

Sample Size vs. Subgroup Size

On an ASQ (American Society for Quality) web site, a group member posted the following inquiry:

Can you let me know if the sample size and subgroup size mean the same? This is because in some websites and books the selection of control charts is based on sample size whereas in others it is based on the subgroup size even though the charts mentioned are the same.

The post generated a number of both helpful and convoluted responses, among them:

- Sample Size is the number of data points that you plot on the chart! Each data point could be an average of the number of measurements taken at the same time frame. Subgroup size is normally 5 and sample size normally 25-30.
- You will take samples from a group to understand the group. [This respondent's profile trumpeted that he's an "expert in Six Sigma."]
- The way I have always dealt with it is that sample size is the number of replicates and subgroup size is the number of individuals in that sample.
- Subgroup size come into picture when we are taking samples in regular intervals like during control charts while sample size is irrespective of regular intervals after the lot is completed. [Comment posted by a self-described "certified Lean Six Sigma Black Belt."]
- Don Wheeler has a good paper here: <http://www.spcpress.com/pdf/DJW260.pdf>.
- In Xbar-S/R chart you select the subgroup size (5-30 samples) from your entire sample size to create the chart. If not economically feasible, use individuals chart. [I'd like to ask this Lean Six Sigma Black Belt, "But what if the individuals aren't normally distributed?"]

One would think that such a rudimentary question would generate better responses from a bunch of Black Belts and quality professionals. I decided pitch in and posted the following:

Walter Shewhart taught that minimum sample size should be 30-50 measurements. The data should be collected over a sufficient period of time to allow the process (common cause variation) to be represented. He determined that after plotting the representative measurements on a histogram, the distribution will take shape (unimodal, skewed, bimodal, etc.) When he added more measurements, the distribution grew and spread out, but the shape didn't change.

We make decisions on subgroup size based on the shape of the distribution. The more heavily skewed or non-normal the distribution of individuals, the larger must be the subgroup size. The more the distribution approximates a bell-shaped or normal distribution, the smaller the subgroup size – because you won't need much help to get the Central Limit Theorem working for you.

The late Dr. David Chambers once told me that once he figured everything out, Shewhart always used subgroups of size $n = 5$ and X-bar and R charts. He did so for three reasons:

1. He was supporting a high-volume telephone production line; plenty of opportunity to get the measurements.
2. His measurement process was neither difficult nor destructive.
3. By then, he had figured out that – regardless of the shape of the distribution of individual measurements – averages from subgroups of size $n = 5$ would get enough of the Central Limit Theorem working for him to lend sufficient validity to his calculated control limits.

Of course, when dealing with a process that yields slowly-evolving variables data, the X-bar and R charts may not be applicable. They may require more data than the process is providing. In such cases, the Individuals/Moving R charts can be used if the individual measures plot in a bell shape, or the Moving X-bar/Moving R charts can be deployed using larger moving subgroups if the distribution of individuals is non-normal.

© 2017 James F. Leonard. All rights reserved.