ASSIGNMENT BOOKLET

Post Graduate Diploma in Applied Statistics

(Specialisation in Industrial Statistics)

MST-001 to MSTL-002

(Valid from 1st January, 2018 to 31st December, 2018)

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

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

Dear Student,

Please read the information on assignments in the Programme Guide that we have sent you after your enrolment. A weightage of 30%, as you are aware, has been earmarked for continuous evaluation, **which would consist of one tutor-marked assignment** for this course. The assignments for MST-001 to MSTL-002 have been given in this booklet.

Instructions for Formatting Your Assignments

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

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

ENROLLMENT NO :……………………………………………

NAME :……………………………………………

ADDRESS :……………………………………………

……………………………………………

……………………………………………

PROGRAMME CODE: ………………………..

COURSE CODE: ……………………………….

COURSE TITLE: ………………………………

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

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

2) Use only foolscap size writing paper (but not of very thin variety) for writing your answers.

3) Leave 4 cm margin on the left, top and bottom of your answer sheet.

4) Your answers should be precise.

5) This assignment is to be submitted at the Study Centre.

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

6) This assignment is valid up to December 31, 2018.

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 2018,** you should submit your TMAs by **March 31, 2018.** Similarly, If you wish to appear in the **TEE, December 2018,** you should submit your TMAs by **September 30, 2018.**

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**. Give reason in support of your answer: **(5×2=10)**

    (a) Which term of the A.P. $20,\ 19\frac{1}{2},\ 19,\ 18\frac{1}{2},\ 18,\ 17\frac{1}{2},\ ...$ is the first negative term?

    (b) $\dfrac{dy}{dx}$ of the functions $x = 5 + 3t^2,\ y = \log t$ is 6.

    (c) The rule f shown in the following figure is a function.
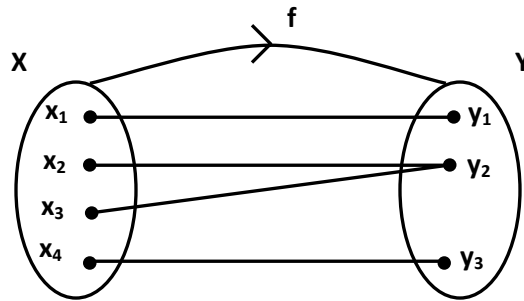
    (d) Measurement of blood group comes under ordinal scale of measurement.

    (e) Pie diagram/chart is used when the change in the values of the components of one variable is important.

2 **(a)** A town has total population of 50,000. Out of it 28,000 read Hindustan times and 23,000 read Times of India while 4,000 read both the papers. Indicate how many read neither Hindustan times nor Times of India?

   (b) Find the number of terms required to make the sum equal to zero for the following series 80, 75, 70, ...

   (c) Find A G.P. whose 3rd and 6th terms are 1 and $-\dfrac{1}{8}$ respectively. **(4+3+3)**

3 **(a)** The question paper of MST-001 contains 5 questions. Question no. 1 has no choice but question no. 2 to 5 each having an internal choice. In how many ways can an examinee answer all the 5 questions?

   (b) In how many different ways can 8 players be arranged in a line so that the best and worst players are never together? **(5+5)**

4. **(a)** Prove that $\lim\limits_{x \to 0} \dfrac{\sqrt{2+x^2} - \sqrt{2-x^2}}{x^2} = \dfrac{1}{\sqrt{2}}$

(b) The cost C of manufacturing a certain article is given by the formula

$$C = 5 + \frac{48}{x} + 3x^2$$

Find minimum value of C. **(5+5)**

5. Evaluate the following integrals:

i) $\int \frac{3x}{(x^2+k^2)^5} \, dx$

ii) $\int_0^6 f(x)dx$ where $f(x) = \begin{cases} x^2 + 3, \ 0 \le x < 3 \\ 2\sqrt{x}, \ \ 3 \le x < 4 \\ e^{-2x} - x^{3/2}, \ 4 \le x \le 6 \end{cases}$ **(4+6)**

6. (a) Show that

$$\begin{vmatrix} x-y & 1 & x \\ y-z & 1 & y \\ z-x & 1 & z \end{vmatrix} = \begin{vmatrix} x & 1 & y \\ y & 1 & z \\ z & 1 & x \end{vmatrix}$$

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

Player	102, 61, 82, 88, 90, 63, 69, 85, 105, 93, 65, 94, 107, 97, 67

7. (a) The daily cost of operating a hospital C is a linear function of the number of in-patients I, out-patients P and a fixed cost **a**, i.e.,

$$C = a + bP + dI$$

Given the following data from 3 days, find the value of a, b and d by setting up a linear system of equations and using the matrix method.

Day	Cost in Rs.	No. of In-patients (I)	No. of Out-patients (P)
1	6,950	40	10
2	6,725	35	9
3	7,100	40	12

(Hint: 40d + 10b + a = 6950) **(10)**

(b) A researcher collected the following data about different points of households. Determine whether the data thus obtained come under nominal, ordinal, interval or ratio scale.

     i)   The caste of the family residing in a household

     ii)   Number of members in a household

     iii)   Age of the oldest person of a household

     iv)   Highest education of the family members of a household

v)   Monthly income of a household

vi)   Number of mobile phones in a household

vii)   Length of the longest long-distance call made in a month

viii) Whether there is a landline telephone in a household or not

ix)   Whether there is a high-speed internet connection in the household

x)   Monthly mobile bill of a household

(10×1=10)

8. The following table shows the life (in months) of a sample of eighty 7-watt LED bulbs produced by a manufacturer:

Life Time of LED				
68	97	20	33	21
31	35	48	52	52
59	60	08	70	36
34	83	05	09	11
22	24	26	26	38
39	43	46	54	41
72	77	74	05	14
16	41	52	80	13
19	36	88	27	03
07	12	18	62	43
52	59	62	16	52
94	04	05	07	15
16	12	20	22	34
38	22	70	77	12
10	20	13	13	16
15	15	74	18	30

i)   Form a frequency distribution by taking suitable width,

ii) Construct histogram and stem-and-leaf diagram,

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

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

(5+6+6+3)

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: (5×2=10)

   a) The SD of 10 values of a data is 50. If each value is divided by 5, then the New SD will be 10.

   b) The values of regression on coefficient b_{xy} and b_{xy} are $\dfrac{1}{10}$ and 20, respectively.

   c) If $\dfrac{(A)}{N} = x, \dfrac{(B)}{N} = 2x$ and $\dfrac{(C)}{N} = 3x$ and $\dfrac{(AB)}{N} = \dfrac{(BC)}{N} = \dfrac{(AC)}{N} = y$, then neither x nor y exceeds 1/4, if the data is consistent.

   d) If Y is proportional to X, then $r(X, Y) = 1$.

   e) The mean weight of 150 students in a class is 60 kg. The mean weight of boys in the class is 70 kg and that of girls is 55 kg. Then the ratio of boys and girls in the class is 3:2.

2. a) A number of particular articles has been classified according to their weights. After drying for two weeks the same articles have again been weighted and similarly classified. It is known that median weight in the First Weighing was 20.83 gm, while in the Second Weighing it was 17.35 gm. Some frequencies a and b in the First Weighing and x and y in the Second Weighing are missing. It is known that $a = \dfrac{1}{3}x$ and $b = \dfrac{1}{2}y$. Find out the values of the missing frequencies.

Class:	0-5	5-10	10-15	15-20	20-25	25-30
Ist Weighing:	a	b	11	52	75	22
IInd Weighing:	x	y	40	50	30	28

   (06)

   b) An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following results:

	Firm A	Firm B
Number of Workers	500	600
Average daily wage	Rs. 186	Rs. 175
Variance of Distribution of wages	81	100

   i) Which firm A or B has a larger wage bill?
   ii) In which firm, A or B, is there greater variability in individual wages?
   iii) Calculate (a) the average daily wage, and (b) the variance of the distribution of wages of all the workers in the firms A and B taken together.

   (08)

3. For the following frequency distribution of scores in Statistics of 50 students selected at random from among those appearing in a certain examination, compute the first four moments about mean:

Score:	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140
Frequency:	1	0	0	1	1	2	1	0	4
Score	140-150	150-160	160-170	170-180	180-190	190-200	200-210	210-220	220-230
Frequency:	4	2	5	10	11	4	1	1	2

Find also the corrected values of the moments after Sheppard's corrections are applied. Also obtain moment coefficients of skewness and kurtosis and comment on the nature of the distribution. (10)

4. a) Calculate Karl Pearson's coefficient of correlation between X and Y for the following data:

$$N = 12, \sum X = 120, \sum Y = 130, \sum (X-8)^2 = 50, \sum (y-10)^2 = 200 \text{ and}$$
$$\sum (X-8)(Y-10) = 50.$$

(07)

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

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

(07)

5. The following table gives frequency distribution of 45 clerks in a business office according to age and pay:

Pay Age	600-700	700-800	800-900	900-1000	1000-1100	Total
20-30	4	3	1	–	–	8
30-40	2	5	2	1	–	10
40-50	1	2	3	2	1	9
50-60	–	1	3	5	2	11
60-70	–	–	1	1	5	7
Total	7	11	10	9	8	45

Calculate the correlation between Age and Pay for the above bivariate data. (10)

6. In a statistical study related to the price (in of two shares X and Y, the following two regression lines were found:

Y = 4X – 35 and 9X = Y + 135

The standard deviation of the price of Share Y is Rs12. You are required to calculate:

i) Average Price of shares X and Y
ii) Standard deviation of the price of Share X and
iii) Coefficient of correlation between the price of Shares X and Y.

(10)

7. (a) In a trivariate distribution:

$\sigma_1 = 2$, $\sigma_2 = \sigma_3 = 3$, $r_{12} = 0.7$, $r_{23} = r_{31} = 0.5$

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

(b) 800 candidates of both sexes appeared at an examination. The boys outnumbered the girls by 15% of the total. The number of candidates who passed exceeds the number failed by 480. Equal number of boys and girls failed in examination. Prepare a 2×2 table and find the coefficient of association. Also comment. (06)

8. (a) Among the adult population of a certain town 50 percent are males, 60 percent are wage earners and 50 percent are 45 years of age or over, 10 percent of the males are not wage-earners, and 40 per cent of the males are under 45 years. Make the best possible inference about the limits within which the percentage of persons (male or female) of 45 years or over are wage earners. (06)

(b) The following table gives the IQ values of 1000 students at a college according to the economic conditions of their parents:

Economic Condition	IQ Level		Total
	High	Low	
Rich	460	140	600
Poor	240	160	400
Total	700	300	1000

Use the coefficient of contingency to determine the amount of Association between economic condition of parents and IQ level of students. (10)

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

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

   (a) If odds in favour of an event A are 2:5 then probability of happening of the event A will be 2/7.

   (b) If two independent events A and B are such that $P(A) = 2/3$ and $P(B) = 1/3$, then events A and $\overline{B}$ are independent.

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

X	0	1	2
P[X=x]	$\dfrac{1}{4}$	$\dfrac{1}{2}$	**a**

   then value of **a** will be 1.

   (d) An oil company conducts a geological study that indicates an exploratory oil well should have a 20% chance of striking oil. To find the probability that the third strike comes on the seventh well drilled, we use geometric distribution?

   (e) Suppose that you spin the dial shown in the figure so that it comes to rest at a random position.
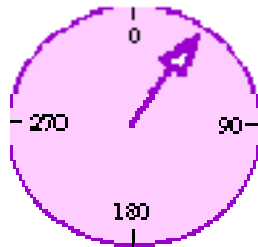
   The probability that the dial will land somewhere between 0° and 90° will be 1/4.

2. The distribution function F of a continuous variable is given by

   $$F(x) = 0, \quad \text{if } x < 0$$

   $$= x^2, \text{ if } 0 \le x \le \frac{1}{2}$$

   $$= 1 - \frac{3(3-x)^2}{25}, \text{ if } \frac{1}{2} \le x < 3$$

   $$= 1, \quad \text{if } x \ge 3$$

   (i) Find the pdf of X,

(ii) Evaluate $P[|X| \leq 1/2]$, using distribution function F,

(iii) Evaluate $P[|X| \leq 1/2]$, using pdf. **(2+4+4)**

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

$$f(x, y) = 4xye^{-(x^2+y^2)}, x \geq 0, y \geq 0$$

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

(b) A bag contains 5 balls which are numbered from 1 to 5. Two balls are drawn without replacement from this. Find the expectation of the sum of the numbers on these balls. **(4)**

4 **(a)** Ms. X thinks that there is a 30% chance that her company will set up a branch office in New Delhi. If it is so, then she is 60% certain that she will be assigned the responsibilities of a branch manager there. What is the probability that Ms. X will be branch manager at the New Delhi office? **(5)**

(b) A smoke-detector system uses two devices A and B. If smoke is present the probability that it will be detected by device A is 0.95; by device B, 0.98 and by both devices, 0.94. If smoke is present find the probability that it will be detected. **(5)**

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

6. **(a)** The probability that a newly designed electronic component will survive a given shock test is 5/6. Find the probability that exactly 3 of the next 4 components tested survive. **(5)**

(b) The probability that a person living in a certain city owns a dog is estimated to be 0.3. Find the probability that the tenth person randomly interviewed in that city is the fifth one to own a dog. **(5)**

7 **(a)** The metro trains on a certain section run every 10 minutes between 5 a.m. to 10 p.m. What is the probability that a commuter entering the station at a random time during this period will have to wait

(i) at least five minutes?

(ii) zero minute?

(iii) 2 to 5 minutes? **(10)**

(b) Three technicians X, Y and Z service respectively 20%, 30% and 50% breakdowns when occur on an automated production line. The technician X makes an incomplete repair 1 time in 20, Y makes an incomplete repair 1 time in 10 and Z makes an incomplete 1 time in 15. For the next breakdown a repair made was found to be incomplete, find the probability that this repair was made by technician

(i) X,

(ii) Y, and

(iii) Z **(10)**

8.(a) The daily consumption of milk in a city, in excess of 20,000 litres, is approximately distributed as a gamma variate with parameters $r = 2$ and $\lambda = 1/10,000$. The city has a daily stock of 30,000 litres. What is the probability that the stock is insufficient on a particular day? **(5)**

(b) A company has installed 10,000 electric lamps in a metro. Assume that these lamps have an average life of 1,000 burning hours with a S.D. of 200 hours. Assuming normality, what number of lamps might be expected to fail.

i) in the first 800 burning hours.

ii) between 800 and 1200 burning hours.

iii) more than 1000 burning hours **(4+6+5)**

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: **(5×2=10)**

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

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

   then degrees of freedom of the distribution will be (2,2).

   (b) The number of runs in the sequence HTTTHTHTTTHT will be 6.

   (c) A patient suffering from fever reaches to a doctor and suppose the doctor formulate the hypotheses as

   H_0 : The patient is a chikunguniya patient

   H_1 : The patient is not a chikunguniya patient

   If the doctor rejects H_0 when the patient is actually a chikunguniya patient, then the doctor commits type II error.

   (d) In a random sample of 525 families owning television set in the region of New Delhi, it is found that 370 subscribe to Star Plus. A 95% confidence interval for the actual proportion of such families in New Delhi which subscribe to Star Plus will be (0.66, 0.74).

   (e) For testing the pulse rate of one month old child is to be 120 per minute, a researcher measures (per minute) the pulse rate of 10 one month old children and applies sign test and Wilcoxon signed-rank test, then power of the Wilcoxon signed-rank test is more than the sign test.

2. The lifespan (in '000 hours) of five LED bulbs of 07 watts are as follows:

   40, 46, 48, 50, 42

   (i) How many samples of size 2 are possible without replacement? Write them.

   (ii) Compute mean of all samples of size 2 and set up the sampling distribution of the sample mean.

   (iii) Compute the expected value and standard error of the sample mean. **(2+4+4)**

3(a) An ambulance agency claims that the mean length of service times is 10 minutes with SD 3 minutes. Investigator suspects that this claim is wrong. She takes a random sample of 20 services and finds the mean length of service times is 15 minutes. Assume that the service time of the ambulance follows normal distribution.

   i) What is the probability that the mean length of service times is more than and equal to 10 minutes.

ii) Find the 95% confidence interval for the standard deviation of the mean length of service times. **(3+3)**

(b) A sample of 500 shops was selected in a large metropolitan area to determine various information concerning the consumer behaviour. One question, among the questions, asked, was "Do you enjoy shopping for clothing?" Out of 240 males,136 answered yes. Out of 260 females, 224 answered yes. Find 95% confidence interval for the difference of the proportions for enjoys shopping for clothing. **(4)**

4. An engineer conducted an experiment to compare two metals: iron and copper, 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 breaking pressures are given in the following table:

| S. No. | Breaking Pressure | |
	Iron	Copper
1	72.7	73
2	69.6	67.2
3	83.4	75.3
4	78.9	61.4
5	75	74
6	71.6	69.5
7	85.7	69.8
8	73.5	73.8
9	70.4	68
10	84.2	76.1

If the breaking pressures for both iron and copper are normally distributed, are the variances of the distributions of the breaking pressure of iron and copper equal at 5 % level of significance? **(10)**

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

S. No.	Test For	Name of the Test	Test Statistic	Assumptions for Applying the Test	Test Type
1	Population mean when population variance is known and population is normal	Z-test	$Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	1. Sample observations should be independent. 2. The measurement scale should be at least interval scale.	Parametric Test
2	Population mean when population variance is unknown and population is normal				
3	Population mean/median when the form of population is not known and sample size is small				
4	Population mean/median when the form of population is not known and the data is in ordinal scale and sample				

	size is small				
5	Population variance when the population is normal distributed and sample size is small				
6	Independence of Two Attributes				

(10)

6. To study whether or not the level of earning is affected by educational attainment, a social scientist randomly selected 100 people from each of three income categories 'low', middle', 'high' and then recorded their educational attainment as in the following table:

Educational attainment	Income Categories			Total
	Low	Middle	High	
No college	32	20	23	75
UG	13	16	1	30
PG	43	51	60	154
Doctoral	12	13	15	41
Total	100	100	100	300

Do these data indicate that the level of earning is affected by educational attainment? Test at $\alpha = 0.01$. **(10)**

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

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

then find

i)   maximum likelihood estimator of the parameter θ,

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

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

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

(8+2+10)

8.   Results of a survey by the National Centre for Health Statistics indicated that people between 65 and 74 years of age contact a physician an average of 9.8 times per year. People 75 and older contact physicians an average of 12.9 times per year. Suppose you want to validate these results by taking your own samples and obtained the following data represent the number of annual contacts people make with a physician. The samples are independent.

Number of annual contacts people make with a physician	
65 to 74	**75 and Older**
12	16
13	15

8	10
11	17
9	13
6	12
11	14
	9
	13

i) Formulate the hypotheses that average number of contacts with physician by people 75 and older is greater than the number by people 65 to 74 years old.

ii) What assumptions are necessary to apply the t-test.

iii) Apply the t-test for testing the hypotheses formulated in (i) using $\alpha = 0.05$.

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

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

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 the SSE in a two way ANOVA having 4 rows and 5 columns is given as 48, then the MSSE will be 4.

   b) If there is one missing value in RBD with 3 blocks and 4 treatments, then the total of degrees of freedom will be 10.

   c) In a sequence of random numbers, generated through LCG" $x_i = (5x_{(i-1)} + 3)$ Mod(16), with $x_0 = 5$" the x_1, x_2 and x_3 will be 11, 14, 13.

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

   e) In 2^3 factorial design, there 2 factors each at 3 levels.

2. Below are given the weights of 5 students:

S. No	1	2	3	4	5
Weights (Kg)	56	49	66	64	55

   a) Calculate Population Mean, Population mean square and Population Variance.
   b) Enumerate all possible samples of size 2 without replacement and show that
      i) Sample mean $(\bar{y})$ is unbiased for Population mean $\bar{Y}$
      ii) Sample mean square (s^2) is unbiased for the Population Mean Square (S^2)
      iii) $\text{Var}(\bar{y}) = \dfrac{N-n}{N} \cdot \dfrac{S^2}{n}$ (10)

3. Given the following population, allocate a sample of size n = 10 by optimum allocation:

Stratum I :	2	4	4	6	6	8
Stratum II :	8	8	12	16	16	18
Stratum III :	16	16	18	22	22	26

   (6)

4. Suppose from a total of 20 bearing trees of Guava in a village, 5 clusters of size 4 trees each were selected and Yield (in Kgs) is a given in the following table:

Cluster	1st Tree	2nd Tree	3rd Tree	4th Tree
1	5	4	1	15
2	11	1	4	7
3	36	10	19	11
4	7	15	12	10
5	2	22	8	6

a) i) $Y_i^s (i = 1,2,...,5)$ cluster totals

   ii) $\overline{Y}_i^s (i = 1,2,...,-5)$ cluster Means

   iii) $\overline{Y}_C$ Population Mean per Cluster and

   iv) $\overline{Y}$, i.e., Population Mean Per Unit

b) Select a sample of 3 cluster and calculate the sample cluster mean $\overline{Y}_c$ (08)

5. The following data represent the number of units of production per day turned out by 5 different workers using four different types of machines:

Workers	Machine Type			
	A	B	C	D
1	54	48	57	46
2	56	50	62	53
3	44	46	54	42
4	53	48	56	43
5	48	52	59	49

a) Test whether the mean productivity is the same for the different machine types.

b) Test whether the 5 workers differ with respect to mean productivity.   (12)

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

Blocks	Heights of Plants (cms)					
I	(1) 24.7	(3) 27.7	(2) 20.6	(4) 16.2	(5) 16.2	(6) 24.9
II	(3) 22.7	(2) 28.8	(1) 27.3	(4) 15	(6) 22.5	(5) 17
III	(6) 26.3	(4) 19.6	(1) 38.5	(3) 36.8	(2) 39.5	(5) 15.4
IV	(5) 17.7	(2) 31	(1) 28.5	(4) 14.1	(3) 34.9	(6) 22.6

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

7(a) Below is given the plan and yields of 2^2 factorial experiment involving 2 factors A and B each at two levels 0 and 1:

Block I	(1) 117	b 106	ab 129	a 124
Block II	ab 124	(1) 120	b 117	a 124
Block III	(1) 111	a 127	b 114	ab 126
Block IV	ab 125	a 131	b 112	(1) 108
Block V	ab 95	b 97	(1) 73	a 128
Block VI	a 158	(1) 81	ab 125	b 117

Analyse the design. Does treatment effects A differ from treatment effect B significantly? (14)

(b) An Investigator is interested in finding the level of knowledge about the history of India of 4 different schools in a city. A test is given to the 5, 6, 7, 6 students of 8th class of 4 schools, respectively. Their scores out of 10 are given as below:

School I:	8	6	7	5	9		
School II:	6	4	6	5	6	7	
School III:	6	5	5	6	7	8	5
School IV:	5	6	6	7	6	7	

Test the equality of Average Scores of the 4 Schools at 5% level of significance.   (08)

8 a) Generate a complete cycle for the LCG given $x_i = (12x_{i-1} + 6)$ Mod 11 with $x_0 = 06$. A man tosses an unbiased coin 10 times. Using the first 10 random numbers generated above, obtain a sequence of heads and tails.   (10)

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

251	769	153	575	390	335	066	104	200	019
597	729	012	922	691	817	064	539	419	305
449	998	919	470	372	851	643	509	913	445
464	447	279	413	494	972	983	432	368	574

Apply chi-square test to test the fit of the distribution   (12)

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: **(5×2=10)**

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

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

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

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

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

		Player B				
		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

   Then second column will be dominated by the first column.

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

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

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

   (b) A process of manufacturing an item is in statistical control with $\overline{\overline{X}} = 250$ and $\overline{R} = 3.5$. The specification limits of the process are 200 ± 81.90. Calculate the process capability. Does it appear that the manufacturing process is capable of meeting the specification requirements? $(d_2 = 2.059)$ **(3)**

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

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

4. An office supply company ordered a lot of 4,000 pens. When the lot arrives the company inspector will randomly inspect 15 pens. If more than three pens in the sample are non-conforming, the lot will be rejected. If fewer than two pens are non-conforming, the lot will be accepted. If two, three pens are non-conforming, a second sample of size 10 will be taken. The value of c_2 is 3. Suppose the inspector finds two nonconforming items in the first sample and two in the second sample. Also AQL and LTPD are 0.05 and 0.10 respectively. Let incoming quality be 4%.

    i) What type of acceptance plan is used here?

    ii) What is the probability of accepting the lot at the first sample?

    iii) What is the probability of accepting the lot at the second sample?   **(2+3+5)**

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

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

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

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

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

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

7. (a) The system shown below is made up of ten components. Components 2, 3 and 4 are not identical and at least two component of this group must be available for system success.
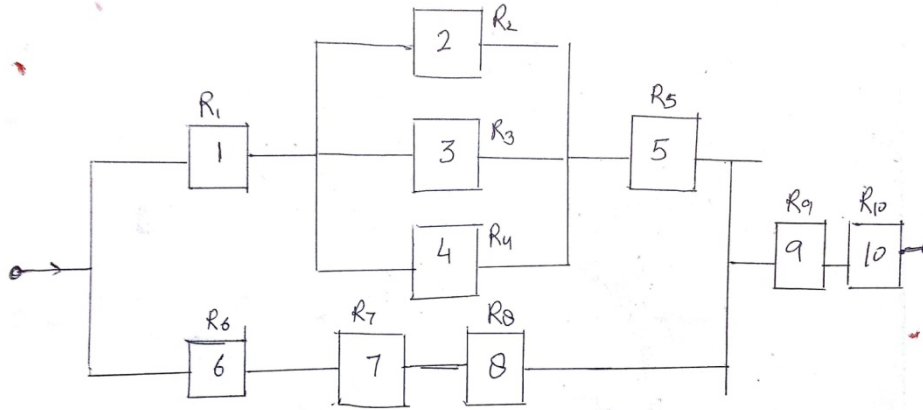
What is the system reliability if $R_1 = R_5 = 0.85$, $R_6 = R_7 = R_8 = 0.80$, $R_9 = R_{10} = 0.70$ $R_2 = 0.40$, $R_3 = 0.50$ and $R_4 = 0.60$. **(10)**

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

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

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

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

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

i)   Estimate the process mean.

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

iii) Plot the charts on the graph papers.

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

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

TUTOR MARKED ASSIGNMENT
MSTE-002: Industrial Statistics-II

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

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

    i)   The number of possible basic solutions of an LPP of order 2 × 3 is 5.

    ii) If there are 3 jobs to be performed one at a time on each of the 2 machines, then the possible number of sequences would be 36.

    iii) If the arrival rate is 6 per hour and service rate is 3 per hour then the probability of no customer in queue is 0.7

    iv) The estimated value of 'b' in the regression line Y = a + bX + e in terms of variance-covariance is $\left[\mathrm{cov}(X,Y)/\sigma_X^2.\sigma_Y^2\right]$.

    v)   Given the trend equation, $\hat{Y} = 108 + 2.88X$ with 1980 as origin and yearly data from 1980 to 1982, the estimated value for 1985 is 119.52.

2. A private company has two bottling plants located in city I and City II. Each plant produces three cold drinks Drink A, Drink B and Drink C. The number of bottles produced per day are as follows:

Drink	Plant at	
	City I	City II
Drink A	1,500	1,500
Drink B	3,000	1,000
Drink C	2,000	5,000

    A market survey indicates that during the month of May, there will be a demand of 20,000 bottles of Drink A, 40,000 bottles of Drink B and 44000 bottles of Drink C The operating cost per day for plants at city I and City II are Rs 600 and Rs 400. For how many days should each plant be run in May so to minimise the production cost, while till meeting the market demand? Solve graphically. (6)

3. A company makes two kinds of leather belts. Belt A is a high quality belt and Belt B is of lower quality. The respective profits are Rs 4 and Rs 3 per belt. The production of each of type A requires twice as much time as a belt of type B, and if all belts were of type B, the company could make 1000 per day. The supply of leather is sufficient for only 800 belts per day (Both A and B combined). Belt A requires a fancy buckle and only 400 per day are available. There are only 700 buckles a day available for Belt B. What should be the daily production of each type of belt? Formulate this problem as an LP Model and solve it using Simplex Method. (8)

4. (a) A department has five employees with five jobs to be performed. The time (in hours) each man will take to perform each job is given in the effectiveness matrix:

Jobs	I	II	III	IV	V
Employees					
A	10	5	13	15	16
B	3	9	18	13	6
C	10	7	2	2	2
D	7	11	9	7	12
E	7	9	10	4	12

How should the Job be allocated, one per employee, so as to minimise the total man-hours?

(08)

(b) Find the sequence that minimizes the total elapsed time required to complete the following tasks on two machines:

Tasks:	A	B	C	D	E	F	G	H	I
Machine I:	2	5	4	9	6	8	7	5	4
Machine II:	6	8	7	4	3	9	3	8	11

(06)

5. A company has factories at F_1, F_2 and F_3 which supply to Ware houses at W_1, W_2 and W_3. Weekly factory capacities are 200, 160 and 90 units, respectively. Weekly warehouse requirements are 180, 120 and 150 units, respectively. Unit shipping costs (in Rs) are as follows:

Factory	Warehouse			Capacity
	W_1	W_2	W_3	
F_1	16	20	12	200
F_2	14	8	18	160
F_3	26	24	16	90
Demand	180	120	150	450

Obtain the Initial basic feasible solution using North-west Corner Rule and also determine the optimal distribution for this company to minimize total shipping cost.

(10)

6. a) The production department for a company requires 3600 kg of raw material for manufacturing a particular item per year. It has been estimated that the cost of placing an order is Rs 36 and the cost of carrying inventory is 25 percent of the investment in the inventories. The price is Rs10 per kg. The purchase manager wishes to determine an ordering policy for raw material.

b) Arrivals at telephone booth are considered to be Poisson with an average time of 10 minutes between are arrival and the next. The length of phone call is assumed to be distributed exponentially with mean 3 minutes.

i) What is the probability that a person arriving at the booth will have to wait?

ii) The telephone department will install a second booth when convinced that an arrival would expect waiting for at least 3 minutes for phone call. By how much should the flow of arrivals increases in order to justify a second booth?

iii) What is the average length of the queen that forms from time to time?

iv) What is the probability that it will take him or her more than 10 minutes altogether to wait for the phone and complete his or her call?

(4+8)

7. (a) The annual sales revenue (in lakhs of Rs) of a product as a function of sales force (number of salesmen) and annual adverting expenditure (in Thousands of Rs) for the past 10 years are summarised in following table:

Annual Sales Revenue Y(in Lakhs)	Sales Force X_1 (in Number)	Annual Advertising Expenditure X_2 (in Thousand)
100	40	140
115	65	115
125	40	190
135	90	80
105	115	100
145	80	140
110	50	115
120	60	150
135	70	130
175	100	160

Obtain a regression model to forecast the annual sales revenue of the product using Matrix Method. (12)

(b) A statistician collected data of 78 values with two independent variable X_1 and X_2. The four models considered are: (a) $Y = B_0 + e$, (b) $Y = B_0 + B_1X_1 + e$ (c) $Y = B_0 + B_1X_1 + B_2X_2 + e$ and (d) $Y = B_0 + B_2X_2 + e$.
The results are
$SS(B_0) = 6524.2$, $SS(B_0, B_1) = 6793.4$, $SS(B_0, B_2) = 6540.0$, $SS(B_0, B_1, B_2) = 6877.9$ and $\hat{\sigma}^2 = 9.1$. Find 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. (08)

8. (a) The demand (in thousands of units) of refrigerators manufactured by a company are summarised in the following table:

Year	Quarter			
	I	II	III	IV
2011	124	120	176	140
2012	104	128	148	132
2013	80	152	168	156
2014	116	156	184	180

Obtain the seasonal Indices by moving average method and also obtain the deseasonalized values. (14)

(b) For the following Auto regressive model

$$X_t = 0.7X_{t-1} - 0.4X_{t-2} + a_t$$

i)   Verify whether the series is Stationary
ii) Obtain $\rho_k : k = 1, 2, 3, 4 \text{ and } 5$
iii) Plot the Correlogram. (06)

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 hypotheses, interpretations, etc.**

Q 1. A packaging process is supposed to fill small boxes of raisins and claimed to fill the boxes with approximately 50 raisins. However, the number of raisins in each box will vary. Suppose 100 boxes of raisins are randomly sampled and the number of raisins in each box is counted. The data are given in the following table:

57	51	53	52	50	60	51	51	52	52
44	53	45	57	39	53	58	47	51	48
49	49	44	54	46	52	55	54	47	53
49	52	49	54	57	52	52	53	49	47
51	48	55	53	55	47	53	43	48	46
54	46	51	48	53	56	48	47	49	57
55	53	50	47	57	49	43	58	52	44
46	59	57	47	61	60	49	53	41	48
59	53	45	45	56	40	46	49	50	57
47	52	48	50	45	56	47	47	48	46

Answer the following:

i) Compute the skewness and kurtosis for the given data and interpret the results.

ii) Compute suitable width of the class interval and construct continuous frequency distribution for the given data.

iii) Also, construct histogram. What does the histogram reveal about the filled boxes?

iv) If the number of raisins in each box is normally distributed, then check whether there is a variation in the number of raisins in boxes at 5 % level of significance?

(6+6+5+8)

Q 2(i). In order to measure the effect of a storewide sales campaign on non-saleable items, the research director of a national supermarket chain took a random sample of 13 pairs of stores which were matched according to average weekly sales volume. One store of each

pair (the experimental group) was exposed to the sales campaign, while the other member of the pair (the control group) was not. The following data indicate the results over a weekly period:

Store	Sales (in Rs. '000)	
	With sales campaign	Without sales campaign
1	67.2	65.3
2	59.4	54.7
3	80.1	81.3
4	47.6	39.8
5	97.8	92.5
6	38.4	37.9
7	57.3	52.4
8	75.2	69.9
9	94.7	89
10	64.3	58.4
11	31.7	33
12	49.3	41.7
13	54	53.6

a) What assumption is necessary to perform the parametric test to test that there is evidence that the sales campaign has increased the sales of the items?

b) Perform the test at the 0.05 level of significance

(ii)   Before increasing in the excise duty on tea, 800 people out of a sample of 1000 were the consumers of tea. After increasing in the excise duty, 800 people were the consumers of tea in a sample of 1200 persons. Check whether there is a significant decrease in the consumption of tea after increasing the excise duty.

(2+13+10)

Q 3(i). The ranks given to twenty branches of a bank on the basis of their deposit mobilisation and recovery of loans, are shown in the following table:

Deposit Mobilisation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Loan Recovery	4	2	9	1	7	10	8	13	5	3	11	6	12	15	14	12	12	14	15	15

Compute the rank correlation coefficient and interpret the result.

(10)

(ii)   A scientist wants to check the yield of wheat from different varieties of wheat. In this regards, the yields of wheat from 16 plots, all of approximately equal fertility, when 4

varieties of wheat (Say, A, B, C and D) were cultivated are recorded in the following table:

Plot No.	1	2	3	4	5	6	7	8
Variety:	A	B	D	C	B	C	A	D
Yield:	32	34	29	31	33	34	34	26
Plot No.:	9	10	11	12	13	14	15	16
Variety:	B	D	A	C	B	A	B	C
Yield:	36	30	33	35	37	35	35	32

Assuming that the yields of wheat are normally distributed in each variety with approximately equal variances, test the hypothesis that the average yields of wheat in all four varieties of wheat are the same at 1% level of significance.

(15)

Q 4. A researcher wanted to study the effects of different types of cooling fans on the temperature of the different types of computers. The following table gives the observation on temperature of a computer chip when five different types of cooling fans were tried on each of the six different computers:

Cooling Fan	Computer					
	A	B	C	D	E	F
I	26	18	23	12	21	15
II	26	22	28	21	28	22
III	24	19	22	21	24	22
IV	24	21	23	18	19	20
V	25	19	24	15	22	18

Assuming that the effects of each cooling fan and computer are normally distributed with approximately equal variances, analyse the design at 5% level of significance. Test whether the effect on the temperature due to different types of (i) cooling fans and (ii) computers are significant or not at 5% level of significant. If significant, do the pair-wise comparison between them.

(25)

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 hypotheses, interpretations, etc.**

Q 1. A manufacturing company produces bearings. The diameter specified for the bearings is 5 millimeters. Every 30 minutes, twelve bearings are sampled and their diameters are measured and recorded. In this way, forty samples of twelve bearings are selected and the data are displayed in the following table:

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
5.01	4.96	5.01	5.02	4.98	4.98
4.92	4.98	4.87	5.01	5.08	5.02
5.01	4.95	5.02	4.99	5.09	4.97
4.88	4.96	5.01	5.02	4.89	4.99
5.05	5.01	5.02	5.03	5.06	4.98
4.97	4.89	5.04	5.01	5.02	4.92
5.03	4.96	5.01	5.02	4.96	4.98
4.92	4.98	4.97	5.02	5.08	5.02
5.01	4.95	5.02	4.99	4.82	4.97
4.88	4.96	5.02	5.02	5.01	5.00
5.05	5.01	5.12	5.03	5.06	4.98
4.97	5.01	5.04	5.01	5.20	4.99

Sample 7	Sample 8	Sample 9	Sample 10	Sample 11	Sample 12
4.99	4.96	4.96	5.03	4.97	5.00
5.08	5.01	5.00	4.99	4.91	4.96
5.00	5.02	4.91	4.96	5.02	5.05
5.02	5.05	4.87	4.99	4.93	4.98
5.01	5.04	4.96	5.11	4.95	5.06
5.01	5.02	5.01	5.04	4.96	5.01
4.99	4.96	4.96	5.03	4.97	5.02
5.00	5.01	5.00	4.99	5.02	4.96
5.00	5.02	4.91	4.96	5.02	5.05
5.02	5.05	5.04	4.98	4.93	4.94
5.01	5.04	4.96	4.97	4.95	5.06

5.01	5.02	5.01	5.04	4.96	5.01
Sample 13	**Sample 14**	**Sample 15**	**Sample 16**	**Sample 17**	**Sample 18**
4.96	4.99	4.96	4.96	4.96	5.04
4.99	4.97	4.99	4.99	4.93	5.03
4.92	5.01	4.92	4.92	4.97	4.97
5.03	4.98	5.03	5.03	5.01	4.99
5.00	4.96	5.00	5.00	4.98	5.05
4.96	5.02	4.96	4.96	4.92	5.06
4.96	4.99	4.96	4.96	4.96	5.04
4.99	4.97	4.99	4.99	4.93	5.03
4.82	5.01	4.82	4.82	4.97	4.97
5.03	4.98	5.03	5.03	5.01	4.99
5.00	4.96	5.00	5.00	4.98	5.05
4.96	5.02	4.96	4.96	4.92	4.90
Sample 19	**Sample 20**	**Sample 21**	**Sample 22**	**Sample 23**	**Sample 24**
5.04	5.03	5.02	5.11	5.02	5.02
5.03	4.92	4.98	4.97	4.98	4.98
4.97	4.91	4.99	4.92	4.99	4.99
4.99	4.98	5.03	4.98	5.03	5.03
5.05	4.95	4.96	5.02	4.96	4.96
5.06	4.90	5.03	4.94	5.03	5.03
5.04	5.03	5.02	5.11	5.02	5.02
5.03	4.99	4.98	4.99	4.98	4.98
4.97	4.91	4.99	4.92	4.99	4.99
4.99	4.98	5.03	4.98	5.03	5.03
5.05	4.95	4.96	5.02	4.96	4.96
4.90	5.02	5.03	4.94	5.03	5.03
Sample 25	**Sample 26**	**Sample 27**	**Sample 28**	**Sample 29**	**Sample 30**
4.98	4.96	4.96	5.03	4.97	5.09
5.02	5.01	5.00	4.99	4.91	4.96
4.92	5.02	4.91	4.96	5.02	5.05
4.99	5.05	4.87	4.87	4.93	4.99
4.98	5.04	4.96	4.98	4.95	5.06
4.99	5.02	5.01	5.04	4.96	5.01
4.98	4.96	4.96	5.03	4.97	5.09
5.02	5.01	5.00	4.99	4.91	4.96
4.97	5.02	4.91	4.96	5.02	5.05
4.99	5.05	5.02	5.10	4.93	4.91
4.98	5.04	4.96	5.11	4.95	5.06
4.99	5.02	5.01	5.04	4.96	5.01
Sample 31	**Sample 32**	**Sample 33**	**Sample 34**	**Sample 35**	**Sample 36**

4.96	4.99	5.01	5.05	4.96	4.90
4.99	4.97	5.04	4.97	4.93	4.95
4.82	5.01	4.98	5.04	4.97	5.04
5.03	4.98	5.07	5.03	5.01	5.01
5.00	4.96	4.96	5.09	4.98	4.88
4.96	5.02	4.91	5.01	4.92	5.09
4.96	4.99	5.01	5.05	4.96	4.90
4.99	4.97	5.04	4.97	4.93	4.85
5.02	5.01	5.09	5.04	4.97	5.17
5.03	4.98	5.07	5.03	5.01	5.12
5.00	4.96	4.92	4.97	4.98	4.88
4.96	5.02	5.13	5.01	4.92	4.86

Sample 37	Sample 38	Sample 39	Sample 40		
5.04	5.05	4.96	5.04		
5.03	4.97	4.93	5.03		
4.97	5.04	4.97	4.97		
4.99	5.03	5.01	4.99		
5.05	5.09	4.98	5.05		
5.06	5.01	4.92	5.06		
5.04	5.05	4.96	5.04		
5.03	4.97	4.93	5.03		
4.97	5.04	4.97	4.97		
4.99	5.03	5.01	4.99		
5.05	4.97	4.98	5.05		
4.90	5.01	4.92	4.90		

Construct the suitable control charts for variability as well as average to infer whether the diameter of the process is under control or not. If it is out-of-control, construct the revised control charts.

(25)

Q 2. A manufacturer produces decorative wooden pieces. Each decorative wooden piece is supposed to meet company standard. To check and control the process of cutting, the company quality people randomly sampled wooden pieces varies in length from every shift and determined the number of defects on the wooden pieces. The data collected from 40 such samples are shown below:

Sample	Length of Wooden Piece (in inches)	Number of Defects
1	15	1
2	15	2
3	17	2
4	15	7
5	15	2
6	14	1
7	15	2
8	16	3

9	15	2
10	15	1
11	16	3
12	18	1
13	16	3
14	16	2
15	15	3
16	14	2
17	15	1
18	16	3
19	15	1
20	15	2
21	15	1
22	16	3
23	15	2
24	15	3
25	14	2
26	12	1
27	13	2
28	12	2
29	12	3
30	12	2
31	13	1
32	13	2
33	12	3
34	15	3
35	13	1
36	12	6
37	12	1
38	14	3
39	13	3
40	12	1

Construct a suitable control chart 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 Personnel Manager of a company that produces an industrial product, is interested in analysing the salaries of workers involved in the production process. To accomplish this, she/he has decided to develop a multiple regression model to predict their weekly salaries. For this purpose, he/she has selected a random sample of 50 workers involved in the production process. The information on their current monthly salaries in hundreds (Y), lengths of employment (X_1), ages (X_2), sex $(X_3; 0$ for female and 1 for male$)$, and job classifications $(X_4; 0$ for technical jobs and 1 for clerical jobs$)$ are summarised in the following table:

Employee	Y	X_1	X_2	X_3	X_4
1	395	69	47	1	0
2	306	46	40	0	0
3	467	125	39	1	0
4	423	20	45	1	0
5	475	173	56	1	0
6	337	37	25	0	0
7	564	237	48	1	0
8	391	52	28	1	0
9	372	67	46	0	0
10	307	124	30	1	0
11	278	12	20	1	0
12	625	313	46	1	0
13	500	291	47	0	0
14	340	34	23	0	0
15	562	275	48	1	0
16	423	111	56	1	0
17	328	14	27	1	0
18	435	89	29	1	0
19	433	188	58	0	0
20	428	44	34	1	0
21	346	21	24	0	0
22	428	35	26	1	0
23	398	46	21	1	0
24	378	43	25	1	0
25	407	27	22	0	0
26	378	19	24	1	0
27	545	229	58	1	0
28	477	276	58	0	0
29	454	330	52	1	0
30	554	331	60	1	0
31	466	72	41	1	0
32	333	85	27	0	0
33	366	84	47	1	0
34	265	25	21	0	1
35	577	220	39	0	1
36	373	31	25	1	1
37	544	300	55	1	1
38	585	311	50	1	1
39	256	6	32	0	1
40	344	18	44	1	1

41	378	89	46	0	1
42	308	76	40	0	1
43	356	53	47	0	1
44	309	17	53	0	1
45	591	354	58	0	1
46	363	64	42	0	1
47	336	88	34	0	1
48	313	11	21	1	1
49	634	407	53	0	1
50	363	64	42	0	1

i) Prepare a scatter plots considering Y and (a) X_1 and X_3, (b) X_1 and X_4 (c) X_2 and X_3 and (d) X_2 and X_4 to get an idea about the relationship among the variables.

ii) Develop the best fitted multiple regression model using stepwise regression procedure.

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

(8+9+8)

Q 4 An agricultural institute is interested in analysing the pattern of the employment in agriculture in a particular state in past years and then forecasting the expected number of employment for the upcoming years. The number of persons (in millions) employed in agriculture on a monthly basis from January 2008 through December 2017. The following monthly data are available for past 10 years:

	Year									
Month	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
January	2.877	2.869	2.955	3.197	2.888	2.853	2.672	2.868	2.762	2.782
February	2.846	2.909	2.956	3.283	2.89	2.802	2.709	2.771	2.796	2.836
March	3.042	3.094	3.131	3.334	2.988	2.897	2.804	2.913	2.925	2.962
April	3.505	3.287	3.295	3.437	3.171	3.273	3.14	3.151	3.074	3.081
May	3.598	3.531	3.467	3.604	3.622	3.415	3.478	3.369	3.309	3.436
June	3.92	3.976	4.053	3.895	3.869	3.78	3.82	3.983	3.785	3.737
July	3.971	4.061	4.165	4.024	4.09	3.931	3.79	3.997	3.857	3.853
August	3.764	4.031	3.826	3.851	3.886	3.842	3.682	3.856	3.795	3.636
September	3.444	3.658	3.436	3.563	3.626	3.396	3.326	3.549	3.545	3.635
October	3.47	3.721	3.525	3.536	3.524	3.447	3.408	3.553	3.467	3.501
November	3.262	3.363	3.419	3.224	3.156	3.081	3.181	3.1	3.257	3.214
December	2.948	3.163	3.202	2.959	2.856	2.85	2.914	2.99	2.995	3.057

a) Compute the seasonal indices using ratio to moving average method.

b) Obtain the deseasonalised values and then fit a linear trend line to the average annual employment using least squares method.

c) Convert the annual least-squared trend equation to a monthly trend equation.
d) Use the monthly trend equation and seasonal indices to forecast the number of persons employed in agriculture for all twelve months of 2018 and 2019.
e) Plot the original data, deseasonalised data and trend values.

(8+6+5+4+2)