

Lab 1: Hypothesis Testing

w203 Teaching Team

Overview

The American National Election Studies (ANES) conducts surveys of voters in the United States, with a flagship survey occurring immediately before and after each presidential election. While the post-election data for 2020 is not yet available, pre-election data is available as a preliminary release. In this lab, you will use the ANES data to answer questions about voters in the US.

This lab consists of 3 parts. Each part is centered around a research question. For each question, you will conduct a statistical analysis and generate a written report in pdf format. This means that you will create three separate reports, each one a complete analysis on its own (This is especially important since different graders may grade each of your responses).

This is an exercise in both statistics and professional communication. It is important that your techniques are properly executed, but equally important that your writing is clear, well defended, and organized.

Your instructor will divide you into teams to work on this lab. To maximize your learning, we ask that you do not use a divide-and-conquer approach to finish the lab. Instead, all students should participate in all parts as much as possible.

This one week lab is due before your Unit 9 live session. You will find a separate place on Gradescope to submit each of your three responses, along with the source file used to create your pdf.

Data

Data for the lab should be drawn from the 2020 American National Election Studies (ANES). You can access this data at <https://electionstudies.org>. This is the official site of the ANES, a project that has been ongoing since 1948, and federally funded by the National Science Foundation since 1977.

To access the data, you will need to register for an account, confirm this account, and then login. The data that you need should come from the **2020 Time Series Study**.

You will note that there are two forms of data that are available, data that is stored in a **.dta** format, and data that is stored in a **.sav** format. Both of these are proprietary data formats (**.dta** for STATA, and **.sav** for SPSS). You will need to find an appropriate library to read this data into R.

While you're at the ANES website, you will also want to download the codebook, because all of the variables are marked as something like, **V200002** – which isn't very descriptive without the codebook.

For a glimpse into some of the intricacies that go into the design of this study, take a look at the introduction to the codebook.

The Research Questions

The research question for each of the three parts of the lab are as follows:

Part 1: Are Democratic voters older or younger than Republican voters in 2020?

Part 2: Are Democratic voters more enthusiastic about Joe Biden or Kamala Harris?

Part 3: Are people who have had Covid-19 infect someone in their home more likely to disapprove of the way their governor is handling the pandemic?

Report Guidelines

- Each report must stand alone as a separate document: a separate pdf with a separate source Rmd file.
- We have created a template Rmd file for each question, including prompts to guide you through the parts of an analysis. Make sure you fill in each prompt with all information requested.
- Page limit for each report: 3 pages in standard latex formatting.
- Each question should include either a plot or table that is informative about the distribution of the data, and this plot or table should be discussed in the text.
- You're going to have to clean, tidy, and structure the data for the questions that you're using; this code should be included in your report .Rmd file, but should not be shown in your final report. To accomplish this, you can use an `echo=FALSE` argument in your .Rmd file.
- You're going to have to write code to produce tables and figures. This code should be included in your report .Rmd file, but should not be shown in your final report. To accomplish this, you can use an `echo=FALSE` argument in your .Rmd file.
- You're going to have to execute tests, for example, using `t.test()`. This code *should* be shown in your report, because it makes very clear the specific test that you're conducting.
- While you can choose to display the results of your test in the report, you also *certainly* need to write about these results. This should be accomplished using inline code chunks, rather than by hard-coding / hard-writing output into your written report.