Some Misconceptions about Confidence Intervals

By Keith M. Bower

Reprinted with permission from the American Society for Quality

Six Sigma practitioners often require a reasonable range of values for some characteristic of a process (e.g., the process mean, \( \mu \)). The confidence interval (CI) procedure, developed by Jerzy Neyman in the early 1930s, is typically employed. Confidence intervals (CIs) for the population mean (\( \mu \)) or standard deviation (\( \sigma \)) are widely used in manufacturing and business.

This article discusses three misconceptions associated with confidence intervals:

1. Overlapping confidence intervals imply “no difference.”
2. Confidence intervals are always appropriate.
3. Confidence intervals address the spread of individual values.

The CI Procedure

Consider a population that is well modeled by a normal (Gaussian) distribution. To obtain a reasonable estimate of the arithmetic mean for the population, \( \mu \), assume a sample of size \( n \) is randomly drawn from the population. The sample mean (\( \bar{x} \)) and sample standard deviation (\( s \)) are calculated. As discussed by Robert V. Hogg and Eliot Tanis, a CI for \( \mu \) may be constructed using the formula:

\[
\left[ \bar{x} - t_{\frac{\alpha}{2}} (n - 1) \left( \frac{s}{\sqrt{n}} \right), \bar{x} + t_{\frac{\alpha}{2}} (n - 1) \left( \frac{s}{\sqrt{n}} \right) \right]
\]

where \( t_{\frac{\alpha}{2}} (n - 1) \) is the inverse cumulative probability of a t-distribution with \( n-1 \) degrees of freedom at \( 1 - \frac{\alpha}{2} \). This value may be obtained from a t-distribution table or by using a statistical software package. Clearly, \( \mu \) either will or will not be covered in the interval; the theory is that with repeated samples, \( \mu \) would be contained in \( 100\% (1-\alpha) \) of all intervals in the long run. Note that the value of \( \alpha \) typically used is 0.05, leading to 95% CIs.
Misconception 1:

*Overlapping confidence intervals imply “no difference.”*

When comparing multiple means, practitioners are sometimes advised to compare the results from CIs and determine whether the intervals overlap. When 95% CIs for the means of two independent populations do not overlap, there will indeed be a statistically significant difference between the means (at the 0.05 level of significance). **However, the opposite is not necessarily true.**

As discussed by Nathanielt Schenker and Jane F. Gentleman, CIs may overlap yet there could be a statistically significant difference between the means. This is illustrated in Figure 1 below, where the p-value for the two-sample t-test is 0.04 (<0.05), yet the CIs overlap considerably.

**Fig. 1 Two-Sample T-Test and CI: A, B**

Two-sample T for A vs B

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
<td>0.05</td>
<td>1.06</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>-0.471</td>
<td>0.860</td>
</tr>
</tbody>
</table>

Difference = mu A - mu B

Estimate for difference: 0.525

95% CI for difference: (0.025, 1.025)

T-Test of difference = 0 (vs not =): T-Value = 2.10 P-Value = 0.040 DF = 58

Both use Pooled StDev = 0.967

Individual 95% CIs For Mean Based on Pooled StDev

| Level | N  | Mean   | StDev  | --------+---------+---------+---------+-- |
|-------|----|--------|--------|----------+---------+---------+---------+-- |
| A     | 30 | 0.0543 | 1.0642 | (----------*---------) |
| B     | 30 | -0.4707| 0.8600 | (----------*---------) |

Pooled StDev = 0.9675

-0.70 -0.35 0.00 0.35

N.B. When there are 3 or more levels of a factor, multiple comparison procedures are more appropriate to determine whether means are significantly different.

Misconception 2:

*Confidence intervals are always appropriate.*

If all observations in a population are measured, then inferential procedures (including confidence intervals) are without merit. For such a study, the sample mean would be the population mean, μ.
Importantly, however, many studies conducted by Six Sigma practitioners are analytic, a term discussed by W. Edwards Deming. Analytic studies are designed to assess the underlying causal system of a process. As such, the population under inspection therefore includes future production. Even with 100% inspection, CIs may therefore have some legitimacy (in particular, provided the process exhibits stability).

**Misconception 3:**

*Confidence intervals address the spread of individual values.*

Occasionally, practitioners are led to believe that a CI covers a particular proportion of the population (e.g., 95%). However, this is not the case for the most widely used CIs. In particular, the CI is a region one may expect a parameter (e.g., the population mean) to occupy. For the situation in which a reasonable upper and lower bound is required to cover a specific proportion of a population, tolerance intervals (TIs) may be appropriate.

**Tolerance Intervals**

TIs are constructed such that a specific proportion (p) of the population will be contained with a prescribed confidence level (1-α). For example, consider a manufacturer of machine parts that have a required diameter of 0.42 cm and specification limits of 0.42 ±0.02 cm. Assume 20 machine parts are randomly sampled from a stable process that is well modeled by a normal (Gaussian) distribution. Figure 2 shows the 95% CI for μ.

**Fig. 2**

**One-Sample T: Diameter**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95.0% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter</td>
<td>20</td>
<td>0.42328</td>
<td>0.01776</td>
<td>0.00397</td>
<td>(0.41497, 0.43160)</td>
</tr>
</tbody>
</table>

Since values between 0.40 cm and 0.44 cm are within specification limits, the results obtained from the sample (0.4150 cm to 0.4316 cm) appear to indicate that the process would be acceptable. However, it is crucial to note that the CI obtained is assessing the mean of the population – not the individual values. The results from a 95% confidence interval that covers 99% of the population are shown in Figure 3.

**Figure 3**

**Tolerance Interval for Diameter**

<table>
<thead>
<tr>
<th>Tolerance (Confidence) Level:</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of Population Covered:</td>
<td>99%</td>
</tr>
<tr>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>20</td>
<td>0.423285</td>
</tr>
</tbody>
</table>
These results (0.3591 cm to 0.4875 cm) indicate that the process itself is actually doing poorly; i.e., a high proportion falls outside the specification limits.

Implications

Though CIs are very useful tools for gaining process knowledge, it is crucial to consider drawbacks in their interpretation. For more information on the relationship between CIs and TIs, see Robert V. Hogg and Johannes Ledolter.  

About the Author

Keith M. Bower is a technical training specialist with Minitab Inc. He received a bachelor’s degree in mathematics with economics from Strathclyde University in Great Britain and a master’s degree in quality management and productivity from the University of Iowa in Iowa City, USA. Bower has conducted training courses for firms including GE Capital, American Express, and Motorola. He is a member of ASQ and the Six Sigma Forum.

Technical Information: For more information on the macro for tolerance intervals, refer to Answer ID #1216 at www.minitab.com/support/answers

References