

## IE 2028 Project

**Project Purpose:** to create a portfolio of evidence of what you have learned in 2028.

**General overview:** In this project, you will choose a topic, develop a list of questions to investigate, collect publicly available data, conduct analysis, and draw conclusions. You will imagine a client who has hired you to conduct the investigation. You may also invent your own consulting firm.

### Requirements:

1. You will work in a team of two students and are jointly responsible for the entire project or you may work alone. If you work in a team, each person will receive the same grade on the project, unless there are serious mitigating circumstances.
2. You may only use data which is publicly available. This is to avoid needing IRB approval to collect your data.
3. Your dataset must have at least 20 observations, but it is better to have about 50.
4. Your dataset must have at least 3 continuous variables and at least one of each of the following types: binary, categorical, ordinal. In total, you must have at least 7 different variables.
5. The different modules will be integrated into a single, coherent report. You are encouraged to use an Rmd (R markdown) file or R notebook to keep your work organized.
6. Please use APA style for referencing, figure titles, etc. Any specific instructions by Dr. A will supersede APA style.
7. The report should have a title page, a table of contents, and be in a three-ring binder or be spiral bound. It must lay flat for grading.
8. Please use Times New Roman, 12 pt type with 1 inch borders. Use 1.5 line-spacing so that we have room to give you feedback.

**Writing Style** The writing style should be professional, not colloquial. Use first person (I or we) and do not address the reader (e.g. “you can see...”). Use active voice with strong verbs (no “it can be seen...” and no “we will attempt to...” and no “we looked at”). Yes to verbs like “analyze”, “investigate”, “examine”). A thesaurus is very useful.

The report is NOT a list of answers to the following questions. The questions are to guide your thinking and organize your work. For each paragraph/sentence, think about what purpose it has in the report --- what is it doing for your reader, e.g. giving a framework, orienting the reader, giving information, connecting two ideas, summarizing, etc.

A great resource is the Online Writing Lab (OWL) at Purdue University.

**To maximize your grade:** As you are choosing what questions you will answer and how you will model your variables, choose a variety of concepts and procedures in order to demonstrate the breadth of your understanding. For example, stronger project will include both a hypothesis test on the population mean as well as one on the population proportion instead of two tests on means.

### **Module 1: Introduction and Data Collection**

This module is to orient your reader to the motivation and structure of your project. It should form a “funnel”, beginning with the big picture and context, then funneling down to the specific questions. Remember that you have a generic reader. For example, even though you may have chosen sports data because you are a big fan, your reader may not even know the sport of your context. Connecting your project to larger issues of economy, society, education, etc. makes the motivation stronger.

1. What is the context of your project? Why is it important and interesting?
2. What “big” questions are you investigating?
3. Data collection
  - a. Where did you get it?
  - b. How was it collected?
  - c. Did you need to clean it? If so, how did you clean it?
  - d. What is your observational unit? Sample? Population? Is it a random sample? Why or why not?
  - e. Variable table: name you use in report (not necessarily the coding name), the concept it represents, variable type, range of reasonable values
4. Give the reader a guide to the rest of the project and how it is laid out.

Note: This section should not be a list of questions and their answers. The questions are just to guide you and suggest areas for discussion.

### **Module 2: Modeling data with Probability Distributions**

In this module, you demonstrate your ability to model data with probability distributions. Different types of data (binary, categorical, ordinal, and continuous) may have different approaches to choosing and justifying a model.

For one variable of each type, summarize the data numerically and graphically, then model it with a specific probability distribution. Justify your parameter estimates. Examine the fit of your model to your data. Only use the procedures which are appropriate for that data type; for example, do not compute the mean of nominal data and do not fit a Normal distribution to categorical data.

In this section, be sure to demonstrate your understanding of different types of graphs, measures of central tendency and dispersion, parameter estimation, confidence intervals, QQ plots, and goodness of fit tests. In order to earn an A, you do not have to find a perfectly fitting model for each variable, but you do have to test a reasonable one thoroughly. If you want to model more than one variable of any type, you may; but you are only required to include one variable of each type.

### **Module 3: Single sample analysis**

In this module, you will demonstrate the analyses you have learned that involve only one sample from one population. You will tell a story by posing questions, doing the analysis, and interpreting the results.

Hypotheses tests form the basis of this module. The constants that you use (e.g.  $\mu_0$ ) should be justified, not pulled out of the air. You can base these constants on different data sets (e.g. historical data produces the constant; while you use current data as your sample) or on research (e.g. median household income reported from U.S. Census).

One hypothesis test must be done “by hand”, which means that you show every step of the process (but can be confirmed via R). This one must still be typed up. At least one analysis should include the power calculation for a particular alternative (done by hand or via R or via G\*Power or a combination of these). An “A level” project will include at least one two-sided hypothesis and at least

one one-sided hypothesis. All HT discussions should include null and alternative hypothesis, test statistics, alpha, and conclusions. Assumptions for HTs should be justified.

#### **Module 4: Two sample analysis**

In this module, you will demonstrate the analyses you have learned that involve two populations or two samples from the same population. You will tell a story (or stories) by posing three questions, doing the corresponding analyses, and interpreting the results. Module 3 guidance also applies to this module.

#### **Module 5: Prediction**

In this module, you will demonstrate the analyses you have learned to predict the value of one variable using the values of one (or more) variable(s). You will tell a story (or stories) by doing a linear regression and analyzing the results. All computations can be done using R. Illustrate the use of the model to predict the value of the response variable for a value of the predictor for which you do not have data. Include an analysis of the residuals and a discussion of the reasonableness of the assumptions.

#### **Module 6: Summary and Conclusion**

In this module, you will summarize your investigation's findings and draw conclusions. You will return to the "big" questions and project motivation and discuss how your investigation partially answers the big questions.

**References:** Use APA reference guidelines, except for web sites. When you reference web sites, give the complete URL in addition to what APA requires. Include in your references, any and all sources that help you complete this project. Conversations and peer reviews should be included as "personal communication."

#### **Appendices:**

- A. One page of your data spreadsheet. I need to see all variables (names and a few observational units' values), but not all observations.
- B. A billing invoice from you to a fictitious client who hired you to do this analysis.
- C. A consulting log summary that documents, on a weekly basis, the time you spent and what you accomplished on the project.
- D. Acknowledgements – thank people who have helped you accomplish this project

#### **Reflection**

In addition, each project partner will separately turn in a personal, private reflection. Instructions for the reflection are in a separate document.