اونيۏرسيتـي مليسيا ڤهڠ
UNIVERSITI MALAYSIA PAHANG

FINAL EXAMINATION

COURSE	**:**	**APPLIED STATISTICS**
COURSE CODE	**:**	**BUM2413**
COURSE COORDINATOR	**:**	**NOR HAFIZAH BINTI MOSLIM**
DATE	**:**	**7 FEBRUARY 2022**
DURATION	**:**	**3 HOURS**
SESSION/SEMESTER	**:**	**SESSION 2021/2022 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.
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 **TWELVE (12)** printed pages including the front page.

QUESTION 1 [7 MARKS]

A previous record shows that drivers get an average of 52300 km on a set of Maxim A51 radial tires with a standard deviation of 6400 km. Hoping to improve that figure, the tire company has added a new polymer to the rubber that should help protect the tires from deterioration caused by extreme temperatures. Twenty drivers who tested the new tires have reported getting an average and standard deviation of 54400 km and 5010 km respectively. Assume that the new tire mileage is normally distributed, can the tire company claim that the addition of the new polymer in the new tire produces a more consistent tire mileage? Test at 10% level of significance.

[7 Marks]

QUESTION 2 [13 MARKS]

A research is conducted to investigate the prevalence of airborne bacteria in carpeted and uncarpeted rooms in a hospital. Eight samples of room air from both rooms were pumped over a Petri dish and incubated for three hours before the number of bacteria formed on each Petri dish were measured (in bacteria per cubic foot of air). A statistical analysis of the data using *Microsoft Excel* is illustrated in **Figure 1**. Assume that the number of bacteria has a normal distribution.

t-Test: Two-Sample Assuming Equal Variances

	Carpeted	*Uncarpeted*
Mean	10.95	8.325
Variance	6.5257	15.2193
Observations	8	8
Pooled Variance	10.8725	
Hypothesized Mean Difference	0	
df	14	
t Stat	1.5922	
P(T<=t) one-tail	0.0668	
t Critical one-tail	1.8875	
P(T<=t) two-tail	0.1337	
t Critical two-tail	2.2638	

Figure 1

i) Conduct a necessary analysis to show that the number of bacteria formed in both carpeted and uncarpeted rooms have equal population variances at 4% level of significance. Use confidence interval approach.

[8 Marks]

ii) It is claimed that the carpeted rooms are more likely to accumulate more bacteria compared to uncarpeted rooms. Test this claim at 4% level of significance using *P-value* approach.

[5 Marks]

QUESTION 3 [15 Marks]

An experiment to determine the effect of four different types of engine oil (A, B, C, and D) on the rolling friction coefficient of a car speed has been conducted. Three brands of car (Honda, Toyota, and Mazda) were chosen, and each engine oil was tested twice on each car, producing the following ANOVA output in **Figure 2**.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.00012	3	3.8E-05	1.1203	[redacted]	3.4903
Columns	0.00036	2	0.00018	5.2110	[redacted]	3.8853
Interaction	0.00019	6	3.1E-05	**Y**	[redacted]	**Z**
Within	**W**	**X**	3.4E-05			
Total	0.00107	23				

Figure 2

i) How many treatments involved? Write down all the treatments.

[2 Marks]

ii) Identify the number of replication for each treatment.

[1 Mark]

iii) Based on the ANOVA table above, find the values of **W**, **X**, **Y** and **Z**.

[6 Marks]

iv) Test the interaction effect on rolling friction coefficient of car speed between the four different types of engine oil and the three different brands of car.

[4 Marks]

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

[2 Marks]

QUESTION 4 [20 Marks]

A medical researcher conducted an observational study to understand the recovery rate for patients infected with the COVID-19 in Malaysia. The researcher contacted thirteen COVID-19 survivors and interviewed them regarding their recovery experience. Two of the questions asked are the recovery period (in days) and the number of days that have passed since receiving the second dose of COVID-19 vaccine prior to infection. The recorded data is summarized as

$$\sum x = 1499, \quad \sum y = 133, \quad x^2 = 244523, \quad y^2 = 1431, \quad xy = 16678,$$
$$S_{xx} = 71676.7692, \quad S_{yy} = 70.3077, \quad S_{xy} = 1342.0769$$

i) Identify the dependent variable in the study.

[1 Mark]

ii) Calculate the correlation coefficient and interpret its value.

[3 Marks]

iii) Estimate the regression model parameters and write the estimated linear regression model.

[5 Marks]

iv) Based on your answer in **iii)**, predict the recovery period if a person is infected with COVID-19 after 200 days of receiving second dose of COVID-19 vaccine.

[2 Marks]

v) **Table 1** represents the incomplete ANOVA table of the study. Find the values of **P, Q, R** and **S**.

Table 1

Source of Variations	**Sum of Squares**	**Degrees of Freedom**	**Mean of Squares**	f_{test}
Regression	25.0968	1	**Q**	**S**
Residual	**P**	11	**R**	
Total	70.3077	12		

[4 Marks]

vi) Test the linearity between the two variables at 4% significance level.

[5 Marks]

QUESTION 5 [20 Marks]

A distributor of frozen dessert pies wants to develop a new product. Before that the general director (GD) needs to evaluate the factors influencing the demand of frozen pies. The company collected the data of weekly pie sales (in unit sold), the unit price, *P* (in RM), the investment in advertising, *A* (in RM hundreds) and the number of good reviews received, *G*. The complete statistical analysis at 8% significance level using *Microsoft Excel* are presented in **Figure 3 to Figure 9**.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.4433
R Square	0.1965
Adjusted R Square	0.1347
Standard Error	59.0911
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	11100.4380	11100.4380	3.1790	0.0979
Residual	13	45392.8953	3491.7612		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	558.2772	90.4410	6.1728	0.0000	362.8912	753.6633	386.5411	730.0134
Price	-24.0339	13.4796	-1.7830	0.0979	-53.1547	5.0870	-49.6299	1.5621

Figure 3

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.5563
R Square	0.3095
Adjusted R Square	0.2564
Standard Error	54.7787
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	17484.2225	17484.2225	5.8267	0.0313
Residual	13	39009.1108	3000.7008		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	147.6994	105.2007	1.4040	0.1838	-79.5728	374.9716	-52.0635	347.4622
Advertising	72.3086	29.9556	2.4139	0.0313	7.5934	137.0238	15.4267	129.1906

Figure 4

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.6688
R Square	0.4473
Adjusted R Square	0.4048
Standard Error	49.0070
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	25271.3588	25271.3588	10.5223	0.0064
Residual	13	31221.9745	2401.6903		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	304.3631	31.8948	9.5427	0.0000	235.4586	373.2675	243.7989	364.9272
Good reviews	17.3726	5.3556	3.2438	0.0064	5.8025	28.9427	7.2030	27.5423

Figure 5

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7221
R Square	0.5215
Adjusted R Square	0.4417
Standard Error	47.4634
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	29460.0269	14730.0134	6.5386	0.0120
Residual	12	27033.3065	2252.7755		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	306.5262	114.2539	2.6829	0.0199	57.5883	555.4640	88.0370	525.0154
Price	-24.9751	10.8321	-2.3057	0.0398	-48.5763	-1.3739	-45.6895	-4.2607
Advertising	74.1310	25.9673	2.8548	0.0145	17.5530	130.7089	24.4733	123.7886

Figure 6

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7232
R Square	0.5230
Adjusted R Square	0.4435
Standard Error	47.3897
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	29543.9117	14771.9559	6.5776	0.0118
Residual	12	26949.4216	2245.7851		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	184.2708	92.3686	1.9949	0.0693	-16.9832	385.5248	7.6330	360.9086
Advertising	40.4603	29.3339	1.3793	0.1930	-23.4527	104.3732	-15.6353	96.5558
Good reviews	13.5843	5.8621	2.3173	0.0390	0.8119	26.3567	2.3741	24.7945

Figure 7

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7466
R Square	0.5574
Adjusted R Square	0.4836
Standard Error	45.6475
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	31488.9837	15744.4919	7.5560	0.0075
Residual	12	25004.3496	2083.6958		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	433.5092	80.4492	5.3886	0.0002	258.2254	608.7930	279.6650	587.3534
Price	-18.2670	10.5748	-1.7274	0.1097	-41.3076	4.7735	-38.4894	1.9553
Good reviews	15.8470	5.0661	3.1281	0.0087	4.8090	26.8850	6.1591	25.5349

Figure 8

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8124
R Square	0.6601
Adjusted R Square	0.5674
Standard Error	41.7825
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	37289.7459	12429.9153	7.1200	0.0063
Residual	11	19203.5874	1745.7807		
Total	14	56493.3333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 92.0%*	*Upper 92.0%*
Intercept	308.6138	100.5837	3.0682	0.0107	87.2305	529.9971	114.6454	502.5821
Price	-20.5602	9.7609	-2.1064	0.0589	-42.0437	0.9233	-39.3833	-1.7371
Advertising	47.5406	26.0806	1.8228	0.0956	-9.8623	104.9436	-2.7539	97.8351
Good reviews	11.2042	5.2906	2.1178	0.0578	-0.4403	22.8486	1.0017	21.4067

Figure 9

Based on **Figure 9**,

i) state the coefficient of determination and interpret its value.

[2 Marks]

ii) can the distributor conclude that not all independent variables in the analysis related to the dependent variable?

[5 Marks]

iii) write a multiple linear regression model in predicting the weekly pie sales by considering the three variables.

[1 Mark]

iv) interpret the coefficient of investment in advertising.

[2 Marks]

v) identify the independent variable that should not be considered in constructing the linear regression model for predicting the weekly pie sales.

[3 Marks]

vi) Based on **Figure 3** to **Figure 9**, complete the summary of the regression analysis for the study in **Table 2**.

Table 2

Predictor(s)	***P*-value**	r^2	**Adjusted** r^2	**Regression model**
P	0.0979	0.1965	0.1347	$\hat{y} = 558.2772 - 24.0339P$
A				
G	0.0064	0.4473	0.4048	$\hat{y} = 304.3631 + 17.3726G$
P, A				
A, G	0.0118	0.5230	0.4435	$\hat{y} = 184.2708 + 40.4603A + 13.5843G$
P, G	0.0075	0.5574	0.4836	$\hat{y} = 433.5092 - 18.2670P + 15.8470G$
P, A, G				

[3 Marks]

vii) Determine the best regression model to predict the weekly pie sales. Justify your answer.

[2 Marks]

viii) Based on your regression model in **vii)**, how many pie will be sold if the price per unit is RM5.50, the distributor invest RM300 in advertising and they received 15 good reviews?

[2 Marks]

QUESTION 6 [12 Marks]

Dr. Atikah is an editor of Data Analytics and Applied Mathematics (DAAM) journal. In order to maintain the quality of the journal content, she is responsible to check the format of the journal prior to publication. **Table 3** shows the number of mistakes per manuscript observed in the journal for the first three editions.

Table 3

No. of mistakes	0	1	2	3	4
No. of manuscript	211	90	19	5	0
Probability	0.6703	0.2681	0.0536	0.0072	0.0008

Test that the number of mistakes per manuscript follows a Poisson distribution with $\lambda = 0.4$ at 1% level of significance.

[12 Marks]

QUESTION 7 [13 Marks]

A test of body reaction to the Malaysian citizens who have taken their vaccine was conducted. Selected citizens who got side effect are categorised according to their blood type and the type of vaccine received. From a sample of 250 citizens, 150 of them were administered with Sinovac while the rest received AstraZeneca. **Table 4** shows the number of citizens who had side effect after being vaccinated.

Table 4

Types of Vaccine	**Blood type**			
	A	**B**	**AB**	**O**
Sinovac	50	**M**	50	20
AstraZeneca	30	20	20	**N**

i) Complete **Table 4** by finding the values of **M** and **N**.

(1 Mark)

ii) Does different blood type among people who had side effect relate to the type of vaccine received at $\alpha = 0.025$?

(12 Marks)

END OF QUESTION PAPER