Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2022/2023

COURSE NAME	: STATISTICS I
COURSE CODE	: DAS 10503
PROGRAMME CODE	: DAU
EXAMINATION DATE	: JULY / AUGUST 2023
DURATION	: 3 HOURS
INSTRUCTION	: 1. ANSWERS **ALL** QUESTIONS.
	2. THIS FINAL EXAMINATION IS CONDUCTED VIA **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 **EIGHT (8)** PAGES

TERBUKA

Q1 A study is conducted to assess the quality of staffs in a company based on their quality of letters of recommendation. There are three predictor variables used to determine the quality of letters of recommendation, namely college GPA, high school GPA and TOEFL total. The following Excel output (**Figure Q1**) shows a multiple regression analysis when all the three predictor variables are used to predict the quality of letters of recommendation.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6304
R Square	0.3974
Adjusted R Square	0.3785
Standard Error	1.179
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	a	87.9735	29.3245	c	1.3967E-10
Residual	b	133.4165	1.3898		
Total	99	221.39			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.8793	0.5760	4.9991	0.0000
College GPA	0.0907	0.2039	0.4449	0.6574
High School GPA	1.3192	0.2000	6.5967	0.0000
TOEFL Total	-0.0006	0.0007	-0.8616	0.3911

Figure Q1: Summary output

(a) Identify the dependent and independent variables.

(4 marks)

(b) State the coefficient of correlation, r for this study and interpret this value.

(2 marks)

(c) Write the multiple regression equation.

(1 mark)

(d) Interpret the intercept value and the coefficient of TOEFL total when the other variables are remained constant.

(2 marks)

(e) Complete the ANOVA table.

(4 marks)

(f) At $\alpha = 0.02$, can we conclude that neither of the independent variables is related to the dependent variables

(4 marks)

(g) State the value of the coefficient of determination and interpret this value.

(3 marks)

TERBUKA

Q2 The manager of a slimming center has decided to illustrate a theoretical approach on how aerobic exercise can affect weight loss. Twelve of the club members carefully recorded the total number of minutes of aerobic exercises they did for one week and the amount of weight they had loss during the week. **Table Q2** indicates the data recorded.

Table Q2: The total number of minutes of aerobic exercise and the weight loss

Weight Loss	Aerobic exercise (in minutes)
1.2	112
2.8	190
2.3	171
1.9	148
3.0	193
3.8	235
4.1	237
2.5	176
2.6	185
2.5	186
3.6	228
0.4	65

(a) Identify the dependent and independent variables.

(2 marks)

(b) Sketch a scatter diagram and give a comment.

(3 marks)

(c) Calculate the correlation between aerobic exercise and weight loss by using

(i) Pearson product moment correlation coefficient.

(6 marks)

(ii) Spearman's rank correlation coefficient.

(7 marks)

(d) Interpret the correlation value that get from **Q2 (c)(i)** and **Q2 (c)(ii).**

(2 marks)

TERBUKA

Q3 Five Sdn Bhd is a manufacturing company producing oil gaskets for automobile and small tractor engine. The factory is located in Terengganu industrial area. Presently the company employs 500 employees. Three hundred of them are locals and have been working with the company since it started its operation in 1989. The rest of them are foreign workers from Bangladesh and Indonesia. Recently the management had been getting informal information that the employees are dissatisfied with the new policy that have been implemented for the past two months and this situation seems to be affecting their productivity. The general manager requested a survey research be conducted. Among the issues of concern is how much job satisfaction has been affected by the new policy. Other related factors include age, length of service, gender, opportunity for advancement and origin (local or immigrant). The sample size for this survey is 100. Based on the given information,

(a) Identify the target population and sample. Briefly explain how the sampling frame is acquired.

(4 marks)

(b) State the problem statement of the study.

(4 marks)

(c) State **three (3)** objective of the study.

(3 marks)

(d) Describe a suitable research design for the study and its purpose.

(4 marks)

(e) Identify the dependent and independent variable in this study.

(3 marks)

(f) State **two (2)** suitable hypotheses for this study.

(2 marks)

TERBUKA

Q4 As part of a quality control study aimed at improving a production line, the weights (in ounces) of 50 bars of soap are measured. The results are as follows, sorted from smallest to largest.

11.6	**12.6**	**12.7**	**12.8**	**13.1**	**13.3**	**13.6**	**13.7**	**13.8**	**14.1**
14.3	**14.3**	**14.3**	**14.8**	**15.1**	**15.2**	**15.6**	**15.6**	**15.7**	**15.8**
15.8	**15.9**	**15.9**	**16.1**	**16.2**	**16.2**	**16.3**	**16.4**	**16.5**	**16.5**
16.5	**16.6**	**17.0**	**17.1**	**17.3**	**17.3**	**17.4**	**17.4**	**17.4**	**17.6**
17.7	**18.1**	**18.3**	**18.3**	**18.3**	**18.5**	**18.5**	**18.8**	**19.2**	**20.3**

(a) Find the value of range of data set.

(1 mark)

(b) Construct frequency table start with 11.6 – 12.5, 12.6 – 13.5, 13.6 – 14.5 and so on. Calculate the mean, median, 1st quartile and 3rd quartile.

(12 marks)

(c) Construct a stem and leaf plot for these data.

(3 marks)

(d) Construct a box plot and comment the distribution.

(4 marks)

TERBUKA

Q5 (a) A research can use both primary and secondary data.

(i) Explain the definition of primary and secondary data and give **two (2)** example for both of it.

(6 marks)

(ii) List and briefly explain **two (2)** methods of collecting primary data.

(4 marks)

(b) Describe **four (4)** level of measurements in research. Give examples of each.

(6 marks)

(c) A group of researchers plan to carry out a survey on the repair workshops in Kota Bharu, Kelantan. These workshops can be categorized according to the type of vehicles of being repaired. There are 35 bicycle workshops, 50 motorcycle workshops, 85 car workshops and 20 heavy vehicle workshops. In order to save time and cost, the researchers plan to survey only 57 of these workshops.

(i) State the population in the study.

(1 mark)

(ii) State source of the data used in the study.

(1 mark)

(iii) Give ONE most suitable method of data collection to be used in the study and give the advantages of this method.

(2 marks)

- END OF QUESTIONS –

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM II 2022/2023
COURSE NAME : STATISTICS I
PROGRAMME CODE : DAU
COURSE CODE : DAS 10503

Formula

$k = 1 + 3.3 \log n$

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

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

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

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

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

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

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

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM II 2022/2023
COURSE NAME : STATISTICS I
PROGRAMME CODE : DAU
COURSE CODE : DAS 10503

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

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

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

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

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

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

$$\sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2$$
$$\sum x_1 y = b_0 \sum x_1 + b_1 \sum {x_1}^2 + b_2 \sum x_1 x_2$$
$$\sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum {x_2}^2$$

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

TERBUKA