UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	STATISTICS FOR BUSINESS AND SOCIAL SCIENCES
COURSE CODE	:	STA404
EXAMINATION	:	JUNE 2018
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 Statistical Tables – 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 8 printed pages

QUESTION 1 Chapter 1 - Introduction

SS Airlines has implemented a new boarding policy. In order to determine its customers' opinion of this new policy, a group of researchers made a list of all its flights and randomly selected 30 flights. All of the passengers on those flights were invited to answer a questionnaire during a certain week. One of the survey items was "Please rate your overall boarding experience today based on the following scale: 1- Excellent; 2- Good; 3-Fair; 4-Poor; 5- Very poor".

a) State the population and the sample of this study.
Population: all SS Airlines customers
Sample: 30 flights that were randomly selected (or All passengers of the 30 flights randomly selected)
(2 marks)

b) Name the sampling method used in this study. Give a reason for your answer.
Cluster Sampling.
Reason: Only 30 flights were selected and all passengers in the selected flights were invited to perform the survey.
(3 marks)

c) Identify the type of variable and the scale of measurement for the variable "boarding experience rating".
Variable Type: Qualitative
Scale of Measurement: Ordinal
(2 marks)

QUESTION 2 Chapter 2 - Descriptive Statistics

The following chart shows the recorded weekly milk yield (in the nearest kg) for each cow selected at random from Farm A.

12	9	9				
13	2	4	5	6	8	8
14	1	2	3	7	7	
15	1	3	5	8		
16	2	2	2	3	7	
17	2	3				

129, 129,
132, 134, 135, 136, 138, 138,
...

Key: 12 | 9 means 129

a) State the name of the above chart.
stem and leaf display
(1 mark)

b) Find the median weekly milk yield recorded at Farm A. Hence, interpret the result.
median location = (n+1)/2 = (24+1)/2 = 12.5
Median = (X12 + X13)/2 = (147 + 147)/2 = 147
$X_{12} = 147$ $X_{13} = 147$
Half of the cows yielded weekly milk of less than 147 kg.
(2 marks)

c) The statistics for the weekly milk yields for Farm A and Farm B are summarized in the following table. Using an appropriate measure, determine which farm has more consistent weekly milk yield.

Descriptive Statistics

	N	Mean	Std Deviation
Farm A	24	148.7	13.7
Farm B	28	158.5	8.3

(5 marks)

Farm A
CV = 13.7/148.7 x 100% = 9.21%

Farm B
CV = 8.3/158.5 x 100% = 5.24%

Farm B has a more consistent weekly milk yield

QUESTION 3 Chapter 3 - Estimation

The table below shows the summary statistics for the traveled distance from home to work (in km) by 28 employees of ABC company.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Distance	28	14.3	**X**	0.4914

$\frac{s}{\sqrt{n}}$

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

$SE = \frac{s}{\sqrt{n}} \rightarrow 0.4914 = \frac{x}{\sqrt{28}} \rightarrow x = 2.6002$

(3 marks)

b) Construct a 90% confidence interval for the true mean traveled distance from home to work (in km) by the employees of ABC company. $t_{\alpha/2, 27} = 1.703$

$90\% \rightarrow \alpha = 0.1$

$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$, $df = n-1$

Thus, 90% confidence interval is given by,
14.3 - 1.703(0.4914) < mu < 14.3 + 1.703(0.4914
13.4631 < mu < 15.1369

$\alpha/2 = 0.05$

(4 marks)

c) Based on the confidence interval in (b), is the mean traveled distance from home to work by the employees of ABC company different from 15 km? Give a reason to support your answer.

13.46 — 15 — 15.14

There is enough evidence to support the mean traveled distance is 15 km since the value is included in the confidence interval.

(2 marks)

QUESTION 4 Chapter 4 - Independent sample t test

Two groups of drivers are surveyed to see how many kilometers per week they drive for pleasure trips. The recorded data were analyzed using SPSS software. The output of statistical analysis is shown in the following tables.

Group Statistics

	Type of drivers	N	Mean	Std. Deviation	Std. Error Mean
Distance in km	Single drivers	35	196.4779	32.62732	5.51502
	Married drivers	35	189.1209	25.94217	4.38503

Independent Samples Test

			Distance in km	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		1.102	
	Sig.		.298	
t-test for Equality of Means	t		1.044	1.044
	df		68	64.714
	Sig. (2-tailed)		.300	.300
	Mean Difference		7.35700	7.35700
	Std. Error Difference		7.04585	7.04585
	95% Confidence Interval of the Difference	Lower	-6.70277	-6.71570
		Upper	21.41677	21.42971

a) Based on p-value in the Levene's Test, test the equality of variances in this study. Use $\alpha = 0.05$.

H0: Equal variances assumed; H1: Equal variances not assumed
Reject H0 if p-value < 0.05. Here we have p-value = 0.3 > 0.05, therefore we failed to reject the H0. Hence, the equal variance assumption is assumed.

(3 marks)

b) State the null and alternative hypotheses to test whether single drivers do more driving on average than married drivers for pleasure trips.

Null: Driving for pleasure trips is the same for single and married drivers
Alternative: Driving pleasure trips for single drivers is more than married drivers

$H_0: \mu_s = \mu_m$ $\quad H_1: \mu_s > \mu_m$ $\quad H_0: \mu_s - \mu_m = 0$ $\quad H_1: \mu_s - \mu_m > 0$

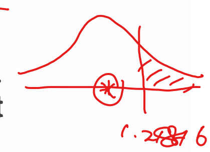
(2 marks)

c) At 10% significance level, can it be concluded that the single drivers do more driving for pleasure trips on average than married drivers?

(3 marks)

① $CV = z_\alpha = z_{0.1} = 1.2816$ ② $z_{stat} = \dfrac{(\bar{x}_s - \bar{x}_m) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \dfrac{7.357}{\sqrt{\frac{32.62732^2}{35} + \frac{25.96217^2}{35}}} = 1.044$

④ Decision: Reject H_0 if $z_{stat} > z_{cv}$



QUESTION 5

A secretarial training school is experimenting with four different manuals for a typing course. The school divided 20 students into four classes, and each class used a different manual. At the end of the training session, a test was given, and the scores shown in table below.

Manual			
A	B	C	D
67	89	75	78
79	86	95	74
85	87	69	95
86	73	94	85
79	87	60	93
79.2	84.4	78.6	85

Gm = 81.8

$P = 5(79.2-81.8)^2 + 5(84.4-81.8)^2 + 5(78.6-81.8)^2 + 5(85-81.8)^2$
$= 170$

The data were analyzed by using SPSS and the following results were produced.

$MS = \frac{SS}{df}$

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**P** 170	**Q** 3	56.667	**T**	.663
Within Groups	1687.2	**R** 16	**S** 104.8875		
Total	1857.2	19			

$SS = MS \times df$
$= 56.667 \times 3$
$= 170$

$\frac{56.667}{104.8875} = 0.5403$

a) Compute the values of **P**, **Q**, **R**, **S** and **T**.

(5 marks)

P = Total − Within Group = 1857.2 − 1687.2 = 170.

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

(2 marks)

c) Based on the p-value, test at the 5% level of significance whether the four different manuals create different effects.

(3 marks)

p-value = 0.663 > 0.05 ⇒ Fail to reject Ho.
There is not enough evidence that the four diff manual create different effects.

QUESTION 6

A travel agency was curious about whether the service a guest receives is related to the size of the hotel. A sample of 300 customers was selected at random to gather the information. The data were analyzed using SPSS and the following tables were obtained.

Opinion on services * Size of Hotel Crosstabulation

			Size of Hotel			Total
			Large	Mid-size	Small	
Opinion	Satisfied	Count	80	40	**E** = 30	150
		Expected Count	80.0	45.0	25.0	150.0
	So-so	Count	60	30	10	100
		Expected Count	53.3	30.0	16.7	100.0
	Dissatisfied	Count	20	20	10	50
		Expected Count	**F**	15.0	8.3	50.0
Total		Count	160	90	50	300
		Expected Count	160.0	90.0	50.0	300.0

$\frac{50 \times 160}{300} = 26.7$

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.722[a]	G	.068
Likelihood Ratio	9.081	4	.059
Linear-by-Linear Association	.117	1	.732
N of Valid Cases	300		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.33.

$(r-1)(c-1) = (3-1)(3-1) = 4$

a) Compute the values of **E**, **F** and **G**.

(3 marks)

b) State the null and alternative hypotheses to test whether the customers' opinion and the size of the hotel are related.

Null hypothesis: customers' opinion and the size of hotel are not related.
Alternative hypothesis: customers' opinion and the size of hotel are related.

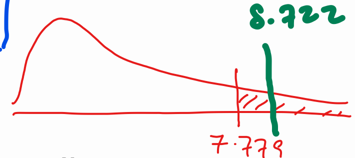
(2 marks)

c) Based on the p-value, state your decision and conclusion for the above test. Use α = 0.10.

(3 marks)

$\chi^2_{0.1,4} = 7.779$

obtain from table

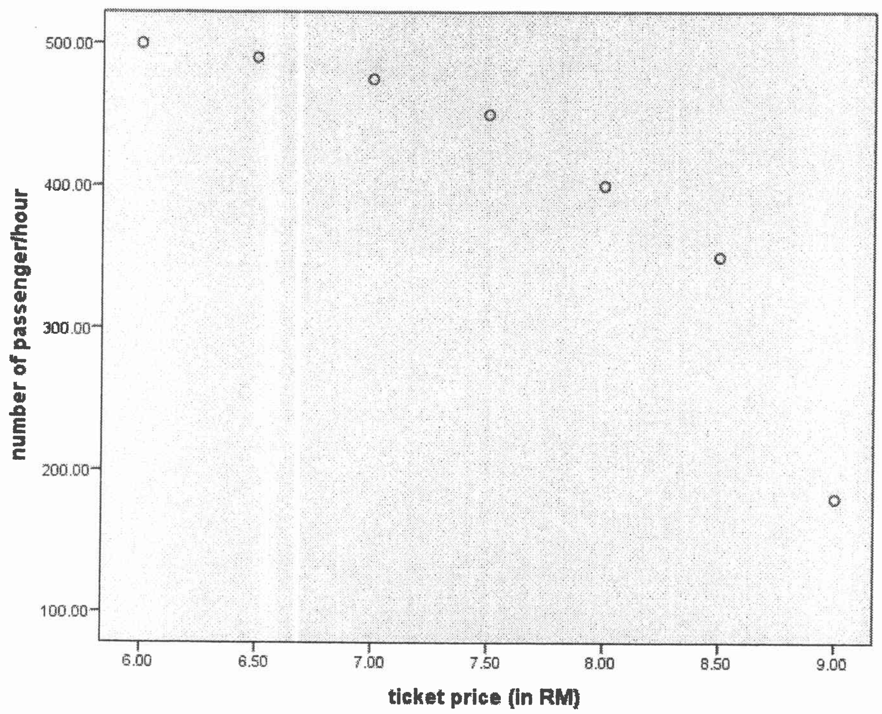


reject null

QUESTION 7

A study was conducted to investigate the effects of train ticket prices on the number of passengers per hour. The data are plotted on a scatter diagram and the SPSS outputs are given below.

Ticket Price (in RM)	6.00	6.50	7.00	7.50	8.00	8.50	9.00
Passengers/Hour	500	490	475	450	400	350	180



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.897[a]	.804	.765	54.91389

a. Predictors: (Constant), ticket price (in RM)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1110.893	157.044		7.074	.001
	ticket price (in RM)	-93.929	20.756	-.897	-4.525	.006

a. Dependent Variable: number of passenger/hour

a) Based on the scatter diagram above, briefly describe on the relationship between the two variables.

There is a strong negative relationship between number of passengers and ticket price (2 marks)

b) Using SPSS output, write the linear regression equation.

(2 marks)

y = 1110.893 - 93.929x
where y = number of passenger/hour
x = ticket price (in RM)

c) What does the slope tell you about the ticket price and the passenger/hour?

slope = -93.929. Foe each additional increase in ticket price by one unit (RM1), the total number of passengers per hour will decrease by approximately 94 passengers.

(1 mark)

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

r-square = 0.804. Interpretation: 80.4% variations in number of passengers per hour can be explained by the amount of ticket price.

(3 marks)

e) Estimate the number of passenger/hour if the train ticket price is RM7.90.

x = 7.9

y = 1110.893 - 93.929(7.9) = 368.9 passengers/hour

(2 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	$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$
Mean μ, for small samples, variance σ^2 unknown	$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$; df = n – 1
Difference in means of two normal distributions $\mu_1 - \mu_2$, variances $\sigma_1^2 = \sigma_2^2$ 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 $\mu_1 - \mu_2$, variances $\sigma_1^2 \neq \sigma_2^2$ 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

HYPOTHESIS TESTING

Null Hypothesis	Test statistic
$H_0 : \mu = \mu_0$ σ^2 known, large samples	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ or $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
$H_0 : \mu = \mu_0$ σ^2 known, 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	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $df = \dfrac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$
$H_0 : \mu_d = 0$	$t = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$; $df = n - 1$, 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: $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$

