

IE 361 Module 18

Process Capability Analysis: Part 2

Prof.s Stephen B. Vardeman and Max D. Morris

Reading: Section 5.3 *Statistical Quality Assurance Methods for Engineers*

In this module we consider methods of characterizing process output that focus on the values of individual future process outcomes rather than on process summary measures.

Inference for One More Value from the Process and for the Location of "Most" of the Distribution of Process Values

A way of characterizing process output different from trying to pin down process summary parameters like the mean and standard deviation (or even functions

of those parameters like 6σ , C_p , and C_{pk}) is to instead try to quantify what one has learned about future process outcomes from data in hand. If I KNOW process parameters, making statements about future individual values generated by the process is a matter of simple probability calculation. Suppose, for example, that I model individual values as normal with $\mu = 7$ and $\sigma = 1$.

Then doing simple normal distribution calculations, both

- there's a "90% chance" the next x is between 5.355 and 8.645
- 90% of the process distribution is between 5.355 and 8.645

But what if I only have a sample, and not the process parameters? What then can I say? When one has to use a sample to get an approximate picture

of a process, it is important to hedge statements in light of sample variability/uncertainty ... this can be done

- for normal processes using \bar{x} and s
- in general, using the sample minimum and/or maximum values

Methods for Normal Processes

We consider first methods for normal processes. (Just as we cautioned in Module 17 that the methods for estimating capabilities are completely unreliable

unless the data-generating process is adequately described by a normal model, so too does the effectiveness of the next 2 formulas depend critically on the normal assumption being appropriate.)

For normal processes, "**prediction limits**" for a single additional individual are

$$\bar{x} \pm ts\sqrt{1 + \frac{1}{n}}$$

Prediction limits are sometimes met in Stat 231 in the context of regression analysis, but the simple one-sample limits above are even more basic and unfortunately not always taught in an introductory course. They are intended to capture a single additional observation from the process that generated \bar{x} and s .

Example 18-1 (EDM drilling) What do the $n = 50$ measured angles in Table 5.7 tell us *about additional angles drilled by the process*? One possible answer

can be phrased in terms of a 95% prediction interval for a single additional output. (Recall that for the hole angle data $\bar{x} = 44.117^\circ$ and $s = .984^\circ$.)

Using the fact that the upper 2.5% point of the t distribution for df $\nu = 50 - 1 = 49$ is 2.010, 95% prediction limits for a single additional output are

$$44.117 \pm 2.010 (.984) \sqrt{1 + \frac{1}{50}}$$

i.e.

$$44.117 \pm 1.998$$

One can in some sense be 95% sure that the next angle drilled will be at least 42.119° and no more than 46.115° . This is a way of quantifying what the sample tells us about the angles drilled by the process different from confidence limits for quantities like μ , σ , 6σ , C_p and C_{pk} . The 95% figure is a "lifetime batting average" that is associated with a long series of repetitions of the

whole business of selecting n , making the interval, selecting one more value, and checking for success. (In any given application of the method one is either 100% right or 100% wrong.)

Another way to identify what to expect from future process outcomes might be to locate not just a single outcome, but some large fraction (p) of all future outcomes (under the current process conditions). For a normal process, **two-sided "tolerance limits" for a large fraction (p) of all additional individuals** are

$$\bar{x} \pm \tau_2 s$$

(the values τ_2 are special constants tabled in Table A.9.a of *SQAME*). (One-sided limits are similar, but use constants τ_1 from Table A.9.b of *SQAME*.) Tolerance limits are not commonly taught in an introductory statistics course, so most students will not have seen this idea.

Example 18-1 continued A second possible answer to the question "What do the $n = 50$ measured angles in Table 5.7 tell us *about additional angles drilled by the process?*" can be phrased in terms of a 99% tolerance interval for 95% of all values from the process. (This would be an interval that one is "99% sure" contains "95% of all future values.") Reading directly in Table A.9.a of *SQAME* produces a multiplier of $\tau_2 = 2.58$ (note that the table is set up in terms of sample size, NOT degrees of freedom) and therefore the two-sided tolerance limits

$$44.117 \pm 2.58 (.984)$$

or

$$44.117^\circ \pm 2.54^\circ$$

for the bulk of all future angles (assuming, of course, that current process conditions are maintained into the future). (A one-sided tolerance limit can be had by replacing τ_2 with a value τ_1 from Table A.9.b of *SQAME*.)

A "thought experiment" illustrating the meaning of "confidence" associated with a tolerance interval method involves

1. drawing multiple samples
2. for each one computing the limits $\bar{x} \pm \tau_2 s$
3. for each one using a normal distribution calculation based on the true process parameters to ascertain the fraction of the population covered by the sample interval
4. checking to see if the fraction in 3. is at least the desired value p ... if it is, the interval is a success, if it is not, the interval is a failure

The confidence level for the method is then the lifetime batting average of the method.

Prediction and Tolerance Intervals Based on Sample Minima and Maxima

A second approach to making prediction and tolerance interval (that **doesn't depend upon normality** of the data generating process for its validity, only on process stability) involves simply using the smallest and largest data values in hand to state limits on future individuals. That is, one may use the interval

$$(\min x_i, \max x_i)$$

as either a prediction interval or a tolerance interval. Provided the "random sampling from a fixed (not necessarily normal) universe" model is sensible, used as a **prediction interval for one more observation** this has confidence level

$$\frac{n - 1}{n + 1}$$

and used as a **tolerance interval for a fraction p of all future observations** from the process it has associated confidence level

$$1 - p^n - n(1 - p)p^{n-1}$$

Example 18-1 continued The smallest and largest angles among the $n = 50$ in the data set of Table 5.7 are respectively 42.017 and 46.050. We consider the interval

$$(42.017, 46.050)$$

for locating future observed angles. As a prediction interval for the next one, the appropriate confidence level is

$$\frac{50 - 1}{50 + 1} = .961 = 96.1\%$$

And, for example, as a tolerance for 95% of EDM drilled angles, the appropriate confidence level is

$$1 - (.95)^{50} - 50 (.05) (.95)^{49} = .721 = 72.1\%$$

The interpretation of confidence level associated with these intervals based on sample minimum and maximum values is exactly the same as for the normal distribution prediction and tolerance intervals based on \bar{x} and s . The news here is that these levels are guaranteed for *any* (continuous) process distribution ... normal or not.

Section Formula Summary

Below is a table summarizing the formulas of Section 5.3 of *SQAME*.

		Two-Sided Intervals	One-Sided Intervals
Normal Process	PI	$\left(\bar{x} - ts\sqrt{1 + \frac{1}{n}}, \bar{x} + ts\sqrt{1 + \frac{1}{n}}\right)$	$\left(\bar{x} - ts\sqrt{1 + \frac{1}{n}}, \infty\right)$ $\left(-\infty, \bar{x} + ts\sqrt{1 + \frac{1}{n}}\right)$
	TI	$(\bar{x} - \tau_2s, \bar{x} + \tau_2s)$	$(\bar{x} - \tau_1s, \infty)$ $(-\infty, \bar{x} + \tau_1s)$
Any Stable Process	PI	$(\min x_i, \max x_i)$ <i>confidence</i> = $\frac{n-1}{n+1}$	$(\min x_i, \infty)$ or $(-\infty, \max x_i)$ <i>confidence</i> = $\frac{n}{n+1}$
	TI	$(\min x_i, \max x_i)$ <i>confidence</i> = $1 - p^{n-n} (1 - p) p^{n-1}$	$(\min x_i, \infty)$ or $(-\infty, \max x_i)$ <i>confidence</i> = $1 - p^n$