INDEX NO: .....

SECTION 2: 6 QUESTIONS (60) MARKS

Choose 04 out of the 06 questions given below and answer using the provided answer sheets.

2.1 The chart shows the number of tear gas (lachrymator) canisters used in a certain country during the period 1998-2008.

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Canisters	130	94	40	40	555	54	35	38	67	81	125

a) Arrange the data in ascending order and calculate the summation of number of canisters. *(3 marks)*

b) Calculate the average number of canisters used over the 11 years. *(2 marks)*

c) Calculate the median (Q_2) and quartiles (Q_1, Q_3) of number of canisters. (Clearly show steps of the calculations). *(5 marks)*

d) Find the range and Inter quartile range (IQR) of the number of canisters. *(2 marks)*

e) Sketch the box plot for above obtained results. *(3 marks)*

2.2 On average Sri Lankan University students score 100 in an IQ test. A team of psychologists tested whether providing a two-month training makes a difference in the IQ scores. A sample of 30 participants who had gone through the training has a mean of 140 with a standard deviation of 20. Did the training affect IQ scores? Test your hypothesis at a 5% alpha level.

a) Write the null and alternative hypothesis. *(2 marks)*

b) Calculate the test statistics value using following formula. *(5 marks)*

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} =$$

c) Find the degree of freedom to determine the t-table value. *(2 marks)*

d) Write down the decision rule. (t-table value at a 5% alpha level is 2.0452) *(3 marks)*

e) Write the conclusion. *(3 marks)*

2.3 This question is based on the following data set of variables time spent in fuel queue (X) and the level of perceived frustration (Y) based on a self-reported questionnaire

X (Minutes)	60	34	40	50	45	41	22	43
Y (Perceived frustration out of 100)	75	32	34	40	45	33	12	30

a) Sketch the scatter plot using data of X and Y variable. (Take X variable as the horizontal axis and Y variable as the vertical axis.) *(3 marks)*

b) Interpret the graph output. *(2 marks)*

c) Fill the following table. *(5 marks)*

No.	X	Y	X^2	Y^2	XY
1	60	75			
2	34	32			
3	40	34			
4	50	40			
5	45	45			
6	41	33			
7	22	12			
8	43	30			
Total					
Average					

d) Calculate the correlation coefficient using following formula. *(3 marks)*

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} =$$

e) Comment on the value obtained in the part (d). *(2 marks)*

2.4 Briefly explain the following concepts and use examples where needed.

a) Inferential Statistics *(3 marks)*

b) Descriptive Statistics *(3 marks)*

c) Control variables *(3 marks)*

d) Parametric Test Vs Non-Parametric Tests *(3 marks)*

e) Null hypothesis *(3 marks)*

2.5 Provide answers based on the following study.

A study was conducted to examine whether wearing a helmet reduced the chances of a rider suffering a head injury after meeting with an accident. The objective of the analysis is to test whether there is an association between the two variables. Use 5% level of significance.

Table 2.5.1

Head Injury	Wearing Helmet		Total
	Yes	No	
Yes	17	218	235
No	130	428	558
Total	147	646	793

a) State a suitable hypothesis and a null hypothesis for this study. *(5 marks)*

Figure 2.5.2 Output of the Chi-Square test

Head Injury * Wearing Helmet Crosstabulation

			Wearing Helmet		
			No	Yes	Total
Head Injury	No	Count	428	130	558
		Expected Count	454.6	103.4	558.0
	Yes	Count	218	17	235
		Expected Count	191.4	43.6	235.0
Total		Count	646	147	793
		Expected Count	646.0	147.0	793.0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	28.255[a]	1	.000		
Continuity Correction[b]	27.202	1	.000		
Likelihood Ratio	32.543	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	28.220	1	.000		
N of Valid Cases	793				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 43.56.

b. Computed only for a 2x2 table

a) Is the Chi-Square test valid? Provide reasons for your answer. *(5 marks)*

b) State your statistical decision and conclusion *(5 marks)*

2.6 A researcher conducts a study to test the relationship between 'Depression level at the beginning of the semester' and the 'Depression level at the end of the semester'.

Table 2.6.1: Depression at the start and end of the semester

	Shapiro-Wilk		
	Statistic	df	Sig.
Depression level at Time 0	.945	12	0.572
Depression level at Time 1	.938	12	0.475

a) What Correlational analysis would you select based on the Shapiro-Wilk test? *(5 marks)*

b) Based on table 2.6.2, what can you say about the relationship between depression level at the beginning of the semester and the depression level at the end of the semester? *(5 marks)*

Table 2.6.2: Results of the Correlation Analysis

		Depression level at	
		Time0	Time1
Depression level at Time0	Pearson Correlation	1	.860**
	Sig. (2-tailed)		.000
	N	12	12
Depression level at Time1	Pearson Correlation	.860**	1
	Sig. (2-tailed)	.000	
	N	12	12
**. Correlation is significant at the 0.01 level (2-tailed).			

c) What is the appropriate statistical test to test the hypothesis for the following research question if the data were normally distributed?

"Do males and females differ in terms of their means of exam scores" *(2 marks)*

d) Mention at least **3 conditions** for test you mentioned for question 'c'. *(3 marks)*

----------END OF QUESTION PAPER----------