B.Sc. DEGREE END SEMESTER EXAMINATION - MARCH 2019

SEMESTER – 6: STATISTICS FOR (B. Sc. COMPUTER APPLICATIONS)

COURSE: **15U6CRCST7: COMPUTER AIDED DATA ANALYSIS USING EXCEL & R**

(Common For Regular - 2016 Admission / Supplementary-Improvement 2015 admission)

Time: Three Hours   Max. Marks: 75

PART A

Each question carries five marks, maximum marks from this part is 25

1.   Explain the advantages of R as software.

2.   Explain how to import a data to R from Excel.

3.   For the following data set

0.5	1.2	2.1	2.5	2.5	2.5	3.0	3.8	4.0	4.2	4.5	5.0	5.0
6.5	7.0	8.0	9.5	13.0	5.0	5.0	6.0					

   (a)   Enter the data using scan function and store it in variable x.

   (b)   Find the mean, median, mode and Variance of x using R.

4.   Annual sales of a pharmaceutical firm for six years are given below

Year	1995	1996	1997	1998	1999	2000
Annual sales	15.0	25.0	27.0	28.0	26.0	26.6

   Represent the data by bar chart in R.

5.   Explain the method of finding Pearson's correlation coefficient in Excel. Also calculate the correlation coefficient for the following data using Excel

x	2	3	4	5	6	7	8
y	4	5	6	12	9	5	4

6.   Write the R code to generate 50 binomial random numbers with parameters n=10 and P=0.75. Also find the mean and Variance of the generated data.

7.   Explain the purpose of the following functions given in Excel with examples.
     NORMDIST, NORMINV.

8.   For the following data of the scores of two batsmen A and B in six innings the coefficient of variation using Excel. Comment on the consistency of the two batsmen in scoring

Batsmen A	12	12	80	70	60	100	0	4
Batsmen B	8	9	7	10	5	9	10	8

PART B

Each question carries ten marks, maximum marks from this part is 50

9.   (a) Following data pertain to measurements of cholesterol (X) and triglyceride (Y) levels for 10 patients in suitable units

X	5.12	6.18	6.77	6.65	6.36	5.90	5.48	6.02	10.34	8.51
Y	2.30	2.54	2.95	3.77	4.18	5.31	5.53	8.83	9.48	14.20

   Calculate Pearson's correlation coefficient using R.

(b) Obtain the equations of lines of regression X on Y for the following data. Also obtain the estimate of X for Y=70 using R.

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

10. Obtain mean, median and mode using R for the following data

X	1	2	3	4	5	6	7	8	9
Frequency	8	10	11	16	20	25	15	9	6

11. Calculate the mean, standard deviation, mean deviation for the following table giving age distribution of 542 members using R.

Age	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of members	3	61	132	153	140	51	2

12. An insurance agent has claimed that the average age of policy holders who insure through him is less than the average for all agents, which is 30.5 years. A random sample of 100 policy holders who had insured through him give the following age distribution

Age	16-20	21-25	26-30	31-35	36-40
No. of persons	12	22	20	30	16

Calculate the arithmetic mean and standard deviation of this distribution and use these values to test his claim at 5% level of significance using Excel.

13. In a study of cardio vascular risk factors among young adults, the proportion of hyper-tension was reported in men and women between ages 26 and 32 years.

Hypertension

Gender	Abnormal	Normal
Male	41	1120
Women	15	1042
Total	56	2162

Using MS- Excel can we conclude on the basis of these data that in the proportion of hyper-tension is same in men and women?

14. To study the effect of a drug on anaemia patients the drug was administered to 6 patients and their Hb were noted before and after the administration of the drug. Test whether the drug is effective at 5% level of significance. Use MS – Excel. (Two Tail Test)

Patient	1	2	3	4	5	6
Hb before	10	12	8	11	10	9
Hb after	12	11	9	10	11	8

15. Create a data set of numbers 1 to 50 in excel and read that data in R and compute its Mean, variance and standard deviation.

****