Problems on Statistics

Prof. Jaydip Sen

IIDT Business Analytics Course

Frequency Distribution

1. The data for average inventory (in days) for 20 convenient stores is given below.

   2.0 3.8 4.1 4.7 5.5 3.4 4.0 4.2 4.8 5.5
   3.4 4.1 4.3 4.9 5.5 3.8 4.1 4.7 4.9 5.5

   Represent the data in frequency distribution with 6 classes and 12 classes.

Solution: (6 classes)

Class (Group of similar values of data points)	Frequency (# of observations in each class)
2.0 – 2.5	1
2.6 – 3.1	0
3.2 – 3.7	2
3.8 – 4.3	8
4.4 – 4.9	5
5.0 – 5.5	4

12 classes

Class	Frequency	Class	Frequency
2.0 – 2.2	1	3.8 – 4.0	3
2.3 – 2.5	0	4. – 4.3	5
2.6 – 2.8	0	4.4 – 4.6	0
2.9 – 3.1	0	4.7 – 4.9	5
3.2 – 3.4	2	5.0 - 5.2	0
3.5 – 3.7	0	5.3 - 5.5	4

Relative Frequency Distribution

Class	Frequency	Relative Frequency
2.0 – 2.5	1	1/20 = 0.05
2.6 – 3.1	0	0/20 = 0.00
3.2 – 3.7	2	2/20 = 0.10
3.8 – 4.3	8	8/20 = 0.40
4.4 – 4.9	5	5/20 = 0.25
5.0 – 5.5	4	4/20 = 0.20
	Total: 20	Total : 1.00

Steps in Constructing a Frequency Distribution

1. Decide on the type(equal or unequal) and number of classes for dividing the data.

2. Sort the data points into classes and count the number of points in each class.
3. Illustrate the data in a table and then in the form of a suitable chart.

2. The data array of daily production in yards of 30 carpet looms (in ascending order of magnitude) is given below.

15.2 15.4 15.6 15.6 15.6 15.7 15.7 15.8 15.8 15.8 15.9 15.9
15.9 15.9 16.0 16.1 16.2 16.2 16.3 16.3 16.3 16.4 16.4 16.6
16.8 16.8 16.9

Daily Production in a Sample of 30 Carpet Looms with 0.3 Yard Class Interval

Class	Frequency
15.2 – 15.4	2
15.5 – 15.7	5
15.8 – 16.0	11
16.1 – 16.3	6
16.4 – 16.6	3
16.7 – 16.9	3

Now draw the frequency bar diagram (y axis showing the frequencies and the x axis showing the class intervals). We could also have used two classes with 15.1 – 16.0 the first class and 16.1 – 17.0 the second class. The two classes having frequencies 18 and 12 respectively. We could also have drawn the frequency diagram depicting the two classes.

Measures of Central Tendency

1. The service time in minutes at the reservation counter of a railway station is given by the following frequency distribution. This frequency distribution was constructed from the raw data that included service to twenty five customers. Calculate the average service time.

Service time (in min)	2.1 – 2.7	2.7 – 3.3	3.3 – 3.9	3.9 – 4.5	4.5 – 5.1	5.1 - 5.7
Frequency	2	6	7	5	3	2

Solution: The average can be computed using the mid values of class intervals and using the formula: Mean(x)v = Sum(fi*xi)/Sum(fi).

Class Interval	Class Marks (xi)	Frequency (fi)	xi*fi
2.1 – 2.7	2.4	2	4.8
2.7 – 3.3	3.0	6	18.0
3.3 – 3.9	3.6	7	25.2
3.9 – 4.5	4.2	5	21.0
4.5 – 5.1	4.8	3	14.4
5.1 – 5.7	5.4	2	10.8
		Total: 25	Total: 94.2

Average service time = 94.2/25 = 3.8 min (approx)

2. The average rate of dividend for engineering, chemical, and textile industry for a year are 25%, 40% and 20%. The number of observations on which these averages are based on are 40, 50 and 30. Calculate the pooled arithmetic mean.

   Solution: $\overline{X_1}$ = 25, $\overline{X_2}$ = 40, $\overline{X_3}$ = 20, n1 = 40 n2 = 50, n3 = 30.
   Pooled average = (25*40 + 40*50 + 20*30)/(40+50+30) = 30

3. Calculate the crude and standardized death rates of two populations A and B from the following data:

Age group	Standard population Town A			Standard population Town B		
	Population	Death	Death per thousand	Population	Death	Death per thousand
Below 10	30.000	720	24	80,000	2000	25
10 – 30	40.000	800	20	1,04,000	2080	20
Above 50	20.000	560	28	16,000	480	30

   Solution: Crude (or corrected) death rate of town A = (30000*24 + 40000*20 + 20000*28)/90000 = 23.1
   Crude (or general) death rate of town B = (80000*25 + 104000*20 + 16000*30)/200000 = 22.8
   Standardized or corrected death rate of town B = (30000*25 + 40000*20 + 20000*30)/90000 = 23.9
   Note: It may be observed that on the basis of comparison of crude rates, the town B has a lower death rate and therefore is a healthier town. However, this comparison may be misleading. Therefore, a better procedure would be to compare the standardized (corrected) death rates of both the towns. This comparison indicates that the standardized death rate is higher for town B than town A and therefore town A is healthier.

4. Calculate the median for the following data that relates to number of patients treated in an emergency ward on ten consecutive days: 20, 68, 50, 15, 35, 48, 33, 44, 56, 64.

   Solution: We arrange the data in ascending order:

Data item# :	1	2	3	4	5	6	7	8	9	10
No of patients:	15	20	33	35	44	48	50	56	64	68

   The central items are 5^{th} and $6^{th,}$ and so the median is: (44+48)/2 = 46

5. Data on weekly wages of 265 industrial workers is tabulated below. Compute the median wages.

Weekly wages (Rs)	50 - 59	60 - 69	70 – 79	80 - 89	90 – 99	100 - 109	100 - 119
No. of	15	40	50	60	45	40	15

workers							

Solution: The median is the (265+1) th item, i.e., 133rd item.
Hence the median class (i.e., class in which the median lies) is 80 -90 based on the cumulative frequency figures. Hence the lower boundary of the median class is 79.5.

Therefore, $Median = l + \frac{\frac{N}{2} - F}{f} * h$ where l is the lower boundary of the median class, N is the total number of observation, F is the cumulative frequency till the class preceding the median class, f is the frequency of the median class and h is the width of the median class.

Here, l = 79.5, N = 265, F = 105, f = 60 and h = 10. Median = 79.5 + $\frac{\frac{265}{2} - 105}{6}$*10 = 79.5 + 4.58 = 84.08.

Weekly Wages (Class Limits)	Class Boundaries	No of Workers (fi) (Frequencies)	Cumulative Frequencies
50 – 59	49.5 – 59.5	15	15
60 – 69	59.5 – 69.5	40	55
70 – 79	69.5 – 79.5	50	105
80 – 89	79.5 – 89.5	60	165
90 – 99	89.5 – 99.5	45	210
100 – 109	99.5 – 109.5	40	250
110 - 119	109.5 – 119.5	15	265

6. Calculate the median from the following data:

Land Holding (Acres)	0 - 5	5 - 15	15 - 45	45 - 60	60 - 70
# of Farmers	16	30	60	30	10

Solution: Even in cases where class intervals are unequal, the formula for computation of the median remains unchanged and therefore, there is no need to adjust frequencies to make the class intervals equal.

Land Holding	Class Boundaries	# of Farmers (Frequencies)	Cumulative Frequencies
0 - 15	0 – 15	16	16
5 – 15	5 – 15	30	46
15 – 45	15 – 45	60	106
45 – 60	45 – 60	30	136
60 - 70	60 - 70	10	146

The median is the N/2 th term i.e., 146/2 = 73rd item. Hence the median lies in the class 15 – 45.
The median is = 15 + ((73 – 46)*30)/60 = 15 + 13.5 = 28.

7. Calculate the mode for the following distribution:

Gross profit as %	0 - 7	7 - 14	14 - 21	21 - 28	28 - 35	35 - 42	42 - 49

of sales							
No of companies	19	25	36	72	51	43	28

Solution: The largest frequency is 72 and it lies in the class 21 – 28. So the "Modal Class" is 21 – 28. Note that the classes are represented by Class Boundaries. Mode is given by the following formula:

$$Mode = l + \frac{f - f_1}{2f - f_1 - f_2} * h$$

Where l is the lower boundary of the modal class, f is the frequency of the modal class, f1 and f2 are the frequencies of the classes preceding and following the modal class respectively and h is the width of the modal class.

In this problem, l = 21, f = 72, f1 = 36, f2 = 51, h = 7. Mode = 21 + ((72 -36)/(144-36-51))*7 = 21 + 4.4 = 25.4

8. Calculate the lower and upper quartiles, fourth decile and the 60th percentile for the following distribution:

Percentage of dividend declared	Number of companies	Percentage of dividend declared	Number of companies
5 -10	5	25 - 30	5
10 -15	6	30 -35	4
15 - 20	15	35 - 40	2
20 - 25	10	40 - 45	2

Solution: We first construct the cumulative frequency distribution table

Class Boundaries	Frequencies	Cumulative Frequencies
5 – 10	5	5
10 – 15	6	11
15 – 20	15	26
20 - 25	10	36
25 – 30	5	41
30 – 35	4	45
35 – 40	2	47
40 - 45	2	49

For computing the 1st Quartile, we note that N/4 = 49/4 = 12.25 which lies in the class 15 – 20. Hence, l = 15, F = 11, f = 15 and h = 5. Therefore, the 1st Quartile is given by

$$Q_1 = l + \frac{\frac{N}{4} - F}{f} * h \quad = 15 + ((12.25 - 11)/15)*5 = 15.41$$

For computing the 3rd Quartile, we have 3N/4 = 36.75, which lies in the class 25 – 30 and l = 25, F = 36, f = 5 and h = 5. Therefore, Q3 = 25 + ((36.75 – 36)*5)/5 = 25.75.

For the 4th Decile (D4), we have, 4*(N/10) = 4*49/10 = 19.6. This lies in the class 15 – 20. Thus l = 15, F = 11, f = 15 and h = 5. D4 = 15 + ((19.6 – 11)*5)/15 = 17.86

For the 60th Percentile (P60), we have 60(N/100) = 60 *49/100 = 29.4. This lies in the class 20 - 25. Here, l = 20, F = 26, h = 5, f = 10. There P60 = 20 +((29.4 – 26)/10)*5 = 20 + 1.7 = 21.7

9. Find the geometric mean of the following distribution:

Dividend declared (in %)	0 - 10	10 - 20	20 - 30	30 - 40
No of companies	5	8	3	4

Solution: We construct the following table for computing the geometric mean.

Class Boundaries	Class Marks (xi)	Frequencies (fi)	log10xi	fi *log10xi
0 -10	5	5	0.6990	3.4950
10 – 20	15	8	1.1761	9.4088
20 – 30	25	3	1.3979	4.1937
30 - 40	35	4	1,5441	6.1764
		Sum = 20		Sum = 23.2739

$$\log G = \frac{1}{N}\sum_{i=1}^{N} f_i \log_{10} x_i = \frac{23.2739}{20} = 1.1637$$

G = antilog(1.1637) = 14.58

Therefore, average dividend in terms of geometric mean is 14.58%.

10. Find the average rate of increase in prices which in the first year increased by 20%, in the next year by 25%, and in the third year by 44%.

Solution: log G = (log 20 + log 25 + log 44)/3 = (1.3010 + 1.3979 + 1.6434)/3 = 4.3423/3 = 1.4474. Therefore, G = antilog(1.4474) = 28.02. The average rate of increase in prices is 28.02%.

11. Find the arithmetic mean and the median of the frequency distribution of monthly wages of 100 laborers of a coal mine.

Wages (in Rs)	260 - 269	270 - 279	280 - 289	290 - 299	300 - 309	310 - 319	320 - 329
No of laborers	6	16	27	23	16	10	2

12. A man travelled by car for 3 days. He covered 480 kms each day. On the first day, he drove for 10 hours at 48 kms per hour, on the second day, he drove for 12 hours at 40 kms per hour and on the last day, he drove for 15 hours at 32 kms per hour. What was his average speed?

Solution: Since the distance travelled each day is the same, the appropriate average is the harmonic mean which is given by: $H.M. = \frac{n}{\sum \frac{1}{x_i}}$ = 3 /(1/48 + 1/40 + 1/32) = 3/37/480 =3*480/37 = 38.92. Hence, the average speed is 39 kmph (approx).

13. The average weekly wages of laborers in a Tea Garden is Rs 125. If the average wages paid to the male laborer is Rs 140 while that to a female laborer is Rs 100, find the ratio of male and female laborers employed in the Tea Garden.

Solution: Let n1 and n2 represent the male and female laborers employed in the Tea Garden. Now n1 + n2 = 100. Now mean(x1) = 140, Mean(x2) = 100 and Mean(combined) = 125.

$$\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

Therefore, 125 = [n1 *140 + (100 – n1)*100] * 100

Hence, n1 = 2500/40 = 62.5 and n2 = 37.5. Thus the ratio of male to female = 62.5/37.5 = 10:6.

14. A product is purchased at a price of Rs. 1.20, Rs. 1.40 and Rs. 1.50 per unit respectively for three months. If expenditure for this product for three months is Rs. 240, Rs. 259 and Rs. 270 respectively, calculate the average price at which the product is purchased.

Measures of Dispersion

1. Find the mean deviation about median for the following distribution:

Quantity demanded (in units)	6	12	18	24	30	36	42
Frequency	4	7	10	18	12	7	2

Solution: First we compute the median. The median is 24. So we will consider absolute deviations from 24.

xi	fi	xi - med	fi|xi – med|
6	4	-18	72
12	7	-12	84
18	10	-6	60
24	18	0	0
30	12	6	72
36	7	12	84
42	2	18	36
	Sum = 60		Sum = 408

Therefore, the mean deviation around the median = 408/60 = 6.8 (Note that mean deviation around median is the minimum).

2. In a survey of 50 chemical industries, following data was collected. Calculate the variance in 'level of profit'.

Level of profit (xi)	10	15	20	25	30
No of companies (fi)	15	10	15	6	4

Solution: We construct the following table for computing the variance in the data.

xi	fi	xi2	fixi	fixi2
10	15	100	150	1500
15	10	225	150	2250
20	15	400	300	6000
25	6	625	150	3750
30	4	900	120	3600
	Sum = 50		Sum = 870	Sum = 17 100

Now mean and the variance of x are computed as follow:

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i} = 870/50 = 17.4$$

$$\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \bar{X}^2 = 17000/50 - 17.4*17.4 = 37.24$$

3. A security analyst studied hundred companies and obtained the following data for the year 1983. Calculate the standard deviation of the dividend declared.

Dividend declared (%)	0 - 8	8 - 16	16 -24	24 - 32	32 - 40
No of companies	5	30	40	10	15

Solution: We know that value of SD does not change due to change in origin but changes due to change in scale. We assume c = 20 and d = 8 so that the value of the variable is changed from x to y = (x-c)/d. SD(y) = SD(x) / |d| .We construct the following table for this purpose.

Class Boundaries	Class marks (xi)	yi = (xi – c)/d	fi	fiyi	fiyi2
0 – 8	4	-2	5	-10	20
8 – 16	12	-1	30	-30	30
16 – 24	20	0	40	0	0
24 – 32	28	1	10	10	10
32 - 40	36	2	15	30	60
		Sum = 0	Sum = 100	0	120

Now, SD(y) = Sq rt[Sum(fiyi2)/fi) – Mean(y)*Mean(y)] = Sqrt(120/100 – 0) = Sqrt(120/100) = 1.10. Therefore, SD(X) = 8* SD(Y) = 8* 1.10 = 8.8.

Hence, the standard deviation of the dividend series is 8.8%.

4. The portfolio manager of an investment corporation proposed the following portfolio of securities. Calculate the average return and the standard deviation for the portfolio.

Security	Proportion of funds to be invested (zi)	Return (in percentage) (xi)
51	0.15	8
52	0.10	15
53	0.35	25
54	0.30	20
55	0.10	10

Solution: The average return (Mean(x)) = $\sum z_i x_i$ = 0.15*8 + 0.10*15 + 0.35*25 + 0.30*20 + 0.10*10 = 1.2 + 1.5 + 8.75 + 6 + 1 = 18.45.

Variance = $\sum z_i((x_i - \overline{x})^2$ = 0.15(8-18.45)^2 + 0.10(15 – 18.45)^2 + 0.35(25 – 18.45)^2 + 0.30(20 – 18.45)^2 + 0.10(10 – 18.45)^2 = 40.44

Therefore, the standard deviation = Sqrt(40.44) = 6.36. Thus the average return on the portfolio under consideration is 18.45% and the standard deviation is 6.36%.

5. The daily sales of two products, X and Y are give below. Find out which of the two shows greater fluctuations in sales.

   Daily sales for Product X: 620, 624, 622, 625, 622, 618, 619, 616, 623, 625, 626, 625.

   Daily sales for Product Y: 2152, 2134, 2132, 2145, 2132, 2142, 2146, 2130, 2146, 2142, 2150, 2135, 2152.

6. The following data is related to the orders booked by salesmen during a given period for two similar manufacturing divisions under the same management having large sales-force in its marketing department. Calculate the combined standard deviation for both the division and also the combined variability.

	Division 1	Division 2
No of Salesmen	25	18
Average number of orders booked	72	64
Variance of the distribution	8	6

7. The following table gives the distribution of income of household in large town. (i) What are the problems involved in calculating the standard deviation from the data below? (ii) Compute the quartile deviation and the coefficient of quartile deviation? (iii) Would the skewness of the data be affected if the income of everyone was increased by a certain proportion?

Income	% of households	Income	% of households
Under 800	1.5	1200 – 1299	16.2
800 – 899	6.2	1300 – 1399	15.6
900 – 999	9.6	1400 – 1499	10.4
1000 – 1099	12.3	1500 – 1599	9.5
1100 - 1199	14.8	1600 and above	3.9

8. Calculate the appropriate measure of dispersion from the following data on quantity of goods handled in metric ton by a local operator.

Quantity of goods handled (tons)	Less than 35	35 - 37	38 - 40	41 - 43	Over 44
No of days	14	62	99	18	7

9. Particulars regarding income of two villages are given below. (i) In which village is the variation is income greater? (ii) What is the total income of both the villages together? (iii) What is the

average income of households in both villages put together? (iv) What is the combined standard deviation and combined variability?

	Village	
	Naveli	Virpur
No of household	650	525
Average income per household per month	950	990
Variance of income	100	81

10. An industrial dispute arising out of wage disputes was referred to the industrial tribunal. The data related to number of employees, average wages, median wages and standard deviation is given below before the award and after the award. Comment on the benefit of award to the workers.

	Before the award	After the award
Number of employees	1560	1525
Mean wages per month	980	1030
Median wages per month	1040	965
Standard deviation	250	225

11. Find the missing information from the following data:

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

12. The arithmetic mean and the standard deviation of a series of 20 items were calculated by a student as 20 cm and 5 cm respectively. But while calculating them an item 13 was misread as 30. Find the correct mean and the standard deviation.
13. The Personnel Manager at the corporate headquarters of a large engineering firm wanted to know about the pattern of availing of leaves in the manufacturing units. The data related to leave taken by the staff is as follows. Calculate the average number of days of leaves taken by the staff members in each unit. Also calculate the average and standard deviation for both the units s taken together.

	Number of staff members	
	Manufacturing Unit I	Manufacturing Unit 2
Upto 5	20	18
5 – 10	22	
10 – 15	32	
15 – 20	36	
20 -25	30	
25 - 30	16	

Skewness, Kurtosis and Moments

1. Calculate the second, third, and fourth moments for the following distribution of service time at the registration counter of a local post office.

Service time (in minutes)	2.0	2.5	3.0	3.5	4.0	4.5
Number of frequencies	5	30	40	15	5	5

2. Calculate the coefficient of skewness for the following distribution and conclude on the type of skewness of the data.

Debt as % of total capitalization	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
No of companies	10	17	19	27	19	8

3. In a hotel having 70 rooms, data was collected for 104 days, for the number of rooms occupied on a day. This data was classified and is given below. Calculate the measures of skewness of the data.

No of rooms occupied	0 -10	10 - 20	20 -30	30 - 40	40 - 50	50 - 60	60 -70
No of days	10	12	18	25	16	15	8

4. Consider two distributions with the following summarized details. (i) Distributions A has the same degree of variation as the distribution B. Do you agree? Give reasons. (ii) Both the distributions have the same degree of skewness. TRU/FALSE. Give reasons.

Distribution	A	B
Mean	100	90
Median	90	80
Standard deviation	10	10

Correlation Analysis

1. From the following data on marks in Statistics and Accountancy, calculate Pearson's product moment correlation coefficient.

Marks in Stats	90	80	70	60	50	40	30	20	10
Marks in Acc.	75	80	70	65	55	60	50	40	45

2. Calculate the correlation coefficient for the following data:

X	700	600	500	100	200	300	400
Y	130	110	100	30	50	60	80

3. Data from 52 medium sized companies, belonging to engineering industry, regarding their annual gross sales (Rs in lacs) and the annual dividend (percentage) declared by them are given below. Find the degree of association (correlation coefficient) between annual sales (X) and the annual dividend (Y).

X \ Y	14 - 16	16 - 18	18 - 20	20 - 22	Total (fx)
150 – 250	2	1	1		4
250 – 350	3	2	3	2	10
350 – 450	3	4	5	6	18
450 – 550	2	2	3	4	11
550 – 650		1	2	2	5
650 - 750		1	2	1	4
Total (fy)	10	11	16	15	52

4. In a beauty competition, two judges ranked 12 participants as follows. What is the degree of agreement between the judgments of two judges?

Judge 1	3	4	1	5	2	10	6	9	8	7	12	11
Judge 2	6	10	12	3	9	2	5	8	7	4	1	11

5. In a selection committee meeting, two experts rank the 10 candidates in the following order. Is there any concurrence between the two judges?

Candidate #	1	2	3	4	5	6	7	8	9	10
Judge 1 rank	6	4	3	1	2	7	9	8	9	4
Judge 2 rank	1	1	6	7	5	8	10	9	4	3

6. Data given below indicate the ranks given to 15 branches of a nationalized bank on the basis of their deposit mobilization (DM) and recovery of loans (LR). Calculate the rank correlation and test its significance.

DM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LR	4	2	9	1	7	10	8	13	5	3	11	6	12	15	14

7. Calculate the coefficient of rank correlation coefficient from the following data.

X	48	33	40	9	16	16	65	24	16	57
Y	13	13	24	6	15	4	20	9	9	19

8. The coefficient of rank correlation as computed using Spearman's rank difference method, of the ranks obtained by 10 students in Statistics and Accountancy was found to be 0.8. Subsequently it was discovered that the difference in the ranks in the two subjects obtained by one of the students was wrongly recorded as 6 instead of 8. Calculate the correct coefficient of rank correlation.

Regression Analysis

1. Data related to batch size (X) and production cost (Y) is as follows. Find the regression equation of Y on X.

X	11	13	18	24	28	32	38	42	47	53
Y	2.1	2.7	2.9	2.9	3.1	3.0	3.3	3.7	4.0	4.4

2. The following table gives the age of cars of a certain make and annual maintenance costs. Obtain the regression equation for costs related to age. Estimate the maintenance cost for a ten year old car.

Age of cars (yrs)	2	4	6	8
Maint cost (Rs '000)	10	20	25	30

3. In a correlation study, the following values are obtained. Find the two regression equations that are associated with these values.
   Mean(X) = 65, Mean(Y) = 67, SD(X) = 1.5, SD(Y) = 3.5, Coefficient of correlation between X and Y = 0.8.
4. Data related to Yield on Security (Y) and Yield on Market Portfolio (X) is given below. Calculate standard error of estimate. Also compute the coefficient of determination.

Period	Yield on Security (Y)	Yield on Market Portfolio (X)
1	6.2	5.4
2	7.0	6.7
3	7.2	6.8
4	7.8	8.0
5	7.0	5.02
6	7.2	5.0
7	7.4	5.4
8	7.2	6.2
9	7.3	6.5
10	7,1	6.1

5. Given the following bivariate data: (i) Fit a regression equation of Y on X and hence predict Y if X = 8. (ii) Fit a regression equation of X on Y and hence predict X if Y = 3. (iii) Calculate the correlation coefficient between X and Y.

X	1	5	3	2	1	1	7	3
Y	6	1	0	0	1	2	1	5

6. The equations of two lines of regression obtained in a correlation analysis are the following: (i) $2x + 3y - 8 = 0$ and (ii) $x + 2y - 5 = 0$. Obtain the value of correlation coefficient and the variance of Y, given that the variance of X is 12.
7. In a production process, the operators performance ratings as given by the number of good items turned out per 100 units and their experience (in number of years) were compiled for ten randomly selected operators. Develop a least square regression line of performance ratings on experience and estimate the performance of an operator who has nine yeas' experience. Estimate the standard error.

Operator Experience (X)	Performance Ratings (Y)
8	80
6	76

16	87
12	88
18	89
4	68
3	78
10	80
5	75
12	83

8. In a crop production experiment conducted to study the relationship between yield per acre of a crop (Y) and the dose of NPK fertilizer (X), following results were obtained:
   Sample size (n) = 20, Mean(X) = 12.8, Mean(Y) = 130.7, Variance(X) = 70.6/20, Variance(Y) = 98.5/20, Covariance(X, Y) = 68.3/20.
   Assuming linear relationship: (i) develop the least squares regression lines , (ii) compute the standard error of estimate, (iii) estimate yield corresponding to X = 12.

Theory of Probability

1. What is the probability that all 3 children in a family have different birthdays? (Assume 1 year = 365 days)
2. A batch contains 10 articles of which 4 are defective. If 3 articles are chosen at random, what is the probability that none of them is defective?
3. 10 distinguishable balls are distributed at random into 4 boxes. What is the probability that a specified box contains exactly 2 balls?
4. A lady declares that by tasting a cup of tea with milk, she can discriminate whether milk or tea infusion was first added to the cup. In order to test this assertion, 10 cups of tea are prepared – 5 in one way, and 5 in the other, and presented to the lady for judgment in a random order. Assuming that the lady has no discrimination power, calculate the probability that she would judge correctly all the cups, it being known to her that 5 are of each kind.
5. A group of 2n boys and 2n girls is divided at random into two equal batches. Find the probability that each batch will be equally divided into boys and girls.
6. 10 dissimilar balls are distributed at random into 4 boxes marked A, B, C, D. Find the probability that these boxes contain respectively 2, 4, 4, o balls.
7. 10 identical balls are distributed at random into 4 boxes marked A, B, C, D. Find the probability that these boxes contain respectively 2, 4, 4, 0 balls.
8. What is the probability of obtaining a sum of 10 in a single throw with 5 dice?
9. Five different objects numbered 1,2,3,4,5 are placed at random into 5 places also marked 1,2,3,4,5. What is the probability that (i) no object occupies the place corresponding to its number; (ii) exactly 2 objects are in their correct places?
10. The odds in favor of an event A are 3:4. The odds against another independent event B are 7:4. What is the probability that at least one of the events will happen?
11. One urn contains 2 white and 2 black balls; a second urn contains 2 white and 4 black balls. (i) If one ball is chosen from each urn, what is the probability that they will be of the same color? (ii)

If an urn is selected at random and one ball is drawn from it, what is the probability that it will be a white ball?

12. There is a 50-50 chance that a contractor's firm A will bid for the construction of a multi-storied building. Another firm B submits a bid and the probability is ¾ that it will get the job, provided firm A does not bid. If firm A submits a bid, the probability that firm B will get the job is only 1/3. What is the probability that firm B will get the job?
13. Two players A and B toss a die alternatively. He who first throws a "six" wins the game. If A begins, what is the probability that he wins? What is the probability of B winning the game?
14. A bag contains 8 red and 5 white balls. Two successive draws of 3 balls are made without replacement. Find the probability that the first drawing will give 3 white balls and the second 3 red balls?
15. A coin is tossed 10 times. Find the probability of getting (i) exactly 6 heads, and (ii) 9 heads and 1 tail?
16. A machine produces on the average 2% defectives. If 4 articles are chosen randomly, find the probability that there are at least 2 defective articles.
17. Two boxes contain respectively 4 white and 2 black, and 1 white and 3 balck balls. One ball is transferred from the first box into the second, and then one ball is drawn from the latter. It turns out to be black. What is the probability that the transferred ball was white?
18. In a bolt manufacturing factory, machines M1, M2, M3 manufacture respectively 25, 35 and 40 per cent of the total product. Of their output 5, 4 and 2 per cent respectively are defective bolts. One bolt is drawn at random from the product and is found to be defective. What is the probability that it was manufactured by machine m3?
19. Two dice, with faces numbered 1 to 6, are thrown and their [points are added. The thrower is given Rs. 40 for a score of 12, but he has to pay Rs. 2 if the score is less than 12. Find the expectation per throw.
20. A coin is tossed repeatedly until a head appears. Find the expected number of tosses required to obtain the first head.
21. The probability density function (pdf) of a continuous random variable is given by: $y = k(x – 1)(2 – x)$ where $1 <= x <= 2$. Determine: (i) the value of the constant k, (ii) the cumulative distribution function, (iii) the probability that x is less than 5/4, (iv) the probability that x is greater than 3/2, (v) the probability that x lies between 5/4 and 3/2.
22. Four unbiased coins are tossed. If X and Y denote respectively the "number of heads" and "longest run of heads", construct the joint distribution of x and y.

Probability Distributions: Binomial, Poisson and Normal

1. A random variable x is defined as follows: Prob (x=1) = p, Prob(x=0) = 1- p, where $0<p<1$. Find the mean, variance and the central moments $\mu 2$, $\mu 3$ and $\mu 4$ of the distribution.
2. If a discrete random variable x follows uniform distribution and assumes only the values 8, 9, 11, 15, 18, 20, find the probabilities: (i) P(x=9), (ii) P(x=12), (iii) P(x<15), (iv) P(x<= 15), (v) P(x>15), (vi) P(|x-4| <5).

3. Five coins are tossed 3200 times. Find the expected frequencies of the distribution of heads and tails, and tabulate the result.
4. The overall percentage of failures in a certain examination is 40. What is the probability that out of a group of 6 candidates at least 4 passed the examination?
5. In 10 independent throws of a defective die, the probability that an even number will appear 5 times is twice the probability that an even number will appear 4 times. Find the probability that an even number will not appear at all in 10 independent throws of the die.
6. A business firm receives on an average 2.5 telephone calls per day during the time period 10.00 – 10.05 AM. Find the probability that on a certain day, the firm receives (i) no call, (ii) exactly 4 calls, during the same period. (Assume Poisson distribution; given exp(-2.5) = 0.0821)
7. In turning out certain toys in a manufacturing process in a factory, the average number of defectives is 10%. What is the probability of getting exactly 3 defectives in a sample of 10 toys chosen at random, by using the Poisson approximation to the Binomial distribution? (Take e = 2.72).
8. Two defective tube-lights are mixed with 8 non-defective tube-lights by mistake. A sample of 3 tubes is taken at random from the lot and tested for the number of defectives. Find the probability distribution of the 'number of defective tube-lights in the sample' and tabulate the probabilities. What are the mean and standard deviation of the number of defectives obtained in the sample?
9. A box contains 4 white, 3 black and 5 red balls. A ball is drawn from the box at random, its color noted, and then the ball is replaced. If 6 balls are drawn in this manner (i.e., with replacement), find the probability that (i) 3 are white, 1 black and 2 red, (ii) 2 are white and 4 red.
10. The following table gives the joint distribution of x and y. Answer the following questions: (i) How many pairs of values of x and y are possible? Write them down and show the probability of each pair separately. (ii) Show the marginal distributions of x and y. (iii) Show the conditional distribution of x, given y = 2 (iv) Show the conditional distribution of y, given x = 3. (v) Find the probabilities P(x<y), P(2x+y>=9).

x \ y	2	3	7	Total
1	0.10	0.25	0.05	0.40
3	0.30	0.15	0.15	0.60
Total	0.40	0.40	0.20	1.00

11. The following table gives the joint distribution of x and y. (i) Are x and y independent? (ii) Determine the correlation coefficient between x and y.

y / x	0	1	2
1	0.3	0.2	0.1
2	0.1	0.0	0,3

12. If a continuous random variable x follows rectangular distribution in the range (2, 7), find the probabilities: (i) $P(2.5 \leq x \leq 4)$, (ii) $P(x \leq 3.4)$, (iii) $P(x > 6)$, (iv) $P(x = 4.5)$.

13. The height distribution of a group of 10,000 men is normal with mean height 64.5" and SD 4.5". Find the number of men whose height is (a) less than 69" but greater than 55.5", (b) less than 55.5", and (c) more than 73.5".
14. The mean height of 500 males students at a certain college is 151 lbs and the standard deviation is 15 lbs. Assuming that the weights are normally distributed, find how many students weigh (i) between 120 and 155 lbs, (ii) more than 155 lbs (Given $\Phi'(0.27) = 0.6064$ and $\Phi'(2.07) = 0.9808$, where $\Phi'(t)$ denotes the area under standard normal curve to the left of the ordinate at t.)
15. A sample of 100 dry battery cells tested to find the length of life produced the following results: Mean(x) = 12 hours, SD(x) = 3 hours. Assuming that the data are normally distributed, what percentage of battery cells are expected to have life (i) more than 15 hours, (ii) less than 6 hours, and (iii) between 10 and 14 hours?
16. A fair coin is tossed 400 times. Using Normal approximation, find the probability of obtaining (i) exactly 200 heads, (ii) less than 200 heads, (iii) between 190 and 210 heads, both inclusive. Given that the area under standard Normal curve

between z = 0 and z = 0.05 is 0.0199
between z = 0 and z = 0.95 is 0.3289
between z = 0 and z = 1.05 is 0.3531

Sampling Theory

1. A simple random sample of size 5 is drawn without replacement from a finite population consisting of 41 units. If the population standard deviation is 6.25, what is the standard error of sample mean? (use finite population correction).
2. The safety limit of a crane is known to be 32 tons. The mean weight and the standard deviation of a large number of iron rods are 0.3 ton and 0.2 ton respectively. 100 rods are lifted at a time. Find the probability of an accident.
3. It has been found that 2% of the tools produced by a certain machines are defective. What is the probability that a shipment of 400 such tools, 3% or more will be defective? (Probability that the normal deviate lies between 0 and 1.43 is 0.4236).
4. For a population of six units, the values of a characteristics x are given as follows: 3, 9, 6, 5, 7, 10. Consider all possible samples of size two from the above population and show that the mean of the sample means is exactly equal to the population mean.
5. A population consists of the four members 3, 7, 11, 15. Consider all possible samples of size two which can be drawn with replacement from this population. Find (i) the population mean, (ii) the population standard deviation, (iii) the mean of the sampling distribution of means, (iv) the standard deviation of the sampling distribution of means. Verify (iii) and (iv) directly from (i) and (ii) by use of suitable formulae (which you have to mention).

Theory of Estimation and Test of Significance

1. A sample of 600 screws is taken from a large consignment and 75 are found to be defective. Estimate the percentage of defectives in the consignment and assign limits within which the percentage lies.

2. A random sample of 100 ball bearings selected from a shipment of 2000 ball bearings has an average diameter of 0.354 inch with an SD = 0.048 inch. Find 95% confidence interval for the average diameter of these 2000 ball bearings.
3. A random sample of 100 articles taken from a large batch of articles contains 5 defective articles. (a) Set up 96 per cent confidence limits for the proportion of defective articles in the batch. (b) If the batch contains 2696 articles, set up 95% confidence interval for the proportion of defective articles.
4. A random sample of size 10 was drawn from a normal population with an unknown mean and a variance of 44.1 $(\text{inch})^2$. If the observations are (in inches): 65, 71, 80, 76, 78, 82, 68, 72, 65 and 81, obtain the 95% confidence interval for the population mean.
5. A random sample of 10 students of class II was selected in a certain region, and their weights recorded are shown below (in lbs): 38, 46, 45, 40, 35, 39, 44, 45, 33, 37. Find 95% confidence limits within which the mean weight of all such students in the region is expected to lie. (Given $t_{.025} = 2.262$ for 9 df and 2.228 for 10 df).
6. The standard deviation of a random sample of size 12 drawn from a normal population is 5.5. Calculate the 95% confidence limits for the standard deviation (σ) in the population (Given $\chi^2_{.975} = 3.82$ and $\chi^2_{.025} = 21.92$ for 11 degrees of freedom).
7. A sample of size 8 from a normal population yields as the unbiased estimate of population variance the value 4.4. Obtain the 90% confidence limits for the σ^2 (Given $\chi^2_{.995} = 0.99$ and $\chi^2_{.005} = 20.3$ for 7 df).

Theory of Test of Significance

1. The fraction of defective items in a large lot is P. To test Null Hypothesis H_0 : P =0.2, one considers the number f of defectives in a sample of 8 items and accepts the hypothesis if $f \leq 6$ and rejects the hypothesis otherwise. What is the probability of Type I Error of this test? What is the probability of Type II Error corresponding to P = 0.1?
2. A die was thrown 400 times and 'six' resulted in 80 times. Do this data justify the hypothesis of an unbiased die?
3. In a sample of 400 parts manufactured by a factory, the number of defective parts was found to be 30. The company, however, claimed that only 5% of their product is defective.
4. A manufacturer claimed that at least 90% of the components which he supplied conformed to specifications. A random sample of 200 components showed that only 164 were up to the standard. Test his claim at 1% level of significance.
5. In a large city A, 20 per cent of a random sample of 900 school children had defective eyesight. In another large city B, 15 per cent of a random sample of 1600 children had the same defect. Is this difference between the two proportions significant? Obtain 90% confidence limits for the difference in the population proportions.
6. Before an increase in excise duty on tea, 400 people out of a sample of 500 persons were found to be tea drinkers. After an increase in duty, 400 people were tea drinkers in a sample of 600 people. Using standard error of proportion, state whether there is a significant decrease in the consumption of tea.

7. In a certain city 100 men in a sample of 400 were found to be smokers. In another city, the number of smokers was 300 in a random sample of 800. Does this indicate that there is a greater proportion of smokers in the second city than in the first?
8. Is it likely that a sample of 300 items whose mean is 16.0 is a random sample from a large population whose mean is 16.8 and standard deviation 5.2?
9. An automatic machine was designed to pack exactly 2.0 kg of Vanaspati. A sample of 100 tins was examined to test the machine. The average weight was found to be 1.94 kg, with standard deviation 0.10 kg. Is the machine working properly?
10. A machine part was designed to withstand an average pressure of 120 units. A random sample of 100 from a large batch was tested and it was found that the average pressure which these parts could withstand is 105 units with a standard deviation of 20 units. Test whether the batch meets the specification.
11. In a certain factory there are two different processes of manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 gm with an SD of 12 gm; the corresponding figures in the sample of 400 items from the other process are 124 and 14. Compute the Standard Error (SE) of the difference between the two sample means. Is this difference significant? Also find the 95% confidence limits for the difference in the average weights of items produced by the two processes.
12. The mean yield of wheat from a district A was 210 lbs with SD = 10 lbs per acre from a sample of 100 plots. In another district B, the mean yield was 220 lbs with SD = 12 lbs from a sample of 150 plots. Assuming that the standard deviation of yield in the entire state was 11 lbs, test whether there is any significant difference between the mean yields of crops in the two districts.
13. In his experiments on the pea breeding, Mendel obtained the following frequencies of seeds: Bound and Yellow – 315; Wrinkled and Yellow – 101; Round and Green – 108; Wrinkled and Green – 32; Total – 556. Theory predicts that the frequencies should be in the proportions-(:3:3:1. Examine the correspondence between theory and observations. (Given that 5% value of chi-square or 3 df is 7.815).
14. A die was thrown 60 times with the following results:

Face	1	2	3	4	5	6	Total
Frequency	6	10	8	13	11	12	60

Are the data consistent with the hypothesis that the die is unbiased? (Given $\chi^2_{.01}$ = 15.09 for 5 degrees of freedom).

15. A set of 5 identical coins is tossed 320 times, and the number of heads appearing each time is recorded. The results are as follows. Would you conclude that the coins are biased? (Given $\chi^2_{.05}$ = 11.07 and $\chi^2_{.01}$ = 15.09 for 5 degrees of freedom).

No of Heads	0	1	2	3	4	5	Total
Frequency	14	45	80	112	61	8	320

16. A random sample of 500 students were classified according to economic condition of their family and also according to merit as shown below:. Test whether the two attributes Merit and Economic Condition are associated or not. (Given $\chi^2_{.05}$ = 5.99 and $\chi^2_{.01}$ = 9.21 for 2 df).

Merit	Economic Condition			Total
	Rich	Middle Calss	Poor	

Meritorious	42	137	61	240
Not meritorious	58	113	89	260
	100	250	150	500

17. The following data show the type of leaves and the color of flowers from a random sample of 160 plants. Test whether the color of flowers is dependent on the type of leaves. (Given that 5% value of χ^2 for 1 degree of freedom is 3.841)

	Flat Leaves	Curled Leaves	Total
White Flowers	99	36	135
Red Flowers	20	5	25
Total	119	41	160

18. A random sample of size 10 was taken from a normal population, whose variance is known to be 7.056 sq inches. If the observations are (in inches) 65, 71, 64, 71, 70. 69. 64, 63, 67 and 68, test the hypothesis that the population mean is 69 inches. Also obtain 99% confidence limits for the population mean.
19. It is required to estimate the mean of a normal population using a sample sufficiently large so that the probability will be 0.95 that the sample mean will not differ from the population mean by more than 25% of the population SD. How large should the sample be?
20. Two independent random samples were taken from two normal populations and the following information is given: below. Is it likely that the mean of Population I is smaller than that of Population II? (Use 1 % level of significance). Also find 99% confidence limits for the difference of population means.

	Population I	Population II
Sample Size	n1 = 10	n2 = 12
Sample Mean	20	27
Population SD	8	6

21. A random sample of size 20 from a normal population gives a sample mean of 42 and sample standard deviation of 6. Test the hypothesis that the population mean is 44. State clearly the alternative hypothesis you allow for and the level of significance adopted.
22. A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozens. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective?
23. In a certain test, one section of 10 students had an average score of 40 with standard deviation 5. Another section of 12 students had an average of 46 with standard deviation 4. Examine whether there is a significant difference in the average scores of the two groups. What assumption do you make in applying the test? (You may use the information that 5% value of t for 20 df is 2.086).
24. A group of 5 patients treated with medicine A weigh 42, 39, 48, 60, 41 kg; a second group of 7 patients from the same hospital treated with medicine B weigh 38, 42, 56, 64, 68. 69, 62 kg. Do

you agree with the claim that medicine B increases the weight significantly? (The value of 't' at 5% level of significance for 10 df is 2.2281).

25. A random sample of size 20 from a normal population gives a sample mean of 42 and a sample standard deviation of 6. Test the hypothesis that the population SD is 9. Clearly state the alternative hypothesis you allow for and the level of significance adopted.
26. Weights (in kg) of 10 students are as follows: 38, 40, 45, 53, 47, 43, 55, 48, 52, 49. Can we say that variance of the distribution of weights of all students from which the above sample of 10 students were drawn, is equal to 20 square kg? (Given that for DF 9, Chi-square.05 = 16.92 and Chi-square.01= 21.67. For DF 10, Chi-square.05 = 18.31 and Chi-square.01 = 23.21)
27. The standard deviation calculated from two random samples of sizes 9 and 13 are 2.1 and 1.8 respectively. May the samples be regarded as drawn from Normal population with the same SD? (The 5% value of F from tables with df 8 and 12 is F.05 = 2.85).
28. In a sample of 8 observations, the sum of the squared deviations of items from the mean was 94.5. In another sample of 10 observations, the value was found to be 101.7. Test whether the difference is significant at 5% level. (You are given that at 5% level, critical value of F for v1 = 7 and v2 = 9 degrees of freedom is 3.29 and for v1 = 8 and v2 = 10 df its value is 3.07).