3 (Sem–4/CBCS) STA SE 1/2/3

2023

STATISTICS

(Skill Enhancement Course)

Answer the Questions from any one Option.

OPTION – A

(Statistical Data Analysis using R)

Paper : STA–SE–4014

OPTION – B

(Database Management System)

Paper : STA–SE–4024

OPTION – C

(Statistical Techniques for Research Methods)

Paper : STA–SE–4034

Full Marks : 50

Time : Two hours

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

Answer **either** in English **or** in Assamese.

OPTION – A

(Statistical Data Analysis using R)

Paper : STA–SE–4014

1. Answer the following questions :

$1 \times 4 = 4$

(a) What is R ?

   (i) An object-oriented programming language

   (ii) An open source project from CRAN

   (iii) A programming language for statistical computing

   (iv) All of the above

   (Choose the correct option)

(b) What is primary file type of R ?

   (i) Vector

   (ii) Text file

   (iii) R scripts

   (iv) Statistical file

   (Choose the correct option)

(c) _____ function is used for reading the .csv file in R language.

   (i) write.csv( )

   (ii) read.csv( )

   (iii) let.csv( )

   (iv) table.csv( )

   (Choose the correct option)

(d) If a command is not complete at the end of a line, R will give a different prompt, by default it is

   (i) *

   (ii) –

   (iii) +

   (iv) 1

   (Choose the correct option)

2. Answer the following questions : $2 \times 3 = 6$

(a) Discuss the steps required for saving data in R.

(b) Write the summary and table function used in R.

(c) Write R program for addition of two 2×2 matrices.

Contd.

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

(a) What are the data frames ? Write its significance in R language.

(b) Write a note on import data into R program from various data sources.

(c) Write a note on different data types in R.

(d) Write R program to find the correlation matrix of a multivariate dataset.

4. Answer the following questions : $10 \times 3 = 30$

(a) The frequency distribution of number of heads obtained in an experiment of tossing 5 coins 110 times is given below :

Number of heads :	0	1	2	3	4	5
Frequency :	6	15	25	42	18	4

Fit a binomial distribution to the above data and find the expected frequency. Plot observed and expected frequencies and comment on the adequency of model.

Or

Write R code to

(i) draw histogram for a given data. 5

(ii) draw ogives with graphical summary of a given data. 5

(b) Write R code to calculate arithmetic mean, geometric mean, harmonic mean, mode, median, sixth decile for given $x_i / f_i \ (i = 1, 2, ..., 5)$.

Or

(i) Write an R program for a two-sample t-test. 5

(ii) Write R program for confidence interval of population mean of a normal distribution. 5

(c) Write R program to fit a normal distribution for $x_i / f_i (i = 1, 2, 3, 4, 5)$ and also test the goodness of fit.

Or

Fit a second degree parabola

$y = a + bx + cx^2$ to the following data :

Year *(X)* : 1999 2000 2001 2002 2003 2004 2005

Profit *(Y)* : 24 27 32 38 45 20 17
(in'000)

Estimate profit for the year 2006. Find estimated values of *Y* for each *X*. Plot estimated and observed values and represent parabola by graph.

----

1. Answer the following questions as directed : 1×4=4

   (i) What is information about data called ?

   (a) Hyperdata

   (b) Teradata

   (c) Metadata

   (d) Relations

   (Choose the correct option)

   (ii) _____ is a hardware component that is most important for the operation of database management system.

   (a) Microphone

   (b) High speed large capacity disk to store data

   (c) High resolution video display

   (d) Printer

   (Choose the correct option)

(iii) The ability to query data, as well as insert, delete, and alter tuples, is offered by

(a) TCL (Transaction Control Language)

(b) DCL (Data Control Language)

(d) DDL (Data Definition Language)

(e) DML (Data Manipulation Language)

(Choose the correct option)

(iv) Which command is used to remove a relation from an SQL?

(a) Drop table

(b) Delete

(c) Purge

(d) Remove

(Choose the correct option)

2. Answer the following questions : 2×3=6

(a) What do you mean by database management system ?

(b) Give syntax for create table.

(c) Define multivalued attributes.

3. Answer *any two* of the following : 5×2=10

(a) What are the major advantages and disadvantages of database management system ?

(b) Explain 2-tier architecture in DBMS with example.

(c) Explain different levels of data abstractions in DBMS.

(d) Define a foreign key and superkey with proper example.

4. Answer *any three* of the following :

$10 \times 3 = 30$

(a) Explain different types of keys in DBMS with example.

Or

Consider the following relational schema and solve the queries :

$2+2+2+2+2=10$

Student (Roll no, Name, Address, Course)

(i) Give syntax to create table.

(ii) Give syntax to insert data into a table.

(iii) List the name of students who live in "Guwahati".

(iv) List the name of those students who are from "Delhi" and enrolled in course "BA".

(v) Give syntax to change the name of Roll no. 10 to Amar.

(b) Define database. What do you mean by database management system ? What are the major advantages and disadvantages of DBMS ?

$2+2+(3+3)=10$

Or

Write down the steps used for conversion of ER diagram to relational database.

(c) What is normalisation ? Define 1NF, 2NF and 3NF with examples.

$1+(3+3+3)=10$

Or

What is SQL ? What are the major SQL commands ? Explain them.

———————

OPTION – C

(Statistical Techniques for Research Methods)

Paper : STA-SE-4034

1. Answer the following questions :   1×4=4

তলত দিয়া প্ৰশ্নবোৰৰ উত্তৰ দিয়া ঃ

(a) To collect data through questionnaire is relatively costly.   *(True or False)*

প্ৰশ্নপত্ৰৰ জৰিয়তে তথ্য সংগ্ৰহ কৰাটো তুলনামূলকভাৱে ব্যয়বহুল।   *(সঁচা বা মিছা)*

(b) What is data cleaning ?

তথ্য পৰিষ্কাৰ কৰা মানে কি?

(c) In which method of data collection respondents should be literate and cooperative ?

তথ্য সংগ্ৰহৰ কোনটো পদ্ধতিত উত্তৰদাতাসকল সাক্ষৰ আৰু সহযোগিতামূলক হ'ব লাগে?

(d) Sampling errors are absent in a census survey.   *(True or False)*

লোকপিয়লৰ সমীক্ষাত প্ৰতিচয়ন ক্ৰটি অনুপস্থিত।   *(সঁচা বা মিছা)*

2. Answer the following questions :   2×3=6

তলত দিয়া প্ৰশ্নবোৰৰ উত্তৰ দিয়া ঃ

(a) What is the difference between quantitative and qualitative research ?

পৰিমাণগত আৰু গুণগত গৱেষণাৰ মাজত পাৰ্থক্য কি?

(b) What do you mean by sample ?

নমুনা বুলিলে কি বুজায়?

(c) What are type I and type II error ?

প্ৰথম আৰু দ্বিতীয় বিধৰ ক্ৰটি কি কি?

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

তলত দিয়া প্ৰশ্নবোৰৰ উত্তৰ দিয়া ঃ *(যিকোনো দুটা)*

(a) Write a note on quantitative research approach.

পৰিমাণগত গৱেষণা পদ্ধতিৰ ওপৰত এটা টোকা লিখা।

(b) Describe some important points related to significance of research.

গৱেষণাৰ তাৎপৰ্যৰ সৈতে জড়িত কিছুমান গুৰুত্বপূৰ্ণ কথা বৰ্ণনা কৰা।

(c) Define sampling errors.

প্ৰতিচয়ন ক্ৰটিৰ সংজ্ঞা দিয়া।

(d) What are the differences between questionnaire and schedule ?

প্ৰশ্নপত্ৰ আৰু অনুসূচীৰ মাজত কি. কি পাৰ্থক্য আছে?

4. Answer the following questions : *(any three)*   10×3=30

তলত দিয়া প্ৰশ্নবোৰব উত্তৰ দিয়া ঃ *(যিকোনো তিনিটা)*

(a) Write a note on formulating of research problem.

গৱেষণা সমস্যাৰ প্ৰণয়নৰ ওপৰত এটা টোকা লিখা।

Or

Describe the methods of collecting primary data.

প্ৰাথমিক তথ্য সংগ্ৰহৰ পদ্ধতিসমূহ বৰ্ণনা কৰা।

(b) Examine the advantages and limitations of the observation method in collecting data.

তথ্য সংগ্ৰহৰ ক্ষেত্ৰত পৰ্যবেক্ষণ পদ্ধতিৰ সুবিধা আৰু সীমাবদ্ধতা পৰীক্ষা কৰা।

Or

"Processing of data implies editing, coding, classification and tabulation." Describe in brief these four operations pointing out the significance of each in the context of research study.

"তথ্যৰ প্ৰক্ৰিয়াকৰণৰ অৰ্থ হৈছে সম্পাদনা, কডিং, শ্ৰেণীবিভাজন আৰু টেবুলেচন।" গৱেষণা অধ্যয়নৰ প্ৰসংগত প্ৰতিটোৰ তাৎপৰ্য আঙুলিয়াই দি এই চাৰিটা অপাৰেচনৰ চমু বৰ্ণনা কৰা।

(c) Explain the important steps involved in the technique of data interpretation.

তথ্য ব্যাখ্যাৰ কৌশলৰ লগত জড়িত গুৰুত্বপূৰ্ণ পদক্ষেপসমূহ ব্যাখ্যা কৰা।

Or

Explain the guidelines for constructing a questionnaire.

প্ৰশ্নপত্ৰ তৈয়াৰ কৰাৰ বাবে থকা নিৰ্দেশনাসমূহ ব্যাখ্যা কৰা।

———————