PEARSON NEW INTERNATIONAL EDITION

Statistics for the Behavioral and Social Sciences: A Brief Course

Aron Coups Aron

Fifth Edition

ALWAYS LEARNING

PEARSON

Pearson New International Edition

Statistics for the Behavioral
and Social Sciences: A Brief Course
Aron Coups Aron
Fifth Edition

PEARSON

Examples

Here are three examples. Once again, we will use IQ for our examples, where $M = 100$ and $SD = 16$.

Example 1: What IQ score would a person need to be in the top 5%?

❶ **Draw a picture of the normal curve and shade in the approximate area for your percentage using the 50%–34%–14% percentages.** We wanted the top 5%. Thus, the shading has to begin above (to the right of) 1 *SD* (there are 16% of scores above 1 *SD*). However, it cannot start above 2 *SD* because only 2% of all the scores are above 2 *SD*. But 5% is a lot closer to 2% than to 16%. Thus, you would start shading a small way to the left of the 2 *SD* point. This is shown in Figure 7.

❷ **Make a rough estimate of the *Z* score where the shaded area stops.** The *Z* score has to be between +1 and +2.

❸ **Find the exact *Z* score using the normal curve table (subtracting 50% from your percentage if necessary before looking up the *Z* score).** We want the top 5%, which means we can use the "% in Tail" column of the normal curve table. Looking in that column, the closest percentage to 5% is 5.05% (or you could use 4.95%). This goes with a *Z* score of 1.64 in the "*Z*" column.

❹ **Check that your exact *Z* score is within the range of your rough estimate from Step ❷.** As we estimated, +1.64 is between +1 and +2 (and closer to 2).

❺ **If you want to find a raw score, change it from the *Z* score.** Using the formula, $X = (Z)(SD) + M = (1.64)(16) + 100 = 126.24$. In sum, to be in the top 5%, a person would need an IQ of at least 126.24.

Example 2: What IQ score would a person need to be in the top 55%?

❶ **Draw a picture of the normal curve and shade in the approximate area for your percentage using the 50%–34%–14% percentages.** You want the top 55%. There are 50% of scores above the mean. So, the shading has to begin below (to the left of) the mean. There are 34% of scores between the mean and 1 *SD* below the mean, so the score is between the mean and 1 *SD* below the mean. You would shade the area to the right of that point. This is shown in Figure 8.

❷ **Make a rough estimate of the *Z* score where the shaded area stops.** The *Z* score has to be between 0 and −1.

❸ **Find the exact *Z* score using the normal curve table (subtracting 50% from your percentage if necessary before looking up the *Z* score).** Being in the top 55% means that 5% of people have IQs between this IQ and the mean (that is, 55% − 50% = 5%). In the normal curve table, the closest percentage to 5% in the "% Mean to *Z*" column is 5.17%, which goes with a *Z* score of .13. Because you are below the mean, this becomes −.13.

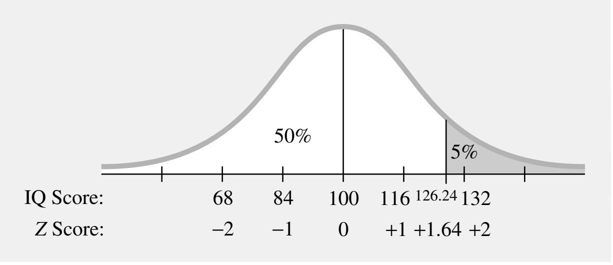

Figure 7 Finding the *Z* score and IQ raw score for where the top 5% starts.

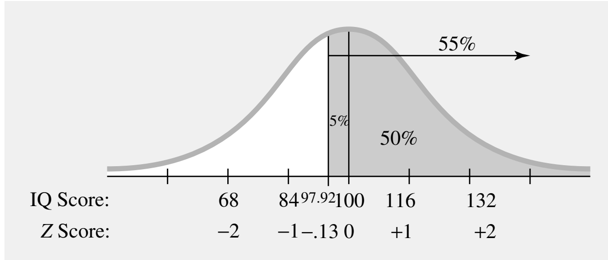

Figure 8 Finding the IQ score for where the top 55% start.

❹ **Check that your exact Z score is within the range of your rough estimate from Step ❷.** As we estimated, $-.13$ is between 0 and -1.

❺ **If you want to find a raw score, change it from the Z score.** Using the usual formula, $X = (-.13)(16) + 100 = 97.92$. So, to be in the top 55% on IQ, a person needs an IQ score of 97.92 or higher.

Example 3: What range of IQ scores includes the 95% of people in the middle range of IQ scores?

This kind of problem, of finding the middle percentage, may seem odd at first. However, it is actually a very common situation used in procedures. Think of this kind of problem in terms of finding the scores that go with the upper and lower ends of this percentage. Thus, in this example, you are trying to find the points where the bottom 2.5% ends and the top 2.5% begins (which, out of 100%, leaves the middle 95%).

❶ **Draw a picture of the normal curve, and shade in the approximate area for your percentage using the 50%–34%–14% percentages.** Let's start where the top 2.5% begins. This point has to be higher than 1 *SD* (16% of scores are higher than 1 *SD*). However, it cannot start above 2 *SD* because there are only 2% of scores above 2 *SD*. But 2.5% is very close to 2%. Thus, the top 2.5% starts just to the left of the 2 *SD* point. Similarly, the point where the bottom 2.5% comes in is just to the right of -2 *SD*. The result of all this is that we will shade in the area starting just above -2 *SD* and continue shading up to just below $+2$ *SD*. This is shown in Figure 9.

❷ **Make a rough estimate of the Z score where the shaded area stops.** You can see from the picture that the Z score for where the shaded area stops above the mean is just below $+2$. Similarly, the Z score for where the shaded area stops below the mean is just above -2.

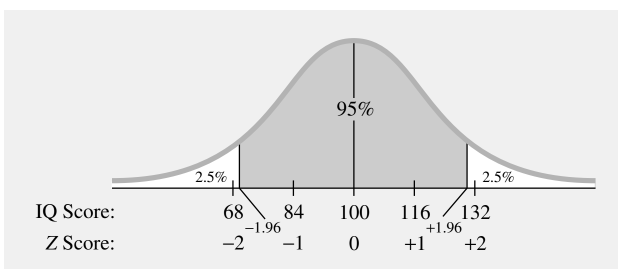

Figure 9 Finding the IQ scores for where the middle 95% of scores begin and end.

❸ **Find the exact Z score using the normal curve table (subtracting 50% from your percentage if necessary before looking up the Z score).** Being in the top 2.5% means that 2.5% of the IQ scores are in the upper tail. In the normal curve table, the closest percentage to 2.5% in the "% in Tail" column is exactly 2.50%, which goes with a Z score of +1.96. The normal curve is symmetrical. Thus, the Z score for the lower tail is −1.96.

❹ **Check that your exact Z score is within the range of your rough estimate from Step ❷.** As we estimated, +1.96 is between +1 and +2 and is very close to +2, and −1.96 is between −1 and −2 and very close to −2.

❺ **If you want to find a raw score, change it from the Z score.** For the high end, using the usual formula, $X = (1.96)(16) + 100 = 131.36$. For the low end, $X = (-1.96)(16) + 100 = 68.64$. In sum, the middle 95% of IQ scores run from 68.64 to 131.36.

How are you doing?

1. Why is the normal curve (or at least a curve that is symmetrical and unimodal) so common in nature?
2. Without using a normal curve table, about what percentage of scores on a normal curve are (a) above the mean, (b) between the mean and 1 *SD* above the mean, (c) between 1 and 2 *SD* above the mean, (d) below the mean, (e) between the mean and 1 *SD* below the mean, and (f) between 1 and 2 *SD* below the mean?
3. Without using a normal curve table, about what percentage of scores on a normal curve are (a) between the mean and 2 *SD* above the mean, (b) below 1 *SD* above the mean, (c) above 2 *SD* below the mean?
4. Without using a normal curve table, about what Z score would a person have who is at the start of the top (a) 50%, (b) 16%, (c) 84%, (d) 2%?
5. Using the normal curve table, what percentage of scores are (a) between the mean and a Z score of 2.14, (b) above 2.14, (c) below 2.14?
6. Using the normal curve table, what Z score would you have if (a) 20% are above you, (b) 80% are below you?

Answers

1. It is common because any particular score is the result of the random combination of many effects, some of which make the score larger and some of which make the score smaller. Thus, on average these effects balance out to produce scores near the middle, with relatively few scores at each extreme, because it is unlikely for most of the increasing or decreasing effects to come out in the same direction.
2. (a) Above the mean: 50%, (b) between the mean and 1 *SD* above the mean: 34%, (c) between 1 and 2 *SD* above the mean: 14%, (d) below the mean: 50%, (e) between the mean and 1 *SD* below the mean: 34%, (f) between 1 and 2 *SD* below the mean: 14%.
3. (a) Between the mean and 2 *SD* above the mean: 48%, (b) below 1 *SD* above the mean: 84%, (c) above 2 *SD* below the mean: 98%.
4. Above start of top (a) 50%: 0, (b) 16%: 1, (c) 84%: −1, (d) 2%: 2.
5. (a) Between the mean and a Z score of 2.14: 48.38%, (b) above a Z score of 2.14: 1.62%, (c) below a Z score of 2.14: 98.38%.
6. (a) 20% above you: .84, (b) 80% below you: .84.

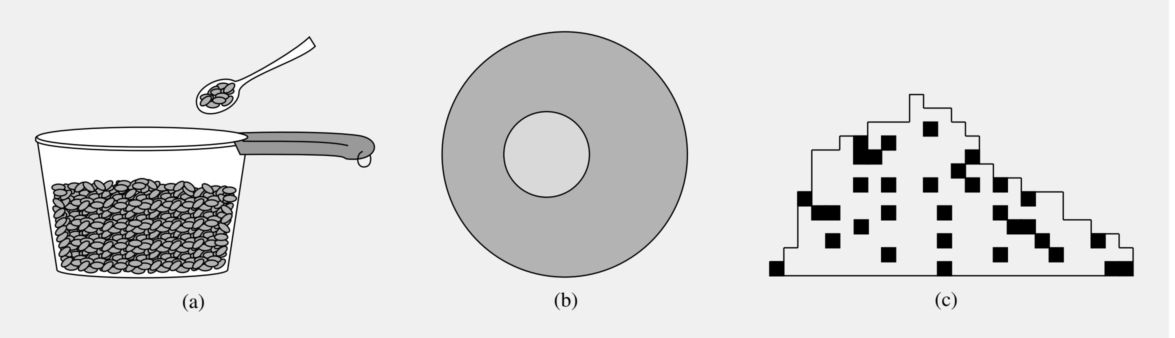

Figure 10 Populations and samples: In (a), the entire pot of beans is the population and the spoonful is a sample. In (b), the entire larger circle is the population and the circle within it is the sample. In (c), the histogram is of the population and the particular shaded scores together make up the sample.

Sample and Population

We are going to introduce you to some important ideas by thinking of beans. Suppose you are cooking a pot of beans and taste a spoonful to see if they are done. In this example, the pot of beans is a **population,** the entire set of things of interest. The spoonful is a **sample,** the part of the population about which you actually have information. This is shown in Figure 10a. Figures 10b and 10c are other ways of showing the relation of a sample to a population.

In behavioral and social science research, we typically study samples not of beans but of individuals to make inferences about some larger group. A sample might consist of 50 Canadian women who participate in a particular experiment; but the population might be intended to be all Canadian women. In an opinion survey, 1,000 people might be selected from the voting-age population and asked for whom they plan to vote. The opinions of these 1,000 people are the sample. The opinions of the larger voting public, to which the pollsters hope to apply their results, is the population (see Figure 11).[2]

Why Samples Instead of Populations Are Studied

If you want to know something about a population, your results would be most accurate if you could study the *entire population* rather than a *subgroup* from that population. However, in most research situations this is not practical. More important, the whole point of research is usually to be able to make generalizations or predictions about events beyond our reach. We would not call it research if we tested three particular cars to see which gets better gas mileage—unless you hoped to say something about the gas mileage of those models of cars in general. In other words, a researcher might do an experiment on the effect of a particular new method of teaching geography using 40 eighth-grade students as participants in the experiment. But the purpose of the experiment is not to find out how these *particular 40 students* respond to the experimental condition. Rather, the purpose is to learn something about how *eighth-grade students in general* respond to the new method of teaching geography.

population Entire group of people to which a researcher intends the results of a study to apply; the larger group to which inferences are made on the basis of the particular set of people (sample) studied.

sample Scores of the particular group of people studied; usually considered to be representative of the scores in some larger population.

[2]Strictly speaking, *population* and *sample* refer to scores (numbers or measurements), not to the people who have those scores. In the first example, the sample is really the scores of the 50 Canadian women, not the 50 women themselves, and the population is really what the *scores* would be if all Canadian women were tested.

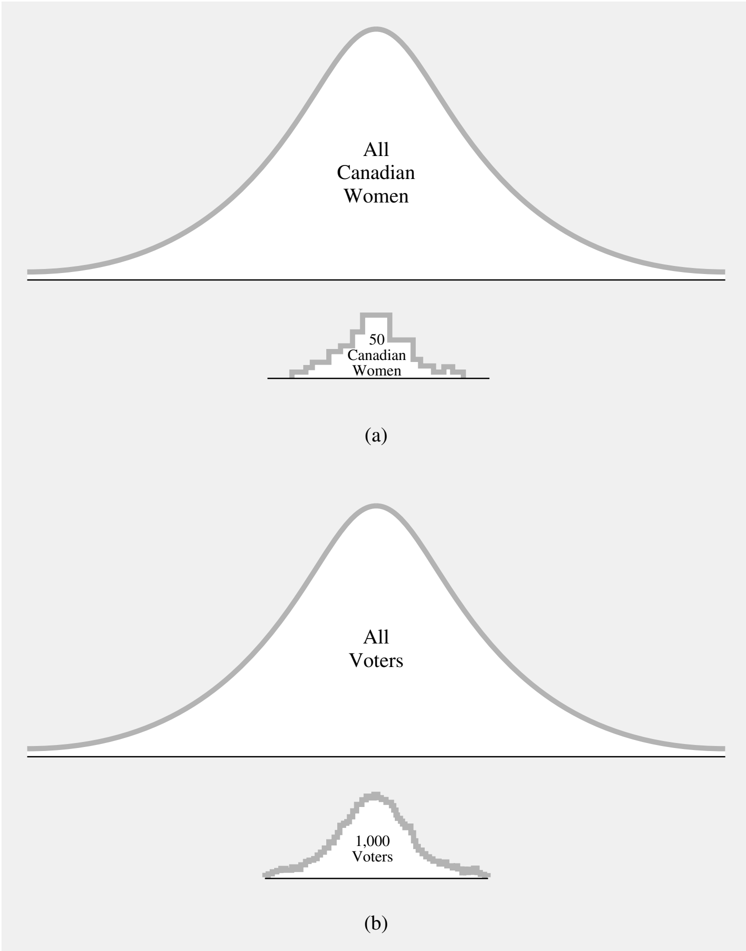

Figure 11 Additional examples of populations and samples. In (a), the population is the scores of all Canadian women and the sample is the scores of the 50 particular Canadian women studied. In (b), the population is the voting preferences of the entire voting-age population and the sample is the voting preferences of the 1,000 voting-age people who were surveyed.

The strategy in almost all behavioral and social science research is to study a sample of individuals who are believed to be representative of the general population (or of some particular population of interest). More realistically, researchers try to study people who do not differ from the general population in any systematic way that should matter for that topic of research.

The sample is what is studied, and the population is an unknown about which researchers draw conclusions based on the sample. Most of what you learn in the rest of this text is about the important work of drawing conclusions about populations based on information from samples.

Methods of Sampling

Usually, the ideal method of picking out a sample to study is called **random selection.** The researcher starts with a complete list of the population and randomly selects some of them to study. An example of a random method of selection would be to put each

random selection Method for selecting a sample that uses truly random procedures (usually meaning that each person in the population has an equal chance of being selected); one procedure is for the researcher to begin with a complete list of all the people in the population and select a group of them to study using a table of random numbers.

name on a table tennis ball, put all the balls into a big hopper, shake it up, and have a blindfolded person select as many as are needed. (In practice, most researchers use a computer-generated list of random numbers.)

It is important not to confuse truly random selection with what might be called **haphazard selection;** for example, just taking whoever is available or happens to be first on a list. When using haphazard selection, it is surprisingly easy to pick accidentally a group of people that is very different from the population as a whole. Consider a survey of attitudes about your statistics instructor. Suppose you give your questionnaire only to other students sitting near you in class. Such a survey would be affected by all the things that influence where students choose to sit, some of which have to do with just what your survey is about—how much they like the instructor or the class. Thus, asking students who sit near you would likely result in opinions more like your own than a truly random sample would.

Unfortunately, it is often impractical or impossible to study a truly random sample. One problem is that we do not usually have a complete list of the full population. Another problem is that not everyone a researcher approaches agrees to participate in the study. Yet another problem is the cost and difficulty of tracking down people who live in remote places. For these and other reasons as well, behavioral and social scientists use various approximations to truly random samples that are more practical. However, researchers are consistently careful to rule out, as much as possible in advance, any systematic influence on who gets selected. Once the study has begun, researchers are constantly on the alert for any ways in which the sample may be systematically different from the population. For example, in much experimental research in education and psychology, it is common to use volunteer college students as the participants in the study. This is done for practical reasons and because often the topics studied in these experiments (for example, how short-term memory works) are thought to be relatively consistent across different types of people. Yet even in these cases, researchers avoid, for example, selecting people with a special interest in their research topic. Such researchers are also very aware that their results may not apply beyond college students, volunteers, people from their region, and so forth.

Methods of sampling is a complex topic that is discussed in detail in research methods textbooks (also see Box 1).

Statistical Terminology for Samples and Populations

The mean, variance, and standard deviation of a population are called **population parameters.** A population parameter usually is unknown and can be estimated only from what you know about a sample taken from that population. You do not taste all the beans, just the spoonful. “The beans are done” is an inference about the whole pot.

In this text, when referring to the population mean, standard deviation, or variance, even in formulas, we use the word “Population”[3] before the *M*, SD^2, or *SD*. The mean, variance, and standard deviation you figure for the scores in a sample are called **sample statistics.** A sample statistic is figured from known information. Sample statistics are what we have been calculating all along. Sample statistics use the symbols we have been using all along: *M*, SD^2, and *SD*.

haphazard selection Procedure of selecting a sample of individuals to study by taking whoever is available or happens to be first on a list.

population parameter Actual value of the mean, standard deviation, and so on, for the population (usually population parameters are not known, though often they are estimated based on information in samples).

sample statistic Descriptive statistic, such as the mean or standard deviation, figured from the scores in a particular group of people studied.

[3]In statistics writing, it is common to use Greek letters for population parameters. For example, the population mean is μ (“mu”) and the population standard deviation is σ (lowercase “sigma”). However, we have not used these symbols in this text, wanting to make it easier for students to grasp the formulas without also having to deal with Greek letters.