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 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 SO

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

QUESTION 1

a. A retail company examines how product selling price (RM) affects monthly quantity demanded (units sold) for one of its products. Data from 12 months were analyzed using Excel, and the regression output is summarized in Tables 1 and 2 below. The selling price ranges from RM10 to RM40.

Table 1

Regression Statistics	
Multiple R	0.8650
R Square	***X***
Adjusted R Square	0.7200
Standard Error	120.0
Observations	12

Table 2

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1520.0	95.0	16.000	0.0000
Selling price (RM)	-28.5	6.2	-4.597	0.0011

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 a graph that illustrates the relationship between independent and dependent variables.

(1 mark)

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

(3 marks)

v. State the y-intercept, ***a*** and slope coefficient, ***b***. Hence, write the regression equation.

(4 marks)

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

(2 marks)

vii. Estimate the monthly quantity demanded when the selling price is RM25. Is the estimation reliable? Explain.

(3 marks)

b. A marketing manager wants to understand customer service efficiency. For n = 30 days, the average customer waiting time (minutes) at a service counter was recorded and analyzed after removing one extreme recording error. The summary statistics are shown in Table 3.

Table 3

Customer waiting time (minutes)	
Mean	18.2
Standard Error	1.1
Median	17.9
Mode	17.5
Standard Deviation	6.0
Sample Variance	36.0
Kurtosis	0.15
Skewness	0.28
Range	26.0
Minimum	6.0
Maximum	32.0
Sum	546.0
Count	30

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. Based on the Kolmogorov–Smirnov test, when can the data be considered normally distributed? State your answer in terms of the *p*-value.

(2 marks)

(Total: 25 marks)

QUESTION 2

a. A food delivery company has introduced a new rider allocation system to improve service efficiency. Management claims that the mean delivery time for completed orders should be 30 minutes. To evaluate this claim, the operations team randomly sampled 15 completed deliveries from a busy evening period. The delivery times were analyzed using statistical software, and the summary output is shown in Table 4.

Table 4

One-Sample Statistics

Delivery time (minutes)	n	Mean	Std. Deviation	Std. Error Mean
	15	33.2	6.0	Y

i. Find the value of Y.

(2 marks)

ii. State the null hypothesis and alternative hypothesis to test whether the population mean delivery time is 30 minutes.

(2 marks)

iii. Construct a 90% confidence interval for the true population mean delivery time. (Given $t_{0.05,14} = \pm 1.761$). Explain the values obtained.

(5 marks)

iv. Based on the confidence interval in (iii), comment on whether the company's mean delivery time target of 30 minutes is achieved.

(3 marks)

b. A bank tested two different customer service methods over four weeks to evaluate their effect on customer satisfaction scores (0–100). Method A involved face-to-face service with personalized attention, while Method B involved standard digital service via kiosks. Teams handled identical tasks, and weekly satisfaction scores were averaged per employee. Preliminary analysis indicated similar variability, so a two-sample t-test assuming equal variances is appropriate. The results are summarized in Table 5.

Table 5

t-Test: Two-Sample Assuming Equal Variances

	Method A	*Method B*
Mean	85.000	78.000
Variance	18.000	20.000
Observations	18	18
Pooled Variance	19.0	
Hypothesized Mean Difference	0	
df	34	
t Stat	3.122	
P(T<=t) one-tail	0.002	
t Critical one-tail	1.690	
P(T<=t) two-tail	0.004	
t Critical two-tail	2.032	

i. Which service method has the higher sample mean, and by how many points? (2 marks)

ii. State the null and alternative hypotheses for testing whether there is a difference in mean customer satisfaction scores between the two methods. (2 marks)

iii. Using the critical method approach, is there is a significant difference in mean satisfaction scores between the two methods. Justify your answer. (Use $\alpha = 0.05$). (3 marks)

c. A call center implemented a new customer service script to improve customer satisfaction. Twenty-five customer service agents were evaluated before and after using the script. Each agent handled the same set of calls in both periods, and their average customer satisfaction scores (0–100) were recorded. The summary statistics are shown in Table 6.

Table 6

t-Test: Paired Two Sample for Means

	Before	*After*
Mean	70.4	75.2
Variance	36.0	30.0
Observations	25	25
Pearson Correlation	0.68	
Hypothesized Mean Difference	0	
df	24	
t Stat	***M***	
P(T<=t) one-tail	0.002	
t Critical one-tail	1.711	
P(T<=t) two-tail	0.004	
t Critical two-tail	2.064	

i. Calculate the value of ***M***. (Given the standard deviation of the mean differences, $s_d = 5.00$).

(3 marks)

ii. Using the *p*-value approach, decide whether the new customer service script changed mean customer satisfaction scores. (Use $\alpha = 0.10$)

(3 marks)

(Total : 25 marks)

QUESTION 3

a. A company wants to compare the consistency of customer call handling times (in minutes) between two call centre, Centre X and Centre Y, over 20 randomly selected days. Management wants to determine whether the variability in daily average call handling times differs between the two centres. Use a 1% significance level.

Table 7

F-Test Two-Sample for Variances

	Centre X	*Centre Y*
Mean	8.5	8.2
Variance	1.44	0.81
Observations	20	20
df	19	19
F	***K***	
P(F<=f) one-tail	0.038	
F Critical one-tail	3.00	

i. State the null and alternative hypotheses for testing whether the variability in call handling times is the same between Centre X and Centre Y.

(2 marks)

ii. Calculate the value of ***K***.

(2 marks)

iii. Using the critical-value approach $\alpha = 0.01$, determine whether there is a significant difference in variability between Centre X and Centre Y. Explain your conclusion.

(3 marks)

b. A retail company wants to evaluate whether three different customer loyalty programs result in different average monthly purchase amounts (RM) over a 2-month trial. Customers are randomly assigned to one of the three programs: Points Program, Cashback Program and Discount Program. The results are analyzed using ANOVA at the 10% significance level. The summary tables are shown below.

Table 8

Anova : Single Factor
SUMMARY

Program	*Count*	*Sum*	*Average*	*Variance*
Points	12	1200	100	225
Cashback	12	1080	90	180
Discount	12	1020	85	150

Table 9

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	***A***	2	***C***	***E***	0.045	2.96
Within Groups	6105	***B***	***D***			

i. Calculate the Sum of Squares Between Groups (***A***). Then determine the values of ***B, C, D*** and ***E***.

(11 marks)

ii. State the null and alternative hypotheses for the ANOVA test.

(2 marks)

iii. Using the critical value approach at $\alpha = 0.10$, determine whether there is a significant difference in mean monthly purchases among the three loyalty programs. Justify your answer.

(3 marks)

iv. Based on the summary statistics, by how many RM does the highest average purchase differ from the lowest average purchase?

(2 marks)

(Total: 25 marks)

QUESTION 4

a. A company wants to analyze which communication channel customers prefer when contacting customer service. The channels are: Email, Phone, Live Chat, Social Media, Mobile App and In-Person. They want to test whether the observed customer preferences differ from an equal distribution across the six channels. Data were collected from 180 customers and analyzed in MS Excel at $\alpha = 0.05$. The summary tables are shown below.

Table 10

Channel	Observed N	Expected N	Residual
Email	35	30	***V***
Phone	***T***	30	5
Live Chat	28	30	-2
Social Media	22	30	***U***
Mobile App	35	30	5
In-Person	25	30	-5

Table 11

Pearson Chi-Square	***W***
p-value	0.347
Critical value	11.071

i. For the above goodness-of-fit test (no parameters estimated), what is the value of degree of freedom (df)?

(1 mark)

ii. Determine the values of ***T***, ***U*** and ***V***.

(3 marks)

iii. Calculate Pearson Chi-Square value, ***W***.

(3 marks)

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

(2 marks)

v. Based on the *p*-value in Table 11, decide whether there is a significant difference in customer preferences at the 5% significance level?

(3 marks)

b. A city health researcher investigates whether type of exercise is related to preferred time of day for exercise in a sample of 200 adults. Each respondent is classified by exercise type (Cardio, Strength, Yoga) and by preferred time of day (Morning, Afternoon, Evening). Use $\alpha = 0.01$. The data and Excel outputs are summarized below.

Table 12

<table>
<tr><td colspan="3" rowspan="2"></td><td colspan="3">Time of Day</td></tr>
<tr><td>Morning</td><td>Afternoon</td><td>Evening</td></tr>
<tr><td rowspan="6">Exercise Type</td><td rowspan="2">Cardio</td><td>Observed</td><td>35</td><td>30</td><td>15</td></tr>
<tr><td>Expected</td><td>28</td><td>P</td><td>Q</td></tr>
<tr><td rowspan="2">Strength</td><td>Observed</td><td>25</td><td>20</td><td>15</td></tr>
<tr><td>Expected</td><td>R</td><td>18.6</td><td>20.4</td></tr>
<tr><td rowspan="2">Yoga</td><td>Observed</td><td>10</td><td>12</td><td>38</td></tr>
<tr><td>Expected</td><td>21</td><td>S</td><td>20.4</td></tr>
</table>

Table 13

Pearson Chi-Square	33.90
p-value	0.000
Critical value	13.277

i. Compute the values of ***P***, ***Q***, ***R*** and ***S***.

(8 marks)

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

(2 marks)

iii. Based on the critical value in Table 13, determine whether there is a significant relationship between exercise type and preferred time of day at $\alpha = 0.01$.

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