

Statistical Intervals

Part 1: The Confidence Interval

BY STEPHEN N. LUKO AND DEAN V. NEUBAUER

Statistical intervals play an important role in statistical inference. But practitioners often confuse the difference among confidence, prediction and tolerance type intervals. This article series will clarify these differences and indicate when each type of interval should be used.

Q: What is a confidence interval?

A: The most common type of statistical interval used in practice is the confidence interval (CI).¹ CIs apply to unknown parameters of either a probability distribution or some other statistical model, such as the parameters in a regression model. Common examples of parameters to which this might be applied include a mean, a standard deviation, a percentile or a proportion. To construct a two-sided CI for an unknown parameter, we start with some observed data representing the model we are using and proceed to calculate the confidence limits a and b (with $a < b$) such that it may be claimed that the parameter lies between a and b with the stated confidence. We say that the interval would capture or cover the parameter of interest with the stated confidence. The term confidence refers to the long run proportion of the time (probability) that the interval would capture the parameter of interest. For example, if the unknown parameter is the mean, μ , of a normal distribution, the two-sided CI would take the form $a \leq \mu \leq b$. For one-sided intervals, this would take the form $\mu \leq b$ or $\mu \geq a$. It is common practice to call the confidence value a confidence coefficient and denote it as $100(1 - \alpha)\%$, where $0 \leq \alpha \leq 1$. In other words, there is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ . Confidence coefficients of 90 percent, 95 percent and 99 percent are commonly used.

The manner in which the CI end points are calculated is context dependent. If the parameter of interest is a mean and the standard deviation is known, then the CI calculation is from Equation 1, where the standard normal distribution is used to produce the desired confidence level.

$$\bar{x} \pm z_{\alpha/2} \sigma / \sqrt{n}, \text{ with confidence } 100(1 - \alpha)\% \quad (1)$$

The quantity $z_{\alpha/2}$ is the upper $100(\alpha/2)$ percentage point from the standard normal distribution, such that $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$. Values of $z_{\alpha/2}$ for 90 percent, 95 percent and 99 percent confidence are 1.645, 1.960 and 2.575, respectively.

EXAMPLE 1

ASTM E23, Test Methods for Notched Bar Impact Testing of Metallic Materials, defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch method measures impact energy (J) and is used to determine if a material has a ductile-to-brittle transition with decreasing temperature. Suppose that $n = 10$ measurements of impact energy on specimens of A238 steel cut at 60°C are 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2 and 64.3. If we assume that impact energy is normally distributed with $\sigma = 1$ J, a 95 percent CI for the mean impact energy, μ , can be determined from Equation 1.

$$\begin{aligned} \bar{x} - z_{\alpha/2} \sigma / \sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2} \sigma / \sqrt{n} \\ 64.46 - 1.96(1/\sqrt{10}) \leq \mu \leq 64.46 + 1.96(1/\sqrt{10}) \\ 63.84 \leq \mu \leq 65.08 \end{aligned}$$



This CI indicates that the true value for the mean impact energy for A238 steel at 60°C lies between 63.84J and 65.08J with 95 percent confidence. In practice, any CI either contains the true value or it does not. The correct interpretation of the CI is to say that if many random samples were taken, and a 95 percent CI were computed from each sample, then 95 percent of these intervals would contain the true value of the mean impact energy, μ . In other words, 5 percent of these CIs would fail to contain μ .

Since σ is typically unknown, we must estimate it with the sample standard deviation s and use Equation 2, where the student's t distribution is used to produce the desired confidence level.

$$\bar{x} \pm t_{\alpha/2} s / \sqrt{n}, \text{ with confidence } 100(1 - \alpha)\% \quad (2)$$

The quantity $t_{\alpha/2}$ is the upper 100($\alpha/2$) percentage point from student's t distribution with $n - 1$ degrees of freedom such that $P(t \leq t_{\alpha/2}) = 1 - \alpha/2$. If the parent population from which the data originated is a normal distribution, then Equation 2 is always valid for any sample size. If the parent population is not normally distributed, then Equation 2 is still valid if the sample size is large enough, say at least 30. This is due to the central limit theorem, which states that averages from any parent population will be normally distributed.

The interval construction in Equation 2 is said to capture or contain the unknown mean, μ , with the stated confidence 100($1 - \alpha$)%. For sample mean and sample standard deviation s , we can write a probability statement for Equation 2, assuming that the sample comes from a normal distribution or that the sample size is at least 30. This statement takes the form of Equation 3.

$$P(\bar{x} - t_{\alpha/2} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2} s / \sqrt{n}) = 1 - \alpha \quad (3)$$

Equation 3 is a true probability statement because \bar{x} and s are random variables in this expression. We appeal to the process of implementing Equation 3 many times and define confidence as the long run proportion of the cases that the parameter would be captured by these intervals. In such a process, approximately 100($1 - \alpha$)% of the time the interval in Equation 3 would capture the true mean, μ .

EXAMPLE 2

Suppose that $n = 22$ tensile adhesion tests were made on U-700 alloy specimens. The load of each specimen at failure in megapascals is given in Table 1.

Table 1 – Alloy Specimen Load at Failure, in Megapascals

19.8	10.1	14.9	7.5	15.4	15.4
15.4	18.5	7.9	12.7	11.9	11.4
11.4	14.1	17.6	16.7	15.8	19.5
8.8	13.6	11.9	11.4	–	–

The sample statistics are: $\bar{x} = 13.71$ and $s = 3.55$. The value of student's t using 21 degrees of freedom and two sided 95 percent confidence is $t = 2.08$. Using Equation 2, the 95 percent CI is

$$\begin{aligned} \bar{x} - t_{\alpha/2} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2} s / \sqrt{n} \\ 13.71 - 2.080(3.55/\sqrt{22}) \leq \mu \leq 13.71 + 2.080(3.55/\sqrt{22}) \\ 12.14 \leq \mu \leq 15.28 \end{aligned}$$

Based on the central limit theorem, we can also state a more general form for a confidence interval for any parameter of interest, say θ . Let $\hat{\theta}$ be the estimate of the parameter and let $SE(\hat{\theta})$ be its estimated standard error. If the sample size is at least 30, then the following is an ap-

proximate general expression for a confidence interval construction for the parameter θ . Here the standard normal value z appropriate for the two-sided $100(1 - \alpha)\%$ confidence interval is used.

$$\hat{\theta} \pm z_{\alpha/2} SE(\hat{\theta}) \quad (4)$$

Suppose that we wish to construct a confidence interval for a population proportion, p . When we observe x observations in the sample of size n that belong to a class of interest, we can estimate the population proportion as $\hat{p}=x/n$. For large n , the sampling distribution of \hat{p} is approximately normal with mean p and standard deviation approximated as $\sqrt{\hat{p}(1-\hat{p})/n}$ whenever both $n\hat{p}$ and $n(1-\hat{p})$ are both at least 5. The confidence interval construction for p is shown in Equation 5.

$$P\left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha \quad (5)$$

EXAMPLE 3

A random sample of $n = 85$ automobile crankshaft bearings, in which $x = 10$ have a rejectable surface finish, is examined. The point estimate of the proportion of crankshaft bearings in the population outside the roughness specification is $\hat{p}=x/n=10/85=0.12$. A 95 percent CI for p is computed from Equation 5.

$$\begin{aligned} \hat{p} - z_{.05/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &\leq p \leq \hat{p} + z_{.05/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ 0.12 - 1.96 \sqrt{\frac{0.12(0.88)}{85}} &\leq p \leq 0.12 + 1.96 \sqrt{\frac{0.12(0.88)}{85}} \\ 0.05 &\leq p \leq 0.19 \end{aligned}$$

To summarize, a confidence interval for an unknown parameter of a statistical model is a random interval that probably covers or contains the true value of the parameter for the sample we are using. The containment probability, called confidence, does not apply to the particular interval we are calculating. Confidence applies to the run proportion of such intervals that would be expected to cover the parameter if we were to use many samples of the same size under the same conditions.

REFERENCE

1. Most textbooks in general statistics have confidence interval formulations for various parameters and distribution assumptions. Two of these texts are: Montgomery, D.C., and Runger, G.C., *Applied Statistics and Probability for Engineers*, 4th ed., Hoboken, N.J., Wiley, 2006; and Duncan, A. J., *Quality Control and Industrial Statistics*, 5th edition, Homewood, Ill., Irwin, 1992.

STEPHEN N. LUKO, Hamilton Sundstrand, Windsor Locks, Conn., is the immediate past chairman of Committee E11 on Quality and Statistics, and a fellow of ASTM International.

DEAN V. NEUBAUER, Corning Inc., Corning, NY, is an ASTM fellow; he serves as vice chairman of Committee E11 on Quality and Statistics, chairman of Subcommittee E11.30 on Statistical Quality Control, chairman of E11.90.03 on Publications and coordinator of the DataPoints column.

In the next article in this series, we will discuss prediction intervals for a future observation.

Statistics play an important role in the ASTM International standards you write, and a panel of experts is ready to answer your questions about how to use statistical principles in ASTM standards. Please send your questions to SN Editor in Chief Maryann Gorman at mgorman@astm.org or ASTM International, 100 Barr Harbor Drive, P.O. Box C700, West Conshohocken, PA 19428-2959.