HW 5 - due 05/26 at 11:59 pm.

Math 181B, Spring 23, 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

[7 points] Assume that $Y_{ij} = \mu_j + \epsilon_{ij}$ for $j = 1, \ldots, k$, $i = 1, \ldots, n_j$, where $\epsilon_{ij} \sim N(0, \sigma^2)$. Assume that all the ϵ_{ij} are independent. This is the setting we used in class. Find $E(SSTOT)$, where $SSTOT = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{..})^2$ is the total sum of squares defined in class.

2 Exercise 2

You are curious to know if there is any difference in average daily screen time between adolescents (13-19 years), adults (20-39 years) and middle age adults (40-59 years). You interview random adolescents, adults, and middle age adults and you record their daily screen time. The data, in hours, are collected in the three CSV files 'Adolescent.csv', 'Adult.csv', and 'MiddleAge.csv', that are available on Canvas.

a) [6 points] Conduct an HT with significance level 0.05. You can use R to compute the value of the test statistic and the p-value as shown in class. However, make sure to report every line of code you use, also those lines you use to create the appropriate dataset. Report the complete R output for your test. Make sure to define parameters, hypotheses, assumptions and write a meaningful conclusion.

b) [4 points] From the test conducted in part a), can you conclude that on average adolescents spend more time in front of the screen than both adults and middle age adults? Can you conclude that on average adults spend more time in front of the screen than middle age adults? If your answer is yes, explain why you can make such conclusions. If your answer is no run the appropriate HT. Make sure that the probability of committing at least one type I error is 0.05. Feel free to use built-in functions of R to run the tests.

3 Exercise 3

[6 points] Assume you are interested in measuring the value of Y in m different groups. Let μ_j be the mean of Y in group j, for $j = 1, \ldots, m$. Consider a multiple testing problem where you have $\binom{m}{2}$ pair of hypotheses $H_0^{ij} : \mu_i = \mu_j$ vs $H_1^{ij} : \mu_i \neq \mu_j$, for $i, j = 1, \ldots, m$ and $i \neq j$. Suppose you have m independent samples, one for each group and you can assume equal variance of Y across groups. In class we have learned that if we were to conduct $\binom{m}{2}$ two-sample t-tests with $\alpha = 0.05$ the probability of committing at least one type I error would be greater than 0.05. In class we have talked about the Bonferroni correction, that consists in conducting $\binom{m}{2}$ two-sample t-tests with $\alpha = 0.05/m$. This method guarantees that the probability of committing at least one type I error is less than 0.05. In this exercise we use simulations to see that this is indeed the case.

In R sample $m = 3$ samples of size $n = 50$ from a normal distribution with mean 1 and standard deviation 1. Apply the two-sample t-test three times, to first and second sample, to second and third sample and to first and third sample. Use the built-in function of R *t.test* to do so (remember to pass as argument $var.equal = T$). For each test extract the $p - value$ and check whether the $p - value < 0.05$, if this is the case, you reject the null. Record in a vector the results of the three tests. TRUE/1 means that you rejected the null, FALSE/0 means that you have failed to reject the null. Check whether in your vector there is at least one TRUE, that is whether you have committed at least one type I error.
Now apply to the same data the Bonferroni correction, that is, for each test extract the $p - value$ and check whether the $p - value < 0.05/m$. Again collect the result of the three tests in a vector and check whether in your vector there is at least one TRUE.
Repeat the process 1000 times. (Hint: To do so you can create a function in R that does what we have described and return a vector of two elements, the first element is TRUE if you have committed at least one type I error without using the Bonferroni correction, the second is TRUE if you have committed at least one type I error using the Bonferroni correction. Then you can apply the function *replicate* to the function you created. If you are not comfortable with the function *replicate* you can instead use a for loop.)
At this point you should have two vector of 1000 elements each, one for each procedure. TRUE/1 means that you committed at least one type I error, FALSE/0 means that you have not committed any type I errors. Check the proportion of TRUE for both procedures and briefly comment the results.