UNIVERSITI TEKNOLOGI MARA
FINAL ASSESSMENT (ASSESSMENT 4)

COURSE	:	**STATISTICS FOR BUSINESS AND SOCIAL SCIENCES**
COURSE CODE	:	**STA404**
EXAMINATION	:	**16TH FEB 2021**
TIME	:	**(2.15 PM – 4.15) 2 HOURS**

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of **EIGHT (8)** 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 answers in one PDF file before submission **(<FULLNAME_GROUP>.pdf).**
5. Candidates are given 30 minutes to email their completed answers to the respective lecturers.
6. Candidates are required to attach the front page that includes the following details:

   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 13 printed pages

SELF DECLARATION FINAL ASSESSMENT:

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. This assessment is my own work.
3. I have not been involved and will not allow anyone to copy my work with the intention of passing it off as their own work.
4. I acknowledge that copying someone else's work (or part of it) is wrong and declare that my assessment is my own work.

QUESTION 1

The management of a supermarket wanted to investigate whether the male customers spend less money, on average than the female customers. The data were collected and analyzed as shown in the following outputs. Assume the variances are equal.

Independent Sample Test

			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		.151	
	Sig.		.703	
t-test for Equality of Means	t		-1.906	-1.919
	df		18	17.622
	Sig. (2-tailed)		.073	.071
	Mean Difference		-3.495	-3.495
	Std. Error Difference		**A**	1.821
	95% Confidence Interval of the Difference	Lower	**B**	-7.328
		Upper	**C**	.338

a) Proof the value **A = 1.8337**. Then, find the values of **B** and **C**.

(6 marks)

b) Based on the 95% confidence interval obtained, is there a significant difference in the average amount spend between the male customers and the female customers. Give a reason for your answer.

(2 marks)

QUESTION 2

The research manager of CMD Bank observed four tellers in the number of customers served per hour. The following table gives the number of customers served by the four tellers during each of the observed hours.

Teller W	Teller X	Teller Y	Teller Z
17	14	10	24
20	15	13	18
25	13	20	20
23	16	14	25
17	13	17	20

ANOVA

Number_customer					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**D**	**E**	69.4	**G**	0.01
Within Groups	152	16	9.5		
Total	360.2	19			

a) By using the SSB formula, proof the value of **D= 208.2**. Then, find the value of **E** and **G**.
(6 marks)

b) At the 5% significance level, test whether the mean number of customers served per hour by each of these four tellers is the same.
(3 marks)

QUESTION 3

An English class was given a trial 1 on writing composition. After trial 1, the class was given a four-page review handout to study for a week. Then, the class was given another trial. The table below shows the analysis of the result for trial 1 and trial 2.

Paired Sample Test

		Paired Differences			t	df
		Mean	Std. Deviation	Std. Error Mean		
Pair 1	Trial_1 – Trial_2	-10.167	2.787	1.138	-8.936	5

a) How many samples involved in this study?

(1 mark)

b) Use $\alpha = 0.05$, is there sufficient evidence that the scores improved?

(3 marks)

QUESTION 4

A researcher would like to see the dependency of working mothers and the school education level in homework responses in an online class. A random sample of 60 students each in pre-school, primary school, and secondary school are selected, and outputs are shown below.

Crosstabulation Mother * Level

			Level			Total
			Pre School	Primary School	Secondary school	
Working Mother	Yes	Count	29	38	51	118
		Expected Count	39.3	39.3	39.3	118.0
	No	Count	31	22	9	62
		Expected Count	20.7	20.7	20.7	62.0
Total		Count	60	60	60	180
		Expected Count	60.0	60.0	60.0	180.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**18.046**	**A**	.000
Likelihood Ratio	19.123	2	.000
Linear-by-Linear Association	17.763	1	.000
N of Valid Cases	180		

a) Find the value of **A**.

(2 marks)

b) Show that the value of Pearson Chi-Square value is **18.046**.

(3 marks)

c) At α = 0.10, test whether the homework response in an online class is dependent between working mothers and the school education level.

(4 marks)

QUESTION 5

In the automobile industry, customer service is a crucial factor affecting car sales. The management of a reputed automobile company is interested in determining the level of customer satisfaction with the service provided by the company's service centres. The company has altogether 40 service centres in the Northern Region. A sample of four centres was selected at random. Then all customers, who service their cars at these four service centres were selected for the study. A questionnaire was emailed to these customers.

a) State the population and sample for this study.

(2 marks)

b) State the variable for this study. Hence, identify the type of variable and the level of measurement.

(3 marks)

c) What is the best sampling technique to be used in this study? Justify your answer.

(2 marks)

QUESTION 6

The performances measured by the time taken (in seconds) for 20 finishers in the women's 1000-meter event at a championship were analyzed and the SPSS output is given below:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
time_taken	20	308.35	23.282	**P**

One-Sample Test

	Test Value = 300					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
time_taken	**Q**	19	.125	8.350	-6.54	23.24

a) Determine the values of **P** and **Q.**

(4 marks)

b) Based on the p-value in the SPSS output, is there any evidence that the average performance is significantly different from 300 seconds? Use $\alpha = 0.01$.

(3 marks)

QUESTION 7

An insurance company researcher has conducted a survey on the number of car thefts in several large cities for a period of 15 days last year. The data presented in the form of stem and leaf display is as follows:

5	0	1	3	4	4	7	
6	2	3	5	5	5	7	9
7	2	3					

a) Find the mode for the number of car thefts.

(1 mark)

b) Compute the mean and standard deviation of the number of car thefts**.**

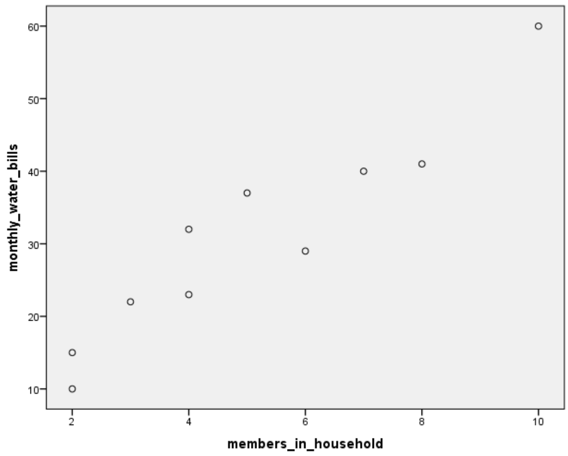
(4 marks)

c) Determine the shape of the distribution using an appropriate calculation?

(3 marks)

QUESTION 8

A research was carried out to identify the relationship between the members in a household and their monthly expenditure on water bills (to the nearest RM). The data collected randomly from ten households in area Bandar Laguna Merbok. The results are shown as follows:



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.945[a]	**X**	.880	5.048

a. Predictors: (Constant), members_in_household

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.378	3.618		1.210	.261
	members_household	5.200	.637	.945	8.170	.000

a. Dependent Variable: monthly_water_bills

a) Based on the scatter diagram, comment on the relationship between the two variables. (2 marks)

b) Determine the coefficient of determination. Interpret its meaning. (3 marks)

c) Estimate the monthly water bills if the members in the household is 9. (3 marks)

END OF QUESTION PAPER

APPENDIX 1 (1)

SAMPLE MEASUREMENTS

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

APPENDIX 1 (2)

CONFIDENCE INTERVAL

Parameter and description	A (1 - α) 100% confidence interval
Mean μ, for large samples, σ^2 unknown	$\bar{x} \pm z_{\alpha/2} \dfrac{s}{\sqrt{n}}$
Mean μ, for small samples, σ^2 unknown	$\bar{x} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$; df = n – 1
Difference in means of two normal distributions, μ_1 - μ_2 $\sigma_1^2 = \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$; df = $n_1 + n_2 - 2$ $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$
Difference in means of two normal distributions, μ_1 - μ_2 , $\sigma_1^2 \neq \sigma_2^2$ and unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$; $df = \dfrac{\left[s_1^2/n_1 + s_2^2/n_2 \right]^2}{\dfrac{\left(s_1^2/n_1 \right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2 \right)^2}{n_2 - 1}}$
Mean difference of two normal distributions for paired samples, μ_d	$\bar{d} \pm t_{\alpha/2} \dfrac{s_d}{\sqrt{n}}$; 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: $SST = \sum x^2 - \frac{(\sum x)^2}{n}$

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