TRANSFORMATIONS

The deviation score

It is often very useful to know not the raw score of a particular person in a distribution (i.e., the actual numerical value achieved by that person on the variable in question), but the relationship of the raw score to some special point of reference. A common approach to this issue is to transform raw scores by expressing them as distances from the mean of the distribution in which they occur. Suppose, for instance, that a particular individual obtained a score of 62 on a maths test. Is this score high or low? You do not have sufficient information to give an answer. If I tell you that the mean score on the test in question was 54, it is now possible to say whether the score was above or below average. We can transform the raw score of 62 by subtracting the mean score from it: 62 - 54 = +8. This person obtained a score 8 points above the mean (the + sign makes it clear that the person's score was above the mean). Another person obtained a raw score of 49; this person was 5 points below the mean and the result of the calculation 49 - 54 was -5 (the -sign shows that the score was below the mean).

This transformation of raw scores yields what are called deviation scores. Deviation scores are written with small letters, as against capital letters for raw scores. Thus, X_i is the raw score of person i, whereas x_i is the deviation score of person i. The formula for the transformation of raw scores to deviation scores is

$$x_i = X_i - X$$

The z-score

The deviation score of an individual tells you whether this person was above or below the mean, and gives you an idea of whether the distance from the mean was large or small. However, in order to be able to say accurately, not only whether the person was above or below the mean, but also whether he or she was a long way above or below it, we need some way of calculating whether a given deviation from the mean is large or small. For instance, on a maths test a student obtained a deviation score of +12, on an English test a score of +10. We see at once that: this person was above the mean on both tests, but the question now arises of whether he or she was further above the mean on maths or English. Simply looking at the value of the deviation score suggests that the maths score of +12 was further above the mean than the English mark of +10. However, if the really good maths students obtained deviation scores of +15, +20 and so on, whereas most of the better English students obtained scores of +2, +6, and so on, the English mark of +10 would be an exceptional performance, whereas +12 in maths would not be particularly high.

In order to answer the question of whether a deviation score is particularly high or low, we thus need to take account not only of the distance of the raw score from the mean, but also of the distribution of scores around the mean; if lots of people got scores well above (or below) the mean, a deviation score just above the mean would not be particularly high, but if most people were close to the mean, a deviation score which was not all that far from the mean could, nonetheless, be exceptional. Thus, we can further transform deviation scores by comparing the size of the deviation of a particular score with the standard deviation. If a particular deviation is large relative to the standard deviation, the score is very high (or low), if the deviation is small relative to the standard deviation, the score is only small. The formula for making this comparison is as follows:

$$z_i = \frac{X_i - X}{SD}$$

Since X_i - X defines the deviation score, as explained in the preceeding section, you could also say that the z-score is obtained by dividing x_i by the standard deviation.

The z-score is usually referred to as a standard score, and is abbreviated as What it is, in effect is the number of standard deviations (or fractions thereof) by which a raw score deviates from the mean. Standard scores always have a mean of 0.00 and a standard deviation of 1.00.

1. Determine the mean and standard deviation of your distribution.
2. Subtract the mean from X_i to find how many raw-score points X_i is from the mean.
3. Divide this difference by SD to determine how many standard deviation units this difference is equal to.

Example 1: For an introductory Psychology class at Bendigo a mean aptitude test score is 48, the standard deviation is 8. What is the z-score equivalent of a score of 43?

To compute a z score:

1. Determine the mean and standard deviation for your distribution.
2. Subtract the mean from Xi to find how many raw score points Xi is from the mean.
3. Divide this difference by SD to determine how many standard deviation units this difference is equal to.

What is the z score good for?

In terms of descriptive statistics, it makes it possible to compare directly scores from distributions with markedly different means and standard deviations. Consider the following data:

X SD

Maths	65	10
English	50	8

A student obtained 64 for English, 72 for maths. In which subject did the student do better?

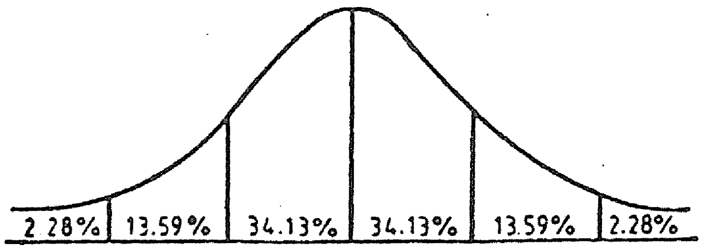
According to the raw scores, in maths. However, the z score for maths was 72-65, which equals + 0.7. The z score for English was 64-50 = +1.175. The student actually did better in English, in the sense of being further above average in English than in Maths.

The use of the normal curve to answer questions regarding relative standing in a distribution

In working with the normal curve, we typically use the mean as our starting point and work up or down from there. In addition, we need a unit of measurement to apply to distances above or below our point of departure. This unit is the standard deviation, which gives us a yardstick for siting points along the baseline of the curve.

Now we are ready to apply these elements - the mean as a starting point and the standard deviation as a unit of measurement above and below that point on the base line of the normal curve. Some very useful facts emerge. We find that at one standard deviation above the mean, the portion of the area under the curve enclosed by this, the mean, and the base line is 34.13 per cent of the total area under the curve. The same portion of the area is cut off by one unit below the mean. Similarly, two units above or below the mean cut off an additional 13.59 per cent of the area under the curve, and three units in either direction cut off another 2.28 per cent. These portions pertain to all curves which fit the criteria of normality, even though the distribution may be rather widely dispersed (as shown in Figure 7.1)

Figure 7.1: Portions of area under the normal curve


2.28%
13.59%
34.13%
34.13%
13.59%
2.28%