

# Control Charts for Non-Normal Data: Illustrative Example from the Construction Industry Business

M. AICHOUNI <sup>(1,\*)</sup>, A. I. AL-GHONAMY <sup>(2)</sup> and L. BACHIOUA <sup>(3)</sup>

<sup>(1,2)</sup> BinLaden Research Chair on Quality and Productivity Improvement in the Construction Industry,  
College of Engineering

<sup>(3)</sup> Department of Mathematics, Preparatory Year College  
University of Hail  
P.O. Box. 2440, Hail  
SAUDI ARABIA

<sup>(\*)</sup> m.aichouni@uoh.edu.sa    <http://faculty.uoh.edu.sa/m.aichouni/>

*Abstract:* - Statistical Process Control charts widely used in industry and services by quality professionals require that the quality characteristic being monitored is normally distributed. If, in contrast, the distribution of this characteristic is not normal, any conclusion drawn from control charts on the stability of the process may be misleading and erroneous. In this paper, an alternative approach has been suggested that is based on the identification of the best distribution that would fit the data. Specifically, the Johnson distribution was used as a model to normalize real field data that showed departure from normality. Real field data from the construction industry was used as a case study to illustrate the proposed analysis.

*Key-Words:* - Statistical Process Control, Shewhart control charts, non-normal data, Johnson System of distributions

## 1 Introduction

Statistical Process Control is a process improvement methodology widely used by modern manufacturing and service organizations. This methodology is mainly based on the use of control charts and frequency distributions of process and quality characteristics data. Common and well established control charts include the Shewhart control chart ( $\bar{x}$ -R and  $\bar{x}$ -s charts), the cumulative sum control chart (CUSUM) and the exponentially weighted moving average control chart (EWMA). In process improvement strategies, these control charts are used to monitor product quality and detect special events occurring in the process that may cause out-of-control situations that would lead to an unstable and unpredictable process. Such processes deliver poor quality products to customers. Customers expect suppliers of products and services to provide proof of process control and process capability. Control charts help organizations management to continuously improve processes, by making them more stable and capable to produce high quality, to meet customer specifications, and to achieve business excellence.

Standards (Shewhart) control charts are designed on the assumption that the process being monitored produces a quality characteristic that can be approximated by a symmetrical normal distribution, when only the innate sources of variability are present in the system. The central limit theorem can be used to approximate distributions to the normal distribution provided that the samples being measured and monitored would be large enough. However, in many industrial situations, this can not be assured and the process output is not normally distributed and heavy tailed and skewed. Experience has showed that in some manufacturing processes, such as chemical processes parameters, cutting tool wear processes and some concrete production processes, the distribution are usually skewed. In this case, standard control charts based on normality assumptions can lead to erroneous conclusions regarding the stability and the capability of the process. Such wrong conclusions would cost manufacturing and service organization big financial losses and lost customers to competitors.

With the advents in statistical theories and computing facilities, this can be easily solved, by understanding of distributions that provide good model for most non-normal quality characteristics. Such an approach has been reported in the technical

literature (Farnaum, (1996), Chou et al., (1998), Sherill and Johnson (2009), and Derya and Canan (2012)). Derya and Canan (2012) developed standards control charts based using Weibull, Gamma and lognormal distributions. Sherill and Johnson (2009) showed the possibility to use exponential, Weibull and Lognormal distribution for transforming non-normal data for process control and process capability calculations. The objective of the present paper is to examine the use of the Johnson's family of distributions to model control charts that can be used for process improvement purposes. A real field case study is presented for ready mixed concrete production plants where process distribution showed a skewed non-normal distribution.

## 2 Johnson's Distributions in Quality Improvement

Statisticians and quality professionals are often faced with the problem of summarizing a set of data by means of a mathematical function which fits the data and allow obtaining estimates of percentiles. Frequently, statisticians and quality professionals usually have insufficient theoretical grounds for selecting a model like normal, gamma or extreme-value distributions for a "real world" data set (Johnson, 1949). Usually data are obtained and empirical methods are used to draw conclusions and make decisions on process and quality improvement in real business situations. The fitting of empirical distributions to data has a long history, and many different procedures have been advocated. The most common of these is the use of normal distribution. The central limit theorem leads one to expect this distribution to provide reasonable representation for many, but not all, physical phenomena (Gerald and Samuel, 1967).

Although models like gamma, log-normal and beta distributions do lead to a wide diversity of distribution shapes, they still do not provide the degree of generality that is frequently desirable. In 1949, Johnson derived a system of curves that has the flexibility of covering a wide variety of shapes. This system has the practical and theoretical advantages of being able to transform these curves to the normal distribution. The Johnson system is able to closely approximate many of the standard continuous distributions through one of the three functional forms and is thus highly flexible. The Johnson system provides one distribution 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  $S_U$ ,  $S_L$ , and  $S_B$ . This motivated us to use Johnson system for the analysis of micro array data (Johnson, Kotz & Balakrishnan 1994). This family of distributions, published by the statistician N.L. Johnson in 1949, is perhaps the most versatile choice. It is based on a transformation of the standard normal variable, and includes four forms:

1. Unbounded: the set of distributions that go to infinity in both the upper or lower tail.
2. Bounded: the set of distributions that have a fixed boundary on either the upper or lower tail, or both.
3. Log Normal: a border between the Unbounded and Bounded distribution forms.
4. 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 estimates of the percentiles of the fitted distribution to be calculated from the Normal distribution percentiles, for use in control limit calculations (on the Individual-X chart or the  $\bar{x}$ -R charts) 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 (Johnson, Kotz & Balakrishnan 1994). In this study, the authors applied the Johnson system, which includes the  $S_U$ ,  $S_L$ , and  $S_B$  distributions, as the Johnson system exhibits the key property of being able to accommodate all theoretically feasible skewness-kurtosis combinations (Figure 1).

The standard process capability analysis is one of many statistical process control widely used in manufacturing and services engineering. It is based on the assumption that process data are normally distributed. When this condition cannot be guaranteed, either capability indices should be computed based on distributions other than normal, or the data should be transformed so that it conforms better to the normal distribution (Farnum, 1996). Sherill and Johnson (2009), and many others showed that the use of Box-Cox and the Johnson transformations would help the quality professional to perform correct process analysis using both control charts for process stability and capability indices for process capability to meet customer specifications. In addition, it is worth mentioning here that in a recent study, Kilink et al (2012), showed compressive strength of concrete elements in buildings are best modeled using log-normal and the Johnson SB distributions.

### 3 Mathematical Formulation of the Johnson's Distributions

As stated earlier, when process data exhibit non-normal distribution, it is erroneous to draw standards control charts for process improvement and perform capability analysis. The practical solution is to transform the data and drive them towards normality, using common and well established probability distributions, such as Box-Cox, log-normal or the Johnson distribution. Such an approach has been used in the open literature. Basically the Johnson transformation computes an optimal transformation function from three flexible distribution families (SU, SB, and SL). This makes this transformation more powerful than other distribution (Sherill and Johnson, 2009).

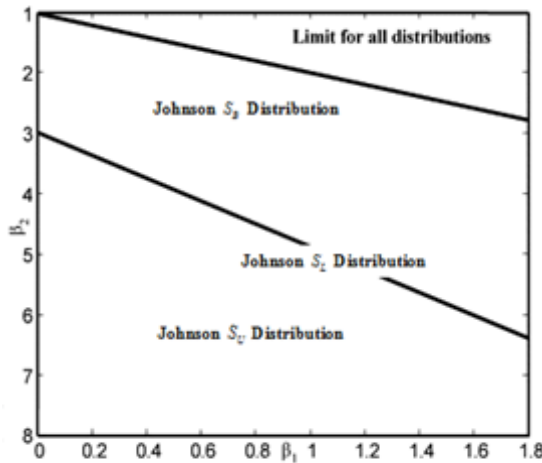


Figure (1): The Skewness and Kurtosis Plane for the Johnson Distributions

These translations transform any continuous random variable  $X$  into a standard normal variable  $Z$  using general form:

$$Z = a + bg\left(\frac{X - \mu}{\sigma}\right) \dots \dots \dots (1)$$

Where:  $a$  and  $b$  are shape parameters,  $\mu$  is a location parameter, and  $g(x)$  is a function defining the Johnson system of families, determined as:

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

As discussed in [Johnson, 1949], the above system

has the flexibility to match any feasible set of values for the mean, variance, skewness, and kurtosis coefficients. With this system, the skewness and kurtosis also uniquely identify the appropriate form for the (g) function.

#### 3.1 Johnson's Translation System:

Johnson proposed three normalizing transformations having the general form:

$$Z = \gamma + \sigma f\left(\frac{X - \mu}{\lambda}\right) \dots \dots \dots (1)$$

Where  $f(\cdot)$  denotes the transformation function,  $Z$  is a standard normal random variable  $\gamma$  and  $\sigma$  are shape parameters,  $\lambda$  is a scale parameter and  $\mu$  is a location parameter. Without loss of generality, it is assumed that  $\sigma > 0$  and  $\lambda > 0$ .

The first transformation proposed by Johnson defines the lognormal system of distributions denoted by  $S_L$ :

$$Z = \gamma + \sigma \ln\left(\frac{X - \mu}{\lambda}\right) = \gamma + \sigma \ln(X - \mu), \quad X > \mu, \dots (2)$$

The bounded system of distributions  $S_B$  is defined by:

$$Z = \gamma + \sigma \ln\left(\frac{X - \mu}{\mu + \lambda - X}\right) = \mu < X < \mu + \lambda, \dots \dots \dots (3)$$

$S_B$  curves cover bounded distributions. The distributions can be bounded on either lower end, or the upper end, or both. This family covers gamma distributions, beta distributions and many others.

The unbounded system of distributions  $S_U$  is defined by:

$$Z = \gamma + \sigma \ln\left[\left(\frac{X - \mu}{\lambda}\right) + \left\{\left(\frac{X - \mu}{\lambda}\right)^2 + 1\right\}^{1/2}\right] = \gamma + \sigma \sinh^{-1}\left(\frac{X - \mu}{\lambda}\right), \dots \dots \dots -\infty < X < +\infty \dots \dots \dots (4)$$

The  $S_U$  curves are unbounded and cover the  $t$  and normal distributions, among others.

#### 3.2 Johnson's Family of Distributions:

The Johnson family of distributions is made up of three distributions, Johnson  $S_U$ , Johnson  $S_B$  and lognormal. It covers any specified average, standard deviation, skewness and kurtosis. Together they form 4-parameter family distributions that

cover the entire skewness-kurtosis region other than the impossible region. The Johnson  $S_U$  distribution covers the area above the lognormal curve and the Johnson  $S_B$  covers the area below the normal curve. A family of distributions is several distributions combined so that they cover a well defined region in a skewness and kurtosis plot (lognormal family of distributions, negative lognormal and normal distributions,...). Readers can find detailed developments about the Johnson family of distributions in reference books (Gerald and Samuel, 1967).

This family of distributions is usually parameterized as a function of skewness and kurtosis. Skewness is a measure of non symmetry in the data, so for a normal distribution it takes the value of zero. Negative values for the skewness indicate that data are skewed left, and positive values indicate that data are skewed right. On the other hand, kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. The kurtosis for a normal distribution is 3.0. A kurtosis value larger than 3.0 indicates a “peaked” distribution and a kurtosis value less than 3.0 indicates a “flat” distribution. Thus, both can be seen as measures of shape of the distributions.

#### 4 Application of Johnson's System of Distributions for Real Field Data

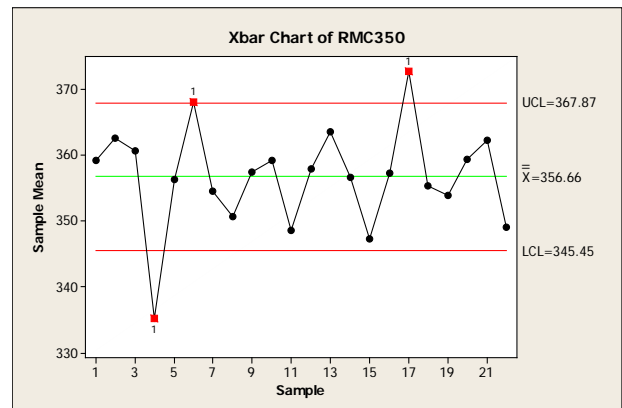
To illustrate the above analysis, real field data from the construction industry business was chosen as a case study. Data from Ready mixed concrete plants were gathered and analyzed using Minitab 16 statistical software. The observed quality characteristic was the compressive strength (kgf/cm<sup>2</sup>) of concrete as defined by international quality standards (ACI-214). The gathered data consisted of 22 samples of concrete with a nominal specification 350 kgf/cm<sup>2</sup>. The sampling process consists of a sample size of 3 spanning over a period of 30 days. These data are presented in table (1).

Initial analysis of the data of the concrete using standard  $\bar{x}$ -chart (figure 2) showed that the process is out of statistical control; this would mean the existence of special causes of variation affecting the process.

Table 1 – Data for compressive strength for Ready Mixed Concrete (Kgf/cm<sup>2</sup>)

Sample	Cylinder 1	Cylinder 2	Cylinder 3
1	353.8	363	360.6
2	357.8	358.7	370.9
3	365.2	360	356.6
4	340.4	335.2	330.1
5	359.6	358.1	351.2
6	368.1	366.7	369.3
7	357.9	355.0	350.6
8	337.8	352.6	361.6
9	359.1	349.2	363.7
10	361.1	358.2	358.3
11	358.3	345.7	341.7
12	357.3	359.2	356.9
13	352.6	363.1	374.6
14	360.8	356.2	352.7
15	347.5	339.8	354.3
16	358.2	359.5	353.9
17	375.2	372.5	370.2
18	357.5	359.5	348.9
19	343.2	355.8	362.4
20	362.1	356.6	359.1
21	365.2	362	359.4
22	361.3	346.8	339.0

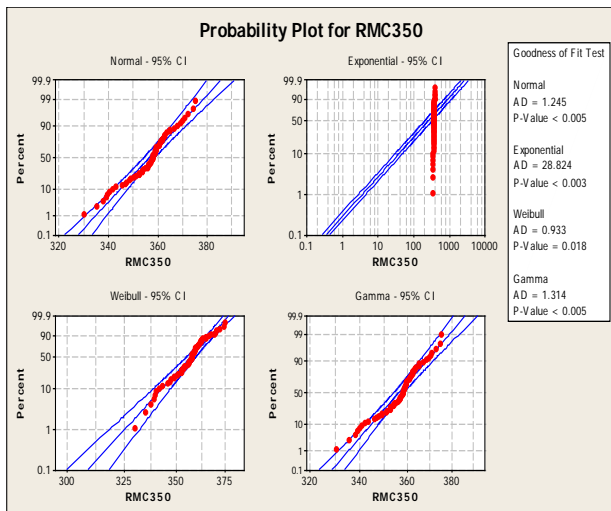
Figure 2 – Standards  $\bar{x}$  chart for the concrete compressive strength



The observation of an out of control situation shown from the  $\bar{x}$  control chart was drawn based on the assumption of normally distributed concrete data. Is this assumption correct? If not what would be the best distribution that fits these real field data. To answer this question, distributions identification was carried out for the data, and the outcome is presented in figure 3 as probability plots. From this figure, it can be seen that the compressive strength of concrete does fit neither the normal, nor the exponential, nor the Weibull, nor the lognormal

distributions. It is very obvious that the exponential distribution is a poor model for the concrete data. The Johnson distribution would be an alternative for the model (Kilink et al, 2012, Sherill and Johnson, 2009). The transformed data by the Johnson system are illustrated in figure 4, where it can be seen that this distribution shown as a mixture would be the best model of these concrete data. From this figure, it can be seen that within the interval percentile ranging from 1.054 to 98.94, would be the best fit of the data. Normality within this interval can be guaranteed. These correspond to the lower control limit and the upper control limit for the normalized data which are UCL=375.2 (kgf/cm<sup>2</sup>) and LCL=330.1 (kgf/cm<sup>2</sup>). These control limits will be used as the new control limits for the  $\bar{x}$  chart as shown in figure (5). It is clearly shown that the control chart with the new control limits indicate totally the opposite of the early conclusion drawn from the standard control chart. The process is shown to be in statistical control.

Figure 3 – Probability Plots for the Concrete Compressive Strength



### 5 Conclusions

Most statistical process control charts require that the quality characteristic being monitored is normally distributed. If, in contrast, the quality distribution of the quality characteristic of interest is not normal, the conclusions drawn from control charts on the stability of the process may be misleading and highly erroneous. In this paper, an alternative approach has been suggested that is based on the identification of the best distribution that would fit the data. Specifically, the Johnson distribution was used as a model to normalize real field data that showed departure from normality.

Real field data from the construction industry was used as a case study to illustrate the analysis. The assumption of normality when the data were not normally distributed led to conclude that the monitored process was out of statistical control, indicating that some special causes are present in the process, which would require some intervention from management on the process to get rid of the special cause of variation to occur again. This would certainly cost the organization some cost. However, when the data were transformed and brought to normality through Johnson transformations, and the new control limits calculated, the new control chart indicated no sign of special causes of variation.

Figure 4 – Probability Plots for the Johnson Transformed data of Concrete Strength

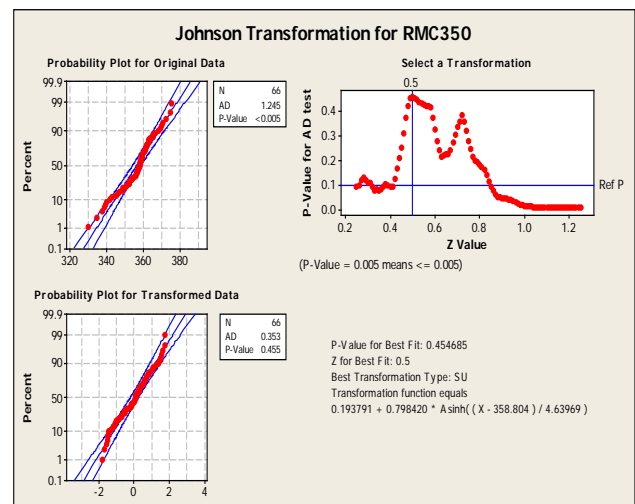
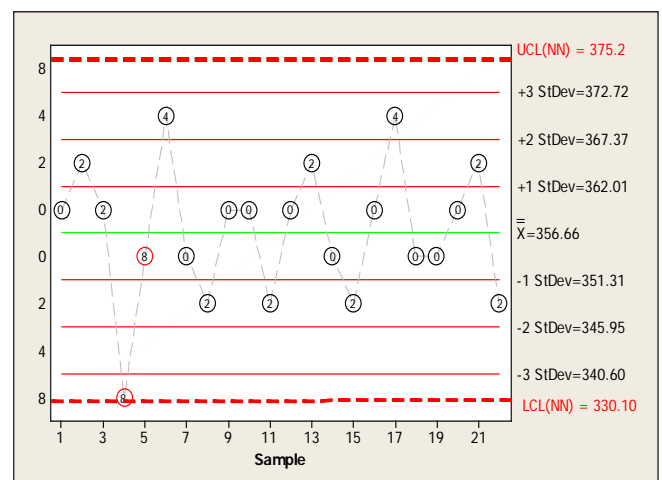


Figure 5 –  $\bar{x}$  -chart with the new Control Limits using Johnson transformations



## Acknowledgment

The present research work has been undertaken within the Bin Laden Research Chair on Quality and Productivity Improvement in the Construction Industry funded by the Saudi Bin Laden Constructions Group; this is gratefully acknowledged. The opinions and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the sponsoring organization.

### References:

- [1] Nicholas R. Farnum., Using Johnson Curves to Describe Non-Normal Process Data, *Quality Engineering*, Vol. 9, No. 2, December, 1996, 329-336.
- [2] Chou.Y, A.M. Polansky, and R.L. Mason, Transforming Non-normal Data to Normality in Statistical Process Control", *Journal of Quality Technology*, Vol. 30, April, 1998pp 133-141.
- [3] Sherill, R. W. and Johnson, L. A., Calculated Decisions, *Quality Progress*, Vol. 42 (1), 2009, pp. 30-35.
- [4] Derya, K and Canan, H., Control Charts for Skewed Distributions: Weibull, Gamma and Lognormal, *Metodoloski zvezki*, Vol. 9, N. 2, 2012 pp. 95-106.
- [5] Johnson, N.L., Systems of frequency curves generated by methods of translation, *Biometrika*, Vol. 36, 1949, 149-176.
- [6] Hahn J. Gerald and Shapiro S. Samuel, *Statistical models in Engineering*, John Wiley and Sons, 1967.
- [7] Johnson, N. L., Kotz, S., and Balakrishnan, N., *Continuous Univariate Distributions*, Second Edition, New York: John Wiley & Sons., 1994.
- [8] Kilink, K, Celik, A., Tuncan, M., Tuncan, A., Arslan, G. and Ario, O., Statistical distributions of in situ microcore concrete strength, *Construction & Building Materials*, Vol. 26 Issue 1, Jan2012, p393-403.
- [9] ACI Committee 214, *Evaluation of Strength Test Results of Concrete (ACI 214R-02)*, 2005.