

## Improved Control Charts for Attributes

---

David B. Laney\*

3109 Brookhill Drive, Birmingham, AL 35242

### ABSTRACT

*The classic control charts for attribute data (p-charts, u-charts, etc.), are based on assumptions about the underlying distribution of their data (binomial or Poisson). Inherent in those assumptions is the further assumption that the “parameter” (mean) of the distribution is constant over time. In real applications, this is not always true (some days it rains and some days it does not). This is especially noticeable when the subgroup sizes are very large. Until now, the solution has been to treat the observations as variables in an individual’s chart. Unfortunately, this produces flat control limits even if the subgroup sizes vary. This article presents a new tool, the p’-chart, which solves that problem. In fact, it is a universal technique that is applicable whether the parameter is stable or not.*

**Key Words:** Attribute control chart; p-chart; u-chart; Individuals chart; Batch-to-batch variation

### THE PROBLEM

In control charts for attributes (p-, np-, u-, and c-charts), the standard deviation is obtained from a formula based on the overall mean. It is *assumed* that such data come from a binomial or Poisson probability distribution. For instance, in a p-chart:

$n_i$  = sample size, subgroup  $i$  ( $i = 1, \dots, k$ )

$x_i$  = number of occurrences of the attribute of interest

$p_i = x_i/n_i$

$$\bar{p} = \sum x_i / \sum n_i$$

$$\sigma_{p_i} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_i}}$$

$$\text{CL} = \text{Center Line} = \bar{p}$$

$$\text{UCL/LCL} = \text{Upper/Lower Control Limits}$$

$$= \bar{p} \pm 3\sigma_{p_i}$$

Note that if the subgroup sizes vary, the control limits are different.

\*E-mail: david\_laney@bellsouth.net

Unfortunately, we sometimes forget a critical assumption in this approach: the distributional parameters (the underlying probabilities) must remain *fixed over time*. We assume we always roll the same pair of dice. In real processes, this is seldom true—the dice often change over time. In a telephone company, for instance, whether there is rain or lightning today will have a profound effect on the repair work today and tomorrow. A chart over many days might suggest that rainy days exhibit special causes of variation. In the Sahara perhaps! For most of us rainy days are part of our normal existence (common causes of variation). Experience suggests that this is a pervasive problem, especially among service industries, where total process variation is so heavily influenced by environmental and other external influences, and where subgroup sizes are often quite large (in the thousands or even millions).

So long as the plotted points are based on samples of only a few hundred, this is seldom a problem; the uncertainty due to sampling error is usually so large that any variation in the parameter over time is practically invisible. (This is why applications in manufacturing, where subgroup sizes are usually small, have failed to reveal this phenomenon.) The problem with the distributional assumption is usually seen only when the sample sizes are extremely large, rendering sampling error virtually nonexistent. Then the “batch-to-batch” variation [(1), pp. 157–165] is almost all there is. Unfortunately, the classical control limit formulas depend entirely on sampling variation, so when the sample sizes are very large, the limits

squeeze in toward the center line of the chart. If the data points have variation in excess of the assumed probability distribution, they will “hang out.” Jones and Govindaraju (2) call this “overdispersion.” For example, see Fig. 1.

The data for Fig. 1 were based on subgroups averaging over 7600 observations each. The control limits react to that number by tightening around the center line, but there is obviously a lot more variation present. *When most of the data are outside the control limits, how can we say it is due to special causes?* The large sample sizes do not cause this phenomenon; they merely reveal it. Larger sample sizes improve the sampling precision of our estimates, thus revealing more clearly the invalidity of the binomial (or Poisson) assumption.

In private correspondence on this subject, Roger Hoerl described this very well: “With very large sample sizes (in the thousands, for example) the statistical uncertainty associated with within-sample variation is ‘averaged out’ by the large subgroup size, resulting in limits that are right on top of each other. This is not a problem theoretically. Practically, however, since no process is perfectly stable, the chart is so sensitive that most points fall outside the limits. They are statistically detectable, but represent such minor shifts or trends that they are not worth going after. Of course, Shewhart’s original purpose was to segregate special causes that were economical to remove, so in this case, the chart loses its original motivation.”

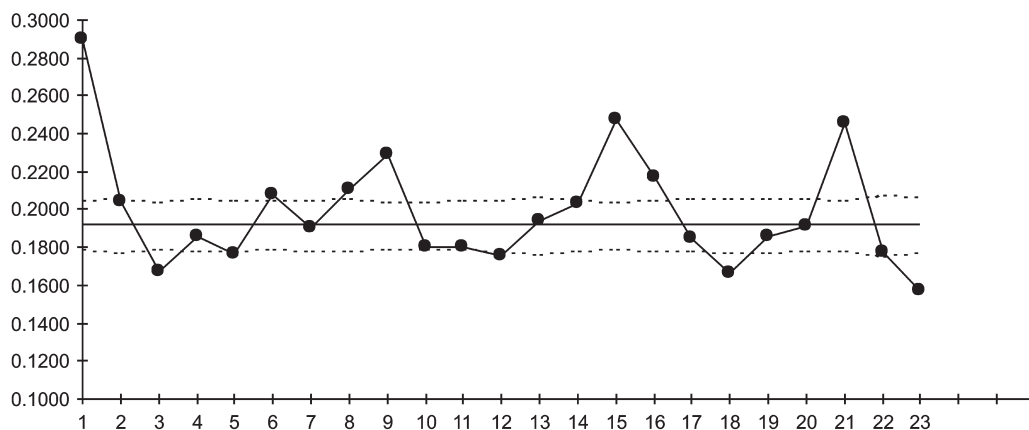


Figure 1. Heimann’s Fig. 5 (p-chart).

**THE USUAL REMEDY: THE X-CHART**

The usual solution to this problem is to ignore the fact that the data are attributes, and simply plot them with an individual's chart or X-chart. Justification for this practice can be found in Ref. (3), pp. 196–197; Ref. (4), p. 136; and Ref. (1), p. 259. It is done as follows:

$$R_i = |p_i - p_{i-1}| \quad (i = 2, \dots, k)$$

$$\bar{R} = \frac{1}{k-1} \sum_{i=2}^k R_i$$

$$\sigma_p = \bar{R}/1.128$$

$$CL = \bar{p}$$

$$UCL/LCL = \bar{p} \pm 3\sigma_p = \bar{p} \pm 2.66\bar{R}$$

Figure 2 shows the same data from Fig. 1 displayed in an X-chart. Clearly, these control limits make a lot more sense.

The data for these examples came from Peter Heimann at AT&T in Ref. (5), which represents a diagnostic test to help us know when the use of an X-chart is warranted.

**A NEW APPROACH: THE p'-CHART**

The problem with the X-chart is that the resulting flat control limits no longer account for the effects of varying subgroup sizes. A new instrument, the *p'*-chart developed here, solves this problem. The data from the

previous example are shown in a *p'*-chart in Fig. 3. Note that the control limits are about where they were in the X-chart, but now they vary with changes in subgroup sizes, as one would expect.

Is this a big problem? As it happens, it was a major problem for another data set in Dr. Heimann's paper. Figure 4 shows the original p-chart; Fig. 5 the corresponding X-chart. Precisely because of the wide fluctuations in subgroup size, Dr. Heimann was unable to decide which chart was better. Is point 15 in or out of control?

In a *p'*-chart (Fig. 6), the necessity of variable control limits becomes obvious. Point 15 is definitely out of control. The very fact that it was based on a relatively large sample helps to show that its distance above the center line is indeed significant.

The *p'*-chart does not have to choose between intra-subgroup variation (as in the p-chart) or inter-subgroup variation (as in the X-chart). It uses *all* the variation in the data. If there is any batch-to-batch variation, its control limits are appropriately farther away from the center line than in a p-chart. In addition, if there is a variation in subgroup sizes, its control limits will vary, unlike the X-chart. The following describes how the *p'*-chart is constructed.

**STEP 1: THE Z-CHART**

A standard method for handling attribute data, usually when the display of variable control limits is undesirable, is to convert each *p*-value to a *z*-score (the number of sample standard deviations between that point and the overall mean), and then plot these numbers on a "z-chart" [(3), p. 197]. Since the theoretical mean of the *z*-scores is

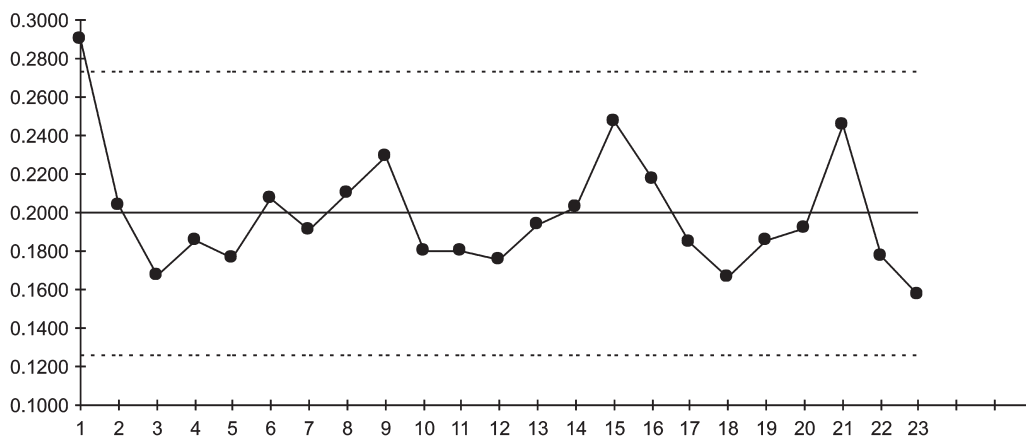


Figure 2. Heimann's Fig. 5 (X-chart).

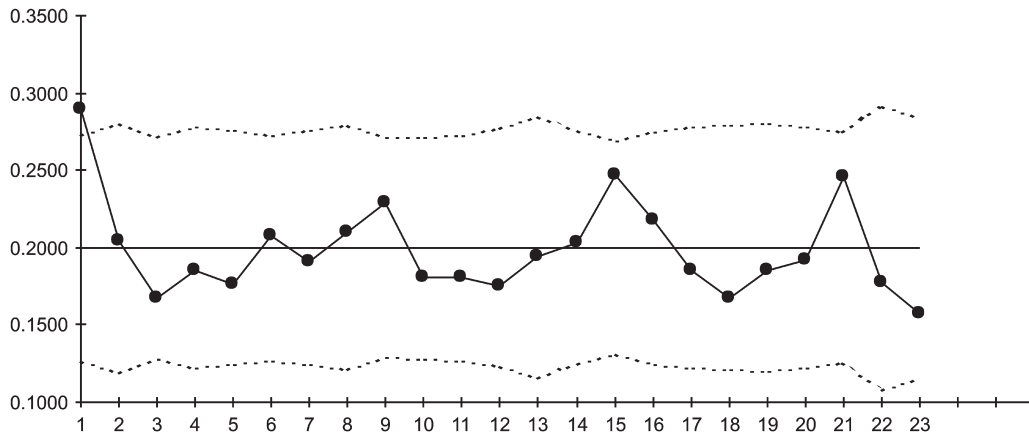


Figure 3. Heimann's Fig. 5 (p'-chart).

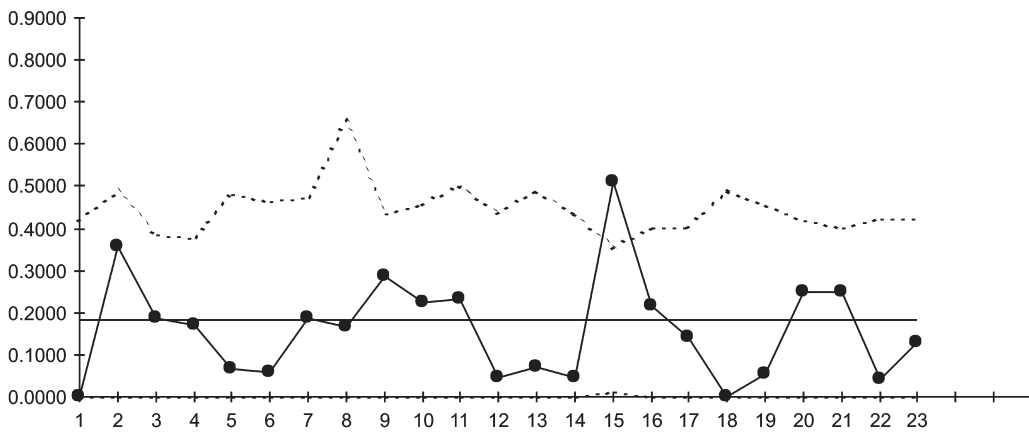


Figure 4. Heimann's Fig. 7 (p-chart).

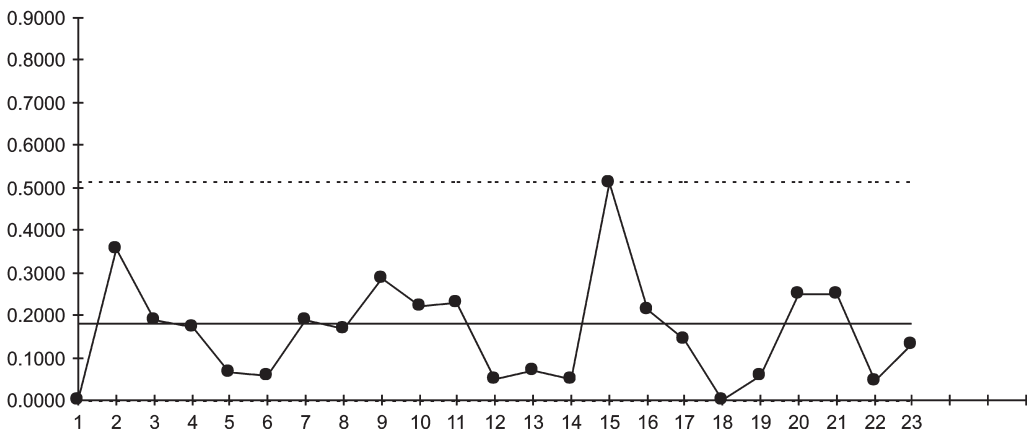


Figure 5. Heimann's Fig. 7 (X-chart).

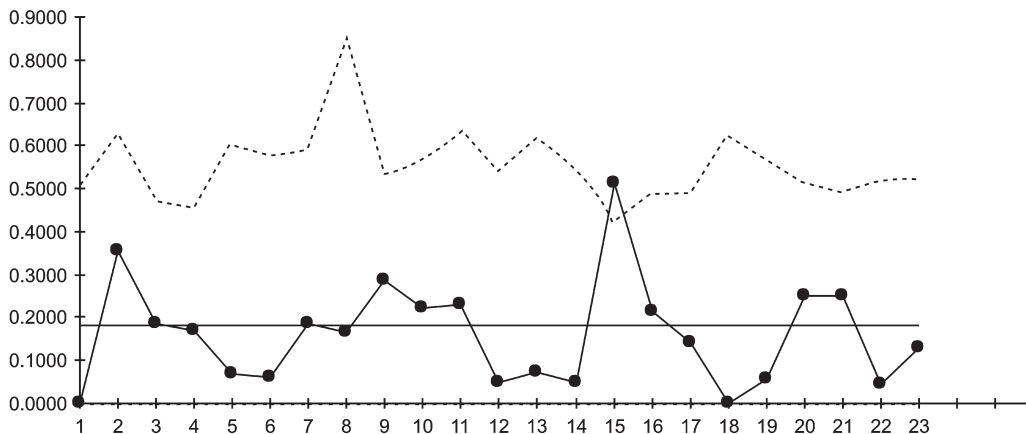


Figure 6. Heimann’s Fig. 7 (p’-chart).

zero, this is what is used for the center line of the chart. Moreover, since the standard deviation of  $z$  is *assumed to be unity*, the control limits are set at  $+3$  and  $-3$ . The “ $z$ -transformation” automatically adjusts each point for its unique intra-subgroup variation, thus producing flat control limits:

$$z_i = \frac{p_i - \bar{p}}{\sigma_{p_i}}$$

$$CL = 0$$

$$UCL/LCL = \pm 3$$

But is the “unit variance” assumption valid? Since that assumption relies solely on the intra-subgroup variation, it is *not* correct when batch-to-batch variation is present. As Dr. Donald Wheeler has asked in numerous books and speeches, “Why *assume* the variation when you can *measure* it?”

**STEP 2: AN IMPROVED Z-CHART**

What we should do is put together the concepts of the X-chart and the  $z$ -chart: Convert the  $p$ -values to  $z$ -scores (thus correcting *in advance* for variable sample sizes) and then plot the  $z$ ’s in an individuals chart:

$$R_i = |z_i - z_{i-1}| \quad (i = 2, \dots, k)$$

$$\bar{R}' = \frac{1}{k-1} \sum_{i=2}^k R'_i$$

$$\sigma_z = \bar{R}' / 1.128$$

$$CL = 0$$

$$UCL/LCL = \pm 3\sigma_z$$

We no longer *assume* that the standard deviation of the  $z$ -values is equal to one. We *measure* it to find out what it actually is.

**THE FINAL STEP**

All that remains now is to unravel the  $z$ -transformation and put our results back into the meaningful units of the “ $p$ -plane”:

$$z_i = \frac{p_i - \bar{p}}{\sigma_{p_i}}$$

$$p_i = \bar{p} + \sigma_{p_i} z_i$$

$$sd(p_i) = \sigma_{p_i} \sigma_z$$

$$CL = \bar{p}$$

$$UCL/LCL = \bar{p} \pm 3\sigma_{p_i} \sigma_z$$

From this last line, we can see what  $\sigma_z$  really is. It is the relative amount of process variation not explained by the binomial assumption alone. As  $n$  increases, the variation due to sampling diminishes, thus making the

batch-to-batch component relatively larger. That is why applications with large subgroup sizes reveal this situation very often.

In the first example (Figs. 1 and 2),  $\sigma_z$  was 5.6—there was 460% more variation in the data than a classical p-chart could account for. In the second (Figs. 3–5), it was only 1.4, but enough to justify abandoning the p-chart to allow for the 40% additional variation in the process. In addition, this case illustrates that the use of variable control limits makes the p'-chart superior to the X-chart.

What this method does was well described by Roger Hoerl on seeing an early draft of this manuscript: “We may choose to define both within-sample and also the variation between one sample and those immediately before and after it (i.e., using moving ranges) as common cause variation. In this case, only variation that is above and beyond the normal point-to-point variation shows up as special causes. The key point here is that we have consciously changed the definition of common and special cause. This is OK, as long as we realize we have done it. In such cases where there are unequal subgroup sizes, your approach seems to be the most logical thing to do. I like it!”

### SOME OBSERVATIONS

1. If a data set is in fact binomially distributed (no batch-to-batch variation is present),  $\sigma_z = 1$ . Then the p'-chart is the same as the p-chart.

2. If all subgroup sizes are equal, the p'-chart is the same as the X-chart. Therefore, the p'-chart is merely an extension of the X-chart to the case with varying subgroup sizes.

3. While this discussion has examined the binomial case (p-chart), the extension to the Poisson case (u-chart) is very simple. Just replace  $\sigma_{p_i} = \sqrt{\bar{p}(1 - \bar{p})/n_i}$  with  $\sigma_{u_i} = \sqrt{\bar{u}/n_i}$ .

4.  $\sigma_z < 1$  (“underdispersion” per Jones and Govindaraju) shows that there is positive autocorrelation in the data. The classical limits of p- and u-charts will be too wide.

5. As stated above, Heimann (5) presented a diagnostic test to determine when the problem exists. More recently, Jones and Govindaraju (2) gave us another insightful article on this subject. Unfortunately, neither paper went beyond diagnosis to cure. The only remedy remained the X-chart.

6. The only attempt at a cure (known to this author) was offered by Wheeler and Poling (6), suggesting

control limits given by:

$$UCL/LCL = \bar{p} \pm 2.66\bar{R}\sqrt{\frac{\bar{n}}{n_i}}$$

This simple adjustment term affixed to the common X-chart formula does adjust the levels of the limits according to the varying subgroup sizes. Unfortunately, the very existence of such differences suggests that the average moving range (which gives equal weight to all its constituents) is a biased estimate of process variation. The p' chart does not have that problem.

### SUMMARY

In control charts for attributes, the presence of batch-to-batch variation suggests the use of X-charts instead of p-charts (or u-charts). Such additional variation (over and above that which the binomial and Poisson assumptions can detect) is widespread, especially in service company applications. Further, it is in those very cases that subgroup sizes are often quite large, a situation in which the presence of inter-subgroup variation is clearly revealed. Unfortunately, the X-chart ignores the intra-subgroup variation and is therefore unable to produce variable control limits when the subgroup sizes vary. The p'- and u'-charts, by addressing *both* sources of variation, put the control limits in the right place *and* show how they vary with subgroup sizes. Since the X-chart has long been the method of choice when there is batch-to-batch variation, since the p'- and u'-charts are merely extensions of the X-chart to handle the case of varying subgroup sizes, and since the p'- and u'-charts are the same as the p- and u-charts when there is no batch-to-batch variation, the p'- and u'-charts should be universally adopted as the new standards for plotting attribute data.

### ACKNOWLEDGMENTS

My thanks to Larry Aft (Southern Polytechnic State University), Forrest Breyfogle (Smarter Solutions, Inc.), Roger Hoerl (General Electric), Keith Johnson (Bell-South), Thomas Pyzdek, John Ramberg (University of Arizona), Donald Wheeler (SPC, Inc.), and William Woodall (Virginia Polytechnic Institute and State University) for their helpful suggestions and encouragement in this work.

### ABOUT THE AUTHOR

Mr. Laney is the Director of Statistical Methodology at BellSouth Corporation. He is a senior member of the American Society for Quality, a Certified Quality Engineer, and a Certified Six Sigma Black Belt.

### REFERENCES

1. Wheeler, D.J. *Advanced Topics in Statistical Process Control*; SPC Press: Knoxville, TN, 1995.
2. Jones, G.; Govindaraju, K.A. Graphical Method for Checking Attribute Control Chart Assumptions. *Qual. Eng.* **2000**, *13* (1), 19–26.
3. Western Electric. *Statistical Quality Control Handbook*; AT&T Customer Information Center: Indianapolis, IN, 1956.
4. Wheeler, D.J. *Understanding Variation—The Key to Managing Chaos*; SPC Press: Knoxville, TN, 1993.
5. Heimann, P.A. Attributes Control Charts with Large Sample Sizes. *J. Qual. Technol.* **1996**, (28), 451–459.
6. Wheeler, D.J.; Poling, S. *Building Continual Improvement*; SPC Press: Knoxville, TN, 1998.





