Chapter 6: Standard Error

I. Overview

a. Whenever a sample is selected from a population, there may be a difference between the characteristics of the sample relative to the population.
   i. This difference is caused by the process of sampling: When a random sample is selected, one never knows exactly what the sample will look like.
      1. This is known as *random sampling error,* often just called *error.*

b. To help researchers gain a sense of how well the sample represents the population, a standard error is calculated.
   i. The standard error is the average amount of difference between the sample statistic and the population parameter when the samples are randomly selected and of a given size.

c. The sample size and the sample standard deviation are the two key elements of the standard error of the mean.

d. Any statistic calculated from sample data can have a standard error
   i. E.g., the mean, the difference between two sample means, the correlation coefficient, the regression coefficient, etc.

e. The standard error is usually the denominator of the formulas for calculating inferential statistics (e.g., *t* values, *F* values, Tukey values, etc.

II. The Standard Error of the Mean

a. Definition: The standard deviation of the sampling distribution of the mean.
   i. When random samples of the same size are repeatedly selected from a population, the means of these samples will form a distribution.
   ii. This distribution will have a mean, called the expected value.
      1. The expected value is the same as the population mean.
   iii. This distribution will also have a standard deviation, called a standard error.
   iv. The standard error is the average difference between the population mean (i.e., the expected value) and the sample means when the samples are randomly selected and of a given size.

b. How to calculate it: It is the standard deviation divided by the square root of the sample size.

When the population standard deviation is known: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

When the population standard deviation is not known: $s_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

c. Effects of sample size: Notice that, assuming the standard deviation stays the same, the larger the sample size, the smaller the standard error.

d. What it is used for: The standard error of the mean is used as the denominator in *z* score and *t* value formulas when the researcher is calculating the probability of obtaining a difference of a given size between a sample mean and a population mean.

III. Central Limit Theorem

a. Definition: This theorem states that random samples selected from a population will produce normally distributed sampling distributions of the mean, as long as the samples are not too small (i.e., $n > 30$).

b. Importance: This means that even for relatively small samples, we can calculate probabilities of obtaining sample means by chance associated with the normal distribution.

IV. Calculating *z* and *t* values

a. Standard errors can be used to calculate *z* and *t* values

b. If the population standard deviation is known, it can be used in the numerator of the standard error of the mean formula, and Appendix A can be used to calculate the probability of obtaining sample means by chance.

$$z = \frac{\overline{X} - \mu}{\sigma_{\bar{x}}}$$

c. If the population standard deviation is not known, the sample standard deviation must be used in the standard error formula, and the family of *t* distributions must be used to calculate probabilities of obtaining sample means of a given size by chance (Appendix B).

$$t = \frac{\overline{X} - \mu}{s_{\bar{x}}}$$

i. *t* distributions are like the normal distribution, but they take sample size into account. Smaller samples produce flatter sampling distributions of the mean with more extreme means, whereas larger samples produce more normal-looking distributions, with higher peaks in the middle and fewer extreme sample means in the tails.

V. Summary

a. The standard error is a critical statistic for inferential statistics.

b. There is a standard error for virtually every sample statistic.

c. The standard error tells us how different we can expect our sample statistics to differ from their respective population parameters.

d. Without standard errors, we could not determine whether the results generated with sample data are statistically significant.
    i. i.e., whether they represent a genuine phenomenon in the population or are just a fluke, an artifact of random sampling error.