CHAPTER I

DIFFERENT ASPECTS OF ROBUSTNESS

1.0 INTRODUCTION : In this chapter the different aspects of robustness are discussed. The concept of robustness is explained in the section 1.1. In the section 1.2. the achivements expected from a robust procedure is considered. The definition of robust sequence and some general discussion is given in the section 1.3. Some examples concerning robust estimator and robust test are given in the last section of this chapter.

1.1 CONCEPT OF ROBUSTNESS : In any statistical analysis problem, one applyies various statistical methods to arrive at a conclusion. One always tryies to set up a theoratical model for experiment so that standard techniques become applicable. While fitting appropriate or the best possible model, various 'errors' can be committed so that final results could be inconsistent. These inconsistency could be due to imposing unnecessary conditions or excluding necessary conditions or ignoring more appropriate model etc.

For example, (i) if a linear regression model is fitted to a data which is known to have quadratic regression, the results may not be consistent and (ii) in fitting of a distribution to a given data (which perhaps contain erronnious observations that can be identified easily based

on the nature of observations) if the necessary conditions are not imposed then again the result could be inconsistent.

Generally, statistical inferences are based upon sample observations as well as prior assumptions or beliefs about underlying situation. The prior assumptions may be about form of distribution (distributions of variable of interest) or independence of sample observations etc. These assumptions are mathematically convenient for further developement of the theory. The validity of such assumptions in a particular problem has a great value.

To quote from Huber(1981)Page 1; a mathematical model can be justified by stability principle. This principle states that a minor error in choosing the mathematical model should cause only a small error in the final conclusions.

Here, small has relative meaning. The actual decision about smallness depends upon the importance of quantity of interest in particular situation.

One can observed that most common statistical procedures (especially those which are 'optimal' for underlying model) are sensitive to minor deviations from prior assumptions. When a model is fixed, there are various methods to proceed and reach at conclusion. The high degree of sensitivity of a test procedure may not be always desirable (By sensitivity of a test we mean the performance (power) of the test gets drastically changed when the model is slightly changed). When

assumptions required for validity of a test procedure are not satisfied, one cannot go for applying that test procedure. In such case, one has to look for those procedures which would be insensitive to small aberration (departure) from postulated assumptions. This leads to concept of robustness.

In the following we discuss the various aspects of robustness:

i) A statistical procedure is described as robust if it is not very sensitive to departure from the assumptions on which it depends (Kendall and Stuart, Vol.II, (1960),P.483)

ii) Any test or estimate that performs well under modifications of underlying assumptions is usually referred to as robust (Rohatgi (1976),P.580)

iii) Robustness is a sign of (or reflects) the insensitivity of test procedures or estimators to small deriations from underlying assumptions (Ray (1981), P.1).

iv) Robustness signifies insensitivity to small deviations from the assumptions (Huber (1981), P.1).

v) Robust statistics is a body of knowledge, partly formalized into "theories of robustness", relating to deviations from idealized assumptions in statistics (Hample (1986), P.6).

vi) Robust statistics, as a collection of related theories, is the statistics of approximate parametric models (Hample (1986), P.7).

Main feature of the robust methods is their reduced sensitivity to a departure from the assumptions. Robustness theory helps us to understand the behaviour of statistical procedures in real-life situations.

For example, if we consider normal distributions with location parameter θ and variance 1, then it is known that the mean $\bar{x}$ is minimal sufficient statistic. But however, if the model is slightly enlarged to the class of all normal distributions with location parameter θ and variance σ^2 such that $1-\epsilon_1<\sigma^2<1+\epsilon_2$ where ϵ_1 and ϵ_2 are arbitrary very small, then $\bar{x}$ will no more be sufficient. That is the sufficiency of $\bar{x}$ is lost when the model is slightly changed and ii) in the aspect of robustness, discussed above, the word departure is used. This departure may be of the following types :

i) departure from normality.

ii) departure from independence.

iii) departure from true value of mean.

iv) departure from correctness in recording the observations.

In the following we shall explain the above four kind of departures in the context of t-test.

The distribution of the t-statistic defined by

$$t = \sqrt{n}\left|\frac{\bar{x} - \mu_o}{S}\right| \text{ or } t^2 = \frac{n(\bar{x} - \mu_o)^2}{S^2}, \qquad (1.1.1)$$

where $\bar{x}$ and S are computed from n observations $x_1, x_2, \ldots, x_n$ on a random variable x, is obtained on the following assumptions :

a) The distribution of the random variable x is normal.

b) The observations drawn are mutually independent.

c) The mean in the population is exactly μ_o.

d) There are no errors in recording the observations.

Suppose we wish to use the t-distribution to test the hypothesis that $H_o: \mu = \mu_o$ against $H_1: \mu \neq \mu_o$ on the basis of n observations. If the value of t is large, that is, significant at level α(small). Then one or more of the assumptions (a) to (d) may be wrong.

We know that for large n, the statistic t has an asymptotic normal distribution irrespective of the population being sampled (explained in example(1.4.2), that is, for large n, the distribution of t is independent of the form of the population. In other words, the t-distribution is insensitive to moderate departures from normality. Thus a significant value of t may not be interpreted as indicating departure from normality of the observations.

Now, let us consider the effect of departure from independence on the t-distribution. Suppose that the

observations $x_1, x_2, \ldots, x_n$ have multivariate normal distribution with $E(x_i)=\mu, V(x_i)=\sigma^2$, and ρ is the common correlation coefficent between any x_i and x_j, $i \neq j$. Therefore, the expected values of the numerator and denominator of t^2 of (1.1.1) are

$$\frac{nE(\bar{x} - \mu_o)^2}{E(S^2)} = \frac{\sigma^2[1+(n-1)\rho]}{\sigma^2(1-\rho)} = 1 + \frac{n\rho}{1-\rho} ,$$

where $V(\bar{x}) = E(\bar{x} - \mu_o)^2 = \frac{\sigma^2}{n}[1+(n-1)\rho]$ and $E(S^2) = \sigma^2(1-\rho)$

Thus, the ratio $\frac{nE(\bar{x} - \mu_o)^2}{E(S^2)}$ $= 1$, if $\rho=0$

> 1, if $\rho>0$

$\rightarrow \infty$, if $\rho \rightarrow 1$

It follows that a large value of t is expected to occur when ρ is positively large ($\rho \rightarrow +1$), evenwhen μ_o is the true value of the mean. Thus a significant value of t may be due to departure from independence.

When the assumptions (a),(b) and (d) are true and the true value of the mean is $\mu \neq \mu_o$. The ratio of the expected values of the numerator and denominator of t^2 of (1.1.1) is

$$\frac{n(\mu - \mu_o)^2}{\sigma^2} + 1 ,$$

and is equal to 1 when $\mu = \mu_o$.

Thus the large value of |t| do occur when assumption (c) is wrong that is mean of the population is departure from its

true value.

There is departure from the correctness in recording the observations but there is no way to study the effect of recording errors on the distribution of t. With some care in recording the value of observations departure from assumption (d) can be avoided. ∎

1.2 ACHIVEMENTS EXPECTED FROM A ROBUST PROCEDURE : To start with we have a model which hopefully is a godd approximation to the true set up, but we cannot and do not assume that it is exactly correct. Any statistical procedure should satisfy the following desirable features:

i) It should have a reasonably good (or optimal) efficiency at the assumed model.

ii) It should justify stability principle, that is, it should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly.

iii) It should safeguard against gross error or outliers (Huber(1981), P.5).

There is another term which is frequently used in literature concerning robustness study namely "outlier". The outliers affect the estimates of unknown parameters and may make them insignificant.

When a sample contains outliers or more generally to get a sample which is free from outliers, a two-step procedure

can be applied.

i) To apply some methods which will prevent inclusion of outliers in the sample, that is, clean the data by applying some rule for outlier rejection.

ii) To apply usual statistical methods on cleaned or outlier free data.

Such procedures are called outlier resistant procedures and results obtained from such procedures are called outlier resistant conclusions.

The above procedure apperently seems to be a reasonable one but it may not be applicable in practice due to some reasons. It is not possible always to carryout two-steps procedure sequentially. For example, in multiparameter regression problems outliers are difficult to recognize unless we have reliable, robust estimates for the parameters. Even though the original batch of data along with some outliers has normal behaviour, the cleaned data may not have normal behaviour (there will be statstical errors of both kinds, false rejections and false retentions).

Due to these reasons, it is difficult to justify the performance of such outlier resistant procedure. Hence, if needed, one should prefer robust procedures over outlier preventing procedures. ∎

1.3 SOME VIEWS OF ROBUSTNESS : In this section we shall discuss some views of robustness.

i) Robustness due to change in the shape of the distribution : When there is aberration (departure from the right path) from postulated assumptions, the shape of original distribution of variable is affected. Generally, the changes in the tails of the distribution are vital or dominating. The tails are either shortened or lengthened. Usually, the lengthening the tails of the underlying distribution increase the variance or error in using concerned test statistic increases significantly. On the other hand, shortening the tails of the underlying distribution produces quite negligible effects on the distributions of the estimates (Huber(1981),P.4).

A test procedure which is insensitive to such small changes in the shapes of distributions is a robust one.

ii) Robustness due to contamination of distributions : Many times, it is not easy to justify the exact distributional form of underlying variable. In such cases, one can mix a distribution whose exact form is known with actual distribution. Such mixing is called the contaminalion of the distributions. The proportion in which exactly known distribution is mixed is called the degree of contamination.

One obvious question arises, when one can ignore the

effect of contamination on behaviour of test?. When contamination effects are significant, one must look for procedures which will be insensitive to small degrees of contamination. Such procedures are also called robust procedures.

iii) Let us consider an estimator T_n (for a fixed sample size n) and a distribution F. T_n has a certain distribution $F^{(T_n)}$ under F (that is, if all observations are distributed independently according to F). Robustness of T_n required that the distribution of T_n changes slightly whenever the distribution F is change slightly. This slight change in distribution F may correspond to

i) The diffusion of a small probability mass over an arbitrary range.

ii) The diffusion of the whole mass into a small neighborhood(Refe.Hample(1968),P.6). ∎

In the following we shall define a robust sequence.

Definition(1.3.1): Robust Sequence: A sequence of estimators $\{T_n\}$ is robust (qualitatively robust) at a probability measure F iff for every $\varepsilon>0$ there exists $\delta>0$ such that for all G in class of distributions and for every n:

$$\Pi(F,G)<\delta \Rightarrow (F^{(T_n)},G^{(T_n)}) <\varepsilon, \qquad (1.3.1)$$

where Π is a suitable distance measure.

We illustrate this definition as follows.

Let distance measure be

$$\Pi(F_{\theta_1}, F_{\theta_2}) = \left|\frac{\theta_1-\theta_1}{\sigma}\right|$$

Let $x \sim N(0,\sigma^2) \equiv F_0$

$y \sim N(\theta,\sigma^2) \equiv F_\theta$

and

$T_n = \bar{x}$ = sample mean.

Therefore,

$$F_{0,T_n} \equiv F_{T_n} \sim N(0,\frac{\sigma^2}{n})$$

and

$$F_{\theta,T_n} \sim N(\theta,\frac{\sigma^2}{n})$$

$$\Pi(F_0,F_\theta) = \frac{|\theta|}{\sigma} < \varepsilon$$

and

$$\Pi(F_{0,\bar{x}}\ ,\ F_{\theta,\bar{x}}) = \frac{|\theta|}{\sigma/\sqrt{n}} = \frac{|\theta|\sqrt{n}}{\sigma} < \varepsilon$$

This implies,

$$|\theta| < \frac{\varepsilon}{\sqrt{n}}\,\sigma$$

Therefore,

$$\frac{\theta}{\sigma} < \frac{\varepsilon}{\sqrt{n}} = \delta$$

observe that for $\delta < \frac{\varepsilon}{\sqrt{n}}$, the condition (1.3.1) is satisfied. Thus, $\bar{x}$ is robust at F_{θ_1}.

Definition(1.3.2): A sequence $\{T_n\}$ is robust in a neighborhood of F iff there exists $\eta>0$ such that

$\Pi(F,G)<\eta \Rightarrow \{T_n\}$ is robust at G. ∎

1.4 ILLUSTRATIVE EXAMPLES : In this section we shall discuss three examples of these the first one is to show that the median is robust as compare to the mean and the second one is to show the t-test is robust. In the third example different robust estimators are computed for Cushny and Peebles data.

Example(1.4.1): Let $x_1, x_2, \ldots, x_n$ be a sample with mean $\bar{x}$ drawn from $N(\mu,\sigma^2)$. Here estimate for the population $\hat{\mu}_{mean}$ is the sample mean and it has the property of unbiasedness for all normal populations with finite mean. We know that for normal distribution mean and median are same. Therefore, μ = mean = median. Also we know that the sample mean is affected by extreme observations. A single observation that is either to large or to small may make $\bar{x}$ worthless as an estimate of μ.

Since the sample $x_1, x_2, \ldots, x_n$ is from normal population. Occasionally something happens to the system and a wild observation is obtained, that is, suppose $x_i (i=1,2,\ldots,n)$ coming from $N(\mu,\sigma^2)$ with probability t and from $N(\mu,K\sigma^2)$ with probability (1−t), where $K>1$ and $0\leq t\leq 1$. In other words all observations have the same mean, but the errors of some are increased by a factor of $\sqrt{K}$.

Equivalently, we could say that the $x_i (i=1,2,\ldots,n)$ are i.i.d. (independent, identically distributed) with the common underlying density function

$$f(x) = tf_1(x)+(1-t)f_2(x), \qquad (1.4.1)$$

where $f_1(x)$ is the p.d.f. of $N(\mu,\sigma^2)$ and $f_2(x)$, the p.d.f. of $N(\mu,K\sigma^2)$.

Ofcourse, here also $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is again unbiased for μ. That is

$$E(x) = tE(x_1)+(1-t)E(x_2) = \alpha\mu+(1-\alpha)\mu = \mu$$

Here both μ and σ^2 are unknown and one wishes to estimate μ.

Now, $V(x) = \int (x-\mu)^2 \, d\left[tF_1+(1-t)F_2\right](x)$

$$= t\int (x-\mu)^2 dF_1(x) + (1-t)\int (x-\mu)^2 dF_2(x)$$

$$= t\sigma^2+(1-t)K\sigma^2$$

$$= \sigma^2\left[t+(1-t)K\right]$$

Therefore,

$$V_{\xi}(\bar{x}) = \frac{1}{n} V(x_i)$$

$$= \frac{\sigma^2}{n}\left[t+(1-t)K\right]$$

Thus,

$$V_{\xi}(\bar{x}) \longrightarrow \infty \text{ as } K\rightarrow\infty \qquad (1.4.2)$$

That is, if K(1−t) is large, then $V_\epsilon(\bar{x})$ is large and we see that even an occasional wild observation makes $\bar{x}$ subject to a sizable error.

We know that the sample median is a much better estimate than the mean in the presence of extreme values. In the contamination model (1.4.1), if we use M the sample median of the $x_i (i=1,2,\ldots,n)$, as an estimate of μ (which is the population median), then for large n we have

$$V(M) = E(M-\mu)^2$$

$$\approx \frac{1}{4n\left[f(\mu)\right]^2}$$

(Refe. Rohatgi(1976), P.310, Theorem 7.3.7)

But $f(\mu) = tf_1(\mu)+(1-t)f_2(\mu)$

$$= \frac{t}{\sigma\sqrt{2\pi}} + \frac{1-t}{\sigma\sqrt{2\pi K}}$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\left[t + \frac{1-t}{\sqrt{K}}\right]$$

Therefore,

$$V(M) \approx \frac{\pi\sigma^2}{2n} \quad \frac{1}{\left[t+(1-t)/\sqrt{K}\right]^2}$$

$$V(M) \approx \frac{\pi\sigma^2}{2nt^2}, \text{ as } K\longrightarrow\infty \qquad (1.4.3)$$

This implies, the estimate M will not be greatly affected by how large K is, that is, sample median will not be greatly affected in presence of a wild observation.

Now,

$$\frac{V(\bar{x})}{V(M)} = \frac{2}{\pi}\left[t+(1-t)K\right]\left[t+\frac{1-t}{\sqrt{K}}\right]^2$$

Thus, $\frac{V(\bar{x})}{V(M)} \longrightarrow \infty$, as $K \longrightarrow \infty$ (1.4.4)

Note that this ratio is independent of μ and σ^2. From (1.4.4) and using (1.4.2) and (1.4.3), we conclude that the sample median M becomes a better estimate of μ than the sampe mean. That is M is robust.

Example(1.4.2): The tests on population means (that is, student's t-tests for the mean of a normal population and for the difference between the means of two normal populations with the same variance) are rather insensitive to departures from normality. The tests on variances (that is, the χ^2-test for the variance of a normal population, the F-test for the ratio of two normal population variances) are very sensitive to departures from normality.

Let $x_1, x_2, \ldots, x_n$ be a sample from a population with mean μ and finite variance σ^2. Let $\bar{x}$ denote the sample mean and S^2, the sample variance.

where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$.

The t-statistic for testing the mean of a normal population with unknown variance is given by

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \qquad (1.4.5)$$

Here the numerator and denominator are independent. We will observe here that, if the observations are coming from normal parent population and as $n \longrightarrow \infty$, then the distribution of t itself approaches normality.

Moreover, if we are sampling from non-normal populations, the central limit theorem assures us that sample mean and that of variance will be asymptotically normally distributed and this can be shown as fallows.

We know that

ρ = correlation coefficient between $\bar{x}$ and S^2

$$= \frac{cov(\bar{x}, S^2)}{\sqrt{V(\bar{x})\ V(S^2)}}$$

Now,

$$V(\bar{x}) = \frac{\mu_2}{n}$$

and

$$V(S^2) = \frac{\mu_4 - \mu_2^2}{n} \qquad \text{(for large n)}$$

and

$$Cov(\bar{x}, S^2) = \frac{\mu_3}{n},$$

where μ_2, μ_3 and μ_4 are central moments.

Therefore,

Asymptotic correlation coefficient between $\bar{x}$ and S^2 is

$$\rho = \frac{\mu_3}{\sqrt{\mu_2(\mu_4 - \mu_2^2)}} .$$

If the non-normal population is symmetrical, then $\mu_3 = 0$ (odd order central moments are zero).

Therefore,

$$\rho = 0$$

and hence $\bar{x}$ and S^2 are exactly uncorrelated and asymptotically independent. So that normal theory will be hold for n large enough.

If $\mu_3 \neq 0$ and when μ_4 is large, then ρ will be smaller but will remain non-zero and students t-distribution itself approaches normality as $n \longrightarrow \infty$.

Thus, whatever the parent distribution the statistic (1.4.5) tends to normality, that is, for sufficiently large n, the statistic t has an asymptotic normal distribution irrespective of the population being sampled and it has finite variance. In other words we can say that, for large n, the distribution of t is independent of the form of the population and hence t-test is robust.

It can show that the χ^2-test is not robust (Refe.Rohatgi (1976)).

Example(1.4.3): With the various aspects of robust statistics, we start with a simple example. Let us see the data, given in Hample(1986),P.79, by Cushny and Peebles (1905) on the prolongation of sleep by means of two drugs. For ten subjects, two different values were recorded (one for each drug). The ten pairwise differences (that is, the set of differences

between drug effects per subject) are as follows:

0.0, 0.8, 1.0, 1.2, 1.3, 1.3, 1.4, 1.8, 2.4, 4.6.

At the begining, this example considered as a normally distributed sample, but if we look at these observations, it reveals that the normality assumption is questionable, due to the occurrence of 4.6 which appears to be an outlier.

The ordinary arithmetic mean for above data equals 1.58. If we observe all these values in the data and the corresponding mean of these values, we come to the conclusion that, this mean is not a good representative of the data and hence not a robust. The other estimates which are given below are robust, although not all to the same extent. The 10%-trimmed mean(defined in third chapter) is 1.4 and the 20%-trimmed mean corresponds to the average of the middle six numbers, yields 1.33. The 10% and 20%-Winsorized means(defined in third chapter)are 1.44 and 1.36 respectively. The median (50%) equals 1.30. The Hodges-Lehman(1963) estimator H/L, which is defined as the median of all pairwise averages $(x_i+x_j)/2$ for $i,j=1,2,\ldots,10$ amounts to 1.32.

Now, by just looking at the data and making a subjective decision that an observation 4.6 is 'far away' from the other nine observations. So it is an 'outlier'. The average of the remaining nine observations is 1.24. The estimate obtained by such procedure can be considered a robust estimator.

Summarizing these results, we note that all the robust estimates range from 1.24 to 1.44, leaving a clear gap up to the arithmetic mean 1.58. In general all these robust estimators give quite different answers, because some of the estimators are more robust than others (Here the median is much more robust than the others). ∎