4-1-2006

Multivariate Control Charts for Attribute Data

Donald S. Holmes
Stochos, Inc.

A. Erhan Mergen
Rochester Institute of Technology

2006 Northeast Decision Sciences Institute

Thirty-Fifth Annual Meeting
March 30 - April 1, 2006
Caribe Hilton San Juan
San Juan, Puerto Rico

Proceedings Editor
Doug White
Gabelli School of Business
Roger Williams University

MULTIVARIATE CONTROL CHARTS FOR ATTRIBUTE DATA

Donald S. Holmes, Stochos Inc. 14 N. College Street, Schenectady, N.Y. 12305.
(518) 372-5426, dsholmes@stochos.com
A. Erhan Mergen, Rochester Institute of Technology, College of Business
Decision Sciences, 107 Lomb Memorial Drive, Rochester, N.Y. 14623-5608.
(585) 475-6143, emergen@cob.rit.edu

ABSTRACT

In this paper the use of multivariate control charts for attribute data is proposed. These charts are based on chi-square statistics. Data from various categories can be summarized into a multivariate statistic, i.e., the chi-square statistic, and then the process can be monitored by plotting this statistic on a control chart. A numerical example is provided.

Keywords: Attribute data, multivariate analysis, chi-square statistics.

DISCUSSION

Multivariate statistical process control (SPC) techniques, such as T^2 Control charts (see, for example, Montgomery [5]), are widely available for variable data. The major advantage of the multivariate techniques is the ability to monitor multiple continuous variables on the same control chart. This helps reduce type I and type II errors. In a process where various categories of attribute data are collected, the stability check of the process through several individual attribute control charts could be quite time consuming. In addition, the desired alpha level, i.e., type I error, could be inflated with the use of multiple control charts. Thus, if the goal is to check of the overall stability of the process, one summary chart which includes all the attribute variables, i.e., a multivariate control chart, will be more appropriate. This way the desired type I error would be maintained.

In this paper we are using a multivariate approach to answer the question, "Is the process changed?" by looking at all the available attribute data together in one control chart. We are proposing to use Chi Square statistics as the summary statistics. Similar approaches have been proposed by Duncan [2], Marcucci [4] and Nelson [7] to monitor the stability of the percentages and proportions. Holmes and Mergen [3] also proposed a Chi Square control chart to check the stability of variable data from a process by looking at the entire distribution of data, rather than just looking at mean and standard deviation.

The Chi Square statistic is one of the multivariate statistics which detects change. We therefore use this statistic to build a multivariate chart for multiple groups of attribute data to monitor changes in the process. The attribute data should be classified into several categories, such as types of defects. The statistic to be monitored, i.e., Chi square statistics (χ^2), will be determined as follows for each batch of size n:

$$\chi^2 = \sum_{i=1}^{k} \frac{(Y_i - \overline{Y}_i)^2}{\overline{Y}_i} \qquad (1)$$

where k = number of categories of the attribute data

Y_i = observed frequency of attribute data in category i in a batch or lot

$\overline{Y}_i$ = expected frequency of attribute data in category i.

This statistic will have k-1 degrees of freedom (Dixon and Massey [1]). For example, if there are ten categories, there will be nine degrees of freedom. The expected value and the variance of this statistic are given as

$$E(\chi^2) = k - 1 \tag{2}$$

$$Var(\chi^2) = 2(k - 1) \tag{3}$$

respectively (Mood, Graybill and Boes [6]).

After the chi-square statistics are obtained as defined above, they can be monitored to check the stability of the process by using a Chi-Square control chart (Holmes and Mergen [3]) or using an ordinary control chart, such as Cumulative Sum (Cusum), an exponentially weighted moving average (EWMA) control chart, etc.

Chi-square control chart limits for about 99.7% can be built as follows:

Central line (CL), i.e., the average of the chi-square statistics, of the chart will be:

$$CL = k - 1 \tag{4}$$

$$\text{Upper control limit (UCL)} = CL + 3\sqrt{2(k-1)} \tag{5}$$

$$\text{Lower control limit (LCL)} = CL - 3\sqrt{2(k-1)} \tag{6}$$

If the sample chi-square statistics fall within the above defined limits with no non-random pattern, then the process is said to be stable. Otherwise, it is declared unstable (i.e., out-of-statistical control). However, getting an out-of control signal from the chi-square chart will not indicate which one(s) of the k categories is causing the problem; all it will say is that there is significant evidence that the process has changed. To find out which one(s) of the k categories is the source of the problem, each category has to be analyzed individually; this is a necessary step in any multivariate statistical process control application when there is an out of control signal. Of course, a large sample size would be the other requirement to calculate the sample chi-square statistics. The requirement for a large sample size, along with the relatively difficult calculation of the chi-square statistics, necessitates the use of computerized SPC systems for such charts.

The alternative to chi-square control charts would be to use EWMA or Cusum charts to monitor the chi-square statistics to check for the stability. These charts are known to be very sensitive to small changes in the process (see, for example, Montgomery [5]); they also have small average run lengths (ARL) detecting changes in the process. EWMA is commonly used in process industries (i.e., low data environments). The EWMA chart is similar to a Moving X-bar chart. In a Moving X-bar chart, each data has equal weight. In EWMA the weights are distributed geometrically:

$$EWMA_t = 0.1X_t + (0.9)(0.1)X_{t-1} + (0.9)(.9)(.1)X_{t-2} + (0.9)(.9)(.9)(.1)X_{t-3} + \ldots \tag{7}$$

The generic form of an EWMA chart is:

$$\text{EWMA}_t = \alpha X_t + (1-\alpha)\alpha X_{t-1} + (1-\alpha)^2 \alpha X_{t-2} + (1-\alpha)^3 \alpha X_{t-3} + \ldots \tag{8}$$

where $0 < \alpha \leq 1$

The EWMA may also be calculated by the equation below.

$$\text{EWMA}_t = \alpha X_t + (1-\alpha)\text{EWMA}_{t-1} \tag{9}$$

The standard deviation of the set of EWMAs is:

$$\sigma_{\text{EWMA}} = \frac{\sigma_x}{\sqrt{\frac{2-\alpha}{\alpha}}} \text{ or } \sigma_x\sqrt{\frac{\alpha}{2-\alpha}} \tag{10}$$

where σ_x is the estimate of the process capability standard deviation, which can be determined, for example, as $= \frac{\overline{R}}{d_2}$ or $\frac{\overline{s}}{c_4}$, MSSD, etc.

The EWMA Control limits are then given as

$$\overline{\text{EWMA}} \pm k\sigma_{\text{EWMA}} \tag{11}$$

where k is a constant and is usually equal to 3. See, for example, Montgomery [5] for these charts.

The chi-square approach to multiple groups of attribute data would make it easier to monitor different categories of attribute data in one chart. This saves time and resources. Of course, in the case of an out-of-control signal, additional tests will need to be done to find the category (or the categories) which triggered the signal.

EXAMPLE

To demonstrate the proposed approach we will use an attribute data set collected on five different categories of defects over 35 batches (see the Appendix). Chi-square statistics were calculated for each batch. An EWMA chart with a smoothing constant, i.e., weighting factor, (α) of 0.1 is built using these chi-square values. Starting value for the EWMA is taken as the average of the first 16 chi-square values. Control limits of the chart are determined by using the first 16 chi-square values so that the first half and the second half of the data can be compared. As you notice the chi-square values are getting smaller in the second half of the data indicating that batch to batch differences from the standard distribution of defects in a batch ($\overline{Y}_1, \overline{Y}_2, .. \overline{Y}_k$) are getting smaller and smaller (see Figure 1 below).

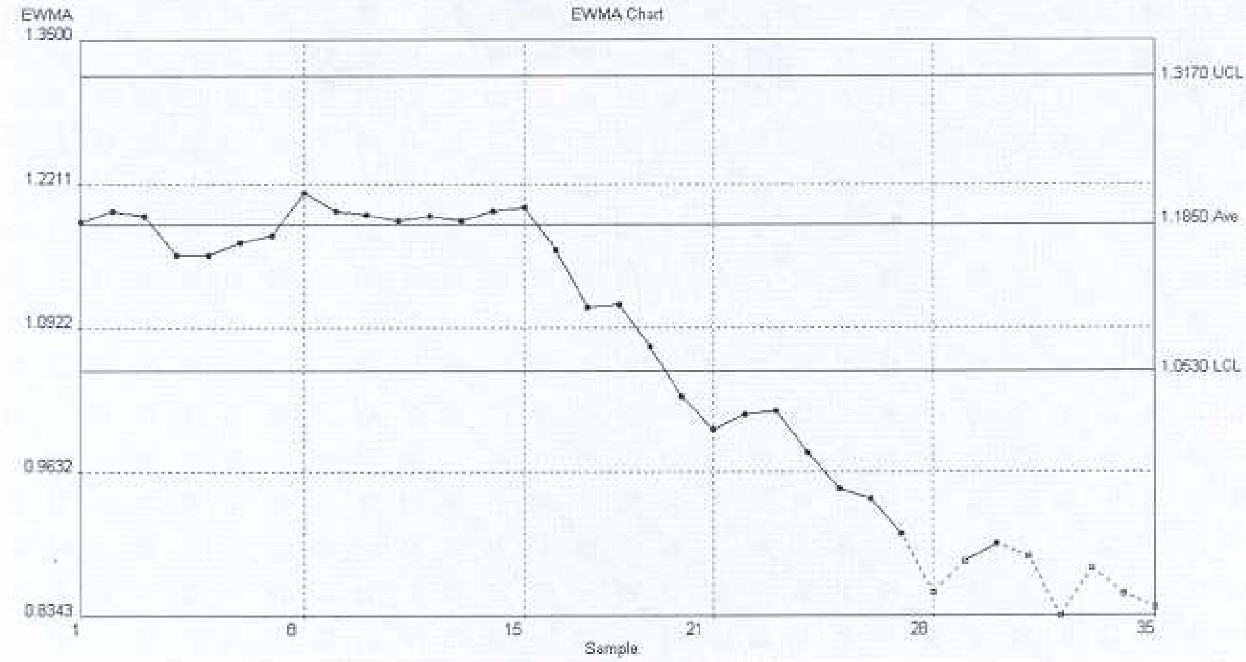

Figure 1. EWMA for Chi square values.

CONCLUSION

In this paper we proposed a multivariate approach to process control for attribute data. Using such an approach would help reduce the type I and type II errors and reduce the number of charts required to monitor the process. Thus, if the question to answer is a check of the overall stability of the process, one summary chart which includes all of the (attribute) variables, i.e., a multivariate control chart, will be most appropriate, which would help save time and resources.

APPENDIX

CATEGORIES OF DEFECTS

C1	C2	C3	C4	C5
8	20	24	25	14
6	16	19	20	11
7	17	21	22	12
7	16	20	21	12
7	17	21	22	12
6	16	20	21	11
7	16	20	21	11
7	18	22	23	12
8	19	24	25	14
7	17	21	22	12
7	18	22	23	13
7	16	20	21	11
7	18	22	23	13
6	16	20	21	11
7	16	20	21	11
6	11	21	19	16
7	12	22	21	17
6	10	20	19	16
7	11	21	20	16
7	13	23	22	18
7	11	21	20	16
6	10	20	19	16
7	11	21	20	17
7	12	22	21	17
8	13	23	22	18
9	17	27	23	21
7	13	23	22	18
7	13	23	22	17
6	10	20	19	16
7	11	21	20	17
6	10	19	18	15
8	15	24	27	17
6	9	18	17	15
7	12	22	21	17
7	11	21	20	16

REFERENCES

[1] Dixon, W.J., Massey, F.J. *Introduction to statistical analysis*. New York, NY: McGraw-Hill, 1969.

[2] Duncan, A.J. "A chi-square chart for controlling a set of percentages." *Industrial Quality Control,* 1950, 7(3), 11-15.

[3] Holmes, D.S., Mergen, A.E. "Chi-square vs. X-bar and R chart." *Quality*, 1986, 25(2), 60-61.

[4] Marcucci, M. "Monitoring multinomial processes." *Journal of Quality Technology*, 1985, 17(2), 86-91.

[5] Montgomery, D.C. *Introduction to statistical quality control.* New York, NY: Wiley, 2001.

[6] Mood, A.M., Graybill, F.A., Boes, D.C. *Introduction to the theory of statistics*. New York, NY: McGraw-Hill, 1974.

[7] Nelson, L.S. "A chi-square control chart for several proportions." *Journal of Quality Technology*, 1987, 19(4), 229-231.