# MATH108: Elementary Stats

*Spring 2024*

## Final Project: Applying Your Statistical Skills

*Assignment is DUE as indicated on the course schedule.*

*This is a **group assignment**. Work with 2-3 classmates, and submit as a group on Gradescope. **Individual assignments will not be accepted unless prior approval is obtained.***

### Overview

For your final project you will design, conduct, and analyze an experiment. You can investigate any topic of interest to you *as long as your experiment can be analyzed using hypothesis testing techniques we have covered in this class*.

This project will be a large portion of your final grade, and is broken up into milestones, described below. Be sure to submit each milestone on time and to put your best effort into all pieces. Your final grade will be based on your grades for all milestones.

**All milestones must be well formatted and readable**. Text must be proofread, use good grammar, and communicate clearly.

### Milestone 1: Proposal – 20 points

For your project proposal you will identify your project group, a plan for working together, your topic of interest, and potential roadblocks. Type up a document that answers the following questions. Your document should be 1 – 2 pages long and well formatted. ==Submit as a group on Gradescope before class on the proposal due date==. Your proposal will be reviewed in class so that you get quick feedback.

1. Group
   a. With whom do you plan to work? Groups must be 2 – 3 members (you must speak with me if you would like to work with a different size group).
   b. Talk about your schedules. Detail a plan for coordinating your work throughout the project.

2. Topic
    a. What topic area will you work with? Why is investigating this area important/interesting?
    b. What is the broader impact of your analysis? (What will other people use it for?)
3. Building Blocks
    a. What is the specific research question for your analysis? Remember, your research question needs to be a question that you can investigate using methods we have learned in class.
    b. What variables will you need in your dataset to answer your research question?
    c. What will the types of those variables be?
    d. What specific kind of hypothesis test (z-test, t-test, paired t-test, chi-square test) will you perform to answer your research question?
4. Roadblocks
    a. What roadblocks do you anticipate as you complete your project?
    b. Detail your plan for dealing with these roadblocks. How will you overcome them, or if you cannot, how will you modify your project?

Points will be awarded for answering each question above fully (each question is 2 points, 0 points will be awarded for missing or nonsensical answers, 1 point for partial answers, and 2 points for complete and correct answers).

## Milestone 2: Check-in – 20 points

Your first check-in will focus on finding a dataset to work with, understanding the data, and making substantial progress on your analysis. **The work you submit for your analysis is not expected to be complete**; it is expected to show that you have made significant progress towards completing your project.

1. Data Source
    a. Find or collect data that you can use to perform your analysis. If you find data, provide a link to your data source. If you collect data, provide a link to your data table in a spreadsheet.
    b. Investigate the source of your data (even if it was you). Who collected it? Who provided the funds for data collection? Who published it? What possible data biases do the answers to these questions raise?
    c. How was data collected? What sampling technique was used? Does the sampling technique indicate any potential downstream issues to your analysis?

2. Exploratory Data Analysis
   a. What variables are included in your dataset? What is the type of each variable (type here means mathematical type)?
   b. Which variables will you use for your analysis? Which of these are explanatory variables and which are response variables?
   c. For each individual variable you plan to use, perform EDA. What is the distribution of the variable? Is it what you expected? Do there appear to be any missing values or distribution issues that could affect your analysis?
   d. Continue performing EDA to investigate any patterns between variables in your dataset. Do you notice anything that impacts your analysis plan?
   e. What data issues and biases did your EDA reveal? Ex. Do some variables have a lot of missing data? Is that missing data concentrated to specific observations?
3. Data Cleaning [**You only need to show that you have started this section, it does not need to be complete for this milestone**.]
   a. Which of the issues and biases that you found during EDA will you correct for your analysis? Why are these important to correct?
   b. Correct the issues / biases you identified above. Clearly record the steps you take so that your analysis is replicable by other researchers.
4. Hypothesis test [**You only need to show that you have started this section, it does not need to be complete for this milestone**.]
   a. Perform the appropriate hypothesis test given your research question and data. Remember that your test must include:
      - Hypotheses
      - Test statistic
      - Null distribution
      - P-value

Points will be awarded for answering questions 1 and 2 above fully (each question is 2 points, 0 points will be awarded for missing or nonsensical answers, 1 point for partial answers, and 2 points for complete and correct answers). Zero to 4 points will be awarded for questions 3 and 4 depending on the amount and clarity of progress shown.

## Milestone 4: Final Report and Presentation – 20 points

Your final milestone includes all the final deliverables for your project. Your submissions should be well formatted, proofread, and clear. **Submit on as a group Gradescope before class on the day presentations begin. No late submissions will be accepted. There will be no exceptions to this rule.** You will submit a final report, and give a presentation to the class.

1. Your **final report** should include:
    a. Clear formatting.
    b. Clear, proofread writing.
    c. An introduction to your project with your research question and a brief explanation of why your analysis is interesting and important.
    d. A data overview explaining your data source (who collected the data, who funded collection, who distributed the data), and any biases, ethical issues, or general issues with the data.
    e. An overview of your data cleaning process. How did you clean the data and why did you choose to do it that way?
    f. Walk the reader through the hypothesis test. What calculations did you do? What did the output of each tell you in the context of your research question?
    g. A conclusion. What did you find with respect to your research question? What are the implications of this finding?
    h. Citations for any sources referenced.
2. Your **presentation** should include:
    a. Clear formatting.
    b. Appropriate visualizations.
    c. Clear speaking.
    d. Contributions from all group members (each person must talk).
    e. An introduction to your analysis, including motivation for your research question.
    f. An overview of your data source.
    g. A summarized version of the EDA you performed on your dataset.
    h. A discussion of any data issues you encountered and how you addressed them.
    i. Walk the audience through the hypothesis test. What calculations did you do? What did the output of each tell you in the context of your research question?
    j. A conclusion. What did you find with respect to your research question? What are the implications of this finding?
    k. Future work. What new questions do you have? How would you build on or follow up on this project?
    l. Time for Q&A. [In total, your presentation should be ~15 minutes.]

Points will be awarded for addressing each bullet point above fully (each question is 1 point, 0 points will be awarded for missing or nonsensical work, 1 point for complete and correct work).

## Reflection

Your reflection is an individual portion of the project that you will submit on Gradescope individually. The purpose is to reflect on your own work, and how your group worked together.

Write a few short paragraphs that address these points:

1. Your specific contributions to the project.
2. Your teammates' specific contributions to the project.
3. Whether you navigated any conflict or discrepancy in workloads with your teammates.
4. How you navigated those conflicts or redistributed work.

You will not receive points for your reflection; however answers may be used to adjust individual's project grades if the distribution of work was not even.