اونيۏرسيتي مليسيا قهڠ السلطان عبد الله
UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

FINAL EXAMINATION

COURSE	:	**APPLIED STATISTICS**
COURSE CODE	:	**BUM2413**
COURSE COORDINATOR	:	**FARAHANIM BINTI MISNI**
DATE	:	**24 JANUARY 2024**
DURATION	:	**3 HOURS**
SESSION/SEMESTER	:	**SESSION 2023/2024 SEMESTER I**

INSTRUCTIONS TO CANDIDATES:

1. This examination paper consists of **SEVEN (7)** questions. Answer **ALL** questions.
2. All answers to a new question should start on a new page.
3. All the calculations and assumptions must be clearly stated and in **FOUR (4) decimal places**.
4. Candidates are not allowed to bring any materials other than those allowed by the invigilator into the examination room.

EXAMINATION REQUIREMENTS:

1. Statistical Tables & Formulae 2.0

APPENDIX:

1. None

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO

This examination paper consists of **NINE (9)** printed pages including the front page.

QUESTION 1 [8 MARKS]

Ikan Kembung or Indian Mackerel, is a vital protein source in Southeast Asia. A researcher claims that the population variance of the length for Indian Mackerel caught on rainy days exceeded 15. To test this claim, the researcher randomly selected 13 Indian Mackerel specimens caught by a local fisherman and the lengths (in cm) of these fishes are recorded as in **Table 1**.

Table 1

13.1	5.1	18.0	8.7	16.5	9.8	6.8	12.0	17.8	25.4	19.2	15.8	23.0

Assume that the length of Indian Mackerel follows normal distribution, verify the researcher's claim at a 0.01 level of significance.

[8 Marks]

QUESTION 2 [13 MARKS]

Plants utilize atmospheric CO_2, water, and sunlight to fuel their growth and reproduction. An experiment was conducted to assess the impact of different CO_2 concentrations on plant growth. Both groups of plants, one group exposed to normal air conditions and the other group exposed to elevated CO_2 levels. Both groups received equal amounts of water and light over a four-week period. **Figure 1** presents a *Microsoft Excel* output for the related test on the weight of plant (in grams).

t-Test: Two-Sample Assuming Equal Variances

	Normal Air	*Enriched Air*
Mean	4.1633	5.1050
Variance	0.9143	2.5915
Observations	12	8
Pooled Variance	1.5665	
Hypothesized Mean Difference	0	
df	18	
t Stat	-1.6484	
P(T<=t) one-tail	0.0583	
t Critical one-tail	2.2137	
P(T<=t) two-tail	0.1166	
t Critical two-tail	2.5524	

Figure 1

Based on the **Figure 1**,

i) show that there is no difference in variability of the population plants' weights at 98% confidence level.

[8 Marks]

ii) test whether a CO_2-enriched atmosphere increases the average of plant growth at 0.02 level of significance by using the *P*-value approach.

[5 Marks]

QUESTION 3 [15 MARKS]

A botanist is interested in learning four different conditions of exposure to sunlight (none, low, medium and high) and watering frequency (daily and weekly) that affect plant growth. She plants 40 seeds and observe the growth of the plants for two months. Then, she records the height (in cm) of each plant as in **Table 2**.

Table 2

Watering Frequency	Sunlight Exposure			
	None	Low	Medium	High
Daily	12.19	12.70	16.26	16.00
	11.18	13.21	15.75	16.26
	8.13	14.22	11.94	14.22
	9.91	10.92	13.97	12.19
	11.18	12.19	14.73	14.73
Weekly	11.18	12.45	14.73	15.24
	10.67	13.46	15.75	12.45
	9.65	14.48	16.00	11.68
	9.40	13.72	16.51	14.22
	9.91	12.19	13.97	13.97

i) State the response variable.

[1 Mark]

ii) **Figure 2** shows the output of analysis of variance (ANOVA) for the growth of the plants for two months. Find the values of **X** and **Y** in **Figure 2**.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.0016	1	0.0016	0.0009	0.9760	4.1491
Columns	121.0627	3	40.3542	Y	0.0000	2.9011
Interaction	6.5210	3	2.1737	1.2415	0.3109	2.9011
Within	56.0257	32	X			
Total	183.6109	39				

Figure 2

[4 Marks]

iii) Given that there is no interaction effect between the conditions of exposure to sunlight and watering frequency on the plant growth. Test the marginal effect of the conditions of exposure to sunlight and watering frequency at 5% significance level.

[10 Marks]

QUESTION 4 [16 MARKS]

A manufacturer of an electronic device wishes to produce a device that can be used in a very wide temperature range. It is known that increasing in temperature shortens the life time of the device. Thus, a study is performed and the summarized values for the data are given as follows:

$$\hat{\beta}_0 = 453.5556, S_{xy} = -31880, S_{xx} = 6000, \sum y^2 = 489857, \sum y = 1691 \text{ and } n = 9$$

i) Determine the dependent and independent variables.

[1 Mark]

ii) **Figure 3** presents the scatter diagram for the two variables. Give comment on the trend of the scatter diagram.

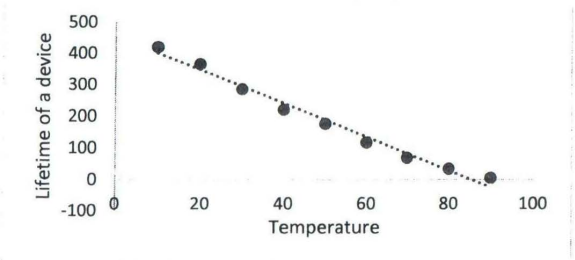

Figure 3

[1 Mark]

iii) Write the estimated regression model.

[3 Marks]

iv) Based on your answer in **iii)**, interpret the value of the regression coefficient obtained.

[1 Mark]

v) Test the linear relationship between the range of the temperature and the lifetime of the device at 2% significance level using t_{test}.

[10 Marks]

QUESTION 5 [23 MARKS]

A researcher conducted an analysis to predict the salary of programmer (in thousands of RM per year) based on number of years of experience of the programmer, x_1 and aptitude test score, x_2 using a software. The researcher collected the data from a sample of 20 programmers and generated the output for the analysis using *Microsoft Excel* as in **Figure 4**.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9133
R Square	0.8342
Adjusted R Square	0.8147
Standard Error	2.4188
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	500.3285	250.1643	42.7601	0.0000
Residual	17	99.4570	5.8504		
Total	19	599.7855			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.1739	6.1561	0.5156	0.6128	-9.8142	16.1621	-9.8142	16.1621
Experience	1.4039	0.1986	7.0702	0.0000	0.9850	1.8228	0.9850	1.8228
Test Score	0.2509	0.0774	3.2433	0.0048	0.0877	0.4141	0.0877	0.4141

Figure 4

i) What is the main difference between simple linear regression analysis and multiple linear regression analysis?

[1 Mark]

ii) What is the dependent variable in this regression analysis?

[1 Mark]

iii) Test the hypothesis that at least one of the independent variables has a significant relationship with the dependent variable at 2.5% significance level. Use the *P*-value method.

[5 Marks]

iv) State the coefficient of determination and interpret the value.

[2 Marks]

v) Identify significant independent variables using 2.5% significance level.

[4 Marks]

vi) Interpret the coefficient associated with the test score variable and the coefficient of the regression intercept.

[4 Marks]

vii) **Table 3** displays an incomplete summary of the multiple linear regression analysis. Complete the table for all predictors involved.

Table 3

Predictor	*P*-value	*r*	r^2	Adjusted r^2	Regression model
x_1	0.0000	0.8553	0.7316	0.7167	$\hat{y} = 22.8111 + 1.6200\,x_1$
x_2	0.0063	0.5887	0.3466	0.3103	$\hat{y} = -4.7097 + 0.4344\,x_2$
x_1, x_2					

[2 Marks]

viii) Determine the best regression model to predict the salary of programmer. Give the appropriate reasons for the chosen model.

[2 Marks]

ix) Suppose a programmer scores 90 in an aptitude test and currently earns a salary of RM40,000 per year. Based on your answer in **viii)**, predict how many years of experience he has.

[2 Marks]

QUESTION 6 [12 MARKS]

The proportions of motorcycle accidents at a certain district for a particular month of January, February, March, and April in year 2023 are known to be in the ratio 4:3:2:1, respectively. **Table 4** shows a sample of 100 accidents occurred in the four particular months.

Table 4

Month	January	February	March	April	Total
No. of accidents	30	**A**	**B**	20	100

i) Given that the number of accidents occurred in February and March are equal, find the values of **A** and **B**.

[2 Marks]

ii) Test the hypothesis that the proportions of number of motorcycle accidents do not differ significantly from those with the stated ratio at 0.5% level of significance.

[9 Marks]

iii) Based on your result in **ii)**, does the data shows a good fit?

[1 Mark]

QUESTION 7 [13 MARKS]

(a) State **ONE** difference between contingency table and linear regression analyses.

[1 Mark]

(b) A sports teacher conducts a survey on the sport preference among students in primary and secondary schools. **Table 5** shows the data of 160 students from both schools according to their sports preference.

Table 5

	Archery	Football	Cycling
Primary school	40	10	20
Secondary school	10	**W**	30

i) Find the value of **W** in **Table 5**.

[2 Marks]

ii) Test the claims that the sport preference is independent of school level using 0.01 significance level.

[10 Marks]

END OF QUESTION PAPER