ASSIGNMENT BOOKLET

Post Graduate Diploma in Applied Statistics

(Specialisations in Industrial Statistics/Biostatistics)

MST-001 to MSTL-003

(Valid from 1st January 2023 to 31th December, 2023)

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
(2023)

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-003 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, 2023.

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 2023,** you should submit your TMAs by **March 31, 2023.** Similarly, If you wish to appear in the **TEE, December 2023,** you should submit your TMAs by **September 30, 2023.**

We wish you good luck.

TUTOR MARKED ASSIGNMENT

MST-001: Foundation in Mathematics and Statistics

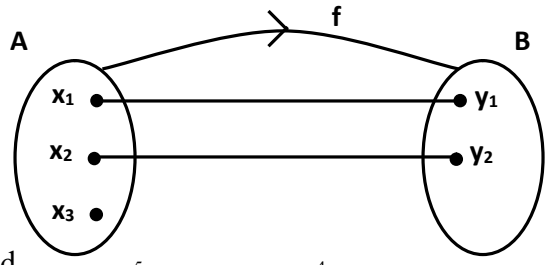
Course Code: MST-001

Assignment Code: MST-001/TMA/2023

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 3×2.

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/2023

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, (αB) = 15, (Aβ) = 20 and (αβ) = 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 σ_y if $\sigma_x = 3$. **(10)**

5. 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)**

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:

	Union Locals			
Reaction	A	B	C	Total
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/2023

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

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

   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:

a) Age lying between 18 – 25 and meet 3 accidents

b) Age lying between 18 – 40 and meet 1 accident

c) Age more than 25 years and meet at most one accident

d) Having no accident in the year

e) 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
   a) Scooter driver
   b) Car driver **(5 + 5)**

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

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

   Find
   a) The marginal distributions.
   b) The conditional distribution of Y given X = 2 **(5 + 5)**

6. a) 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?
   b) 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. a) (i) 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 $P[X + Y > 1]$
   (ii) Comment on the statement: "The mean of a binomial distribution is 4 and variance 5".

   b) 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
   (i) exactly 2
   (ii) more than 3
   (iii) at least one
   individuals suffer from bad reaction. **(10 + 10)**

8. a) A die is rolled. If the outcome is a number greater than 2, what is the probability that it is an odd prime number?
   b) 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.
   c) 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/2023

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 θ such that $Var(T_1) = 1/2n$ and $Var(T_2) = 2/n$ 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 θ for the following distribution :

   $$f(x,\theta)=\theta e^{-\theta x}; \quad 0 \le x,\ \theta > 0$$ **(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 followed normal distribution by two workers Rahul and Ramesh for a number of days:

Rahul	19	22	24	27	24	18	20	19	25	
Ramesh	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, 36 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 6 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					

(20)

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)

TUTOR MARKED ASSIGNMENT
MST-005: Statistical Techniques

Course Code: MST-005

Assignment Code: MST-005/TMA/2023

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)

   c) In SRSWOR, the possible numbers of sample of size n from a population of size N if sampling is done with replacement is N^n.

   d) One-way analysis of variance is a generalization of the two sample t-test.

   e) If experimental error is reduced considerably and the efficiency of the design is decreased.

   f) If strata are heterogeneous then stratified sampling schemes provides estimates with greater precision.

   g) If one wants to convert random numbers selected from two digit numbers 00-99 to uniformly distributed U (0, 1) variables then one has to divide them by 99.

2. Assume that you have to perform a sample survey for Family expenditure of the faculty of Indira Gandhi National Open University. Then explain the main steps involved in the planning and execution of that sample survey. (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	8	7	12
2	10	5	9
3	7	10	13
4	14	9	12
5	11	9	14

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

5. An experiment was planned to study the effect of Sulphate, Potash and Super Phosphate on the yield of potatoes. All the combinations of 2 levels of Super Phosphate [0 cent (p_0) and 5 cent (p_1)/ acre] and two levels of Sulphate and Potash [0 cent (k_0) and 5 cent (k_1)/acre] were studied in a randomised block design with 4 replications each. The (1/70) yields [lb per plot = (1/70) acre] obtained are given in table below:

Blocks	**Yields (lbs per plot)**			
I	(1) 23	k 25	p 22	kp 38
II	p 40	(1) 26	k 36	kp 38
III	(1) 29	k 20	pk 30	p 20
IV	kp 34	k 31	p 24	(1) 28

Analyse the data and give your conclusions. (10)

6. By generating 10 uniform random variate U (0, 1) estimate the integral

$$\theta = \frac{1}{\sqrt{2\pi}} \int_{-1}^{2} e^{-x^2/2}\, dx$$

Recognizing this function as probability density function of N (0, 1), compare the value of $\hat{\theta}$ with θ. (10)

7. A sample of 100 villagers is to be drawn from a population of villages 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

Find the number of 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 that of SRSWOR.

(20)

8. A manufacturer wishes to determine the effectiveness of four types of machines (A, B, C and D) in the production of bolts. To accumulate this, the numbers of defective bolts produced for each of two shifts in the results are shown in the following table:

Machine	First shift					Second Shift				
	M	T	W	Th	F	M	T	W	Th	F
A	6	4	5	5	4	5	7	4	6	8
B	10	8	7	7	9	7	9	12	8	8
C	7	5	6	5	9	9	7	5	4	6
D	8	4	6	5	5	5	7	9	7	10

Perform an analysis of variance to determine at 5% level of significance, whether there is a difference (a) Between the machines and (b) Between the shifts. (20)

TUTOR MARKED ASSIGNMENT
MSTE-001: Industrial Statistics-I

Course Code: MSTE-001
Assignment Code: MSTE-001/TMA/2023
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 the average number of defects in an item is 4, the upper control limit of the c-chart will be 12.

b) The specification limits and natural tolerance limits are same in statistical quality control.

c) If the probability of making a decision about acceptance or rejection of a lot on the first sample is 0.80 and the sizes of the first and second samples are 10 and 15, respectively, then the average sample number for the double sampling plan will be 25.

d) Two independent components of a system are connected in series configuration. If the reliabilities of these components are 0.1 and 0.30, respectively then the reliability of the system will be 0.65.

e) A point in the pictorial representation of a decision tree having states of nature as immediate sub-branches is known as decision point.

2. To monitor the manufacturing process of mobile phones, a quality controller randomly selected 100 mobile phones from the production line, each day over 15 days. The mobile phones were inspected for defectives and the number of defective mobile phones found each day was recorded. The data are given below:

Subgroup Number	Number of Mobile Phones Inspected	Number of Defective Mobile Phones
1	100	3
2	100	6
3	100	4
4	100	6
5	100	20
6	100	2
7	100	6
8	100	7
9	100	3
10	100	0
11	100	6
12	100	15
13	100	5
14	100	7
15	100	6

i) Determine the trial centre line and control limits for the fraction defective using the above data. (4)

ii) Contract the control chart on graph paper and determine that the process is stable or not. If there is any out-of-control point, determine the revised centre line and control limits.(6)

3. A shirt manufacturing company supplies shirts in lots of size 250 to the buyer. A single sampling plan with n = 20 and c = 1 is being used for the lot inspection. The company and the buyer decide that AQL = 0.04 and LTPD = 0.10. If there are 15 defective in each lot, compute the

i) probability of accepting the lot. (2)

ii) producer's risk and consumer's risk. (4)

iii) average outgoing quality (AOQ), if the rejected lots are screened and all defective shirts are replaced by non-defectives. (2)

iv) average total inspection (ATI). (2)

4. The failure density function of a random variable T is given by

$$f(t) = \begin{cases} 0.011\, e^{-0.011t}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Calculate, the

i) reliability of the component. (2)

ii) reliability of the component for a 100 hour mission time. (2)

iii) mean time to failure (MTTF). (2)

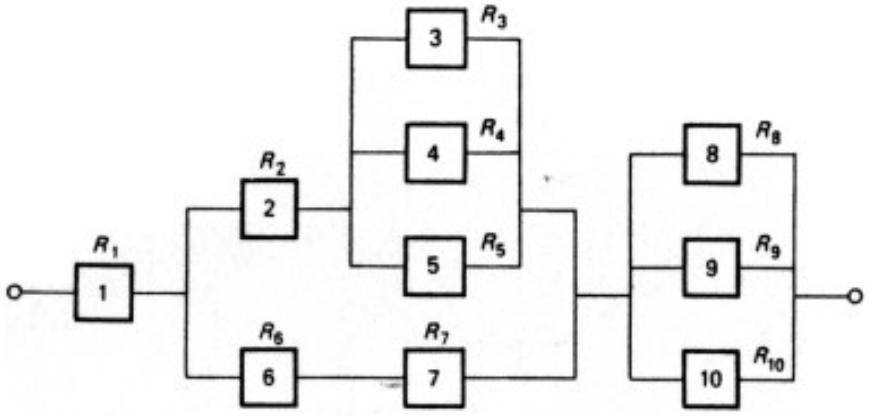
iv) median of the random variable T. (2)

v) life of the component, if the reliability of 0.96 is desired. (2)

5. Solve the two-person zero-sum game having the following payoff matrix for player A: (10)

		Player B				
		B_1	**B_2**	**B_3**	**B_4**	**B_5**
Player A	**A_1**	3	4	5	–2	3
	A_2	1	6	–3	3	7

6. The system shown below is made up of ten components. Components 3, 4 and 5 are not identical and at least one component of this group must be available for system success. Components 8, 9 and 10 are identical and for this particular group it is necessary that two out of the three components functions.



What is the system reliability if $R_1 = R_3 = R_5 = R_7 = R_9 = 0.85$ and $R_2 = R_4 = R_6 = R_8 = R_{10} = 0.95$ (10)

7. A small electronic device is designed to emit a timing signal of 200 milliseconds (ms) duration. In the production of this device, 10 subgroups of four units are taken at periodic intervals and tested. The results are shown in the following table:

Subgroup Number	Duration of Automatic Signal (in ms)			
	a	b	c	d
1	195	201	194	201
2	204	190	199	195
3	195	197	205	201
4	211	198	193	180
5	204	193	197	200
6	200	202	195	200
7	196	198	197	196
8	201	197	206	207
9	200	202	204	192
10	203	201	209	192

i) Estimate the process mean and standard deviation. (4)

ii) Determine the centre line and control limits for the process mean and process variability.(4)

iii) By plotting the charts on graph paper, determine that the process is stable or not with respect to the process mean and process variability. If necessary, compute revised control limits. (12)

8. The failure data of 10 electronic components are shown in the table given below:

Failure Number	1	2	3	4	5	6	7	8	9	10
Operating Time (in hours)	3	5	31	51	76	116	140	182	250	302

Estimate, the

i) reliability. (5)

ii) cumulative failure distribution. (5)

iii) failure density. (5)

iv) failure rate functions. (5)

TUTOR MARKED ASSIGNMENT

MSTE-002: Industrial Statistics-II

Course Code: MSTE-002

Assignment Code: MSTE-002/TMA/2023

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) The Set $ ={(x, y) : $0 \le y \le 5$ when $0 \le x \le 2$ and $3 \le y \le 5$ when $2 \le x \le 7$ } is not a convex set.

   b) 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.

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

   d) The Value $d_i \ge 3$ indicates an outlying observation in regression analysis.

   e) Variations which occur due to natural forces and operate in a regular and periodic manner over a span of less than or equal to one year are termed as cyclic variations.

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

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

   Subject to the Constraints:

   $$-2x_1 + 4x_2 \le 4$$
   $$x_1 + 2x_2 + x_3 \ge 5$$
   $$2x_1 + 3x_3 \le 2$$
   $$x_1 \ge 0, x_2 \ge 0, x_3 \ge 0$$ (5)

   (b) Solve the following LPP using graphical method:

   Maximize $Z = 3x_1 + 2x_2$

   Subject to the Constraints:

   $$-2x_1 + x_2 \le 1$$
   $$x_1 \le 2$$
   $$x_1 + x_2 \le 3$$
   $$x_1, x_2 \ge 0$$ (5)

3. Solve the following LPP using Simplex method:

   Maximize $Z = x_1 + 2x_2$

   Subject to the Constraints:

   $$-x_1 + 2x_2 \le 8$$
   $$x_1 + 2x_2 \le 12$$
   $$x_1 - x_2 \le 3$$
   $$x_1, x_2 \ge 0$$ (10)

4. 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)

5. 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. (4)

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

$Y = 2.6185 + 0.4369\ X$ (6)

6. a) A Statistician collected the data of 78 values with two independent variable X_1 and X_2, and considered the four models:

(i) $Y = B_0 + e$; (ii) $Y = B_0 + B_1 X_1 + e$; (iii) $Y = B_0 + B_1 X_1 + e$ and
(iv) $Y = B_0 + B_1 X_1 + B_2 X_2 + e$.

The results obtained are: $\hat{\sigma}^2 = 0.91$, $SS(B_0) = 652.42$, $SS(B_0, B_1) = 679.34$, $SS(B_0, B_2) = 654.00$, and $SS(B_0, B_1, B_2) = 687.79$. Find the additional contribution of (i) X_2 over X_1 and (ii) X_1 over X_2. Test whether their inclusion in the model is justified. (5)

b) Fifteen successive observations on a stationary time series are as follows:

34, 24 23 31 38 34 35 31 29 28 25 27
32 33 30

Calculate r_6, r_7, r_8 and r_9 and plot the correlogram. (5)

7. 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

(20)

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:

(a) Basic Feasible Solution by Vogel's Approximation Method.

(b) Optimum solution using MODI method. (20)

TUTOR MARKED ASSIGNMENT
MSTE-003: Biostatistics-I

Course Code: MSTE-003
Assignment Code: MSTE-003/TMA/2023
Maximum Marks: 100

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

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

   (a) Suppose A is the exposure and B is a confounding factor for outcome C, then there will be a path from A to C via B.

   (b) Doing exercise may also be a regimen.

   (c) In clinical trials, a control only may be: treatment or no treatment.

   (d) In Greville's method, the central death rate is more in the life table than the population.

   (e) In a slope ratio assay, both regression lines have common slope.

2. Explain with examples:

   (a) Various sources of demographic data in India

   (b) Types of bioassays

   (c) LD50 and ED50

(5+5+5)

3.(a) Differentiate between complete and abridged life tables.

(b) The data on population and number of deaths for different age groups of Districts A and B in the year 2001 were collected in the following table:

Age Group (Years)	District A		District B	
	Population	No. of Deaths	Population	No. of Deaths
0 - 5	55,300	385	51,165	805
5 –15	109,125	410	98,170	510
15-35	1,72,050	675	1,68,450	790
35-50	1,15,600	1625	1,40,625	990
50 & above	2,65,775	3288	2,40,900	2485

Calculate standardised death rate by direct method, taking population of District A as the standard population.

(5+5)

3. In a parallel-line assay, total 18 guinea pigs, 4 guinea pigs each from 3 different breeds were selected and classified into 4 groups for each breed. Two groups were administered with two doses of the standard preparation and remaining two groups with two doses of the test preparation. The responses of these doses are recorded in the following table:

	Dose (Standard)			**Dose (Test)**		
Breed	**10 (in µL)**	**15 (in µL)**	**20 (in µL)**	**5 (in µL)**	**10 (in µL)**	**15 (in µL)**
1	25	42	55	20	43	64
2	23	47	52	23	42	66
3	22	38	58	24	44	67

(i) Determine the dose-response regression models for both preparations.

(ii) Test whether the dose-response curves of both preparations are linear and parallel to each other or not.

(iii) Interpret whether the relative potency and its confidence interval can be computed or not.

(8+15+2)

5.(a) If D^+ and D^- denote presence and absence, respectively, of a disease and T^+ and T^- denote test result as positive and negative, respectively, then on the basis of the following information:

	D^+	D^-	Total
T^+	145	2000	2145
T^-	15	48000	48015
Total	160	50000	50160

Find: (i) sensitivity (ii) specificity (iii) positive and negative predictive values.

(b) Obtain sample size for the following given information:

$\delta = 0.04,\ \pi_1 = 0.72,\ \pi_2 = 0.84,\ \alpha = 0.01,\ \beta = 0.25$ **(5+5)**

6.(a) Explain design and analysis of data of case control study in detail.

(b) Creatinine excretion is a parameter of kidney function. Generally speaking, lower values indicate better health. This depends on body weight. A researcher conducted a study on creatinine excretion in test group and control group to find the efficacy of a new drug. The subjects were randomly divided. He included 100 subjects in each group but for this exercise consider only 10 subjects in each group. The data obtained on creatinine level in these 10 subjects are as follows:

Test group:	16.6	19.8	17.1	17.0	15.6	20.3	24.7	18.5	17.6	22.0
Control group:	23.2	22.0	21.9	16.7	14.2	23.2	24.8	25.5	28.1	21.8

Do you think that creatinine excretion was really lower in the test group on average?

(15+5)

7. Suppose you try a regimen A on 1000 subjects and regimen B on 1600 subjects. Results of the trial show that efficacies of regimen A and B are 76% and 82% respectively. Suppose doctor determines 4% as superiority margin. Can you consider regimen B as superior to regimen A.

(10)

TUTOR MARKED ASSIGNMENT

MSTE-004: Biostatistics-II

Course Code: MSTE-004

Assignment Code: MSTE-004/TMA/2023

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) The value of sensitivity of the following results of a diagnostic test is 0.85.

Disease	Test result		Total
	+	–	
Present	170	30	200
Absent	20	280	300

(b) For the following cohort study, the relative risk for the lung cancer among smokers is 3.5.

	Lung Cancer	No Lung Cancer	Total
Smokers	100	1220	1320
Non-smokers	50	2260	2310

(c) The logit link function is $\log\left[-\log(1-\pi)\right]$.

(d) We define three indicator/dummy variables for a regressor variable with three categories.

(e) Left censoring occurs whenever the exact time of occurrence of an event is not known.

2. Differentiate between Chi-square tests for association and homogeneity of proportions. Also mention the assumptions of these tests.

(10)

3. A random sample of 250 patients was selected and their workout timing and diabetes status were recorded. The following table shows the workout timing and severity of diabetes:

Workout (in minutes)	Severity of diabetes status		
	Low	**Moderate**	**High**
0 −15	06	27	19
15 to 30	08	36	17
30 to 45	21	45	33
≥ 45	14	18	06

Test at 5% level of significance whether workout habit and diabetes are associated with to each other or not.

(10)

4.(a) Explain the assumptions underlying multiple linear regression model.

(b) Suppose a researcher wants to evaluate the effect of cholesterol on the blood pressure. The following data on serum cholesterol (in mg/dL) and systolic blood pressure (in mm/Hg) were obtained for 15 patients to explore the relationship between cholesterol and blood pressure:

S. No.	Cholesterol (mg/dL)	SBP (mm/Hg)
1	300	150
2	410	270
3	380	210
4	530	310
5	570	350
6	490	310
7	340	210
8	320	150
9	280	110
10	550	320
11	340	220
12	350	170
13	410	260
14	390	230
15	450	270

(i) Fit a linear regression model using the method of least squares.

(ii) Construct the normal probability plot for the data on serum cholesterol and systolic blood pressure.

(iii) Test the significance of the fitted regression model.

(5+15)

4. Write a short note on the following:

(i) Polytomous logistic models

(ii) Poisson regression

(iii) Kaplan and Meier method

(12)

6. The following data on diagnosis of coronary heart disease (where 0 indicating absence and 1 indicating presence), serum cholesterol (in mg/dl), resting blood pressure (in mmHg) and weight (in kg) were obtained for 80 patients to explore the relationship of coronary heart disease with cholesterol and weight.

S. No.	Serum Cholesterol (mg/dl)	Weight (kg)	Number of Patients having CHD	Total Number of Patients
1	420	60	10	20
2	450	68	15	30
3	400	54	4	15
4	510	74	2	10
5	480	62	1	5

(i) Fit a multiple logistic model for the dependence of coronary heart disease on the average serum cholesterol and weight considering $\hat{\beta}_0^0 = 4.279$, $\hat{\beta}_1^0 = -0.035$ and $\hat{\beta}_2^0 = 0.172$ as the initial values of the parameters (solve only for one Iteration).

(ii) Test the significance of the fitted model using Hosmer-Lemeshow test at 5% level of significance.

(12+8)

7.(a) Describe censoring and differentiate between different types of censoring with the help of examples which are not considered in Block 4 of MSTE-004.

(b) A study was conducted on 185 patients aged more than 45 years which are followed until the time of death or up to 10 years, whichever comes first. The patients have different covariates: age, gender (male/female), systolic blood pressure, smoking (yes/no), total serum cholesterol and diabetes (yes/no). The objective of this study is to determine which covariate influences the survival time. An analysis is conducted to investigate differences in all-cause mortality between men and women participating in the study. Suppose we obtain the following results after applying the Cox regression hazard model analyses:

Risk Factor	Parameter Estimate	SE
Age	0.150	0.010
Gender	0.450	0.150
Systolic Blood Pressure	0.015	0.008
Smoking	0.650	0.170
Total Serum Cholesterol	0.002	0.004
Diabetes	-0.350	0.250

(i) Obtain hazard ratio and interpret the results.

(ii) Find the 99% confidence interval for the hazard ratio.

(iii) Test whether the covariates are significant or not at 1% level of significance.

(8+10)

TUTOR MARKED ASSIGNMENT
MSTL-001: Basic Statistics Lab

Course Code: MSTL-001
Assignment Code: MSTL-001/TMA/2023
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 assignment booklets with all necessary interpretation and steps.**

Q 1 A cooking oil supplier distributed two types of oils, say Oil A and Oil B to a large numbers of retail stores. The supplier wants to compare the popularity of both oils. For this purpose, he selects a sample of 100 stores and tracks record of the sold oils (in litres) of each type at each store. The data are noted in the following table:

Store No.	Oil A	Oil B	Store No.	Oil A	Oil B
1	161	419	**51**	478	196
2	285	411	**52**	284	241
3	219	168	**53**	488	182
4	321	241	**54**	447	132
5	435	125	**55**	384	322
6	325	261	**56**	267	341
7	463	119	**57**	390	139
8	319	285	**58**	270	462
9	108	441	**59**	381	227
10	328	213	**60**	252	140
11	479	116	**61**	245	420
12	285	319	**62**	196	474
13	489	135	**63**	201	392
14	448	187	**64**	227	452
15	385	349	**65**	181	406
16	268	279	**66**	441	397
17	391	306	**67**	130	375
18	271	296	**68**	213	455
19	382	269	**69**	373	367
20	253	403	**70**	190	503
21	246	309	**71**	280	366
22	197	424	**72**	236	486

23	202	349	**73**	297	171
24	228	250	**74**	421	219
25	182	457	**75**	340	173
26	442	196	**76**	380	418
27	131	240	**77**	308	454
28	214	337	**78**	361	228
29	374	252	**79**	183	432
30	191	423	**80**	121	468
31	281	322	**81**	162	231
32	237	406	**82**	286	252
33	298	146	**83**	220	283
34	422	175	**84**	322	114
35	341	487	**85**	436	325
36	381	278	**86**	326	213
37	309	442	**87**	464	229
38	362	326	**88**	320	183
39	184	414	**89**	120	291
40	122	377	**90**	329	175
41	160	250	**91**	480	141
42	284	272	**92**	286	394
43	218	356	**93**	490	163
44	320	366	**94**	449	134
45	434	170	**95**	386	130
46	324	213	**96**	134	459
47	462	147	**97**	392	363
48	318	195	**98**	272	315
49	118	452	**99**	383	338
50	327	385	**100**	254	365

Answer the following:

i) Which type of oil has more average sales?

ii) Which oil shows greater variability in the sales?

iii) Determine the correlation between both types of oils.

iv) Compute suitable width of the class intervals for both oils,

v) Construct the continuous frequency distribution for both oils.

(4+6+3+4+8)

Q 2 The scores (out of 100) secured by 60 employees of three different departments D1, D2 and D3 who participated in a study, are presented in the following table:

Employee No.	Scores of D1	Scores of D2	Scores of D3	Employee No.	Scores of D1	Scores of D2	Scores of D3
1	54	78	56	31	59	76	57
2	49	73	55	32	57	87	66
3	36	72	52	33	46	80	62
4	64	87	67	34	57	82	61
5	47	85	65	35	48	78	59
6	46	75	58	36	65	90	66
7	61	94	70	37	69	94	70
8	56	88	67	38	43	73	54
9	57	81	59	39	36	68	48
10	43	73	56	40	43	66	48
11	60	89	69	41	56	90	66
12	54	92	70	42	52	73	56
13	56	96	75	43	57	83	61
14	55	85	62	44	45	69	51
15	53	89	66	45	46	75	58
16	63	85	64	46	58	88	64
17	50	67	47	47	49	73	53
18	67	96	71	48	60	92	68
19	50	67	49	49	63	81	59
20	54	87	64	50	51	78	57
21	41	69	49	51	53	76	58
22	53	83	60	52	47	76	56
23	55	85	64	53	38	68	52
24	58	76	59	54	46	82	63
25	36	70	54	55	39	66	47
26	49	71	51	56	67	91	71
27	62	95	74	57	61	82	61
28	66	88	65	58	56	83	60
29	53	75	56	59	48	67	50
30	49	88	64	60	35	68	50

i) Compute the correlation coefficient between scores of the employees working in department D1 and the joint effects of scores of the employees of departments D1 and D2.

ii) Compute the correlation coefficient between scores of the employees working in departments D1 and D2 after eliminating the linear effect of the scores of departments D3.

iii) Also represent the scores obtained by departments D1, D2 and D3 using box plot.

(7+7+11)

Q 3 An experiment was conducted to compare two metals: A and B, as bonding agents for an alloy material. Components of the alloy were bonded using the metals as bonding agents, and the pressures required to break the bonds were measured. The data for the pressures required for breaking the metal are given in the following table:

S. No.	Breaking Pressure		S. No.	Breaking Pressure	
	Metal A	Metal B		Metal A	Metal B
1	71.9	72.2	21	86.5	70.6
2	68.8	66.4	22	74.3	74.6
3	82.6	74.5	23	71.2	68.8
4	78.1	60.6	24	85	76.9
5	74.2	73.2	25	80.5	63
6	70.8	68.7	26	76.6	75.6
7	84.9	69	27	73.2	71.1
8	72.7	73	28	87.3	71.4
9	69.6	67.2	29	75.1	75.4
10	83.4	75.3	30	72	69.6
11	78.9	61.4	31	85.8	77.3
12	75	74	32	81.3	63.4
13	71.6	69.5	33	77.4	76
14	85.7	69.8	34	74	71.5
15	73.5	73.8	35	88.1	71.8
16	70.4	68	36	75.9	75.8
17	84.2	76.1	37	72.8	70
18	79.7	62.2	38	86.6	77.7
19	75.8	74.8	39	82.1	63.8
20	72.4	70.3	40	78.2	76.4

If the pressure required to break both metals are normally distributed, then answers the following questions:

i) Are the variances of the distributions of the pressure of Metals A and B equal at 5 % level of significance?

ii) If yes, check whether the average pressure for Metal A is more than the Metal B at 5 % level of significance?

(15+10)

Q 4 An investigation was performed to study the impacts of different types of machines on the production of a particular variety of toys. The six machines (A, B, C, D, E and F) are assigned at random to 36 cells of the square with the restriction that each machine is used only once by each operator and in each time-period. The following design was obtained in which 6 operators are arranged in “columns" and 6 time-periods are in “rows":

		Operator					
		1	2	3	4	5	6
Time Period	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 production in a day is given as follows:

		Operator					
		1	**2**	**3**	**4**	**5**	**6**
Time Period	**1**	142	148	149	149	154	147
	2	145	150	152	155	148	151
	3	149	147	151	148	148	150
	4	138	141	146	145	149	147
	5	141	153	152	151	151	149
	6	147	149	150	146	150	148

Assuming that the effect of each operator, time-period and machine are normally distributed with approximately equal variances, analyse the design at 1% level of significance. Test whether the effect of the different operators, time periods and machines on the production 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/2023

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 assignment booklets with all necessary interpretation and steps.**

Q 1 A manager of an amusement park wanted to study the waiting times of visitors for issuing entry tickets during a peak hour. A subgroup of 15 visitors was selected (one at each ten minutes interval during an hour) and the time (in minutes) was measured from the point each visitor entered in the line to when he or she began to be attended. The results of 40 days period are recorded in the following table:

Sample No.	1	2	3	4	5	6	7	8
Obs. 1	9.3	7.2	7.3	6.2	7.5	8.8	7.9	6.3
Obs. 2	7.0	6.9	8.1	9.3	7.7	7.3	9.4	8.7
Obs. 3	9.4	7.9	10.1	7.7	10.6	7.8	7.7	5.8
Obs. 4	6.7	7.3	8.7	9.7	8.1	7.7	9.3	9.1
Obs. 5	9.2	9.1	9.9	7.5	10.4	7.7	9.5	7.6
Obs. 6	6.2	6.1	8.1	8.5	6.9	7.5	8.6	7.9
Obs. 7	8.6	8.5	9.3	6.9	9.8	7.1	8.9	5.0
Obs. 8	5.9	6.5	7.9	8.9	7.3	8.6	8.5	8.3
Obs. 9	6.4	6.4	8.3	8.7	7.1	8.3	8.8	8.1
Obs. 10	8.8	8.7	9.5	7.1	10.0	7.3	9.1	6.9
Obs. 11	6.1	6.7	8.1	9.1	7.5	7.1	8.7	8.5
Obs. 12	8.7	8.5	9.4	6.9	9.8	9.2	8.9	7.0
Obs. 13	7.1	7.0	9.3	9.7	7.9	7.4	9.5	9.0
Obs. 14	9.9	9.8	10.7	7.9	11.3	8.1	9.5	5.6
Obs. 15	6.7	7.4	9.0	10.2	8.3	7.9	9.8	9.5

Sample No.	9	10	11	12	13	14	15	16
Obs. 1	9.9	9.6	9.6	6.4	7.7	6.6	9.1	7.3
Obs. 2	9.1	8.8	7.8	8.5	8.3	7.9	11.3	8.6
Obs. 3	8.1	6.0	7.8	9.1	9.7	8.6	7.0	7.1
Obs. 4	9.3	9.2	8.3	8.9	9.1	8.3	11.7	9.0
Obs. 5	8.7	7.9	7.6	8.9	8.1	8.5	6.8	6.9
Obs. 6	8.4	8.0	7.1	7.7	9.2	7.1	10.5	7.8

Obs. 7	9.0	7.3	7.0	8.3	8.4	7.8	6.2	6.3
Obs. 8	7.9	8.4	7.5	8.1	8.7	7.5	10.9	8.2
Obs. 9	8.7	8.2	7.3	7.9	9.4	7.3	10.7	8.0
Obs. 10	9.2	5.4	7.2	8.6	8.6	8.1	6.4	6.5
Obs. 11	8.8	8.6	7.7	8.4	8.9	7.7	11.1	8.4
Obs. 12	8.1	5.2	7.0	8.4	9.4	7.9	6.2	6.4
Obs. 13	9.7	9.2	8.1	8.8	10.6	8.1	12.1	8.9
Obs. 14	10.3	8.3	8.0	9.6	9.6	9.0	7.1	7.2
Obs. 15	9.0	9.7	8.5	9.3	8.4	8.6	12.5	9.4

Sample No.	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Obs. 1	9.2	6.9	7.2	6.1	7.4	8.7	7.8	6.2
Obs. 2	6.2	7.5	9.1	9.2	7.6	7.2	9.3	8.6
Obs. 3	9.3	9.2	8.1	7.6	10.5	7.7	9.6	5.7
Obs. 4	6.6	5.8	9.2	9.6	8.0	7.6	8.5	9.0
Obs. 5	9.1	9.0	7.4	7.4	10.3	7.6	9.4	7.5
Obs. 6	7.2	6.0	8.2	8.4	6.8	8.6	8.5	7.8
Obs. 7	8.5	8.4	9.2	6.8	9.7	7.0	8.8	4.9
Obs. 8	5.8	6.4	8.4	8.8	7.2	6.8	7.7	8.2
Obs. 9	5.6	7.0	8.4	8.6	7.0	6.6	8.7	8.0
Obs. 10	8.7	8.6	9.4	7.0	9.9	7.2	9.0	5.1
Obs. 11	6.0	5.2	8.6	9.0	7.4	7.0	8.0	8.4
Obs. 12	8.6	8.4	9.3	6.8	9.7	7.0	8.8	6.9
Obs. 13	8.2	6.9	9.4	9.6	7.7	7.3	9.7	8.9
Obs. 14	9.8	9.6	10.6	7.7	11.1	7.9	10.1	8.9
Obs. 15	6.6	7.3	9.6	10.1	8.2	7.8	8.9	9.4

Sample No.	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**
Obs. 1	8.1	9.8	9.5	6.3	7.6	9.1	9.0	7.2
Obs. 2	8.7	8.7	7.7	8.4	9.9	7.8	11.2	8.5
Obs. 3	9.7	7.3	7.7	9.0	8.0	8.5	5.8	7.0
Obs. 4	8.5	9.1	8.2	8.8	8.7	8.2	12.3	8.9
Obs. 5	9.5	6.5	7.5	8.8	9.0	8.4	6.7	6.8
Obs. 6	7.9	7.9	7.0	7.6	9.1	7.0	10.4	7.7
Obs. 7	8.9	6.5	6.9	8.2	9.3	7.7	5.0	6.2
Obs. 8	7.8	8.3	7.4	8.0	7.9	7.4	11.5	8.1
Obs. 9	8.1	8.1	7.2	7.8	9.3	7.2	10.6	7.9
Obs. 10	9.1	6.8	7.1	8.5	9.5	8.0	5.2	6.4

Obs. 11	8.0	8.5	7.6	8.3	8.1	7.6	11.7	8.3
Obs. 12	8.9	5.9	6.9	8.3	8.5	7.8	6.1	6.3
Obs. 13	9.0	9.1	7.9	8.7	10.5	8.0	11.9	8.8
Obs. 14	10.2	7.5	7.8	9.4	9.5	8.9	5.6	7.1
Obs. 15	8.9	9.5	8.4	9.2	9.0	8.5	13.2	9.3

Sample No.	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
Obs. 1	9.1	5.8	7.1	10.3	6.8	8.2	7.1	9.7
Obs. 2	8.7	6.3	8.0	8.4	9.1	8.9	8.4	12.1
Obs. 3	9.2	9.1	9.9	8.3	9.7	7.4	9.2	7.4
Obs. 4	6.5	7.2	7.7	8.8	9.5	9.7	8.9	12.5
Obs. 5	9.0	8.9	9.7	8.1	9.6	8.6	9.0	7.2
Obs. 6	7.9	5.5	7.9	7.5	8.2	6.8	7.6	11.2
Obs. 7	8.4	8.3	9.1	7.4	8.9	8.9	8.4	6.6
Obs. 8	5.7	6.4	7.0	8.0	8.7	9.3	8.0	11.7
Obs. 9	8.1	5.8	10.4	7.7	8.5	9.4	7.8	11.4
Obs. 10	8.6	8.5	9.3	7.7	9.1	9.2	8.6	6.8
Obs. 11	5.9	6.7	7.2	8.2	8.9	9.5	8.2	11.9
Obs. 12	8.5	8.3	9.2	7.5	8.9	9.9	8.4	6.6
Obs. 13	9.0	8.6	9.0	7.2	8.0	9.8	11.0	9.0
Obs. 14	9.7	9.5	10.5	7.9	7.8	10.3	10.8	8.7
Obs. 15	6.5	7.3	7.9	11.0	10.8	7.9	8.8	12.5

The manager of this amusement park needs to construct suitable control charts for variability as well as average to infer whether the waiting times of visitors for getting entry tickets is under statistical control or not. If it is out-of-control, she also computes the revised control limits, if necessary.

(25)

Q 2 A company designs decorative glass wall panels. Each panel is supposed to meet company standards for such things as glass thickness, ability to reflect, size of panel, quality of glass, colour, and so on. To control these features, the company quality people randomly sampled the panels from every shift and determined how many of the panels are out of compliance on at least one feature. The data collected from 25 such samples are shown below:

Sample No.	**Sampled Panels**	**Out of Compliance Panels**
1	69	2
2	71	3
3	66	3
4	65	9

5	69	3
6	67	2
7	70	4
8	73	5
9	71	3
10	69	2
11	74	5
12	79	2
13	74	4
14	74	3
15	71	2
16	67	3
17	69	2
18	75	4
19	71	2
20	72	4
21	69	3
22	74	2
23	69	4
24	69	2
25	66	3

Construct a suitable control chart for fraction of out of compliance panels to check whether the process is said 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 is interested in studying the impact of the weekly working hours and type of machine used (0 for Machine A and 1 for Machine B) on the number of produced items of a particular type. The data were collected for 40 weeks and shown in the following table:

Week	Produced Item	Working Hours	Machine Type
1	9	48	1
2	15	67	1
3	12	61	1
4	17	86	0
5	19	93	1
6	17	80	1
7	12	55	0
8	9	51	1

9	7	44	0
10	18	89	0
11	13	55	1
12	10	56	0
13	15	67	1
14	13	63	1
15	15	73	0
16	15	73	0
17	14	70	0
18	15	67	1
19	12	57	1
20	14	68	0
21	13	57	1
22	11	57	0
23	11	64	0
24	13	67	0
25	10	56	0
26	7	47	0
27	8	47	0
28	12	64	0
29	7	42	0
30	11	60	0
31	15	67	1
32	13	60	1
33	16	69	1
34	10	44	1
35	18	83	1
36	20	94	1
37	17	82	0
38	19	93	1
39	10	57	0
40	7	35	0

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

ii) Fit a linear regression model and its related analysis at 1% level of significance.

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

iv) Also, draw both fitted regression lines on the scatter plot.

(5+5+10+5)

Q 4 A popular café chain wishes to improve customer service and its employee scheduling based on the daily customers' footfall during past 10 weeks. The numbers of customers served in the restaurants during that period are given as follow:

Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	443	608	371	341	544	460	332
2	279	358	312	377	438	277	402
3	219	288	349	223	375	208	199
4	264	343	190	362	423	202	387
5	204	273	334	208	373	216	392
6	379	292	417	234	303	364	238
7	332	241	348	377	252	432	441
8	321	478	499	478	327	604	429
9	588	649	523	699	499	569	772
10	658	848	843	793	751	975	941

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 customers on Wednesday of the 22^{th} week.

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

(8+6+5+4+2)

TUTOR MARKED ASSIGNMENT
MSTL-003: Biostatistics Lab

Course Code: MSTL-003

Assignment Code: MSTL-003/TMA/2023

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 assignment's booklets with all necessary interpretation and steps.**

Q1(a) A random sample of 440 patients of cardiology department of a hospital was taken and their workout timing and severity of heart disease status were recorded. The following table shows the workout timing and severity of heart disease:

Workout (in minutes)	Severity of Heart Disease				
	Low	Mild	Moderate	High	Very High
No workout	5	13	26	21	23
0- 15	6	15	19	19	21
15 to 30	16	17	14	16	12
30 to 45	18	17	13	11	9
45 to 60	20	19	15	13	7
≥ 60	16	22	6	5	6

Test at 5% level of significance whether workout habit and heart disease are associated with to each other or not.

(b) To study the association between the diabetic patients and their family history of diabetes, the following data were obtained on 70 subjects.

Diabetes in Family	Diabetes in Subject		Total
	Yes	No	
Yes	14	3	**17**
No	3	50	**53**
Total	**17**	**53**	**70**

Which test is appropriate in this situation? Check whether the diabetes runs with generations in families or not at 5% level of significance using appropriate test.

(10+15)

Q2 A researcher is interested to check the relationship between the serum creatinine (in mg/dL) with the weight (in kg) and gender (0 if female and 1 if male). The data were collected from the hospital records to examine the contribution of these variables to serum creatinine. A total of 40 patients were sampled and the data are shown in the following table:

S. No.	Serum Creatinine	Weight	Gender	S. No.	Serum Creatinine	Weight	Gender
1	0.7	46	1	**21**	1.1	55	1
2	1.3	65	1	**22**	0.9	55	0
3	1	59	1	**23**	0.9	62	0
4	1.5	84	0	**24**	1.1	65	0
5	1.7	91	1	**25**	0.8	54	0
6	1.5	78	1	**26**	0.5	45	0
7	1	53	0	**27**	0.6	45	0
8	0.7	49	1	**28**	1	62	0
9	0.5	42	0	**29**	0.5	40	0
10	1.6	87	0	**30**	0.9	58	0
11	1.1	53	1	**31**	1.3	65	1
12	0.8	54	0	**32**	1.1	58	1
13	1.3	65	1	**33**	1.4	67	1
14	1.1	61	1	**34**	0.8	42	1
15	1.3	71	0	**35**	1.6	81	1
16	1.3	71	0	**36**	1.8	92	1
17	1.2	68	0	**37**	1.5	80	0
18	1.3	65	1	**38**	1.7	91	1
19	1	55	1	**39**	0.8	55	0
20	1.2	66	0	**40**	0.5	33	0

(i) Prepare a scatter plot to get an idea about the relationship among the variables.

(ii) Fit a linear regression model and its related analysis at 1% level of significance.

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

(iv) Also, draw both fitted regression lines on the scatter plot.

(5+5+10+5)

Q3 A hypothetical data of 40 patients on age (in years), weight (in kgs) and systolic blood pressure (in mm/hg) denoting 1 for high SBP and 0 for normal SBP are given in the following table:

S. No.	Age	Weight	SBP	S. No.	Age	Weight	SBP
1	52	60	0	**21**	47	48	0
2	56	68	1	**22**	42	45	0
3	51	54	0	**23**	45	57	0
4	63	74	1	**24**	56	83	1
5	54	62	0	**25**	49	63	0
6	51	67	0	**26**	56	94	1
7	51	66	0	**27**	55	87	1
8	54	65	0	**28**	53	67	0
9	59	71	1	**29**	65	70	1
10	51	89	1	**30**	44	70	0

11	56	72	1	**31**	48	54	0
12	55	72	1	**32**	61	79	1
13	46	57	0	**33**	45	85	0
14	42	54	0	**34**	63	98	1
15	52	63	0	**35**	49	78	0
16	65	67	1	**36**	65	80	1
17	50	67	0	**37**	60	70	1
18	42	53	0	**38**	53	98	1
19	50	68	1	**39**	41	53	0
20	39	55	0	**40**	50	70	1

For this data:

(i) Fit a multiple logistic regression model.

(ii) Test the significance of the individual model coefficients β_1 and β_2 at 5% level of significance.

(iii) Obtain the 95% confidence intervals for β_1 and β_2.

(iv) Determine the Nagelkerke pseudo R-squared.

(11+6+4+4)

Q4 A clinical study was conducted on individuals with advanced stage of Hepatocellular Carcinoma to test three lines of Treatments: T1, T2 and T3. Thirthy-six patients with stage III Hepatocellular Carcinoma who agreed to take part in the experiment were randomly allocated one of three line of Treatments T1, T2 and T3. The primary outcome was mortality, and patients were monitored for up to 60 months (5 years) after recruitment. The data (in months) so obtained are given as follows:

Patient ID	Survival time	Outcome	Treatment	Patient ID	Survival time	Outcome	Treatment
ID001	14	Died	T3	**ID019**	50	Died	T1
ID002	27	Unknown	T1	**ID020**	54	Unknown	T3
ID003	37	Unknown	T3	**ID021**	57	Died	T2
ID004	44	Died	T1	**ID022**	60	Survived	T3
ID005	27	Died	T2	**ID023**	20	Died	T1
ID006	29	Died	T3	**ID024**	22	Unknown	T2
ID007	50	Died	T2	**ID025**	11	Unknown	T2
ID008	31	Died	T1	**ID026**	12	Unknown	T1
ID009	54	Died	T2	**ID027**	57	Unknown	T3
ID010	32	Died	T2	**ID028**	60	Survived	T3
ID011	32	Unknown	T2	**ID029**	44	Died	T1

ID012	60	Unknown	T1	**ID030**	47	Unknown	T2
ID013	2	Unknown	T2	**ID031**	32	Died	T2
ID014	42	Died	T3	**ID032**	34	Died	T1
ID015	42	Unknown	T2	**ID033**	17	Died	T2
ID016	60	Died	T3	**ID034**	6	Died	T1
ID017	60	Survived	T3	**ID035**	50	Unknown	T3
ID018	47	Died	T3	**ID036**	14	Unknown	T2

For this data,

(i) Construct Kaplan and Meier survival curves.

(ii) Test whether there is a significant difference between the survival distributions of the patients under all treatments at 5% level of significance.