Z-scores

Hervé Abdi[1]

1 Overview

A score by itself does not tell much. If we are told that we have obtained a score of 85 on a beauty test, this could be very good news if most people have a score of 50, but less so if most people have a score of 100. In other words a score is meaningful only relative to the means of the sample or the population. Another problem occurs when we want to compare scores measured with different units or on different population. How to compare, for example a score of 85 on the beauty test with a score of 100 on an I.Q. test?

Scores from different distributions, such as the ones in our example, can be *standardized* in order to provide a way of comparing them that includes consideration of their respective distributions. This is done by transforming the scores into Z-scores which are expressed as standardized deviations from their means. These Z-scores have a mean of 0 and a standard deviation equal to 1. Z-scores computed from different samples with different units can be directly compared because these numbers do not express the original unit of measurement.

[1]In: Neil Salkind (Ed.) (2007). *Encyclopedia of Measurement and Statistics.* Thousand Oaks (CA): Sage.
Address correspondence to: Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

2 Definition of Z-scores

In order to compute a Z-score, we start with an original score (called Y) obtained from a sample (or a population) with a mean of M_Y and a standard deviation of S_Y. We eliminate the mean by subtracting it from the score, this transforms the original score into a deviation from its mean. We eliminate the original unit of measurement by dividing the score deviation by the standard deviation. Specifically, the formula for calculating a Z-score is

$$Z = \frac{Y - M_Y}{S_Y} \; . \tag{1}$$

We say that subtracting the mean *centers* the distribution, and that dividing by the standard deviation *normalizes* the distribution. The interesting properties of the Z-scores are that they have a zero mean (effect of "centering") and a variance and standard deviation of 1 (effect of "normalizing"). This is because all distributions expressed in Z-scores have the same mean (0) and the same variance (1) that we can use Z-scores to compare observations coming from different distributions

The fact that Z-scores have a zero mean and a unitary variance can be shown by developing the formulas for the sum of Z-scores and for the sum of the squares of Z-scores. This is done in the appendix.

3 An example

Applying the formula for a Z score to a score of $Y = 85$ coming from a sample of mean $M_Y = 75$ and standard deviation $S_Y = 17$ gives

$$Z = \frac{Y - M_Y}{S_Y} = \frac{85 - 75}{17} = \frac{10}{17} = .59 \; . \tag{2}$$

4 Effect of Z-scores

When a distribution of numbers is transformed into Z-scores, the *shape*, of the distribution is unchanged but this shape is translated in order to be centered on the value 0, and it is scaled such that its area is now equal to 1.

As a practical guide, when a distribution is normal most (*i.e.*, more than 99%) of the Z-scores lay between the values −3 and plus +3. Also, because of the central limit theorem, a Z-score with a magnitude larger than 6 is extremely unlikely to occur regardless of the shape of the original distribution.

Appendix: Z-scores have a mean of 0, and a variance of 1

In order to show that the mean of the Z-scores is equal to 0, it suffices to show that the sum of the Z-scores is equal to 0. This is shown by developing the formula for the sum of the Z-scores:

$$\begin{aligned}\sum Z &= \sum \frac{Y - M_Y}{S_Y} \\ &= \frac{1}{S_Y} \sum (Y - M_Y) \\ &= \frac{1}{S_Y} (\sum Y - NM_Y) \\ &= \frac{1}{S_Y} (NM_Y - NM_Y) \\ &= 0 \,. \end{aligned} \tag{3}$$

In order to show that the variance of the Z-scores is equal to 1, it suffices to show that the sum of the squared Z-scores is equal to $(N - 1)$ (where N is the number of scores). This is shown by developing the formula for the sum of the squared Z-scores:

$$\begin{aligned}\sum Z^2 &= \sum \left(\frac{Y - M_Y}{S_Y} \right)^2 \\ &= \frac{1}{S_Y^2} \sum (Y - M_Y)^2 \end{aligned}$$

But $(N-1)S_Y^2 = \sum (Y - M_Y)^2$, hence:

$$\sum Z^2 = \frac{1}{S_Y^2} \times (N-1)S_Y^2$$
$$= (N-1) \ . \tag{4}$$

And this shows that the mean of the Z-scores is equal to 0 and that their variance and standard deviation are equal to 1.