

Math 150 - Methods in Biostatistics - Homework 10

your name here

Monday, April 24, 2023

Assignment Summary (Goals)

- understanding PPV
- type I and type II errors
- adjusting p-values to control for multiple comparisons

Q1. Collaborative Learning Describe one thing you learned from someone (a fellow student or mentor) in our class this week (it could be: content, logistical help, background material, R information, etc.) 1-3 sentences.

Q2. Positive Predictive Value Using the Ioannidis paper, explain the details of PPV for the model with multiple researchers. That is, derive the entire PPV equation (you may need to derive most of Table 3).

Q3. Evaluating errors part 1 Consider the following claim (which may or may not be true):

If a null hypothesis is NOT rejected in multiple studies, then we have good evidence that the null is likely to be true.

Five datasets are created to study the same phenomenon. For each dataset there is a treatment group and a control group. Do the two groups differ, on average, with respect to the continuous response variable? There are pairs of columns representing each of the 5 studies. Your task is to:

- Ascertain whether the response variable is different across the treatment and control. Look at p-values and confidence intervals. Consider the datasets separately, and consider the datasets combined into one analysis.
- Respond to the claim above.

You might consider R code like the following:

```
errordata <- read_delim("https://pomona.box.com/shared/static/acsduo30e4yzrc05mi2nnr56vj8qn3pb",
                      delim= "\t")
t.test(resp1~group1, data=errordata) %>% tidy()

response <- c(errordata$resp1, errordata$resp2, errordata$resp3, errordata$resp4, errordata$resp5)
group <- c(errordata$group1, errordata$group2, errordata$group3, errordata$group4, errordata$group5)
t.test(response~group) %>% tidy()
```

Solution:

Q4. Evaluating errors part 2 Consider a large randomized controlled trial designed to investigate problem drinking in Australian university students (Kypri et al., *Randomized controlled trial of proactive web-based alcohol screening and brief intervention for university students.*, 2009). They specified 7 outcomes in advance, 3 were primary and 4 were secondary. No adjustments for multiple comparisons were made, and the p-values were reported to be 0.001, 0.02, 0.001 (primary endpoints), 0.59, 0.87, 0.22, 0.001 (secondary endpoints).

- a. Adjust the p-values using Bonferroni, Holm, and Benjamini-Hochberg. Do all 3 methods give the same conclusions with respect to significance? Explain.
- b. Note that the Bonferroni and Holm adjusted p-values report the smallest familywise error under which each of the tests would reject the null hypothesis. Benjamini-Hochberg report the experiment wide FDR if all tests below a critical value are rejected. Explain why (within each of the methods) some of the adjusted p-values are repeated.
- c. Explain how adding 100 more tests (all of whose unadjusted p-values are greater than 0.1) would change each of the adjusted p-values (and corresponding conclusions).

```
praise()
```

```
## [1] "You are tiptop!"
```