CHAPTER I

INTRODUCTION

1.1 Truncation and Censoring

Definition: Both forms of censoring for curtailing sample data are different from truncation, where the range of the population from which the sample drawn is restricted and the distribution within the untruncated range (after suitable normalization) is called the truncated distribution. Thus truncation is the property of distribution or population where as censoring is a property of sample.

Censoring may be done both at right and left. Censored samples are categorized in two ways. If the observations below or above at a given point (censoring point) may be censored, the sample consisting of uncensored observations is called Type I censored sample. If (n−k) the smallest or the greatest observations out of a sample of size n may be censored, it is called Type II censored sample.

Suppose we put n electric bulbs for their life testing and terminate the experiment when the preassigned number of bulbs, say r (<n) have failed, the samples obtained from such an experiment are called failure censored samples, denoted by $x_{(1)} < x_{(2)} < \ldots < x_{(r)}$ and the fact is that (n−r) bulbs have survived beyond $x_{(r)}$. In this case the number of bulbs that are failed is fixed and the point or the time at which the experiment terminates is a random variable. This is an example of Type II censored sample.

If the time of termination or the point at which the experiment terminates is fixed and the number of bulbs that have failed before a specified time is a random variable, the sample consisting of failure censored observations is called Type I censored sample.

Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be order statistics of complete sample such that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. If the order statistics $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n-r_2)}$ are available, while the largest $r_2$ observations are missing, it is called Right
censored sample. If only order statistics \( x_{(r_1+1)} \leq \ldots \leq x_{(n)} \) are available, while the smallest \( r_1 \) observations are missing, it is called Left censored sample.

Either of these cases are called singly censored sample. If only order statistics \( x_{(r_1+1)} \leq \ldots \leq x_{(n-r_2)} \) are available, where \( r_1 \) smallest and \( r_2 \) largest observations are missing, it is called doubly censored sample.

The present study is based on Type II censored samples from location and scale families.

1.2 Location and Scale Families

Definition: Let \( X \) be a Random variable (R.V) with the distribution function \( F \) and let \( F_\mu \) be the distribution function of the R.V \( X+\mu \). The family \( \mathcal{F}_L : \{ F_\mu : -\infty < \mu < \infty \} \) is called location parameter family and \( \mu \) is called location parameter. It is customary to say that \( X \) generates \( \mathcal{F}_L \).

Let \( F^*_\sigma \) be the distribution function of a r.v \( \sigma X \) where \( \sigma > 0 \). The family \( \mathcal{F}_S : \{ F^*_\sigma : \sigma > 0 \} \) is called a scale parameter family and \( \sigma \) is called scale parameter.

If \( F_{\mu,\sigma} \) is the distribution function of \( \sigma X+\mu \), then the family \( \mathcal{F}_{LS} : \{ F_{\mu,\sigma} : -\infty < \mu < \infty ; \sigma > 0 \} \) is called a location-scale parameter family, \( \mu \) is called a location parameter and \( \sigma \) is called a scale parameter.

The family of Normal distributions \( N(\mu, \sigma^2) \) and the family of Logistic distributions \( F(\mu, \sigma^2) \) are called location-scale families.

Generally location parameter has been estimated by the sample mean or sample median. Their respective asymptotic variances constitute a natural basis for a comparison of the efficiencies of these estimators. Scale parameters have been assessed by three estimators: the standard deviation, the mean deviation to the mean and the mean deviation to the median. A fourth estimator, the semi-interquartile range, has also been used, but the results
being the same with the mean deviation to the median. Due to lack of a natural definition for the scale parameter, it is appropriate to compare the estimators on the basis of their respective coefficient of variations, the ratio of their asymptotic standard deviation to their means.

1.3 Normal Distribution

**Definition:** A continuous random variable $X$ is said to have a Normal distribution, if its probability density function is of the form

$$
\phi(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right]; \quad -\infty < x < \infty; \quad -\infty < \mu < \infty; \quad \sigma > 0.
$$

and distribution function $\Phi(x) = \int_{-\infty}^{x} \phi(u) \, du$.

Here $\mu$, the mean, is a location parameter and $\sigma$, the standard deviation, is a scale parameter.

When the sample is censored, the estimation of location and scale parameters of Normal family are widely discussed by several authors with several methods of estimation of parameters including method of Least-Squares and Maximum Likelihood (ML) method.

Cohen (1950) used ML method of estimation to estimate the parameters of Normal populations from singly and doubly censored samples. Lloyd (1952) applied the theory of Least - Squares estimation to an ordered sample from the distribution depending upon location and scale parameters. Gupta (1952) derived best linear estimators ($n < 10$) for the mean and standard deviation using singly censored samples from Normal population.

He also proposed alternative estimators for $\mu$ and $\sigma$ as

$$
\hat{\mu} = \sum_{i=1}^{k} \beta_i x_i \\
\hat{\sigma} = \sum_{i=1}^{k} \gamma_i x_i
$$
The coefficients $\beta_i$ and $\gamma_i$ are calculated to construct these estimators. For large $n$, these alternative linear estimators are found to be more useful since the above coefficients are available.

Halperin (1952) proved that under mild regularity conditions the ML estimators from a singly censored samples are consistent, asymptotically normal and efficient for large sample size.

Breakwell (1953) also obtained ML estimators for singly censored samples. The asymptotic distribution of these estimators and their asymptotic bias are worked out.

Walsh (1956) obtained distribution-free estimators for the population mean and population variance using doubly censored samples.

Plackett (1958) showed that ML estimators are asymptotically linear and best linear unbiased estimators are asymptotically normal and efficient.

Harter and Moore (1968) gave iterative procedures for joint ML estimators, based on singly and doubly censored samples from Normal populations.


The other distribution which is closely connected to Normal distribution is Logistic distribution.
1.4 Logistic Distribution

**Definition:** The Logistic distribution is given by the cumulative distribution function

\[
F(x) = \left(1 + \exp \left[ -\frac{\pi}{\sqrt{3}} \left( \frac{x - \mu}{\sigma} \right) \right] \right)^{-1}
\]

for \(-\infty < x < \infty ; \ -\infty < \mu < \infty ; \ \sigma > 0, \)

... (1.4.1)

where \(\mu\) and \(\sigma\) are location and scale parameters respectively.

"If cumulative distribution functions \(\Phi(x)\) and \(F(x)\) of the standard normal and logistic distributions are compared, \(\text{Max} \mid \Phi(x) - F(x)\mid\) is about 0.0228, attained when \(x = 0.7\). This maximum may be reduced to a value less than 0.01 by changing the scale of \(x\) in \(\Phi(x)\) and using \(\Phi \left( \frac{16x}{15} \right)\) as an approximation to \(F(x)\)."

Halley (1952) has showed that the distance

\[
\delta(\beta) = \text{Max} \mid F(\beta y) - \Phi(y)\mid
\]

between the logistic distribution \(F(\beta y)\) with zero mean (\(\beta\), a scale parameter) and the standard normal distribution \(\Phi(y)\) is minimised by taking \(\beta = 0.5875\) which gives \(\delta(\beta) < 0.01,\)

Gumbel and Keeney (1950), Gumbel (1958), Gumbel and Pickands (1967) widely studied the Logistic distribution. Shah and Dave (1963) defined log-logistic distribution in a manner similar to the definition of log-normal distribution.

1.5 Definition

The ordered sequence \((x_{r_1}, ..., x_{r_n})\) of observations \((x_1 ... x_n)\) where

\[
x_{r_1} \leq x_{r_2} \leq ..... \leq x_{r_n}
\]

is called order statistics and is generally denoted by

\[
\{ x_{(1)}, x_{(2)}, \ldots, x_{(n)} \}.\]
If \( x_{(1)}, \ldots, x_{(n)} \) are sample observations from normal population with \( N(0,1) \), then

\[
E(x_{(i)}) = \frac{n!}{(i-1)! (n-i)!} \int_{-\infty}^{\infty} x [\Phi(x)]^{i-1} (1 - \Phi(x))^{n-i} \phi(x) \, dx
\]

\[
E(x_{(i)}^2) = \frac{n!}{(i-1)! (n-i)!} \int_{-\infty}^{\infty} x^2 [\Phi(x)]^{i-1} (1 - \Phi(x))^{n-i} \phi(x) \, dx
\]

\[
E(x_{(i)} x_{(j)}) = \frac{n!}{(i-1)! (j-i-1)! (n-j)!} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy [\Phi(x)]^{i-1} [\Phi(y)]^{j-i-1} (1 - \Phi(x))^{n-i} (1 - \Phi(y))^{n-j} \phi(x) \phi(y) \, dx \, dy
\]

Inter-relationships:

i) \( E(x_{(i)}) = -E(x_{(n-i+1)}) \)

ii) If \( n \) is odd, then the centre order statistic has its mean equal to zero.

iii) \( E(x_{(i)} x_{(j)}) = E(x_{(i)} x_{(j)}) = E(x_{(n-i+1)} x_{(n-j+1)}) \).

1.6 Estimation Procedures

When the sample is censored, the estimation procedures like ML method, ML type method, Least-Squares method etc. lead to complicated equations which are time consuming to solve. According to Tiku (1967), if \( r_1 - r_2 \) censored samples are from Normal population, the equations by ML method to solve are
\[
\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=r_1+1}^{n-r_2} (x_i - \mu) - \frac{r_1}{\sigma} \left[ \frac{\phi \left( \frac{x_{r_1+1} - \mu / \sigma}{\Phi (x_{r_1+1} - \mu / \sigma)} \right)}{\Phi (x_{r_1+1} - \mu / \sigma)} \right] \\
+ \frac{r_2}{\sigma} \left[ \frac{\phi \left( \frac{x_{n-r_2} - \mu / \sigma}{\Phi (x_{n-r_2} - \mu / \sigma)} \right)}{1-\Phi (x_{n-r_2} - \mu / \sigma)} \right] = 0 \quad \ldots \ldots \ (1.6.1)
\]

\[
\frac{\partial \log L}{\partial \sigma} = -\frac{m}{\sigma} + \frac{1}{\sigma^3} \sum_{i=r_1+1}^{n-r_2} (x_i - \mu)^2 - r_1 \frac{[(x_{r_1+1} - \mu / \sigma)]^2 \phi (x_{r_1+1} - \mu / \sigma)}{\Phi (x_{r_1+1} - \mu / \sigma)} \\
+ \frac{r_2 \phi (x_{n-r_2} - \mu / \sigma)}{\Phi (x_{n-r_2} - \mu / \sigma)} = 0 \quad \ldots \ldots \ (1.6.2)
\]

where \( m = n-r_1-r_2 \) and \( r_1, r_2 \) are number of observations censored below and above of potential sample size \( n \). If \( m = n \) or \( r_1 = 0 = r_2 \) then

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \\
\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2}
\]

According to Harter and Moore (1967), if censored samples are from Logistic population, the equations by ML method to solve are

\[
\frac{\partial \log L}{\partial \mu} = \left( -\frac{\pi}{\sqrt{3}} \sigma \right) \left( n-r_1-r_2 \right) - 2 \sum_{i=r_1+1}^{n-r_2} \frac{f(z_i)}{F(z_i)} - r_1 \frac{f(z_{r_1+1})}{F(z_{r_1+1})} + r_2 \frac{f(z_{n-r_2})}{1-F(z_{n-r_2})} = 0 \quad \ldots \ldots \ (1.6.3)
\]
\[
\frac{\partial \log L}{\partial \sigma} = \frac{1}{\sigma} \left\{ \sum z_i - 2 \sum z_i \frac{f(z_i)}{F(z_i)} - (n-1-r_2) - r_1 z_{r_1+1} \frac{f(z_{r_1+1})}{F(z_{r_1+1})} + r_2 z_{n-r_2} \frac{f(z_{n-r_2})}{1 - F(z_{n-r_2})} \right\} = 0 \quad \ldots \quad (1.6.4)
\]

where \( z_i = \prod_i \frac{(x_i - \mu)}{\sqrt{3} \sigma} \).

To avoid difficulties in solving the above equations, Huber (1967) suggested the method called 'ML Type Method'. In this method, he proposed to estimate location parameter \( \mu \) and scale parameter \( \sigma \) by solving the simultaneous equations

\[
\sum \psi \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = 0, \quad \sum \eta \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) = 0
\]

for \( \mu \) and \( \sigma \). Here \( \psi \) is an odd function and \( \eta \) is an even function, both being monotonic and increasing for positive arguments.

The function

\[
\psi(x) = \begin{cases} 
  x & \text{if } |x| \leq k \\
  k & \text{if } x > k \\
  -k & \text{if } x < -k.
\end{cases}
\]

One reasonable estimator for \( \sigma \) is

\[
\hat{\sigma} = \text{Med}_{1 \leq i \leq n} \frac{|x_i - \hat{\mu}|}{0.67}
\]

Thus Huber estimators for location parameter with suitable choice of \( \hat{\sigma} \) show good efficiency when compared to the trimmed means.
Besides ML methods and ML type methods, the problem of finding location parameters which are ‘Robust’ against deviations from normality have received increasing attention for the past three decades.

1.7 Robust Estimators

**Definition:** Let \((x_1, ..., x_n)\) be a sample from a population with a distribution function \(F(x)\) and let \(\hat{\mu}\) be the estimator of the unknown parameter \(\mu\).

Further, let \(\Phi(\hat{\mu}, F)\) be the sampling distribution of the estimator \(\hat{\mu}\), depending on \(F(x)\). Let \(G(x)\) be nearly valid function of \(F(x)\). The estimator \(\hat{\mu}\) is said to be Robust, if it scarcely depends upon the difference between \(F(x)\) and \(G(x)\) or the sampling distributions \(\Phi(\hat{\mu}, F)\) and \(\Phi(\hat{\mu}, G)\).

According to Rieder (1980), the estimator \(\hat{\mu}\) is said to be Robust with respect to the distribution \(G\) (and to \(F\)) if

\[
d(F, G) < \eta \Rightarrow d[\Phi(\hat{\mu}, F), \Phi(\hat{\mu}, G)] < \varepsilon \quad \cdots \quad (1.7.1)
\]

where \(\varepsilon > 0\), \(\eta > 0\) and \(d\) is the difference function. This definition is not quite realistic since the condition \(d(F, G) < \eta\) covers any distribution function \(F(x)\) in the \(\eta\)-neighbourhood of \(G(x)\).

In the above definition, the number \(\varepsilon\) remains small whatever is \(\eta\); and it remains inferior to some critical value \(\eta^*\). This value \(\eta^*\) is called the break-down point. Hence it is advised first to select \(\varepsilon\) and \(\eta\) such that

\[
\eta^* = \text{Max} \{\eta : d(F, G) < \eta\} \Rightarrow d[\Phi(\hat{\mu}, F), \Phi(\hat{\mu}, G)] < \varepsilon \quad \cdots \quad (1.7.2)
\]

Thus the Robustness of any estimator \(\hat{\mu}\) of \(\mu\) belonging to family \(F(x)\) is measured by the distance function defined between \(F\) and \(G\).

Hampel (1971) defines the Robustness in order to relate the estimators based on different sample sizes, despite some lack of independence between samples.
In estimating the location or scale parameters which are robust, the following assumptions are made.

(i) The distribution function $F(x)$ exists and symmetric
   \[ i.e., \ F(x) = F(-x) \ \forall \ x. \]

(ii) $F(x)$ is monotonic

(iii) $F(x)$ is continuous at least in some interval, say $\pm \xi_p$

(iv) The first two derivatives of $F(x)$ exist and are finite.

1.8 Robust estimation of location parameter $\mu$ when the underlying distribution is continuous

Let $(X_{(r+1)}, \ldots, X_{(n-r)})$ be the ordered sample of observations of $(X_{r+1}, \ldots, X_{n-r})$ from location - scale family $F(x; \mu, \sigma; -\infty < \mu < \infty; \sigma > 0)$; $\hat{\mu}$ be an estimator of $\mu$ and $G(x)$ be a related function of $F(x)$. One can assume $G(x)$ as empirical distribution function and $g(x)$ as its density.

Alloting weights $w_{(i)}$ to each ordered observation $X_{(i)}$ such that

(i) $w_{(i)} > 0$ and

(ii) $w_{(i)} \leq w_{(i+1)} \ \forall \ i = r+1, \ldots, n-r$

where $0 < \sum w_{(i)} < 1$, one can write $g(x)$ in terms of weight function as

\[ g(x) = \sum w_{(i)} \delta(x - X_{(i)}) \]

Here $\delta(x - X_{(i)})$ stands for Dirac function concentrated at $X_{(i)}$. Using Dirac function and weight function, the estimator $\hat{\mu}$ can be written as

\[ \hat{\mu} = \mu + \frac{1}{1!(\sum w_{(i)})} \sum w_{(i)} \psi(X_{(i)}) + \frac{1}{2!(\sum w_{(i)})^2} \sum w_{(i)} w_{(j)} \phi(x_{(i)}, x_{(j)}) \]

\[ + \ldots \ldots \]

The coefficients $\psi(X_{(i)})$ and $\phi(x_{(i)}, x_{(j)})$ are defined with respect to the distribution function $F(x)$ and are independent of empirical density.
The estimator \( \hat{\mu} \) unbiased for \( \mu \) implies that
\[
E\{ \psi(x_{(i)}) \} = \delta_i = 0 = E\{ \psi(x_{(i)} - x_{(j)}) \}.
\]
Further,
\[
\text{Var}(\hat{\mu}) = \frac{1}{n'(n'-1)(n'-2)(n'-3)} \sum_{i=r+1}^{n-r} \delta_i^2
\]
and
\[
\mu_3(\hat{\mu}) = \frac{n-r}{n'(n'-1)(n'-2)(n'-3)} \sum_{i=r+1}^{n-r} \delta_i^3
\]
where \( n' = n - 2r \), the size of uncensored sample.

Since the finite sample size properties of many proposed estimators are difficult to study analytically, the research has focused on their asymptotic properties which will provide useful approximation to the finite sample size. Most of the estimators commonly used under certain regularity conditions are asymptotically normal about the centre of symmetry, with the asymptotic variances depending upon their parent distribution. The validity of estimators are generally judged on the grounds of their asymptotic variances.

1.9 Need for Study of Order Statistics

To avoid the above difficulties in solving equations by ML method or ML Type method or Robust estimation methods, most of the researchers are concentrating on the usage of order statistics which provide valid results for the estimation of location and scale parameters. The use of order statistics saves the time since the expected values, variances and covariance of order statistics are already available.

Hastings (1947) calculated means, variances and covariances of order statistics up to sample of size 10. Godwin (1949) calculated these quantities more accurately and extend them to more decimal places. From his tables he
was able to calculate the best linear systematic statistics of standard deviation using all sample elements for sample size ≤ 10. Since the sum of the rows of the variance matrix of order statistic (symmetric form) is equal for any sample size, the best linear systematic statistics of population mean is the sample mean.

Gupta (1952) used Godwin’s tables of means, variances and covariances to find the best linear systematic statistics of the population mean and population standard deviation from singly censored samples.

Teichroew (1956) calculated the expected values of the product of ith and jth order statistics for samples of size ≤ 20 drawn from normal population. Using these values, Sarhan and Greenberg (1957) calculated the variances and covariances of order statistics for samples from normal distribution up to ten decimal places to the samples of size up to 20.

Sarhan and Greenberg (1962) calculated the coefficient $\alpha_{1i}$ and $\alpha_{2i}$ for the best linear systematic statistic of mean and standard deviation for singly and doubly censored samples of size ≤ 10 drawn from normal population. Further when $r_1 = r_2 = 0$, these tabulated results are found to be more accurate than Godwin or Gupta results.

The best linear systematic statistics for $\mu$ and $\sigma$ are

$$\hat{\mu} = \sum_{i=r_1+1}^{n-r_2} \alpha_{1i} y_{(i)}$$

$$\hat{\sigma} = \sum_{i=r_1+1}^{n-r_2} \alpha_{2i} y_{(i)}$$

Further Rossberg (1963) tabulated the expected values of standardized order statistic $x_{(i)}$ where $y_{(i)} = \mu + \sigma x_{(i)}$ up to nineteen decimal places.
Gupta and Shah (1965) have given a table for $n = 1(1) 10$ and for selected order statistics $n = 11(1) 25$ and percentage points 0.50, 0.75, 0.90, 0.975, 0.99 to four decimal places. They also provide explicit formulae for the first four moments of order statistics for samples of size $n = 1(1) 10$. Birnbaum and Dudman worked out the tables of numerical values of the mean and standard deviation to 5 decimal places of each order statistics for $n = 1(1) 10(5) 25$ and to some selected order statistics for $n = 100$. Gupta et al. (1967) calculated variances and covariances to 8 decimal places for all order statistics in samples of size $n = 10(5) 25$.

1.10 Class of Estimators for Location and Scale Parameters

If $(x_{(r+1)}^{(r+1)}, \ldots, x_{(n-r)}^{(n-r)})$ are order statistics of sample observations $(x_1, \ldots, x_n)$ from a continuous population where $r$ observations below and above are censored, the winsorized mean proposed by Dixon (1960) is

$$\hat{\mu}_w = \frac{rx_{(r+1)} + \sum_{i=r+1}^{n-r} x_{(i)} + r x_{(n-r)}}{n} \quad \ldots \quad (1.10.1)$$

and the trimmed mean proposed by Gupta (1952) is

$$\hat{\mu}_t = \frac{\sum_{i=r+1}^{n-r} x_{(i)}}{n-2r} \quad \ldots \quad (1.10.2)$$

Tukey (1963) corrected the formula (1.10.1) by considering $n$ odd or even and his modified winsorized mean is
\[ \hat{\mu}_{\text{mw}} = \frac{(r+1) x_{(r+1)} + \sum_{i=r+2}^{n-r-1} x_{(i)} + (r+1) x_{(n-r-2)}}{n} \text{ if } r < \frac{n-1}{2}; \text{ if } n \text{ even} \]

\[ = x_{(l+1)} \text{ if } n = 2l + 1 \]

Crow and Siddiqui (1967) suggested linearly weighted means for \( n \) even and \( n \) odd as

\[ \hat{\mu}_{\text{lw}} = \frac{1}{2} (l-r)^{-2} \left[ x_{(r+1)} + x_{(n-r)} \right] + 3 \left[ x_{(r+2)} + x_{(n-r-1)} \right] + \ldots \]

\[ + (2l-2r-1) \left[ x_{(l)} + x_{(l+1)} \right] \text{ for } n = 2l. \]

In the standard one-dimensional location - scale problems, medians are frequently used as being insensitive to outlying observations. The median is an estimator which has attracted a great deal of attention for it is by nature consistent. It can be much superior to mean in assessing the location of a contaminated normal distribution. One of the reasons why potential users of the median estimators are discouraged is the fact that it is difficult to evaluate on computer or at least relatively expensive in that it takes a long computer time.

Hodges and Lehmann (1967) proposed estimator \( \hat{\mu}_{\text{HL}} \) as the median of pairwise means of \( r \)-\( r \) symmetrically censored sample observations \( (x_{r+1} \ldots x_{n-r}) \).
\[ \hat{\mu}_{HL} = \text{Median of } [\bar{x}_1, \ldots, \bar{x}_l] \] \hspace{1cm} (1.10.7)

Where \( \bar{x}_i = \frac{1}{2} (x_i + x_j) \forall i, i = r+1, \ldots, n-r; l = \left( \frac{n-2r}{2} \right) \)

An estimator for scale parameter is given by the Mean Absolute Deviation (MAD) estimator as

\[ \text{MAD estimator} = \text{median } [ |x_i - \text{med}(x_i) | ] \]

A more general structure for a location estimator due to Edgeworth (1893) is

\[ \hat{\mu}_E = \left[ 5q_{1/4} + 6q_{1/2} + 5q_{3/4} \right] + 16 \hspace{1cm} (1.10.8) \]

and Tukey (1958) estimator is

\[ \hat{\mu}_T = \left[ q_{1/4} + 2q_{1/2} + q_{3/4} \right] + 4 \hspace{1cm} (1.10.9) \]

and Gastwirth (1966) estimator is

\[ \hat{\mu}_G = \left[ 3q_{1/3} + 4q_{1/2} + q_{3/3} \right] + 10 \hspace{1cm} (1.10.10) \]

Here \( q_\alpha \) stands for the empirical quantile of level \( \alpha \). All these estimators are linear combinations of the order statistics of the available sample. Among all these estimators, most of the researchers paid their attention to the deep study of winsorized estimators and Trimmed estimators.

### 1.11 Winsorized Means and Trimmed Means

Dixon (1960) showed that winsorized means are highly efficient (at least for limited censoring) and more efficient than trimmed means for sampling from normal distribution. But in practice trimmed means and linearly weighted means are intuitively more attractive than winsorized means. The relative efficiency of trimmed means with respect to Best Linear Systematic Statistics (BLSS) is never less than 0.99 and a similar study proved that
when symmetry is maintained, winsorization results in estimators of the mean whose efficiencies are scarcely distinguishable from those of Best Linear Systematic Statistics and $\hat{\mu}_w$ has a relative efficiency of 0.99912 with respect to BLSS for any $r$ up to the greatest integer $\frac{n-2}{2}$ and for $n \leq 20$.

Authors such as Lambert (1985), Welch (1987), Welch and Gutierrez (1988), Spino and Pagano (1991) made deep study on asymptotic efficiencies of trimmed means. Welch and Gutierrez (1988) considered trimmed mean as a replacement for the mean pair difference. Bickel (1965), Crow and Siddiquie (1967), Dixon and Tukey (1963), Jackel (1971) and Lehmann (1983), made comparisons among alternative robust estimators such as winsorized mean trimmed mean, median and HL estimators and suggest that a moderately trimmed mean had high efficiency when the underlying distribution is normal. Even in small sample situations Gastwirth and Cohen (1970) proved that the moderately trimmed mean is suitable if the underlying distribution is nearly normal.

The reason for many researchers emphasising trimmed mean is its robustness in particular. The trimmed mean is less sensitive to extreme deviation and heavy-tailed distribution than the ordinary sample mean. The necessary and sufficient condition for the asymptotic normality of the trimmed mean is that the trimming may be done at proportions corresponding to uniquely defined percentiles of the population distribution.

If this condition does not hold, the limiting distribution may not be normal and the use of trimmed mean may become to invalid when sampling is from discrete populations or from continuous populations with gaps or from grouped data.

Stigler (1973) derived the asymptotic distribution of trimmed mean when sampling was from any arbitrary continuous distribution.
Relationship Among $\hat{\mu}_w$, $\hat{\mu}_t$ and $\hat{\mu}_w$

(i) If sample observations are large and $r$ is small comparatively, then for symmetrical censoring, the winsorized mean is approximately equivalent to trimmed mean.

i.e., $\hat{\mu}_w = \hat{\mu}_t$ if $n$ is large and $\lim_{n \to \infty} \frac{r}{n} = 0$.

(ii) If $\hat{\mu}_w$ is given weight $\frac{1}{n}$ rather than one in averaging, then the resulting estimator $\hat{\mu}_w$ becomes $\hat{\mu}_{lw} (\frac{1}{2})$ for $n = 2l+1$

1.12 Asymptotic Properties of Trimmed Mean

If the number of observations ($r$) censored at either end are proportional to the sample size $n$, i.e., $p = r/n$ then

$$\hat{\mu}_w = \frac{[np] x_{(np+1)} + \sum_{i=[np]+1}^{n-[np]} x_{(i)} + [np] x_{(n-[np])}}{n} \quad \ldots \quad (1.12.1)$$

and

$$\hat{\mu}_t = \frac{\sum_{i=[np]+1}^{n-[np]} x_{(i)}}{n-2[np]} \quad \ldots \quad (1.12.2)$$

where $[np]$ is the greatest integer $\leq (np)$ and $p$ satisfies the inequality $0 \leq p < 1/2$.

Authors such as Dixon (1960), Huber (1964), Bickel (1965) paid attention on asymptotic properties of $\hat{\mu}_t$ in r-r symmetrically censored Type II samples.
If \((np) = [(n-1)p]\) and \(\mu = 0\), i.e., the location parameter is taken at the origin, the sensitivity of trimmed mean is observed by Bickel. He worked out minimum efficiencies with respect to the families of all symmetric unimodel distributions of the winsorized and trimmed means.

He proved that

\[
\hat{\mu}_t \sim N(\mu, s(\hat{\mu}_t)) 
\]

where

\[
s(\hat{\mu}_t) = \sqrt{\frac{SSD(\hat{\mu}_t)}{h(h-1)}} \quad \text{and} \quad h = n - 2[np],
\]

if the censored observations are from normal population with

\[
\phi(x) = N(\mu, \sigma^2) .
\]

Making use of (1.12.3) and (1.12.4) one can observe the robustness of the estimator \(\hat{\mu}_t\).

Assumptions like population being normal and the estimator following asymptotic normality are very common. In fact any model similar to Normal could have been responsible for drawing of sample at hand. Assuming that the underlying distribution is contaminated normal as

\[
f(x) = (1-\eta) N(\mu, \sigma^2) + \eta N(\alpha, \sigma^2) .
\]

with \(\eta\) being small and \((\alpha - \mu)^2 > n\sigma^2\),

\[
\phi(\hat{\mu}_t, f) = \sum_{l=0}^{h} \binom{h}{l} \eta^l (1-\eta)^{h-l} N\left[\mu + l(\alpha - \mu)/h\right], s^2(\hat{\mu}_t)/h\right].
\]

In such case the distance function

\[
d[\phi(\hat{\mu}_t, f), \Phi(\hat{\mu}_t, \phi)] \geq 1 - (1-\eta)^h \quad \text{and leads to}
\]

\[
\varepsilon \geq 1 - (1-\eta)^h > 1 - e^{-hn}\eta
\]

Hence by considering the definition of robustness, a small \(\eta\) does not necessarily imply a small \(\varepsilon\) and the trimmed mean \(\hat{\mu}_t\) is not a robust
estimator when the underlying distribution is asymmetric, contaminated normal. But the robustness can be restored if the underlying distribution is symmetric, contaminated normal. In general, there is a possibility of committing mistakes in applying statistical techniques to the data without studying the basic assumptions behind it. This has been explained by Tukey (1963) in his study on contaminated normal distribution. His model is

\[ f(x) : (1- \varepsilon) N(\mu, \sigma^2) + \varepsilon N(\mu_1, \sigma_1^2) \]

This model is called symmetric, contaminated normal if \( \mu = \mu_1 \) and asymmetric contaminated normal if \( \mu \neq \mu_1 \).