

Assignment Booklet  
for  
**M.Sc. (Applied Statistics)**  
(MSCAST)



**SCHOOL OF SCIENCES**

Indira Gandhi National Open University  
New Delhi - 110068

**MSCAST**

# **Assignment Booklet**

## **M.SC. (APPLIED STATISTICS)**

**Valid from July 2023 to June 2024**

### **ASSIGNMENTS**

for

### **Semester-I Courses**

**\* MST-011**

**\* MST-012**

**\* MST-013**

**\* MST-014**

**\* MST-015**

**It is compulsory to submit Assignments before filling the  
Term-End Examination (TEE) Form**



**School of Sciences**

**Indira Gandhi National Open University**

**Maidan Garhi, New Delhi-110068 (INDIA)**

**(2023-2024)**

Dear Learner,

Welcome to the M.Sc. (Applied Statistics) Programme.

As per the laid down guidelines of the University, you need to complete the assignment for each theory course. All the questions given in an assignment are compulsory. It is important that you should write the answers to all questions in your own words. You should remember that writing answers to the assignment questions will improve your writing skills and prepare you for the term-end examination.

This booklet includes assignments for the first-semester courses:

**MST-011: Real Analysis, Calculus and Geometry**

**MST-012: Probability and Probability Distributions**

**MST-013: Survey Sampling and Design of Experiments-I**

**MST-014: Statistical Quality Control and Time Series Analysis**

**MST-015: Introduction to R Software**

**It is compulsory to submit the assignments within the stipulated time to be eligible for appearing in the term-end examination.** You will not be allowed to appear for the term-end examination for a course if you do not submit the assignment for that course within the due date. As per the University guidelines, if you appear in the term-end examination of a course without submitting its assignment, the result of the term-end examination is liable to be cancelled/ withheld.

**The assignments constitute the continuous component of the evaluation process and have 30% weightage in the final grading.**

Before you write the assignments, first go through the course material and then prepare the assignments carefully by following the instructions pertaining to assignments. Your responses should not be a verbatim reproduction of the textual materials provided for self-learning purposes, but it should be in your own words.

If you have any doubt or problem pertaining to the course material and assignments, contact the concerned Programme in-charge or Academic Counsellor at your Study Centre. If you still have problems, do feel free to contact us at the School of Sciences, IGNOU, New Delhi.

Wishing you all the best to successfully complete the programme.

**Programme Team  
MSCAST**

**Email: [mscast@ignou.ac.in](mailto:mscast@ignou.ac.in)**

## INSTRUCTIONS

1. Read the instructions related to assignments given in the Programme Guide.
2. Please note that unless you submit the assignments contained in this booklet within the stipulated time, you would not be permitted to appear for the term-end examination.
3. You are advised to mention the following information on the first page of the assignment response sheet:

NAME: .....
ENROLLMENT NO.....
CYCLE OF ADMISSION:.....
PROGRAMME CODE:.....
ASSIGNMENT CODE:.....
COURSE CODE:.....
COURSE TITLE:.....
REGIONAL CENTRE CODE:.....
STUDY CENTRE:.....
ADDRESS: .....
.....
.....
CONTACT NUMBER:.....
DATE OF SUBMISSION:.....

You are advised to strictly follow the above format. If you do not follow this format, your assignment response sheet will be returned to you, and you will be asked for re-submission.

**Note the following points before you start writing the assignments:**

- Use only A-4 size paper for writing your responses. Only handwritten assignments will be accepted. Typed or printed copies of the assignments will not be accepted.
- Tie the pages after numbering them carefully.
- Write the question number for each answer.
- All questions are compulsory.
- **Keep a copy of the assignment answer sheets with you before submission for future reference.**
- Answer each assignment on separate sheets.
- It is mandatory to write all assignments neatly in your own handwriting. Write Your Name, Course Code, Enrollment No. and Cycle of admission on all the assignments in bold letters.
- Express your response in your own words. You are advised to restrict your response based on the marks assigned to it. This will also help you to distribute your time in writing or completing your assignments on time.
- The assignment must be submitted at your Study Centre.

You will have to submit your solved assignments at the Study Centre allotted to you before the due date as set by the University.

It is desirable to keep with you a photocopy of the assignment(s) submitted by you.

**\*You will have to submit the assignments to the Study Centre by 31<sup>st</sup> March, 2024 or by the due date as specified by the University.**

**Due Date of Submission\*: March 31, 2024**

\*Please note that last date of submission may be changed by the University. Please check IGNOU website for updated information regarding due date of assignment submission.

## TUTOR MARKED ASSIGNMENT

### MST-011: Real Analysis, Calculus and Geometry

Course Code: MST-011

Assignment Code: MST-011/TMA/2023

Maximum Marks: 50

**Note: All questions are compulsory. Answer in your own words.**

1. Solve the following problems.

- (a) A ball is thrown in an upward direction. If the variable  $x$  represents the velocity of the ball when it strikes the ground. Classify variable  $x$  as discrete or continuous. Justify your answer with a proper explanation.
- (b) In R we have a built-in data set "trees". A screenshot of the first four rows together with the R code to obtain it is given as follows. To get more detail about this data set you can run? trees command on R console.

```
> head(trees, 4)
  Girth Height Volume
1  8.3     70  10.3
2  8.6     65  10.3
3  8.8     63  10.2
4 10.5     72  16.4
```

Note that all the three variables of this data set are numeric. So, assuming each row of this data set is a point in 3-dimension. Find the distances between the points corresponding to the first and the third rows using the Manhattan and Chebyshev distance formula.

- (c) Find the equation of a line passing through points A(2, 3, 5) and B(5, 8, 9). Also, find the coordinates of a point on this line which is at a distance of 10 units from point A opposite to the side of point B.
- (d) Give an example of a set which is convex but not affine. Justify your claim with a proper explanation.

**(2 + 3 + 2 + 3)**

2. (a) Test the convergence of the series

$$\frac{3}{5.9.7}x + \frac{5}{10.12.11}x^2 + \frac{7}{15.15.15}x^3 + \frac{9}{20.18.19}x^4 + \dots, x > 0.$$

- (b) If  $f : [0, 5] \rightarrow \mathbb{R}$  be a function defined by  $f(x) = x^2 + 2x + 1$ ,  $x \in [0, 5]$ . Show that  $f$  is Riemann integrable using both definitions. Also, verify that the results of both definition match.

**(10 + 10)**

3. (a) Evaluate the integral  $\iint_D e^{4x+5y} dx dy$ , where  $D = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$ , by considering  $D$  as a region of Type I and then as a region of Type II.

- (b) Evaluate the integral  $\int_0^4 \frac{dx}{\sqrt{4x - x^2}}$  using beta and gamma functions. **(14 + 6)**

## TUTOR MARKED ASSIGNMENT

### MST-012: Probability and Probability Distributions

Course Code: MST-012

Assignment Code: MST-012/TMA/2023

Maximum Marks: 100

**Note: All questions are compulsory. Answer in your own words.**

**1(a)** Suppose two friends Anjali and Prabhat trying to meet for a date to have lunch say between 2 pm to 3 pm. Suppose they follow the following rules for this meeting:

- Each of them will reach either on time or 10 minutes late or 20 minutes late or 30 minutes late or 40 minutes late or 50 minutes late or 1 hour late. All these arrival times are equally likely for both of them.
- Whoever of them reaches first will wait for the other to meet only for 10 minutes. If within 10 minutes the other does not reach, he/she leaves the place and they will not meet.

Find the probability of their meeting.

**(b)** In the study learning material (SLM), you have seen many situations where Poisson distribution is suitable and discussed some examples of such situations. Create your own example for a situation other than which are discussed in SLM. If you denote your created random variable by  $X$  then find the probability that  $X$  is less than 2.

**(8 + 17)**

**2(a)** In an election there are two candidates. Being a statistician, you are interested in predicting the result of the election. So, you plan to conduct a survey. Using the learning skill of this course answer the following question. How many people should be surveyed to be at least 90% sure that the estimate is within 0.03 of the true value?

**(b)** If  $X_1, X_2, X_3, X_4, \dots$  be a sequence of continuous random variables such that

$$f_{X_n}(x) = \frac{n}{2} e^{-n|x|} \text{ then prove that } X_n \xrightarrow{P} 0. \quad (15 + 10)$$

**3.** Explain the procedure of assigning probability in a continuous world of probability theory. **(25)**

**4(a)** If  $X \sim \text{Gamma}(\theta, \alpha)$  and  $Y \sim \text{Gamma}(\theta, \beta)$  be two independent gamma distributions and  $U = \frac{X}{X+Y}$  and  $V = X+Y$  then find the distribution of  $U$ .

**(b)** A hospital specialising in heart surgery. In 2022 total of 2000 patients were admitted for treatment. The average payment made by a patient was Rs 1, 50,000 with a standard deviation of Rs 25000. Under the assumption that payments follow a normal distribution, answer the following questions.

**(i)** The number of patients who paid between Rs 1,40,000 and Rs 1,70,000.

**(ii)** The probability that a patient bill exceeds Rs 1,00,000.

**(iii)** Maximum amount paid by the lowest paying one-third of patients.

**(10 + 15)**

# TUTOR MARKED ASSIGNMENT

## MST-013: Survey Sampling and Design of Experiments-I

Course Code: MST-013

Assignment Code: MST-013/TMA/2023-24

Maximum Marks: 100

**Note: All questions are compulsory. Answer in your own words.**

1. State whether the following statements are true or false and also give the reason in support of your answer: **(5×2=10)**

- (a)  $V_{\text{opt}}(\bar{x}_{\text{st}})$  lies between  $V_{\text{prop}}(\bar{x}_{\text{st}})$  and  $V_{\text{Random}}(\bar{x}_{\text{st}})$ .
- (b) The total number of all possible samples of size 3 without replacement from a population of size 7 is 21.
- (c) While analysing the data of a  $5 \times 5$  Latin Square design the d.f. for ESS is equal to 16.
- (d) In a Two-way Analysis of Variance test with 5 observations per cell having 4 blocks and 4 treatments the degree of freedom for the total variation is 64.
- (e) The probability of selection of a sample of  $n$  from the population by SRSWOR is  $1/N$ .

- 2(a) A sample of 100 employees is to be drawn from a population of collages A and B. The population means and population mean squares of their monthly wages are given below:

Village	$N_i$	$\bar{X}_i$	$S_i^2$
Collage A	400	60	20
Collage B	200	120	80

Draw the samples using Proportional and Neyman allocation techniques and compare. Obtain the sample mean and variances for the Proportional Allocation and SRSWOR for the given information. Then Find the percentage gain in precision of variances of sample mean under the Proportional Allocation over the that of SRSWOR.

**(15)**

- (b) A population consists of 10 villages with a total of 212 households. The second column of the accompanying table shows the number of households corresponding to each village. Select a PPS with replacement sample of 6 villages by using the Cumulative Total method:

Village	1	2	3	4	5	6	7	8	9	10
No. of Households	35	28	20	25	30	19	10	12	18	15

**(10)**

- 3(a) In a population of size  $N = 5$ , the values of the population characteristics are 1, 3, 5, 7, 9, a sample of size 2 is drawn. Verify that  $\bar{y}$  is an unbiased estimate of  $\bar{Y}$  and  $V(\bar{y})$  is equal to



$$V(\bar{y}) = \frac{N-n}{N.n} S^2. \quad (12)$$

- (b) In order to compare the mileage yields of 3 kinds of Gasoline, several tests were run, and the following results were obtained:

<b>Gasoline A:</b>	19	21	20	18	21	21
<b>Gasoline B:</b>	23	20	22	20	24	23
<b>Gasoline C:</b>	20	17	21	19	20	17

Carry out the Analysis of Variance test and test whether there is significant differences between the average mileage of 3 kinds of gasoline at 5% level of significance.

(8)

- 4(a) A manurial trial with six levels of Farmyard Manure (FYM) was carried out in a randomised block design with 4 replications at the experimental station Junagarh with a new study the rate of decomposition of organic matters in soil and its synthetic capacity in soil on cotton crop. The yield per plot in kg for different levels of FYM and replications is given below:

**Cotton Yield Per Plot (in Kg)**

Levels of FYM	Replications			
	I	II	III	IV
1	6.90	4.60	4.40	4.81
2	6.48	5.57	4.28	4.45
3	6.52	7.60	5.30	5.30
4	6.90	6.65	6.75	7.75
5	6.00	6.18	6.50	5.50
6	7.90	7.57	6.80	6.62

Carry out the Analysis of Variance test and draw the conclusions. (15)

- (b) For the given data the yield of the treatment 2 in 3<sup>rd</sup> block is missing. Estimate the missing value and analyse the data.

Treatments	Blocks			
	I	II	III	IV
<b>1</b>	105	114	108	109
<b>2</b>	112	113	Y	112
<b>3</b>	106	114	105	109

(10)

- 5(a) The following data is the data pertaining to a feeding trial on sheep. Treatments

- A: Grazing only
- B: Grazing + Maize Supplements
- C: Grazing + Maize + Protein Supplement P<sub>1</sub>
- D: Grazing + Maize + Protein Supplement P<sub>2</sub>
- E: Grazing + Maize + Protein Supplement P<sub>3</sub>

Layout and Wool Yield (100 gm) is given as:

32(D)	33(E)	30(C)	28(B)	24(A)
51(C)	45(D)	41(A)	45(E)	29(B)
41 (E)	29 (A)	24 (B)	36 (D)	35 (C)
38 (B)	39(C)	42(E)	23(A)	37(D)
38(A)	24(B)	21(D)	29(C)	26(E)

Analyse the design with appropriate method and calculate the Critical Difference for the treatment mean yield. **(15)**

- (b)** In a class of Statistics, total number of students is 30. Select the linear and circular systematic random samples of 12 students. The age of 30 students is given below:

**Age:** 22    25    22    21    22    25    24    23    22    21    20  
           21    22    23    25    23    24    22    24    24    21    20  
           23    21    22    20    20    21    22    25

**(5)**

## TUTOR MARKED ASSIGNMENT

### MST-014: Statistical Quality Control and Time Series

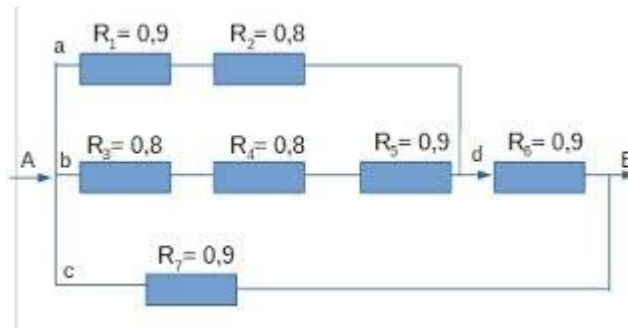
Course Code: MST-014

Assignment Code: MST-014/TMA/2023

Maximum Marks: 100

**Note: All questions are compulsory. Answer in your own words.**

1. (a) State whether the following statements are **True** or **False**. Give reason in support of your answer: **(5×2=10)**
- (i) The R- chart is suitable when subgroup size is greater than 10.
  - (ii) In single sampling plan, if we increase acceptance number then the OC curve will be steeper.
  - (iii) If the effect of summer and winter is not constant on the sale of AC then we use the additive model of the time series.
  - (iv) If a researcher wants to find the relationship between today's unemployment and that of 5 years ago without considering what happens in between then the partial autocorrelation is the better way in comparison to autocorrelation.
  - (v) A system has four components connected in parallel configuration with reliability 0.2, 0.4, 0.5, 0.8. To improve the reliability of the system most, we have to replace the component which reliability is 0.2.
- (b) Differentiate between the autoregressive and moving average models of time series. **(10)**
- 2(a) A manufacturer of men's jeans purchases zippers in lots of 500. The jeans manufacturer uses single-sample acceptance sampling with a sample size of 10 to determine whether to accept the lot. The manufacturer uses  $c = 2$  as the acceptance number. Suppose 3% nonconforming zippers are acceptable to the manufacturer and 8% nonconforming zippers are not acceptable. Find
- (i) Probability of accepting a lot of incoming quality 0.04.
  - (ii) Average outgoing quality (AOQ), if the rejected lots are screened and all defective zippers are replaced by non-defectives.
  - (iii) Average total inspection (ATI). **(6+2+2)**
- (b) An office supply company ordered a lot of 400 printers. When the lot arrives the company inspector will randomly inspect 12 printers. If more than three printers in the sample are non-conforming, the lot will be rejected. If fewer than two printers are non-conforming, the lot will be accepted. Otherwise, a second sample of size 8 will be taken. Suppose the inspector finds two non-conforming printers in the first sample and two in the second sample. Also AQL and LTPD are 0.05 and 0.10 respectively. Let incoming quality be 4%.
- (i) What is the probability of accepting the lot at the first sample?
  - (ii) What is the probability of accepting the lot at the second sample? **(2+8)**
- 3(a) A system has seven independent components and reliability block diagram of it shown as follows:



Find reliability of the system.

(10)

(b) The failure data for 40 electronic components is shown below:

<b>Operating Time (in hours)</b>	0-5	5-10	10-15	15-20	20-25	25-30
<b>Number of Failures</b>	5	7	6	4	5	4
<b>Operating Time (in hours)</b>	30-35	35-40	40-45	45-50	≥50	
<b>Number of Failures</b>	4	0	2	1	2	

Estimate the reliability, cumulative failure distribution, failure density and failure rate functions.

(10)

4. At a call centre, callers have to wait till an operator is ready to take their call. To monitor this process, 5 calls were recorded every hour for the 8-hour working day. The data below shows the waiting time in seconds:

Time	Sample Number				
	1	2	3	4	5
9 a.m	8	9	15	4	11
10	7	10	7	6	8
11	11	12	10	9	10
12	12	8	6	9	12
1 p.m.	11	10	6	14	11
2	7	7	10	4	11
3	10	7	4	10	10
4	8	11	11	11	7

- Use the data to construct control charts for mean and comments about the process. If process is out of control, then calculate the revised control limits.
- Construct the CUSUM chart when the process is under control and draw the conclusion about the process.
- If the specification limits as the  $8 \pm 2$ , then calculate the process capability index  $C_{pk}$  and interpret the result.
- Also find the percentage of calls lie outside the specification limits assuming that calls follow the normal distribution.

(20)

5(a) Consider the time series model

$$y_t = 10 + 0.5y_{t-1} - 0.8y_{t-2} + \varepsilon_t$$

where  $\varepsilon_t \sim N[0,1]$

- Is this a stationary time series?
- What are the mean and variance of the time series?
- Calculate the autocorrelation function.
- Plot the correlogram.

(10)

(b) The marketing manager of a company recorded the number of mobiles sold quarterly for which are given in the following table:

<b>Quarter Year</b>	<b>Q<sub>1</sub></b>	<b>Q<sub>2</sub></b>	<b>Q<sub>3</sub></b>	<b>Q<sub>4</sub></b>
<b>2018</b>	48	41	60	65
<b>2019</b>	58	52	68	74
<b>2020</b>	60	56	75	78

(i) Find the quarterly seasonal indexes for the mobile sold using the ratio to trend method.

(ii) Do seasonal forces significantly influence the sale of mobile? Comment.

(iii) Also find the deseasonalised values.

**(10)**

**TUTOR MARKED ASSIGNMENT**  
**MST-015: INTRODUCTION TO R SOFTWARE**

Course Code: MST-015  
Assignment Code: MST-015/TMA/2023  
Maximum Marks: 50

**Note: All questions are compulsory. Answer in your own words.**

1. Attempt the following:

(a) Write the output of the following statements:

(i) `rep(x=c(T,F,T,F), times=c(2,1,2,3))`

(ii) `5%/3; diag(3)`

(b) Differentiate between the use of the `sep` and `collapse` arguments of the `paste()` function.

(c) Write R commands to create a bar plot of the following data by using arguments of the used function for filling up the bars and to give labels to the axis:

5, 10, 8, 7, 8, 5, 8, 7, 5, 8, 9, 6, 8, 8, 8

(d) Check whether the given loop is finite or infinite. If infinite, do the necessary changes in the written loop to make it finite.

```
x<-0
repeat{
  print(x^2)
  x<-x+1
  if(x<5) print(x) }
```

**(1×3+2=5)**

2. The following data relates to the number of items produced per shift by two workers for a number of days.

<b>Worker A</b>	19	22	24	27	24	18
<b>Worker B</b>	26	37	40	35	NA	NA

(a) Write R command to create a list named `LT` with worker's data. Also, after creating the list, do the following tasks:

(i) Use a suitable loop function to compute the mean of number of items produced by each worker in a single line command.

(ii) Extract the worker A data from it by using two different approaches.

(b) Write R command to create a data frame named `DF` with worker's data and do the following tasks:

(i) Use suitable function to remove `NA` from the data and then create a scatter plot.

(ii) Write the known data obtained in step (i) to a .txt file named "WORK".

**(8+7=15)**

3. Write R commands to:

(a) Create a function to compute ranks (in case of tied ranks) of the given data.

(b) Create a date object named `Ddata` consisting of the following dates.

26Jan2023, 15Aug2023, 02Oct2023, 05Sep2023

(c) Create an array of two dimension with following elements.

$$\begin{pmatrix} -2 & 4 \\ 0 & 1 \\ 9 & 2 \end{pmatrix}$$

Also, extract the row shown in the rectangular box.

(d) Create the graph of the following function.

$$f(x) = |x|, -5 \leq x \leq 5$$

(4x3+3=15)

4. (a) Create following two matrices A and B with following elements.

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} -3 & 1 \\ 2 & -1 \end{pmatrix}$$

Write R commands to do the following tasks:

- (i) Multiply the two matrices.
- (ii) Combine the two matrices row-wise.
- (iii) Create a function that computes the following expression:

$$A^2 + 3*B$$

(b) Create a data frame named **RData** consisting of the following data:

```
x      y      z w
0.04 0.16 0.53 A
0.82 0.87 0.84 A
0.32 0.65 0.99 A
0.39 0.83 0.42 A
0.31 0.93 0.78 A
0.83 0.31 0.41 A
0.73 0.74 0.88 A
0.32 0.39 0.50 B
0.60 0.85 0.68 B
0.65 0.28 0.86 B
0.55 0.95 0.77 B
0.53 0.35 0.32 B
0.91 0.01 0.91 B
0.84 0.81 0.37 B
```

Write R commands to:

- (i) Compute the group wise means of **x**, **y** and **z** according to the groups defined by **w** column using apply family function.
- (ii) Sort **RData** according the **y** column of it.

(8+7=15)