

The Standard Error Concept

Measuring a Random Error in a Reported Statistic

BY STEPHEN N. LUKO

Q: What is a standard error and how is this used in practice?

A. One of the most useful concepts in statistical practice is the “standard error” concept. This term was originally defined by British statistician Udny Yule in the early 20th century. ASTM E2586, Practice for Calculating and Using Basic Statistics, defines standard error as “The standard deviation of the population of values of a sample statistic in repeated sampling, or an estimate of it.” The term “uncertainty” is closely related to standard error and has been given considerable treatment in recent decades. Standard error measures random error in a reported statistic: the kind of error due to random sampling variation in repeating a test under the same conditions. Uncertainty is a broader concept that considers additional components of potential error besides random error. ASTM E2655, Guide for Reporting Uncertainty of Test Results and Use of the Term Measurement Uncertainty in ASTM Test Methods, describes the use of the uncertainty concept as applied to a test result.

Practitioners and decision makers who use data are typically more concerned with statistics than with individual measurements in a set of data. Users of data want to see means, variances, ranges, proportions, maximum or minimum values, percentiles or other statistics. What is often not fully appreciated is that statistics also behave in a random fashion similar to individual measurements, and this is measured by the standard error. When a sample mean is reported, it is not the “true” average but an estimate of it. The sample statistic may be somewhat higher or lower than the unknown true value. The standard error of the mean measures how different the true mean might be from the statistic that is reported. More generally, we can speak of the “standard error of the estimate” anytime an estimated statistical quantity is reported.

When a single statistic is calculated, it is possible to calculate the standard error of the estimate. Generally, the larger the sample size, the smaller the standard error of an estimated quantity.

To see how this works, consider a sample mean. From a sample of size n , the sample mean and standard deviation are calculated. In reality there is a true mean, μ , and true standard deviation σ , and these are unknown. The sample gives us the estimates \bar{x} and S . If we were to sample repeatedly from the population/process from which the sample came and calculate the sample mean again and again, the standard deviation of the distribution of means would be the true standard error of the mean. Theoretically, this is Equation 1:

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} \quad (1)$$

Since we only have one estimated mean, and we do not know the true σ , we can only estimate the standard error as:

$$SE(\bar{x}) \approx \frac{S}{\sqrt{n}} \quad (2)$$

The error in a reported result is called sampling error, and this is measured as an absolute deviation from the unknown true value. Thus, for a mean, the sampling error can be thought of as the deviation $|\bar{x} - \mu|$. Approximately 68 percent of the time the sampling error will be at most one standard error in size; and 95 percent of the time it will be at most two standard errors in size. We can state this more concisely as:

$$|\bar{x} - \mu| \leq S / \sqrt{n} \quad \text{with 68 percent confidence} \quad (3)$$

$$|\bar{x} - \mu| \leq 2S / \sqrt{n} \quad \text{with 95 percent confidence} \quad (4)$$

This gives the user of a statistic some idea of the magnitude of the difference that might have



been realized in practice, how the sample size affects the possible error of the estimate and with what approximate probability (confidence). Here we are assuming a sample size of 20 or more and are using the theory of the normal distribution. Some readers will also recognize this as akin to the construction of a confidence interval for an unknown mean. A full treatment of confidence intervals is discussed in ASTM E2586 and an earlier DataPoints article on this topic has been published.¹

EXAMPLES

Suppose in a sample of size $n = 20$ the sample mean and standard deviation were found to be 162 and 11.5, respectively. The estimated standard error of the mean is from Equation 2: $11.5/4.47 = 2.57$. Thus, the potential for an error in the reported result is no more than ± 2.57 (68 percent confidence) or no more than $2(2.57) = \pm 5.14$ (at 95 percent confidence).

One of the most commonly used statistics is a simple proportion. There is a sample of objects size n , and we observe each object for the occurrence of an attribute. Each object either has or has not the attribute. This is the situation, for example, in quality control sampling or in public opinion polling. The statistic, denoted \hat{p} , is the proportion in the sample having the attribute. The true and unknown proportion of all objects is p . The theoretical standard error of the estimate is:

$$SE(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} \quad (5)$$

In practice we never know the true p so we substitute the statistic \hat{p} and obtain an estimate of the standard error. Using Equation 5, the estimated standard error is:

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (6)$$

When this technique is used with a political or marketing research poll, the quantity $2SE(\hat{p})$ is referred to as the margin of error for the poll. Suppose in a sample of $n = 200$ inspected metal components, 23 were classified as defective. The estimate of the process proportion defective is $\hat{p} = 23/200 = 0.115$ or 11.5 percent.

The standard error of this estimate is, using Equation 6, is 0.0226 or 2.26 percent. Should we want to claim an approximately 95 percent confidence in the possible error in the result we should report the result using two standard errors or as 11.5 percent ± 4.52 percent. In any case we should at least report the standard error (2.26 percent) along with the estimate.

Standard error formulas are available for several common cases in ASTM E2586. Other cases and methods are available in the literature of statistical science.

REFERENCE

1. Stephen N. Luko and Dean V. Neubauer, "Statistical Intervals, Part 1: The Confidence Interval," *ASTM Standardization News*, Vol. 39, No. 4, July/Aug. 2011.

STEPHEN N. LUKO, *United Technologies Aerospace Systems, Windsor Locks, Conn.*, is an ASTM fellow; a past chairman of Committee E11 on Quality and Statistics, he is current chairman of Subcommittee E11.30 on Statistical Quality Control.

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

snonline

Get more tips for ASTM standards development at www.astm.org/sn-tips.

Find other DataPoints articles at www.astm.org/standardization-news/datapoints.