UMP

اونيۏرسيتي مليسيا ڤهڠ

UNIVERSITI MALAYSIA PAHANG

FINAL EXAMINATION

COURSE	:	**APPLIED STATISTICS**
COURSE CODE	:	**BUM2413/BUM2443**
COURSE COORDINATOR	:	**NOR HAFIZAH BINTI MOSLIM**
DATE	:	**3 JULY 2023**
DURATION	:	**3 HOURS**
SESSION/SEMESTER	:	**SESSION 2022/2023 SEMESTER II**

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.
4. All calculations must be in **FOUR (4) decimal places**.

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 [14 MARKS]

Two types of instruments for measuring the amount of sulfur monoxide (measured in parts per million, ppm) in the atmosphere are being used in an air-pollution experiment. **Table 1** shows the numerical summary for the recorded data.

Table 1

Instrument	**Sample size,** n	**Sample mean,** $\bar{x}$	**Sample standard deviation,** s
A	9	0.7456	0.1041
B	9	0.6711	0.1117

i) A researcher wishes to determine whether the recorded data have the same variability. Assuming the data are approximately normally distributed, perform an appropriate test at $\alpha = 0.10$.

[7 Marks]

ii) The manufacturer of **instrument A** claims that the standard deviation of the data using their instrument should not be more than 0.1 ppm. Is there enough evidence to support the claim at $\alpha = 0.05$?

[7 Marks]

QUESTION 2 [5 MARKS]

The management of a taxi cab company is trying to decide if they should switch from bias tires to radial tires in order to improve fuel economy (measured in km/liter). Ten cabs were first fitted with bias tires and driven on a test course. Without changing drivers, tires of the same ten cabs were then switched to radial tires and the test course was repeated. **Figure 1** presents the *Microsoft Excel* output for the related test on the fuel economy for the 10 cabs at 5% significance level.

t-Test: Paired Two Sample for Means

	Bias tires	*Radial tires*
Mean	10.2200	10.8100
Variance	0.0551	0.0277
Observations	10	10
Pearson Correlation	-0.6602	
Hypothesized Mean Difference	0	
df	9	
t Stat	-5.0905	
P(T<=t) one-tail	0.0003	
t Critical one-tail	1.8331	
P(T<=t) two-tail	0.0007	
t Critical two-tail	2.2622	

Figure 1

The management does not want to switch unless a hypothesis test shows that the average mileage is improved. Should the management switch to radial tires? Based on **Figure 1**, perform the necessary test using the p-value method.

[5 Marks]

QUESTION 3 [16 MARKS]

A medical researcher conducts a clinical trial to determine the effects of four types of drugs (A, B, C and D) in reducing LDL cholesterol of male and female patients. He selects 16 male patients and 16 female patients to participate in this clinical trial. Both male and female patients are divided equally into four groups, respectively, for each type of drug. The LDL cholesterol levels (in *mmol/L*) are then measured one month after the patients have taken the drug. **Figure 2** shows the ANOVA output for the recorded measurements.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.3828	1	0.3828	Z	0.2060	4.2597
Columns	0.2134	3	0.0711	0.3140	0.8151	3.0088
Interaction	0.5059	3	0.1686	0.7444	0.5362	3.0088
Within	5.4375	X	Y			
Total	6.5397	31				

Figure 2

i) State the independent and dependent variables.

[3 Marks]

ii) How many treatments are involved? List down all the treatments.

[2 Marks]

iii) Find the values of **X**, **Y** and **Z**.

[4 Marks]

iv) Is there any interaction effect between type of drugs and gender on the LDL cholesterol level?

[5 Marks]

v) Do we need to test for marginal effect? Give a reason.

[2 Marks]

QUESTION 4 [20 MARKS]

A car rental company is interested to increase its revenue by increasing the number of cars available for rent. The marketing department conducts a study and records the number of cars (in ten thousand) available for rent and its corresponding revenue (in RM billion) as shown in **Table 2**.

Table 2

Number of cars (in ten thousand)	Revenue (in RM billion)
10.5	1.6
13.8	2.1
20.7	2.3
25.6	2.1
32.6	2.4
38.8	3.4
42.1	3.1
49.9	3.5
55.6	4.1
60.1	3.6

i) Given that $\sum x = 349.7$, $\sum y = 28.2$, $\sum x^2 = 14918.53$, $\sum y^2 = 85.62$ and $\sum xy = 1106.79$, find the correlation coefficient.

[8 Marks]

ii) Based on your answer in **i)**, what is the trend of revenue when the number of cars increases?

[1 Mark]

iii) Given that $\hat{y} = 1.2515 + 0.0449x$, interpret the slope of the regression equation.

[1 Mark]

iv) Predict the revenue if the number of cars available for rent is 786,000.

[2 Marks]

v) Test the relationship between number of cars available for rent and revenue at 5% significant level.

[8 Marks]

QUESTION 5 [20 MARKS]

A real estate agency wants to predict the prices of properties in a particular city based on various features. The agency collects a dataset of 24 properties in the city, each with their features and corresponding prices. The dataset contains the following features such that prices of properties (in RM thousand), number of bedrooms, *R*, size of the property, *S* (in square feet), and distance from city centre, *T* (in kilometres). The *Microsoft Excel* output of multiple linear regression analysis is given in **Figure 3**.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8697
R Square	0.7564
Adjusted R Square	0.7199
Standard Error	220.9736
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	3032609.07	1010870	20.7021	0.0000
Residual	20	976586.7631	48829.34		
Total	23	4009195.833			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-155.4333	304.2201	-0.5109	0.6150	-790.0254	479.1588	-790.0254	479.1588
R	3.1354	81.8342	0.0383	0.9698	-167.5677	173.8386	-167.5677	173.8386
S	0.3977	0.1049	3.7916	0.0011	0.1789	0.6165	0.1789	0.6165
T	-9.6819	26.6360	-0.3635	0.7201	-65.2436	45.8798	-65.2436	45.8798

Figure 3

i) State and interpret the coefficient of determination value.

[2 Marks]

ii) Interpret the distance from city centre coefficient, *T*.

[2 Marks]

iii) Based on the *p*-value, is there any evidence to conclude that at least one of the independent variables is related to the dependent variable?

[5 Marks]

iv) Based on the coefficient table, determine the significance of predictors which gives the best fit to the data at 5% significance level.

[3 Marks]

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

[4 Marks]

Table 3

Predictor	*P*-value	*r*	r^2	Adjusted r^2	Regression model
R, S, T		0.8697			
R, S	0.0000	0.8687	0.7548	0.7315	$\hat{y} = -246.2562 + 0.4096S + 9.1297R$
R, T	0.0001	0.7624	0.5813	0.5415	$\hat{y} = 55.5639 + 227.7763R - 41.1427T$
S, T	0.0000	0.8697	0.7563	0.7333	$\hat{y} = -149.4361 + 0.4006S - 9.8876R$
R	0.0000	0.7410	0.5491	0.5287	$\hat{y} = -343.2065 + 55.5510R$
S	0.0000	0.8687	0.7546	0.7435	$\hat{y} = -233.91 + 0.4192S$
T	0.0013	0.6189	0.3830	0.3550	$\hat{y} = 1206.8456 - 107.6811T$

vi) Determine the best regression model to predict the prices of properties in a particular city based on various features. Give the appropriate reasons for the chosen model.

[2 Marks]

vii) Predict the prices of properties in a particular city based on 5 bedrooms, 2670 square feet in property size and 3.5 km from city centre using the best regression model obtain in **vi)**.

[2 Marks]

QUESTION 6 [12 MARKS]

The customer service department of a telco company received a lot of complaints regarding the product issues. The number of complaint letters received in a day for 100 days is presented in **Table 4**.

Table 4

No. of complaint letters	0	1	2	3	4
No. of days	30	19	24	**P**	10
Probability	0.2491	0.3460	0.2406	0.1256	0.0387

i) Find the value of **P**.

[2 Marks]

ii) Test the hypothesis that the number of complaint letters received per day follows Poisson distribution at 1% significance level.

[10 Marks]

QUESTION 7 [13 MARKS]

a) Give one difference between contingency table and linear regression.

[2 Marks]

b) A manufacturing factory that produces electronic components operates three machines with two work shifts. Machine breakdown always happen during the operation. Data in **Table 5** shows 100 breakdowns with respect to the machine type and shift.

Table 5: Number of breakdowns with respect to the machine type and shift

		Shift	
		Day	**Night**
Machine	**A**	15	18
	B	10	12
	C	**Q**	25

i) Find the value of **Q**.

[2 Marks]

ii) Would you conclude that type of machine and work shift are independent on the number of machine breakdowns at $\alpha = 0.025$.

[9 Marks]

END OF QUESTION PAPER