ASSIGNMENT BOOKLET

Post Graduate Diploma in Applied Statistics

(Specialisation in Industrial Statistics)

MST-001 to MSTL-002

(Valid from 1ˢᵗ January, 2017 to 31ˢᵗ December, 2017)

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

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, 2017.

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 2017,** you should submit your TMAs by **March 31, 2017.** Similarly, If you wish to appear in the **TEE, December 2017,** you should submit your TMAs by **September 30, 2017.**

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) If $A = \{x : x \text{ is a Prime Minister of India and belongs to Gujarat}\}$ and $B = \{x : x \text{ is a woman Prime Minister of India}\}$, then A and B are equal set.

   (b) If F: $N \rightarrow N$ defined by $f(x) = (2 + x)^x$, $x \in N$, then the function is one-one function.

   (c) The pollution level in Delhi is a discrete variable.

   (d) $\dfrac{dy}{dx}$ of the functions $x = 1 + t^2$, $y = 5 + t$ is 2t.

   (e) Stem-and-leaf display is better than histogram.

2. (a) In a survey conducted on 200 clerks in an office, it was found that 48% preferred coffee , 54% liked tea and 64% used to smoke. Of the total, 28% used coffee and tea, 32% used tea and smoke and 30% preferred coffee and smoke. Only 6% did none of these. Find

   i) the number of clerks having all the three,

   ii) tea and smoke but not coffee, and

   iii) only coffee. **(2+2+2)**

   (b) Show that the sum of all odd numbers between 2 and 1000 which are divisible by 3 is 83,667. **(4)**

3. (a) Three numbers whose sum is 15 are in A.P.. If 1, 4 and 19 are added to them, respectively, the results are in G.P. Find the numbers. **(4)**

   (b) The Prime Minister of India calls a meeting of Chief Ministers of the states whose state boundaries touch Pakistan to discuss on the terrorism. In how many ways can they seat themselves at a round table if Gujarat and Rajasthan Chief Ministers choose to sit together? **(3)**

   (c) There are five routes for journey from Station A to Station B. In how many different ways, can a man go from A to B and return, if for returning

   i) any of the five routes is taken,

   ii) the same route is taken,

   iii) the same route is not taken. **(1+1+1)**

4. (a) Prove that $\lim\limits_{x \to 0} \dfrac{\sqrt{a + x^2} - \sqrt{a - x^2}}{x^2} = \dfrac{1}{\sqrt{a}}$ **(5)**

(b) A company notices that higher sales of a particular item which it produces are achieved by lowering its selling price. As a result the total revenue from the sales at first rises as the number of units sold increases, reaches the highest point and then falls off. This pattern of total revenue is described by the relation

$$y = 40,00,000 - (x - 2000)^2$$

where y is the total revenue and x is the number of units sold.

i) What number of units sold maximises total revenue?

ii) What is the amount of this maximum revenue?

iii) What would be the total revenue if 2500 units were sold? **(3+2+1)**

5. Evaluate the following integrals:

i) $\int \dfrac{1}{x \log x \log (\log x)} \, dx$

ii) $\int_{10}^{20} \dfrac{1}{x - x^3} dx$ **(5+5)**

6. **(a)** Prove that

$$\begin{vmatrix} a + b + 2c & a & b \\ c & b + c + 2a & b \\ c & a & c + a + 2b \end{vmatrix} = 2(a + b + c)^3$$

(5)

(b) A survey of 705 workers asked how much they used the Internet at workplace. The results were as follows:

Use of the Internet at Workplace	%
Too much	5
More than I should	4
Within limits	60
Very little	5
Do not use	26

Which graphical method do you think is best to represent these data? Construct it. **(5)**

7. **(a)** Solve the following equations by matrix method

$$5x - 6y + 4z = 15$$
$$7x + 4y - 3z = 19$$
$$2x + y + 6z = 46$$

(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. If the data come under the interval or ratio scale, determine whether the data are discrete or continuous.

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

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

x) Monthly mobile bill of a household   **(10×1=10)**

8. The ordered arrays in the following table show the life (in days) of a sample of forty 7-watt LED bulbs produced by each Manufacturer A and Manufacturer B.

Manufacturer A					Manufacturer B				
684	697	720	773	821	819	836	888	897	903
831	835	848	852	852	907	912	918	942	943
859	860	868	870	876	952	959	962	986	992
893	899	905	909	911	994	1,004	1,005	1,007	1,015
922	924	926	926	938	1,016	1,072	1,020	1,022	1,034
939	943	946	954	971	1,038	1,072	1,077	1,077	1,082
972	977	984	1,005	1,014	1,096	1,100	1,113	1,113	1,116
1,016	1,041	1,052	1,080	1,093	1,153	1,154	1,174	1,188	1,230

i) Form a frequency distribution for each manufacturer, using the following class-interval widths for each distribution:

- Manufacturer A: 650 but less than 750, 750 but less than 850, and so on.

- Manufacturer B: 750 but less than 850, 850 but less than 950, and so on.

ii) Construct histograms and stem-and-leaf diagrams on separate graphs

iii) Form cumulative distributions and plot both cumulative frequency curves on one graph.

iv) Who has LED bulbs with a longer life-Manufacturer A or Manufacturer B? Explain.

v) Find the median of the life of LED bulbs produced by each manufacturer with the help of curve.   **(4+6+6+2+2)**

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 average of n observations $x_1, x_2, x_3,..., x_n$ is $\overline{x}$. If x_1 is replaced by x', then the new average is $\dfrac{\overline{x} - x_1 + x'}{n}$.

   b) If each value of the variable X is divided by 5, then b'_{yx} from coded values will be 5 times of the original value of b_{yx}.

   c) If r(x, y) is the correlation coefficient between X and Y, the correlation coefficient between (ax + b) and y is "a×r(x, y) + b".

   d) The correlation between five paired observations $(3, 6), \left(\dfrac{1}{2}, 1\right), (2, 4), (1, 2)$ and $(4, 8)$ for X and Y variables is equal to 0.5.

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

2. a) In the following table, distribution of students is shown according to their weights (in kg). Determine which class shows greater variation in weights?

Weight (kg):	20-30	30-40	40-50	50-60	60-70
Class A:	07	10	20	18	07
Class B:	05	09	21	15	06

   Also find the coefficient of skewness for both series. (7)

   b) Find the average rate of

   a) motion in case of a person who rides the first mile at 10 mile/hour, the next mile at 8 mile/hour and the third mile at 6 mile/hour ; and

   b) increase in population which in the first decade has increased 20%, in the second 25% and in the third 44%. (3)

3. a) Ten competitors in a dance competition were ranked by three judges in the following orders:

I[st] Judge	1	6	5	10	3	2	4	9	7	8
IInd Judge	3	5	8	4	7	10	2	1	6	9
IIIrd Judge	6	4	9	8	1	2	3	10	5	7

   Determine which pair of Judges have the nearest approach to the common likings in dance. (4)

   b) When studying the relationship between two variables with 11 observations each, it was found that

$\bar{x} = 7, \quad \sigma_x = 2, \quad \bar{y} = 9, \quad \sigma_y = 4, \quad r = 0.5$

On verification one value of x(=7) and one value of y(=9) were found inaccurate and hence were removed from data. Find that how much the original value of r be affected?

(6)

4. a) In a state with a total population of 70,000 adults, 34, 000 are males and out of a total of 5000 graduates, 700 are females. Out of 1200 graduate employees of the state 200 are females. Is there any sex bias in education among people? The state holds that no distribution is made in appointments with respect of sex. How far this claim is substantiated by the data given above? (4)

   b) A census revealed the following figures of the blind and of the insane in two age groups in a certain population.

	Age group	
	15-25 years	over 25 years
Total population	2,70,000	1,60,000
Number of blinds	1,000	2,000
Number of insanes	6,000	1,000
Number of insane among the blinds	19	9

   a) obtain a measure of the association between blindness and insanity in each of the age groups.
   b) Do you consider that blindness and insanity are equally associated or disassociated with each other in two age groups or more in one-group than in other? (6)

5. Two random variables have the regression lines

   $3x - 2y = 26$ and $6x + y = 31$

   i) Find the mean values of x & y and r_{xy}.
   ii) Also find out the ratio of two variances. In particular is $\sigma_x = 4$, what do you conclude about σ_y?
   iii) Also show that the regression estimate of y for x = 0 is 13, where as the regression estimate of x corresponding to y = 13 is –3. Explain the cause of difference. (10)

6. Find the equation of best fitted suitable curve of the form $y = ae^{bx}$ for the following data:

x:	1	2	3	4	5	6
y:	1.6	4.5	13.8	40.2	125.3	300

   (10)

7. a) Following are the observation made on 18 cotton plants with respect to the yield of cotton x_1, number of seed-vessels x_2 and height of the plant (x_3)

   x_1: 10 51 30 42 25 18 44 56 38 32 25 10 20 27 13 49 27 55

   x_2: 8 15 11 21 07 05 10 13 12 13 5 6 4   8 7 12 6 16

   x_3: 4 5 3 3 2 1 4 6 3 4 2 3 4 4 3 5 3 7

   Find the multiple and partial correlations. Also find the regression equation of x_1 on x_2 and x_3. (12)

b) In a very hotly fought battle atleast 70% of the combatants lost an eye, atleast 75% lost an ear, atleast 80% lost an arm and atleast 85% lost a leg. How many atleast must have lost all four? (4)

c) A market investigator returns the following data:

Of 1000 people consulted, 811 liked chocolates, 752 liked toffee and 418 liked boiled sweets, 570 liked chocolates and toffee, 356 liked chocolates and boiled sweets and 348 liked toffee and boiled sweets, 297 liked all three. Show that this information as it stands must be incorrect. (4)

8. a) The first four moments of a distribution about the value 4 of the variable are -1.5, 17, -30 and 108. Find the moments about mean, β_1 and β_2.

Find also the moments about (i) the origin, and (ii) the point $x = 2$. (14)

b) A number of school boys were examined for the presence or absence of the three physical defectives, viz., development defect denoted by A, nerve signs denoted by B and low nutrition denoted by C. Following data in terms of the ultimate class frequencies were observed: $(ABC) = 149$, $(AB\gamma) = 738$, $(A\beta C) = 225$, $(A\beta\gamma) = 1196$, $(\alpha BC) = 204$ $(\alpha B\gamma) = 1762$, $(\alpha\beta C) = 171$ and $(\alpha\beta\gamma) = 21842$. Find all the class frequencies which are for the presence of defects. (6)

# TUTOR MARKED ASSIGNMENT
MST-003: Probability Theory

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 3:5 then odds against of the event A will be 5:3.

(b) If two independent events A and B are such that $P(A) = 2/3$ and $P(B) = 1/3$, then events $\overline{A}$ and $\overline{A} \cap B$ are not independent.

(c) If a random variable X takes values 0, 1, 2 with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ respectively, then $F(1) = 3/4$.

(d) If X denotes the waiting time in minutes until the 3rd customer arrives in a mobile showroom, then X follows exponential distribution.

(e) The probability that a shooter hits a target in Olympic is 1/3. If she fires 6 times, then to calculate the probability of hitting the target at least 3 times we can use geometric distribution.

2. (a) For answering a multiple choice question in a test, a student has two possibility: he/she knows the answer or guesses. Let $\frac{3}{4}$ be the probability that he/she knows the answer and $\frac{1}{4}$ be the probability that he/she guesses. Assuming that a student who guesses the answer will be correct with probability $\frac{1}{4}$. What is the probability that a student who knows the answer, given that he/she answered it correctly? **(5)**

(b) A student has a questions bank consisting of 300 easy True/False questions, 200 difficult True/False questions, 500 easy multiple choice questions and 400 difficult multiple choice questions. If a question is selected at random from questions bank, what is the probability that it will be an easy question given that it is a multiple choice question? **(5)**

3. (a) A continuous random variable X has density function $f(x) = x + \frac{1}{2}$ for $0 < x < 1$ and jointly distributed with a continuous random variable Y. If conditional density function of Y given X = x is

$$f_{Y|X}(y|X = x) = \frac{x + y}{x + \frac{1}{2}} \text{ for } 0 < x < 1 \text{ and } 0 < y < 1.$$

then find $f_Y(y)$. **(6)**

(b) An unbiased die and an unbiased coin (having two numbers 1 and 2 on its faces) are thrown simultaneously. Find the expected value of the sum of numbers on them. **(5)**

4. (a) A hospital has 20 kidney dialysis machines and the chance of malfunctioning any one of them during any day is 0.06. If someone want to find the probability that exactly 3 machines will be out of service on the same day. Then,

    i) can he/she use the binomial distribution to find this probability? If yes, calculate the probability.

    ii) can he/she use the Poisson distribution to find this? If yes, calculate the probability.

    iii) if both, then compare the probabilities   **(5)**

(b) During a particular period, a University's Information Technology Office received 20 service orders for problems with printers, of which 8 were laser printers and 12 were inkjet printers. A sample of 5 of these service orders is to be selected in a completely random fashion in a customer satisfaction survey. What is the probability that exactly 2 of the selected service orders were for inkjet printers?   **(5)**

5. An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil.

    i) What is the probability that the first strike comes on the third well drilled?

    ii) What is the probability that the third strike comes on the seventh well drilled?

    iii) What is the mean and variance of the number of wells that must be drilled if the oil company wants to set up three producing wells?   **(4+4+2)**

6. (a) The magnitude of earthquakes recorded in a region modelled as having an exponential distribution, with a mean of 2.4 (as measured on the Richter Scale). Find the probability that an earthquake striking this region will:

    i) exceed 3.0 on the Richter Scale and

    ii) all between 2.0 and 3.0 on the Richter Scale   **(3+3)**

(b) Suppose that you spin the dial shown in the figure so that it comes to rest at a random position.
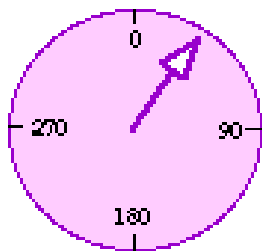
What is the suitable probability density function of this? Use it, to find the probability that the dial will land somewhere between 5° and 300°.   **(4)**

7. (a) In a college, it is known that 60% students reside in hostel and 40% are day scholars (not residing in hostel). Previous year results report that 30% of all the students who reside in hostel attained A grade and 20% of day scholar attained A grade in their annual examination. At the end of the year, one student is chosen at random from the college and he has an A grade, what is the probability that the student is hosteller?   **(8)**

(b) A set of final examination grades in PGDAST programme is normally distributed, with a mean of 73 and a standard deviation of 8.

    i) What is the probability of getting a grade below 91 on this exam?

    ii) What is the probability that a student scored between 65 and 89?

    iii) Compute the grade of a student, taking the test scores higher than that grade for which the probability is 5%. **(4+4+4)**

8. A student is appearing in a multiple-choice exam in which each question has four choices. Assuming that she has no knowledge of the correct answers to any of the questions, she has decided on a strategy in which she will place four balls (marked A, B, C and D) into a box. The marking on the ball will determine her answer to the question. She randomly selects one ball for each question and replaces the ball in the box. There are five multiple-choice questions on the exam.

    i) Find sample space of the correct answers.

    ii) If X represents the numbers of correct answers, then find the probability distribution of X.

    iii) Find the distribution function of the random variable X and draw its graph.

    iv) What is the probability that she will get more than four questions correct?

    v) What is the probability that she will get no more than two questions correct?

    vi) If each question is of 2 marks, what is the expected value of the correct answers?

(2+4+6+2+2+4)

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 sample size of a survey is increased 4 times, then the standard error will be doubled.

   (b) If p-value and level of significance for testing a hypothesis, $H_0 : P_1 \geq P_2$ against $H_1 : P_1 < P_2$, are 0.06 and 0.05, respectively, then the H_0 may be accepted.

   (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) A survey is planned to determine the mean annual family medical expenses of employees of a large company. For determining the minimum sample size, the management of the company requires only the knowledge of standard deviation of the annual family medical expenses of the previous years.

   (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 as 120,118,125,122,117,120,121, 126,125,115 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 following data represent the number of absent days in a year of 5 employees in a small company:

   1 3 6 7 9

   i) Assuming that you sample without replacement, select all possible sample of n = 2 and construct the sampling distribution of the mean.

   ii) Repeat (i) for all possible samples of n = 3.

   iii) Compare the shape of the sampling distribution of the mean in (i) and (ii). Which sampling distribution has less variability? Why? **(4+4+2)**

3. **(a)** Find the mean and variance of the following sampling distributions:

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

   ii) $f(\chi^2) = \dfrac{1}{2}e^{-\chi^2/2}; \quad 0 < \chi^2 < \infty$ **(4)**

   (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. Is there evidence of a significant difference

between males and females in the proportion that they enjoy shopping for clothing at the 5% level of significance? **(6)**

4. An ambulance agency claims that the standard deviation of the length of service times is 5 minutes. Investigator suspects that this claim is wrong. She takes a random sample of 20 services and finds the standard deviation as 6 minutes. Assume that the service time of the ambulance follows normal distribution.

   i) What is the probability that the standard deviation of the length of service times is more than and equal to 5 minutes.

   ii) Find the 95% confidence interval for the standard deviation of the length of service times.

   iii) Test, at $\alpha = 0.01$, is there enough evidence to reject the agency's claim? **(4+2+4)**

5. The following data represent the distribution of number of sons among the children of 100 families containing 5 children each:

Number of Sons	0	1	2	3	4	5
Number of Families	4	23	38	23	10	2

   Apply the suitable test to test whether the number of sons in a family follows uniform distribution at 5% level of significance. **(10)**

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

S. No.	Test For	Name of the Test	Test Statistic	Assumptions for Applying the Test	Hypotheses and Decision Rule (in short)
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.	1. $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$ Reject H_0, if $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ 2. $H_0 : \mu \leq \mu_0$ and $H_1 : \mu > \mu_0$ Reject H_0, if $z \geq z_\alpha$ 3. $H_0 : \mu \geq \mu_0$ and $H_1 : \mu < \mu_0$ Reject H_0, if $z \leq -z_\alpha$
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 when the form of population is not known and sample size is large				
5	Population mean/median when the form of population is not known and the data is in ordinal scale and sample size is small				

(10)

7. If the magnitude of the earthquakes recorded in a region of a country follows an exponential distribution with parameter θ, then find

i) the maximum likelihood estimator of the parameter

ii) the 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 obtained estimator is also sufficient estimator using invariance property of sufficiency.　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　**(6+6+8)**

8. A bank with a branch located in a commercial district of a city has developed an improved process for serving customers during the lunch period (noon-to-1 P.M). The waiting time (operationally defined as the time elapsed from when the customer enters in the line until he/she reaches the teller window) needs to be shortened to increase customer satisfaction. A random sample of 15 customers is selected and measured the waiting time. The results (in minutes) are as follows:

4.21	5.55	3.02	5.13	4.77	2.34	3.54	3.20

4.50	6.10	0.38	5.12	6.46	6.19	3.79

Suppose that another branch, located in a residential area, is also concerned with the lunch period (noon –to-1 P.M.). A random sample of 15 customers is selected and the results (in minutes) are as follows:

9.66	5.90	8.02	5.79	8.73	3.82	8.01	8.33

10.49	6.68	5.64	4.08	6.17	9.91	5.47

i) Assuming that the population variances of the waiting time for both banks are equal, is there evidence of a difference in the mean waiting times between the two branches? (Use $\alpha = 0.05$.)

ii) In addition to equal variances, what other assumption is necessary in (i)?

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

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

    a) $V_{opt}(\bar{x}_{st})$ lies between $V_{prop}(\bar{x}_{st})$ and $V_{Random}(\bar{x}_{st})$.

    b) In cluster sampling the variance within clusters is greater than the variance between clusters.

    c) While analysing the data of a 5×5 Latin Square design the error d.f. is equal to 12.

    d) In a Two way analysis of variance with 4 blocks & 4 treatments the degree of freedom for the total variation is 15.

    e) Suppose a random number generated by Middle. Square Method is 16, then the next random number will be 44. $(2\times5=10)$

2. A sampler proposes to take a stratified random sample. He expects that his field costs will be of the form $\sum c_i n_i$ His advance estimate of relevant quantities for the two strata are as follows:

Stratum	W_i	S_i	C_i
1	0.4	10	Rs. 4/-
2	0.6	20	Rs. 9/-

    Then

    i) find the value of $\dfrac{n_1}{n}$ and $\dfrac{n_2}{n}$ that minimise the total cost for a given value of $V(\bar{y}_{st})$,

    ii) find the sample size required under this optimum allocation to make $V(\bar{y}_{st}) = 1$.

    iii) how much will the total field cost be? (10)

3. Consider the following four samples

Sample			
A	B	C	D
21	25	25	22
25	22	23	20
22	27	23	33
28	21	26	35
20	23	31	20

    Perform the analysis of variance and test the hypothesis at 0.05 level of significance, that the samples come from population having the same mean. (10)

4. A farmer applies three types of fertilizers on 4 separate plots. The figure on yield per acre are tabulated below:

Fertilizer/Plot	A	B	C	D	Total
Yield					
Nitrogen	6	4	8	6	24
Potash	7	6	6	9	28
Phosphate	8	5	10	9	32
	21	15	24	24	84

Find out if the plots are materially different in fertility and also, if the three fertilizers make any material difference in yields. (10)

5. Following U(0,1) were generated by a generator.

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

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

6. a) Obtain a systematic sample of 25 units from a list comprising the population of 150 individuals. (4)

b) The inter-arrival times of customers arriving in a Pizza shop have an exponential distribution with rate $\alpha = 0.3$ per minute. Simulate the times of 10 customers arriving in the shop. Also give the number of customers arriving in first 30 minutes. (6)

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

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

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

b) Generate a complete cycle for the LCG given below: $x_i (10 x_{i-1} + 5) \mod 20$ with $x_0 = 8$.

A man tosses an unbiased coin 10 times. Using the first 10 random numbers generated above, obtain a sequence of heads and tails. (6)

8. a) How does sampling with replacement differs form that without replacement? Which of them gives a lower value of the standard deviation of the sample mean? Explain by considering sample mean and sample variance are unbiased estimates of the population mean and population variance, respectively, and show that its variances is given by

$\dfrac{N-n}{Nn} S^2$, where the terms have their usual meaning. (8)

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

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

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

TUTOR MARKED ASSIGNMENT
MST-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) Twenty pieces of different length of cloth contained 2, 4, 1, 3, 5, 4, 2, 7, 3, 5, 2, 2, 4, 5, 6, 4, 2, 1, 2, 4 defects respectively. To check the process is under control with respect to the number of defects, we should use p-chart.

   (b) If density function of the time to failure of an appliance is $f(t) = \dfrac{32}{(t+4)^3}$; $t > 0$ then reliability of the appliance for two years will be 0.444.

   (c) If the probabilities are associated wit the occurrence of different states of nature, then the situation s known as decision making under uncertainty.

   (d) If there are 5% defectives notes of Rs.500 in a lot of 10000 notes of Rs 500, then the lot quality will be 95%.

   (e) A two-person zero-sum game having the following payoff matrix for player A

		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	Sample	Number out of Compliance
1	2	6	0	11	6
2	0	7	4	12	1
3	6	8	3	13	0
4	3	9	2	14	2
5	1	10	2	15	1

   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 $\overline{\overline{X}} = 199$ 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? **(3)**

3. A manufacturer of men's jeans purchases zippers in lots of 800. The jeans manufacturer uses single-sample acceptance sampling with a sample size of 10 to determine whether to accept the lot. The manufacturer uses $c = 2$ as the acceptance number. Suppose 10% nonconforming zippers are acceptable to the manufacturer and 30% nonconforming zippers are not acceptable.

    i)   Construct an OC curve.

    ii) Average outing quality (AOQ), if the rejected lots are screened and all defective zippers are replaced by non-defectives.

    iii) Average total inspection (ATI). **(6+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 four 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, or four 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.

    i)   What type of acceptance plan is used here?

    ii) What is the probability of accepting the lot?

    iii) What is the inspector's decision?

    iv) What is the consumer's risk

    v) What is the producer's risk **(1+4+1+2+2)**

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	170	180	190	200
Probability	0.1	0.2	0.4	0.3

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?

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-Diwali sales during the first week of November. Sales are advertised through a local newspaper, radio and television media. With the aid of an advertising firm store ABC constructed the game matrix given below.(Figures in the matrix represent a gain or loss of customers).

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

Determine optimal strategies and the worth of such strategies for both ABC and XYZ.

7. (a) The system shown below is made up of ten components. Components 3, 4 and 5 are not identical and at least two component of this group must be available for system success.

Components 8, 9 and 10 are not identical and for this particular group it is necessary that one out of the three components functions.
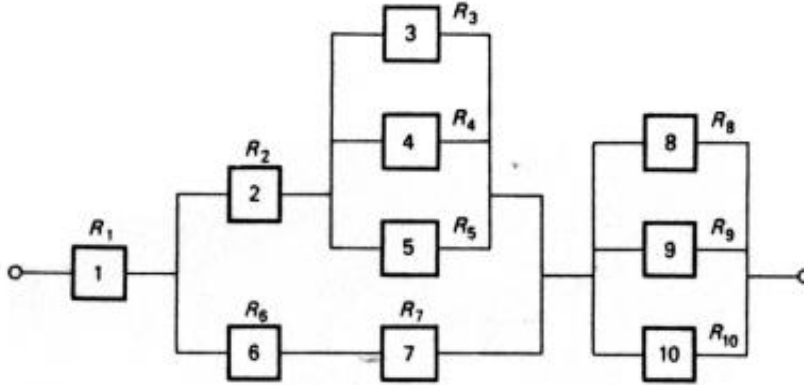
What is the system reliability if $R_1 = R_2 = R_3 = R_4 = R_5 = 0.85$, $R_6 = R_7 = R_8 = 0.95$ $R_9 = 0.80$ and $R_{10} = 0.70$. **(10)**

(b) If density function of time to failure of an appliance is

$$f(t) = \frac{32}{(t+4)^3}; \, t > 0$$

Then find

i)   reliability function,

ii)   the failure rate, and

iii) Mean tine to failure **(4+3+3)**

7. A toy-manufacturing company has been given a large order for small plastic whistles that will be given away by large fast-food hamburger chain with its kid's meal. Seven random samples of four whistles have been taken. The weight of each whistle has been ascertained in grams. The data are shown here:

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7
4.1	3.6	4.0	4.6	3.9	5.1	4.6
5.2	4.3	4.8	4.8	3.8	4.7	4.4
3.9	3.9	5.1	4.7	4.6	4.8	4.0
5.0	4.6	5.3	4.7	4.9	4.3	4.5

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) If there are 2 equations with 4 variables in a LPP then the maximum number of possible basic solutions is are 6.
   ii) If the arrival rate, is 6 per hour and service rate is 2 per hour, then the probability of no customer in queue is 0.7.
   iii) A time series is a set of values arranged in geographical order.
   iv) In a regression equation $Y = 5X + 4 + e$, the intercept of the line is 5.
   v) If the coefficient of determination is 0.933, the number of observations and independent variables are 10 and 2, respectively, then Adjusted R^2 will be 0.84.

2. Solve the following LPP using simplex method:

   Maximize $\quad Z = 107 x_1 + x_2 + 2x_3$

   Subject to the constraints:

   $14x_1 + x_2 - 6x_3 + 3x_4 = 7$

   $16x_1 + \dfrac{1}{2}x_2 - 6x_3 \leq 05$

   $3 x_1 - x_2 - x_3 \leq 0$

   $x_1, x_2, x_3, x_4 \geq 0$ (10)

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

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

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

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

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

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

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

vii)   What is the probability that more than 3 customers are in the system?   (10)

4. Let X is the advertisement expenditures (in Lakh Rs.) and Y is the sales (in Lakh Rs.). Let the data are

X:	1182	1172	1264	1493	1571	1711	1804	1840	1956	1954
Y:	129	135	147	160	171	184	198	223	240	293

Estimate the parameters and find the estimated linear equation. Whether the advertisement influences the sale of product? Test and comment on the goodness of fit of the model.   (10)

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

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

(10)

6. A solicitors' firm employs typists on hourly price-rate basis for their daily work. There are five typists and their charges and speed are different. According to an earlier understanding only one job is given to one typist and the typist is paid for a full hour even if he works for a fraction of an hour. Find the least cost allocation for the following data:

Typist	Jobs				
	P	Q	R	S	T
A	85	75	65	125	75
B	90	78	66	132	78
C	75	66	57	114	69
D	80	72	60	120	72
E	76	64	56	112	68

(10)

7. a) Obtain seasonal Indices by the "Moving average" method from the following data:

Quarterly output of a Factory			
Year	I	II	III	IV
2010	65	58	56	61
2011	68	63	63	67
2012	70	59	56	52
2013	60	55	51	58

(10)

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

$\hat{B}_0 = 52.38$, $\hat{B}_1 = 31.6161$, $\hat{B}_2 = 0.0414$, $SS(B_0) = 125.26$, $SS(B_0; B_1) = 179.08$
$SS(B_0, B_2) = 171.25$, $SS(B_0, B_1, B_2) = 180.79$

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

8. a) A company has three production factories S_1, S_2 and S_3 with production capacity of 7, 9 and 18 units (in 100 s) per week of a product, respectively. These units are to be shipped to four warehouses D_1, D_2, D_3 and D_4 with requirements of 5, 8, 7 and 14 units (in 100's) per week, respectively. The transportation costs (in rupees) per unit between factories to ware houses are given in table below.

Warehouses / Factories	D_1	D_2	D_3	D_4	Capacity
S_1	19	30	50	10	7
S_2	70	30	40	60	9
S_3	40	8	70	20	18
Demand	5	8	7	14	34

Obtain the initial basic solution using LC Method and also obtain the optimum solution using MODI method. (14)

b) Twenty Five successive observations on a stationary time series are given as follows:

30, 33, 32, 27, 25, 28, 29, 31, 35, 34, 38, 31, 23, 24, 34, 36, 29, 32, 38, 27, 22, 29, 20, 40, 37.

Calculate r_1, r_2, ..., r_{10} and 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, interpretation, etc.**

Q 1 An electronic company manufactured a new battery which is supposed to be better than the battery of the reputed brand. 100 batteries of both brands were tested to determine how long they would last. The results, to the nearest minute, were recorded as follows:

S. No.	Life of New Battery	Life of Old Battery	S. No.	Life of New Battery	Life of Old Battery
1	308	335	51	455	546
2	328	383	52	433	520
3	390	455	53	336	403
4	364	425	54	352	423
5	407	325	55	405	486
6	324	378	56	427	512
7	374	436	57	319	383
8	390	455	58	370	445
9	312	365	59	427	512
10	321	375	60	325	390
11	382	446	61	342	356
12	407	475	62	316	380
13	328	383	63	376	451
14	339	395	64	351	421
15	429	515	65	393	471
16	316	380	66	312	374
17	312	351	67	361	433
18	450	540	68	376	451
19	342	364	69	321	361
20	363	436	70	332	371
21	456	548	71	369	443
22	434	521	72	393	471
23	337	405	73	316	380
24	354	425	74	326	392
25	406	488	75	426	511
26	428	514	76	313	376
27	320	384	77	346	348
28	372	446	78	447	536
29	428	514	79	326	361
30	326	391	80	361	433
31	387	358	81	454	545
32	318	381	82	431	518
33	377	453	83	334	401

34	352	423	84	351	421
35	394	473	85	404	485
36	313	376	86	425	511
37	362	435	87	317	381
38	377	453	88	369	443
39	316	363	89	425	511
40	311	373	90	323	388
41	370	445	91	352	355
42	394	473	92	315	378
43	318	381	93	375	450
44	328	394	94	350	420
45	427	513	95	391	470
46	315	378	96	327	373
47	354	349	97	359	431
48	448	538	98	375	450
49	312	363	99	314	360
50	362	435	100	351	370

Answer the following:

i) Which brand of battery has more average life?

ii) Which brand of battery is more consistent?

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

iv) Also, construct the histogram for both brands of batteries and interpret the results.

(2+5+10+8)

Q 2 The marks obtained by 60 students of PGDAST programme in the assignments of MST-002, MST-004 and MSTL-001 courses consisting of maximum 100 marks, are presented in the table given below:

S. No.	Marks in MST-002	Marks in MST-004	Marks in MSTL-001	S. No.	Marks in MST-002	Marks in MST-004	Marks in MSTL-001
1	58	82	60	31	63	80	61
2	53	77	59	32	61	91	70
3	40	76	56	33	50	84	66
4	68	91	71	34	61	86	65
5	51	89	69	35	52	82	63
6	50	79	62	36	69	94	70
7	65	98	74	37	73	98	74
8	60	92	71	38	47	77	58
9	61	85	63	39	40	72	52
10	47	77	60	40	47	70	52
11	64	93	73	41	60	94	70
12	58	96	74	42	56	77	60
13	60	100	79	43	61	87	65
14	59	89	66	44	49	73	55
15	57	93	70	45	50	79	62
16	67	89	68	46	62	92	68

17	54	71	51	47	53	77	57
18	71	100	75	48	64	96	72
19	54	71	53	49	67	85	63
20	58	91	68	50	55	82	61
21	45	73	53	51	57	80	62
22	57	87	64	52	51	80	60
23	59	89	68	53	42	72	56
24	62	80	63	54	50	86	67
25	40	74	58	55	43	70	51
26	53	75	55	56	71	95	75
27	66	99	78	57	65	86	65
28	70	92	69	58	60	87	64
29	57	79	60	59	52	71	54
30	53	92	68	60	39	72	54

i) Compute the correlation coefficient between the marks in MSTL-001 and the joint effect of the marks in MST-002 and MST-004.

ii) Compute the correlation coefficient between the marks in MST-002 and MST-004 after eliminating the linear effect of the marks in MSTL-001.

iii) Also represent the marks obtained in MST-002, MST-004 and MSTL-001 using box plot.

(7+7+11)

Q 3 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		S. No.	Breaking Pressure	
	Iron	Copper		Iron	Copper
1	72.7	73	21	87.3	71.4
2	69.6	67.2	22	75.1	75.4
3	83.4	75.3	23	72	69.6
4	78.9	61.4	24	85.8	77.7
5	75	74	25	81.3	63.8
6	71.6	69.5	26	77.4	76.4
7	85.7	69.8	27	74	71.9
8	73.5	73.8	28	88.1	72.2
9	70.4	68	29	75.9	76.2
10	84.2	76.1	30	72.8	70.4
11	79.7	62.2	31	86.6	78.1
12	75.8	74.8	32	82.1	64.2
13	72.4	70.3	33	78.2	76.8
14	86.5	70.6	34	74.8	72.3
15	74.3	74.6	35	88.9	72.6
16	71.2	68.8	36	76.7	76.6
17	85	76.9	37	73.6	70.8

18	80.5	63	38	87.4	78.5
19	76.6	75.6	39	82.9	64.6
20	73.2	71.1	40	79	77.2

If the breaking pressures for both iron and copper are normally distributed, then answers the following questions:

i) Are the variances of the distributions of the breaking pressure of iron and copper equal at 5 % level of significance?

ii) If yes, check whether the average breaking pressure for iron is more than the copper at 5 % level of significance? **(15+10)**

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

	Operator					
	1	2	3	4	5	6
1	A	B	C	D	E	F
2	B	C	D	E	F	A
3	C	D	E	F	A	B
4	D	E	F	A	B	C
5	E	F	A	B	C	D
6	F	A	B	C	D	E

(Time Period in rows)

The average hourly production in a day is given as follows:

	Operator					
	1	2	3	4	5	6
1	133	139	140	140	145	138
2	136	141	143	146	139	142
3	140	138	142	139	139	141
4	129	132	137	136	140	138
5	132	144	143	142	142	140
6	138	140	141	137	141	139

(Time Period in rows)

Assuming that the effect of each operator, time period and machine are normally distributed with approximately equal variances, analyse the design at 1% level of significance. Test whether the effect of the different operators, time periods and machines on the production is significant or not. 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, interpretation, etc.**

Q 1   An engineer of a chemical plant has responsibility of observing pH level of a particular process to monitor stability of a system. To be consistent with the data, he/she takes four pH readings every hour. The data are displayed in the following table:

Time (hours)	pH Values			
1	8	8.2	7.99	8
2	8.12	8.08	8.12	7.9
3	8.07	8.03	8	8.01
4	7.98	8	7.94	8
5	7.99	8	7.99	7.98
6	8	7.93	8.02	7.93
7	7.92	8	7.92	8.02
8	7.88	8	7.94	7.99
9	8.1	8	8	8
10	7.9	8.02	8.12	8.04
11	8.01	7.86	8.01	7.9
12	7.86	7.98	7.9	7.98
13	7.9	8	7.87	8
14	8.01	8.04	8.01	8.05
15	8	7.95	8	7.99
16	8.09	8.02	8.03	8.1
17	7.89	8.14	7.87	8.2
18	7.98	7.8	7.98	7.89
19	8	7.9	8	7.9
20	8	8	8.03	8
21	8.04	8.03	8.08	8
22	7.9	7.92	7.98	7.92
23	8	7.99	8.1	8
24	8	7.97	8.01	7.98
25	7.98	7.8	7.98	7.89

The engineer of this plant needs to construct suitable control charts for variability as well as average to infer whether the pH of the process is under control or not. If it is out-of-control, construct the revised control charts. **(25)**

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

Sample No.	Sampled Mirrors	Out of Compliance Mirrors
1	75	1
2	77	2
3	72	2
4	71	8
5	75	2
6	73	1
7	76	3
8	79	4
9	77	2
10	75	1
11	80	4
12	85	1
13	80	3
14	80	2
15	77	1
16	73	2
17	75	1
18	81	3
19	77	1
20	78	3
21	75	2
22	80	1
23	75	3
24	75	1
25	72	2

Construct a suitable control chart for fraction of out of compliance mirrors to check whether the process is said to be in a state of control or not using both approaches. Also construct the revised control charts, if necessary. **(25)**

Q 3 A researcher is interested to check the relationship between the number of accidents on a particular stretch of a highway with the average speed of the vehicles and the accident time (0 if day time and 1 if night time). The accident-related data were collected from traffic police records in order to examine the contribution of several variables to the number of accidents. A total of 40 subjects involved in accidents were sampled. The data are shown in the following table:

S. No.	No. of Accidents	Average Speed (km/h)	Accident Time
1	7	46	1
2	13	65	1
3	10	59	1
4	15	84	0
5	17	91	1
6	15	78	1
7	10	53	0
8	7	49	1
9	5	42	0
10	16	87	0
11	11	53	1
12	8	54	0
13	13	65	1
14	11	61	1
15	13	71	0
16	13	71	0
17	12	68	0
18	13	65	1
19	10	55	1
20	12	66	0
21	11	55	1
22	9	55	0
23	9	62	0
24	11	65	0
25	8	54	0
26	5	45	0
27	6	45	0
28	10	62	0
29	5	40	0
30	9	58	0
31	13	65	1
32	11	58	1
33	14	67	1
34	8	42	1
35	16	81	1
36	18	92	1
37	15	80	0
38	17	91	1
39	8	55	0
40	5	33	0

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

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

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

iv) Also, draw both fitted regression lines on the scatter plot.   **(5+5+10+5)**

Q 4　An archaeologist is interested in analysing the pattern of the number of visitors in a particular monument in past years and then forecasting the expected number of visitors for the upcoming years. The following quarterly data are available for past 10 years:

Year	Quarter	No. of Visitors	Year	Quarter	No. of Visitors
2006	I	72	2011	I	221
	II	79		II	246
	III	116		III	309
	IV	99		IV	278
2007	I	99	2012	I	233
	II	112		II	258
	III	172		III	322
	IV	149		IV	291
2008	I	135	2013	I	246
	II	157		II	271
	III	212		III	334
	IV	193		IV	303
2009	I	173	2014	I	258
	II	191		II	283
	III	256		III	347
	IV	224		IV	316
2010	I	208	2015	I	271
	II	233		II	296
	III	297		III	359
	IV	266		IV	328

i) Determine the seasonal indices for the data using ratio-to-trend method.

ii) Obtain deseasonalised values.

iii) Fit the appropriate trend on the deseasonalised data using the least-squares method by matrix approach which best describes the data.

iv) How many visitors would you expect for each quarter of the year 2016?

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

(8+2+8+2+5)