Standardization and z-scores

In this *Guide*, we will learn how to calculate z-scores. Let's go ahead and generate some data for us to practice z-scoring with. These data are anxiety levels on a 1 (*"not anxious at all"*) to 14 (*"extremely anxious"*) scale of 15 office employees.

Name	Anxiety
Jim	1
Pam	7
Kelly	12
Ryan	8
Michael	9
Holly	5
Kevin	4
Dwight	9
Angela	7
Toby	4
Oscar	6
Andy	13
Stanley	2
Phyllis	6

z-score formula

$$z_i = \frac{x_i - \mu}{\sigma}$$

Let's say that we want to know how Dwight's anxiety level compares to that of his colleagues. We can answer this question by converting his raw score (9) into a z-score using the formula above. Specifically, we'll need to subtract Dwight's raw score from the mean anxiety level of the entire group (the "population" mean, μ, since this group is our entire population of interest) and then divide by the standard deviation of this group (σ).

Calculating the mean is easy, as we just need to add up all of the anxiety values and divide by N. If you do this, you'll end up with a mean anxiety level of 6.64.

Next, we'll need to calculate the standard deviation of anxiety scores using the population formula for standard deviation that we've seen in this class previously. As a reminder, here it is:

Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{N}}$$

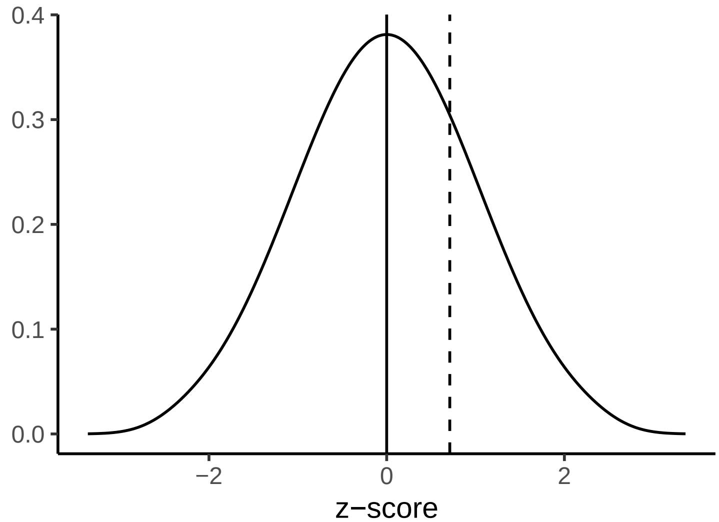
I'll show my work for calculating the standard deviation of this population on the last page, if you're interested. For now, let me just tell you that $\sigma = 153.22$.

Now, this becomes an easy problem to solve. We can simply plug in Dwight's anxiety score (9, which we represent by x), the average level of anxiety in the group (6.64, represented by μ), and the standard deviation of anxiety in the group (3.31, represented by σ). Let's do that now:

$$z_i = \frac{x_i - \mu}{\sigma} = \frac{9.00 - 6.64}{3.31} = 0.71$$

A *z*-score of 0.71 can be interpreted as follows: Dwight is 0.71 standard deviations above the mean of anxiety, meaning he is slightly more anxious than average.

Note that, if you were to calculate *z*-scores for every single person in this group, you'll have an entirely new column of data containing only *z*-scores. We call this *"standardizing"* data, because the new, "standardized" scores will always have a mean of 0 and a standard deviation of 1. This is useful because then each person's score is a *z*-score that shows how they compare to their peers. This makes it easy to quickly tell who is scoring very high compared to their peers (large, positive *z*-scores) and who is scoring very low compared to their peers (negative *z*-scores).



Dashed line is Dwight's z−score for anxiety

Calculating Population Standard Deviation

Let's calculate population standard deviation of this group of office employees. As always, we'll need to start by finding the mean, which we found was 6.64. Then, we calculate the deviations of each score from that mean, and then squaring those deviations from the mean. Let's do that now:

Name	Anxiety	deviations	squares
Jim	1	-5.643	31.842
Pam	7	0.357	0.128
Kelly	12	5.357	28.699
Ryan	8	1.357	1.842
Michael	9	2.357	5.556
Holly	5	-1.643	2.699
Kevin	4	-2.643	6.985
Dwight	9	2.357	5.556
Angela	7	0.357	0.128
Toby	4	-2.643	6.985
Oscar	6	-0.643	0.413
Andy	13	6.357	40.413
Stanley	2	-4.643	21.556
Phyllis	6	-0.643	0.413

Now, we need to calculate our *Sum of Squares* ($\sum(x_i - \overline{x})^2$) by adding up all of the values in the *"squares"* column, which gives us a value of 153.22.

Finally, to calculate population standard deviation, we need to divide the Sum of Squares by N and then square root that value. Let's do that now:

$$\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{N}} = \sqrt{\frac{153.22}{14}} = 3.31$$

So, our population standard deviation (σ) is 3.31.