HW 7 - due 06/02 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

[6 points] Assume that $Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$ for $j = 1, \ldots, k$, $i = 1, \ldots, b$ where $\epsilon_{ij} \sim N(0, \sigma^2)$. Assume ϵ_{ij} are independent. Prove that $E(SSB) = \sigma^2(b-1) + \sum_{i=1}^{b} k(\beta_i - \beta)$, where $SSB = \sum_{i=1}^{b} k(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$ is the block sum of squares defined in class.

2 Exercise 2

You and a couple of friends took your dogs to puppy class. The three dogs did a pretty good job at learning different commands. Do they take on average the same time to respond to commands? Each one of you give your dogs 3 different commands. The time, in seconds, it takes each dog to respond to the commands is recorded. Here are the results:

	Dog 1	Dog 2	Dog 3
'Sit'	5	6.1	7.2
'Wait'	6	5.9	7
'Drop it'	6.5	6.3	7.3

a) [5 points] Conduct an HT with significance level 0.05. You can use R to run the test. However, make sure to report every line of code you use. Report the complete R output for your test. Make sure to define parameters and hypotheses and to draw a conclusion. Don't worry about assumptions.

b) [3 points] Do you think that the commands chosen have an important impact on the analysis? Define parameters and write hypotheses that express this idea, and then use the RBD printout to decide this issue using $\alpha = 0.05$.

3 Exercise 3

In this exercise you will empirically compare the power of the Anova procedure with the power of the RBD procedure explained in class. Consider $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$. In R sample $Y_{ij} = \mu_j + \beta_i + \epsilon_{ij}$, with $\epsilon_{ij} \sim N(0, 0.4)$. Construct a data frame with 3 columns, a numeric one containing the observations, a categorical one containing the treatment, and a categorical one indicating the block. (Make sure that the second and the third column are indeed categorical, that is they contain elements such as '1' and not 1). Run on this data frame both the Anova procedure, which ignores the blocking and the RBD procedure, which takes into account the blocking. Save the p-values for testing the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$. Check if your Anova $p - value < 0.05$ and report 1/TRUE if it does, 0/FALSE if it doesn't. Do the same for the RBD $p - value$. TRUE/1 means that you reject the null, FALSE/0 means that you fail to reject the null. Repeat the process 1000 times and record in a matrix/data frame 1000x2 the results of the 1000 Anova tests in the first column and the results of the 1000 RBD tests in the second column. Compute the proportion of 1/TRUE in both columns. Since μ_j are different from each other, the proportion of the TRUE in the first column will approximate the power of the Anova test while the proportion of the TRUE in the second column will approximate the power of the RBD test.

a) [6 points] Run the procedure described using $\beta_1 = \beta_2 = \beta_3 = 1$. Compare the power of the two tests. What do you notice?

b) [4 points] Run the procedure described using $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$. Compare the power of the two tests. What do you notice?