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 ($n \geq 30$ 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

$$point\ estimate \pm z^{\star} \times SE,$$

,

where $z^{\star}$ corresponds to the cutoff points in the standard normal distribution to capture the middle XX% of the data, where XX% is the desired confidence level.

- For means this is: $\bar{x} \pm z^\star \frac{\sigma}{\sqrt{n}}$
- Note that $z^\star$ is always positive.

LO 6. Define margin of error as the distance required to travel in either direction away from the point estimate when constructing a confidence interval, i.e.$z^\star \frac{\sigma}{\sqrt{n}}$.

- Notice that this corresponds to half the width of the confidence interval.

LO 7. Interpret a confidence interval as “We are XX% confident that the true population parameter is in this interval”, where XX% is the desired confidence level.

- Note that your interpretation must always be in context of the data – mention what the population is and what the parameter is (mean or proportion).