Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.169-181

Using the kurtosis correction method to design quality control limits for asymmetrical distributions: A comparative study in Kirkuk city

AbdulRahim K. Rahi Alharithy[1] , Wakkaa Ali Hadba[2], Ahlam Ahmed Juma[3]
[1] College of Administration & Economics, Mustansiriya University, Iraq
[2] College of Science, University of Kirkuk
[3] College of Arts, University of Baghdad

ABSTRACT

In this paper, the limits of the mean and the range are constructed when the normal distribution condition is not available for the measured quality (relative humidity in Kirkuk city). By using kurtosis correction as well as Johnson's transform to convert the data to normal distribution, it is appeared to be subjected to the smallest extreme value distribution and then compared the new charts with Shewhart chart using process capability. The results show that the new charts are more efficient and sensitive to changes in the production process, which makes the proposed method of dealing with asymmetric data more benefit.

Keywords: Shewhart chart, kurtosis correction, Johnson's transform, extreme value distribution.

Corresponding Author:
AbdulRahim K. Rahi Alharithy
College of Administration & Economics
Mustansiriya University, Iraq
E-mail: rahistat@yahoo.com

1. Introduction

Control charts are important tools used in industry to control of production process by statistical control. Shewhart charts include the average chart ($\bar{X}$), the range chart (R) and the standard deviation chart (S). Shewhart assumed when constructing the average chart based on the normal distribution of quality characteristic data, and this assumption was widely used in statistical process control (SPC) of the average. In many cases this condition is not available especially when the quality characteristic variable has a skewed (asymmetric) distribution which means that the process output is inaccurate when using the Shewhart chart which causes an error of type I and the false alarm rate rises [1].

Tadikamalla and Popescu [2], proposed a kurtosis correction method for constructing control charts for symmetrical long-tailed distributions. Metlapalli [3], investigated the performance of control limits and compared through a simulation study. This comparison is based on in Laplace Distribution. Goswami and Dutta [4], have investigated process capability indices based on normal and non-normal processing data. The results evidently have shown that the process capability investigation of data for non-normal distribution may considerably fluctuate based on approaches of analysis for process capability. The authors of [5-6] , proposed Johnson distribution as a model for normalizing real field data showing departure from the normality. Aydin [7], has considered different estimates of the quantiles of two-parameter Gumbel distribution and modeling of the data to estimate the minor and higher quantiles of Gumbel distribution in the application of wind speed data. In [8], some important issues about process capability and performance have been highlighted based on the non-normal distribution of a characteristic process.

In this paper, we applied Johnson's transform for the data to calculate the limits of control charts before and after transformation. Also, the process capability criteria is used to compare among the charts of Shewhart, before and after transformation using Johnson's transformation and kurtosis correction.

2. Shewhart charts

Shewhart chart consists of three parallel horizontal lines which are upper control limit (UCL), center line or target line, lower control limit (LCL), and the chart put at $(\pm 3.09\sigma)$ of target line as in Figure 1 [1].

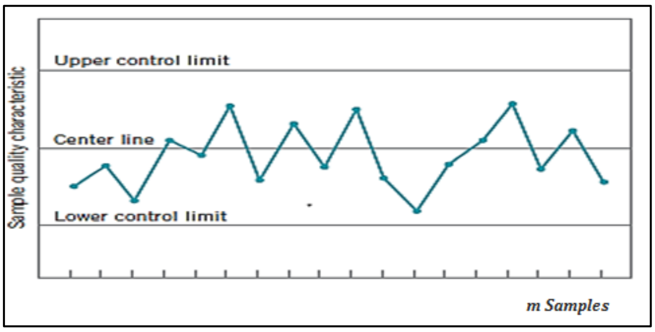

Figure 1. Shewhart chart

3. Control chart for variables

In the control chart of the variables where the variable is quantitative and has a unit of measurement such as weight in kilograms or liters where several samples of production are withdrawn at regular intervals. The most important charts are the average and range charts and the mean deviation chart as well as the accumulated total charts and the moving averages of various types [1].

3.1 Control chart for mean ($\bar{X}$ *Chart*)

When building the control chart of the mean of the production process, a number (m) of samples is drawn so that each sample includes (n) units. The limits of this chart as follow:

$$\left.\begin{aligned} UCL_{\bar{X}} &= \bar{\bar{X}} + A_2\bar{R} \\ CL_{\bar{X}} &= \bar{\bar{X}} \\ LCL_{\bar{X}} &= \bar{\bar{X}} - A_2\bar{R} \end{aligned}\right\} \quad (1)$$

where,

$\bar{\bar{X}}$: The general mean of all samples.

$\bar{R}$: Average range for all samples.

A_2: The constant control is extracted from the tables of control charts or equal to $\frac{3\sigma}{\sqrt{n}}$.

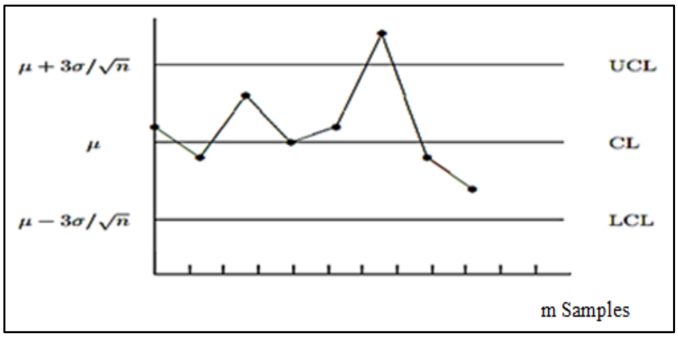

Figure 2. The control chart of the mean

3.2 Range control chart (*R*-Chart)

When the control chart is built for the production process, a number (m) of samples is drawn so that each sample includes (n) units. In general, this chart consists of three parallel horizontal lines representing the limits of control as in Figure 3 based on:

$$\left.\begin{array}{r}\mathrm{UCL}_R = \mathrm{D}_3\overline{\mathrm{R}} \\ \mathrm{CL}_R = \overline{\mathrm{R}} \\ \mathrm{LCL}_R = \mathrm{D}_4\overline{\mathrm{R}}\end{array}\right\} \qquad (2)$$

where,

$\bar{R}$: Represents the average range of all samples
$\mathrm{D}_3, \mathrm{D}_4$: Control constants are extracted from the tables of control charts.

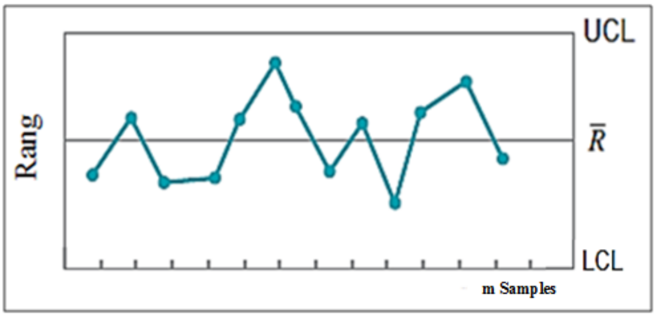

Figure 3. The control chart of the range

4. Johnson's transform

Johnson's transform has a wide range of applications and is also applicable to negative values. Johnson proposed three transformations for normal distribution and the general formula are [5]:

$$Z = \gamma + \sigma f\left(\frac{X-\mu}{\lambda}\right), \qquad \sigma > \lambda > 0 \qquad (3)$$

Where,
$f(\cdot)$ stands for the transform function, Z is the random variable of the standard normal distribution, γ and σ are the shape parameters, λ is the scale parameter, and μ is the location parameter [5, 6].

I. The first transformation defines the lognormal system of distributions denoted by (LS) and is used to transform the logarithmic normal distribution and its formula is equal to:

$$Z = \gamma + \sigma ln\left(\frac{X-\mu}{\lambda}\right) = \gamma + \sigma ln(X-\mu), \qquad X > \mu \qquad (4)$$

II. The second transformation of the specified random variable denoted by (SB) and its formula is:

$$Z = \gamma + \sigma ln\left(\frac{X-\mu}{\mu+\lambda-X}\right), \qquad \mu < X < \mu + \lambda \qquad (5)$$

III. The third transformation of the indeterminate random variable denoted by (SU) and its formula is:

$$Z = \gamma + \sigma \ln\left[\left(\frac{X-\mu}{\lambda}\right) + \left\{\left(\frac{X-\mu}{\lambda}\right)^2 + 1\right\}^{\frac{1}{2}}\right]$$

$$= \gamma + \sigma\, sinh^{-1}\left(\frac{X-\mu}{\lambda}\right), \qquad -\infty < X < +\infty \qquad (6)$$

(SU) is used for data transformation of undefined curves such as normal distribution and distribution of (t).

5. Gaussian distribution

Gaussian distribution is the most common statistical distribution because the normal state almost appears in many social, vital and other situations. Many statistical analyzes require that the data follow the normal distribution formulated by [9]:

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right), \qquad -\infty < X < \infty \qquad (7)$$

where,

$EX = \mu, \qquad V(X) = \sigma^2$

6. Gumbel distribution (smallest extreme value distribution)

The random variable X is said to be distributed according to Gumbel distribution if its probability density function is equal to [7]:

$$f(X:;\mu,\beta) = \frac{1}{\beta} e^{-\left(\frac{X-\mu}{\beta}+e^{-\left(\frac{X-\mu}{\beta}\right)}\right)}, \ -\infty < X < \infty, -\infty < \mu < \infty \qquad (8)$$

since,

$\beta > 0$ (*scale parameter*), μ : *location parameter*

The standard distribution of $Z = \frac{X-\mu}{\beta}$ is equal to:

$$f(z) = \frac{1}{\beta} e^{-(z+e^{-Z})}, \qquad -\infty < Z < \infty \qquad (9)$$

$$V(X) = \frac{\pi^2}{6}\beta^2, \qquad \mu = \bar{X} - 0.5772\,\beta$$

7. Kurtosis in quality control

Kurtosis is defined as the amount of flatness or fatigue of the distribution curve of a random variable. The concept of kurtosis is closely related to the concept of dispersion. The higher the dispersion of the values of the variable, it is an indicator of the flatness of the probability distribution curve [2, 3].

When the value of kurtosis equal to (3), it means that the kurtosis of the curve is normal. When the value of kurtosis is greater than (3), the curve is tapered. Finally, when the kurtosis value is less than (3), it is an indication that the curve is flat.

The degree of kurtosis can be measured by:

$$\alpha_4(\bar{\mathrm{X}}) = \frac{1}{\mathrm{nm}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{m}} \left(\frac{\mathrm{X_{ij}} - \bar{\bar{\mathrm{X}}}}{\mathrm{S_{nm}}}\right)^4 \qquad (10)$$

The standard deviation of the range is equal to:

$$\mathrm{S_R} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(Ri - \bar{R})^2} \qquad (11)$$

The kurtosis coefficient of the range is equal to:

$$\alpha_4(R) = \frac{1}{m} \sum_{\mathrm{i}=1}^{\mathrm{m}} \left(\frac{Ri - \bar{R}}{\mathrm{S_R}}\right)^4 \qquad (12)$$

8. Kurtosis correction method

Assume that the observations for the quality characteristic variable X_{ij} are $(i = 1,2,\cdots m \ and \ j = 1,2,\cdots,n)$ consists of m of subunits the size of each is equal to n, if $\bar{X}_i$ represents the mean for each subunit, and $\bar{\bar{X}}$ represents the general mean of all subunits based on the Karl Pearson's kurtosis coefficient for sample[10].

In 2007, Tadikamalla and Popescu developed a control chart for the mean and range based on the kurtosis coefficient as follows [2]:

<u>$\bar{X}$ *Chart*</u>

$$\left.\begin{aligned} UCL_{kc} &= \bar{\bar{X}} + \left[3 + \frac{\alpha_4(\bar{X})}{1+0.33\alpha_4(\bar{X})}\right]\frac{\bar{R}}{\sqrt{n}d_2^{kc}} = \bar{\bar{X}} + A_{\mathrm{U}}^{\mathrm{kc}}\bar{R} \\ CL_{kc} &= \bar{\bar{X}} \\ LCL_{kc} &= \bar{\bar{X}} - \left[3 + \frac{\alpha_4(\bar{X})}{1+0.33\alpha_4(\bar{X})}\right]\frac{\bar{R}}{\sqrt{n}d_2^{kc}} = \bar{\bar{X}} - A_{\mathrm{L}}^{\mathrm{kc}}\bar{R} \end{aligned}\right\} \qquad (13)$$

where,
d_2^{KC} is also the mean of relative range R/σ_x.
$\alpha_4(\bar{X})$ is given in eq.10.
$A_{\mathrm{U}}^{\mathrm{kc}}, A_{\mathrm{L}}^{\mathrm{kc}}$ are given by Tadikamalla and Popescu.

Tadikamalla and Popescu (2007) had given the representative values of d_2^{KC}'s corresponding to different kurtosises and sample sizes.
The kurtosis correction method (KC) of the R-chart is equal to:

$$\left.\begin{aligned}
UCL_{kc} &= \bar{R} + \left(3 + \frac{\alpha_4(R)}{1+0.33\alpha_4(R)}\right)\bar{R}\frac{d_3^{kc}}{d_2^{kc}} \\
&= \left[1 + \left(3 + \frac{\alpha_4(R)}{1+0.33\alpha_4(R)}\right)\frac{d_3^{kc}}{d_2^{kc}}\right]\bar{R} = \bar{R}D_4^{kc} \\
CL_{kc} &= \bar{R} \\
LCL_{kc} &= \bar{R} - \left(3 + \frac{\alpha_4(R)}{1+0.33\alpha_4(R)}\right)\bar{R}\frac{d_3^{kc}}{d_2^{kc}} \\
&= \left[1 - \left(3 + \frac{\alpha_4(R)}{1+0.334\alpha_4(R)}\right)\frac{d_3^{kc}}{d_2^{kc}}\right]\bar{R} = \bar{R}D_3^{kc}
\end{aligned}\right\} \quad (14)$$

where,
d_3^{kc}: is the standard deviation of the relative R/σ_x.
$\alpha_4(R)$: is the kurtosis coefficient of the range given in eq.12.
$D_3^{kc}$and D_4^{kc} : are given by Tadikamalla and Popescu.

9. Process capability indices (PCI)

The process capability indicator is an effective tool for the purpose of providing us with a numerical scale used to determine the relationship between the dispersion of the process and the limits of specifications, this indicator also shows the efficiency of the process as well as the evaluation and improvement of the process and the production of units according to specifications, although it is easy to understand and apply where it reflects the performance of the process and the summary of what is happening in production in numerical value [8].
The formulas for this index in the normal distribution are equal to [4]:

$$\left.\begin{aligned}
C_P &= \frac{USL-LSL}{6\sigma} \\
C_P^U &= \frac{USL-\mu}{3\sigma} \\
C_P^L &= \frac{\mu-LSL}{3\sigma} \\
C_{PK} &= Min(C_P^U, C_P^L)
\end{aligned}\right\} \quad (15)$$

USL: Upper specification limit

LSL: Lower specification limit

The formulas for this index in the non-normal distribution are equal to:

$$P_P = \frac{USL-LSL}{X_{99.865}-X_{0.135}} \quad (16)$$

$$\left.\begin{array}{l} P_P^U = \frac{\mathrm{USL}-\mathrm{X}_{50}}{\mathrm{X}_{99.865}-\mathrm{X}_{50}} \\ P_P^L = \frac{\mathrm{X}_{50}-\mathrm{LSL}}{\mathrm{X}_{50}-\mathrm{X}_{0.135}} \\ P_{PK} = Min(P_P^U, P_P^L) \end{array}\right\} \quad (16)$$

where:

$X_{0.135}$: Quantile of appropriate distribution.
$X_{99.865\%}$: Quantile of appropriate distribution.
$X_{50\%}$: Median of appropriate distribution.

The interpretation of C_P and C_{PK} can be summarized as follows [4]:
1. If $C_P = C_{PK}$, then it implies that the process is centered at the midpoint of the specification.
2. If $C_{PK} < C_P$, then it implies that the process is off-center. The user should try to center the process.
3. If $C_{PK} = 0$, then the process is exactly equal to the one of the specification limit.
4. If $C_P < 0$, then the process mean is outside the specification limits.
5. If $C_P = 1$, it implies that the process is centered
6. If $C_P < 1$, it implies that the process is not fully capable.
7. If $C_P > 1$, capable of meeting the specification.

10. The practical side

Data were collected representing the humidity rates in Kirkuk city for three months (December, January, and February) within (1988-2018) period. It stands for 31 values at each of the three months has been analyzed using the Minitab V.18. Figure 4 shows the data.

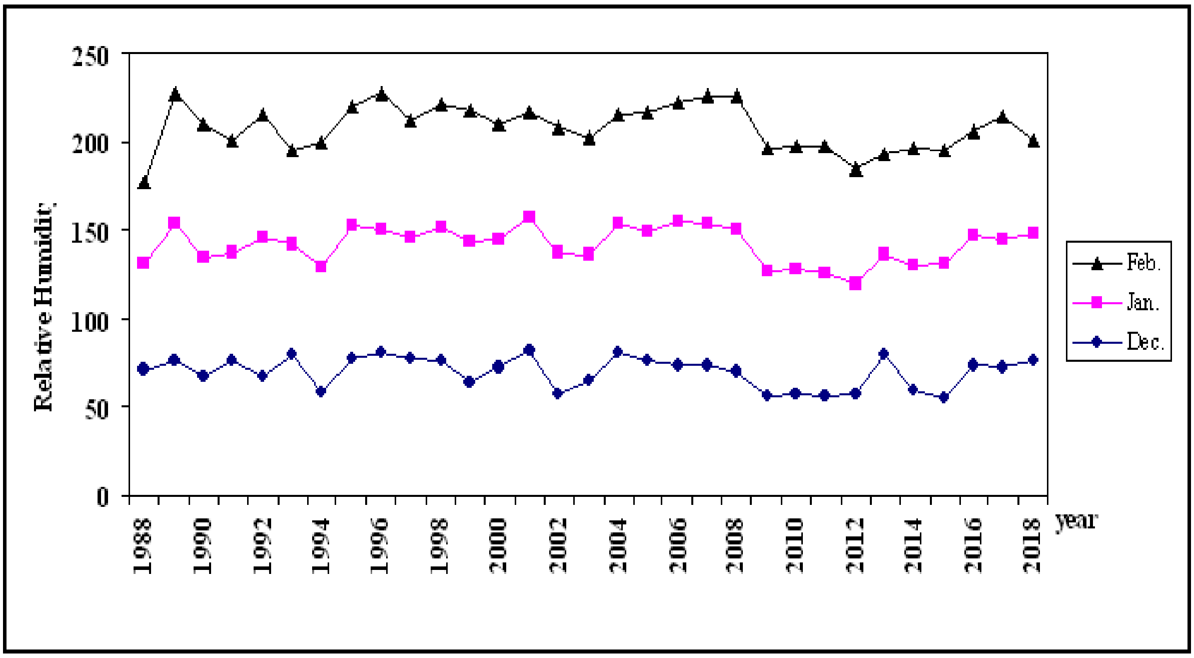

Figure 4. Humidity rates in Kirkuk city

Testing the data shows that it does not follow the normal distribution, but it follows the distribution of smallest extreme values as shown in Figure 5.

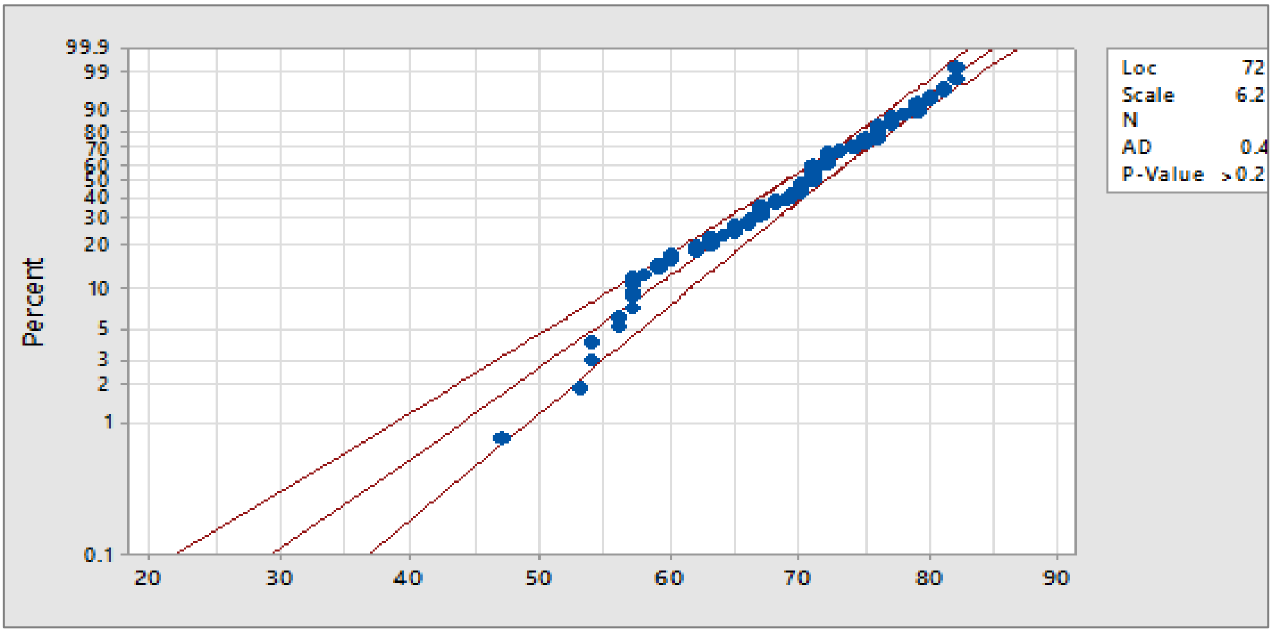

Figure 5. Test of the distribution of smallest extreme values

The mean chart was plotted and there was a point out of control. Figure 6 shows the control limits of the Shewhart chart with LCL = 56.92 and UCL = 81.55

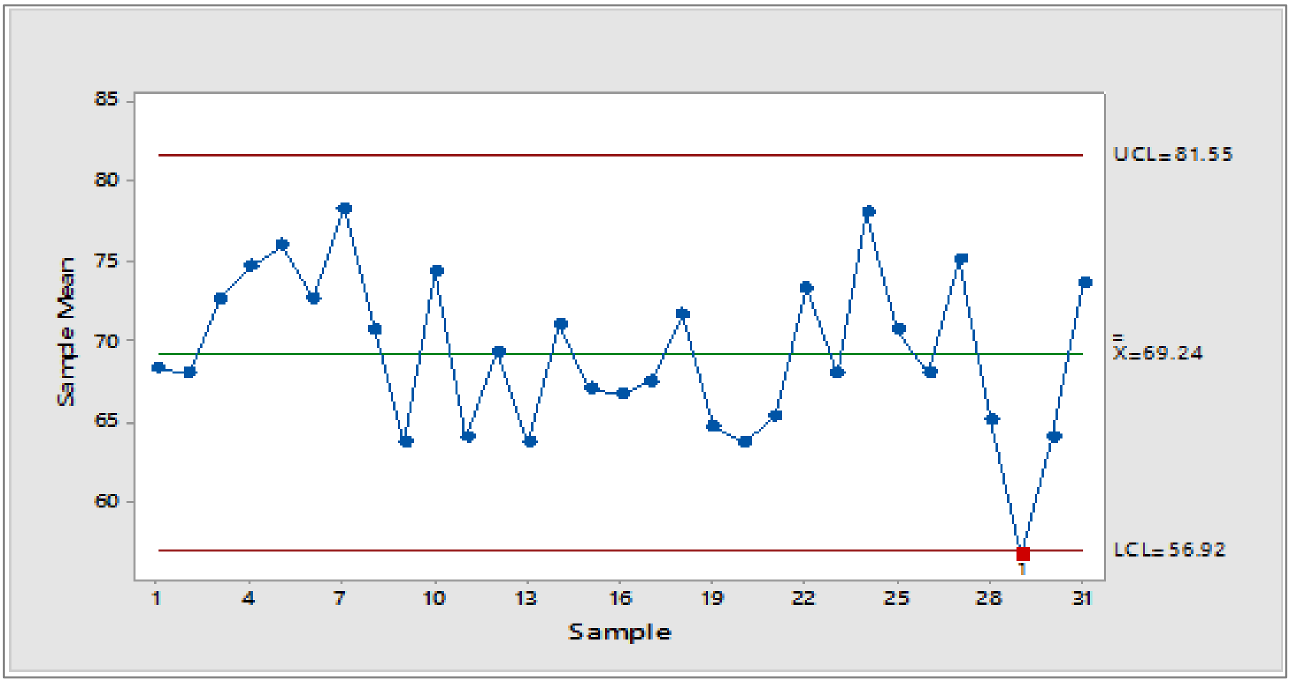

Figure 6: Shows the Shewhart chart for the mean

The transform data using the Johnson system is shown in Figure 7. Note that the best model is (SB) for Johnson's distribution to the data of humidity rates, as well as the best "fit" for data is within the period (0.749 – 0.9925). In other words, the normality is certainly within this period, which allows the determination of both the lower and upper limits of the data for the normal type and their values (LCL = 47, UCL = 82).

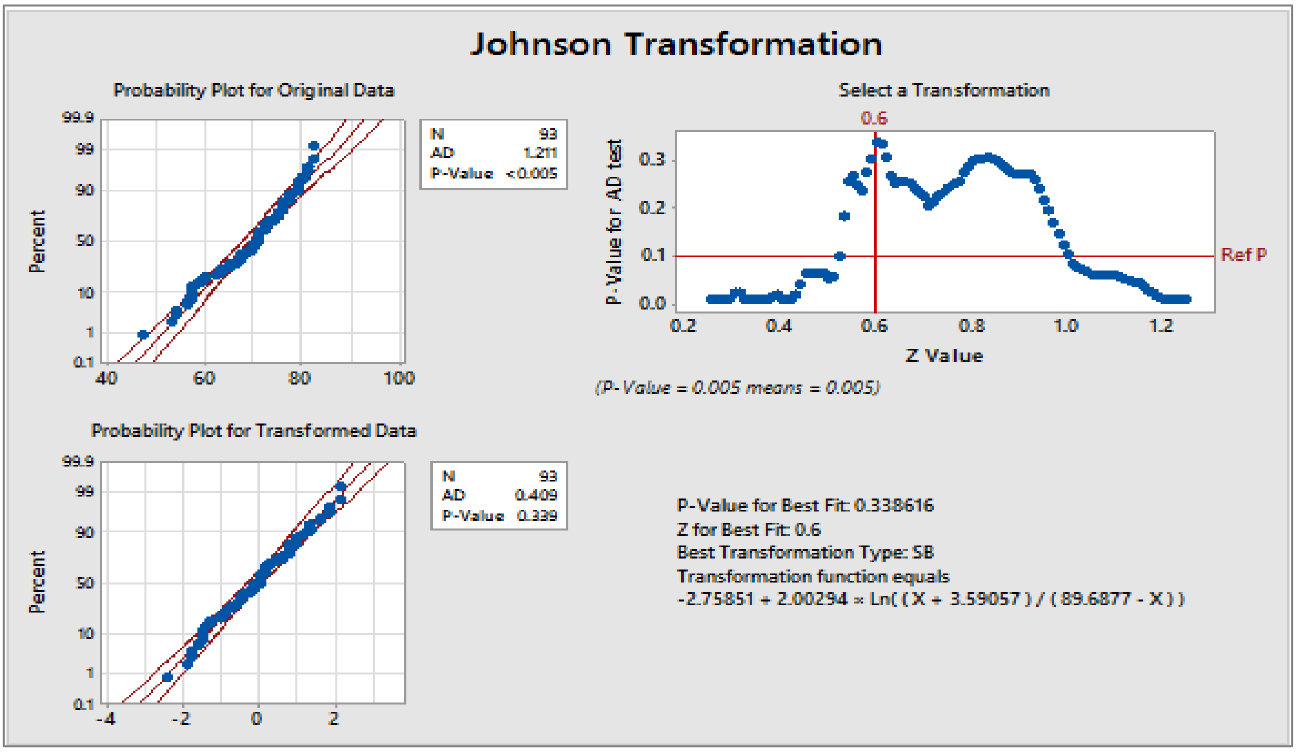

Figure 7. The data transformation by Johnson's method

These limits are considered to be new limits for the mean chart as in Figure 8. It is clear that they are broader than the standard control limits of Shewhart which relies on the normal distribution to contain all points. The process after using the Johnson transformation has become under statistical control and the point outside the control limits in the Shewhart chart. It can be considered a false alarm for the process. Therefore, the use of Johnson's data transformation avoided the process of the emergence of false alarms in terms of stability and productivity.

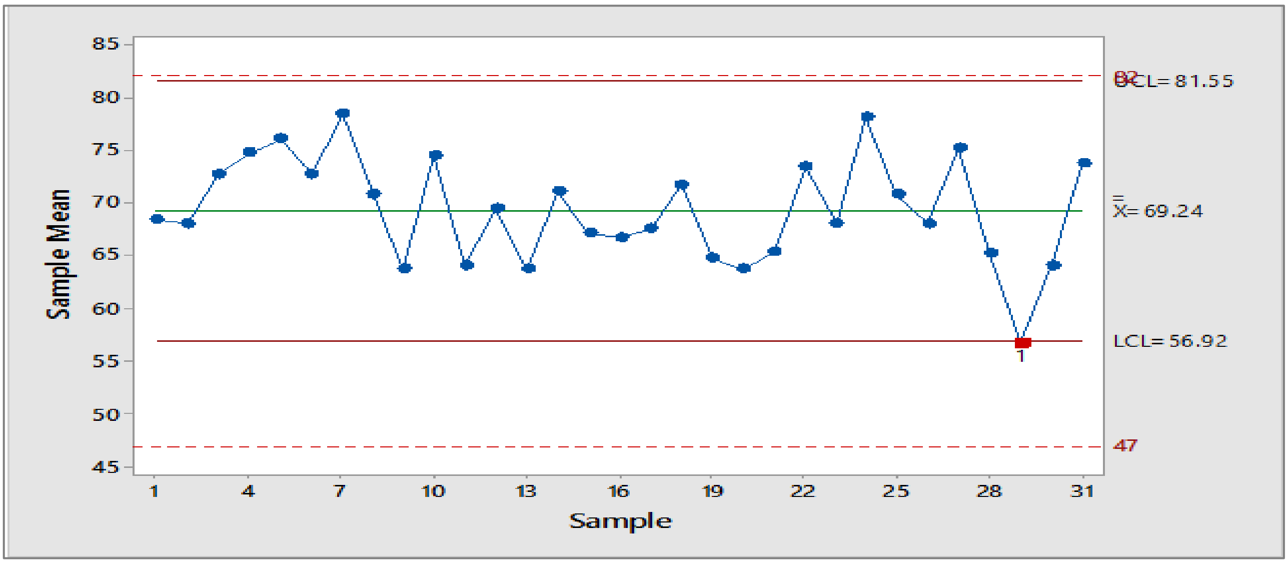

Figure 8. The mean chart after the data transformation using Johnson's method

Using the formulas in (13) to find the control limits of the mean by the method of kurtosis correction, the lower limit is equal to (54.78) and the upper limit is equal to 84. Figure 9 shows that the control limits have expanded and included the point outside the control limits and therefore the process under statistical control.

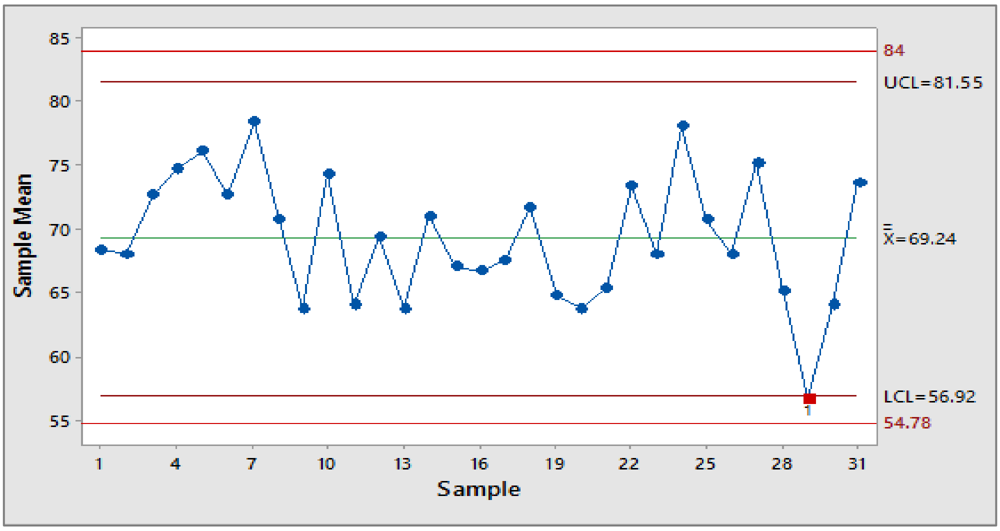

Figure 9: Control limits using kurtosis correction method

The range chart is illustrated in Figure 10 as all points are within the control limits.

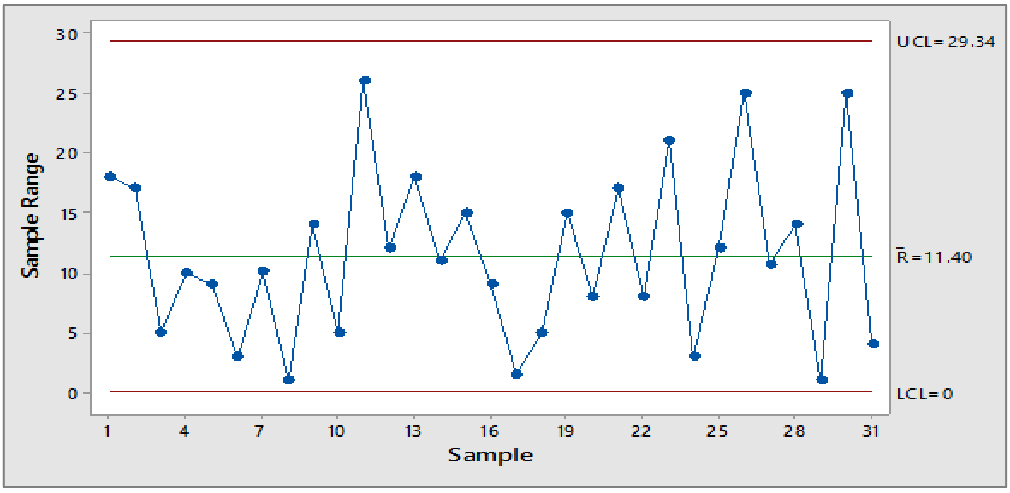

Figure 10. Shewhart chart for the range

When applying the formulas in (14), it has found that the lower control limit is equal to (0) and the upper control limit for the method of kurtosis correction is equal to 43.11 as in Figure 11.

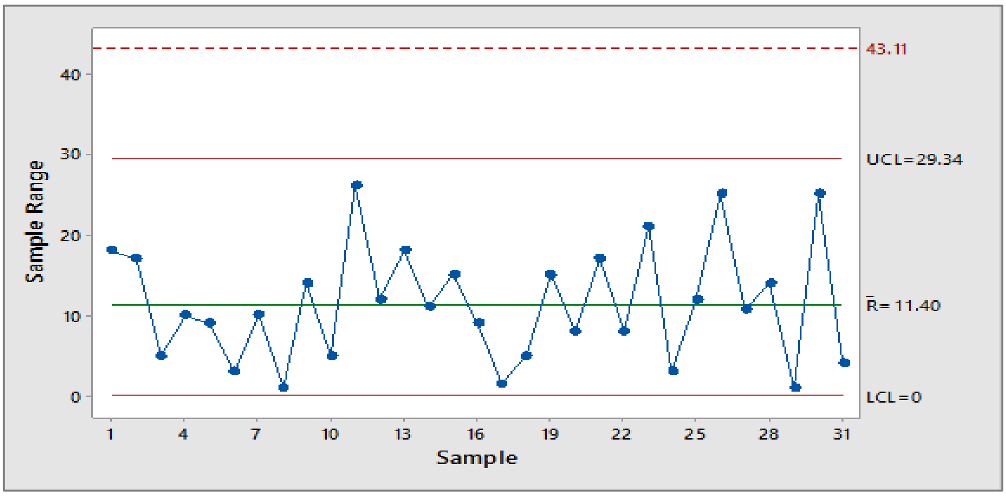

Figure 11. The range chart in the kurtosis correction method

When calculating the process capability of the Shewhart chart before the data transformation shown in Figure 12, the values of $P_{PK} = 0.54$ and $C_{PK} = 0.57$

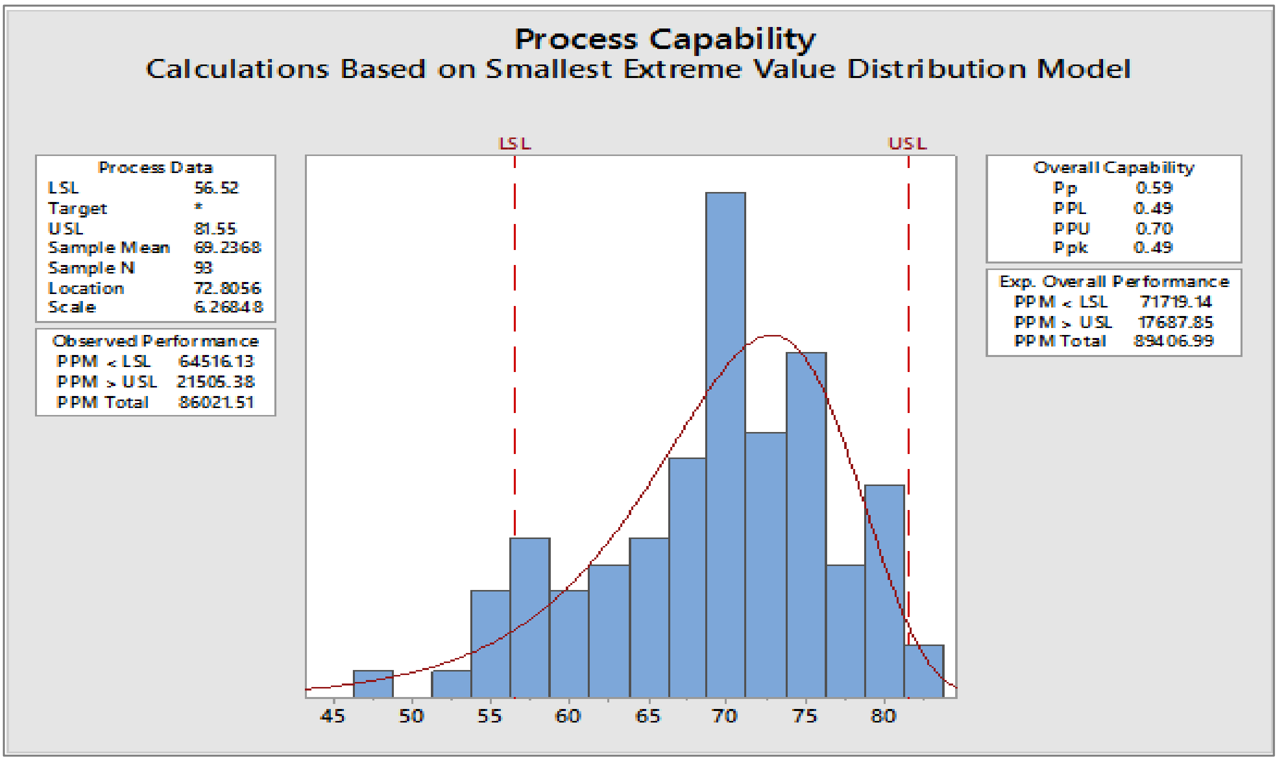

Figure 12. The process capability of the Shewhart chart before the transfer of data for normal distribution

When calculating the process capability of the Shewhart chart before transformation the data of the distribution of extreme values, Figure 13 shows that $P_{PK} = 0.49$

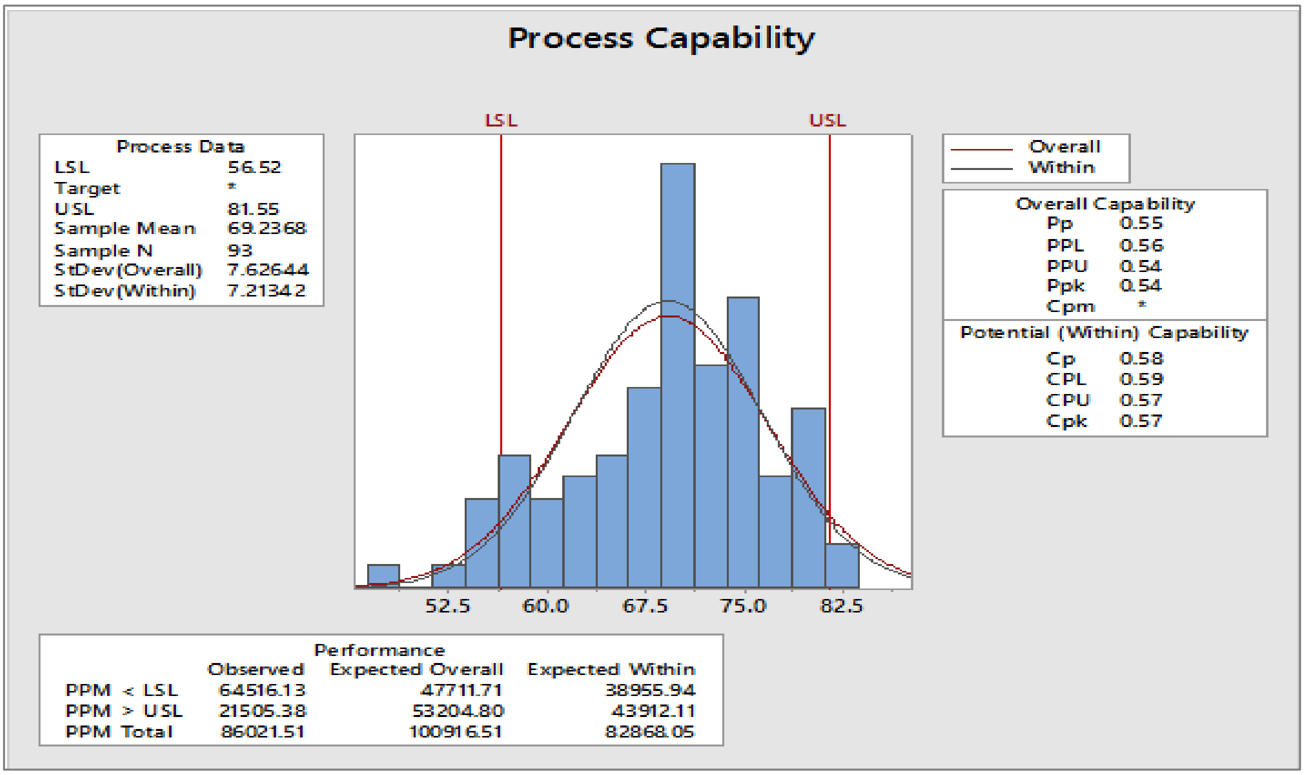

Figure 13. The process capability of the Shewhart chart for extreme values distribution

When calculating the process capability after the data transformation in Johnson's method, Figure 14 shows that $P_{PK} = 0.56$ *and* $C_{PK} = 0.60$

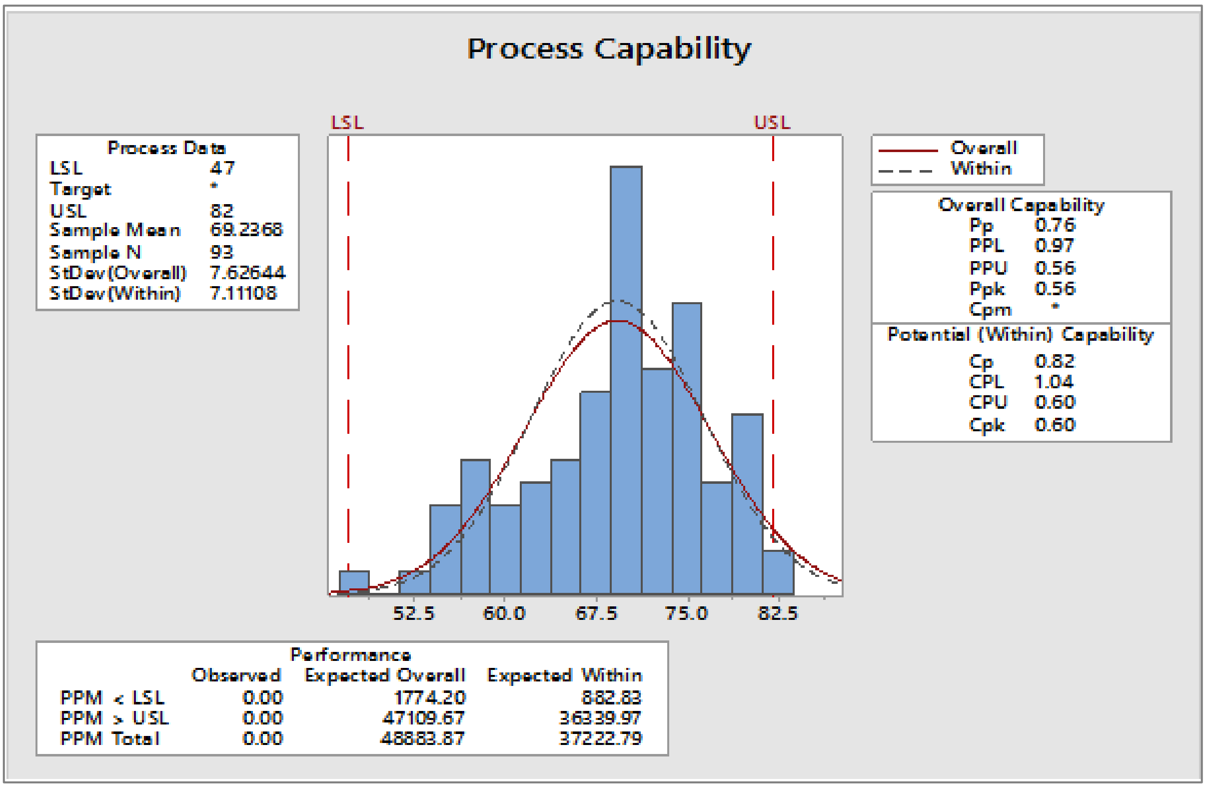

Figure 14. The process capability after the data transformation in Johnson's method

The values of the process capability of the kurtosis correction method are $P_{PK} = 0.63$, $C_{PK} = 0.67$, as in Figure 15.

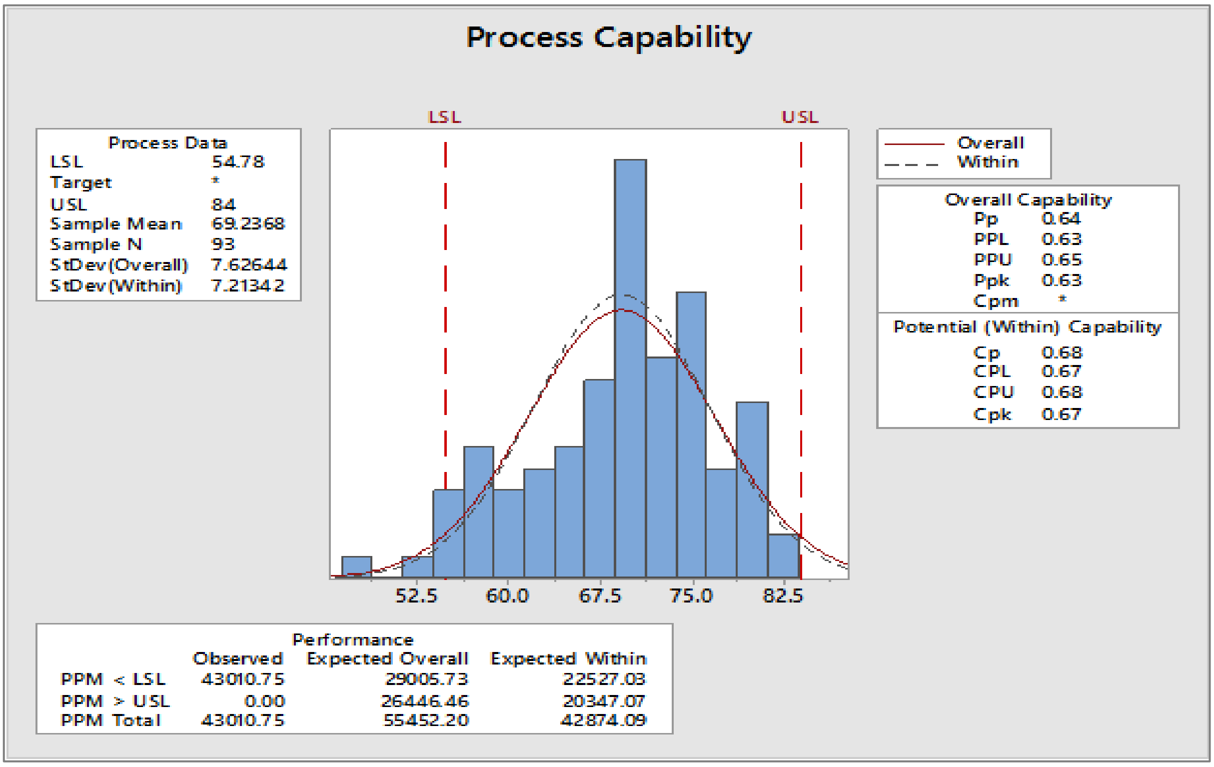

Figure 15. The process capability in kurtosis correction method

11. Results comparison

Table (1) shows the comparison among charts using Shewhart, Johnson and Kurtosis Correction as well as the values of process capability of these methods

Table 1. Lower and upper limits for mean chart and process capability of three methods

Method	UCL	LCL	C_{PK}	P_{PK}
Shewhart Chart	81.55	56.92	0.57	0.54
Johnson Chart	82	47	0.60	0.56
Kurtosis Correction Chart	84	54.78	0.67	0.63

12. Conclusions and recommendations

12.1 Conclusions

1. The use of Johnson's data transformation has avoided the process the false alarms in terms of stability and productivity
2. The process capability has improved after Johnson transformation and the use of the kurtosis correction chart.
3. The limits of quality control have expanded and all points within the limits as in the humidity rates in Kirkuk city for the period of study in the case of statistical control after the use of Johnson transformation.
4. When using the kurtosis correction chart, all points were within the control. The limits of statistical control have changed and the outgoing points are contained.
5. Applying the kurtosis correction method to construct the limits of R-chart to expand the width of limits. The process in both cases (with and without using KC) is under statistical control.
6. Calculating C_P, C_{PK} and P_{PK} shows that the process is not centered and not fully capable before and after using Johnson's transformation method.

12.2 Recommendations

1- We suggest dealing with probability limits for control charts rather than fixed limits because of the nature of data available to companies and factories.

2- Probability distributions and the number of parameters more than two parameters for high flexibility in the description of many cases can be used.

References

[1] D.C. Montgomery, "Introduction to Statistical Quality Control", sixth edition, John Wiley& Sons, United State of America, 2009.

[2] P.R. Tadikamalla and D.G. Popescu, "Kurtosis Correction Method for $\bar{X}$ and R Control Charts for Long-Tailed Symmetrical Distributions", *Naval Research Logistic*, vol.54, pp.371-383, 2007.

[3] C.P. Metlapalli , "Kurtosis Correction Method for Variable Control Charts – A Comparison in Laplace Distribution", *Pakistan Journal of Statistics and Operation Research*, vol.2, no.1, pp:51-54,2011.

[4] A. Goswami1 and N. H. Dutta , "Some Studies on Normal and Non-Normal Process Capability Indices" , *International Journal of Mathematics and Statistics Invention*, vol.1, no. 2, pp.31-40 , 2013.

[5] M. Aichouni , N.A. Messaoudene ,A. Al-Ghonamy , “On the Use of Johnson's Distribution in Quality and Process Improvement” , *International Journal of Economics and Statistics*, Vol. 2, pp.362-367, 2014.

[6] [6] M. Aichouni, A. I. AL-Ghonamy and L. Bachioua, “Control Charts for Non-Normal Data: Illustrative Example from the Construction Industry Business”, Conference Paper, *Mathematical and Computational Methods in Science and Engineering*, pp.71-76, April 2014.

[7] D. Aydin, “Estimation of the Lower and Upper Quantiles of Gumbel Distribution: An Application to Wind Speed Data”, *Applied Ecology and Environmental Research, vol.16*, no.1, pp.1-15, 2016.

[8] Lahcene Bachioua, “Process Capability Analysis with Non-Normal Data used in Quality Control”, pp.1-17, September 2018, https://www.researchgate.net/publication/327651445

[9] C. Walck, “Hand-Book on Statistical Distributions for Experimentalist”, University of Stockholm, 2007, http://www.stat.rice.edu/~dobelman/textfiles/DistributionsHandbook.pdf.

[10] B.S. Wang and C.S. Jyh, “Skewness and Kurtosis Correction for $\bar{X}$ and R Control Charts, Institute of Statistics”, National University of Kaohsiung, Taiwan 811 R.O.C, 2009.