

Total number of printed pages–12

3 (Sem–4/CBCS) STA SEC

2021

STATISTICS

(Skill Enhancement Course)

Full Marks : 50

Time : Two hours

The figures in the margin indicate full marks for the questions.

Paper : STA–SE–4014

(Statistical Data Analysis Using R)

1. Answer the following questions : $1 \times 6 = 6$
 - (a) For a given vector $x = c(2,7,4,5,9,10)$, the values obtained by using `cummax(x)` are _____ . *(Fill in the blank)*
 - (b) A command used to extract 3rd and 5th element from a vector x of 6 elements is _____ . *(Fill in the blank)*

Contd.

- (c) In R, missing values are represented by _____ which should be in capital letters. *(Fill in the blank)*
- (d) What is represented by R command `abline()` ?
- (e) Write down the argument used in graphical representation of R for the style of the lines plotted.
- (f) What is represented by the R command `length(x)` ?

2. Answer the following questions : $2 \times 5 = 10$

- (a) Write the output of the following R codes :
`x <- seq(-2, 10, 2)`
`x`
- (b) Write R codes to obtain $P(2 \leq X \leq 15)$ where $X \sim \text{Normal}(0,1)$.
- (c) What is the significance of having a plotting character (`pch`) in R ?
- (d) If nine people are to be assigned into three committees of sizes two, three and four respectively, write R codes to calculate how many possible assignments exist.

(e) Write down comparison operator in R for the following :

(i) Is not equal to

(ii) Logical no.

3. Answer the following questions : **(any two)**
5×2=10

(a) Write R codes to draw a histogram for a grouped frequency distribution with unequal class intervals.

(b) Write R codes to construct box plot and *qq* plot to check the normality of the parent population.

(c) Telephone calls to a local 108 number are known to follow a Poisson distribution with an average of two calls per minute. Write R codes to compute the probability that —

(i) there will be zero calls during a one minute period.

(ii) there will be less than five calls in a one minute period.

(iii) there will be less than six calls in one hour.

(d) Write R codes to —

(i) draw frequency polygon for a data given as a vector r

(ii) draw frequency pie chart for a data given as a vector r .

4. Answer the following questions : **(any three)**
8×3=24

(a) Write down R codes to compute Q_1, Q_2 and Q_3 for the values

$$x_{(1)} = 1, x_{(2)} = 4, x_{(3)} = 7, x_{(4)} = 9,$$

$$x_{(5)} = 10, x_{(6)} = 14, x_{(7)} = 15, x_{(8)} = 16,$$

$$x_{(9)} = 20 \text{ and } x_{(10)} = 21.$$

(b) A random sample of size 12 is taken from a population that follows a $N(\mu, \sigma)$ distribution, where, the value of σ is unknown.

$$\text{Given } \sum_{i=1}^{12} x_i = 61.9 \text{ and } \sum_{i=1}^{12} x_i^2 = 450$$

Write R codes to determine a 90% confidence interval for the population mean.

(c) A random sample of size $n = 30$ is taken from a distribution known to be $N(\mu, 2)$.

If the $\sum_{i=1}^{30} x_i = 56$, write R codes to

(i) test the null hypothesis $H_0 : \mu = 1.8$ versus the alternative hypothesis $H_1 : \mu > 1.8$ at $\alpha = 0.05$ significance level.

(ii) Find $\beta(\mu_1 = 3)$ and power ($\mu_1 = 3$).

(d) Write R codes to fit a binomial distribution for given $x_i|f_i, (i = 1, 2, \dots, 5)$ and also test the goodness of fit.

(e) For the given vectors x and y , write R codes to fit a line of regression and plot a graph.

Paper : STA-SE-4024

(Database Management System)

Group-A

1. Answer the following : 1×6=6
- (a) Which of the following is an advantage of DBMS ?
- (i) Data dependency
 - (ii) Reduction of data redundancy
 - (iii) Data inconsistency
 - (iv) None of the above.
- (b) In relational database, every table is a _____.
- (i) set
 - (ii) attribute
 - (iii) relation
 - (iv) ratio
- (c) Which of the following is not a data model in DBMS ?
- (i) Relational data model
 - (ii) Structured data model
 - (iii) Network data model
 - (iv) Hierarchical data model.

- (d) Which of the following is not a RDBMS ?
- (i) MySQL
 - (ii) SQL SERVER
 - (iii) ORACLE
 - (iv) C
- (e) Which of the following is not a component of DBMS ?
- (i) Data administrator
 - (ii) File manager
 - (iii) DML compiler
 - (iv) None of the above
- (f) DML stands for —
- (i) Data Manipulation Language
 - (ii) Data Mark-up Language
 - (iii) Data Management Language
 - (iv) All of the above.

2. Write the following : 2×5=10

- (a) Define RDBMS. What are tuples and relation ?

- (b) What do you mean by data redundancy in DBMS ?
 - (c) Define primary key with example.
 - (d) Define Relational Model with example.
 - (e) What are object oriented databases ?
3. Write the following : (*Answer **any two***)
2×5=10
- (a) What is Data Manager ? Describe in brief.
 - (b) Define the terms DDL and DML. What are the difference between them ?
 - (c) Write the advantages and disadvantages of RDBMS.
 - (d) Explain in detail about Database Management System advantages over file management system.

Group-B

4. Write the following : (*Answer **any three***)
3×3=24
- (a) Create student table with inclusion of Name, ID, Course, Year, Phone Number. Insert at least five data into the student table.

- (b) With a neat block diagram, explain the architecture of a typical DBMS.
- (c) Define Foreign Key. Explain all possible options that can be specified when a referential integrity constraint is violated.
- (d) What is meant by Integrity Constraints ? Explain the importance of referential integrity constraints. How referential integrity constraints is implemented in SQL ?
- (e) Write the SQL syntax with example for the following :
- (i) ALTER
 - (ii) INSERT
 - (iii) UPDATE
 - (iv) CREATE
-

Paper : STA-SE-4034

(Statistical Techniques for Research Methods)

1. Answer the following questions as directed :
1×6=6

(a) The research problem undertaken for study must be carefully _____.
(Fill in the blank)

(b) Data collected for survey from warranty cards are secondary data.
(True **or** False)

(c) Collection of data through questionnaires is free from the bias of the interviewer. (True **or** False)

(d) What is editing in processing and analysis of data ?

(e) What is coding ?

(f) Define unemployment rates.

2. Answer the following questions : 2×5=10

(a) What is research ?

(b) What do you mean by research problem ?

(c) Write a note on questionnaire.

(d) What is data processing ?

(e) What do you mean by removal of subsidy ?

3. Answer the following questions : **(any two)**
5×2=10

(a) Distinguish between questionnaire and schedule.

(b) Write a note on the process of selecting a research problem.

(c) Elucidate the research approaches.

(d) Write on different techniques of data interpretation.

4. Answer the following questions : **(any three)**
8×3=24

(a) Write a note on different types of research.

- (b) What precautions one should take while interpreting the data ?
 - (c) Write a note on methods of data collection.
 - (d) Write different steps of data processing. Also write a note on different types of data processing tools.
 - (e) Develop a questionnaire to collect survey data pertaining to a research problem (gender discriminations in private *vs* government sector).
-