UNIVERSITI TEKNOLOGI MARA
ASSESSMENT 1

COURSE	:	**STATISTICS FOR BUSINESS AND SOCIAL SCIENCES**
COURSE CODE	:	**STA404**
DATE	:	**1 DEC 2021**
TIME	:	**8.00 – 9.00 PM (1 HOUR)**

INSTRUCTIONS TO STUDENTS

1. This question paper consists of two (2) parts : PART A (2 Questions)
   PART B (3 Questions)

2. Answer ALL questions on your own Answer Sheet using pen.

3. Student must ensure that their test papers are readable. Ensure that your answers are written clearly with your name, group, student ID and lecturer's name are provided.

4. Upload all your answer as **ONE (1) PDF** document to the platform assigned by your lecturer. Please save your file name as "**NAME_GROUP**". Example "**AHMAD BIN ALI_KAC2205A**".

5. Students are given **30 minutes to submit** their completed answer to the respective lecturers.

5. If you face problem uploading the PDF document, please inform/notify your lecturer before the submission timeline ends (**9.30 PM**). All late submission WILL NOT be accepted.

6. Please check to make sure that this assesment pack consists of:
   i) the Question paper
   ii) a two - page Appendix 1 (List of Formulae)

7. Answer ALL questions in English.

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO

This examination paper consists of 5 printed pages

PART A

QUESTION 1

a) State whether the following statements are **TRUE** or **FALSE**.

i) Selecting sample every j^{th} element of the population is a method of systematic sampling technique.
ii) Nominal data is categorical data that can be ranked.
iii) Using information of sample to make inference about population is a descriptive statistics approach.
iv) One of the advantages of telephone survey is that respondents have more time to think of proper responses.
v) A list of all units in a population is called sampling frame.

(5 marks)

b) Research was conducted to examine the negative implication of online education to higher education institution undergraduate students from the educator's perspective. The research was carried out to all universities located in the Klang Valley. There are a total of 12 universities (both public and private) listed under the Ministry of Education database. An online questionnaire was distributed through email to randomly selected respondents from all the 12 universities. Among the information recorded were the respondents age, number of years with the current university, job position (Lecturer, Senior Lecturer, Associate Professor and Professor) and highest education level (Bachelors, Masters and PhD). In addition, the respondents were asked a YES/NO response on 15 items statement related to their perceptions towards the negative implication of online education.

Answer the following statements **TRUE** or **FALSE** based on the above research.

i) The population is all the undergraduate students at higher education institutions in Klang Valley.
ii) The sampling frame is the list of all public and private university in Klang Valley.
iii) The most appropriate sampling technique for the above research is stratified sampling.
iv) Job position takes the nominal level of measurement.
v) All the variables mentioned above are quantitative.

(5 marks)

QUESTION 2

In the recent year, the price of vegetables has increased drastically all over the country. The following table shows the price (in RM/kg) of 8 vegetables recorded in November 2021.

16	20	9	8	15	6	19	14

Based on this data,

a) Compute the mean and standard deviation.

(5 marks)

b) Given that the median price of vegetable in November 2021 is RM14.50/kg, calculate the Pearson's Measure of Skewness.

(2 marks)

c) The mean and variance price per kg for the same vegetables in November 2020 were recorded as RM7.25 and RM10.79, respectively. Determine which year has a more consistent distribution of vegetables price.

(3 marks)

PART B

QUESTION 1

The table below shows the summary statistics for the amount of money (in RM) spent by 30 customers at a grocery shop.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Amount spent (RM)	30	53.6892	16.7128	3.0513

a) Show that the standard error of the sample mean is 3.0513.

(2 marks)

b) Construct a 95% confidence interval for the true mean amount of money (in RM) spent by the customers at a grocery shop.

(3 marks)

QUESTION 2

A researcher wants to know if there is a significant difference in the Mathematics score between gender of a group of students in a private university. The data were analyzed using SPSS and the output are given below. The Mathematics score are assumed to be normally distributed.

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Math Score	Male	13	52.3846	7.0774	1.96292
	Female	12	53.7500	9.2846	2.68024

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Math Score	Equal variances assumed	1.085	.308	-.416	23	.682	-1.3654	-8.1633	5.4325
	Equal variances not assumed			-.411	20.546	.685	-1.3654	-8.2389	5.5081

a) Using Levene's test, determine whether there is any difference between the variances of the population at 5% level of significance.

(2 marks)

b) Calculate the standard error difference for this study.

(4 marks)

c) Construct the 95% confidence interval for the difference between mean score of Mathematics between gender.

(3 marks)

d) Based on the values obtained in (c), is there any enough evidence to indicate there is significant different in the Mathematics score between gender? Give a reason to support your answer.

(1 mark)

QUESTION 3

The management of a large plastic manufacturer is experiencing a major loss in income due to factory accidents. Manger decided to hire a trainer to educate all the workers on industrial safety. Below is the output of weekly loss of workings hours resulting from factory accidents in six of its factories both before and after safety program.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before	22.83	6	9.908	4.045
	After	21.83	6	9.152	3.736

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	98% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Before - After	**A**	1.673	.683			1.464	**B**	.203

a) Determine the value of **A** and **B**.

(2 marks)

b) Construct the 98% confidence interval of the mean difference between working hours lost due to the factory accidents before and after training program. Interpret.

(3 marks)

END OF QUESTION PAPER

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 $\dfrac{\text{mean} - \text{mode}}{\text{standard deviation}}$

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	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; $df = \frac{\left[s_1^2/n_1 + s_2^2/n_2 \right]^2}{\frac{\left(s_1^2/n_1 \right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2 \right)^2}{n_2 - 1}}$
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

