

Understanding Control Charting: Techniques and Assumptions

Charles H. Darby, Health Services Evaluation, Mayo Clinic, Rochester, Minnesota

Abstract

Every process has some component of variation, and by measuring that variation directly via control charts it becomes possible to determine when that variation is attributable to randomness (typically termed common-cause variation), or to some identifiable cause (usually referred to as special-cause variation). Two such control charts--the proportion nonconforming or p-chart and the mean value or x-bar chart are important tools for quantitative efforts at quality improvement. This paper will address the components of control charting, the assumptions which form the basis for these chart's construction, the classes of control charts, techniques for quick access to the SAS-QC® Control Charting Module via SAS-ASSIST®, and how to recognize changes in chart patterns which might indicate "special cause" variation.

Introduction

Control charting has become a widely used technique in quality improvement since its introduction by Dr. Walter Shewhart of Bell Laboratories in the 1920's. Although other

techniques exist which are useful in quality efforts, control charts are among those which are most technically sophisticated. The earliest uses were in the monitoring of production and manufacturing processes, however in the most recent years these techniques have been introduced to practically every sector of our economy. Control charts display a single series of measurements taken in sequence from a process. By viewing the data in a set of control charts, investigators can gain a better understanding of variation in a given process or set of processes.

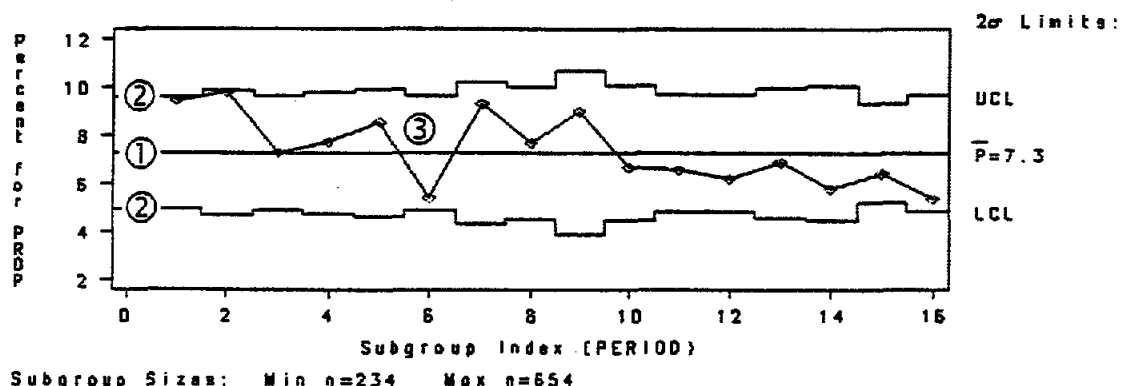
Charting Components

At a minimum, each constructed control chart has the following components:

1. The center line,
2. Upper and Lower Control Limits
3. Data points and line, and
4. Warning Limits (in some instances).

These components are highlighted in Figure 1 below.

Figure 1-- Control Chart Components



Control Chart Initiation in SAS-QC

P-Charting

At the first SAS-ASSIST screen, choose Data Analysis, Quality Control, and Control Charts. Identify the data set containing the data values, and select Proportions of nonconforming items under the Chart for Attributes listing on Types of Control Charts Menu. The system will prompt for a fixed subgroup size for all observations in the and for the subgroup variable name whose values describe the components to be averaged for each data point.

SAS will also prompt for the process variable name (in this case the proportion variable in the p-chart) and one of the four methods identified previously to obtain the control limits. If this data begins a new process measurement, I would recommend the first choice--compute control limits from the data. Options may be selected on the Additional Options screens. The options screen entitled Tests for Special Causes identifies Zone A, B, and C. These keywords describe values within 3-, 2-, and 1-sigma distances of the center line respectively.

Once the options and customization choices are completed, click on the Run button to prepare the chart. If the value of the center line is "in the neighborhood" of 0 or 1, and the process variability is high, or limits are set broadly, then the possibility exists that an upper or lower control limit will exceed one or fall below zero. The SAS system will automatically adjust these limits to reflect these limiting values should this occur. This should be the only occasion in which the p-chart limits are non-symmetric with respect to the center line.

X-Bar Charting

The process for x-bar charting initiation is virtually identical to that for p-charting. Select Mean and Range Charts under the Type of Control Chart button. The range chart which is paired to the mean (x-bar) chart indicates the difference between the minimum value of the process variable and the maximum value in each

subgroup. The range chart is considered to be a subsidiary chart to the x-bar, but can be used to profile how variability itself changes between identified subgroups. Again, once chart options have been selected (if any), choose the Run button to prepare the chart.

Pattern Recognition

Once statistical control has been established, if special-cause variation is present, some type of non-random pattern will emerge from the sequence of points which are formed from the subgroups. For charts having 2-sigma (3-sigma) control limits, any one of the patterns in Table 2 below provides sufficient justification to search for one or more assignable causes. Items A, B, C, and D have will occur by chance alone approximately 1 in 20 (1 in 300) times for normally distributed data.

Table 2--Patterns indicative of "special cause" variation for 2-sigma and (3-sigma) control limits.

- A. Any single point outside a plotted 2-sigma (3-sigma) control limit,
 - B. A run of at least 5 (8) points showing an increasing trend or a decreasing trend,
 - C. Any 3 of 5 (4 of 5) consecutive points falling outside either 1-sigma control limit (in Zone B or further),
 - D. Any run of 5 (15) consecutive points on either side of the center line.
 - E. Any unusual or non-random pattern in the data (such as a cyclical or rapid swinging pattern).
-

See Figures 2A-2E for examples of these patterns. For additional information, see Reference [5] beginning on page 117. Once chart stability has been verified and a non-cyclical trend of data is observed, consider using test of hypotheses to verify the trend's significance. If verified, it will be appropriate to recast the chart on a new center line and control limits for continued monitoring.

process had attained "statistical control" and future measurements will inform appropriately.

All control limits (including warning limits) may be set in SAS in one of four ways:

1. Computed directly from the data,
2. Manually entered based on process mean and standard deviation,
3. Loaded from an existing data set, or
4. Manually entered as LCL, UCL and/or average with multiple sigma values.

The prompts within SAS-ASSIST guide users to make one of these selections and to allow selection of other processing and display parameters.

Charting Assumptions

In designing any control chart for monitoring a process, three assumptions about the process measurement are made. The first is that the process from which the data is being taken is stable, (i.e. that the data are independent of each other and identically distributed in each subgroup).

The second assumption is that real-valued data is at least approximately normal in distribution, or in the case of binomial (yes/no) data, that the data is well-approximated by the normal distribution. If these assumption are not supported by the first 25-30 values taken from the process, additional steps may be necessary to adjust the data so that this expectation is justified. See References [1], [2], and [3] for additional information.

The third assumption is that a sampled component from the process is included in only one sampling subgroup. If the same component can be selected so as to be included in more than one subgroup, autocorrelation will be introduced into the chart.

Autocorrelation, depending on the extent to which multiple subgroups have the same data, can range in severity from a minor problem, to a major one. If present, the chart will display greater variability than is exists in the process. If severe, it could lead a user to believe

that the process is out of statistical control more often than it really is.

For example, suppose we constructed an x-bar chart with 3-sigma control limits. We would normally expect that about 1 subgroup in 300 would fall outside the control limits purely by chance. Severe autocorrelation might make this series of measurements appear "out of control" perhaps three to five times as often. For additional information see References [3], or [4].

Control Chart Styles

There are two broadly-defined classes of control charts. One class, known as Variable Control Charts track some real-valued process. Examples might include the length of a manufactured component or the time-to-failure of an electrical circuit. The most widely used variable chart is the x-bar chart.

The second class of charts is the Attribute Control Chart. A common chart in this class, the p-chart plots a subgroup proportion of items which conform or fail to conform to a defined standard. Examples would include the proportion of parts meeting a certain number of quality characteristics, or the proportion of patients improving after a medical procedure has been performed.

Attribute charts in general lack the sensitivity to process change that variable charts provide for equivalent sample sizes. However, in many instances, attribute charts are a better choice due to economic or time constraints. Setting up variable control charting sampling may require additional expense and time for prototyping data acquisition methods. For additional details see Reference [5].

The center line of a control chart can be thought of as describing the true or estimated process mean. Frequently, little is known about the process to be measured, so a choice is made to estimate the true process mean typically as the weighted average of the data contributing to the first 5 to 8 data points. In the SAS® system, each data point represents the average of the process variable values in groups defined by values of the subgroup variable. Table 1 shows how SAS determines subgroup values for a p-chart.

Table 1-SAS System Variable Interpretation for P-Chart Construction Example.

Period	Sample Size	Defectives	%
1	350	67	19.1%
2	435	54	12.4%
3	254	32	12.6%
4	376	75	19.9%
5	254	32	12.6%
6	374	63	16.8%
7	387	55	14.2%
8	406	48	11.8%

Period describes the sequential subgroup for plotting. Sample size is the value on which control limits are determined, the ratio of defectives to sample size produces the percentage which is the process variable.

Once plotted, these data values are then connected by a series of line segments to form the data line. When completed to a given measurement, trends in the chart may become apparent. It is these trends which allow inference on the effect of a process improvement effort. For clarity purposes, each process variable of interest in a data set is plotted on its own control chart.

The Upper (UCL) and Lower (LCL) Control Limits are set as thresholds. Should data points fall at or beyond either limit, "Special Cause" variation is indicated. As seen in Figure 1, the UCL and LCL are parallel and symmetric with respect to the center line. When the center line itself is near an upper or lower tolerance or specification limit, the control limits may not be equidistant from the center line. This concept will be further discussed in the section on p-charting. Control limits generally need to be chosen with care. Limits which are too close to

the center line will falsely indicate that a process is "out of statistical control" and limits which are set too distant from the center line may falsely indicate that a process is "in control" when it is not.

Typically, new control chart limits are set at some multiple of the process sampling standard error (in the SAS-QC module standard errors are denoted as sigmas). This multiple is usually based upon the size and frequency of the samples intended to be plotted on the chart. As sample size and/or frequency of sampling increase, the sensitivity of a given chart to process change also increases. Small to moderate sample sizes may be best plotted on charts having control limits set at 2-sigmas. Three-sigma charts are most useful when the samples are large or fairly frequent (minutes, hours, days).

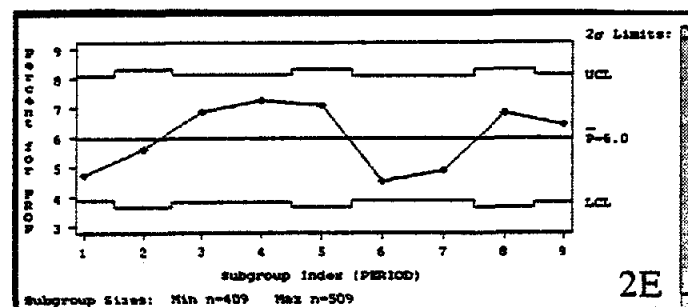
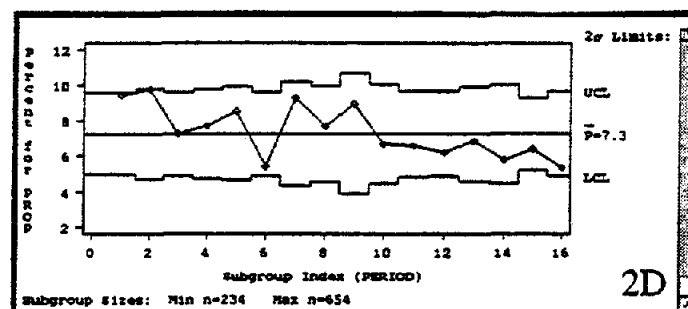
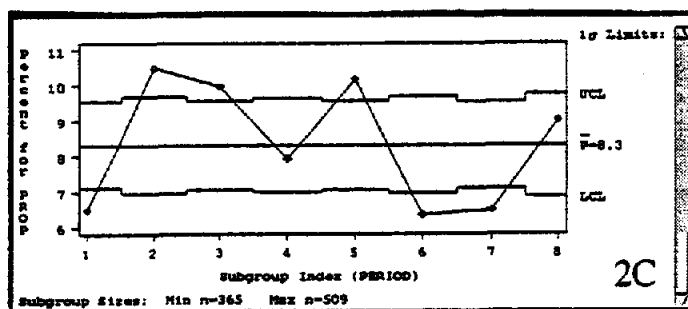
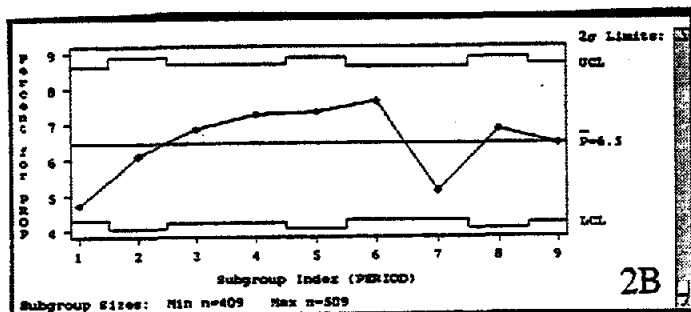
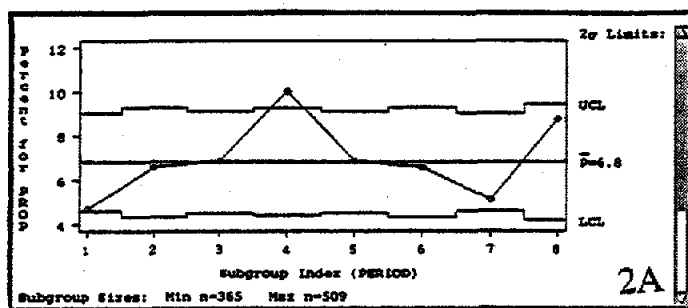
In ongoing control charting, increased sensitivity may also be obtained by plotting warning limits within the control limits. These are usually set at 1-, and or 2-sigmas from the true or estimated process mean.

By selecting a large sample more information about a process becomes available, so control limits can be more precise. Control limits will vary across subgroups in inverse proportion to the sample size. That is, the larger the sample, the closer the control limits will be to the center line. For example, suppose a p-chart control limit is set at 2-sigmas. If the subgroup size is 30, its control limits will appear more narrow than the same chart in which the subgroup size is 15.

Every new process which is measured via a control chart method requires an initialization period. Five to eight data points are usually required as a minimum to establish that the process is in "statistical control". That is, until a center line can be determined and control limits set, process data trends should remain suspect. Additionally, until the pattern of data points appears to stabilize, no inference about changes in process should be made.

Based on visual inspection of the first 5-8 subgroups, if the pattern appears random and the data points do not exhibit a regular pattern identified in the Pattern Recognition section below, then we can reasonably assume the

Figure 2--Chart Patterns described in Table 2



Conclusion

Control Charting for Attribute and Variables Data is an endeavor which relies on proper construction and administration techniques. This paper addressed some of these techniques: control charting elements, chart classes and concerns, sample size effects, and pattern recognition.

Understanding how increasing sample sizes promotes improvement in control charts ability to detect real changes and how data autocorrelation can influence chart performance is vital to reliable inference making. Understanding key patterns in control charts improves our abilities to detect change when it does occur. Inferences made about a process can only be as reliable as are the data which support that inference.

References

- [1] Alwan, Layth C., and Roberts, Harry V., The Problem of Misplaced Control Limits, Journal of Applied Statistics, Vol. 44, No.3, pp-269-278, 1995.
- [2] Caulcutt, Roland, The Rights and Wrongs of Control Charts, Journal of Applied Statistics, Vol.44, No. 3, pp.279-288, 1995.
- [3] SAS Institute, Inc., SAS/QC® Software: Usage and Reference, Volumes 1 & 2, Version 6, Cary, NC., SAS Institute, Inc., 1995.
- [4] Alwan, Layth C., Effects of Autocorrelation on Control Chart Performance, Communications in Statistics--Theory and Methods, V.21, No.4, pp. 1025-1029, 1992.
- [5] Montgomery, Douglas C., Introduction to Statistical Quality Control, Second Edition, John Wiley & Sons, New York, 1991.

Acknowledgements

SAS is a registered trademark of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration.

Author

The author may be contacted at:

Charles H. Darby
Health Services Evaluation
Mayo Clinic
200 First St. SW
Rochester, MN 55905
(507) 284-1230
E-mail: darby@mayo.edu