Stat 139: section 7 handout

Jameson Quinn

Multiple comparisons

For each of the following situations, would a multiple comparison procedure be appropriate? If so, which one? Are there any other special considerations for the analysis?

As discussed in the book, a study is conducted to see if 4 various (easily-discernable) handicaps affect third-party assessments of qualifications for a non-physical job.

Subjects randomly drawn from two different populations are given a battery of 5 tests, and the question is whether scores on any of the tests differ between the two populations.

A study is investigating the relationship between household income and levels of stress hormones. A significant relationship is found, but investigators notice that the greatest differences are between employed and unemployed subjects. They decide to do two further comparisons: one which compares stress hormone levels using employment as a binary variable, and another which takes only the employed subjects and regresses hormone levels against (log of) income.

The number of parasites in 4 different species of oak trees are compared, using samples of 5-10 trees of each species.

As above, but there is a post-hoc decision to compare the two “riparian” to the two “non-riparian” species.

Brief R package-installation clinic

One student on Piazza has suggested that there’s an issue installing the asbio package in R-studio. We’re going to make sure everyone in the section can install it.

Hypothesis statements for permutation and randomization

Say I am comparing the effectiveness of two headache drugs, A and B. I ask my subjects to rate the intensity of their headache on a 1-10 scale, one hour after taking the drug. For drug A, I get the responses 2, 5, and 6; for drug B, I get 1, 3, and 4.

I think that most of you would have an easy time deciding between a permutation test (for observational data) and a randomization test (for randomly-assigned treatments); and actually carrying out that test to get a p-value. However, I think there’s confusion about stating the null hypothesis in cases like this.

For the randomization test, we’re dealing with causal inference, so as always, it’s helpful to think in terms of potential outcomes: each subject has a potential headache-level for each drug, but only one of those two levels is observed. If the two drugs were in fact exactly equivealent, those two levels would be the same; that’s Fisher’s sharp null. To state that succinctly with symbols, we can assume that there is a constant treatment effect δ, and then say “$H_0 : \delta = 0$”. Once we’ve done that, the alternative is obvious: “$H_A : \delta \neq 0$”.

For the permutation test, we’re dealing with populations. In that case, the null hypothesis is that the distribution of post-treatment headache strength is independent of drug taken: “$H_0 : Y \perp\!\!\!\perp X$”. The alternative hypothesis, then, is just the negation of that. If you want a one-sided alternative, it is difficult to state it succinctly in symbols, though you could do so by using the CDF of the headache strength distribution for the two groups.