PLSC 502: "Statistical Methods for Political Research"

Exercise One
September 16, 2024

Data

This first exercise is all about graphing / visualizing data, and interpreting those plots. The data in question – which are available in the "Exercises" folder on the Github repository in `.csv` format – consist of 14 variables taken from the December 19, 1998 House of Representatives votes on the impeachment of President William Clinton (so, N = 435, give or take some missing data). The variables in the data include:

- `Name` – The member of Congress's name.
- `State` – The two-letter postal code for the state from which the member hails.
- `District` – The member's district number within their state.
- `Vote1` – The member's vote on the first article of impeachment (0 = "No," 1 = "Yes").
- `Vote2` – The member's vote on the second article of impeachment (0 = "No," 1 = "Yes").
- `Vote3` – The member's vote on the third article of impeachment (0 = "No," 1 = "Yes").
- `Vote4` – The member's vote on the fourth article of impeachment (0 = "No," 1 = "Yes").
- `VoteSum` – The sum of the member's four impeachment votes (from zero to four).
- `GOP` – Whether (= 1) or not (= 0) the member was a Republican.
- `ADA98` – The member's 1998 Americans for Democratic Action (ADA) voting score; this ranges from zero to 100, with higher values indicating a more liberal voting record.
- `CC98` – The member's 1998 Christian Coalition voting score; this ranges from zero to 100, with higher values indicating more conservative voting.
- `ClintonVote96` – The percentage of the two-party vote received by President Clinton in the 1996 general election in that member's district.
- `UnionPct` – The percentage of all workers in the member's district that were members of organized labor.
- `MemberVote98` – The percentage of the two-party vote received by the member him- or herself in the 1998 Congressional general election.

Exercise

For each of the above items, write up your response using whatever combination of prose, tables, and figures you feel is appropriate. ***Do not use regression analysis or some other multivariate method***; instead, answer all questions as completely as possible using graphical displays of data *only*.

1. Begin by picking <u>two</u> of the following variables: `ADA98`, `CC98`, `Clinton96`, `unionpct`, or `Member98`.
   (a) Plot and describe the distribution of the variables you chose.
   (b) Are (and to what extent are) the variables you chose normally distributed? Again, use plots to answer the question.
2. Next, pick <u>one</u> of the four `vote` variables, and plot and describe its distribution, as well as those of the `GOP` and `votesum` variables.
3. Answer the following three questions, remembering the old adage that "a picture is worth a thousand words":
   (a) What is the relationship between district union membership and the percentage of the vote received by President Clinton in 1996?
   (b) How would you characterize the relationship between members' ideology (measured through their `ADA98` and/or `CC98` scores) and their propensity to vote for or against each of the articles of impeachment?
   (c) What is the relationship between Clinton's *1996* vote share and each member's *1998* general election vote shares? How is that relationship moderated by the party identification of the member (that is, is the relationship different for Democrats than for Republicans)?

Further Instructions

1. For each figure, follow the guidelines for making "good" figures (axis labels, etc.).
2. Submit your answers **in PDF format**. For each answer, provide both the figure(s) and a short textual description of your "answer."
3. In addition to your answers, please include a copy of all computer code used to generate your figures. This can be in any form – a separate `.R` or `.do` file, an appendix in the PDF, or as a `.Rmd` or similar format containing both content and code.
4. Submit your materials electronically – via e-mail attachment – to Morrgan (`mth5492@psu.edu`) *and* to me (`zorn@psu.edu`).
5. This exercise is due at or before 11:59 p.m. ET on Wednesday, September 25, 2024, and is worth 50 possible points.