UNIVERSITI POLYTECH MALAYSIA
UPTM

FINAL EXAMINATION

BACHELOR OF EDUCATION(HONOURS) IN TEACHING ENGLISH AS A SECOND LANGUAGE
BACHELOR OF COMMUNICATION (HONS) IN CORPORATE COMMUNICATION
BACHELOR OF BUSINESS ADMINISTRATION (HONOURS) HUMAN RESOURCE MANAGEMENT
BACHELOR OF BUSINESS ADMINISTRATION (HONOURS)
BACHELOR OF ACCOUNTANCY (HONOURS)

COURSE	:	**STATISTICS FOR SOCIAL SCIENCES**
COURSE CODE	:	**STA2113/STA3073**
DURATION	:	**3 HOURS**

INSTRUCTIONS TO CANDIDATES:

1. This question paper consists of **FOUR (4)** questions.
2. Answer **ALL** questions in the Answer Booklet provided.
3. Please check to make sure that this examination pack consists of:
   i. The Question Paper
   ii. an Answer Booklet
   iii. Appendix 1(1) and 1(2)
4. Do not bring any material into the examination hall. Electronic calculator is allowed.
5. Please write your answer using permanent ink.

MYKAD/ PASSPORT NO : _____

ID. NO. : _____

LECTURER : _____

SECTION : _____

DO NOT OPEN THIS QUESTION PAPER UNTIL YOU ARE TOLD TO DO

This question paper consists of 10 printed pages including the front page

QUESTION 1

a. An energy consultant investigates whether daily appliance usage time (hours) influences household electricity consumption (kWh per day) for 14 households. The data were analyzed using Microsoft Excel, and the regression output is summarized in Tables 1 and 2. Daily appliance usage time ranges from 0.5 to 6 hours per day.

Table 1

Regression Statistics	
Multiple R	0.785
R Square	***X***
Adjusted R Square	0.580
Standard Error	5.200
Observations	14

Table 2

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	78.400	4.200	18.667	0.0000
Usage time(hours)	−3.250	0.900	−3.611	0.003

i. State the dependent and independent variables.

(2 marks)

ii. Name a measure of association between independent and dependent variables. Then, identify and interpret the value from Table 1.

(4 marks)

iii. Name an appropriate graph that can be used to illustrate the relationship between appliance usage time and electricity consumption.

(1 mark)

iv. Compute the value of ***X***, in Table 1 and explain its meaning.

(3 marks)

v. State the values of the intercept and slope coefficient, and hence write the regression equation.

(4 marks)

vi. Interpret the slope coefficient in practical terms.

(2 marks)

vii. Predict the household electricity consumption for a household with 4 hours of daily appliance usage. Comment on whether this prediction is reliable. Give a reason for your answer.

(3 marks)

b. An event organizer evaluates the entry waiting time (minutes) for 30 attendees at a conference registration counter. The summary statistics of the waiting times are presented in Table 3 below.

Table 3

Entry Waiting time (minutes)	
Mean	26.8
Standard Error	1.4
Median	25.9
Mode	24.0
Standard Deviation	7.5
Sample Variance	56.25
Kurtosis	0.12
Skewness	0.61
Range	32.0
Minimum	12.0
Maximum	44.0
Sum	804.0
Count	30

i. State **TWO (2)** methods that can be used to assess whether a dataset follows a normal distribution.

(2 marks)

ii. Based on the information in Table 3, discuss whether the distribution of entry waiting times can be considered approximately normally.

(2 marks)

iii. If the distribution were perfectly normal, describe the relative positions of the mean, median, and mode.

(2 marks)

(Total: 25 marks)

QUESTION 2

a. A university is reviewing whether students' average nightly sleep duration during the semester differs from the recommended 7 hours per night. A random sample of 25 undergraduate students reports their average sleep duration (hours) on weekdays. The summary statistics are shown in Table 4.

Table 4

One-Sample Statistics

	n	Mean	Std. Deviation	Std. Error Mean
Sleep duration	25	6.5	1.2	*Y*

i. Calculate the value of ***Y***.

(2 marks)

ii. State the null hypothesis and alternative hypothesis to test whether the average sleep duration differs from 7 hours per night.

(2 marks)

iii. Construct a 95% confidence interval for the true average sleep duration of students. (Given $t_{0.025,24} = \pm 2.064$). Explain the values obtained.

(5 marks)

iv. Based on the confidence interval in part (iii), comment on whether the recommended average sleep duration of 7 hours is supported.

(3 marks)

b. A lecturer wants to compare the average weekly study time (hours) of students who attend in-person revision classes and students who rely mainly on online recorded lectures. Random samples were taken from each group, and the results are summarized in Table 5. The variances of study time for the two groups can be assumed to be equal.

Table 5

t-Test: Two-Sample Assuming Equal Variances

	In-person	*Online*
Mean	9.200	7.800
Variance	4.000	3.600
Observations	15	15
Pooled Variance	3.800	
Hypothesized Mean Difference	0	
df	28	
t Stat	2.041	
P(T<=t) one-tail	0.0254	
t Critical one-tail	1.701	
P(T<=t) two-tail	0.0508	
t Critical two-tail	2.048	

i. Identify which group shows a greater average weekly study time, and state the numerical difference between the two groups.

(2 marks)

ii. State the null hypothesis and alternative hypothesis for comparing the two learning modes.

(2 marks)

iii. Using the critical method approach, state whether the observed difference in mean weekly study time is statistically meaningful. Support your answer using $\alpha = 0.05$.

(3 marks)

c. Fifteen first-year students attended a study skills workshop focusing on effective learning techniques. Each student completed the same course quiz before and after the workshop, and the scores were recorded. The data and analysis results are presented in Table 6 below.

Table 6

t-Test: Paired Two Sample for Means

	Before	*After*
Mean	64.5	70.0
Variance	36.000	30.250
Observations	15	15
Pearson Correlation	0.680	
Hypothesized Mean Difference	0	
df	14	
t Stat	***M***	
P(T<=t) one-tail	0.010	
t Critical one-tail	1.761	
P(T<=t) two-tail	0.020	
t Critical two-tail	2.145	

i. Determine the value of ***M***. (Given the standard deviation of the paired differences, $s_d = 6.00$).

(3 marks)

ii. Using the p-value approach, decide whether there is sufficient evidence that the workshop affected the mean quiz score. Use $\alpha = 0.10$ and state your conclusion clearly.

(3 marks)

(Total: 25 marks)

QUESTION 3

a. A car manufacturer is comparing the consistency of fuel efficiency (km per litre) between Car Model A and Car Model B. Random samples of vehicles from each model were tested under similar driving conditions. The manufacturer is interested in determining whether the variability in fuel efficiency differs between the two models. Use a 5% significance level. The results are summarized in Table 7 below.

Table 7

F-Test Two-Sample for Variances

	Model A	*Model B*
Mean	18.6	18.9
Variance	1.44	0.81
Observations	18	18
df	17	17
F	***K***	
P(F<=f) one-tail	0.171	
F Critical one-tail	2.35	

i. State the null hypothesis and alternative hypothesis for this study.

(2 marks)

ii. Calculate the value of ***K***, the test statistic for the F-test.

(2 marks)

iii. Using the critical value approach, determine whether there is a significant difference in fuel efficiency variability between Car Model A and Car Model B at $\alpha = 0.05$.

(3 marks)

b. A music streaming platform analyses whether average song popularity ratings differ among tracks released by artists at three different career stages: New artists, Mid-career artists and Established artists. The data were analyzed using a one-way ANOVA at the 5% significance level. The summary statistics and ANOVA results are presented in Tables 8 and 9 below.

Table 8

Anova : Single Factor SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
New artists	9	612	68.0	40.0
Mid-career artists	9	684	76.0	36.0
Established artists	9	729	81.0	25.0

Table 9

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	***A***	***C***	387.0	***E***	0.004	3.40
Within Groups	***B***	24	***D***			
Total	1260	26				

i. Using the Sum of Squares Between (SSB) formula, show that ***A*** = 774. Hence, determine the values of ***B, C, D*** and ***E***.

(11 marks)

ii. State the null hypothesis and alternative hypothesis for this study.

(2 marks)

iii. Using the critical value approach, determine whether there is a significant difference in mean song popularity ratings among the three artist groups. (Use $\alpha = 0.05$).

(3 marks)

iv. Rank the artist groups according to their mean popularity ratings, starting with the highest.

(2 marks)

(Total: 25 marks)

QUESTION 4

a. A tourism board conducts a survey to understand visitors' preferred holiday activity. The activities considered are: Sightseeing, Beach activities, Shopping, Food tasting, and Relaxing at resorts. The tourism board would like to test whether visitors' preferences are equally distributed among the five activities. The data were analyzed using Microsoft Excel at a 10% significance level, and the results are summarized in Tables 10 and 11 below.

Table 10

Activity	Observed N	Expected N	Residual
Sightseeing	55	45	***W***
Beach activities	***T***	45	5
Shopping	40	45	-5
Food tasting	52	***U***	7
Relaxing at resorts	28	45	-17

Table 11

Pearson Chi-Square	***V***
p-value	0.0284
Critical value	7.779

i. State the formula used to calculate the degrees of freedom for a chi-square goodness-of-fit test.

(1 mark)

ii. Determine the values of ***W, T*** and ***U*** shown in Table 10.

(3 marks)

iii. Using the Pearson Chi-Square test formula, show that the value of the test statistic, ***V*** is 10.84.

(3 marks)

iv. State the null hypothesis and alternative hypothesis for this study.

(2 marks)

v. Based on the *p*-value given in Table 11, determine whether there is a significant difference in visitors' holiday activity preferences at the 10% significance level.

(3 marks)

b. A psychologist investigates whether personality type is related to preferred work style. Each participant is classified according to their dominant personality type (Introvert, Ambivert, Extrovert) and their preferred work style (Working alone, Small group, Large group). A survey of 180 adults was conducted, and the results were analyzed using Microsoft Excel at the 5% significance level. The observed and expected frequencies are summarized in Tables 12 and 13 below.

Table 12

			Work Style			
			Working alone	Small group	Large group	Total
Personality type	Introvert	Observed	30	18	12	60
		Expected	20	***P***	***Q***	60
	Ambivert	Observed	20	25	15	60
		Expected	***R***	23.33	16.67	60
	Extrovert	Observed	10	27	23	60
		Expected	20	***S***	16.67	60
Total			60	70	50	180

Table 13

Pearson Chi-Square	14.347
p-value	0.0062
Critical value	9.488

i. Calculate the values of ***P***, ***Q***, ***R*** and ***S*** shown in Table 12.

(8 marks)

ii. State the null hypothesis and alternative hypothesis for this study.

(2 marks)

iii. Based on the critical value in Table 13, determine whether there is a significant relationship between personality type and preferred work style at the 5% significance level.

(3 marks)

(Total: 25 marks)

(TOTAL: 100 MARKS)

END OF QUESTION PAPER

APPENDIX 1(1)

Correlation and Regression

i. Pearson's Product Moment Correlation Coefficient

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[n\left(\sum X^2\right) - \left(\sum X\right)^2\right]\left[n\left(\sum Y^2\right) - \left(\sum Y\right)^2\right]}}$$

ii. The least-squares regression line of Y against X, Y = a + bX

$$b = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{n\left(\sum X^2\right) - \left(\sum X\right)^2}$$

$$a = \frac{\left(\sum y\right)}{n} - b\frac{\left(\sum x\right)}{n}$$

Confidence Intervals

i. Mean μ, for large samples

$$\bar{x} \pm z\alpha_{/2}\frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x} \pm z\alpha_{/2}\frac{s}{\sqrt{n}}$$

ii. Mean μ, for small samples

$$\bar{x} \pm t\alpha_{/2}\frac{s}{\sqrt{n}} \qquad df = n - 1$$

iii. Test statistic for two population mean, paired observation

$$t = \frac{\bar{d} - 0}{\frac{s_d}{\sqrt{n}}}$$

iv. Difference in means, $\mu_1 - \mu_2$, , for large and independent samples

$$(\bar{x}_1 - \bar{x}_2) \pm z\alpha_{/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{or} \quad (\bar{x}_1 - \bar{x}_2) \pm z\alpha_{/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

v. Difference in means, $\mu_1 - \mu_2$, , for small and independent samples

$$(\bar{x}_1 - \bar{x}_2) \pm t\alpha_{/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad df = n_1 + n_2 - 2$$

vi. Difference in means, $\mu_1 - \mu_2 = \mu_d$ for paired samples

$$\bar{d} \pm t\alpha_{/2}\frac{s_d}{\sqrt{n}} \qquad df = n - 1$$

APPENDIX 1(2)

Analysis of Variance (ANOVA)

i. Degrees of freedom for the numerator = k – 1

ii. Degrees of freedom for the denominator = n – k

iii. Total sum of squares:

$$SST = \sum x^2 - \frac{(\sum x)^2}{n}$$

iv. Between-samples sum of squares:

$$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_1} + \frac{T_3^2}{n_1} + \cdots\right) - \frac{(\sum x)^2}{n}$$

v. Within-samples sum of squares = $SST - SSB$

vi. Variance between samples:

$$MSB = \frac{SSB}{(k-1)}$$

vii. Variance within samples:

$$MSW = \frac{SSW}{(n-k)}$$

viii. Test statistic for a one-way ANOVA test:

$$F = \frac{MSB}{MSW}$$

Chi-square Test

i. Test statistics for Chi-square test

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

ii. $$Expected\ value = \frac{row\ sum \times column\ sum}{grand\ total}$$