3-31-2016

Analyzing Data Utilized in Process Control and Continuous Improvement

Jeffrey E. Jarrett

University of Rhode Island, jejarrett133@outlook.com

Follow this and additional works at: https://digitalcommons.uri.edu/cba_facpubs

Analyzing Data Utilized in Process Control and Continuous Improvement

Creative Commons License

Analyzing Data utilized in Process Control and Continuous Improvement

Jeffrey E. Jarrett*

College of Business Administration, University of Rhode Island, Kingston, RI 02881, USA

Abstract

Process control under the guidelines of reducing the costs of operations, production and inspection and produce high quality service and operations is examine to see if data from various parent populations can change the results and interpretations. With the purpose of achieving a minimal amount or error generation from the examination of control charts in univariate SPC, we compare data taken from three parent populations, i.e., Normal, Poisson and Exponential. Data and graphical analysis of Statistics permits one to visualize problems associated with where data comes from and whether it is satisfactory to use. We focus of univariate application but discuss multivariate application. This research will enable us to evaluate the conditions brought on by serial correlation and time series characteristics of models.

Keywords: Process control; Operations; Production costs; Data

Introduction

The powerful method to control product and service quality and to lower operations and production costs require controlling, monitoring and improving processes. The applications originally designed for industrial applications by the like of Shewhart called a process that operates under common causes of variation is in statistical control and those that operate under assignable causes of variation as being not in control (or out-of-control) [1]. Statistical process control (SPC) assumes that the quality characteristic(s) are either normal or in the case of multivariate process control are distributed according to the Hotelling T-square or in some applications a multivariate exponentially weighted moving average (MEWMA) or a multivariate cumulative sum (MCUSUM).

The assumptions of univariate SPC and multivariate SPC (MPC) is that the observations are serially uncorrelated or statistically independent over time. Previously, many (Alwan etc.) indicated that SPC and MPC applications consider the authenticated serial correlation in data for use in quality control applications. For example, many manufacturing chemical, health processes yield data containing serial correlation and in the case of multivariate processes cross-correlation of variables.

Literature Review

In the literature, two general approaches to deal with serial correlation in the process. One method for univariate processes is to fit autoregressive moving average (ARMA) models to the sampled data and then apply monitoring to the residuals. Although this approach is very promising, they may not be good enough to process mean shift. Harris and Ross [2] and Zhang and Hibler [3] may have inefficient detective power to find significant shifts in the process mean. Literature on this subject is rich and will be discussed later. A second approach is to apply changes to the MPC model to solve problems of serial correlation in set of time series data. This procedure lacks some ability to solve the problem of a time series characterized by serial correlation. Others studied this problem, including Vassilopoulos and Stamboulis [4], Alwan and Roberts [5], Harris and Ross [2], Montgomery and Mastrangelo [6] Maragah and Woodall [7], Wardell, Moskowitz and Plante [8], Superville and Adams [9], Lu and Reynolds [10], Schmid [11-14] West, Dellana and Jarrett [15] West and Jarrett [16] and Pan and Jarrett [17].

Residual Charts differ from control charts for means in that they need only one joint control limit which are based on independent and identically distributed case. Hence, residual charts have an advantage on the construction of control limits than adjusting the control limits.

SPC applications focus on the residuals from the univariate serially correlated case. A number of authors have extended this analysis to the multivariate case. For example, Pan and Jarrett [18] propose using multivariate processes (VAR models) of residuals of the models having serial correlation. In addition, Kalgonda and Kulkarni [19], Jarrett and Pan [20], Wang et al. [21] and Snoussi [22] proposed other solutions to this case using multivariate analyses of residuals.

In comparison, Harris and Ross, to that of Zhang and Hibler suggested improvement to the univariate case. Hence, we propose another study where we are aware that the normality assumption may be the source of the problem. That is, we wish to compare this assumption when we know the data are distributed by another probability distribution function (PDF). In the next section, we examine the literature on assumptions of SPC control charts. Do the data result from another PDF which may characterize time series data from relatively large samples? The answer to this question will enable quality control practitioners to better understand the importance of examining one's time series data to reduce the number of false signals from control charts and improve the quality of product or service under control.

Data Analysis and Experiment

To examine the influence of the underlying PDF on the univariate SPC process, we created by simulation data from a normal PDF, Poison PDF and Exponential PDF of size 500. The purpose is to allow experimentation with univariate SPC to see if signals differ and how the differences may be avoided. The analysis will point how to process controllers whether industrial, commercial or health related professions may handle the differences when known.

First, we generate three data sets by standard simulation techniques design in *Minitab®* software producing the time series data. [Each data set contains 500 time series observations assuming the parent population was either normal having a mean of 9, the variance will vary depending on the characteristics of the variances from populations that are Normal, Poisson and Exponential.]

*Corresponding author: Jeffrey E. Jarrett, College of Business Administration, University of Rhode Island, USA, Tel: +401 (874) 4169, Fax: 401 (874) 4312; E-mail: jejarrett133@gmail.com

Table 1 which follows the descript statistics of the three sets of time series data.

From Table 1, observe that the for the Normal distribution the mean and median are very close to each other as one may expect from having a normal parent population. In turn, for the Poison case, one may observe that the mean is 9.122 and standard deviation (2.9424) if squared (8.66) would also be very close to the to the value of the mean. This is the main characteristic of the Poisson model. Last, for the exponential model of the Parent Population the distribution is highly skewed towards greater values, that is, the data if graphed would have a very long tail to the right. All these characteristics indicate the time series are not alike but do conform the characteristics of the parent population from they were generated.

Another way in which one may examine the data simulated from each of the parent populations is to compare the Boxplots (sometimes referred to as a Box-Whisker Plot) of the time series data. Figure 1 contains the three Boxplots. Note the Normal has the greatest variation and the Poisson a smaller variation and on the same graph the Exponential appear to have very little variation.

All of this is expected when we deal with data from three very distinct parent populations. Another simple graphical analysis would be to observe the time series data.

Figures 2A-2C contains the time series plot of the foo observations of data having normal, poison and exponential parent populations. Observe the differences in the manner in which these data are distributed from left (the beginning) to the right (the ending). We know their means and variation measured previously by the sample standard deviation and noted in Table 1. We should note that measures of skewness for the distributions are 0.06 for Normal, 0.23 for the Poisson

and 0.60 for the Exponential distribution. Obvious, the skewness for the Exponential is much greater in comparison to the data for the other two time series. Much of this was observed in Table 1 ad Figure 1 before but the evidence is so exemplified in Figures 2A-2C.

Residual plots- Normal probability plot of the residuals

Minitab graphs plots the residuals versus their expected values when the distribution is normal. The residuals from the analysis should be normally distributed. In practice, large sample data with moderate departures from normality do not seriously affect the results.

The normal probability plot of the residuals should roughly follow a straight line. One uses this plot to look for the following patterns:

1. Non-normality if it not a straight line.

2. U-shaped curve in the tail areas, if skewness is observed.

3. An Outlier observation is one that is very distant from the remaining observations.

4. An additional or unidentified variable is affecting the data if there is a change in the trend of the data.

If the data has fewer than 50 observations, the plot may display curvature in the tails even if the residuals are normally distributed. As

Parent Population	Mean	Standard Deviation	Minimum	Q1	Median	Q3	Maximum
Normal	8.901	5.88	-8.667	5.214	8.911	12.816	26.369
Poisson	9.122	2.9424	1	7	9	11	19
Exponential	0.411	0.0251	0	0.0232	0.0378	0.0556	0.1099

Table 1: Descriptive statistics: Normal, poisson, exponential.
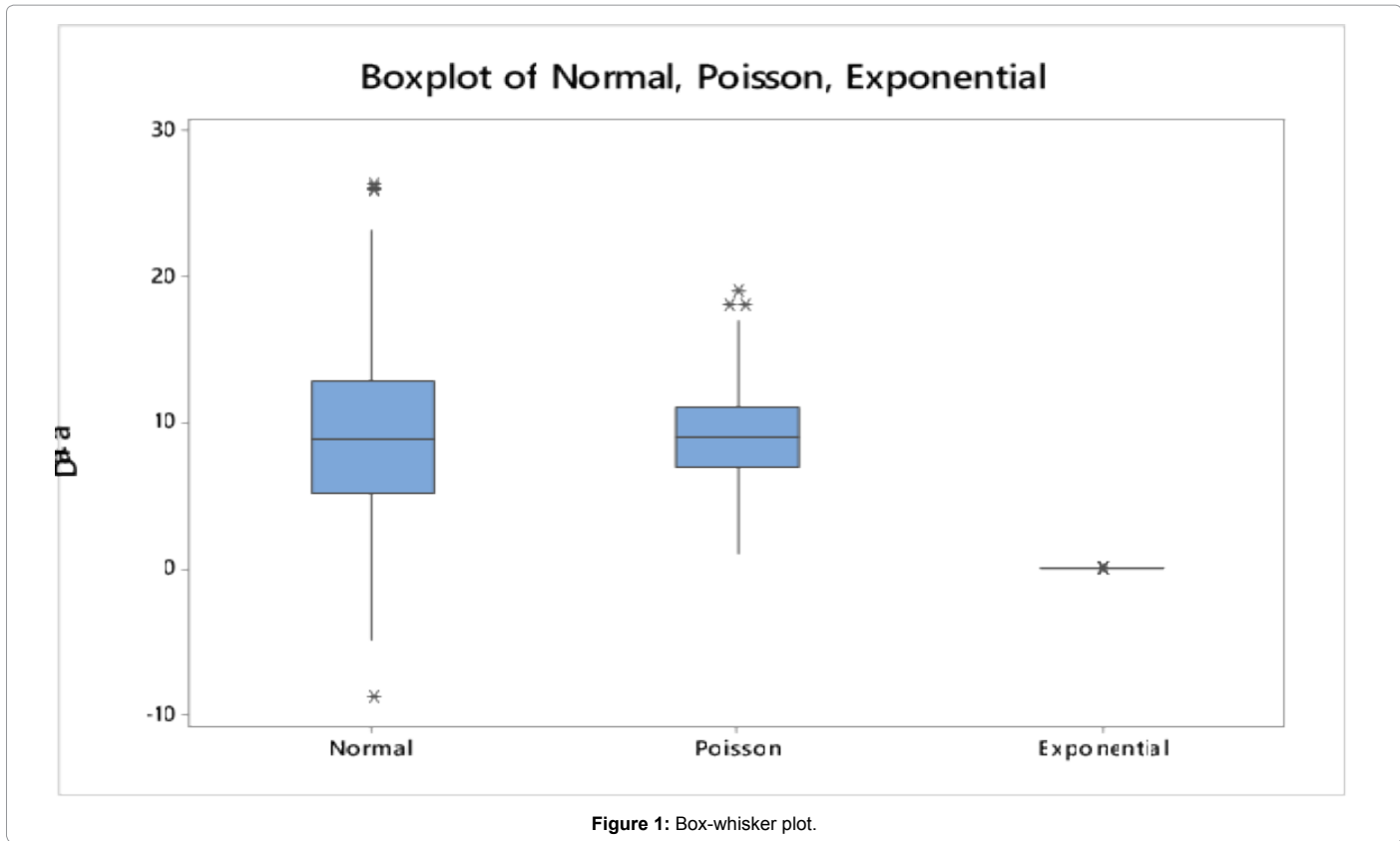
Figure 1: Box-whisker plot.

Figure 2(A): Time series plot of normal.

Figure 2(B): Time series plot of poisson.

Figure 2(C): Time series plot of exponential.
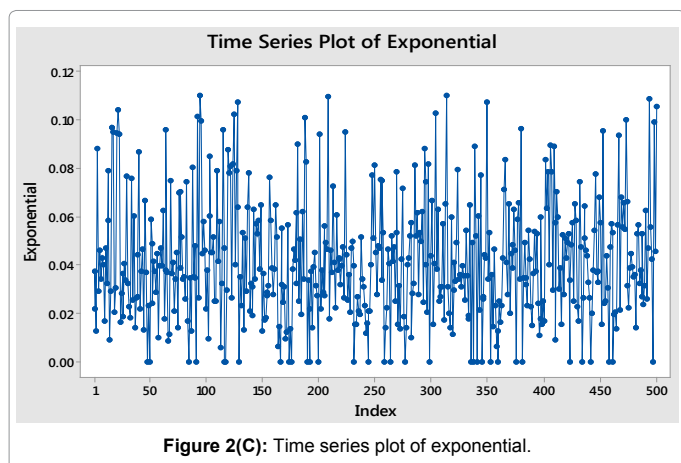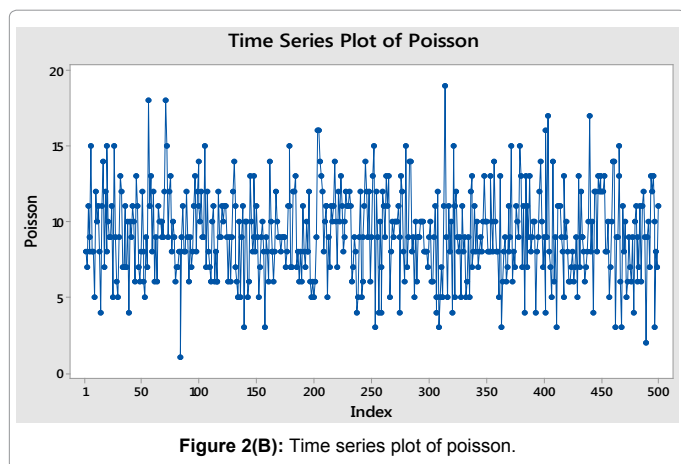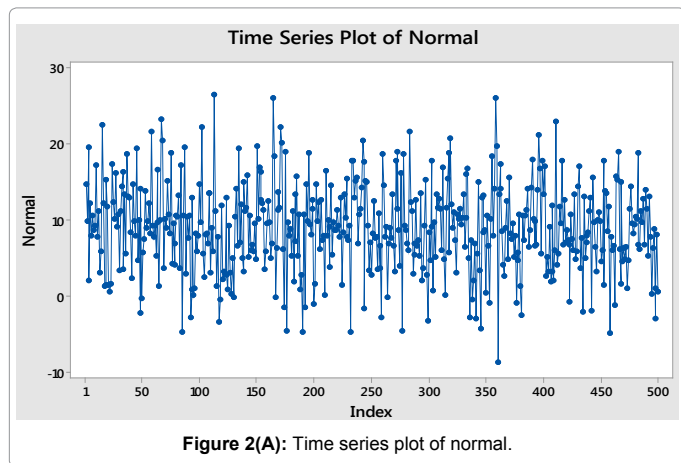
the number of observations decreases, the probability plot may show even greater variation and nonlinearity. Even though probability plots perform worse with small data sets, histograms are worse. Hence, we utilize the normal probability plot and goodness-of-fit tests to assess the normality of residuals in small data sets. Our data sets contain 500 observations per set, so any tests would probably have more than sufficient robustness.

We obtain in Figures 3A-3C the probability plots for the three times series and observe the patterns in the time series. In Figure 3A we

observe only two points outside the control limits at the very extremes of the graph. This would not be unusual since 95 percent control limits; one would expect 25 outside the confidence limits. Hence, the data would be near perfectly normally distributed.

By observing Figure 3B, we view a very different picture. The data points do indicate U-shaped pattern. This indicates that the data contains skewness and standard Shewhart control charts based on normal distribution of data would be inappropriate. Many points are outside the 95 percent control limits which also indicate that the data does not come from a normal parent population.

Observing the probability plot of the of the time series having an Exponential parent population provides evidence that these data are highly skewed and normal probability analysis are inappropriate. The sample observations and control limits on standard Shewhart control limits would be a fruitless undertaking with many false limits.

In the next section, we observe the use of Shewhart designed control charts.

Univariate SPC control charts

The univariate control chart that depicts the data in our sample is called the I-MR chart. This univariate chart is a is a combined chart that consists of an Individuals (I) chart, which plots the values of each individual observation, and provides a means to assess the process center and a moving range (MR) chart, which plots the range calculated from artificial subgroups created from successive observations, and
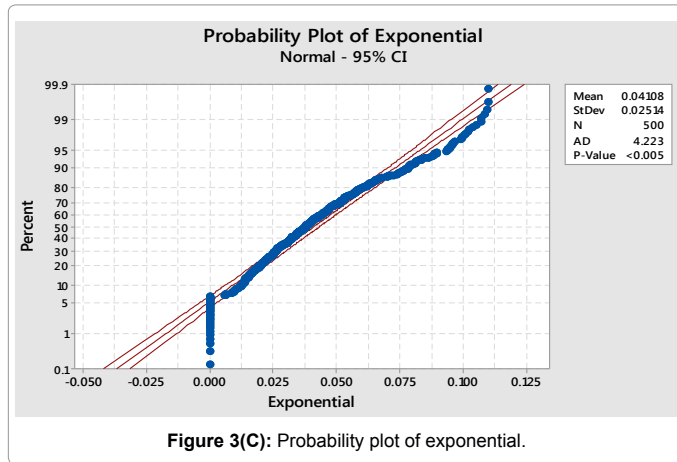

Figure 3(A): Probability plot of normal.

Figure 3(B): Probability plot of poisson.
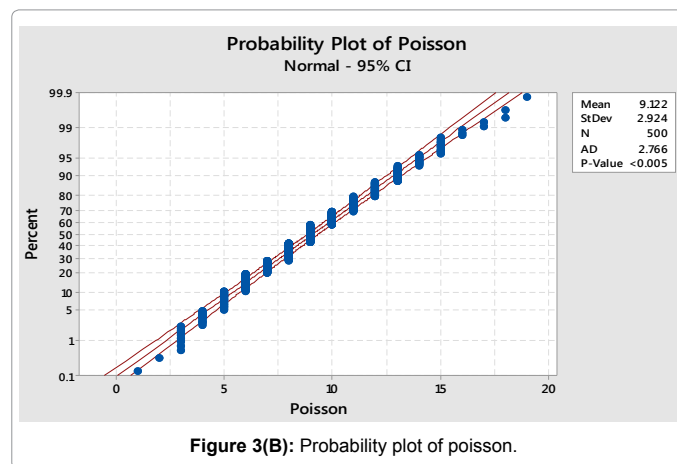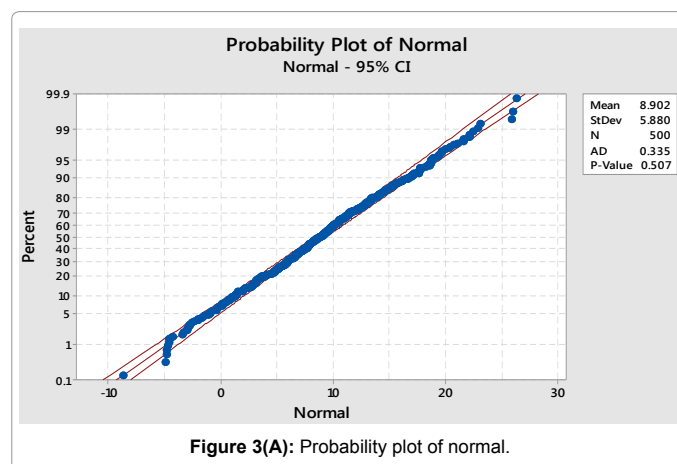
Figure 3(C): Probability plot of exponential.

provides a means to calculate the moving range (a measure of variation) in the process. Using the I-MR chart to draw a combined control chart for assessing whether process center and variation are in control when your data are individual observations. An **in-control** process exhibits only random variation within the established control limits. On the other hand, an **out-of-control** process exhibits unusual variation, which may be due to the presence of special causes. The MR chart must be in control before you can interpret the **I**-chart. This is because the **I**-chart control limits are calculated considering both process variation and *mean*.

When the MR chart is out of control, the control limits on the **I**-chart may be inaccurate and may not correctly signal an out-of-control condition. In this case, the lack of control will be due to unstable variation rather than actual changes in the process mean. Similarly, when the MR chart is in control, you can be sure that an out-of-control **I**- chart is due to changes in the process mean. The results follow for these charts for data from Normal parent population.

Observe in Figure 4A there were no observations out-of-control for the I-Chart, but there were a number of points out of control in the MR-Chart indicating that there may be inaccurate limits in the I-Chart. However, since there are only few such observations, the number of observations would in turn be small. Figure 4B has three observations out-of-control for the I-Chart and five observations out-of-control for the MR Chart. Hence there probably is some inaccuracy in the control limits and more observations would be out-of-control in the MR-Chart. Figure 4C results indicate that there may be a much large number of out-of-controls for the Poisson data then for the Normal data. Finally, the exponential charts show zero observations out-of-control for the I-Chart, but five points out of control in the MR-Chart. Again, the I-Chart limits are inaccurate. Many of the observations in the I-Chart for these data that are close to the limits should be out-of-control if the limits were correctly calculated. False signals are bad because they indicate that incorrect decisions will occur when process managers make decision and incur Type I and Type II errors in total quality management. Adjustments in decision making and process control for managers to make correct decisions. Incorrect interpretation of the data analysis and decision analytics produce unsavory conditions for decision making.

Conclusions and Further Study

This study aims at understanding some pitfall in univariate SPC as applied in industrial and engineering applications as well as monitoring

service application and in the field of health care and prevention, the data analysis from hospital acute care treatments and diagnosis of health care prevention and treatments. All of these analyses are utilized in these applications to achieve quality control and improvement activities often referred to as statistical process control by Deming [23,24].

Our research entailed an examination of some of the limiting assumptions of SPC and use of control charts to produce quality control and improvement. We reviewed literature on control charts both univariate and multivariate and focused on data from three sources. Stated differently, we generated samples of 500 from three sources in groups of 1 from a normally distributed parent population, a Poisson distributed parent population and an exponentially distributed parent population. In turn, we examined the descriptive statistics of three samples of time series data for its properties, the boxplots of the three time series to understand their secondary characteristics of asymmetry, variance and range. The three data sets were different in all these characteristics. The figures indicated the differences visually from normal probability plots and time series plot of each data set.
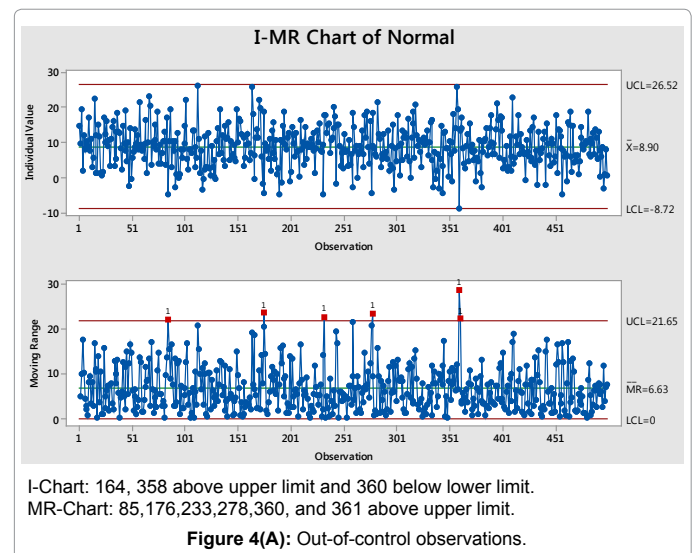

I-Chart: 164, 358 above upper limit and 360 below lower limit.
MR-Chart: 85,176,233,278,360, and 361 above upper limit.
Figure 4(A): Out-of-control observations.

I-Chart: 56, 71 and 314 are above upper limit and none below lower limit.
MR-Chart: 56, 253, 321, and 462 above upper limit.
Figure 4(B): Out-of-control observations.

I-Chart: None are above upper limit or below lower limit.
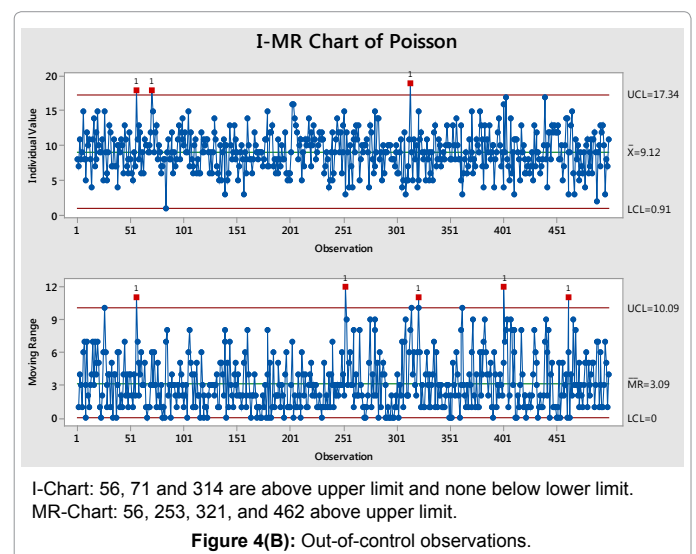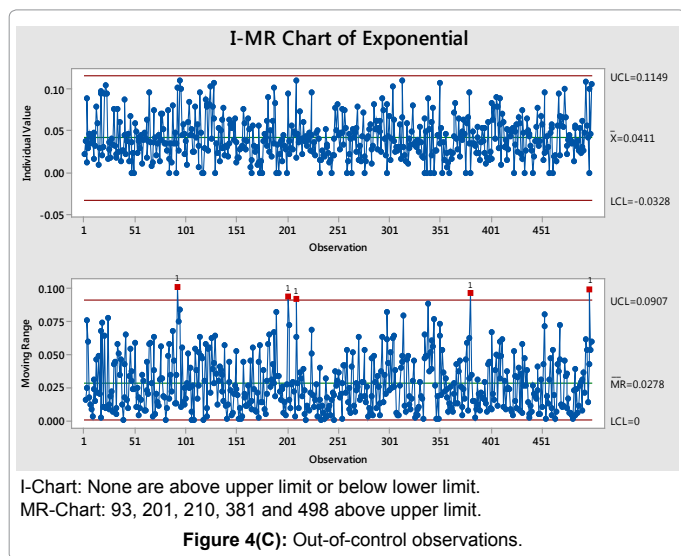MR-Chart: 93, 201, 210, 381 and 498 above upper limit.

Figure 4(C): Out-of-control observations.

Last, we analyzed the results after control charts based on + and – "three sigma" limits. The combinations of I_MR Control Charts indicated the fundamental difference in the results. The probability distribution of the parent populations produces varying results. Investigations into the characteristics of the parent population is necessary to insure that practitioners of SPC fully understand what it is to incur Type I and II errors.

In the future, this research will enable us to evaluate the conditions brought on by serial correlation and time series characteristics of models. As noted earlier by Alwan and later by West, Delana and Jarrett we must study the effects of autoregressive moving average models generating time series utilized in SPC. Finally, Pan and Jarrett [25,26] proposed using an operations research method called "Golden Ratio" to find an optimal solution in SPC.

References

1. Shewhart WA (1931) Economic control of quality of manufactured product. D. Van Nostrand Company, New York.

2. Harris TJ, Ross WH (1991) Statistical process procedures for correlated observations. Canadian Journal of Chemical Engineering 69: 48- 57.

3. Zhang J, Hibler WD (1997) On an efficient numerical method for modeling sea ice dynamics, Journal geophysical research 102: 8691- 8702.

4. Vassilopoulos AV, Stamboulis AP (1978) Modification of control charts limits in the presence of data correlation. Journal of Quality Technology 10: 20-30.

5. Alwan BM, Roberts HV (1988) Time Series Modeling for Detecting Level Shifts of Auto correlated Processes. Journal of Business and Economics Statistics 6: 87- 95.

6. Montgomery DC, Mastrangelo CM (1991) Some statistical process control methods for auto correlated data. Journal of Quality Technology 23: 179- 193.

7. Maragah H, Woodall WH (1992) The effect of autocorrelation on the retrospective. Journal of Statistical Computation and Simulation 40: 29- 42.

8. Wardell D, Moskowitz H, Plante RD (1994) Using run- length distributions of control charts to detect False alarms. Production and operations management 3: 219-239.

9. Superville CR, Adams BM (1994) An evaluation of forecast-based quality control schemes. Communications in Statistics - Simulation and Computation 23: 645-661.

10. Lu CW, Reynolds MR (1999) Control Charts for Monitoring the Mean and Variance of Auto correlated Processes. Journal of Quality Technology 31: 259-274.

11. Schmid W (1994) On EWMA Charts for Time Series. Proceedings and the American Society for Quality of 5th International Workshop on Intelligent Statistical Quality Control 161-184.

12. Schmid W (1997) On EWMA Charts for Time Series. Frontiers of Statistical Quality Control 5: 115- 137.

13. Schmid W (1997) CUSUM Control Schemes for Gaussian Processes. Statistical Papers 38: 191- 217.

14. Schmid W, Schone A (1997) Some Properties of the EWMA Control Chart in the Presence of Auto correlation. Anna Statist 25: 1277- 1283.

15. West D, Delana S, Jarrett JE (2002) Transfer function modeling of processes with dynamic inputs. Journal of Quality Technology 34: 315-326.

16. West D, Jarrett J (2004) The impact of first order positive auto-regression on process control. International journal of business and economics 3: 29-37.

17. Pan X, Jarrett JE (2008) Estimating σ and Improving Control Limits with RWAV for Normal and Non-normal Processes. Quality Engineering 21: 72- 78.

18. Pan X, Jarrett JE (2004) Applying State Space into SPC: Monitoring Multivariate Time Series. Appl Stat

19. Kalgonda AA, Kulkarni SR (2004) Multivariate Quality Control Chart for Auto correlated Processes. Journal of Applied Statistics 31: 317- 627.

20. Jarrett JE, Pan X (2007) Monitoring Variability and Multivariate Auto correlated Process. Appl Stat 34: 459- 469.

21. Wang F, Hwang Y, Quinn P, Wang X (2010) A Statistics-guided approach to precise characterization of nanowire morphology. ACS Nano 4: 885-860.

22. Snoussi A (2011) SPC for short-run multivariate autocorrelated processes. Appl Stat 38: 2303-2312.

23. Deming WE (1982) Out of the Crisis, MIT Center for Advanced Engineering Study: Cambridge, UK.

24. Deming WE (1993) The New Economics; for Industry, Government, and Education MIT Center for Advanced Engineering Study: Cambridge, UK.

25. Hastie T, Ribshirani R, Friedman J (2000) The Elements of Statistical Learning (2ndedn), Springer, Stanford, California.

26. Pan X, Jarrett JE (2013) Using Golden Ratio Search to Improve Paired Construction of Quality Control Charts. International Journal of Economics and Management Engineering 3: 107- 111.