Lesson Learning Objectives

LO 1. Define a confidence interval as the plausible range of values for a population parameter.

LO 2. Define the confidence level as the percentage of random samples which yield confidence intervals that capture the true population parameter.

LO 3. Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal.

- In the case of the mean the CLT tells us that if

(1a) the sample size is sufficiently large (n ≥ 30 or larger if the data are considerably skewed), or

(1b) the population is known to have a normal distribution, and

(2) the observations in the sample are independent,

then the distribution of the sample mean will be nearly normal, centered at the true population mean and with a standard error of $\frac{\sigma}{\sqrt{n}}$:

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

When the population distribution is unknown, condition (1a) can be checked using a histogram or some other visualization of the distribution of the observed data in the sample.

The larger the sample size (n), the less important the shape of the distribution becomes, i.e. when n is very large the sampling distribution will be nearly normal regardless of the shape of the population distribution.

LO 4. Recall that independence of observations in a sample is provided by random sampling (in the case of observational studies) or random assignment (in the case of experiments).

In addition, the sample should not be too large compared to the population, or more precisely, should be smaller than 10% of the population, since samples that are too large will likely contain observations that are not independent.

LO 5. Recognize that the nearly normal distribution of the point estimate (as suggested by the CLT) implies that a confidence interval can be calculated as