**Math 58B: Preliminary Data Analysis**

**The Islands Activity[[1]](#footnote-1)**

**Due: Tuesday, April 18, 2023**



|  |
| --- |
| The goal of the preliminary data analysis assignment is to start working with the variables including visualizations, numerical summaries, and modeling. |

You will now begin to work in your project groups of approximately 3 people. If you have a group already, great, let me know! If you do not have a group, please let me know that you’d like me to put you in a group. You must complete this assignment with the same people who will be in your final Island project group.

Reminder for logging on: Go to the Islands (<https://theislands.umn.edu/login.php>). If you have forgotten your password, click on ‘Need to set or reset your password?’, enter your 5C email address (most likely of the form aaaa20xx@mymail.pomona.edu), and the Islands will resend your password to you. Let me know if you cannot access the Islands.

Note: The Visitor Center on the Islands (click on the square just south of Arcadia) is a great source for information about the Islands.

Don’t forget about the **Academy**. What is there? Are there any studies that seem interesting? Is the response variable quantitative or categorical? How many explanatory variables are there? Are they each quantitative or categorical? How many observational units total were collected in the study?

**FINAL REPORTS / OUTCOMES**

3/28/23: **Pilot Study** -- research question, methods section, preliminary data

4/4/23: **Power Analysis** – calculation of sample size from the pilot data

4/18/23: **Preliminary Data Analysis** -- data summaries / visualizations, first draft of analysis, including regression

5/2/23: **Complete Draft** – including multiple regression

5/4 + 12/23: **Presentation** -- 2 slides, < 5 min presentation (either during our final time or during reading days)

5/12/23: **Final** (edited) **Project**

**TO TURN IN: a preliminary data analysis study**

* Remind the reader your research question and enough information about the study that the variables are clear to the reader.
* Is there anything about how you sampled which could make your results difficult to generalize to all Islanders? Or to Islanders other than those you measured?
* 3-4 (more?) visualizations that describe the variables collected as part of the research project. All plots should be accurate and descriptive labels.
* For every visualization, at least one sentence (maybe more!) describing the plot. Do not create any naked plots that are not described.
* Any relevant numerical summaries. For all numerical summaries, provide at least one sentence describing the numerical summaries. No naked tables.
* At least one of the plots and/or numerical summaries should speak to the issue of whether or not your randomization worked. Remember, if your randomization worked then the distribution of the demographic variables should be the same across your treatment groups.
* The output of at least one regression model. One sentence describing the model predicting the response variable at hand.

As before, each individual student should turn in the project on Canvas. Each member of the group will all turn in an identical assignment.

For this assignment, **turn in an .Rmd file and a .pdf**. Having a single .Rmd file will be complicated for collaboration, but a single document will synthesize your group effort. Make sure that everyone is contributing to both the writing and to the computational analysis. It is important to do the work in the markdown file so that the R code can be integrated with the narrative.

**PRELIMINARIES**

* Do all of your work through \*one\* person’s login to the Islands (that is, share the login information!). Make sure the person’s password is not a password they use for any other accounts.
* Data collection should take place on a shared Google Sheet. Everyone should participate in the data collection process. [Keep data tidy: each row is a person; each column is a variable.]
* Writing should be done through a shared Google Doc. Everyone should participate in the writing process.
* The response variable must be quantitative.
* The explanatory variable of interest must be categorical (binary is great, but 3-4 levels is also okay).
* You will also collect demographic information on the Islanders you measure. Ideally, at least one of the demographic variables will be numeric.
* **Note: Multiple islanders can be completing tasks at the same time, but each islander can only complete one task at a time.**

**ROLES**

Each individual will have two roles in the project. They will have an organizational role (to facilitate the group dynamics) as well as a technical role (to facilitate the project). See the bottom of this handout for more details. The roles can change over the life of the project, but should be clearly designated for each assignment / step in the process.

* Group Dynamic Roles: Project Manager, Task Manager, Facilitator
* Project Roles: Director of Research, Director of Computation, Reporter

**RESEARCH QUESTION**

Develop a research question that is both interesting and feasible for an Island project. Check out the PICO structure which helps implement evidence-based practice in medicine[[2]](#footnote-2). Keep in mind that your question could be either experimental or observational in nature.

Your final research question should be in the form of a sentence or two describing: the population, the treatment or exposure, the comparison, and the outcome measurement.

**METHODS SECTION**

Along with the research question, you should start writing the methods you will use to investigate your research question. Your methods section should address the following (prepare as much as possible for this report). Eventually, your methods section should fully describe:

* the planned study design (e.g., experimental study, case-control study, cross-sectional study, or cohort study),
* the target population/participants,
* the inclusion and exclusion criteria (kids? pregnant women? …),
* the sampling methods (if any),
* the intervention(s) (or exposure(s)): explanatory variable, should be categorical,
* the outcome(s) to be measured: response variable, should be numeric,
* other demographic variables collected: at least 4, two of which must be numeric,
* the sample size calculation and justification (will be in next project),
* the randomization (random assignment) methods (if any),
* the blinding methods (if any),
* the study implementation (including timing), and
* the statistical methods used to analyze the data (will be in final project).

Your methods section should be as long as it needs to be to cover the items above but no longer. Avoid repetition and write clearly and concisely.

**PRELIMINARY DATA COLLECTION**

* Consider/report on how you selected people to be in the sample. You do not need to come up with a pure random sample (i.e., having all Islanders in a hat and pulling out 20 names), but you should think carefully about how you will find the individuals. What population can you represent?
* Measure at least 20 Islanders
* The response variable of interest must be quantitative.
* The explanatory variable of interest must be categorical (2-4 levels).
* Collect at least 4 additional pieces of information on each Islander you measure (2 of the variables must be quantitative).

**GROUP PROJECT ROLES:**

**Norms for the Class and Projects**

1. No one individual is as smart as the entire group.
2. Everyone has something to offer.
3. Everyone has a responsibility to help others in their group.
4. Everyone has a responsibility to understand the full project and to ask for help if they need it.

**Roles for the Projects**

Each individual will have (at least) 2 roles; one role to help foster the group dynamics and the other role to divvy up the responsibilities involved in completing the project.

**Group Dynamic Roles:**

* *Project Manager:* Makes sure that the group is organized and has a clear plan for completing the project. This includes scheduling meetings and having a plan for what needs to be done before the next meeting.
* *Task Manager:* Makes sure that everyone knows what they are expected to do before the next meeting and makes sure that they do it. This might involve calling or emailing each person between meetings to discuss what they have/haven’t done. If someone can’t do the work that needs to be done, the task manager is responsible for calling another meeting if needed.
* *Facilitator:* Makes sure that every member of the group is participating and being listened to and heard. This might involve asking questions of a member that’s been silent and stopping others when someone’s comment is being overlooked.

**Project Roles:** In each case,the person assuming the role is responsible for that aspect of the project. **It doesn’t mean that they will do all that part of the project by themselves**; it means that they are responsible for **dividing** that work up among the members of the group and ensuring that it is done and recorded correctly.

* *Director of Research:* Is responsible for the literature searches / background research. The Director of Research identifies what needs to be searched for in the literature / the Academy, divvies up the background work to be performed among the group members, and coordinates changes in the searches based on information gathered and changes in direction. They are also responsible for making sure that the citations in the project are complete and accurate.
* *Director of Computation:* Is responsible for the computer programs involved in the project. The Director of Computation is responsible for designing the code so that different people can write different parts of the code. The programmer is responsible for making sure that any code written by different people can be integrated.
* *Reporter:* Is responsible for the written report. This involves taking notes during the complete process in order to keep a record of what has been done. The reporter may also gather everyone’s individual notes and put them together. The reporter is also responsible for editing the final report and making sure that the various pieces (that may have been written by different people) fit well together.

**Things to think about:**

1. If other members of a group think that one member isn’t pulling their weight, then you can come and talk to me about it. We will figure out what to do!
2. (Generally) All members of the group get the same grade for the project.
3. Grade on projects (I will send out a rubric for grading toward the end of the semester, but generally the grade will be based on the following):
* Grade for technical depth and sophistication.
* Grade for quality of write-up (organization, clarity, grammatical correctness, appropriate use of graphs, tables, formulae).
* Grade for quality of oral presentation (organization, clarity, appropriate use of graphs, tables formulae, ability to answer questions).
* Grade for quality of group work and distribution of labor.
1. Attendance: you should keep track of who is or isn’t showing up to group project meetings. Students should keep a record of all the times that they worked on the project and the work that they did. Group members should be spending roughly the same amount of time on the project.
1. Thanks to Laura Le and Ann Brearley at UMN for sharing resources. Much of the assignment is taken directly from their excellent resources. [↑](#footnote-ref-1)
2. Description: <https://guides.mclibrary.duke.edu/ebm/pico>; Sample PICO questions: <https://libraryguides.nau.edu/c.php?g=665927&p=4682772> [↑](#footnote-ref-2)