Lebanese University
Faculty of information
Section 1

Final Exam: Statistics 3

Time: 90 minutes
Dr. Abbas RAMMAL

Problem 1

The following measurements have been obtained in a study:

No.	1	2	3	4	5	6	7	8	9	10	11	12	13
y	1.45	1.93	0.81	0.61	1.55	0.95	0.45	1.14	0.74	0.98	1.41	0.81	0.89
x_1	0.58	0.86	0.29	0.20	0.56	0.28	0.08	0.41	0.22	0.35	0.59	0.22	0.26
x_2	0.71	0.13	0.79	0.20	0.56	0.92	0.01	0.60	0.70	0.73	0.13	0.96	0.27

No.	14	15	16	17	18	19	20	21	22	23	24	25
y	0.68	1.39	1.53	0.91	1.49	1.38	1.73	1.11	1.68	0.66	0.69	1.98
x_1	0.12	0.65	0.70	0.30	0.70	0.39	0.72	0.45	0.81	0.04	0.20	0.95
x_2	0.21	0.88	0.30	0.15	0.09	0.17	0.25	0.30	0.32	0.82	0.98	0.00

It is expected that the response variable y can be described by the independent variables x_1 and x_2. This imply that the parameters of the following model should be estimated and tested

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

You can copy the following lines to R to load the data:

```
D <- data.frame(
  x1=c(0.58, 0.86, 0.29, 0.20, 0.56, 0.28, 0.08, 0.41, 0.22,
       0.35, 0.59, 0.22, 0.26, 0.12, 0.65, 0.70, 0.30, 0.70,
       0.39, 0.72, 0.45, 0.81, 0.04, 0.20, 0.95),
  x2=c(0.71, 0.13, 0.79, 0.20, 0.56, 0.92, 0.01, 0.60, 0.70,
       0.73, 0.13, 0.96, 0.27, 0.21, 0.88, 0.30, 0.15, 0.09,
       0.17, 0.25, 0.30, 0.32, 0.82, 0.98, 0.00),
  y=c(1.45, 1.93, 0.81, 0.61, 1.55, 0.95, 0.45, 1.14, 0.74,
      0.98, 1.41, 0.81, 0.89, 0.68, 1.39, 1.53, 0.91, 1.49,
      1.38, 1.73, 1.11, 1.68, 0.66, 0.69, 1.98)
)
```

1. Calculate the parameter estimates β_0, β_1, β_2 and $\hat{\sigma}^2$ in addition find the usual 95% confidence intervals for β_0, β_1 and β_2. (3 points)
2. Calculate the residuals and verify the property that the mean of the residuals is zero. (1 points)
3. Calculate the variance estimators of β_0, β_1 and β_2. (1 points)

4. Determine the coefficient of determination R^2 and interpret the result. (1 points)

5. Calculate the correlation coefficients between these variables taken two by two. Conclude. (1 points)

6. Test the correlation between these variables taken two by two. Conclude. (2 points)

7. Test the following null hypotheses, at the significance level α = 5%, using an appropriate F ratio: (2 points)

   a) $H_0: \beta_1 = 0$ in the model $y = \beta_0 + \beta_1 x_1 + \varepsilon$

   b) $H_0: \beta_2 = 0$ in the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

8. Establish the ANOVA table associated with this regression. What can we conclude about parameter β_1 and β_2? (1 points)

9. The multiple regression model can be written in matrix form as follows:

$$Y = \beta X + \varepsilon$$

Give the matrix X. (1 points)

10. Calculate the levers of the observations: (1 points)

$$h_{ii} = x_i (X'X)^{-1} x_i' \text{ avec } i = 1, \cdots, 50.$$

11. Prove that the sum of the levers is equal to p + 1. (1 points)

12. Give the number of levers which are greater than $2\frac{p+1}{n}$. (1 points)

13. Calculate the internal studentized residuals. How many points are suspected? (1.5 points)

14. Calculate the external studentized residuals. How many points are suspected? (1.5 points)

15. Calculate Cook's distance. (1 points)