

**FINAL EXAMINATION****DIPLOMA OF ACCOUNTANCY****COURSE : BUSINESS STATISTICS****COURSE CODE : STA1123****DURATION : 2 HOURS 30 MINUTES****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 11 printed pages including the front page

QUESTION 1

- a. Regression analysis is a statistical technique used to examine the relationship between variables and to predict future outcomes.
- i. State a graph that can display the relationship between the independent and dependent variables.
(1 mark)
 - ii. Describe the purpose of the correlation coefficient in regression analysis and state its possible range of values.
(3 marks)
 - iii. Describe a negative linear relationship between two variables by providing a suitable real-life example. Explain why the relationship is considered negative.
(2 marks)
 - iv. Identify and elaborate on **TWO (2)** types of prediction used in regression analysis.
(4 marks)

- b. A manager at a retail company conducted a study to examine the relationship between the number of customer service workshops attended per year and employees' customer satisfaction scores (measured out of 100). The following regression data were obtained based on 35 employees.

Table 1	
<i>Regression Statistics</i>	
Multiple R	0.61
R Square	X
Adjusted R Square	0.34
Standard Error	6.5
Observations	35

Table 2				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	52.0	3.5	14.86	0.000
Number of customer service workshops	2.4	0.9	2.67	0.012

- i. Determine the independent and dependent variables. (2 marks)
- ii. Using Table 1, identify the correlation coefficient and interpret its meaning. (2 marks)
- iii. Calculate the value of **X** that represents the percentage of variation in the dependent variable explained by the independent variable. Then, interpret the result in the context of the study. (3 marks)
- iv. Referring to Table 2, state the values of the y-intercept, **a** and the slope coefficient, **b**. Then, write the regression equation. (2 marks)
- v. Interpret the value of slope coefficient, **b** from part (iv). (2 marks)
- vi. Predict the customer satisfaction scores for employees who attended
 - a. 5 customer service workshops per year.
 - b. 12 customer service workshops per year.

(4 marks)

(Total: 25 marks)

QUESTION 2

- a. Table 3 summarizes the statistical measures of weekly study hours for university students enrolled in a Statistics course at a college in Johor.

Table 3	
<i>Weekly Study Hours</i>	
Mean	12.500
Standard Error	0.842
Median	12
Mode	11
Standard Deviation	3.174
Sample Variance	10.074
Kurtosis	-0.865
Skewness	0.156
Range	11
Minimum	7
Maximum	18
Sum	175
Count	14

- i. Based on Table 3, can we indicate that the weekly study hours for the students are normally distributed? Justify your answer. (3 marks)
- ii. Besides graphical methods and descriptive statistical measures, state and explain **ONE (1)** other method that can be used to assess whether the data is normally distributed. (3 marks)

- b. Table 4 summarizes the outcomes of a Kolmogorov-Smirnov test performed to evaluate the normality of weekly customer waiting time data across three different bank branches in Kuala Lumpur: Central Branch, Suburban Branch and University Branch.

Table 4

	Branch	Kolmogorov-Smirnov		
		Statistic	df	Sig.
Weekly customer waiting time	Central	0.143	15	0.074
	Suburban	0.218	15	0.019
	University	0.117	15	0.126

Based on Table 4, using $\alpha = 0.05$, determine whether the weekly customer waiting time at the Suburban Branch follows a normal distribution. Provide a clear explanation for your conclusion.

(2 marks)

- c. Besides the Kolmogorov-Smirnov test, the Shapiro-Wilk test is another method used to evaluate the normality of a dataset. Explain what a p -value greater than 0.05 indicates in the Shapiro-Wilk test for the weekly customer waiting time data across the bank branches.

(2 marks)

(Total: 10 marks)

QUESTION 3

- a. A hotel manager investigates the average daily room occupancy score (%) of selected rooms in a resort. He collects data from a random sample of 20 rooms. The analysis results are shown in Table 5.

Table 5
One-Sample Statistics

Occupancy score (%)	n	Mean	Std. Deviation	Std. Error Mean
	20	72	8.4	Y

- i. Find the value of Y. (2 marks)
- ii. State the null and alternative hypotheses to test whether the mean occupancy score of all rooms in the resort is 75%. (2 marks)
- iii. Assuming a 90% confidence level, calculate the confidence interval for the mean occupancy score of all rooms in the resort. Interpret the values obtained. (Given $t_{0.05,19} = \pm 1.729$) (4 marks)
- iv. Based on the confidence interval obtained in (iii), can we conclude that the mean occupancy score is 75%? (3 marks)

- b. A retail company investigates whether there is a significant difference in average weekly sales (in RM'000) between employees who received sales training and those who did not receive training. Weekly sales data were collected and analyzed using a statistical method at $\alpha = 0.05$, and the results are presented in Table 6 below.

Table 6
t-Test: Two-Sample Assuming Equal Variances

	<i>Trained Employees</i>	<i>Untrained Employees</i>
Mean	78.50	71.333
Variance	144.667	192.267
Observations	6	6
Pooled Variance	170.467	
Hypothesized Mean Difference	0	
df	<i>M</i>	
t Stat	0.957	
P(T<=t) one-tail	0.181	
t Critical one-tail	1.812	
P(T<=t) two-tail	0.361	
t Critical two-tail	2.228	

- i. State the null and alternative hypotheses. (2 marks)
- ii. Calculate the degree of freedom, ***M***. (2 marks)
- iii. Explain why a *t*-test was used to compare weekly sales between employees who received sales training and those who did not receive training. (3 marks)
- iv. Using the critical value approach, is there a significant difference in average weekly sales (in RM'000) between employees who received sales training and those who did not receive training? ($\alpha = 0.05$). (3 marks)
- v. Is the decision in (iv) consistent with your decision if the *p*-value approach is used? Use $\alpha = 0.05$. (4 marks)

- c. A smart logistics company is evaluating the effect of introducing an automated warehouse sorting system on parcel handling efficiency. The company recorded the average parcel processing time (in seconds) for the same set of delivery hubs before and after implementing the new system over a fixed period. The data and analysis results are presented in Table 7 below using $\alpha = 0.05$.

Table 7		
	<i>Before automation</i>	<i>After automation</i>
Mean	128.6	119.4
Variance	31.245	28.916
Observations	10	10
Pearson Correlation	0.992	
Hypothesized Mean Difference	0	
Df	9	
t Stat	Z	
P(T<=t) one-tail	0.000	
t Critical one-tail	1.833	
P(T<=t) two-tail	0.000	
t Critical two-tail	2.262	

- i. Based on the table output, what type of analysis is being used? (2 marks)
- ii. State the null and alternative hypotheses. (2 marks)
- iii. Find the value of **Z**, given the standard deviation of the differences, $s_d = 0.8452$. (3 marks)
- iv. Using the p -value approach, can we conclude that there is a significant difference in the mean processing times before and after implementing the automated warehouse sorting system at a 5% significance level? (3 marks)

(Total: 35 marks)

QUESTION 4

- a. A logistics and e-commerce company is comparing two different delivery methods namely, drone delivery and motorbike delivery in order to determine whether there is a significant difference in the variability of delivery times (in minutes). Delivery time data were collected from both methods and analyzed using statistical software at $\alpha = 0.10$, and the output is presented in Table 8.

Table 8
F-Test Two-Sample for Variances

	<i>Drone delivery</i>	<i>Motorbike delivery</i>
Mean	19.5	21.0
Variance	18.49	4.41
Observations	9	9
df	8	8
F	4.19	
P(F<=f) one-tail	0.015	
F Critical one-tail	2.589	

- i. Why F-test is used in this study? (2 marks)
- ii. State the null and alternative hypotheses. (2 marks)
- iii. Based on the critical value, test at the 10% significance level whether there is a significant difference in the variability of delivery times (in minutes) between Drone delivery and Motorbike delivery. (3 marks)

- b. A healthcare researcher is conducting a study to determine whether different workplace wellness programs have an impact on employee stress reduction scores. Four types of wellness programs were evaluated: Yoga & Mindfulness Program, Gym Membership Program, Flexible Working Hours Program, and No Wellness Program. Stress reduction scores of employees participating in these programs were collected and analyzed using analysis of variance (ANOVA). The results are presented in Tables 9 and 10 using $\alpha = 0.05$.

Table 9

Anova: Single Factor
SUMMARY

Groups	Count	Sum	Average	Variance
Yoga & Mindfulness	6	486	81.0	36.5
Gym Membership	6	444	74.0	52.3
Flexible Working Hours	6	510	85.0	28.7
No Wellness	6	378	63.0	61.4

Table 10

ANOVA

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	A	B	416.2	D	0.000312	3.098
Within Groups	1145.6	C	47.73			

- Calculate the values of **A**, **B**, **C** and **D**.
(10 marks)
- Explain why ANOVA is a more appropriate test for this study compared to t-test. Provide reasons to support your answer.
(3 marks)
- State the null and alternative hypotheses.
(2 marks)
- Based on the critical value approach, what conclusion can be drawn about the impact of different workplace wellness programs on employee mean stress reduction scores?
(3 marks)

- v. What does a p -value less than 0.05 indicate in the context of this ANOVA test?
(3 marks)
- vi. Given the average stress reduction scores in different wellness programs, which program appears to have the highest and lowest impact on employee well-being?
(2 marks)

(Total: 30 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, $Y = a + bX$

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

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

Hypothesis Testing

- i. Test statistic for one sample t-test

$$t = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- ii. Test statistic for two population mean, paired observation

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

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:

$$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \dots \right) - \frac{(\sum x)^2}{n}$$

- 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:

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

