Final Exam for MATH 354 – Data Analysis I
Submission and Formatting Instructions

First Author
Affiliation
Department
`email@domain`

Abstract

This document provides a basic template for the 2-page extended abstract and some submission guidelines.

Keywords: Optional, Yet Potentially, Helpful

1 Introduction

1.1 What we've done

On the midterm exam, you should have been able to show that wellbeing measures increased, and illbeing measures decreased. The effects were small, but several of them were significant. I don't know about you, but that's enough for me to want to play with some dogs.

We hadn't discussed regression techniques as of the midterm, so we could not do the analysis while "controlling" for other variables; e.g., gender, age, whether they have pets at home, etc. For the final analysis, we take another step forward!

1.2 Prompt

The prompt for this exam is based on my joint work with Erin Cooley (Colgate), Jazmin Brown-Iannuzzi (UVA), Ryan Lei (Haverford), and Lauren Philbrook (Colgate). Our most recent research explored how beliefs about White poverty play a role in forming White Americans' policy attitudes.

The data consists of observations on 837 non-Hispanic Black and White Americans who were not receiveing welfare. We measured their beliefs that White people are poor; their beliefs that Black people are poor; their perception that welfare recipients are hardworking; their general support for welfare policy; and information about the race, political affiliation, income, and education of participants.

The study results support the hypothesis that beliefs White people are poor predict beliefs that White people benefit from welfare. This belief mediates a greater tendency to humanize welfare recipients, with downstream consequences for greater support for agentic welfare policy (i.e., cash assistance). Further, we found that these processes may be motivated: White Americans high in zero-sum beliefs, the beliefs that one person's or group's loss is anothers gain, showed a stronger association between beliefs that White people benefit from welfare and the humanization of welfare recipients.

Use regression models and posthoc evaluations to evaluate the data from Study 1 (described below). Ensure that the statistical procedures you selected are appropriate for these data. You will note that this template is just two pages long, not including bibligrapy and apendix – this ensures that you can determine what is necessary for succinctly, but fully, answering this question.

1.3 Template

This template is based on the SDSS conference template, which is based the template for the annual conference of the Association for Computational Linguistics (ACL), whose template includes the following paragraph of credits:

This document has been adapted from the instructions for earlier ACL proceedings, including those for ACL-2012 by Maggie Li and Michael White, those from ACL-2010 by Jing-Shing Chang and Philipp Koehn, those for ACL-2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, those for ACL-2005 by Hwee Tou Ng and Kemal Oflazer, those for ACL-2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats. Those versions were written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence*.

The goal of this two-page template is to emphasize extracting what's essential.

In preparing your final analysis I strongly encourage using the same four section headings seen here, namely, Introduction, Methods, Data (or Results), and Discussion (or Conclusions). Use of subsections is at your discretion.

You submission *must* be two-pages long, with the option of a 1 page appendix showing your most important code chunks.

1.4 Citations

In LaTeX, please use the BibTeX bibliography style, as demonstrated by this example document. Review the paragraph below in the .RnW file; itshows how the LaTeX and BibTeX code work together to generate the citations.

Citations within the text appear in parentheses as (Nussbaum, 2018) or, if the author's name appears in the text itself, as Nussbaum (2018). Treat double authors as in Wasserstein and Lazar (2016). Collapse multiple citations as in

(Nussbaum, 2018; Wasserstein and Lazar, 2016). As alluded to above, full citations should not be used as sentence constituents. For instance, instead of

"Arbuthnot (1710) derived the earliest p-value ..."

use

"Arbuthnot (1710) derived the earliest p-value ...".

Please make references complete, presentable, and consistent. When using BibTeX, protect capital letters of names and abbreviations in titles, as in {B}ayesian or {EM} algorithm, in your .bib file. For an example, notice that the "FDA" in the title of LaVange (2019) is typed as `{FDA}` in the .bib file or else it would be incorrectly shown as "fda" in the references section.

The LaTeX and BibTeX style files provided roughly fit the American Psychological Association format, allowing regular citations, short citations and multiple citations as described above.

1.5 Graphics and Tables

Place figures and tables in the paper near where they are first discussed, rather than at the end.
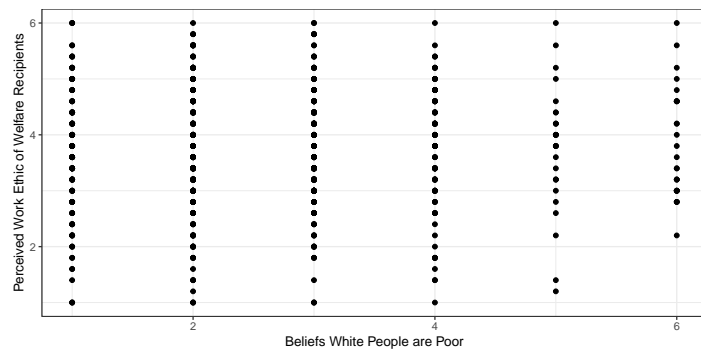

Figure 1: An example plot using the data.

2 Methods

You may use this prompt to help build your introduction and methods sections. You want to provide some essential information about what you're addressing and where the data came from.

3 Data/Results

3.1 Models

Fit the following models to help answer the questions of the researcher with our new-found tools.

3.1.1 Work Ethic

In Study 1 of the paper, we build a regression model for the perceived work ethic of welfare recipients (`HardworkingWelf`) based on beliefs White people are poor (`Wpoor`), race of the participant (`PsWhite`), and their interaction. Create and interpret this model in the context of the question. Ensure to control for education (`edu`), income (`income`), political affiliation (`Democrat`), and beliefs Black people are poor (`Bpoor`).

3.1.2 A Challenge

A main goal of this course is to give you a solid foundation. It's less important to memorize every detail of the course than it is to learn how to approach a data problem. I hope that you've learned enough to come across something new, and tackle it with the help of documentation and vignettes. As a challenge, you might want to consider mediation analysis, which quantifies the effect of one independent variable "through" another. See the `mediation` package for R (Tingley et al., 2014; Imai et al., 2010).

To demonstrate that you've learned enough to pick up new modeling techniques, try using mediation to explore Study 1. Specifically, does the race of the participant moderate the effect of beliefs White people are poor through the perceived work ethic of welfare recipients (the possible mediator)?

3.1.3 An Optional Bonus

Locate data documenting the percentage of people below the poverty line who are White by zip code and add that as a variable in the data set. Check if this variable (percentage of white people in poverty in the participant's location) predicts the degree of beliefs that White people are poor. In other words, are these race/class stereotypes shaped by sensitivity to the racial distribution of poverty in one's area?

4 Discussion/Conclusions

Discuss the results of your analyses here. This is where you tie all of your work together into a nice, succinct summary.

4.1 Submissions

This part of the final is "self-scheduled." I know y'all have projects and finals coming up. It is expected that the final will be submitted by the end of finals week, Friday, December 17 by 5p. I will not accept late submissions.

4.2 Short Answer Part

The short answer questions for the final will look similar to the short answer question on the midterm. While the data analysis aims to measure "what you can do," the short answer questions aim to measure "what you're walking around with." These questions will be available to complete as a quiz on Moodle between 3:00p and 5:00p on December 13th. This will take 35 minutes. Keep in mind these are meant to be questions about "big picture" understanding rather than small technicalities.

References

Arbuthnot, J. (1710). An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions of the Royal Society of London*, 27:186–190.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2010). Causal mediation analysis using r. In Vinod, H. D., editor, *Advances in Social Science Research Using R*. Springer-Verlag, New York.

LaVange, L. M. (2019). Statistics at the FDA: Reflections on the past six years. *Statistics in Biopharmaceutical Research*, 11(1):1–12.

Nussbaum, B. D. (2018). Statistics: Essential now more than ever. *Journal of the American Statistical Association*, 113(522):489–493.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5):1–38.

Wasserstein, R. L. and Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2):129–133.

5   Appendix

Below are examples of how to incorporate your code.

Loaded Data.

```
library(tidyverse,quietly = T)
suppressMessages(dat <- read_csv("final.csv", na="#NULL!")) # Study 1
```

Plot plot.

```
pdf("figure/workethic.pdf",width=8, height=4)   #save ggplot to a pdf
ggplot(data=dat,aes(x=Wpoor, y=HardworkingWelf))+
  geom_point()+
  theme_bw()+
  xlab("Beliefs White People are Poor")+
  ylab("Perceived Work Ethic of Welfare Recipients")
dev.off()
```