

**FINAL EXAMINATION**

**BACHELOR OF COMMUNICATION (HONS) IN CORPORATE COMMUNICATION
BACHELOR OF BUSINESS ADMINISTRATION (HONOURS)
HUMAN RESOURCE MANAGEMENT
BACHELOR OF BUSINESS ADMINISTRATION (HONOURS)**

COURSE : STATISTICS FOR SOCIAL SCIENCES

COURSE CODE : STA2113

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 café owner wants to examine the relationship between advertising cost (RM) and monthly sales revenue (RM'000) of one of its products. Data from 10 months were analyzed using Excel, and the regression output is summarized in Tables 1 and 2 below. The advertising cost ranges from RM500 to RM3000.

Table 1

<i>Regression Statistics</i>	
Multiple R	0.9120
R Square	X
Adjusted R Square	0.8100
Standard Error	3.25
Observations	10

Table 2

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	18.50	4.20	4.405	0.0023
Advertising cost (RM)	0.012	0.003	4.000	0.0039

- 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 sales revenue when the advertising cost is RM2000. Is the estimation reliable? Explain. (3 marks)

- b. A customer service supervisor wants to examine the pattern of daily customer complaint resolution time. The average time taken to resolve complaints (hours) was recorded. The summary statistics are shown in Table 3.

Table 3

<i>Complaint resolution time (hours)</i>	
Mean	4.8
Standard Error	0.35
Median	4.6
Mode	4.5
Standard Deviation	1.75
Sample Variance	3.06
Kurtosis	0.20
Skewness	0.32
Range	7.0
Minimum	2.0
Maximum	9.0
Sum	120.0
Count	25

- i. State any two methods that can be used to assess the normality of data. (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 logistics company wants to evaluate whether its average parcel processing time meets the company target. Management claims that the mean processing time should be 45 minutes. To evaluate this claim, the quality control team randomly selected 20 parcels processed during peak hours. The processing times were analyzed using statistical software, and the summary output is shown in Table 4.

Table 4

One-Sample Statistics

Processing time (minutes)	n	Mean	Std. Deviation	Std. Error Mean
	20	47.5	8.0	Y

- i. Find the value of Y. (2 marks)
- ii. State the null hypothesis and alternative hypothesis. (2 marks)
- iii. Construct a 95% confidence interval for the true population mean processing time. (Given $t_{0.025,19} = \pm 2.093$). Explain the values obtained. (5 marks)
- iv. Based on the confidence interval in (iii), comment on whether the company's mean processing time target of 45 minutes is achieved. (3 marks)

- b. A training manager wants to compare the effectiveness of two training methods on employees' performance scores. Method A uses classroom training, while Method B uses online self-paced training. Both groups completed the same assessment after training. The data were recorded and analysed using Microsoft Excel, and the results are summarized in Table 5.

Table 5
t-Test: Two-Sample Assuming Equal Variances

	<i>Method A</i>	<i>Method B</i>
Mean	82.000	76.000
Variance	16.000	18.000
Observations	20	20
Pooled Variance	17.0	
Hypothesized Mean Difference	0	
df	38	
t Stat	4.620	
P(T<=t) one-tail	0.00002	
t Critical one-tail	1.686	
P(T<=t) two-tail	0.00004	
t Critical two-tail	2.024	

- i. Which training method has the higher sample mean, and by how many points?
(2 marks)
- ii. State the null and alternative hypotheses.
(2 marks)
- iii. Using the critical method approach at $\alpha = 0.01$, is there a significant difference in mean performance scores between the two methods? Justify your answer.
(3 marks)

- c. A fitness center introduced a new 4-week exercise programme to improve members' fitness level. Twenty members were evaluated before and after joining the programme. Each member completed the same fitness assessment, and their fitness scores were recorded. The summary statistics are shown in Table 6.

Table 6

t-Test: Paired Two Sample for Means

	<i>Before</i>	<i>After</i>
Mean	62.5	68.0
Variance	25.0	20.0
Observations	20	20
Pearson Correlation	0.72	
Hypothesized Mean Difference	0	
df	19	
t Stat	<i>M</i>	
P(T<=t) one-tail	0.001	
t Critical one-tail	1.729	
P(T<=t) two-tail	0.002	
t Critical two-tail	2.093	

- i. Calculate the value of ***M***. (Given the standard deviation of the mean differences, $s_d = 5.50$).
- ii. Using the p -value approach at $\alpha = 0.05$, decide whether there is a significance difference of mean fitness score between before and after joining the programme.

(3 marks)

(3 marks)

(Total : 25 marks)

QUESTION 3

- a. A company wants to compare the consistency of daily delivery times (in minutes) between two courier branches, Branch A and Branch B. Management wants to determine whether the variability in daily delivery times differs between the two branches.

Table 7
F-Test Two-Sample for Variances

	<i>Branch A</i>	<i>Branch B</i>
Mean	42.5	40.8
Variance	6.25	4.00
Observations	15	15
df	14	14
F	<i>K</i>	
P(F<=f) one-tail	0.210	
F Critical one-tail	2.48	

- i. State the null and alternative hypotheses. (2 marks)
- ii. Calculate the value of *K*. (2 marks)
- iii. Using the critical-value approach $\alpha = 0.10$, determine whether there is a significant difference in variability of daily delivery times between Branch A and Branch B. Explain your conclusion. (3 marks)

- b. A college wants to evaluate whether three different teaching methods result in different average test scores among students. Students were randomly assigned to one of the three methods which are Lecture, Video based and Interactive. The results were analyzed using ANOVA at the 5% significance level. The summary tables are shown below.

Table 8

Anova: Single Factor
SUMMARY

<i>Program</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Lecture	10	800	80	100
Video based	10	720	72	90
Interactive	10	680	68	110

Table 9

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>
Between Groups	A	2	C	E	0.036	3.35
Within Groups	2700	B	D			

- Calculate the Sum of Squares Between Groups, **A**. Then determine the values of **B**, **C**, **D** and **E**.
(11 marks)
- State the null and alternative hypotheses for the ANOVA test.
(2 marks)
- Using the critical value approach at $\alpha = 0.05$, determine whether there is a significant difference in mean test scores among the three teaching methods. Justify your answer.
(3 marks)
- Based on the summary statistics, by how many marks does the highest average test score differ from the lowest average test score?
(2 marks)

(Total: 25 marks)

QUESTION 4

- a. A university wants to analyze which learning platform students prefer for online learning. The platforms are Google Classroom, Microsoft Teams, Moodle, Zoom, You Tube and WhatsApp. The university wants to test whether student preferences differ from an equal distribution across the six platforms. Data were collected from 240 students and analyzed in MS Excel at $\alpha = 0.05$. The summary tables are shown below.

Table 10

Platform	Observed N	Expected N	Residual
Google Classroom	50	40	V
Microsoft Teams	T	40	5
Moodle	35	40	-5
Zoom	30	40	U
You Tube	45	40	5
WhatsApp	35	40	-5

Table 11

Pearson Chi-Square	W
p-value	0.186
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 of online learning platforms, decide whether there is a significant difference in student preferences of online learning platforms at the 5% significance level?
(3 marks)

- b. A university researcher investigates whether preferred payment method is related to student age group. A sample of 180 students was classified by payment method and age group. Use $\alpha = 0.05$. The data and Excel outputs are summarized below.

Table 12

			Age Group		
			18-25	26-35	36-45
Payment Method	Cash	Observed	30	20	10
		Expected	P	23.33	Q
	Debit Card	Observed	20	30	25
		Expected	25.00	R	20.83
	E-wallet	Observed	10	20	15
		Expected	15	S	12.50

Table 13

Pearson Chi-Square	12.526
p-value	0.014
Critical value	5.991

- Compute the values of P , Q , R and S .
(8 marks)
- State the null and alternative hypotheses for this study.
(2 marks)
- Based on the critical value in Table 13, determine whether there is a significant relationship between preferred payment method and student age group at $\alpha = 0.05$.
(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

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

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

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

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

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

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad df = n - 1$$

- 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

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad df = n_1 + n_2 - 2$$

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

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} \quad df = n - 1$$

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:

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

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

- vi. Mean Square between samples:

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

- vii. Mean Square 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* = $\frac{\text{row sum} \times \text{column sum}}{\text{grand total}}$