Pearson New International Edition

Elementary Statistics in Social Research
Essentials

Jack Levin James Alan Fox

Third Edition

Pearson New International Edition

Elementary Statistics in Social Research
Essentials
Jack Levin James Alan Fox
Third Edition

PEARSON®

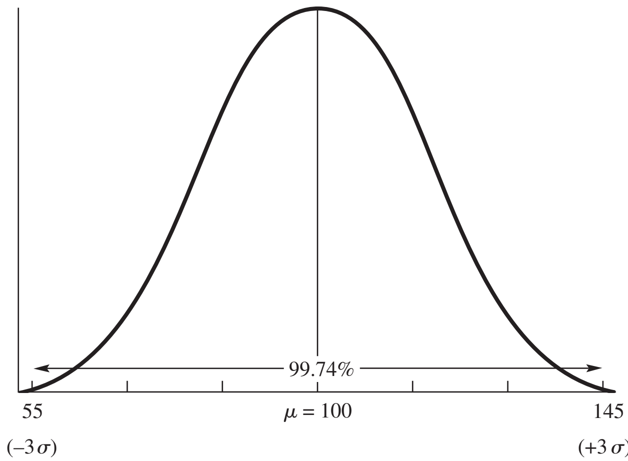

FIGURE 9 A distribution of male IQ scores

Using Table A

In discussing the normal distribution, we have so far treated only those distances from the mean that are exact multiples of the standard deviation. That is, they were precisely one, two, or three standard deviations either above or below the mean. This question now arises: What must we do to determine the percent of cases for distances lying between any two score values? For instance, suppose we wish to determine the percent of total area that falls between the mean and, say, a raw score located 1.40σ above the mean. As illustrated in Figure 11, a raw score 1.40σ above the mean is obviously greater than 1σ but less

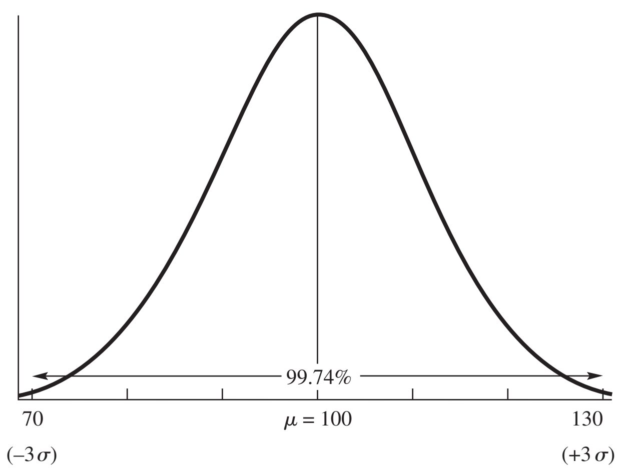

FIGURE 10 A distribution of female IQ scores

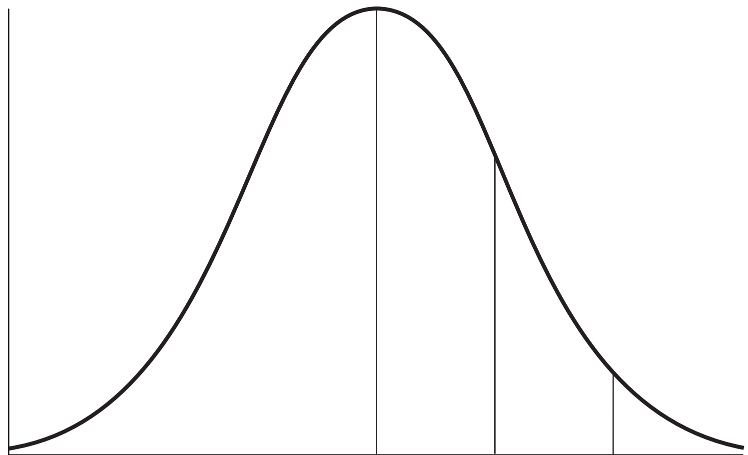

FIGURE 11 The position of a raw score that lies 1.40σ above μ

than 2σ from the mean. Thus, we know this distance from the mean would include more than 34.13% but less than 47.72% of the total area under the normal curve.

To determine the exact percentage within this interval, we must employ the table below. This shows the percent under the normal curve (1) between the mean and various sigma distances from the mean (in column b) and (2) at or beyond various scores toward either tail of the distribution (in column c). These sigma distances (from 0.00 to 4.00) are labeled *z* in the left-hand column (column a) and have been given to two decimal places.

Notice that the symmetry of the normal curve makes it possible to give percentages for only one side of the mean, that is, only one-half of the curve (50%). Values in the table below represent either side.

(a) *z*	**(b)** **Area between Mean and** *z*	**(c)** **Area beyond** *z*
.00	.00	50.00
.01	.40	49.60
.02	.80	49.20
.03	1.20	48.80
.04	1.60	48.40
.05	1.99	48.01
.06	2.39	47.61
.07	2.79	47.21
.08	3.19	46.81
.09	3.59	46.41

When learning to use and understand this table, we might first attempt to locate the percent of cases between a sigma distance of 1.00 and the mean (the reason being that we already know that 34.13% of the total area falls between these points on the base line). Looking at column b, we see it indeed indicates that exactly 34.13% of the total frequency falls between the mean and a sigma distance of 1.00. Likewise, we see that the area between the mean and the sigma distance 2.00 includes exactly 47.72% of the total area under the curve.

But what about finding the percent of cases between the mean and a sigma distance of 1.40? This was the problem in Figure 11, which necessitated the use of the table in the first place. The entry in column b corresponding to a sigma distance of 1.40 includes 41.92% of the total area under the curve. Finally, how do we determine the percent of cases at or beyond 1.40 standard deviations from the mean? Without a table to help us, we might locate the percentage in this area under the normal curve by simply subtracting our earlier answer from 50%, because this is the total area lying on either side of the mean. However, this has already been done in column c, where we see that exactly 8.08% $(50 - 41.92 = 8.08)$ of the cases fall at or above the score that is 1.40 standard deviations from the mean.

Standard Scores and the Normal Curve

We are now prepared to find the percent of the total area under the normal curve associated with any given sigma distance from the mean. However, at least one more important question remains to be answered: How do we determine the sigma distance of any given raw score? That is, how do we translate our raw score—the score that we originally collected from our respondents—into units of standard deviation? If we wished to translate feet into yards, we would simply divide the number of feet by 3, because there are 3 ft in a yard. Likewise, if we were translating minutes into hours, we would divide the number of minutes by 60, because there are 60 min in every hour. In precisely the same manner, we can translate any given raw score into sigma units by dividing the distance of the raw score from the mean by the standard deviation. To illustrate, let us imagine a raw score of 16 from a distribution in which μ is 13 and σ is 2. Taking the difference between the raw score and the mean and obtaining a deviation $(16 - 13)$, we see that a raw score of 16 is 3 raw-score units above the mean. Dividing this raw-score distance by $\sigma = 2$, we see that this raw score is 1.5 (one and one-half) standard deviations above the mean. In other words, the sigma distance of a raw score of 16 *in this particular distribution* is 1.5 standard deviations above the mean. We should note that regardless of the measurement situation, there are always 3 ft in a yard and 60 min in an hour. The constancy that marks these other standard measures is not shared by the standard deviation. It changes from one distribution to another. For this reason, we must know the standard deviation of a distribution by calculating it, estimating it, or being given it by someone else before we are able to translate any particular raw score into units of standard deviation.

The process that we have just illustrated—that of finding sigma distance from the mean, yields a value called a *z score* or *standard score*, which indicates the *direction and degree that any given raw score deviates from the mean of a distribution on a scale of sigma units.* Thus, a z score of $+1.4$ indicates that the raw score lies 1.4σ (or almost $1\frac{1}{2}\sigma$) *above* the mean,

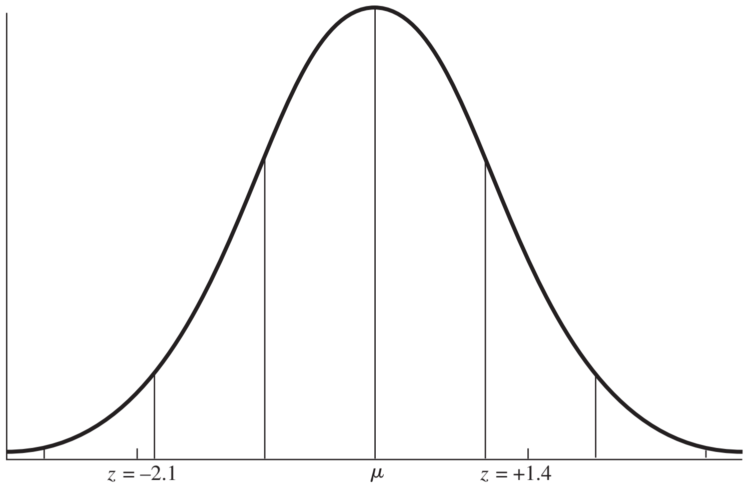

FIGURE 12 The position of $z = -2.1$ and $z = +1.4$ in a normal distribution

whereas a z score of -2.1 means that the raw score falls slightly more than 2σ *below* the mean (see Figure 12).

We obtain a z score by finding the deviation $(X - \mu)$, which gives the distance of the raw score from the mean, and then dividing this raw-score deviation by the standard deviation.

Computed by formula,

$$z = \frac{X - \mu}{\sigma}$$

where μ = mean of a distribution
σ = standard deviation of a distribution
z = standard score

As an example, suppose we are studying the distribution of annual income for home healthcare workers in a large agency in which the mean annual income is \$20,000 and the standard deviation is \$1,500. Assuming that the distribution of annual income is normally distributed we can translate the raw score from this distribution, \$22,000, into a standard score in the following manner:

$$z = \frac{22{,}000 - 20{,}000}{1{,}500} = +1.33$$

Thus, an annual income of \$22,000 is 1.33 standard deviations above the mean annual income of \$20,000 (see Figure 13).

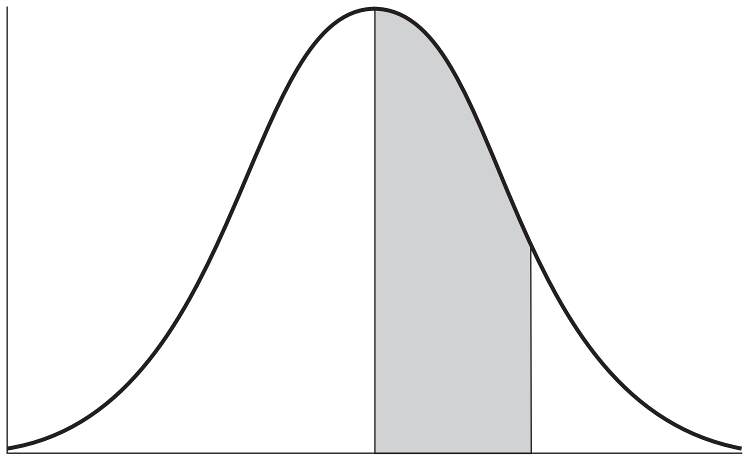

FIGURE 13 The position of $z = +1.33$ for the raw score $22,000

As another example, suppose that we are working with a normal distribution of scores representing job satisfaction among a group of city workers. The scale ranges from 0 to 20, with higher scores representing greater satisfaction with the job.

Let us say this distribution has a mean of 10 and a standard deviation of 3. To determine how many standard deviations a score of 3 lies from the mean of 10, we obtain the difference between this score and the mean, that is,

$$\begin{aligned} X - \mu &= 3 - 10 \\ &= -7 \end{aligned}$$

We then divide by the standard deviation:

$$\begin{aligned} z &= \frac{X - \mu}{\sigma} \\ &= \frac{-7}{3} \\ &= -2.33 \end{aligned}$$

Thus, as shown in Figure 14, a raw score of 3 in this distribution of scores falls 2.33 standard deviations below the mean.

Finding Probability under the Normal Curve

As we shall now see, the normal curve can be used in conjunction with *z* scores and the table shown earlier to determine the probability of obtaining any raw score in a distribution. In the

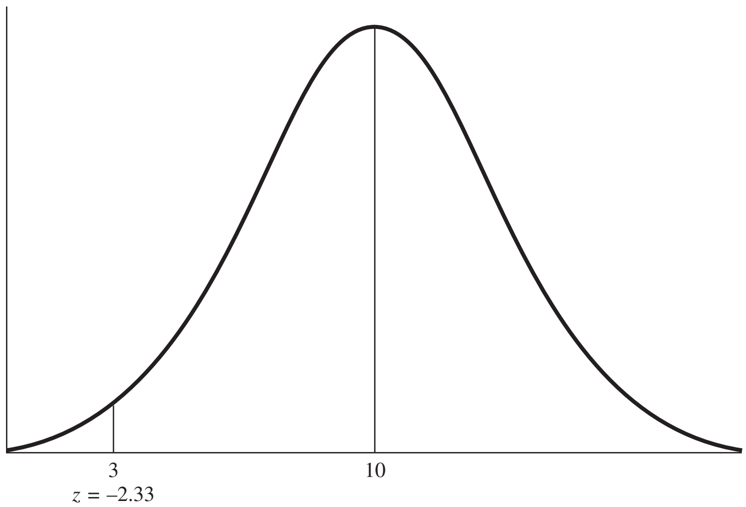

FIGURE 14 The position of $z = -2.33$ for the raw score 3

present context, the normal curve is a distribution in which it is possible to determine probabilities associated with various points along its baseline. As noted earlier, the normal curve is a *probability distribution* in which the total area under the curve equals 100%; it contains a central area surrounding the mean, where scores occur most frequently, and smaller areas toward either end, where there is a gradual flattening out and thus a smaller proportion of extremely high and low scores. In probability terms, then, we can say that probability decreases as we travel along the baseline away from the mean in either direction. Thus, to say that 68.26% of the total frequency under the normal curve falls between -1σ and $+1\sigma$ from the mean is to say that the probability is approximately 68 in 100 that any given raw score will fall within this interval. Similarly, to say that 95.44% of the total frequency under the normal curve falls between -2σ and $+2\sigma$ from the mean is also to say that the probability is approximately 95 in 100 that any raw score will fall within this interval, and so on.

This is precisely the same concept of probability or *relative frequency* that we saw in operation when flipping pairs of coins. Note, however, that the probabilities associated with areas under the normal curve are always given relative to 100%, which is the entire area under the curve (for example, 68 in 100, 95 in 100, 99 in 100).

BOX 1 • *Step-by-Step Illustration: Probability under the Normal Curve*

To apply the concept of probability in relation to the normal distribution, let us return to an earlier example. We were then asked to translate into its z-score equivalent a raw score from an agency's distribution of annual salary for healthcare workers, which we assumed approximated a normal curve. This distribution of income had a mean of \$20,000 with a standard deviation of \$1,500.