Shewhart Charts for Variables

Dr. Austin Brown

Kennesaw State University

Table of Contents

Introduction to Control Charts

- As mentioned in the last class session, one of the most powerful tools in quality monitoring is a control chart.

- A control chart is a graphical tool rooted in statistical theory, and all of which contain the same key elements:
  - A plotting statistic, which is a summary statistic derived from random samples of the monitored process
  - A target or in-control value that the process ideally adheres to
  - Control limits (sort of like critical values in hypothesis testing)

- A good control chart has two main desirable characteristics: (1) if the process has shifted away from target, then the control chart will signal to the operator in as few samples as possible, and (2) if the process is in-control, then the chart doesn't give false alarms too frequently

Introduction to Control Charts

▶ There are a variety of control charts which have been developed in order to monitor specific types of processes, and there is none more famous than the original control chart, the Shewhart $\bar{X}$-chart

▶ Walter Shewhart is the inventor of the control charts which bear his name and were developed during his work at Bell Labs in the 1920s. He was a friend and mentor to the godfather of quality management, Dr. William Edwards Deming.

Introduction to the Shewhart $\bar{X}$ and R Charts

▶ Typically when we're monitoring a process for a variable, we want to make sure that the process location (like the mean or median) as well as the process variability remain in a state of control.

▶ This is where the Shewhart $\bar{X}$ and R Charts come into play;
  ▶ The $\bar{X}$ chart monitors the process' mean
  ▶ The R Chart ("R" for range) monitors the process' variability

Introduction to the Shewhart $\bar{X}$ and R Charts

▶ Let's assume that the process we're monitoring follows a Normal distribution with in-control mean, μ_0 and in-control standard deviation, σ_0.

▶ At time point, t ($t = 1, 2, ...$) we take a random sample of size n and calculate the sample mean:

$$\bar{x}_t = \frac{1}{n} \sum_{i=1}^{n} x_i$$

▶ We know from the Central Limit Theorem that, so long as the random variable X has finite variance:

$$\bar{x} \dot\sim N\left(\mu_0, \frac{\sigma_0}{\sqrt{n}}\right)$$

Introduction to the Shewhart $\bar{X}$ and R Charts

▶ And thus we can calculate an upper and lower control limit like:

$$UCL = \mu_0 + L\frac{\sigma_0}{\sqrt{n}}$$

$$LCL = \mu_0 - L\frac{\sigma_0}{\sqrt{n}}$$

Center Line (CL) $= \mu_0$

Introduction to the Shewhart $\bar{X}$ and R Charts

▶ We can choose L the same way that we choose critical values for a two-sided Z-test – by specifying the α level we are willing to live with.

  ▶ Recall, α is the probability of making a Type I error at any given time point, which is the probability of observing a false alarm when the process is actually in-control.

▶ The choice of α is not an arbitrary one. Select a value overly large and you'll have a control limits that are narrow and thus the chart will be very sensitive and give false alarms more frequently than we'd like. The opposite is true if we select a value too small.

▶ Typically, control charts are operated with $\alpha = 0.0027$, which yields $L = 3.00$, and is where the term *Six Sigma* originates.

Introduction to the Shewhart $\bar{X}$ and R Charts

- ► However, in the previous example, we knew the in-control values of μ_0 and σ_0. While there may be instances where μ_0 is known, it's unlikely that σ_0 would be, and many times, neither are known and both have to be estimated (part of the analyze phase of DMAIC).

- ► In such cases, in Phase I, we take m historical samples of size n and estimate both the mean as well as the standard deviation.

$$\hat{\mu}_0 = \bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^{m} \bar{x}_i$$

$$\hat{\sigma}_0 = \bar{R} = \frac{1}{m} \sum_{i=1}^{m} R_i$$

Introduction to Shewhart $\bar{X}$ and R Charts

▶ Now, our control limits are going to change:

$$UCL = \bar{\bar{x}} + \frac{3}{d_2\sqrt{n}}\bar{R}$$

$$CL = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - \frac{3}{d_2\sqrt{n}}\bar{R}$$

▶ Note, d_2 is a control charting constant that can be found in the back of your book (Table VI) or at the following link: https://web.mit.edu/2.810/www/files/readings/ ControlChartConstantsAndFormulae.pdf

Introduction to Shewhart $\bar{X}$ and R Charts

- Now, since we're estimating σ_0 by using the range of the samples, we also want to monitor the variability to make sure it stays in a state of control, too. The R-chart works exactly like the $\bar{X}$-chart, except we're using the sample range as our plotting statistic and therefore we'll have different control limits:

$$UCL = D_4 \bar{R}$$
$$CL = \bar{R}$$
$$LCL = D_3 \bar{R}$$

- where again, D_3 and D_4 are control charting constants found in Table VI of your book or at the following link: https://web.mit.edu/2.810/www/files/readings/ ControlChartConstantsAndFormulae.pdf.

Introduction to Shewhart $\bar{X}$ and R Charts

- What's probably clear here is the importance of this estimation component in Phase I if we're going to be successful in Phase II monitoring.

- For starters, how big should n and m be?
  - n doesn't need to be overly large, samples of size 4 or 5 are perfectly fine.

  - m is a bit more contentious. Your text suggests 20 or 30 samples are good enough, but many studies refute this as insufficiently small and I agree. m should be more like 100 and some studies have gone as far as to suggest 500.

Introduction to Shewhart $\bar{X}$ and R Charts

▶ A second question that might be raised here is: What if the process we're monitoring isn't in-control right now? Aren't we building error into the control limits?

▶ Great question and the answer is yes! Here in Phase I, after we've calculated the control limits, we plot the $\bar{x}_i$'s and R_i's against the control limits. If a point plots beyond the limits, then this implies the process is not currently in statistical control.

  ▶ We need to improve our process by identifying potential causes, correcting them, omitting the out-of-control point(s) and recalculating the control limits.
  ▶ Once all of the points plot inside of the control limits, then we can feel confident that our process is now in a state of statistical control and we can move to Phase II monitoring.

Introduction to Shewhart $\bar{X}$ and R Charts

▶ Okay, well what if we can't find an assignable cause? Your text suggests two things: (1) Drop the point as if you had found an assignable cause or (2) Include the out-of-control point in the control limit calculation, and just assume it was a false alarm.

▶ Both of these suggestions aren't ideal, but if m is big enough, we won't miss those couple of points too badly.

▶ However, if we have a lot of points that are out-of-control, it isn't really feasible to investigate each individually, so your text suggests attempting to interpret the pattern of plotted points as this would likely aid in understanding why the process is OOC
   ▶ Patterns could be things like, all the points are greater than or less than the CL or maybe they have a "W" type of pattern among other things.

Introduction to Shewhart $\bar{X}$ and R Charts

▶ Okay, now that we know a little bit about how these charts are formed, let's go through an example using both R and Python.

▶ In Table 6.1 of your text, there contains data for $m = 25$ samples of size $n = 5$ where the monitored process is the "Flow Width Measurements (in microns) of a Hard-Break process." I've created an Excel file that contains this data for us.
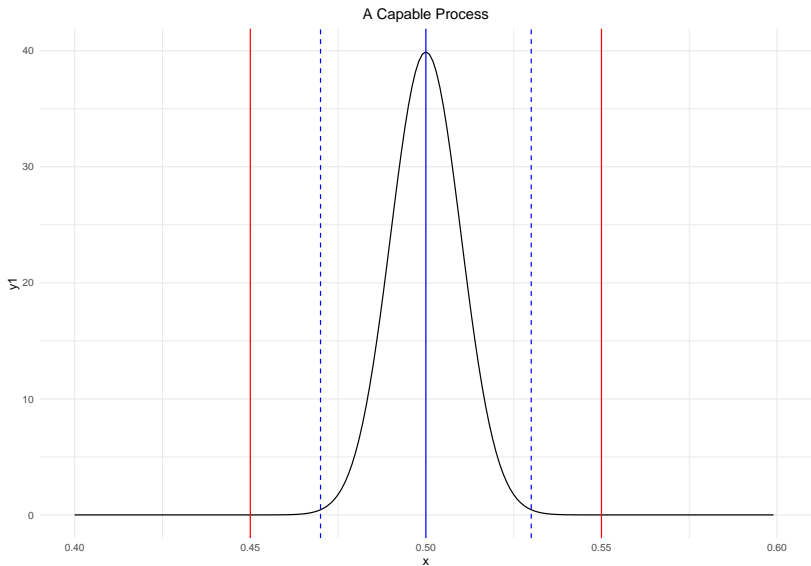
Process Capability

▶ As mentioned last week, when we're performing Phase I analysis, we're interested in answering two main questions:
  1. Is the process in a state of statistical control?
  2. Is my process capable?

▶ For us, we've just learned how to estimate our process parameters (e.g., the mean and standard deviation) and how to determine if the process is in-control.
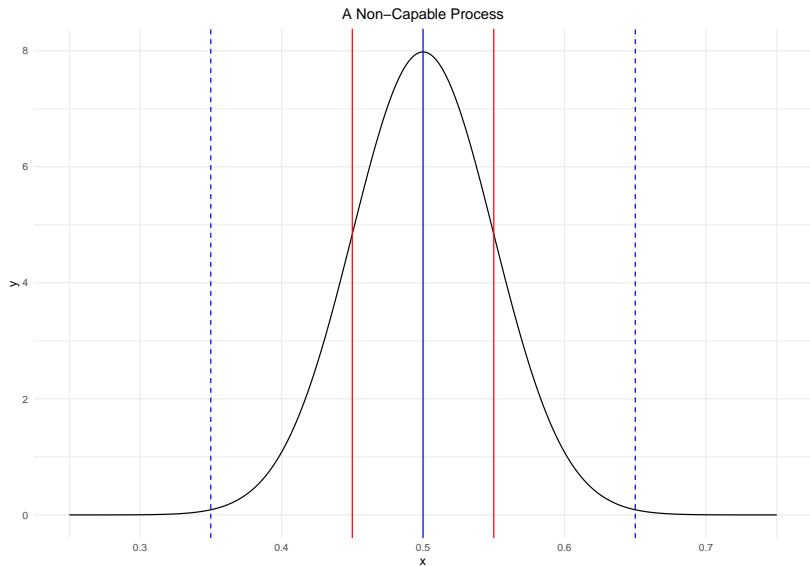
▶ But what do we mean by process capability?

Process Capability

▶ A process is considered capable if it is able to produce items that meet the specifications of the end-user, customer, or engineer.

▶ For instance, if we're monitoring the width of a USB port, the specification limits might be 0.5 ± 0.05 mm.

▶ If our inherent process variability, as quantified by the standard deviation $\hat{\sigma}$, is too large, then we will produce a lot of nonconforming units. Let's look at a visual example of this.

Process Capability

A Capable Process

Process Capability

A Non–Capable Process

Process Capability

▶ In the first plot, the process is capable because the natural tolerance limits (i.e., the process mean plus/minus 3σ) are inside of the specification limits.

▶ In the second plot, the process is not capable because the natural tolerance limits are outside of the specification limits.

▶ We can see this visually, but how do we quantify capability?
  ▶ We have a few different measures available to us!

Process Capability

▶ One way of measuring capability is by estimating the **fraction non-conforming**.

▶ The fraction non-conforming is the proportion of samples we would expect to fall outside of the *specification limits* (which are the physical bounds an item has to conform to in order to work) despite the process being in a state of statistical control.

Process Capability

▶ In the Wafer example, suppose our specification limits are 1.50 ± 0.50. Let's also get an estimate of the process standard deviation by using:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0.3252}{2.326} \approx 0.1398$$

▶ Now the FNC is just the probability of an observation falling outside of the specification limits:

$$P[(x < 1.00) \cup (x > 2.00)] = P[x < 1.00] + (1 - P[x < 2.00])$$

$$\implies P\left[Z < \frac{1.00 - 1.5056}{0.1398}\right] + \left(1 - P\left[\frac{2.00 - 1.5056}{0.1398}\right]\right)$$

$$\implies P[Z < -3.6166] + (1 - P[Z < 3.5365]) \approx 0.00035$$

Process Capability

- The FNC of 0.035% implies that for every 100,000 wafers manufactured, we would expect about 35 to not conform.

- Another more common way of expressing process capability is by using the aptly named "Process Capability Ratio," which is denoted by C_p and is calculated by:

$$C_p = \frac{USL - LSL}{6\sigma}$$

- Since we don't know σ and have to estimate it, we can plug our estimate, $\hat{\sigma}$ in the denominator:

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}}$$

Process Capability

- So in our example:

$$\hat{C}_p = \frac{2.00 - 1.00}{6(0.1398)} \approx 1.192$$

- Since this number is greater than 1, it implies that process' natural tolerance limits (i.e., process mean plus/minus 3σ) are inside of the specification limits (which is good!).
  - In other words, the specification range is 1.192 times the length of the natural tolerance range.

- The reciprocal of $\hat{C}_p$ shows the converse:

$$\hat{p} = \left(\frac{1}{\hat{C}_p}\right) \times 100\% = \left(\frac{1}{1.192} \times 100\% \approx 83.89\%\right)$$

Process Capability

- This value tells you the proportion of the specification band ($USL - LSL$) the natural tolerance limits are taking up.

- If $\hat{p} > 100\%$ this means we will be producing a lot of nonconforming units. If it is less than 100%, then this is good and means we will be producing a low number of nonconforming units (like we saw in the FNC).

Process Capability

- One of the key assumptions in using the Shewhart $\bar{X}$ and R Charts is that the underlying monitored process follows a Normal distribution

- In practice, this may or may not be the case. Fortunately, simulation studies have shown that the charts, and particularly the $\bar{X}$ chart, are robust to slight and moderate deviations from normality (the R chart is much more sensitive).

- However, in the presence of highly skewed data (such as time-to-event data), neither chart performs well.

- In such cases when normality is unknown or cannot be reasonably assumed, it is better to use a nonparametric chart for variables, of which there are many (see Chakraborti et al (2001) for a thorough overview of many of the classics).

Chart Efficiency: Average Run Length and Operating Characteristic Curves

▶ To this point, we have learned some techniques for evaluating the capability of a process, but we haven't yet explored how to determine the efficacy of a control charting scheme in general.

▶ Remember, there are two desirable characteristics of a control chart:
  1. If the process is OOC, we find out in as few samples as possible
  2. If the process is IC, we don't get false alarms too frequently

▶ We have two tools available to us for graphically and quantitatively assessing these questions called **Operating Characteristic Curves** and **Average Run Length**.

Chart Efficiency: Average Run Length and Operating Characteristic Curves

► Regarding the OC-Curves, what they basically do is for a given sample size and shift away from target, they plot the probability of making a Type II error, denoted β, against an increase in the magnitude of the shift.

► For the $\bar{X}$-chart, the probability of making a Type II error given a particular shift away from target on the first sample after the shift has occurred can be denoted:

$$\beta = P[LCL \leq \bar{x} \leq UCL | \mu_1 = \mu_0 + k\sigma]$$

► With a little algebra, we can show that:

$$\beta = \Phi(L - k/\sqrt{n}) - \Phi(-L - k/\sqrt{n})$$

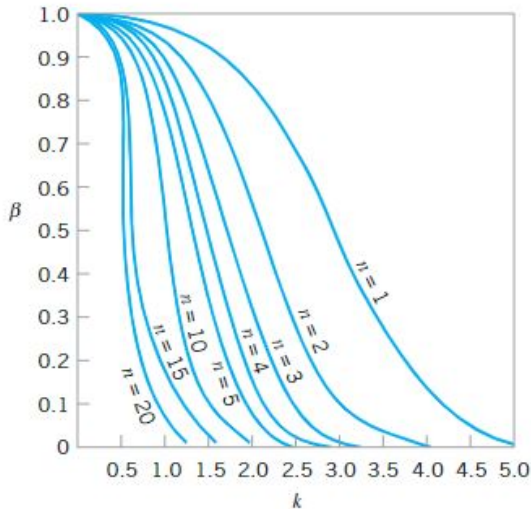# Chart Efficiency: Average Run Length and Operating Characteristic Curves

Figure 1: From Figure 6.13 of your text

Chart Efficiency: Average Run Length and Operating Characteristic Curves

▶ Now this leads us to the second concept of chart performance, Average Run Length or ARL.

▶ What this value tells us, is on average, if we have experienced a particular shift away from target, how many samples can we expect to take before the chart signals OOC?

▶ We know from the OC-curves that the probability of making a Type II error on the first sample after a shift has occurred (i.e., the chart fails to signal even though the mean has shifted away from target) is β.

▶ So then what's the probability of the chart correctly signaling on the second sample?

Chart Efficiency: Average Run Length and Operating Characteristic Curves

$$P[\text{Detect on 2nd Sample}|OOC] = \beta(1 - \beta)$$

▶ And thus the probability of not detecting on the second sample is β^2. This process iterates recursively, in theory, a countably infinite number of times. Thus, the run length itself can be considered a geometric random variable with a PDF:

$$P[RL = r] = \sum_{i=1}^{r} \beta^{r-1}(1 - \beta)$$

▶ Therefore, for a given β, the ARL is just the expectation of the geometric distribution:

$$ARL = E[RL] = \sum_{i=1}^{\infty} r\beta^{r-1}(1 - \beta) = \frac{1}{1 - \beta}$$

Chart Efficiency: Average Run Length and Operating Characteristic Curves

▶ This ARL value tells us that given our process is out of control, and also given the width of our control limits as well as our sample size, here's how many samples on average it would take for the chart to signal.

▶ On the flip side, we know we don't want this chart to be giving us a bunch of false alarms either. That is, we want a long in-control ARL sometimes denoted ARL_0.

▶ Using a similar logic as before, the IC-RL is also a geometric random variable, but the parameters are flipped as α denotes the probability of a false alarm and $(1 - \alpha)$ denotes the probability of not observing a false alarm.

Chart Efficiency: Average Run Length and Operating Characteristic Curves

$$ARL_0 = E[RL_0] = \sum_{r=1}^{\infty} r(1-\alpha)^{r-1}\alpha = \frac{1}{1-(1-\alpha)}$$

$$\implies ARL_0 = \frac{1}{\alpha}$$

▶ Thus, for $\alpha = 0.0027$, $ARL_0 \approx 371$, which is a commonly desired ARL_0 value as it corresponds with $\pm 3\sigma$ control limits.