DTSC 2302 Spring 2021

Drs. Berardinelli and Windett

Project 2 Introduction

Generally speaking, your team will have three options for how to approach project 2:

1. You can revise Project 1. You will have to add some new components to meet the specifications for Project 2, but you can work with the same themes and some of the same research questions and analyses.
   This option is not recommended for teams that scored >80% on Project 1.
2. You can work with data similar in theme to Project 1, but at the state level instead of the individual level. More information about this option is below.
3. If your team is particularly inspired by some other topic and you are prepared to do the extra work of finding your own data, you may pursue a completely self-designed project. **This option requires a complete project proposal submitted via email** to both Dr. Berardinelli (aberardinelli@uncc.edu) and Dr. Windett (jwindett@uncc.edu) **before noon on Sunday, March 28, 2021**.
   This proposal must include the research question, the data source(s), the reason for interest, and how you plan to put in the time and effort to do the legwork required for a project of your own design (rather than using the scaffolding provided in the class assignments).

Project 2 Requirements

These requirements apply to all project options.

- Paper structure must include the following sections with these exact headings:
    - Introduction
    - Theoretical Framework
    - Data and Methods
    - Results
    - Conclusion
- Research question must be clearly stated in the introduction and clearly motivated. (Why should we care?)
- Theoretical framework must include literature review and theory, just as in project 1, but must also address the following questions:
    - Drawing from **Leonelli 2020**, in what ways did your research for this project respond to the following themes of how big data impacts scientific research? [You must refer to at least 3 of Leonelli's main themes in your paper.]

a. how statistics, formal and computational models help to extrapolate patterns from data, and with which consequences;
b. the role of critical scrutiny (human intelligence) in machine learning, and its relation to the intelligibility of research processes;
c. the nature of data as research components;
d. the relation between data and evidence, and the role of data as source of empirical insight;
e. the view of knowledge as theory-centric;
f. understandings of the relation between prediction and causality;
g. the separation of fact and value; and
h. the risks and ethics of data science.
- ○ **Pasquale 2016, Eubanks 2016, Benjamin 2019, Hu 2020, and Caplan and Boyd 2016:** Each of these authors addresses some dimension of how data-driven practices can lead to ethical harms, including the exacerbation of forms of economic inequality and the erosion of public trust in basic democractic processes. Cite and discuss at least two readings, and describe how the conclusions from your research addresses the forms of ethical harm described by the authors you chose.
- In choosing models for your analysis you must use at least one of the models covered in DTSC 2302 (linear regression, nonlinear regression, classification, decision). You may also use other methods of your choosing.
- In addition to addressing the questions about the data and variables that were stressed in previous projects, your Data and Methods section must also answer the following questions:
  - ○ What type of model(s) are you running?
  - ○ Justify the choice of your model(s).

Additional Background for Project Option #2

In January of 2020, the Pew Research Center released a report on economic inequality in the United States based on a large survey of Americans, as well as Current Population estimates from the census. The report concludes that Americans view economic inequality is structured on many personal characteristics of the respondent, such as their political party, race, sex, income status, and education level.

Economic inequality is not a new area of evaluation, but one where political ideology, social context, as well as individual beliefs will structure opinions, outcomes, and policy changes to make society more equal in terms of wealth and income distribution. Furthermore, there is a strong normative component of economic policymaking that discusses whether or not the government should intervene in the so-called free labor market to level the economic playing field.

In this project option, you will look at individual components that influence economic inequality at the state level within the United States. We will focus primarily on the application of economic data, political data, population data, as well as numerous surveys of US residents to evaluate responses to economic inequality, as well as their effectiveness. DTSC 2302 will have studio assignments aimed to familiarize you with measurement and statistics specific to each domain area. You will download, process and warehouse diverse sets of indicators that are either publicly available or provided by the instructors. You will then run a series of exploratory analyses, leading up to a multivariate regression.