Homework #1
Stat 212A, Fall 2015: Topics in Selective Inference
Instructor: Will Fithian
Assigned Sep. 17, 2015. Due 11:59pm Oct. 6, 2015

You are welcome to work with each other or consult articles or textbooks online, but you should then go away and write up the problem by yourself. If you collaborate or use other resources, please list your collaborators and cite the resources you used. Please show your work and include code where appropriate.

You can turn in the problem set in class Oct. 6 or under my door (Evans 301) Tuesday night.

1. Derived Intervals for $\overline{\mu}$ You are working with a scientist, who has one-way layout data:

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, \ldots, m, \quad \text{with} \quad \varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 1),$$

The scientist asks you to come up with FWER-controlling confidence intervals for $\mu_1, \ldots, \mu_m$. You construct

$$C_i = y_i \pm z_{\tilde{\alpha}_m/2}, \quad \text{where} \quad \tilde{\alpha}_m = 1 - (1 - \alpha)^{1/m}. \tag{1}$$

(a) After the scientist sees the results, she notices an interesting fact: even though only a few of the intervals exclude 0, most of the y_i are larger than zero. This makes her curious about the value of the parameter

$$\overline{\mu} = \frac{1}{m} \sum_{i=1}^{m} \mu_i,$$

and she expresses regret that she didn't think of asking about it before seeing the data. "Aha!" you exclaim, "but we *can* use the intervals we just constructed to derive an interval for $\overline{\mu}$."

- (i) Give an explicit expression akin to (1) for the interval you report.
- (ii) What is the approximate asymptotic radius of this interval as $m \to \infty$?[1]
- (iii) What is its radius for $\alpha = 0.05$ and $m = 3$, 5, 10, and 100? (Give numbers e.g. 3.45).

(b) Your collaborator explains in her paper that she got interested in $\overline{\mu}$ only after looking at the Šidák intervals. One of the referees cries foul: he claims that she is guilty of data dredging and she should remove the interval for $\overline{\mu}$ from the paper. Do you agree that the finding is not properly adjusted for multiplicity? Why or why not?

(c) Same question as (a), except suppose that instead of Šidák intervals for all μ_i, the scientist asked you initially to construct Scheffé intervals for all linear contrasts. Then, as in (a), she gets interested after the experiment in $\overline{\mu}$ and wants an interval for it.

- (i) Give an explicit expression akin to (1) for the interval you report.
- (ii) What is the approximate asymptotic radius of this interval as $m \to \infty$?
- (iii) What is its radius for $\alpha = 0.05$ and $m = 3$, 5, 10, and 100?

(d) If you were to redo the analysis for a new data set, knowing ahead of time that the scientist is going to be interested in $\overline{\mu}$ as well as the univariate means μ_i, how could you devise a more powerful FWER-controlling procedure than the one you used here? (**Note:** There could be more than one right answer).

Expanded Note: This is intentionally open-ended and there really are many right answers; I want you to think about how to cook up a method that behaves "well" or "sensibly," just as you would in a real collaboration. Here's one "nice" (but not necessarily mandatory) property: for $i \geq 1$, your intervals are asymptotically no wider than the Šidák intervals (i.e. the ratio of the two lengths goes to 1), but the length of your C_0 tends to 0 as $m \to \infty$.

[1] Remember, the radius of the interval is half its width; e.g. the width of the Šidák interval is $2z_{\tilde{\alpha}/2}$ while the radius is $z_{\tilde{\alpha}/2}$.

(i) Explain how you would construct the intervals C_i for μ_i and C_0 for $\overline{\mu}$, and give a relatively explicit expression for their lengths (e.g. in terms of a quantile of a random variable you can simulate).

(ii) What is the approximate asymptotic radius of this interval as $m \to \infty$?

(iii) What is its radius for $\alpha = 0.05$ and $m = 3$, 5, 10, and 100?

(e) Same question as (a), except suppose that instead of Šidák intervals for all μ_i, the scientist asked you initially to construct Tukey's HSD intervals for all pairwise comparisons. Show that the derived confidence interval for $\overline{\mu}$ has infinite length.

2. PoSI vs. Scheffé For regression with p variables and n observations, with known $\sigma^2 = 1$, prove that the PoSI interval radius r_α is always strictly smaller than $\chi_p(\alpha)$ (which is roughly $\sqrt{p}$).

Note / Hint: Please be sure to show that $r_\alpha < \chi_p(\alpha)$, where p is the number of variables to choose from — not just that $r_\alpha < \chi_n(\alpha)$, where n is the dimension of μ and ε. You'll need to use some fact about the set of all PoSI contrasts.

3. Closing ANOVA We saw that closing the Simes and Bonferroni procedures resulted in pretty good FWER-controlling multiple-testing procedures (Hochberg's and Holm's procedures, respectively). A natural question to ask is, what if we closed the χ^2 test of the intersection null?

It turns out this is a pretty bad idea! Assume we have the scenario in class where

$$\mu_1 = \cdots = \mu_{k_m} = \rho_m, \quad \mu_{k_m+1} = \cdots = \mu_m = 0,$$

and $y_i = \mu_i + \varepsilon_i$ for $i = 1, \ldots, m$, with $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, I)$.

(a) Show that even with $O(m)$ non-nulls (quite dense), we need ρ_m to be on the order of $\sqrt{m}$ to get any rejections. More precisely, assume that $k_m = m/2$, and that $\rho_m = o(\sqrt{m})$. Show that $\mathbb{P}(\text{any rejections}) \to 0$ as $m \to \infty$.

Hint: Recall that we are closing the intersection-null test

$$\phi_I(y) = 1\left\{\sum_{i\in I} y_i^2 > \chi^2_{|I|}(\alpha)\right\},$$

and $\chi^2_d(\alpha)/d = d\left(1 + O(\sqrt{d})\right)$ as $d \to \infty$.

Now, suppose there is some "bad set" I_{bad} for which $S_{\text{bad}} = \sum_{i\in I_{\text{bad}}} y_i^2$ is substantially less than $|I_{\text{bad}}|$. If such a bad set exists, we won't be able to reject $\{j\} \cup I_{\text{bad}}$ unless y_j^2 is big enough to make up the deficit between $|I_{\text{bad}}$ and S_{bad}.

(b) Show that if instead we used Bonferroni's procedure with $k_m = m/2$ and $\rho_m \geq \delta\sqrt{2\log m}$ for any constant $\delta > 0$, then $\mathbb{P}(\text{any rejections}) \to 1$ as $m \to \infty$.

4. Testing Hypotheses in Fixed Order Suppose someone gives us an *a priori* ordering on hypotheses $H_{0,1}, \ldots, H_{0,m}$ with p-values $p_1, \ldots p_m$ (i.e. the order is specified in advance of looking at the data). We then use the following procedure: If $p_1 \geq \alpha$, stop and accept all null hypotheses. Otherwise, reject $H_{0,1}$ and keep going. Then, if $p_2 \geq \alpha$, stop and accept $H_{0,2}$ through $H_{0,m}$. Otherwise, reject $H_{0,2}$ and keep going, etc. In other words, if k is the index of the first p-value that is larger than α, we reject $H_{0,i}$ for each $i < k$ and accept the rest.

(a) Prove that this procedure controls the FWER, regardless of the dependence structure of the p-values.

(b) **Challenge** (Optional) Can you formulate this problem as a special case of a closed-test procedure? That is, what intersection-null test is it the closure of? (**Note:** this part suffices to prove part (a) so you can just write "see answer to (b)").

5. A Bit of Philosophy (**Note:** Graded for completion only; write as little or as much as you want, but write something. Also note I don't know the answer to this question!)

Suppose a journal decides to embrace statistical rigor and requires that in each submitted paper, all of the hypothesis tests / confidence intervals, taken together, must control the FWER at level $\alpha = 0.05$. In other words, if ten confidence intervals appear in your paper, they must have been generated according to a procedure guaranteeing that, with 95% probability, all ten cover their true parameters.

In a meeting of the editors, one particularly conservative editor pipes up saying "this is a good start, but really we should be controlling the FWER across all of the inferences in all of the articles in each issue of the journal." Discuss the feasibility of this proposal. Aside from feasibility, do you think this is a good goal? Why stop at FWER for each issue, as opposed to FWER control for each year, or over the entire life of the journal? If you think these proposals are too conservative, is there a principled reason to require FWER control for each article but not for each issue of the journal?