Ch 7: Sampling and Sampling Distributions

- Analysis of a population is often based on a random sample taken from the population.
- Simple random sampling, the most basic sampling technique, uses random numbers to randomly select \( n \) items from the population in an unbiased way.
- A point estimate is a number calculated from sample data to estimate a population parameter.
- Sample mean \( \bar{x} \) estimates pop. mean \( \mu \), while sample proportion \( \bar{p} \) estimates pop. proportion \( p \).

Sampling Distribution: The sampling distribution of a statistic (such as \( \bar{X} \) and \( \bar{p} \)) is the distribution of values taken by the statistic in all possible random samples of a given size.

Central Limit Theorems (CLT)

- CLT #1: Let \( X \) be a Normal population with mean \( \mu \) and standard deviation \( \sigma \). Then the sampling distribution of \( \bar{X} \) is Normal(mean \( \mu \), standard error \( \sigma / \sqrt{n} \)) for any sample size \( n \).
- CLT #2: Let \( X \) be any population with mean \( \mu \) and standard deviation \( \sigma \). Then the sampling distribution of \( \bar{X} \) becomes Normal(\( \mu \), \( \sigma / \sqrt{n} \)) as the sample size \( n \) gets large.
- CLT #3: Suppose that \( np \geq 5 \) and \( n(1-p) \geq 5 \). Then the sampling distribution of \( \bar{p} \) is approximately Normal(mean \( p \), standard error \( \sqrt{p(1-p)/n} \)).

Ch. 8: Interval Estimation

- Point estimates (\( \bar{x} \) and \( \bar{p} \)) don’t say how much uncertainty is associated with them.
- Interval estimates are preferable; they are stated with some level of confidence.
- A Confidence Interval (CI) has the form: CI = [Point Estimate \( \pm \) Margin of Error]
- \( n \) = sample size, can be set to achieve a desired MOE & confidence level.

Case 1: CI for population mean \( \mu \): large sample case (\( n \geq 30 \))
- Margin of Error = \( z_{a/2} (\sigma / \sqrt{n}) \). If \( \sigma \) is unknown, use sample standard deviation \( s \) instead.
- \( z_{a/2} \) = z-value corresponding to \( \alpha/2 \) area in the right tail of the Z distribution.
- \( 1 - \alpha \) = Confidence level; \( \alpha \) (alpha) = Significance level.
- \( z \) = # of standard errors from the mean to give the desired confidence level 1 – \( \alpha \).

Case 2: CI for the population mean \( \mu \): small sample case (\( n < 30 \))
- Parent population from which we’re sampling must be nearly Normal in this case.
- Margin of Error = \( t_{df} (s / \sqrt{n}) \), where \( s \) = sample standard deviation.
- \( t_{df} \) = # of standard errors to go from the mean to give the desired confidence level.
- \( df \) = degrees of freedom = \( n - 1 \); as \( df \) increases, t distribution looks more like Z distribution.

Case 3: CI for the population proportion \( p \) = % of the population with a certain yes/no trait.
- Point Estimate = sample proportion \( \bar{p} \) = # of sampled items with the trait / sample size.
- Check that \( np \geq 5 \) & \( n(1-p) \geq 5 \) so that distribution of \( \bar{p} \) is approximately Normal.
- Margin of Error = \( z_{a/2} \sqrt{\bar{p}(1-\bar{p})/n} \)

Ch. 9: Hypothesis Testing (or “Significance Testing”)

- Null hypothesis \( H_0 \): \( \mu \) (or \( p \)) =, \( \leq \) or \( \geq \mu_0 \) (or \( p_0 \)) – often reflects the status quo.
- Alt. Hypothesis \( H_1 \): \( \mu \) (or \( p \)) \( \neq \), > or < \( \mu_0 \) (or \( p_0 \)) – often states what tester hopes to show.