

# **ASSIGNMENT BOOKLET**

## **Post Graduate Diploma in Applied Statistics (Specialisation in Industrial Statistics)**

**MST-001 to MST-004**

**(Valid from 1<sup>st</sup> January, 2016 to 31<sup>st</sup> December, 2016)**

**It is compulsory to submit the assignments  
before filling the Examination Form.**



**School of Sciences  
Indira Gandhi National Open University  
Maidan Garhi, New Delhi-110068  
(2016)**

Dear Student,

Please read the information on assignments in the Programme Guide that we have sent you after your enrolment. A weightage of 30%, as you are aware, has been earmarked for continuous evaluation, **which would consist of one tutor-marked assignment** for this course. The assignments for MST-001 to MST-004 have been given in this booklet.

### **Instructions for Formatting Your Assignments**

Before attempting the assignment, please read the following instructions carefully:

- 1) On top of the first page of your answer sheet, please write the details exactly in the following format:

---

ENROLLMENT NO : .....

NAME : .....

ADDRESS : .....

.....

.....

PROGRAMME CODE: .....

COURSE CODE: .....

COURSE TITLE: .....

STUDY CENTRE: ..... DATE: .....

---

**PLEASE FOLLOW THE ABOVE FORMAT STRICTLY TO FACILITATE EVALUATION AND TO AVOID DELAY.**

- 2) Use only foolscap size writing paper (but not of very thin variety) for writing your answers.
- 3) Leave 4 cm margin on the left, top and bottom of your answer sheet.
- 4) Your answers should be precise.
- 5) This assignment is to be submitted at the Study Centre.

**We strongly suggest that you should retain a copy of your answer sheets.**

- 6) This assignment is valid up to December 31, 2016.
- 7) **You cannot fill the Exam Form for these courses** till you have submitted the assignments for these courses. So solve it and **submit it to your study centre at the earliest**. If you wish to appear in the **TEE, June 2016**, you should submit your TMAs by **March 31, 2016**. Similarly, if you wish to appear in the **TEE, December 2016**, you should submit your TMAs by **September 30, 2016**.

We wish you good luck.

# TUTOR MARKED ASSIGNMENT

## MST-001: Foundation in Mathematics and Statistics

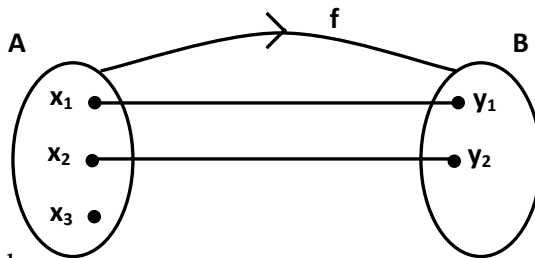
Course Code: MST-001

Assignment Code: MST-001/TMA/2016

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. (2×5=10)
  - a) Collection of rich persons in India forms a set.
  - b) Following rule is a function from A to B.



- c)  $\frac{d}{dx}(9-7x)^5 = 45(9-7x)^4$
  - d) In exclusive method, upper limit of a class is included in the same class.
  - e) The order of the matrix  $\begin{bmatrix} 2 & 5 & 6 \\ 4 & 3 & 1 \end{bmatrix}$  is  $3 \times 2$ .
2. If four cards are chosen from a pack of 52 playing cards then find the number of ways that all four cards are:
    - a) of same suit
    - b) red
    - c) face cards
    - d) king
    - e) of different suit (2 + 2 + 2 + 2 + 2)
  3. Arrange the numbers 49, 36, 42, 19, 22, 27, 14, 13, 24, 48, 23, 28, 17, 42, 39, 45, 22, 24, 17, 41, 18, 42, 38, 43, 11, 27, 36, 13, 40, 30, 24, 10, 18, 47, 18, 19, 23, 12, 27 in stretched stem-and-leaf display that has single-digit starting parts and leaves, but has stem width of 5. **(10)**
  4. If the universal set is  $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$  and  $A = \{2, 3, 6, 7\}$ ,  $B = \{4, 6, 8\}$ ,  $C = \{6, 7, 8\}$  are the subsets of  $U$ , then verify
    - a) De-Morgan's laws
    - b) left distributive law (5 + 5)
  5. Evaluate the following:
    - a)  $\int x^2 e^{2x} dx$
    - b)  $\frac{dy}{dx}$ , where  $y = (4x + 5)^4 (9x + 4)^5$  (5 + 5)

6. a) Prove that 
$$\begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix} = (a+b+c)(ab+bc+ca - a^2 - b^2 - c^2)$$

b) What do you mean by primary data and secondary data? Also give an example in each case. (5 + 5)

7. Draw a box plot with whisker, +ve sign and outliers for the following data:

42, 37, 28, 23, 32, 25, 26, 39, 38, 41, 22, 38, 21, 31, 26, 36, 42, 52, 50, 47, 24, 53, 28 (20)

8. a) Find the values of **a** and **b**, if the function **f** given below is continuous at  $x = 2$

$$f(x) = \begin{cases} a + b, & x < 2 \\ a + bx + 4, & x = 2 \\ 5, & x > 2 \end{cases}$$

b) Draw a histogram for the following data:

Wages	40 – 49	50 – 69	70 – 99	100 – 109	110 – 119
No of workers	2	20	60	35	4

Also draw frequency polygon in the same graph. (8 + 12)

# TUTOR MARKED ASSIGNMENT

## MST-002: Descriptive Statistics

Course Code: MST-002

Assignment Code: MST-002/TMA/2016

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. (2×5=10)
  - a) If standard deviation of  $x$  is 5, standard deviation of  $y = 2x-3$  is 7.
  - b) Mean deviation is least when calculated from the median.
  - c) The correlation coefficient between  $x$  and  $(a - x)$  is  $-1$ .
  - d) The regression coefficients  $b_{yx}$  and  $b_{xy}$  of a data are 1.2 and 0.8, respectively.
  - e) If  $(AB) = 10$ ,  $(\alpha B) = 15$ ,  $(A\beta) = 20$  and  $(\alpha\beta) = 30$  then  $A$  and  $B$  are associated.

2. a) Find the missing information from the following data:

	Group I	Group II	Group III	Combined
Number	50	?	90	200
Standard Deviation	6	7	?	7.746
Mean	113	?	115	116

- b) If AM and GM of two numbers are 30 and 18, respectively, find the numbers. (7+3)
3. a) The frequency distribution of the marks obtained by the 25 students each of the two sections is given as follows:

Marks:	10-20	20-30	30-40	40-50	50-60
Section A:	2	5	10	5	3
Section B:	3	7	8	5	2

Find which section is more consistent.

- b) Mean and Standard deviation of 18 observations are found to be 7 and 4, respectively. On comparing the original data, it was found that an observation 12 was miscopied as 21 in the calculations. Calculate correct mean and standard deviation. (7+3)
4. The equations of two regression lines are given as follows:

$$4x - 5y + 30 = 0$$

$$20x - 9y - 107 = 0$$

Calculate (i) regression coefficients,  $b_{yx}$  and  $b_{xy}$ ; (ii) correlation coefficient  $r(x, y)$ ; (iii) Mean of  $X$  and  $Y$ ; and (iv) the value of  $\sigma_y$  if  $\sigma_x = 3$ . (10)

5. A researcher collects the following information for two variables  $x$  and  $y$ :

$$n = 20, r = 0.5, \text{mean}(x) = 15, \text{mean}(y) = , \sigma_x = 4 \text{ and } \sigma_y = 5$$

Later it was found that one pair of values  $(x, y)$  has been wrongly taken as  $(16, 30)$  whereas the correct values were  $(26, 35)$ . Find the correct value of  $r(x, y)$ . (10)

6. a) If  $a, b, c, d$  are constants, then show that the coefficient of correlation between  $ax+b$  and  $cy+d$  is numerically equal to that between  $x$  and  $y$ .
- b) A statistician wanted to compare two methods A and B of teaching. He selected a random sample of 22 students. He grouped them into 11 pairs so that the students in pair have approximately equal scores on an intelligence test. In each pair one student was taught by method A and the other by method B and examined after the course. The marks obtained by both methods are given as:

Methods	1	2	3	4	5	6	7	8	9	10	11
Method A	24	29	19	14	30	19	27	30	20	28	11
Method B	37	35	16	26	23	27	19	20	16	11	21

Find the rank correlation coefficient. **(3+7)**

7. a) Fit an exponential curve of the form  $Y = ab^X$  to the following data:

X:	1	2	3	4	5	6	7	8
Y:	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

- b) Calculate the first, second and third quartile for the following data:

Class:	Below 30	30-40	40-50	50-60	60-70	70-80	80 and above
Frequency:	69	167	207	65	58	27	10

Also find the quartile deviation and coefficient of quartile deviation. **(10+10)**

8. a) Board of Directors of Labour Union wishes to sample the opinion of its members before submitting a change in its contribution at a forthcoming annual meeting. Questionnaires are sent to a random sample of 200 members in three union locals. The results of the survey are as follows:

Reaction	Union Locals			Total
	A	B	C	
Favour Change	35	45	20	100
Against Change	15	25	16	56
No Response	10	10	24	44
Total	60	80	60	200

Determine the amount of association between the Union locals and their reactions using coefficient of contingency and interpret the result.

- b) 600 candidates were appeared in an examination. The boys outnumbered girls by 15% of all candidates. Number of passed exceeded the number of failed candidates by 300. Boys failing in the examination numbered 80. Determine the coefficient of association. **(12+8)**

# TUTOR MARKED ASSIGNMENT

## MST-003: Probability Theory

Course Code: MST-003

Assignment Code: MST-003/TMA/2016

Maximum Marks: 100

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

1. Which of the following statements are true or false? Give reason in support of your answer. (2×5 = 10)

a) When two dice are thrown simultaneously then total number of sample points in the sample space will be 12.

b) Expected value of a continuous random variable X is defined as  $E(X) = \int_{-\infty}^x x f(x) dx$ .

c) If X and Y are independent random variable then  $V(X - Y) = V(X) - V(Y)$ .

d) If  $X \sim B(4, 3)$  then variance of X is 12.

e) If probability density function of a normally distributed random variable X is

$$f(x) = \frac{1}{6\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-46}{6}\right)^2}, \quad -\infty < x < \infty$$

then variance of X is 36.

2. An insurance company selected 6000 drivers from a city at random in order to find a relationship between age and accidents. The following table shows the results to these 6000 drivers.

Age of drivers (in years) Class Interval	Accidents in one year				
	0	1	2	3	4 or more
18 – 25	700	310	225	110	85
25 – 40	1100	290	200	105	80
40 – 50	1200	235	175	80	55
50 and above	600	205	140	70	35

If a driver from the city is selected at random, find the probability of the following events:

- Age lying between 18 – 25 and meet 3 accidents
- Age lying between 18 – 40 and meet 1 accident
- Age more than 25 years and meet at most one accident
- Having no accident in the year
- Age lying between 18 – 40 and meet at least 3 accidents

(2 + 2 + 2 + 2 + 2)

3. Determine the constant k such that the function  $f(x) = kx^2(1-x)^5$ ,  $0 < x < 1$  is a beta distribution of first kind. Also, find its mean and variance. (10)

4. An insurance company insured 2000 scooter drivers, 3000 car drivers and 5000 truck drivers. The probabilities that scooter, car and truck drivers meet an accident are 0.02, 0.04, and 0.25 respectively. One of the insured persons meets with an accident. What is the probability that he is a
- Scooter driver
  - Car driver
- (5 + 5)**

5. The following table represents the joint probability distribution of the discrete random variable (X, Y):

	Y	1	2	3
X				
1		0.2	0.2	0.1
2		0.1	0.3	0.1

Find

- The marginal distributions.
  - The conditional distribution of Y given  $X = 2$
- (5 + 5)**
6.
  - A rain coat dealer can earn Rs 800 per day during a rainy day. If it is a dry day, he can lose Rs 150 per day. What is his expectation, if the probability of rain is 0.6?
  - A player tosses two unbiased coins. He wins Rs. 10 if 2 heads appear, Rs. 5 if one head appears and Rs 1 if no head appears. Find the expected value of the amount won by him.

**(5 + 5)**

7. 
  - Let X and Y be two independent random variables such that  $X \sim B(5, 0.06)$  and  $Y \sim B(4, 0.6)$ . Find  $P[X + Y > 1]$
    - Comment on the statement: "The mean of a binomial distribution is 4 and variance 5".
  - If the probability that an individual suffers a bad reaction from an injection of a given serum is 0.002, determine the probability that out of 400 individuals
    - exactly 2
    - more than 3
    - at least one
 individuals suffer from bad reaction.
 

**(10 + 10)**

8. 
  - A die is rolled. If the outcome is a number greater than 2, what is the probability that it is an odd prime number?
  - A person is known to hit the target in 3 out of 4 shots whereas another person is known to hit 2 out of 5 shots. Find the probability that the target being hit when they both try.
  - Events A, B, C are mutually exclusive and exhaustive. If odds against A are 4 : 1 and against B are 3 : 2. Find the odds against event C.

**(7 + 7 + 6)**



# TUTOR MARKED ASSIGNMENT

## MST-004: Statistical Inference

Course Code: MST-004

Assignment Code: MST-004/TMA/2016

Maximum Marks: 100

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

1. State whether the following statements are True or False. Give reason in support of your answer. **(2×5=10)**

- a) If probability density function of a t-distribution is  $f(t) = \frac{1}{\pi(1+t^2)}$ ;  $-\infty < t < \infty$  then degrees of freedom of the distribution will be 1.
- b) If  $T_1$  and  $T_2$  are two estimators of an parameter  $\theta$  such that  $\text{Var}(T_1) = 1/2n$  and  $\text{Var}(T_2) = 2/n$  then  $T_1$  is more efficient than  $T_2$ .
- c) If the probability of non rejection of  $H_0$  when  $H_1$  is true is 0.4 then power of the test will be 0.6.
- d) The Wilcoxon signed-rank test is more powerful than the sign test.
- e) The t-test is used for testing the independence of two attributes.

2. a) A random sample of nine college students yielded the following data concerning the number of hours per day each student spent in using mobile phone:

5, 2, 7, 5.5, 3.5, 4, 5, 4.5, 4

Estimate the average number of hours per day spent in using mobile phone by the college students.

- b) If the sample values are 3, 5, 2, 7, and 0 then obtain the ML estimate for parameter  $\theta$  for the following distribution :

$$f(x, \theta) = \theta e^{-\theta x}; \quad 0 \leq x, \theta > 0 \quad \text{(5+5)}$$

3. A sample of 100 tyres is taken from a lot. The mean life of the tyres selected is the sample is found to be 40,000 kms with a standard deviation of 3200 kms. Is it reasonable to suppose the mean life of tyres in the lot as 41,000 kms at 5% level of significance? Also establish 95% confidence limits within which the mean life of tyres in the lot is expected to lie. **(10)**

4. The blood cholesterol levels of a population of workers have mean 202 mg/dl and standard deviation 14 mg/dl. If a sample of 36 workers is selected from the population and sample mean is calculated then find

- i) mean and standard error of the sampling distribution of the mean.
- ii) approximate the probability that the sample mean of their blood cholesterol levels will lie between 198 mg/dl and 206 mg/dl. **(5+5)**

5. The following data relate to the number of items produced per shift by two workers Rahul and Ramesh for a number of days:

<b>Rahul</b>	19	22	24	27	24	18	20	19	25	
<b>Ramesh</b>	26	37	40	35	30	40	26	30	35	45

Can it be inferred that Rahul is more stable worker compared to Ramesh by testing the variation in the item produced by them at 5% level of significance. **(10)**

- 6 a) In a city, 16 out of a random sample of 500 men were found to drinkers at a certain date. After the heavy increase in tax on intoxicants, another sample of 100 men in the same city included 3 drinkers. Do you feel that the observed proportion of drinkers decreasing significantly at 1% level?
- b) In a locality, 100 persons were randomly selected and asked about their educational achievements. The results were as follows:

Sex	Education			Total
	Middle	High School	College	
Male	12	13	25	50
Female	22	13	15	50
Total	34	26	40	100

Can we say that education depends on sex at 5% level of significance? **(5+5)**

7. Complete the following table:

S. No.	Test for	Name of the Test	Null and Alternative Hypotheses	Test Statistic	Assumptions for Applying the Test	Decision Rule (in short)
1	Population mean when population variation is known					
2	Population mean when population variation is unknown					
3	Population proportion					
4	Difference of two population means					
5	Two population standard deviation					
6	Difference of two population proportion					
7	Independence of two attributes					

8. A company is trying to improve the work efficiency of its employees. It has organized a special training programme for all employees. In order to assess the effectiveness of the training programme, the company has selected 10 employees randomly and administered a well-structured questionnaire. The scores (out of 100) obtained by the employees are given in the following table:

S. No	Before Training	After Training
1	60	68
2	62	70
3	67	80
4	64	74
5	66	66
6	63	72
7	69	84
8	63	60
9	60	65
10	62	90

To examine whether the training programme has improved efficiency of the employees, give the answer of the following:

- i) Are both samples are paired or independent?
- ii) Formulate the null and alternative hypotheses.
- iii) Which parametric test is used for testing the null hypothesis if it is known that the scores of the employees before and after the training programme follow the normal distribution? Conduct the test at 1% level of significance and conclude the result.
- iv) Which non-parametric test is used for testing the null hypothesis if it is known that the scores of the employees before and after the training programme do not follow the normal distribution but the distribution of the differences of scores before and after the training is symmetrical about its median? Conduct the test at 1% level of significance and conclude the result.

**(2+2+8+8)**