Booklet-I

ASSIGNMENT BOOKLET

Post Graduate Diploma in Applied Statistics

(Specialisation in Industrial Statistics)

MST-001 to MSTL-002

(Valid from 1st January, 2019 to 31st December, 2019)

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

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 MSTL-002 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, 2019.

7) **You cannot fill the Exam Form for this course** till you have submitted this assignment. So solve it and **submit it to your study centre at the earliest.** If you wish to appear in the **TEE, June 2019,** you should submit your TMAs by **March 31, 2019.** Similarly, If you wish to appear in the **TEE, December 2019,** you should submit your TMAs by **September 30, 2019.**

We wish you good luck.

TUTOR MARKED ASSIGNMENT

MST-001: Foundation in Mathematics and Statistics

Course Code: MST-001

Assignment Code: MST-001/TMA/2019

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: **(5×2=10)**

(a) The n^{th} term of the sequence $\frac{1}{4}, -\frac{1}{2}, 1, -2, \ldots$ is $\frac{1}{4}(-2)^n$.

(b) Symmetric difference of two sets A = {1, 2, 3} and B = {2, 3, 7, 9} is {1, 3, 7}.

(c) $\lim_{x\to 5}\frac{|x-5|}{x-5}=1$

(d) Ratio scale is highest level of measurement scale because here measurements of the characteristic under study can be positive or negative or both.

(e) The range of the data shown in the following frequency distribution is 350:

Classes	0-50	50-100	100-150	150-200	200-250	250-300	300-350
Frequencies	10	20	30	40	30	20	0

2 (a) In a city having total population of 1,00,000, out of which 10,000 can read and speak English, 88000 can read and speak Hindi and 80000 can read and speak Hindi only. Find how many of them can read and speak:

(i) Both the languages.

(ii) English only.

(iii) Neither Hindi nor English. **(2+1+2)**

(b) Show that the set $\{1, -2, 4, -8, 16, -32, \ldots\}$ is enumerable. **(3)**

(c) How many 5 digits numbers are possible using 8 digits 2 to 9 such that three digits 2, 5 and 9 are always included? **(2)**

3. (a) Expand $(8-3x)^{1/3}$ by binomial theorem.

(b) Expand $(2-ax)^5$ by binomial theorem for any index. **(5+5)**

4. Discuss the continuity and differentiability of the following function at x = 9.

$$f(x)=\begin{cases}|x-9|, & x\neq 9\\ 0, & x=9\end{cases}$$ **(5+5)**

5. Evaluate the following integrals:

i) $\int \frac{3x}{(x+2)(x-1)^2}\,dx$

ii) $\int_2^9 \frac{f(x)}{f(x)+f(11-x)}dx$ **(6+4)**

6. (a) Without expanding prove that

$$\begin{vmatrix} 4 & 5 & 1 \\ -4 & -8 & 8 \\ 7 & 3 & 19 \end{vmatrix} = 0$$

(b) Draw box plot for batting scores of the player given below: **(5+5)**

Player	10, 62, 22, 8, 90, 13, 105, 155, 25, 53, 6, 4, 52, 57, 27

7. (a) Find a matrix B such that $B\begin{vmatrix} 1 & 3 & 2 \\ -1 & -4 & 5 \\ 2 & 5 & 6 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 \\ -1 & -4 & 5 \\ 2 & 5 & 6 \end{vmatrix} B = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$

(10)

(b) A researcher visits 100 families and collects the information. He/she has ten questions in his/her questionnaire listed below from i to x. Assume that he/she gets information from all 100 families for all the 10 questions. He/she arranges this information in a data frame having 100 rows (each corresponding to response of a family) and 10 columns (each corresponding to a question in his questionnaire). To analyse this data first he/she has to find out scale (nominal, ordinal, interval or ratio) of each column. If he/she is your friend then write scale of each of 10 variables in his/her data frame.

i) Number of members in the family

ii) Age of the oldest person of the family

iii) Sex of the oldest person of the family

iv) Highest education qualification among the family members of the family

v) Monthly income of the family

vi) Saving (Income in that particular month − expenditure in the same month) of the family. Keep in mind that expenditure may be more that income in that particular month of some family(ies)

vii) Number of mobile phones in the family

viii) Height of tallest person of the family

ix) Does the family have landline telephone?

x) Monthly mobile bill of the family **(10×1=10)**

8. The following table shows the life (in weeks) of a sample of 15-watt LED bulbs produced by a manufacturer:

Life Time of LED				
50	33	47	73	15
11	98	24	72	37
19	82	07	53	51
36	61	02	25	34
21	42	45	08	15
32	23	26	06	17
15	44	53	55	41

71	77	74	28	13
18	13	55	81	06
05	35	85	17	53
52	03	60	64	42
14	58	04	21	32
91	21	21	09	17
35	11	72	12	18
32	20	14	76	11
13	15	34	23	32

i) Form a frequency distribution by taking suitable width,

ii) Form cumulative frequency curves (ogives) on one graph.

iii) Find the average (median) of the life of LED bulbs with the help of ogives.

(5+12+3)

TUTOR MARKED ASSIGNMENT

MST-002: Descriptive Statistics

Course Code: MST-002

Assignment Code: MST-002/TMA/2019

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) If the regression coefficients b_{YX} and b_{XY} of a data are 1.6 and 0.4, respectively, then the value of r(X, Y) is 0.8.

b) If each value of X is multiplied by 10 and each value of Y is multiplied by 20, then the modified regression coefficient b_{XY} would be the half of previous one.

c) If (AB) = 10, (αB) = 15, (Aβ) = 20 and (αβ) = 30 then A and B are associated.

d) If standard deviation of x is 5, standard deviation of y = 2x–3 is 7.

e) If with usual notations for two attributes the inequality $(AB)(\alpha\beta) < (\alpha B)(A\beta)$ holds, then $-1 \le Q \le 1$.

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) For the following distribution, calculate first four central moments using recurrence relations:

Marks:	2.7-7.5	7.5-12.5	12.5-17.5	17.5-22.5	22.5-27.5	27.5-32.5	32.5-37.5
Frequency:	06	13	23	39	19	15	05

Also find the coefficients of skewness and kurtosis. (8)

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

Class:	0-10	10-20	20-30	30-40	40-50
Frequency:	05	15	30	12	08

Also find the quartile deviation and coefficient of quartile deviation. (6)

5 a) A researcher collects the following information for two variables x and y:

n = 20, r = 0.5, mean (x) = 15, mean (y) = 20, $\sigma_x = 4$ 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)

b) Calculate the coefficient of rank correlation for the following data:

X:	48	33	40	09	16	16	65	24	16	57
Y:	13	13	24	06	15	04	20	09	06	19

6 a) Explain the method of least squares. Fit a straight line Y = a +b X to the following data:

X:	1	3	5	7	9	10
Y:	5	8	12	15	18	22

(5)

b) The equations of two regression lines are given as follows:

$$5x - 15y = 30$$

$$10x - 20y = 15$$

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

7. (a) In a trivariate distribution:

$\sigma_1 = 4,\ \sigma_2 = \sigma_3 = 6,\ r_{12} = 0.5,\ r_{23} = r_{31} = 0.8$

Find (i) $r_{23.1}$, (ii) $R_{1.23}$, (iii) $b_{12.3}, b_{13.2}$ and (iv) $\sigma_{1.23}$ (10)

ii. Suppose a computer has found for a given set of values of X_1, X_2 and X_3: r_{12}=0.90, r_{13}=0.30 and r_{23}=0.70. Examine whether these computations are error free. (4)

8 a) A company is interested in determining the strength of association between the communication time of their employees and the level of stress-related problems observed on job. A study of 120 assembly line workers reveals the following data:

	Stress			
	High	Moderate	Low	Total
Under 20 min.	10	10	15	35
20-50 min	15	10	25	50
Over 50 min	15	10	10	35
Total	40	30	50	120

Determine the amount of association between the communication time of their employees and the level of stress using coefficient of contingency and interpret the result. (12)

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. (8)

TUTOR MARKED ASSIGNMENT

MST-003: Probability Theory

Course Code: MST-003

Assignment Code: MST-003/TMA/2019

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) If A and B are any two events defined on a sample space S then $P(A \cup B) = P(S)$ always holds.

(b) Cumulative distribution function of a discrete random variable is always strictly increasing.

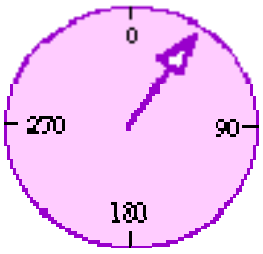
(c) If X is a discrete random variable with probability mass function (pmf)

X	0	1	2	3
P[X = x]	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	a

then value of a will be 1.

(d) If X and Y are two independent random variables then $P[X \geq 2|Y > 1] < P[X \geq 2]$.

(e) Suppose that you spin the dial shown in the figure so that it comes to rest at a random position.



The probability that the dial will land somewhere between 0 and 45 will be 1/4.

2. First check whether the following function is a valid density function? If it is a valid density then obtain its cumulative probability function F(x). If it is a valid density then finally calculate $P(7 \leq X \leq 8)$ either using f(x) or F(x).

$$f(x) = \begin{cases} \frac{2}{25}(x-5), & 5 \leq X \leq 10 \\ 0, & \text{otherwise} \end{cases} \quad \text{if } x < 0$$

(3+3+4)

3. **(a)** The joint density function of random variables X and Y is given by

$$f(x,y)=\begin{cases}14e^{-2x-7y}, & x\geq 0, y\geq 0\\ 0, & \text{otherwise}\end{cases}$$

Are X and Y independent? **(6)**

(b) A particular game is played where the contestant spins a wheel that can land on the number 1, 5, 30 with probabilities of 0.50, 0.45 and 0.05, respectively. The contestant pays INR5 to play the game and is awarded the amount of money indicated by the number where the spinner lands. Is this a fair game? [By fair, it is meant that the contestant should have an expected return equal to the price she pays to play the game.]

(4)

4. (a) Suppose two fair dice are tossed where each of the 36 possible outcomes is equally likely to occur. Knowing that the first die shows a 4, what is the probability that the sum of the two dice equals at least 7. **(5)**

(b) Suppose that there are m students in a room. What is the probability that at least two of them have the same birthday? Assume that every day of the year is equally likely to be a birthday, and disregard leap years. That is, assume there are always 365 days to a year. [Hint: Attack the problem by first calculating probability of complement event and then use $P(E) = 1 - P(\overline{E})$] **(5)**

5. The A taxi cab company has 12 Ambassadors and 8 Fiats. If 5 of these taxi cabs are in the workshop for repair and an Ambassador is as likely to be in for repair as a Fiat, what is the probability that (i) 3 of them are Ambassadors and 2 are Fiats, (ii) at least 3 of them are Ambassadors, and (iii) all 5 are of the same make? **(2+4+4)**

6. (a) The probability that a player hits a target is 0.24. He fires 6 times. What is the probability of hitting the target exactly twice? **(5)**

(b) What is the probability that 5^{th} success is obtained in 9^{th} trail if probability of success and failure do not vary from trial to trail. **(5)**

7. (a) Metro trains in a certain city run every 9 minutes between 6.15 a.m. to 11.15 p.m. What is the probability that a commuter entering the station at a random time during this period will have to wait at least five minutes? **(5)**

(b) Obtain mean and variance for the beta distribution whose density is given by

$$f(x)=\frac{280x^3}{(1+x)^9},\quad 0<x<\infty$$ **(5)**

8. (a) A car manufacturer purchases car batteries from two different suppliers A and B. Suppose supplier A provides 60% of the batteries and supplier B provides the rest. If 6% of all batteries from supplier A are defective and 4% of the batteries from supplier B are defective. Determine the probability that a randomly selected battery is not defective.

(b) An item is produced by a machine in large numbers. The machine is known to produce 5% defectives. A quality control engineer is testing the items randomly. What is the probability that at least 5 items are examined in order to get 2 defectives? **(10+10)**

TUTOR MARKED ASSIGNMENT

MST-004: Statistical Inference

Course Code: MST-004

Assignment Code: MST-004/TMA/2019

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: **(5×2=10)**

   (a) If the probability density function of a random variable X follows t-distribution is

   $$f(x) = \frac{1}{\pi(1+x^2)}, x \geq 0$$

   then degrees of freedom of the distribution will be 2.

   (b) The cars that enter in a Metro parking are classified either Indian-made or Foreign-made. To check that the car ender in the parking is in random order, we use Mann Whitney U test.

   (c) A random variable has the pdf

   $$f(x) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta$$

   If the critical region of testing the null and alternative hypotheses $H_0 : \theta = 2$ and $H_0 : \theta \neq 2$ is $X > 1$ then type-I error will be 0.5.

   (d) If sample mean ($\overline{X}$) is consistent estimator of the parameter θ then $\log(\overline{X})$ also consistent for log(θ).

   (e) In a random sample of 525 families owning television set in the region of New Delhi, it is found that 370 subscribe to Star Plus. A 99% confidence interval for the actual proportion of such families in New Delhi which subscribe to Star Plus will be (0.66, 0.74).

2. A baby-sister has 6 children under her supervision. The age of each child is as follows:

Child	Age(in years)
Sonu	10
Rishi	8
Lavnik	6
Chiya	4
Aman	2
Avishi	6

   i) Find the mean and SD of this finite population.

   ii) List all possible sample of size 3 from this population without replacement.

   iii) Construct the sampling distribution of mean.

iv) Compute the mean and standard error of the mean of the sampling distribution obtained in (iii). **(2+3+2+3)**

3. **(a)** A Pizza company would like to determine the average delivery time it can promise its customers. How large should the sample size be if it wants to be 95% confident that the sample estimate would not differ from the actual average delivery time by more than 1.5 minutes? The previous studies have shown the SD to be 7 minutes. **(4)**

(b) A sample of 400 shops was selected in a large metropolitan area to determine various information concerning to the consumer behaviour. One question, among the questions, asked, was **"Do you enjoy shopping for clothing?"** Out of 200 males,170 answered yes. Out of 250 females, 224 answered yes. Find 95% confidence interval for the difference of the proportions for enjoys shopping for clothing. **(6)**

4. The following data relate to the number of items produced in a shift by two workers A and B for some days:

A	26	37	40	35	30	30	40	26	30	35	45
B	19	22	24	27	24	18	20	19	25		

Assuming that the parent populations are normal, can it be inferred that B is more stable or consistent) worker compared to A? **(10)**

5. An economist wants to undertake a survey to establish if there is any relationship between the age of the person and his/her attitude towards the economy of the nation during current administration as compared to the economy during the previous administration. A random sample of 500 persons across the country was selected and they were put into categories of their age group and their respective opinions regarding the economy. These categories are shown in the following table:

Age Group	**Opinion on Economy**		
	Improve	**Remained Same**	**Worsened**
Below 30 years	40	64	96
30 to 50 years	84	52	24
Above 5o years	30	40	70

Test whether the age group and attitude towards economy are dependent at 5 % level of significance? **(10)**

6. Complete the following table, one is done for you:

S. No.	**Test For**	**Name of the Test**	**Test Statistic**	**Assumptions for Applying the Test**	**Test Type**
1	Population mean when population variance is known and population is normal	Z-test	$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$	1. Sample observations should be independent. 2. The measurement scale should be at least interval scale.	Parametric Test
2	Population mean when population variance is unknown and population is normal				

3	Two population means/medians when the form of populations is normal, samples are independent, population SDs are known and sample size is small				
4	Two population means/medians when the form of populations is normal, samples are independent, population SDs are unknown and sample size is small				
5	Two population means/medians when the form of populations is not known, samples are independent and sample size is small				
6	Independence of Two Attributes				

(10)

7. If the magnitude of the earthquakes recorded in a region of a country follows a distribution with parameter θ

$$f(x) = \frac{1}{\theta^2} xe^{-x/\theta}, x \geq 0, \theta \geq 0$$

then find

i) maximum likelihood estimator of the parameter θ,

ii) maximum likelihood estimate of the parameter on the basis of the following data:

Magnitude of the Earthquakes (on the Richter scale)	6.7	7.7	5.6	7.3	6.7	6.6	7.8	6.7	6.2	5.2

iii) Show that maximum likelihood estimator is unbiased and sufficient for parameter θ.

(8+2+10)

8. A medical insurance company is encouraging its subscribers to buy generic drugs because they are cheaper rather than brand name drugs. In order to test the company's claim, random samples of prices of 8 different drugs are compared. The following table represents the prices of both types of drugs bought from eight different pharmacies at random:

Pharmacy	**Generic Drug**	**Brand name Drug**
1	20	21
2	8	11
3	15	15
4	32	40
5	25	22
6	12	15

7	18	17
8	20	25

i) Formulate the hypotheses to test the claim

ii) What assumptions are necessary to apply the parametric test to test the claim.

iii) Apply the parametric test for testing the hypotheses formulated in (i)and under the assumptions (ii) using $\alpha = 0.05$.

iv) If assumptions in (ii) are not fulfilled, then which test is applied in this situation and why?

v) Apply the test for (iv) and compare the result with (iii). **(2+2+8+2+6)**

TUTOR MARKED ASSIGNMENT
MST-005: Statistical Techniques

Course Code: MST-005

Assignment Code: MST-005/TMA/2019

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) In a sequence of random numbers, generated through LCG" $x_i = (12x_{i-1} + 6)$ Mod 11 with $x_0 = 06$ the x_1, x_2 and x_3 will be 11, 17, 12.

   b) The probability of selection of a sample of n from the population by SRSWOR is $1/N$.

   c) While analysing the data of a 4 × 4 Latin Square design the error d.f. is equal to 10.

   d) In a Two way analysis of variance with 5 blocks & 5 treatments the degree of freedom for the total variation is 14.

   e) Suppose a random number generated by Middle. Square Method is 15, then the next random number will be 22.

2. 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	$\overline{X}_i$	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 Proprtional 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. (10)

3 a) In a class of Statistics, total number of students is 30. Select the linear and circulur systematic random samples of 10 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)

b) To determine the yield rate of wheat in a district of Punjab, 6 groups were constructed of 6 plots each. The data is given in the following table:

Plot No.	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
1	8	6	18	13	17	12
2	13	5	8	7	15	15
3	11	16	6	13	10	11
4	26	5	10	6	21	17

5	13	16	16	7	20	8
6	31	5	20	2	25	10

Select a cluster sample of 3 clusters from the given data and find sample mean. (5)

4. Three varieties A, B and C of wheat are shown in five plots each of the following fields per acre as obtained:

Plots	A	B	C
1	40	35	60
2	50	25	45
3	35	50	65
4	70	45	60
5	55	45	70

Set up a table of analysis of variance and find out whether there is significant difference between the fields of these varieties. (10)

5. The heights of the trees are given in the table given below for an experiment involving six treatments in four randomised blocks. The treatments are indicated by numbers with brackets:

Blocks	**Heights of Trees (cms)**					
I	(1) 247	(3) 277	(2) 206	(4) 162	(5) 162	(6) 249
II	(3) 227	(2) 288	(1) 273	(4) 150	(6) 225	(5) 170
III	(6) 263	(4) 196	(1) 385	(3) 368	(2) 395	(5) 154
IV	(5) 177	(2) 310	(1) 285	(4) 141	(3) 349	(6) 226

Analysis the data using suitable design and draw the conclusion. (10)

6. The following table gives the layout and the results of a 2^3 factorial design laid out in four replicates:

Block –I	nk	kp	p	np	1	k	n	nkp
	291	391	312	373	101	265	106	450
Block-II	kp	p	k	nk	n	nkp	np	1
	407	324	272	306	89	449	338	106
Block-III	p	1	np	kp	nk	k	n	nkp
	323	87	324	423	334	279	128	471
Block-IV	np	nk	n	p	k	1	nkp	kp
	361	272	103	324	302	131	437	435

Determine the effect the 3 kinds of fertilizer; Nitrogen N, Potash K and phosphate P on potato crop yield. (14)

7. Four doctors each test four treatments for a certain disease and observe the number of days each patient takes to recover. The results are as follows (recovery time in days)

	Treatment			
Doctor	**1**	**2**	**3**	**4**
A	10	14	19	20
B	11	15	17	21
C	9	12	16	19
D	8	13	17	20

Determine whether there is significant difference in terms of (i) doctor's and (ii) treatments. (12)

8 a) Generate a complete cycle for the LCG given $x_i = (5x_{(i-1)} + 3)$ Mod(16), with $x_0 = 5"$. A man tosses an unbiased coin 10 times. Using the first 10 random numbers generated above, obtain a sequence of heads and tails. (10)

b) Following U(0,1) were generated by a random number generation method:

0.251	0.769	0.153	0.575	0.390	0.335	0.066	0.104	0.200
0.019	0.597	0.729	0.012	0.922	0.691	0.817	0.064	0.539
0.419	0.305	0.449	0.998	0.919	0.470	0.372	0.851	0.643
0.509	0.913	0.445	0.464	0.447	0.279	0.413	0.494	0.972
0.983	0.432	0.368	0.574					

Apply chi-square test to test the fit the distribution. (10)

TUTOR MARKED ASSIGNMENT
MSTE-001: Industrial Statistics-I

Course Code: MSTE-001

Assignment Code: MSTE-001/TMA/2019

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: **(5×2=10)**

(a) The process capability of a manufacturing process of a certain type of bolt, with mean diameter 2 inches and standard deviation 0.05 inches, will be 0.30

(b) If the probability of accepting a lot of satisfactory quality is 0.9401, then the producer's risk will be 0.9401.

(c) Hurwicz criterion is a method to solve the problems that involve decision-making under certainty.

(d) Two independent components of a system are connected in series. If the reliability of these components are 0.8 and 0.6 respectively, then the reliability of the system will be 0.92.

(e) For a two person zero sum game, the payoff matrix is given below:

		Player B				
		$\mathbf{B_1}$	$\mathbf{B_2}$	$\mathbf{B_3}$	$\mathbf{B_4}$	$\mathbf{B_5}$
Player A	$\mathbf{A_1}$	3	4	5	–2	3
	$\mathbf{A_2}$	1	6	–3	3	7

Then second column will be dominated by the first column.

2. (a) A glass manufacturer produces hand mirrors. Each mirror is supposed to meet company standards for such things as glass, color of handle, and so on. To control for these features, the company quality people randomly sample 40 mirrors in every shift and determine how many of the mirrors are out of compliance on at least one feature. Shown here are the data for 15 such samples.

Sample	**Number out of Compliance**	**Sample**	**Number out of Compliance**
1	2	6	0
2	0	7	4
3	6	8	3
4	3	9	2
5	1	10	2

Use the data to construct a suitable chart. Observe the results and comment on the control of the process as indicated by the chart. **(7)**

(b) A process of manufacturing an item is in statistical control with $\overline{\overline{X}} = 250$ and $\overline{R} = 3.5$. The specification limits of the process are 200 ± 8 1.90. Calculate the process capability. Does it appear that the manufacturing process is capable of meeting the specification requirements? $(d_2 = 2.059)$ **(3)**

3. A leather bag manufacturing company supplies bags in lots of size 150 to a buyer. A single sampling plan with n = 10 and c = 1 is being used for the lot inspection. The company and the buyer decide that AQL = 0.08 and LTDD = 0.16. If there are 15 defective bags in each lit, compute the

i) Probability of accepting the lot,
ii) Producer's risk and consumer's risk,
iii) AOQ, if the rejected lots are screened and all defective bags are replaced by non-defectives, and
iv) Average total inspection. **(2+4+2+2)**

4. An office supply company ordered a lot of 4,000 pens. When the lot arrives the company inspector will randomly inspect 15 pens. If more than three pens in the sample are non-conforming, the lot will be rejected. If fewer than two pens are non-conforming, the lot will be accepted. If two, three pens are non-conforming, a second sample of size 10 will be taken. The value of c_2 is 3. Suppose the inspector finds two nonconforming items 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 type of acceptance plan is used here?
ii) What is the probability of accepting the lot at the first sample?
iii) What is the probability of accepting the lot at the second sample? **(2+3+5)**

5. The manager of a Flower Shop promises its customers delivery within four hours on all flower orders. All flowers are purchased on the previous day and delivered to Parker by 8.00 AM the next morning. The daily demand for roses is as follows:

Dozens of roses	150	160	170	200
Probability	0.2	0.3	0.4	0.1

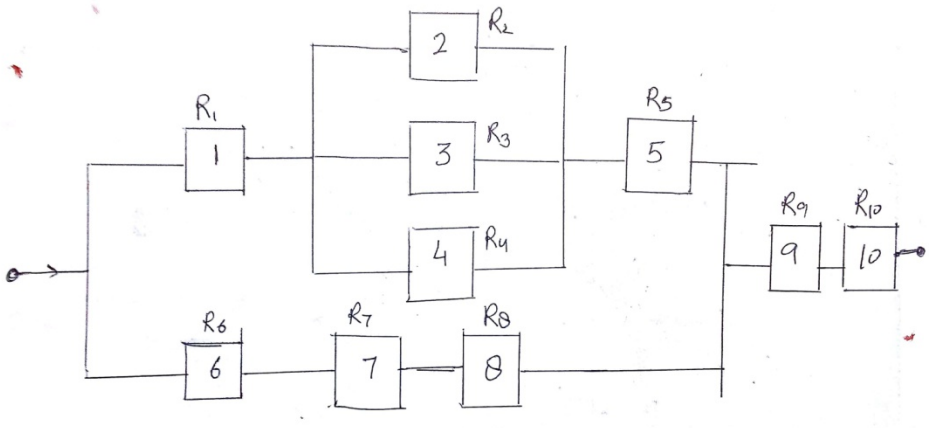
The manager purchases roses Rs 50 per dozen and sells them for Rs 150. All unsold roses are donated to a local hospital. How many dozens of roses should Parker order each evening to maximize its profits? What is the optimum expected profit? **(10)**

6. In a small town, there are only two stores, ABC and XYZ that handle sundry goods. The total number of customers is equally divided between the two, because price and quality of goods sold are equal. Both stores have good reputation in the community, and they render quality good customer run annual pre-New Year sales during the first week of January. Sales are advertised through a local newspaper, radio and television media. With the aid of an advertising firm store ABC constructed the game matrix given below.(Figures in the matrix represent a gain or loss of customers).

	Strategy of XYZ		
Strategy of ABC	**Newspaper**	**Radio**	**Television**
Newspaper	30	40	– 80
Radio	0	15	– 20
Television	90	20	50

Determine optimal strategies for both ABC and XYZ. **(10)**

7. (a) The system shown below is made up of ten components. Components 2, 3 and 4 are not identical and at least two component of this group must be available for system success.



What is the system reliability if $R_1 = R_5 = 0.85$, $R_6 = R_7 = R_8 = 0.80$, $R_9 = R_{10} = 0.70$ $R_2 = 0.40$, $R_3 = 0.50$ and $R_4 = 0.60$. **(10)**

(b) A system having Weibull failure distribution with pdf as

$$f(t) = \frac{1}{\theta}\bar{e}^{t/\theta}\ t \geq 0$$

i) Compute the reliability function of the system,
ii) find the reliability for a 100 days mission for $\theta = 2$,
iii) show that the hazard rate is constant,
iv) find the MTTF, and
v) what is the life of the system if reliability of 0.90 is desired. **(2+1+2+2+3)**

8. A factory producing dry-cells wanted to test the life of cells produced daily. The cells will be considered satisfactory if their life is 25 hours. For this, a sample of 4 cells was drawn on 7 consecutive days. The results are as follows:

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
24	16	20	26	20	25	26
20	23	25	18	22	23	24
25	23	24	28	24	25	20
20	20	21	20	22	20	25

i) Estimate the process mean.

ii) Determine the centre line and control limits of the control charts for controlling the process mean and process variability.

iii) Plot the charts on the graph papers.

iv) What managerial decisions should be made on the basis of these charts?

v) If necessary, compute revised control limits. **(2+4 +6 +4+4)**

TUTOR MARKED ASSIGNMENT
MSTE-002: Industrial Statistics-II

Course Code: MSTE-002

Assignment Code: MSTE-002/TMA/2019

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)

   (i) If 10 is added to each of the entries of the cost matrix of a 3 x 3 assignment problem, then the total cost of an optimal assignment for the changed cost matrix will increase by 10.

   (ii) The solution to a transportation problem with m-rows (supplies) and n-columns (destinations) is feasible if number of positive allocations is m + n.

   (iii) If the arrival rate is 6 per hour and service rate is 2 per hour, then the probability of no customer in queue is 0.7.

   (iv) A time series is a set of values arranged in geographical order.

   (v) If the coefficient of determination is 0.933, the number of observations and independent variables are 10 and 2, respectively, then Adjusted R^2 will be 0.84.

2(a) Rewrite the following linear programming problem in Standard form:

Minimise $Z = 2x_1 + x_2 + 4x_3$

Subject to the Constraints:

$$-2x_1 + 4x_2 \leq 4$$
$$x_1 + 2x_2 + x_3 \geq 5$$
$$2x_1 + 3x_3 \leq 2$$
$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$$ (4)

(b) Solve the following LPP using graphical method:

Maximize $Z = 3x_1 + 2x_2$

Subject to the Constraints:

$$-2x_1 + x_2 \leq 1$$
$$x_1 \leq 2$$
$$x_1 + x_2 \leq 3$$
$$x_1, x_2 \geq 0$$ (8)

3 A department head has four subordinates, and four tasks to be performed. The subordinates differ in efficiency, and the tasks differ in their intrinsic difficulty. His estimate, of the time each man would take to perform each task, is given in the table below:

Tasks	Subordinates			
	E	F	G	H
A	18	26	17	11
B	13	28	14	26

C	38	19	18	15
D	19	26	24	10

How should the tasks be allocated, one to a subordinate, so as to minimise the total man hour? (10)

4 a) Use graphical method to minimise the time added to process the following jobs on the machines shown:

Job 1:	Sequence	A	B	C	D	E
	Time	3	4	2	6	2
Job 2:	Sequence	B	C	A	D	E
	Time	5	4	3	2	6

Calculate the total time elapsed to complete both the jobs. (5)

b) The following data comprising the number of customers (in hundred) and monthly sales (in thousand Rupees):

Number of Customers (in hundred)	4	6	6	8	10	14	18	20	22	26	28	30
Monthly Sales (in thousand Rs)	1.8	3.5	5.8	7.8	8.7	9.8	10.7	11.5	12.9	13.6	14.2	15

Calculate the residuals and determine the standardised residuals for the model

$Y = 2.6185 + 0.4369\ X$ (10)

5(a) For the following series of observations, verify that the 4-year centered moving average is equivalent to a 5-year weighted moving average with weights 1, 2, 2, 2, 1, respectively:

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Annual Sales (in 000 Rs.)	20	60	10	50	30	70	20	60	40	80	30

(6)

(b) A statistician collected data of 50 values with two independent variables X_1 and X_2, in the process of fitting the following models (i) $Y = B_0 + e$ (ii) $B_0 + B_1X_1 + e$ (iii) $Y = B_0 + B_2X_2 + e$ and (iv) $Y = B_0 + B_1X_1 + B_2X_2 + e$. The following results obtained:

$\hat{B}_0 = 52.38,\ \hat{B}_1 = 31.6161, \hat{B}_2 = 0.0414,\ SS(B_0) = 125.26,\ SS(B_0; B_1) = 179.08$

$SS(B_0, B_2) = 171.25,\ SS(B_0, B_1, B_2) = 180.79$ and $\hat{\sigma}^2 = 0.91$

Apply all the three selection procedures to choose an appropriate model. (10)

6 Customers arrive at a one-man barber shop according to the Poisson process with a mean inter arrival time of 12 minute. Customers spend an average of 10 minute. in the barber's chair.

i) What is the expected number of customers in the barber shop and in the queue?

ii) Calculate the percentage of time an arrival can walk straight into the barber's chair without having to wait.

iii) How much time can a customer expect to spend in the barber's shop?

iv) Management will provide another chair and hire another barber, when a customer's waiting time in the shop exceeds 1.25hour. How much must the average rate of arrivals increase to warrant a second barber?

v) What is the average time customers spend in the queue?

vi) Calculate the percentage of customers who have to wait prior to getting into the barber's chair.

vii) What is the probability that more than 3 customers are in the system? (7)

7 a) Calculate seasonal indices by the ratio to moving average method from the following data:

Year / Quarter	2001	2002	2003	2004
Q_1	750	860	900	1000
Q_2	600	650	720	780
Q_3	540	630	660	720
Q_4	590	800	850	930

(10)

b) For the following Auto regressive model

$$X_t = 0.6X_{t-1} - 0.3X_{t-2} + a_t$$

i. Verify whether the series is Stationary.
ii. Obtain $\rho_k : k = 1,2,3,4 \text{ and } 5$
iii. Plot the Correlogram. (08)

8. Consider the following Transportation problem:

Factory	Godowns						Stock Available
	1	2	3	4	5	6	
A	7	5	7	7	5	3	60
B	9	11	6	11	-	5	20
C	11	10	6	2	2	8	90
D	9	10	9	6	9	12	50
Demand	60	20	40	20	40	40	

It is not possible to transport any quantity from Factory B to Godown 5. Determine basic Feasible Solution by Vogel's Approximation Method and optimum solution using MODI method. (12)

TUTOR MARKED ASSIGNMENT

MSTL-001: Basic Statistics Lab

Course Code: MSTL-001

Assignment Code: MSTL-001/TMA/2019

Maximum Marks: 100

Note:

1. **All questions are compulsory.**
2. **Solve the following questions in MS Excel.**
3. **Take the screenshots of the final output/spreadsheet.**
4. **Paste all screenshots in the assignments booklets with all necessary hypotheses, interpretation, etc.**

Q 1 A researcher conducted a stimulus reaction experiment on 50 subjects to a threatening stimulus and 50 subjects to a nonthreatening stimulus. The reaction times of all 100 students to the nearest tenth of a secondwere recorded as follows:

Reaction Times of Students

Nonthreatening Stimulus									
2.4	2.2	2.7	2.5	2.4	2.6	2.5	2.6	2.5	2.5
2.4	2.4	2.2	2.3	2.6	2.4	2.6	2.8	2.5	2.4
2.6	2.5	2.6	2.3	2.1	2.4	2.4	2.7	2.5	2.3
2.4	2.6	2	2.5	2.7	2.4	2.4	2.4	2.6	3
2.4	2.4	2.3	2.3	2.6	2.7	2.2	2.1	2.1	2.2
Threatening Stimulus									
2.2	2.1	1.8	2.5	1.7	1.9	2	2.2	1.9	1.8
1.8	2.4	1.9	2.2	1.8	2.1	2.1	2.1	1.8	2.3
2.3	2.1	2	2.9	2	2	2.2	2.1	2.3	2.3
1.9	2.2	2	2.3	1.7	1.9	2	1.9	2	1.9
1.7	2.1	1.7	2.1	2.1	2.2	2	2.1	2.1	2.1

Answer the following:

i) Which stimulus has more average reaction?

ii) Which stimulus showed more consistent result?

iii) Compute suitable width of the class intervals for both types of stimulus and construct the continuous frequency distribution for both.

iv) Also, construct the histogram for both types of stimulus and interpret the results.

(2+5+10+8)

Q 2 The marks (out of 100) given to the twenty candidates in a job interview on the basis of various parameters of the selection, are presented in the table given below:

S. No.	Examiner I	Examiner II	Examiner III
1	58	82	60

2	53	77	59
3	40	76	56
4	68	91	71
5	51	89	69
6	50	79	62
7	65	98	74
8	60	92	71
9	61	85	63
10	47	77	60
11	64	93	73
12	58	96	74
13	60	100	79
14	59	89	66
15	57	93	70
16	67	89	68
17	54	71	51
18	71	100	75
19	54	71	53
20	58	91	68

i) Use the rank correlation coefficient to determine which pair of examiners had the nearest approach in selection of candidates.

ii) Also represent the marks given by three examiners using box plot.

(15+10)

Q 3 For the data given in question 1, if the reaction timeforboth stimulus are normally distributed, then answers the following questions:

i) Are the variances of the distributions of the reaction time for both stimulusequal at 5 % level of significance?

ii) Also, check whether the average reaction time for nonthreatening stimulusis more than the threatening stimulusat 5 % level of significance? **(15+10)**

Q 4 A researcher wanted to compare the mean distance covered by four brands of golf balls struck by various types of clubs. The six golfers(A, B, C, D, E and F) are assigned at random to 36 cells of the square with the restrictionthat each golfer will hit each brand of ball only once usingeach type ofclub. The following design was obtained in which 6 brands of balls are arranged in “columns" and 6types of clubs are in “rows":

		Brand of Ball					
		1	2	3	4	5	6
Type of Club	1	A	B	C	D	E	F
	2	B	C	D	E	F	A
	3	C	D	E	F	A	B
	4	D	E	F	A	B	C
	5	E	F	A	B	C	D
	6	F	A	B	C	D	E

The average distance coveredby each ball is given as follows:

		Brand of Ball					
		1	**2**	**3**	**4**	**5**	**6**
Type of Club	**1**	133	139	140	140	145	138
	2	136	141	143	146	139	142
	3	140	138	142	139	139	141
	4	129	132	137	136	140	138
	5	132	144	143	142	142	140
	6	138	140	141	137	141	139

Assuming that the effect of each golfer, ball and club are normally distributed with approximately equal variances; analyse the design at 1% level of significance. Test whether the effect of the different golfer, ball and club on the distance covered are significant or not. If significant, do the pair-wise comparison between them.

(25)

TUTOR MARKED ASSIGNMENT
MSTL-002: Industrial Statistics Lab

Course Code: MSTL-002

Assignment Code: MSTL-002/TMA/2019

Maximum Marks: 100

Note:

1. **All questions are compulsory.**
2. **Solve the following questions in MS Excel.**
3. **Take the screenshots of the final output/spreadsheet.**
4. **Paste all screenshots in the assignments booklets with all necessary hypotheses, interpretation, etc.**

Q1 A Company uses a process to paint refrigerators with a coat of enamel. During each shift, a sample of 5 refrigerators is selected (2 hours apart) and the thickness of the paint (in mm) is determined. If the enamel is too thin, it will not provide enough protection. If it is too thick, it will result in an uneven appearance with running and wasted paint.In this regards, total 20 subgroups consisting of a sample of 5 refrigerators in each subgroupwere selected. The following table lists the measurements from 20 consecutive shifts:

Shift No.	Thickness (in mm)				
1	2.4	2	2.3	2.1	2.4
2	2.3	2.1	2.3	2	2.5
3	2	2	2.1	2.2	2.1
4	2.5	2	2.1	2.3	2.4
5	2.3	2.2	2.3	1.8	2.5
6	1.9	2	2.4	1.9	2.3
7	1.9	2.3	2.1	1.7	2
8	2.5	2.3	2.3	2.4	2.2
9	2.1	2.5	2.1	1.9	2
10	2.3	2	1.7	2.2	2.1
11	2.8	2.7	3.2	2.5	2.7
12	2.1	2.5	1.9	2.6	2.2
13	1.8	2.9	2.2	2.3	2.5
14	1.9	2.5	1.8	1.9	2.1
15	2.1	2.7	2.2	2.2	1.7
16	2.8	2.3	2.3	2.5	1.8
17	2.6	2.1	2.6	1	1.5
18	1.6	1.3	2.3	3	3
19	2	2.3	2.4	2.5	2.9
20	1.5	2.5	2	1.7	2.6

The quality manager of this company needs to construct suitable control charts for variability as well as average to infer whether the thickness of paint process is under control or not. If it is out-of-control,construct the revised control charts.

(25)

Q 2 A company is monitoring the percentage of line items that are shipped correctly from a major supplier. The company's quality people randomly sampled the number of line items shipped each weekfrom the supplier as well as the number of line items shipped correctly. They have collected data for the past 25 weeks.The data are displayed in the following table:

Week	Number of Shipped Items	Number of Correctly Shipped Items
1	86	74
2	76	76
3	42	34
4	94	93
5	115	115
6	37	34
7	79	71
8	89	83
9	64	54
10	60	58
11	116	115
12	120	98
13	46	44
14	103	102
15	81	79
16	126	122
17	117	102
18	17	13
19	64	52
20	60	59
21	92	87
22	104	103
23	118	103
24	89	87
25	77	74

Construct a suitable control chart for fraction of incorrectly shipped itemsto check whether the process issaid to be in a state of control or not using both approaches. Also construct the revised control charts, if necessary.

(25)

Q 3 A researcher wants to study the varieties of cigarettes according to their tar, nicotine, and carbon monoxide content as each of these three substances considered hazardous to a

smoker's health. The following table presents data on tar, nicotine, and carbon monoxide content (in milligrams) and weight (in grams) for a sample of 25 brand of cigarettes in a recent year:

Tar	Nicotine	Weight	Carbon Monoxide
12.9	0.74	0.87	12.4
15.2	0.94	0.98	15.4
28.6	1.91	1.05	22.3
6.8	0.55	0.81	9
2.9	0.28	0.83	4.2
13.8	0.92	0.77	13.8
7.6	0.64	0.91	7.8
11.2	0.83	0.81	11.1
15.4	1	0.82	15.1
13.7	0.9	0.77	14.2
12.5	0.89	0.85	11.8
13.9	0.78	0.82	13.2
6.6	0.45	0.86	8.8
10.2	0.66	1.01	9
7.8	0.62	0.74	8.3
-0.2	0.01	0.67	0.3
15.8	1.14	0.8	17.3
11.6	0.96	0.92	11.4
14.6	0.84	0.84	16.3
3.3	0.3	0.8	3.7
13.3	0.89	0.89	14.7
6.1	0.49	0.87	7.3
7.4	0.57	0.85	9.4
14	0.9	0.83	12.7
10.8	0.7	1	13.7

i) Prepare a scatter plots matrix to get an idea about the relationship among the variables.

ii) Develop the best fitted multiple regression modelconsidering carbon monoxide content as a function of tar content, nicotine content, and weight using stepwise regression procedure.

iii) Does the fitted regression model satisfy the linearity and normality assumptions?

(8+9+8)

Q 4 An amusement park manager wishes to improve the quality and activities based on the total number of kids visiteddaily. The scheduling will be prepared based on the daily levels of customers in the past 10 weeks. The numbers of kids visited in the park during that period were given below:

Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday

1	359	524	287	257	460	376	248
2	195	274	228	293	354	193	318
3	135	204	265	139	291	124	115
4	180	259	106	278	339	118	303
5	120	189	250	124	289	132	308
6	295	208	333	150	219	280	154
7	248	157	264	293	168	348	357
8	237	394	415	394	243	520	345
9	504	565	439	615	415	485	688
10	574	764	759	709	667	891	857

i) Determine the seasonal indices for these data using a 7-day moving averages.

ii) Obtain the deseasonalised values.

iii) Fit the appropriate trend for the deseasonalised data using the least-squares method by matrix approach that best describes these data.

iv) Project the number of kids visited on Wednesday of the 52^{th} week.

v) Plot the original data, the deseasonalised data, and the trend.

(8+2+8+2+5)