PHP2510

CATEGORICAL COMPARISONS WEEK 11

AN INTERVENTION FOR OPIOID USE DISORDER

This worksheet builds on our study of statistical inference, moving from continuous to categorical outcomes.

SCENARIO

You are an epidemiologist working for Rhode Island's Department of Public Health. They recently completed a pilot study on a new, intensive case-management program for individuals with Opioid Use Disorder who are starting buprenorphine treatment.

The study recruited participants from three different types of clinical settings and tracked their treatment retention at 6 months, a critical outcome measure.

Clinical Setting:

- "Specialty": Dedicated substance use treatment centers.
- "Integrated": Primary care clinics that have integrated behavioral health services.
- "PCP-Only": Standard primary care clinics (controls) that offer buprenorphine but without specialized co-located support.

Outcome:

- "Retained": Still in treatment at 6 months.
- "Relapsed/Lost": Disengaged from treatment before 6 months (includes relapse or being lost to follow-up).
- "Transferred": Formally transferred to a different treatment facility (e.g., inpatient) and thus censored from the analysis.

The director wants to know: "Is there an association between the clinical setting and a patient's 6-month retention status?"

DATA

You are given the following data:

Clinical Setting	Retained	Relapsed/Lost	Transferred
Specialty	31	20	4
Integrated	40	18	5
PCP-Only	16	29	2

ANALYSIS PLAN

The Standard Chi-sq Test

- State the hypotheses (null and alternative)
- Calculate all the expected counts – do it "by hand" or with your own R code
- Check your assumptions; specifically the "rule-of-thumb" for this test
- Perform the test in R via chisq.test(); report the test statistic, df, and p-value.
- Check that your calculations for the observed count are correct (they are stored, if you save the chisq.test() object)

The Permutation Alternative

- Explain why this alternative is worthwhile in this context
- Describe the algorithm (i.e. in pseudocode)
    - What is your test statistic?
    - What data do you shuffle/permute to generate the null distribution? Note: you need to "format" the data differently than it is given to you. Describe this.
    - How do you calculate the p-value?
- Run the test. Note that you can just use chisq.test() with simulate.p.value = T
- Is your conclusion any different from above?

ONE STEP FURTHER

If we do indeed detect an association, can we explain why? Yes! Each cell in our contingency table contributes its own information into the test statistic (inspect the form of the test statistic and convince yourself of this fact!); and large values of the test statistic are evidence against the null hypothesis. Therefore, the specific cells that contribute more are giving us more information to reject the null. These deviations from the null are stored in a chisq.test() object as $stdres. You can view the standardized residual functions as a "Z-score" for that cell (being non-0):

- A large positive residual (e.g., > 2) means we observed significantly more cases in that cell than expected by chance
- A large negative residual (e.g., < -2) means we observed significantly fewer cases in that cell than expected by chance
- Residuals close to 0 indicate that the observed count is very close to what we would expect under the null

Take a look at your standardized residuals.

- What cells are driving the effect?
- Can you interpret these to provide more insights to your director?
- Given these findings, do you have any intuition as to why we observed the results we did, in terms of consistency between the standard and permutation approach?

HELPFUL CODE

```
observed_counts <- matrix(c(31, 20, 4,40, 18, 5, 16, 29, 2), nrow = 3, byrow = TRUE)
dimnames(observed_counts) <- list(
  Setting = c("Specialty", "Integrated", "PCP-Only"),
  Outcome = c("Retained", "Relapsed/Lost", "Transferred")
)
```