Lebanese University
Faculty of information
Section 1

Final Exam: Statistics 3

Time: 90 minutes
Dr. Abbas RAMMAL

EXERCISE 1

A researcher wants to study the effect of unemployment on divorce. He collects data from ten cities and proposes the following regression model:

$$D_i = \beta_0 + \beta_1 C_i + \epsilon_i$$

where D_i is the divorce rate per 10,000 married couples in 2002, and C_i is the unemployment rate (%) in 2002. The random variables ϵ_i are independent and normally distributed for every individual i.
Note that:

$$\sum_{i=1}^{10} D_i = 258, \quad \sum_{i=1}^{10} D_i^2 = 6698, \quad \sum_{i=1}^{10} C_i = 48, \quad \sum_{i=1}^{10} C_i^2 = 270, \quad \sum_{i=1}^{10} D_i C_i = 1245.$$

SSE

The residual sum of squares is: $SC_{res} = 40.5$.

1. Is this a simple or multiple regression? Identify the dependent and independent variables.

2. Find the estimated values of the unknown coefficients β_0 and β_1.

3. Compute the sum of squares due to the regression.

EXERCISE 2

The "**freeny**" database in R is a data frame with the following variables:
y, **lag.quarterly.revenue**, **price.index**, **income.level**, and **market.potential**.
Type "?freeny" for more information.
We aim to explain **y** as a function of **lag.quarterly.revenue**, **price.index**, **income.level**, and **market.potential**.

1. Construct the multiple linear regression of **y** on **lag.quarterly.revenue**, **price.index**, **income.level**, and **market.potential**.

2. Test whether the ϵ_i are normally distributed and test the homogeneity of variances.

3. Test the hypothesis $H_0 : \beta_j = 0$ (against $H_1 : \beta_j \neq 0$) with a significance level of $\alpha = 5\%$ for $j = 1, 2, 3, 4$. Conclude.

4. Identify the number of outliers using the methods of externally studentized residuals and leverage points.

Exercise 3

The productivity Y of a variety of trees is evaluated according to its population density X (in plants per m2). The following table provides a summary of the values (x_i, y_i).

X	1.11	1.22	1.49	2.01	2.46	3	3.22	3.67	4.02
Y	1.73	1.49	1.1	0.7	0.52	0.39	0.31	0.2	0.17

1. Construct the simple linear regression of Y as a function of X.
2. Graph the residuals against the estimated values. What do we notice?
3. We would like to estimate productivity as a function of density using a relationship of the form $Y = \alpha X^{\beta}$. To do this, we set Z = ln(X) and T = ln(Y).

   a) Determine the linear correlation coefficient between Z and T. Interpret.

   b) Estimate the coefficients of the fitting line from T to Z. Give the regression equation.

   c) Plot the QQ-plot of the residuals.