UNIVERSITI POLY-TECH 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.
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. A biologist is investigating the relationship between the average daily temperature (in degrees Celsius) and the number of active bees in a hive. The goal is to determine if higher temperatures are associated with increased bee activity. Data is collected over several days, recording both temperature and the number of active bees. The range of daily temperature is between 24 to 33 degrees Celsius. The data is analyzed using statistical software; the output is shown in Tables 1 and 2 below.

Table 1

Regression Statistics	
Multiple R	0.9859
R Square	***X***
Adjusted R Square	0.9673
Standard Error	5.8560
Observations	8

Table 2

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-185.282	22.201	-8.346	0.000
Daily temperature	10.810	0.749	14.427	0.000

i. List the dependent and independent variables.

(2 marks)

ii. State a type of graph that can be used to identify the relationship between the independent and dependent variables. Subsequently, identify and interpret the relationship between the variables obtained in Table 1.

(3 marks)

iii. Name a measure that can identify the relationship between the independent and dependent variables.

(2 marks)

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

(3 marks)

v. Referring to Table 2, state the y-intercept, ***a*** and slope coefficient, ***b***. Hence, determine the regression equation for the independent and dependent variables.

(4 marks)

vi. Interpret the meaning of slope coefficient, ***b*** obtained in (v).

(2 marks)

vii. Predict the number of active bees in a hive if the average daily temperature is 34 degrees Celsius. Is the estimation reliable? Explain your answer.

(3 marks)

b. A study examines the effect of a new learning app on students' vocabulary skills. Over a month, 13 students used the app, and their improvement in vocabulary test scores was measured. The summary statistics of the data are presented in Table 3 below.

Table 3

Students' scores	
Mean	58.231
Standard Error	7.115
Median	68
Mode	#N/A
Standard Deviation	25.652
Sample Variance	658.026
Kurtosis	0.725
Skewness	-1.253
Range	78
Minimum	5
Maximum	83
Sum	757
Count	13

i. Name two types of graphical methods that can be used to identify the normality of the distribution.

(2 marks)

ii. Based on Table 3 above, can we indicate that the distribution is normally distributed? Explain.

(2 marks)

iii. State **ONE (1)** property that defines a symmetric distribution.

(2 marks)

(Total: 25 marks)

QUESTION 2

a. A supermarket manager investigates the average checkout speed (items scanned per minute) of cashiers during peak hours. They collect data from a random sample of 18 cashiers. The analysis results are presented in Table 4 below.

Table 4

One-Sample Statistics

	n	Mean	Std. Deviation	Std. Error Mean
Checkout speed	18	42	5.8	***M***

i. Prove that ***M*** is 1.37.

(2 marks)

ii. State the null and alternative hypotheses that the mean checkout speed for all cashiers during peak hours is 46 items per minute.

(4 marks)

iii. Assuming a 95% confidence level, calculate the confidence interval for all cashiers' true average checkout speed during peak hours.
(Given $t_{0.025,17} = \pm 2.11$).

(3 marks)

iv. Based on the confidence interval in (iii), can we conclude that the mean checkout speed is 46 items per minute?

(3 marks)

b. A company investigates whether there is a significant difference in productivity between employees working in the office and at home. Productivity scores were collected and analyzed using a statistical method, and the results are presented in Table 5 below ($\alpha = 0.10$).

Table 5

t-Test: Two-Sample Assuming Equal Variances

	Office	*Home*
Mean	72	69
Variance	141.2	222.8
Observations	6	6
Pooled Variance	182	
Hypothesized Mean Difference	0	
df	10	
t Stat	0.385	
P(T<=t) one-tail	0.354	
t Critical one-tail	1.372	
P(T<=t) two-tail	0.708	
t Critical two-tail	1.812	

i. State the null and alternative hypotheses.

(4 marks)

ii. Using the critical method approach, is there a significant difference in the average productivity between employees working in the office and at home? ($\alpha = 0.10$).

(3 marks)

c. A manufacturing company is testing the impact of a new machine lubricant on the operating temperature of its machines. The company measured the operating temperature of each machine before and after applying the new lubricant over a specific time frame. The data and analysis results are presented in Table 6 below. ($\alpha = 0.01$).

Table 6

t-Test: Paired Two Sample for Means

	Pre -Lubricant temperature	*Post - Lubricant temperature*
Mean	85.336	82.345
Variance	39.729	27.567
Observations	11	11
Pearson Correlation	0.585	
Hypothesized Mean Difference	0.000	
Df	10.000	
t Stat	***X***	
P(T<=t) one-tail	0.047	
t Critical one-tail	2.764	
P(T<=t) two-tail	0.093	
t Critical two-tail	3.169	

i. Calculate the value of ***X***. (Given the standard deviation of the mean differences, $s_d = 5.43$).

(3 marks)

ii. Using the p-value approach, can we conclude that there is a significant difference in the mean operating temperatures before and after applying the new lubricant at a 1% significance level?

(3 marks)

(Total: 25 marks)

QUESTION 3

a. A pharmaceutical company is testing two different drug formulations to determine if there is a significant difference in the variance of their effectiveness. They conducted a test to compare the variances of effectiveness scores from both drug formulations. Data were recorded and analyzed using statistical software at $\alpha = 0.10$, and the output is presented in Table 7.

Table 7

F-Test Two-Sample for Variances

	Drug *Formulation A*	Drug *Formulation B*
Mean	75.333	74
Variance	70.5	55.75
Observations	9	9
df	8	8
F	1.265	
P(F<=f) one-tail	0.374	
F Critical one-tail	2.589	

i. Write the null and alternative hypotheses.

(4 marks)

ii. Based on the p-value in Table 7, test at the 10% significance level whether there is a significant difference in the variance of effectiveness scores between Drug Formulation A and Drug Formulation B.

(3 marks)

b. An energy company is conducting a study to determine the effectiveness of three different energy-saving programs in reducing household energy consumption. They divided a sample of households into three groups, each following a different energy-saving program. After a certain period, the company measures the households' energy consumption, and the data are presented in Tables 8 and 9 using analysis of variance (ANOVA) at a significance level of 0.05.

Table 8

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Program Eco	5	719	143.8	24.2
Program Green	5	706	141.2	27.7
Program Save	5	679	135.8	26.2

Table 9

ANOVA

Source of Variation	*SS*	*Df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	166.533	***B***	83.267	***D***	0.077	3.885
Within Groups	***A***	***C***	26.033			
Total	478.933	14				

i. Calculate the values of ***A***, ***B***, ***C*** and ***D***.

(9 marks)

ii. State the null and alternative hypotheses.

(4 marks)

iii. Based on the critical value in Table 9, is there a significant difference in the mean energy consumption among the three energy-saving programs? ($\alpha = 0.05$).

(3 marks)

iv. Which energy-saving program shows the highest and the lowest average energy consumption?

(2 marks)

(Total: 25 marks)

QUESTION 4

a. A sociologist conducts a survey to examine whether there is a significant preference for social media platforms among teenagers. The data were analysed at $\alpha = 0.05$ using MS Excel and the results are shown in Tables 10 and 11.

Table 10

Social Media Platform	Observed N	Expected N	Residual
Instagram	40	23.4	16.6
Snapchat	15	23.4	-8.4
TikTok	32	23.4	8.6
Twitter	17	23.4	-6.4
Facebook	13	23.4	-10.4

Table 11

Pearson Chi-Square	24.325
p-value	0.000
Critical value	9.488

i. Prove that $\chi^2 = 24.325$.

(4 marks)

ii. State the null and alternative hypotheses for the above study.

(4 marks)

iii. Based on the p-value, can we conclude that there is a significant difference in preference for social media platforms at a 5% significance level?

(3 marks)

b. A researcher is investigating the relationship between coffee consumption and productivity levels among office workers. Data is gathered and classified based on coffee consumption (Yes, No) and productivity levels (High, Low) as shown in Table 12. Table 13 displays Chi-square test results at $\alpha = 0.10$.

Table 12

			Productivity Level		Total
			High	Low	
Coffee Consumption	Yes	Observed	35	15	50
		Expected	***X***	***Y***	50
	No	Observed	30	42	72
		Expected	***Z***	33.64	72
Total			65	57	122

Table 13

Pearson Chi-Square	9.516
p-value	0.002
Critical value	2.706

i. Compute the values of ***X***, ***Y*** and ***Z***.

(6 marks)

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

(4 marks)

iii. Based on the critical value, can we conclude that there is a significant relationship between coffee consumption and productivity levels at $\alpha = 0.10$?

(4 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

i. Test statistic for Chi-square:

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

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