Final Project Guidelines

Stat 313

Write-Up Guidelines

You will be summarizing the results of your analysis in a written report. The results **must** be written in the RMarkdown template provided. You can include all pertinent plots inline in the typed document, **not** in an appendix. Your group will be submitting both the RMarkdown file used to generate your report and the knitted HTML file to Canvas by **Sunday, December 5 at 11:59pm**.

Begin by accessing the Final 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

For this project, you are working with the *same* `evals` data from the Midterm Project. You were provided background information about these data for that project. You are expected to go back and re-read the introduction to these data **before** you begin your project. Additionally, the article where these data are from has been posted on Canvas for your reference.

Components of this project will require that you revise portions of your previous project. Thus, it will be important for you to read (hopefully, re-read) the comments I left your group on your previous submission.

Data Wrangling

You have been provided code which has two purposes. The first three lines of code take the original `evals` dataset, groups by the professor, and then randomly samples **one** observation for each professor. The next portion of code (in the `mutate()` function) creates a new variable, `attr_cat`, which contains three distinct groups of attractiveness ratings: low, medium, and high. The groups were chosen based on the following criteria:

- "low" – average attractiveness was scored below a 4
- "medium" – average attractiveness was scored between a 4 and a 7
- "high" – average attractiveness was scored above a 7

Your Focus

In this project, you will will explore the relationship between course evaluation scores and attractiveness using ANOVA instead of multiple linear regression. Your analysis consists of two components: a one-way ANVOA and a two-way ANOVA.

In part one, your task is to use the three attractiveness groups to investigate if there is evidence of a difference in the mean course evaluation scores for at least one of these groups.

Next, in part two, you will include a second categorical explanatory variable into your ANOVA. You will use the **same** explanatory variable you used in your Midterm Project. Your investigation will focus on whether there is evidence that the relationship between course evaluation scores and attractiveness differs based on your categorical variable. This should sound **extremely** familiar, as it is the *same* investigation you undertook for the Midterm Project, except now both of your explanatory variables are categorical.

Introduction (5 pts)

- (2 pts) Take your original Introduction from the Midterm Project and spice it up! At a bare minimum your revised Introduction should include:

  - a persuasive argument as to *why* this investigation is worthwhile
  - substantial background of the data (e.g., year of collection, institutional context, the unit of study, number of observations)

- (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.

- (2 pts) Provide a summary table of the mean, standard deviation, and sample size for the course evaluation scores of each attractiveness group.

- (1.5 pts) Produce a data visualization exploring the relationship between course evaluation score and attractiveness group.

- (1.5 pts) Produce another visualization that incorporates your chosen categorical variable into your previous visualization.

  - Hint: Use either `fill` or `facet` to accomplish this!

- (1 pt) Describe what you see in the visualizations, making direct references to the plots!

- (4 pts) For each of your research questions, outline what statistical model you will use to address the question. Saying "to address our first question we will use an ANOVA" will earn you very few points. I expect you to be specific about **why** the method being used are appropriate for the question at hand (e.g. types of variables).

- (5 pts) Check all model conditions for the statistical methods you used.

  - Describe what each condition is **in the context of these data**.
  - Reference and include appropriate plots necessary for checking the model conditions.
  - Justify your conclusions regarding each condition!

Summary of Statistical Findings (10 pts)

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

- (2 pts) What is your conclusion for the first research question? Namely, "Does the mean course evaluation score differ for at least one attractiveness group?"

  - Base your conclusion on the visualizations you created **and** the one-way ANOVA model you fit to these data.
  - When communicating the results of the statistical test make certain you include the value of the test statistic and the distribution it follows under the null hypothesis.
  - If you make a rejection decision, you **must** specify an α !
  - **Do not** use the word "significant"!

Example for an ANOVA: There is strong evidence that at least one credit rating has a different mean income (p-value = 0.027, F-stat = 3.067 on an F(3, 1616) distribution).

- (3 pts) Interpret **in the context of the data** the p-value associated with `attr_cat` in your ANOVA table.

- (2 pts) What is your conclusion for the second research question? Namely, "Does the relationship between course evaluation scores and attractiveness differ based on ___?"

  - Base your conclusion on the visualizations you created **and** the two-way ANOVA model you fit to these data.
  - Your conclusion should include all of the components discussed above (test statistic, distribution it follows under the null hypothesis, α threshold)

- (3 pts) Interpret **in the context of the data** the p-value associated with the interaction term in your ANOVA table.

  - Keep in mind what the p-value is *conditional* on!

Scope of Inference (5 pts)

Take your original Scope of Inference from the Midterm Project and spice it up!

You can find the sampling methodology used for this study in the *Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity* article posted on Canvas. Specifically, section 2 "Measuring teaching productivity and its determinants" describes how professors were selected for this study.

At a bare minimum, your revisions should incorporate:

- the feedback I provided
- a description of the population of interest
- a description of the sample selected
- a justification as to who the results of the study can be inferred to
- a description of the relationships found in the analysis
- a justification as to whether causal statements can be made
- a conclusion about what type of relationships can be inferred

Study Limitations (5 pts)

In this new section you will address some of the limitations of the study, not addressed in the previous section.

This study used six undergraduate students to rate the attractiveness of the professor's. There were three women and three men, with one student of each gender being a lower division division student, and two being upper-division students.

- (2 pts) What limitations do you see from how attractiveness was scored for each professor?

Additionally, the study used faculty photos to classify each professor as "female" or "male," as well as whether each faculty was a "minority" or "not a minority."

- (1.5 pts) What limitations do you see from how gender was included in this study?

- (1.5 pts) What limitations do you see from how ethnicity was included in this study?

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 `broom()` function from the **broom** package and the `kable()` function from the **knitr** package to tidy up the output of the `anova()` function. The code should look like the following:

```
anova(model_name) %>%
  tidy() %>%
  kable()
```

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.