

5.3: z-Scores

A z-score is used to summarize how far an individual case is from the mean in standard deviations. The z-score is often also referred to as the standard score. Each z-score indicates two important things about the location of a raw score: its direction and distance from the mean.

Direction

The sign of a z-score indicates its direction. When a raw score does not deviate from the mean, it is equal to the mean. However, raw scores can also deviate by being below the mean or above the mean. When $z = 0$, it means the raw score did not deviate from the mean and is, thus, equal to the mean. Keep in mind that 0 is not considered a signed number because it represents the dividing point between signed numbers (which are either positive or negative). When a z-score is positive, it means that the raw score was greater than the mean. When a z score is negative, it means a raw score was lesser than the mean. Therefore, by looking at the sign of any z-score, you can instantly deduce whether the raw score was greater, less than, or equal to the mean.

Distance

The size of the z-score indicates distance from the mean. Because z-scores are computed to show how many standard deviations a raw score is from a mean, the value of the z-score can be used to quickly deduce how far a raw score is from a mean. The smaller the absolute value of a z-score, the closer a raw score is to the mean. Conversely, the larger the absolute value of a z score, the further a raw score is to the mean. When a raw score is equal to the mean, $z = 0$; zero represents the absence of distance from the mean.

z-Score Formula

Because z-scores are quite useful, raw scores are often converted to z-scores. This is done using a simple two step formula:

Z-Score Computations

Formula	Calculation Steps
$z = \frac{x - \mu}{\sigma}$	<ol style="list-style-type: none">1. Subtract the mean from a given raw score to find the deviation.2. Divide the deviation by the standard deviation.

Notice that the z-score formula uses population symbols. This is the standard way to write the formula because the normal curve is focused on broader truths that are presumed to apply to populations. However, sometimes the formula is written with sample symbols when working with sample data. The steps are the same regardless of whether the sample or population symbols are used. The z-score formula written with the sample symbols looks like this:

$$z = \frac{x - \bar{x}}{s}$$

The formula requires knowing the mean and standard deviation for a set of scores before plugging those in to convert any given raw score to its z-score. If the mean and standard deviation are already known, they can be plugged in with a given raw score and used to evaluate z. When we say we are “evaluating” with a formula, it means we are simply computing a result by following the steps of a formula or algebraic expression.

Finding z with a Given M and SD

Let’s try with an example. Suppose data were gathered on the amount of coffee beans customers purchased measured in ounces. The quantitative variable being measured is *weight*. The unit of measurement is *ounces*. The sample is *bags of coffee sold*. Weight is our X variable and each individual raw score is an example of X. Suppose that when the data were summarized, it was found that the mean was 16.00 ounces with a standard deviation of 4.00. These summaries are commonly provided in APA-formatted sentences like this:

The mean weight for the sample of coffee bags was 16.00 ounces ($SD = 4.00$)

Now that we are given the mean and standard deviation, we can convert any given raw score to its corresponding z-score. Suppose we want to know the z-score for the first bag of coffee sold and that this bag weighed 12.00 ounces. We could summarize this by saying $X_1 = 12.00$. You can read this as saying “The raw score of X for case 1 was 12.00.” Now we have all that we need to find the z-score for the raw score of 12.00. Let’s organize our information to get ready for using the formula:

$$\bar{x} = 16.00$$

$$s = 4.00$$

$$X_1 = 12.00$$

Now we can plug these into the z-score formula to find z. We will use the sample symbols to remind ourselves that we are estimating using sample information rather than known population values.

Tip

Always start by writing the formula down before plugging in any values. This helps you get familiar with the formula and its parts. This can help you to both recognize and memorize the formulas.

$$z = \frac{x - \bar{x}}{s}$$

$$z = \frac{12.00 - 16.00}{4.00}$$

$$z = \frac{-4.00}{4.00}$$

$$z = -1.00$$

We can follow this same process to convert any raw score to a z-score. Here are three examples of different raw scores for weights of coffee bags being converted; this example is focused on a sample and, thus, sample symbols are used:

Example 1	Example 2	Example 3
Given: $\bar{x} = 16.00$ $s = 4.00$ $X_1 = 21.00$ Find z_1 : $z = \frac{x - \bar{x}}{s}$ $z = \frac{21.00 - 16.00}{4.00}$ $z = \frac{5.00}{4.00}$ $z = 1.25$	Given: $\bar{x} = 16.00$ $s = 4.00$ $X_2 = 8.80$ Find z_2 : $z = \frac{x - \bar{x}}{s}$ $z = \frac{8.80 - 16.00}{4.00}$ $z = \frac{-7.20}{4.00}$ $z = -1.80$	Given: $\bar{x} = 16.00$ $s = 4.00$ $X_3 = 16.40$ Find z_3 : $z = \frac{x - \bar{x}}{s}$ $z = \frac{16.40 - 16.00}{4.00}$ $z = \frac{0.40}{4.00}$ $z = 0.10$

Finding z Starting with Raw Data.

Sometimes raw data are provided which require the statistician to do a little more work before using the formula. When raw scores are provided, the statistician must first find the mean and standard deviation before plugging them in to solve for the z-score of a given raw score.

Let's take a look at Data Set 5.1; this data set is for the variable *volume of coffee consumed* measured in ounces for a sample of 22 customers. First, we can start by finding sample size, then the mean, and finally, the standard deviation. A summary of these calculations is shown at the bottom of Data Set 5.1. The standard deviation is shown rounded to the fourth decimal place (also known as rounded to the ten-thousandths place). For a detailed review of how to calculate the mean, see Chapter 3. For a detailed review of how to calculate the standard deviation, see Chapter 4.

Once the mean and standard deviation have been computed for a sample, the z-score formula can be used to calculate the z-score for any raw score. Let's find the z-score for the first person in the data set who had a raw score of 12. Thus, $X_1 = 12.00$.

Data Set 5.1

Volume of Coffee Consumed in Ounces	
12	8

Volume of Coffee Consumed in Ounces	
11	8
10	7
10	7
9	7
9	7
9	7
9	6
9	6
8	5
8	4
Descriptive Statistics	
$n = 22$	
$\bar{x} = 8.00$	
$s \approx 1.9024$	

Calculations with Data Set 5.1

Given:

$$\bar{x} = 8.00$$

$$s = 1.9024$$

$$X_1 = 12.00$$

Find z_1 :

$$z = \frac{x - \bar{x}}{s}$$

$$z = \frac{12.00 - 8.00}{1.9024}$$

$$z = \frac{4.00}{1.9024}$$

$$z \approx 2.1026$$

The z-score for X_1 is 2.10 when rounded to the hundredths place. It can be summarized as: $Z_1 = 2.10$

Thus, the only difference in finding a z-score when given raw data is that you must first calculate the mean and standard deviation. After that, values can be plugged into the z-score formula to solve for z.

Converting from z-Scores to Raw Scores

Sometimes z-scores are given yet a statistician wants to know the raw scores instead. When z-scores, the mean, and the standard deviation are provided, you can plug what is known into the formula and solve for x. This becomes an algebra problem for which we need to isolate the unknown (x) on one side of the formula to solve, as shown in the example below. For this section we will presume we have population data and, thus, will use population symbols for the formula.

Formula and Steps	Explanation
Given: $\mu = 16.00$ $\sigma = 4.00$ $z_1 = 2.00$	Start by identifying what is known and the formula needed.

Formula and Steps	Explanation
Find X_1 : $z = \frac{x - \mu}{\sigma}$	
$2.00 = \frac{x - 16.00}{4.00}$	Plug in what is known. The remaining unknown is x . You can see that the value being solved for (which is x) is not isolated on one side of the equation. Thus, steps need to be used, one at a time, to isolate the unknown.
$2.00(4) = \frac{x - 16.00}{4.00}(4.00)$	Remove divide by 4 by doing its opposite to both sides. This means we must multiple each side of the equation by 4.
$8.00 = x - 16.00$	Things are looking simpler but the x is still not isolated so an additional step is needed
$8.00 + 16.00 = x - 16.00 + 16.00$	Remove subtract 16 by doing its opposite to both sides. This means we must add 16 to each side of the equation.
$24.00 = x$	We have solved for x and now know it was 24.00
$x = 24.00$	Optional: Some people prefer to reorder the result so it reads from left to right the same way we would typically say it " x equals 24.00"

We can follow this same process to convert any z-score to its corresponding raw score for a data set. Here are three examples of different raw scores being converted following the example from above:

Example 1	Example 2	Example 3
Given: $\mu = 16.00$ $\sigma = 4.00$ $z_1 = 21.00$ Find X_1 : $z = \frac{x - \mu}{\sigma}$ $2.00 = \frac{x - 16.00}{4.00}$ $2.00(4.00) = \frac{x - 16.00}{4.00}(4.00)$ $8.00 = x - 16.00$ $8.00 + 16.00 = x - 16.00 + 16.00$ $24.00 = x$	Given: $\mu = 16.00$ $\sigma = 4.00$ $z_1 = 8.80$ Find X_1 : $z = \frac{x - \mu}{\sigma}$ $0.40 = \frac{x - 86.00}{7.50}$ $0.40(7.50) = \frac{x - 86.00}{7.50}(7.50)$ $3.00 = x - 86.00$ $3.00 + 86.00 = x - 86.00 + 86.00$ $89.00 = x$	Given: $\mu = 16.00$ $\sigma = 4.00$ $z_1 = 16.40$ Find X_1 : $z = \frac{x - \mu}{\sigma}$ $-1.50 = \frac{x - 55.00}{1.62}$ $-1.50(1.62) = \frac{x - 55.00}{1.62}(1.62)$ $-2.43 = x - 55.00$ $-2.43 + 55 = x - 55.00 + 55.00$ $52.57 = x$

The Purpose of z-Scores

Raw scores are converted to z-scores and used to create something known as a standard distribution. Standardizing a distribution refers to the act of transforming from a normal curve with raw scores to an equivalent normal curve with z-scores. Thus, a standard distribution is a normal curve with z-score locations as the anchors for the x-axis in place of their raw score counterparts. Standardizing in this way has several practical applications.

First, it makes it easy to immediately deduce whether the raw score was greater or less than the mean or other measures of central tendency including the median and the mode. By simply checking whether the z-score is positive or negative you can deduce whether the raw score was greater than or less than the mean, respectively.

Second, knowing the z-score makes it easy to tell how common and likely or rare and unlikely a score is. The closer the z-score is to 0, the more common it and its corresponding raw score are. The farther the z-scores is from 0, regardless of whether it is positive or negative, the rarer it and its corresponding raw score are.

Third, and relatedly, z-scores help us identify outliers. Outliers are extreme, rare scores which would be found in the tails of the graph. Outliers have large z-scores indicating they are rare and far from most other scores in a normal distribution. However, the boundaries for which z scores are considered outliers are a bit flexible and depend upon aspects of the variable and data such as sample size and importance of skew (which is caused by outliers). Because there is flexibility, it is useful to review a few z-scores and why they are or are not generally used as outlier boundaries.

Let's consider the boundaries for outliers in a normal distribution. Approximately 68.26% of raw scores are within one standard deviation of the mean and are not considered outliers in most circumstances. This means z-scores between -1.00 and 1.00 are not considered outliers. However, z-scores of ± 1.96 are sometimes used to identify outliers. This is because approximately 95.00% of raw scores are within 1.96 standard deviations of the mean; sometimes, therefore, z-scores beyond this range are considered outliers because they represent the rarest 5% of scores. Thus, when ± 1.96 is used as the cutoff or boundary for identifying rare scores, z-scores outside this range (such as z-score of -2.00 or 2.00) are considered outliers. The use of this boundary can be more useful when a sample size is small. However, you may also see other cutoffs recommended.

Two other cutoffs are also often considered: z-scores of ± 2.58 and z-scores of ± 3.00 . Though there is some gray area about what is considered an outlier, z-scores at or beyond -2.58 and 2.58 are often considered outliers. This is because approximately 99.00% of scores are within ± 2.58 standard deviations of the mean (meaning they would be between z-scores of -2.58 and 2.58). When this boundary is being used, the statistician is defining the rarest 1% of scores as outliers.

However, others prefer using ± 3.00 . This would mean that the 0.26% of scores, which are found outside the boundaries of -3.00 and 3.00, would be considered outliers and the other 99.74% of scores within the boundaries of -3.00 and 3.00 would not be considered outliers. Though this is a bit of a strange cutoff (because it defines the rarest 0.26% as outliers), it is sometimes used because of its ease. When a normal curve is sketched, the x-axis is often drawn showing z-score locations starting from -3.00 and ending at 3.00. Thus, when the normal curve is sketched, these boundaries are used because nearly all scores fall between them and it is simpler to show integers rather than decimal numbers such as 1.96 or 2.58. Thus, ± 3.00 is sometimes used as a simple rule of thumb for identifying outliers visually. Nevertheless, for some statisticians ± 2.58 is preferred over 3.00 when working with larger samples because it is based more on mathematical logic than convenience.

Fourth, z-scores allow for comparison of scores within a group/sample. By knowing the z-scores of two individuals in a group, it is easy to identify which is rarer (when data are normally distributed). Suppose the mean is 88.00 and the *SD* is 6.60. You have two raw scores of 89.65 and 84.70. Suppose you want to know which score is rarer. This is hard to quickly deduce from the information provided (but it is possible). However, suppose that instead of the two raw scores you have their z-scores which are 0.25 and -0.50, respectively. You don't need to look at the mean, the *SD*, or even the raw scores themselves to know which is more common (i.e. the z-score of 0.25) and which is rarer (i.e. the z-score of -0.50). Thus, using z-scores makes it easy to quickly compare scores within a group or sample.

The fifth reason z-scores get used is a bit more complex but is especially useful: z-scores allow us to compare scores of different quantitative variables to each other. For example, we may want to compare a student's score on an English exam to their score on a psychology exam to see if their relative performance on each was approximately equivalent. Relative here refers to how well they performed compared to others on each exam and on each exam compared to the other exam. Making a direct comparison of raw scores for each exam is limited but useful comparisons of standardized scores (z-scores) on each exam are possible. Suppose the student's score on the English exam was 80 points and their score on the psychology exam was 65 points. We may assume that the English score indicates the higher relative skill simply because is the larger number. However, we have no indication here as to how many points were possible on each exam nor how well others did on each exam. If the English exam was out of 200 points and the psychology exam was out of 100 points, then the grade as a percentage would actually be higher for the psychology exam (65%) than for the English exam (40%). This is a way of comparing data many are accustomed to but it cannot tell us how well the student did relative to others. For that, a standardized score, such as a z-score, is needed.

Using a z-score for each exam allows a statistician to identify on which exam, if either, a student had better performance relative to others. Let's take the example of the student with scores of 80 and 65 on an English exam and a psychology exam, respectively. Suppose the mean score on the English exam was 70 with a standard deviation of 10 and that the mean score of the psychology exam was 50 with a standard deviation of 5. Presume the means and *SDs* represent the population of students. We can find the student's z-score for each exam as follows:

Psychology Exam	Psychology Exam	Comparison

Psychology Exam	Psychology Exam	Comparison
<p>Given:</p> $\mu = 70.00$ $\sigma = 10.00$ $x_1 = 80.00$ Find X_1 : $z = \frac{x - \mu}{\sigma}$ $z = \frac{80.00 - 70.00}{10.00}$ $z = \frac{10.00}{10.00}$ $z = 1.00$	<p>Given:</p> $\mu = 50.00$ $\sigma = 5.00$ $x_1 = 65.00$ Find X_1 : $z = \frac{x - \mu}{\sigma}$ $z = \frac{65.00 - 50.00}{10.00}$ $z = \frac{15.00}{5.00}$ $z = 3.00$	<p>The student had a higher relative score on the psychology exam than on the English exam.</p>

Reading Review 5.3

Try to answer the following questions using the graph:

1. What are the symbols in the z-score formula and what does each stand for?
2. What are the steps to computing a z-score from a given data set?
3. What are five uses for z-scores?

This page titled [5.3: z-Scores](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .