Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2023/2024

COURSE NAME	:	STATISTICS I
COURSE CODE	:	DAS 10503
PROGRAMME CODE	:	DAU
EXAMINATION DATE	:	JULY 2024
DURATION	:	3 HOURS
INSTRUCTIONS	:	1. ANSWER ALL QUESTIONS 2. THIS FINAL EXAMINATION IS CONDUCTED VIA ☐ Open book ☒ Closed book 3. STUDENTS ARE **PROHIBITED** TO CONSULT THEIR OWN MATERIAL OR ANY EXTERNAL RESOURCES DURING THE EXAMINATION CONDUCTED VIA CLOSED BOOK

THIS QUESTION PAPER CONSISTS OF **SEVEN (7)** PAGES

Q1 **Table Q1.1** shows the selected UTHM diploma students' height with respective shoe sizes.

Table Q1.1: Students' height with respective shoe sizes.

Student	A	B	C	D	E
Height (cm)	165	172	169	160	159
Shoe sizes	6.5	9.5	7.5	7	5

(a) Find the correlation, r, by using

(i) Pearson's Product Moment Correlation Coefficient

(13 marks)

(ii) Spearman's Coefficient of Rank Correlation.

(5 marks)

(b) Interpret the correlation value obtained from the Pearson and Spearman Coefficient of Rank.

(2 marks)

Q2 A university administration is interested in determining the duration of time employees take to travel from home to campus each day. A random sample of employees is chosen, and their travelling time (in minutes) is shown in **Table Q2.1.**

Table Q2.1: Travelling time of employees.

Travelling time	Numbers of employees
0 and less than 10	12
10 and less than 20	14
20 and less than 30	23
30 and less than 40	26
40 and less than 50	16
50 and less than 60	9

(a) Find class boundary, midpoint, and cumulative frequency.

(3 marks)

(b) Calculate the median and mode.

(8 marks)

(c) Calculate standard deviation.

(9 marks)

Q3 **Table Q3.1** shows the time taken by students to play sports and leisure activities on weekends in minutes.

Table Q3.1: Time taken by students for sports and leisure activities on weekends.

132	118	124	109	101	125	83	99	104
131	98	125	97	112	92	106	120	103
111	117	135	143	112	116	112	106	117
119	110	105	128	126	105	112	102	

(a) Differentiate quartiles and percentiles.

(2 marks)

(b) Construct a stem and leaf plot.

(3 marks)

(c) Arrange the data in a frequency distribution table starting with 80-89, 90-99 and so on.

(3 marks)

(d) Calculate the grouped data properties as follows.

(i) 1^{st} quartile

(4 marks)

(ii) 3^{rd} quartile

(4 marks)

(iii) 65^{th} percentile.

(4 marks)

Q4 A scientist wishes to study the time taken for a chemical reaction to synthesize a new compound based on applied temperature, as shown in **Table Q4.1.**

Table Q4.1: Temperature versus time taken for a chemical reaction.

Temperature (°C)	18	25	30	35	40	45
Time taken (hour)	9	6	4	3	4	1

(a) Regression analysis is a statistical method used in research. State the function of regression analysis.

(2 mark)

(b) Given $S_{xx} = 490.83$, $S_{yy} = 37.5$ and $S_{xx} = -126.5$. Based on **Table Q4.1**

(i) state the dependent and independent variables.

(2 marks)

(ii) find $\hat{\beta}_0$ and $\hat{\beta}_1$.

(5 marks)

(iii) find the estimated regression line, $\hat{y}$.

(2 marks)

(c) Calculate the Coefficient of Determination (R^2) by using the following equation and interpret the result.

$$R^2 = \frac{SSR}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}$$

(7 marks)

(d) Estimate the time taken if the required temperature applied is 20 °C.

(2 marks)

Q5 A boarding international school's hostel has 500 students. Four hundred students are local, while the remaining are from other countries such as India, China, and Japan. All students stayed in 4 blocks, namely blocks A, B, C, and D. Local foods were served daily in the cafeteria. However, some international students are dissatisfied with the existing menu as most of them cannot eat spicy food or rice in the morning and request that the food selection be improved. Furthermore, the students claimed that more breakfast selection could improve their focus in class and overall study experience.

Hostel administrations conduct online votes to decide on the breakfast menu involving 250 students from blocks A and B. Among other parameters are the origin of the students, gender, current academic performance, and food suggestions. The options were nasi lemak, toast, noodles, cereal, eggs, and fruits. The results show that only 1/5 of the students selected nasi lemak as their preferred breakfast menu.

(a) Identify the target population and sample of the study.

(2 marks)

(b) Briefly explain how the sampling frame of the study is acquired.

(2 marks)

(c) State the problem statements of the study.

(4 marks)

(d) State three (3) objectives of the study.

(3 marks)

(e) Identify the response and predictor variable of this study.

(2 marks)

(f) Describe a suitable research design for the study and, in your opinion, provide two (2) purposes of the study.

(5 marks)

(g) Give a suitable hypothesis that could be obtained from the study conducted.

(2 marks)

- END OF QUESTIONS -

FORMULA

$k = 1 + 3.3 \log n$

$$\bar{x} = \frac{\sum x}{n} \qquad \bar{x} = \frac{\sum fx}{\sum f}$$

$$M = L_M + C\left(\frac{\frac{n}{2} - F}{f_M}\right)$$

$$M_0 = L + C\left(\frac{d_b}{d_b + d_a}\right)$$

$$Q_1 = L_{Q1} + C\left(\frac{\frac{n}{4} - \sum f_{Q_1 - 1}}{f_{Q_1}}\right)$$

$$Q_3 = L_{Q3} + C\left(\frac{\frac{3n}{4} - \sum f_{Q_3 - 1}}{f_{Q_3}}\right)$$

$$P_k = L_{Pk} + C\left(\frac{\frac{kn}{100} - \sum f_{P_k - 1}}{f_{P_k}}\right)$$

$$Mean\ deviation = \frac{\sum |x_i - \bar{x}|}{n} \qquad Mean\ deviation = \frac{\sum f|x - \bar{x}|}{n}$$

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \qquad s^2 = \frac{1}{n-1}\left[\sum fx^2 - \frac{(\sum fx)^2}{n}\right]$$

$$r = \frac{mean - mode}{standard\ deviation}$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad b = (X'X)^{-1}X'y$$

$$a = \bar{y} - b\bar{x}$$

$$\sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2$$

$$\sum x_1 y = b_0 \sum x_1 + b_1 \sum {x_1}^2 + b_2 \sum x_1 x_2$$

$$\sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum {x_2}^2$$

$$R^2 = \frac{SSR}{SST}$$

$$SSE = S_{yy} - \hat{\beta}_1 S_{xy}$$

$$\tilde{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\hat{\beta}_0 = \frac{S_{xy}}{S_{xx}}$$