UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

FINAL EXAMINATION

COURSE	**:**	**APPLIED STATISTICS**
COURSE CODE	**:**	**BUM2413**
COURSE COORDINATOR	**:**	**MS. NUR ZAHIRAH BINTI MD NOOR**
DATE	**:**	**24 JUNE 2024**
DURATION	**:**	**3 HOURS**
SESSION/SEMESTER	**:**	**SESSION 2023/2024 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 starts on a new page.
3. All calculations and assumptions must be clearly stated, and in **FOUR (4) decimal places**.
4. Candidates are not allowed to bring any material 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 **ELEVEN (11)** printed pages including the front page.

QUESTION 1 [13 MARKS]

The Malaysian Nurses Association conducted a survey on women who received inadequate prenatal care. Mothers were selected based on birth certificate information and categorized into two groups:

Group 1: 14 mothers reporting 5 or fewer prenatal visits.
Group 2: 14 mothers reporting 6 or more prenatal visits.

The survey involved sampling birth weights (in kg) from babies born to these two independent populations, both assumed to follow a normal distribution. **Figure 1** illustrates the hypothesis testing results for comparing two population variances of the birth weights from Group 1 and Group 2. The statistical analysis, conducted with a significance level of 5% provides insights into the average and variability comparison between the groups.

F-Test Two-Sample for Variances

	Group 1	*Group 2*
Mean	2.6022	3.4838
Variance	0.5130	0.4069
Observations	14	14
df	13	13
F	1.2608	
P(F<=f) one-tail	0.3411	
F Critical one-tail	2.5769	

Figure 1

i) Based on **Figure 1**, can we conclude that there is a difference in the population variance of the birth weights from Group 1 and Group 2? (Use the *P*-value approach).

ii) Based on the results from **i)**, conduct appropriate hypothesis testing to determine whether the population mean birth weights of babies from Group 1 is significantly lower than Group 2.

[13 Marks]

[CO2, PO3, C4]

QUESTION 2 [7 MARKS]

A quality controller at Factel, a Malaysian yogurt producer, assesses the variability of sugar content in their product. To comply with company regulations, the variance of sugar content should ideally be 0.5. In a recent study, a random sample of 20 Factel yogurt servings weighing 470g each was selected.

i) Based on the sample, the 90% confidence interval for the variance of sugar content in the Factel's yogurt servings lies between 0.3022 and 0.9005. Using the confidence interval approach, test the hypothesis whether the variance of sugar content in Factel's 470g yogurt meets the company's regulation.

ii) Identify the type of error (Type I, Type II or no error) exists in **i)**. Justify your answer.

[7 Marks]

[CO2, PO3, C4]

QUESTION 3 [19 MARKS]

A team of environmental researchers is studying the effects of soil type and fertilizer type on the growth of a particular species of plants. Three types of soil (Sandy, Loamy, and Clay) and three different types of fertilizer (Organic, Inorganic, and Control) were chosen. Three experimental plots are randomly assigned to each combination of soil and fertilizer type, and the height of the plants (in centimeters) is measured after a specified period. The output of analysis of variance (ANOVA) for the height of the plants is shown in **Figure 2**.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	*x*	2	11.1715	113.3947	0.0000	3.5546
Columns	20.0719	2	10.0359	101.8684	0.0000	3.5546
Interaction	3.0726	*y*	0.7681	*z*	0.0008	2.9277
Within	1.7733	18	0.0985			
Total	47.26074	26				

Figure 2

i) Identify the control and response variables.

ii) How many treatments and replications for each treatment involved in this experiment? List all the possible treatments.

iii) Find the values of *x*, *y* and *z* in **Figure 2**.

iv) Test if there is an interaction effect between soil type and fertilizer type on the height of the plants.

v) Based on your answer in **iv)**, does the team of researchers need to conduct the test on the effect of soil type and effect of fertilizer type, respectively? Support your answer with a reason.

[19 Marks]

[CO2, PO3, C4]

QUESTION 4 [16 MARKS]

In laboratory research, the effects of Organophosphate (OP) pesticides on species are investigated due to their widespread use as a pesticide. To assess these effects, an experiment was conducted involving five groups of five female mice of similar age and condition. Each group received different dosages (in mg/kg body weight) of a particular OP pesticide, while one group served as a control with no chemical administrated. The primary measure of response was the activity level in the brain (in moles/litre/minute). It was hypothesized that brain activity would decrease in linear correlation with higher dosages of OP pesticides. **Figure 3** displays the scatter diagram, while **Figure 4** presents the *Microsoft Excel* output analyzed from the collected data during the experiment.

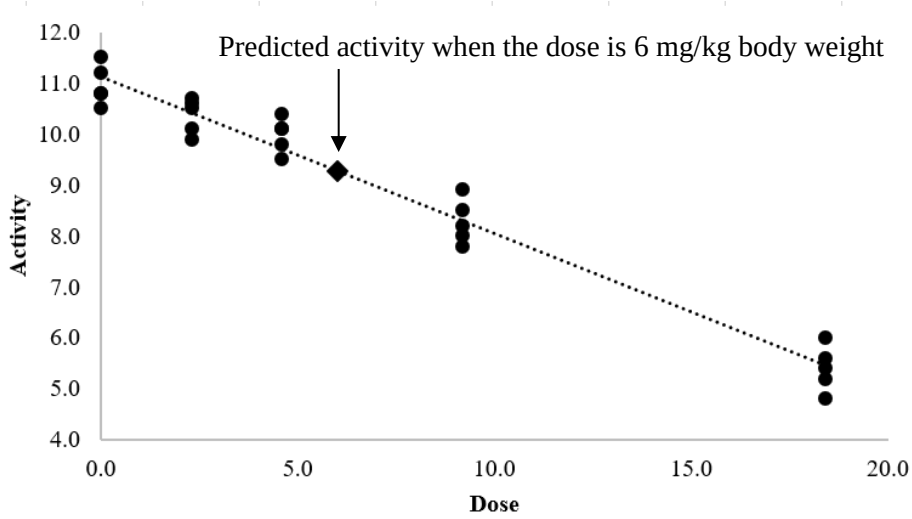

Figure 3

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9823
R Square	0.9649
Adjusted R Square	0.9634
Standard Error	0.3977
Observations	25

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	100.1113	100.1113	632.8580	0.0000
Residual	23	3.6384	0.1582		
Total	24	103.7496			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11.1185	0.1160	95.8845	0.0000	10.8786	11.3584	10.8786	11.3584
Dose	-0.3076	0.0122	-25.1567	0.0000	-0.3329	-0.2823	-0.3329	-0.2823

Figure 4

i) Interpret the correlation coefficient between the level of activity in the brain and the dosage of the pesticide.

ii) Briefly describe whether the answer in **i)** and the hypothesized statement can be supported based on the scatter diagram in **Figure 3**.

iii) Test the linearity between the level of activity in the brain and the dose of pesticide by using the *P*-value approach.

iv) Identify and interpret the coefficient of determination between the level of activity in the brain and the dosage of the pesticides.

v) Predict the level of activity in the brain when the dose is 6 mg/kg body weight by using an appropriate estimated simple linear regression line.

vi) Based on **Figure 3**, do you think the predicted level of activity in the brain in **v)** is reasonable? State your reason.

[16 Marks]

[CO2, PO3, C4]

QUESTION 5 [20 MARKS]

A kinematic viscosity of a certain solvent (in 10^{-6} m^2/s) depends on the ratio of the two solvents and the temperature (in °C). An engineer conducted a number of tests to determine the relationship between these variables. The following *Microsoft Excel* output shown in **Figure 5, 6** and **7** are multiple linear regression analysis done by the engineer.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.5986
R Square	0.3584
Adjusted R Square	0.3292
Standard Error	0.5883
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	4.2525	4.2525	12.288	0.0020
Residual	22	7.6136	0.3461		
Total	23	11.8662			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 98.0%*	*Upper 98.0%*
Intercept	-0.2064	0.3940	-0.5237	0.6057	-1.0236	0.6108	-1.1948	0.7820
ratio	2.0216	0.5767	3.5054	0.0020	0.8256	3.2176	0.5750	3.4681

Figure 5

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.7903
R Square	0.6246
Adjusted R Square	0.6076
Standard Error	0.4500
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7.4118	7.4118	36.607	0.0000
Residual	22	4.4544	0.2025		
Total	23	11.8662			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 98.0%*	*Upper 98.0%*
Intercept	1.7535	0.1406	12.4691	0.0000	1.4619	2.0451	1.4008	2.1062
temperature	-0.0182	0.0030	-6.0504	0.0000	-0.0244	-0.0120	-0.0257	-0.0107

Figure 6

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8998
R Square	0.8097
Adjusted R Square	0.7916
Standard Error	0.3279
Observations	24

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	9.6081	4.8041	44.678	0.0000
Residual	21	2.2580	0.1075		
Total	23	11.8662			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 98.0%*	*Upper 98.0%*
Intercept	0.7007	0.2545	2.7535	0.0119	0.1715	1.2300	0.0600	1.3414
ratio	1.4919	0.3301	4.5195	0.0002	0.8054	2.1784	0.6608	2.3230
temperature	-0.0159	0.0023	-7.0575	0.0000	-0.0206	-0.0112	-0.0215	-0.0102

Figure 7

Based on **Figure 7**, answer question **i)** to **vi)**.

i) Identify the predictor variables in the analysis.

ii) Determine the number of observations in the analysis.

iii) State the coefficient of determination and interpret its value.

iv) State the estimated regression constant ($\hat{\beta}_0$) and interpret its value.

v) Test at 2% level of significance to determine the linear relationship between predictor variables and response variable involved in the analysis. Use the *P*-value approach.

vi) Determine significant predictors using 2% level of significance.

vii) The summary of the multiple linear regression analysis is given by **Table 1**. Complete the table for the single predictors based on **Figure 5** and **Figure 6**.

Table 1

Predictor	***P*-value**	**R Square**	**Adjusted R Square**	**Regression Equation**
x_1				
x_2				
x_1, x_2	0.0000	0.8097	0.7916	$\hat{y} = 0.7007 + 1.4919x_1 - 0.0159x_2$

viii) Based on your answer in **vii)**, select the best regression model for analyzing the kinematic viscosity of a certain solvent. Justify your answer for the selection.

[20 Marks]

[CO2, PO3, C4]

QUESTION 6 [12 MARKS]

The number of times that a machine malfunction in a week at a factory was observed for 28 weeks. The relative frequency of machine malfunction was recorded in **Table 2**.

Table 2

Number of times a machine malfunction	**0**	**1**	**2**	**3**
Frequency	15	7	3	3
Probability	0.4558	0.3581	0.1407	0.0454

i) Find the average value, λ for the number of times that a machine malfunction in a week at a factory.

ii) Can we conclude that the number of times that a machine malfunction in a week at a factory follows the Poisson distribution at 0.5% significance level?

[12 Marks]

[CO2, PO3, C4]

QUESTION 7 [13 MARKS]

A survey was conducted on shopping preferences between physical and online store for various products. A random sample of 300 people participated in the survey resulted the information shown in **Table 3**.

Table 3

Shopping preferences	**Type of products**			
	Electrical Appliances	**Handphone Accessories**	**Clothing**	**Furniture**
Physical store	36	40	35	30
Online store	55	24	*m*	15

i) Identify one of the variables involved in this survey.

ii) Determine the value of *m* in **Table 3**.

iii) Test at 1% significance level that the proportion of shopping preferences is the same for all types of products.

[13 Marks]

[CO2, PO3, C4]

END OF QUESTION PAPER