# ROBUST CONTROL CHARTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYSUN ÇETİNYÜREK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

DECEMBER 2006

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. H. Öztaş Ayhan
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Birdal Şenoğlu
Co-Supervisor

Dr. Barış Sürücü
Supervisor

**Examining Committee Members**

Assoc. Prof. Dr. İnci Batmaz (METU, STAT)

Dr. Barış Sürücü (METU,STAT)

Assoc. Prof. Dr. Birdal Şenoğlu (Ankara Unv.,STAT)

Asst. Prof. Dr. Güvenç Arslan (Başkent Unv., STAT)

Dr. Özlem İlk (METU, STAT)

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Aysun Çetinyürek

# ABSTRACT

## ROBUST CONTROL CHARTS

Çetinyürek, Aysun
M. Sc., Department of Statistics
Supervisor: Dr. Barış Sürücü
Co-Supervisor: Assoc. Prof. Dr. Birdal Şenoğlu

December 2006, 82 pages

Control charts are one of the most commonly used tools in statistical process control. A prominent feature of the statistical process control is the Shewhart control chart that depends on the assumption of normality. However, violations of underlying normality assumption are common in practice. For this reason, control charts for symmetric distributions for both long- and short-tailed distributions are constructed by using least squares estimators and the robust estimators -modified maximum likelihood, trim, MAD and wave. In order to evaluate the performance of the charts under the assumed distribution and investigate robustness properties, the probability of plotting outside the control limits is calculated via Monte Carlo simulation technique.

**Keywords:** Control chart, long-tailed symmetric distribution, short-tailed symmetric distribution, robustness, least squares estimators, modified maximum likelihood estimators, trim estimators, MAD estimators, wave estimators, outliers

# ÖZ

## SAĞLAM KONTROL GRAFİKLERİ

ÇETİNYÜREK, Aysun
Yüksek Lisans, İstatistik Bölümü
Tez Yöneticisi: Dr. Barış SÜRÜCÜ
Ortak Tez Yöneticisi: Doç. Dr. Birdal ŞENOĞLU

Aralık 2006, 82 sayfa

Kalite kontrol grafikleri, istatistiksel kalite kontrolünde en çok kullanılan araçlardan biridir. İstatistiksel kalite kontrolünün en popüler özelliği normallik varsayımına dayanan Shewhart kontrol grafikleridir. Fakat normallik varsayımının ihlali ile uygulamada sık karşılaşılır. Bu nedenle, en küçük kareler tahmin edicileri ve sağlam (robust) tahmin edicilerden uyarlanmış en çok olabilirlik, budanmış, ortalama mutlak sapma (MAD) ve wave tahmin edicileri kullanılarak uzun ve kısa kuyruklu simetrik dağılımlar için kontrol grafikleri oluşturulmuştur. Bu grafiklerin, varsayılan dağılım için performansı ve istatistiksel sağlamlık özellikleri, Monte Carlo simulasyonu ile kontrol sınırları dışında olma olasılıkları elde edilerek değerlendirilmiştir.

**Anahtar Kelimeler:** Kontrol grafikleri, uzun kuyruklu simetrik dağılım, kısa kuyruklu simetrik dağılım, istatistiksel sağlamlık, en küçük kareler tahmin edicileri, uyarlanmış en çok olabilirlik tahmin edicileri, budanmış tahmin edicileri, ortalama mutlak sapma tahmin edicileri, wave tahmin edicileri, aykırı değerler

*To my family for their everlasting love and support*,

## ACKNOWLEDGEMENTS

I want to acknowledge many people who have helped me throughout this work.

I would first like to thank my supervisors Dr. Barış Sürücü and Assoc. Prof. Dr. Birdal Şenoğlu, for their guidance, criticism, encouragement throughout this study.

I would like to thank to the members of Statistics and Computer Science Department, Başkent University for their help, support and encouragement.

A very special thank goes to my ex and new housemates Evrim, Gamze, Fatma Gül and Esen for helping me to come over the hard times.

Another special thank goes to Assoc. Prof. Dr. Atilla Halil Elhan for his kind support, encouragement and love during my master's study, he has always been more than an professor to me.

Thanks to my friend İbrahim for his help and friendship through the years we sent at METU.

Last but not the least thanks go to my family for their being with me all the time, for their constant support, tremendous patience and endless love. I am truly thankful. I could not have accomplished what I have had without you.

# TABLE OF CONTENTS

# LIST OF TABLES

TABLES

# LIST OF FIGURES

FIGURES

# CHAPTER 1

# INTRODUCTION

History of quality control goes back to 1900s. However, the usage of statistical methods begun with Dr. Walter Shewhart. After the World War II, statistical quality control gained popularity in manufacturing industries. During the war, the need for statistical techniques for quality control purposes were realized. In 1950s and 1960s, reliability engineering begun to emerge. In 1980s, United States started to use more statistical methods for quality control purposes.

Statistical process control has many advantages (Montgomery, 2001; Wheeler, 1995). These advantages can be listed as follows:

- When the process is stable, its behaviour can be predicted.
- A process in statistical control operates with less variability than a process having special causes. Lower variability has become an important tool of competition.
- A process having special causes is unstable, and the excessive variation may hide the effect of changes being introduced to achieve improvement.
- Knowing that a process is in statistical control is helpful for workers running the process.

There are many statistical methods used for quality control purposes. Control charts are online monitoring techniques used for quality control. It was in 1920s when control charts were used for the first time in the Bell laboratories. Dr. Walter

Shewhart designed a control chart for quality control purpose and since then these type of charts are known as Shewhart charts and have wide usage in statistical process control. Sower et. al. (1999) stated that "Shewhart recognized the variation as the enemy of quality". The control charts for the mean have common usage in statistical process control. "The control chart is a tool which provides a means of what type of variation is present in a process and whether a process is performing predictably" (Sower et. al. , 1999). The main function of a control chart is to define the variation present in the process and to find out if some adjustments can be done for the process.

If a product is to meet requirements, the process should be operating with little variability around the target or nominal value of the product's quality characteristic (Montgomergy, 2001). In any production process, there is an inherent variability regardless of how well the process is designed. This inherent variability cannot be economically identified and corrected (chance causes) (Alloway and Raghavachari, 1991). The process that is operating with chance causes only, is said to be **in-control** (Montgomery, 2001). If, in contrast, there are some other causes affecting the process and making the process produce nonconforming units, these causes are known as assignable causes. This variability may be due to defective material, unadjusted or unproperly controlled machines or etc. The variability caused by assignable causes is not inherent in the process and can be taken under control. A process operating with assignable causes is said to be **out-of control**. For long time, production processes will operate in-control state and produce conforming products. Eventually, some assignable causes will occur and make the process produce out-of-control state. At this time, the assignable causes are tracked down, emergent precaution is taken to encounter the causes and the process is restored to its in-control state. Control chart is the most popular technique used for this purpose. Control charts are commonly used for the reasons listed below:

- Effective in prevention of defects
- Improve productivity
- Provide diagnostic information
- Prevent unnecessary process adjustment

One of the main ideas used in control charts is rational subgrouping. Here the idea is to minimize the variability within the subgroups and maximize the variability between the subgroups, so it would be easy to detect the occurence of assignable causes. Time order is a good basis for rational subgrouping. Not only the means plotting outside the control limits but also the pattern of the sample means plotted between the control limits is important. Even if all the points are between the limits, the pattern of the points should be random without a trend.

A control chart is a graph of a quality measurement plotted against time with control lines superimposed to show statistically significant deviations from the normal level of performance (Wood, 1995). A control chart consists of a center line, an upper control limit and a lower control limit. Center line represents average value of the quality characteristic when the process is in-control. The probability of plotting outside the control limits is 0.0027 if the underlying process distribution is normal. For other process distributions this probability more or less depends on the magnitude of the tails (Chan et. al. , 1988). This procedure allows one to find unusual subgroups that may indicate sporadic problems and also to detect components of variation that are not reflected in the within group variation (Rocke, 1989). Sometimes, even if the entire points plot inside the limits, most of the points may follow a pattern or may take place below or above the center line; this is an indication of something wrong with the process (Montgomery, 2001).
Control charts are similar to hypothesis testing procedures (Chakraborti et. al., 2001). Control charts are means of testing the process for being in-control state. If a point plots outside the limits, this is equivalent to rejection of the null hypothesis. The

similarity of hypothesis testing and control charts is used to assess the performance of the control charts.

There are two types of control charts which are determined with respect to the type of distribution of the quality characteristic of interest. For quality characteristics with discrete distributions, attributes control chart; for quality characteristics with continuous distributions, variables control chart is used.

The variables control charts have the assumption that the quality characteristic of interest has normal distribution. However, normal distribution assumption is rarely met in practice (Tiku et. al. , 1986; Montgomery, 2001). For this reason some robust and nonparametric estimators are used in the literature to make the control charting procedure robust to deviations from normality assumption.

Control charts are one of the seven tools used in statistical process control (SPC). These tools are very important in SPC to make improvements in quality. There are many control charts used for the process control. These are $\bar{x}$ **c**hart**,** CUSUM charts, EWMA charts, s chart, R chart, attributes control charts, etc. Control chart for $\bar{x}$ is used for widget processes and for high value processes moving mange chart is used (Caulcutt, 1995). The main focus of this thesis is on $\bar{x}$ control chart.

### 1.1 $\bar{x}$ Control Chart

A control chart is defined as a graphical display of a quality characteristic, which is measured from a sample, versus time. The control charts are constructed in such a way that each has a central line (CL), which is the mean quality of the quality characteristic, lower control limit (LCL), and an upper control limit (UCL). In the quality control, for the quality characteristic of interest, it is assumed that the random variable X which represents the quality characteristic has normal distribution with known mean $\mu$ and variance $\sigma^2$. For a normally distributed random variable, 99.73% of the values lie within $3\sigma$ limits. That is, for a normally distributed random quality characteristic, incorrect out-of-control signal or false alarm will be generated in only 27 points out of 10,000 points. Hence, $3\sigma$ limits above and below the mean are constructed as shown below:

$$\text{LCL}=\mu-\frac{3\sigma}{\sqrt{n}}$$

$$\text{CL}=\mu \qquad (1.1)$$

$$\text{UCL}=\mu+\frac{3\sigma}{\sqrt{n}}$$

However, the parameters $\mu$ and $\sigma$ are rarely known in practice. For this reason, $\mu$ and $\sigma$ are estimated from samples. For a normal distribution, $\bar{x}$ is unbiased estimator of $\mu$ and is the most efficient estimator of $\mu$. Thus $\mu$ is replaced by $\bar{x}$. For estimating $\sigma$, there are two different estimators used, standard deviation, s and range, R. R chart was formerly more popular due to easy computation. However, the technological advances help to overcome this problem. Nowadays, it is an easy task to estimate s. $s^2$ is an unbiased estimator of $\sigma^2$, whereas $s$ is not an unbiased estimator of $\sigma$. When the distribution is normal, s estimates $c_4\sigma$, where $c_4$ depends only on n. That is, $c_4$ is tabulated constant which is used to make the scale estimator unbiased.

$$\text{E}\left(\frac{s}{c_4}\right)=\sigma .$$

Values of $c_4$ can be obtained using the following exact and the approximate formulas:

$$c_4 = \sqrt{\frac{2}{n-1}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \text{ and } c_4 \cong \frac{4(n-1)}{4n-3}, \text{ respectively.}$$

The control charts are formed as follows:

$$\text{UCL}=\bar{x}+\frac{3s}{c_4\sqrt{n}},$$

$$\text{CL}=\bar{x},$$

$$\text{LCL}=\bar{x}-\frac{3s}{c_4\sqrt{n}}.$$

Shewhart suggested the use of rational subgroups to strengthen the control limits. Thus, g subsamples are taken so that the variation between the samples is maximized and variation within the samples is minimized, and hence, assignable causes would be easily detected. Then, mean of subsample means can be written as

$$\bar{\bar{x}}=\sum_{i=1}^{g}\bar{x}_i, \qquad \bar{s}=\sum_{i=1}^{g}s_i,$$

where $\bar{x}_i$ and $s_i$ are the sample mean and standard deviation. These are unbiased estimates of the corresponding parameters. Then, the control limits are constructed as:

$$\begin{aligned} &\text{UCL}=\bar{\bar{x}}+3V(\bar{x})=\bar{\bar{x}}+\frac{3\bar{s}}{c_4\sqrt{n}},\\ &\text{CL}=\bar{\bar{x}}, \qquad (1.2)\\ &\text{LCL}=\bar{\bar{x}}-3V(\bar{x})=\bar{\bar{x}}-\frac{3\bar{s}}{c_4\sqrt{n}}. \end{aligned}$$

As all the above calculations show, the procedure of deriving control limits entirely depends on the assumption of normality. When the distribution of quality characteristic of interest is normal, the control limits are located so that the probability of a sample mean of a quality characteristic plotting outside the control limits is 0.0027. This is obtained from the fact that for a normally distributed random variable, 99.73% of the observations lie inside the 6σ limits.

Generally, Type I error rate is used to compare the chart performances. There is also another measure for comparing the performances of the chart performances, in-control average run length (ARL) value,

$$\text{In-control ARL}=\frac{1}{\alpha}, \tag{1.3}$$

where

$$\alpha = P\{\bar{x} > UCL\} + P\{\bar{x} < LCL\}.$$

For an in-control process comparing the ARL or type I errors are the same. ARL value indicates the expected number of subsamples that should be taken before an out-of control signal occurs.

In this thesis, while comparing the performances of the robust control limits and Shewhart control limits, both type I error rate and ARL are used.

# CHAPTER 2

# LITERATURE REVIEW

The fact that the distribution of many engineering processes is not normal, has been realized (Spedding and Rawlings, 1993; Castagliola, 2000). Despite this fact, the distribution of sample means follow the normal distribution. However, this is just proven for rectangular and right triangular distributions which do not show significant departures from normality (Spedding and Rawlings, 1993). Larger sample sizes are needed for more extreme population distributions to achieve normality. There is not much knowledge about the accuracy of the Central Limit Theorem with respect to sample size and the degree of non-normality (Spedding and Rawlings, 1993).

Caulcutt (1995) stated that violations of assumptions were the rule rather than the exception. Many distributions of process attributes that are likely to be encountered in practice have more or less mass in the tails than the normal distribution (Alloway and Raghavachari, 1991; Chakraborti et. al., 2001). If quality control charts are set up for samples drawn from non-normal populations using normal theory, probability levels will not equal to the true probability level (Moore, 1957). Thus performing the tests or procedures based on normality assumption, erroneous inferences could easily be made. In this case, there are different courses of action:

- Transformations
- Nonparametric Control Charts
- Robust Control Charts

Transformations are not preferred in statistics because it makes difference in the scale of the distribution, the inferences made for the transformed data are not valid for untransformed data.

For constructing nonparametric control charts the user does not need to specify any particular distribution for the underlying process. This robustness with respect to distribution could be an advantage (Chakraborti, 2004). However, nonparametric control charts perform better than parametric control charts in only certain conditions (when sampling from skewed or heavy tailed distributions) (Chakraborti, 2004). Chakraborti et. al. (2001) argued that using charts with the Hodges-Lehmann estimator, the mostly known nonparametric chart, as the control statistic was problematic, and they concluded that the Hodges-Lehmann estimator is of limited usefulness.

Robust methods are preferred due to certain reasons. Robust estimators use the original data and hence the inferences are valid for the original data and also the robust estimators provide robustness to possible distortions, e.g. outliers and contaminations, and violations of assumptions which are very common in practice. Sometimes, the assumed distribution and the true distribution may be somewhat different. Therefore, a statistical procedure should be robust under plausible alternatives. Besides, robust procedures are not based on the assumption of normality like many of the other statistical tests. Parametric procedures which are based on the assumption of normality perform worse when the distribution is not normal; however, the robust procedures which are not based on the normality assumption perform well when the distribution is not met.

In 1920s, Shewhart control chart was proposed by Dr. Walter Shewhart. Since then, they have had wide usage in industry (Montgomery, 2001).

Pearson and Haines (1935) made a research to study in when to replace standard deviation safely by range. They concluded that the standard deviation could easily be replaced by range, due to easy calculation and less loss of efficiency. However, they stated that this was made for practical purposes, and yet standard deviation was theoretically a more efficient estimator of variation.

There are many studies on robustness of the control charts. Langenberg and Iglewitz (1986) proposed a control chart using trimmed mean and trimmed mean of sample ranges for controlling a quality characteristic. Lack of resistance of $\bar{x}$ to extreme observation caused the control chart perform worse. They replaced mean of subgroup means ($\bar{\bar{x}}$) with trimmed mean of the subgroup means and $\bar{R}$ with trimmed mean of the subgroup ranges. The proposed control chart was a Shewhart type chart, but they used an adjustment factor. They found that the proposed control limits and Shewhart charts lead to similar results when deviation from normality was moderate. When extreme departures from normality are observed, the proposed control charts are quicker to detect and signal an alarm for corrective measures.

Choobineh and Ballard (1987) proposed a new heuristic method for setting limits. Their method led to same limits as Shewhart method when the underlying population was symmetric. When the underlying population was skewed, the new limits were adjusted in accordance with the direction of skewness by weighting the variance. They split the skewed probability density function (pdf) into two parts at its mean. By each segment they constructed symmetrical pdfs. The weighted variance method uses the two symmetrical distributions to establish the limits of the control chart. Their method provided an acceptable probability of coverage.

Chan et al. (1988) made a research to examine the effect of violation of the assumption of normality. They used Tukey's $\lambda$ family of symmetric distributions while comparing the probabilities lying outside 3-standard deviation and 2-standard deviation control limits. The reason for using Tukey's $\lambda$ family of symmetric distributions was that the family contains distributions with variety tail areas. The constant required for constructing the control charts are obtained for Tukey's $\lambda$ family. It was found that control charts which based on the normality assumption do not give appropriate results when the distribution is heavy tailed.

Rocke (1989) compared six procedures for control limits from the point of robustness. The six procedures were mean and range, trimmed mean and range, median of means and range, mean and IQR, trimmed mean and IQR, mean of medians and IQR. Mean and IQR is found to be the most efficient and robust of the six procedures. Robust control charts do better when there are outliers but the distribution is normal. He mentioned that when the error distribution is symmetric but long tailed and the extreme observations are not to be considered outliers, then the robust procedures used in this paper will reject too many subgroups when there is no special cause.

Alloway and Raghavachari (1991) proposed a control chart based on the Hodges-Lehmann estimator. The proposed chart was nonparametric and maintained the nominal type I error rate specified. Calculations of some control limits were simplified for certain conditions. The proposed control chart performs better than the traditional approach in the case of moderate sample sizes from long-tailed symmetric distributions. These properties make it well suited for early production runs of limited size when the distribution of the process is unknown. Not all processes are normally distributed. For this reason, nonparametric methods hold over a larger class of underlying distributions and maintain a specified type I error rate. Although calculations to determine the non-parametric estimators are usually simple, they are often tedious. The more efficient estimators tend to involve more work.

Yourstone and Zimmer (1992) examined the effect of skewness and kurtosis on the performance of control charts. A new method was proposed for non-normal distributions. They examined the effect of skewness and kurtosis on average run length (ARL) of symmetrical control charts and then proposed non-symmetrical control limits. They used Burr distribution while comparing symmetrical and non-symmetrical control limits. They concluded that kurtosis has a more dramatic effect on the in-control ARL than skewness does. This is due to the effect of kurtosis in the area of tails.

Spedding and Rawlings (1993) stated the problem of non-normality by using a system of distribution which is known as Johnson System of Distributions. The use of this family made many engineering processes to be modeled. The system allows transformations to and from normality. They concluded that in small samples where the skewness and kurtosis has large sampling errors, the use of traditional methods are recommended and it has been shown that assuming normal for a non-normal process has significant errors. Thus, they recommended the use of generalized systems of distributions such as Johnson. They also pointed out that Johnson family of distributions is more efficient than Pearson system of distributions because of transforming to and from normality.

Wu (1996) proposed asymmetric control limits for skewed process distributions. Wu's eventual goal was to reduce the scrap products falling outside the limits. The optimization was conducted based on the statistical distribution of the process and took both the control limits and specification limits into consideration.

Castagliola (2000) proposed a chart for monitoring the skew distributions. Scaled weighted variance chart is an improvement of the weighted variance method which was proposed by Choobineh and Ballard (1987). He compared the performance of the proposed control chart with Shewhart and weighted variance control chart for the process the distributions lognormal, gamma and Weibull. It was concluded that scaled weighted variance method gave a type I error closer to the nominal value and could replace the weighted variance method without increasing the amount of calculation.

Chakraborti et. al. (2001) made an overview about the nonparametric or distribution free control charts. They pointed out both advantages and disadvantages of nonparametric control charts. Their aim was that their article led to a wider acceptance of distribution free control charts among practitioners.

Chakraborti et. al. (2004) introduced some distribution free control charts. The control charts have all the same in-control run length distribution for all continuous process distributions. Distribution free or non-parametric control charts can be useful in statistical process control problems. Their advantage is that one does not need to assume a particular distribution for the process and the in-control probability calculations remains the valid for all continuous distributions. The distribution robustness is an advantage especially when the distribution is not known. Gibbons and Chakraborti (2003) showed that distribution free statistical tests could be more efficient than their parametric counterparts when sampling from skewed or heavy tailed distributions. Chakraborti et. al. examined the robustness of the charts via ARL. The charts they proposed have attractive ARL properties and may be useful where standard Shewhart charts are used.

Shore (2004) used sample estimates of the first four moments to approximate the distribution and as an alternative he used the first two moments to deliver a better representation to the underlying distribution. His aim was to address the problem of skew populations in process control. Three control charts were compared in terms of MSE and it was found that the mostly used method, Clements' method needed to be re-thinking while using.

Kocherlakota and Kocherlakota (1995) made a comparison study of 8 different control charts under normality. In their study, for constructing these control charts, seven estimators out of 25 estimators from Andrews' et. al. (1972) study and Tiku's modified maximum likelihood (MML) estimators are used. Their aim was to investigate the performance of the control charts under normality and to assess the effect of non-normality on these control charts. They compared eight control charts in terms of probability of plotting outside the control limits and width of control limits via Monte Carlo swindle technique. As would be expected, in the case that the quality characteristic has the normal distribution, least squares (LS) estimators and $\bar{x}$ chart using R have the shortest width and the probability of plotting outside the control limits is equal to the prescribed value. However, robust control charts are not good in terms of probability of false alarms and width of the intervals. Under non-normal situations, performance of estimators varies. For t distribution with 3 degrees of freedom, all control charts have poor performance when n=5. When n increases, control chart using range gives probability of false alarm equal to prescribed value. For this model, MAD has the worst performance. When t distribution with 9 degrees of freedom is used, LS, Winsorized, MML and Wave estimators have good performance. For outlier model, LS, Winsorized, MML and Wave estimators have probability of false alarm close to 0.0027. For mixture distributions, none of the control charts produce acceptable levels of false alarm rates. For slash distribution, range and winsorized charts produce false alarm close to nominal value.

Kocherlakota and Kocherlakota (1995), compared the performances of the LS, Trimmed mean and winsorized variance, MML, MAD estimators and Wave estimators (k=2.4) under normality via Monte Carlo simulation. Different from

Kocherlakota and Koceherlakota (1995), in this study, performances of control charts using the LS estimators, Trimmed mean and winsorized variance, MML, MAD and wave estimators (k=2.4) are compared for long-tailed symmetric distributions. However, we used the MML estimators specific to long-tailed symmetric family and they are different from Kocherlakota and Kocherlakota (1995). Also, control charts for short-tailed symmetric family are also compared along the same lines as the long-tailed symmetric family. In this case, all the estimators are the same with Kocherlakota and Kocherlakota (1995), but MML estimators are specific to short-tailed symmetric distribution.

Long-tailed family is a family which has kurtosis more than 3 for all $p$ and it covers symmetric distributions with kurtosis greater than 3. Short-tailed family always has kurtosis less than 3, depending on parameters, $r$ and $d$, short-tailed symmetric family covers symmetric distributions with kurtosis less than 3. By obtaining control limits for these two distributions, results will be generalized from normal to all symmetric distributions either with heavy or light tails. As will be noted in the later chapters, normal distribution is just a special case of long-tailed symmetric family for $p = \infty$.

# CHAPTER 3

# CONTROL CHARTING PROCEDURE FOR LONG-TAILED SYMMETRIC DISTRIBUTIONS

## 3.1. Long-Tailed Symmetric Distributions

Many of the processes are not normally distributed (Ferrell, 1958, Langenberg and Iglewitz, 1986). Generally, the distributions of process attributes to be encountered in practice have more mass in the tails than the normal distribution. "Noble (1951) provides an excellent motivation for heavy tailed distributions by observing that assignable causes may not always economically feasible to eliminate" (Alloway and Raghavachari, 1991). Control charts based on the assumption of normality give inaccurate results when the tails of the underlying distribution are thin or thick (Chan et. al., 1988).

For modeling heavy tailed distributions, long-tailed symmetric (LTS) distributions given by the following pdf

$$f(y,p) \propto \frac{1}{\sigma}\left\{1+\frac{(y-\mu)^2}{k\sigma^2}\right\}^{-p}, \quad -\infty < \mathbf{y} < \infty\,; \tag{3.1.1}$$

$k$=2 $p$-3 and $p \geq 2$, are used (Tiku and Suresh, 1992). Here, mean and variance of the family are given in equation 3.1.2,

$$\text{E(y)}=\mu \text{ and V(y)}=\sigma^2 . \qquad (3.1.2)$$

The kurtosis is defined as

$$\frac{\mu_4}{\mu_2^2} = 3\frac{(p-3/2)}{(p-5/2)} . \qquad (3.1.3)$$

The kurtosis values of the LTS distribution is shown in table 3.1 below.

**Table 3.1** Kurtosis values of the LTS distribution for different values of $p$

| $p$ | 2.5 | 3.0 | 3.5 | 4.0 | 5.0 | 6.0 | 10.0 | $\infty$ |
|---|---|---|---|---|---|---|---|---|
| *Kurtosis* | - | 9.0 | 6.0 | 5.0 | 4.2 | 3.86 | 3.4 | 3.0 |

For $1 \leq p < 2$, V(y) does not exist in which case $\sigma$ is a scale parameter. For $p$=1, E(y) does not exist and $\mu$ is a location parameter. The distribution of

$\sqrt{\frac{v}{k}}\left(\frac{y-\mu}{\sigma}\right)$ is the Student's t with $v=2p-1$ degrees of freedom. LTS can approximate many symmetric distributions. LTS family is a symmetric family which has different tail areas depending on the parameter $p$. For $p=\infty$, LTS family reduces to the normal distribution. For $p$=5, it has the same first four moments as the logistic density. Thus, by achieving control limits for symmetric p family, the results will be generalized from normal to alternative symmetric distributions which have kurtosis greater than 3.

In quality control charting procedure, least squares (LS) estimators are frequently used. Despite this common usage, these estimators do not provide robustness to deviations from underlying assumptions. For this reason, modified maximum likelihood (MML) estimators, trimmed mean and winsorized variance (TRIM), MAD estimators and wave (W24) estimators are used for studying the behavior of the control charts under the assumed model and for deviations from the assumed model. In this study, four estimators' performances are compared with LSE. The reason why these estimators are chosen is that W24, TRIM, MML and MAD estimators are used for LTS distributions and they provide robustness.

### 3.2. Least Squares Estimators

LS estimators of μ and σ are given in equation 3.2.1.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \qquad s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}. \tag{3.2.1}$$

$\bar{y}$ is an unbiased estimator of μ and has variance V( $\bar{y}$ )= $\sigma^2 / n$ . $\bar{y}$ is fully efficient and an ideal estimator of μ under normality assumption. Besides, $s^2$ is an unbiased estimator of $\sigma^2$.

### 3.3. Modified Maximum Likelihood Estimators

For the $p$ family (or long-tailed symmetric family), ML estimators can not be obtained because of the intractable terms in the likelihood equation. They can be solved by iterative methods but these methods may be problematic, especially for small sample sizes. In this case, MML estimators are obtained. The MML estimators are:

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i y_{(i)} / m \, , \left( m = \sum_{i=1}^{n} \beta_i \right), \tag{3.3.1}$$

$$\hat{\sigma} = \left\{ B + \sqrt{B^2 + 4nC} \right\} / 2\sqrt{n(n-1)} \tag{3.3.2}$$

where

$$M = \frac{2pm}{k}, \tag{3.3.3}$$

$$B = \frac{2p}{k} \sum_{i=1}^{n} \alpha_i y_{(i)} \text{ and } C = \frac{2p}{k} \sum_{i=1}^{n} \beta_i (y_{(i)} - \hat{\mu})^2 = \frac{2p}{k} \left\{ \sum_{i=1}^{n} \beta_i y_{(i)}^2 - m\hat{\mu}^2 \right\}.$$

In order to obtain MML estimators, coefficients obtained from the expected values $t_i$ are

$$\alpha_i = \frac{(2/k)t_i^{\,2}}{\left[1 + (1/k)t_i^2\right]^2} \text{ and } \beta_i = \frac{1 - (1/k)t_i^2}{\left[1 + (1/k)t_i^2\right]^2}. \tag{3.3.4}$$

The values of $t_i$ are available in Tiku and Kumra (1981) for p=2(0.5)10, $n \le 20$; Vaughan (1992b) for $p$ =1.5, $n \le 20$.

For $p$ =1 (Cauchy distribution), the expected values and variances and covariances of the order statistics $y_{(i)}$, $3 \le i \le n-1$ ($n \ge 6$); are given in Vaughan (1994), the expected values of the first two (and the last two) order statistics are infinite (Tiku and Akkaya, 2004).

It is shown that $\hat{\mu}$ is asymptotically fully efficient and it is also unbiased for all n. $s^2/M$ is the asymptotic variance of $\hat{\mu}$ (Tiku and Akkaya, 2004).

The estimator $\hat{\sigma}$ is real and positive provided that $\beta_i \geq 0$ for all i=1,2,….,n, and is highly efficient for all n. In fact, $\hat{\sigma}$ is asymptotically fully efficient (Tiku and Suresh, 1992; Vaughan 1992a) . If $\beta_i$<0 for some i, $\hat{\sigma}$ can cease to be real. Since $t_i$ is decreasing sequence of values until the middle value and $|t_{(i)}| = |t_{(n-i+1)}|$, it follows that if $\beta_i$ is positive, then all the remaining $\beta_i$ coefficients are positive. For small values of $p$ ( $p \leq 3$) and large n, however, a few $\beta_i$ coefficients can be negative as a result of which $\hat{\sigma}$ can cease to be real and positive. If $\beta_i < 0$ for some i, then coefficients $\alpha_i$ and $\beta_i$ are replaced by (Tiku et. al., 2000)

$$\alpha_i^* = 0 \text{ and } \beta_i^* = 1/\left\{1 + (1/k)t_{(i)}{}^2\right\}. \quad (3.3.5)$$

The constants given in equations 3.3.4 and 3.3.5 give smaller weights to tails and larger weights to middle observations and thus achieve robustness to long-tailedness.

Asymptotic Properties:

1. For large n, $\dfrac{\sqrt{M}(\hat{\mu} - \mu)}{\sigma}$ is distributed as N (0, 1).
2. $\dfrac{(n-1)\hat{\sigma}^2}{\sigma^2}$ is distributed as chi-square with n-1 degrees of freedom.

Tiku and Akkaya (2004) states that "non-normality comes from the tails and once the extreme order statistics are censored, there is hardly any difference between a normal and a non-normal sample.". Thus they recommend the use of MML estimators in the framework of robustness.

### 3.4. Trimmed Mean and Winsorized Standard Deviation

Here, $y_{(i)}$ denotes the ith order statistic in a random sample of size n. Then ρ% trimmed sample mean is given by,

$$\hat{\mu}_T = \frac{\left(\sum_{i=r+1}^{n-r} y_{(i)}\right)}{(n-2r)}. \tag{3.4.1}$$

Here r=[$\rho n$ +0.5] where [a] is the greatest integer less than or equal to a. When ρ=0.1, $\hat{\mu}_T$ is called 10% trimmed sample mean. The estimator of σ which matches with $\hat{\mu}_T$ is

$$\hat{\sigma}_T^2 = \frac{1}{n-2r-1}\left\{\sum_{i=r+1}^{n-r}(y_i - \hat{\mu}_T)^2 + r\left[(y_{r+1}-\hat{\mu}_T)^2 + (y_{n-r}-\hat{\mu}_T)^2\right]\right\}. \tag{3.4.2}$$

Huber (1970) showed that under certain regularity conditions, $\lim_{n\to\infty}(\sqrt{(n-2r)\,\mathrm{var}(\hat{\mu}_T)} / E(\hat{\sigma}_T) = 1$ for fixed q=r/n.

### 3.5. Median and Median Absolute Deviation

MAD estimators of m and s are given in 3.5.1,

$$T = median\ (y_i) \text{ and } S = MAD = median\ \{|y_i - T|\}. \tag{3.5.1}$$

Median is defined as the middle value of a sorted data set. Calculation of median depends on the sample size as given below:

$$Median = \begin{Bmatrix} x_{\frac{n+1}{2}} & if\ n\ odd \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n+2}{2}}}{2} & if\ n\ even \end{Bmatrix}.$$

### 3.6. Wave Estimators

Huber M-estimators are proposed under the assumption that the underlying distribution is a symmetric and long-tailed distribution. Huber M estimators use different functional forms. To illustrate, W24 estimators are a special form of Huber estimators obtained by using the wave function given below:

$$\xi(z) = \begin{Bmatrix} \sin(z) & if\ |z| \le \pi \\ 0 & if\ |z| > \pi \end{Bmatrix}.$$

W24 estimators are given in equations 3.6.1 and 3.6.2 below:

$$T = T_0 + k * S_0 \tan^{-1}\left[\sum \sin(z_i) / \sum \cos(z_i)\right], \tag{3.6.1}$$

$$S^2 = (k * S_0)^2 n \left[\sin^2(z_i) \left\{\sum \cos(z_i)\right\}^2\right], \tag{3.6.2}$$

$k$ =2.4 and only the terms that ensure the condition $|z_i| < \pi$ where $z_i = (y_i - T_0)/(k * S_0)$, are included in the summations in the equation 3.6.1 and 3.6.2. Here $T_0$ and $S_0$ represent the median and Median Absolute Deviation as given in Section 3.5.

Other popular Huber M-estimators are bisquare estimators (BS82) and Hampel estimators (H22). W24, BS82 and H22 are found to be the most efficient of the Huber M-estimators as a result of a simulation study designed by Gross (1976). W24, BS82 and H22 have similar efficiency and robustness properties; hence only W24 is used in our simulation study.

W24 estimators are remarkably efficient for heavy tail distributions but not for skew or short-tailed distributions (Tiku and Akkaya, 2004). W24 provides robustness since it censors observations implicitly and the number of observations censored is not known. For symmetric distributions, Huber estimators of $\mu$ are unbiased and uncorrelated with the matching estimators of $\sigma$ (Tiku et. al. , 1986).
TRIM and W24 estimators are equally efficient for LTS distributions (Tiku et. al., 1986).

### 3.7. Construction of Control Limits

Let T denote a robust location estimator and S the corresponding scale estimator for any estimation method, the control limits given in equation 1.2 become

$$UCL = \overline{T} + 3\sqrt{V(T)},$$
$$CL = \overline{T}, \qquad (3.7.1)$$
$$LCL = \overline{T} - 3\sqrt{V(T)}.$$

Here V(T) can be calculated as $\frac{cS}{A\sqrt{n}}$ where $A$ is the constant to make the scale estimator unbiased and c is the adjusting factor for standard deviation of location parameter. The reason for using c and A in the definition of *V(T)* is that S is not unbiased estimator of $\sigma$ and V(T) is not equal to the exact variance of location estimator T for small samples.

$$UCL = \overline{T} + 3\frac{c\overline{S}}{A\sqrt{n}},$$
$$CL = \overline{T}, \tag{3.7.2}$$
$$LCL = \overline{T} - 3\frac{c\overline{S}}{A\sqrt{n}}.$$

The estimators of $\mu$ and $\sigma$, *T* and *S*, are calculated for each method given in Sections 3.2-3.6. To obtain the control limits, constant *A* particular to estimators are obtained via simulation by using the fact that *E(S/A)=σ*. In each case, a sample of size n is taken and value of *S*, the scale estimator, is determined. The constant *A* is found by averaging these values of *S* over 10,000 repetitions.

## 3.8. Simulation Study

The simulation study is performed for sample of sizes n=5, 7, 10, 15 and 20. In each simulation, number of subgroups denoted by g is taken as 20. Random numbers are generated from LTS distribution in equation 3.1.1 where σ=1.0, using t-distributed random variables. Simulation results are obtained over 10,000 repetitions. The estimation techniques taken into account are LS, MML, TRIM, MAD and W24. The formulas for calculating these estimators are explained in Sections 3.2-3.6.

In this simulation study, control charts are constructed using the simulation procedure below and probability of plotting outside the control limits (Type I error) and in-control average run length (ARL) values are computed.

Simulation procedure consists of the steps given below:

1. g samples of size n are generated.
2. For each sample, location and scale parameters, T and S, are calculated for all estimation techniques.
3. Averages of each of the estimators are obtained over g groups.
4. The control limits are formed as follows:

$$\text{UCL}=\overline{T}+3\frac{c\overline{S}}{A\sqrt{n}},$$

$$\text{CL}=\overline{T},$$

$$\text{LCL}=\overline{T}-3\frac{c\overline{S}}{A\sqrt{n}}.$$

5. Steps 1-4 are repeated N= 10,000 times. Thus, there are 10,000 control limits for each estimation procedure. Means of UCL and LCL are obtained.
6. A further independent set of 10,000 samples of size n are generated which is called as testing sample.
7. Mean, $T$, of each testing sample is calculated.
8. To study the performances of control charts, the tail probabilities are calculated by comparing mean of each comparison sample with the mean of control limits obtained before using the following formula:

$$\alpha_i = P\{T_i > UCL\} + P(T_i < LCL) .$$

9. ARL values are calculated by taking the inverse of Type I error rates obtained in this step using equation 1.2.

Type I error rates and ARL values are not only calculated for some selected values of $p$ (i.e. $p$ =2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 10.0) and n but also for some plausible models. These alternative models are given in Section 3.8.

Firstly, expectations of scale estimators should be calculated before beginning the simulation study in order to make the scale estimators unbiased estimators of σ. Expectations of scale estimators are obtained via simulation by calculating the mean of each scale estimator. These constants are given in Table 3.2. Then, constants for adjusting the variance of location estimators for small sample sizes are obtained and they are given in Table 3.3. After obtaining the constants, type I error rates and ARL values are obtained by simulation procedure described in this section. Type I error rates and ARL values are given in Table 3.4.

**Table 3. 2** Expectations of Scale Estimators Using LTS Distributions for Parameter $p$

| | n= | 5 | 7 | 10 | 15 | 20 |
|---|---|---|---|---|---|---|
| p=2.5 | LSE | 0.8748 | 0.9078 | 0.9289 | 0.9465 | 0.9559 |
| | MML | 1.0959 | 1.0899 | 1.0880 | 1.0857 | 1.0683 |
| | TRIM | 0.7493 | 0.8343 | 0.7866 | 0.8692 | 0.8604 |
| | MAD | 0.4648 | 0.4905 | 0.4974 | 0.5086 | 0.5125 |
| | W24 | 0.7029 | 0.7641 | 0.7880 | 0.8144 | 0.8258 |
| p=3.0 | LSE | 0.9060 | 0.9258 | 0.9519 | 0.9620 | 0.9698 |
| | MML | 1.0660 | 1.0639 | 1.0538 | 1.0528 | 1.0514 |
| | TRIM | 0.7466 | 0.8291 | 0.7868 | 0.8683 | 0.8581 |
| | MAD | 0.4950 | 0.5167 | 0.5323 | 0.5429 | 0.5481 |
| | W24 | 0.7444 | 0.7985 | 0.8382 | 0.8625 | 0.8734 |
| p=3.5 | LSE | 0.9137 | 0.9380 | 0.9591 | 0.9702 | 0.9747 |
| | MML | 1.0424 | 1.0410 | 1.0398 | 1.0308 | 1.0232 |
| | TRIM | 0.7446 | 0.8316 | 0.7860 | 0.8701 | 0.8603 |
| | MAD | 0.5086 | 0.5412 | 0.5530 | 0.5657 | 0.5705 |
| | W24 | 0.7618 | 0.8268 | 0.8642 | 0.8909 | 0.8997 |
| p=4.0 | LSE | 0.9204 | 0.9437 | 0.9563 | 0.9748 | 0.9780 |
| | MML | 1.0315 | 1.0286 | 1.0260 | 1.0253 | 1.0194 |
| | TRIM | 0.7443 | 0.8329 | 0.7846 | 0.8714 | 0.8596 |
| | MAD | 0.5170 | 0.5465 | 0.5600 | 0.5809 | 0.5841 |
| | W24 | 0.7731 | 0.8364 | 0.8695 | 0.9091 | 0.9161 |
| p=5.0 | LSE | 0.9265 | 0.9448 | 0.9602 | 0.9770 | 0.9802 |
| | MML | 1.0040 | 1.0172 | 1.0160 | 1.0157 | 1.0149 |
| | TRIM | 0.7403 | 0.8290 | 0.7886 | 0.8698 | 0.8614 |
| | MAD | 0.5283 | 0.5556 | 0.5751 | 0.5950 | 0.5999 |
| | W24 | 0.7872 | 0.8481 | 0.8887 | 0.9243 | 0.9327 |
| p=6.0 | LSE | 0.9305 | 0.9484 | 0.9650 | 0.9774 | 0.9814 |
| | MML | 0.9985 | 1.0050 | 1.0122 | 1.0102 | 1.0010 |
| | TRIM | 0.7454 | 0.8303 | 0.7864 | 0.8696 | 0.8575 |
| | MAD | 0.5357 | 0.5654 | 0.5849 | 0.6042 | 0.6087 |
| | W24 | 0.7941 | 0.8589 | 0.9003 | 0.9332 | 0.9426 |
| p=10.0 | LSE | 0.9352 | 0.9551 | 0.9667 | 0.9809 | 0.9844 |
| | MML | 0.9741 | 0.9854 | 0.9909 | 0.9987 | 1.0000 |
| | TRIM | 0.7442 | 0.8228 | 0.7858 | 0.8679 | 0.8600 |
| | MAD | 0.5437 | 0.5774 | 0.5978 | 0.6175 | 0.6243 |
| | W24 | 0.8049 | 0.8754 | 0.9121 | 0.9491 | 0.9592 |

**Table 3.3** Constants for Adjusting the Standard Deviation of Location Parameter for LTS Distribution

| *p* | *n* | *MML* | *TRIM* | *MAD* | *WAVE* |
|---|---|---|---|---|---|
| 2.5 | 5 | 0.9008 | 1.2518 | 0.9268 | 0.9689 |
| | 7 | 0.8800 | 1.0036 | 0.8994 | 0.9560 |
| | 10 | 0.9154 | 0.8792 | 0.8819 | 0.9182 |
| | 15 | 0.8867 | 0.9382 | 0.8687 | 0.9477 |
| | 20 | 0.8852 | 0.8667 | 0.8695 | 0.9326 |
| 3.0 | 5 | 0.9424 | 1.2710 | 0.9698 | 1.0213 |
| | 7 | 0.9273 | 1.0467 | 0.9505 | 1.0305 |
| | 10 | 0.9041 | 0.9172 | 0.9275 | 0.9831 |
| | 15 | 0.9260 | 0.9901 | 0.9146 | 1.0121 |
| | 20 | 0.9282 | 0.9037 | 0.9100 | 0.9966 |
| 3.5 | 5 | 0.9576 | 1.2777 | 0.9830 | 1.0437 |
| | 7 | 0.9587 | 1.0650 | 0.9758 | 1.0675 |
| | 10 | 0.9469 | 0.9489 | 0.9587 | 1.0276 |
| | 15 | 0.9226 | 1.0022 | 0.9290 | 1.0490 |
| | 20 | 0.9296 | 0.9260 | 0.9381 | 1.0397 |
| 4.0 | 5 | 0.9766 | 1.2917 | 1.0131 | 1.0876 |
| | 7 | 0.9625 | 1.0813 | 0.9937 | 1.0836 |
| | 10 | 0.9590 | 0.9436 | 0.9575 | 1.0448 |
| | 15 | 0.9368 | 1.0212 | 0.9463 | 1.0738 |
| | 20 | 0.9477 | 0.9576 | 0.9528 | 1.0722 |
| 5.0 | 5 | 0.9923 | 1.3130 | 1.0293 | 1.1121 |
| | 7 | 0.9758 | 1.0957 | 1.0097 | 1.1182 |
| | 10 | 0.9732 | 0.9747 | 0.9876 | 1.0797 |
| | 15 | 0.9643 | 1.0429 | 0.9873 | 1.1145 |
| | 20 | 0.9704 | 0.9844 | 0.9788 | 1.1064 |
| 6.0 | 5 | 0.9874 | 1.3179 | 1.0431 | 1.1333 |
| | 7 | 0.9846 | 1.1003 | 1.0147 | 1.1326 |
| | 10 | 0.9839 | 0.9905 | 0.9982 | 1.1013 |
| | 15 | 0.9682 | 1.0724 | 0.9921 | 1.1426 |
| | 20 | 0.9863 | 0.9752 | 0.9791 | 1.1231 |
| 10.0 | 5 | 0.9965 | 1.3347 | 1.0639 | 1.1633 |
| | 7 | 0.9915 | 1.1221 | 1.0400 | 1.1729 |
| | 10 | 0.9951 | 1.0113 | 1.0262 | 1.1366 |
| | 15 | 0.9885 | 1.0871 | 1.0059 | 1.1823 |
| | 20 | 0.9975 | 1.0011 | 1.0104 | 1.1609 |

**Table 3.4** Type I Errors and ARL Values for LTS Distribution ( $g$ =20)

| *n=* | *5* | | *7* | | *10* | | *15* | | *20* | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Control Chart* | *Type I error* | *ARL* | *Type I error* | *ARL* | *Type I error* | *ARL* | *Type I error* | *ARL* | *Type I error* | *ARL* |
| | | | | | *p=2.5* | | | | | |
| *LSE* | 0.0089 | 112.4 | 0.0075 | 133.3 | 0.0066 | 151.9 | 0.0047 | 214.3 | 0.0044 | 228.3 |
| *MML* | 0.0058 | 172.4 | 0.0048 | 208.3 | 0.0041 | 243.9 | 0.0040 | 250.0 | 0.0038 | 263.2 |
| *TRIM* | 0.0126 | 79.4 | 0.0179 | 55.8 | 0.0058 | 172.4 | 0.0164 | 61.0 | 0.0085 | 118.1 |
| *MAD* | 0.0068 | 147.1 | 0.0055 | 181.8 | 0.0048 | 208.3 | 0.0044 | 228.3 | 0.0040 | 250.0 |
| *W24* | 0.0067 | 149.3 | 0.0056 | 178.6 | 0.0047 | 212.8 | 0.0043 | 232.6 | 0.0039 | 256.4 |
| | | | | | *p=3.0* | | | | | |
| *LSE* | 0.0071 | 141.5 | 0.0060 | 166.7 | 0.0048 | 210.5 | 0.0045 | 221.3 | 0.0043 | 234.2 |
| *MML* | 0.0056 | 178.6 | 0.0047 | 212.8 | 0.0039 | 256.4 | 0.0038 | 263.2 | 0.0034 | 294.1 |
| *TRIM* | 0.0083 | 120.5 | 0.0111 | 90.1 | 0.0037 | 270.3 | 0.0098 | 102.0 | 0.0069 | 144.9 |
| *MAD* | 0.0062 | 161.3 | 0.0055 | 181.8 | 0.0048 | 208.3 | 0.0041 | 243.9 | 0.0040 | 250.0 |
| *W24* | 0.0060 | 166.7 | 0.0053 | 188.7 | 0.0041 | 243.9 | 0.0039 | 256.4 | 0.0038 | 263.2 |
| | | | | | *p=3.5* | | | | | |
| *LSE* | 0.0063 | 158.7 | 0.0045 | 222.2 | 0.0042 | 238.1 | 0.0034 | 298.5 | 0.0032 | 312.5 |
| *MML* | 0.0050 | 200.0 | 0.0046 | 217.4 | 0.0038 | 263.2 | 0.0036 | 277.8 | 0.0033 | 303.0 |
| *TRIM* | 0.0061 | 163.9 | 0.0100 | 100.0 | 0.0034 | 294.1 | 0.0077 | 129.9 | 0.0057 | 175.4 |
| *MAD* | 0.0056 | 178.6 | 0.0053 | 188.7 | 0.0049 | 204.1 | 0.0044 | 227.3 | 0.0039 | 256.4 |
| *W24* | 0.0054 | 185.2 | 0.0051 | 196.1 | 0.0047 | 212.8 | 0.0040 | 250.0 | 0.0038 | 263.2 |
| | | | | | *p=4.0* | | | | | |
| *LSE* | 0.0055 | 181.2 | 0.0045 | 224.7 | 0.0042 | 240.9 | 0.0031 | 322.6 | 0.0030 | 333.3 |
| *MML* | 0.0043 | 232.6 | 0.0041 | 243.9 | 0.0037 | 270.3 | 0.0034 | 294.1 | 0.0029 | 344.8 |
| *TRIM* | 0.0053 | 188.7 | 0.0076 | 131.6 | 0.0029 | 344.8 | 0.0068 | 147.1 | 0.0041 | 243.9 |
| *MAD* | 0.0056 | 178.6 | 0.0049 | 204.1 | 0.0044 | 227.3 | 0.0040 | 250.0 | 0.0039 | 256.4 |
| *W24* | 0.0052 | 192.3 | 0.0043 | 232.6 | 0.0041 | 243.9 | 0.0036 | 277.8 | 0.0032 | 312.5 |
| | | | | | *p=5.0* | | | | | |
| *LSE* | 0.0048 | 208.3 | 0.0041 | 243.9 | 0.0035 | 285.7 | 0.0031 | 327.8 | 0.0030 | 333.3 |
| *MML* | 0.0041 | 243.9 | 0.0037 | 270.3 | 0.0034 | 294.1 | 0.0031 | 327.8 | 0.0029 | 344.8 |
| *TRIM* | 0.0047 | 212.8 | 0.0072 | 138.9 | 0.0020 | 500.0 | 0.0054 | 185.2 | 0.0033 | 303.0 |
| *MAD* | 0.0046 | 217.4 | 0.0043 | 232.6 | 0.0039 | 256.4 | 0.0037 | 270.3 | 0.0035 | 285.7 |
| *W24* | 0.0049 | 204.1 | 0.0042 | 238.1 | 0.0037 | 270.3 | 0.0034 | 294.1 | 0.0028 | 357.1 |
| | | | | | *p=6.0* | | | | | |
| *LSE* | 0.0036 | 277.8 | 0.0035 | 285.7 | 0.0031 | 307.7 | 0.0030 | 333.3 | 0.0027 | 370.4 |
| *MML* | 0.0038 | 263.2 | 0.0036 | 277.8 | 0.0032 | 312.5 | 0.0031 | 327.8 | 0.0028 | 357.1 |
| *TRIM* | 0.0040 | 250.0 | 0.0058 | 172.4 | 0.0019 | 526.3 | 0.0051 | 196.1 | 0.0031 | 322.6 |
| *MAD* | 0.0042 | 238.1 | 0.0041 | 243.9 | 0.0038 | 263.2 | 0.0034 | 294.1 | 0.0033 | 303.0 |
| *W24* | 0.0047 | 212.8 | 0.0041 | 243.9 | 0.0037 | 270.3 | 0.0030 | 333.3 | 0.0029 | 344.8 |
| | | | | | *p=10.0* | | | | | |
| *LSE* | 0.0032 | 315.8 | 0.0029 | 344.8 | 0.0028 | 358.4 | 0.0027 | 375.0 | 0.0026 | 384.6 |
| *MML* | 0.0036 | 277.8 | 0.0032 | 312.5 | 0.0031 | 322.6 | 0.0029 | 344.8 | 0.0027 | 370.4 |
| *TRIM* | 0.0034 | 294.1 | 0.0042 | 238.1 | 0.0010 | 100.0 | 0.0044 | 227.3 | 0.0020 | 500.0 |
| *MAD* | 0.0039 | 256.4 | 0.0036 | 277.8 | 0.0034 | 294.1 | 0.0033 | 303.0 | 0.0031 | 322.6 |
| *W24* | 0.0045 | 222.2 | 0.0036 | 277.8 | 0.0033 | 303.0 | 0.0029 | 344.8 | 0.0028 | 357.1 |

LTS distribution is far from normal for small $p$. As $p$ increases LTS distribution gets closer to normal distribution. When $p$ =2.5, where LTS distribution is far from normal distribution, all the control charts using 5 different estimators show some levels of inflated Type I error rates and hence produce less ARL values for small n. While for n less than 10, MML estimators are preferable, for n greater than 10, probability of plotting outside the limits decreases to acceptable levels especially for MAD and WAVE estimators. When n equals 20, MML estimators has the best performance with ARL value 263.2 which is less than nominal ARL value, 370; other estimators which does not have good performance for smaller n, have Type I error rates much closer to the nominal value.

As seen in Table 3.4, the major effect of large kurtosis values for small sample sizes is an inappropriately large values of the false alarm rate. LTS ( $p$ =3.0,σ) has kurtosis 9, ARL values of each control chart is much less than the specified level. For smaller n, MML estimators have the best performance with probability of plotting outside the control limits close to 0.0027. However, an increase in sample size lead control charts using MAD and WAVE estimators to have better performance in terms of type I errors and ARL values. For n=20, MML, MAD and WAVE estimators are superior to other methods, where LS and TRIM estimators have greater Type I error rate and W24 has lower probability of plotting outside the limits than the nominal value.

When LTS ( $p$ =3.5, σ) is considered, where the kurtosis of the distribution is 6, the performance of control charts are poor for small n. Particularly for n less than 15, MML is superior to other estimation methods. Following MML, WV24 has the second best performance. For n greater than 15, LSE and MML has similar type I error rates which is a little larger than the specified level. Probability of plotting outside the limits is less than the nominal values for W24 estimators when n is greater than 10.

If $p$ increases to 4, the performance of all estimators improved when judged in terms of ARL even for small n. However, the probabilities of false alarm, while being reduced to a large extent, deviate considerably from the prescribed value of 0.0027 when n is small. While the best performance belongs to W24 estimator for n less than 15, MML and LSE estimators have closer ARL values to nominal value than other methods for n greater than 15.

When $p=5.0$, LTS distribution has the same first four moments with logistic distribution, LSE, W24 and MAD have the smallest type I error rates but still greater than the prespecified level for n=5. For n greater than 10, MML and W24 estimators have probabilities of type I error close to nominal value of 0.0027.

For $p=6.0$, LS estimators are superior to other methods when judged in terms of ARL and type I error rates when n is less than 10, however as n increases to 20, the performance of, MML and W24 improves to the same level of LSE obtaining type I error rates very close to nominal value of 0.0027.

LTS($p=10$, σ) is similar to normal distribution with kurtosis 3.4. In this case, the performance of LS estimators is superior to other methods which is not an interesting result. Also, performance of MML method improves as n gets closer to 20. LS, MML, TRIM and W24 estimators give acceptable levels of type I error rates for n greater than 10.

### 3.9. Robustness

Statistical inferences are based on the assumptions about the underlying situation. However these assumptions are not to be exactly true and many statistical procedures are sensitive to deviations from the assumptions. Therefore, robust procedures have been proposed. A robust estimator is one that performs well even if its assumptions are violated by the true model from which the data were generated. Robust procedures should have reasonably good efficiency at the assumed model and they should maintain, in our case, Type I error rate close to nominal value when small deviations from the model assumptions is observed. Also, robust procedures should not lead a catastrophe for larger deviations from the assumed model (Huber, 1981).

The performance of robust estimators is often better than traditional estimators for heavy tailed distributions. The estimators which are used in this study are compared in terms of robustness using the models below:

The distribution with $p$=3.5 will be taken as the population model and will be denoted by $LTS(3.5,\sigma)$.

Misspecification of the distribution:

1. $p$=3.0
2. $p$=5.0

Dixon's outlier model: n-$r$ come from $LTS(3.5,\sigma)$ and r (we do not know which) come from

3. $LTS(3.5,2\sigma)$

r=[$\rho n$]+1 where [a] is the greatest integer less than or equal to a.

Mixture Model:

4. 0.90 $LTS(3.5,\sigma)+0.10\,LTS(3.5,2\sigma)$
5. 0.90 $LTS(3.5,\sigma)+0.10\ LTS(3.5,4\sigma)$

Contamination Model:

6. 0.90$\,LTS(3.5,\sigma)$+0.10$\,LTS(\infty,\sigma)$

**Table 3.5** Type I Errors and ARL Values for LTS Distribution under Sample Models ( $g$ =20)

| *n=* | *5* | | *7* | | *10* | | *15* | | *20* | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Control Chart* | *Type I error* | *ARL* | *Type I error* | *ARL* | *Type I error* | *ARL* | *Type I error* | *ARL* | *Type I error* | *ARL* |
| *Model 1:Misspecification p=3.0* | | | | | | | | | | |
| *LSE* | 0.0057 | 175.4 | 0.0056 | 178.6 | 0.0054 | 185.2 | 0.0041 | 243.9 | 0.0030 | 333.3 |
| *MML* | 0.0051 | 196.1 | 0.0046 | 217.4 | 0.0040 | 250.0 | 0.0035 | 285.7 | 0.0026 | 384.6 |
| *TRIM* | 0.0069 | 144.9 | 0.0094 | 106.4 | 0.0039 | 256.4 | 0.0104 | 96.2 | 0.0051 | 196.1 |
| *MAD* | 0.0046 | 217.4 | 0.0043 | 232.6 | 0.0039 | 256.4 | 0.0037 | 270.3 | 0.0030 | 333.3 |
| *W24* | 0.0054 | 185.2 | 0.0045 | 222.2 | 0.0040 | 250.0 | 0.0037 | 270.3 | 0.0029 | 344.8 |
| *Model 2: Misspecification p=5.0* | | | | | | | | | | |
| *LSE* | 0.0048 | 208.3 | 0.0043 | 232.6 | 0.0037 | 270.3 | 0.0031 | 322.6 | 0.0025 | 400.0 |
| *MML* | 0.0050 | 200.0 | 0.0043 | 232.6 | 0.0035 | 285.7 | 0.0030 | 333.3 | 0.0027 | 370.4 |
| *TRIM* | 0.0037 | 270.3 | 0.0063 | 158.7 | 0.0019 | 526.3 | 0.0068 | 147.1 | 0.0043 | 232.6 |
| *MAD* | 0.0049 | 204.1 | 0.0046 | 217.4 | 0.0041 | 243.9 | 0.0037 | 270.3 | 0.0035 | 285.7 |
| *W24* | 0.0042 | 238.1 | 0.0041 | 243.9 | 0.0039 | 256.4 | 0.0034 | 294.1 | 0.0031 | 322.6 |
| *Model 3: Dixon's Outlier Model* | | | | | | | | | | |
| *LSE* | 0.0072 | 138.9 | 0.0059 | 169.5 | 0.0049 | 204.1 | 0.0043 | 232.6 | 0.0037 | 270.3 |
| *MML* | 0.0043 | 232.6 | 0.0033 | 303.0 | 0.0028 | 357.1 | 0.0027 | 370.4 | 0.0024 | 416.7 |
| *TRIM* | 0.0070 | 142.9 | 0.0087 | 114.9 | 0.0032 | 312.5 | 0.0087 | 114.9 | 0.0054 | 185.2 |
| *MAD* | 0.0055 | 181.8 | 0.0046 | 217.4 | 0.0036 | 277.8 | 0.0034 | 294.1 | 0.0031 | 322.6 |
| *W24* | 0.0054 | 185.2 | 0.0039 | 256.4 | 0.0034 | 294.1 | 0.0033 | 303.0 | 0.0031 | 322.6 |
| *Model 4: Mixture1* | | | | | | | | | | |
| *LSE* | 0.0079 | 126.6 | 0.0072 | 138.9 | 0.0058 | 172.4 | 0.0052 | 192.3 | 0.0048 | 208.3 |
| *MML* | 0.0054 | 185.2 | 0.0049 | 204.1 | 0.0043 | 232.6 | 0.0035 | 285.7 | 0.0029 | 344.8 |
| *TRIM* | 0.0087 | 114.9 | 0.0117 | 85.5 | 0.0031 | 322.6 | 0.0106 | 94.3 | 0.0045 | 222.2 |
| *MAD* | 0.0069 | 144.9 | 0.0059 | 169.5 | 0.0050 | 200.0 | 0.0049 | 204.1 | 0.0043 | 232.6 |
| *W24* | 0.0068 | 147.1 | 0.0054 | 185.2 | 0.0045 | 222.2 | 0.0040 | 250.0 | 0.0036 | 277.8 |
| *Model 5: Mixture2* | | | | | | | | | | |
| *LSE* | 0.0247 | 40.5 | 0.0201 | 49.8 | 0.0160 | 62.5 | 0.0140 | 71.4 | 0.0098 | 102.0 |
| *MML* | 0.0086 | 116.3 | 0.0058 | 172.4 | 0.0030 | 333.3 | 0.0025 | 400.0 | 0.0021 | 476.2 |
| *TRIM* | 0.0182 | 54.9 | 0.0194 | 51.5 | 0.0058 | 172.4 | 0.0168 | 59.5 | 0.0084 | 119.0 |
| *MAD* | 0.0057 | 175.4 | 0.0050 | 200.0 | 0.0045 | 222.2 | 0.0043 | 232.6 | 0.0041 | 243.9 |
| *W24* | 0.0070 | 142.9 | 0.0068 | 147.1 | 0.0062 | 161.3 | 0.0059 | 169.5 | 0.0057 | 175.4 |
| *Model 6: Contamination* | | | | | | | | | | |
| *LSE* | 0.0048 | 208.3 | 0.0047 | 212.8 | 0.0044 | 227.3 | 0.0036 | 277.8 | 0.0029 | 344.8 |
| *MML* | 0.0053 | 188.7 | 0.0043 | 232.6 | 0.0034 | 294.1 | 0.0031 | 322.6 | 0.0026 | 384.6 |
| *TRIM* | 0.0178 | 56.2 | 0.0076 | 131.6 | 0.0033 | 303.0 | 0.0076 | 131.6 | 0.0041 | 243.9 |
| *MAD* | 0.0058 | 172.4 | 0.0050 | 200.0 | 0.0045 | 222.2 | 0.0043 | 232.6 | 0.0040 | 250.0 |
| *W24* | 0.0075 | 133.3 | 0.0045 | 222.2 | 0.0040 | 250.0 | 0.0036 | 277.8 | 0.0029 | 344.8 |

As mentioned before, the control limits constructed should not be sensitive to deviations from underlying assumptions. The assumed model may always not be correct: it may be misspecified or outliers may be observed. Hence, the control limits constructed here should have robustness to such kind of situations.

For investigating the effect of misspecification of the process distribution, $p$ =3.0 and $p$ =5.0 are used. For the first misspecification model, where the population is assumed to be LTS( $p$ =3.5) and the sample come is assumed to come from LTS( $p$ =3.0), MAD estimators are superior to other methods for small n. When n increases the performance of MML, W24 and MAD are similar and they give type I error very close to nominal value.

Another scenario is that the population model is assumed to be LTS ( $p$ =3.5) whereas the sample comes from LTS ( $p$ =5.0). In this situation, all estimators give type I errors close to prespecified value. LS has the best performance for n less than 10, MML estimators and LS perform well for n greater than 10, whereas TRIM is inferior to other methods for all n.

In the outlier model, when n is small MML is superior to other methods. However, as n increases, performance of W24 and LS improves. In this case for large n, W24 and MAD estimators may be preferred since it causes a little less ARL value than the nominal value.

In the mixture model, two types of mixture distributions are used. In the first mixture model, 90% of the observations come from the assumed model, however, 10% of the observations has variance 2σ. In the second model, 10% of the observations come from long-tailed symmetric distribution with variance 4σ. In the first model, performance of MML estimators is good in terms of ARL values for n less than 20.

When n is equal to 20, MML has type I error rate close to nominal value, however, type I error rate for W24 decreases to acceptable level; besides, TRIM performs poorly for all n.

For the second mixture model, when n is small, MAD has the best performance among other estimators. As n gets larger, performance of MML, TRIM and W24 improves. For moderate sample sizes, MML is superior to other methods. For larger sample sizes, MML can be used with ARL value close to nominal value of 370.

With the disturbance induced by the contamination model from normal distribution with proportion of the contamination in the distribution being 0.1, MAD and MML seems to provide largest ARL value for small n. As n increases, performances of W24 and LS improves. The performance of LS improves to the nominal value for n equal to 20.

### 3.10. Conclusion

Heavy tailed distributions are commonly confronted in practice. If a marked departure from normality is observed, there could be a serious effect on the performance of control limits derived from normality assumption. For this reason, estimators particularly designed for heavy tailed distributions should be preferred to LS estimators especially for small values of $p$, for all sample sizes. W24 estimators perform well relative to other methods used in this study for small sample size. For large sample sizes MML and W24 methods perform well.

Hence, appropriate control chart should be chosen in order not to cause producing system give more false alarms than expected and continue properly.

# CHAPTER 4

# CONTROL CHARTING PROCEDURE FOR SHORT-TAILED SYMMETRIC DISTRIBUTIONS

## 4.1 Short-Tailed Symmetric Distribution

In the quality control literature, many robustness studies are limited to heavy tailed distributions. However, some of the process distributions have lighter tails than normal distribution. There is not much work dealing with short tailed symmetric (STS) distributions but many data sets have short tailed symmetric distributions (Akkaya and Tiku, 2005). When the data sets from quality control literature are analyzed, some data sets are found to follow STS distributions.

STS family is introduced by Tiku and Vaughan (1999) to model the samples containing inliers (Tiku et. al. , 2001).

$$f(y) \propto \frac{1}{\sigma}\left\{1+\frac{\lambda}{2r}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^r\left[1+\frac{1}{2k_1}\left(\frac{y-\mu}{\sigma}\right)^2\right]^{-p} \quad , -\infty < y < \infty \quad (4.1.1)$$

where $r$ is an integer, $\lambda = \frac{r}{r-d}, r > d, k_1 = p - \frac{3}{2}, and\ p > r + \frac{3}{2}$. $\sigma$ is scale parameter for the STS distribution family.

The mean of the distribution is zero and the variance is given by

$$\mu_2 = \frac{\left[\sum_{j=0}^{r}\binom{r}{j}\left(\frac{\lambda}{2r}\right)^j \frac{\{2(j+1)!\}}{2^{j+1}(j+1)!}\right]}{\left[\sum_{j=0}^{r}\binom{r}{j}\left(\frac{\lambda}{2r}\right)^j \frac{(2j)!}{2^j (j)!}\right]}. \tag{4.1.2}$$

STS family represents a wide variety of symmetric distributions. For a given $r$, when $d$ decreases, the kurtosis increases. While for $d$ <0, the distributions are unimodal, for $d$ >0, they are generally multimodal. The kurtosis values for this family are less than 3 for all $p$ (Tiku and Vaughan, 1999; Tiku et. al., 2001).

In this thesis, our aim is to compare the performances of robust control limits for STS distributions using LS estimators and some robust estimators.

## 4.2 Least Squares Estimators

$\bar{x}$ is calculated as usual in the normal case. However, $s^2$ is calculated using an adjustment with the square root of the variance of short tailed symmetric random variates. LS estimators are given as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and } s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)\,\mu_2}}, \tag{4.2.1}$$

where $\mu_2$ is the variance of the short tailed symmetric family and given in equation 4.1.2.

### 4.3 Modified Maximum Likelihood Estimators

The MML estimators of μ and σ are obtained exactly in the same way as the LTS distributions (Tiku and Akkaya, 2004) and they are given as

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i y_{(i)} / m , \quad \left( m = \sum_{i=1}^{n} \beta_i \right), \tag{4.3.1}$$

$$\hat{\sigma} = \left\{ -\lambda B + \sqrt{(\lambda B)^2 + 4nC} \right\} / 2\sqrt{n(n-1)} , \tag{4.3.2}$$

where

$$B = \sum_{i=1}^{n} \alpha_i y_{(i)} \text{ and}$$

$$C = \sum_{i=1}^{n} \beta_i (y_{(i)} - \hat{\mu})^2 = \frac{2p}{k} \left\{ \sum_{=1}^{n} \beta_i y_{(i)}^2 - m\hat{\mu}^2 \right\} .$$

The coefficients used in the calculation of $\hat{\mu}$ and $\hat{\sigma}$ are given in the equations below:

$$\alpha_i = \frac{(\lambda / r) t_i^3}{\left[1 + (\lambda / 2r) t_i^2\right]^2} \quad \text{and} \quad \gamma_i = \frac{1 - (\lambda / 2r) t_i^2}{\left[1 + (\lambda / 2r) t_i^2\right]^2} . \tag{4.3.3}$$

From 4.1.9, $\beta_i$ can be written as

$$\beta_i = 1 - \lambda \gamma_i . \tag{4.3.4}$$

For $\lambda \le 1$, all $\beta_i \ge 0$; and hence the coefficients $\alpha_i$ and $\beta_i$ in equations 4.3.3 and 4.3.4 are used. However, for $\lambda > 1$, some $\beta_i < 0$ (for some i), and hence $\alpha_i$ and $\beta_i$ are replaced by $\alpha_i^*$ and $\beta_i^*$ (Tiku et. al. , 2001):

$$\alpha_i^* = \frac{(\lambda / r)t_i^3 + (1 - 1/\lambda)t_i}{\left\{1 + (\lambda / 2r)t_i^2\right\}^2} \quad \text{and} \quad \gamma_i^* = \frac{(1/\lambda) - (\lambda / 2r)t_i^2}{\left\{1 + (\lambda / 2r)t_i^2\right\}^2}. \tag{4.3.5}$$

Realizing that $\alpha_i^* + \gamma_i^* z_{(i)} \cong \alpha_i + \gamma_i z_{(i)}$ since $z_{(i)} - t_{(i)} \cong 0$ (asymptotically),

$$\beta_i^* = 1 - \lambda \gamma_i^* \tag{4.3.6}$$

is always positive if $\lambda > 1$. This alternative coefficients guarantee that the MML estimator of $\sigma$ is always positive and real and higly efficient for all n. Otherwise $\hat{\sigma}$ fail to be real (Tiku and Suresh, 1992; Vaughan 1992a).

The coefficients $\beta_i (1 \le i \le n)$ have inverted umbrella ordering, such that they decrease until the middle value and then increase in a symmetric fashion. Thus the order statistics in the middle receive small weights (Tiku and Akkaya, 2004). As mentioned before, STS family is used to model the distributions which contain inliers. Giving small weight to middle order statistics is helpful, handy in acquiring robustness to inliers.

It is shown that $\hat{\mu}$ is asymptotically fully efficient and it is also unbiased for all n (Tiku and Akkaya, 2004). For large n, $\hat{\mu}$ is the minimum variance bound estimator of $\mu$ with variance $V(\hat{\mu}) \cong \frac{\sigma^2}{m}$ which is a direct result of asymptotic equivalence of likelihood and modified likelihood equations. MML estimators are known to be essentially as efficient as the LS estimators (Tan, 1985; Vaughan, 1992).

For STS distribution, calculation of TRIM, MAD and W24 estimators are explained in Chapters 3.4-3.6. In this study, LS, MML, TRIM, MAD and W24 estimators are used.

Construction of control limits are exactly the same as LTS distribution.

## 4.4 Simulation Study

Similar to LTS, n=5, 7, 10, 15 and 20 are used in the comparisons. For number of subgroups 15-30 are available in the literature (Montgomery, 2001; Wheeler, 1995). Hence, number of subgroups is taken as 20 and simulation results are obtained over 10,000 repetitions. The estimation techniques taken into account are LS, MML, TRIM, MAD and W24. The formulas for calculating these estimators are explained in Sections 4.2, 4.3 and 3.4-3.6.

The control limits for STS are obtained along the same lines LTS distribution and probability of plotting outside the control limits (Type I error) and in-control average run length (ARL) values are computed for some selected values of $r$ and $d$.

Before calculating probabilities of plotting outside the limits, expectations of scale estimators are calculated to make the scale estimators unbiased estimators of σ. Expectations of scale estimators are obtained by simulating mean of each scale estimator. These constants are given in Table 4.1. After obtaining these constants to make the scale estimators unbiased estimator of σ, Type I error rates are obtained via simulation. Type I error rates and ARL values are given in Table 4.2.

**Table 4.1** Expectations of Scale Estimators using STS with Parameters $r$ and $d$ .

| | *n=5* | *n=7* | *n=10* | *n=15* | *n=20* |
|---|---|---|---|---|---|
| | | | *r=2 d=-0.5* | | |
| *LSE* | 0.948149 | 0.964682 | 0.979187 | 0.985825 | 0.990019 |
| *MML* | 0.876329 | 0.923379 | 0.952044 | 0.962654 | 0.977416 |
| *TRIM* | 1.823936 | 1.607458 | 1.480476 | 1.419968 | 1.372860 |
| *MAD* | 0.774599 | 0.831473 | 0.876297 | 0.919198 | 0.935494 |
| *W24* | 1.125504 | 1.22971 | 1.291806 | 1.339228 | 1.355088 |
| | | | *r=2 d=0.0* | | |
| *LSE* | 0.949064 | 0.965838 | 0.978200 | 0.986098 | 0.990027 |
| *MML* | 0.879606 | 0.925729 | 0.954881 | 0.964812 | 0.979318 |
| *TRIM* | 1.922570 | 1.699013 | 1.568870 | 1.507582 | 1.460271 |
| *MAD* | 0.820690 | 0.895735 | 0.934673 | 0.956013 | 0.987924 |
| *W24* | 1.191472 | 1.298046 | 1.385436 | 1.416484 | 1.431415 |
| | | | *r=2 d=0.5* | | |
| *LSE* | 0.954072 | 0.970133 | 0.977666 | 0.987832 | 0.990810 |
| *MML* | 0.921706 | 0.952678 | 0.968368 | 0.983072 | 0.986742 |
| *TRIM* | 2.065557 | 1.819197 | 1.704690 | 1.629959 | 1.587649 |
| *MAD* | 0.876092 | 0.959068 | 1.020148 | 1.080301 | 1.106305 |
| *W24* | 1.281163 | 1.402298 | 1.466126 | 1.528138 | 1.542025 |
| | | | *r=4 d=-0.5* | | |
| *LSE* | 0.949516 | 0.965716 | 0.979127 | 0.987322 | 0.990424 |
| *MML* | 0.870615 | 0.912057 | 0.943474 | 0.964175 | 0.973664 |
| *TRIM* | 2.067250 | 1.821165 | 1.689701 | 1.619329 | 1.568485 |
| *MAD* | 0.876076 | 0.953651 | 1.003589 | 1.056566 | 1.074157 |
| *W24* | 1.280013 | 1.398822 | 1.464674 | 1.524238 | 1.54002 |
| | | | *r=4 d=0.0* | | |
| *LSE* | 0.952904 | 0.971152 | 0.979792 | 0.988287 | 0.992095 |
| *MML* | 0.877834 | 0.929433 | 0.945995 | 0.969760 | 0.974618 |
| *TRIM* | 2.168294 | 1.910452 | 1.785012 | 1.703680 | 1.652444 |
| *MAD* | 0.932458 | 1.003725 | 1.063603 | 1.081733 | 1.119905 |
| *W24* | 1.347111 | 1.472202 | 1.536487 | 1.600763 | 1.617588 |
| | | | *r=4 d=0.5* | | |
| *LSE* | 0.955704 | 0.972088 | 0.980914 | 0.988386 | 0.993087 |
| *MML* | 0.895427 | 0.933731 | 0.958700 | 0.974111 | 0.980579 |
| *TRIM* | 2.290560 | 2.017799 | 1.894165 | 1.805072 | 1.759280 |
| *MAD* | 0.977275 | 1.071531 | 1.128718 | 1.195550 | 1.224432 |
| *W24* | 1.420078 | 1.561532 | 1.626511 | 1.692053 | 1.707677 |

**Table 4.2** Constants for Adjusting the Standard Deviation of Location Parameter for STS Distribution

| | *n* | *MML* | *TRIM* | *MAD* | *WAVE* |
|---|---|---|---|---|---|
| *r=2 d=-0.5* | *5* | 1.3419 | 1.8743 | 1.7341 | 1.5394 |
| | *7* | 1.3471 | 1.5956 | 1.7714 | 1.5015 |
| | *10* | 1.3367 | 1.4424 | 1.7311 | 1.4582 |
| | *15* | 1.3228 | 1.5705 | 1.8332 | 1.4398 |
| | *20* | 1.3246 | 1.4360 | 1.8002 | 1.4176 |
| *r=2 d=0.0* | *5* | 1.4044 | 1.9515 | 1.8463 | 1.6180 |
| | *7* | 1.3981 | 1.6746 | 1.8961 | 1.5831 |
| | *10* | 1.3697 | 1.5223 | 1.8804 | 1.5442 |
| | *15* | 1.3731 | 1.6814 | 2.0196 | 1.5416 |
| | *20* | 1.3522 | 1.5447 | 1.9972 | 1.5184 |
| *r=2 d=0.5* | *5* | 1.4584 | 2.1283 | 2.0666 | 1.7871 |
| | *7* | 1.4411 | 1.8197 | 2.1361 | 1.7486 |
| | *10* | 1.4288 | 1.6580 | 2.1051 | 1.6833 |
| | *15* | 1.4073 | 1.7938 | 2.2773 | 1.6442 |
| | *20* | 1.3967 | 1.6844 | 2.2895 | 1.6491 |
| *r=4 d=-0.5* | *5* | 1.5047 | 2.1214 | 1.9868 | 1.7551 |
| | *7* | 1.5004 | 1.7880 | 2.0111 | 1.6883 |
| | *10* | 1.4964 | 1.6278 | 1.9872 | 1.6443 |
| | *15* | 1.4706 | 1.7920 | 2.1233 | 1.6387 |
| | *20* | 1.4759 | 1.6623 | 2.1037 | 1.6335 |
| *r=4 d=0.0* | *5* | 1.5518 | 2.2096 | 2.1192 | 1.8474 |
| | *7* | 1.5431 | 1.8986 | 2.1819 | 1.8074 |
| | *10* | 1.5339 | 1.7169 | 2.1340 | 1.7404 |
| | *15* | 1.5194 | 1.8687 | 2.2708 | 1.7075 |
| | *20* | 1.5074 | 1.7446 | 2.2624 | 1.7080 |
| *r=4 d=0.5* | *5* | 1.6105 | 2.3399 | 2.2776 | 1.9648 |
| | *7* | 1.5899 | 2.0300 | 2.3797 | 1.9411 |
| | *10* | 1.5458 | 1.8372 | 2.3193 | 1.8641 |
| | *15* | 1.5473 | 1.9890 | 2.4882 | 1.8175 |
| | *20* | 1.5383 | 1.8469 | 2.4739 | 1.8012 |

**Table 4.3** Type I Errors and ARL Values for STS Distribution ( $g$ =20)

| n= | 5 | | 7 | | 10 | | 15 | | 20 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Type I Error | ARL | Type I Error | ARL | Type I Error | ARL | Type I Error | ARL | Type I Error | ARL |
| *r=2 d=-0.5 (Kurtosis=2.56)* | | | | | | | | | | |
| LSE | 0.0276 | 36.2 | 0.0276 | 36.2 | 0.0275 | 36.4 | 0.0256 | 39.1 | 0.0256 | 39.1 |
| MML | 0.0030 | 333.3 | 0.0028 | 357.1 | 0.0028 | 357.1 | 0.0026 | 384.6 | 0.0025 | 400.0 |
| TRIM | 0.0695 | 14.4 | 0.0283 | 35.3 | 0.0101 | 99.0 | 0.0084 | 119.0 | 0.0063 | 158.7 |
| MAD | 0.0010 | 1000.0 | 0.0014 | 714.3 | 0.0018 | 555.6 | 0.0024 | 416.7 | 0.0026 | 384.6 |
| W24 | 0.0027 | 370.4 | 0.0031 | 322.6 | 0.0033 | 303.0 | 0.0037 | 270.3 | 0.0040 | 250.0 |
| *r=2 d=0.0 (Kurtosis=2.43)* | | | | | | | | | | |
| LSE | 0.0364 | 27.5 | 0.0358 | 27.9 | 0.0344 | 29.1 | 0.0344 | 29.1 | 0.0342 | 29.2 |
| MML | 0.0031 | 322.6 | 0.0030 | 333.3 | 0.0028 | 357.1 | 0.0027 | 370.4 | 0.0026 | 384.6 |
| TRIM | 0.0674 | 14.8 | 0.0253 | 39.5 | 0.0091 | 109.9 | 0.0078 | 128.2 | 0.0059 | 169.5 |
| MAD | 0.0010 | 1000.0 | 0.0012 | 833.3 | 0.0013 | 769.2 | 0.0015 | 666.7 | 0.0016 | 625.0 |
| W24 | 0.0039 | 256.4 | 0.0036 | 277.8 | 0.0034 | 294.1 | 0.0033 | 303.0 | 0.0031 | 322.6 |
| *r=2 d=0.5 (Kurtosis=2.26)* | | | | | | | | | | |
| LSE | 0.0503 | 19.9 | 0.0465 | 21.5 | 0.0465 | 21.5 | 0.0464 | 21.6 | 0.0462 | 21.6 |
| MML | 0.0039 | 256.4 | 0.0036 | 277.8 | 0.0034 | 294.1 | 0.0030 | 333.3 | 0.0029 | 344.8 |
| TRIM | 0.0604 | 16.6 | 0.0123 | 81.3 | 0.0091 | 109.9 | 0.0082 | 122.0 | 0.0051 | 196.1 |
| MAD | 0.0003 | 3333.3 | 0.0007 | 1428.6 | 0.0009 | 1111.1 | 0.0010 | 1000.0 | 0.0011 | 909.1 |
| W24 | 0.0040 | 250.0 | 0.0036 | 277.8 | 0.0034 | 294.1 | 0.0034 | 294.1 | 0.0031 | 322.6 |
| *r=4 d=-0.5 (Kurtosis=2.46)* | | | | | | | | | | |
| LSE | 0.0507 | 19.7 | 0.0505 | 19.8 | 0.0497 | 20.1 | 0.0492 | 20.3 | 0.0485 | 20.6 |
| MML | 0.0031 | 322.6 | 0.0030 | 333.3 | 0.0029 | 344.8 | 0.0026 | 384.6 | 0.0025 | 400.0 |
| TRIM | 0.0694 | 14.4 | 0.0345 | 29.0 | 0.0107 | 93.5 | 0.0087 | 114.9 | 0.0034 | 294.1 |
| MAD | 0.0011 | 909.1 | 0.0013 | 769.2 | 0.0015 | 666.7 | 0.0017 | 588.2 | 0.0022 | 454.5 |
| W24 | 0.0043 | 232.6 | 0.0042 | 238.1 | 0.0040 | 250.0 | 0.0031 | 322.6 | 0.0021 | 476.2 |
| *r=4 d=0.0 (Kurtosis=2.37)* | | | | | | | | | | |
| LSE | 0.0627 | 15.9 | 0.0619 | 16.2 | 0.0618 | 16.2 | 0.0614 | 16.3 | 0.0613 | 16.3 |
| MML | 0.0037 | 270.3 | 0.0034 | 294.1 | 0.0033 | 303.0 | 0.0029 | 344.8 | 0.0027 | 370.4 |
| TRIM | 0.0693 | 14.4 | 0.0289 | 34.6 | 0.0100 | 100.0 | 0.0075 | 133.3 | 0.0055 | 181.8 |
| MAD | 0.0006 | 1666.7 | 0.0010 | 1000.0 | 0.0011 | 909.1 | 0.0013 | 769.2 | 0.0014 | 714.3 |
| W24 | 0.0041 | 243.9 | 0.0038 | 263.2 | 0.0036 | 277.8 | 0.0034 | 294.1 | 0.0033 | 303.0 |
| *r=4 d=0.5(Kurtosis=2.25)* | | | | | | | | | | |
| LSE | 0.0786 | 12.7 | 0.0744 | 13.4 | 0.0741 | 13.5 | 0.0740 | 13.5 | 0.0704 | 14.2 |
| MML | 0.0042 | 238.1 | 0.0038 | 263.2 | 0.0037 | 270.3 | 0.0033 | 303.0 | 0.0029 | 344.8 |
| TRIM | 0.0646 | 15.5 | 0.0188 | 53.2 | 0.0089 | 112.4 | 0.0057 | 175.4 | 0.0036 | 277.8 |
| MAD | 0.0004 | 2500.0 | 0.0007 | 1428.6 | 0.0009 | 1111.1 | 0.0010 | 1000.0 | 0.0011 | 909.1 |
| W24 | 0.0022 | 454.5 | 0.0026 | 384.6 | 0.0031 | 322.6 | 0.0038 | 263.2 | 0.0040 | 250.0 |

A variety of short-tailed symmetric distributions are used in order to make a comparison of estimators in terms of probabilities of plotting outside the control limits and ARL values. Among these models, the one with parameters $r$ =2, $d$ =-0.5 is closest to normal distribution with kurtosis 2.56. For constant $r$ , as $d$ increases kurtosis of the distribution diminishes. For the first distribution ( $r$ =2, $d$ =-0.5), even for small n, type I error rate is close to prespecified level for MML and W24 estimators. When n gets closer to 20, type I error rate for MML gets closer to 0.0027, also the performance of other estimators improves but still much more than the nominal value. Among these estimators LS estimators has the worst performance in terms of type I error rates and MAD has good performance for n=20. For $r$=2, $d$=0.0, which has kurtosis 2.43, type I error rates obtained increases, but still MML has the best performance and W24 has good performance for all n. For large n, type I error rate MML achieves is close to nominal type I error rate, 0.0027. For $r$=2, $d$=0.5, the performance of MML worsens, however, LS method gives unacceptably high type I error rate.

For $r$=4, $d$=-0.5, short-tailed distribution has kurtosis 2.46 and close to normal distribution. For this distribution, MML has the best performance for all n, especially for n greater than 10, type I error rate for MML is close to nominal value. LS estimators have the worst performance when compared with other methods in terms of ARL values and probabilities of plotting outside the control limits.

For $r$=4, as $d$ increases to 0.0, the probability of plotting outside the control limits increases; however MML has still the best performance. Particularly, for large n (n greater than 10), type I error rates are closer to nominal value.

When $d$=0.5 and $r$=4, performance of control charts diminishes, but MML still is superior to other methods in terms of probabilities of plotting outside the limits especially large n.

### 4.5 Robustness

The main reason for studying robustness is that assuming a particular distribution and believing that it is exactly correct may lead to erroneous conclusions. As mentioned before, “robustness” provides efficiency for the assumed model, and maintains high efficiency for the plausible alternatives (Tiku et. al., 1986).

While studying the robustness of the control limits, several models are used. The process distribution is assumed to be STS ( $r$=4, $d$=-0.5). Tukey’s λ family has been widely used in robustness studies (Chan et. al., 1988) since it can cover many short-tailed distributions with kurtosis less than 3 with a variety of tail areas. Tukey’s λ family defined by transformation in (4.5.1) is used.

$$x = \left[u^l - (1-u)^l\right]/l \text{ where u is uniform(0,1).} \qquad (4.5.1)$$

This assumed model has kurtosis 2.46. Listed distributions are chosen as plausible alternatives to assumed model:

(1) $r$=4, $d$=0.0

(2) $r$=2, $d$=-0.5

(3) $l$=0.585

(4) $l$=1.0

(5) Normal(0,1)

**Table 4.4** Type I Errors and ARL Values for STS Distribution under Sample Models ( $g$ =20)

| n= | 5 | | 7 | | 10 | | 15 | | 20 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Type I Error | ARL | Type I Error | ARL | Type I Error | ARL | Type I Error | ARL | Type I Error | ARL |
| | | | | | *r=4 d=0.0 (Kurtosis=)* | | | | | |
| *LSE* | 0.0511 | 19.6 | 0.0509 | 19.6 | 0.0508 | 19.7 | 0.0493 | 20.3 | 0.0488 | 20.5 |
| *MML* | 0.0034 | 294.1 | 0.0034 | 294.1 | 0.0030 | 333.3 | 0.0029 | 344.8 | 0.0024 | 416.7 |
| *TRIM* | 0.0644 | 15.5 | 0.0232 | 43.1 | 0.0119 | 84.0 | 0.0087 | 114.9 | 0.0049 | 204.1 |
| *MAD* | 0.0005 | 2000.0 | 0.0005 | 2000.0 | 0.0012 | 833.3 | 0.0020 | 500.0 | 0.0022 | 454.5 |
| *W24* | 0.0046 | 217.4 | 0.0040 | 250.0 | 0.0037 | 270.3 | 0.0033 | 303.0 | 0.0025 | 400.0 |
| | | | | | *r=2 d=-0.5 (Kurtosis=2.43)* | | | | | |
| *LSE* | 0.0543 | 18.4 | 0.0513 | 19.5 | 0.0512 | 19.5 | 0.0475 | 21.1 | 0.0470 | 21.3 |
| *MML* | 0.0039 | 256.4 | 0.0038 | 263.2 | 0.0035 | 285.7 | 0.0035 | 285.7 | 0.0030 | 333.3 |
| *TRIM* | 0.0714 | 14.0 | 0.0289 | 34.6 | 0.0107 | 93.5 | 0.0093 | 107.5 | 0.0039 | 256.4 |
| *MAD* | 0.0011 | 909.1 | 0.0018 | 555.6 | 0.0019 | 526.3 | 0.0020 | 500.0 | 0.0024 | 416.7 |
| *W24* | 0.0033 | 303.0 | 0.0033 | 303.0 | 0.0034 | 294.1 | 0.0031 | 322.6 | 0.0026 | 384.6 |
| | | | | | *Tukey λ=0.585* | | | | | |
| *LSE* | 0.0499 | 20.0 | 0.0498 | 20.1 | 0.0497 | 20.1 | 0.0490 | 20.4 | 0.0489 | 20.4 |
| *MML* | 0.0012 | 833.3 | 0.0018 | 555.6 | 0.0020 | 500.0 | 0.0022 | 454.5 | 0.0023 | 434.8 |
| *TRIM* | 0.0600 | 16.7 | 0.0249 | 40.2 | 0.0101 | 99.0 | 0.0063 | 158.7 | 0.0043 | 232.6 |
| *MAD* | 0.0001 | 10000.0 | 0.0010 | 1000.0 | 0.0015 | 666.7 | 0.0021 | 476.2 | 0.0027 | 370.4 |
| *W24* | 0.0061 | 163.9 | 0.0055 | 181.8 | 0.0043 | 232.6 | 0.0036 | 277.8 | 0.0034 | 294.1 |
| | | | | | *Tukey λ=1.00* | | | | | |
| *LSE* | 0.0469 | 21.3 | 0.0468 | 21.4 | 0.0465 | 21.5 | 0.0449 | 22.3 | 0.0436 | 22.9 |
| *MML* | 0.0008 | 1250.0 | 0.0010 | 1000.0 | 0.0011 | 909.1 | 0.0012 | 833.3 | 0.0015 | 666.7 |
| *TRIM* | 0.0583 | 17.2 | 0.0206 | 48.5 | 0.0082 | 122.0 | 0.0042 | 238.1 | 0.0032 | 312.5 |
| *MAD* | 0.0001 | 10000.0 | 0.0004 | 2500.0 | 0.0008 | 1250.0 | 0.0016 | 625.0 | 0.0023 | 434.8 |
| *W24* | 0.0068 | 147.1 | 0.054 | 18.5 | 0.0049 | 204.1 | 0.0037 | 270.3 | 0.0025 | 400.0 |
| | | | | | *Normal (0,1)* | | | | | |
| *LSE* | 0.0558 | 17.9 | 0.0556 | 18.0 | 0.0501 | 20.0 | 0.0495 | 20.2 | 0.0485 | 20.6 |
| *MML* | 0.0058 | 172.4 | 0.0054 | 185.2 | 0.0047 | 212.8 | 0.0044 | 250.0 | 0.0033 | 303.0 |
| *TRIM* | 0.0756 | 13.2 | 0.0313 | 31.9 | 0.0135 | 74.1 | 0.0098 | 102.0 | 0.0043 | 232.6 |
| *MAD* | 0.0017 | 588.2 | 0.0019 | 526.3 | 0.0020 | 500.0 | 0.0021 | 476.2 | 0.0024 | 416.7 |
| *W24* | 0.0038 | 263.2 | 0.0033 | 303.0 | 0.0030 | 333.3 | 0.0028 | 357.1 | 0.0024 | 416.7 |

The performances of estimators are compared in terms of robustness under the models above. For the first model, even for small n, MML is superior to other estimators. As n increases to 20, probability of plotting outside the control limits gets closer to 0.0027 for MML. Also W24 has good performance as n increases. Probability of plotting outside the limits is much larger than the nominal value for other estimators.

When the second model is considered, ARL value is close to nominal value of 370 even for small n for W24, as n increases to 20, ARL value of MML and W24 get closer to 370. Especially LS estimators perform poorly for all n.

For Tukey lambda family ($l$=0.585), whereas, type I error rate is less than the nominal value for n equal to 5 for MML, for other estimators, it is much more than the prespecified level for LS estimators.

The results for model 4 ($l$=1.0) does not produce satisfactory ARL values. For n=20, MAD, W24, TRIM have ARL values close to 370.4. LS has ARL value much less than 370, should not be used.

When normal distribution is used as plausible alternative to short-tailed distribution ($r$=4, $d$=-0.5), type I error rates of W24 and MML are larger than the nominal value for all n, however, type I error rates for MAD and W24 are close to nominal value.

To sum up, STS distributions are frequently confronted in practice. While obtaining control limits for these distributions, LS estimators give inappropriately large values of Type I error rates. MML and W24 may be preferred to other estimators for all sample sizes for various values of parameters $r$ and $d$.

# CHAPTER 5

# APPLICATIONS

In this chapter, data sets from literature are analyzed in order to determine whther they follow long- or short-tailed symmetric distributions and control limits are constructed. Several data sets are investigated whether they follow normal distribution or not. For checking normality, firstly, skewness and kurtosis values are calculated and then Q-Q plots are plotted for suitable parameters. For a random sample $x_1, x_2, ...., x_n$ from a location-scale distribution, population quantiles obtained from $F(\theta_i) = i/(n+1)$ $1 \le i \le n$. In order to construct Q-Q plot, $x_i$ is plotted against $\theta_i$ $1 \le i \le n$. Under the assumption of the underlying distribution, all points would lie on a straight line to be a plausible model for data.

Many data sets are found to follow non-normal distributions. Control limits are constructed for data sets which have symmetric distributions using the robust estimators and ordinary LS method.

**Application 1: (DeVor, Cheng and Sutherland, 1992)**

The first data set represents the depths of keyways collected in 28 samples of size n=5. Sample skewness estimate equals to -0.10072 and sample kurtosis estimate is 2.18. For this data set, Q-Q plot based on short tailed symmetric distribution with parameters r=4 and d=0.5 gives “closest to straight line” pattern for the data set (Figure 5.1).

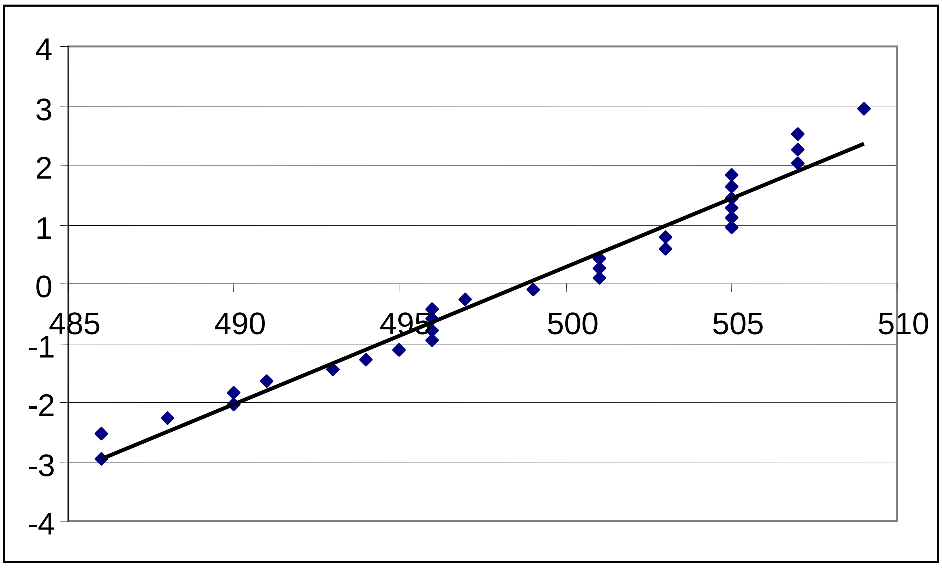


**Figure 5.1**: Q-Q Plot Based on STS Density for the Data of Application 1.

For calculating control limits constants obtained in Table 4.1 are used. Then the control limits are constructed for the assumed distribution by using LS, MML, TRIM, MAD and W24 estimators respectively and given in Table 5.1.

**Table 5.1** Control Limits and Number of Samples Outside the Control Limits for Application 1

| *Estimator* | *LCL* | *CL* | *UCL* | *Number outside the limits* |
|---|---|---|---|---|
| *LSE* | 494.79 | 498.69 | 502.13 | 6 |
| *MML* | 492.96 | 498.58 | 503.84 | 2 |
| *TRIM* | 495.47 | 498.71 | 501.19 | 8 |
| *MAD* | 496.98 | 499.25 | 506.53 | 0 |
| *WAVE* | 494.25 | 498.94 | 504.02 | 1 |

Here the underlying distribution seems to be STS with r=4 and d=0.5. As mentioned before, for this distribution, MML and W24 have satisfactory performance of type I error rate and potentially effective in evaluating the process. Other methods produce more false alarms that cause the process to be stopped frequently.

**Application 2: (Montgomery, 2001)**

The second data set represents the readings from a chemical process on successive days. It consists of 18 subsamples of size 5. Sample kurtosis estimate equals to 1.998 and sample skewness estimate is 0.255416. Q-Q plot based on short tailed symmetric distribution with parameters r=4 and d=0.5 gives "closest to straight line" pattern for the data set which indicates that STS($r$=4, $d$=0.5) is a plausible model for the data (Figure 5.2).

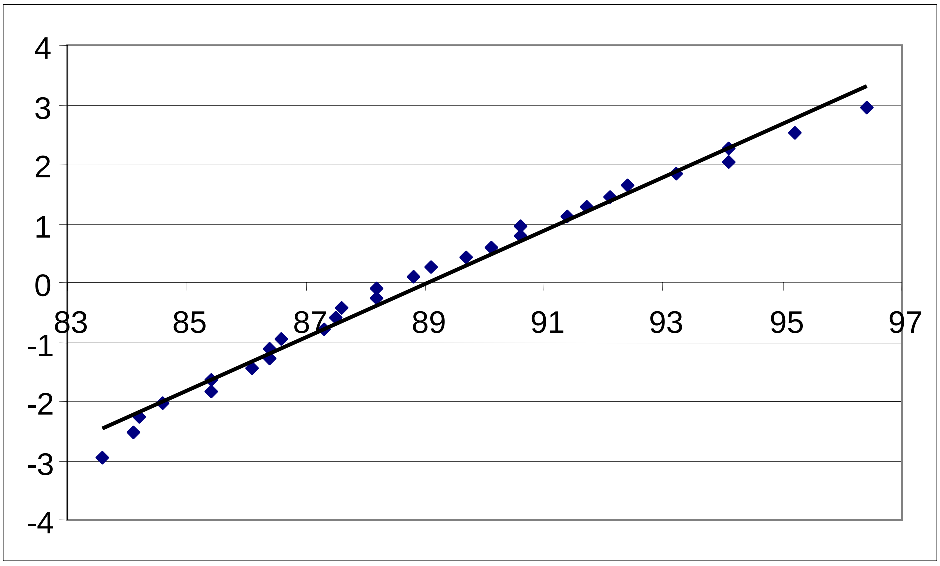


**Figure 5.2**: Q-Q Plot Based on STS Density for the Data of Application 2.

The control limits for all estimators are obtained and given in Table 5.2 below.

**Table 5.2** Control Limits and Number of Samples Outside the Control Limits for Application 2

| *Estimator* | *LCL* | *CL* | *UCL* | *Number outside the limits* |
|---|---|---|---|---|
| *LSE* | 86.23 | 89.47 | 93.01 | 3 |
| *MML* | 85.44 | 89.45 | 93.51 | 1 |
| *TRIM* | 87.32 | 89.48 | 91.63 | 4 |
| *MAD* | 89.42 | 89.53 | 93.98 | 0 |
| *W24* | 86.28 | 89.22 | 93.96 | 1 |

In this application, the underlying distribution seems to be STS with $r$ =4 and $d$ =0.5. For this distribution, MML and W24 have good performance among other methods. MML and W24 do not produce false alarms which leads to unnecessary process adjustment by stopping the process that cause time and money consuming.

**Application 3: (DeVor et. al., 1992)**

The third data set represents the thickness of sheets in samples of size 5 (g=14). Sample kurtosis estimate equals to 2.22 and sample skewness estimate is 0.020816. Q-Q plot based on short tailed symmetric distribution with parameters r=4 and d=0 gives “closest to straight line” pattern for the data set (Figure 5.3).

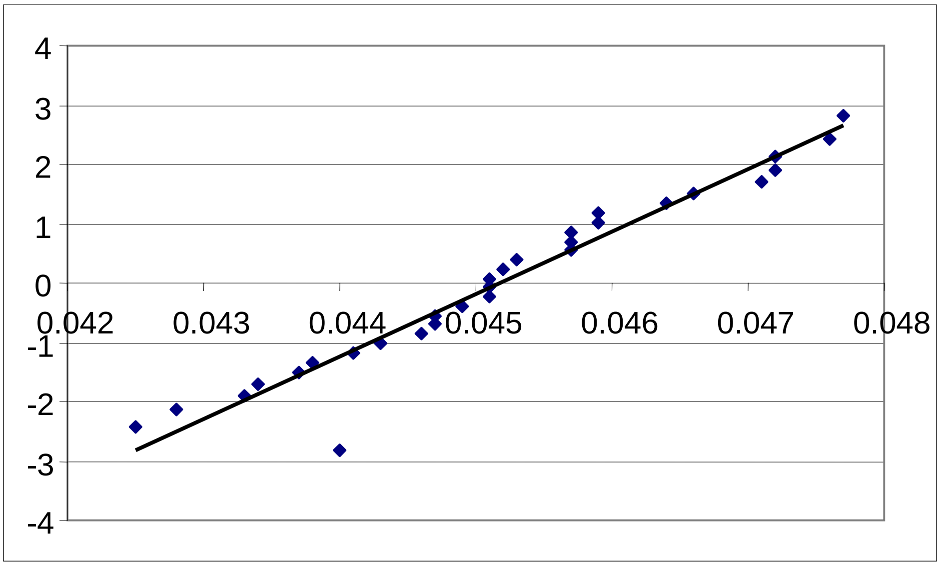


**Figure 5.3** Q-Q Plot Based on STS Density for the Data of Application 3.

**Table 5.3** Control Limits and Number of Samples Outside the Control Limits for Application 3

| *Estimator* | *LCL* | *CL* | *UCL* | *Number outside the limits* |
|---|---|---|---|---|
| *LSE* | 0.04421 | 0.04496 | 0.04359 | 7 |
| *MML* | 0.04367 | 0.04491 | 0.04615 | 2 |
| *TRIM* | 0.04432 | 0.04497 | 0.04563 | 8 |
| *MAD* | 0.04115 | 0.04507 | 0.04813 | 0 |
| *WAVE* | NAN | NAN | NAN | 0[1] |

For this example, STS distribution with r=4 d=0.0 seems to be a reasonable model. For this model, when n is small, MML produces a little more false alarms than

[1] Due to ties in data set, W24 can not be calculated.

prespecified value, however, other methods give type I error rates much more than nominal value.

**Application 4: (Montgomery, 2001)**

The data set represents the thickness of a printed circuit board in inches. Fifteen samples of size 5 are given. Sample kurtosis estimate equals to 3.77476 and sample skewness estimate is -0.2938. Q-Q plot based on long tailed symmetric distribution with parameter p=5.0 gives "closest to straight line" pattern for the data set which indicates that the model is a plausible model (Figure 5.4).

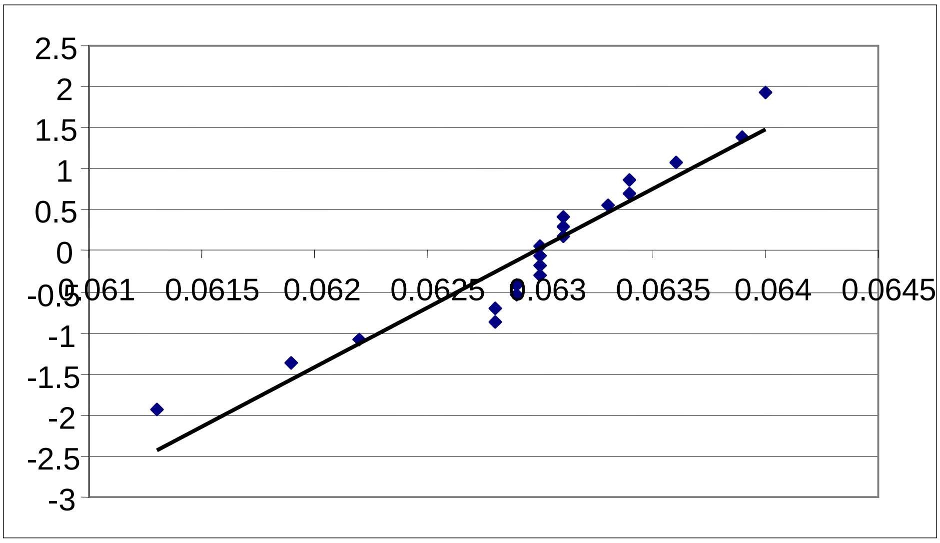


**Figure 5.4**: Q-Q Plot Based on STS Density for the Data of Application 4.

**Table 5.4:** Control Limits and Number of Samples Outside the Control Limits for Application 4

| *Estimator* | *LCL* | *CL* | *UCL* | *Number outside the limits* |
|---|---|---|---|---|
| *LSE* | 0.06216 | 0.06295 | 0.06375 | 0 |
| *MML* | 0.06220 | 0.06296 | 0.06371 | 1 |
| *TRIM* | 0.06228 | 0.06297 | 0.06365 | 1 |
| *MAD* | 0.06222 | 0.06295 | 0.06367 | 0 |
| *W24* | NAN | NAN | NAN | 0 |

For LTS distribution, when n=5, all methods lead to similar type I error rates. A disadvantage of W24 is that it can not be calculated when there are ties in the data and sample size is small. MML and TRIM have similar performances.

# CHAPTER 5

# CONCLUSIONS

Statistical process control is used for an in-process or real time monitoring tool for a process. SPC makes sure that the process is in control and prevents the production of defects. $\bar{x}$ control chart is one of the most commonly used techniques in SPC, the underlying assumption that may not be met in practical grounds; since, many process distributions are not normally distributed. In this thesis, quality control charts for the mean of process are constructed for non-normal symmetric distributions, long- and short-tailed symmetric distributions. The performances of these control charts are investigated via Monte Carlo simulation. Furthermore, some examples from literature are analyzed which either have long or short tailed symmetric distribution and control charts are constructed for these data sets.

Ordinary control charting procedure is very easy and involved in much SPC software. The control charts using robust estimators lead to Type I error rates close to nominal value for long-tailed symmetric distributions. When the distribution is long-tailed symmetric, MML estimators should be preferred for small n, and MML and W24 estimators are preferable for large n. For short-tailed symmetric distribution, MML and W24 have good performance for all values of sample size and parameters $r$ and $d$.

Only the symmetric distributions are considered in this thesis, as a future work control charts for auto-correlated observations and skewed distributions will be studied.

# REFERENCES

Akkaya, A., Tiku, M.L. (2005). Robust Estimation and Hypothesis Testing under Short-tailedness and Inliers, *Test* 14(1): 129-150.

Alloway, J.A., Raghavachari, M. (1991). Control Chart Based on the Hodges-Lehmann Estimator, *Journal of Quality Technology* **23(4)**: 336-347.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Rogers, W.H., Tukey, J.W. (1972). *Robust Estimation of Location: Survey and Advances*, Princeton University Press, Princeton, NJ.

Borrows, P.M. (1962). X-Bar Control Schemes for a Production Variable with Skewed Distribution, *The Statistician* **12(4)**: 296-312.

Canavos G.C., Koutrouvelis, I.A. (1984). The Robustness of Two-Sided Tolerance Limits for Normal Distributions, *Journal of Quality Technology* **16(3)**: 144-149.

Castagliola, P. (2000). $\bar{x}$ Control Chart for Skewed Populations Using a Scaled Weighted Variance Method, *International Journal of Reliability, Quality and Safety Engineering* **7 (3)**: 237-252.

Caulcutt, R. (1995). The Rights and Wrongs of Control Chart, *Applied Statistics* **44(3)**: 279-288.

Chan, L.K., Hapuarachchi, K.P., Macpherson, B.D. (1988). Robustness of $\bar{X}$ and R Charts, *IEEE Transactions on Reliability* **37(1)**: 117-123.

Chakraborti, S., Van der Laan P., Bakır S.T. (2001). Nonparametric Control Charts. *Journal of Quality Technology* **33**: 304-315.

Chakraborti, S., Van der Laan, P., Bakır S.T. (2004). A Class of Distribution Free Control Charts, *Appl. Statist.* **53**: 443- 462.

Chan, L. K., Hapuarachchi, K. P., Macpherson, B.D. (1988). Robustness of $\bar{x}$ and R Charts. *IEEE Transactions on Reliability* **37**: 117-123.

Choobineh, F., Ballard, J.L. (1987). Control Limits of QC Charts for Skewed Distributions Using Weighted Variance. *IEEE Transactions on Reliability* **R-36**: 473-477.

Cryer, J.D., Ryan, T.P. (1990). The Estimation of Sigma for an X Chart: $\overline{MR}/d_2$ or $S/c_4$?. *Journal of Quality Technology* **22 (3)**: 187-189.

David, H.A., Hartley, H.O., Pearson, E.S. (1954). The Distribution of the Ratio, in a single Normal Sample, of Range to Standard Deviation. *Biometrika* **41**: 482-493.

Devor, R.E., Chang, T., Sutherland, J.W. (1992). *Statistical Quality Control*. Prentice Hall, New Jersey.

Ferrell, E.B. (1958). Control Charts for Log-Normal Universe, *Industrial Quality Control*, **15**: 4-6.

Gibbons, J. D. and S. Chakraborti (2003): *Nonparametric Statistical Inference*, 4th Edition, Marcel Dekker, New York.

Grimshaw, S.D., Alt, F.B. (1997). Control Charts for Quantile Function Values, *Journal of Quality Technology* **29(1)**: 1-7.

Harter, H.L. (1960). Tables of Range and Studentized Range. *The Annals of Mathematical Statistics* **31**: 1122-1147.

Harter, H. L. (1961). Expected Values of Normal Order Statistics. *Biometrika* **48**: 118-121**.**

Kendall, M.G., Stuart, A. (1979). *The Advanced Theory of Statistics,* Vol. 3, McMillan: New York.

Klein, M. (2000). Two Alternatives to the Shewhart $\overline{X}$ Control Chart, *Journal of Quality Technology* **32(4)**: 427-431.

Kocherlakota, K., Kocherlakota, S. (1995). Control charts Using Robust Estimators. *Total Quality Management* **6(1)**: 79-90.

Langenberg, P., Iglewitz, B. (1986). Trimmed Mean $\bar{x}$ and R Charts. *Journal of Quality Technology* **18**: 152-161 .

Montgomery, D.C. (2001). *Introduction to Statistical Quality Control*, John Wiley & Sons, Inc., NewYork.

Moore, P.G. (1957). Normality in Quality Control Charts, *Applied Statistics* **6(3)**: 171-179.

Nedumaran, G., Pignatiello, J.J. (2001). On Estimating $\overline{X}$ Chart Limits, Journal of Qaulity Technology 33(2): 206-212.

Page, E.S. (1954). Control Charts for the Mean of a Normal Population, *Journal of the Royal Statistical Society, Series B(Methodological)* **16(1)**: 131-135.

Pearson, E.S., Haines, J. (1935). "The use of Range in Place of Standard Deviation in Small Samples", Supplement to the Journal of the Royal Statistical Society, Vol. 2.

Rocke, M. D. (1989). Robust Control Charts. *Technometrics* **31:** 173-184.

Shore, H. (2004). Non-normal Populations in Quality Applications: A Revisited Perspective, *Quality and Reliability Engineering International* **20**: 375-382.

Simon, G. (1976). Computer Simulation Swindles, with Applications to Estimates of Location and Dispersion. *Applied Statistics* **25:** 266-274 .

Sower, V.E., Savoie, M.J., Renick S. (1999). *An Introduction to Quality Management and Engineering Body of Knowledge*, John Wiley and Sons, New York.

Spedding, T.A., Rawlings, P.L. (1993). Non-normality in Statistical Process Control Measurements. *International Journal of Quality and Reliability Management* **11**: 27-37.

Tan, W.Y. (1985). On Tiku's robust procedure- a Bayesian insight, *J. Stat. Plann. Inf.* **11**: 329-340.

Taylor, H.M. (1967). Statistical Control of a Gaussian Process, *Technometrics* **9(1)**: 29-41.

Tiku, M.L., Kumra, S. (1981). Expected Values and Variances And Covariances Of Order Statistics for a Family of Symmetric Distributions. *Selected Tables in Mathematical Statistics* **8**: 141-270.

Tiku, M.L., Tan, W.Y., Balakrishnan N. (1986). *Robust Inference*, Marcel Dekker, New York.

Tiku, M.L., Suresh, R.P. (1992). A New Method of Estimation for Location and Scale Parameters, *J. Stat. Plann. Inf.* **30**: 281-292.

Tiku, M.L., Vaughan, D.C. (1999). A Family of Short-tailed Symmetric Distributions, Technical Report: McMaster University, Canada.

Tiku, M.L., İslam, M.Q., Selçuk, S. (2001). Nonnormal Regression. II. Symmetric Distributions, *Communications in Statistics: Theory and Methods* **30(6)**, 1021-1045.

Tiku, M.L., Akkaya, A.D. (2004). *Robust Estimation and Hypothesis Testing*. New Age International Limited Publishers, New Delhi.

Vardeman, S.B. (1999). A Brief Tutorial on the Estimation of the Process Standard Deviation, *IEEE Transactions* **31**: 503-507.

Vaughan, D.C. (1992a). On the Tiku-Suresh Method of Estimation. *Commun. Theory Meth.,* **21**: 451-469.

Vaughan, D.C. (1992b). Expected Values, Variances and Covariances of Order Statistics for Student's t Distribution with Two Degrees of Freedom. *Commun. Stat. Simula.*, **21**: 391-404.

Vaughan, D.C. (1994). The Exact Values, Variances and Covariances of the Order Statistics from the Cauchy Distribution. *J. Stat. Comput. Simul.*, **49**: 21-32.

Yourstone, S.A., Zimmer, W. J. (1992). Non-Normality and the Design of Control Charts for Averages. *Decision Sciences* **23**, 1099-1113.

Wheeler, D.J. (1995). *Advanced Topics in Statistical Process Control*, Statistical Process Controls, Inc., Knoxville, TN.

Wood, M. (1995). Three Suggestions for Improving Control Charting Procedures, *International Journal of Quality and Reliability Management* **12(5)**: 61-74.

Wu, Z. (1996). Asymmetric Control Limits of the x-bar Chart for Skewed Process Distributions, *International Journal of Quality and Reliability Management* **13(9)**: 49-60.

# APPENDIX A

# EXPECTED STS VARIATES FOR SHORT TAILED SYMMETRIC DISTRIBUTION

The tables are taken from Tiku and Akkaya (2004).

**Table A.1. Expected STS variates**

| *r=2 a=0* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *n* | i | t(i) | *n* | i | t(i) | *n* | i | t(i) |
| *3* | 1 | -1,07286 | *10* | 1 | -1,93809 | *20* | 5 | -1,12721 |
| | 2 | 0 | | 2 | -1,39774 | | 6 | -0,91359 |
| *4* | 1 | -1,30734 | | 3 | -0,97094 | | 7 | -0,70706 |
| | 2 | -0,42316 | | 4 | -0,57758 | | 8 | -0,50396 |
| *5* | 1 | -1,47603 | | 5 | -0,19227 | | 9 | -0,30217 |
| | 2 | -0,70706 | *12* | 1 | -2,04653 | | 10 | -0,10071 |
| | 3 | 0 | | 2 | -1,54485 | *30* | 1 | -2,5292210 |
| *6* | 1 | -1,60608 | | 3 | -1,16107 | | 2 | -2,1542550 |
| | 2 | -0,91359 | | 4 | -0,81765 | | 3 | -1,8957710 |
| | 3 | -0,30217 | | 5 | -0,48841 | | 4 | -1,6866680 |
| *7* | 1 | -1,71103 | | 6 | -0,16269 | | 5 | -1,5045740 |
| | 2 | -1,07286 | *15* | 1 | -2,17306 | | 6 | -1,3390250 |
| | 3 | -0,52924 | | 2 | -1,71103 | | 7 | -1,1841870 |
| | 4 | 0 | | 3 | -1,36911 | | 8 | -1,0364340 |
| *8* | 1 | -1,798496 | | 4 | -1,07286 | | 9 | -0,8933640 |
| | 2 | -1,200991 | | 5 | -0,79682 | | 10 | -0,7532978 |
| | 3 | -0,707064 | | 6 | -0,52924 | | 11 | -0,6150723 |
| | 4 | -0,235004 | | 7 | -0,26439 | | 12 | -0,4778767 |
| *9* | 1 | -1,87316 | | 8 | 0 | | 13 | -0,3411770 |
| | 2 | -1,30734 | *20* | 1 | -2,32714 | | 14 | -0,2046776 |
| | 3 | -0,85109 | | 2 | -1,90669 | | 15 | -0,0682259 |
| | 4 | -0,42316 | | 3 | -1,60608 | | | |
| | 5 | 0 | | 4 | -1,35427 | | | |

**Table A.1. Expected STS variates (continued)**

| *r=4 a=0* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *n* | i | t(i) | *n* | i | t(i) | *n* | i | t(i) |
| *3* | 1 | -1,221075 | *10* | 1 | -2,186975 | *20* | 5 | -1,282530 |
| | 2 | 0,0 | | 2 | -1,587048 | | 6 | -1,040592 |
| *4* | 1 | -1,485558 | | 3 | -1,105642 | | 7 | -0,805864 |
| | 2 | -0,482454 | | 4 | -0,658426 | | 8 | -0,574541 |
| *5* | 1 | -1,674709 | | 5 | -0,219231 | | 9 | -0,344524 |
| | 2 | -0,805864 | *12* | 1 | -2,305889 | | 10 | -0,114832 |
| | 3 | 0,000000 | | 2 | -1,751556 | *30* | 1 | -2,829094 |
| *6* | 1 | -1,819782 | | 3 | -1,320763 | | 2 | -2,423515 |
| | 2 | -1,040592 | | 4 | -0,931635 | | 3 | -2,140427 |
| | 3 | -0,344524 | | 5 | -0,556822 | | 4 | -1,909332 |
| *7* | 1 | -1,936359 | | 6 | -0,185499 | | 5 | -1,706600 |
| | 2 | -1,221075 | *15* | 1 | -2,444000 | | 6 | -1,521168 |
| | 3 | -0,603352 | | 2 | -1,936359 | | 7 | -1,346855 |
| | 4 | 0,000000 | | 3 | -1,554947 | | 8 | -1,179857 |
| *8* | 1 | -2,033157 | | 4 | -1,221075 | | 9 | -1,017637 |
| | 2 | -1,365805 | | 5 | -0,907955 | | 10 | -0,858460 |
| | 3 | -0,805864 | | 6 | -0,603352 | | 11 | -0,701132 |
| | 4 | -0,267954 | | 7 | -0,301447 | | 12 | -0,544815 |
| *9* | 1 | -2,115536 | | 8 | 0,000000 | | 13 | -0,389004 |
| | 2 | -1,485558 | *20* | 1 | -2,611237 | | 14 | -0,233374 |
| | 3 | -0,969629 | | 2 | -2,152462 | | 15 | -0,077791 |
| | 4 | -0,482454 | | 3 | -1,819782 | | | |
| | 5 | 0,000000 | | 4 | -1,538277 | | | |

# APPENDIX B

## MATLAB PROGRAM FOR CALCULATION OF TAIL PROBAILITIES FOR LONG-TAILED SYMMETRIC DISTRIBUTION

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% A MATLAB PROGRAM WRITTEN BY
% AYSUN ÇETINYUREK
% ANKARA, 2006.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear
g=20
p=2.50

nnn=[5 7 10 15 20]
format long
for i=5:5
n=nnn(i)
r=floor(0.5+0.1*n);
simno=10000
yno=10000

%The "A" Constants Required To Calculate The Control Limits
else if p==2.5
    if n==5
     c4=0.8748;
     ctr=0.7493011;
     cm=0.4648;
     cw=0.7029;
     cmml=1.0959;
         else if n==7
           c4=0.9078;
           ctr=0.834351;
           cm=0.4905;
           cw=0.7641;
           cmml=1.0899;
               else if n==10
                  c4=0.9289;
                  ctr=0.786652;
                  cm=0.4974;
                  cw=0.7880;
                  cmml=1.0880;
                     else if n==15
```

```
                  c4=0.9465;
                  ctr=0.8692489;
                  cm=0.5086;
                  cw=0.8144;
                  cmml=1.0857;
                else if n==20
                  c4=0.9559;
                  ctr=0.860451;
                  cm=0.5125;
                  cw=0.8258;
                  cmml=1.0683;
                end
              end
            end

else if p==3
   if n==5
        c4=0.9060;
        ctr=0.746567;
        cm=0.4950;
        cw=0.7444;
        cmml=1.0660;
          else if n==7
          c4=0.9258;
          ctr=0.829133;
          cm=0.5167;
          cw=0.7985;
          cmml=1.0639;
             else if n==10
                c4=0.9519;
                ctr=0.786778;
                cm=0.5323;
                cw=0.8382;
                cmml=1.0538;
                    else if n==15
                    c4=0.9620;
                    ctr=0.868291;
                    cm=0.5429;
                    cw=0.8625;
                    cmml=1.0528;
                  else if n==20
                      c4=0.9698;
                      ctr=0.85810756;
                      cm=0.5481;
                      cw=0.8734;
                      cmml=1.0514;
                    end
                  end
                end

else if p==3.5
      if n==5
        c4=0.9137;
        ctr=0.744573;
        cm=0.5086;
        cw=0.7618;
        cmml=1.0424;
        else if n==6
```

```
c4=0.9138
ctr=1.0462;
cm=0.5117;
cw=0.7954;
   else if n==7
   c4=0.9380;
   ctr=0.831633;
   cm=0.5412;
   cw=0.8268;
   cmml=1.0410;
         else if n==10
            c4=0.9591;
            ctr=0.786035;
            cm=0.5530;
            cw=0.8642;
            cmml=1.0398;
                  else if n==15
                  c4=0.9702;
                  ctr=0.8701316;
                  cm=0.5657;
                  cw=0.8909;
                  cmml=1.0308;
               else if n==20
                     c4=0.9747;
                     ctr=0.86027;
                     cm=0.5705;
                     cw=0.8997;
                     cmml=1.0232;
                  end
               end
            end
         end
else if p==4.0
if n==5
      c4=0.9204;
      ctr=0.744317;
      cm=0.5170;
      cw=0.7731;
      cmml=1.0315;
         else if n==7
            c4=0.9437;
            ctr=0.832949;
            cm=0.5465;
            cw=0.8364;
            cmml=1.0286;
            else if n==10
                  c4=0.9563;
                  ctr=0.7846089;
                  cm=0.5600;
                  cw=0.8695;
                  cmml=1.0260;
               else if n==15
                     c4=0.9748;
                     ctr=0.8714;
                     cm=0.5809;
                     cw=0.9091;
                     cmml=1.0253;
                  else if n==20
```

```
                                    c4=0.9780;
                                    ctr=0.85963;
                                    cm=0.5841;
                                    cw=0.9161;
                                    cmml=1.0194;
                                  end
                                end
                             end
                           end

else if p==5
if n==5
                c4=0.9265;
                ctr=0.740333;
                cm=0.5283;
                cw=0.7872;
                cmml=1.0040;
                   else if n==7
                       c4=0.9448;
                       ctr=0.829012;
                       cm=0.5556;
                       cw=0.8481;
                            else if n==10
                                c4=0.9602;
                                ctr=0.788649;
                                cm=0.5751;
                                cw=0.8887;
                                cmml=1.0160;
                                    else if n==15
                                    c4=0.9770;
                                    ctr=0.86981;
                                    cm=0.5950;
                                    cw=0.9243;
                                    cmml=1.0157;
                                else if n==20
                                       c4=0.9802;
                                       ctr=0.86136;
                                       cm=0.5999;
                                       cw=0.9327;
                                       cmml=1.0149;
                                    end
                                end
                              end
                             end
  else if p==6
       if n==5
                   c4=0.9305;
                   ctr=0.7453958;
                   cm=0.5357;
                   cw=0.7941;
                   cmml=0.9985;
                      else if n==7
                          c4=0.9484;
                          ctr= 0.8303169;
                          cm=0.5654;
                          cw=0.8589;
                          cmml=1.0050;
                              else if n==10
```

```
                        c4=0.9650;
                        ctr=0.7865487;
                        cm=0.5849;
                        cw=0.9003;
                        cmml=1.0122;
                         else if n==15
                           c4=0.9774;
                           ctr=0.8695905;
                           cm=0.6042;
                           cw=0.9332;
                           cmml=1.0102;
                        else if n==20
                              c4=0.9814;
                              ctr=0.8574635;
                              cm=0.6087;
                              cw=0.9426;
                              cmml=1.0010;
                           end
                        end
                     end
                  end
else if p==10
        if n==5
           c4=0.9352;
           ctr=0.7441905;
           cm=0.5437;
           cw=0.8049;
           cmml=0.9741;
              else if n==7
              c4=0.9551;
              ctr=0.8228023;
              cm=0.5774;
              cw=0.8754;
              cmml=0.9854;
                 else if n==10
                    c4=0.9667;
                    ctr=0.7858148;
                    cm=0.5978;
                    cw=0.9121;
                    cmml=0.9909;
                       else if n==15
                       c4=0.9809;
                       ctr=0.86793;
                       cm=0.6175;
                       cw=0.9491;
                       cmml=0.9987;
                    else if n==20
                          c4=0.9844;
                          ctr=0.8599975;
                          cm=0.6243;
                          cw=0.9592;
                          cmml=1.0000;
                       end
                    end
                 end

end
end
```

```
end
end
end
end
end
end

v=2*p-1;
k=2*p-3;

%%%%% RANDOM SAMPLE TO COMPARE WITH THE LIMITS%%%%%%%%%%%%
for j=1:yno
    y=trnd(v,1,n);
    y=sqrt(k/v)*y;
   ym(j)=mean(y);
   ymLTS(j)=LTS(y,p);
   ymtr(j)=trim(y);
   ymMAD(j)=MAD(y);
   ymwave(j)=wave(y);

end

%%Calculation of Control Limits
for i=1:simno
    for  j=1:g

  %%Generating LTS random numbers
    x=trnd(v,1,n);
    x=sqrt(k/v)*x;

   xbar(j)=mean(x);
   sdev(j)=std(x);

    [m_mml v_mml M]=LTS(x,p);
    mean_mml(j)=m_mml;
    var_mml(j)=v_mml;
    MMM=M;

    [m_MAD v_MAD]=MAD(x);
    meanMAD(j)=m_MAD;
    varMAD(j)=v_MAD;

   [m_tr v_tr]=trim(x);
   meantr(j)=m_tr;
    vartr(j)=v_tr;

    [m_wv v_wv]=wave(x);
    meanwv(j)=m_wv;
    varwv(j)=v_wv;
    end

 xdbar=mean(xbar);
 sbar=mean(sdev);

%Control Limits Using LS Estimators
 Ulse(i)=xdbar+3*sbar/(sqrt(n)*c4);
 Llse(i)=xdbar-3*sbar/(sqrt(n)*c4);
```

```
meanmml_bar=mean(mean_mml);
varmml_bar=mean(var_mml);

varmuhed(i)=varmml_bar/sqrt(MMM);

%Control Limits Using MML Estimators
Umml(i)=meanmml_bar+3*varmml_bar/(sqrt(MMM)*cmml);
Lmml(i)=meanmml_bar-3*varmml_bar/(sqrt(MMM)*cmml);

meantrbar=mean(meantr);
vartrbar=mean(vartr);

%Control Limits Using TRIM Estimators
Utr(i)=meantrbar+3*vartrbar/(sqrt(n)*ctr);
Ltr(i)=meantrbar-3*vartrbar/(sqrt(n)*ctr);

meanMADbar=mean(meanMAD);
varMADbar=mean(varMAD);

%Control Limits Using MAD Estimators
UMAD(i)=meanMADbar+3*varMADbar/(sqrt(n)*cm);
LMAD(i)=meanMADbar-3*varMADbar/(sqrt(n)*cm);

meanwvbar=mean(meanwv);
varwvbar=mean(varwv);

%Control Limits Using W24 Estimators
Uwv(i)=meanwvbar+3*varwvbar/(sqrt(n)*cw);
Lwv(i)=meanwvbar-3*varwvbar/(sqrt(n)*cw);

end

%% MEAN OF CONTROL LIMITs will be found%%%

%mUlse=mean(Ulse);
%mLlse=mean(Llse);

mUmml=mean(Umml);
mLmml=mean(Lmml);

%mUtr=mean(Utr);
%mLtr=mean(Ltr);

%mUMAD=mean(UMAD);
%mLMAD=mean(LMAD);

%mUwv=mean(Uwv);
%mLwv=mean(Lwv);

%%%%%%%%%% end of control limits %%%%%%%%%

  alpu=find(ym>mUlse);
  alpd=find(ym<mLlse);
  alp=length(alpu)+length(alpd);

  alpmmlu=find(ymLTS>mUmml);
  alpmmld=find(ymLTS<mLmml);
  alpmml=length(alpmmlu)+length(alpmmld);
```

```
    alptru=find(ymtr>mUtr);
    alptrd=find(ymtr<mLtr);
    alptr=length(alptru)+length(alptrd);

   alpMADu=find(ymMAD>mUMAD);
   alpMADd=find(ymMAD<mLMAD);
    alpMAD=length(alpMADu)+length(alpMADd);

    alpwaveu=find(ymwave>mUwv);
    alpwaved=find(ymwave<mLwv);
    alpwave=length(alpwaveu)+length(alpwaved);

%Type I error rates and ARL values will be found

alp=alp/yno
ARLalp=1/alp

alpmml=alpmml/yno
ARLalpmml=1/alpmml

alptr=alptr/yno
ARLtr=1/alptr

alpMAD=alpMAD/yno
ARLMAD=1/alpMAD

alpwave=alpwave/yno
ARLwave=1/alpwave

end
```

# APPENDIX C

# MATLAB PROGRAM FOR CALCULATION OF TAIL PROBAILITIES FOR SHORT-TAILED SYMMETRIC DISTRIBUTION

```
n=20
rr=2
aa=-0.50
g=20
format long
if (rr==2&&aa==0)
    if n==5
        tvar=[-1.476030 -0.7070637 0.0 0.7070637 1.476030];
        c4=0.949064
        tr4=1.92257
        md4=0.820690
        cw=1.191472
        cmml=0.8796060

        else if n==7
            tvar=[-1.711025 -1.072855 -0.5292416 0.0 0.5292416 1.072855 1.711025];
            c4=0.965838
            tr4=1.699013
            md4=0.895735
            cw=1.298046
            cmml=0.925729
          else if n==10
                tvar=[-1.938086 -1.397743 -0.9709358 -0.5775833 -0.1922703 0.1922703 0.5775833
0.9709358 1.397743 1.938086];
                c4=0.978200
                tr4=1.56887
                md4=0.934673
                cw=1.385435845
                cmml=0.954881

                else if n==15
tvar=[-2.173061 -1.711025 -1.369114 -1.072855 -0.7968235 -0.5292416 -0.2643871 0.0 0.2643871
0.5292416 0.7968235 1.072855 1.369114 1.711025 2.173061];
                    c4=0.986098
                    tr4=1.507582
                    md4=0.956013
                    cw=1.416484
```

```
                    cmml=0.964812
                  else if n==20
tvar=[-2.327137 -1.906691 -1.606083 -1.354265 -1.127214  -0.9135866 -0.7070637 -0.5039597 -0.3021717
-0.1007128 0.1007128 0.3021717 0.5039597 0.7070637  0.9135866 1.127214 1.354265  1.606083  1.906691
2.327137];

                      c4=0.990027
                      tr4=1.4602713
                      md4=0.987924
                      cw=1.431415
                      cmml=0.979318
                    end
                    end
                  end
                end
              end
   else if (rr==2&&aa==-0.5)
    if n==5
        tvar=[-1.382351 -0.6446648 0.0 0.6446648 1.382351];
        c4=0.9481489
        tr4=1.823936
        md4=0.774599
        cw=1.125504
        cmml=0.886329
        else if n==7
            tvar=[-1.614885 -0.9904289 -0.4801178 0.0 0.4801178 0.9904289 1.614885];
            c4=0.9646825
            tr4=1.607458
            md4=0.8314726
            cw=1.22971
            cmml=0.923379
          else if n==10
tvar=[-1.841602  -1.305475  -0.8930874  -0.5246353  -0.17733589  0.17733589   0.5246353  0.8930874
1.305475 1.841602];
                c4=0.979187
                tr4=1.480476
                md4=0.8762695
                cw=1.2918064
                cmml=0.952044
              else if n==15
tvar=[-2.077694  -1.614885  -1.277447  -0.9904289  -0.7285786  -0.4801178  -0.2385902  0.0  0.2385902
0.4801178 0.7285786 0.9904289 1.277447 1.614885 2.077694 ];
                    c4=0.985
                    tr4=1.458035
                    md4=0.956013
                    cw=1.416484
                    cmml=0.962654
                  else if n==20
tvar=[-2.233047 -1.81015 -1.510735 -1.262932 -1.042652  -0.8386421 -0.6446648 -0.4568958 -0.2728367 -
0.090742111  0.090742111  0.2728367  0.4568958  0.6446648  0.8386421  1.042652  1.262932  1.510735
1.81015 2.233047];
                      c4=0.990027
                      tr4=1.542014
                      md4=0.987924
                      cw=1.431415
                      cmml=0.977416

                    end
```

```
                              end
                           end
                        end
                     end

else if (rr==2 && aa==0.5)
    if n==5
          ty=[-1.611099 -0.8123684 ];
          tvar=[ty 0 sort(-ty)]
          c4=0.954072
          tr4=2.065557
          md4=0.876092
          cw=1.281163
          cmml=0.921706
      else if n==7
             ty=[-1.844959 -1.201048 -0.615387];
             tvar=[ty 0 sort(-ty)]
                c4=0.970133
                tr4=1.819197
                md4=0.959068
                cw=1.402298
                cmml=0.952678
             else if n==10
                   ty=[-2.06892 -1.532526 -1.094713 -0.6694984 -0.227232];
                   tvar=[ty sort(-ty)]
                      c4=0.977666
                      tr4=1.70469
                      md4=1.020148
                      cw=1.466126
                      cmml=0.968368
                   else if n==15
  ty=[-2.299519 -1.844959 -1.503687 -1.201048 -0.9096718 -0.615387 -0.3116989]
                              tvar=[ty 0 sort(-ty)]
                              c4=0.987832
                                 tr4=1.629959
                                 md4=1.080301
                                 cw=1.528138
                                 cmml=0.983072
                              else if n==20
ty=[-2.450352  -2.03804  -1.740847  -1.488695  -1.257238  -1.034269  -0.8123684  -0.5869102  -0.3557014  -
0.1192665]
                                    tvar=[ty sort(-ty)]
                                    c4=0.98981
                                    tr4=1.587649
                                    md4=1.106305
                                    cw=1.5420255
                                    cmml=0.986742
                                 end
                              end
                           end
                        end
                     end

  elseif (rr==4 && aa==0)
    if n==5
          tvar=[-1.674709 -0.8058643 0.0 0.8058643  1.674709];
          c4=0.952904
          tr4=2.168294
```

```
          md4=0.932458
          cw=1.342044
          cmml=0.877834
      else if n==7
              tvar=[-1.936359 -1.221075 -0.6033516 0.0 0.6033516 1.221075 1.936359];
              c4=0.971152
              tr4=1.910452
              md4=1.003725
              cw=1.472202
              cmml=0.929433
            else if n==10
tvar=[-2.186975 -1.587048 -1.105642  -0.6584263 -0.2192307 0.2192307  0.6584263 1.105642 1.587048
2.186975];
                    c4=0.978792
                    tr4=1.7850122
                    md4=1.063603
                    cw=1.536487
                    cmml=0.945995
                  else if n==15
tvar=[-2.444000  -1.936359  -1.554947  -1.221075  -0.9079552  -0.6033516  -0.3014469  0.0  0.3014469
0.6033516 0.9079552  1.221075  1.554947 1.936359  2.444000];
                          c4=0.988287
                          tr4=1.7036698
                          md4=1.081733
                          cw=1.600763
                          cmml=0.96976
                          else if n==20
tvar=[-2.611237          -2.152462 -1.819782          -1.538277 -1.282530 -1.040592 -0.805864 -0.574541 -
0.344524 -0.114832 0.114832 0.344524 0.574541 0.805864    1.040592    1.282530    1.538277    1.819782
      2.152462 2.611237]
                            c4=0.992095
                            tr4=1.652444
                            md4=1.119905
                            cw=1.617588
                            cmml=0.974618
                          end
                        end
                      end
                    end
                  end

elseif (rr==4 && aa==-0.5)
    if n==5
       ty=[-1.583300 -0.7460564]
       tvar=[ty 0 sort(-ty)]
          c4=0.949516
          tr4=2.06725
          md4=0.876076
          cw=1.280013
          cmml=0.870615
          else if n==7
              ty=[-1.842079 -1.141405 -0.5564499]
              tvar=[ty 0 sort(-ty)]
              c4=0.965716
              tr4=1.821165
              md4=0.953651
              cw=1.3988222
              cmml=0.912057
```

```
                else if n==10
ty=[-2.091894 -1.497192 -1.030579 -0.6078243 -0.2012539]
                    tvar=[ty sort(-ty)]
                    c4=0.979127
                    tr4=1.689701
                    md4=1.003589
                    cw=1.464674
                    cmml=0.943474
                  else if n==15
ty=[-2.349967 -1.842079 -1.465731 -1.141405 -0.8424282 -0.5564499 -0.2769089]
                      tvar=[ty 0 sort(-ty)]
                        c4=0.987322
                        tr4=1.619329
                        md4=1.056566
                        cw=1.524238
                        cmml=0.964175
                        else if n==20
ty=[-2.517815 -2.05739 -1.726494 -1.449413 -1.200695 -0.9684277 -0.7460594 -0.529623 -0.3166103 -
0.1053619]
                            tvar=[ty sort(-ty)]
                          c4=0.990424
                          tr4=1.568485
                          md4=1.074157
                          cw=1.54002
                          cmml=0.973664
                        end
                      end
                    end
                  end
                end
 elseif (rr==4 && aa==0.5)
   if n==5
     ty=[-1.789312 -0.8887577]
     tvar=[ty 0 sort(-ty)]
       c4=0.955704
       tr4=2.29056
       md4=0.977275
       cw=1.420078
       cmml=0.895427
       else if n==7
           ty=[-2.05183 -1.326084 -0.6698513];
           tvar=[ty 0 sort(-ty)]
           c4=0.972088
           tr4=2.017799
           md4=1.0715308
           cw=1.561532
           cmml=0.933731
             else if n==10
                 ty=[-2.301235 -1.700726 -1.205978 -0.7297516 -0.2454662]
                 tvar=[ty sort(-ty)]
                 c4=0.980914
                 tr4=1.894165
                 md4=1.128718
                 cw=1.626511
                 cmml=0.9587
               else if n==15
 ty=[-2.555676 -2.05183 -1.668177 -1.326084 -0.9976959 -0.6698513 -0.3371143]
                     tvar=[ty 0 sort(-ty)]
```

```
                    c4=0.987386
                      tr4=1.805072
                      md4=1.19555
                      cw=1.692053
                      cmml=0.974111
                      else if n==20
  ty=[-2.720776 -2.266978 -1.935196 -1.651258 -1.389618 -1.13781 -0.8887577 -0.6384087 -0.3849792 -
0.1287079]
                          tvar=[ty sort(-ty)]
                        c4=0.990087
                        tr4=1.75928
                        md4=1.224432
                        cw=1.707677
                        cmml=0.980579
                      end
                    end
                  end
                end
              end


end
end
end
end

if (rr==2 && aa==-0.5)
  mu2=1.84199184;
else if (rr==2&& aa==0.0)
    mu2=2.03689984;
  else if (rr==2&&aa==0.5)
      mu2=2.33325625;
    else if (rr==2&& aa==1)
        mu2=2.81803369;
      else if(rr==2&& aa==1.5)
          mu2=3.66684201;
        else if(rr==4&&aa==-0.5)
            mu2=2.35960321;
          else if(rr==4&& aa==0)
              mu2=2.57634601;
            else if(rr==4&&aa==0.5)
                mu2=2.859481;
              else if(rr==4&&aa==1)
                  mu2=3.236401;
                else if(rr==4&&aa==1.5)
                    mu2=3.74732164;
                  else if (rr==4&&aa==2.0)
                      mu2=4.444541;
                  else if(rr==4&&aa==2.5)
                      mu2=5.37729721;
                    end
                  end
                end
              end
            end
          end
        end
      end
```

```
      end
    end
  end
end

lamda=rr/(rr-aa);

gmma1=1-(lamda/(2*rr))*tvar.^2;
gmma2=(1+(lamda/(2*rr)).*(tvar.^2)).^2;
gmma=gmma1./gmma2;
b=1-lamda.*gmma;
ff=find(b<0)
count=length(ff)

if (count==0)
      ag=tvar.^2;
      ag1=(1+(lamda/(2*rr)).*tvar.^2).^2;
       alfa=((lamda/rr)*ag.*tvar)./ag1;
      gmma=(1-(lamda/(2*rr))*tvar.^2)./ag1;
       b=1-lamda.*gmma
    else
          ag3=tvar.^2;
                ag2=(lamda/rr)*ag3.*tvar+(1-1/lamda).*tvar;
                ag4=(1+(lamda/(2*rr))*tvar.^2).^2;
                alfa=ag2./ag4;
          ag5=(1/lamda)-(lamda/(2*rr))*tvar.^2;
                gmma=ag5./ag4;
                b=1-lamda.*gmma
   end

%%% Comparison sample are generated from STS
yno=10000

xc=stsrnd(aa,rr,n,yno);
xxc=sort(xc);

sumbbet=sum(b)
for j=1:yno
xxc1=xxc(:,j);

sum3x=sum(b'.*xxc1);

xmu(j)=sum3x/sumbbet;
xbar(j)=mean(xxc1);
xtr(j)=trim(xxc1);
xMAD(j)=MAD(xxc1);
xwave(j)=wave(xxc1);
end

simno=10000

  y=stsrnd1(aa,rr,n,g,simno);
  yy1=sort(y);

%%%   CALCULATES MML FOR STS

for k=1:simno
  for j=1:g
```

```
        yy=yy1(:,j,k);

sumbbet=sum(b);
sum3=sum(b'.*yy);
emu(j)=sum3/sumbbet;
sum4=sum(alfa'.*yy);

ob=lamda*sum4;
sum5=sum(b'.*(yy.^2));
CC=sum5-sumbbet*emu(j).^2;
sgm(j)=(-ob+sqrt((ob)^2+4*n*CC))/(2*sqrt(n*(n-1)));

% CALCULATES LSE FOR STS
my(j)=mean(yy);
stdy(j)=sqrt(var(yy)/mu2);

% CALCULATES TRIM FOR STS
 [meantr stdtr]=trim(yy);
mtr(j)=meantr;
str(j)=stdtr;

 %CALCULATES MADFOR STS
 [meanMAD stdMAD]=MAD(yy);
 mMAD(j)=meanMAD;
 sMAD(j)=stdMAD;

% CALCULATES W24 FOR STS
 [meanwv stdwv]=wave(yy);
 mwave(j)=meanwv;
 swave(j)=stdwv;

end

meanmu=mean(emu);
meansgm=mean(sgm);

varmuhed(k)=meansgm/sqrt(sumbbet);

%Control Limits Using MML);
LCLmml(k)=meanmu-3.0*meansgm/(sqrt(sumbbet)*cmml);
UCLmml(k)=meanmu+3.0*meansgm/(sqrt(sumbbet)*cmml);

meanybar=mean(my);
meanstd=mean(stdy);

%Control Limits Using LSE;
LCLlse(k)=meanybar-3*meanstd/(sqrt(n)*c4);
UCLlse(k)=meanybar+3*meanstd/(sqrt(n)*c4);

meanmtr=mean(mtr);
meanstr=mean(str);

%Control Limits Using TRIM;
LCLtr(k)=meanmtr-3*meanstr/(sqrt(n)*tr4);
UCLtr(k)=meanmtr+3*meanstr/(sqrt(n)*tr4);

 meanmMAD=mean(mMAD);
 meansMAD=mean(sMAD);
```

```
%Control Limits Using MAD;
LCLMAD(k)=meanmMAD-3*meansMAD/(sqrt(n)*md4);
UCLMAD(k)=meanmMAD+3*meansMAD/(sqrt(n)*md4);

meanmwave=mean(mwave);
meanswave=mean(swave);

%Control Limits Using W24;
LCLwv(k)=meanmwave-3*meanswave/(sqrt(n)*cw);
UCLwv(k)=meanmwave+3*meanswave/(sqrt(n)*cw);

end

%Mean of Control Limits will be found
mLCLmml=mean(LCLmml);
mUCLmml=mean(UCLmml);

mUCLlse=mean(UCLlse);
mLCLlse=mean(LCLlse);
mUCLtr=mean(UCLtr);
mLCLtr=mean(LCLtr);

mUCLMAD=mean(UCLMAD);
mLCLMAD=mean(LCLMAD);

mLCLwv=mean(LCLwv);
mUCLwv=mean(UCLwv);

% Probability Plotting Outside The limits will be calculated
  ammlu=find(xmu>mUCLmml);
  ammll=find(xmu<mLCLmml);
  alpmml=length(ammlu)+length(ammll);
   alseu=find(xbar>mUCLlse);
   alsel=find(xbar<mLCLlse);
   alp=length(alseu)+length(alsel);
  atru=find(xtr>mUCLtr);
  atrl=find(xtr<mLCLtr);
  alptr=length(atru)+length(atrl);
  aMADu=find(xMAD>mUCLMAD);
  aMADl=find(xMAD<mLCLMAD);
  alpMAD=length(aMADu)+length(aMADl);
  awaveu=find(xwave>mUCLwv);
   awavel=find(xwave<mLCLwv);
   alpwave=length(awaveu)+length(awavel);

%ARL Values and Type I errors are calculated
alp=alp/yno
ARLlse=1/alp

alpmml=alpmml/yno
ARLmml=1/alpmml

alptr=alptr/yno
ARLtr=1/alptr

alpMAD=alpMAD/yno
ARLMAD=1/alpMAD
```

```
alpwave=alpwave/yno
ARLwave=1/alpwave
```