Stats 210A, Fall 2017
Bonus Homework

Not due

1. Derived Intervals for $\bar{\mu}$

You are working with a scientist, who has one-way layout data:

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

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

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

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 $\bar{\mu} = n^{-1} \sum_{i=1}^n \mu_i$, and she expresses regret that she didn't think of asking about it before seeing the data.

"Aha!" you cry, "but we can use the intervals we just constructed to derive an interval for μ."

(a) Use the simultaneous confidence intervals above to derive a confidence region for $\mu \in \mathbb{R}^n$.

(b) Give an explicit expression akin to (1) for the interval for $\bar{\mu}$ based on this region.

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

(d) What is its radius for $\alpha = 0.05$ and $n = 3$, 5, 10, and 100? (Give numbers e.g. 3.45).

4. Testing Hypotheses in Fixed Order

Suppose someone gives us an a priori ordering on hypotheses $H_{0,1}, \ldots, H_{0,n}$ 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 $p_1 > \alpha$, stop and accept all null hypotheses. Otherwise, reject $H_{0,1}$ and keep going. Then, if $p_2 > \alpha$, 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.

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