SCS 2211: Laboratory II – Statistical Methods using R

Take-home Assignment - 2019

Instructions:

- Type the answers and paste the corresponding outputs for the questions in a Word document where necessary and you should also submit the complete R code (in **.R** format)
- Name the Word document and R file with your index number and put them into a zipped folder.
- Finally submit the zipped folder named with your index number.

Questions:

1) Load the data set "birthwt" in the R package "MASS" and attach it.

2) Prepare a table to list down the variables in the data set with their corresponding scale of measurement. Mention whether the variable is discrete or continuous where appropriate.

3) Choose 3 qualitative variables from the data set and provide appropriate visualizations for each of those variables.

4) Draw appropriate graphical summaries for the quantitative variables and comment on the shape of the distributions.

5) Calculate suitable descriptive statistics for the quantitative variables and identify the most suitable measure of central location for each quantitative variable.

6) Is there any relationship between the age of the mother and the birth weight of the baby? Briefly explain your answer by using a suitable graphical summary.

7) Set the seed in R to your index number

8) Draw a random sample of 50 observations from the "birthwt" data set and save it as a new data set named "birthwt_sample". The sample you take should contain equal number of mothers who smoked [smoke=1] and did not smoke [smoke=0] during pregnancy.
   (Hint: You may first separate the "birthwt" data set into smoking and non-smoking mothers and then randomly select 25 from each group)

Use the newly created data set to answer the following questions:

9) Find mean and standard deviation of birth weight for the mothers who smoked and did not smoked separately.

10) Test whether the mean birth weight of babies born to mothers who smoked during pregnancy is different to mothers who did not smoked, at 5% level of significance.

    Assume that the two population variances are equal.

    (Do not use the inbuilt function in R to conduct the test. Calculation of the test statistic and the p-value should be done using your own logical R code. You should clearly state all the steps of hypothesis testing in the Word document.)

11) Fit a Simple Linear Regression model to predict the birth weight of a baby using the variable "lwt" (mother's weight at last menstrual period). You should use the entire dataset for this part.

    - Obtain the parameter estimates.
    - Write down the model equation.
    - Using a suitable measure, comment on the goodness of fit of the model.

Deadline for assignment submission: 15th December 2019 11.55 PM