086

The Open University of Sri Lanka
B.Sc/B.Ed. DEGREE, CONTINUING EDUCATION PROGRAMME
Open Book Test (OBT) 2023/2024
Level 05 - Applied Mathematics
ADU5301– Regression Analysis I

Date: 22.12.2023 **9.00 a.m. to 10.00 a.m.**

Instructions

- This examination is of **one-hour** duration.
- Answer **all** questions.
- Each of the two questions is allocated equal marks, distributed as indicated.
- Non programmable calculators are permitted.

1. a) Briefly describe two situations that the Pearson correlation coefficient cannot correctly measure the strength of the association between two variables.

(20 marks)

b) A researcher recorded the total area examined (in square meters) and the dried weight of litter collected (in grams), at a site. Following summary statistics were calculated from a sample of 60 observations.

$$\sum x_i = 156, \ \sum y_i = 1427, \ \sum x_i^2 = 540 \ \sum y_i^2 = 44313, \ \sum x_i y_i = 4819.$$

i) Calculate the Pearson correlation coefficient between the two variables.

(20 marks)

ii) Suppose that the data satisfy the conditions that permit correct interpretation of the Pearson correlation coefficient. Based on the value you obtained in part (i), what can you say about the association between the two variables.

(10 marks)

2. A researcher added known amounts (in mg) of a catalyst to 40 samples of a chemical solution, and measured the time taken for the completion of a chemical reaction (in minutes). Suppose the researcher seeks your advice to fit a regression model to the data.

i) Among the two variables, which one would you recommend as the response variable? Give reasons for your answer.

(5 marks)

ii) In relation to this study, briefly explain what is meant by the random error in the response.

(10 marks)

iii) In relation to this study, clearly state the assumptions you would make about the data to build a regression model, using the method of least squares.

(15 marks)

iv) Two scatter plots (Figure1, Figure2) constructed from the data collected are given below. Among the two plots, which plot would you recommend to select a candidate regression function to model the relationship between the response and the predictor variable you identified in part (i)?

Figure1

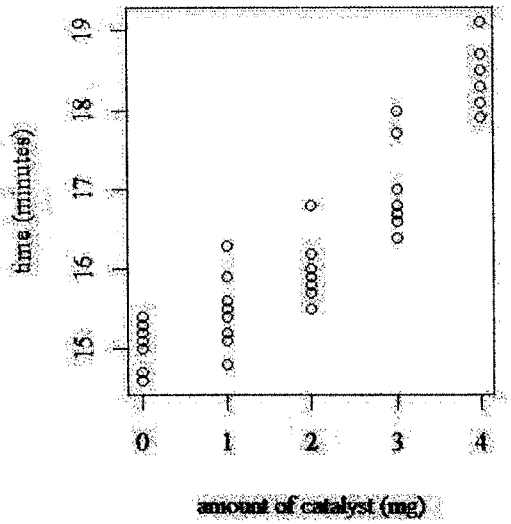
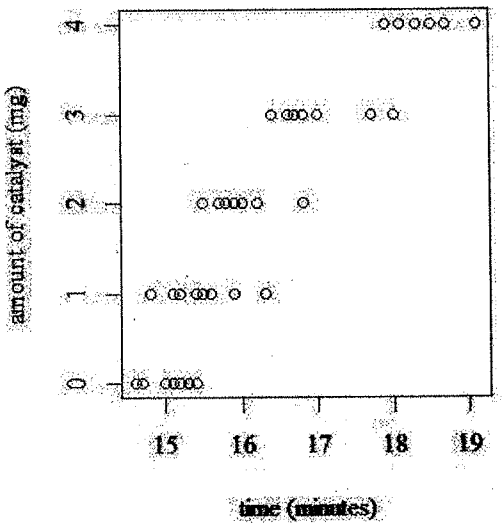


Figure2



(10 marks)

v) Does the plot you recommended in part (iv), indicate any violations of the model assumptions stated in part (iii). Give reasons for your answer.

(10 marks)