Simulations Starts
for ( i in 1:1000){
```

```
## Error is supposed to be normally distributed
e<-rnorm(16)
## Error is corrupted with an outlier
e[8]<-e[8]+(9+runif(1))
y<-x%*%bta+e
bols[i,]<-coef(lm(y~(a+b+c+d)^2))
bhub[i,]<-coef(rlm(y~(a+b+c+d)^2,
psi=psi.huber,
method=c("M"),scale.est="MAD",maxit=150),k=1)
bham[i,]<-coef(rlm(y~(a+b+c+d)^2,
psi=psi.hampel,
method=c("M"),scale.est="MAD",maxit=150),init="ls")
bres[i,]<-coef(ltsreg(y~(a+b+c+d)^2,nsamp="exact"))
}
## Density Plots for Empirical Distributions
par(mfrow=c(3,4))
for (i in 1:11) {
        plot(density(bols[,i]),main=" ")
        lines(density(bhub[,i]),lty=2,col=2)
        lines(density(bham[,i]),lty=3,col=3)
        lines(density(bres[,i]),lty=4,col=4)
        abline(v=bta[i])
}
par(mfrow=c(1,1))
```

REFERENCE

1. Andrews, D.F. and A.M. Herzberg, [1979] Robustness \& Optimality of Response Surface Designs. Journal Of Statistics, Planning & Inference, 3: p. 249-257.

2. Anscombe, F.J., [1960] Rejection of Outliers. Technometrics, 2: p. 123-147.

3. Barnett, V. and T. Lewis, [1994] Outliers in Statistical Data. 1994, New York: Wiley.

4. Box, G.E.P. and N.R. Draper., [1959] A Basis for the Selection of a Response Surface Designs. Journal of American Statistical Association, 54: p. 622-654.

5. Box, G.E.P. and N.R. Draper, [1975] Robust Designs. Biometrika, 62: p. 347-352.

6. Collins, J.R., [1976] Robust Estimation of a Location Parameter in the Presence of Asymmetry. Annals of Statistics, 4: p. 68-85.

7. Cook, R.D. and S. Weisberg, [1982] Residuals & Influence in Regression. 1982, New York: Chapman & Hall.

8. Daniel, C., [1959] Use of Half Normal Plots in Interpreting Factorial 2 Level Experiments. Technometrics, 1: p. 311-341.

9. Davies, P.L., [1993] Aspects of Robust Linear Regression. Annals of Statistics, 21: p. 1843-1899.

10. Draper, N.R. and A.M. Hertzberg, [1979] Designs to Guard against Outliers in the Presence or Absence of Model Bias. Canadian Journal Of Statistics, 7: p. 127-135.

11. Faraway, J., [2000] Practical Regression & Anova Using R, University of Michigan, Ann Arbor.

12. Fisher, R.A., [1935] The Design of Experiments. 1935, London: Oliver & Boyd.

13. Goldsmith, P.L. and R. Boddy, [1973] Critical Analysis of Factorial Experiments & Orthogonal Fractions. Applied Statistics, 22(2): p. 141-160.

14. Hampel, F., [1974] The Influence Curve & Its Role in Robust Estimation. Journal of the American Statististical Association, 69: p. 383-393.

15. Hawkins, D.M., [1980] Identification of Outliers. 1980, New York: Chapman & Hall.

16. Herzberg, A.M. and D.F. Andrews, [1976] Some Considerations in the Optimal Design of Experiment in Non-Optimal Situations. Journal Of Royal Statistical Society, Ser. B, 38: p. 284-289.

17. Hettmansperger, T.P. and S.J. Sheather, [1992] A Cautionary Note on the Method of Median Squares. American Statistician, 46: p. 79-83.

18. Hogg, R.V., [1975] Estimates of Percentile Regression Lines Using Salary Data. Journal of the American Statistical Association, 70: p. 56-59.

19. Huber, P.J., [1964] Robust Estimation of a Location Parameter. Annals of Mathematical Statistics, 35: p. 73-101.

20. Huber, P.J., [1973] Robust Regression; Asymptotics, Conjectures & Monte Carlo. Annals of Statistics, 1: p. 799-821.

21. Huber, P.J. [1978] Robust Statistical Procedures. in SIAM. Philadelphia.

22. Marazzi, A., [1993] Algorithms, Routines, & S Functions for Robust Statistics. 1993, London: Wadsworth & Brooks.

23. Morgenthaler, S. and M.M. Schumacher, [1999] Robust Analysis of a Response Surface Designs. Chemometrics & Intelligent Laboratory Systems, 47: p. 127-141.

24. Pena, D. and V. Yohai, [1999] A Fast Procedure for Outlier Diagnostics in Large Regression Problems. Journal of the American Statistical Association, 94: p. 434-445.

25. Rey, W.J.J., [1983] Introduction to Robust & Quasi Robust Statistical Methods. 1983, New York: Springler-Verlag.

26. Rousseeuw, P.J., [1984] Least Median of Squares Regression. Journal of the American Statistical Association, 79: p. 871-880.

27. Rousseeuw, P.J. and V.J. Yohai, [1984] Robust Regression by Means of S-Estimators, in Robust & Nonlinear Time Series - Lecture Notes in Statistics No. 26, J. Franke, W. Hardle, and R.D. Martin, Editors. 1984, Springer-Verlag: New York. p. 256-272.

28. Rousseeuw, P.J. and A. Leroy, [1987] Robust Regression and Outlier Detection. 1987, New York: Wiley.

29. Taguchi, G. and Y. Wu, [1980] Introduction to Off-Line Quality Control. 1980, Nagoya, Japan: Central Japan Quality Control Association.

30. Taguchi, G., [1986] Introduction to Quality Engineering. 1986, New York: Asian Productivity Organization, UNIPUB.

31. Taguchi, G., [1987] System of Experimental Design: Engineering Methods to Optimize Quality \& Minimize Cost. 1987, New York: UNIPUB/Kraus International, White Plans.

32. Venables, W.N. and B.D. Ripley, [2001] Modern Applied Statistics with S-Plus. 2001, New York: Springer.

33. Yohai, V.J., [1987] High Breakdown Point & High Efficiency Robust Estimates for Regression. Annals of Statistics, 15: p. 642-656.

34. Yohai, V.J. and R. Zamar, [1988] High Breakdown Point Estimates \& Inference in Linear Regression, in Directions in Robust Statistics & Diagnostics, Part Ii, W. Stahel and S. Weisberg, Editors. 1988, Springer-Verlag: New York.

- A marketing researcher has so much to do in order to survey a population's perceptions of quality dimensions of products and services.

So, whether one likes it or not, quantitative techniques, statistics in particular, often finds their applications in all areas of management may mention that a recent survey in the USA revealed that more than 80% of American organization uses one or other quantitative technique. In Pakistan as well, the trend in this regard has been visibly rising.

Information is fundamental for change and improvement. Without useful information, our approach to problems remains shallow, superficial and vague, and one may walk aimlessly and pointlessly. With the growing awareness about quality management, the importance of reliable information is now no longer under-emphasized. Information may be qualitative or quantitative, it is conceived as an important assets, a powerful basis to comprehend a business phenomenon, situation or event. When information involved is poor, we all know the management functions suffer. It is like when at a road crossing the false information may send a traveler to opposite direction, a manager experiences similar disappointments after using fallacious information. But at times we do encounter situations where even when reliable information is available the managers may formulate biased policies, or conduct the operations dishonestly to suit their certain interests. It is do with one's behaviour.

Ethics refer to principles of moral duty and obligations relating to right behaviour of an individual or a group. Adherence to these principles provides security, direction and integrity. These principles derive their support from a religion, from common law, or from the standards of conduct accepted as right over a long period of time. Ethical principles display impartiality and do not allow, encourage or approve special exceptions benefiting or harming specific persons or groups if not all, most of these principles have their applications in all professions, including Management and statistics.

Regarding managers, it depends as to what their areas of management are. Marketing ethics except fair treatment of consumers, selling quality product according to marketing claims at a reasonable price. A production manager is expected to produce a product according to the stated specifications and not mislead his customers. An inventory manager should protect his inventory from damage and destruction. The advertising managers portray women as advertising strategies for the promotion of industrial products, which may seen unethical to many. The ethics of management accountants and auditors emphasize that they should communicate information fairly and objectively disclosing all information. Human resource management is expected to hire, fire, promote employees in view of well written rules. Corruption, bribery, discrimination, coercion, deception thefts, nonperformance of duties are all unethical aspects of

INFORMATION FOR THE AUTHORS

1. Research Papers:

The main aim of the journal is to publish research papers containing original work in all fields of statistics.

2. Manuscript:

Articles submitted for publication should be typewritten and double spaced with a margin of 1½ inches each on both sides. Standard notations should be used. Hand drawn symbols are satisfactory if clearly done. Figures, diagrams and table should be prepared on separate pages with proper title and number.

3. Format of Manuscript:

(i) **Language**: All articles should be in English.

(ii) **Abstract**: A brief summary of the article not exceeding 200 words should be given below the title of the article. It should include all the results in the article.

(iii) **Headings**: Each section/sub–section should be given a heading and a number.

(vi) **References**: References should be listed at the end of the article, e.g.

[1] **Pervaiz, M.K. (1989)** "Asymptotically robust tests for the equality of two covariance matrices in complex surveys – I." *PROC. ICCS – I VOL. II P. 861 – 876*

4. Submission of Manuscripts:

The manuscripts should be submitted in triplicate to any member of the editorial board. All the articles are referred. The decision of editorial board is final.

5. Reprints and page charges:

Each author is to be charged $10 per page to meet the mail expenses. One copy of the Journal together with 25 reprints of the article are supplied to the authors. Order for additional reprints should be made in advance. Each additional reprint will cost $ 5.00.

management, Some researchers exploit their subjects when doing their research; and it may happen anywhere, in a developed or developing country. For instance, in 1990 the prestigious John Hopkins University displayed a highly unethical behaviour in "exposing hundreds of poor infants to major health risks without alerting their parents, with the consequence that some of them suffer learning disabilities".

No doubt workers in management organizations follow, or are also expected to follow ethical principles. In this regard, I refer to an important survey conducted in the USA. Henry Fountain published an article "Of White Lies and Yellow Pads", in New York Times on July 6, 1997. His research comprised the following questions which he posed Times on July 6, 1997. His research comprised the following questions which he posed to 1.300 workers who said they were involved in unethical activities in their work. The percentage of the respondents who admitted the unethical behaviour is also given:

1.	Cut corners on quality control That is, they do not strictly follow the prescribed specifications In producing quality products or services	16%
2.	Covered up incidents. That is, if they do anything wrong or fallacious the conceal it.	13%
3.	Abused or lied about sick days. It is again an unethical tendency among the workers.	11%
4.	Lied to or deceived customers. It refers to misleading of customers	9%
5.	Put inappropriate pressure on others. It reflects selfishness to achieve certain unethical motives.	7%
6.	Falsified numbers or reports.	6%
7.	Dismissed or promoted an employee unfairly	6%
8.	Lied to or deceived superiors on serious matters	5%
9.	Withheld important information.	5%
10.	Misused or stole company property	4%
11.	Engaged in or overlooked environmental infraction.	4%
12.	Took credit for someone's work or idea.	4%
13.	Discriminated against a co-work or idea.	4%
14.	Abused drugs or alcohol.	4%

15.	Engaged in copyright or software infringement.	3%
16.	Lied to or deceived subordinates on serious matters.	3%
17.	Overlooked or paid or accepted bribes.	4%
18.	Had extramarital affair with business associate.	3%
19.	Abused an expense account.	2%
20.	Abused or leaked proprietary information.	2%
21.	Forged name without person's knowledge	2%
22.	Accepted inappropriate gifts or services.	1%
23.	Filed false regulatory or government reports.	1%
24.	Engaged in insider trading.	1%

Some more unethical activities can also be entered depending on the conditions prevailing in an organization.

Unethical activities in an organization dwindle the quality of its products and services, reduce the volume of its productivity, adversely affect its goodwill and image. The environment may also lose its decency, grace and its overall impact.

STATISTICIANS AND THEIR BEHAVIOURS

Apart from the above kind of behaviour relating to management we have the behaviour of statisticians. When statistical information is deliberately fabricated, distorted, inflated or deflated, we leave the domain of statistical ethics. It is an area where most of the people take liberty in criticizing against those who produce statistical information for a management activity. Political leaders when in opposition quite often accuse the government to justify their management policies based on false statistics.

Generally speaking, errors for whatever reason occur in all professional activities. Statistical work is not is an exception. To understand the nature of errors involved in any profession we need to know the purpose of that profession. In 1869 Adolphe Quetlet gave a number of definitions of statistics including the one that became popular because of its pleasant circularity. He said statistics is what statisticians do. But then one may wonder with a sense of curiosity as to what do they do. And whatever they do how ethical they are.

A statistician whether meeting the needs of management or other disciplines has two main functions.

One is that he collects quantitative information as reliable as possible through surveys, experiments, trials or observation in consistence with