**Sampling Distributions**

*How do we get from a sample statistic to an estimate of the population parameter? A crucial midway concept is the* ***sampling distribution****. Imagine that instead of just taking a single sample like we do in a typical study, we took three independent samples of the same population. And furthermore, imagine that for each of our three samples, we collected a single response and computed a single statistic—the mean (or proportion) of the response. Even though all three samples came from the same population, we wouldn't expect to get the exact same statistic from each. They would differ slightly just due to the random "luck of the draw" or to the natural fluctuations of drawing a sample. We would expect that all three samples would yield a similar statistic because they were drawn from the same population.*

*Now, for the big leap of imagination! Imagine that we did an infinite number of samples from the same population and computed the mean for each one. If we plotted them on a histogram or bar graph we should find that most of them converge on the same central value and that we get fewer and fewer samples that have means farther away up or down from that central value. In other words, the bar graph would be well described by the bell curve shape that is an indication of a "normal" distribution in statistics.* ***The distribution of an infinite number of samples of the same size*** *as the sample in our study* ***is known as the sampling distribution.***

*We don't ever actually construct a sampling distribution. Why not? Because to construct it we would have to take an infinite number of samples.*

*So why do we even talk about a sampling distribution? Because we need to realize that our sample is just one of a potentially infinite number of samples that we could have taken.*

*When we keep the sampling distribution in mind, we realize that while the statistic (mean) we got from our sample is probably near the center of the sampling distribution (because most of the samples would be there), we could have gotten one of the extreme samples just by the luck of the draw. If we take the average of the sampling distribution—the average of the averages of an infinite number of samples—we would be much closer to the true population average—the parameter of interest.* ***The average of the sampling distribution is essentially equivalent to the parameter.***



*But what is the standard deviation of the sampling distribution? The standard deviation of the sampling distribution tells us something about how different samples would be distributed. It is referred to as the standard error so we can keep it separate in our minds from a standard deviation. A* ***standard deviation*** *is the spread of the scores around the average in a single sample. The* ***standard error*** *is the spread of the sample means around the average of the sample means in a sampling distribution***. *We can find the standard error of the sampling distribution by taking the sample standard deviation divided by the square root of the sample size.***

**Normal Distributions**

The normal distribution refers to a particular way in which data is spread out or distributed. Data can be spread out in many ways. For example:

**Data can be spread out to the left**

60

40

20

 0

**Data can be spread out to the right**

30

20

10

0

But in many situations, data tends to be spread around a central value with no bias left or right, and it gets close to a "Normal Distribution" like this:

**Data can be all jumbled up**

30

20

10

 0

So, a normal distribution is a distribution or spread of data where the data is evenly spread in a bell-shaped curve around the mean. In a normal distribution:

* Fifty percent of data is above the mean, and fifty percent below it.
* The distribution of data is symmetric about the mean.
* The mean, median and mode are all equal.

Many distributions are not exactly normal, but approximate a normal distribution, as shown in the graph above, where the histogram is not exactly symmetrical and bell- shaped, but is roughly symmetrical and bell-shaped. Examples of things that approximate a normal distribution are:

* Heights of people
* Size of things produced by machines
* Blood pressure
* Marks on a test
* Errors in measurement

The normal distribution is a continuous distribution, meaning that it describes variables that can be graphed in a histogram. A continuous variable is a variable that can take on any value in an interval. For example, the measurement of a group of people’s heights is continuous because it can be any part of a whole unit: 165.97cm, for example. On the other hand, counting the number of heads/tails in a collection of coin tosses is not continuous (it is discrete) because the result can only be an integer number; it is not possible to have 3.5 heads.

The normal distribution also has other interesting properties, which are discussed in more detail later in this worksheet.