

2020

## STATISTICS — HONOURS

Paper : SEC-A-1

(Statistical Data Analysis using R)

Full Marks : 80

*The figures in the margin indicate full marks.**Candidates are required to give their answers in their own words as far as practicable.*1. Answer *any fifteen* questions :

1×15

- (a) Define open source software.
- (b) How will you search R for all commands containing the term 'test'?
- (c) What will be the value of  $\text{atan}(1)$ ?
- (d) Using an example, distinguish between  $\sin(x)$  and  $\sin\pi(x)$ .
- (e) Write down the output for "Hello" > "Good bye". Justify your answer.
- (f) Why do you use  $\text{getwd}()$ ?
- (g) Give the syntax to create a bit of 2 variables in R.
- (h) Suppose,  $\text{day} \leftarrow c(\text{'Mon'}, \text{'Tue'}, \text{'Wed'}, \text{'Thu'})$ . Find the output for  $\text{day}[-3]$ . Explain your answer.
- (i) Explain how to use 'rep' to get the output  $c(1,2,3,1,2,3)$ .
- (j) Suppose,  $x \leftarrow c(2,3,5)$  and  $y \leftarrow c(0,3,6)$ . Write down the output of  $x*y$  and explain.
- (k) What is the use of 'header' in read table?
- (l) Write a code to generate the matrix  $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}$
- (m) Give an example showing the application of 'rbind'.
- (n) Write a code to get the value of  $\pi$  rounded upto 5 decimal places.
- (o) Write a code to find out the eigen vectors of a matrix.
- (p) What code will you use to check the list of commands in the MASS library?
- (q) Create a factor variable with levels 2, 4 and 6.
- (r) What is the function of  $\text{type} = 'l'$  in R graphics?

Please Turn Over

2. Answer *any six* questions :

- (a) How will you fit a simple linear regression in R on a categorical covariate? How will you extract the regression coefficients from the summary?
- (b) Suppose the square root value of 32 is 5.6568. Write down the outputs of (i) `sqrt(-32)`, (ii) `sqrt(-32 + 0i)`, (iii) `sqrt(as.complex(-32))`.
- (c) Suppose `x <- c(1,2,3,4)`. What will be the output of  
 (i)  $\frac{1}{x}$ , (ii) `c(x, 0, x)`, (iii) `x % / % 2`?
- (d) Give the R code using 'Seq' to generate the vectors  
 (i) `(-2, -1, 0, 1, 2)` (ii) `(3, 3.5, 4, 4.5, 5)` (iii) `(3,....., 5)` of length 100 with equispaced inputs between 3 and 5.
- (e) Give an example of the application of 'abline' in R graphics. How will you superimpose the regression line on a scatterplot using 'abline'?
- (f) Consider `u <- c(3, 5, 7, 5, 3, 2, 8, 5, 6, 9, 8, 8)`. Write codes to find  
 (i) length of the vector, (ii) unique elements (iii) frequency distribution.
- (g) The following table gives the actual expenditure in the public sectors during the 4th five year plan. Write a code to draw a pie chart for the data.

Sectors	Expenditure (₹ in crores)
Agriculture	2217
Power	2085
Large Industry	3090
Small Industry	295
Communication	3173
Miscellaneous	2575

- (h) How will you generate a sample of size 100 from the standard normal distribution and then find the quantiles of the generated data?

3. Answer *any two* questions :

- (a) Give an example of how to use the 'byrow' option of two vectors. Write the codes for  
 (i) matrix multiplication  
 (ii) elementwise multiplication  
 (iii) cross product of two vectors  
 (iv) solution to a linear system  $A\mathbf{x} = \mathbf{b}$ .

2+2+2+2+2

- (b) An inbuilt dataset 'women' gives heights and weights for American women aged 30–39 in the year 1975.
- (i) Write code(s) to get the names of the columns and the first few rows of the dataset along with the header.
  - (ii) Write code to change the inbuilt variable names to "V1", "V2",...
  - (iii) Compute the median of height, variance of weight and correlation coefficient between the two variables.
  - (iv) Write R codes to setup a linear regression model taking 'height' as the independent variable and estimate the regression coefficients. 2+3+3+2
- (c) (i) The following table gives the number of students in the colleges of Kolkata in the year 2001 according to their birth places. Write a code to graphically represent this data.

Place of birth	No. of students
Kolkata	22400
Outside Kolkata but within W.B.	14600
Outside W.B.	6000

Add a legend to the plot

- (ii) A typical standard normal table has a column with  $x$  values and another with  $\Phi(x)$  values. Write an R code to generate the table. What code will you write to get the lower tail C.D.F. value for the probability 0.5 and d.f. = 3 for a t-distribution? 5+5
-