QC Report

December 1996

Published by Stochos Inc., 14 N. College Street, Schenectady, New York 12305

Sample Size Questions and Answers

(by Donald S. Holmes)

Questions about sample size required to achieve some desired level of accuracy arise very often. There is a rather wide variety of these questions. Examples include (and you may have others - let us know):

Question #1: How large a sample do I need to estimate the average value of my process?
Question #2: What sample size do I need to decide whether or not to accept a lot?
Question #3: How large a sample do I need from a lot generated by a multiple stream process to be confident that the sample includes an item from each stream?

I'll discuss these sample size questions (and others if you send them) in separate QC Reports.

(Continued on page 3)

Several Definitions:

n = Sample size
μ = Process average - a measure of process center
xbar = sample estimate of process average is defined as:
xbar = Sum(x)/n *example:* data x's 1, 2, 3, 4, 5
xbar = (1 + 2 + 3 + 4 + 5)/5 = 3
σ (Sigma) = Process standard deviation - a measure of process variability
Sample estimate of process standard deviation is defined as:
s = Square root (Sum ((x-xbar)^2) / (n-1))
example: data x 1, 2, 3, 4, 5 xbar=3
x-xbar -2,-1, 0, 1, 2
(x-xbar)^2 4, 1, 0, 1, 4 sum ((x-xbar)^2) = 10
sample standard deviation = Sqr (10/4) = 1.58
Z(x) = The number of standard deviations an x value is from the average.
Z(x) = (x-μ)/σ *example:* mu = 50, sigma = 2, x = 53
z (53) = (53 - 50)/2 = 1.5 (53 is one and one half standard deviations away from 50)

For a Normal curve, 99.7% of the x's are within 3 standard deviations of the average, 95% are within 2 standard deviations etc. So Z = 2 is associated with "95% confidence". Other confidence percentages can be found in textbooks on statistics or using Custom/QC's Frequency Curve module.

Statistical Process Control Training

The Quality professionals at Stochos offer customized SPC training for either technical personnel (engineers, technicians, etc.) or shop-floor operators. We will tailor the class to your industry by using data from your processes for case studies and examples where applicable.

Technical Personnel (Quality Manager / Quality Technician/ Supervisor Training):
The training for technical personnel is available on-site or at Stochos' facility in Schenectady, NY. A typical syllabus for a two-day class is as follows:

Objective: Refresh technical personnel on the general principles of Statistical Process Control (SPC) and train them in the advanced techniques that are relevant to your specific industry. Case study examples will be covered using plant/industry specific data. This class is recommended for supervisors that demonstrate a high level of technical competence and interest in quality control application and techniques.

Day 1 - Morning Session:

- Why is quality important?
- What is Statistical Process Control? - An overview
- Types of Data used in Quality Analysis
- Pareto Charts - Construction & Interpretation (Case Study)
- Histogram Construction (Case Study)
- Descriptive Statistics: Measures of Process Center - Your Data
- Measures of Process Variability - Your Data
- Normal Curves and their Characteristics
- Random vs. Assignable Cause Variation (Common Cause vs. Special Cause)

Day 1 - Afternoon Session:

- What is 'Over-Control'? (Case Study: Quincunx)
- Process Capability vs. Process Performance Standard Deviations
- Review the meaning of Specifications (Nominal, USL, LSL)
- Relative Quality Measures: Process Performance Indices (Pp, Ppk, etc.)
- Process Capability Indices (Cp, Cpk, etc.) - Case Study

Day 2 - Morning Session:

- Introduction to Control Charts: XBar and Range Chart
- A New Capability Standard Deviation
- Case Study - Your Data
- Why do we need chart pairs?
- Additional Control Tests
- Discussion of "In-Control" vs. "In-Specification"
- Different uses for Control Charts - Specific Examples
- Common Control Chart Pitfalls

Day 2 - Afternoon Session:

- Attribute Control Charts
- Advanced Control Chart Types: CuSum Charts
- EWMA Charts (Parallel Limits & Parabolic Mask)

☞ ***Follow the lead of the other companies that have used our SPC training expertise:***

- Georgia-Pacific • Sealed Air • Proctor & Gamble
- Beechnut • Waldorf • Jefferson Smurfit

Key Variable Identification Service

A refractory business identified 5 out of 75 variables that impacted on the lifetime of their products.

Results: 40% improvement in lifetime with no anticipated investment in DOE.

A metal film manufacturer identified 3 out of 22 control variables that impacted on the tensile strength of its sheets.

Results: 25% increase in the tensile strength of the conductive sheet using only historical data.

A motor manufacturer identified 5 out of 60 control variables that were related to vibration of the motor.

Result: Fixing one key component reduced the variation by 10% - again with only existing historical data.

A wire manufacturer found 3 out of 15 variables that were related to the uniformity of its products.

Result: Applying new control methodology to two of the variables improved the uniformity by 15%. Again, the expected investment in DOE was delayed until the "first fruit" was harvested.

A container manufacturer found 2 out of 6 incoming material variables were the major causes of variation in its final product.

Result: The supplier and the manufacturer worked together to reduce the incoming variation and to reduce the sensitivity of the process to variation of incoming material.
Result: Reduction in container customer complaints by about 20%.

Would you like to have similar results in your plant?

Call us for further information about how you can improve your processes using our process detectives.

Sample Question #1

Cont'd from page 1

How large a sample do I need to be relatively confident (e.g. 95%) that the sample average does not miss the true value by more than some specified amount (d)?

Solution: n = (Z*Sigma)^2 / d^2

Example: If d = 0.35 millimeters, then you have decided to accept a difference of 0.35 millimeters between the sample average and the actual average of the process.

Suppose you would like to be 95% confident that the difference between the sample average and the actual average of the process is no larger than the value you assigned to d (0.35 millimeters). Since the Xbars are approximately Normally distributed, a value of 2 will be assigned to Z. (If you wanted 99.7% confidence, you would use Z=3.)

The next thing you need is an estimate of the standard deviation (sigma). Suppose you estimate sigma=0.40 millimeters. Then n = (2*0.40)^2/ (0.35)^2 = 0.64 / 0.1225 = 5.2 which rounds up, in this application, to 6.

If you don't have an estimate of sigma, take a sample of some size and plug the sample standard deviation (s) in equation (1). If the new sample size is larger than the one used, then take additional samples to make the total sample size about the value calculated. If the new sample size is about the same as the one you used, then report your average. You can cycle through this step several times until you are satisfied that the sample size is large enough to give you your required confidence level.

To complete the example, suppose you took your six samples and got the following results: 9.8, 10, 10.5, 11, 10.8, 10.7. Using this data we find:

Xbar = 10.37 and s = 0.45.

The standard deviation is a little larger than you thought, so let's do the next cycle using s = 0.45:

n = (2*0.45)^2 / (0.35)^2 = 0.81 / 0.1225 = 6.6 or 7

You could take one more data point and merge the result with those you already have. Suppose you do that and the value is 10.4.

Xbar = 10.37 and s = 0.41.

Putting this value for s in the equation will result in:

n = (2*0.41)^2 / (0.35)^2 = 0.71 / 0.1225 = 5.8 or 6

We can now report the true mean value is 10.37 with about 95% confidence that error is 0.35 or less.

More precision about the confidence level can be obtained using student's t as follows:

Calculate: $$t = \frac{t * \sqrt{n}}{s} = \frac{0.35x}{0.41} = 2.25$$

Look up 2.25 in the student's t table using n-1 = 6 degrees of freedom. You would find that you have about 93% confidence rather than the 95%.

You would also solve for d: $$d = t\frac{s}{\sqrt{n}}$$

Using t for 95% (with 6 degrees of freedom) you will find the value of d for which you have 95% confidence.

$$d = \frac{2.447 * 2.41}{\sqrt{7}} = \frac{1.003}{2.646} = 0.38$$

You can have either 93% confidence that your maximum error is 0.35 or 95% confidence that your maximum error is 0.38.