

PLSC 502: “Statistical Methods for Political Research”

Exercise Seven

November 15, 2022

Part I

Here, 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, and ordinal data. We’ll consider data on the 2002 Swedish election to the national parliament.² Those data consist of roughly 7000 respondents³ to a series of exit polls there following that election. The data contain a range of variables; we’ll focus on eight of them:

- `gender` – a poorly-named variable, coded “0” for men, “1” for women.
- `citizen` - coded “0” for non-citizens, “1” for citizens.
- `employed` – a four-category nominal variable reflecting employment status: “1” = employed, “2” = unemployed, “3” = student/in training, and “4” = retired.

¹Such continuous correlated data are easy to generate using the `mvtnorm` package; see the code from November 10 for examples.

²Why the 2002 Swedish election? Well... why not?

³There are actually more than 10000 respondents, but there is significant missing data in a few of the variables.

- `union` – a four-category nominal variable, coding whether the respondent was a member of one of three types of unions (1,2, or 3) or not (4).
- `uniondummy` – a dichotomous version of `union`, coded “1” if the respondent is a member of *any* union, and “0” otherwise.
- `eu` – another poorly-named variable, variable coded “0” if the respondent thought that Sweden should resign from the EU, and “1” if s/he believed Sweden should stay in the EU.
- `partychoice` – a four-category ordinal variable, coded “1” for conservative parties, “2” for liberals, “3” for social democrats, and “4” for parties of the left.
- `euro` – coded “1” if the respondent favored Sweden’s adopting of the Euro as its currency and “0” if s/he did not.

Exercise

1. Assess whether there are any differences in party identification, employment status, union membership, and opinions toward the Euro and EU membership between men and women.
2. Are there systematic differences between members of different political parties in views on Euro adoption? on EU membership?
3. What, if any, is the relationship between support for EU membership and opinion on the Euro? Is that relationship different for men than it is for women? for citizens than for non-citizens? for union members versus non-union members?

As is typically the case, this homework exercise is worth 50 possible points. It is due by 5:00 p.m. EST on Wednesday, November 30, 2022, and should be submitted electronically – via e-mail attachment – to Tuba (tzs5636@psu.edu) *and* to me (zorn@psu.edu).