What Are Z-Scores?

A Simple but Useful Statistical Computation

BY STEPHEN N. LUKO

Q: What exactly are z-scores and how are these statistics used in practice?

A. Given a set of sample data of size n, suppose we calculate the sample mean, $\bar{x}$, and standard deviation, s, in the ordinary way. For review, these formulas are:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \qquad (1)$$

Now let x_i be any individual sample value. The z-score associated with x_i is calculated as:

$$z_i = \frac{x_i-\bar{x}}{s} \qquad (2)$$

For example, suppose $\bar{x}$ = 120 and s = 7. What is the z-score associated with the sample value x = 105.8? Using formula 2 gives us z = (105.8-120)/7 = -2.03, rounded to 2 significant digits to the right of the decimal point. The value of z = -2.03 means that the sample value x = 105.8 is approximately 2.03 standard deviations to the left of the sample mean. For a sample value of x = 129.2, we find, in the same way, z = 1.31, approximately. In this case, the value of z = 1.31 says that x = 129.2 lies 1.31 standard deviations to the right of the sample mean. That is the essence of the z-score; it indicates how far from the mean, in the units of standard deviations, a particular sample value lies.

A negative z-score indicates a value to the left of the mean, and a positive z-score indicates a value to the right of the mean. The magnitude of the z-score tells you how many standard deviations the associated value of x is from the mean. One interesting property of z-scores is the fact that the sample z-scores will always have a mean of 0 and standard deviation of 1. This means that you can compute z-scores for several datasets and compare them directly as to the relative variation of each set of data values about their respective means.

The term z-score is usually reserved for sample data, but we can perform similar calculations for a theoretical distribution. If a variable x has some distribution with mean μ and standard deviation σ, we say $z = (x - \mu)/\sigma$ is its standardized form. In particular, if x has a normal distribution we say that z, as defined here, has a standard normal distribution – one with mean 0 and standard deviation 1. In the case of the standard normal distribution, we can calculate exactly the theoretical probability that z exceeds or is less than a particular value. For this we need a table of the standard normal distribution, or we can use any of a number of computer programs. The following intervals were calculated for the standard normal distribution using Microsoft Excel 2010.

Table 1 – Standard Normal Distribution for Selected Intervals

Interval	Probability	Interval	Probability	Interval	Probability
Z < -1	0.1587	Z > 1	0.1587	-3 < Z < 3	0.9973
Z < -2	0.0228	Z > 2	0.0228	-2 < Z < 2	0.9545
Z < -3	0.0013	Z > 3	0.0013	-1 < Z < 1	0.6827

The three entries in the far right column of Table 1 form the basis for the well-known "empirical rule" in classical statistics. This rule says that for mound-shaped and approximately symmetric distributions, approximately 68.3 percent of the

time the variable falls within 1 standard deviation of the mean; 95 percent within two standard deviations of the mean and 99.7 percent within three standard deviations of the mean. This rule is exactly true for a theoretical normal distribution with known mean and standard deviation. In the case of statistics, though, we do not know the values of μ and σ, nor do we know with certainty if the normal distribution is the correct model for the data we have. It is common practice, though, to apply Table 1 to sample data when we have a mound-shaped symmetric distribution and are using a reasonably large sample size, say 100 or more. The rule is approximate with sample data. For more accurate statements, other methods such as tolerance or prediction intervals, both distribution-dependent and nonparametric (no distribution assumed), can be used.

Many practitioners use sample z-scores as an informal test for detecting outliers or unusually large (or small) sample values. From Table 1 we see that an interval of ±3 sigma has a probability of 99.73 percent of containing random values from the distribution (assuming normality). Then there is a probability of 0.27 percent (or about 3/1000) of observing values beyond 3 sigma in either direction. Sample z-scores greater than (less than) 3 (-3) are said to be potential outliers. For example, this is the basis of the ±3 sigma limits on a control chart. Some caution is advised in using this technique. Many people ignore the sample size in using z-scores for outlier detection. Caution is advised, and it may be shown (see Reference 1) that a sample z-score beyond ±3 is not possible unless that sample size is at least 11. On the other extreme, for very large sample sizes, we expect some z-scores beyond ±3. For example, if $n = 2000$ and the normal distribution applies, we would expect between five and six observations indicating a z-score beyond ±3. ASTM E178, Practice for Dealing with Outlying Observations, does contain a more precise method that uses extreme z-scores in a sample of size n. Here, tables are provided that give an exact critical value, at some significance level, for the extreme z-score when the normal distribution applies.

In other quarters, z-scores are used as a ranking metric. This is common practice, for example, in education where test scores are often presented in standardized form (as a z-score). Bearing in mind the empirical rule, this approach gives us an approximate measure of how extreme any score is with respect to the other sample values. Ranking in this way also makes sense for other disciplines such as interlaboratory studies.

The z-score is one of the simplest computations used in elementary statistics, but it can be a very useful device for many applications.

REFERENCE
1. Shiffler, R. E., "Maximum Z Scores and Outliers," The American Statistician, Vol. 42, No. 1, Feb. 1988, pp. 79-80.

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, N.Y., 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.*

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.*