

Stat 313: Project 1

Write-Up Guidelines

You will be summarizing your results in a (guided) written report following the **Project Writing Guidelines** posted on Canvas. The results **must** be written in the RMarkdown template provided. You can include all pertinent plots inline in the typed document, rather than in the appendix. Your group will be submitting both the RMarkdown file used to generate your report and the knitted HTML file to Canvas by **Sunday, October 25 at 11:59pm**.

Begin by downloading the RMarkdown and data files from Canvas and saving them on your computer in the **same** folder. Then open the RMarkdown file in RStudio and run the first chunk of code reading the data in. If you encounter an error, make sure that the data are located in the same folder as the RMarkdown file!

Data Background

The National Center for Education Statistics (NCES) administers the Integrated Postsecondary Education Data System (IPEDS) ([web link](#)), which is a large-scale survey that collects institution-level data from postsecondary institutions in the United States (50 states and the District of Columbia) and other U.S. jurisdictions. IPEDS defines a postsecondary institution as an organization that is open to the public and has the provision of postsecondary education or training beyond the high school level as one of its primary missions.

Currently, there are 12 survey components that comprise the annual IPEDS data collection cycle. The 12 survey components are separated into one of three seasonal reporting periods—fall, winter, or spring. We will be considering data from the 2018 - 2019 fall and spring collection periods, namely data on Institutional Completions and Student Financial Aid. In the data from the surveys, each university is given a unique identifying code. For example, Cal Poly is 110422. For more information about the IPEDS survey, see <https://nces.ed.gov/ipeds/ReportYourData/IpedsSurveyMethodology>.

As is typical in educational data, students are not randomly assigned to receive financial aid, but instead receive grants based on merit and need. As universities focus more on improving the diversity of their graduates, they will be looking into what may be driving factors of students' success.

For this project, we will consider the following data for the universities surveyed by IPEDS:

- the percentage of graduates who identify as Black
- the degree of the graduates (Associates, Bachelors, Masters)
- the percentage of undergraduate students who receive federal, state, local, or institutional grants or support

Specifically, we will explore the relationship between the percentage of graduates who identify as Black and the percentage of undergraduates who receive some form of grant support. Additionally, we will investigate if and how this relationship differs across degrees.

Introduction (5 pts)

- (2 pts) Give a brief background of the research problem and how the data were collected.
- (3 pts) Clearly outline the question(s) of interest you will address with the statistical analysis. The more specific you define the question of interest here, the easier the rest of the analysis and report will be. The research questions should start with, “What is the relationship between...” and should be as specific as possible. Your *Summary of Statistical Findings* should directly answer the question(s) you pose here.

Statistical Methods (15 pts)

This section should lay out the steps, decisions, and logic leading to the statistical model you will use to answer the research question of interest.

- (3 pt) Describe the response and explanatory variables.
- (3 pts) Produce data visualizations exploring the relationship(s) you are interested in investigating. Describe what you see in the visualizations, making direct references to the plots!

Note: Keep in mind that there are 1654 observations in the dataset, so you may want to use tools that help to alleviate overplotting!

- (5 pts) Describe the appropriate statistical method you will use to answer the question(s) of interest that you stated previously. Be specific about why the method being used are appropriate for the investigation at hand (e.g. types of variables).
- (4 pts) Check all model conditions of the statistical method you used. Describe what each condition is **in the context of these data**. Reference and include appropriate plots necessary for checking the model conditions. What are your conclusions regarding the conditions? Justify your conclusions!

Summary of Statistical Findings (10 pts)

In this section you will write up your findings for each research question of interest.

- (5 pts) What is your conclusion for the questions of interest? Namely, “What is the relationship between the percentage of graduates who identify as Black and the percentage of undergraduates who receive some form of grant support?” and “How does the relationship differ by degree awarded?”. Base your conclusion on the visualizations you created and the regression model you found. Tailor your conclusion to communicate your results to a university administrator.
- (5 pts) Interpret **in the context of the data** each of the coefficient estimates you got from R.

Scope of Inference (4 pts)

Write a brief Scope of Inference statement. Specifically, answer these two questions and comment on their implications:

- (2 pts) Were the observations randomly selected from some larger population? Based on the sampling method used, what larger population can you infer the results to?
- (2 pts) Was the explanatory variable randomly assigned to observations? Based on the study design, are cause-and-effect statements justified?

Make sure you write the scope of inference specific to the language of the data (not just generic statements)!

Project Presentation (3 pts)

- (1 pts) Your report should not have any spelling errors! To check for spelling errors in RStudio, click the green check mark button next to the “Knit” button.
- (2 pts) Your report should look as neat and professional as possible. Make sure that your figures don’t end up in the middle of your paragraphs, and that your sections have headings. If you would like to fine tune the appearance of your report, please post questions to Canvas and I will respond ASAP.

Note: Please try to make the output from the statistical model look as nice as possible. Output from `summary()` looks terrible in a statistical report! Try using the `get_regression_table()` function from the **moderndive** package or the `tidy()` function from the **broom** package.

Group Evaluation (3 pts)

Each member of your group will fill out a group evaluation form detailing each member’s contributions, cooperation, communication, and participation.