Control Charts for Number of Nonconformities

Dr. Austin Brown

Kennesaw State University

Introduction

- Last week, we introduced a Shewhart-style control chart for the fraction nonconforming, the p chart.
- Obviously, this can be quite useful in instances when the quality characteristic we are monitoring is inherently dichotomous (e.g., a student graduates within five years (conforming) or they don't (nonconforming)).
- However, a potential <u>limitation</u> to the widespread adoption of the p chart is the ease of its interpretability.

Introduction

- While one may not think the fraction nonconforming is a particularly difficult statistic to interpret, one that is perhaps more straightforward is the simple count or frequency of nonconformities in a given sample.

- A count is a positively valued integer. For small counts, the distribution tends to be positively skewed and thus reliance upon the CLT to use the $\bar{x}$ chart will likely result in a less efficient control chart (less efficient here meaning it will be slower to signal out-of-control when a shift has taken place).

- Thus, it would be of value to use a chart specifically designed for count data.

Introduction

- Fortunately for us, we have a few options. The first, the np chart, works much like the p chart whereas the C and the U chart are used in a slightly different way.

- We will discuss the appropriate situations to use each one as well as potential limitations of their implementation.

The *np* Chart

- Perhaps the most straightforward control chart for counts is a simple extension of the p chart.
- If I know the fraction nonconforming is p, then I also know that the number of nonconformities would simply be my sample size n times p, or np.
- Much like the p chart, we will take samples of size n at a particular time point and count the number of units which are nonconforming.

The *np* Chart

- Okay, this seems easy enough, but how do we calculate control limits?
- We already know the sampling distribution of $\hat{p}$:

$$\hat{p} \dot{\sim} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

The *np* Chart

- If we multiply a normally distributed random variable by a constant, say n, we know from math stats that:

$$E[n\hat{p}] = nE[\hat{p}] = np$$

$$Var[n\hat{p}] = n^2 Var[\hat{p}] = n^2 \frac{p(1-p)}{n} = np(1-p)$$

The *np* Chart

- Therefore:

$$n\hat{p} \dot{\sim} N\left(np, \sqrt{np(1-p)}\right)$$

- Knowing this, if p is known, we can create Shewhart-style control limits:

$$UCL = np + 3\sqrt{np(1-p)}$$

$$LCL = np - 3\sqrt{np(1-p)}$$

The np Chart

- However, most of the time, we aren't going to know p, so we have to estimate it using Phase I sample data.

- Following the same procedures are before, we will take m samples of size n and calculate an estimate of np:

$$n\bar{p} = \frac{n}{m} \sum_{i=1}^{m} \hat{p}_i$$

The *np* Chart

- Then, we simply replace the np's in the control limits with whatever we calculated for the estimate:

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$$

The *np* Chart

- Let's go through the examples from last week, except using the *np* chart instead of the *p* chart.
- One thing you'll notice is that the behaviors exhibited by the *p* charts, specifically in the orange juice cans examples, are identical to those exhibited by the *np* charts.
- The only difference now is a slight difference in the interpretation. We're monitoring the number of nonconformities per sample rather than the proportion of nonconformities.

The *np* Chart

- Let's think about what exactly the *np* chart is doing:
- We take a sample of n units and our plotting statistic (say X), is the subset of those n units which do not conform to whatever standard.
- In some cases, like the purchase order example from last week, nonconformities aren't always the same. In fact, in the purchase order example, we're omitting key information by only considering whether the PO is acceptable or not.

The *C* Chart

- Specifically, we're not considering the number of errors on a given PO.

- If there was a single error, we said it was nonconforming. But it is likely of interest to us to differentiate between nonconforming POs with a single error (like a missing signature or a small typo) and those with multiple errors.

- In such a situation, what we're actually interested in monitoring is the *number of nonconformities per unit*, not the fraction nonconforming or the number of nonconformities in a sample. This leads to the necessity of a new chart called the *C* chart.

The C Chart

- Typically when we're monitoring counts, we can make use of the Poisson distribution as an adequate model for calculating probabilities.
- Here, we have a couple of basic assumptions:
  1. The sampling unit (PO, car door, etc.) is consistent from sample to sample
  2. The probability of observing a nonconformity is fixed from sample to sample

The C Chart

- If we let X be a Poisson random variable, and the mean number of nonconformities per unit is C. Then, the probability of observing x number of defects is:

$$P[X = x] = \frac{e^{-C} C^x}{x!}$$

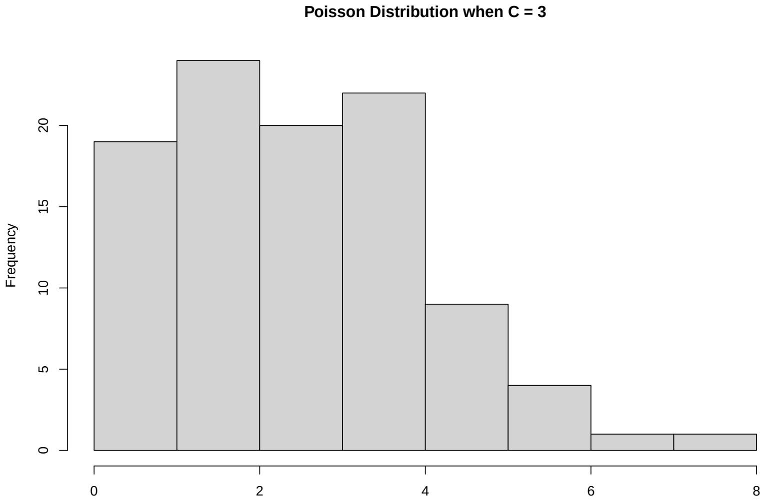
- From here, if C is known, we can calculate Shewhart-style control limits.

$$UCL = C + 3\sqrt{C}$$
$$LCL = C - 3\sqrt{C}$$

The C Chart

- One practical consideration with the C chart is the control limit width. So far, we've mostly been using 3 σ control limits.
- However, in mostly all of those situations, we were working with a variable assumed to be normally distributed.
- The Poisson distribution, for a small value of C, is positively skewked.

The C Chart



The C Chart

- Resultingly, it is sometimes more appropriate to use α derived limits (like what we did for the s^2 chart) rather than σ based limits.

- Remember from previous sessions that to calculate α derived limits, we want to choose them such that:

$$P[LCL \leq X < UCL|IC] = 1 - \alpha$$
$$\implies P[X < UCL|IC] - P[X \leq LCL] = 1 - \alpha$$

The C Chart

- So using software, we can figure out what our values of UCL and LCL ought to be for a specified α and known C.
- As before, though, we most likely won't know what the mean number of nonconformities per sampled unit is, and will have to estimate it from Phase I data.

The C Chart

- By now, we know the drill. We will take m samples and calculate the average number of nonconformities across those m samples:

$$\bar{C} = \frac{1}{m} \sum_{i=1}^{m} C_i$$

- As you may recall from math stats, $\bar{C}$ is both the maximum likelihood estimator as well as the method of moments estimator for C. So now our 3 σ control limits are:

$$UCL = \bar{C} + 3\sqrt{\bar{C}}$$
$$LCL = \bar{C} - 3\sqrt{\bar{C}}$$

The C Chart

- Let's say we manufacture circuit boards. Our sampling unit is not a single board, but rather 100 boards. We are taking $m = 26$ samples of 100 boards and counting up the number of nonconformities observed.

- These data are contained in sheet one of the Number of Nonconformities Datasets Excel sheet in D2L.

The C Chart

- Remember back from our first class session where we used Pareto analysis with our coffee shop data?
- This tool is a perfect compliment to the C chart as it can help us see what specific type of nonconformities we are primarily dealing with.
- Additionally, use of a cause-and-effect diagram can further aid in understanding specific parts of a process which are causing the nonconformities in our sampling units.

The C Chart

- Let's do another example. Let's say we are the chair of a statistics department which teaches several presumably identical introductory statistics courses.

- Let's say the same 10 item midterm exam is given to each class. We want to monitor, semester to semester, the number of incorrect responses per semester as a means of monitoring and controlling class consistency (obviously, there are a lot more factors at play here, but let's roll with it).

- Assuming that there are 20 sections taught each semester (including summer), data from the past 10 years are contained in the third sheet of the Number of Nonconformities Datasets Excel file.

The C Chart

- As we saw from the Phase II data, it appears that we have experienced an upward shift in the process mean, although we only have one point plotting slightly above the upper control limit.

- As we did in the last class, when we see a non-random looking pattern, we can use other statistical tools, like t-tests, to determine if a difference exists between what we determined to be in-control in Phase I to what seems to be out-of-control in Phase II.

The C Chart

- When we did this in a previous class, we assumed our plotting statistic was normally distributed. Here, we for sure know it isn't.

- However, a characteristic of the Poisson distribution is that as the rate parameter, C, gets relatively large, the shape of the distribution becomes bell-shaped.

- This is to say that for a large C ($>$ 30ish), it's probably okay to use the t-test. For small values of C the nonparametric Mann-Whitney U test would be more appropriate.

The C Chart

- One potential limitation of using the C chart would be the distributional assumption.
- We have two potential issues. The first one we've already discussed: for a small C, the Poisson distribution is asymmetric and thus using σ limits is inappropriate. For a very small C, we may run into the same issue we had before where $LCL = 0$. There is an easy fix for this (use α derived limits).
- The other issue is that of overdispersion. One of the unique characteristics of the Poisson is that its mean and its variance are equal to C.

The *C* Chart

- If in reality, $\sigma^2 > \mu$, then the process doesn't follow a Poisson distribution. We can also run into situations where $\sigma^2 < \mu$.
- In the literature, researchers have developed overdispersed/underdispersed Poisson control charts for such situations.
- While we won't get into them in this class, it's important to know that such charts exist (see Sellers, K. F. (2012). A generalized statistical control chart for over or under-dispersed data. Quality and Reliability Engineering International, 28(1), 59-65.).

The U Chart

- For the C chart, we are typically examining a single unit, whether that by a classroom, a refrigerator door, or what have you.

- However, why restrict our sampling to just a single unit? Instead of examining the number of nonconformities *per unit*, we could instead monitor the *average number of nonconformities per batch or sample* of size n.

- Here, we are increasing our sample size without necessarily losing information as we can still collect nonconformity data on a per unit basis. But this leads to the development of a slightly different chart referred to as the U chart.

The U Chart

- Suppose we take individual samples of size n (e.g., daily purchase orders). Let x denote the total number of observed nonconformities across all n units.

- Now let's define our plotting statistic as:

$$u = \frac{x}{n}$$

- Let's assume that x follows the sum of n independently and identically distributed Poisson random variables, with rate parameters C.

The U Chart

- Now, it can be shown that:

$$E[X] = nC \quad \text{and} \quad Var[X] = nC$$

- Then it is clear to see:

$$E[u] = E\left[\frac{x}{n}\right] = \frac{nC}{n} = C$$

$$Var[u] = Var\left[\frac{x}{n}\right] = \frac{nC}{n^2} = \frac{C}{n}$$

The U Chart

- Now, the control limits for a known C, we would use:

$$UCL = C + 3\sqrt{\frac{C}{n}}$$

$$LCL = C - 3\sqrt{\frac{C}{n}}$$

- But of course, we likely won't know C. So we'll have to estimate it using Phase I data.

The U Chart

- Let's take m samples of size n in Phase I. We'll define our estimate of C as:

$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m} u_i$$

- Thus, the control limits would just replace the parameter with the estimator:

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{n}}$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{n}}$$

The U Chart

- Let's look at an example. Let's say we're monitoring the supply chain of a company. Each week, we take a random sample of 50 shipments to monitor the number of errors, whether that be in documentation, actually product delivered, or what have you.

- These data are contained in the fifth sheet of the Number of Nonconformities Datasets Excel spreadsheet.

The U Chart

- One of the nice characteristics of the U Chart is that it works with variable sample sizes much in the same way that the $\bar{x}$ and p charts did.
- In the sixth sheet of this week's Excel workbook, we have data representing the average number of defects of a textile per 50 square meters of a roll of cloth. 10 rolls are inspected.

The U Chart

- Alternatively, like the p chart, we can also deal with variable sample sizes by calculating our control limits based on average sample size:

$$\bar{n} = \frac{1}{m}\sum_{i=1}^{m} n_i$$

- Or we can also standardize each u_i:

$$Z_i = \frac{u_i - \bar{u}}{\sqrt{\frac{\bar{u}}{n_i}}}$$

OC Curves for C and U Charts

- So far, we've learned about a few new types of charts for monitoring the number of nonconformities in a sample, including the *np*, *C* and *U* charts.
- We learned how to calculate Phase I control limits for each of the charts using R as well as how to take the Phase I limits and use them in Phase II monitoring.
- But just like with our other control charts, there are decisions for us to make in designing the chart to make them useful in efficiently detecting shifts of various magnitudes with high probability.
- As before, we'll be using our old friends, the *OC* curves.

OC Curves for the C and U Charts

- Remember, OC curves are built by modeling the probability of making a Type II error, that is, having a point plot inside the control limits when the process is really out-of-control.

- For example, with the C chart, assume the true process mean, c, has shifted from an in-control mean value, c_0, to a new value c_1.

$$P[LCL < X < UCL | c = c_1] = \beta$$

$$P[X < UCL | c = c_1] - P[X < LCL | c = c_1] = \beta$$

OC Curves for the C and U Charts

- Let's look at the circuit boards example from before.
- After removing the two OOC points, we estimated $\bar{c} = 19.67$, $UCL = 32.97$ and $LCL = 6.36$.
- Let's use R to help us do the rest of the heavy lifting.

OC Curves for C and U Charts

- The U chart is a little bit different. Technically, u is not distributed as a Poisson random variable, so we have to manipulate the formula for calculating β before to get it into a known distributional form.

- Remember $u = x/n$ and $x \sim POI(nC)$. Thus:

$$P[u < UCL|OOC] - P[u < LCL|OOC] = \beta$$

$$P[x/n < UCL|OOC] - P[x/n < LCL|OOC] = \beta$$

$$P[x < nUCL|OOC] - P[x < nLCL|OOC] = \beta$$

OC Curves for C and U Charts

- Because the control limits multiplied by n are likely not an integer, we practically round the LCL up and the UCL down.
- Let's take a look at an example of plotting the OC Curve using the supply chain data from before.