E-BUSINESS WEB SITE MONITORING

Mary R. Lind, NCA&T State University, lindm@ncat.edu

Abstract

Using control chart methodology, the service quality of a commercial web site is examined. The web site statistics were obtained using a web monitoring tool and then analyzed using x-charts. Analysis of the usage statistics across time periods displayed much variance in the results for the March time period while the April period was more in control.

Keywords: e-business, web site monitoring, xchart, control chart

INTRODUCTION

In web based systems, service quality involves a complex interplay between the provider's service design / execution and the customer's expectations and perceptions in a single service encounter. In web based customer systems, sound measures of quality/ productivity outcomes should summarize operating data in a format that is of interest to management - including customer usage, customer buying behavior, and efficiency ratios of web usage (5).

Given the potential difficulties in examining web based service outcome data, there is a need for alternate analytical techniques which preserve - rather than obscure- the richness of time series quality related data, while still assuring statistical rigor (8). This paper proposes an alternate methodology for analyzing the quality of web based service quality using control chart theory. If a process is out of control, or unsettled, then the limits of variation in process data are not predictable, making it impossible to anticipate how the process will behave in the future. This research paper will use X-chart methodology to address the service quality stability of a customer web site.

THE *X CHART*

Shewart control charts (i.e. charts for which control limits are proportional to the standard deviation of the random variable plotted) have been a basic tool in the statistical control of industrial processes for over fifty years (3). A special type of Shewart control chart is the *x chart* or the chart of individual measurements. However, the x chart is actually **more** sensitive to certain non-random data patterns, including trends, short cycles and clustering effects, than the other Shewart charts because the individual data observations are not aggregated as they are in the *X-bar* and *R charts* (10,7,9).

CONSTRUCTION OF THE *X CHART*

Since the *x chart* is a specific type of *Shewart control chart*, the basic formulas for setting the upper and lower control limits are where $\hat{m}$ is a statistical estimate of the true mean and $\hat{s}$ is an

$$UCL_x = \hat{m} + k\hat{s}$$
$$UCL_x = \hat{m} - k\hat{s}$$

estimate of the standard deviation. The scalar k is set equal to the number of standard deviations necessary to guarantee a particular confidence interval. (Setting $k = 3$ is conventional for a 99.7 percent confidence interval when normality of the data is assumed.) The arithmetic mean $\bar{x}$ is commonly used to provide the estimate of the population mean m. In constructing the *x chart*, the arithmetic mean of an initial series of no fewer than 10 (and ideally 20 or more observations) is computed (10). The computed mean is substituted for $\hat{m}$ from the above equations becomes the center line for the chart.

As the base sequence of observations may exhibit non-random patterns, a common approach is to use the mean of moving ranges of successive observations (1). The average moving range is computed as

$$\overline{MR} = \sum_{i=1}^{n} \frac{| x_i - x_{i-1} |}{n - 1}$$

Thus, expressions for UCL become

$$UCL_x = \bar{x} + \overline{MR}/1.128 = \bar{x} + 2.66\overline{MR}$$

or equivalently,

$$LCL_x = \bar{x} - 3\overline{MR}/1.128 = \bar{x} - 2.66\overline{MR}$$

Nelson (6, p.173) notes that the "moving range rides trends and slow oscillations and is minimally affected by them; and a moving range of two is least affected of all."

AN APPLICATION OF THE X CHART

The application context involves a web site that provided a unique resource for its users providing them with information that could not be found from any other computer based source. The web site statistics were obtained from a web monitoring tool that collects on a daily basis data on usage of the site and the users who are accessing the site.

To explore the use of X Charts in web site tracking (4), two measures were used in this exploratory study: percentage of new visitors/total visitors and number of web page views. These were

examined over a time span of five months. Each set of measures was tested for autocorrelation; no serial correlation existed within either data set. An x chart was then constructed for each ratio. Each set of weekly measures was then plotted on its respective chart; it was concluded that the period 1/15 – 2/9 for each measure was in control (i.e., no significant outliers, no significant runs above or below the mean, and no significant run lengths). The analysis of the plots of the baseline data was based on decision rules for identifying non-random data patterns in control charts. *Shewart's Criterion I* which states that the data pattern is unnatural if one or more points falls outside a control limit. The presence of points beyond the control limits indicates process instability (i.e., abnormal causes of variation have entered the process) (10). The second and third decision rules involved the theory of runs. Duncan (2, p.428) defines a run as "a succession of items of the same class." Thus, runs may be analyzed by determining the number of runs belonging to a particular class or the length of the longest run of a specific type. In both cases, a greater number of runs belonging to one class or a particularly long run indicate non-random behavior. The runs analyzed in this data were the number of runs above or below the mean (decision rule 2) and the number of runs up or down (i.e. increases or decreases), which comprised the third decision rule.

Since analysis of the January *x charts* of the baseline data indicated process stability, the quality of the website was then studied by plotting measures from the following January period on their respective charts. Seasonal variation in web access was accounted for by not including any holiday periods. The post baseline data were plotted on their respective *x charts*; the plots are displayed in Figures 1 - 4.

RESULTS DISCUSSION

The results of the *x chart* of the percentage of new visitors displayed in Figure 1 reveals a significant run downward for the percentage of new visitors for 3/5 – 3/23 on the web site. Thus, indicating an abnormal cause of variation in the percentage of new visitors in Figure 1. Figure 2 displays the percentage of new visitors for the 4/2 – 4/20 period. While two points were plotted beyond the control limits, the other observations were plotted around the center line. Application of *rule 2* showed that the number of runs above/below the mean were not more than expected so the number of runs up is not significant. Thus, not indicating an abnormal cause of variation in the percentage of new visitors in Figure 2. Thus the visits to the web site over one of these two periods exhibit unstable behavior. Looking at the total page views for the same two time periods (Tables 3 and 4) shows that for the 3/5 – 3/23 time period that while there was clearly more new visitors there was also a greater number of page views enough in fact to indicate a significant run in the viewing process. For the period 4/2-4/20, the page views appear to be fairly stable with runs of two but not enough to be significant.

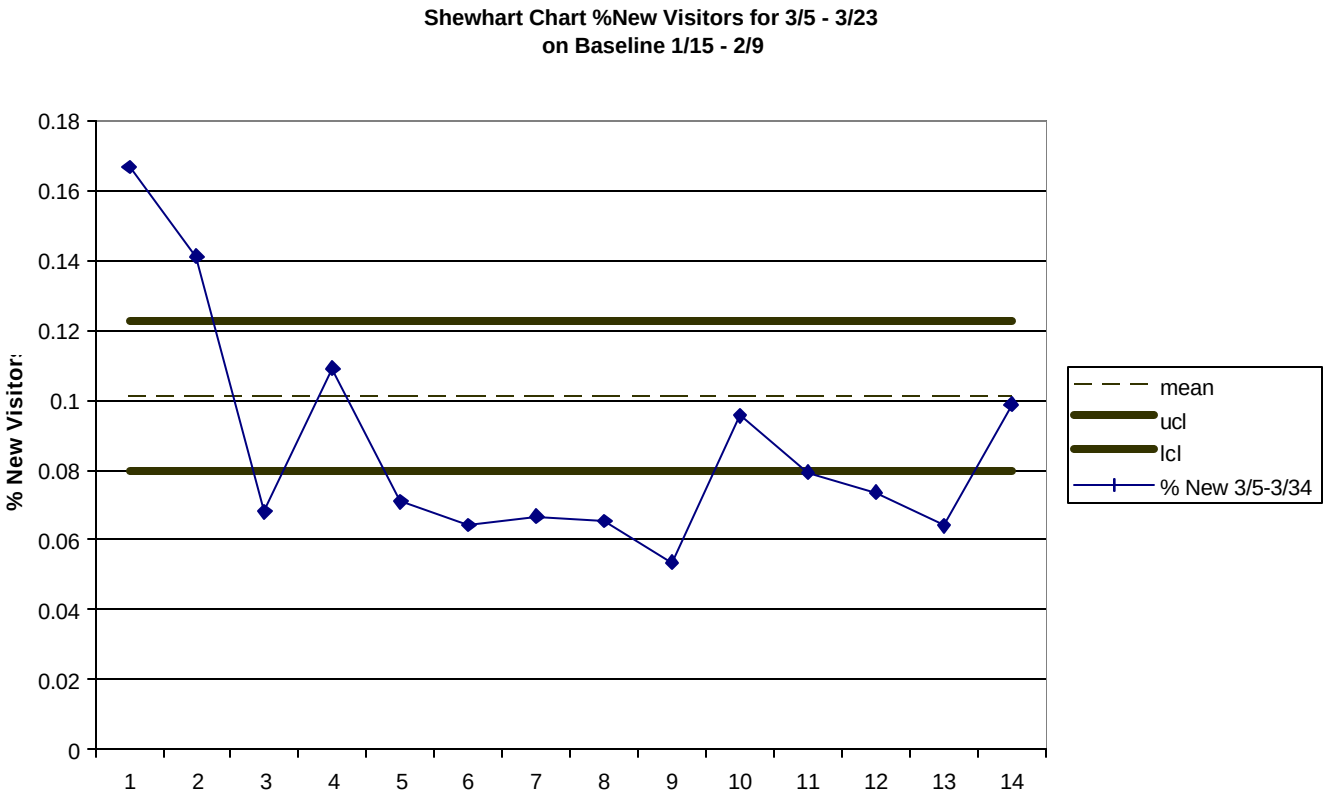
This discussion of the *x chart* for web monitoring indicates that it is important for management to use proper tools for evaluating website data. The relative simplicity of *x chart* construction and interpretation makes it an attractive alternative to more complex statistical techniques; furthermore, the *x chart* can be used when the available data does not fit the assumptions of the statistical models or when the data requires aggregation. Just as importantly, the *x chart* methodology can be used to determine if there are no other abnormal causes of variation in the process. In conclusion, the *x chart* methodology is a set of straightforward yet powerful tools for

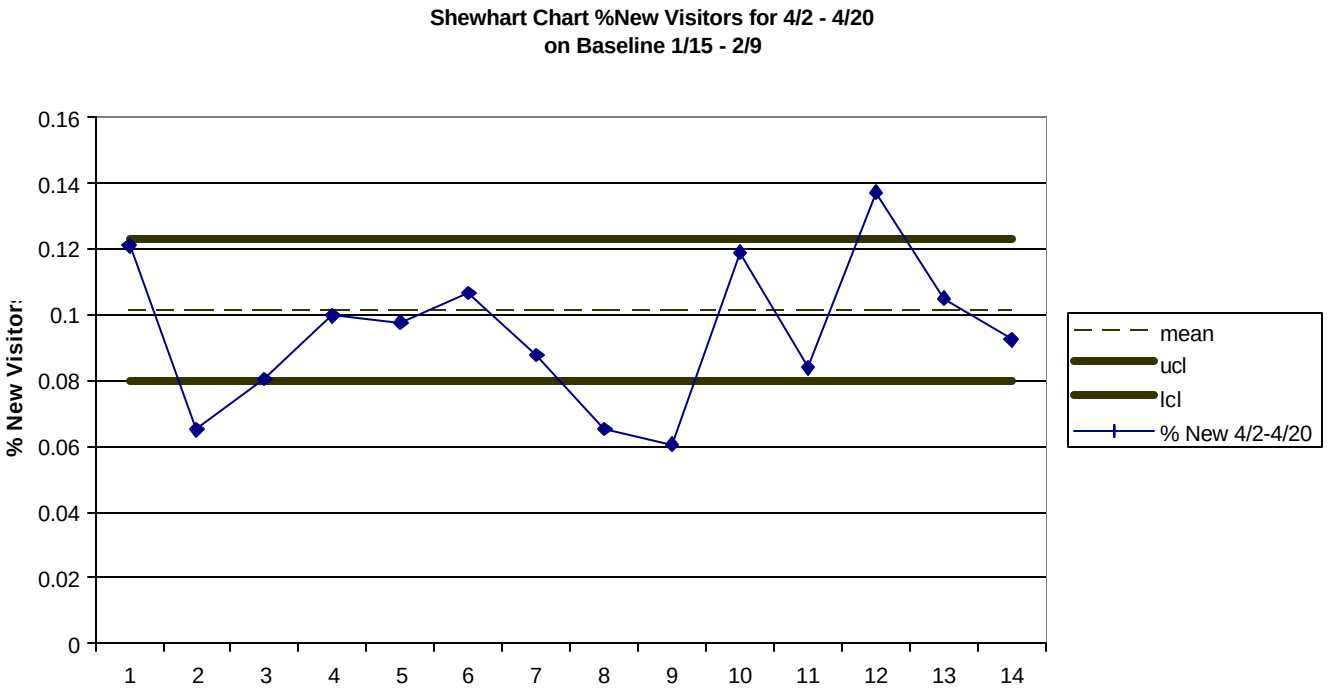
management to monitor their web site usage measures. As more data becomes available the control chart methodology will be used to analyze multiple periods month's of web statistics with control chargs. This should provide a clearer picture of the process stability of the web site with longer time spans in which to examine runs or trends in the results.

REFERENCES

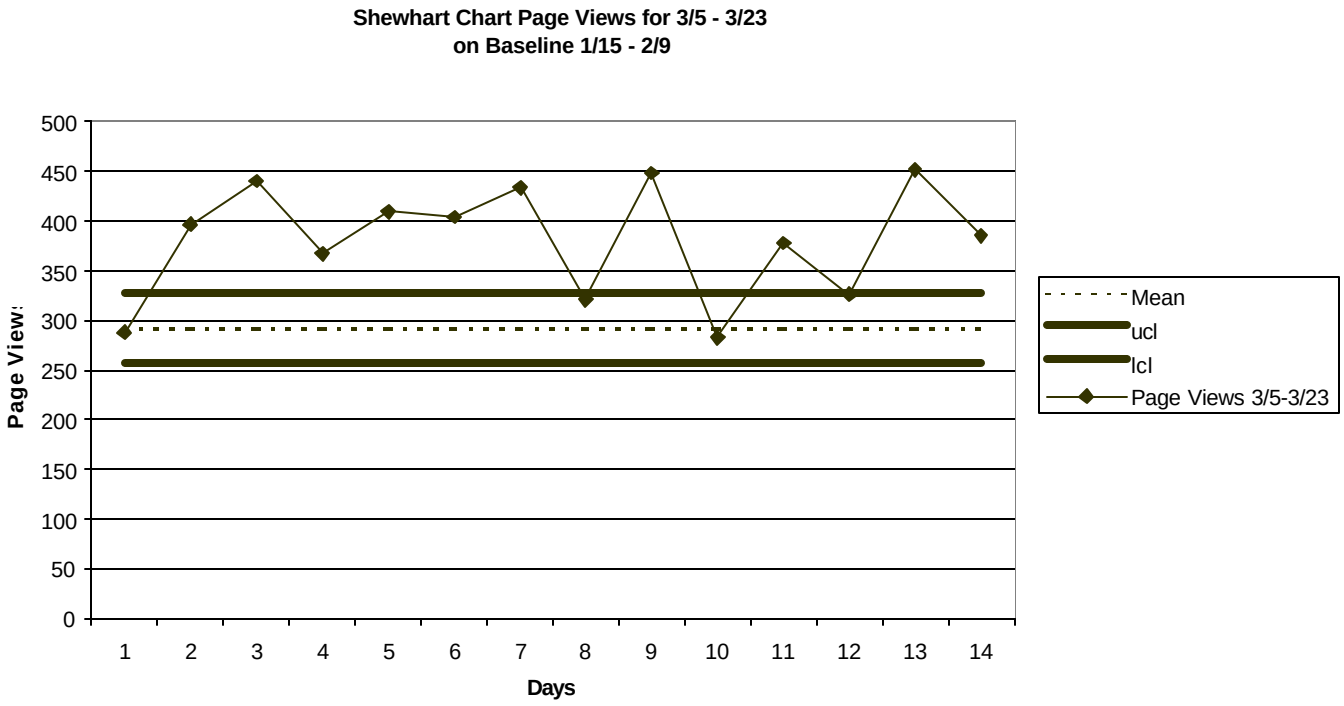
1. Alwan, L. C. and Roberts, H.V. (1988). Time-Series Modeling for Statistical Process Control, Journal of Business and Economics, 6(1), 87-95.
2. Duncan, A.J. (1986). Quality Control and Industrial Statistics, (5^{th} ed.). Irwin, Homewood, Ill.
3. Hart, M.K., Hart, R.F., Philip, G.C. (1990). Control Charts: False Indications of Lack of Control, Presented at the ASA Winter Conference, Orlando, FL, January.
4. Liebmann, L., Websidestory's Got Game, InternetWeek, Nov. 22, 39-47.
5. Mitra, D. Sahin, K.E., Sethi, R. and Silverschatz, A. (2000). New Directions in Services Management, Bell Labs Technical Journal, January-March, 17-34.
6. Nelson, L. S. (1982). Technical Aids: Control Charts for Individual Measurement, Journal of Quality Technology, 14(3), 172-173.
7. Roberts, H.V. and Tsay, R.S. (1990) Making Control Charts More Effective by Time-Series Analysis: Two Illustrative Applications, University of Chicago Graduate School of Business Working Papers. Presented at the Fifth Annual Conference Making Statistics More Effective in Schools of Business, Lawrence, KS.
8. Rogers, A. (1999) Keeping a Close Eye on Web Usage, Computer Reseller News, January, 25, 75.
9. Wardell, D.G., Moskowitz, H., and Plante, R.D. (1992) Control Charts in the Presence of Data Correlation, Management Science, 38(8), 1084-1105.
10. Western Electric Company (1958) Statistical Quality Control Handbook, Mack Printing Company, Easton, PA.

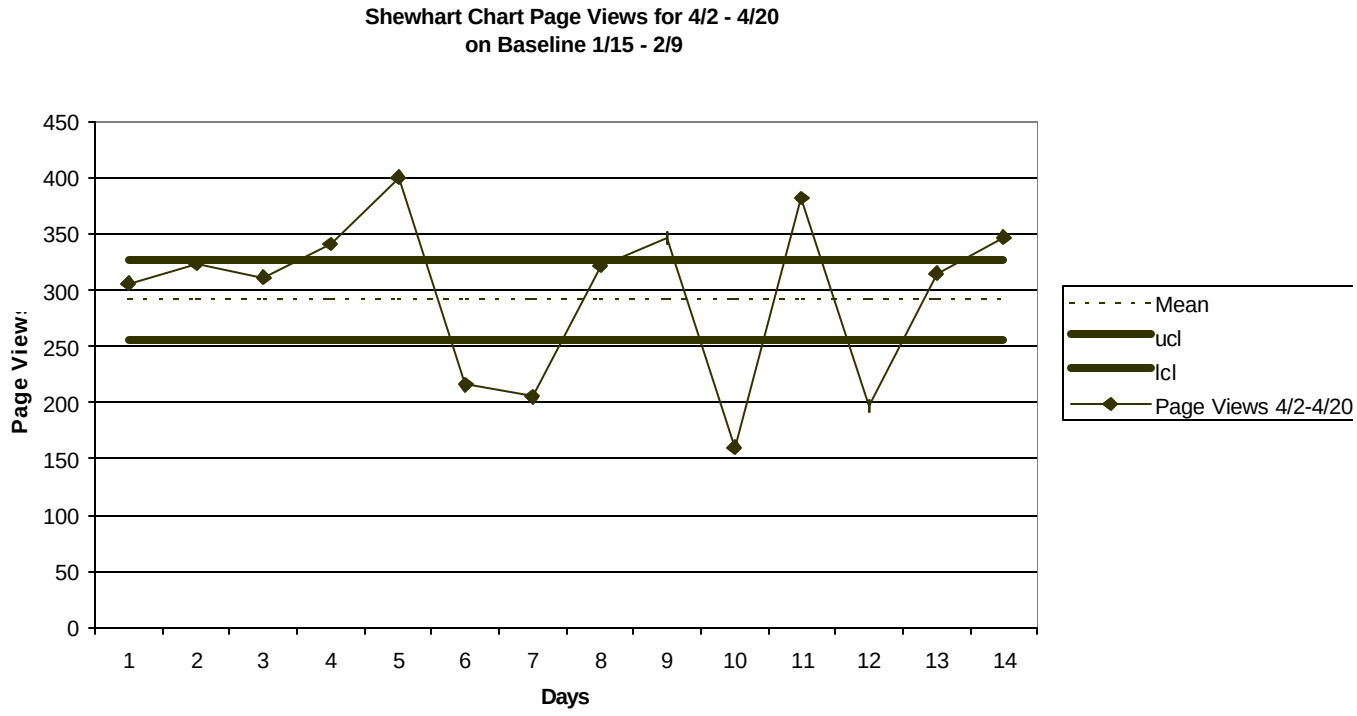
Table 1 & 2


Shewhart Chart %New Visitors for 3/5 - 3/23
on Baseline 1/15 - 2/9
% New Visitors
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
mean
ucl
lcl
% New 3/5-3/34


Shewhart Chart %New Visitors for 4/2 - 4/20
on Baseline 1/15 - 2/9
% New Visitors
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
mean
ucl
lcl
% New 4/2-4/20

Figures 3 & 4


Shewhart Chart Page Views for 3/5 - 3/23
on Baseline 1/15 - 2/9
Page Views
500
450
400
350
300
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Days
Mean
ucl
lcl
Page Views 3/5-3/23


Shewhart Chart Page Views for 4/2 - 4/20
on Baseline 1/15 - 2/9
Page Views
450
400
350
300
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Days
Mean
ucl
lcl
Page Views 4/2-4/20