Midterm Project Guidelines

Stat 313

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 24 at 11:59pm**.

Begin by accessing the Midterm Project RMarkdown template on RStudio Cloud. Open the RMarkdown file in RStudio and run the first code chunks loading packages and loading the data.

Data Background

Today, most colleges require the teaching of faculty to be evaluated through anonymous student evaluations. Historically, student evaluations date back to the 1920s, where faculty were in favor of this form of evaluation replacing the casual judgments of lone administrators. Faculty were largely dissatisfied with the alternatives, and eventually most accepted the premise of student evaluations. *Grading the College* (Gelber, 2020) suggests student questionnaires actually appealed to professors because they generated large numbers of mostly-positive reviews, required little labor to implement, and yielded results that seemed to be statistically reliable. By the mid-1970s Gelber estimates nearly three-quarters of faculty members believed course evaluations should be used in tenure and promotion proceedings.

Professors today, however, look a lot different than professors in the 1920s or even the 1970s. Numerous researchers have found teaching evaluations to be biased against women and people of color (Fan, Shepherd, Slavich, Waters, Stone & Abel, 2019; MacNell, Driscoll & Hunt, 2015; Mengel, Sauermann & Zölitz, 2018; Storage, Horne, Cimpian & Leslie, 2016; Mitchell & Martin, 2018). It is believed that student evaluations often reflect biases in favor of nonteaching-related characteristics, such as the physical appearance of the instructor. The 2005 Economics of Education Review article titled *Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity* found that instructors who are viewed to be better looking receive higher instructional ratings.

In this project, we will explore the relationship between professor's teaching evaluation scores their attractiveness. The data come from end of semester student evaluations for a large sample of professors from The University of Texas at Austin. These data have been merged with descriptors of the professors and the classes. In addition, six students rate the professors' physical appearance.

The original data set published in *Beauty in the Classroom* (Hamermesh & Parker, 2005) are available through the **openintro** R package. Below is a list of select variables in the data set and their descriptions. There were six total students who rated the attractiveness of each professor based on their picture. Attractiveness ratings ranged from 1 (lowest) to 10 (highest). The `bty_avg` score is an average of these six attractiveness scores.

variable	description
course_id	Variable identifying the course (out of 463 courses)
prof_id	Variable identifying the professor who taught the course (out of 94 professors)
score	Average professor evaluation score: (1) very unsatisfactory - (5) excellent
rank	Rank of professor: teaching, tenure track, tenured
ethnicity	Self-identified ethnicity of professor: not minority, minority
gender	Self-identified gender of professor: female, male
language	Language of school where professor received education: English or non-English
cls_level	Class level: lower, upper
cls_profs	Number of professors teaching sections in course in sample: single, multiple
cls_credits	Number of credits of class: one credit (lab, PE, etc.), multi credit
bty_avg	Average beauty rating of professor

Your Task

For this project, you are expected to investigate the relationship between a professor's course evaluation score and their average beauty score. You will also explore how this relationship differs across the levels of a categorical variable.

You have the flexibility to use *any* of the categorical variables listed above in your multiple regression model (e.g., gender, rank, ethnicity, language, etc.). You will be expected to justify *why* you chose the variable you chose in the *Introduction* section of your report. You **cannot** use prof_id or course_id, as those are unique identifiers for each professor and each course and would result in a perfect model.

Introduction (5 pts)

- (2 pts) Give a brief background of the research problem and how the data were collected. Make sure to describe who the unit of study is (e.g. professors? classes? universities?)!

- (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 (10 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.

- (1 pt) Describe the response and explanatory variables.

- (2 pts) Provide a summary table of both the mean and median of evaluation scores and average beauty scores.
- (2 pts) Produce data visualizations exploring the relationship(s) you are interested in investigating.
    - (1 pt) Describe what you see in the visualizations, making direct references to the plots!

Note: Keep in mind the tools that help to alleviate overplotting!

- (4 pts) Outline the appropriate statistical model 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).

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 course evaluation scores and average beauty scores?" and "How does the relationship differ by _____?" Base your conclusion on the visualizations you created **and** the regression model you found. There should be **no** mention of p-values!
- (2 pts) Using the coefficients you obtained, write out the estimated regression equation for each level of your categorical variable.
    - (3 pts) Interpret **in the context of the data** the coefficients from each regression equation.

Scope of Inference (5 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 these results to?
- (2 pts) Was the explanatory variable randomly assigned to observations? Based on the study design, are cause-and-effect statements justified?
- (1 pt) 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 with ABC over it, 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 Discord and I will respond ASAP.

Note on Figures: I expect that the figures *are included in the section they are discussed* not at the end of the report.

Note on Model Output: Please try to make the output from the statistical models look as nice as possible. Use the `get_regression_table()` function from the **moderndive** 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.

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

- It is not expected that every group member is an expert on these topics.
- Rather, it is expected that every group member articulates what they are and are not comfortable contributing.

- Every member of the group can (and should) contribute to proof reading your final report!

If you take control of your group's project and do not let others contribute, your grade will be deducted 20%.

If you fail to contribute to your group's project, your grade will be subject to my discretion.

References

Boring, A, K Ottoboni, and P. B. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research*.

Fan, Y, LJ Shepherd, E Slavich, D Waters, M Stone, and R Abel. 2019. "Gender and Cultural Bias in Student Evaluations: Why Representation Matters." *PLoS ONE*, no. 2, 14.

Hamermesh, Daniel S., and Amy Parker. 2005. "Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity." *Economics of Education Review* 24 (4): 369–76.

MacNell, L, A Driscoll, and A. N. Hunt. 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education*, 291--303.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2018. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association* 17 (2): 535–66.

Mitchell, K M W, and J Martin. 2018. "Gender Bias in Student Evaluations." *Political Science & Politics*, no. 3, 51: 648--652.

Peterson, Lori A. AND Andersen, David A. M. AND Biederman. 2019. "Mitigating Gender Bias in Student Evaluations of Teaching." *PLOS ONE* 14 (5): 1–10.

Spooren, Pieter, Bert Brockx, and Dimitri Mortelmans. 2013. "On the Validity of Student Evaluation of Teaching: The State of the Art." *Review of Educational Research* 83 (4): 598–642.

Storage, D, Z Horne, A Cimpian, and S-J Leslie. 2016. "The Frequency of "Brilliant" and "Genius" in Teaching Evaluations Predicts the Representation of Women and African Americans Across Fields." *PLoS ONE* 11 (3).

Uttl, Bob, Carmela A. White, and Daniela Wong Gonzalez. 2017. "Meta-Analysis of Faculty's Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning Are Not Related." *Studies in Educational Evaluation* 54: 22–42.