



RAJARATA UNIVERSITY OF SRI LANKA
FACULTY OF APPLIED SCIENCES, MIHINTALE

B.Sc. (General Degree) in Applied Sciences

Third Year Semester II– End Semester Examination – October 2014
Statistical Computing SAS – MAT3209

Answer ALL questions.

Time allowed: 3 hours

Note:

- You should prepare a text document (**MS-WORD**) to include all the necessary outputs and relevant graphs. Save the file with the name **XXXX.doc**, where **XXXX** represents your index number.
(E.g.: Your index number is “2000” then the text document should read as “2000.doc”)
- Write down the relevant answers and their interpretations in the answer scripts provided to you.
- Save the SAS program as **XXXXQX.sas**, 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 loss of data due to power failure.

1. a). The reaction times for two different stimuli in a psychological word-association experiment were compared by using an independent random sample of 8 subjects for each stimulus. Assume data is approximated by the normal distribution and the variances of two distributions are equal.

Stimulus 1	1	3	2	1	2	1	3	2
Stimulus 2	4	2	3	3	1	2	3	3

- (i). Define the hypothesis of interest difference in the mean reaction time of the two stimuli are statistically significant.
- (ii). Test the above mentioned hypothesis at the 5% significance level? Interpret your results briefly.
- b). In order to study the number of hours to relief by five different brands (A, B, C, D, E) of pain relievers for different causes of pain provided [headache (H), muscle pain (M), pain due to cuts and bruises (CB)], doses are administered to five subjects each having similar types of pain.

		Types of Pain		
		H	M	CB
Brands	A	3	5	1
	B	1	4	2
	C	2	3	4
	D	5	1	3
	E	4	2	5

- (i). Identify the experimental design that can be applied in to above situation.
- (ii). Write down the SAS program to print the dataset.
- (iii). Check whether there is a difference in mean hours to relief provided by five different brands with respect to types of pain at 5% significance level and interpret the results.
2. a). Accident victims are sometimes taken by helicopter from the accident scene to a hospital. The helicopter may save time and also brings medical care to the accident scene. The below data represents percent of accident victims who die with helicopter evacuation and with the usual transport to hospital by road.

Serious Accidents		
	Road	Helicopter
Died	60	48
Survived	40	52

Less Serious Accidents		
	Road	Helicopter
Died	200	16
Survived	800	84

- (i). Using the "Do" statement, write down the SAS program to print the two contingency tables.
- (ii). Determine the risk of transporting victims by road in comparison to helicopter using the relative risk.

[Continued...]

- b). In a hospital, patients admitted with chronic kidney disease were asked whether they used purified water or non-purified water. At the same time the researcher asked the same questions from the patients who do not have chronic kidney disease. The following table represents the count of each category.

		Kidney Disease	
		Yes	No
Type of Water	Non-Purified	175	80
	Purified	95	190

- (i). Write down the relevant hypothesis to check the dependency between the two variables.
 - (ii). Perform the relevant test to identify the dependencies between the two variables at 5% significance level and interpret the result.
 - (iii). The researcher claims that the people who used non purified water are more likely to suffer chronic kidney disease than who used the purified water. Confirm the researcher's claim using a suitable statistic.
3. Details of the mortality rates of selected hospitals in United States are listed in the dataset "*Hospital.txt*" in your home folder (H:/). The dataset consist of seven variables. Descriptions of the variables are given below.

Variable	Description
Name	Name of the Hospital
State	Name of the state where the hospital is located (1=Alaska, 2=California, 3=Florida, 4=Georgia, 5=Indiana, 6=Kansas)
Ownership	Ownership of the hospital (Government, Private, Church)
MR1	Mortality rate from heart attack.
No_HA	Number of patients admitted for heart attack
MR2	Mortality rate from heart failure.
No_HF	Number of patients admitted for heart failure

- (i). Write a SAS program to print the dataset without the observation numbers.
- (ii). Code the variable "State" according to the given descriptions above, using suitable SAS statements.
- (iii). Find the state where the highest numbers of hospitals are located by using a suitable SAS procedure.
- (iv). Find the mean of mortality rates that are from heart attack and heart failure and store it as a new variable "Mortality".

[Continue...]

[P.T.O.]

[Continued ...]

- (v). According to the newly created variable "Mortality" in part (iv), the government of United States wants to categorize the performance of the hospitals into groups. The table below shows how the government wishes to categorize the hospitals. Find the group of the hospitals according to the table given below.

Mortality	Performance
Less than 12	Good
$12 \leq \text{Mortality} < 15$	Better
Greater than or equal to 15	Bad

- (vi). Obtain a suitable graph to represent the number of hospitals in each state. Add a relevant title to the graph.
4. The regional sales manager for Toys, recently collected data on weekly sales (in Rupees) for the 15 stores in his region. He also collected data on the number of sales clerk working hours during the week for each of the stores. The data are given in the text file "*Sales.txt*".
- Construct a suitable plot to explore the relationship between sales and sales clerk working hours and interpret the result.
 - Using a relevant correlation measurement identify the strength of the above relationship.
 - Obtain a least square regression model by identifying the independent and dependent variables.
 - Write down the regression model assumptions.
 - Using relevant graphs and procedures validate above mentioned assumptions.
 - Interpret the fit of the model using related statistics.

*****END*****