

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# **Applications of Control Charts Arima for Autocorrelated Data**

---

Suzana Leitão Russo, Maria Emilia Camargo and  
Jonas Pedro Fabris

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/50990>

---

## **1. Introduction**

The traditional methodology of Statistical Quality Control (SEQ) is based on a fundamental supposition that the process of the data is independent statistically, however, the data not always are independent. When a process follows an adaptable model, or when the process is a deterministic function, the data will be autocorrelated.

Drawing the process of data is extremely valuable, however, under such circumstances, there isn't any scientific reason to use the traditional techniques of statistical control of quality, because it will induce erroneous conclusions and facilitate a safety absence that the process is under statistical control with flaw in the identification of systematic variation of the process.

Thus, the theme here proposed is to investigate the acting and the adaptation of the traditional use of the statistical control of process methods in no-stationary processes, and to discuss the use of time series methodologies to work with correlated observations.

## **2. Theoretical Review**

History of Quality Control is as old as the history of the industry itself. Before the Industrial Revolution, the quality was controlled by the vast experience of the artisans of the time, which guarantee product quality. The industrial system has suffered a new technical era, where the production process split complex operations into simple tasks that could be performed by workers with specific skills. Thus, the worker is no longer responsible for all product manufacturing, leaving the responsibility of only a part of it (Juran, 1993).

It is within this context that the inspection, which sought to separate the non-conforming items from the establishment of specifications and tolerances. A simple inspection did not

improve the quality of products, only provided information on the quality level of these and pick the items conform, those not complying. The constant concern with costs and productivity has led to the question: how to use information obtained through inspection to improve the quality of products?

The solution of this question led to the recognition that variability was a factor inherent in industrial processes and could be understood through the statistics and probability, noting that could be measurements made during the manufacturing process without having to wait for the completion of the production cycle.

In 1924, Dr. Walter. A. Shewhart of Bell Telephone Laboratories, developed a statistical graph to monitor and control the production process, being one of the tools of Statistical Quality Control. The purpose of these graphs was differentiate between aleatórias<sup>1</sup> causes unavoidable and causes a remarkable process. According to Shewhart (1931), if the random causes were present, one should not tamper with the process, if assignable causes are present, one should detect them and eliminate them. In other words, these graphics monitor the change or lack of instability in the process thus ensuring quality products.

Studies by Johnson and Basgshaw (1974) and Harris and Ross (1991) showed that the graphics Shewhart and cumulative sums (CUSUM) are sensitive to the presence of autocorrelated data (data that are not independent of each other over time), especially when the autocorrelation is extreme, ie tools are not suitable for the process control.

You will need to process the data first and then control them statistically. The presence of autocorrelation in the data leads to growth in the number of false alarms. Alwan and Roberts (1988) show that many false alarms (signals of special causes) may occur in the presence of moderate levels of autocorrelation, and the resulting measurement system, the dynamics of the process or both aspects, and conventional control charts are used without knowing the presence or absence of correlation, much effort can be spent in vain.

Many methods have been proposed to deal with statistical data autocorrelation. The interest in the area was stimulated by the work of Box and Jenkins, published in 1970 work entitled Time Series Analysis: Forecasting and Control, where it was presented among several quantitative methods, methodology used to analyze the behavior of the time series. The method of Box and Jenkins uses the concept of filter composed of three components: component autoregressive (AR), the integration filter (I) component and the moving average (MA).

The reason for monitoring residual processes is that they are independent and identically distributed with mean zero, when the process is controlled and remains independent of possible differences in the mean when the process gets out of control. Zhang (1998), the traditional graphics Shewhart, CUSUM graphics, the graphics may be applied to the EWMA waste, since the use of graphics residual control has the advantage that they can be applied to autocorrelated data, even if the data is nonstationary processes. When a graph of residual control is applied to a non stationary, it can only be concluded that the process has some deviation in the system because of a non stationary there is no constant average and / or constant variance.

### 3. Statistical Quality Control

The statistical quality control (SQC) is a technique of analyzing the process, setting standards, comparing performance, verify and study deviations, to seek and implement solutions, analyze the process again after the changes, seeking the best performance of machinery and / or persons (Montgomery, 1997).

Another definition is given by Triola (1999), which states that the SQC is a preventive method where the results are compared continuously through statistical data, identifying trends for significant changes, and eliminating or controlling these changes in order to reduce them more and more.

SPC charts are designed to detect shifts among natural fluctuations caused by chance noises. For example, the Shewhart chart utilizes the standard deviation (SD) statistic to measure the size of the in-control process variability. By graphically contrasting the observed deviations against a multiple (usually, triple) of SDs, the control chart is intended to identify unusual departures of the process from its normal state (controlled state).

Under certain assumptions, when the observed deviation from the mean exceeds three SDs, it is said that the process is out of control since there is only a probability of 0.0026 for the observation to fall outside the three SD limits given an unshifted mean chance the process mean is shifted. This Shewhart chart scheme is in effect a statistical hypothesis testing that reveals only whether the process is still in-control (Chen and Elsayed, 2000).

To better understand the technical statistical quality control, it is necessary to bear in mind that the quality of a product manufactured by a process is inevitably subject to variation, and which can be described in terms of two types concerned.

The *special cause* is a factor that generates variations that affect the process behavior in unpredictable ways, it is therefore possible to obtain a standard or a probability distribution.

The *common cause* is defined as a source of variation that affects all the individual values of a process. It results from various sources, without having any predominance over the other.

When these variations are significant in relation to the specifications, it runs the risk of having non-compliant products, ie products that do not meet specifications. The elimination of requiring special causes a local action, which can be made by people close to the process, for example, workers. Since the common causes require actions on the system of work that can only be taken by the administration, since the process is itself consistent, but still unable to meet specifications (Ramos, 2000).

According to Woodall et al (2004), Statistical Quality Control is a collection of tools that are essential in quality improvement activities.

#### *Descriptive Statistics*

According to Reid and Sanders (2002), descriptive statistics can be helpful in describing certain characteristics of a product and a process. The most important descriptive statistics are

measures of central tendency such as the mean, measures of variability such as the standard deviation and range, and measures of the distribution of data. We first review these descriptive statistics and then see how we can measure their changes.

*The mean:* To compute the mean we simply sum all the observations and divide by the total number of observations. The equation for computing the mean is:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

where:  $\bar{x}$  = mean;

$x_i$  = the observation  $i$ ,  $i = 1, 2, \dots, n$ ;

$n$  = number of observation.

*The range and standard deviation:* There are two measures that can be used to determine the amount of variation in the data. The first measure is the *range*, which is the difference between the largest and smallest observations in a set of data. Another measure of variation is the *standard deviation*. *Standard deviation* is a statistic that measures the amount of data dispersion around the mean. The equation for computing the standard deviation is (Reid and Sanders, 2002),:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

where:  $\sigma$  = standard deviation of a sample

$\bar{x}$  = the mean;

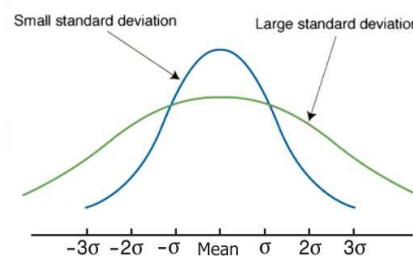
$x_i$  = the observation  $i$ ,  $i = 1, 2, \dots, n$ ;

$n$  = number of observation in the sample

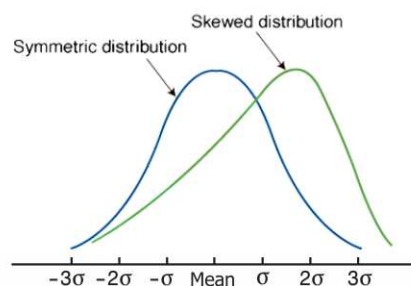
Small values of the range and standard deviation mean that the observations are closely clustered around the mean. Large values of the range and standard deviation mean that the observations are spread out around the mean.

#### *Distribution of the data*

A third descriptive statistic used to measure quality characteristics is the shape of the distribution of the observed data. When a distribution is symmetric, there are the same number of observations below and above the mean. This is what we commonly find when only normal variation is present in the data. When a disproportionate number of observations are either above or below the mean, we say that the data has a skewed distribution.



**Figure 1.** Normal distributions with varying standard deviations (adapted of Reid and Sanders, 2002).



**Figure 2.** Differences between symmetric and skewed distributions (adapted of Reid and Sanders, 2002).

### *Control Charts*

In any production process, no matter how well designed or carefully maintained it is, a certain amount of inherent or natural variability will always exist. Natural variability is the cumulative effect of many causes small, essentially unavoidable. When this variation is relatively small, generally considered an acceptable level of performance of the process. In the context of statistical quality control, this natural variability often called "a stable system of special causes" is said to be in statistical control. Control charts are used to examine whether or not the process is under control, ie, indicate only random causes are acting on this process. Synthesize a wide range of data using statistical methods to observe the variability within the process, based on sampling data. Can inform us at any given time as the process is behaving, if it is within prescribed limits, signaling thus the need to seek the cause of variation, but not showing us how to eliminate it (Ryan, 1989).

It was W. A. Shewhart (1931) which introduced control charts in 1924 with the intention to eliminate variations to distinguish them from the common causes and special causes. A control chart consists of three parallel lines: a line that reflects the average level of process operation, and two external lines called upper control limit (UCL) and lower control limit (LCL), calculated according to the standard deviation of a process variable (Shewhart, 1931).

There are several types of control charts, as the characteristic values or purpose, and we can divide them by attribute control charts and control charts for each variable.



### *Control charts by attributes*

A control chart for attributes, on the other hand, is used to monitor characteristics that have discrete values and can be counted. Often they can be evaluated with a simple yes or no decision (Reid and Sanders, 2002).

There are two broad categories of control charts for attributes: those who classify items into compliance or non-compliant, as is the case of graphs of the fraction of the number of faulty or defective, and those who consider the number (amount) of nonconformity existing graphics such as the number of defects in the sample or per unit.

According to Ramos (2000), the difficulties are:

- a) due to the small size of the batch, the approximation of binomial and Poisson by the normal distribution may no longer be valid, in which case the limits of control charts can not be determined by standard formulas;
- b) the probability distributions Binomial and Poisson may not adequately represent the studied phenomenon. This occurs when the parts are manufactured simultaneously (multiple mold cavities, for example), in which the incidence of defects or defects is not independent, statistically speaking.

### *Control charts for variable*

Control charts for variables monitor characteristics that can be measured and have a continuous scale, such as height, weight, volume, or width. When an item is inspected, the variable being monitored is measured and recorded (Reid and Sanders, 2002).

They may not be used for quality characteristics that cannot be measured because the control of the process requires monitoring of the mean and variability of measures. The graphics control variables used to data that can be measured or which undergo a continuous variation.

Some of the methods suitable for the construction of different control charts are the *Shewhart chart*, *Chart MOSUM - Moving Sum*, the *EWMA Chart - Weight Exponential Moving Average (Exponentially Weighted Moving Averages)* and *CUSUM Chart - Cumulative Sum (Cumulative Sum)*.

### *Shewhart control charts*

The first formal model of control chart was proposed by Dr. Walter A. Shewhart (1931), which now bears his name. Let  $X$  a statistical sample which measures a characteristic of the process used to control a production line. Suppose that  $\mu$  is the population mean of  $X$  and  $\sigma$  is the population standard deviation.

The following equations are used to describe the three parameters that characterize the Shewhart control charts (Montgomery, 1997)

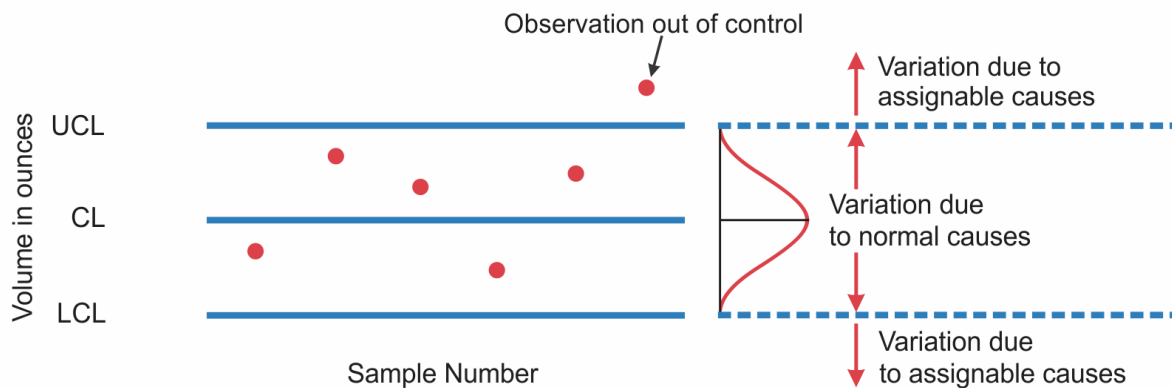
$$UCL = \mu + k\sigma_{\bar{x}} \quad (1)$$

$$CL = \mu \quad (2)$$

$$LCL = \mu - k\sigma_{\bar{x}} \quad (3)$$

where UCL is the upper control limit, CL is the center line or the average of the process, LCL is the lower control limit of the process, and  $k$  is the distance the control limits by the center line, which is expressed as a multiple of the  $\sigma$  standard deviation. The value of  $k$  is 3 most widely used.

The control graph is divided into zones (Figure 3). If a data point falls outside the control limits, we assume that the process is probably out of control and that an investigation is warranted to find and eliminate the cause or causes.



**Figure 3.** Control chart (adapted of Reid and Sanders, 2002).

### $\bar{X} - S$ Control Charts

A mean control chart is often referred to as an  $\bar{X}$  chart. It is used to monitor changes in the mean of a process. The  $\bar{X} - S$  control charts are generally preferred over the  $\bar{X} - R$  charts when  $n \geq 10$  or 12, since for larger samples the amplitude sampling  $R$  loses the efficiency to estimate  $\sigma$ , when compared to the sample standard deviation. The  $\bar{X}$  control charts is used in order to control the mean of the considered process. The two charts should be used simultaneously (Werkema, 1995).

The limits of the  $\bar{X} - S$  control charts are obtained in a similar manner, calculated under the assumption that the quality feature of interest ( $x$ ) has a normal distribution with ( $\mu$ ) mean and ( $\sigma$ ) standard deviation, ie, in abbreviated form (Panagiotidou and Nenes, 2009; Werkema, 1995).

$$x \sim N(\mu, \sigma)$$

However, satisfactory results are obtained even when this assumption is not true and distribution of  $x$  can only be considered approximately normal. In practice the  $\mu$  and  $\sigma$  parameters are unknown and must be estimated from sample data. The method of estimation of  $\mu$



and  $\sigma$  again involves taking  $m$  samples (subgroups rational) primary, each containing  $n$  observations of the quality characteristic considered.

Estimation of  $\mu$ :

The ( $\mu$ ) mean is estimate through the overall average of the sample ( $\bar{\bar{x}}$ ) as defined in the equation:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i \quad (4)$$

where  $\bar{x}_i$ ,  $i=1,2,\dots,m$  is the  $i$ -ésima sample mean:

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + \dots + x_{in}}{n} \quad (5)$$

Estimation of  $\sigma$  based on sample standard deviation:

The ( $\sigma$ ) standard deviation is estimate based in the ( $\bar{s}$ ) standard deviation mean as defined by:

$$\bar{s} = \frac{s_1 + s_2 + \dots + s_m}{m} = \frac{1}{m} \sum_{i=1}^m s_i \quad (6)$$

where  $s_i$ ,  $i=1,2,\dots,m$  is the  $i$ -ésima sample of the standard deviation:

$$s_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \quad (7)$$

It can be shown that the standard deviation sigma must be estimated by  $\hat{\sigma} = \frac{\bar{s}}{c_4}$ , where  $c_4$  is a correction factor, tabulated as a function of size  $n$  of each sample.

Expressions for calculating the limits of  $\bar{X} - S$  control charts:

$\bar{X}$  control charts -

$$UCL = \bar{\bar{x}} + 3\bar{s} / c_4 \sqrt{n} = \bar{\bar{x}} + A_3 \bar{s} \quad (8)$$

$$CL = \bar{\bar{x}} \quad (9)$$

$$LCL = \bar{\bar{x}} - 3\bar{s} / c_4 \sqrt{n} = \bar{\bar{x}} - A_3 \bar{s} \quad (10)$$

where  $A_3 = 3 / c_4 \sqrt{n}$  is a constant tabulated as a function of size  $n$  of each sample.

S control charts -

$$UCL = \bar{s} + 3\hat{\sigma}_s = B_4\bar{s} \quad (11)$$

$$CL = \bar{s} \quad (12)$$

$$LCL = \bar{s} - 3\hat{\sigma}_s = B_3\bar{s} \quad (13)$$

where  $\hat{\sigma}_s$  is a estimative of the standard deviation of the distribution of the  $S$  and  $B_3$  and  $B_4$  are constants tabulated in function of size  $n$  of each sample (Panagiotidou and Nenes, 2009; Werkema, 1995).

#### *Identification of Process in Control*

It is understood that the process is controlled to:

- a) all points on the chart are within the control limits;
- b) the arrangement of points within the control limits is random.

#### *Identification of Process out of Control*

According Montgomery (2009), various criteria may be simultaneously applied to a control graph for determining whether the process is under control. The basic criterion is one or more points outside the control limits. The additional criteria are sometimes used to increase the sensitivity of the control graphs when there is a small change in the process, so as to respond quickly to an assignable cause.

The Shewhart control charts have some rules sensitizers (Montgomery, 2009):

1. One or more points outside the control limits;
2. Two or three consecutive points outside the warning limits of 2-sigma;
3. Four or five consecutive points above of the limits of one-sigma;
4. A sequence of eight consecutive points of a same side of the center line;
5. Six points in a sequence is always increasing or decreasing;
6. Fifteen points in sequence in the area C (both above and below the center line);
7. Fourteen points alternately in sequence up or down;
8. Sequence of eight points on both sides of the center line CL;
9. A standard non-random data;
10. One or more points near a limit or control.

Typical patterns of behavior are non-random (Lourenço Filho, 1964):

- a) *Periodicity* - increases and decreases at regular intervals of time. The periodicity appears as one of the operating conditions of the process suffers periodic changes or when regular exchange of machines or operators.
- b) *Trend* - when the points are directed substantially upwards, or downwards. The general trend indicates a gradual deterioration of a critical process. This "decay" can be a tool wear and operator fatigue.
- c) *Shift* - changes in performance of the process. The cause of the change can be introduction of new machinery, new operators, new methods or even a quality program, which usually brings motivation and improves performance.

## 4. Time Series

The time series analysis aims to: investigate the mechanism generating the time series; to forecast future values of the series, to describe the behavior of the series; seek relevant periodicities in the data. A model that describes a series does not necessarily lead to a procedure (or formula) prediction. You need to specify a function-loss, beyond the model, to get the procedure. A function-loss, which is often used, is the mean square error, although on some occasions, other criteria or loss functions are more appropriate (Morettin and Toloi, 2006; Camargo and Russo, 2011).

### *Autocorrelation*

The autocorrelation is a measure of dependency between observations Same series separated by a given range named retardation.

Be a time series  $Y_t$ . The ratio between the covariance ( $Y_t, Y_{t-k}$ ) and variance ( $Y_t$ ) defines a autocorrelation coefficient simple ( $r_k$ ), while the sequence of  $r_k$  values is called autocorrelation function simple (AFS) (Camargo and Russo, 2006).

The graphical representation of this function is called correlogram. Formally, the autocorrelation coefficients simple between  $Y_t$  and their  $Y_{t-k}$  lagged values, are defined by:

$$r_k = \frac{\text{cov}(Y_t, Y_{t-k})}{\text{var}(Y_t)} = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (14)$$

We can see the existence of unit root if the values of the autocorrelation function begin near to unit and decline slowly and gradually as increases the distance (number of lags,  $k$ ) between the two sets of observations to which they concern, calling himself, not stationary and follows a random walk. If these coefficients decline rapidly as this distance increases, there is a series of characteristics of stationary (Morettin and Toloi, 2006; Russo et al. 2006).

### *Stationary Processes*

A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. A process is considered stationary if its statistical characteristics do not change with time.

Stationarity is a assumption in time series analysis. It means that the main statistical properties of the series remain unchanged over time. More precisely, a process  $\{Y_t\}$  is said to be completely stationary or strict sense stationary (abbreviated as SSS) if the process  $Y_t$  and  $Y_{t+n}$  have the same statistics for any  $n$ . So, the characteristics  $Y_{(t+n)}$  for all  $n$ , will be the same as  $Y_t$ .

### *Non-Stationary Processes*

A big reason for using a stationary data sequence instead of a non-stationary sequence is that non-stationary sequences, usually, are more complex and take more calculations when forecasting is applied to a data series (Beusekom, 2003).

Where a series submit over time variation in your parameters, so, we have a series non-stationary, which when submitted to differentiation process becomes stationary. If the time series is not stationary, we can often transform it to stationarity with one of the following way:

a) Difference the data, by create the new series

$$Y_t = X_t - X_{t-1}$$

The differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.

b) If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit.

c) For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

### *White noise*

In according of Cochrane (2005), The building block for our time series models is the white noise process, which I'll denote  $\varepsilon_t$ . In the least general case,

$$\varepsilon_t \sim i.d.d.N(0, \sigma_{\varepsilon_t}^2)$$

Notice three implications of this assumption:

1.  $E(\varepsilon_t) = E(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = E(\varepsilon_t \mid \text{all information at } t-1) = 0$
2.  $E(\varepsilon_t \varepsilon_{t-j}) = \text{cov}(\varepsilon_t \varepsilon_j) = 0$
3.  $\text{var}(\varepsilon_t) = \text{var}(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \text{var}(\varepsilon_t \mid \text{all information at } t-1) = \sigma^2$

The first and second properties are the absence of any serial correlation or predictability. The third property is conditional homoscedasticity or a constant conditional variance. Later, we will generalize the building block process. For example, we may assume property 2 and 3 without normality, in which case the  $\varepsilon_t$  need not be independent. We may also assume the first property only, in which case  $\varepsilon_t$  is a martingale difference sequence (Cochrane, 2005).

*Summary of time series models:*

*Autoregressive models - AR(p)*

The class of models purely autoregressive is defined by:

$$Y_t = \frac{a_t}{\varphi_p(B)} \quad (15)$$

where  $\varphi_p(B)$  has  $p$  coefficients. The AR ( $p$ ) assumes that the result is the weighted sum of its  $p$  past values than white noise.

The condition of stationarity of the AR ( $p$ ) states that all the  $p$  roots of the characteristic equation fall outside the unit circle (Russo, et al, 2006).

*Moving average models - MA(q)*

According to Russo, et al (2009), the class of moving averages models is defined by

$$Y_t = \theta_q(B)a_t \quad (16)$$

where  $\theta_q(B)$  has  $q$  coefficients. The models MA ( $q$ ) resulting from the linear combination of random shocks that occurred during the current and past periods.

The invertibility condition requires that all roots of the characteristic equation fall outside the unit circle.

*Autoregressive and moving average models - ARMA ( $p, q$ )*

The class of models, autoregressive-moving average is of type

$$Y_t = \frac{\theta_q(B)a_t}{\varphi_p(B)} \quad (17)$$

where  $\varphi_p(B)$  has  $p$  coefficients and  $\theta_q(B)$  has  $q$  coefficients. With a combination of models AR ( $p$ ) and MA ( $q$ ), it is expected that the models ARMA ( $p, q$ ) be models extremely parsimonious, using few coefficients to explain the same serie.

From the standpoint of adjustment, it is very important because you can adjust more quickly. The condition of stationary and invertibility of a ARMA ( $p, q$ ) require that all  $p$  roots of  $\varphi(B) = 0$  and all the  $q$  roots of  $\theta(B) = 0$  fall outside the unit circle (Russo, et al, 2009).

#### *Autoregressive Integrated Moving Averages Models - ARIMA ( $p, d, q$ )*

The class of autoregressive-integrated-moving-average models are defined by the equation,

$$Y_t = \frac{\theta_q(B)a_t}{\varphi_p(B)(1-B)^d} \quad (18)$$

to an integrator positive  $d$ . Made the differentiation of the series  $d$  times necessary to make it stationary, then the ARIMA( $p, d, q$ ) model can be adjusted through the ARMA( $p, q$ ) model (Russo, et al, 2009).

#### *Sazonal Model - SARIMA*

According to Fischer (1982), the appearance of some short-term cyclical behavior is called seasonality. For a full treatment about series of time, need to characterize and eliminate this cyclic function of time to become the condition of stationarity.

Seasonality means a tendency to repeat a certain behavior of the variable that occurs with some regularity in time. That is, are those series that have variations of a similar amount of time to another, characterized by showing high serial correlation between observations of the variable spaced by the period of seasonality, and, of course, the serial correlation between the next observations.

Similar to the process ARIMA ( $p, d, q$ ) this process develops the model in one of three basic forms of description of each value of  $Y_t$ , and applies the same procedures developed for a model where the seasonal component is not present. After establishing the value of the variable in period  $t+h$ , then applies the expectancy operator. Forecast errors, confidence intervals and updating are treated similarly to the ARIMA model (Fischer, 1982).

#### *Box-Jenkins Methodology*

This method for the prediction is based on the setting called tentative ARIMA models, has a flexible modeling methodology that forecasts are made from the current and past values of these series. Therefore, describing both the stationary behavior as the non-stationary zero. ARIMA models are able to describe the process of generating a variety of series for forecasters (corresponding to the filters) without taking into account the economic relations, for example that generated the series (Morretin and Toloi, 2006).

The determination of the best model for "Box and Jenkins" methodology following this steps (Leroy, 2006):



### *Identification*

Identification is the most critical phase of the "Box and Jenkins" methodology, it is possible that several researchers to identify different models for the same series, using different criteria of choice (ACF, PACF, Akaike, etc..). Typically, the models should be parsimonious. The study analyzes the ACF and PACF, and attempts to identify the model. The process seeks to determine the order of (p,d,q), based on the behavior of the Autocorrelation Functions (ACF) and Partial Autocorrelation (PACF), as well as their respective correlograms.

### *Estimation:*

After identifying the best model should then adjust and examine it. The adjusted models are compared using several criteria. One of the criteria is the of parsimony, in which it appears that the incorporation of coefficients additional improves the degree of adjustment (increases the  $R^2$  and reduces the sum of squared residuals) model, but you reduces the degrees of freedom. One of ways to improve the degree of adjustment of this model to time series data is to include lags additional in Cases AR (p), MA (q), ARMA (p, q) and ARIMA.

The inclusion of additional lags implies increasing the number of repressors, which leads to a reduction in the sum of squared residuals estimated. Currently, there are several criteria for selection of models that generate a trade-off between reductions in the sum of squared residuals and estimated a more parsimonious model.

Generally, when working with lagged variables are lost about the time series under study. Therefore, to compare alternative models (or competitors) should remain fixed number of information used for all models compared.

### *Checking:*

Aspiring to know the efficacy of the model found, takes place waste analysis. If the residuals are autocorrelated, then the dynamics of the series is not completely explained by the coefficients of the fitted model. It should be excluded from the process of choosing the model(s) with this feature.

An analysis of existence (or not) of serial autocorrelation of waste is made based on the functions of autocorrelation and partial autocorrelation of waste and their respective correlograms. It is noteworthy that, when estimating a model, it is desired that the error produced by it have characteristic "white noise" that is, this will be independent and identically distributed (i.i.d. condition).

### *Forecast:*

Predictions can be ex-ante, made to calculate future values of short-term variable in the study. Or, ex-post held to generate values within the sample period. The better these last, the more efficient the model estimated. We choose the best model throught the lower Mean Absolute Percentage Error (MAPE). It is a formal measure of the quality of forecasts ex-post. Therefore, the lower value of the MAPE is the best fit of forecasts of the model to time series data.

## 5. Methodology and Results

In this work we analyzed the Têxtil Oeste Ltda industry, whose Statistical Control of Processes implantation happened in 1999. Here, we limited to analyze the control charts for continuous variables as tools used for the control of the process. The conventional Shewhart control charts were used added of other appropriated models to transformations of autocorrelations data in data that are independent and usually distributed.

In thread's polypropylene process there are several outputs to consider critical. One of these outputs is the thread's resistance. In an effort to develop a control plan to assure quality of the appropriate surface, it was certain that the resistance has a main impact on surface quality of the thread. So, to verify the quality of the thread, it's resistance should be controlled.

At once, the data used in this study is the daily data of the thread's polypropylene resistances control.

These data are for the models identification and estimation and for the models predictive capacity analysis. Before control charts be applied, three fundamental assumptions must be met: The process is under control; the data are normally distributed; and the observations are independent.

Montgomey (2009) considers that the points out of control are stipulated reasonably well for the controls charts of Shewhart when the normality assumption is somewhat violated, but when observations aren't independent, control charts yield deceiving results. Many processes don't produce independent observations. Alwan (1991) describes a method for control charting with autocorrelated data. The method involves fitting a time series curve and control charting the residuals.

It was made a study that helped to verify where it is the largest instability of the process, so that we can make a better control of the system. It is suspected that the daily thread's resistance data aren't independent, and the result of a plot of these data, as showing in Figure 4, supports this belief.

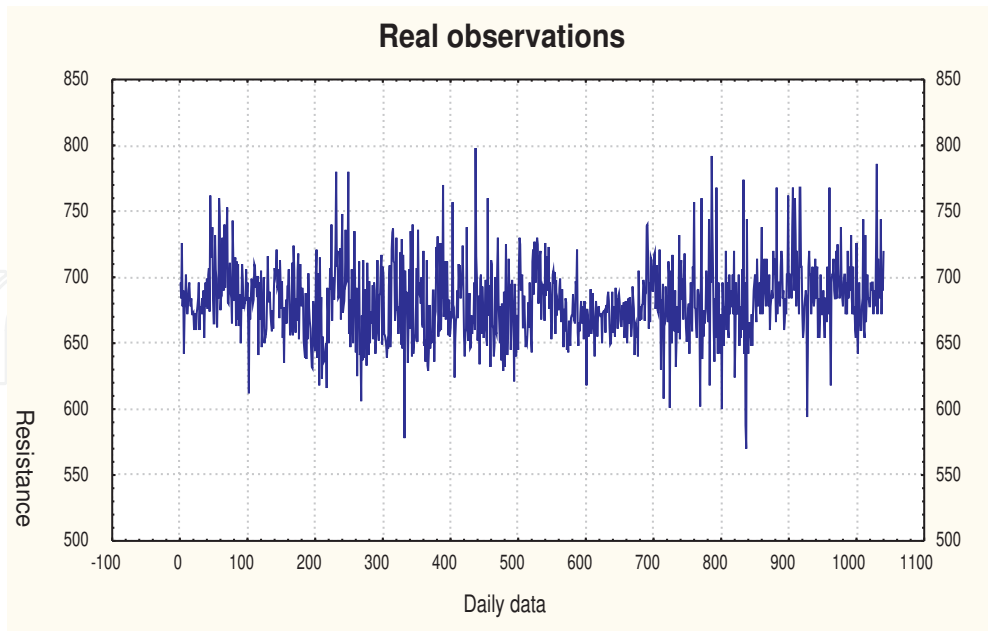
The problem is to implement statistical control for a process that has autocorrelation (Dobson, 1995). The Figure 4 shows us the great data variability. Calculations were performed to confirm the autocorrelation's suspected.

Calculations were done to confirm the suspected autocorrelation. The autocorrelation coefficient for thread's resistance is defined as

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad k=0,1,2, \dots$$

where  $k$  = time periods ahead

$n$  = total number of data



**Figure 4.** Daily data.

The standard error at lag  $k$ , if  $k=1$  is  $Se_k = \sqrt{\frac{1}{n}}$ , and the standard error at lag  $k$ , if  $k > 1$  is

$$Se_k = \sqrt{\frac{1}{n} \left( 1 + 2 \sum_{i=1}^{k-1} r_i^2 \right)}$$

The autocorrelation coefficient for  $k=1$  and  $k=2$  are:

$$r_1 = \frac{160680,24}{971034} = 0,16$$

and

$$r_2 = \frac{100113,10}{971034} = 0,10$$

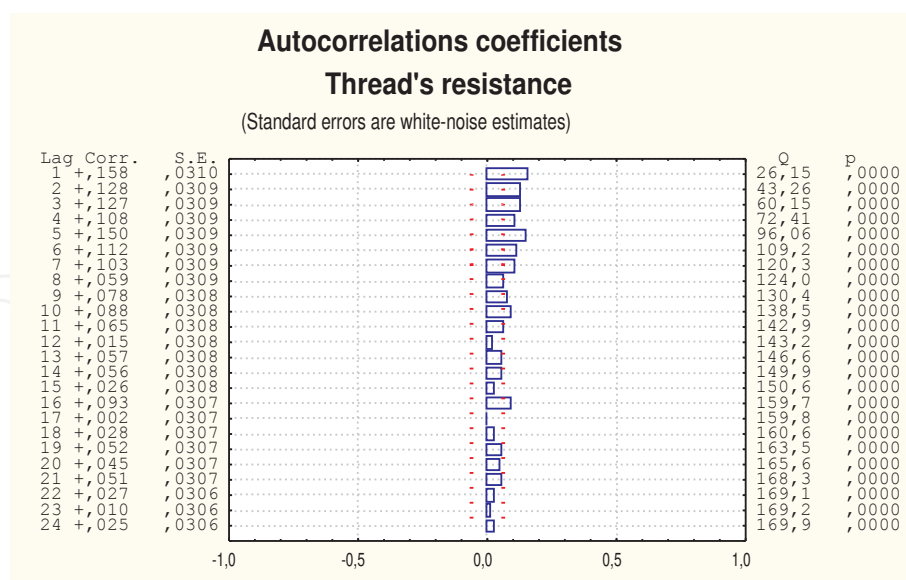
The standard error for  $k=1$  and  $k=2$  are:

$$Se_1 = \sqrt{\frac{1}{1041}} = 0,0310$$

and

$$Se_2 = \sqrt{\frac{1}{1041} [1 + 2(0,16)^2]} = 0,0317$$

The Figure 5 shows the autocorrelation coefficients and 2 standard errors for these coefficients for up to 24 lags, and the Figure 6 shows the partial autocorrelation coefficients and the 2 standard errors for these coefficients for up to 24 lags.

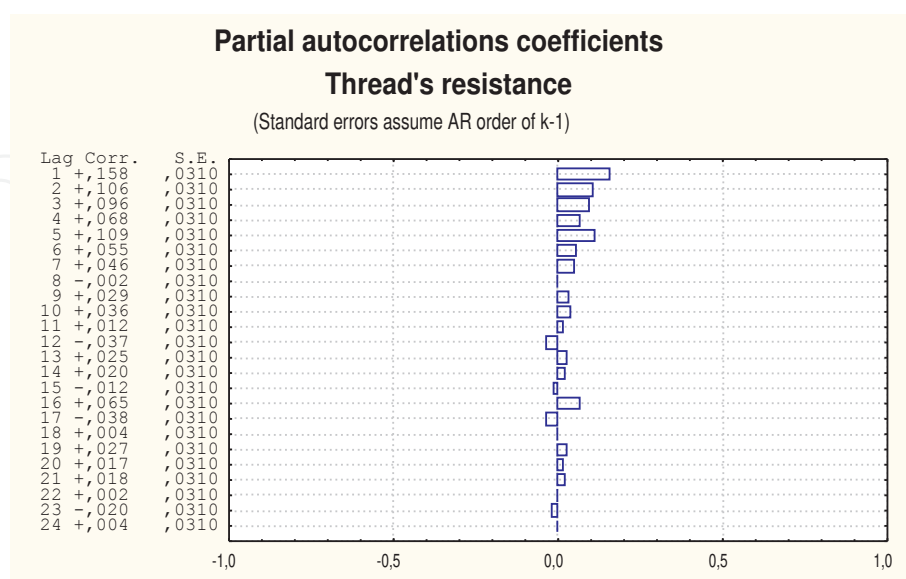


**Figure 5.** Autocorrelations coefficients.

As we can see, the data are highly autocorrelated. The autocorrelation coefficients for lags 1-7 exceed two the standard errors. Before a control charts can be used, these data must be transformed to guarantee the independence of each observation.

To find an independent, normally distributed data set, Montgomery (2009) recommends to model the structure and to develop the control charting of the residuals directly.

The Box & Jenkins's methodology was used, to determine the parameters of the model (Box, Jenkins and Reinsel, 2008).



**Figure 6.** Partial autocorrelations coefficients.

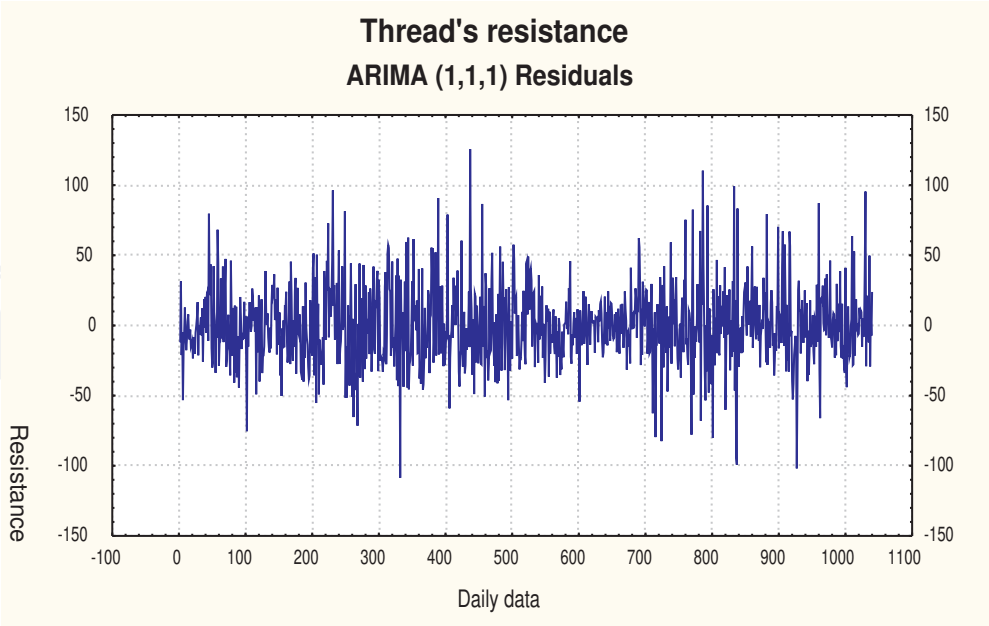


Figure 7. Residuals of thread’s resistance.

The Figures 8 and 9 show that the obtained model is adapted to the resistance data. The autocorrelation coefficients were calculated for the transformed data defined for the model ARIMA (1,1,1), to validate that the autocorrelation has been removed from the data.

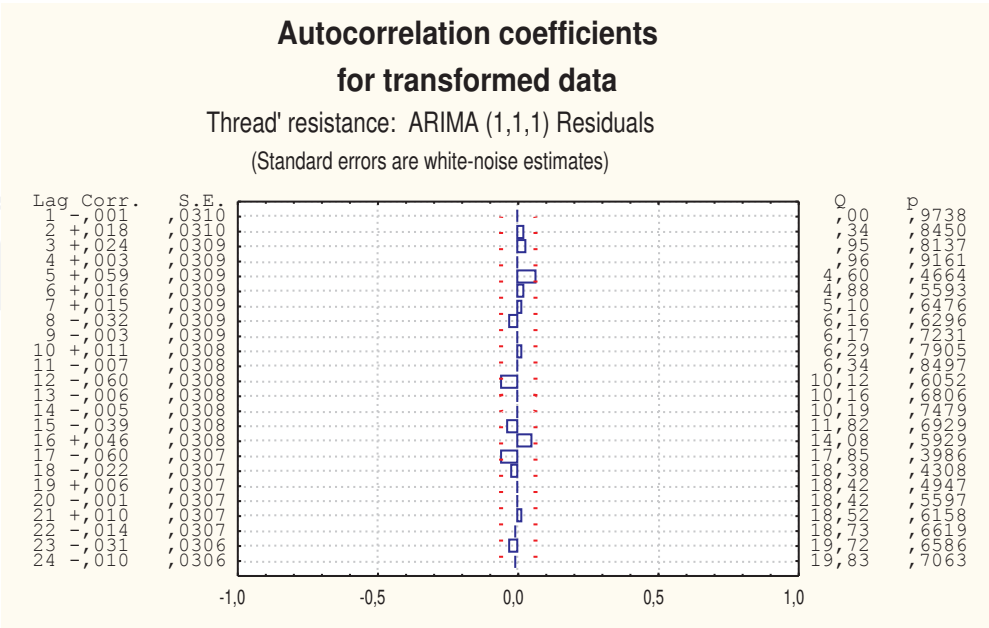
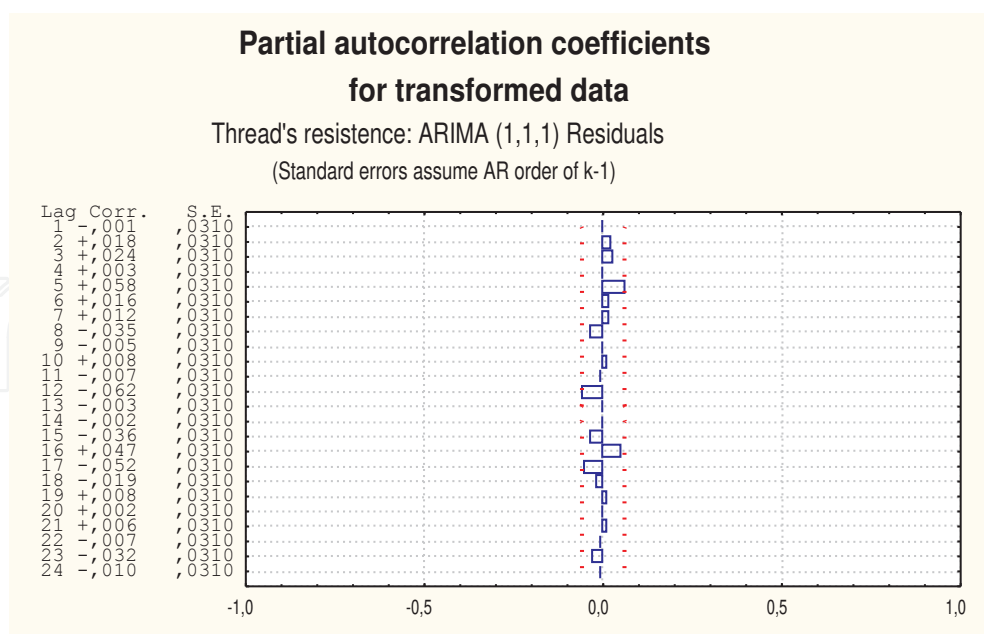


Figure 8. Autocorrelation coefficients.



**Figure 9.** Partial autocorrelations coefficients for transformed data.

Figure 8 and 9 show that the defined data is independent from an observation to another observation. And the table 1 shows the Chi-square test to verify the normality.

For two degrees of freedom,  $\chi^2_{0,05}=5,991$ . As the calculation qui-square value was  $\chi^2=5,0415$ , and it is smaller than the critical value, the data are considered as normal. Now the behavior of the productive process can be verified.

The Chi-square test was executed, to verify the normality:

Lower Limit	Upper Limit	Obs	Exp	(Obs-Exp) <sup>2</sup> / Exp
649,0708	743,3123	744	696,1916	3,2831
648,9230	743,4601	672	696,1916	0,8406
648,7757	743,6074	702	696,1916	0,0485
648,6288	743,7543	690	696,1916	0,0551
648,4823	743,9008	720	696,1916	0,8142
Total				5,0415

**Table 1.** Test the Chi-square.

Figure 10 shows  $(\bar{X})$  and  $(S)$  charts for the real data.



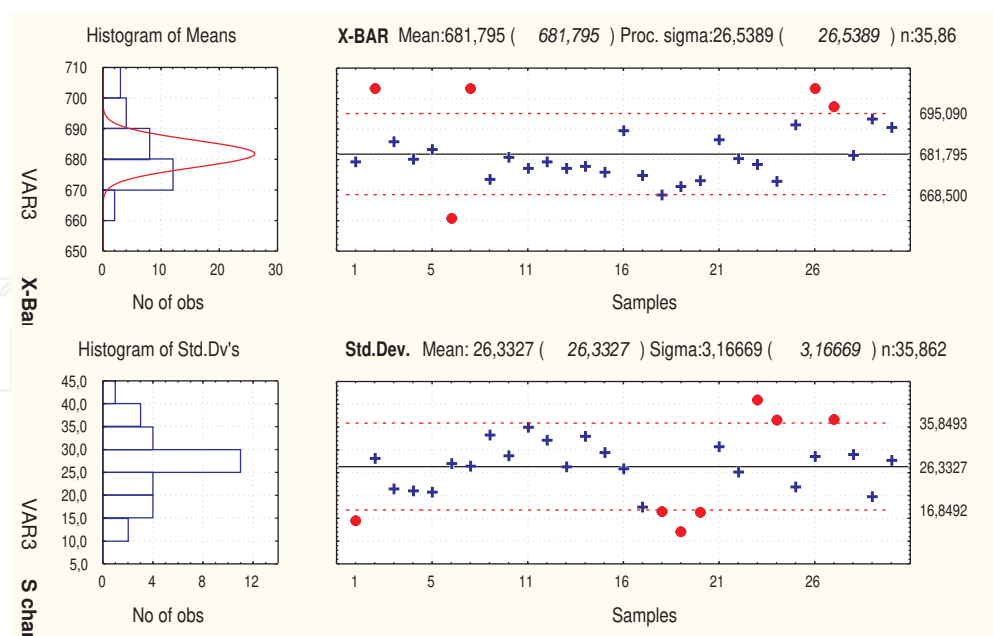


Figure 10.  $\bar{X}$  and  $S$  charts for real data.

Through the illustration 10 we can notice the sequence of observations and limits of the traditional Shewhart charts, where several points were out of the control limits, indicating that the process is apparently out of control. In fact, before the transformation of the data, we found the data really correlated what took us to model for a process ARIMA (Wardell, Moskowitz and Plante, 1994).

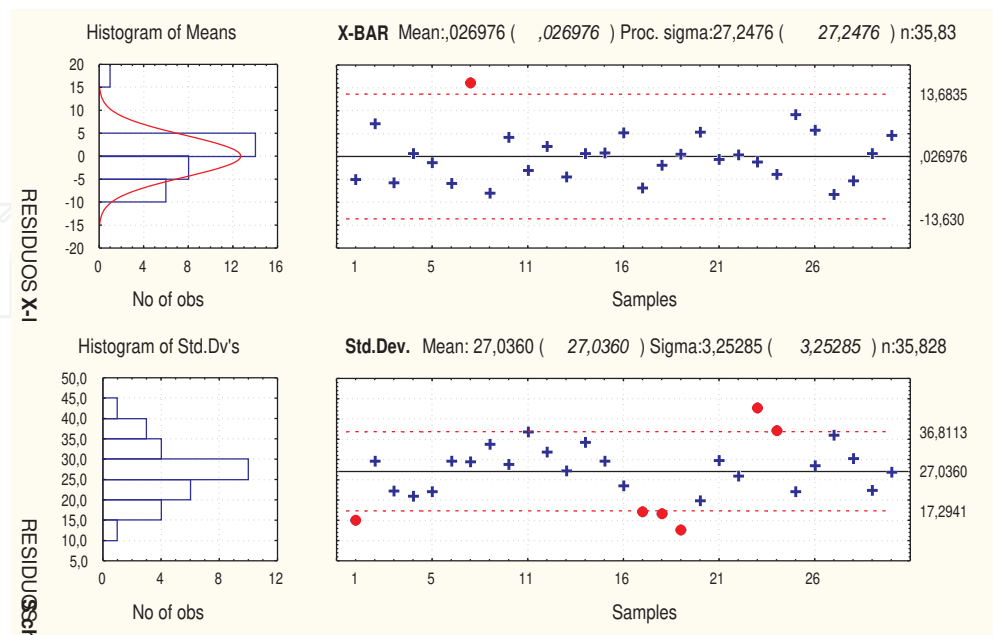


Figure 11.  $(\bar{X})$  and  $(S)$  charts for transformed data.

Figure 11 shows ( $\bar{X}$ ) and ( $S$ ) charts for transformed data. Verifications revealed that the system had been drained during this time period and actions were taken to correct the problem. The problem was in the first observations, which were ignored, the normality condition was verified, and the control charts  $\bar{X}$  and  $S$  were replotted.

Through the Figure 11 we can observe that the control charts for the same data, indicate that the residual values are practically inside of control limits for the average. According to Wardell, Moskowitz and Plant (1994) it is entirely possible in traditional control charts, the points are out of the limits because of the systematic or the common causes and not because of occurrence of special causes.

## 6. Conclusion

According to Reid and Sanders (2002), there are several types of statistical quality control (SQC) techniques. One category of SQC techniques consists of descriptive statistics tools such as the mean, range, and standard deviation. These tools are used to describe quality characteristics and relationships. Another category of SQC techniques consists of statistical process control (SPC) methods that are used to monitor changes in the production process. To understand SPC methods you must understand the differences between common and assignable causes of variation.

Common causes of variation are based on random causes that cannot be identified. A certain amount of common or normal variation occurs in every process due to differences in materials, workers, machines, and other factors. Assignable causes of variation, on the other hand, are variations that can be identified and eliminated. An important part of statistical process control (SPC) is monitoring the production process to make sure that the only variations in the process are those due to common or normal causes. Under these conditions we say that a production process is in a state of control. You should also understand the different types of quality control charts that are used to monitor the production process: x-bar charts, R-range charts, p-charts, and c-charts, Reid and Sanders (2002).

In this chapter we show how to use the techniques of quality control for autocorrelated data. Thus, the data collected were analyzed simultaneously, the continuous variables, to find a possible reason for lack of control in the final stages of production. We presented methods for using the techniques of statistical quality control for correlated observations. It is the autocorrelation data, is modeled by the continuous variables ARIMA. With the residuals obtained in the models, we applied the Shewhart control charts.

The traditional Shewhart control charts can be used for process control, even when the assumptions of independent observations are transgressed, by removing the autocorrelation with a time series models. For applying those techniques, the thread's resistance stayed in control state for the average. The result was a decrease in the variation of surface quality of the polypropylene thread that is produced, while simultaneously it increased the surface quality average.

Many companies, because they believe in the advantages that can be obtained from the practice of SQC, invest many resources in the implementation, especially of the conventional control charts, called Shewhart charts. Since it is not necessary to a thorough knowledge of statistics, is more favorable to the deployment of these graphs by the companies, but not always the results are as expected. There is a concern with the correlation of data.

In this context, the text presented throughout this chapter can serve as a reference to the industries that face difficulties in deploying statistical quality control. However, one must be careful with the type of variables to analyze what is being proposed, which allows us to conclude that this proposed combination of techniques for time series with control charts, claim to be complete and extended to cover all possible difficulties we can find. In the classic model of monitoring, there is no such information to identify an non conform item, in the end of the process, no one knows how to do for the same does not happen, because the variables used in the previous process are autocorrelated.

## Author details

Suzana Leitão Russo<sup>1\*</sup>, Maria Emilia Camargo<sup>2</sup> and Jonas Pedro Fabris<sup>3</sup>

\*Address all correspondence to: [suzana.ufs@hotmail.com](mailto:suzana.ufs@hotmail.com)

1 Department of Statistic - Federal University of Sergipe, Brazil

2 Post-Graduate Program in Administration, University of Caxias of South, Brazil

3 Fabris Industry, Brazil

## References

- [1] Alwan, L. C. (1991). Autocorrelation: fixed versus variable control limits quality and reliability. *Engineering International*, 4(2), 167-188.
- [2] Beusekom, D. P. V. (2003). Forecasting & Detrending Of Time Series Models. University of Amsterdam. Faculty of Sciences Study Bedrijfswiskunde & Informatica. ([www.few.vu.nl/.../werkstuk-beusekom\\_tcm39-91323.html](http://www.few.vu.nl/.../werkstuk-beusekom_tcm39-91323.html) access in Apr, 05, 2012).
- [3] Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics: 4 Ed.
- [4] Chen, A., & Elsayed, E. A. (2000). An alternative mean estimator for processes monitored by SPC charts. *Int. J. Prod. Res.*, 38(13), 3093-3109.

- [5] Camargo, M. E., & Russo, S. L. (2011). Modeling and forecasting tourist traffic in the Mauá and Alba Posse Harbors. *African Journal of Business Management*, 5(2), 433-439, 18 January.
- [6] Cochrane, J. H. (2005). Time Series for Macroeconomics and Finance. Chicago: Spring.
- [7] Dobson, B. (1995). Control charting dependent data: a case study. *Quality Engineering*, 7(4), 757-768.
- [8] Jiang, W., Tsui, K., & Woodall, W. (2000). A new SPC monitoring method: the ARMA chart. *Technometrics*, 399-410.
- [9] Fischer, S. (1982). Séries Univariantes de Tempo-Metodologia Box and Jenkins. Porto Alegre- RS, Fundação de Economia e Estatística.
- [10] Juran, J. M. (1993). Made in U.S.A.: A renaissance in quality. *Harvard Business Review*, July-August, 42-50.
- [11] Loureço Filho, R. C. B. (1964). Controle Estatístico de Qualidade. Rio de Janeiro: Livro Tec. S/A.
- [12] Montgomery, D. C. (2009). Introduction to Statistical Quality Control. Vol. 1, 6 Ed., USA: John Wiley & Sons.
- [13] Morettin, P. A., & Toloi, C. M. (2006). Análise de Séries Temporais. Ed Edgard Blucher. São Paulo.
- [14] Panagiotidou, S. (2009). An economically designed, integrated quality and maintenance model using an adaptive Shewhart chart. *Reliability Engineering & System Safety*, 94(3), 732-741.
- [15] Reid, R. D., & Sanders, N. R. (2002). Operations Management. 3rd Edition. John Wiley & Sons.
- [16] Russo, S., Rodrigues, P. M. M., & Camargo, M. E. (2006). Aplicação de Séries Temporais na Série Teor de Umidade da Areia de Fundação da Indústria Fundimisa. *Revista Gestão Industrial*, 1808-0448, 2(1), 35-45, Universidade Tecnológica Federal do Paraná-UTFPR, Campus Ponta Grossa- Paraná- Brasil, D.O.I.: 10.3895/S1808-04482006000100004.
- [17] Shewhart, W. A. (1931). Economic control of quality of the manufactured product. Van Nostrand, New York.
- [18] Wardell, D. G., Moskowitz, H., & Plante, R. (1994). Run-length distributions of special-cause control charts for correlates process. *Technometrics*, 36(1), 3-16.
- [19] Werkema, M. C. C. (1995). Ferramentas Estatísticas Básicas para o Gerenciamento de Processos. Belo Horizonte: Ed. Littera Maciel Ltda, Vol II.

