(b) Simulate n = 100 values from normal distribution (X) with mean 45 and sd 5 and compute the following :

(i) $P(X < x) = .05$

(ii) $P(X > 50)$

(iii) $P(X \leq 30)$

(iv) $P(35 < X \leq 45)$

Further, using the values of X generated above, write R-code for the following :

- Arrange the values of X in an ascending order.
- Compute values of the PDF and CDF of the corresponding normal distribution, over the values obtained in the step above.
- Generate a histogram using the generated values of X.

Obtain separate plots for these within a single figure window. (5,5)

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 2508A **IC**

Unique Paper Code : 32373902

Name of the Paper : Statistical Data Analysis Using R

Name of the Course : **B.Sc. (Hons.) Statistics : SEC**

Semester : IV

Duration : 2 Hours Maximum Marks : 50

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt **six** questions in all.
3. **Section I** is compulsory.
4. Attempt **four** more questions, selecting **two** questions from each of **Section II** and **III**.
5. Write R codes for each question given in **Section II** and **III**.

Section – I

1. Answer the following :

   (i) Write down the R-command syntax for computing the probability of getting at most five heads in six tosses of a coin.

   (ii) What is the output of the following R-command: >punif(1)?

   (iii) Write down the R-code to create a vector with the following as its content:

   1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5

   (iv) Write the output of the following R code:

   X < – c(5,28,11.8,7,8.7,20,19.3,20.6,19,20.8)

   Y < – X[-(3:6)]; print(Y)

   (v) When external Excel or Text data files, that contains some missing observations, are read into R, then how does R pad out this missing information? (1×5)

2. (a) Write the output of the following R codes :

   > X = c(TRUE,8,9,FALSE); X

   > Y = rep(seq (1,2,by=0.5),2); Y

Assuming some arbitrary values for death rates, give R-code to: (i) enter the data in an appropriate format required to produce the above bar graph; and (ii) write detailed arguments required to produce the above bar diagram. (6,4)

8. (a) Below are given the temperatures of two states :

   tem1 = c(18.7,20,23,20,27,28,29,18,19,20)

   tem2 = c(34,36,38,40,37,38,40,34,45)

   Consider the output below :

   data: tem1 and tem2

   t = -5.2055, df = 14.377, p-value = 0.0001223

   alternative hypothesis: true difference in means is not equal to 0

   95 percent confidence interval:

   -15.94441 -6.65559

   sample estimates:

   mean of x mean of y

   22.7 34.0

   From the output given above, test if the two temperatures differ significantly. Also, interpret your results.

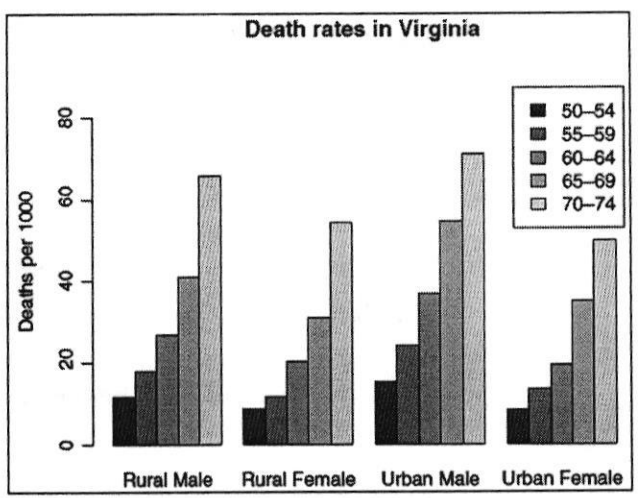
SECTION – III

6. (a) Given the frequency distribution $\{(x_i, f_i), i = 1,2,3...,n\}$, draw the 'less than' cumulative frequency curve.

(b) Distinguish between a Character object and a Factor object in R along with an example. (7,3)

7. (a) Define these data structures (i) List; (ii) Matrix; (iii) Dataframe.

(b) Consider the following bar graph produced in R for a dataset of death rates across various age-groups by various subpopulations.



(b) Write codes for the following output :

[1] "a" "a" "a" "a" "a" "b" "b" "c" "c" "c" "d" "d" "d" "d" "d"

[1] TRUE FALSE TRUE FALSE TRUE FALSE TRUE FALSE TRUE FALSE

(c) Given the following set of R instructions, fill in the blanks with the appropriate R-command.

Time <- c(0, 1, 2, 4, 6, 8, 9, 10, 11, 12, 13,14, 15)

> _____

> _____

> _____

[1] 13

[1] "numeric"

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	4.000	9.000	8.077	12.000	15.000

(d) Write three options of the plotting feature/ argument 'type'.

(e) If x is a vector of length n, then write commands to compute : $\frac{1}{n}\sum x_i^3 - n\bar{x}$. (2×5)

SECTION – II

3. (a) Explain the following functions with examples :

(i) str( )

(ii) sample( )

(iii) rnorm()

(iv) dunif()

(v) tail()

(b) The R-command '*h=hist(x)*' produces a histogram of numerical object x with equal class intervals. Write down the commands to superimpose a frequency polygon on this histogram. (5,2.5)

4. (a) Given the frequency distribution $\{(x_i, f_i), i = 1,2,3...,n\}$, compute mean and variance.

(b) Explain any FOUR features of graphical function par( ) with the help of an example. (4,3.5)

5. (a) A company carried a survey in which motor vibration was measured for five samples of motors,

each sample using a different brand of bearing. Company's interest lies on whether there are differences in the mean vibrations between brands. The data are stored in a data frame called motors. Construct an appropriate chart to make the comparison.

(b) Create data frame with name "distance" using the following :

$x <- c(21, 21, 22.8, 26.5, 18.7, 20, 19.3, 20.6, 19, 20.8)$

$y <- c(3, 3, 3, 6, 5, 7, 7, 4, 4, 4)$

$z <- c("a", "b", "c", "d", "e", "f", "g", "h", "i", "j")$

$w <- c(17, 18, 20, 16, 16, 14, 10, 11.5, 10, 10.8)$

(i) Determine the number of observations in the data frame as well as the number of variables. Also, what are the names of the variables?

(ii) Find the mean of x and w observations for which $y > 4$.

(iii) Construct scatter plot between x and w. (2,5.5)