# 2021

## 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 : 2×15

   (a) Write the name of the website where we can download R software for windows. Also write the installation process of the software after download.

   (b) Write a signle line command to compute $\tan(A)$, where $A = \exp(x^2 + 5x)$ with $x = 1,3,5$.

   (c) Suppose we have, days $<- c$("Sun", "Mon", "Tue", "Wed", "Fri", "Sat"). Insert "Thu" in between "Wed" and "Fri".

   (d) Use '$seq()$' to generate the vector $c(1,0.1,0.01,0.001)$ in a single line command.

   (e) Suppose $x <- c(1,3,4,2,...,5,6,3,5)$ is a vector of length 100. Write a command to display the position of $x$ such that the values are greater than 4.

   (f) Let $x$ be a vector of length 15. Write a single line command to create a sub-vector of $x$ by deleting its 2nd, 5th, 10th and 11th components.

   (g) Suppose $x <- c(2,5,6,7)$. Using the concept of vector multiplication, write a single line command to create the following matrix :

   $$\begin{bmatrix} 2 & 2 & 2 & 2 \\ 5 & 5 & 5 & 5 \\ 6 & 6 & 6 & 6 \\ 7 & 7 & 7 & 7 \end{bmatrix}.$$

   (h) Write a single line command to check : the positive square root of the median of a random sample of size 15 drawn from $\mathcal{N}(50,5^2)$ is greater than 10 or not.

   (i) What are the outputs of (1) $0\wedge 0$ and (2) $\ln f \wedge 0$ in $R$. What is your comment on the $R$ outputs?

   (j) Suppse $(x, y)$ contains a bivariate sample of size $n$. Provide a scatter plot of the given data with the names of the axis are 'Time' (x-axis) and 'Population'.

   (k) Suppose $x$ be a vector that contains number of road accidents per day in last month. Write a single line command to form a frequency table of the dataset.

**Please Turn Over**

(l) Let $x_1, x_2,...x_n$ be the given sample observations, where $n$ is any natural number ($\geq 2$). We want to calculate the variance, $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, in R software. Write a single line command to perform the task.

(m) Let $x_1, x_2,...,x_n$ be the given sample observations. Write a single line command to calculate

$$T = \sum_{i=1}^{n} \psi(x_i) R_i$$

where $R_i = \text{Rank}(x_i)$ within the set $\{x_1, x_2,...,x_n\}$ and $\psi(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ if } x \leq 0 \end{cases}$

(n) Write the command to load the package 'Matrix'.

(o) Let $M$ be 4×4 matrix. Discuss how you use R command to determine whether the matrix is positive definite or not.

(p) Write a single line command to draw a histogram based on a random sample of size 10,000 from *exponential*(*mean* = 10) distribution with 100 class intervals.

(q) Suppose we have a *csv* file 'ABCD.csv' in the location "D:/A/AB/ABC/ABCD.csv" and the first row of the file represents the names of the variables. Write R code to read the file in R.

(r) Distinguish between 'matrix' and 'data frame' in R.

2. Answer *any six* questions : 5×6

(a) A manufacturing company produces ball point pens. In the factory, there are five machines that produces five different types of pens. The following table gives the number of pens produced by the machines on per day basis.

| Machine | M1 | M2 | M3 | M4 | M5 | Total |
|---------|------|------|-----|------|------|-------|
| No. of pens | 1150 | 1600 | 850 | 1950 | 1050 | 6600 |

Write R codes to represent the above data graphically in two different ways with colour in gray scale. Give appropriate title to each graph. Also add legends to your graphs.

(b) Consider a funcion $f(x) = 5x^2 + 2x - 6$. Write R codes to draw the funcion $f(x)$ for different values of $x \in [-2, 2]$. Use *abline*() to mark the real roots of the equation $f(x) = 0$.

(c) Write R codes to draw random sample ($x$) of size 50 from $\mathcal{N}(10,2^2)$ distribution. Let $y$ be a variable such that $y = \exp(-x/2)$. Write command to fit a linear regression of $y$ on $x$. Also provide plot of residual vs fitted.

(d) Let $M$ be a 4×4 matrix. Write R code to create the matrix $M$ with suitable example. Also write R code to find $M_1$ and $M_2$ such that $M = M_1 + M_2$, where $M_1$ and $M_2$ are upper and lower triangular matrices respectively.

(e) Let $d$ be a vector which contains the diagonal elements (all are non-zero) of a diagonal matrix $D$ of order $n$. Write $R$ codes to create the diagonal matrix $D$ in two different ways. Also provide $R$ codes to find inverse of $D$ in two different ways. Create a sub-matrix of $D^{-1}$ after elementing the rows and columns except first and second row-column.

(f) Create a data frame in $R$ using the command *read.table*() based on the following information :

| Serial | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| *Score given by Judge-1* | 25 | 34 | 62 | 48 | 74 | 31 | 82 | 67 | 15 | 45 |
| *Score given by Judge-2* | 34 | 31 | 51 | 49 | 80 | 27 | 75 | 71 | 20 | 41 |

Write $R$ command make the data frame as a matrix. Also provide $R$ code to calculate Spearman's rank correlation.

(g) Write $R$ code to generate two random samples of sizes 120 and 150 from *exponential*(*mean* = 12) and *Log–normal*($\mu = 2$, $\sigma^2 = 1$) distributions. Write $R$ code to find measures of central tendency, dispersion, skewness and kurtosis. Also prodive command to represent the samples by using boxplot in a same graph. Give appropriate names to the boxplots.

(h) Let $y_1$ and $y_2$ be two variables depending on a single variable $x$. Suppose we are given $n$ such values of the variables. Write $R$ code to plot $(x, y_1)$ and $(x, y_2)$ on the same graph. Discuss how you set the range of x-axis and y-axis. Also explain the use of the command *legend*() in this regard.

3. Answer *any two* questions :

(a) Use $R$ code to find the values of the following probabilities :

(i) $P(1 \le X_1 < 2 \mid X_2 > 0)$ where $X_1 \sim N(0, 2)$

(ii) $P(1 < X_2 < 2 \mid -3 < X_2 < 3)$ where $X_2 \sim N(0, 1)$

Write $R$ code to draw two random samples each of size 10,000 from the distributions of $X_1$ and $X_2$. On the basis of the drawn samples, give estimates of the two probabilities. Find the absolute errors of the estimated probabilities.    2+2+2+2+2

(b) Write $R$ code to create the following partitioned matrix of order 5×5

$$A = \begin{bmatrix} 1 & | & 2 & 6 & 5 & 9 \\ - & | & - & - & - & - \\ 2 & | & 6 & 2 & 7 & 2 \\ 6 & | & 2 & 5 & 4 & 1 \\ 5 & | & 7 & 4 & 8 & 3 \\ 9 & | & 2 & 1 & 3 & 7 \end{bmatrix}$$

Using the formula of partition matrix, write $R$ code to calculate the determinant and inverse of the matrix $A$. For a vector $b$ of length 5, write $R$ code to find the solution of the system of linear equation $Ax = b$ in two different ways.    2+(2+2)+4

**Please Turn Over**

(c) Let $X$ be discrete random variable which takes values from $\{1,3,7,11,17,19\}$ with the corresponding probabilities $\{0.2, 0.15, 0.3, 0.05, 0.1, 0.2\}$. Without using any inbuilt function, write single line command to calculate (i) $E(X)$, (ii) $Var(X)$ and (iii) $MD(\mu)$ where $\mu = E(X)$. Also provide $R$ code to compute comulative distribution function $F(x)$ and store in a vector. Draw the graph of $(x, F(x))$ with an appropriate title. 5+2+3