



**RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES, MIHINTALE**

B.Sc. (General) Degree Examination

**Third Year Semester II– End Semester Examination – November 2014
Data Analysis Using Statistical Software – MAT3206**

Answer ALL questions.

Time allowed: 3 hours

Note:

- You should prepare a text document (**MS-WORD**) including all the necessary outputs and relevant graphs. Save the file with the name **XXXX.doc**, where **XXXX** represents your index number.

(E.g.: If the index number is “**2000**” then the text document should be “**2000.doc**”)

- Write down the relevant answers and their interpretations in the answer scripts provided to you.
- Save the following files, naming as indicated below.
 - Minitab worksheet as “**XXXXQX.MTW**”
 - Minitab project as “**XXXXQX.MPJ**”

Where, “**XXXX**” represents your index number and ‘**X**’ in ‘**QX**’ represents the question number.

- **Save all necessary files in the home directory ‘H:\’**
- **You are advised to save your works regularly to avoid losses of data due to power failures.**

1. a). An experiment was conducted to determine the effect of three methods of soil preparation on the first-year growth of slash pine seedlings. Four locations (state forest lands) were selected, and each location was divided into three plots as soil fertility within a location was likely to be more homogeneous than between locations. The methods of soil preparation were A (no preparation), B (light fertilization), and C (burning). Each soil preparation was randomly applied to a plot within each location. On each plot the same number of seedlings was planted, and the observation recorded was the average first-year growth (cm) of the seedlings on each plot. These observations are reproduced in the table below. (Use $\alpha = 0.05$)

Soil Preparation	Location			
	Location 1	Location 2	Location 3	Location 4
A	11	13	16	10
B	15	17	20	12
C	10	15	13	10

- (i). Import the "*Q1a.txt*" dataset and identify the experimental design that can be applied in the above situation.
- (ii). Do the data provide sufficient evidence to indicate differences in the mean growth for the three types of soil preparations? Clearly state the hypothesis and interpret results.
- (iii). Perform the Tukey's multiple comparison test to identify the different soil preparations.
- b). An electronics engineer is interested in tube conductivity of five different types of coating for cathode ray tubes in a telecommunications system display device. The following conductivity data are obtained. (Use the "*Q1b.txt*" dataset)

Coating Type	Conductivity			
A	143	141	150	146
B	152	149	137	143
C	134	133	132	127
D	129	127	132	129
E	147	148	144	142

- (i). Identify the different types of coating using a suitable plot.
- (ii). Is there any difference in conductivity due to coating type at the 5% significance level?
- (iii). Using a suitable test identify the different coating types, if there are any.
2. a). Arsenic in public drinking water supplies is a potential health risk. The "*Q2a.txt*" file contains the reported drinking water arsenic concentrations in parts per billion (ppb) for 10 metropolitan communities and 10 communities in rural area.
- (i). Define the hypothesis of interest to check the normality of data. Perform the relevant test and interpret the results.
- (ii). Define hypotheses for the equality of the variances, perform relevant test, and interpret the results.
- (iii). Using the results of part (i) and (ii), identify and perform the required test to check whether there is any significant difference in the mean arsenic concentration in two locations at the 5% significance level.

[Continued...]

64

- b). The following are the compressive strengths of a material in N/mm^2 , manufactured by two different methods A and B.

A	60.3	50.2	56.5	60.6	59.3	49.7	50.8	59.8	52.5	57.4	55.8	54.5	53.6	56.7	57.1
B	56.0	56.2	55.1	59.2	62.3	54.5	56.5	57.1	56.2	56.1	58.5	63.5	58.2	48.9	53.0

- (i). Identify one parametric and one non-parametric test that can be applied to test whether the differences between compression strengths of materials manufactured by the two methods are significant. .
- (ii). Compare the median of two series using a suitable graphical method and interpret it. (Use the "Q2b.txt" dataset)
- (iii). Assuming that above data are non-normal, apply an appropriate non-parametric test to investigate whether method B is superior to method A. Did you obtain the same conclusion as in the part (ii). (Use $\alpha = 0.05$)
3. a). In an extensive orchard (*enclosed land*) there are 1000 trees grown under two different conditions, shaded and non shaded. The trees are low yielding and high yielding. The researcher wants to identify the effect shading on yield

Conditions	Yield	
	Low Yielding	High Yielding
Shaded	250	350
Non shaded	195	205

Library
Faculty of Applied Science
Rajshree University of Sri Lanka
Minimale.

- (i). Clearly state the hypothesis for the above scenario.
- (ii). Perform relevant test to test the above hypothesis at 5% significance level and interpret your results.
- b). The diameter of a dot, produced by a printer is normally distributed with a mean diameter of 0.002 inches and a standard deviation of 0.0004 inches.
- (i). What is the probability that the diameter of a dot **does not** exceed 0.0026 inches?
- (ii). Find the median of the above distribution.
- c). The "Q3c.txt" file represents the travelling distances for employees of the Pay-and-Carry Department store of a particular company.
- (i). The manager of the company wants to categorize the data into three groups according to the distances travelled by employees. Categorize the dataset into groups and name the newly created variable as "Group".

Distance	Group
Distance < 3 miles	A
Distance < 6 miles	B
Distance \geq 6 miles	C

- (ii). Construct a pie chart to illustrate the frequency and percentage of employees in each category.

4. a). Data available at "*Q4a.txt*" described by X_1 = *systolic blood pressure* , X_2 = *age in years* , X_3 = *weight in pounds*
- Using suitable graphical methods comment on the relationship between the above variables.
 - Statistically test the above explanations at the 5% significance level
- b). A researcher wishes to examine the factors effecting the population density of small cities of a particular state. The "*Q4b.txt*" dataset contains the relevant data by small cities and the variable description is as follows.

Variable	Description
X_1	Death rate per 1000 residents
X_2	Doctor availability per 100,000 residents
X_3	Hospital availability per 100,000 residents
X_4	Annual per capita income in thousands of dollars
X_5	Population density people per square mile

- Use a suitable graphical method to identify the relations between each variable.
- Identify the relevant dependent variable and independent variables for each of the relationships above.
- Obtain the multiple linear regression model to predict the population density of cities using all independent variables.
- Define the hypothesis for the appropriateness of the model and interpret the results.
- Use the forward selection method to find the best model and state the best regression model.
- How do you identify the best model using the model selection statistics in the output?

*****END*****