Final Project Instructions

You may complete this project independently or with a partner.

Data Sources:

- Use data from your textbook data files, your area of study (in your major), or other sources such as WHO Global Health Observatory (https://www.who.int/data/gho), Kaggle (https://www.kaggle.com/datasets), UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php) , Government Open Data Portals (https://www.data.gov/) , World Bank (https://data.worldbank.org/)

Data Requirements:

- Sample size should be at least 25

Statistical Methods:

For each of the following inferential statistics methods, write research questions and perform the relevant analysis:

- t-test - compares 2 means
- z-test -compare two proportions
- Chi-square test for goodness of fit or association between two categorical variable or comparing more than two proportions
- Regression - uses a data set to find the adjusted effect of a factor by controlling other factors

Structure of Report

- Introduction
  - Background/Purpose
    - Briefly introduce the context and purpose of your study.
    - Why is this analysis important or interesting?
  - Research Questions
    - List the specific research questions you aim to answer using the statistical methods listed above.
- Methodology
  - Data Description
    - Describe your dataset: From where data are collected, how many observations are in the dataset, what variables are included? What type of data is it (e.g., quantitative, categorical)?
  - Data Analysis
    - Briefly explain the steps you will take to analyze the data (e.g., performing t-tests, regression)
  - Hypotheses
    - State your hypotheses for each inferential method.
      - Example: For a t-test: *Null Hypothesis (H_0)*: There is no difference between the means of the two groups.

    - *Alternative Hypothesis (H_1)*: There is a significant difference between the means of the two groups.

- Results
    - Descriptive Statistics: provide basic statistics such as mean, median, standard deviation, etc.
    - Inferential Statistics: include results from the tests (e.g., p-values, confidence intervals).
    - Include appropriate charts and graphs for each type of statistic (e.g., bar graphs, box plots, scatter plots) to help communicate what the statistics show.
    - Write brief sentences interpreting what the statistics reveal.
    - Example: The t-test results show that there is a significant difference between the means of Group A and Group B (p-value < 0.05).

- Discussion
    - Meaning/Implications of Results
        - Discuss the significance of your findings.
        - What do these results mean in the context of your research question?
    - Limitations
        - Discuss potential limitations in your study (e.g., sample size, biases in data, potential confounding variables).