Shewhart Individual Control Charts for Monitoring Monthly Rainfall in South-West Nigeria

Olubisi L. Aako & Ilesanmi A. Ajibode
The Federal Polytechnic
Ilaro, Ogun State, Nigeria
E-mails: olubisi.aako@federalpolyilaro.edu.ng, ilesanmi.ajibode@federalpolyilaro.edu.ng

ABSTRACT

The Shewhart classic control charts are based on the normality assumption and would fail to properly monitor an asymmetric distribution. Data non-normality is a common problem encountered in statistical process control. Most data that occur naturally tends to depart from normality. Transformation is one of the convenient and easy ways to remedy the problem. In this research, Box-cox and Johnson transformation were used out of many existing normalizing transformation tools to transform the data to approximately normal distribution. The monthly rainfall data in South West, Nigeria, were collected. The individual Shewhart control charts were constructed for Abeokuta, Oyo, Akure and Ikeja transformed rainfall data. The false alarm rate was used to measure the performance of the control charts under each transformation. Box-cox transformation gives lower false alarm rate for Abeokuta, Oyo and Ikeja while Johnson transformation gives lower false alarm rate for Akure.

Keywords: Shewharts control charts, Non-normality, transformation, Box-cox and Johnson transformation, False alarm rate.

iSTEAMS Proceedings Reference Format
Olubisi, L.A. & Ilesanmi, A.A. (2019): Shewhart Individual Control Charts for Monitoring Monthly Rainfall in South-West Nigeria. Proceedings of the 16th iSTEAMS Multidisciplinary Research Nexus Conference, The Federal Polytechnic, Ilaro, Ogun State, Nigeria, 9th – 11th June, 2019. Pp 113-122. www.isteams.net - DOI Affix - https://doi.org/ 10.22624/AIMS/iSTEAMS-2019/V16N2P15

1. INTRODUCTION

Statistical control chart is widely used by modern manufacturing and service organizations. A control chart is a statistical scheme usually allowing graphical implementation devised for the purpose of checking and monitoring the statistical stability of a process. Control charts are useful for tracking statistics over time and detecting the presence of special cause. The special cause results in variation that can be detected and controlled. Common cause variation, on the other hand, is variation that is inherent in the process. A process is in control when only common causes- not special causes affect the process output. Shewhart type control charts are the most commonly used method to test whether or not a process is in control. Its component include a center line and two control limits. The basic idea of shewhart type control chart is that given a quality measurement which is independently and identically follows normal distribution, k-sigma limits would be used to detect an out of control signal. However, the assumption of having independently identical distributed normal population is invalid in many cases, especially encountered frequently in real application. Burr (1967) studies the effect of non-normality on the $\bar{X}$control chart considering various degrees of skewness and kurtosis. He determines constants for each degree of non-normality.

He discussed the effect of non-normality and provided appropriate control constant for different non-normal population. Padgett *et al.* (1992) examine the impact of non-normality on the design scheme in (i) when μ and σ are estimated by their usual estimators, i.e. for μ the mean of the sample means and for σ the mean sample standard deviation or the mean sample range. They also conclude that the in-control probability of signaling of both charts greatly increases under non-normality. Alwan *et al.* (1995) examine 235 quality control applications and find that in most cases the assumptions of normality and independence are not fulfilled, resulting in incorrect control limits. The impact of non-normality on the performance of the control chart can be substantial. Woodal (1997) pointed that, since the statistics used for monitoring with control chart data usually have underlying distributions which are skewed to the right, the traditional k-sigma limits for the Shewhart chart may be inappropriate. Schwertman (1997) proposed some probability based methods for determining control limits in order to improve the control charts performance, they pointed out that control chart properties are determined by the reciprocals of the tails areas, but most approximations, including normal approximation, perform the poorest in the tails of a distribution. But control chart properties are determined by the reciprocal of the actual tail areas. Chen and Cheng (2007) utilized the Johnson distribution and William's (1989) cost model to model an X-bar economic statistical control chart within Weibull distribution failure mechanism for non-normal data.

To overcome the problem of non-normality, simplified transformation tool is applied. Normalizing transformations do not intrinsically change the relative positioning of the data values, but re-express the data while preserving the rank order, to a scale that allows the normal distribution to serve as a benchmark for interpretation and judgment. Among the transformation methods commonly used are the power transformation, Johnson or Pearson distribution system and Box-Cox procedure (Alwan, 2000). This research work examines how to handle non-normal data, transforming it into normal distribution using appropriate transformation tools. By using the statistics of monthly rainfall in south-west, Nigeria from the year 1981-2015, we test whether this variation conform to normal distribution. If not the data are transformed using the appropriate transformation tool to show the variation in measurements during the time period.

2. RELATED LITERATURE

2.1 Shewhart Control Charts

A control chart is one of the primary techniques of statistical process control (SPC). It is a graphical display of a quality characteristic that has been measured or computed from a sample, versus the sample number or time. A control chart always has a central line (CL) for the average, an upper line for the upper control limit (UCL) and a lower line for the lower control limit (LCL). These lines are specified from historical data. Comparing current data to these lines, we can decide whether the process variation is in-control or out of control. In the first case the process is assumed to be in control, and no action is necessary, whereas in the second one, investigation and corrective action are required to find and eliminate the assignable cause or causes responsible for this behaviour. The control chart is also a very useful process monitoring technique, when unusual sources of variability are present. Thus, the systematic use of a control chart is an excellent way to reduce variability. A general model for a control chart is called Shewhart control chart (Maragos, 2015).

Control charts have four key features:

1. Data points are either averages of subgroup measurements or individual measurements plotted on the cartesian axis and joined by a line. Time is always on the x-axis.
2. The Center Line is the average or mean of the data points and is drawn across the middle section of the graph, usually as a heavy or solid line.
3. The Upper Control Limit (UCL) is drawn above the center line and often marked as "UCL". This is often called the "+ 3 sigma" line.

4. The Lower Control Limit (LCL) is drawn below the center line and often marked as "LCL". This is called the "-3 sigma" line.

The general formula for control limits are:

$$\begin{cases} \mathrm{UCL} = \mu_y + k\sigma_y \\ \mathrm{CL} = \mu_y \\ \mathrm{LCL} = \mu_y - k\sigma_y \end{cases} \quad (1)$$

where y is a sample statistic that measures some quality characteristic of interest, μ_y is the mean of y and σ_y the standard deviation of y. *k* is the "distance" of the upper and lower control limits from the center line in terms of the standard deviation units. *k* is often taken as 3, which means that the 99.73 % of all the observations will fall within the control limits under the normality assumption (Sharma, 2003).

2.2 Individual Control Chart

There are many process monitoring problems where application of the rational subgrouping principal leads to a sample size of n = 1. The traditional method of dealing with the case is to use the shewhart control chart to monitor the process mean. The individual control chart, although, as indicated by Borr *et al.* (1999), is easily implemented and can assist in identifying shifts and drifts in the process over time, one of its two widely cited disadvantages is that the performance of the chart can be adversely affected if the observations are not normally distributed. Thus, the individual chart is not robust at all to the normality assumption if false alarms are a concern. To enhance the traditional chart, the main purpose of which is to have a quicker signal. Kittlitz (1999) made the long – tailed, positively skewed exponential distribution by taking the fourth root of the data. The transformed data thus can be plotted conveniently on an individual charts. The rationale for the use fourth – root transformation of the exponential distribution is that it produces essentially a bell-shaped distribution. The usual interpretations can then be easily made for prompt attention if deterioration occurs or captured quickly for an improvement. Borr *et al.* (1999) showed that average run length (ARL) performance of the shewhart individuals control chart when the process is in control is very sensitive to the assumption of normality.

2.3 Box-Cox Transformation

The most common approach nowadays to deal with non- normal data in quality related applications involves the use of the Box – cox transformation, as articulated by Box & Cox (1964). Box-Cox transformation mainly applies a deterministic power function to the raw data by using the estimate of the power transformation parameter, λ. Therefore, the estimation of λ is crucial. The original proposal of the methodology (Box and Cox, 1964) involved the maximum likelihood estimation (MLE).

2.4 Johnson Transformation Method

Johnson (1949) provided an alternative to the Pearson system of curves for modeling non- normal distributions. This approach was to start with a small set of curves capable of approximating the shape of a wide spectrum of probability distributions and then to find simple transformation that would convert these curves into the standard normal distribution. The Johnson system is able to closely approximate many of the standard continuous distributions through one of three functional forms and is thus highly flexible. The corresponding to each pair of mathematically possible values of skewness and kurtosis, any data set can be fitted by a member of the Johnson families such as Su, S_L, and S_B.

This family of distributions is perhaps the most versatile choice. It is based on a transformation of the standard normal variable, and includes four forms.

i. Unbounded (S_U): The set of distributions that go to infinity in both the upper or lower tail.
ii. Bounded (S_B): The set of distribution that have a fixed boundary on either the upper or lower tail, or both.
iii. Log Normal (S_L): A border between the unbounded and bounded distribution forms.
iv. Normal: A special case of the unbounded form.

The fact that the Johnson system involves a transformation of the raw variable to a normal variable allows estimating of the percentile of the fitted distribution to be calculated from the normal distribution percentiles, for use in control limits calculations on the individuals chart or for capability analysis thus, although capability indices and control limits are generally only defined for normal variables, this approach allows their calculation for all distribution types. Many others have shown that the use of Box-cox and Johnson transformations would help quality professionals to perform correct process analysis using both control charts for process stability and capability indices for process capability to meet customer specifications Kilink *et al.* (1999).

3. METHODOLOGY

This section presents the methodology adopted in this research study to transform non-normal data to normal data using Box-cox transformation method and Johnson transformation method. The probability plot is used to assess whether or not the data set is approximately normally distributed. The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality. The individual control chart is then plotted to know if the process is within control or out of control.

3.1 Method of Data Collection

The study uses secondary source of data collection. The data collection for this research was compiled for future use and was collected from the Nigeria Meteorological Agency. The data collected is monthly rainfall in some selected states in south-west, Nigeria. Four states were randomly selected which include Ondo, Ogun, Lagos and Oyo.

3.2 The Individual control charts

Individual control charts are used to monitor individual value.

The central limit and sigma (σ) are estimated from the data. The chart central limit is estimated using the formula:

$$\bar{x} = \frac{\sum xi}{k} \qquad (2)$$

The standard deviation is estimated as:

$$\sigma = \frac{\bar{R}}{d_2} \text{ where} \qquad (3)$$

where $\bar{R} = \frac{\sum_{i=1}^{k} Ri}{k}$, $R_i = / x_i - x_{i-1}/$ and $d_2 = \frac{E(R)}{\sigma} = \frac{(\mu_R)}{\sigma}$

The lower and upper control limits for the individual chart are calculated using the formulas:

$$\begin{cases} \mathrm{LCL} = \overline{X} - \mathrm{m}\hat{\sigma} \\ \mathrm{UCL} = \overline{X} + \mathrm{m}\hat{\sigma} \end{cases} \tag{4}$$

where m is a multiplier (usually set to 3) chosen to control the likelihood of false alarm (out-of-control the signals when the process is in control)

3.4 Box-Cox Transformation

We use λ to estimate the parameter of the Box-cox transformation as well as an alternative method to determine plausible value for it. The former is accomplished by defining a grid of value for λ and further perform a normality test on the λ transformed data . The optimum value of λ, say λ^*, is such that the p-value from the normality test is the highest. The set plausible value is determined using the inverse probability method after plotting the p-value against the value of λ on the grid.

Let y = (y_1 , y_2 … y_n) be the data on which Box-cox transformation is to be applied. The transformation used herein is defined as:

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda}; & \lambda \neq 0 \\ \log y_i & \lambda = 0 \end{cases} \tag{5}$$

Such that, for unknown λ,

$$y^{\lambda} = x\,\beta + \in \tag{6}$$

Where y^{λ} is the λ-transformation data, x is the design matrix (possible covariates of interest), β is the set of parameters associated with the λ -transformed data, and $\in$ is the error term. The aim of equation (5) is that $y_i^{(\lambda)} \sim N(x\,\beta, ln\sigma^2)$ and $\in \sim N(0, \sigma^2)$.

Minitab estimates the optimal value of λ from the data and displays a plot.

3.5 Johnson Transformation Method

The practical solution is to transform the data and drive them towards normality. Basically, the Johnson transformation computes an optimal transformation function from three flexible distribution families (S_U, S_B, S_L). This transformation transforms any continuous random variable into a standard normal variable Z using general form:

$$Z = a + b\, g\left(\frac{x-\mu}{\sigma}\right) \tag{7}$$

a and b are shape parameter, μ is a location parameter and g(x) is a function defining the Johnson system of families, determine as:

$$g(x) = \begin{cases} \ln(x), & \textit{for the lognormal family} \\ \ln\left(x + \sqrt{x^2+1}\right), & \textit{for the unbounded family} \\ \ln\frac{x}{1-x}, & \text{for the bounded family} \\ x, & \text{for the normal family} \end{cases} \tag{8}$$

3.6 Anderson-Darling Normality test

The Anderson-Darling test (Anderson and Darling, 1954) is a statistical test of whether a given sample of data is drawn from a specific distribution. It is one of the most powerful statistical tools for detecting most departures from normality. The test reject the hypothesis of normality when the p-value is less than or equal to 0.05 failing the normality test allows you to state with 95% confidence that data does not fit the normal distribution. Passing the normality test only allows you to state no significant departure from normality was found. The Anderson-Darling test statistic is defined as:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)lnF[Y_i + ln(1 - F\{Y_{n+1-i}\})] \qquad (9)$$

where *n*=the sample size, $F(Y)$ is the cumulative distribution function for the specified distribution.

4. RESULTS

In this section, the monthly rainfall in Abeokuta, Oyo, Ikeja and Akure would be analyzed. The sample sizes taken for each location were large enough for the possibility of a good knowledge of probability distribution.

4.1 Tests for Normality

Figure 1 shows the plot of histogram and the normal distribution curve for data from each state.

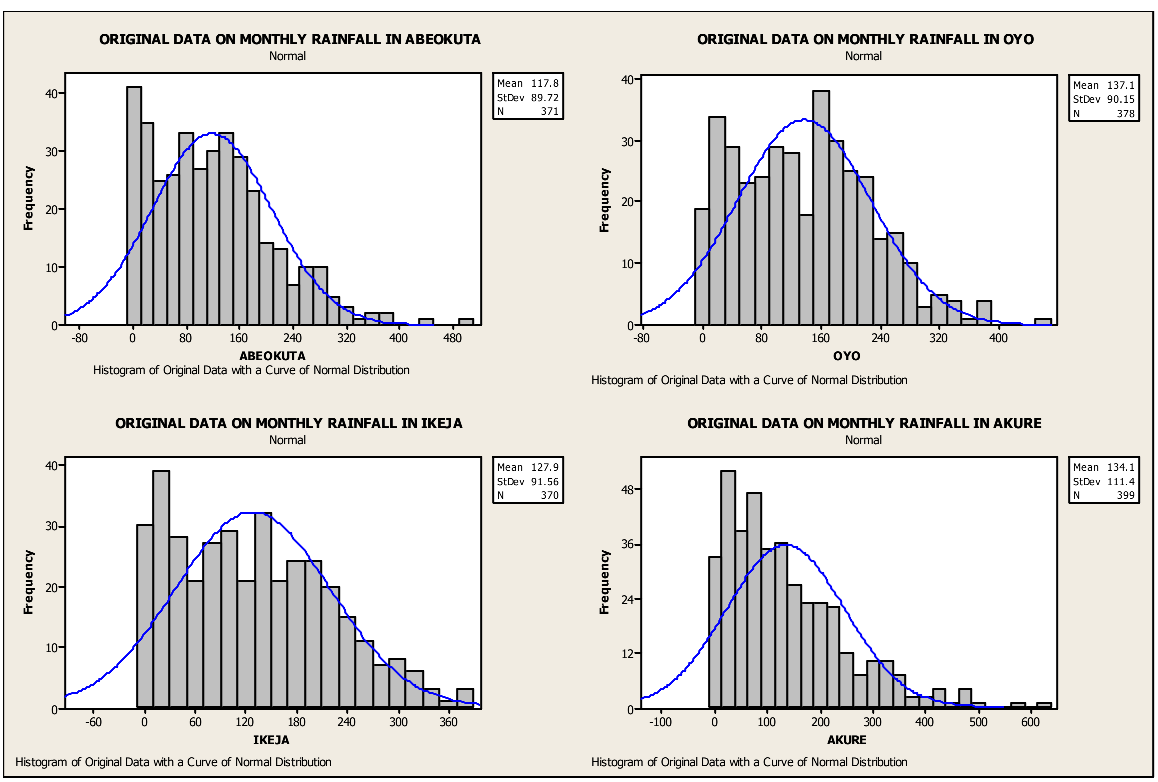

Fig1: Histogram of various states.

From Figure 1, it can be seen that the original rainfall data from various states are not normal, since not all the bars fall under the normal distribution curve. The mean value of rainfall in Abeokuta is 117.8 with standard deviation of 89.72 out of 371 observation, in Oyo the mean value is 137.1 with standard deviation of 90.15 out of 378 observation, in Ikeja the mean value is 127.9 with standard deviation of 91.56 out of 370 observation, in Akure the mean value is 134.1with standard deviation of 111.4 out of 399 observation.

Figure 2 below shows the probability plot and Anderson Darling test for the original data of the monthly rainfall in the selected states.

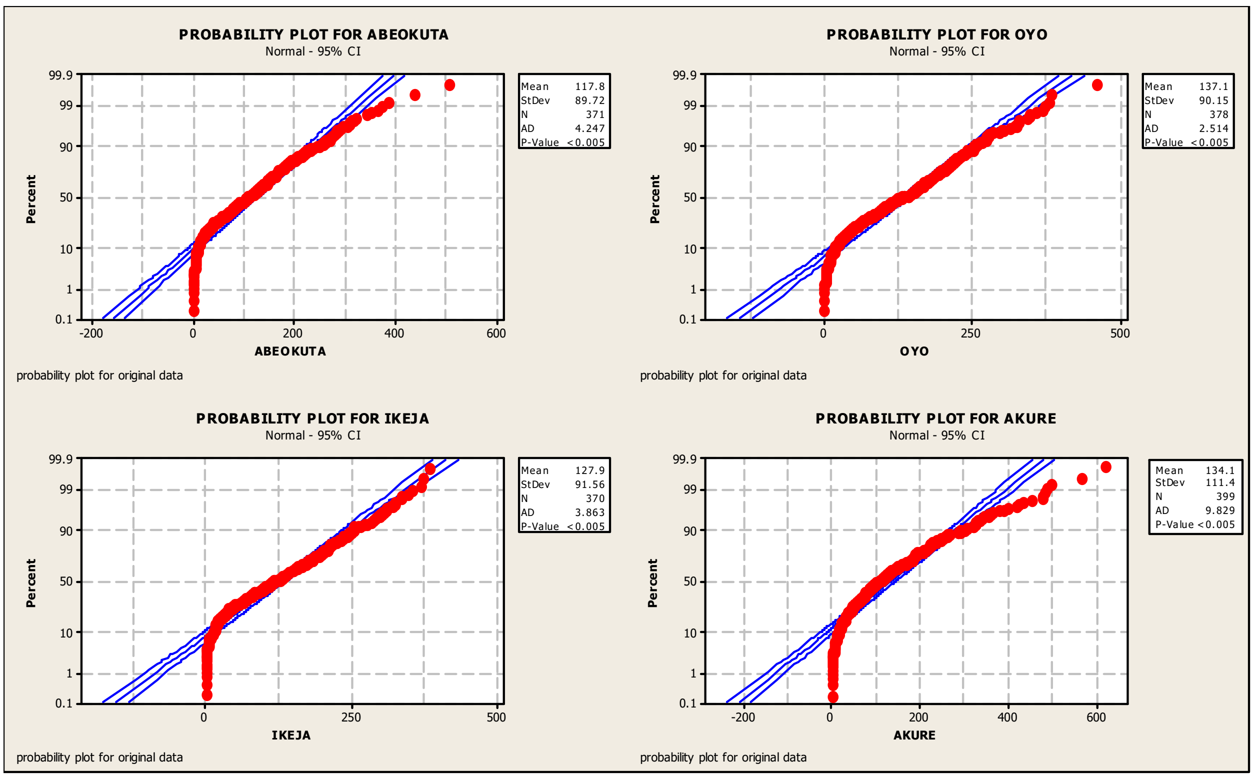

Fig 2: Probability plot and Anderson Darling test for the original data of the monthly rainfall in the selected states.

The Anderson Darling test for Abeokuta is 4.247, Oyo is 2.514, for Ikeja is 3.863, for Akure is 9.829. The Anderson Darling test for each state in south west, Nigeria, shows P-value less than 0.05 which indicates that the set of data are not normal. From the discussion of result above, it is observed that the set of data were not normal, therefore there is need to transform.

4.2 Tranformation

4.2.1 Box-Cox Transformation

Table 1: Box-Cox Optimum Lambda, Original data AD values and Transformed data AD values for each location.

Location	Optima Lambda	Original Data AD Values	Transformed Data AD Values
Abeokuta	0.32	4.247	5.354
Oyo	0.39	2.514	4.657
Ikeja	0.36	3.863	4.948
Akure	0.26	9.829	4.657

Table 1 shows the lambda values that best transform the original data for each location together with the Anderson Darling values for original and transformed data using Box-cox method in Minitab package.

4.2.2 Johnson Transformation

Table 2: Johnson Type, Transformation Functions, Original data AD values and Transformed data AD values for each location.

Location	Type of Transfor-mation	Transformation Function	Origin-al Data AD Values	Transfo-rmed Data AD Values
Abeokuta	S_L	$-95.43 + 13.435 * \ln(X + 1098.83)$	4.247	3.365
Oyo	S_B	$1.166 + 0.785 * \ln\left(\frac{X + 0.418}{480.26 - X}\right)$	2.514	7.283
Ikeja	S_B	$0.864 + 0.682 * \ln\left(\frac{X + 0.350}{391.4 - X}\right)$	3.863	5.841
Akure	S_B	$1.404 + 0.896 * \ln\left(\frac{X + 11.78}{654.53 - X}\right)$	9.829	0.576

Table 2 gives type of Johnson transformation, Transformation function, Original data AD values and Transformed data AD values for each location.

Table 3: Individual Control Charts Limits and False Alarm Rate for the Original data, Box-Cox transformed data and Johnson transformed data for various locations.

	Original Data			**Box-Cox Transformed Data**			**Johnson Transformed Data**		
Location	LCL	UCL	FAR	LCL	UCL	FAR	LCL	UCL	FAR
Abeokuta	-103.0	338.7	0.0189	0.981	7.453	**0.0108**	-2.413	2.365	0.0135
Oyo	1.90	10.99	0.0344	0	5.582	**0.0053**	0	2.702	0.0238
Ikeja	0	3.024	0.0243	0	263.5	**0.0081**	0	5.033	0.0135
Akure	-138.8	407.1	0.0318	1.170	5.374	0.0132	-2.462	2.462	**0.0079**

Table 3 gives Individual Control Charts Limits and False Alarm Rate for the original data, Box-Cox transformed data and Johnson transformed data for various locations. Box-Cox transformation gives the smallest false alarm rate for Abeokuta, Oyo and Ikeja data while Johnson transformation gives the smallest false alarm rate for Akure data. This showed that Box-cox transformation is the best tool to transform Abeokuta, Oyo and Ikeja monthly rainfall data to approximate normal distribution and Johnson transformation is the best for Akure monthly rainfall data.

5. CONCLUSION

The study was set to monitor the monthly rainfall in Abeokuta, Oyo, Ikeja and Akure using Individual Control Charts which is based on the assumption of normality. The observed data was tested for normality using histogram, p-p plot and Anderson-Darling. It was shown that monthly rainfall data from various locations were not normally distributed. Box-cox and Johnson transformation tools were employed for appropriate transformation. Data transformed using Box-cox and Johnson transformation tools were used to construct Individual Control Charts. These were compared on the bases of False Alarm Rate with the control charts of original data. The result showed that the false alarm rate of the transformed data is low when compared with the original monthly rainfall in south west, Nigeria. However, out of the two transformation tools chosen, Box- Cox transformation performed better than the Johnson transformation in three of the four locations.

REFERENCES

1. Alwan, L.C. (2000). Designing an Effective Exponential Chart without the use of Monographs, *Communications in Statistics: Theory and Methods*, *Vol. 29 (12),* pp 2879- 2893.
2. Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. J.Am. Stat. Assoc., Vol. 49, pp 765-769.
3. Borr, C. M. (1999). Robustness of the Variable Control Charts to Non-normality. *Journal of Quality Technology*, *Vol.31, Iss.3*, pp 309-316.
4. Burr, I.W. (1967). The Effect of Non-normality on Constants for $\bar{X}$ and R Charts. *Journal of industrial quality control. Vol 23(10),* pp 563-568.
5. Box, G.E.P. and Cox D.R. (1964). An Analysis of Transformations. *Journal of Royal Statistical Society B, 26(5),* pp 211-243.
6. Chen, H and Cheng, Y. (2007) "Non-normality effects on the economic-statistical design of *X*
7. charts with Weibull in-control time", *European Journal of Operational Research, Vol. 176,* No. 2, pp. 986-998.
8. Johnson, N. L. (1949). Systems of Frequency Curves Generated by Methods of Translation. *Biometrica Vol.36,*(1), pp. 149-176.
9. Kittlitz, R. G., Jr.(1999). Transforming the Exponential for SPC applications. *Journal of Quality Technology*, *Vol.31(3),* pp 301-308.
10. Maragos, I. I (2015). Control charts for skewed distributions. A thesis submitted to the department of statistics of the *Athens University of Economics and Business.*
11. Padgett, C. S., Thombs L. A., Padgett, W. J. (1992). On the alpha-risk for Shewhart control charts. *Communication in Staristics- Simulation and Computation,* 21:1125-1147.
12. Sharma, P. R. (2003). A study of properties and applications of Control Charts for High Yield Processes. A thesis submitted to the department of Industrial and Systems Engineering, *the National University of Singapore*.
13. Shwertman, N.C. and Ryan T.P. (1997). Implementing Optimal Attribute Charts. *Journal of Quality Technology*, *Vol.29*(1), pp 99-104.
14. Williams, T.P. (1989). Economic control chart designs and the in-control time distribution: a sensitivity analysis", Journal of Quality Technology, Vol. 21, No. 2, pp. 103-110.
15. Woodall, W.H. (1997). Control Charts based on Attribute data: Bibliography and Review. *Journal of Quality Technology*, Vol. 29(2), pp172-183.