

SS2864B, 2021
Assignment #1 due to Feb 1, before 12pm, 2021

Instructions Submit an electronic version (pdf, words, etc) of your solutions (appropriately annotated with comments, plots, and explanations) to owl. Save all your R codes in one script file with proper comments and submit it as well to owl.

1. Use R functions **search** and **objects** to find all possible R builtin functions related to normal distribution. **rnorm** should be one of them. **qqnorm** is another one. Explain what the following three lines of codes do:

```
x = rnorm(100)
qqnorm(x)
qqline(x)
```

2. Use **for** loop to find the sum of a random permutation of 1,2, to, 100 as `x=sample(100)` as well as use R function **sum** directly. Modify your **for** loop codes to get the minimum instead of sum and use R function **min** directly. Comment your findings and which way is better?
3. Calculate the sum $\sum_{i=1}^N 1/i$ and compare with $\log(N)+0.6$ for $N = 500, 1000, 2000, 4000, 8000$. Comment your findings.
4. Eight patients are enlisted in a diet experiment. We have a full 2^3 factorial design. Please create an R matrix object representing the design matrix (do not create the matrix manually):

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ 3 & 1 & 2 & 1 \\ 4 & 1 & 2 & 2 \\ 5 & 2 & 1 & 1 \\ 6 & 2 & 1 & 2 \\ 7 & 2 & 2 & 1 \\ 8 & 2 & 2 & 2 \end{pmatrix}$$

Hint: check `help(rep)`

5. Consider the built-in data frame **Formaldehyde**.
 - (a) Extract the elements of the third row of this data frame.
 - (b) Extract the elements of the **carb** (carbohydrate) column.
 - (c) Use the **plot** function to identify the relation between **optden** (optical density) and **carb**. Comment your findings.

6. Consider the data frame **Pima.tr** from the package **MASS**.
 - (a) Determine the number of rows and columns for this data frame.
 - (b) Calculate the median of each column of this data frame with numerical values.
 - (c) Find the average **bmi** for those people who tested **Yes** for diabetes. Also find the average **bmi** for those people who tested **No** for diabetes. Comment on your findings.
7. Consider the data frame **Cars93** from the package **MASS**.
 - (a) Find the number of columns and their names for this data frame.
 - (b) Convert the **Weight** (pounds) column to a factor as "Small" if **Weight** is between 0 and 2500; "Median" if between 2500 and 3500; and "Large" if over 3500. Add such a column with name "**Car.size**" to the data frame **Cars93**.
 - (c) Find the mean of **MPG.city** for those cars in "Small", "Median", or "Large" categories respectively. Commend your findings.