How Should the Sample Size be Selected for an X-bar Chart? (Part I)

Steven Wachs
Principal Statistician
Integral Concepts, Inc.
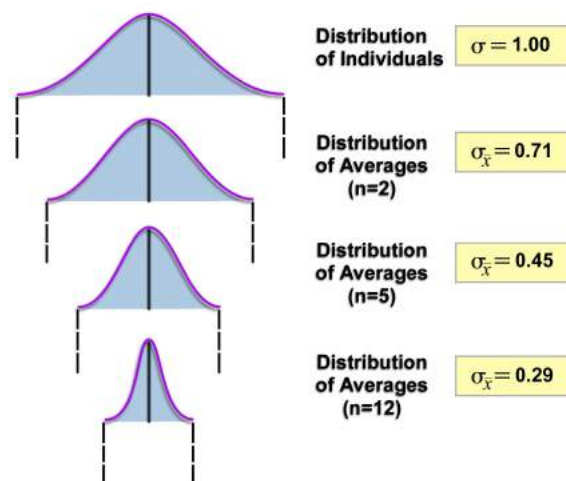
The purpose of control charts is to detect significant process changes *when they occur*. In general, charts that display *averages* of data/measurements (X-bar charts) are more useful than charts of *individual* data points or measurements. Charts of individuals are not nearly as sensitive as charts of averages at detecting process changes quickly. X-bar charts are far superior at detecting process shifts in a timely manner, and the sample size is a crucial element in ensuring that appropriate chart signals are produced.

Often, the sample size is selected without much thought. A sample size of 5 seems to be a common choice. If the sample size is not large enough, then meaningful process shifts may go undetected. On the other hand, if the sample size is too large, then chart signals may be produced from insignificant process shifts. The key is to specify a sample size so that significant shifts (from a practical perspective) are detected with high probability and that insignificant shifts are unlikely to produce a signal.

To understand the concept, it is useful to review the impact that averaging data has on variation. The graphic below compares the distribution of individual values with the distributions of averages of various sample sizes.

Sensitivity

|	
Distribution of Individuals	$\sigma = 1.00$
Distribution of Averages (n=2)	$\sigma_{\bar{x}} = 0.71$
Distribution of Averages (n=5)	$\sigma_{\bar{x}} = 0.45$
Distribution of Averages (n=12)	$\sigma_{\bar{x}} = 0.29$

1

We see that as the sample size increases, the standard deviation of the distribution of *averages* decreases. Specifically, the relationship below relates the standard deviation of averages to the standard deviation of individuals and the sample size:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

On an x-bar control chart, this idea is reflected by control limits getting tighter as the sample size increases.

Now, consider a process that is stable and under statistical control. This curve is represented by the blue curve on the graphic below. Suppose we desire that if the process average shifts by a specified amount (such that it is represented by the red curve below), we would want to obtain a chart signal with very high probability.

Changed Process   Original Process

Distribution of Individuals

Distribution of Averages (n=2)

Distribution of Averages (n=5)

Distribution of Averages (n=12)

2
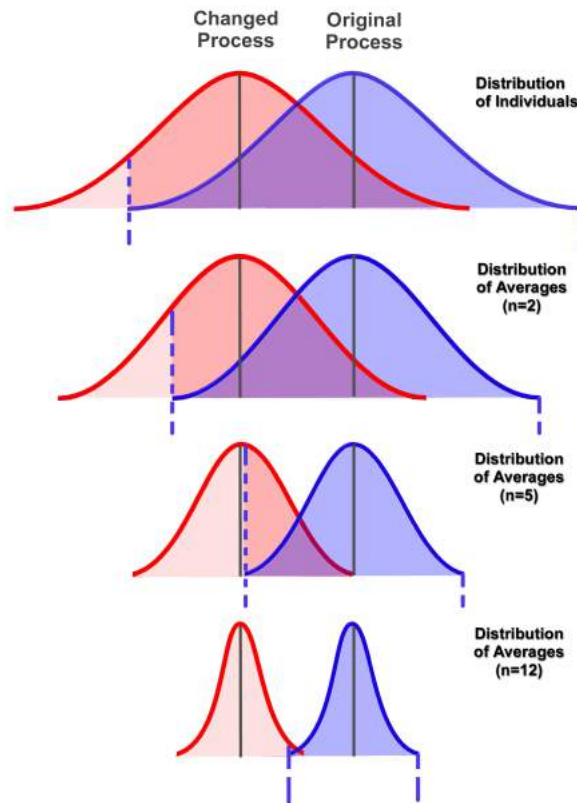
The question is, how likely is it that we will detect the process shift, if in fact the process shifts from the blue curve to the red curve? We consider this question for the 4 cases shown in the above graphic.

In the top set of distributions, we are working with individuals. The control limits for the blue process are represented by the dashed vertical lines. Following the process shift, we will sample from the red curve. Note that most of the red curve still falls inside the control limits for the blue curve. This means that we are very unlikely to see a signal on our chart for the shift indicated.

Now consider, the second set of distributions. They represent the case where we are using an x-bar chart with sample size = 2. A shift of the same size is shown. Here, the process curves are tighter since they represent averages (with n = 2). While things are better than the first case, there is still a significant overlap between these distributions and it is still not very likely that we will detect the shift quickly.

The case with n = 5 (the third case) is better yet, but still not good enough. We want to be able to detect the shift with high probability.

Finally, with n = 12 (the last case), we see that for the same size shift, the two distributions are practically separate. In other words, if the shift occurs, our next sample average will come from the red curve and it will almost certainly be outside of the control limits (based on the blue curve).

Although the above graphics allow us to understand how sample size affects chart sensitivity (the ability to detect desired process changes), a formula is typically used to compute the necessary sample size for a given application. The sample size is a function of the desired sensitivity, the process standard deviation, and the willingness to tolerate type II errors (where the process shifts but the chart fails to detect the shift).

The formula and its application will be discussed in the next article.