UNIVERSITI TEKNOLOGI MARA
FINAL ASSESSMENT

COURSE	**:**	**STATISTICS FOR BUSINESS AND SOCIAL SCIENCES**
COURSE CODE	**:**	**STA404**
EXAMINATION	**:**	**14^{TH} FEBRUARY 2022**
TIME	**:**	**2 HOURS (1415 – 1615)**

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_UiTM ID_GROUP>.pdf).**
5. Candidates are given **30 minutes** to email their 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 12 printed pages

QUESTION 1

A random sample of monthly internet usage bills from three different telecommunication providers were collected among the university students at College H. The result as follow.

ANOVA

Internet Usage

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1613.771	**Q**	**R**	**S**	.069
Within Groups	**P**	12	238.773		
Total	4479.042	14			

a) What are the TWO (2) main assumptions for the model above?
(2 marks)

b) Compute the values of **P, Q, R** and **S.**
(4 marks)

c) Using p-value approach, is there any evidence to indicate the mean internet usage bills among the university students are different from the three telecommunication providers? Use $\alpha=0.05$.
(4 marks)

QUESTION 2

Hashim Motors Sdn. Bhd. specializes in selling a secondhand car. Currently, the company has 12 used cars for sale. The owner of the company wants to investigate the relationship between the age of the car and the mileage of the car. The data were collected and analyzed using SPSS. The results as follow.

Car's age (years)	6	4	2	2.5	3	4	3	5	5.5	4.5	4.5	3.5
Mileage ('000 km)	93	60	33	36	53	59	48	77	79	61	63	50

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975[a]	.951	.946	4.036

a. Predictors: (Constant), Age

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.927	4.134		.950	.365
	Age	13.997	1.002	.975	13.971	.000

a. Dependent Variable: Mileage

a) Prove that the correlation coefficient is 0.975.

(4 marks)

b) Interpret the value of the correlation coefficient in a).

(1 mark)

c) State the value of the coefficient of determination. Interpret its meaning.

(2 marks)

d) Based on the SPSS output, write the complete regression equation.

(1 mark)

e) Based on the regression equation in d), interpret the slope in the context of the problem.

(1 mark)

f) Predict the mileage of the car if the age of the used car is 4.3 years.

(1 mark)

QUESTION 3

A researcher conducted a study in determining the level of satisfaction on the welfare services provided by the employer in the banking sector. This study involved a total of 300 employees taken from five randomly selected banks from a total of 10 banks in Town A. All the employees from the selected banks were asked to answer the questionnaires. Among the questions asked were gender, length of service (years), type of welfare facilities and satisfaction towards welfare services provided by the employer (1=Good, 2=Average and 3=Poor).

a) State the population and sampling frame for the above study.

(2 marks)

b) Name any TWO (2) variables from the study. Hence, state its type of variable.

(2 marks)

c) Name the sampling technique employed in the study.

(1 mark)

d) Suggest the data collection method that suitable to the study. Give ONE (1) advantage of the method used.

(2 marks)

QUESTION 4

The owner of rental houses claims that the average length of occupancy is 3.2 years. A random sample of 35 rental contracts were selected and analysed using SPSS.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Length (years)	35	3.446	1.0551	**T**

a) Show that **T** is 0.1783.

(2 marks)

b) Construct a 95% confidence interval for the average length of occupancy of properties owned by the owner.

(3 marks)

c) Based on the confidence interval in b), is the owner’s claim true? Give a reason to support your answer.

(2 marks)

QUESTION 5

Ten employees were selected at random, and the data on the number of customers per month were recorded before and after the selected employees attending the workshop. The data were collected and analyzed using SPSS. The results as follow.

Paired Samples Test

	Paired Differences						
				95% Confidence Interval of the Difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df
Pair 1 Before - After	2.70000	3.71334	1.17426	.04364	5.35636	2.299	9

Assuming the number of customers is normally distributed.

a) The manager decided to analyze the data using Paired Samples Test. Give a reason to support his decision to use this statistical analysis.

(1 mark)

b) Show the value of *t-statistic* using appropriate formula.

(3 marks)

c) Can the insurance manager conclude that the workshop has increased the number of customers? Use $\alpha=0.10$.

(4 marks)

QUESTION 6

The following stem-and-leaf diagram represents the number of cars sold over a period of 16 days in January.

Stem	Leaf
5	0
6	1 6 9
7	0 1
8	0 5 8 9
9	6
10	0 7
11	2
12	5
13	1

Key: 5|0 means 50 cars

a) Calculate the mean and standard deviation for the number of cars sold in January.

(4 marks)

b) The mean and standard deviation of the cars sold in May and July is summarized in the following table. Using an appropriate measure, determine which month (January, May, or July) is most consistent in selling a car.

Descriptive Statistics

	N	Mean	Std. Deviation
No. of Car Sold in May	16	94.81	19.637
No. of Car Sold in July	16	91.42	20.561

(4 marks)

QUESTION 7

A researcher wants to investigate whether the education level is related to his or her place of residence. Hundred's people were selected randomly. The results are given as follow.

Area * Education Level Crosstabulation

			Education Level			
			SPM	Diploma	Bachelor's Degree	Total
Area	Urban	Count	22	10	18	50
		Expected Count	25.0	8.0	17.0	50.0
	Rural	Count	28	6	16	50
		Expected Count	25.0	8.0	**A**	50.0
Total		Count	50	16	34	100
		Expected Count	50.0	16.0	34.0	100.0

Chi-Square Tests

	Value	df
Pearson Chi-Square	**B**[a]	2
Likelihood Ratio	1.850	2
Linear-by-Linear Association	.778	1
N of Valid Cases	100	

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.00.

a) State a reason of choosing the Chi-Square Test of Independence for this study.
(1 mark)

b) Calculate the values of **A** and **B**.
(5 marks)

c) At the 5% significance level, test whether education level is associated to his or her place of residence.
(4 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 $(1-\alpha)$ 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, $\mu_1 - \mu_2$ $\sigma_1^2 = \sigma_2^2$ 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, $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; $df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$
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
	=	$n_1 + n_2 + n_3 + \ldots$
$\sum x$	=	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 = $\sum x^2 - \frac{(\sum x)^2}{n}$

Sum of squares between groups:

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

Sum of squares within groups = SST - SSB

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

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

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

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