PLSC 502: "Statistical Methods for Political Research"

Exercise Seven
November 11, 2024

Part I

In this part, you're asked to assess the performance of three different measures of association (ϕ, r_{tet}, and τ_b) in recovering the correlation between two continuous variables that have been dichotomized. Consider two latent (unmeasured) variables X_1^* and X_2^*; for simplicity, we'll make each of them simple:

$$X_1^*, X_2^* \sim \mathcal{BVN}(0, 0, 1, 1, \rho).$$

In other words, X_1 and X_2 are both standard normal variables, where the correlation between X_1^* and X_2^* is ρ.[1] For each latent variable, suppose we observe a binary realization X_1 and X_2, such that:

$$X_{1,2} = \begin{cases} 0 \text{ if } X_{1,2}^* \leq 0 \\ 1 \text{ if } X_{1,2}^* > 0 \end{cases}$$

The question is, given binary realizations of the X^*s, which of the three measures of association listed (ϕ, r_{tet}, and τ_b) does "best" at recovering the true underlying correlation ρ? Use simulations to answer this question, being sure to vary ρ widely in formulating your answer.

Part II

This part of the exercise is designed to showcase your mastery of one- and two-way contingency tables, and of measures of association for nominal, binary, ordinal, and continuous data. We'll use data from the American National Election Study's 2024 Pilot Study, a national probability sample of around 1600 Americans fielded in late February 2024. Details of the study are available at the link above, and at the codebook on the course Github repo. We'll use a reduced subset of those data, containing the following 16 variables:

- `ID`: a respondent identifier code; can be safely ignored.

- `Employed`: whether (= 1) or not (= 0) the respondent was employed at the time of the survey.

- `CampaignInterest`: A three category ordinal variable, indicating whether the respondent was "not much" (= 1), "somewhat' (= 2), or "very much" (= 3) interested in the 2024 general election campaign.

- `VoteImportance`: "Thinking now about the people you care about, like your friends and family, how important do they believe it is to vote in elections?"

  1. Not at all important
  2. Slightly important
  3. Moderately important

---

[1]Such continuous correlated data are easy to generate using the `mvrnorm` command in the `mvtnorm` package; see the class code from November 11 for examples.

4. Very important

5. Extremely important

- `ApproveProtest`: "Generally do your friends and family approve or disapprove of participating in public protests?"

    1. Disapprove very strongly

    2. Disapprove moderately

    3. Disapprove a little

    4. Neither approve nor disapprove

    5. Approve a little

    6. Approve moderately

    7. Approve very strongly

- `UseFacebook`: whether (= 1) or not (= 0) the respondent had visited Facebook during the previous year.

- `UseTwitter`: whether (= 1) or not (= 0) the respondent had visited Twitter during the previous year.

- `UseTikTok`: whether (= 1) or not (= 0) the respondent had visited TikTok during the previous year.

- `DemocraticTherm`: The 0 - 100 "feeling thermometer" score that the respondent gave the Democratic Party.

- `RepublicanTherm`: The 0 - 100 "feeling thermometer" score that the respondent gave the Republican Party.

- `Ideology`: Respondent's self-placement on a seven-point ideology scale, defined as:

    1. Extremely liberal

    2. Liberal

    3. Slightly liberal

    4. Moderate; middle of the road

    5. Slightly conservative

    6. Conservative

    7. Extremely conservative

- `PartyID`: Respondents' political party self-identification, coded 1 = *Democrat*, 2 = *Independent*, 3 = *Republican*, 4 = *Other / Not Sure*.

- `Age`: Respondent's age in years.

- `Gender`: Respondent's self-identified gender; one of *Female* or *Male*.

- `Race`: Respondents' self-identified race; one of *Black*, *Hispanic*, *Other*, or *White*.

- `Region`: one of *Midwest*, *Northeast*, *South*, or *West*.

Exercise

Answer the following twelve questions, using the data provided and the statistical tools we discussed in class:

1. What if any is the association between respondents' employment status and their interest in the campaign?

2. To what extent do respondents' friends who think voting is important also tend to support protest activity?

3. Is it the case that old people tend to use Facebook and younger people use TikTok?

4. What is the association between age and Twitter use?

5. Do Democratic respondents feel more warmly toward the Democratic party than Republican respondents feel toward the Republican party?

6. Do Republican respondents feel more coldly toward the Democratic party than Democratic respondents feel toward the Republican party?

7. How do respondents from different regions differ with respect to their ideological self-identification?

8. To what extent are women more liberal than men?

9. Which of the three social media platforms represented here (Facebook, Twitter, TikTok) are middle-aged people (defined as those between 50-65) most likely to use?

10. Are there important regional differences in the race of respondents in this sample?

11. Do unemployed people use more or less social media than the employed?

12. If you treat party identification as ordinal (with Democrats at one extreme, Republicans at the other, and independents / others in the middle), how does it correlate with respondents' age?

As is typically the case, this homework exercise is worth 50 possible points. It is due by 11:59 p.m. EST on Wednesday, November 20, 2024, and should be submitted electronically – via e-mail attachment – to Morrgan (mth5492@psu.edu) *and* to me (zorn@psu.edu).