PLSC 502: “Statistical Methods for Political Research”

Exercise Six
October 30, 2023

Part I

We’ll again start with a simulation-based exercise, this time with a specific purpose. Conventional t-tests are designed for use when the outcome Y comprises *independent* draws from a *Normal* distribution. As we discussed in previous classes, binary variables (where $Y \in \{0,1\}$) are decidedly not Normal; while we still calculate sample arithmetic means in the usual way, binary responses yield means which are proportions, and proportions behave differently than means of unbounded / Normal variables.

Specifically (and as noted in Section 7.6 of Agresti, among other places), for two sample proportions $\hat{\pi}_0$ and $\hat{\pi}_1$ calculated from n_0 and n_1 sampled observations, respectively, the difference in proportions $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$ has sampling variability equal to:

$$\text{Var}(\hat{\delta}) = \sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}$$

where $\hat{\pi}$ is the grand mean $\sum Y_i/N$; this means that the correct statistic for testing $H_0 : \hat{\delta} = 0$ for a difference of proportions is:

$$Z = \frac{(\hat{\pi}_1 - \hat{\pi}_0) - 0}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}$$

which has $Z \sim N(0,1)$, for easy hypothesis testing.

Despite this being a well-known, well-understood, easily-implemented method for two-sample comparisons of proportions, scholars nonetheless often use the conventional Welch t-test to examine differences of means in binary variables. Your assignment here is to explore the implications of this (mal)practice: **What, if any, are the effects of using a standard t-test – rather than a difference of proportions test – on data where Y is binary?**

In completing this part of the exercise, be sure to assess and discuss how the answer to the question above varies by (a) the overall sample size N, (b) the size of the difference $\hat{\delta} = \hat{\pi}_1 - \hat{\pi}_0$, and (c) the degree of skewness in the Ys.[1] Some useful hints:

- The R commands for doing difference-of-proportions and t-tests are `prop.test` and `t.test`, respectively;
- Use two-sample tests in both cases, for comparability;
- One general place to start is to ask “If I have two samples from which I’ve calculated $\hat{\pi}_0$ and $\hat{\pi}_1$, and I do both a t-test and a difference-of-proportions test on those data, how are they different?”

[1] So, for example, when assessing a difference of proportions of size $\hat{\delta} = 0.2$, you should consider both the case where $\hat{\pi}_0 = 0.05$ and $\hat{\pi}_1 = 0.25$ as well as that where $\hat{\pi}_0 = 0.4$ and $\hat{\pi}_1 = 0.6$.

Part II

Part II is about power and sample size analysis. It is, I think, straightforward: Answer each of the questions below, and provide the code and output you used to reach those answers, along with anything else you feel would be appropriate and/or helpful (plots, tables, memes, etc.).

1. For $N = 160$, $P = 0.01$ (two-tailed), and $\mathfrak{P} = 0.90$, plot the relationship between sample size and power for a two-sample t-test.

2. A recent poll of a simple random sample of 380 Penn State - University Park undergraduates finds that the mean "feeling thermometer" score toward President Bendapudi is 78, with a standard deviation of 18. You believe that female students will have a higher mean level of support for the President than male students. If (improbably) the sample is equally divided among men and women, and given $P = 0.05$ (one-tailed) and $\mathfrak{P} = 0.80$, what is the smallest effect size d you can expect to uncover? How does that effect size translate into the scale of the feeling thermometer? That is, what is the smallest feeling thermometer difference you can expect to find in this situation?

3. An existing public opinion survey comprises a simple random sample of $N = 760$ Pennsylvanians. Among other things, it asks five-point Likert-type approval/disapproval questions (1 = strongly disapprove, 2 = disapprove, 3 = neutral, 4 = approve, 5 = strongly approve) about the two major-party candidates for state supreme court, Democrat Daniel McCaffery and Republican Carolyn Carluccio. McCaffrey has a mean approval of 3.4 ($\hat{\sigma} = 0.9$) and had 74 respondents answer "Don't Know;" Carluccio's mean approval is 3.1 ($\hat{\sigma} = 1.1$) but had 158 respondents respond with "Don't Know." If the total sample (and missingness) is divided evenly between Republicans and Democrats, and assuming $P = 0.05$ (one-tailed), with what level of power can you expect to find a one-point difference between Republicans and Democrats in their support for each candidate?

4. You wish to select a random sample of U.S. Supreme Court cases from the Rehnquist Court era (OT1985-OT2004, $N_{\text{votes}} = 22401$) in order to laboriously hand-code petitioners' and respondent's briefs for specific content. You believe that the presence of certain language is likely to make the justices more likely to cast a conservative vote ($Y = 1$) rather than a liberal one ($Y = 0$), and that the presence of such text will increase that probability by at least 0.2, from a baseline liberal vote probability of 0.472. You're willing to set $P = 0.05$ (one-tailed), and you would prefer to to achieve a power level of 90 percent ($\mathfrak{P} = 0.90$) but might be willing to go as low as 80 percent ($\mathfrak{P} = 0.80$). Assuming there are nine votes in each case, how many cases should you code to ensure power levels of 0.80 and 0.90, respectively?

5. In the preceding question, the presence of nine votes per case allows you to code fewer cases than you would if (say) only one justice voted in each case. What are the potential concerns about this approach, what are the potential effects of those concerns, and how might you overcome those concerns?

As usual, use plots, words, or combinations thereof to complete this exercise. Submit your answers **in PDF format**. In addition to your answers, please include a copy of all computer code used to conduct your simulations, generate your figures, etc. 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. This homework exercise is due by 11:59 p.m. ET on Tuesday, November 7, 2023; submit your materials in electronic format – via e-mail attachment – to Nathan (`man@psu.edu`) *and* to me (`zorn@psu.edu`). This exercise is worth 50 possible points.