ASSIGNMENT BOOKLET

Post Graduate Diploma in Applied Statistics

(Specialisation in Industrial Statistics)

MST-001 to MSTL-002

(Valid from 1ˢᵗ January 2016 to 31ᵗʰ December, 2016)

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

ignou
THE PEOPLE'S
UNIVERSITY

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

Dear Student,

Please read the information on assignments in the Programme Guide that we have sent you after your enrolment. A weightage of 30%, as you are aware, has been earmarked for continuous evaluation, **which would consist of one tutor-marked assignment** for this course. The assignments for MST-001 to 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, 2016.

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 2016,** you should submit your TMAs by **March 31, 2016.** Similarly, If you wish to appear in the **TEE, December 2016,** you should submit your TMAs by **September 30, 2016.**

We wish you good luck.

TUTOR MARKED ASSIGNMENT
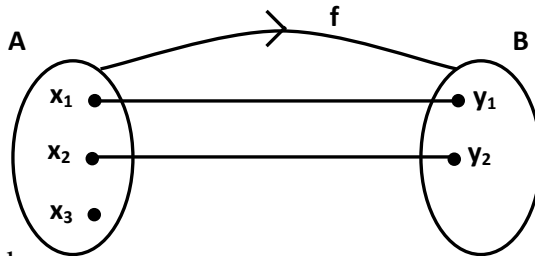## MST-001: Foundation in Mathematics and Statistics

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) $\dfrac{d}{dx}(9-7x)^5 = 45(9-7x)^4$

    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) $\dfrac{dy}{dx}$, 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

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

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

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

   b) Mean deviation is least when calculated from the median.

   c) The correlation coefficient between x and $(a – x)$ is –1.

   d) The regression coefficients b_{yx} and b_{xy} of a data are 1.2 and 0.8, respectively.

   e) If $(AB) = 10$, $(\alpha B) = 15$, $(A\beta) = 20$ and $(\alpha\beta) = 30$ then A and B are associated.

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

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

   b) If AM and GM of two numbers are 30 and 18, respectively, find the numbers. **(7+3)**

3. a) The frequency distribution of the marks obtained by the 25 students each of the two sections is given as follows:

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

   Find which section is more consistent.

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

4. The equations of two regression lines are given as follows:

   $$4x – 5y + 30 = 0$$
   $$20x − 9y – 107 = 0$$

   Calculate (i) regression coefficients, b_{yx} and b_{xy}; (ii) correlation coefficient $r(x, y)$; (iii) Mean of X and Y; and (iv) the value of 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:

Reaction	Union Locals			
	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)**

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):

Y\X	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 ~ B(5, 0.06) and

        Y ~ 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

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 $f(t) = \dfrac{1}{\pi(1+t^2)}; \; -\infty < t < \infty$ 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 :

   $$f(x,\theta) = \theta e^{-\theta x}; \;\; 0 \le x, \, \theta > 0 \tag{5+5}$$

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

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

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

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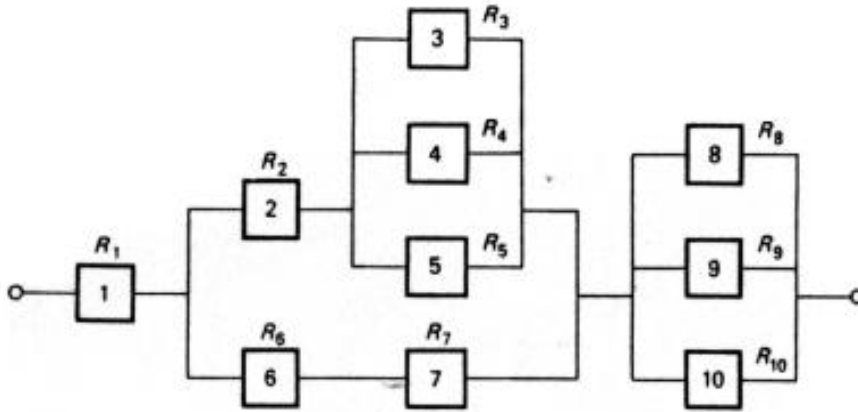
What is the system reliability if $R_1 = R_3 = R_5 = R_7 = R_9 = 0.85$ and
$R_2 = R_4 = R_6 = R_8 = R_{10} = 0.95$ (10)

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)

MSTE-002: Industrial Statistics-II

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. $(2\times5=10)$

   a) The Set $\$ = \{(x, y) : 0 \leq y \leq 5$ when $0 \leq x \leq 2$ and $3 \leq y \leq 5$ when $2 \leq x \leq 7 \}$ 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 $d_i \geq 3$ 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:

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

    (5)

    (b) Solve the following LPP using graphical method:

    Maximize $Z = 3x_1 + 2x_2$
    Subject to the Constraints:

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

    (5)

3. Solve the following LPP using Simplex method:

    Maximize $Z = x_1 + 2x_2$

    Subject to the Constraints: $-x_1 + 2x_2 \leq 8$
    $$x_1 + 2x_2 \leq 12$$
    $$x_1 - x_2 \leq 3$$
    $$x_1, x_2 \geq 0$$

    (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: $\hat{s}^2 = 0.91$, $SS(B_0) = 652.42$, $SS(B_0, B_1) = 679.34$, $SS(B_0, B_2) = 654.00$, 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
Q₁	750	860	900	1000
Q₂	600	650	720	780
Q₃	540	630	660	720
Q₄	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
MSTL-001: Basic Statistics Lab

Note:

1. **All questions are compulsory.**
2. **Solve the following questions in MS Excel 2007.**
3. **Take the screenshots of the final output/spreadsheet.**
4. **Paste all screenshots in the assignments booklets with all necessary interpretation and steps.**

Q 1. A Soap manufacturing company was distributing a particular type of brand, say Brand A through a large number of retail stores. These stores also sell another famous brand of soap, say Brand B. The manager of the company wants to compare the popularity of the newly launched soap (brand A) with the old popular Brand B of the soap. For this purpose, she selects a sample of 100 stores of each brand. The data are recorded in the following table:

Store No.	Brand A	Brand B	Store No.	Brand A	Brand B
1	154	412	51	471	189
2	278	404	52	277	234
3	212	161	53	481	175
4	314	234	54	440	125
5	428	118	55	377	315
6	318	254	56	260	334
7	456	112	57	383	132
8	312	278	58	263	455
9	101	434	59	374	220
10	321	206	60	245	133
11	472	109	61	238	413
12	278	312	62	189	467
13	482	128	63	194	385
14	441	180	64	220	445
15	378	342	65	174	399
16	261	272	66	434	390
17	384	299	67	123	368
18	264	289	68	206	448
19	375	262	69	366	360
20	246	396	70	183	496
21	239	302	71	273	359
22	190	417	72	229	479
23	195	342	73	290	164
24	221	243	74	414	212
25	175	450	75	333	166
26	435	189	76	373	411
27	124	233	77	301	447
28	207	330	78	354	221
29	367	245	79	176	425
30	184	416	80	114	461
31	274	315	81	155	224
32	230	399	82	279	245

33	291	139	83	213	276
34	415	168	84	315	107
35	334	480	85	429	318
36	374	271	86	319	206
37	302	435	87	457	222
38	355	319	88	313	176
39	177	407	89	113	284
40	115	370	90	322	168
41	153	243	91	473	134
42	277	265	92	279	387
43	211	349	93	483	156
44	313	359	94	442	127
45	427	163	95	379	123
46	317	206	96	127	452
47	455	140	97	385	356
48	311	188	98	265	308
49	111	445	99	376	331
50	320	378	100	247	358

Answer the following: (4+8+3+4+6)

i) Which soap brand has more average sales?

ii) Which soap brand shows greater variability using coefficient of variation in the sales?

iii) Determine the correlation between both brands of soap.

iv) Compute suitable width of the class intervals for both brands.

v) Construct the continuous frequency distribution for both brands.

Q 2. A Company has 20 operators in a particular department. The long-serving operators feel that they should have a promotion based on length of service built into their job structure. An assessment of their efficiency by their department manager and the personnel department produces a score of efficiency based on different parameters. This is shown below together with their length of service.

S. No.	Length of Service	Efficiency Scores
1	3	67
2	10	85
3	6	69
4	10	83
5	5	65
6	6	72
7	7	72
8	6	73
9	8	82
10	7	75
11	5	67
12	9	78
13	9	79
14	8	79
15	9	80
16	8	85
17	4	64

18	6	84
19	8	85
20	6	74

Compute the Spearman's rank correlation coefficient between the length of service and efficiency scores? (25)

Q 3. For the data given in Question 1, the manager also wishes to compare both brands to get the answers of the following questions manually as well as using data analysis toolpak:
   i) Is there enough evidence that the average sale of soap for brand A is more than the average sales of the Brand B at 5 % level of significance?
   ii) Are the variances of the distributions of soaps of brand A and brand B equal at 5 % level of significance? (15+10)

Q 4. Suppose that a fast-food chain wants to evaluate the service of the eight restaurants. The customer service director for the chain hires seven evaluators with varied experience in food-service evaluation to act as raters. To reduce the effect of the variability from rater to rater you use a randomised block design, with raters serving as the blocks. The eight restaurants are the groups of interest.

The seven raters evaluate the service of each of the eight restaurants in a random order. A rating scale from 0 (low) to 100 (high) is used. The following table summarises the results:

Raters	Restaurants							
	A	B	C	D	E	F	G	H
1	75	66	87	79	73	64	85	86
2	82	80	93	81	80	78	91	92
3	81	72	95	85	79	70	87	93
4	85	68	92	81	83	66	90	91
5	89	71	97	89	87	69	95	96
6	83	73	99	91	81	71	85	97
7	82	80	93	81	80	78	91	89

The effect of evaluation of each rater on the service of four restaurants is normally distributed with approximately equal variances.

i) Analyse the design at 2% level of significance.
ii) Is the average service of the eight restaurants significantly different? If the difference between the averages services of the eight restaurants is significant, do the pair-wise comparison between them.

(15+10)

TUTOR MARKED ASSIGNMENT
MSTL-002: Industrial Statistics Lab

Note:

1. **All questions are compulsory.**
2. **Solve the following questions in MS Excel 2007.**
3. **Take the screenshots of the final output/spreadsheet.**
4. **Paste all screenshots in the assignments booklets with all necessary interpretation and steps.**

Q 1 The branch manager of a bank wanted to study the waiting times of customers for issuing the demand draft during the peak 11 am to 1:30 pm. A subgroup of 15 customers was selected (one at each ten minutes interval during the hour) and the time in minutes was measured from the point each customer entered the line to when he or she began to be served. The results of 40 days period were as under.

Sample No.	1	2	3	4	5	6	7	8
Obs. 1	8.2	6.1	6.2	5.1	6.4	7.7	6.8	5.2
Obs. 2	5.9	5.8	7	8.2	6.6	6.2	8.3	7.6
Obs. 3	8.3	6.8	9	6.6	9.5	6.7	6.6	4.7
Obs. 4	5.6	6.2	7.6	8.6	7	6.6	8.2	8
Obs. 5	8.1	8	8.8	6.4	9.3	6.6	8.4	6.5
Obs. 6	5.1	5	7	7.4	5.8	6.4	7.5	6.8
Obs. 7	7.5	7.4	8.2	5.8	8.7	6	7.8	3.9
Obs. 8	4.8	5.4	6.8	7.8	6.2	7.5	7.4	7.2
Obs. 9	5.3	5.3	7.2	7.6	6	7.2	7.7	7
Obs. 10	7.7	7.6	8.4	6	8.9	6.2	8	5.8
Obs. 11	5	5.6	7	8	6.4	6	7.6	7.4
Obs. 12	7.6	7.4	8.3	5.8	8.7	8.1	7.8	5.9
Obs. 13	6	5.9	8.2	8.6	6.8	6.3	8.4	7.9
Obs. 14	8.8	8.7	9.6	6.8	10.2	7	8.4	4.5
Obs. 15	5.6	6.3	7.9	9.1	7.2	6.8	8.7	8.4

Sample No.	9	10	11	12	13	14	15	16
Obs. 1	8.8	8.5	8.5	5.3	6.6	5.5	8	6.2
Obs. 2	8	7.7	6.7	7.4	7.2	6.8	10.2	7.5
Obs. 3	7	4.9	6.7	8	8.6	7.5	5.9	6
Obs. 4	8.2	8.1	7.2	7.8	8	7.2	10.6	7.9
Obs. 5	7.6	6.8	6.5	7.8	7	7.4	5.7	5.8
Obs. 6	7.3	6.9	6	6.6	8.1	6	9.4	6.7
Obs. 7	7.9	6.2	5.9	7.2	7.3	6.7	5.1	5.2
Obs. 8	6.8	7.3	6.4	7	7.6	6.4	9.8	7.1

Obs. 9	7.6	7.1	6.2	6.8	8.3	6.2	9.6	6.9
Obs. 10	8.1	4.3	6.1	7.5	7.5	7	5.3	5.4
Obs. 11	7.7	7.5	6.6	7.3	7.8	6.6	10	7.3
Obs. 12	7	4.1	5.9	7.3	8.3	6.8	5.1	5.3
Obs. 13	8.6	8.1	7	7.7	9.5	7	11	7.8
Obs. 14	9.2	7.2	6.9	8.5	8.5	7.9	6	6.1
Obs. 15	7.9	8.6	7.4	8.2	7.3	7.5	11.4	8.3

Sample No.	17	18	19	20	21	22	23	24
Obs. 1	8.1	5.8	6.1	5	6.3	7.6	6.7	5.1
Obs. 2	5.1	6.4	8	8.1	6.5	6.1	8.2	7.5
Obs. 3	8.2	8.1	7	6.5	9.4	6.6	8.5	4.6
Obs. 4	5.5	4.7	8.1	8.5	6.9	6.5	7.4	7.9
Obs. 5	8	7.9	6.3	6.3	9.2	6.5	8.3	6.4
Obs. 6	6.1	4.9	7.1	7.3	5.7	7.5	7.4	6.7
Obs. 7	7.4	7.3	8.1	5.7	8.6	5.9	7.7	3.8
Obs. 8	4.7	5.3	7.3	7.7	6.1	5.7	6.6	7.1
Obs. 9	4.5	5.9	7.3	7.5	5.9	5.5	7.6	6.9
Obs. 10	7.6	7.5	8.3	5.9	8.8	6.1	7.9	4
Obs. 11	4.9	4.1	7.5	7.9	6.3	5.9	6.9	7.3
Obs. 12	7.5	7.3	8.2	5.7	8.6	5.9	7.7	5.8
Obs. 13	7.1	5.8	8.3	8.5	6.6	6.2	8.6	7.8
Obs. 14	8.7	8.5	9.5	6.6	10	6.8	9	7.8
Obs. 15	5.5	6.2	8.5	9	7.1	6.7	7.8	8.3

Sample No.	25	26	27	28	29	30	31	32
Obs. 1	7	8.7	8.4	5.2	6.5	8	7.9	6.1
Obs. 2	7.6	7.6	6.6	7.3	8.8	6.7	10.1	7.4
Obs. 3	8.6	6.2	6.6	7.9	6.9	7.4	4.7	5.9
Obs. 4	7.4	8	7.1	7.7	7.6	7.1	11.2	7.8
Obs. 5	8.4	5.4	6.4	7.7	7.9	7.3	5.6	5.7
Obs. 6	6.8	6.8	5.9	6.5	8	5.9	9.3	6.6
Obs. 7	7.8	5.4	5.8	7.1	8.2	6.6	3.9	5.1
Obs. 8	6.7	7.2	6.3	6.9	6.8	6.3	10.4	7
Obs. 9	7	7	6.1	6.7	8.2	6.1	9.5	6.8
Obs. 10	8	5.7	6	7.4	8.4	6.9	4.1	5.3
Obs. 11	6.9	7.4	6.5	7.2	7	6.5	10.6	7.2
Obs. 12	7.8	4.8	5.8	7.2	7.4	6.7	5	5.2
Obs. 13	7.9	8	6.8	7.6	9.4	6.9	10.8	7.7
Obs. 14	9.1	6.4	6.7	8.3	8.4	7.8	4.5	6
Obs. 15	7.8	8.4	7.3	8.1	7.9	7.4	12.1	8.2

Sample No.	33	34	35	36	37	38	39	40
Obs. 1	8	4.7	6	9.2	5.7	7.1	6	8.6
Obs. 2	7.6	5.2	6.9	7.3	8	7.8	7.3	11
Obs. 3	8.1	8	8.8	7.2	8.6	6.3	8.1	6.3
Obs. 4	5.4	6.1	6.6	7.7	8.4	8.6	7.8	11.4
Obs. 5	7.9	7.8	8.6	7	8.5	7.5	7.9	6.1
Obs. 6	6.8	4.4	6.8	6.4	7.1	5.7	6.5	10.1
Obs. 7	7.3	7.2	8	6.3	7.8	7.8	7.3	5.5
Obs. 8	4.6	5.3	5.9	6.9	7.6	8.2	6.9	10.6
Obs. 9	7	4.7	9.3	6.6	7.4	8.3	6.7	10.3
Obs. 10	7.5	7.4	8.2	6.6	8	8.1	7.5	5.7
Obs. 11	4.8	5.6	6.1	7.1	7.8	8.4	7.1	10.8
Obs. 12	7.4	7.2	8.1	6.4	7.8	8.8	7.3	5.5
Obs. 13	7.9	7.5	7.9	6.1	6.9	8.7	9.9	7.9
Obs. 14	8.6	8.4	9.4	6.8	6.7	9.2	9.7	7.6
Obs. 15	5.4	6.2	6.8	9.9	9.7	6.8	7.7	11.4

The manager of this bank needs to construct the suitable control charts for variability as well as average to infer whether the waiting times of customers for issuing the demand draft is under control or not. Construct the suitable control chart. Also compute the revised control limits, if necessary. (25)

Q 2 The branch manager of the bank recorded the total number of draft issued for last 25 months and also the number of committed errors of a particular type that has been made in preparing the demand drafts. The results are given in the following table:

Day	Number of Issued Drafts	Errors in Drafts
1	150	4
2	156	8
3	171	8
4	150	18
5	150	8
6	144	4
7	153	8
8	162	12
9	156	10
10	150	2
11	165	12
12	180	6
13	165	12
14	165	10
15	156	2
16	144	8
17	150	6
18	168	12
19	156	4

20	159	8
21	150	6
22	165	2
23	150	10
24	150	6
25	141	8
26	125	5
27	130	10
28	125	10
29	125	14
30	120	10
31	135	5
32	130	10
33	125	12
34	150	13
35	130	3
36	120	15
37	125	7
38	140	15
39	130	12
40	125	4

The manager needs to set up a suitable control chart for the number of errors to check whether the number of errors in a state of control or not. Also computes the revised control limits, if necessary. (25)

Q 3 A company conducted a study on its employees to see the relationship of several variables with an employ's IQ. For this purpose, fifty employees were selected and one IQ as well as five different personality tests were given to them. Each employ's IQ was recorded along with scores on five tests. The data are shown in the following table:

Employee	Test 1	Test 2	Test 3	Test 4	Test 5	IQ
1	83	80	78	77	67	99
2	73	85	67	80	63	92
3	81	80	71	81	68	94
4	96	86	82	83	56	99
5	84	73	75	75	68	94
6	72	74	71	67	59	79
7	84	79	84	84	69	97
8	54	86	61	69	53	92
9	86	85	79	78	76	94
10	42	71	60	80	56	86
11	83	72	72	78	74	98
12	63	86	65	85	56	83
13	69	76	64	85	61	98

14	81	84	65	79	64	96
15	50	85	71	65	75	76
16	82	79	82	79	71	98
17	72	78	82	65	61	91
18	80	84	61	74	52	93
19	89	77	81	78	52	98
20	83	84	76	79	70	93
21	71	85	72	64	56	78
22	83	74	81	63	76	96
23	79	79	77	67	75	91
24	85	80	61	68	61	93
25	41	78	84	69	53	85
26	82	82	83	69	74	97
27	62	75	62	67	69	82
28	68	83	81	82	59	97
29	55	76	81	83	62	95
30	49	77	69	63	55	75
31	84	73	73	67	68	76
32	74	85	82	75	66	93
33	82	76	86	64	69	95
34	69	75	60	61	63	72
35	85	82	85	83	62	95
36	73	79	68	75	51	80
37	85	82	81	82	74	98
38	78	79	83	77	75	93
39	87	86	78	79	65	95
40	43	72	62	64	64	87
41	84	75	76	74	72	99
42	64	78	62	75	55	84
43	70	86	81	81	57	99
44	78	84	82	85	62	97
45	51	74	77	81	73	77
46	84	80	82	65	67	97
47	82	70	78	76	58	92
48	70	74	67	60	51	77
49	89	76	79	76	61	95
50	76	77	70	72	64	90

Now determine the most appropriate regression model for the employee's IQ using stepwise approach at 5 % level of significance and interpret the results. Does the final regression model satisfy the linearity and normality assumptions? (25)

Q 4   A restaurant manager wishes to improve customer service and employee scheduling based on the daily levels of customers in the past 10 weeks. The scheduling based on the daily levels of customers in the past 10 weeks. The numbers of customers served in the restaurants during that period were given below:

Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	439	604	367	337	540	456	328
2	275	354	308	373	434	273	398
3	215	284	345	219	371	204	195
4	260	339	186	358	419	198	383
5	200	269	330	204	369	212	388
6	375	288	413	230	299	360	234
7	328	237	344	373	248	428	437
8	317	474	495	474	323	600	425
9	584	645	519	695	495	565	768
10	654	844	839	789	747	971	937

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 52th week.
v)   Plot the original data, the deseasonalised data, and the trend.   (6+4+10+2+3)