5 SAMPLING DISTRIBUTIONS

5.07 Example

As you might have noticed, Hipsters often wear oversized glasses. Yet many Hipsters don't really need glasses. Their glasses only serve as a celebration of their Hipsterness. Suppose you study the strength of the glasses of Hipsters in, say, Italy. You know that the variable "strength of glasses" has a population distribution that is skewed to the left. The peak will be around zero, because most glasses are 'fake' glasses. But, of course, not all glasses are fake. Because Hipsters are generally relatively young, and young people are more often short-sighted than long-sighted, there will be more Hipsters who are short-sighted than Hipsters who are long-sighted. Those who are short-sighted have a negative score on "strength of glasses" and those who are long-sighted have a positive score. You know that the mean in the population is -0.75 and that the standard deviation is 2.89.

We would like to know three things. First, what does the population distribution look like? We would like to see shape, mean and standard deviation. Second, what does the sampling distribution of the sample mean look like based on a sample size of n equals 3000? Third, what does the sample or data distribution look like if you draw a simple random sample of 3000 cases from this population? And fourth, what is the probability of selecting such a sample from this population with a sample mean between -0.71 and -0.81?

Let's start with the first question. The distribution would look something like *this*. The peak is around 0 and the distribution is skewed to the left. The population mean is left of the population mode. We know that the score of this parameter, symbolized by Mu, is -0.75. The population standard deviation, symbolized by Sigma, is 2.89.

The second question is what the sampling distribution of the sample mean looks like. We know that when the sample size is sufficiently large (which is, with an n of 3000, clearly the case in this example), the sampling distribution is bell-shaped with a mean that equals the population mean. We can therefore conclude that Mu-X-bar equals Mu equals -0.75. The standard deviation of the sampling distribution, symbolized by Sigma-X-bar, can easily be computed. It is the standard deviation in the population divided by the square root of n. That is 2.89 / SQRT(3000) = 0.05. The cases in the sampling distribution of the sample mean are not individual cases, as in the population distribution, but an infinite number of sample means. This is symbolized by the X-bar next to the Mu and the Sigma.

The third question is what the sample or data distribution looks like. *This* is what the distribution of scores looks like in our sample of 3000 Italian hipsters. Because we are dealing with a simple random sample with a fairly large sample size, we can be pretty confident that the sample resembles the population. The shape of its distribution will be very similar to the shape of the population distribution, and the sample mean, symbolized by X-bar, will be close to the population mean of -0.75. The sample standard deviation, symbolized by s, will be close to the population standard deviation of 2.89. However, also note that it is very unlikely that the sample statistics are *exactly* the same as the population parameters. You will probably have a sample mean of, say, -0.70 or -0.78. Similarly, your sample standard deviation will probably have a value slightly different from, but very close to 2.89. It could, for instance, be 2.77, or 3.01.

In fact, based on the sampling distribution of the sample mean, we can compute what the probability is of finding particular sample means. The fourth question we want to answer is what the probability is of selecting a sample from this population with a sample mean between -0.71 and -

0.81? We are dealing with a sample mean here, so we should look at the sampling distribution. This is what the sampling distribution looks like. We're interested in the probability of finding a value *between* these two values. So we're interested in *this* surface. To find this probability we first have to convert the original scores of -0.71 and -0.81 into z-scores. *This* is the relevant formula. Let's first look at the score of -0.71. We subtract -0.75 (which is the mean of the sampling distribution) from -0.71 (which is the sample mean score we're interested in) and divide the outcome by the standard deviation of the sampling distribution. We have already computed this value. It's the population mean divided by the square root of the sample size. That equals 0.05. The outcome is 0.8. This means that the original score of -0.71 corresponds to a z-score of 0.8. We also do that with the score of -0.81. So we subtract -0.75 from -0.81 and divide it by 0.05. That makes -1.2. So the original score of -0.81 corresponds to a z-score of -1.2. We thus have to find the probability between a z-value of -1.2 and a z-value of 0.8.

If we look at the z-table, we see that the probability of finding a value lower than a z-score of 0.8 is 0.7881. That corresponds to *this* surface. The probability of finding a value lower than a z-score of -1.2 is 0.1151. That corresponds to *this* surface. We're interested in *this* surface, so we have to subtract *this* part from *this* part. That means 0.7881 – 0.1151. That is about 0.67. Or 67 percent. The probability of selecting a sample of n equals 3000 with a sample mean between -0.71 and -0.81 is 67 percent.