This question paper contains 4+1 printed pages]

Roll No. | | | | | | | | | | | |

S. No. of Question Paper : 7210

Unique Paper Code : 32373902 HC

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

*(Write your Roll No. on the top immediately on receipt of this question paper.)*

Attempt *six* questions in all. Section A is compulsory.

Attempt *four* more questions, selecting

*two* questions from each of Section B and C.

Write R codes for each question given in Section B and C.

## Section A

1. Answer the following :

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

P.T.O.

(*ii*) What is the output of the following R-command :

> punif(1) ?

(*iii*) What does the *y*-axis represent in a histogram with unequal class intervals produced in R ?

(*iv*) When external Excel or Text data files, that contains some missing observations, are read into R, then how does R pads out this missing information ?

(*v*) Given the following set of R instructions, fill in the blank with the appropriate R-command.

>grass.df = data.frame(rich, graze)

> _____

'data.frame': 9 obs. of 2 variables:

$ rich :int 12 15 17 11 15 8 9 7 9

$ graze: Factor *w/2* levels "mow","unmow": 1 1 1 1 1 2 2 2 2. 1×5

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

> X = *c*(4.5, "*a*", 6, "*b*"); X

> Y = seq(1, 10, by = 0.5); Y



(*b*) Write codes for the following output:

[1] 3 3 3 4 4 4 5 5 5

[1] "mark1", "mark1", "mark1", "mark1", "mark2", "mark2", "mark2".

(*c*) Explain the output of the following R-code segment:

rm(list = ls( ))

F<-numeric(10)

F[1] < –F[2] < –1

for (*i* in 3 : 10) F[*i*] < –F[*i* – 1] + F[*i* – 2]

print(*i*); print(F)

(*d*) What is the significance of having a plotting feature/ argument '*type*' ?

(*e*) If *x* is a vector of length *n*, write R-commands to compute :

$$\frac{1}{n}\sum(x_i - \bar{x})^3.$$

## Section B

3. (*a*) Explain the functions scan( ) and structure( ) with the help of example.

(*b*) The R-command '$h$ = *hist*($x$)' produces a histogram of numerical object $x$ with equal class intervals. Then write down the commands to superimpose a frequency polygon on this histogram. 4+3.5

4. Given the frequency distribution $X_i|f_i$ :

(*i*) Compute mean and variance.

(*ii*) Draw the 'less than' cumulative frequency curve. 4+3.5

5. (*a*) Explain any *four* features of graphical function *par*( ).

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

*mpg* < –*c*(21, 21, 22.8, 26.5, 18.7, 20, 19.3, 20.6, 19, 20.8)

*cyl* < –*c*(3, 3, 3, 4, 4, 3, 3, 4, 4, 4)

Create a boxplot graph for the relation between *mpg* (miles per gallon) and *cyl* (number of cylinders). 4+3.5

## Section C

6. For any given vectors $x$, $y$ :

(*i*) Fit the regression line $y$ on $x$.

(*ii*) Estimate $y$ for any given value of $x$.

(*iii*) Create data frame to store $x$, $y$, fitted, residual.

(*iv*) Plot regression line of $y$ on $x$. 10



7. Below are given the temperatures of *two* states :

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

sam_tem2 = *c*(34, 36, 38, 40, 37, 38, 40, 34, 45)

Consider the output below :

$t$ = 13.819, $df$ = 18.12, $p$-value = 4.585e–11

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

95 percent confidence interval :

23.95498 32.53976

sample estimates :

mean of $x$ mean of $y$

29.721053 1.473684

From the output given above, test if the two temperatures differ significantly. Also, interpret your results. 10

8. Fit a Poisson distribution to the given data $X_i|f_i$ ($i$ = 1, 2,.....6) and test its goodness of fit. 10