



**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 **TEN (10)** printed pages including the front page.

QUESTION 1 [13 MARKS]

A dietitian student from a local university conducted a study to compare the average and variability of weights (in grams) for two flavors of yogurt drink, Low-Fat Strawberry Kudo and Low-Fat Blueberry Kudo. Nine samples of Low-Fat Strawberry Kudo and 13 samples of Low-Fat Blueberry Kudo were randomly selected and purchased. The study, assuming normal distribution weights, used hypothesis testing to compare two population variances as shown in **Figure 1**. It merely provides insights into the weight characteristics of these Kudo varieties for dietary considerations.

F-Test Two-Sample for Variances

	<i>Low-Fat Strawberry Kudo</i>	<i>Low-Fat Blueberry Kudo</i>
Mean	21.0333	20.8923
Variance	0.3675	1.0141
Observations	9	13
df	8	12
F	0.3624	
P(F<=f) one-tail	0.0787	
F Critical one-tail	0.3045	

Figure 1

- i) Based on **Figure 1**, can we conclude that there is a difference in population variance of the weights of both Low-Fat Strawberry Kudo and Low-Fat Blueberry Kudo? (Use the *P*-value approach).
- ii) Based on the results from **i)**, conduct appropriate hypothesis testing to determine whether the population mean weights of Low-Fat Strawberry Kudo is not similar to Low-Fat Blueberry Kudo.

[13 Marks]**[CO2, PO3, C4]**

QUESTION 2 [8 MARKS]

A quality controller at Everedy, a battery company, is evaluating the variability in the lengths of their batteries. To comply with company regulations, the variance of battery length should ideally be 0.4. In the study, a random sample of 20 AAA batteries was selected for analysis.

- i) Based on the sample, the 90% confidence interval for the standard deviation of battery length lies between 0.3394 mm and 0.5859 mm. Using the confidence interval approach, conduct a hypothesis test to determine if the variance complies with the company's regulation.

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

[8 Marks]

[CO2, PO3, C4]

QUESTION 3 [20 MARKS]

Agricultural researchers are conducting an experiment to determine the effects fertilizer type and watering frequency on the yield of a certain crop. They tested on three types of fertilizers (A, B, and C) and three watering frequencies (daily, every three days, and weekly). They want to analyze if there are any significant differences in crop yield (in kilograms) based on these two factors. The output of analysis from *Microsoft Excel* at 2% level of significance is given in **Figure 2**.

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Sample	0.9630	2	0.4815	0.1121	0.8946	4.9001
Columns	442.7407	2	221.3704	51.5259	0.0000	4.9001
Interaction	13.9259	4	3.4815	0.8103	0.5348	3.8369
Within	77.3333	18	4.2963			
Total	534.9630	26				

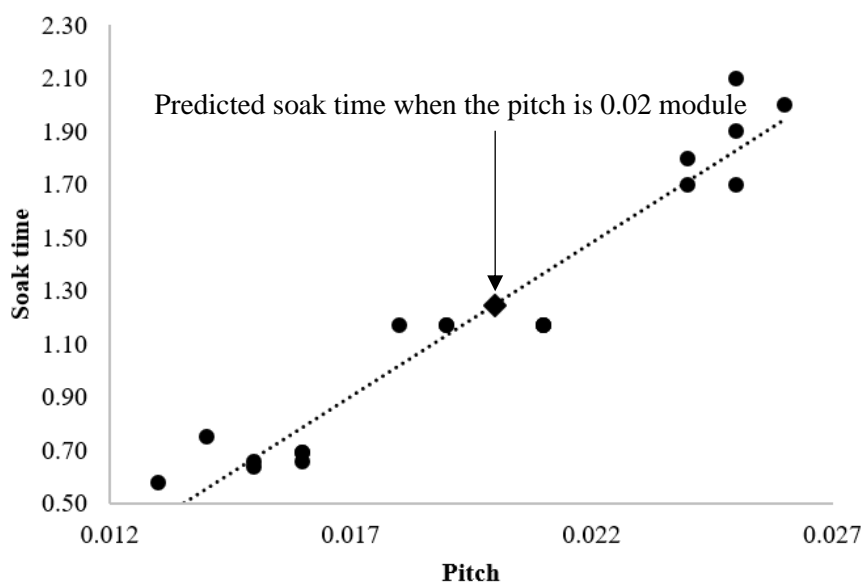
Figure 2

- i) State the independent and dependent variables.
- ii) How many treatments are involved in this experiment? List all possible treatments.
- iii) Test at 2% level of significance whether there is an interaction effect between type of fertilizer and watering frequency using the *P*-value approach.
- iv) Based on your answer in **iii**), conduct the row effect and column effect tests.

[20 Marks]**[CO2, PO3, C4]**

QUESTION 4 [16 MARKS]

Heat treating is a typical method employed in carburizing metal parts such as gears, where the thickness of the carburized layer is crucial for reliability. To ensure the quality, a lab test is conducted on each furnace load. The test is destructive, involving cross-sectioning an actual part and soaking it in a chemical for a specified period. The aim of this lab test is to investigate the dependency of soak time (in seconds) on the pitch (in the module). **Figure 3** displays the scatter diagram, while **Figure 4** presents the *Microsoft Excel* output derived from the sample data collected during the lab test.

**Figure 3**

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.9640
R Square	0.9293
Adjusted R Square	0.9252
Standard Error	0.1414
Observations	19

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	4.4690	4.4690	223.5018	0.0000
Residual	17	0.3399	0.0200		
Total	18	4.8089			

	<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 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-1.0715	0.1557	-6.8829	0.0000	-1.3999	-0.7430	-1.3999	-0.7430
Pitch	115.9468	7.7557	14.9500	0.0000	99.5838	132.3098	99.5838	132.3098

Figure 4

- i) Interpret the correlation coefficient between the soak time and the pitch.
- ii) Briefly describe whether the answer in i) can be supported based on the scatter diagram in **Figure 3**.
- iii) Test the linearity between the soak time and the pitch by using the *P*-value approach.
- iv) Identify and interpret the coefficient of determination between the soak time and the pitch.
- v) Predict the soak time when the pitch is 0.02 module by using an appropriate estimated simple linear regression line.
- vi) Based on **Figure 3**, do you think the predicted soak time in v) is reasonable? State your reason.

[16 Marks]

[CO2, PO3, C4]

QUESTION 5 [18 MARKS]

An engineer performed an experiment to determine an effect of carbon dioxide (CO₂) temperature (in °C), peanut moisture (in %) and peanut particle size (in mm) on the total yield of peanut oil. An experiment was conducted on 20 batches of peanut and a multiple linear regression analysis was performed. The following *Microsoft Excel* output shown in **Figure 5** is a multiple linear regression analysis done by the engineer.

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.8569
R Square	0.7343
Adjusted R Square	0.6845
Standard Error	14.1754
Observations	20

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	3	8887.4643	2962.5	14.7429	0.0001
Residual	16	3215.0857	200.94		
Total	19	12102.55			

	<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 90.0%</i>	<i>Upper 90.0%</i>
Intercept	65.5829	9.3086	7.0454	0.0000	45.8495	85.3163	49.3312	81.8347
CO2 temperature	0.3577	0.0937	3.8153	0.0015	0.1589	0.5564	0.1940	0.5213
peanut moisture	0.6536	0.6562	0.9960	0.3341	-0.7375	2.0446	-0.4921	1.7992
peanut size	-14.1568	2.3689	-5.9760	0.0000	-19.1787	-9.1349	-18.2927	-10.0209

Figure 5

- i) What is the dependent variable in this regression analysis?
- ii) Determine the number of independent variables involved in this analysis.
- iii) State the correlation coefficient and interpret its value.
- iv) State the regression coefficient of peanut size ($\hat{\beta}_3$) and interpret its value.

- v) Test the hypothesis that at least one of the independent variables has a significant relationship with the dependent variable at 10% significance level. Use the P -value method.
- vi) The summary of the multiple linear regression analysis is given in **Table 1**. Complete the table for the three predictors based on **Figure 5**.

Table 1

Predictor	P -value	R Square	Adjusted R Square	Regression Equation
x_1	0.0923	0.1414	0.0937	$\hat{y} = 39.9929 + 0.2643x_1$
x_2	0.7527	0.0057	-0.0496	$\hat{y} = 52.15 + 0.37x_2$
x_3	0.0010	0.4472	0.4165	$\hat{y} = 87.5029 - 11.877x_3$
x_1, x_2	0.2737	0.1414	0.0404	$\hat{y} = 39.9929 + 0.2643x_1 + 0.0000x_2$
x_1, x_3	0.0000	0.7179	0.6847	$\hat{y} = 70.1370 + 0.3732x_1 - 13.7635x_3$
x_2, x_3	0.0031	0.4927	0.4330	$\hat{y} = 78.8550 + 1.0708x_2 - 12.6504x_3$
x_1, x_2, x_3				

- vii) Based on your answer in **vi**), select the best regression model for studying the yield of peanut oil. Justify your answer for the selection.
- viii) Based on your answer in **vii**), predict the yield of peanut oil if the batch of peanut used has a size of 3mm with 17% moisture and set at 75°C CO₂ temperature.

[18 Marks]

[CO₂, PO₃, C₄]

QUESTION 6 [12 MARKS]

A Bluetooth speaker manufacturing company has a defective rate of 7%. The number of defective speakers produced by the company in a week was observed for a year (52 weeks). The relative frequencies and probability for defective speakers are recorded in **Table 2**.

Table 2

Number of defective speakers	0	1	2	3	4
Frequency	26	m	7	3	5
Probability	0.0081	0.0756	0.2646	0.4116	0.2401

- i) Determine the value of m in **Table 2**.
- ii) Can we conclude that the number of defective speakers produced by the company in a week follows the Binomial distribution with parameter $n=4$ and $p=0.07$ at 5% significance level?

[12 Marks]**[CO2, PO3, C4]**

QUESTION 7 [13 MARKS]

A survey was conducted on holiday destinations preferences between man and woman. A random sample of 165 man and 185 woman participated in the survey resulted the information shown in **Table 3**.

Table 3

Gender	Holiday destinations			
	Islands and Beaches	Forests and Mountain	Town and City	Theme Park
Man	34	p	26	50
Woman	70	28	42	45

- i) Identify one of the variables involved in this survey.
- ii) Determine the value of p in **Table 3**.
- iii) Test at 2.5% significance level that the holiday destinations preference is dependent of gender.

[13 Marks]**[CO2, PO3, C4]****END OF QUESTION PAPER**