Chapter 7 - Estimation

Estimation

Estimation:

- inferential statistics
- **Estimation** is a process where we select a random sample from a population & use a sample statistic to estimate a population parameter.
    - Parameter - measure used to describe an entire population
    - Statistic - any measure used to describe the sample pulled from population

- 2 types of **estimation**:
    - **Point estimation** - attempting to estimate the exact value of a population parameter.
    - **Interval estimation** - Creating a range of values which we believe that the population parameter is going to fall into.

- **Confidence interval** - range of values defined by the confidence level within which the population parameter is estimated to fall
    - for 90% confidence level, z value is 1.65
    - for 95% confidence level, z value is 1.96
    - for 99% confidence level, z value is 2.58
    - These were the **top 3 confidence levels** used in social sciences
    - You need confidence level to construct confidence interval
- **Confidence level** - Likelihood, expressed as a percent or probability that a specified interval will contain the population parameter. Expressed as a percentage or a probability. Basically, evaluates accuracy of the estimate of population parameters falling into the confidence interval.
    - To have more confidence, expand confidence interval
    - To be more precise, decrease confident level which results in a less expanded confidence interval
- To construct a **confidence interval** is:
    - $$CI = \bar{Y} \pm Z(\sigma_{\bar{y}})$$
    - Confidence Interval = Sample Mean & add & subtract the z value (z value is based on our confidence level) times the standard error.
        - → Calculate the product of the z value & the standard error first.
        - → Line over the y indicates standard error. Without the line over the y, it indicates standard deviation.
        - → **Standard error** is calculated by taking *population standard deviation* & divide it by the square root of the sample size, here is the formula:
            - $$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{N}}$$
        - → You may have to estimate the standard error instead of finding the actual standard error because the standard error isn't generally known. To calculate the **estimate standard error**, you divide the *sample standard deviation* by the square root of the sample size. Here is the formula:

- $$s_{\overline{Y}} = \frac{s_Y}{\sqrt{N}}$$

    - When calculating the confidence interval, the reason you add & subtract the values is to find the **upperbound** & the **lowerbound**

-

Reducing Risk

Reducing Risk:

- To reduce the risk of being incorrect in terms of our interval not containing the true population mean, we can increase the level of confidence being used
    - By increasing the level of confidence, this widens our confidence interval
- When we use the highest level of confidence (99% confident), there is only a 1% risk that we're wrong that the specified interval doesn't contain the true population mean

Sample Size & Confidence Intervals:

- By increasing the sample size, researchers increase the precision of their estimate, and as a result decreases the confidence interval size

Confidence Intervals for Proportions & Percentages

For Proportions:

- For Proportions, the formula is:
    - $$CI = p \pm Z\,(S_p)$$
    - CI = Confidence interval
    - P = observed sample proportion
    - Z = Z corresponding to the confidence level
    - Sp = Estimated standard error of proportions

→ Formula used for calculating **estimated standard error of proportions**: $S_p =$

Interpreting Results:

- We are x% confident that the true population proportion is somewhere between x and x
- Express the results in percentages as it's easier to understand percents

For Percentages:

- For percents, convert it into a proportion then use the proportions method of calculating the confidence interval.