UNIVERSITI TEKNOLOGI MARA
FINAL ASSESSMENT

COURSE	**:**	**STATISTICS FOR BUSINESS AND SOCIAL SCIENCES**
COURSE CODE	**:**	**STA404**
EXAMINATION	**:**	**JULY 2022**
TIME	**:**	**2 HOURS**

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of **SEVEN (7)** questions.

2. Answer ALL questions in the foolscap paper. Start each answer on a new page.

3. Candidates must accomplish this assessment within 2 hours.

4. Candidates are required to convert their completed answer in one PDF file before submission **(<FULLNAME STUDENTNO GROUP>.pdf).**

5. Candidates are given 30 minutes to email their finalized and completed answer to the respective lecturers.

6. Candidates are required to attach the following details in every page of the answer script :

   i) Full Name
   ii) Student Number
   iii) Group
   iv) HP Number

7. Please check to make sure that this assessment pack consists of :

   i) the Question Paper
   ii) a five – page Appendix 1

8. Answer ALL questions in English.

PLEASE READ THE INSTRUCTIONS CAREFULLY BEFORE START THE EXAMINATION

This assessment paper consists of 11 printed pages

QUESTION 1

A researcher is interested to study the E-wallet usage among customers of Pasaraya Intan Belian. The researcher intended to obtain information from 50 respondents by interviewing every 5th customer of Pasaraya Intan Belian on a particular day. The respondents are asked on their marital status, age, the frequency of using E-wallet (never, seldom, often, very often), the preferred E-wallet provider (Boast, Grabpay, Touch n Go, BigPay) and the last E-wallet transaction amount (RM).

a) State the population and the sample of the study.

(2 marks)

b) Identify **THREE (3)** variables from the study. Hence, state its scales of measurement.

(3 marks)

c) Identify the sampling method used.

(1 mark)

d) Give **ONE (1)** advantage of the data collection method used by the researcher.

(1 mark)

QUESTION 2

The director of a government agency heard that their financial department is receiving an average of 6 complaints from the customers in a week. To solve the problem, he assigned his secretary to collect some data to see if he needs to replace the supervisor of that department. The director will replace the supervisor if the actual mean number of complaints towards the financial department is greater than 6 per week. The secretary gathered data over the next 12 weeks and discovered that the mean number of weekly complaints towards the financial department is 7 with a variance of 3.25.

a) Determine an appropriate statistical analysis to be used in this study.

(1 mark)

b) Calculate the *t-statistic* for this study.

(2 marks)

c) Test at the 5% significance level, is the director going to replace the department supervisor? Show the relevant steps.

(5 marks)

QUESTION 3

The scores of 18 students are summarised as below.

	N	$\sum x$	$\sum x^2$	Mode	Q3
score	18	1116	73366	67.00	74.25

a) Calculate the mean and standard deviation.

(4 marks)

b) Compute the coefficient of skewness. Hence, comment on the shape of the distribution.

(3 marks)

c) Explain the meaning of the value for third quartile **(Q3)** for this study.

(1 mark)

QUESTION 4

A professor at a local university wish to determine whether there is a significant difference in the average of final examination marks between the students who took his STA404 course online and face-to-face. Fifteen students were randomly selected from each group and the final examination marks were recorded. Hence, he analysed the data using IBM SPSS and the results are as follows.

Independent Samples Test

			Mark	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		2.041	
	Sig.		.164	
t-test for Equality of Means	t		-1.524	-1.524
	df		**W**	24.625
	Sig. (2-tailed)		.139	.140
	Mean Difference		-6.42000	-6.42000
	Std. Error Difference		4.21320	4.21320
	95% Confidence Interval of the Difference	Lower	**X**	-15.10390
		Upper	**Y**	2.26390

a) Are the variances of the two populations equal? Use α=0.05.

(3 marks)

b) Find the value of **W**.

(1 mark)

c) Calculate the values of **X** and **Y**.

(4 marks)

d) Based on the confidence interval obtained in c), is there any evidence to support that the average of final examination marks for students who took online class is different from face-to-face class? Give a reason to support your answer.

(2 marks)

QUESTION 5

A grocery chain wants to know if the three types of advertisements affect the mean sales differently. They used each type of advertisement at four different randomly selected stores for a month and measured the sales (RM '000) for each store at the end of the month. The results are as follow.

Descriptives

	Advertisement		Statistic
Sales	Type 1	Mean	11.5000
		Std. Deviation	3.41565
		Sum	46.00
	Type 2	Mean	10.0000
		Std. Deviation	3.26599
		Sum	40.00
	Type 3	Mean	7.5000
		Std. Deviation	2.51661
		Sum	30.00

ANOVA

Sales

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**A**	2	16.333	**D**	.235
Within Groups	86.000	9	**C**		
Total	**B**	11			

a) Using the sum of squares between groups formula, calculate the value of **A**.

(3 marks)

b) Compute the values of **B**, **C** and **D**.

(3 marks)

c) State the null and alternative hypothesis for the above study.

(1 mark)

d) Using the *p-value* method, is there any evidence to support that the types of advertisements affect the mean sales? Test at α=0.01.

(3 marks)

QUESTION 6

The lecturers of Mathematical Science Department from University M intended to study the association between the stress levels and the hours of online lessons in a week among accounting students. A questionnaire which aimed to assess the stress levels was administered to the respondents of the study. Their responses towards on the stress levels were categorised into low, medium, and high levels. The students were also asked to state the number of hours of their online lessons each week, according to the following category: less than 16 hours, 16 to 18 hours, 19 to 21 hours and more than 21 hours. The data were collected and the results are as follow.

Hours of Online Lessons * Stress Levels Crosstabulation

			Stress Levels			Total
			Low	Medium	High	
Hours of Online Lessons	Less than 16 hours	Count	17	71	18	106
		Expected Count	15.4	69.4	21.2	106.0
	16 - 18 hours	Count	18	92	37	147
		Expected Count	21.3	**S**	29.5	147.0
	19 - 21 hours	Count	22	97	28	147
		Expected Count	21.3	96.2	29.5	147.0
	More than 21 hours	Count	**T**	60	15	89
		Expected Count	12.9	58.2	17.8	89.0
Total		Count	71	320	98	489
		Expected Count	71.0	320.0	98.0	489.0

Chi-Square Tests

	Value	df
Pearson Chi-Square	4.032	**U**
Likelihood Ratio	3.963	6
Linear-by-Linear Association	.148	1
N of Valid Cases	489	

a) Give a reason for conducting the Chi-square Test of Independence for the above study.

(1 mark)

b) Compute the value of **S** using expected value formula.

(1 mark)

c) Calculate the values of **T** and **U**.

(2 marks)

d) State the null and alternative hypothesis for the above study

(1 mark)

e) At the 10% significance level, is there any sufficient evidence to conclude that the stress level is associated with the hours of online lessons in a week among the accounting students?

(4 marks)

QUESTION 7

A study was conducted to investigate the influence of the fathers' height on the sons' height. The heights (cm) of a random sample of fathers and sons were recorded and analysed by using IBM SPSS. The following results were obtained from the bivariate analysis.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.446	.199	.065	6.071

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	96.281	60.053		1.603	.160
	Heights of fathers (cm)	.432	.354	.446	1.220	.268

Answer the following questions based on the above output.

a) Name the independent and dependent variable involved in this study.

(2 marks)

b) State the correlation coefficient value. Hence, interpret the relationship between the variables.

(2 marks)

c) Write the least square regression equation.

(1 mark)

d) Based on the equation in c), comment on the slope value in the context of the above study.

(1 mark)

e) Predict the height of a son if the height of his father is 192 cm.

(2 marks)

END OF QUESTION PAPER

APPENDIX 1 (1)

SAMPLE MEASUREMENTS

Mean	$\bar{x} = \frac{\sum x}{n}$
Standard deviation	$s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]}$ or $s = \sqrt{\frac{1}{n-1}\left[\sum (x-\bar{x})^2\right]}$
Coefficient of Variation	$CV = \frac{s}{\bar{x}} \times 100\%$
Pearson's Measure of Skewness	Coefficient of Skewness = $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$ OR $\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$

APPENDIX 1 (2)

CONFIDENCE INTERVAL

Parameter and description	A (1 - α) 100% confidence interval
Mean μ, for large samples, σ^2 unknown	$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$
Mean μ, for small samples, σ^2 unknown	$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$; df = n – 1
Difference in means of two normal distributions, μ_1 - μ_2 $\sigma_1^2 = \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$; df = $n_1 + n_2 - 2$ $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
Difference in means of two normal distributions, μ_1 - μ_2 , $\sigma_1^2 \neq \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; $df = \frac{\left[s_1^2 / n_1 + s_2^2 / n_2 \right]^2}{\frac{\left(s_1^2 / n_1 \right)^2}{n_1 - 1} + \frac{\left(s_2^2 / n_2 \right)^2}{n_2 - 1}}$
Mean difference of two normal distributions for paired samples, μ_d	$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$; df = n – 1 where n is no. of pairs

APPENDIX 1 (3)

HYPOTHESIS TESTING

Null Hypothesis	Test statistic
$H_0 : \mu = \mu_0$ σ^2 unknown, large samples	$z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
$H_0 : \mu = \mu_0$ σ^2 unknown, small samples	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; df = n – 1
$H_0 : \mu_1 - \mu_2 = 0$ $\sigma_1^2 = \sigma_2^2$ and unknown	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$; df = $n_1 + n_2 - 2$ $s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
$H_0 : \mu_1 - \mu_2 = 0$ $\sigma_1^2 \neq \sigma_2^2$ and unknown	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $df = \dfrac{\left[s_1^2/n_1 + s_2^2/n_2\right]^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$
$H_0 : \mu_d = 0$	$t = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$; df = n – 1, where n is no. of pairs
Hypothesis for categorical data	$\chi^2 = \sum \dfrac{(o_{ij} - e_{ij})^2}{e_{ij}}$

APPENDIX 1 (4)

ANALYSIS OF VARIANCE FOR A COMPLETELY RANDOMIZED DESIGN

Let:

k = the number of different samples (or treatments)

n_i = the size of sample i

T_i = the sum of the values in sample i

n = the number of values in all samples

= $n_1 + n_2 + n_3 + ...$

$\sum x$ = the sum of the values in all samples

= $T_1 + T_2 + T_3 + ...$

$\sum x^2$ = the sum of the squares of values in all samples

Degrees of freedom for the numerator = k – 1
Degrees of freedom for the denominator = n – k

Total sum of squares: $SST = \sum x^2 - \frac{(\sum x)^2}{n}$

Sum of squares between groups:

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

Sum of squares within groups = SST - SSB

Variance between groups: $MSB = \frac{SSB}{(k-1)}$

Variance within groups: $MSW = \frac{SSW}{(n-k)}$

Test statistic for a one-way ANOVA test: $F = \frac{MSB}{MSW}$

APPENDIX 1 (5)

SIMPLE LINEAR REGRESSION

Sum of squares of *xy, xx*, and *yy*:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad \text{and} \quad SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

Least Square Regression Line:

$Y = a + bx$

Least Squares Estimates of *a* and *b*:

$$b = \frac{SS_{xy}}{SS_{xx}} \quad \text{and} \quad a = \bar{y} - b\bar{x}$$

Total sum of squares: $SST = \sum y^2 - \frac{(\sum y)^2}{n}$

Linear correlation coefficient: $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$