UNIVERSITI TEKNOLOGI MARA
TEST

COURSE	**:**	**STATISTICS FOR BUSINESS AND SOCIAL SCIENCES**
COURSE CODE	**:**	**STA404**
EXAMINATION	**:**	
TIME	**:**	**1 HOUR AND 40 MINUTES**

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of **SIX (6)** questions.
2. Answer ALL questions in the foolscap paper. Start each answer on a new page.
3. Candidates are given 1 hour and 40 minutes to accomplish this assessment.
4. Candidates are required to convert their completed answer in one PDF file before submission **(<FULLNAME_GROUP>.pdf**).
5. Candidates are given 30 minutes to email their completed answer to the respective lecturer.
6. Please check to make sure that this examination pack consists of :
   i) the Question Paper
   ii) a three–page Appendix 1
7. Answer ALL questions in English.

NAME:

STUDENT NO:

2	**0**								

GROUP:

	N4AM2253B		**N4AM2253D**		**N4AM2263A**
	N4AM2253C		**N4AM2262A**		**N4AM2263B**

Q1	**/8**
Q2	**/7**
Q3	**/10**
Q4	**/7**
Q5	**/11**
Q6	**/7**
TOTAL	**/50**
CLO 2 - 30%	**%**

PLEASE SUBMIT THIS ASSESSMENT ON THE REQUIRED TIME

This assessment paper consists of 9 printed pages

QUESTION 1

The National Sports Foundation conducted a national survey to assess the physical activity patterns of Malaysian adults. The table below shows the frequency (average number of days in the past year) and duration of time (average number of minutes per single activity) Malaysian adults spent participating in a sample of 11 sports activities.

activity	A	B	C	D	E	F	G	H	I	J	K
frequency	135	68	44	39	30	21	16	19	10	7	5
duration (in minutes)	43	99	61	60	80	100	91	127	249	115	262

Some SPSS output provided as follow.

$\Sigma x = 394$	$\Sigma y = 1287$	$\Sigma x^2 = 28438$	$\Sigma y^2 = 203651$	$\Sigma xy = 30535$

SSxx = 14325.64
SSyy = 53072
SSxy = -15563

NOTE:
In simple linear regression, the standardized coefficient Beta is equal to the correlation coefficient value.

$r = \dfrac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	155.912	26.929		5.790	.000
	frequency	**N**	.530	**M**	-2.051	.070

Assume the frequency and duration of time (in minutes) are normally distributed.

a) Calculate the **M** value. (3 marks)

$M = \dfrac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \dfrac{-15563}{\sqrt{(14325.64)(53072)}} = -0.5644$

b) Compute the **N** value. Hence, interpret the value in the context of the problem. (3 marks)

$N = \dfrac{SS_{xy}}{SS_{xx}} = \dfrac{-15563}{14325.64} = -1.086$

$t = \dfrac{B}{SE} \Rightarrow b = t * se = -2.051(0.53) = -1.08703$

c) Using the regression equation obtained from the output, estimate the duration of time (in minutes) Malaysian adults participate in a sport that they play 25 times a year. (2 marks)

$y = a + bx$ (a: constant, b: slope coef.)

$y = 155.912 - 1.086x$

Given $x = 25$

$y = 155.912 - 1.086(25)$

$= 128.762$ minutes

QUESTION 2

A study was conducted in order to determine whether the time spent per week (in hour) on watching television of the Program A is more than Program B. Hence, a sample of 35 and 40 students were selected to watch Program A and Program B respectively. Assume that the distribution of time spent per week (in hour) watching television is normally distributed. Hence, the result illustrated as follow.

t - stat

z - stat

sd / $\sqrt{n}$

Group Statistics

	Program	N	Mean	Std. Deviation	Std. Error Mean
Time Spent Watching TV (in hours/in week)	Program A	35	26.0857	4.58496	.77500
	Program B	40	13.5500	3.84941	.60864

$\frac{4.58496}{\sqrt{35}}$

a) Compute the test statistic value for the above study. (3 marks)

$z_{stat} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{26.0857 - 13.55}{\sqrt{\frac{4.58496^2}{35} + \frac{3.8494^2}{40}}}$

$z_{stat} = 12.7193$

b) Sate the null and alternative hypotheses for the above study. (1 mark)

$H_0: \mu_A = \mu_B$ $H_1: \mu_A > \mu_B$ (one-tail test)

$H_0: \mu_A - \mu_B = 0$ $H_1: \mu_A - \mu_B > 0$

c) Using the information in a) and b), can it be concluded the time spent per week (in hour) on watching television of the Program A is more than Program B among the students? Use α=0.1. (3 marks)

Step 1: State hypothesis (from part b)
Step 2: alpha = 0.1, Zcrit = Z(0.1) =1.2816 (from Table 4)
Step 3: compute test statistics (from part a)
Step 4: decision - reject H0 if Zstat > Zcrit. Since Zstat = 12.7193 > 1.2816, we reject H0.
Step 5: conclusion - we have enough evidence to support the claim that time spent per week (in hour) on watching TV of the Program A is more than Program B.

QUESTION 3

A group researcher claimed that the mean lifespan of four brands of batteries are equal. The agency randomly selected a few batteries of each brand and tested them. The following table gives the lifespan of these batteries in thousands of hours.

Brand A	Brand B	Brand C	Brand D
74	53	57	56
78	51	71	51
51	47	81	49
56	59	77	43
65		68	
$T_1 = 324$, $n_1 = 5$	$T_2 = 210$, $n_2 = 4$	$T_3 = 354$, $n_3 = 5$	$T_4 = 199$, $n_4 = 4$

$\Sigma x = T_1 + T_2 + T_3 + T_4 = 1087$

$n = 18$

$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4}\right) - \frac{(\Sigma x)^2}{n}$

$= \left(\frac{324^2}{5} + \frac{210^2}{4} + \frac{354^2}{5} + \frac{199^2}{4}\right) - \frac{1087^2}{18} = ?$

Some SPSS provided as follow.

ANOVA

Lifespan

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**P**	k − 1 **Q**	446.976	**R**	.007
Within Groups	1029.350	N − k 14	73.525		
Total	2370.278	N − 1 17			

$F = \frac{MS_B}{MS_W}$; Sig. → p-value

F_{α, v_1, v_2} : α = alpha, v_1 = df_1 = k − 1, v_2 = df_2 = N − k

a) State the response variable and factor for this study. (2 marks)

response = battery lifespan
factor = brands (A,B,C,D)

b) Calculate the value of **P** using the SSB formula. Hence, find **Q** and **R** values. (4 marks)

P = SSB = 1658.178

Q = k - 1 = 3
R = 446.976 / 73.525 = 6.079

c) Assume the lifespan for battery is normally distributed. At the 5% significance level, are the researchers' claimed true? Show the relevant steps. (4 marks)

$H_0: \mu_A = \mu_B = \mu_C = \mu_D$ (claim)
H_1: at least one μ is different.

using the critical value method:
F(0.05,3,14) = 3.34, F-stat = 6.079
decision: reject H0 if F-stat > F-crit. since F-stat = 6.079 > 3.34, we reject H0.
conclusion:

using p-value method:
alpha = 0.05, p-value = 0.007
decision: reject H0 if p-value < alpha. since p-value = 0.007 < 0.05, we reject H0.
conclusion: there is not enough evidence to support the researcher claim that the battery lifespan for the four brands are equal.

QUESTION 4

A researcher believes more than half of the students in School X spent the monthly tuition hours more than 15 hours in a week. Hence, in order to prove her claim, he selects a sample of twelve students. The data recorded as in the following table.

Student	1	2	3	4	5	6	7	8	9	10	11	12
Monthly Tuition (hours/per week)	10	12	25	21	19	15	18	22	17	15	12	23

$\bar{x} = 17.4167$ $s = 4.776$

Assume the monthly tuition hours are normally distributed. Do these data provide sufficient evidence to prove the researcher believes'? Use α=0.05. (7 marks)

$H_0: \mu = 15$
$H_1: \mu > 15$ (claim)

$\alpha = 0.05$, $t_{\alpha, df} = t_{0.05, 11} = 1.796$

$t_{stat} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{17.4167 - 15}{4.776/\sqrt{12}} = 1.7528$

Decision: reject H0 if t-stat > t-crit. since t-stat = 1.7528 < 1.796, we failed to reject the null hypothesis.

Conclusion: There is not enough evidence to support the researcher believes that more than half of the students in School X spent the monthly tuition hours more than 15 hours in a week.

QUESTION 5

The accompanying data were collected from a survey of health status of persons under 18 years of age in Selangor. Interviews concerning heath were conducted to determine whether there was a relationship between family income and health status. The data were collected and analyzed using SPSS. The results showed as below.

Health Status * Family Income Crosstabulation

			Family Income			
			High - Income	Middle - Income	Low - Income	Total
Health Status	Excellent	Count	72	46	32	150
		Expected Count	69.0	51.0	30.0	150.0
	Good	Count	33	32	10	75
		Expected Count	34.5	25.5	15.0	75.0
	Poor	Count	10	7	8	25
		Expected Count	**J**	8.5	5.0	25.0
Total		Count	115	85	50	250
		Expected Count	115.0	85.0	50.0	250.0

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

$$= \frac{(72-69)^2}{69} + \frac{(46-51)^2}{51} + \frac{(32-30)^2}{30} + \frac{(33-34.5)^2}{34.5} + \frac{(32-25.5)^2}{25.5} + \frac{(10-15)^2}{15} + \frac{(10-11.5)^2}{11.5} + \frac{(7-8.5)^2}{8.5} + \frac{(8-5)^2}{5}$$

$$= 6.4031$$

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	**K**[a]	4	.171
Likelihood Ratio	6.261	4	.180
Linear-by-Linear Association	.452	1	.501
N of Valid Cases	250		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.00.

a) What is the condition for this test to be valid?

Cells with expected count of less than 5 should not exceed 20%.

(1 mark)

b) Find the value of **J** and **K**.

$K = \chi^2 = \sum \frac{(O-E)^2}{E} = 6.4031$

J = 25 - 8.5 - 5 = 11.5

(6 marks)

c) Given the information provided in the output, can it be concluded that the health status is related to the family income? Use α=0.05.

(4 marks)

H0: Health status is not related of the family income
H1: Health status is related to family income

alpha = 0.05
p-value = 0.171

decision: reject H0 if p-value < alpha. since p-value = 0.171 > 0.05, we fail to reject H0.
conclusion: health status is NOT related to the family income

using critical value method.

chi-sq-cv = 9.488 (from Table)
chi-sq-stat = 6.4031

decision: reject H0 if chi-sq-stat > chi-sq-cv. since, chi-sq-stat =6.4031 < 9.488, we fail to reject H0.
conclusion: health status is NOT related to the family income

QUESTION 6

A researcher decides to see how effective a medication to the diabetic patients. Hence, the blood sugar level (BSL) reading taken from a sample of 13 randomly selected diabetic patients before the medication given to them. Then, after six hours, the researcher took the reading on the BSL reading again. The result as follow.

Paired Samples Test

			Pair 1 Before - After
Paired Differences	Mean		8.03846
	Std. Deviation		2.97084
	90% Confidence Interval of the Difference	Lower	6.56992
		Upper	9.50700
df			12
Sig. (2-tailed)			.000

By assuming the BSL reading is normally distributed. Answer the following questions.

a) State ONE (1) assumption to apply this statistical analysis. (1 mark)

1. samples are dependent
2. population is normally distributed

b) Compute t-statistic for this study. (2 marks)

$$t\text{-stat} = \frac{\text{mean diff}}{sd/\sqrt{n}} = \frac{8.03846}{2.97084/\sqrt{13}} = 9.7559$$

c) Using p-value method, at the 10% significance level, test whether the medication is effective. (4 marks)

H0: The medication is NOT effective in reducing BSl
H1: The medication is effective in reducing the BSL

alpha = 0.1
p-value < 0.001

decision: reject H0 if p-value < alpha. since p-value < 0.001, we reject the null hypothesis.
conclusion: there is enough evidence to support the claim that the medication is effective in reducing BSL

END OF QUESTION PAPER

APPENDIX 1 (1)

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	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$; 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	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $df = \dfrac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\dfrac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$
$H_0 : \mu_d = 0$	$t = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$; 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 (2)

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
	=	$n_1 + n_2 + n_3 + \ldots$
$\sum x$	=	the sum of the values in all samples
	=	$T_1 + T_2 + T_3 + \ldots$
$\sum x^2$	=	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: $SST = \sum x^2 - \frac{(\sum x)^2}{n}$

Between-samples sum of squares:

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

Within- samples sum of squares = SST - SSB

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

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

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

APPENDIX 1 (3)

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