Statistics Roundtable

Comparing the Importance of Variation And Mean of a Distribution

An industrial case study

by **LeRoy A. Franklin, Belva J. Cooley and Gary Elrod**

Careful statistical thinking may lead to a different choice than that guided by superficial statistical thinking or conventional wisdom. This point was exemplified when a representative of a manufacturing company approached the authors with a real problem. The following edition of Statistics Roundtable illustrates the use of careful statistical thinking on actual manufacturing data that proved to be somewhat counter intuitive.

After this particular organization acquired a second, smaller company, it realized it had two different methods (method A and method B) for securing a pulley, fan or gear onto a steel hub—a common operation. Both the original organization and its newly acquired company had a preferred method. The two methods did not differ substantially in cost, and the acquisition presented an opportunity to adopt the better method throughout the organization, standardize production and reduce inventories.

TABLE 1 Breaking Strength of Hub-Pulley Assembly (in foot-pounds) of Randomly Selected Items

Method A		Method B	
Sample = 52		Sample = 52	
143	135	84	83
110	110	84	84
103	100	82	82
125	115	82	83
125	90	83	84
135	121	83	84
108	100	82	83
100	148	82	83
105	160	84	82
115	130	83	83
120	130	84	84
110	100	83	82
100	110	83	81
90	130	84	82
75	106	84	82
112	117	83	83
95	100	83	84
95	110	84	84
125	100	83	83
100	90	83	83
107	140	83	84
107	90	82	83
150	120	83	82
100	100	84	82
115	83	82	83
120	135	83	85

Ensure methods are statistically stable

The particular components needed to have a minimum of 50 foot-pounds of torque (holding power) between the steel hub and the pulley or fan. In other words, 50 foot-pounds was a lower specification limit (LSL). The company, trying to incorporate more statistical thinking and quality orientation into its operations, asked two quality technicians to devise a test, collect data and determine which method would be used for this particular process in the future.

The technicians decided to conduct capability studies on both processes, including the calculation of the process index, $CPL = (\mu - LSL)/3\sigma$.[1, 2] Since the items were relatively inexpensive, they collected a fairly large sample (52) from each production method to help ensure sound decisions. Table 1 presents the data collected.

Both technicians were aware that a capability study is only meaningful and valid when the process being studied is, in fact, statistically stable. Thus, they constructed an individuals chart (I) to check the stability of the center of the process and a moving range chart (MR) to check the stability of the variation of the process.[3]

The measurements taken regarding method B presented some peculiarities. The measurement device used was a torque wrench and could only be read to the nearest foot-pound. Thus, the measurements for method B, since they were so consistent, resulted in only three values between the control limits for its range chart.

Several authors have noted that, generally, "If there are fewer than four values on the range chart below the upper control limit, the discrimination of the measuring devices is not adequate."[4] Such lack of discrimination casts some doubt on the chart's ability to evaluate statistical stability.

Faced with practical time constraints, the technicians used input from engineers and operators and historical data about method B to help support their conclusions about its stability. In addition, they expressed a need to procure a more precise measuring device in order to aid future analysis.

Both processes were judged to be statistically stable (in both mean and variation) so they each undertook a capability study.

Capability studies indicate benefits of both methods

Capability studies typically assume the data is normally distributed, but the measurement device was only

FIGURE 1 Method A—Individuals and Moving Range Charts

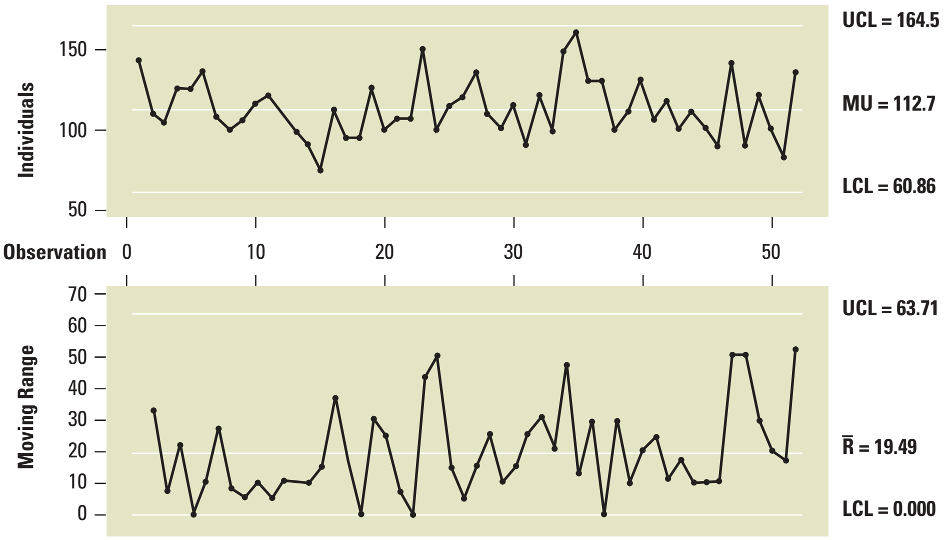
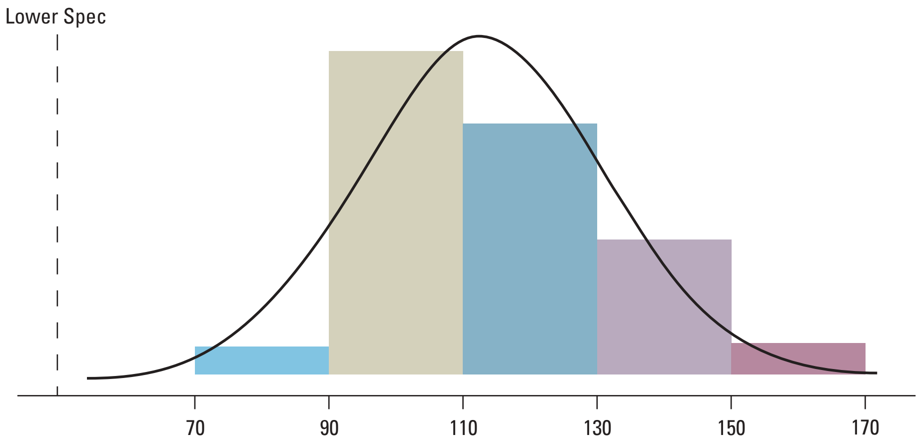


FIGURE 2 Capability Analysis for Method A



Method A									
Cp	*	Targ	*	Mean	112.692	%>USL Exp	*	PPM>USL Exp	*
CPU	*	USL	*	Mean + 3s	166.769	Obs	*	Obs	*
CPL	1.16	LSL	50	Mean - 3s	58.615	%<LSL Exp	0.03	PPM<LSL Exp	253
Cpk	1.16	k	*	s	18.026	Obs	0.00	Obs	0
Cpm	*	n	52						

able to produce five different values for method B (81 to 85 foot-pounds) due to its lack of discrimination. Thus, while the underlying manufacturing process could be normal, the measurement device limitations resulted in categorized data. Tests for normality were performed on both methods (using Minitab software) and neither displayed enough deficiencies to be judged nonnormal.

The results of the capability studies are graphically displayed in Figure 1 (I and MR chart) and Figure 2 (capability study) for method A. Figure 3 (I and MR chart) and Figure 4 (capability study) offer the capability studies results for method B.

Method A had a high mean ($\bar{x}_A$ = 112.7 foot-pounds) with a moderately sized standard deviation (s_A = 18

FIGURE 3 Method B—Individuals and Moving Range Charts

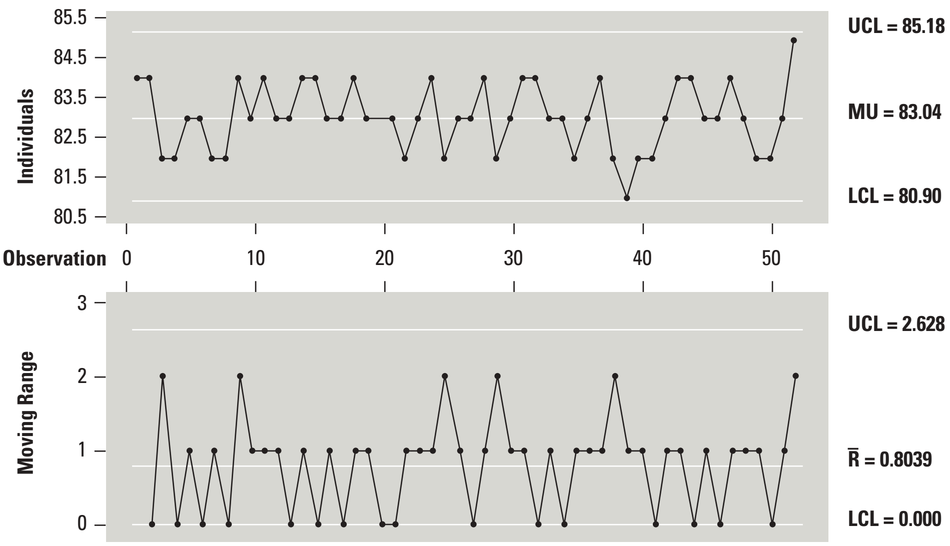
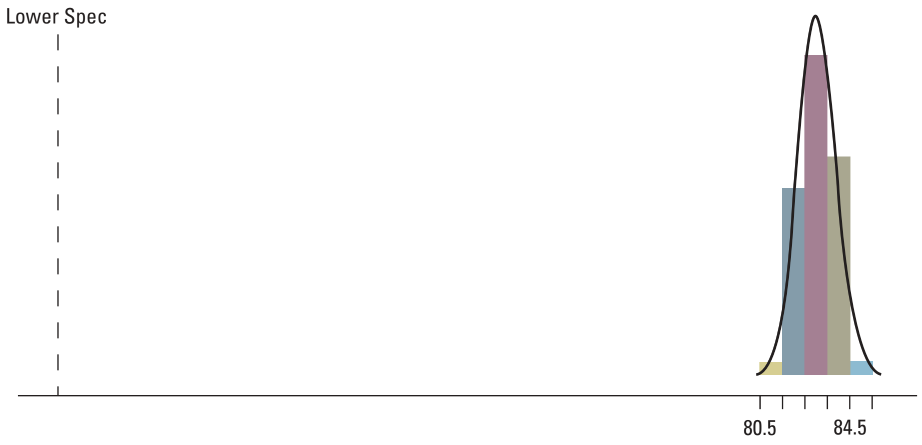


FIGURE 4 Capability Analysis for Method B



Method B									
Cp	*	Targ	*	Mean	83.0385	%>USL Exp	*	PPM>USL Exp	*
CPU	*	USL	*	Mean + 3s	85.5563	Obs	*	Obs	*
CPL	13.12	LSL	50	Mean - 3s	80.5207	%<LSL Exp	0.00	PPM<LSL Exp	0
Cpk	13.12	k	*	s	0.8393	Obs	0.00	Obs	0
Cpm	*	n	52						

foot-pounds), but it seemed to be capable. That is, the value of $CPL_A = 1.16$ indicated a reasonably good process. A CPL of 1 corresponds to a quality level of 3 sigma, therefore, a CPL of 1.16 seems to indicate a slightly better quality level, which was acceptable to the company.[5] This CPL gave rise to an estimated 253 parts per million nonconforming components (assuming a normally distributed process).

Method B had only a moderately high mean ($\bar{x}_B = 83$ foot-pounds) but a small standard deviation ($s_B = .84$

foot-pounds) This gave an extraordinarily large value of $CPL_B = 13.12$.

The technicians had hoped to find one process with a low CPL of .85 or .9 and, thus, eliminate it from further consideration. Method A had the higher mean (which they wanted) but Method B had the lower variation (which they also wanted). Discussion of these results with supervisors found champions for both methods, with method A being particularly attractive to many.

Critical element of analysis missing

The technicians decided to consult other quality professionals for assistance. These quality professionals complimented the technicians on the soundness of their work, but pointed out an often overlooked and certainly underemphasized point crucial to the analysis.

Even though most statistical packages print values like CPL = 1.16, this is not quite accurate. Just as $\bar{x}$ and s are sample point estimates for μ and σ, so also is the value computed and provided for the process index given by $(\bar{x} - LSL)/s$ and thus is actually a point estimate of CPL.

It's important to note that point estimates are provided by virtually all statistical packages of all process indices when such capability studies are done. Precisely written, the quality practitioners actually had $\widehat{CPL}_A = CPL_{A(est)} = (\bar{x}_A - LSL)/s_A$ and $CPL_B = \widehat{CPL}_{B(est)} = (\bar{x}_B - LSL)/s_B$. To provide a better sense of what values for CPL would be possible and reasonable, a confidence interval would be necessary, just as it is for $\bar{x}$ and s.[6]

While some think such an interval is unnecessary for a large sample of 52, in fact, all process capability indices' estimates have much greater variability than typically associated with $\bar{x}$. While still overlooked or misunderstood in practice, this fact has been pointed out by many, including Burt Gunter in his series published in *Quality Progress*.[7, 8, 9,10]

Complex distributions

The precise distributions for most process indices are quite complex. The distribution associated with CPL is a noncentral student's t statistic, for example, when the process is normally distributed. Much of the indices' distributional work has been compiled and developed in S. Kotz and N. L. Johnson's text, *Process Capability Indices*.[11]

Fortunately for the practitioner, Chou, Owen and Borrego have calculated and tabulated 95% lower confidence limits (95% one-sided confidence intervals) for C_p, CPU, CPL and C_{pk}. Table 2 offers a portion of one of their CPL tables.

To understand the use of Table 2, consider an example of a stable process that gave CPL = 1 from a sample of n = 40 observations. Table 2 indicates that with 95% confidence, CPL could be .79 or greater in size. Thus, a misconception that CPL = 1 (when in fact for a sample size of 40, it is $\widehat{CPL} = CPL_{(est)} = 1$) could lead to the conclusion that the process is at a 3-sigma level of quality (marginally acceptable for some processes) when in fact CPL could reasonably be .80, .85 or .90. Almost no one finds this level of quality acceptable!

TABLE 2 A Portion of a CPL Table

n =	40	50	75
CPL = 1.0	.79	.81	.85
1.1	.87	.90	.94
1.2	.96	.98	1.02

REFERENCE

1. Y. Chou, D. Owen and A. S. Borrego, "Lower Confidence Limits on Process Capability Indices," Journal of Quality Technology, July 1990, pp. 223-229.

The results are weighted and a method is chosen

Returning to the example of method A, doing linear interpolation (since n = 52 and $\widehat{CPL} = CPL_{(est)} = 1.16$) gives the following results from Table 2: There is a 95% confidence level that CPL is .9512 or greater in size for method A. This could reasonably allow (assuming a normal distribution and CPL = .9512) 2,200 parts per million nonconforming assemblies for this method. Such a level of quality would be unacceptable to some companies.

All of this work presupposes a forever statistically stable process for method A. It is precisely because this assumption is wildly naive that control charts are necessary to detect when a process becomes statistically unstable.

Most control charts have a reasonable probability of detecting such instability only after a shift of nearly one standard deviation or more in mean or variation. For method A, this could result in levels of nonconforming assemblies of 32,200 parts per million or more before detection would occur. Again, this level of quality would be unacceptable for virtually any company.

Method B has none of these pitfalls. In fact, $\widehat{CPL} = 13.12$ is indicative of very low variation ($s = .84$ foot pounds). While Chou, Owen and Borrego's table does not list such extraordinarily high process index values, a conservative, interpolated estimate would comfortably place CPL = 10 or greater with 95% confidence for method B. Should method B become statistically unstable, a one-standard deviation shift would only result in CPL = 9.67. Recalling that CPL = 2 corresponds to a 6-sigma level of quality and would result in a nonconforming rate of one per billion, it is clear that method B would still be providing conforming parts under circumstances that would be disastrous for method A.

If management were willing to devote the resources to either decrease the variation of method A or increase the mean of method B, such efforts could make method A the more desired method or even improve method B's already preferred status.

Neither the time nor the nature of the methods would allow the company to develop either method. A choice simply had to be made between the two methods as they currently existed.

The technicians and their supervisors, when apprised of the issues, appreciated the misunderstanding often associated with CPL versus $\widehat{CPL} = CPL_{(est)}$ and the potential problems for method A, even if monitored with a control chart. All concurred that method B was the preferred method to reach the levels of quality desired.

REFERENCES

1. Ford Motor Company, *Continuing Process Control and Process Capability Improvement* (Dearborn, MI: Ford Motor Company, 1983).
2. V. E. Kane, "Process Capability Indices," *Journal of Quality Technology*, Vol. 18, 1986, pp. 41-52.
3. Douglas C. Montgomery, I*ntroduction to Statistical Quality Control*, third edition (New York: John Wiley and Sons, 1997).
4. Donald J. Wheeler and Richard W. Lyday, *Evaluating the Measurement Process*, second edition (Knoxville, TN: SPC Press, 1989).
5. H. M. Wadsworthy, K. S. Stephens and A. B. Godfrey, *Modern Methods for Quality Control and Improvement* (New York: John Wiley and Sons, 1986).
6. R. H. Kushler and P. Hurley, "Confidence Bounds for Capability Indices," *Journal of Quality Technology*, October 1992, pp. 188-195.
7. B. Gunter, "The Use and Abuse of C_{pk}: Part 1," *Quality Progress*, January 1989, pp. 72-73.
8. B. Gunter, "The Use and Abuse of C_{pk}: Part 2," *Quality Progress*, March 1989, pp. 108-109.
9. B. Gunter, "The Use and Abuse of C_{pk}: Part 3," Quality Progress, May 1989, pp. 79-80.
10. B. Gunter, "The Use and Abuse of C_{pk}: Part 4," *Quality Progress*, July 1989, pp. 86-87.
11. S. Kotz and N. Johnson, *Process Capability Indices* (New York: Chapman Hall, 1993).

BIBLIOGRAPHY

Minitab: Reference Manual (State College, PA: MINITAB, July 1995).
Minitab: User's Guide, (State College, PA: MINITAB, July 1995).

LEROY A. FRANKLIN *is currently a statistics professor at Rose-Hulman Institute of Technology in Terre Haute, IN. He has a doctorate in statistics and probability from Indiana University-Bloomington. Franklin is an ASQ Senior Member.*

BELVA J. COOLEY *is an associate professor at the University of Montana in Missoula. She has a doctorate in management science from Oklahoma State University-Stillwater. Cooley is an ASQ certified quality engineer and member.*

GARY ELROD *is a quality engineer for Siemens Electromechanical Components in Franklin, KY. He holds a bachelor's degree in computer integrated manufacturing from Indiana State University-Terre Haute. Elrod is an ASQ member.*