

DS 102 Discussion 2  
Wednesday, September 9, 2020

1. **Bonferroni controls FWER.**

Suppose you have a test  $T(X)$ , where  $X$  is your data. Recall that if we observe a test statistic  $t$ , the p-value  $P$  of the test is:

$$P(X) = \mathbb{P}(T(X) > t)$$

, where  $\mathbb{P}$  denotes the probability with respect to data drawn from the null distribution. This is the probability that, under the null distribution, you see a result at least as extreme as from your data. Remember that a  $p$ -value is a random variable, since it is a function of your data.

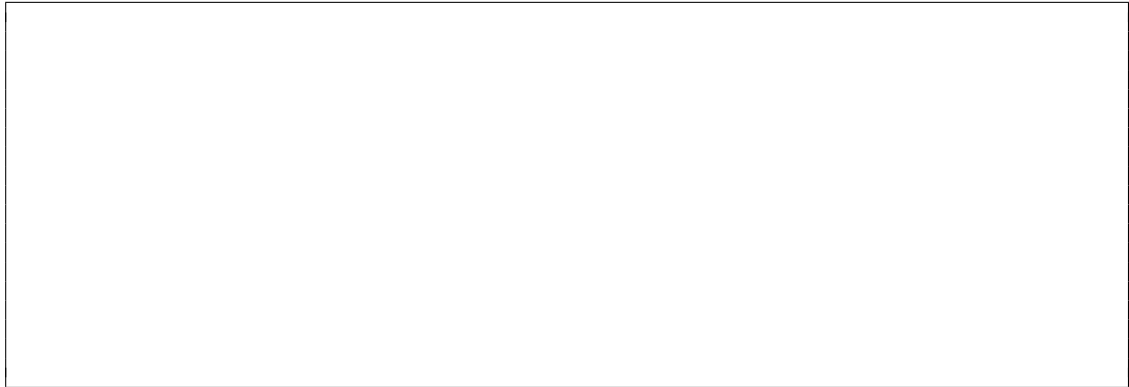
- (a) Suppose you have two independent  $p$ -values  $p_1$  and  $p_2$ . If on both hypotheses you choose a  $0 < \alpha < 1$  and use the naive decision rule:

$$\delta(p; \alpha) = \begin{cases} \text{reject null} & p \leq \alpha \\ \text{accept null} & p > \alpha \end{cases}$$

what is the probability of making at least one false discovery? This probability is also known as the family-wise error rate (FWER).

Hint: recall that the distribution of  $p$ -value under null hypothesis is uniform.

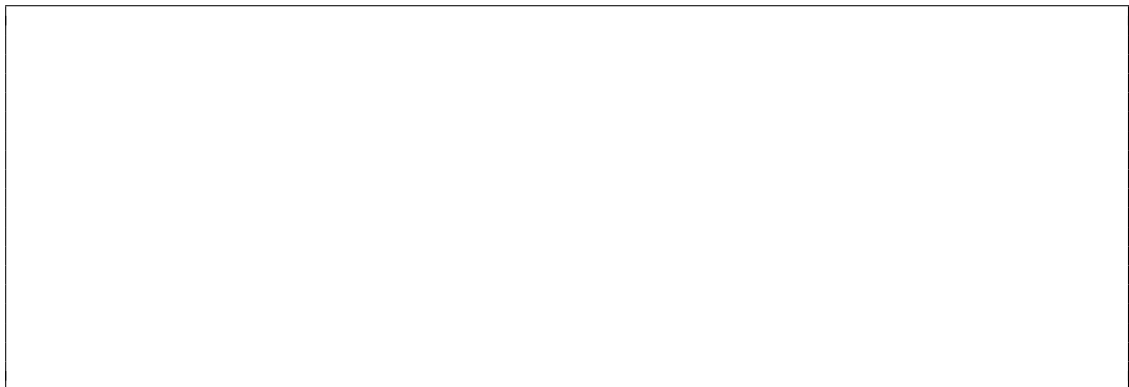
- (b) Does this decision rule keep the probability of a false discovery below  $\alpha$ ?



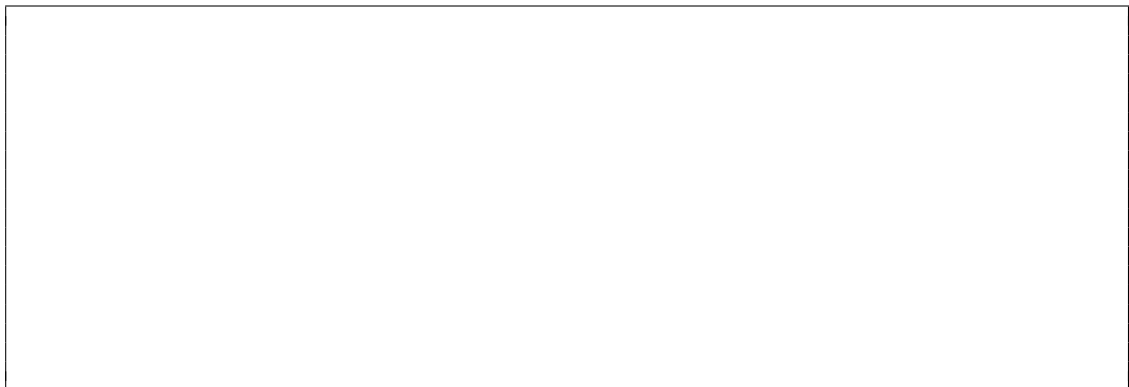
- (c) The Bonferroni correction, which uses the decision rule

$$\delta\left(p; \frac{\alpha}{n}\right)$$

controls the FWER described in the previous problem. Suppose you have  $n$  independent  $p$ -values:  $p_1, \dots, p_n$ . Show that the Bonferroni correction controls the probability of at least one false discovery. *Hint: Let  $E_i$  be the event that  $p_i < \frac{\alpha}{n}$ . This is also shown in the lecture.*



- (d) Given 10  $p$ -values for multiple hypotheses testing: 0.001, 0.003, 0.012, 0.015, 0.08, 0.09, 0.1, 0.14, 0.16, 0.28. What threshold should be used for the decision rule such that the FWER is less than 0.05? How many tests are rejected?



2. **Benjamini-Hochberg procedure** In this question we analyze the properties of the Benjamini-Hochberg (BH) procedure. Recall the steps of the procedure:

---

**Algorithm 1** The Benjamini-Hochberg Procedure

---

**input:** FDR level  $\alpha$ , set of  $n$  p-values  $P_1, \dots, P_n$

Sort the p-values  $P_1, \dots, P_n$  in non-decreasing order  $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(n)}$

Find  $K = \max\{i \in \{1, \dots, n\} : P_{(i)} \leq \frac{\alpha}{n} i\}$

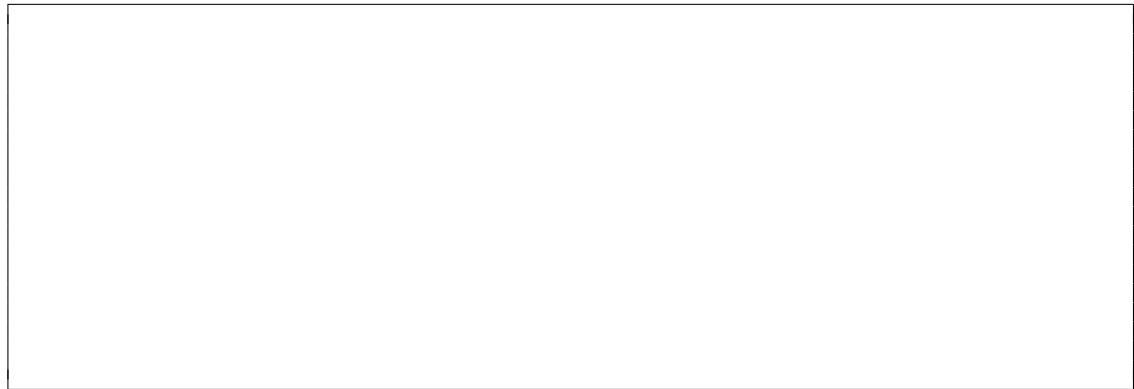
Reject the null hypotheses (declare discoveries) corresponding to  $P_{(1)}, \dots, P_{(K)}$

---

- (a) Given 10  $p$ -values for multiple hypotheses testing: 0.001, 0.003, 0.012, 0.015, 0.08, 0.09, 0.1, 0.14, 0.16, 0.28. Suppose we would like to control the FDR at the level 0.05. How many tests are rejected?

- (b) Suppose  $P_1 = P_2 = \dots = P_n = \alpha$ , and we run BH under level  $\alpha$  on these p-values. How many discoveries does BH make? Explain.

- (c) Suppose  $P_1 = P_2 = \dots = P_{n-1} = \alpha$ ,  $P_n = \alpha + 0.001\alpha$ , and we run BH under level  $\alpha$  on these p-values. How many discoveries does BH make? Explain.



- (d) Suppose we run BH on  $\{P_1, \dots, P_n\}$ , and we make  $R < n$  discoveries. Now suppose we add an extra p-value equal to 0 to this set. Now we run BH on  $\{P_1, \dots, P_n, 0\}$  and get a new number of rejections  $R'$ . Which of the following are possible:  $R' > R$ ,  $R' = R$ ,  $R' < R$ ? If multiple are possible, list all that are possible. Explain why.

