Lab 10: Student Evaluations of Teaching

These questions reference a paper published by researchers, some of whom were here in the Statistics Department at Cal. Anne Boring, Kellie Ottoboni and Philip B. Stark. *Student evaluations of teaching (mostly) do not measure teaching effectiveness.* Science Open Research. Vol. 0(0):1-11. DOI: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Part I

1. **General**: As you read through the article, note here any words or acronyms that you were not familiar with before and provide their definition.
2. **Abstract**: Which finding strikes you as the most important? Why?
3. **Abstract**: Based on the results summarized here, what do you believe was the overarching research question that the scientists were wondering about when they devised this study?
4. **Background**: Why is a student's answer to "How effective was the instructor?" not always helpful in understanding how effective the instructor was?
5. **Background**: What is the general statement of the null hypothesis that is applied to every analysis in this paper?
6. **Background**: What do the results of this study indicate about the relative impact of the teaching effectiveness and perceived instructor gender on SET?
7. **Data**: Why did the researchers in the US experiment have the TAs swap identities for one of the sections that each of them taught? Discuss using the principles of experimental design.
8. **Data**: Based on the description in the paper, sketch/speculate what the US experimental data frame might look like. Be sure to note the unit of observation, the number of rows and columns, the names of the variables, their data type, and the values they can take.
9. **Methods**: What is the test statistic that is used throughout the analysis?
10. **Methods**: Sketch/speculate what a plot could look like of the distribution of Prompt SET scores for each reported instructor gender. This should be a plot of the full data set and should be consistent with the statistics shown in table 8. Repeat the exercise for the Responsive SET scores and reported instructor gender.

Part II

The authors of this manuscript ensured that their analysis is fully reproducible by making their manuscript, data, and code easily available at `https://github.com/kellieotto/SET-and-Gender-Bias`. You can load the data from the US experiment with:

```
library(tidyverse)
set <- read_csv("https://www.dropbox.com/s/jog3lnqjinabe9s/set.csv?dl=1")
```

11. What is the unit of observation in the data frame? What are the dimensions of the data frame? How many students from each section filled out evaluations (at least partially)?

12. Calculate the difference in mean Prompt SET rating for between the two reported TA gender identities. Repeat this exercise for the difference in mean Responsive SET rating. This code should replicate two rows of the difference in means column of table 8.

13. Use the plots that you drew in question 11 as inspiration for constructing two plots of the actual data: the relationship between Prompt SET and reported TA gender, and the relationship between Responsive SET and reported TA gender. Describe each pair of plots: how does the rating for each differ based on the reported instructor gender?

14. Perform a test of the hypothesis that Prompt rating is independent of the reported TA gender id. Include in your analysis:

    a. A statement of the null and alternative hypothesis.
    b. The value of the observed test statistic.
    c. A plot of the distribution of the test statistic under the null hypothesis with your observed test statistic included as a vertical line.
    d. A calculation of the two-sided p-value.
    e. A conclusion regarding the null hypothesis using $\alpha = .05$, stated in the language of SET scores.

15. Perform a test of the hypothesis that Responsive rating is independent of the reported TA gender id. Include in your analysis:

    a. A statement of the null and alternative hypothesis.
    b. The value of the observed test statistic.
    c. A plot of the distribution of the test statistic under the null hypothesis with your observed test statistic included as a vertical line.
    d. A calculation of the two-sided p-value.
    e. A conclusion regarding the null hypothesis using $\alpha = .05$, stated in the language of SET scores.

16. How closely would you expect bias found in this particular experimental setting to transfer to, say, an in-person class at Cal? Would you expect gender bias to be stronger? Weaker? Why?

17. **Conclusion**: Do you find that the arguments in this section of the manuscript are consistent with the results of their data analysis? Do you find that they're consistent with your own experience with evaluating instructors?