

LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – STATISTICS

THIRD SEMESTER – NOVEMBER 2009

ST 3813 - STATISTICAL ANALYSIS USING R

Date & Time: 23/10/2009 / 9:00 - 11:00 Dept. No.

Max. : 100 Marks

SECTION A

Answer any FOUR questions.

(4x5=20)

1. Create two objects x and y where x = 12, 10, 5, 9, 16, 20, 21 and y = 43, 47, 35, 54, 49, 32, 60.
 - a.) Display all values of x greater than 11.
 - b.) Compute $e^{\sqrt{y+x^3}}$.
 - c.) Find the cross product of x and y.
 - d.) Sort the values of x in descending order.
 - e.) Display the minimum and maximum values of y.
2. Create two objects height and gender consisting of the following values.

Height:	121	135	116	142	150	137	128	169
Gender:	M	F	F	F	M	M	F	F

 - a.) Find the average height of males and females.
 - b.) Draw a box plot of the height of males and females. What information does the plot reveal?
3. Attach the data set '*VADeaths*'. Draw a bar diagram of the Rural Male and Rural Female. Give suitable title for the bar diagram and appropriate labels for the x and y axis.
4. Create the following 2 x 5 matrix.

	1	2	3	4	5
X = 1	11	34	56	12	36
2	10	8	5	18	20

Find $X^T X$ and trace of $X^T X$.
5. Generate a random sample of size 1000 from an exponential distribution with mean 3.7. Plot the cumulative distribution function of the above samples. Find the first and third quartile values.

SECTION B

Answer any FIVE questions.

(5x10=50)

6. Draw a histogram of age based on the following data by grouping the age in the classes 15-20, 20-25, 25-30 and 30-35. Super impose the density curve of the normal distribution of age on the histogram. Interpret the result.

Age:	18	25	32	17	34	28	22	24	27	33
------	----	----	----	----	----	----	----	----	----	----

7. a.) Compute $P(25 < X < 75)$, $P(X > 50)$ and $P(X < 30)$ given that $X \sim N(25,5)$.
 b.) Find the expected frequencies of X given that $X \sim B(5,0.5)$.
8. Form a two way frequency table of gender and marks based on the following data by grouping marks in the interval 0-35, 35-60, 60-80 and 80-100.
- | | | | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|----|----|
| Marks : | 88 | 54 | 32 | 21 | 47 | 98 | 67 | 56 | 47 | 77 |
| Gender: | M | M | F | F | F | M | M | M | F | M |
- Perform a Chi-square test to verify whether there is any association between gender and marks.
9. Use '**Orange**' data set. Perform a one way analysis of variance to test whether the mean age of different Trees are equal by assuming (i) equal variances and (ii) unequal variances. Interpret the results.
10. Use the first 4 columns of '**iris**' data set. Perform principal component analysis and compute the principal component scores. Interpret the results. Display the correlation matrix of the principal component scores.
11. The following data gives the diastolic blood pressure of 10 patients before and after the administration of a drug.
- | | | | | | | | | | | |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Before: | 148 | 158 | 152 | 124 | 157 | 148 | 136 | 128 | 140 | 147 |
| After: | 130 | 147 | 154 | 118 | 142 | 136 | 128 | 120 | 127 | 139 |
- Use paired t test to test whether the drug is effective.

SECTION C

Answer any TWO questions.

(2x15=30)

12. Use '**attitude**' data set. Build a linear regression model by taking 'rating' as the dependent variable and the remaining variables as the independent variables using backward method of elimination. Identify the best model. Interpret the results. Find the residuals and fitted values based on the best model.
13. Fit a binary logistic model of Y on X_1 and X_2 based on the following data where Y takes the values 0 (failure) and 1 (success).
- | | | | | | | | | | | |
|---------|----|----|----|----|----|----|----|----|----|----|
| Y: | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| X_1 : | 45 | 55 | 32 | 21 | 18 | 64 | 78 | 21 | 69 | 89 |
| X_2 : | 58 | 60 | 17 | 16 | 5 | 72 | 80 | 33 | 46 | 51 |
- Find the predicted probability of success for all the observations.
14. Perform a two way analysis of variance of the following data and interpret the results.

Blocks	Treatments					
	A	B	C	D	E	F
1	24.7	20.6	27.7	16.2	16.2	24.9
2	27.3	28.8	22.9	15.0	17.0	22.5
3	38.5	39.5	36.8	19.6	15.4	26.3
4	28.5	31.0	34.9	14.1	17.7	22.6