RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES

B. Sc. General Degree
Third Year - Semester II Examination – September/ October 2013

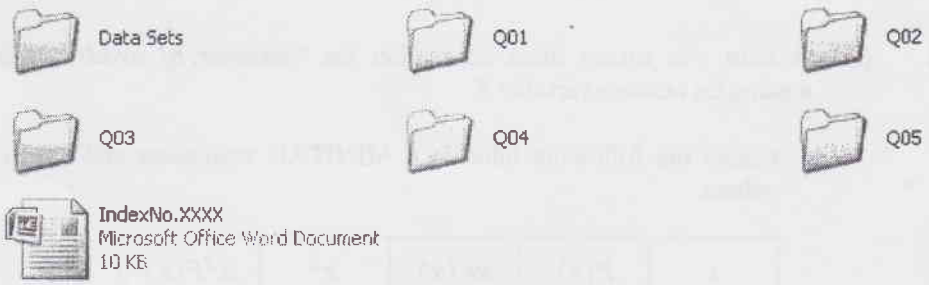
MAT 3206 – DATA ANALYSIS USING COMPUTER SOFTWARE

Answer all questions. **Time: 3 hours**

Carefully read the following instructions before you start answering the paper.

Instructions

- Answer **all** the questions.
- Consider the folder available on your desktop labeled with your index number as your working directory. This folder contains sub-folder for each question and one word document. The contents of the folder are graphically represented bellow.



- You are expected to save all your Minitab files inside the sub-folder relevant to the respective question.
- Write your *index number* in the Microsoft word document titled "*Answerscript.doc*". Include all the relevant answers (relevant commands copied from the session window, relevant outputs and relevant interpretations) in this document file.
- **Your are strongly advised** to save your "*Answerscript.doc*" file regularly to avoid loss of data due to technical problems such as power failures.
- Only the "*Answerscript.doc*" file and Minitab files saved in the relevant folders will considered for evaluation.
- All relevant data sets are available in the Dataset folder in the working directory.
- Clearly state all the hypotheses and assumptions used.

01. Use data *question01.txt* to answer all sections under question 01.

(i). Describe nature of the given data set [10]

(ii). Use a graphical method that is most appropriate to visually depict the relationship between the variables given below. In each case describe the graph you obtained.

a. Distance from the apartments to their working places in relation to nationality (A – Asian B – Black W – White).

b. Rent charged and distance from the apartment to the working place.

c. Workers' marital status and whether they are smokers or non-smokers. [30]

(iii). Use descriptive statistics option to briefly describe the relationship between workers' income and their house rental charges in relation to their nationality. [20]

(iv). Assuming normality and equal variance, statistically test whether the income is affected by the nationality of the workers. Consider **W**-white as Group I and **A, B**–as the Group II.

[Hint: use code data option to form the two groups.] [40]

02. (i). A coin was tossed three times. Let the ***"number of heads"*** obtained during tossing be random variable X.

a. Create the following table in a MINITAB worksheet and fill in the column values.

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$

[20]

b. Using the following equations find the μ and σ of X.

$$Mean = \mu = \sum xP(x)$$

$$Variance = \sigma^2 = \sum x^2 P(x)$$

[10]

[Continued] ***P.T.O***

02. Continued

(ii). Suppose a lady is pregnant and is due in 100 days. Suppose that the probability density distribution function of gestation period is approximately normal with mean 100 and standard deviation 8 days. Her husband has a business trip that will end in 85 days and has planned to go on another business trip on the 107th day.

a. What is the probability that the birth will occur before husband's second trip?

b. What is the probability that the birth will occur after husband's return from his first business trip?

c. What is the probability that husband will be there for the birth?

d. He is able to cancel his second business trip. When should he return home in order to give himself a 99% chance of staying home for the birth? [40]

(iii). Ten recently diagnosed diabetics were tested to determine whether an educational program was effective in increasing their knowledge of diabetes. They were given a test before and after the educational program concerning self-care aspects of diabetes. The scores on the test were as follows:

Patient	1	2	3	4	5	6	7	8	9	10
Before	75	62	67	70	55	59	60	64	72	59
After	77	65	68	72	62	61	60	67	75	68

Using the appropriate statistical test, test whether the educational program has had a significant impact on their knowledge on diabetes.

[30]

03. (i). The amount of money spent on health care is an important issue for workers because many companies provide health insurance that only partially covers many medical procedures. The director of employee benefits at AMZ company wants to determine the amount spent on health care by the typical hourly worker in the company. A random sample of 25 workers is selected and the amount they spent on their families' health care needs during the past year is given here.

400 345 248 1,290 398 218 197 342 208 223 531 172 4,321

143 254 201 3,142 219 276 326 207 225 123 211 108

Graph the data using a normal probability plot and determine whether the population has a normal distribution.

[15]

[Continued] *P.T.O*

03. Continued

(ii). An airline wants to evaluate the depth perception of its pilots over the age of 50.A random sample of n =14 airline pilots over the age of 50 are asked to judge the distance between two markers placed 20 feet apart at the opposite end of the laboratory. The sample data listed here are the pilots' error (recorded in feet) in judging the distance.

2.7 2.4 1.9 2.6 2.4 1.9 2.3
2.2 2.52.3 1.8 2.5 2.0 2.2

Use the above data test whether the average error distance of a pilot is less than 3 feet or not. [20]

(iii) An experiment was conducted to evaluate the effectiveness of a treatment for tapeworm in the stomachs of sheep. A random sample of 24 worm-infected lambs of approximately the same age and health was randomly divided into two groups. Twelve of the lambs were injected with the drug and the remaining twelve were left untreated. After a 6-month period, the lambs were slaughtered and the worm counts were recorded in Table.

Drug-Treated Sheep	18	43	28	50	16	32	13	35	38	33	67	*
Untreated Sheep	40	54	26	63	21	37	39	23	48	58	28	39

a. Generate the box plot and comment on distribution of the data.

b. Statistically check the validity of the distribution suggested by your box plot. [40]

(iv). The data correspond to a sensory analysis where 10 experts have been asked to rate the hardness of four different cheeses during two separate blind sessions, on a scale [0-5]. Here, we consider that different experts rated the cheeses. Our goal is to determine if the difference in hardness between the cheeses is significant or not. Use a suitable statistical test to check the significance of the data in ***chees.xls***. [25]

04. (i). It has been claimed that the mean weight of female students at a college is 54.4 kg. The principal of the college does not believe this claim and sets out to show that the mean weight is not 54.4 kg. To test the claim he collected a random sample of 100female students. The sample means was 53.75 kg. Is this sufficient evidence for principle to reject the statement? Use $\sigma = 5.4\ kg$. [15]

(ii). Using the data available in ***corr.txt***, generate the scatter plot and explain the relationship between the two variables.
Then statistically substantiate the relationship depicted graphically. [25]

(iii). Fit a simple linear regression model to the given data in (ii) and discuss the adequacy of the model. [40]

(iv). Test marks of the ten students in mathematics and science are available in ***exammarks.txt*** file. Using an appropriate test, check whether the science marks are greater than mathematics marks. [20]

05. (i). Generate suitable graphical output to analyze the following data set and interpret the graph. The dataset is on four glassware compositions and their durability.

Composition	Durability	Composition	Durability
A	18.95	D	10.92
C	12.62	B	13.28
C	11.94	B	14.52
B	14.42	C	12.51
C	11.06	D	10.46
D	7.19	A	21.40
D	7.03	A	18.10
B	14.66	A	22.50

[20]

(ii). Design an experiment to assess the durability of four experimental carpet products. A sample of each of the carpet products was placed in four houses and the durability was measured after 60 days. Assuming normality, using a appropriate statistical test, check the equality of means and assess the differences in means. The data set is available in the ***exc.xls*** file.

[***Hint: you can use one of multiple comparisons to assess the difference in means.***] [40]

(iii). Research biologist is studying how *zooplanktons* live in two lakes. He sets up twelve tanks in his laboratory, six each with water from one lake. He adds three nutrient supplements; one each to two tanks of each lake and after 30 days he counts the *zooplankton* in a unit volume of water. He wants to know whether there is any significant effect from supplement, lake or interaction of both of them on the population of *zooplankton.* Counts are available in the ***zoop.xls*** file.

Using appropriate statistical assumptions and procedures provide the result of the analysis to the biologist. [40]

.....***.....