



### FINAL EXAMINATION

<b>COURSE</b>	<b>:</b>	<b>APPLIED STATISTICS</b>
<b>COURSE CODE</b>	<b>:</b>	<b>BUM2413/BUM2443</b>
<b>COURSE COORDINATOR</b>	<b>:</b>	<b>FARAHANIM BINTI MISNI</b>
<b>DATE</b>	<b>:</b>	<b>6 FEBRUARY 2023</b>
<b>DURATION</b>	<b>:</b>	<b>3 HOURS</b>
<b>SESSION/SEMESTER</b>	<b>:</b>	<b>SESSION 2022/2023 SEMESTER I</b>

#### **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 calculations and assumptions must be clearly stated.
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 and Formulae.

#### **APPENDIX:**

1. None

---

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

---

This examination paper consists of **TEN (10)** printed pages including the front page.

**QUESTION 1 [6 MARKS]**

Self-parking technologies in automobile has gained interest in automotive manufacturing. However, drivers are still skeptical and are not confident to use the technology. It is estimated (with 92% confidence level) that the true percentage of Malaysian car drivers who do not trust the self-parking technology lies within 53.8% and 72.2%.

- i) State the estimator used to estimate the true percentage of Malaysian car drivers who do not trust the self-parking technology.

**[1 Mark]**

- ii) Is there enough evidence to claim that the true proportion of Malaysian drivers who do not trust the self-parking technology is 0.75? Use confidence interval approach to test the claim.

**[5 Marks]**

**QUESTION 2 [14 MARKS]**

An experiment was conducted to test the effectiveness of nicotine-patch therapy for smoking cessation in schizophrenia patients. A sample of 41 patients involved in this experiment were divided into two groups, the nicotine-patch group and the control group. The nicotine-patch group used nicotine patch for a duration of eight weeks. At the end of the eight weeks, the carbon monoxide (CO) level was measured for both groups. A statistical analysis was performed for the recorded CO level (in ppm) and the *Microsoft Excel* output for the analysis is given in **Figure 1**.

t-Test: Two-Sample Assuming Equal Variances

	<i>Nicotine-patch</i>	<i>Control</i>
Mean	23.8235	25.7917
Variance	33.7794	46.5199
Observations	17	24
Pooled Variance	41.2930	
Hypothesized Mean Difference	0	
df	39	
t Stat	-0.9662	
P(T<=t) one-tail	0.1700	
t Critical one-tail	1.6849	
P(T<=t) two-tail	0.3399	
t Critical two-tail	2.0227	

**Figure 1**Based on **Figure 1**,

- i) conduct a hypothesis testing to show that the population variances for both groups are equal at 10% significance level. Use traditional approach.

**[8 Marks]**

- ii) does the data indicate that the true mean of CO level for nicotine patch group is different from control group? Use the *P*-value approach at 3% significance level.

**[5 Marks]**

- iii) based on your answer in **ii**), is it possible for the mean of CO level for nicotine patch group to be lower than the control group?

**[1 Mark]**

**QUESTION 3 [15 MARKS]**

Pineapple pulp is believed to reduce the severity of influenza among children due to its high content of vitamin C. Three types of supplements are prepared, each containing either a placebo, a low dose, or a high dose of pineapple pulp. The level of severity of influenza to children may differ based on their gender and type of supplements. Suppose we select 15 children from each gender, who are equally divided into three groups for each supplement. The number of days that children had severe influenza symptoms were recorded. The analysis of variance (ANOVA) was obtained using *Microsoft Excel* and shown in **Figure 2**.

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	16.2000	2	8.1000	6.9427	0.0042	3.4028
Columns	2.1333	1	2.1333	1.8285	0.1889	4.2597
Interaction	0.8667	2	0.4334	Z	0.6936	3.4028
Within	28.0000	X	Y			
Total	47.2000	29				

**Figure 2**

Based on the study,

- i) list down the factors involved.

**[2 Marks]**

- ii) how many treatments involved? List down all the treatments.

**[2 Marks]**

- iii) find the value of **X**, **Y** and **Z**.

**[4 Marks]**

- iv) is there any interaction effect between gender and type of supplements on the number of days the children had severe influenza?

**[5 Marks]**

- v) do we need to test for the marginal effects? Give a reason.

[2 Marks]

#### QUESTION 4 [20 Marks]

A large manufacturing company that produces air filter claimed that the number of air filters sold is related to the reading of air quality index (AQI). A researcher conducted a study and recorded the reading of AQI for each month as shown in **Table 1**:

**Table 1**

Month	Air Quality Index (AQI)	Sale (in thousands unit)
January	61	1.5
February	60	1.6
March	64	1.7
April	67	2.0
May	63	2.4
June	65	2.3
July	81	2.7
August	94	2.9
September	123	3.1
October	120	3.0

- i) State the dependent and independent variables.

[1 Mark]

- ii) Given that  $\sum x = 798$ ,  $\sum y = 23.2$ ,  $\sum x^2 = 69026$ ,  $\sum y^2 = 57.06$  and  $\sum xy = 1963.6$ , find the correlation coefficient.

[8 Marks]

- iii) Based on your answer in **ii)**, what is the trend of sales number for air filter?

[1 Mark]

- iv) Given that  $\hat{y} = 0.6442 + 0.021x$ . If the reading of AQI in November is 78, predict the number of air filters which would be sold.

[2 Marks]

- v) Use the information from the regression equation in **iv)** to test the relationship between AQI and number of sales using 0.5% significant level.

[8 Marks]

**QUESTION 5 [19 MARKS]**

Lake eutrophication refers to an ecological state where proliferation of algal occurs in a lake due to an increase in nutrient loading. A researcher is interested to investigate the factors increasing the risk of eutrophication,  $R$  (in %) in some lakes in Minnesota. He collected data on average lake depth,  $D$  (in m), external phosphorus loading rate,  $L$  (in  $\mu\text{g/L/d}$ ) and flushing rate,  $F$  (in  $10^{-3} \text{ d}^{-1}$ ). The statistical analysis at 1% significance level using *Microsoft Excel* is shown in **Figure 3**.

## SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8736
R Square	0.8287
Adjusted R Square	0.7382
Standard Error	8.7826
Observations	13

## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	36355.5058	12118.5019	8.0570	0.0064
Residual	9	13536.8019	1504.0891		
Total	12	49892.3077			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 99.0%</i>	<i>Upper 99.0%</i>
Intercept	-7.9239	14.3717	-1.5389	0.1582	-21.2668	17.2991	-19.8956	21.2991
D	3.5368	1.1948	2.9602	0.0160	0.8340	6.2395	0.8539	6.5239
L	3.7345	1.1069	3.3738	0.0082	1.2305	6.2385	1.8231	7.9877
F	-2.1661	0.8518	-2.5430	0.0316	-4.0929	-0.2392	-2.0986	-0.0674

**Figure 3**

Based on **Figure 3**, answer the following questions.

- i) State and interpret the coefficient of determination value.

[2 Marks]

- ii) Interpret the coefficient of the flushing rate.

[2 Marks]

- iii) Is there any evidence to conclude that at least one independent variable is related to the dependent variable?

[5 Marks]

- iv) **Table 2** displays an incomplete summary of the multiple linear regression analysis.

**Table 2**

Predictor	<i>P</i> -value	$r^2$	Adjusted $r^2$	Regression model
<i>D</i>	0.0001	0.9146	0.9069	$\hat{R} = -8.3700 + 2.4312D$
<i>L</i>	0.0000	0.9237	0.9188	$\hat{R} = 41.3200 + 38.9761L$
<i>F</i>	0.0038	0.4597	0.4435	$\hat{R} = 17.9800 - 3.1845F$
<i>D, L</i>	0.0000	0.9311	0.9250	$\hat{R} = -4.1900 + 4.1846D + 2.3189L$
<i>D, F</i>	0.0001	0.8643	0.8574	$\hat{R} = -9.5700 + 1.8871D - 2.9861F$
<i>L, F</i>	0.0000	0.9003	0.8921	$\hat{R} = 37.1800 + 28.2512L - 1.8934F$
<i>D, L, F</i>				

Based on **Figure 3**, complete the table for the three predictors.

[4 Marks]

- v) Based on **Table 2**, choose the most appropriate model if one predictor variable is selected.

[1 Marks]

- vi) Then, select the best regression model for predicting risk of lake eutrophication ( $\hat{R}$ ) at 1% significance level. Justify your answer.

[3 Marks]

- vii) Lake Erie with an average lake depth of 19 m in the North America receives external phosphorus loading of 0.125  $\mu\text{g/L/d}$ . The flushing rate of Lake Erie is 0.00105  $\text{d}^{-1}$ . Predict the risk of lake eutrophication in Lake Erie by using the regression model chosen in v).

[2 Marks]

### QUESTION 6 [13 MARKS]

- (a) A goodness of fit test is used to analyze one qualitative variable from a single population that can be categorised.

- i) State one purpose of using Chi-squared test in conducting goodness of fit test.

[1 Mark]

- ii) State one assumption that needs to be satisfied for goodness of fit test.

[1 Mark]

- (b) In a previous study, it was reported that 50% of football fans are willing to pay extra to watch FIFA World cup via online streaming. A summary report of the study is given in **Figure 4**. A similar study was recently conducted involving 240 football fans and the data is summarised in **Table 3**.



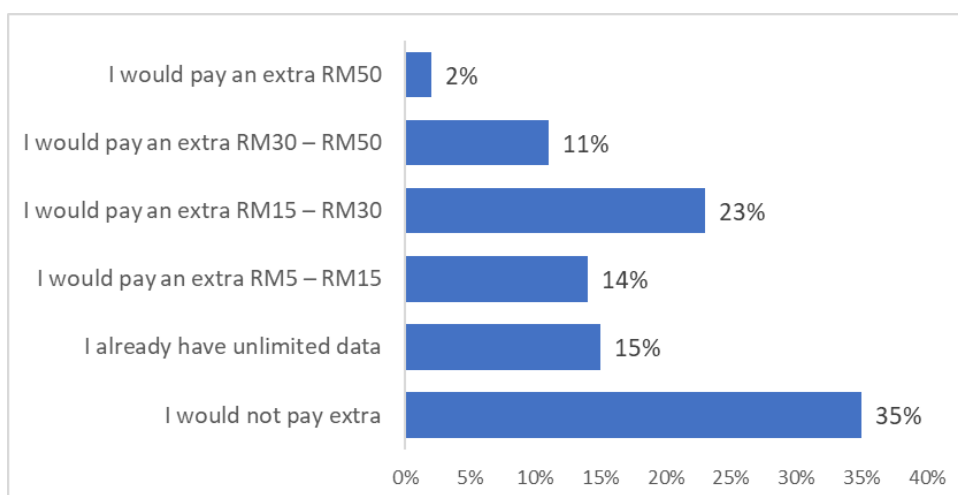


Figure 4

Table 3

Category on willingness to pay extra	Frequency
A: I would not pay extra	73
B: I already have unlimited data	41
C: I would pay an extra RM5 – RM15	40
D: I would pay an extra RM15 – RM30	<b>M</b>
E: I would pay an extra RM30 – RM50	33
F: I would pay an extra RM50	9

- i) Find the value of **M**.

[2 Marks]

- ii) Test the hypothesis that the data from the recent study differs from the previous study. Use 10% significance level.

[9 Marks]

**QUESTION 7 [13 MARKS]**

- (a) Can the result from a contingency table analysis be used for prediction?

[1 Mark]

- (b) A criminologist conducted a survey to determine whether the incidence of certain types of crime varied from different areas of a large city. The particular crimes of interest were assault, burglary and homicide. **Table 4** shows the numbers of crimes committed in three areas of the city during the past years.

**Table 4**

Area	Type of crime		
	Assault	Burglary	Homicide
A	40	60	20
B	50	70	30
C	30	<b>P</b>	<b>Q</b>

- i) Suppose the total crime at area C is 110 cases with equal number of burglary and homicide cases. Find the value of **P** and **Q**.

[3 Marks]

- ii) Can we conclude that at 0.025 level of significance the occurrence of these types of crime is dependent on the city area?

[9 Marks]

**END OF QUESTION PAPER**