QUESTION 1: HYPOTHESIS TESTING

A researcher collect data from all the staffs in a college and only 100 staffs are selected to be the sample in his research. Below is the description of the data that is stored in *Data.xlsx*.

Description of dataset
1. "Age" is in years
2. "Educational Level" values are defined as follows:
   0 = No high school degree
   1 = High School Graduate
   2 = College Graduate
   3 = Graduate Degree
3. "Smoking Status" values are defined as follows:
   0 = Does not smoke
   1 = Smokes less than 1 pack per day 2 = Smokes one or more than 1 pack per day
4. "Exercise" values are defined as follows:
   0 = None 1 = Light
   2 = Moderate
   3 = Heavy
5. "Weight" is given in pounds
6. "Serum Cholesterol" is given in milligram percent (mg%)
7. "Systolic Pressure" is given in millimeters of mercury (mmHg)
8. "IQ" is given in standard IQ test score values
9. "Sodium" is given in milliequivalents per liter (mEq/l)
10. "Gender" is listed as
    1 = Male
    2 = Female
11. "Marital Status" values are defined as follows:
    1 = Single
    2 = Married
    3 = Widowed
    4 = Divorced

Based on the given dataset, using 5% significance level,

(1) Is there any significant difference in the mean of weight on genders?

(2) The researcher wants to review the figures of marital status data to see if all the statuses were equally distributed.

(3) Is there any interaction effect between education level and smoking status on the weights of the staff?

Row effect and column effect should be conducted if there is no interactions exist between the factors

(4) a) is there a significant relationship between Serum Cholesterol, Systolic Pressure, IQ and Sodium on the staff weight at 0.05 level of significance?

    b)   identify the independent variable that contributes to the regression model

    c)   interpret coefficient of correlation and coefficient of determination

    d)   write the regression model

(65 Marks)

QUESTION 2: INDEPENDENCE TEST

A student is interested to test whether there is a relationship between gender and course at her university. She randomly picked some men and women in her campus and asked them if their course was part of natural sciences (NS), social sciences (SS) or engineering (E). Her results can be shown as follows:

	NS	SS	E	Total row
Women	2	16	8	$n_1 = 26$
Men	9	15	10	$n_2 = 34$
Total column	$n_1 = 11$	$n_2 = 31$	$n_3 = 18$	$n = 60$

Can we conclude that these two variables are related to each other by using 10% significance level?

(10 Marks)

– END OF QUESTIONS –