Statistical test homework

Saldivia Gonzatti & Hutter

2024-10-19

Session 1 homework

In order to assess how well you can implement basic statistical analysis in R, we ask you to perform the following tasks. It is important to note that this is not a formal test in the sense that you either passed or you do not pass. Instead, we will provide feedback on your code and discuss with you how your skills might fit with the course or what you can and should do otherwise during the course to improve your skills.

You will work with Eurobarometer survey data in this exemplary exercises. This is not the kind of data we will use during the course, since we will rather focus on protest event data. However, how to approach statistical analysis with different kinds of data is very similar.

The data consist of harmonized Eurobarometer data from 2020 and 2021 covering the countries France and Germany. The data includes 5 variables: country name (country_cl), the year of the survey (year), the age of the respondent (age), trust in parliament (trust_parliament), and trust in government (trust_government). The trust variables are binary, indicating that the individual respondent has no trust (0) or has trust (1) in the respective institution.

1. Load the data from the file `data/ebm_selection.RData` in the Github repository. You can also read the data from the CSV file.
2. First, via R code, extract the mean trust in parliament and trust in government for each country and year.
3. Perform a t-test to compare whether the trust in government is larger in statistically significant terms than the one in parliament in Germany in 2021.
4. Show the distribution of trust in parliament in France in 2020 using the R package `ggplot2`.
5. Last, assume we want to study the relationship between age of the respondents and trust in government in a regression framework. Implement a linear OLS regression model including country fixed effects. Ideally, you would implement a logit model since the dependent variable trust is binary, but for the sake of simplicity, we will stick to OLS.

Please provide the R code for each task by Monday, October 28. Send the R script to Daniel via e-mail.