**Assignment 2**

**FI 4090**

**Group Submission**

**Linear Regression**

***Note:*** *Quiz 2B in iCollege will be based on this Assignment. Please have your R program available in running condition when you take the quiz. You will need solutions of your program to take the quiz. Quiz 2B is not under lock down browser.*

**Part 1 (30 points)**

HousePrices data set is a cross-sectional data set on house prices and other features, e.g., number of bedroom, of houses in Windsor, Ontario. The data were gathered during the summer of 1987.

Use the HousePrices data to perform the following tests using Linear Regression settings:

1. Construct a summary stat for all the variables in the HousePrices data. ( 5 points)
2. What is the percentage of houses in the data with Driveway, Gas-Heat and Air-conditioning present? (Hint: find the mean after creating dummy variables with driveway, gasheat, and aircon variables respectively). (5 points)
3. Construct a linear regression model to test whether number of bedrooms influence house prices. Provide a summary of the linear regression model using summary() function. (10 points)

The online quiz (Q1 to Q4) will be related to the following concepts. You do not have to respond to the following questions in the R program:

* 1. How do you interpret the coefficient of Number of Bedrooms in the model?
  2. What is the null hypothesis related to the model to test the effect of number of bedrooms on house price?
  3. To infer the effect of number of bedrooms on house price, draw your conclusion based on p-value.
  4. Comment on model accuracy: R-square

1. Construct a multiple linear regression model by including all variables as predictors of house prices (response variable) and observe the effect on the house prices. Provide a summary of the regression model using summary() function. (10 points)

**Variable description of HousePrices data**: A data frame containing 546 observations on 12 variables.

price: Sale price of a house.

lotsize: Lot size of a property in square feet.

bedrooms: Number of bedrooms.

bathrooms: Number of full bathrooms.

stories: Number of stories excluding basement.

driveway: Factor. Does the house have a driveway?

recreation: Factor. Does the house have a recreational room?

fullbase: Factor. Does the house have a full finished basement?

gasheat: Factor. Does the house use gas for hot water heating?

aircon: Factor. Is there central air conditioning?

garage: Number of garage places.

prefer: Factor. Is the house located in the preferred neighborhood of the city?

**Part 2: (40 points)**

Use the Credit data to perform the following tests using Linear Regression settings:

1. Perform the following steps: (20 points)
   1. Attach the Credit data to the R environment. (5 points)
   2. Observe the number of rows in the Credit data. Observe the dimension of the Credit data. ( 5 points)
   3. Provide a summary stat for the variables in Credit data. (5 points)
   4. What is the percentage of Student in the Credit data? What is the percentage of Female in the Credit data? (5 points)
2. Construct a linear regression model as follows: (20 points)

**Response variable:** Credit Card Balance

**Predictors:** Credit Rating, Student, Credit Rating \* Student (interaction terms)

Provide a summary of the model using summary() function.

The online quiz (Q5 to Q8) will be related to the following concepts. You do not have to respond to the following questions in the R program:

1. What is the effect of *Student* on *Credit Card Balance*? Explain the coefficient in terms of how *Student* status effect changes in *Credit Card Balance.* Explain the significance level of the coefficient of *Student*.
2. What is the total effect of *Credit Rating* on *Credit Card Balance* for non-students? Explain the coefficient in terms of how changes in *Credit Rating* effect changes in *Credit Card Balance* for non-students*.* Explain the significance level of the coefficient of *Credit Rating*.
3. What is the total effect of *Credit Rating* on *Credit Card Balance* for students? Is the effect of *Credit Rating* on *Credit Card Balance* is significantly different for students vs. non-students? Explain results using relevance test statistics.

**Part 3: (30 points)**

Use the Credit data to perform the following tests using Linear Regression settings: Online quiz (Q9 to Q10) will be based on the results of the regressions performed below.

1. Test whether Age influence Credit Card Balance on the basis of simple linear regression.

(Provide a summary of the model using summary() function).

1. Use Age and Credit Rating as predictors of Credit Card Balance (response variable) in a multiple linear regression setting. (Provide a summary of the model using summary() function). Interpret the effects of both the predictors.
2. Compare effect of Age from part (i) and (ii). **Provide answer to this question in R Code (Script).**

**Deliverables:**

1. Submit R scripts electronically in iCollege in the corresponding Assignment tab.
2. Please submit **one R program (one file)** containing **three parts of the assignment (mark/comment so that each part is separated clearly in the program).** R code should provide comments on each sections of the assignment the code is intended for. **Also indicate which team member contributed to which sections of the code.**
3. The assignment submission grade will be based on whether you have completed each part of the analysis and whether your R code run through all the parts of analysis. Your grade will be based not only on the correctness of the program but also how efficiently the program executes the tasks.
4. Note that **you do not have to write your response** to the above questions (except for Part 3 (iii)) related to the interpretation of the model results in the R code. If you do write the responses in the R program - it will not be part of your grade. Provide your explanation for Part 3 (iii) in R Code.
5. **The quiz 2B will have questions that will test your conceptual understanding of the output/results of the model. Make sure you understand the relevant concepts of the analysis in each part before you take the online quiz (individual submission). You are not allowed to collaborate with your team members in taking the quiz.**
6. Do not submit a separate word document explaining your results.