HW 2 - due 02/05 at 11:59 pm.

Math 185, Winter 25, Rava

Follow closely the 'Hw guide' under Files in the folder 'Course Contents' on how to write, scan and submit your homework.
On any problem involving R, you should include your code and output as part of your answer. You may take a screenshot of the code/output, or write it by hand.
Be careful with notation, remember to define the parameters and the random variables you intend to use.

1 Exercise 1

You want to study the difference between two different brands of light bulbs in terms of their median burnout time. You record the burnout times, in years, of 100 light bulbs of brand A ($X_1, \dots, X_n$) and 80 light bulbs of brand B ($Y_1, \dots, Y_m$). The burnout times are recorded in the dataset 'lightbulb.csv' that is available on Canvas. We consider as estimator of the difference between the two population medians θ, the difference between the two sample medians $\hat{\theta} = median(X_1, \dots, X_n) - median(Y_1, \dots, Y_m)$.

a) [1 points] Load in R the data. Find the value of the estimate $\hat{\theta}$.

b) [3 points] Use bootstrap to estimate the bias of the estimator. Use $B = 200$.

c) [1 points] Use bootstrap to estimate the variance of the estimator. Use $B = 200$.

d) [2 points] Obtain a 95% percentile bootstrap CI for θ. Use $B = 1000$.

e) [2 points] Obtain a 95% Hall's percentile bootstrap CI for θ. Use $B = 1000$.

2 Exercise 2

Let $X_1, \dots, X_n \sim_{i.i.d.} F_X$ and $Y_1, \dots, Y_m \sim_{i.i.d.} F_Y$ be two independent samples. We want to test $H_0 : X \sim Y$ vs $H_1 : X \succeq Y$ using as test statistic $T = \bar{X} - \bar{Y}$. In class, we have talked about the permutation t-test that uses as p-value $p - value = \frac{\#\{\pi : T_\pi \geq T_{obs}\}}{(m+n)!}$ where T_π, T_{obs} have been defined in class. In this exercise, we will use simulations to estimate the probability of type I error and the power of the approximate permutation t-test. In this exercise use B=1000.

a) [4 points] Use 1000 simulations to estimate the probability of type I error of the approximate permutation test. Is this close to the probability of type I error that the test should have? Use $n = m = 10$ and $\alpha = 0.05$. You can choose the distributions of X and Y.

b) [2 points] Repeat exercise a) but this time use $n = m = 50$. Use the same distributions you used in part a). Do the results change? Does it make sense? (It might take a minute to get the results in this case).

c) [2 points] Use 1000 simulations to estimate the power of the approximate permutation test. Use

$n = m = 10$ and $\alpha = 0.05$. You can choose the distributions of X and Y. Avoid situation in which the power get estimated very close to 1.

d) [2 points] Repeat exercise c) but this time use $n = m = 50$. Use the same distributions you used in part c). Do the power increase, decrease or stay the same? Does it make sense? (It might take a minute to get the results in this case)

3 Exercise 3

Let X, Y be continuous RVs and let $\rho = Cor(X, Y)$. We want to test $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$. Given a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ we consider as test statistic the sample correlation $\hat{\rho} = Cor((X_1, Y_1), \ldots, (X_n, Y_n))$. The p-value of the permutation test is defined as $p - value = \frac{\#\{\pi \colon |\hat{\rho}_\pi| \geq |\hat{\rho}_{obs}|\}}{n!}$ where $\hat{\rho}_\pi$ and $\hat{\rho}_{obs}$ have been defined in class. In this exercise we use simulations to investigate the power of the approximate permutation test. We let $n = 100$, $X \sim U(0, 1)$ and $Y = aX + \epsilon$, where $\epsilon \sim N(0, 1)$.

a) [4 points] Use 1000 simulations to estimate the probability of type I error of the approximate permutation test. Is this close to the probability of type I error that the test should have? Use $B = 1000$ and $\alpha = 0.10$. (It might take a minute to get the results in this case).

b) [3 points] Use 1000 simulations to estimate the power of the approximate permutation test for $a = 0.8$, $a = 1$ and $a = 1.2$. Use $B = 1000$ and $\alpha = 0.10$. (It might take a minute to get the results in this case). Compare the three estimates of the power obtained. Say whether they are similar or increase or decrease when a increases. Explain why you think it makes sense for the power to stay the same/increase or decrease when a increases.