



71

Library
Faculty of Applied Sciences
Rajarata University of Sri Lanka
Mihintale.

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

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

MAT 3209 – STATISTICAL COMPUTING

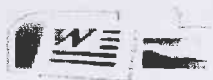
Answer any four (04) questions.

Time: 3 hours

Read the following instructions carefully before commencing to answering the paper.

Instructions

- Please check whether you have permission to access the working directory (H:\IndexNumber XXXX) on your computer.
- A folder labeled with your index number is available in your working directory on your computer. This folder contains two sub-folders and one word document. The contents of the folder are graphically represented below.



Index No



SAS files saving
folder



Data and other
files available
folder

- You are expected to save all your SAS files inside the sub-sub-folders allocated to each question available in the "SAS files saving folder".
- Rename the Microsoft documents labeled "Index No" in your working directory with your index number. All relevant output and statistical statement relevant to each question should be pasted in the Microsoft document.
- Only the "Answerscript.doc" file and SAS files saved in the relevant folders will be considered for evaluation.
- All relevant data set and program files are available in the "Data and other files available folder" in the working directory.
- Clearly state all the hypotheses and assumptions made.
- **Your are strongly advised** to save your "Answerscript.doc" file regularly to avoid loss of data due to technical problems such as power failures.

01. a). Write a SAS program to compare the sales records of three sales people given below by computing the sum and mean for the number of visits, phone calls, and units sold for each salesman.

Salesperson	Target company	Number of visits	Number of phone calls	Units sold
Brown	American	3	12	28000
Johnson	Asian	6	14	33000
Rivera	Europe	2	6	8000
Brown	Europe	0	22	0
Brown	Asian	2	19	12000
Rivera	Asian	4	8	13000
Rivera	American	8	7	27000
Johnson	Europe	3	16	8000
Johnson	American	2	14	2000

[15]

- b). Plot the number of visits against the number of phone calls. Use the value of salesperson (the first character in the name) as the plotting symbol (instead of the usual A, B, C, etc.).

[15]

- c). Make a frequency bar chart for unit sold by each salesperson to target companies.

Hint : Use options Block / GROUP sumvar Type=sum discrete available in PROC CHART

[20]

- d). Following data obtained from an experiment.

Subject	DOSE	REACT	LIVER_WT	SPLEEN
1	1	5.4	10.2	8.9
2	1	5.9	9.8	7.3
3	1	4.8	12.2	9.1
4	1	6.9	11.8	8.8
5	1	15.8	10.9	9.0
6	2	4.9	13.8	6.6
7	2	5.0	12.0	7.9
8	2	6.7	10.5	8.0
9	2	18.2	11.9	6.9
10	2	5.5	9.9	9.1

Use **PROC UNIVARIATE** procedure to produce histograms, normal probability plots, box plots for variables REACT, LIVER_WT, and SPLEEN and test each of these variables for normality. Carryout these procedures for all subjects and then separately for each of the two DOSES.

[50]

02. a). Open the file *q02.sas* and identify the consistency of the file. Then change the variables labeled *id*, *msch* and *region* to string type. [10]
- b). Using PROC MEANS procedure finds the descriptive statistics for available variables. Output should have a meaning full "Title". When numerical statistical output is longer than three decimal places, the values should be rounded to the third decimal place. [15]
- c). Create a new variable *rstate* and assign the values to it using "If" conditions according to the table given below.

Region		R_State
1	-	Hillstate
2	-	Northstate
3	-	Weststate
4	-	Technostate

Then find the descriptive statistical measures for variables *length* and *census* for each state.

Hint: Use CLASS option.

[25]

- d). Develop the code used in part c) to create a new data set called *summarydata* which should include N, MEAN and STD derived from applying the PROC MEANS procedure to the variables *length* and *census*.

Hint: Using these options NOPRINT NWAY NOOBS and (DROP=_TYPE_) to smooth your outputs.

[50]

03. a). The income and team details of 60 randomly selected basketball players were recorded using "," space in *bb.txt* file. Write a SAS program to read the data file. [10]
- b). Make a frequency bar chart for each team for the sum of "players' income."

Hint: PROC CHART procedure and use its option VBAR ... / SUMVAR= TYPE=SUM ;

[10]

- c). Using the data given below, create a SAS data set called DEMOG

Weight	Height	Gender
155	68	M
98	60	F
202	72	M

Weight	Height	Gender
280	75	M
130	63	F

[10]

We want to record weight and height as follows,

Height 0 – 70 = 1
 > 70 = 2

Weight 0 – 100 = 1
 101 – 150 = 2
 152 – 200 = 3
 > 200 = 4

We then want to generate a table of WEIGHT categories (rows) by HEIGHT categories (columns). Recode these variables in two ways:

- (1) with "IF" statements
- (2) with formats

[15]

[30]

Then write the necessary statements to generate the table.

- d). The table below is a summary of the relationship between socio-economic status (SES) and asthma.

Asthma	Yes	No
LOW SES	40	100
HEIGH SES	30	130

Create a SAS data set from these data and statistically check whether there is an association between SES and the occurrence of asthma. [25]

04. a). A survey is conducted, collected and coded. The data layout is shown below:

Variable	Description	Columns	Coding Values
ID	Subject identifier	1-3	-
GENDER	-	4	M=Male F=Female
PARTY	Political party	5	'1' = Republican '2' = Democrat '3' = Not registered
VOTE	Did you vote in the last election	6	O = No 1= Yes
FOREIGN	Do you agree with the government's foreign policy?	7	O = No 1= Yes
SPEND	Should we increase domestic spending?	8	O = No 1= Yes

Collected data are shown next:

007M110
013F2101
137F1001
117F1111
428M3110
017F3101
037M2101

- i. Create a SAS data set, complete with labels for this questionnaire. [10]
 - ii. Generate frequency counts for the variables GENDER, PARTY, VOTE, FOREIGN, and SPEND. [10]
 - iii. Test if there is a relationship between voting in the last election versus agreement with spending and foreign policy (Have SAS compute chi-square for these relationships). [15]
- b). The following table shows the time for subjects to feel relief from headache pain:

Aspirin (Relief time in minutes)	Tylenol
40	35
42	37
48	42
35	22
62	38
35	29

Write a SAS program to read these data and perform a t-test. Is either product significantly faster than the other (at the 0.05 level)?

Hint: use PROC TTEST procedure and CLASS option

[25]

- c). The codes given below are relevant to the computation of Chi-square for any number of 2 X 2 tables,

```
DATA CHISQ;
N + 1;
  DO ROW = 1 TO 2;
    DO COL = 1 TO 2;
      INPUT COUNT @;
      OUTPUT;
    END;
  END;

DATALINES;
3 5 8 6
10 20 30 40
;
RUN;

PROC FREQ DATA=CHISQ;
  BY N;
  TABLES ROW*COL / CHISQ;
  WEIGHT COUNT;
RUN;
```

Use above code and write *SAS* code to analyze the following cases and interpret the results.

Case A

		PIGS	
		M	F
GENOTYPE	A1	55	45
	A2	70	30
	A3	82	18

Case B

		RATS	
		M	F
Hormone level	B1	6	18
	B2	7	21
	B3	21	9

Case C

		RABITS	
		M	F
Cell level	C1	20	15
	C2	70	60
	C3	10	25

Case D

		GOATS	
		M	F
Microbe Type	D1	12	10
	D2	80	50
	D3	8	40

[40]

05. a). We randomly assign 15 subjects to three treatment groups X , Y , and Z (with five subjects per treatment). Each of the three groups has received a different method of speed-reading instruction. A reading test is given, and the number of words per minute is recorded for each subject. The following data are collected:

X	Y	Z
700	480	500
850	460	550
820	500	480
640	570	600
920	580	610

- i. Write a SAS program and compute the Pearson correlation coefficient between X and Y ; X and Z . Interpret your results?

Hint : use option WITH

[10]

- ii. Change the correlation request to produce a correlation matrix, that is, the correlation coefficient between each variable versus every other variable.

[15]

b).

- i. Generate the plot Y versus X [10]
- ii. Compute a regression line (Y on X). Y is the dependent variable, X the independent variable. [20]
- iii. What is the slope and intercept of the fitted line and determine whether it is significantly different from zero or not. [20]
- iv. A plot of the regression line and the original data on the same set of axes. [10]
- v. Conduct residual analysis for the fitted model and comment on it [15]