Homework 6

Data Science Team

Due Tuesday August 9 at 10am

Question 1 In this problem we will compare the number of wrong decisions when using the usual criterion of rejecting when the p-value is less than 5% with the Benjamini Hochberg procedure.

(a) Generate a vector with 1000 normal random variables from N(0,1). Then add 1 to each of the first 100 (so they are N(1, 1)). We can think of these 1000 numbers as the values of 1000 test statistics that result from 1000 hypothesis tests, where in the first 100 tests the null hypothesis (mean = 0) is false, and in the last 900 the null hypothesis is true. Then compute the p-values of these 1000 tests. Hint: use the function `pnorm`.

(b) Use the standard rule for testing one hypothesis: reject if the p-value is less than 0.05. How many type I errors (false positives) did you get? How many type II errors (false negatives)? What is the false discovery proportion?

(c) Write a R function that takes as input a vector of p-values and implements the Benjamini Hochberg procedure at the level 0.05. The output of the function should be the Benjamini Hochberg cutoff, i.e. the largest p-value that gets rejected.

(d) Apply the Benjamini Hochberg procedure from (c) to the p-values from (a). How many type I errors (false positives) did you get? How many type II errors (false negatives)? What is the false discovery proportion?

(e) Compare the result of the standard rule in (b) with that of the Benjamini Hochberg procedure in (d). Which one is preferable in this multiple testing situation?