**Exercise 1 - Submission guidelines:**

\*\*Submission is in pairs –both of your exercises should be submitted in your moodle profiles. Only one of them will be tested. \*\*  
The exercise should be submitted in the relevant place in the Moodle. You can submit the exercise in one of two ways:

1. Using a word file to answer the text questions and an R code for the code. Make sure to note which question exactly you are answering.
2. Using R Markdown, submitting the .Rmd file and the output in html format. Answer the text question between blocks in the code.

-The first lines in the code and word file should **specify your names and IDs**.

-In any case, for a perfect score - **all** parts of code you used to answer the questions should be submitted and it should be clear which code part refers to which question. We highly recommend adding as many comments as possible to your code.

-Before running the code, we will change only the setwd() line to set a new working directory. Other than that, all code should run smoothly.

-Answers to text questions should be concise and to the point. Use no more than 75 words for each answer (3-4 lines in word with font 11). Answers that do not answer the question directly or that include a lot of unnecessary information will be marked as wrong.

**Part 1 (40)**

The dataset attached is from experiment 1 of Merkva, Westfall, & Van Bowen (2019). In short, subjects were asked to detect targets (among non-targets) in an array of pictures and then rated the intensity of emotional response for some of the pictures, targets and non-targets. The authors' main hypothesis was that emotional responses for targets (compared to non-targets) will be rated as more intense. In total, each subject rated 30 pictures.

Please note that the analyses performed in the paper were slightly different from those you will be asked to perform here.

1. Download the dataset "attention-drives-emotion-2019-exp1" and load it in R. Read the methods section from the paper and the codebook attached.
2. Each line describes one rating. What are the relevant columns that describe the rated intensity of emotional response, rated emotional intensity, subject identity, valence of image rated and whether the image was a target or-non target?
3. Create a new variable "rating" that is the average of rated emotional intensity and rated intensity of emotional response, for each image rated (hint: google “how to average two columns?”).
4. Subset the data set to remove:  
   -Participants who answered they thought the experimenter wanted them to "to rate the target images as much more emotionally intense than the other images"  
   -Trials with Neutral images  
   How many participants were removed?
5. Create a new variable "z\_rating" which includes the z-scores of "rating" **within subjects**   
   (hint 1: use "group\_by" as in the TA and then "mutate" instead of "summarise". The line “mutate(data,new\_column\_name=old\_column+1)” creates a copy of the previous data set with the new column added (in this example adding 1 to the previous column) )

(hint 2: "scale" is the function for z-scoring a numeric vector).

1. Use the variable to remove images which were rated +-2 SD from the subject average (**within** each subject).
2. Create a summary dataset which includes the following information for each subject and stimulus type (2 rows per subject, one for targets and one for non-targets):

-mean rating

-sd of rating

-the number of trials left for each participant and stimulus type.   
What is the total number of **target** trials left?

You can use a “for” loop for this and connect subsets of the data using “rbind”, but this could also be performed through “group\_by”

(hint: you can use length(rating) to get the number of trials per subject&stimulus type after you group)

1. Load the gender data, which include the reported gender of each subject. Subjects reported their gender using text so they used different notation to write the same thing. Write a code to fix this dataset so that the "gender" column will include "F" for females, "M" for males and "O" for any other gender.   
   (hint: the "unique" function can assist you in finding all the different variations. You can use this line of code to find the relevant rows to edit for each variation: data[data$gender=="femal"]). How many females, males, and others are in the sample?
2. Merge the summary data with the gender data, so that another column "gender" will be added to the data. What are the means and standard deviations for each gender? Prefarably, report in a table.
3. Conduct the relevant t-test to test the authors' hypothesis using the t.test() function.   
   Assume normality and equal variances, and make sure you get all the arguments for the functions right (there are default settings which won’t necessarily match the kind of test you would like to conduct).  
   Report the results verbally, as you would in a results section of a paper: means, standard error, t-score, degrees of freedom, and p-value. The reader should be able to understand from what you wrote whether the hypothesis was confirmed and what it means.

1. Plot a graph that summarize the result. You can use the "plot" or "hist" function from the base R library, or the function "ggpaired" from the "ggpubr" library. Feel free to use any function of your choice as long as the plot shows all the relevant data to understand the answer for the research question and for the conducted test.

**Part 2 (60 points)**

The goal of this part is to perform a simulation. Simulations are an important tool to perturb our assumptions and find out the actual distributions we would expect to get. In this case, you will use simulations to look into the prevalence of Type I and Type II errors when using t-test under different conditions.

1. Create a function named "t\_test\_sim(M1, M2, S1, S2, N1, N2)" – the function gets as input two expected values and two standard deviations – the parameters of two populations. It also receives two sample sizes for groups sampled from these populations. The function then (15 points):
   1. Samples two different groups from a normal distribution, with the parameters defined for the function as inputs. It performs an independent 2-sample t-test, assuming equal variances between them and save the output of the t.test function into another variable. The result should be a list summarizing the results of the test.
   2. Creates two new variables with the t-statistic and the p-value.
   3. Returns those two variables in a vector/dataframe.

Call the function for conducting a t-test for two samples (with size 50) from the same population and report the t-score, degrees of freedom, and p-value.

1. Run a loop for 1,000 iterations with sample sizes of 30 on each group and equal standard deviations (unless instructed otherwise) (15 points ).  
   In each iteration, call the function and save the output (t-scores and p-values) for simulations of the t-scores received when:
   1. The null hypothesis is correct (Cohen’s d=0, meaning the distance between population means is 0 sd)
   2. Cohen’s d=0 but one of the standard deviations is 3 times larger than the other (sd\_a = 3\*sd\_b)
   3. Cohen’s d = 0 but sample size = 90 in each group
   4. Cohen’s d = 0.5
   5. Same means difference as (d) but one of the standard deviations is 3 times larger than the other (sd\_a = 3\*sd\_b)
   6. Cohen’s d = 0.5 but sample size = 90 in each group

Each time, **visualize (as histograms) and save the distributions of p-values and t-scores** you get. Make sure to give informative names to the variables and graphs you create!

1. Answer the following questions regarding the above distributions (15 points ):
   1. What was the p-value of the t-score you received in Part 1?
   2. What is the shape of the distribution of p-values when no effect is present (i.e., H0 is correct)? How does this distribution change when there a true effect is present (under H1)?
   3. What happens to the distribution of p-values and t-scores when the sample size increases? How does it depend on the existence of an effect? Explain the results using the terms learned in class (e.g., Type I and Type II error, statistical power, sampling distribution).
   4. What happens to the distribution of p-values and t-scores when the variances are unequal? How does it depend on the existence of an effect? Explain the results using the terms learned in class (e.g., Type I and Type II error, statistical power, sampling distribution).
2. (15 points)
   1. Re-write the function you used for the simulations, this time setting the t.test() function to perform welch t-test (try to figure out how using “?t.test" ). How does the answer for question ***3d*** changes when you do that? Explain in your times the influence of using welch t-test instead of regular t-test.
   2. Re-write the function to ask a question of your own using simulation on t-test. It could be the effect of any other kind of change in the sampling process. Make it interesting and be creative! Try to explain the result you got in your simulations.

**Good luck!!**