Fn. Tamkrul Ariffin

UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	**STATISTICS FOR BUSINESS AND SOCIAL SCIENCES**
COURSE CODE	:	**STA404**
EXAMINATION	:	**FEBRUARY 2023**
TIME	:	**2 HOURS**

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of seven (7) questions.

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

3. Do not bring any material into the examination room unless permission is given by the invigilator.

4. Please check to make sure that this examination pack consists of:

   i. the Question Paper
   ii. a five – page Appendix 1
   iii. an Answer Booklet – provided by the Faculty
   iv. a four – page Statistical Table – provided by the Faculty

5. Answer ALL questions in English.

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

This examination paper consists of 6 printed pages

QUESTION 1

The manager of KT Hotel claims that the average bills of guest on weekends are RM700 or less. However, the hotel's accounting department noticed that the total amount of guest bills has increased in recent months. To verify the manager's claim, a sample of weekend guest bills was taken. The results are as follows.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Guest Bills	20	705.800	114.569	25.619

One-Sample Test

Test Value = 700

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Guest Bills	.226	19	.823	5.800	-47.820	59.420

a) Using an appropriate formula, show how the *t-statistic* and degrees of freedom are obtained.

(3 marks)

b) State the null and alternative hypotheses for this study.

(2 marks)

c) Based on the p-value, is there sufficient evidence to support the manager's claim at the 5% significance level?

(4 marks)

QUESTION 2

The owner of a drugstore wants to test the effectiveness of two competing brands of mosquito repellent, X and Y. The test is conducted by the owner of the drugstore. Eight customers who agree to participate in the experiment are randomly selected. Each of these customers will be assigned to apply brand X to one arm and brand Y to the other arm. The table below shows the number of bites on both arms for each customer.

Customer	A	B	C	D	E	F	G	H
brand X	12	23	18	36	8	27	22	32
brand Y	9	20	21	27	6	18	15	25

a) Find the mean difference in the number of bites on the arm with brand X and Y, where the paired difference is defined as the number of bites on the arm with brand X minus the number of bites on the arm with brand Y.

(4 marks)

b) Show that the standard error of the mean difference is 1.463.

(3 marks)

c) Construct a 95% confidence interval for the mean difference.

(4 marks)

d) Name the statistical method that can be used to determine whether the mean difference between number of bites on the arm with brand X and Y is significant.

(1 mark)

QUESTION 3

The owner of a coffee store promoted his business through social media platforms such as Facebook, Instagram, and YouTube to increase the number of potential customers consuming more products from his store. He decided to investigate the relationship between monthly advertising costs spent on social media platforms (in RM hundred) and the monthly profit of store's (in RM thousand) using sample data from January to December 2021. The following information was obtained from the data of monthly advertising cost spent on social media platforms and monthly profit earned by the coffee store.

$$\sum X = 101 \qquad \sum Y - 64 \qquad \sum X^2 = 902.5 \qquad \sum y^2 = 380.5 \qquad \sum XY = 582$$

a) Using an appropriate statistical measure, determine the percentage of variation in the coffee store's monthly profit that can be explained by monthly advertising cost spend on social media platforms.

(4 marks)

b) Name the statistical measure used in a).

(1 mark)

c) It was found that the regression equation used to represent the relationship between monthly advertising costs spent through social media platforms and monthly profit earned by the coffee store is Y = – 1.625 + 0.827X. Based on this regression equation, interpret the meaning of the slope value.

(1 mark)

d) Using the regression equation given in c), estimate the monthly profit of the coffee store if the owner spent RM850 on advertising costs.

(2 marks)

QUESTION 4

A company wishes to investigate the opinion of its 1000 employees on a new salary scheme. A survey is to be conducted to gather some information about the new salary scheme. The company has three main departments, namely the human resources department, the sales and marketing department and the production department. 250 employees are in the human resources department, 450 employees are in the sales and marketing department and 300 employees are in the production department. A random sample of 100 employees is selected from each department.

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

(2 marks)

b) State the appropriate sampling technique. Explain how samples are selected using the stated sampling technique.

(3 marks)

c) State the most appropriate data collection method for the above study.

(1 mark)

d) State the advantage of using the data collection method in c).

(1 mark)

QUESTION 5

A lecturer wishes to investigate whether there is a relationship between students' educational level (diploma, bachelor's, master's) and their preference for a learning method (online distance learning, face-to-face) during the COVID-19 pandemic. To achieve the objective of the study, a random sample of 185 students was selected. The following result shows the findings of this study.

Level of Education* Preference

			Educational Level			
			Diploma	Bachelor's	Master's	Total
Preference	Online Learning Distance	Count	23	42	34	99
		Expected Count	34.2	37.5	27.3	99.0
	Face-to-face	Count	41	28	17	86
		Expected Count	29.8	32.5	K	86.0
Total		Count	64	70	51	185
		Expected Count	64.0	70.0	51.0	185.0

Chi-Square Tests

	Value	df
Pearson Chi-Square	12.579	2
Likelihood Ratio	13.468	2
N of Valid Cases	185	

a) Find the values of **K**.

(1 mark)

b) State the null and alternative hypotheses.

(2 marks)

c) At the 5% significance level, can it be concluded that there is a relationship between students' educational level and their preference for a learning method?

(4 marks)

QUESTION 6

The weight of children (in kg) selected for a study in a day care centre in Kuala Pilah were summarized as follows.

Weight of Children

4	2
5	3 4 5
6	3 3 4 5
7	1 1 2 6 7
8	8 9

Key: 4 | 2 means 4.2

a) Calculate the mean and standard deviation.

(4 marks)

b) Determine the median value.

(2 marks)

c) Interpret the median value obtained in b).

(1 mark)

d) State the name of the plot.

(1 mark)

QUESTION 7

Four different methods of lowering cholesterol were identified. Sixteen individuals with high cholesterol were identified and randomly assigned to the four different methods. After six months, the reduction in cholesterol level of each person is recorded as follows.

Method 1	Method 2	Method 3	Method 4
11	9	10	5
14	1	7	1
8	4	13	2
12	3	2	6

The data were analyzed using SPSS. The results were presented in the following output.

ANOVA

Cholesterol

	Sum of Squares	df	Mean Square	F
Between Groups	P	3	R	4.527
Within Groups	Q	12	11.375	
Total	291.000	15		

a) State the null and alternative hypotheses.

(1 mark)

b) Using the Sum of Square Between (SSB) formula, show that P = 154.500. Then, determine the value of Q and R.

(4 marks)

c) Based on the F-test, is there any evidence that there is a difference in the average reduction in cholesterol level among the four methods? Use $\alpha = 0.05$.

(4 marks)

END OF QUESTION PAPER

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

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, $\mu_1 - \mu_2$ $\sigma_1^2 = \sigma_2^2$ 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

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	$\chi^2 = \sum \dfrac{(o_{ij} - e_{ij})^2}{e_{ij}}$

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: $F = \frac{MSB}{MSW}$

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