Australian Journal of Basic and Applied Sciences, 6(1): 207-212, 2012
ISSN 1991-8178

Multivariate Quality Control Basead On Discriminant Analysis-Ajusted Variables

[1]Maria Emilia Camargo, [2]Angela Isabel dos Santos Dullius, [3]Walter Priesnitz Filho,[4]Suzana Leitão Russo, [5]Márcia Rohr Cruz, [6]Ademar Galelli, [7]Gabriel Francisco da Silva

[1,5,6] Post-Graduate Program in Administration, University of Caxias of South, Brazil
[2] Federal University of Santa Maria, Brazil
[3]CTISM – Federal University of Santa Maria, Brazil
[4,7]Federal University of Sergipe, Brazil

Abstract: This paper presents an analysis of the performance of control charts based on Hotelling wastes obtained in the discriminant analysis for the mean and the variability in multivariate manufacturing process. It shows a real data application process for the production of soybean oil. The control chart performance was assessed based on the analysis of the relative efficiency of the ARL (average run length). The observations were obtained at a company in Rio Grande do Sul, Brazil, whose name will not be disclosed for strategic reasons, which were collected one hundred (100) samples of soybean oil, in two shifts, with the following variables: acidity of oil soybeans and soybean meal moisture. The data are for the period August 2008 to April 2010. Based on the analysis of process control charts, one can observe that the process is not under control. Since the observed fact that the moisture content and acidity, there were points outside the limits of control charts, as well as identify trends.

Key words: Multivariate Process, Hotelling's T^2 Residuals Charts; Residuals Discriminant Analysis.

INTRODUCTION

The quality of products and services is one way that consumers use to get an indication of the quality of the company. It is therefore of paramount importance to adopt innovative ways of quality control, and have this awareness that the pursuit of quality is a dynamic process that requires constant improvement and progress permanent. We use statistical control charts to monitor the performance of production processes. The observations were obtained at a company in Rio Grande do Sul, Brazil, whose name will not be disclosed for strategic reasons, which were collected one hundred (100) samples of soybean oil, referring to the morning and afternoon, with the following variables acidity of soybean oil and soybean meal moisture. The data are for the period August 2008 to April 2010. Based on the analysis of process control charts, one can observe that the process is not under control. Since the observed fact that the acidity and moisture variables, there were points outside the limits of control charts, as well as identify trends.

This paper presents an analysis of the performance of control charts based on Hotelling wastes obtained in the discriminant analysis for the mean and the variability in multivariate manufacturing process.

The paper is organized in four sections: The basic theory related to this paper is presented in section 2. The empirical analysis and discussion on the results are presented in section 3. Final considerations are in section 4.

Basic Theory:

In a univariate process, the Shewhart charts (Johnson, Wichern, 1999), monitor the process average and play a great role, because they can identify when a source of variation is affecting the process, which may be due to common causes or assignable causes. However, it is common that a product has several quality characteristics that need to be monitored together. These graphs, then should no longer be used because the variables have a correlation between them, this will impair the performance of the graph X (Bar) to signal a lack of control.

An alternative to monitor these features, we ignore this correlation and univariate charts to use for each of the characteristics which, in addition to lead to errors, sometimes it becomes impossible by the number of graphs that must be traced.

In these cases, a multivariate chart can provide better results in the monitoring of variables. Hotelling (1947), was the first to realize the influence of multiple univariate control charts when there was a correlation between the variables. The T^2 control chart, based upon Hotelling's T^2 statistic, is used to detect shifts in the process.

Corresponding Author: Maria Emilia Camargo, Post-Graduate Program in Administration, University of Caxias of South, Brazil
E-mail: kamargo@terra.com.br

2.1 Mathematical derivation for the critical region: Q Test:

Let X = [X_1, X_2, ..., X_k] a vector of means for the components k significant features from a subset of n units of the product. Considering the matrix V - variance-covariance matrix of the vector X. By extension of the central limit theorem, the limit distribution of X with n becomes large and can be assumed that the distribution approaches a multivariate quadratic form.

$$Q = [X - \mu]^t V^{-1} [X - \mu] \tag{1}$$

where:

μ: population mean;

X_i: are the values of the population;

V-1: inverse matrix of V (variance-covariance matrix);

The quadratic form of multivariate distribution:

$$f(X) = f(X_1, X_2, ..., X_k) = \frac{[V^{-1}]^{1/2}}{(2\pi 2^{k/2}} e^{-(X-\mu)^t} V^{-2} (X - \mu)/2 \tag{2}$$

can be demonstrated as having a chi-square distribution with k degrees of freedom.

Thus, f(Q) = χ_k^2. With the above, one can define a confidence interval and build up a chart to serve as a test of stability in the pattern of variation.

2.1.1 Mathematical derivation for the critical region - Special case for two characteristic:

We have the bivariate case, when k = 2, where X_1, X_2 are two control characteristic. To get the mathematical expression of Q, one must calculate the variance, covariance and by applying the quadratic, we arrive at an expression of chi-square distribution. For the bivariate plot, we get the equation of an ellipse in the plane. This ellipse will always be of the form:

$$\alpha Y_1^2 + \beta Y_2^2 + \gamma Y_1 Y_2 \leq \chi_{2,\alpha}^2 \tag{3}$$

Having, therefore, outside the center of the origin of the coordinate axes. Its center is the point of intersection of the two averages $\overline{X}_1$ and $\overline{X}_2$ the two characteristics.

When k = 2, the vector X = [X_1, X_2, ..., X_k] takes the form X = [X_1, X_2]. A probability of 99% in the sector can be described by the following equation:

$$\Pr[0 \leq Q \leq \chi_{2;0,99}^2] = 0{,}99 \tag{4}$$

which geometrically represents the plane X_1, X_2, an ellipse. This ellipse is represented in Figure 1, is represented as the limit of real control over the bivariate control chart. This chart has the advantage of not registering a dimensionless Q. The dimensions of the subsets X_1 and X_2 can be plotted directly on the graph. The process is under control when the points representing the subgroup averages are within the ellipse. This is shown graphically for the bivariate case in Figure 1.

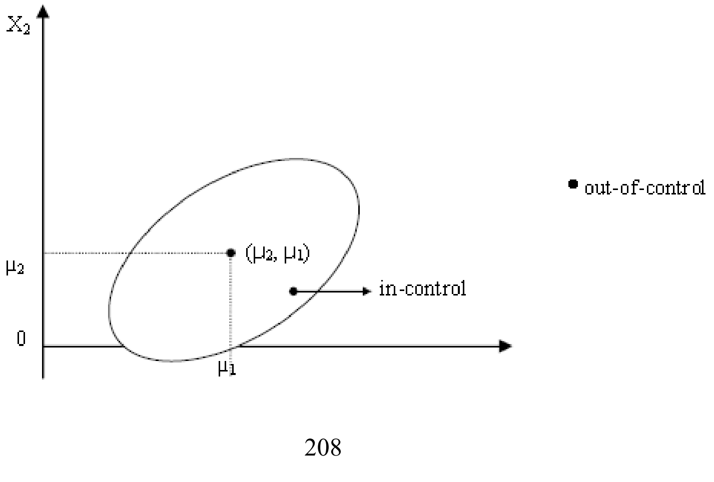

Fig. 1: Critical region of control bivariate

2.2 Discriminant Analysis:

The discriminant analysis, also known as Fisher's linear discriminant. This technique uses information of the categories associated with each pattern to extract linear features more discriminating. Through Fisher's discriminant analysis can also perform the discrimination between classes, supervised by processes (known as the default) or by unsupervised processes, where it is used when one has a known standard. The linear discriminant function is a supervised method in statistical design and should be used when certain conditions are met, such as: (1) classes under investigation are mutually exclusive, (2) each class is obtained from a multivariate normal population, (3) two measures can not be perfectly correlated, among others.

According Khattree and Naik (2000), is a multivariate statistical technique that studies the separation of objects from one population into two or more classes. Discrimination or separation is the first step, being part of exploratory analysis is to be sought and features capable of being used to allocate objects into different groups previously defined. The classification or allocation can be defined as a set of rules that will be used to allocate new objects (Johnson and Wichern, 1999). However, the function that separates objects can also be used to allocate, and the reverse rules that allocate objects can be used to separate. Typically, discrimination and classification overlap in the analysis, and the distinction between separation and allocation is unclear.

The problem of discrimination between two or more groups in order to further classification, was first addressed by Fisher (1936). Is to obtain math functions able to classify an individual X (an observation X) in one of several populations **π_i,** (i=1, 2, ..., g), based on measurements of a number of features p, seeking minimize the probability of misclassification, ie, minimize the probability of erroneously classifying an individual in a population **π_i,**, when it really belongs to the population **π_j,** ($i \neq j$), i, j = 1, 2, ..., g.

2.2.1 Fisher linear discriminant function:

The Fisher linear discriminant function is a linear combination of unique features which are characterized by producing maximum separation between two populations. Considering that μ_i and Σ are known parameters and respectively, the vectors of mean and covariance matrix of the common populations π_i. It is shown that the linear function of the random vector X that produces maximum separation between two populations is given by:

$$D(X) = L^{'} \cdot X = [\mu_1 - \mu_2]^{'} \cdot \Sigma^{-1} \cdot X \tag{5}$$

where,

$X = \lfloor X_1 \; X_2 \cdots X_p \rfloor$ e $\pi = [\pi_1, \pi_2]$

L: discriminant vector;

X_i: randon vector of characteristics of populations;

μ: mean vector;

Σ: common covariance matrix of the populations **π_1** and **π_2;**

The value of the Fisher discriminant function for a given observation X_o is:

$$D(x_o) = [\mu_1 - \mu_2]^{'} \cdot \Sigma^{-1} \cdot x_o \tag{6}$$

The midpoint between the two population means and univariate μ_1 and μ_2 is:

$$m = \frac{1}{2}[\mu_1 - \mu_2]^{'} \cdot \Sigma^{-1} \cdot [\mu_1 + \mu_2] \text{ , ie}$$

$$m = \frac{1}{2}[D(\mu_1) + D(\mu_2)] \tag{7}$$

The classification rule based on Fisher discriminant function is:

Put X_o in π_1 if $D(x_o) = [\mu_1 - \mu_2]^{'} \cdot \Sigma^{-1} \cdot x_o \geq m$

Put X_o in π_2 if $D(x_o) = [\mu_1 - \mu_2]^{'} \cdot \Sigma^{-1} \cdot x_o < m$

Assuming that the populations π_1, π_2 have the same covariance matrix Σ we can then estimate a common covariance matrix Sc:

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)}\right] \cdot S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)}\right] \cdot S_2 \tag{8}$$

where,

S_c: estimate of common covariance matrix Σ;

n_1: number of observations of the population π_1;

n_2: number of observations of the population π_2;

S_1: estimated covariance matrix of the population π_1;

S_2: estimated covariance of the population π_2;

The linear discriminant function of Fisher sampling is obtained by replacing the parameters μ_1, μ_2 and Σ their respective sample quantities $\overline{x}_1$, $\overline{x}_2$ and S_c:

$$D(x) = \hat{L}' \cdot x = [\overline{x}_1 - \overline{x}_2]' \cdot S_c^{-1} \cdot x \tag{9}$$

Where

D(*x*): sample linear discriminant function of Fisher;

$\hat{L}'$: estimate of discriminant vector

$\overline{x}_1$: sample mean of population π_1;

$\overline{x}_2$: sample mean of population π_2

2.3 The average run length (ARL):

The performance of a control chart can be evaluated in terms of sensitivity to detect deviations in the statistics being monitored (Montgomery, 1997).

This sensitivity can be measured by the number of samples collected until the chart signals the occurrence of a deviation.

The number of samples (points) from the re (beginning) of the process until the instant when a signal is out of control, excluding the sample responsible for issuing the sign is the RL (Run Length) and the average number of samples is the ARL (Average Run Length) (Montgomery, 1997).

For the control chart is necessary to send that signal that the time is taken into consideration. If the process is under control, this time should be increased so that the false alarm rate is reduced. If the process is out of control, this time should be short so that the change is detected quickly (Montgomery, 1997).

2.4 Average Time to Signal (ATS):

The Average Time to Signal (ATS) is defined as the mean time since the beginning of the process until the issue of an out of control signal (possibly a false alarm) for the control chart. In the case of a control chart Shewhart, (Shewhart, 1931), type with variable intervals, keeping the steady state of the process, the adaptive intervals are independent and the same distribution of a generic variable D, so that, by Wald's identity (Ross (1970)), we have:

$$ATS = E(D) \times E(Nf) \tag{10}$$

Here is assuming that the first sampling interval has the same distribution of the remaining, which is unusual in real data.

The ATS is a measure used when it is admitted that the process starts already with the presence of one or more assignable causes. However, in most practical situations the process starts under control and a certain future time, as a result of one or more assignable causes, the characteristic parameter of the quality changes. In this case, it is interesting to determine the time interval from the instant at which the change occurs until it is detected by the control chart.

3. Empirical Analysis:

The observations were obtained at a company in Rio Grande do Sul, Brazil, whose name will not be disclosed for strategic reasons, which were collected one hundred (100) samples of soybean oil in two different shifts, with the following variables: acidity of the oil soybeans and soybean meal moisture. The data are for the period August 2008 to April 2010.

Soybean harvest season: March, April and May. The younger the grain (closer to the time of harvest), the lower the acidity of the oil. The higher the percentage of moisture, the lower the percentage of protein in soybean meal.

According to the contract buyer of oil, generally, this acidity should be between 0.8% and 1%, otherwise the load back to the company and is usually sold to poultry farmers to increase the fat in the diet.

3.1 Residual Analysis:

We considered two groups for the morning and afternoon shifts. The discriminant function was found:

$$D(x) = 0{,}49x_1 + 13{,}19x_2$$

For this linear equation was calculated number of residues, which were built to control charts Hotelling T^2, shown in Figure 2.

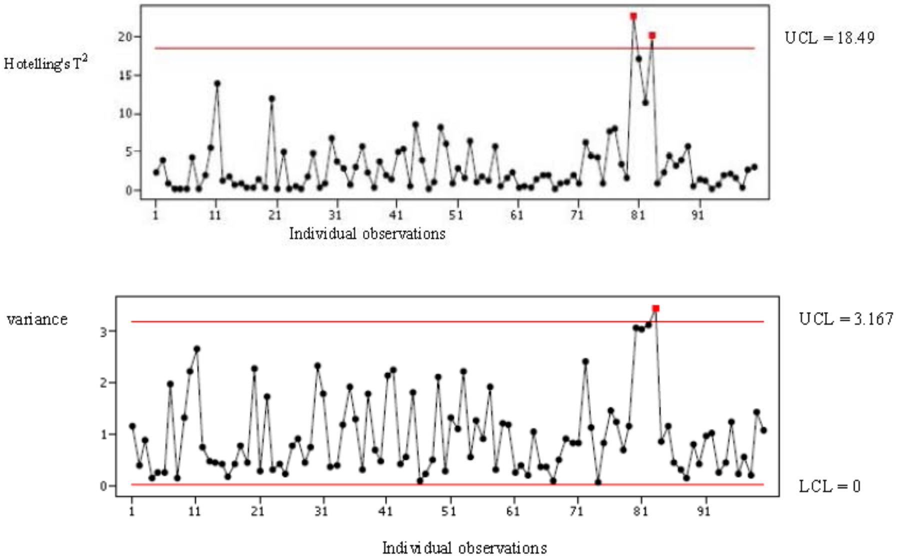

Fig. 2: Hotelling's T^2 control chart for the mean and the variability

Analyzing Figure 2 can be stated that the process is out of control, both the mean and variability, after 83 th days after it was examined each of the variables through the individual graphs, noting that both variables are presented out of control.

3.2 Analysis of the Relative Efficiency of ARL's:

Given a probability of p = 0.05, we have:
ARL=1/p   ARL=20   ATS=ARLX24h (daily values)
ATS=20X24   ATS=480h   480h = 20 days

You could say that the next element will be out of control on the 20th day after the day corresponding to 83th days for both the variable acidity and moisture.

4. Final Considerations:

This paper presented an analysis of the performance of control charts based on Hotelling built waste obtained in the discriminant analysis for the mean and the variability in the production process regarding production of soybean oil, analyzing the variables acidity and moisture.

Acidity in the series of soybean oil was found to average 0.49, range of 0.31 and standard deviation of 0.13.

The humidity range of soybean meal, showed an average of 13.15, a range of 1.93 and a standard deviation of 0.80.

When analyzing the data and then elaborate control charts control, one can observe that the process is not under control. Since the observed fact that the graph T^2, there were out of control points. After the variables were analyzed individually and can be seen that the moisture content and acidity, presented points out the limits of control charts. Thus, we conclude that the process studied is out of control both in the process mean and variability.

REFERENCES

FISHER, R.A., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7: 179-188.

Johnson, R.A., D.W. Wichern, 1999. Applied multivariate statistical analysis. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall.

Khattree, R., D.N. Naik, 2000. Multivariate data reduction and discrimination with SAS software. Cary, NC, USA: SAS Institute Inc.

Hotelling, H., 1947. Multivariate quality control. Techniques of statistical analysis. New York: Mc Graw Hill, p: 111-184.

Montgomery, D.C., 1997. Introduction to statistical quality control. 3 ed. New York: John Wiley & Sons.

Ross, S.M., 1970. Applied Probability Models with Optimization Applications, Holden-Day, San Francisco.

Shewhart, W.A., 1931. Economic control of quality of the manufactured product. Van Nostrand, New York.