SOLVED PROBLEMS INTRODUCTION TO STATISTICAL THEORY

PART I

(Prof. Sher Muhammad Chaudhary

Dr. Shahid Kamal)

CONTENTS

Chapter 1

INTRODUCTION

1.9 (b)(i) The number of students attending a class is a *discrete* variable.

(ii) The amount of milk produced by a cow is a *continuous* variable.

(iii) The number of heads in the toss of 6 coins is a *discrete* variable.

(iv) The yearly income of a College Professor is a *discrete* variable.

(v) The age of a shopkeeper is a *continuous* variable.

(vi) The weight of a college student is a *continuous* variable.

(vii) The number of petals on a flower is a *discrete* variable.

(viii) The life time of television tubes produced by a company is a *continuous* variable.

(ix) Temperature recorded every half hour at a weather bureau is a *continuous* variable.

(x) The number of shares sold each day in the stock market is a *discrete* variable.

1.10 (i) Qualitative, (ii) Quantitative,
(iii) Qualitative, (iv) Quantitative,
(v) Quantitative, (vi) Quantitative,
(vii) Qualitative.

1.11 (i) Ratio, (ii) Ordinal,
(iii) Interval, (iv) Ratio,
(v) Nominal, (vi) Ratio,
(vii) Ordinal, (viii) Ratio,
(ix) Ordinal, (x) Ratio.

1.12 (i) 32.21705 rounded off to four significant digits becomes 32.22.

(ii) 937.05002 rounded off to four significant digits becomes 937.1. We increase zero preceding 5 by 1 as 5 is followed by a non-zero digit.

(iii) 0.003599499 rounded to four significant digits becomes 0.003599 because beginning zeros are not significant but they serve only to locate the decimal point.

(iv) 1.003599499 rounded to four significant digits becomes 1.004.

(v) 0.07000455 rounded to four significant digits becomes 0.07000.

(vi) 22.2500001 rounded to four significant digits becomes 22.26.

Chapter 2

PRESENTATION OF DATA

2.4 Presentation of the data given in the paragraph in the form of a Table.

DISTRIBUTION OF EMPLOYEES OF THE JOHN SMITH MANUFACTURING COMPANY BY SEX AND MEMBERSHIP DURING 1941-44.

	All			Union			Non-Union		
Year	Total	Male	Female	Total	Male	Female	Total	Male	Female
1941	1650	1430	220	1250	1170	80	400	260	140
1942	1725	1500	225	1475	1300	175	250	200	50
1943	1750	1500	250	1700	1460	240	50	40	10
1944	2000	1700	300	1980	1685	295	20	15	5

Source: Census of Manufacturers Report, 1945.

2.5 (b) Determination of class-boundaries, class-limits, etc.

(i) Here the smallest weight = 98 lb, the largest weight = 226 lb.

$\therefore$ Range = 226 − 98 = 128 and n = 300.

Let us take h (class-interval) = 10 lb and the lower limit of the first class as 95 lb. The last class is to include the highest value of 226 lb. The required values are:

Classes	Class-limits	Class-boundaries	Class-mark
First	95 – 104	94.5 – 104.5	99.5
⋮	⋮	⋮	⋮
Last	225 – 234	224.5 – 234.5	229.5

(ii) The smallest observation = 0.421 and the largest observation = 0.563, so that range = 0.563 − 0.421 = 0.142. n = 460. Let us decide to have about 8 classes. Then $h = \frac{0.142}{8} = 0.020$ approximately and we may take the lower limit of the lowest class as 0.420.

The desired figures are given below:

Classes	Class-limits	Class-boundaries	Class-mark
First	0.420–0.439	0.4195–0.4395	0.4295
⋮	⋮	⋮	⋮
Last	0.560 – 0.579	0.5595–0.5795	0.5695

(c) The class-boundaries, the class-limits and class marks are given below:

Class boundaries	Class Limits	Class Marks
199.5 – 219.5	200 – 219	209.5
219.5 – 239.5	220 – 239	229.5
239.5 – 259.5	240 – 259	249.5
259.5 – 279.5	260 – 279	269.5
279.5 – 299.5	280 – 299	289.5
299.5 – 319.5	300 – 319	309.5
319.5 – 339.5	320 – 339	329.5

2.6 (b) Preparation of a Frequency Table.

Classes	Tally	Frequency
35 – 39	II	2
40 – 44	~~IIII~~ I	6
45 – 49	~~IIII~~ I	6
50 – 54	~~IIII~~ III	8
55 – 59	~~IIII~~ III	8
60 – 64	I	1
65 – 69	I	1
70 – 74	III	3
75 – 79	IIII	4
80 – 84	IIII	4
85 – 89	III	3
90 – 94	II	2
95 – 99	I	1
100 – 104	I	1
Total	--	50

Here the smallest value = 36, and the largest value = 100.

Range = 100 – 36 = 64.

The width of class-interval is given equal to 5 units. As the smallest value is 36, we may therefore take 35 (a multiple of 5) as the lower class limit of the lowest class. The frequency table is then constructed as above.

2.7 (b).

Absentees (x)	No. of days (f)	fx
0	5	0
1	7	7
2	9	18
3	6	18
4	4	16
5	2	10
6	1	6
7	1	7
Total	---	82

(i) No. of days on which fewer than 4 people were absent

$$= 5 + 7 + 9 + 6 = 27 \text{ days}$$

(ii) No. of days on which *atleast* 4 people were absent

$$= 4 + 2 + 1 + 1 = 8 \text{ days}$$

(iii) Total number of absences over the whole 35 days $= \sum fx = 82$

2.8. Preparation of the Frequency Distribution.

The lowest marks = 49, highest marks = 121.

$$\therefore \quad \text{Range} = 121 - 49 = 72.$$

Let us take 10 marks as class-interval, *i.e.*, $h = 10$, and place the lower class-limit of the lowest class or group at 40. Then the frequency distribution is constructed as follows:

Frequency Distribution of Marks of 60 Students

Marks	Tally	Frequency
40 – 49	I	1
50 – 59	卌 IIII	9
60 – 69	卌 IIII	9
70 – 79	卌 II	7
80 – 89	卌 卌	10
90 – 99	卌 卌 II	12
100 – 109	卌 I	6
110 – 119	IIII	4
120 – 129	II	2
Total	---	60

2.9 Construction of the frequency distribution

Here the smallest value=61, the largest value=153, so that the range=153–61=92. Class-interval $(h)=5$, (given). Locating the lower class limit of the first group at 60, the frequency distribution is formed as below:

Classes	Tally	Frequency
60 – 64	II	2
65 – 69	I	1
70 – 74	II	2
75 – 79	卌 II	7
80 – 84	II	2
85 – 89	卌 II	7
90 – 94	IIII	4
95 – 99	卌 IIII	9
100 – 104	卌 卌 I	11
105 – 109	卌 卌	10
110 – 114	卌 卌	10
115 – 119	卌 IIII	9
120 – 124	卌	5
125 – 129	卌 II	7
130 – 134	II	2
135 – 139	III	3
140 – 144	IIII	4
145 – 149	III	3
150 – 154	II	2
Total	----	100

2.10 (i) Arrangement of the data in an array.

48.6, 55.9, 58.3, 59.4, 63.9, 64.2, 65.7, 67.6, 68.9, 69.1, 70.8, 71.6, 71.6, 72.1, 73.0, 73.8, 74.2, 74.2, 75.2, 77.6, 77.8, 79.4, 80.7, 81.8, 81.9, 82.7, 82.9, 83.2, 83.5, 88.1, 90.6, 95.5.

(ii) Construction of a frequency distribution using a class-interval of 5.00.

Class-limits	Class-boundaries	Tally	f
45.5 – 50.4	45.45 – 50.45	I	1
50.5 – 55.4	50.45 – 55.45	--	0
55.5 – 60.4	55.45 – 60.45	III	3
60.5 – 65.4	60.45 – 65.45	II	2
65.5 – 70.4	65.45 – 70.45	IIII	4
70.5 – 75.4	70.45 – 75.45	卌 IIII	9
75.5 – 80.4	75.45 – 80.45	III	3
80.5 – 85.4	80.45 – 85.45	卌 II	7
85.5 – 90.4	85.45 – 90.45	I	1
90.5 – 95.4	90.45 – 95.45	I	1
95.5 – 100.4	95.45 – 100.45	I	1
Total	---	--	32

2.11 As the data are discrete, therefore the ungrouped frequency distribution is prepared as below:

No. of children (x)	Tally	No. of women (f)
0	I	1
1	IIII	4
2	卌 III	8
3	卌 卌 IIII	14
4	卌 II	7
5	卌	5
6	IIII	4
7	III	3
8	II	2
9	I	1
10	I	1
Total	--	50

2.12 Total number of letters in each word are counted as below: 2, 7, 2, 6, 2, 5, 10, 2, 2, 1, 4, 2, 8, 2, 2, 6, 1, 4, 2, 8, 2, 2, 4, 2, 5, 7, 2, 2, 7, 4, 2, 10, 3, 4, 4. 2, 3, 2, 9, 3, 2, 5, 1, 6, 9, 2, 8, 5, 7, 8, 3, 3, 3, 8, 2, 6, 6, 7, 2, 2, 3, 8, 2, 3, 3, 3, 7, 3, 3, 4, 3, 9, 2, 5, 11, 3, 4, 4, 1, 3, 4, 1, 6, 2, 5, 2, 3, 7, 4, 2, 7

The desired frequency distribution of word-length is as follows:

Word-Length (x)	Tally	Frequency (f)
1	卌	5
2	卌 卌 卌 卌 卌 II	27
3	卌 卌 卌 I	16
4	卌 卌 I	11
5	卌 I	6
6	卌 I	6
7	卌 III	8
8	卌 I	6
9	III	3
10	II	2
11	I	1
Total	--	91

2.13 Taking the last digit of the numbers as the *leaf* and the rest of the digits as the *stem*, we get the following stem-and-leaf display:

Stem	Leaf (ordered)
19	3
20	2 8
21	2 7 8 9
22	4 5 8
23	0 1 1 4 5 6 6 8
24	0 3 3 5 5 5 6 7 9
25	0 1 2 4 5 5 5 7 8 9 9
26	3 5 8 8 9
27	1 5 7
28	0 3 4 8

Convertion of the *stem-and-leaf* display into a frequency distribution, beginning with 190.

Weight	Frequency
190 – 199	1
200 – 209	2
210 – 219	4
220 – 229	3
230 – 239	8
240 – 249	9
250 – 259	11
260 – 269	5
270 – 279	3
280 – 289	4

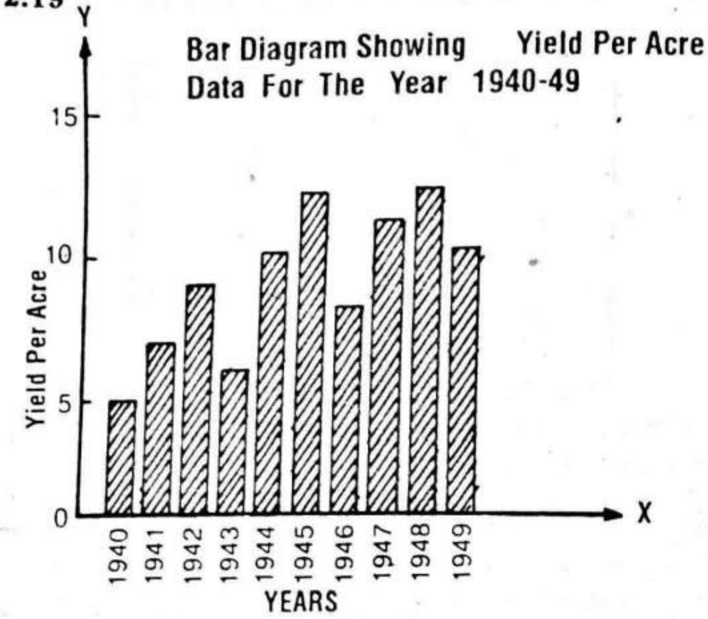
2.14 Using the whole number as the stem and the decimal as the leaf, we get the following *stem-and-leaf display*:

Stem	Leaf (decimals)	Leaf (ordered)
8	0 3 7	0 3 7
9	0 3 7 9 8 1 3 6 9	0 1 3 3 6 7 8 9 9
10	2 7 8 1 0 5 5 6 4 9 6	0 1 2 4 5 5 6 6 7 8 9
11	3 6 0 7 0 5 6 5 2 7 8 8 9 5 8	0 0 2 3 5 5 5 6 6 7 7 8 8 8 9
12	1 0 3 9 5 8 6 6	0 1 3 5 6 6 8 9
13	8 6 7 4 0 2 9 4	0 2 4 4 6 7 8 9
14	1 0 2 7 7 9	0 1 2 7 7 9
15	8 7 1 9 7	1 7 7 8 9
16	4 9 8	4 8 9
17	7 9	7 9

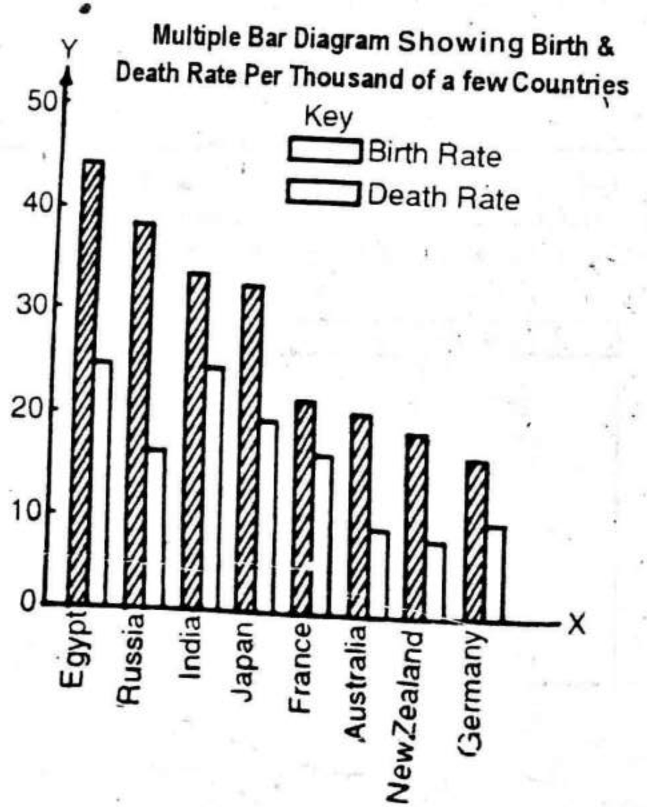
Now the data are very easily converted into a grouped frequency distribution with $h=1$ unit and using 8.0 as the lower limit of the first class. The grouped frequency distribution follows:

Class-limits	Frequency
8.0 – 8.9	3
9.0 – 9.9	9
10.0 – 10.9	11
11.0 – 11.9	15
12.0 – 12.9	8
13.0 – 13.9	8
14.0 – 14.9	6
15.0 – 15.9	5
16.0 – 16.9	3
17.0 – 17.9	2

2.19

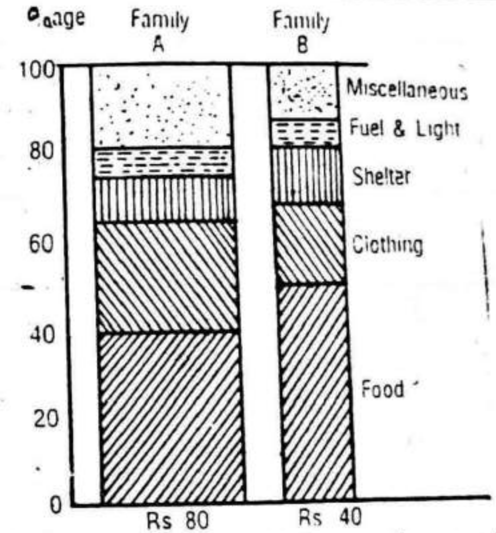


2.20

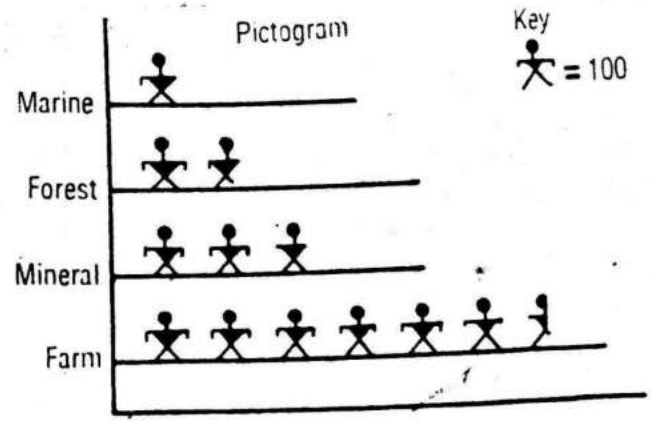


2.21 Representation of the data by rectangular diagram.
Family Budgets of two Families

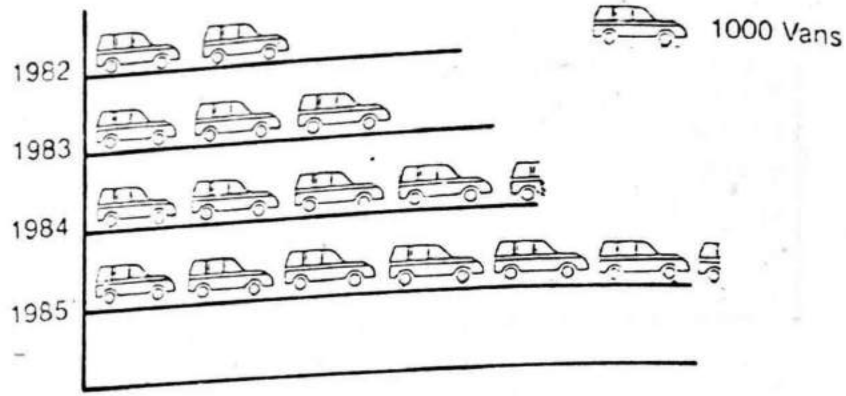
Items of Expenditure	Family *A* (Income Rs. 80)		Family *B* (Income Rs. 40)	
	Actual Expenses	Percentage Expenses	Actual Expenses	Percentage Expenses
Food	Rs. 32	40	Rs. 20	50
Clothing	Rs. 20	25	Rs. 8	20
Shelter	Rs. 8	10	Rs. 4	10
Fuel and Light	Rs. 4	5	Rs. 2	5
Miscellaneous	Rs. 16	20	Rs. 6	15
Total	Rs. 80	100	Rs. 40	100



2.23 (a) Representing 100 employees by one picture, the pictogram is drawn below:



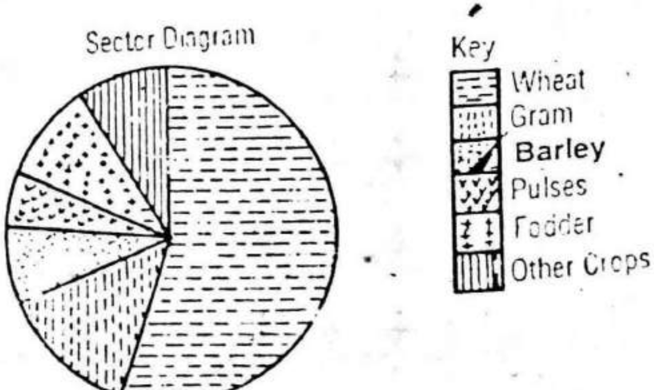
(b) Representing 1000 vans by one symbol, the pictogram is drawn below:



2.25 (a) **Preparation of Pie-chart. The corresponding angles needed to draw the diagram are computed below:**

Crop	Area	Angles of the Sectors (degree)	%age
Wheat	106	° 190.8	53
Gram	30	54	15
Barley	15	27	7.5
Pulses	10	18	5
Fodder	25	25	12.5
Other Crops	14	25.2	7
Total	200	360 °	100

(b) The per cent contribution of each crop to the total Rabi crops appears in the last column of table in (a).

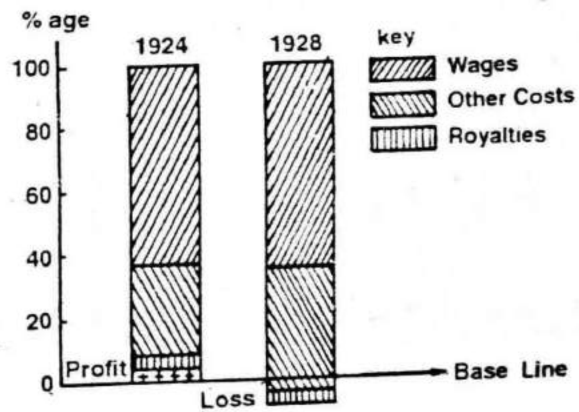


2.26 Representation of the data by (i) Percentage sub-divided Bars, (ii) a Pie-diagram.

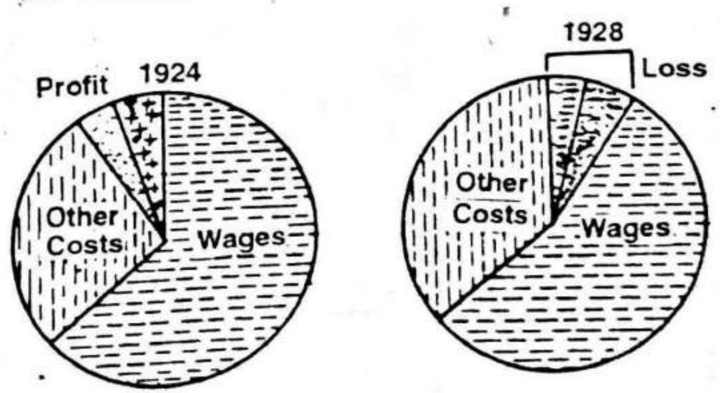
Cost per ton Disposed Commercially

Particulars	1924 Cost	1924 %age	1924 Degrees	1928 Cost	1928 %age	1928 Degrees
Wages	12.74	64	230	7.95	65	235
Other Costs	5.46	27	99	4.51	37	134
Royalties	0.56	3	10	0.50	4	15
Total	18.76			12.96		
Sale proceeds	19.91	100	360	12.16	100	360
Profit (+) or Loss (−)	+1.15	6	+ 21	−0.80	6	−24

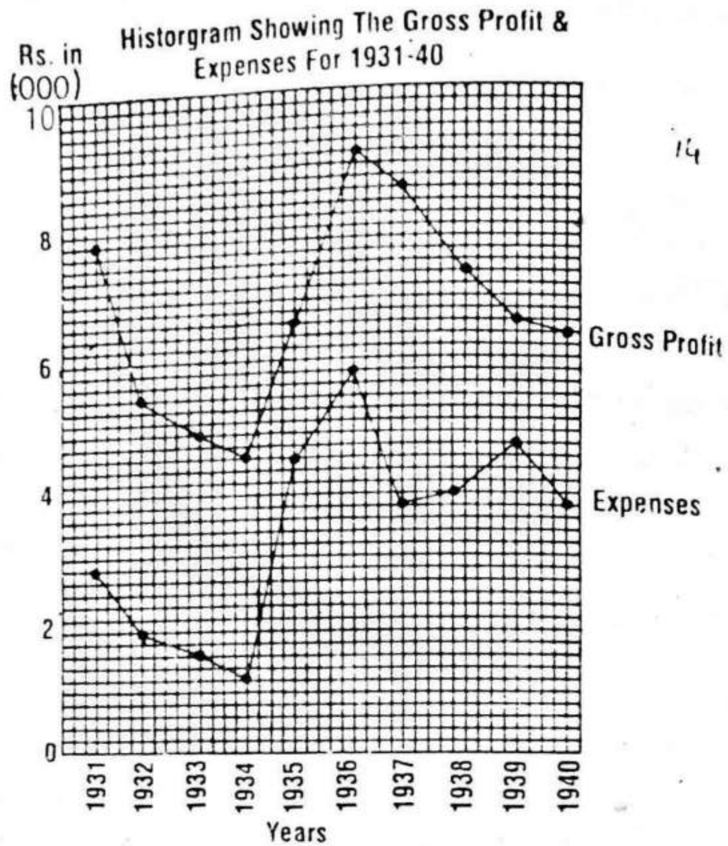
(i) Profit and Loss Chart (% age sub-divided Bars)



(ii) Profit and Loss Chart (Pie-diagram)



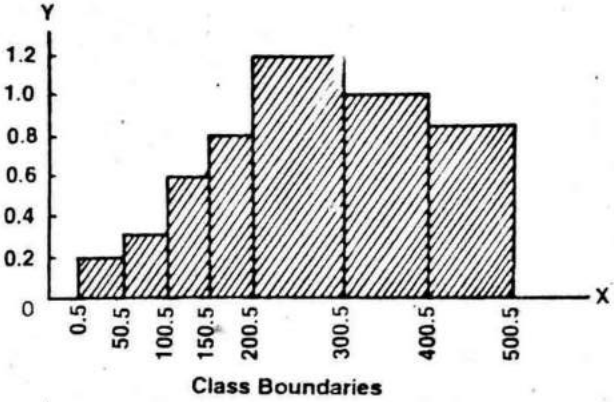
2.28.



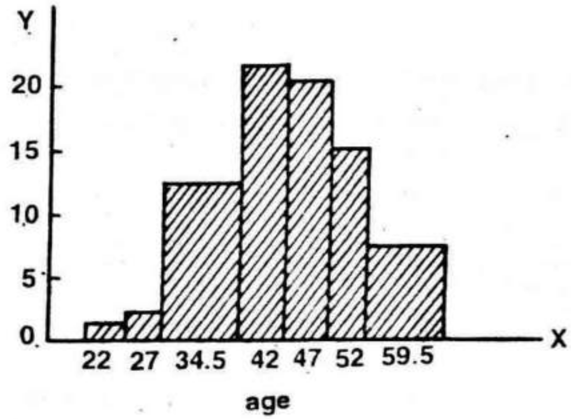
2.32 (c) Drawing of Histogram for unequal class-intervals.

Class-boundaries	Class-Interval	Frequency	Proportional heights
0.5 – 50.5	50	10	0.2
50.5 – 100.5	50	15	0.3
100.5 – 150.5	50	30	0.6
150.5 – 200.5	50	40	0.8
200.5 – 300.5	100	120	1.2
300.5 – 400.5	100	140	1.0
400.5 – 500.5	100	85	0.85

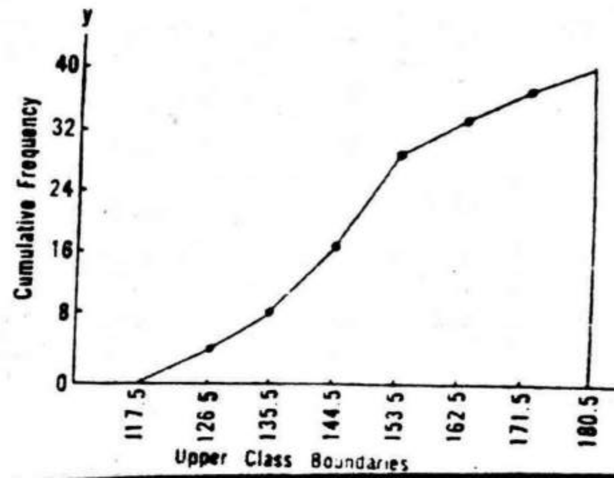
Historgram for the Distribution of Savings Cert



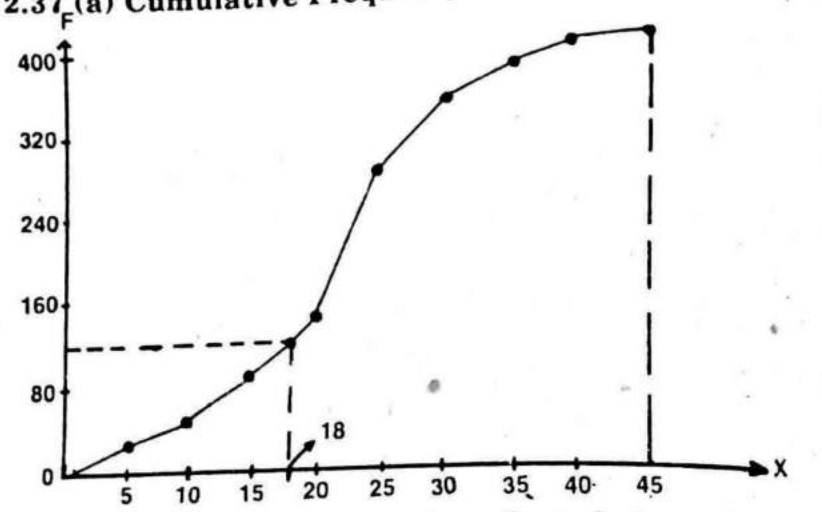
2.34 (a) Histogram illustrating the Age-distribution.



2.36 Ogive for the Frequency Distribution of Weights.



2.37 (a) Cumulative Frequency Curve is drawn below.



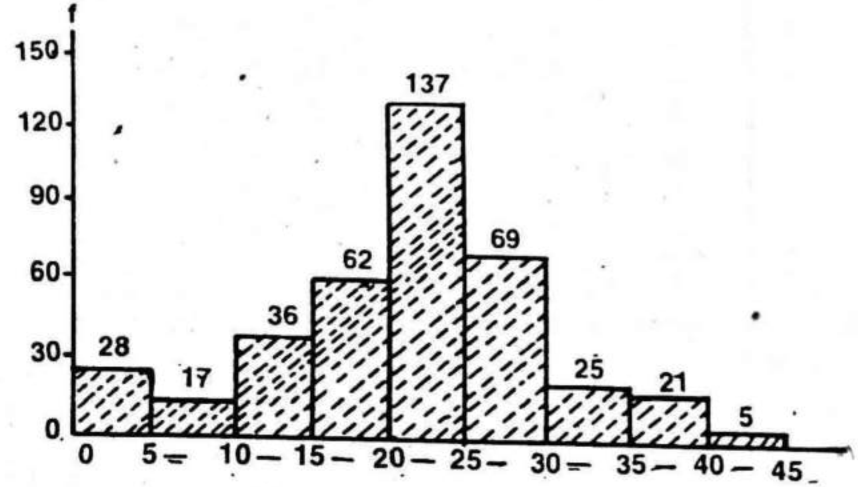
From graph, we estimate that 114 pupils took less than 18 minutes.

(b) 6% of the pupils took x minutes or longer means that 24 pupils took x minutes or longer and $(400-24)=376$ pupils took less than x minutes. From graph $x=36$. Thus 6% of the pupils took 36 minutes or longer.

(c)

ucb	F	Time (min)	Frequency
5	28	0 –	28
10	45	5 –	45–28 = 17
15	81	10 –	81–45 = 36
20	143	15 –	143–81 = 62
25	280	20 –	280–143 = 137
30	349	25 –	349–280 = 69
35	374	30 –	374–349 = 25
40	395	35 –	395–374 = 21
45	400	40 – 45	400–395 = 5

Histogram is:



2.41(b) Here range is $32-8=24$ and number of classes = 5, so the class internal is $\frac{24}{5}=4.8$, *i.e.*, 5. Thus the class

limits are 8 to 12, 13 to 17, ..., 28 to 32. The frequency against 4th class is $15-9=6$. Let total frequency be x. Then $0.30x=6$ gives $x=20$. Relative frequency 0.05 gives a frequency of 1. Proceeding this way, we complete the table as follows:

Class Limits	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Percentage
8 to 12	5	0.25	5	25
13 to 17	1	0.05	6	30
18 to 22	3	0.15	9	45
23 to 27	6	0.30	15	75
28 to 32	5	0.25	20	100
	20	1.00		

2.42 Stem-and-Leaf Display of the given data

Stem	Leaf
0	5 7 7 9
1	0 3 5 5 7
2	9 0 8 4 5 2 2 4 5
3	0 1 0 5 5 3 5
6	3 3
7	8 9 5 0
8	0 0
9	9 0

ii) Condensing the Stem-and-Leaf display, we get

Grouping Stem	Leaf																		f
0 – 2	5	7	7	9	0	3	5	5	7	9	0	8	4	5	2	2	4	5	18
3 – 5	0	1	0	5	5	3	5												7
6 – 9	3	3	8	9	5	0	0	0	9	0									10
																			35

●●●●●●●●●●

Chapter 3

MEASURES OF CENTRAL TENDENCY OR AVERAGES

3.13 (i) The decision is wrong, as an average does not reveal the whole picture.

(ii) The conclusion is wrong, as there can be several brilliant students in the class.

(iii) The conclusion is wrong, as the mean is highly affected by abnormally large or small values.

(iv) The conclusion is absurd as few people walk in the middle of the road.

3.15 (b) Now $\sum(x_i-A)^2 = \sum[x_i-M+M-A]^2$, (adding and subtracting M)

$$= \sum[(x_i-M) + (M-A)]^2$$

$$= \sum[(x_i-M)^2+(M-A)^2+2(M-A)(x_i-M)]$$

$$= \sum[(x_i-M)^2+n(M-A)^2+2(M-A)\sum(x_i-M)]$$

But $\sum x_i-M) = 0$, as the sum of deviations taken from mean is always equal to zero. Therefore the cross product term vanishes.

Hence $\sum(x_i-A)^2 = \sum(x_i-M)^2 + n(M-A)^2$

(c) Let $\bar{x}$ denote the mean of the combined distribution. Then

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$$

where $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ are the means and n_1, n_2, n_3 are the frequencies of the 3 components respectively.

Substitution gives

$$\bar{x} = \frac{3(2) + 4(5.5) + 5(10)}{3 + 4 + 5} = \frac{78}{12} = 6.5$$

3.16. (c) Let $\bar{x}$ denote the mean of the combined distribution. Then

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$$

where $\bar{x}_1, \bar{x}_2, \bar{x}_3$ are the means and n_1, n_2, n_3 are the frequencies of the 3 components respectively.

Hence $$\bar{x} = \frac{45(2) + 40(2.5) + 65(2)}{45 + 40 + 65}$$

$$= \frac{90 + 100 + 130}{150} = \frac{320}{150} = 2.13.$$

3.17. Calculation of the mean, the median and the geometric mean of the values which have been arranged in ascending order.

No.	Values (x)	$\log x$
1	9	0.9542
2	12	1.0792
3	15	1.1761
4	15	1.1761
5	16	1.2041
6	18	1.2553
7	20	1.3010
8	20	1.3010
9	25	1.3979
10	30	1.4771
$\sum$	180	12.3220

Hence (i) Mean, *i.e.* $\bar{x} = \frac{\sum x}{n} = \frac{180}{10} = 18$

(ii) Median = Size of $\frac{1}{2}\left[\left(\frac{n}{2}\right) + \left(\frac{n}{2}+1\right)\right]th$ observation

($\because \frac{n}{2}$ is an integer)

= Size of $\frac{1}{2}$[5th + 6th] observation

$$= \frac{16 + 18}{2} = 17$$

(iii) G.M. = anti-log $\left[\frac{\sum \log x}{n}\right]$

= anti-log $\left[\frac{12.3220}{10}\right]$ = anti-log of 1.2322 = 17.07

3.18 (a) Given two positive numbers a and b.

The A.M., G.M. are H.M. are then defined as

A.M. $= \frac{a + b}{2}$, G.M. $= \sqrt{ab}$ and H.M. $= \frac{2}{1/a + 1/b} = \frac{2ab}{a+b}$

Now A.M. $\geq$ G.M. of $\frac{a + b}{2} \geq \sqrt{ab}$

or $a + b \geq 2\sqrt{ab}$ or $(a + b)^2 \geq 4ab$

or $(a + b)^2 - 4ab \geq 0$

or $(a - b)^2 \geq 0$, which is true.

Again, G.M. $\geq$ H.M., if

$\sqrt{ab} \geq \frac{2ab}{a + b}$, or $ab \geq \frac{4a^2b^2}{(a + b)^2}$

or $(a + b)^2 \geq \frac{4a^2b^2}{ab}$, or $(a + b)^2 \geq 4ab$

or $(a + b)^2 - 4ab \geq 0$ or $(a - b)^2 \geq 0$, which is true.

Hence, for two positive numbers, A.M. $\geq$ G.M. $\geq$ H.M.

(b) Calculation of the arithmetic average, the geometric mean and the harmonic mean.

	Income (x)	$\log x$	$\frac{1}{x}$
	85	1.9294	0.0118
	70	1.8451	0.0143
	10	1.0000	0.1000
	75	1.8751	0.0133
	500	2.6990	0.0020
	8	0.9031	0.1250
	42	1.6232	0.0238
	250	2.3979	0.0040
	40	1.6022	0.0250
	36	1.5563	0.0278
Σ	1116	17.4313	0.3470

$\therefore$ Mean $= \frac{\sum x}{n} =$ Rs. $\frac{1116}{10} =$ Rs. 111.60

G.M. $=$ anti-log $\left[\frac{\sum \log x}{n}\right]$

$=$ anti-log $\left[\frac{17.4313}{10}\right] =$ anti-log $(1.74313) =$ Rs. 55.35

H.M. $= \frac{n}{\Sigma\left(\frac{1}{x}\right)} = \frac{10}{0.3470} =$ Rs. 28.82

Here G.M. is the best average.

3.19. Calculation of the arithmetic mean, the geometric mean and the harmonic mean.

	x	$\log x$	$\frac{1}{x}$
	60	1.7782	0.01667
	80	1.9031	0.01250
	90	1.9542	0.01111
	96	1.9823	0.01042
	120	2.0792	0.00833
	150	2.1761	0.00667
	200	2.3010	0.00500
	360	2.5563	0.00278
	480	2.6812	0.00208
	520	2.7160	0.00192
	1060	3.0253	0.00094
	1200	3.0792	0.00083
	1450	3.1614	0.00069
	2500	3.3979	0.00040
	7200	3.8573	0.00014
Σ	15566	38.6487	0.08048

$\bar{x} = \frac{\sum x}{n} = \frac{15566}{15} = 1037.73$

G.M. $=$ Anti-log $\left[\frac{\sum \log x}{n}\right] =$ Anti-log $\left[\frac{38.6487}{15}\right]$

$=$ Anti-log $(2.5766) = 377.2$

H.M. $= \frac{n}{\Sigma\left(\frac{1}{x}\right)} = \frac{15}{0.08048} = 186.7$

3.20 (a) (i) Mean earnings $= \frac{60(3)+20(2)}{80} = \frac{220}{80} =$ Rs.2.75 per hour

(b) **Calculation of the weighted mean.**

Subject	Marks % (x)	Weight (w)	xw
English	73	4	292
French	82	3	246
Maths	57	3	171
Science	62	1	62
History	60	1	60
Total	---	12	831

Hence weighted mean $= \frac{\sum xw}{\sum w} = \frac{831}{12} = 69.25\%$ marks.

3.21 Calculation of the simple and weighted averages

(1) Piece goods	(2) Price per metre (Rs.) (x)	(3) Quantity (millions metres) (w)	xw
Unbleached	8.37	286	2393.82
Bleached	9.50	255	2422.50
Printed flags	9.16	64	586.24
Other sorts	9.84	172	1692.48
Dyed in piece	13.65	165	2252.25
Of dyed yarn	11.95	80	956.00
Total	62.47	1022	10303.29

(i) $\bar{x} = \frac{\sum x}{n} = \frac{62.47}{6} =$ Rs. 10.41 per metre

(ii) Weighted average $= \frac{\sum xw}{\sum w} = \frac{10303.29}{1022} =$ Rs.10.08 per metre

The weighted average price is more nearly the real average price, because the price of each and every piece goods has been multiplied by the corresponding quantity, *i.e.* properly weighted.

3.22. We first construct the frequency table and then calculate the average bonus paid per employee, which would be the weighted mean.

Monthly salary in rupees	Tally	Frequency (w)	Bonus (x)	xw
Exceeding 60 but not exceeding 75	ǀǀǀ	3	10	30
Exceeding 75 but not exceeding 90	ǀǀǀǀ	4	15	60
Exceeding 90 but not exceeding 105	~~ǀǀǀǀ~~	5	20	100
Exceeding 105 but not exceeding 120	~~ǀǀǀǀ~~	5	25	125
Exceeding 120 but not exceeding 135	~~ǀǀǀǀ~~ ǀǀ	7	30	210
Exceeding 135 but not exceeding 150	~~ǀǀǀǀ~~ ǀ	6	35	210
Total	--	30	--	735

Hence weighted mean $= \frac{\sum xw}{\sum w} = \frac{735}{30} =$ Rs. 24.50.

3.23. Calculation of the average age of the horses.

Age (years)	f (also w_i)	x	D	fD	mean age, $\bar{x}_i$	$\bar{x}_i w_i$
1–4	12	2.5	–9.5	–114	2.7	32.4
5–9	223	7.0	–5	–1115	7.6	1694.8
10–14	435	12.0	0	–1229	12.0	5220.0
15–19	272	17.0	5	1360	16.3	4433.6
20–24	161	22.0	10	1610	20.8	3348.8
25–29	34	27.0	15	510	25.8	877.2
30&over	6	32.0	20	120	31.8	190.8
Total	1143	--	--	+3600 2371	--	15797.6

(a) Average age $_{(simple)} = a + \frac{\sum fD}{n}$

$$= 12.0 + \frac{2371}{1143} = 12.0 + 2.07 = 14.07 \text{ years}$$

(b) Average age $_{(weighted)} = \dfrac{\sum \bar{x}_i w_i}{\sum w_i} = \dfrac{15797.6}{1143} = 13.82$ years

The weighted average age is more nearly the real average age, because the mean age of each and every age-group has been multiplied by the corresponding frequency, *i.e.* properly weighted.

3.24. Calculation of the arithmetic mean, the geometric mean and the harmonic mean of the $(n+1)$ values 1, 2 4, 8, 16, ..., 2^n which are in geometric progression.

The sum of the values in G.P. is obtained by the formula

$$S = \frac{a\,(r^n - 1)}{r-1}, \text{ where } r > 1$$

$$= \frac{a\,(1 - r^n)}{1-r}, \text{ where } r < 1.$$

$$\therefore \text{ Sum of the values} = \frac{1\,(2^{n+1} - 1)}{2-1} \quad (\because\ a = 1, r = 2)$$

$$= 2^{n+1} - 1.$$

$$\text{Product of the values} = 1 \times 2 \times 4 \times 8 \times 16 \times \ldots \times 2^n$$

$$= 2^0 \times 2^1 \times 2^2 \times 2^3 \times 2^4 \times \ldots \times 2^n$$

$$= 2^{0+1+2+3+4+\ldots+n}$$

$$= 2^{n+1(0+n)/2} = 2^{n(n+1)/2}$$

$$\text{Sum of reciprccals} = \frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots + \frac{1}{2^n}$$

$$= \frac{1 - \frac{1}{2^{n+1}}}{1 - \frac{1}{2}} \quad (\because\ a = 1, r = \tfrac{1}{2})$$

$$= 2\left(1 - \frac{1}{2^{n+1}}\right)$$

$$\text{Hence } \bar{x} = \frac{\text{Sum of the values}}{\text{No. of values}} = \frac{2^{n+1} - 1}{n+1};$$

$$\text{G.M.} = (\text{Product of the values})^{1/n+1}$$

$$= [2^{n(n+1)/2}]^{1/n+1} = 2^{n/2}; \text{ and}$$

$$\text{H.M.} = \frac{\text{Number of values}}{\text{Sum of their reciprocals}} = \frac{n+1}{2\left(1 - \frac{1}{2^{n+1}}\right)}$$

3.25. Calculation of the arithmetic mean, the geometric mean and the harmonic mean of the $(n+1)$ values 1, 3, 9, 27, ..., 3^n which are in geometric progression.

The sum of the values in G.P. is obtained by the formula

$$\text{Sum} = \frac{a\,(r^n - 1)}{r-1}, \text{ where } r > 1$$

$$= \frac{a\,(1 - r^n)}{1-r}, \text{ where } r < 1$$

Here $a = 1$ and r (common ratio) $= 3$

$$\therefore \text{ Sum of the values} = \frac{1\,(3^{n+1} - 1)}{3-1} = \frac{1}{2}(3^{n+1} - 1)$$

$$\text{Product of the values} = 1 \times 3 \times 9 \times 27 \times 81 \times \ldots \times 3^n$$

$$= 3^{0+1+2+3+\ldots+n}$$

$$= 3^{(0+n)(n+1)/2} = 3^{n(n+1)/2}$$

$$\text{Sum of reciprocals} = \frac{1}{1} + \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \frac{1}{81} + \ldots + \frac{1}{3^n}$$

$$= \frac{1\left(1 - \frac{1}{3^{n+1}}\right)}{1 - 1/3} = \frac{3}{2}\left(1 - \frac{1}{3^{n+1}}\right)$$

$$(\because\ a = 1, r = \tfrac{1}{3})$$

Hence $\bar{x} = \frac{\text{Sum of all values}}{\text{No. of values}} = \frac{\frac{1}{2}(3^{n+1}-1)}{n+1}$;

G.M. = [Product of the values]$^{1/(n+1)}$

$= [3^{n(n+1)/2}]^{1/(n+1)} = 3^{n/2}$; and

$$\text{H.M.} = \frac{\text{Number of values}}{\text{Sum of their reciprocals}} = \frac{n+1}{\frac{3}{2}\left(1-\frac{1}{3^{n+1}}\right)}.$$

3.27. The rates of his salary after getting rises come to 1.10, 1.20 and 1.25 respectively. The appropriate average of these rates is the geometric mean. Thus

$$\text{G.M.} = \sqrt[3]{1.10 \times 1.20 \times 1.25}$$

Taking logs, we have

$$\log G = \tfrac{1}{3}[\log 1.10 + \log 1.20 + \log 1.25]$$

$$= \tfrac{1}{3}[0.0414+0.0792+0.0969] = \frac{0.2175}{3} = 0.0725$$

$\therefore$ G.M. = Anti-log of 0.0725 = 1.181.

Hence the required annual percentage increase = 18.1%

3.28. (b) Calculation of the Harmonic mean which is the correct average speed of the person in this question.

$$\text{H.M.} = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{2}{\frac{1}{30}+\frac{1}{60}} = \frac{2}{0.05} = 40 \text{ miles per hour}$$

(c) Calculation of the Harmonic mean which is the correct average speed of the person in the question.

$$\text{H.M} = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{3}{\frac{1}{8}+\frac{1}{7.5}+\frac{1}{5.5}}$$

$$= \frac{3}{0.44015} = 6.8 \text{ miles per hour}$$

3.29. (a) As the distances are equal (and hence constant) and the times vary, therefore the correct average is the harmonic mean, which is obtained below:

Speed (x)	$1/x$
10	0.1000
15	0.0667
20	0.0500
25	0.0400
20	0.0500
30	0.0333
40	0.0250
50	0.0200
30	0.0333
40	0.0250
$\sum$	0.4433

Thus $\text{H.M.} = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{10}{0.4433} = 22.56$ kilometres p.h.

(b) (i) Calculation of the harmonic mean, the correct average rate.

$$\text{H.M.} = \frac{3}{\frac{1}{10}+\frac{1}{8}+\frac{1}{6}} = \frac{3}{0.1000+0.1250+0.1667}$$

$$= \frac{3}{0.3917} = 7.66 \text{ m.p.h.}$$

(ii) The rates of increase in population come out to be 1.20, 1.25 and 1.44. The average rate of increase would be their G.M.

$\therefore$ $\text{G.M.} = \sqrt[3]{1.20 \times 1.25 \times 1.44}$

Taking logs, we have

$$\log G = \tfrac{1}{3}[\text{og } 1.20 + \log 1.25 + \log 1.44]$$

$$= \tfrac{1}{3}[0.0792 + 0.0969 + 0.1584]$$

$$= \frac{0.3345}{3} = 0.1115$$

$\therefore$ G.M. = Anti-log (0.1115) = 1.293

Hence the required percentage increase = 29.3%.

3.30. Calculation of the Geometric mean and the Harmonic mean

Weekly income (Rs.)	No. of workers (f)	x	$\log x$	$f \log x$	$f(1/x)$
35–39	15	37	1.5682	23.5230	0.4054
40–44	13	42	1.6232	21.1016	0.3095
45–49	17	47	1.6721	28.4257	0.3617
50–54	29	52	1.7160	49.7640	0.5577
55–59	11	57	1.7559	19.3149	0.1930
60–64	10	62	1.7924	17.9240	0.1613
65–69	5	67	1.8261	9.1305	0.0746
Total	100	--	--	169.1837	2.0632

Now $\quad \text{G.M.} = \text{Anti-log of}\left(\frac{\sum f \log x}{\sum f}\right) = \text{Anti-log of}\left(\frac{169.1837}{100}\right)$

$= \text{Anti-log of } (1.6918) = \text{Rs. } 49.18$, and

$$\text{H.M.} = \frac{n}{\sum f\left(\frac{1}{x}\right)} = \frac{100}{2.0632} = \text{Rs. } 48.47$$

3.31. Calculation of Geometric and Harmonic means.

Variable	f	x	$\log x$	$f \log x$	$\frac{1}{x}$	$f\left(\frac{1}{x}\right)$
0 – 5	2	2.5	0.3979	0.7958	0.4000	0.8000
5 – 10	5	7.5	0.8751	4.3755	0.1333	0.6665
10 – 15	7	12.5	1.0969	7.6783	0.0800	0.5600
15 – 20	13	17.5	1.2430	16.1590	0.0571	0.7423
20 – 25	21	22.5	1.3522	28.3962	0.0444	0.9324
25 – 30	16	27.5	1.4393	23.0288	0.0364	0.5824
30 – 35	8	32.5	1.5119	12.0952	0.0308	0:2464
35 – 40	3	37.5	1.5740	4.7220	0.0267	0.0801
Total	75	--	--	97.2508	---	4.6101

Hence $\text{G.M.} = \text{Anti-log}\left[\frac{\sum f \log x}{n}\right] = \text{Anti-log}\left[\frac{97.2508}{75}\right]$

$= \text{Anti-log } (1.2966) = 19.80$; and

$$\text{H.M.} = \frac{n}{\sum f\left(\frac{1}{x}\right)} = \frac{75}{4.6101} = 16.27.$$

3.32. (i) **The salaries varied greatly, so median is more suitable average.** Arranging the data in ascending order we get Rs.100, Rs. 950, Rs. 1500, Rs. 2100, Rs. 10000.

Thus Median = salary of $\left(\frac{n+1}{2}\right)$th person ($\because \frac{n}{2}$ is not an integer)

= salary of $(5+1)/2$ th, *i.e.* 3rd person = Rs. 1500.

(ii) The given heights do not differ greatly, so arithmetic mean is more suitable average.

Therefore $\bar{x} = \frac{64+65+65+66+66+67}{6} = \frac{393}{6} = 65.5''$

(iii) Here again median is more suitable average. Since $\frac{n}{2}$ is an integer, therefore median is the average value of $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th, *i.e.* 2nd and 3rd observations.

Hence $\quad \text{Median} = \frac{18+18}{2} = 18.$

3.33. Calculation of the median, quartiles, etc.

Size of shoes	No. of pairs (f)	c.f.
5	2	2
$5\frac{1}{2}$	5	7
6	15	22
$6\frac{1}{2}$	30	52
7	60	112
$7\frac{1}{2}$	40	152
8	23	175
$8\frac{1}{2}$	11	186
9	4	190
$9\frac{1}{2}$	1	191
Total	191	--

Here $\frac{n}{2}$, *i.e.* $\frac{191}{2}$ is not an integer, therefore

Median $= \text{Size of}\left(\left[\frac{191}{2}\right] + 1\right)th$ pair

$= \text{Size of } (95 + 1)th$, *i.e.* 96*th* pair = 7

Again $\frac{n}{4}$, *i.e.* $\frac{191}{4}$ is not an integer. Therefore

$Q_1 = \text{Size of}\left(\left[\frac{191}{4}\right] + 1\right)th$ pair,

$= \text{Size of } (47 + 1)th$, *i.e.* 48*th* pair $= 6\frac{1}{2}$

$Q_3 = \text{Size of}\left(\left[\frac{3\times191}{4}\right]+1\right)th$ pair as $\frac{3n}{4}$ is also not an integer

$= \text{Size of } (143 + 1)th$, *i.e.* 144*th* pair $= 7\frac{1}{2}$

Now $\frac{7n}{10}$, *i.e.* $\frac{7 \times 191}{10}$ is not an integer, therefore

$D_7 = \text{Size of}\left(\left[\frac{7 \times 191}{10}\right] + 1\right)th$ pair

$= \text{Size of } (133 + 1)th$, *i.e.* 134*th* pair $= 7\frac{1}{2}$

$P_{64} = \text{Size of}\left(\left[\frac{64 \times 191}{100}\right] + 1\right)th$ pair as $\frac{64n}{100}$ is not an integer

$= \text{Size of } (122 + 1)th$, *i.e.* 123*th* pair $= 7\frac{1}{2}$

3.34. Calculation of the mean, the median and the modal numbers of persons per house.

No. of persons (x)	No. of houses (f)	fx	Cumulative frequency
1	26	26	26
2	113	226	139
3	120	360	259
4	95	380	354
5	60	300	414
6	42	252	456
7	21	147	477
8	14	112	491
9	5	45	496
10	4	40	500
Total	500	1888	---

Hence $\bar{x} = \frac{\sum fx}{n} = \frac{1888}{500} = 3.78,$

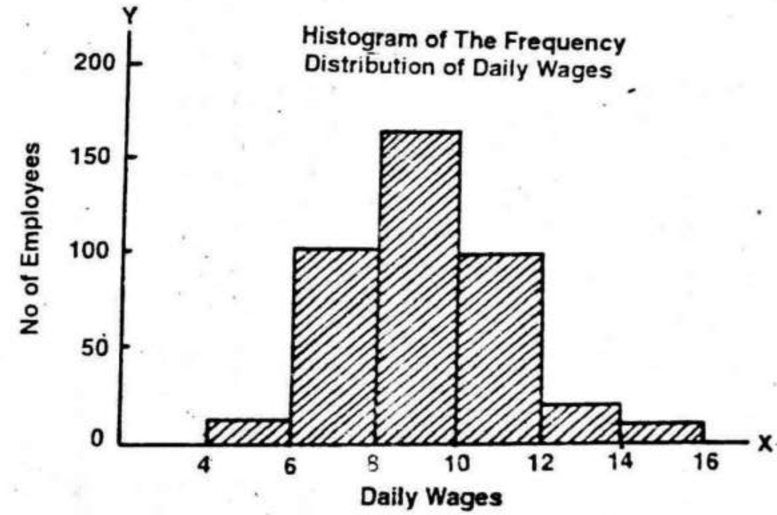
Median = Average of two values with ordinal numbers $\left(\frac{n}{2}\right)$ and $\left(\frac{n}{2}\right) + 1$ as $\frac{n}{2}$, *i.e.* $\frac{500}{2}$ is an integer.

= Average no. of persons of 250*th* and 251*st* houses

= 3, and

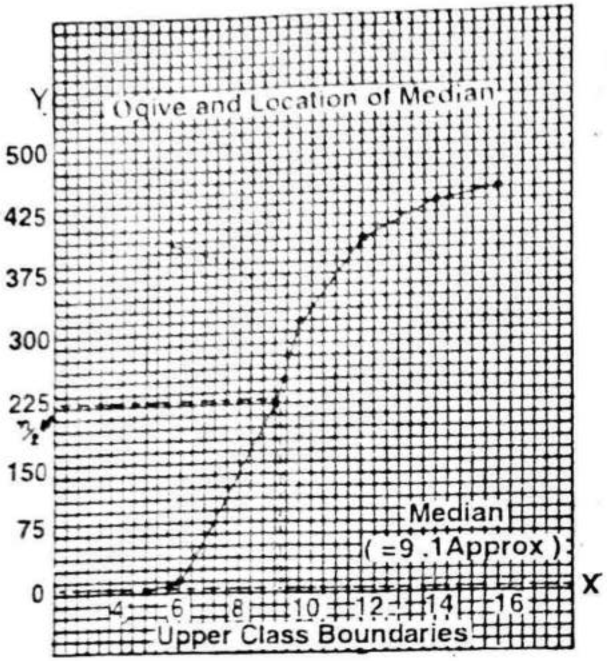
Mode = No. of persons corresponding to the maximum frequency = 3

3.35. (i) Histogram of the daily wages.



(ii) Cumulative Frequency Distribution and its graph.

Daily wages	F
less than or equal to 4	0
less than or equal to 6	13
less than or equal to 8	124
less than or equal to 10	306
less than or equal to 12	411
less than or equal to 14	430
less than or equal to 16	437



Checking the answers by calculations.

Daily wages	f	F
4–6	13	13
6–8	111	124
8–10	182	306
10–12	105	411
12–14	19	430
14–16	7	437
Total	437	--

Median = Daily wage of $\left(\frac{n}{2}\right)th$ employee

= Daily wage of $\frac{437}{2}$ *th, i.e.*, 218.5th employee

which lies in the group 8–10. Therefore

Median $= l + \frac{h}{f}\left(\frac{n}{2} - C\right)$

$= 8 + \frac{2}{182}(218.5 - 124) = 8 + 1.04 =$ Rs. 9.04

3.36 Calculations of the median and quartile ages by formula

Age of head	f	$C.F.$
under 25	44	44
25 and under 30	79	123
30 and under 40	152	275
40 and under 50	122	397
50 and under 60	141	538
60 and under 65	100	638
65 and under 70	58	696
70 and under 75	32	728
75 and under 85	28	756

Median = Age of head of $\left(\frac{n}{2}\right)th$ household

= Age of head of $\left(\frac{756}{2}\right)th$, *i.e.* 378*th* household,

which lies in the age-group 40 and under 50. Therefore

Median $= l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 40 + \frac{10}{122}(378 - 275)$

= 40 + 8.44 = 48.44 years

Q_1 = Age of head of $\left(\frac{n}{4}\right)th$ household

= Age of head of $\left(\frac{756}{4}\right)th$, *i.e.* 189*th* household,

which lies in the age group 30 and under 40. Thus

$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right) = 30 + \frac{10}{152}(189 - 123)$

= 30 + 4.34 = 34.34 years; and

Q_3 = Age of head of $\left(\frac{3n}{4}\right)th$ household

$= \text{Age of head of } \left(\frac{3 \times 756}{4}\right)th$, *i.e.* 567*th* household which lies in the age group 60 and under 65. Thus

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$= 60 + \frac{5}{100}(567 - 538) = 60 + 1.45 = 61.45 \text{ years.}$$

Graphical location is left as an exercise.

3.37. Calculation of the median and the quartiles.

Height (inches)	Number (f)	Cumulative frequency
57.5 – 60.0	6	6
60.0 – 62.5	26	32
62.5 – 65.0	190	222
65.0 – 67.5	281	503
67.5 – 70.0	412	915
70.0 – 72.5	127	1042
72.5 – 75.0	38	1080
Total	1080	---

Median $= \text{Height of } \left(\frac{n}{2}\right)th$ person

$= \text{Height of } \frac{1080}{2}th$, *i.e.* 540*th* person, which lies in the group 67.5 – 70.0

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$= 67.5 + \frac{2.5}{412}(540-503) = 67.5+0.22=67.72 \text{ inches.}$$

Lower quartile $= \text{Height of } \left(\frac{n}{4}\right)th$ person

$= \text{Height of } \frac{1080}{4}th$, *i.e.* 270*th* person, which lies in the group 65.0 – 67.5. Thus

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right)$$

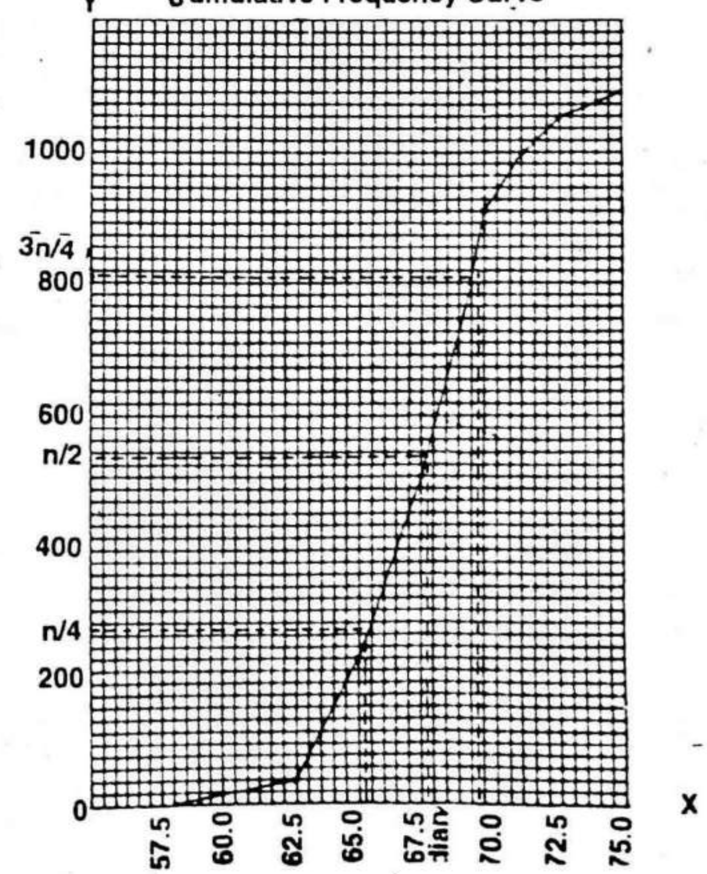
$$= 65.0 + \frac{2.5}{281}(270-222) = 65.0+0.43=65.43 \text{ inches.}$$

Similarly $Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$

$$= 67.5 + \frac{2.5}{412}(810-503) = 67.5+1.86=69.36 \text{ inches.}$$

These results are checked on the following graph

Cumulative Frequency Curve



3.38. Calculation of the Median.

Classes	Class boundaries	Number (f)	F
100–104	99.5–104.5	4	4
105–109	104.5–109.5	14	18
110–114	109.5–114.5	60	78
115–119	114.5–119.5	138	216
120–124	119.5–124.5	236	452
125–129	124.5–129.5	298	750
130–134	129.5–134.5	380	1130
135–139	134.5–139.5	450	1580
140–144	139.5–144.5	500	2080
145–149	144.5–149.5	430	2510
150–154	149.5–154.5	260	2770
155–159	154.5–159.5	128	2898
160–164	159.5–164.5	66	2964
165–169	164.5–169.5	28	2992
170–174	169.5–174.5	12	3004
Total	---	3004	--

Median $= \text{Size of } \left(\frac{n}{2}\right)th$ number

$= \text{Size of } \frac{3004}{2}th$, *i.e.* 1502nd number, which lies in the group 134.5 – 139.5. Thus

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$= 134.5 + \frac{5}{450}(1502-1130) = 134.5+4.13=138.63$$

3.39. Calculation of the mean and median ages.

Age in yrs	No. of persons	F	x	fx
< 1	5	5	0.5	2.5
1–4	10	15	2.5	25.0
5–9	11	26	7.0	77.0
10–19	12	38	14.5	174.0
20–29	22	60	24.5	539.0
30–39	18	78	34.5	621.0
40–59	8	86	49.5	396.0
60–79	7	93	69.5	486.5
Total	93	---	---	2321.0

Mean age $= \frac{\sum fx}{n} = \frac{2321.0}{93} = 24.96$ years.

Median age $=$ age of $\left(\frac{n}{2}\right)th$ person

$=$ age of $\left(\frac{93}{2}\right)th$, *i.e.* 46.5*th* person which lies in the group 19.5 – 29.5

$\therefore$ Median age $= l + \frac{h}{f}\left(\frac{n}{2} - C\right)$

$$= 19.5 + \frac{10}{22}(46.5-38) = 19.5 + \frac{85}{22} = 23.36 \text{ years}$$

3.40. (b) Calculation of the median, the lower and the upper quartiles.

Class Interval	Class boundaries	f	F
Under 25	upto 24.5	222	222
25–29	24.5–29.5	405	627
30–34	29.5–34.5	508	1135
35–39	34.5–39.5	520	1655
40–44	39.5–44.5	525	2180
45–49	44.5–49.5	490	2670
50–54	49.5–54.5	457	3127
55–59	54.5–59.5	416	3543
60 & over	59.5+over	166	3709
Total	---	3709	---

Median $= \text{Size of} \left(\frac{n}{2}\right) th$ observation

$= \text{Size of } \frac{3709}{2}\, th$, *i.e.*. 1854.5*th* observation, which lies in the group 39.5 – 44.5. Thus

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$= 39.5 + \frac{5}{525}(1854.5 - 1655) = 39.5 + 1.9 = 41.4$$

Similarly, $Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right)$

$$= 29.5 + \frac{5}{508}(927.25 - 627) = 29.5 + 2.96 = 32.46; \text{ and}$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$= 49.5 + \frac{5}{457}(2781.75 - 2670) = 49.5 + 1.22 = 50.72$$

3.41. Estimation of the mean, the median and the quartiles.

Consumption in kilowatt hours	No. of consumers (f)	x	fx	F	Class Boundries
5–24	4	14.5	58.0	4	4.5–24.5
25–44	6	34.5	207.0	10	24.5–44.5
45–64	14	54.5	763.0	24	44.5–64.5
65–84	22	74.5	1639.0	46	64.5–84.5
85–104	14	94.5	1323.0	60	84.5–104.5
105–124	5	114.5	572.5	65	104.5–124.5
125–144	7	134.5	941.5	72	124.5–144.5
145–164	3	154.5	463.5	75	144.5–164.5
Total	75	--	5967.5	--	--

$$\text{Mean} = \frac{\sum fx}{n} = \frac{5967.5}{75} = 79.57 \text{ kilowatt hours}$$

Median = Kilowatt hours of $\left(\frac{n}{2}\right) th$ consumer

= Kilowatt hours of $\left(\frac{75}{2}\right) th$, *i.e.* 37.5*th* consumer which lies in the group 64.5–84.5. Thus

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 64.5 + \frac{20}{22}(37.5 - 24)$$

$$= 64.5 + \frac{270}{22} = 76.77 \text{ kilowatt hours}$$

Q_1 = kilowatt hours of $\left(\frac{n}{4}\right) th$ consumer

= kilowatt hours of $\left(\frac{75}{4}\right) th$, *i.e.* 18.75*th* consumer, which lies in the group 44.5–64.5. Therefore

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right) = 44.5 + \frac{20}{14}(18.75 - 10)$$

$$= 44.5 + \frac{175}{14} = 57.0 \text{ kilowatt hours}$$

Similarly, $Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) = 84.5 + \frac{20}{14}(56.25 - 46)$

$$= 84.5 + \frac{205}{14} = 99.14 \text{ kilowatt hours.}$$

3.42. Calculation of the mean, the median and the quartiles.

Yield (lb)	x	f	$u\left(=\frac{x-4.0}{0.2}\right)$	fu	F
2.7–2.9	2.8	4	−6	−24	4
2.9–3.1	3.0	15	−5	−75	19
3.1–3.3	3.2	20	−4	−80	39
3.3–3.5	3.4	47	−3	−141	86
3.5–3.7	3.6	63	−2	−126	149
3.7–3.9	3.8	78	−1	−78	227
3.9–4.1	4.0	88	0	−524	315
4.1–4.3	4.2	69	1	69	384
4.3–4.5	4.4	59	2	118	443
4.5–4.7	4.6	35	3	105	478
4.7–4.9	4.8	10	4	40	488
4.9–5.1	5.0	8	5	40	496
5.1–5.3	5.2	4	6	24	500
Total	--	500	--	+396 −128	---

Now $\bar{x} = a + \frac{\sum fu}{n} \times h$

$= 4.0 + \frac{(-128)}{500} \times 0.2 = 4.0 - 0.05 = 3.95$ lb.

Median $=$ yield of $\left(\frac{n}{2}\right)th$ plot

$=$ yield of $\frac{500}{2}th$, *i.e.* 250*th* plot, which lies in the group 3.9–4.1. Thus

Median $= l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 3.9 + \frac{0.2}{88}(250 - 227)$

$= 3.9 + 0.05 = 3.95$ lb.

First quartile $=$ yield of $\left(\frac{n}{4}\right)th$ plot

$=$ yield of $\frac{500}{4}th$, *i.e.* 125*th* plot, which lies in the group 3.5 – 3.7

$\therefore \quad Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right) = 3.5 + \frac{0.2}{63}(125-86)$

$= 3.5 + 0.12 = 3.62$ lb, and

Third quartile $=$ yield of $\left(\frac{3n}{4}\right)th$ plot

$=$ yield of $\frac{3(500)}{4}th$, *i.e.* 375*th* plot, which lies in the group 4.1 – 4.3. Thus

$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) = 4.1 + \frac{0.2}{69}(375-315)$

$= 4.1 + 0.17 = 4.27$ lb

3.43. Calculation of the average weight, the median weight, the mode, etc.

Class Boundries	No. of Seeds (f)	x	fx	F
9.95–24.95	16	17.45	279.20	16
24.95–39.95	68	32.45	2206.60	84
39.95–54.95	204	47.45	9679.80	288
54.95–69.95	233	62.45	14550.85	521
69.95–84.95	240	77.45	18588.00	761
84.95–99.95	655	92.45	60554.75	1416
99.95–114.95	803	107.45	86282.35	2219
114.95–129.95	294	122.45	36000.30	2513
129.95–144.95	21	137.45	2886.45	2534
144.95–159.95	4	152.45	609.80	2538
Total	2538	---	231638.10	--

(a) Average weight $= \frac{\sum fx}{n} = \frac{231638.10}{2538} = 91.27$ milligrams

Median = weight of $\left(\frac{n}{2}\right)$th seed

$= $ weight of $\left(\frac{2538}{2}\right)$th, *i.e.* 1269th seed which lies in the group 84.95–99.95. Therefore

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 84.95 + \frac{15}{655}(1269-761)$$

$$= 84.95 + 11.63 = 96.58 \text{ milligrams}$$

$$\text{Mode} = l + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 99.95 + \frac{(803-655)}{(803-655)+(803-294)} \times 15$$

$$= 99.95 + \frac{148}{148+509} \times 15 = 99.95 + \frac{148}{657} \times 15$$

$$= 99.95 + 3.38 = 103.33 \text{ milligrams.}$$

(b) Q_1 = weight of $\left(\frac{n}{4}\right)$th seed

$= $ weight of $\left(\frac{2538}{4}\right)$th, *i.e.* 634.5th seed which lies in the group 69.95–84.95. Thus

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right) = 69.95 + \frac{15}{240}(634.5-521)$$

$$= 69.95 + 7.09 = 77.04 \text{ milligrams}$$

Similarly, we estimate

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) = 99.95 + \frac{15}{803}(1903.5-1416)$$

$$= 99.95 + 9.11 = 109.06 \text{ milligrams;}$$

$$D_3 = l + \frac{h}{f}\left(\frac{3n}{10} - C\right) = 84.95 + \frac{15}{655}(761.4-761)$$

$$= 84.95 + 0.01 = 84.96 \text{ milligrams; and}$$

$$P_{45} = l + \frac{h}{f}\left(\frac{45n}{100} - C\right) = 84.95 + \frac{15}{655}(1142.10-761)$$

$$= 84.95 + 8.73 = 93.68 \text{ milligrams.}$$

3.44. Assuming a range of Rs. 55.00 to Rs. 105.00, the frequency distribution would be as below:

Group	f(%)	f	x	fx
Rs. 55.00 and under Rs. 60.00	4	20	57.50	1150.00
Rs. 60.00 and under Rs. 62.50	11	55	61.25	3368.75
Rs. 62.50 and under Rs. 72.75	10	50	67.625	3381.25
Rs. 72.75 and under Rs. 78.75	15	75	75.75	5681.25
Rs. 78.75 and under Rs. 82.25	10	50	80.50	4025.00
Rs. 82.25 and under Rs. 85.25	10	50	83.75	4187.50
Rs. 85.25 and under Rs. 90.50	15	75	87.875	6590.62
Rs. 90.50 and under Rs. 95.00	10	50	92.75	4637.50
Rs. 95.00 and under Rs. 100.00	10	50	97.50	4875.00
Rs.100.00 and under Rs.105.00	5	25	102.50	2562.50
Total	100	500	--	40459.37

Now Mean wage $= \frac{\sum fx}{n}$

$= \text{Rs. } \frac{40459.37}{500} = $ Rs. 80.92 approximately.

3.45. (b) Calculation of the mean, the median and the mode.

Weight	f	x	$u\left(=\frac{x-149}{9}\right)$	fu	F
118–126	3	122	−3	−9	3
127–135	5	131	−2	−10	8
136–144	9	140	−1	−9	17
145–153	12	149	0	−28	29
154–162	5	158	1	5	34
163–171	4	167	2	8	38
172–180	2	176	3	6	40
Total	40	---	---	+19 −9	--

Mean $= a + \frac{\sum fu}{n} \times h$

$= 149 + \frac{(-9)\,(9)}{40} = 149 - 2.025 = 146.975$

Median $=$ Weight of $\left(\frac{n}{2}\right)th$ student

$=$ Weight of $\left(\frac{40}{2}\right)th$, *i.e.* 20*th* student, which lies in the group 144.5–153.5 (class boundaries)

$\therefore$ Median $= l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 144.5 + \frac{9}{12}(20-17)$

$= 144.5 + 2.25 = 146.75.$

Mode $= l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$

$= 144.5 + \frac{12 - 9}{(12 - 9) + (12 - 5)} \times 9$

$= 144.5 + \frac{27}{10} = 147.2.$

3.46. The calculations needed to find the mean, median and mode are given in the following table:

Max. loads (short tons)	Class Boundries	No. of cables (f)	x	fx	F
9.8 – 10.2	9.75 – 10.25	7	10.0	70.0	7
10.3 – 10.7	10.25 – 10.75	12	10.5	126.0	19
10.8 – 11.2	10.75 – 11.25	17	11.0	187.0	36
11.3 – 11.7	11.25 – 11.75	14	11.5	161.0	50
11.8 – 12.2	11.75 – 12.25	6	12.0	72.0	56
12.3 – 12.7	12.25 – 12.75	4	12.5	50.0	60
Total	---	60	--	666.0	--

Now $\bar{x} = \frac{\sum fx}{n} = \frac{666.0}{60} = 11.10$ short tons

Median $=$ Max. load of $\left(\frac{n}{2}\right)th$ cable

$=$ Max. load of $\left(\frac{60}{2}\right)th$ *i.e.* 30*th* cable, which lies in the group 10.75 – 11.25.

$\therefore$ Median $= l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 10.75 + \frac{0.5}{17}(30 - 19)$

$= 10.75 + 0.32 = 11.07$ short tons

Mode $= l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$

$= 10.75 + \frac{(17 - 12)}{(17 - 12) + (17 - 14)} \times 0.5$

$= 10.75 + 0.31 = 11.06$ short tons.

3.47. Calculations needed to find the *modal* and the *median* wages are given in the table below:

Daily Wages (Rs.)	Class Boundries	No. of Employees (f)	F
22	21 – 23	3	3
24	23 – 25	13	16
26	25 – 27	43	59
28	27 – 29	102	161
30	29 – 31	175	336
32	31 – 33	220	556
34	33 – 35	204	760
36	35 – 37	139	899
38	37 – 39	69	968
40	39 – 41	25	993
42	41 – 43	6	999
44	43 – 45	1	1000
Total		1000	--

Mode wage $= l + \dfrac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$

$= 31 + \dfrac{(220-175)}{(220-175)+(220-204)} \times 2$

$= 31 + \dfrac{45}{45+16} \times 2 = 31 + 1.48 =$ Rs. 32.48

Median wage = Daily wage of $\left(\dfrac{n}{2}\right)$th employee

= Daily wage of $\left(\dfrac{1000}{2}\right)$*th, i.e* 500*th* employee, which lies in the group 31–33. Therefore

Median wage $= l + \dfrac{h}{f}\left(\dfrac{n}{2} - C\right) = 31 + \dfrac{2}{220}(500 - 336)$

$= 31 + \dfrac{328}{220} = 31 + 1.49 =$ Rs. 32.49

3.48. (c) Calculation of mode by the empirical relation.

Mode = 3 (Median) – 2 (Mean)

= 3 (36) – 2(40.5) = 108 – 81 = 27.

●●●●●●●●●●

Chapter 4

MEASURES OF DISPERSION, MOMENTS AND SKEWNESS

4.4 (b) (i) To find the quartile deviation graphically, we locate the two quartiles graphically and then apply the relation

$$Q.D. = \frac{Q_3 - Q_1}{2}.$$

(ii) Calculation of Quartile Deviation by using an appropriate formula:

Income per week (Rs.)	Class Boundaries	No. of Earners, f	F
41–50	40.5–50.5	30	30
51–60	50.5–60.5	36	66
61–70	60.5–70.5	43	109
71–80	70.5–80.5	104	213
81–90	80.5–90.5	73	286
91–100	90.5–100.5	14	300
Total	---	300	---

$Q.D. = \dfrac{Q_3 - Q_1}{2}$, where

Q_1 = Income of $\left(\dfrac{n}{4}\right)$*th* earner

= Income of $\left(\dfrac{300}{4}\right)$*th, i.e.* 75*th* **earner which lies in the group 60.5–70.5, therefore**

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right)$$

$$= 60.5 + \frac{10}{43}(75 - 66) = 60.5 + 2.09 = \text{Rs. } 62.59; \text{ and}$$

$$Q_3 = \text{Income of } \left(\frac{3n}{4}\right) th \text{ earner}$$

$= \text{Income of } \frac{3\,(300)}{4} th$, *i.e.* 225*th* earner which lies in the group 80.5–90.5; therefore

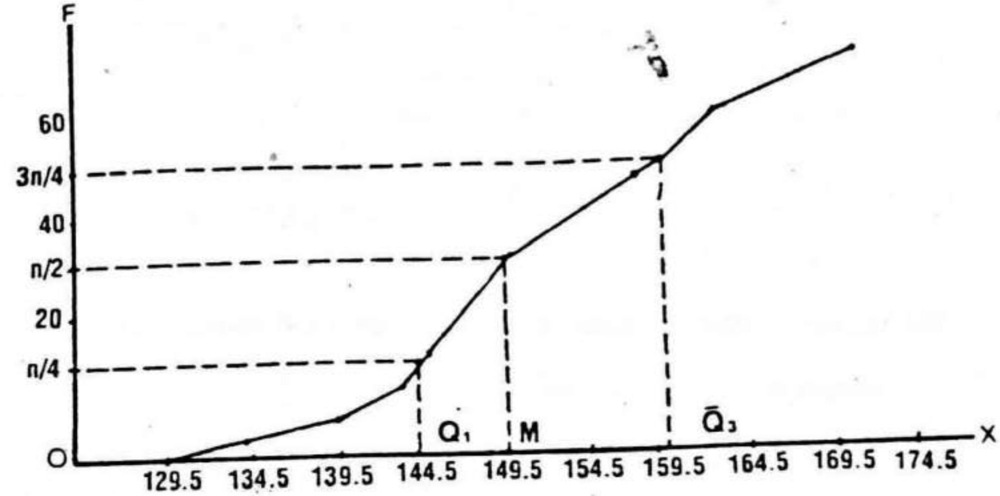
$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$= 80.5 + \frac{10}{73}(225 - 213) = 80.5 + 1.64 = \text{Rs. } 82.14$$

Hence $Q.D. = \frac{Q_3 - Q_1}{2} = \frac{82.14 - 62.59}{2} = \text{Rs. } 9.775 = \text{Rs. } 9.78$

4.5 Construction of a Cumulative Frequency Table

Weight (lbs)	Tally	f	Weight	Cum f
			less than 129.5	0
129.5 – 134.5	III	3	less than 134.5	3
134.5 – 139.5	III	3	less than 139.5	6
139.5 – 144.5	~~IIII~~ I	6	less than 144.5	12
144.5 – 149.5	~~IIII~~ ~~IIII~~ IIII	14	less than 149.5	26
149.5 – 154.5.	~~IIII~~ IIII	9	less than 154.5	35
154.5 – 159.5	~~IIII~~ III	8	less than 159.5	43
159.5 – 164.5	~~IIII~~ II	7	less than 164.5	50
164.5 – 169.5	~~IIII~~ I	6	less than 169.5	54
169.5 – 174.5	IIII	4	less than 174.5	60
Total	---	60	--	--



From graph, we estimate the approximate values of

Median = 152 lb, $Q_1 \simeq$ 146 lb and Q3 $\simeq$ 161 lbs.

$\therefore$ S.I.Q. Range $= \frac{Q_3 - Q_1}{2} = \frac{161 - 146}{2} = 7.5$ lbs.

Calculation of Mean and Standard Deviation, using the grouped data:

Weight	x_i	f	$u(=x-147)/5$	fu	fu^2
129.5–134.5	132	3	−3	−9	27
134.5–139.5	137	3	−2	−6	12
139.5–144.5	142	6	−1	−6	6
144.5–149.5	147	14	0	0	0
149.5–154.5	152	9	1	9	9
154.5–159.5	157	8	2	16	32
159.5–164.5	162	7	3	21	63
164.5–169.5	167	6	4	24	96
169.5–174.5	172	4	5	20	100
Total	--	60	--	69	345

$$\text{Now mean, } \bar{x} = a + \frac{\sum fu}{n} \times h$$

$$= 147 + \frac{69 \times 5}{60} = 147 + 5.75 = 152.75 \text{ lbs, and}$$

$$S = h\sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}$$

$$= 5\sqrt{\frac{345}{60} - \left(\frac{69}{60}\right)^2} = 5\sqrt{5.75 - 1.3225}$$

$$= 5\sqrt{4.4275} = 5\,(2.104) = 10.52 \text{ lbs.}$$

Mean and Standard Deviation of Original Observations

$$\text{Mean} = \frac{\sum X}{n} = \frac{171 + 160 + \dots + 144}{60} = \frac{9175}{60} = 152.917 \text{ lbs.}$$

$$S.D. = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$$

$$= \sqrt{\frac{1409399}{60} - \left(\frac{9175}{60}\right)^2} = \sqrt{23489.98333 - 23383.50694}$$

$$= \sqrt{106.47639} = 10.32 \text{ lbs.}$$

Comparing the means and standard deviations computed from the grouped data and the original observations, we find that they are almost the same.

4.6. (b) Calculation of the Mean Deviation from the mean.

Marks	No. of Students, f	x	fx	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
0–9	2	4.5	9.0	–41.67	41.67	83.34
10–19	3	14.5	43.5	–31.67	31.67	95.01
20–29	8	24.5	196.0	–21.67	21.67	173.36
30–39	24	34.5	828.0	–11.67	11.67	280.08
40–49	27	44.5	1201.5	–1.67	1.67	45.09
50–59	40	54.5	2180.0	8.33	8.33	333.20
60–69	11	64.5	709.5	18.33	18.33	201.63
70–79	5	74.5	372.5	28.33	28.33	141.65
Total	120	--	5540	--	--	1353.36

$$\text{Now } \bar{x} = \frac{\sum fx}{n} = \frac{5540}{120} = 46.17; \text{ and therefore}$$

$$\text{Mean Deviation, } (M.D.) = \frac{\sum f\,\lvert x-\bar{x}\rvert}{n}$$

$$= \frac{1353.36}{120} = 11.28 \text{ marks}$$

4.8. Calculation of the Quartile Deviation, Mean Deviation and their Co-efficients.

Height x	Class Boundaries	Group A				Group B			
		f	F	x–Med	$f\lvert x-$Med$\rvert$	f	F	x–Med	$f\lvert x-$Med$\rvert$
58	57.5–58.5	10	10	–3.35	33.50	15	15	–3.26	48.90
59	58.5–59.5	18	28	–2.35	42.30	20	35	–2.26	45.20
60	59.5–60.5	30	58	–1.35	40.50	32	67	–1.26	40.32
61	60.5–61.5	42	100	–0.35	14.70	35	102	–0.26	9.10
62	61.5–62.5	35	135	0.65	22.75	33	135	0.74	24.42
63	62.5–63.5	28	163	1.65	46.20	22	157	1.74	38.28
64	63.5–64.5	16	179	2.65	42.40	20	177	2.74	54.80
65	64.5–65.5	8	187	3.65	29.20	10	187	3.74	37.40
Total	---	187	--	—	271.55	187	--	--	298.42

Group A.

$$Q_1 = \text{Height of } \left(\frac{n}{4}\right)th \text{ person}$$

$$= \text{Height of } \left(\frac{187}{4}\right)th, \text{ i.e. } 46.75th \text{ person which lies in the group } 59.5 - 60.5,$$

$$\therefore \quad Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right)$$

$$= 59.5 + \frac{1}{30}(46.75 - 28) = 59.5 + 0.62 = 60.12 \text{ inches}$$

$$Q_3 = \text{Height of } \left(\frac{3n}{4}\right)th \text{ person}$$

$$= \text{Height of } \frac{3\,(187)}{4}th, \text{ i.e. } 140.25th \text{ person which lies in the group } 62.5 - 63.5$$

$$\therefore \quad Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) = 62.5 + \frac{1}{28}(140.25-135)$$

$$= 62.5+0.19 = 62.69 \text{ inches}$$

Median = Height of $\left(\frac{n}{2}\right)th$ person

= Height of $\left(\frac{187}{2}\right)th$, *i.e.* 93.5*th* person which lies in the group 60.5 – 61.5.

$$\therefore \text{ Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$= 60.5 + \frac{1}{42}(93.5 - 58) = 60.5 + 0.85 = 61.35 \text{ inches}$$

Thus $Q.D. = \frac{Q_3 - Q_1}{2} = \frac{62.69 - 60.12}{2} = \frac{2.57}{2} = 1.285$ inches,

and Coeff. of $Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{62.69-60.12}{62.69+60.12} = \frac{2.57}{122.81} = 0.02$

$$M.D. = \frac{\sum f\,|x - \text{Med}|}{n} = \frac{271.55}{187} = 1.45 \text{ inches;}$$

$$\text{Coeff of } M.D. = \frac{M.D.}{\text{Median}} = \frac{1.45}{61.35} = 0.024$$

Group B.

Q_1 = Height of $\left(\frac{n}{4}\right)th$ person

= Height of $\left(\frac{187}{4}\right)th$, *i.e.* 46.75*th* person which lies in the group 59.5 – 60.5

$$\therefore \quad Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right)$$

$$= 59.5 + \frac{1}{32}(46.75-35) = 59.5+0.37 = 59.87 \text{ inches;}$$

Q_3 = Height of $\left(\frac{3n}{4}\right)th$ person

= Height of $\frac{3\,(187)}{4}th$, *i.e.* 140.25*th* person which lies in the group 62.5 – 63.5

$$\therefore \quad Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) = 62.5 + \frac{1}{22}(140.25-135)$$

$$= 62.5+0.24 = 62.74 \text{ inches}$$

Similarly, we estimate the median as

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right) = 60.5 + \frac{1}{35}(93.5 - 67)$$

$$= 60.5 + 0.76 = 61.26 \text{ inches}$$

Now $Q.D. = \frac{Q_3 - Q_1}{2} = \frac{62.74 - 59.87}{2} = 1.435$ inches;

$$M.D. = \frac{\sum f\,|x - \text{Med}|}{n} = \frac{298.42}{187} = 1.60 \text{ inches;}$$

Coeff of $Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{62.74-59.87}{62.74+59.87} = \frac{1.435}{122.61} = 0.012$ and

$$\text{Coeff of } M.D. = \frac{M.D.}{\text{Median}} = \frac{1.60}{61.26} = 0.026$$

4.9. (b) Calculation of population variance and standard deviation

X_i	$X_i - \mu$	$(X_i - \mu)^2$
10	+2.5	6.25
8	+0.5	0.25
7	−0.5	0.25
9	+1.5	2.25
5	−2.5	6.25
12	+4.5	20.25
8	+0.5	0.25
6	−1.5	2.25
8	+0.5	0.25
2	−5.5	30.25
75	--	68.5

$$\therefore \quad \mu = \frac{\sum X_i}{N} = \frac{75}{10} = 7.5;$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N} = \frac{68.5}{10} = 6.85; \text{ and}$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = \sqrt{\frac{68.5}{10}} = \sqrt{6.85} = 2.62$$

4.10. (b) Calculation of sample mean and standard deviation.

x_i	$x - \bar{x}$	$(x - \bar{x})^2$
70	10	100
50	−10	100
60	0	0
70	10	100
50	−10	100
300	--	400

$$\therefore \quad \bar{x} = \frac{\sum x}{n} = \frac{300}{5} = 60, \text{ and}$$

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{400}{5}} = \sqrt{80} = 8.944$$

(i) Now, we calculate the sample mean and the standard deviation of the scores obtained by adding 10 points to them:

	$y(=x+10)$	$(y-\bar{y})$	$(y-\bar{y})^2$
	80	10	100
	60	−10	100
	70	0	0
	80	10	100
	60	−10	100
Σ	350	---	400

$$\therefore \quad \bar{y} = \frac{\sum y}{n} = \frac{350}{5} = 70 = \bar{x} + 10; \text{ and}$$

$$S_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} = \sqrt{\frac{400}{5}} = \sqrt{80} = 8.944 = S_x$$

(ii) Increasing all scores by 10% implies that each score is to be multiplied by $\frac{110}{100}$. The calculations then become:

	$y(=\frac{110}{100}x)$	$y-\bar{y}$	$(y-\bar{y})^2$
	77	11	121
	55	−11	121
	66	0	0
	77	11	121
	55	−11	121
Σ	330	---	484

$$\therefore \quad \bar{y} = \frac{\sum y}{n} = \frac{330}{5} = 66 = \frac{110}{100}(60) = a\bar{x}, \text{ where } a = \frac{110}{100};$$

$$S_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} = \sqrt{\frac{484}{5}} = \sqrt{96.8} = 9.838$$

$$= \frac{110}{100}(8.944) = a\, S_x$$

Hence we observe that

(i) when 10 is added to all scores, the mean is increased by 10 but standard deviation remains unchanged; and

(ii) when the scores are increased by 10%, *i.e.* multiplied by $\frac{110}{100}$, both the mean and standard deviation are multiplied by $\frac{110}{100}$. It is based on the following properites. If $y = ax + b$, then $\bar{y} = a\bar{x} + b$, and $S_y = |a|\, S_x$.

4.11. (b) Given $n = 15$, $\sum x = 480$ & $\sum x^2 = 15{,}735$. Therefore

$$\bar{x} = \frac{\sum x}{n} = \frac{480}{15} = 32; \text{ and}$$

$$S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{15735}{15} - \left(\frac{480}{15}\right)^2}$$

$$= \sqrt{1049 - 1024} = \sqrt{25} = 5$$

4.12. (b) Calculation of means and standard deviations

	x	$x-\bar{x}$	$(x-\bar{x})^2$
	3	−1	1
	6	2	4
	2	−2	4
	1	−3	9
	7	3	9
	5	1	1
Σ	24	---	28

$$\therefore \quad \bar{x} = \frac{\sum x}{n} = \frac{24}{6} = 4, \text{ and}$$

$$S_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{28}{6}} = \sqrt{4.6667} = 2.16$$

Now $\quad y = 2x + 5$. Then

	y	$y-\bar{y}$	$(y-\bar{y})^2$
	11	−2	4
	17	4	16
	9	−4	16
	7	−6	36
	19	6	36
	15	2	4
Σ	78	---	112

$$\bar{y} = \frac{\sum y}{n} = \frac{78}{6} = 13 = 2\,(4) + 5 = 2\bar{x} + 5, \text{ and}$$

$$S_y = \sqrt{\frac{\sum(y-\bar{y})^2}{n}} = \sqrt{\frac{112}{6}} = \sqrt{18.6667} = 4.32$$

$$= 2(2.16) = (2)\,S_x.$$

(c) (i) When the age of the youngest child is 1 year, the ages of children are 1, 2, 3, 4, 5, 6 and 7. Therefore

$$\text{mean} = \frac{\sum x}{n} = \frac{1 + 2 + \dots + 7}{7} = \frac{28}{7} = 4 \text{ year; and}$$

$$s.d. = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\left(\frac{1^2 + 2^2 + \dots + 7^2}{7}\right) - \left(\frac{28}{7}\right)^2}$$

$$= \sqrt{20 - 16} = \sqrt{4} = 2 \text{ years.}$$

(ii) When the youngest child is 8 years old, the ages of the children are 8, 9, 10, 11, 12, 13 and 14 years. Therefore

$$\text{Mean} = \frac{8 + 9 + \dots + 14}{7} = \frac{77}{7} = 11 \text{ years; and}$$

$$S.D. = \sqrt{\frac{8^2 + 9^2 + \dots + 14^2}{7} - \left(\frac{77}{7}\right)^2}$$

$$= \sqrt{125 - 121} = \sqrt{4} = 2 \text{ years.}$$

The standard deviations of (i) and (ii) coincide because they remain unaffected if a constant is added to the values of a variable. Here 7 is added to all values of (i).

4.13. Let the sample of size n consists of $x_1, x_2, \dots, x_n$ values. Then, assuming x_1 as the smallest observation and x_n, the largest observation, we compute the Range and the Standard deviation by the relations

Range $= x_n - x_1$, and

$$S.D. = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

When $n = 2$, then assuming $x_2 > x_1$, we have

Range $= x_2 - x_1$; and

$$S.D. = \sqrt{\frac{x_1^2 + x_2^2}{2} - \left(\frac{x_1 + x_2}{2}\right)^2}$$

$$= \sqrt{\frac{1}{4}\left[2x_1^2 + 2x_2^2 - x_1^2 - x_2^2 - 2x_1x_2\right]}$$

$$= \sqrt{\frac{1}{4}(x_1^2 + x_2^2 - 2x_1x_2)}$$

$$= \sqrt{\left(\frac{x_2 - x_1}{2}\right)^2} = \frac{x_2 - x_1}{2} = \frac{1}{2}\text{(Range)}$$

Hence we observe that Range and Standard Deviation are related when $n = 2$.

4.14. Here $M.D. = \dfrac{\sum f_i |x_i - \bar{x}|}{n} = \dfrac{1696.0}{60} = 28.27$ gms

(see Example 4.4)

and $S.D. = 34.87$ gms (see example 4.7 in the text)

$$\therefore \frac{\text{Mean Deviation}}{\text{Standard Deviation}} = \frac{28.27}{34.87} = 0.81$$

The statement is almost correct.

4.15. Calculation of the mean and the standard deviation.

x	f	D	fD	fD^2
30	4	−4	−16	64
31	8	−3	−24	72
32	23	−2	−46	92
33	35	−1	−35	35
34	62	0	−121	0
35	44	1	44	44
36	18	2	36	72
37	4	3	12	36
38	1	4	4	16
39	1	5	5	25
Total	200	--	+101 −20	456

$$\therefore \bar{x} = a + \frac{\sum fD}{n} = 34 + \frac{(-20)}{200} = 33.9, \text{ and}$$

$$s = \sqrt{\frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2} = \sqrt{\frac{456}{200} - \left(\frac{-20}{200}\right)^2}$$

$$= \sqrt{2.28 - 0.01} = \sqrt{2.27} = 1.507$$

4.16. The mean and standard deviation calculated for question 4.15 are

$$\bar{x} = 33.9, \text{ and } S = 1.507,$$

The interval "mean $\pm 2S$" is obtained as below:

$\bar{x} \pm 2S = 33.9 \pm 2(1.507) = 33.9 \pm 3.014$

$= 30.886$ and 36.914

This interval 30.886 − 36.914 will according to Chebyshev's rule, contain *at least* $\left(1 - \frac{1}{2^2}\right)$, *i.e.* $\frac{3}{4}$ of 200 = 150 metal bars. The number of metal bars lying within this interval is 190 which is greater than 150.

4.17. Calculation of means and standard deviations of the expenditure.

Expenditure Rs.	x_i	u_i	Place A			Place B		
			f_1	f_1u	f_1u^2	f_2	f_2u	f_2u^2
30–60	45	−2	28	−56	112	39	−78	156
60–90	75	−1	292	−292	292	284	−284	284
90–120	105	0	389	−348	0	401	−362	0
120–150	135	1	212	212	212	202	202	202
150–180	165	2	59	118	236	48	96	192
180–210	195	3	18	54	162	31	93	279
210–240	225	4	2	8	32	5	20	80
Total	--	--	1000	+392 44	1046	1010	+411 49	1193

Place A:

$$\bar{x} = a + \frac{\sum f_1 u}{n} \times h, \text{ where } u = \frac{x - 105}{30}.$$

$$= \text{Rs. } 105 + \frac{44}{1000} \times 30 = \text{Rs. } 105 + 1.32 = \text{Rs. } 106.32$$

$$s = h \times \sqrt{\frac{\sum f_1 u^2}{n} - \left(\frac{\sum f_1 u}{n}\right)^2}$$

$$= 30 \times \sqrt{\frac{1046}{1000} - \left(\frac{44}{1000}\right)^2} = 30 \times \sqrt{1.046 - 0.0019}$$

$$= 30 \times 1.02 = \text{Rs. } 30.6$$

Place B:

$$\bar{x} = a + \frac{\sum f_2 u}{n} \times h$$

$$= \text{Rs. } 105 + \frac{49}{1010} \times 30 = \text{Rs. } 105 + 1.46 = \text{Rs. } 106.46$$

$$s = h \times \sqrt{\frac{\sum f_2 u^2}{n} - \left(\frac{\sum f_2 u}{n}\right)^2}$$

$$= 30 \times \sqrt{\frac{1193}{1010} - \left(\frac{49}{1010}\right)^2} = 30 \times \sqrt{1.1812 - 0.0025}$$

$$= 30 \times \sqrt{1.1787} = 30 \times 1.09 = \text{Rs. } 32.7$$

4.18. Calculation of the mean wage and the standard deviation.

Classes (Rs.)	Midvalues (x)	u	f	fu	fu^2
4.50–5.50	5.00	–7	6	–42	294
5.50–6.50	6.00	–6	17	–102	612
6.50–7.50	7.00	–5	35	–175	875
7.50–8.50	8.00	–4	48	–192	768
8.50–9.50	9.00	–3	65	–195	585
9.50–10.50	10.00	–2	90	–180	360
10.50–11.50	11.00	–1	131	–131	131
				–1017	
11.50–12.50	12.00	0	173		0
12.50–13.50	13.00	1	155	155	155
13.50–14.50	14.00	2	117	234	468
14.50–15.50	15.00	3	75	225	675
15.50–16.50	16.00	4	52	208	832
16.50–17.50	17.00	5	21	105	525
17.50–18.50	18.00	6	9	54	324
18.50–19.50	19.00	7	6	42	294
Total	----	---	1000	$\frac{+1023}{6}$	6898

Mean wage or $\bar{x}$ $= a + \frac{\sum fu}{n} \times h$, where $u = x - 12.00$ and $h = 1$.

$$= \text{Rs. } 12.00 + \frac{6}{1000} = \text{Rs. } 12.00\dot{6}, \text{ and}$$

$$s = \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}$$

$$= \sqrt{\frac{6898}{1000} - \left(\frac{6}{1000}\right)^2} = \sqrt{6.898 - 0.00036}$$

$$= \sqrt{6.897964} = \text{Rs. } 2.626$$

4.19. Calculation of the mean and the standard deviation.

	x	$u(=x-90)$	u^2
	95	5	25
	73	–17	289
	82	–8	64
	108	18	324
	103	13	169
	78	–12	144
	79	–11	121
	94	4	16
	97	7	49
	95	5	25
	69	–21	441
	87	–3	9
	130	40	1600
	89	–1	1
	67	–23	529
	93	3	9
	96	6	36
	68	–22	484
	83	–7	49
	117	27	729
Total	1803	3	5113

Now $\bar{x} = a + \frac{\sum u}{n} = 90 + \frac{3}{20} = 90.15$; and

$$S = \sqrt{\frac{\sum u^2}{n} - \left(\frac{\sum u}{n}\right)^2}$$

$$= \sqrt{\frac{5113}{20} - \left(\frac{3}{20}\right)^2} = \sqrt{255.65 - 0.0225}$$

$$= \sqrt{255.6275} = 15.99$$

(i) Mean $\pm S = 90.15 \pm 15.99 = 106.14, 74.16$

Observations lying within these limits = 13

$\therefore$ Percentage of observations $= \frac{13}{20} \times 100 = 65\%$

(ii) Mean $\pm 2S = 90.15 \pm 2(15.99)$

$$= 90.15 \pm 31.98 = 122.13, 58.17$$

Observations lying within these limits = 19

$\therefore$ Percentage of observations $= \frac{19}{20} \times 100 = 95\%$

(iii) Mean $\pm 3S = 90.15 \pm 3(15.99)$

$$= 90.15 \pm 47.97 = 138.12, 42.18$$

Observations lying within these limits = 20

Hence %age of observations $= \frac{20}{20} \times 100 = 100\%$.

4.20. Computation of the mean, the standard deviation, etc.

Midvalues (inches) x	No. of Students (f)	u $\left(= \frac{x-14.5}{0.5}\right)$	fu	fu^2
12.5	4	−4	−16	64
13.0	19	−3	−57	171
13.5	30	−2	−60	120
14.0	63	−1	−63	63
14.5	66	0	−196	0
15.0	29	1	29	29
15.5	18	2	36	72
16.0	1	3	3	9
16.5	1	4	4	16
Total	231	--	+72 / −124	544

$$\therefore \quad \bar{x} = a + \frac{\sum fu}{n} \times h = 14.5 + \frac{(-124)}{231} \times 0.5$$

$$= 14.5 - 0.27 = 14.23 \text{ inches}$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}$$

$$= 0.5 \times \sqrt{\frac{544}{231} - \left(\frac{-124}{231}\right)^2} = 0.5 \times \sqrt{2.35 - 0.28}$$

$$= 0.5 \times \sqrt{2.07} = 0.72 \text{ inches.}$$

Using the criterion $\bar{x} \pm 3s$, we get

(i). the largest size of the collars $= \bar{x} + 3s + \frac{3}{4}$ inches

$$= 14.23 + 3(0.72) + 0.75$$

$$= 17.14 \text{ inches}$$

(ii) the smallest size of the collars $= \bar{x} - 3s + \frac{3}{4}$ inches

$= 14.23 - 3(0.72) + 0.75$

$= 12.82$ inches.

4.21. Determination of the actual classes.

u	f	fu	fu^2	Actual Classes
−4	2	−8	32	109.5–115.5
−3	5	−15	45	115.5–121.5
−2	8	−16	32	121.5–127.5
−1	18	−18	18	127.5–133.5
0	22	−57	0	133.5–139.5
1	13	13	13	139.5–145.5
2	8	16	32	145.5–151.5
3	4	12	36	151.5–157.5
Σ	80	$\frac{+41}{6}$	208	---

Substituting the values in the formula

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}, \text{ we get}$$

$$9.6 = h \times \sqrt{\frac{208}{80} - \left(\frac{-16}{80}\right)^2} = h \times \sqrt{2.60 - 0.04} = 1.6h$$

$\therefore \quad h = 9.6 \div 1.6 = 6$, and

mean $= a + \frac{\sum fu}{n} \times h$

i.e. $135.3 = a + \frac{(-16)}{80} \times 6 = a - 1.2$

$\therefore \quad a = 135.3 + 1.2 = 136.5$

Thus the midpoint of the actual class corresponding to the frequency 22 is 136.5. As the lenght of the class-interval is 6, therefore this class is (136.5–3.0) to (136.5+3.0), *i.e.* 133.5–139.5.

The other classes are then determined by adding to and subtracting from these class limits, the width of the class interval repeatedly. The classes thus determined are shown in the last column of the table on page 65.

4.22. Calculation of the mean, standard deviation, etc.

Source A					Source B				
Life (hrs)	No. of components (f)	u	fu	fu^2	Life (hours)	(f)	u	fu	fu^2
1000–1020	40	−3	−120	360	1030–1040	339	−2	−678	1356
1020–1040	96	−2	−192	384	1040–1050	136	−1	−136	136
1040–1060	364	−1	−364	364	1050–1060	25	0	−814	0
1060–1080	372	0	−676	0	1060–1070	20	1	20	20
1080–1100	85	1	85	85	1070–1080	130	2	260	520
1100–1120	43	2	86	172	1080–1090	350	3	1050	3150
Total	1000	--	+171 / −505	1365	---	1000	--	+1330 / +516	5182

Source A:

Mean life, *i.e.* $\bar{x} = a + \frac{\sum fu}{n} \times h = 1070 + \frac{(-505)}{1000} \times 20$

$= 1070 - 10 = 1060$ hours.

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 20 \times \sqrt{\frac{1365}{1000} - \left(\frac{-505}{1000}\right)^2}$$

$$= 20 \times \sqrt{1.365 - 0.255} = 20 \times \sqrt{1.110} = 21.1 \text{ hours}$$

Source B:

$$\bar{x} = a + \frac{\sum fu}{n} \times h$$

$$= 1055 + \frac{516}{1000} \times 10 = 1055 + 5.16 = 1060 \text{ hours}$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 10 \times \sqrt{\frac{5182}{1000} - \left(\frac{516}{1000}\right)^2}$$

$$= 10 \times \sqrt{5.182 - 0.2663} = 22.2 \text{ hours.}$$

We observe that the mean lives in hours of the components of the two sources are equal and the two sources have nearly the same dispersion, but these data give a false impression as the distribution of source B is U-shaped.

4.24. (b) By definition, $C.V. = \frac{s}{\bar{x}} \times 100$

Now $\bar{x} = a + \frac{\sum fu}{n} \times h = 62 + \frac{140}{210} \times 5 \quad (\because u = \frac{x-62}{5})$

$$= 62 + 5.83 = 67.83, \text{ and}$$

$$s^2 = h^2 \left[\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2\right] \text{ where } n = \sum f = 120$$

$$= 25\left[\frac{598}{120} - \left(\frac{140}{120}\right)^2\right] = 25(4.9833 - 1.3611) = 90.555$$

And $s^2(\text{corrected}) = s^2(\text{uncorrected}) - \frac{h^2}{12}$

$$= 90.555 - \frac{25}{12} = 90.555 - 2.0833 = 88.4717, \text{ so that}$$

s (corrected) $= \sqrt{88.4717} = 9.41$

Hence C.V. $= \frac{9.41}{67.83} \times 100 = 13.87\%$

4.25. (b) Computation of standard deviation for the data in locality A.

Income (Rs.)	f	x	u	fu	fu^2
35–39	13	37	–3	–39	117
40–44	15	42	–2	–30	60
45–49	17	47	–1	–17	17
50–54	28	52	0	–86	0
55–59	12	57	1	12	12
60–64	10	62	2	20	40
65–69	5	67	3	15	45
Total	100	--	--	+47 –39	291

Now $\bar{x} = a + \frac{\sum fu}{n} \times h$, where $u = \frac{x_i - 52}{5}$

$$= 52 + \frac{(-39)}{100} \times 5 = 52 - 1.95 = \text{Rs. } 50.05; \text{ and}$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 5 \times \sqrt{\frac{291}{100} - \left(\frac{-39}{100}\right)^2}$$

$$= 5 \times \sqrt{2.91 - 0.1521} = 5 \times \sqrt{2.7579} = \text{Rs. } 8.3.$$

$\therefore$ C.V. for locality $A = \frac{s}{\bar{x}} \times 100 = \frac{8.3}{50.05} \times 100 = 16.58\%$

Again C.V. for locality $B = \frac{s}{\bar{x}} \times 100 = \frac{4.96}{52.28} \times 100 = 9.50\%$

Locality A has greater variability as the coefficient of variation for A is larger than that for B.

4.26. (b) Computation of the co-efficients of variation for the candidates X and Y.

Paper	X	$X-\bar{X}$	$(X-\bar{X})^2$	Y	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
I	58	–4.7	22.09	39	–24.4	595.36
II	49	–13.7	187.69	38	–25.4	645.16
III	76	13.3	176.89	86	22.6	510.76
IV	80	17.3	299.29	72	8.6	73.96
V	47	–15.7	246.49	75	11.6	134.56
VI	72	9.3	86.49	69	5.6	31.36
VII	61	–1.7	2.89	57	–6.4	40.96
VIII	59	–3.7	13.69	49	–14.4	207.36
IX	77	14.3	204.49	83	19.6	384.16
X	48	–14.7	216.09	66	2.6	6.76
Total	627	--	1456.10	634	--	2630.40

Candidate X:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{627}{10} = 62.7 \text{ marks}$$

$$S_x = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}} = \sqrt{\frac{1456.10}{10}} = \sqrt{145.61} = 12.07 \text{ marks}$$

$\therefore$ $C.V.$ for candidate $X = \frac{S_x}{\bar{X}} \times 100 = \frac{12.07}{62.7} \times 100 = 19.25\%$

Candidate Y:

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{634}{10} = 63.4 \text{ marks}$$

$$S_y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n}} = \sqrt{\frac{2630.40}{10}} = \sqrt{263.04} = 16.22 \text{marks}$$

$\therefore$ $C.V.$ for candidate $Y = \frac{S_y}{\bar{Y}} \times 100 = \frac{16.22}{63.4} \times 100 = 25.58\%$

The performance of the candidate X is more consistent than that of the candidate Y as the $C.V.$ for X is smaller than that for Y.

4.27 (a) Here $\bar{x} = 67.45$, s^2 (uncorrected) = 8.5275, and h = 3

$$\therefore \quad s^2 \text{ (corrected)} = s^2 \text{ (uncorrected)} - \frac{h^2}{12}$$

$$= 8.5275 - \frac{9}{12} = 7.7775, \text{ and}$$

$$s \text{ (corrected)} = \sqrt{7.7775} = 2.79.$$

Hence $C.V. \text{ (corrected)} = \frac{s \text{ (corrected)}}{\text{mean}} \times 100$

$$= \frac{2.79}{67.45} \times 100 = 4.14\%$$

(b) Calculation of the means, standard deviations and the coefficients of variation for Batsmen A and B.

	Batsman A		Batsman B	
	Scores x	x^2	Scores y	y^2
	12	144	47	2209
	15	225	12	144
	6	36	76	5776
	73	5329	48	2304
	7	49	4	16
	19	361	51	2601
	199	39601	37	1369
	36	1296	48	2304
	84	7056	13	169
	29	841	0	0
Σ	480	54938	336	16892

Bastman A:

Mean scores $= \frac{\sum x_i}{n} = \frac{480}{10} = 48$ scores,

$$S_x = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{54938}{10} - \left(\frac{480}{10}\right)^2}$$

$$= \sqrt{5493.80 - 2304} = \sqrt{3189.8} = 56.48 \text{ scores, and}$$

$$C.V. = \frac{S_x}{\bar{x}} \times 100 = \frac{56.48}{478} \times 100 = 117.67\%$$

Bastman B:

Mean scores $= \frac{\sum y_i}{n} = \frac{336}{10} = 33.6$ scores,

$$S_y = \sqrt{\frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2} = \sqrt{\frac{16892}{10} - \left(\frac{336}{10}\right)^2}$$

$$= \sqrt{1689.20 - 1128.96} = \sqrt{560.24} = 23.67 \text{ scores, and}$$

$$C.V. = \frac{S_y}{\bar{y}} \times 100 = \frac{23.67}{33.6} \times 100 = 70.45\%$$

Batsman A is better as a run getter as A's average score is 48 and B's average score is 33.6.

Batsman B is a more consistent player as the coefficient of variation for batsman B is smaller than that for A.

4.28. (b) Calculation of coefficient of variation.

Weight (lb)	x	f	u	fu	fu^2
118–126	122	3	–3	–9	27
127–135	131	5	–2	–10	20
136–144	140	9	–1	–9	9
145–153	149	12	0	–28	0
154–162	158	5	1	5	5
163–171	167	4	2	8	16
172–180	176	2	3	6	18
Total	---	40	--	+19 –9	95

$$\bar{x} = a + \frac{\sum fu}{n} \times h, \text{ where } h = 9, \text{ and } u = \frac{x - 149}{9}$$

$$= 149 + \frac{(-9)}{40} \times 9 = 149 - 2.025 = 146.975 \text{ lbs.}$$

$$s^2 \text{ (uncorrected)} = h^2 \left[\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2\right]$$

$$= 81 \left[\frac{95}{40} - \left(\frac{-9}{40}\right)^2\right]$$

$$= 81\,[2.375 - 0.0506] = 81 \times (2.3244)$$

$$s \text{ (uncorrected)} = 9 \times \sqrt{2.3244} = 13.716 \text{ lbs; and}$$

$$C.V. = \frac{s}{\text{mean}} \times 100 = \frac{13.716}{146.975} \times 100 = 9.33\%$$

$$\text{Now, } s^2 \text{ (corrected)} = s^2 \text{ (uncorrected)} - \frac{h^2}{12}$$

$$= 81(2.3244) - \frac{81}{12}$$

$$= 81[2.3244 - 0.0833] = 81 \times (2.2411)$$

$\therefore$ $s \text{ (corrected)} = 9 \times \sqrt{2.2411} = 9 \times 1.497 = 13.473$ lbs; and

$$C.V. \text{(corrected)} = \frac{s \text{ (corrected)}}{\text{mean}} \times 100 = \frac{13.473}{146.975} \times 100 = 9.17\%$$

(c) $C.V. \text{ for tube } A = \frac{280}{1495} \times 100 = 18.8\%$

$C.V. \text{ for tube } B = \frac{310}{1895} \times 100 = 16.4\%$

(i) Tube B has a greater absolute dispersion as $S_B > S_A$

(ii) Tube A has a greater relative dispersion as the co-efficient of variation for A is larger than that for B.

4.29. Computation of co-efficients of variation.

Expenditure			Town A			Town B		
(Rupees)	x	u	f	fu	fu^2	f	fu	fu^2
21–30	25.5	–3	3	–9	27	2	–6	18
31–40	35.5	–2	61	–122	244	14	–28	56
41–50	45.5	–1	132	–132	132	20	–20	20
51–60	55.5	0	153	–263	0	27	–54	0
61–70	65.5	1	140	140	140	28	28	28
71–80	75.5	2	51	102	204	7	14	28
81–90	85.5	3	2	6	18	2	6	18
Total	--	--	542	$\frac{+248}{-1.5}$	765	100	$\frac{+48}{-6}$	168

Town A:

$$\bar{x} = a + \frac{\sum fu}{n} \times h = 55.5 + \frac{(-15)}{542} \times 10$$

$$= 55.5 - 0.28 = \text{Rs. } 55.22;$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}$$

$$= 10 \times \sqrt{\frac{765}{542} - \left(\frac{-15}{542}\right)^2} = 10 \times \sqrt{1.4114 - 0.0009}$$

$$= 10 \times \sqrt{1.4105} = 10 \times (1.1876) = \text{Rs. } 11.88; \text{ and}$$

$$C.V. = \frac{s}{\bar{x}} \times 100 = \frac{\text{Rs. } 11.88}{\text{Rs. } 55.22} \times 100 = 21.51\%$$

Town B:

$$\bar{x} = a + \frac{\sum fu}{n} \times h = 55.5 + \frac{(-6)}{100} \times 10$$

$$= 55.5 - 0.6 = \text{Rs. } 54.9;$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}$$

$$= 10 \times \sqrt{\frac{168}{100} - \left(\frac{-6}{100}\right)^2} = 10 \times \sqrt{1.68 - 0.0036}$$

$$= 10 \times \sqrt{1.6764} = 10 \times (1.2948) = \text{Rs. } 12.95; \text{ and}$$

$$C.V. = \frac{s}{\bar{x}} \times 100 = \frac{12.95}{54.9} \times 100 = 23.59\%$$

The co-efficient of variation for town B is larger than that for town A. Thus there is greater variability in expenditures of families in town B than that of town A.

4.30. Computation of co-efficients of variation.

Weight (kilograms)	u	Class A f	Class A fu	Class A fu^2	Class B f	Class B fu	Class B fu^2	Class C f	Class C fu	Class C fu^2
25	−2	7	−14	28	5	−10	20	6	−12	24
35	−1	10	−10	10	9	−9	9	25	−25	25
45	0	20	−24	0	21	−19	0	24	−37	0
55	1	18	18	18	15	15	15	4	4	4
65	2	7	14	28	6	12	24	3	6	12
Total	--	62	+32 / 8	84	56	+27 / 8	68	62	+10 / −27	65

Class A:

$$\bar{x} = a + \frac{\sum fu}{n} \times h, \text{ where } h = 10 \text{ and } u = \frac{x-45}{10}$$

$$= 45 + \frac{8}{62} \times 10 = 45 + 1.29 = 46.29 \text{ kilograms;}$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 10 \times \sqrt{\frac{84}{62} - \left(\frac{8}{62}\right)^2}$$

$$= 10 \times \sqrt{1.3548 - 0.0166} = 10 \times \sqrt{1.3382} = 11.57 \text{ kg; and}$$

$$C.V. = \frac{s}{\bar{x}} \times 100 = \frac{11.57}{46.29} \times 100 = 24.99\%$$

Class B:

$$\bar{x} = a + \frac{\sum fu}{n} \times h$$

$$= 45 + \frac{8}{56} \times 10 = 45 + 1.43 = 46.43 \text{ kg;}$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 10 \times \sqrt{\frac{68}{56} - \left(\frac{8}{56}\right)^2}$$

$$= 10 \times \sqrt{1.2143 - 0.0204} = 10 \times 1.093 = 10.93 \text{ kg; and}$$

$$C.V. = \frac{s}{\bar{x}} \times 100 = \frac{10.93}{46.43} \times 100 = 23.54\%$$

Class C:

$$\bar{x} = a + \frac{\sum fu}{n} \times h$$

$$= 45 + \frac{(-27)}{62} \times 10 = 45 - 4.35 = 40.65 \text{ kg;}$$

$$s = h \times \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 10 \times \sqrt{\frac{65}{62} - \left(\frac{-27}{62}\right)^2}$$

$$= 10 \times \sqrt{1.0484 - 0.1896} = 10 \times 0.927 = 9.27 \text{ kg; and}$$

$$C.V. = \frac{s}{\bar{x}} \times 100 = \frac{9.27}{40.65} \times 100 = 22.80\%$$

4.31. By definition, the variance, s^2, is

$$s^2 = \frac{\sum n_r (x_r - \bar{x})^2}{\sum n_r}, \text{ where } n_r \text{ denotes the frequency and } \bar{x} = \frac{\sum n_r x_r}{\sum n_r}$$

$$= \frac{\sum n_r (x_r - k + k - \bar{x})^2}{\sum n_r}, \text{ where } k \text{ is any arbitrary number.}$$

$$= \frac{1}{\sum n_r} \left[\sum n_r \{(x_r - k)^2 + (\bar{x} - k)^2 - 2(x_r - k)(\bar{x} - k)\}\right]$$

$$= \frac{\sum n_r (x_r - k)^2}{\sum n_r} + \frac{(\bar{x}_r - k)^2 \sum n_r}{\sum n_r} - 2(\bar{x} - k) \frac{\sum n_r (x_r - k)}{\sum n_r}$$

$$= \frac{\sum n_r (x_r - k)^2}{\sum n_r} + (\bar{x} - k)^2 - 2(\bar{x} - k)^2$$

$$= \frac{\sum n_r (x_r - k)^2}{\sum n_r} - (\bar{x} - k)^2$$

Hence $s = \sqrt{\dfrac{\sum n_r (x_r - k)^2}{\sum n_r} - \delta^2}$, where $\bar{x} = k + \delta$.

4.32. (a) Here $n_1 = 50, \bar{x}_1 = 59.5, S_1 = 8.38,$
$n_2 = 40, \bar{x}_2 = 54.0, S_2 = 8.23.$

Let $\bar{x}$ and S denote the mean and standard deviation respectively of the combined group of children. Then

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

$$= \frac{(50)(59.5) + (40)(54.0)}{50 + 40} = \frac{5135}{90} = 57.06; \text{ and}$$

$$S^2 = \frac{1}{n_1+n_2}\left[n_1S_1^2 + n_1(\bar{x}_1 - \bar{x})^2 + n_2S_2^2 + n_2(\bar{x}_2 - \bar{x})^2\right]$$

$$= \frac{1}{90}[(50)(8.38)^2 + 50(59.5 - 57.06)^2 + (40)(8.23)^2 + 40(54.0 - 57.06)^2]$$

$$= \frac{1}{90}[3511.22 + 297.68 + 2709.32 + 374.54]$$

$$= \frac{6892.76}{90} = 76.5862$$

Hence $S = \sqrt{76.5862} = 8.75$

(b) Here $n_1 = 200, \bar{x}_1 = 25, s_1 = 3,$
$n_2 = 250, \bar{x}_2 = 10, s_2 = 4,$
$n_3 = 300, \bar{x}_3 = 15, s_3 = 5.$

Let $\bar{x}$ and s denote the combined mean and the standard deviation respectively of the combined distribution. Then

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$$

$$= \frac{200(25) + 250(10) + 300(15)}{200 + 250 + 300} = \frac{12000}{750} = 16; \text{ and}$$

$$s^2 = \frac{1}{\sum n_i}\left[\sum n_i \{s_i^2 + (\bar{x}_i - \bar{x})^2\}\right]$$

$$= \frac{1}{750}[200\{9 + (25 - 16)^2 + 250\{16 + (10 - 16)^2\} + 300\{25 + (15 - 16)^2\}]$$

$$= \frac{1}{750}[18000 + 13000 + 7800] = 51.73, \text{ so that}$$

$$s = \sqrt{51.73} = 7.2; \text{ and hence}$$

$$C.V. = \frac{s}{\bar{x}} \times 100 = \frac{7.2}{16} \times 100 = 45\%$$

4.33 (c) Z-score for the top student $= 50 + 10\left(\dfrac{x - \bar{x}}{S}\right)$

$$= 50 + 10\left(\frac{98-63.7}{12.3}\right)$$

$= 50 + 28 = 78$, and

Z-score for the bottom student $= 50 + 10\left(\dfrac{x - \bar{x}}{S}\right)$

$$= 50 + 10\left(\frac{21-63.7}{12.3}\right)$$

$= 50 - 35 = 15.$

4.34. By definition, a standard Z-score $= 50 + 10\left(\dfrac{x - \bar{x}}{S}\right)$

Student A:

Z-score on the 1st test $= 50 + 10\left(\dfrac{70 - 70}{5}\right) = 50$

Z-score on the 2nd test $= 50 + 10\left(\dfrac{90 - 75}{8}\right) = 69$

$$Z\text{-score on the 3rd test} = 50 + 10\left(\frac{70-60}{12}\right) = 58$$

$$\therefore \text{ Average score of student } A = \frac{50 + 69 + 58}{3} = 59$$

Student B:

$$Z\text{-score on the 1st test} = 50 + 10\left(\frac{90-70}{5}\right) = 90$$

$$Z\text{-score on the 2nd test} = 50 + 10\left(\frac{70-75}{8}\right) = 44$$

$$Z\text{-score on the 3rd test} = 50 + 10\left(\frac{70-60}{12}\right) = 58$$

$$\therefore \text{ Average score of student } B = \frac{90 + 44 + 58}{3} = 64$$

4.35 (b) The data ordered from smallest to largest and the two quartiles are found to be

$$42, 43, 58, \underset{\substack{\uparrow \\ Q_1}}{63}, 65, 67, 68, 72, 75, 75, 78, \underset{\substack{\uparrow \\ Q_3}}{79}, 80, 82, 96.$$

To find the trimmed mean and the trimmed standard deviation, we remove the three observations 42, 43 and 58 below the first quartile and the three observations 80, 82 and 96 above the third quartile. Thus we have nine observations 63, 65, 67, 68, 72, 75, 75, 78, 79 as trimmed data set.

$$\therefore \text{ Trimmed mean } = \frac{63 + 65 + \dots + 79}{9} = \frac{642}{9} = 71.33, \text{ and}$$

$$\text{Trimmed } S.D. = \sqrt{\frac{(63)^2 + (65)^2 + \dots + (79)^2}{9} - \left(\frac{642}{9}\right)^2}$$

$$= \sqrt{\frac{46066}{9} - \left(\frac{642}{9}\right)^2}$$

$$= \sqrt{5118.4444 - 5088.4444} = \sqrt{30} = 5.48$$

To find the Winsorized mean and standard deviation, we replace the three observations 42, 43, 58 below Q_1 with 63 and the three observations 80, 82, 96 above Q_3 with 79; and get the Winsorized data set as 63, 63, 63, 63, 65, 67, 68, 72, 75, 75, 78, 79, 79, 79, 79.

$$\text{The Winsorized mean} = \frac{\sum X_i}{n} = \frac{1068}{15} = 71.2, \text{ and}$$

$$\text{the Winsorized } S.D = \sqrt{\frac{76696}{15} - \left(\frac{1068}{15}\right)^2}$$

$$= \sqrt{5113.0667 - 5069.44} = \sqrt{43.6267} = 6.61$$

4.39 (a) The moments about the mean are obtained as below:

$$m_1 = 0$$

$$m_2 = m_2' - (m_1')^2 = 2.5 - (1)^2 = 1.5$$

$$m_3 = m_3' - 3m_2' m_1' + 2(m_1')^3$$

$$= 5.5 - 3(1)(2.5) + 2(1)^3 = 0, \text{ and}$$

$$m_4 = m_4' - 3m_1' m_3' + 6(m_1')^2 m_2' - 3(m_1')^4$$

$$= 16 - 4(1)(5.5) + 6(1)^2(2.5) - 3(1)^4 = 6$$

Again, we have

$$m_1' = \frac{1}{n}\sum f(x-2) \quad \dots (1) \qquad m_2' = \frac{1}{n}\sum f(x-2)^2 \quad \dots (2)$$

$$m_3' = \frac{1}{n}\sum f(x-2)^3 \quad \dots (3) \qquad m_4' = \frac{1}{n}\sum f(x-2)^4 \quad \dots (4)$$

To find the moments about $x=0$, we need the values of $\frac{1}{n}\sum fx$, $\frac{1}{n}\sum fx^2$, $\frac{1}{n}\sum fx^3$ and $\frac{1}{n}\sum fx^4$, which are obtained from the relations (1), (2), (3) and (4).

Thus from (1), we get

$$1 = \frac{\sum fx}{n} - 2 \text{ or } \frac{\sum fx}{n} = 3$$

From (2), we get

$$2.5 = \frac{1}{n}\sum f(x^2 - 4x + 4)$$

or $$2.5 = \frac{1}{n}\sum fx^2 - 4\frac{\sum fx}{n} + 4, \quad \therefore \frac{1}{n}\sum fx^2 = 10.5$$

From (3), we have

$$5.5 = \frac{1}{n}\sum f(x^3 - 6x^2 + 12x - 8)$$

i.e. $$\frac{1}{n}\sum fx^3 - 6\frac{1}{n}\sum fx^2 + 12\frac{\sum fx}{n} - 8 = 5.5$$

or $$\frac{1}{n}\sum fx^3 = 5.5 + 6(10.5) - 12(3) + 8 = 40.5$$

Similarly, from (4) we get $\frac{1}{n}\sum fx^4 = 168$.

Hence the moments about $x = 0$ are

$$m_1' = \frac{\sum fx}{n} = 3, \qquad m_2' = \frac{\sum fx^2}{n} = 10.5,$$

$$m_3' = \frac{\sum fx^3}{n} = 40.5, \text{ and} \qquad m_4' = \frac{\sum fx^4}{n} = 168.$$

(b) $$m_1' = \frac{\sum fu}{\sum f} = \frac{-46}{125} = -0.368$$

$$m_2' = \frac{\sum fu^2}{n} = \frac{306}{125} = 2.448$$

$$m_3' = \frac{\sum fu^3}{n} = \frac{-242}{125} = -1.936$$

$$m_4' = \frac{\sum fu^4}{n} = \frac{1962}{125} = 15.696$$

$$\text{Mean or } \bar{x} = a + \frac{\sum fu}{n} \times h$$

$$= 10 + \left(\frac{-46}{125}\right) \times 5 = 10 - 1.84 = 8.16$$

The moments about mean are

$$m_1 = 0$$

$$m_2 = h^2[m_2' - (m_1')^2] = (5)^2[2.448 - (-0.368)^2]$$

$$= 25(2.313) = 57.825$$

$$m_3 = h^3[m_3' - 3m_1'm_2' + 2(m_1')^3]$$

$$= (5)^3[-1.936 - 3(-0.368)(2.448) + 2(-0.368)^3]$$

$$= 125[-1.936 + 2.703 - 0.100] = 125(0.667) = 83.375$$

$$m_4 = h^4[m_4' - 4m_1'm_3' + 6(m_1')^2 m_2' - 3(m_1')^4]$$

$$= (5)^4[15.696 - 4(-0.368)(-1.936) + 6(-0.368)^2 \times (2.448) - 3(-0.368)^4]$$

$$= 625[15.696 - 2.850 + 1.989 - 0.055] = 625(14.78) = 9237.5$$

Variance $= m_2 = 57.825$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(83.375)^2}{(57.825)^3} = \frac{6951.3906}{193351.22} = 0.04, \text{ and}$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{9237.5}{(57.825)^2} = \frac{9237.5}{3343.7306} = 2.76.$$

As the value of $b_1 \neq 0$ and $b_2 < 3$, *i.e.* the distribution is platy-kurtic, therefore the distribution would not be considered as normal.

4.40. Calculation of the moments, etc.

x	f	D	fD	fD^2	fD^3	fD^4
1	1	−4	−4	16	-64	256
2	6	−3	−18	54	-162	486
3	13	−2	−26	52	-104	208
4	25	−1	−25	25	-25	25
5	30	0	−73	0	-355	0
6	22	1	22	22	22	22
7	9	2	18	36	72	144
8	5	3	15	45	135	405
9	2	4	8	32	128	512
Σ	113	--	+63 / −10	282	357 / +2	2058
$\frac{\text{Sum}}{n}$	1	---	$-0.089 = m_1'$	$2.496 = m_2'$	$0.018 = m_3'$	$18.212 = m_4'$

$\therefore$ Moments about the mean are

$m_1 = 0$

$m_2 = m_2' - (m_1')^2 = 2.496 - (0.089)^2 = 2.49$

$m_3 = m_3' - 3m_1'm_2' + 2\,(m_1')^3$

$= (0.018) - 3(-0.089)\,(2.496) + 2(-0.089)^2 = 0.7$

$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2\,m_2' + 3(m_1')^4$

$= 18.212\text{-}4(-0.089)(0.018)+6(-0.089)^2(2.496)\text{-}3(-0.089)^4$

$= 18.33$

4.41 Calculation of the first four moments.

Weekly wages x	No. of Labourers (f)	D $(x-20)$	fD	fD^2	fD^3	fD^4
15	6	−5	−30	150	-750	3750
16	19	−4	−76	304	-1216	4864
17	13	−3	−39	117	-351	1053
18	18	−2	−36	72	-144	288
19	20	−1	−20	20	-20	20
20	25	0	−201	0	-2481	0
21	28	1	28	28	28	28
22	34	2	68	136	272	544
23	22	3	66	198	594	1782
24	15	4	60	240	960	3840
Total	200	--	+222 / +21	1265	+1854 / -627	14169
$\frac{\text{Sum}}{n}$	1	---	$0.105 = m_1'$	$6.325 = m_2'$	$-3.135 = m_3'$	$70.845 = m_4'$

$\therefore$ **Moments about the mean are:**

$m_1 = 0$

$m_2 = m_2' - (m_1')^2 = 6.325 - (0.105)^2 = 6.314$

$m_3 = m_3' - 3m_1'm_2' + 2\,(m_1')^3$

$= -3.135 - 3(0.105)\,(6.325) + 2(0.105)^3 = -5.125$

$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2\,m_2' - 3(m_1')^4$

$= 70.845 - 4(0.105)(-3.135) + 6(0.105)^2\,(6.325) - 3(0.105)^4$

$= 82.58$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(-5.125)^2}{(6.314)^3} = 0.104, \text{ and}$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{82.58}{(6.314)^2} = 2.071.$$

As $b_1 \neq 0$ and $b_2 < 3$, therefore the distribution is not normal.

4.42. Calculation of the first four moments.

Groups	f	x	$u(=\frac{x-11}{2})$	fu	fu^2	fu^3	fu^4
2–4	18	3	–4	–72	288	–1152	4608
4–6	24	5	–3	–72	216	–648	1944
6–8	47	7	–2	–94	188	–376	752
8–10	80	9	–1	–80	80	–80	80
10–12	102	11	0	–318	0	–2256	0
12–14	66	13	1	66	66	66	66
14–16	40	15	2	80	160	320	640
16–18	21	17	3	63	189	567	1701
18–20	15	19	4	60	240	960	3840
Total	413	--	--	+269 / –49	1427	+1913 / –343	13637
Sum÷n	1	--	--	–0.1186 $= m_1'$	3.4552 $= m_2'$	–0.8305 $= m_3'$	33.0194 $= m_4'$

Hence the moments about the mean and in ordinary units are obtained as below:

$m_1 = 0$

$m_2 = h^2 [m_2' - (m_1')^2]$

$= 4 [3.4552 - (-0.1186)^2] = 4(3.4411) = 13.76$

$m_3 = h^3 [m_3' - 3m_1'm_2' + 2(m_1')^3]$

$= 8 [-0.8305 - 3(-0.1186) (3.4552) + 2(-0.1186)^3]$

$= 8 (0.3956) = 3.16$

$m_4 = h^4 [m_4' - 4m_1'm_3' + 6(m_1')^2 m_2' - 3 (m_1')^4]$

$= 16 [33.0194 - 4 (-0.1186) (-0.8305) + 6(-0.1186)^2 \times (3.4552) - 3(-0.1186)^4]$

$= 528.06$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(3.16)^2}{(13.76)^3} = \frac{9.9856}{2605.2853} = 0.004, \text{ and}$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{528.06}{(13.76)^2} = \frac{528.06}{189.3376} = 2.79.$$

4.43. (b) Calculation of b_1 and b_2.

No. of heads (x)	f	D (x–4)	fD	fD^2	fD^3	fD^4
0	1	–4	–4	16	–64	256
1	7	–3	–21	63	–189	567
2	26	–2	–52	104	–208	416
3	54	–1	–54	54	–54	54
4	74	0	–131	0	–515	0
5	52	1	52	52	52	52
6	32	2	64	128	256	512
7	9	3	27	81	243	729
8	1	4	4	16	64	256
Total	256	--	147 / +16	514	+615 / +100	2842
Sum÷n	1	--	–0.0625 $= m_1'$	2.0078 $= m_2'$	0.3906 $= m_3'$	11.1016 $= m_4'$

The moments about the mean are obtained as below:

$m_1 = 0$

$$m_2 = m_2' - (m_1')^2 = 2.0078 - (0.0625)^2 = 2.0039$$

$$m_3 = m_3' - 3m_1'm_2' + 2\,(m_1)^3$$

$$= 0.3906 - 3(0.0625)\,(2.0078) + 2(0.0625)^3 = 0.0146$$

$$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2\,m_2' - 3(m_1')^4$$

$$= 11.1016 - 4(0.0625)\,(0.3906) + 6(0.0625)^2\,(2.0078) - 3(0.0625)^4 = 11.0510$$

Hence $b_1 = \dfrac{m_3^2}{m_2^3} = \dfrac{(0.0146)^2}{(2.0039)^3} = 0.0003$, and

$$b_2 = \frac{m_4}{m_2^2} = \frac{11.0510}{(2.0039)^2} = 2.75.$$

4.44. Calculation of b_1 and b_2.

Classes	x	f	$u(=\frac{x-19}{2})$	fu	fu^2	fu^3	fu^4
10–12	11	3	–4	–12	48	–192	768
12–14	13	30	–3	–90	270	–810	2430
14–16	15	110	–2	–220	440	–880	1760
16–18	17	218	–1	–218	218	–218	218
18–20	19	275	0	–540	0	–2100	0
20–22	21	222	1	222	222	222	222
22–24	23	108	2	216	432	864	1728
24–26	25	32	3	96	288	864	2592
26–28	27	2	4	8	32	128	512
Total	--	1000	--	+542 +2	1950	2078 –22	10230
Sum÷n	--	1	--	0.002 $= m_1'$	1.950 $= m_2'$	–0.022 $= m_3'$	10.234 $= m_4'$

Now the moments about the mean and in terms of the class interval units are obtained as below:

$$m_1 = 0$$

$$m_2 = m_2' - (m_1')^2 = 1.950 - (0.002)^2 = 1.95$$

$$m_3 = m_3' - 3m_1'm_2' + 2(m_1')^3$$

$$= -0.022 - 3(0.002)\,(1.950) + 2(0.002)^3 = -0.034$$

$$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2\,(m_2') - 3(m_1')^4$$

$$= 10.230 - 4(.002)\,(-0.022) + 6(.002)^2\,(1.95) - 3(.002)^4$$

$$= 10.230$$

Hence $b_1 = \dfrac{m_3^2}{m_2^3} = \dfrac{(-0.034)^2}{(1.95)^3} = 0.0002$, and

$$b_2 = \frac{m_4}{m_2^2} = \frac{10.230}{(1.95)^2} = 2.79.$$

Applying Sheppard's corrections, we have

$m_2(\text{corrected}) = m_2 - \dfrac{h^2}{12}$, where m_2 is in ordinary units.

$$= 2^2(1.95) - \frac{2^2}{12} = 7.47$$

$$m_3(\text{corrected}) = m_3$$

$$= (2)^3\,(-0.034) = -0.27, \text{ and}$$

$$m_4(\text{corrected}) = m_4 - \frac{h^2}{2}m_2 + \frac{7}{240}h^4$$

$$= 2^4\left(10.230 - \tfrac{1}{2} \times 1.95 + \frac{7}{240}\right)$$

$$= 16(10.230 - 0.975 + 0.021) = 148.216$$

Hence $b_1 = \frac{m_3^2}{m_2^3} = \frac{(-0.27)^2}{(7.47)^3} = 0.002$, and

$$b_2 = \frac{m_4}{m_2^2} = \frac{148.216}{(7.47)^2} = 2.66.$$

The following table is constructed to apply Charlier check.

u	f	$u+1$	$f(u+1)$	$f(u+1)^2$	$f(u+1)^3$	$f(u+1)^4$
−4	3	−3	−9	27	−81	243
−3	30	−2	−60	120	−240	480
−2	110	−1	−110	110	−110	110
−1	218	0	−179	0	−431	0
0	275	1	275	275	275	275
1	222	2	444	888	1776	3552
2	108	3	324	972	2916	8748
3	32	4	128	512	2048	8192
4	2	5	10	50	250	1250
Σ	1000	---	$\frac{+1181}{1002}$	2954	$\frac{+7265}{6834}$	22850

Now $\sum f(u+1) = \sum fu + n$

$= 2 + 1000 = 1002;$

$\sum f(u+1)^2 = \sum fu^2 + 2\sum fu + n$

$= 1950 + 2(2) + 1000 = 2954;$

$\sum f(u+1)^3 = \sum fu^3 + 3\sum fu^2 + 3\sum fu + n$

$= -22 + 3(1950) + 3(2) + 1000 = 6834;$ and

$\sum f(u+1)^4 = \sum fu^4 + 4\sum fu^3 + 6\sum fu^2 + 4\sum fu + n$

$= 10230 + 4(-22) + 6(1950) + 4(2) + 1000 = 22,850$

Hence the result.

4.46 (b) Calculation for skewness.

(i) $$Sk = \frac{Q_1 + Q_3 - 2\,\text{Median}}{Q_3 - Q_1} = \frac{37.15 + 61.27 - 2(49.21)}{61.27 - 37.15}$$

$$= \frac{98.42 - 98.42}{24.12} = 0$$

Thus the distribution is symmetrical.

(ii) Since mode ($\neq$1487) is greater than mean (1403), therefore the distribution is negatively skewed.

(iii) Given the first three moments about arbirary origin ($x = 16$) as $m_1' = -0.35$, $m_2' = 2.09$, $m_3' = -1.93$.

We find the third moment about mean to determine the presence of skewness.

$$\therefore \quad m_3 = m_3' - 3\,m_1' m_2' + 2(m_1')^3$$

$$= -1.93 - 3\,(-0.35)\,(2.09) + 2(-0.35)^3 = 0.18$$

As m_3 is not equal to zero and is positive, therefore the distribution is positively skewed.

4.47. Calculation of the coefficient of skewness.

Age (years)	No. of Men (f)	x	fx	fx^2	Class Boundries	F
15–19	29	17	493	8381	14.5–19.5	29
20–24	176	22	3872	85184	19.5–24.5	205
25–29	208	27	5616	151632	24.5–29.5	413
30–34	173	32	5536	177152	29.5–34.5	586
35–39	82	37	3034	112258	34.5–39.5	668
40–44	40	42	1680	70560	39.5–44.5	708
45–49	15	47	705	33135	44.5–49.5	723
50–54	3	52	156	8112	49.5–54.5	726
Total	726	--	21092	646414	--	--

(i) Now, $\bar{x} = \frac{\sum fx}{n} = \frac{21092}{726} = 29.05$ years;

$$\text{Mode} = l + \frac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 24.5 + \frac{(208 - 176)}{(208 - 176) + (208 - 173)} \times 5$$

$$= 24.5 + \frac{32}{67} \times 5 = 24.5 + 2.39 = 26.89 \text{ years; and}$$

$$s = \sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2} = \sqrt{\frac{646414}{726} - \left(\frac{21092}{726}\right)^2}$$

$$= \sqrt{890.3774 - 843.9025} = 6.82 \text{ years.}$$

Applying the Pearsonian measure of skewness, we find

$$Sk = \frac{\text{Mean} - \text{Mode}}{s} = \frac{29.05 - 26.89}{6.82} = 0.32$$

(ii) Median = Age of $\left(\frac{n}{2}\right)th$ person

= Age of $\left(\frac{726}{2}\right)th$, *i.e.* 363*rd* man, which lies in the age-group 24.5 – 29.5.

$\therefore$ $$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$= 24.5 + \frac{5}{208}(363 - 205) = 24.5 + 3.80$$

$$= 28.30 \text{ years}$$

Similarly, we find

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right)$$

$$= 19.5 + \frac{5}{176}(181.5 - 29) = 19.5 + 4.33$$

$$= 23.83 \text{ years, and}$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$= 29.5 + \frac{5}{173}(544.5 - 413) = 29.5 + 3.80 = 33.30 \text{ yrs}$$

Hence using the Bowley's co-efficient of skewness, we get

$$Sk = \frac{Q_1 + Q_3 - 2\,\text{Median}}{Q_3 - Q_1} = \frac{33.30 + 23.83 - 2(28.30)}{33.30 - 23.83}$$

$$= \frac{0.53}{9.47} = 0.06.$$

4.48. Calculation of the first four moments about the mean.

Age (x)	f	$u\left(=\frac{x-40}{5}\right)$	fu	fu^2	fu^3	fu^4
25	2	–3	–6	18	–54	162
30	8	–2	–16	32	–64	128
35	18	–1	–18	18	–18	18
40	27	0	–40	0	–136	0
45	25	1	25	25	25	25
50	16	2	32	64	128	256
55	7	3	21	63	189	567
60	2	4	8	32	128	512
Total	105	--	+86 / +46	252	+470 / 334	1668
Sum÷n	1	--	0.4381 $= m_1'$	2.4000 $= m_2'$	3.1810 $= m_3'$	15.8857 $= m_4'$

Hence the moments about the mean in terms of class-interval units are obtained below:

$$m_1 = 0;$$

$$m_2 = m_2' - (m_1')^2 = 2.4000 - (0.4381)^2 = 2.2081$$

$$m_3 = m_3' - 3m_1'm_2' + 2(m_1')^3$$

$$= 3.1810 - 3(0.4381)(2.4000) + 2(0.4381)^3 = 0.1949;$$

$$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2(m_2') - 3(m_1')^4$$

$$= 15.8857 - 4(0.4331)(3.1810) + 6(0.4381)^2(2.4000) - 3(0.4381)^4 = 12.9646;$$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(0.1949)^2}{(2.2081)^3} = 0.0035, \text{ and}$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{12.9646}{(2.2081)^2} = 2.66.$$

Hence the distribution is slightly positively skewed and platykurtic.

4.49. The necessary computations are given below:

(a) Mean, $\bar{X} = 10 + 5\bar{u} = 10 + 5\frac{(-46)}{125} = 8.16$

$$m_1' = \frac{\sum fu}{n} = \frac{-46}{125} = -0.368;$$

$$m_2' = \frac{\sum fu^2}{n} = \frac{306}{125} = 2.448;$$

$$m_3' = \frac{\sum fu^3}{n} = \frac{-242}{125} = -1.936;$$

$$m_4' = \frac{\sum fu^4}{n} = \frac{1962}{125} = 15.696.$$

$$m_2(\text{variance, } s^2) = h^2[m_2' - (m_1')^2] = 25[2.448 - (-0.368)^2]$$

$$= 25(2.313) = 57.825, \text{ so that } s = 7.60$$

$$m_3 = 125[-1.936 - 3(-0.368)(2.448) + 2(-0.368)^3]$$

$$= 125[-1.936 + 2.703 - 0.100] = 125(0.667) = 83.375$$

$$m_4 = (5)^4 \cdot [15.696 - 4(-0.368)(-1.936) + 6(-0.368)^2(2.448) - 3(-0.368)^4]$$

$$= 625[15.696 - 2.850 + 1.989 - 0.055]$$

$$= 625[14.78] = 9237.5$$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(83.375)^2}{(57.825)^3} = 0.04, \text{ and}$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{9237.5}{(57.825)^2} = 2.76.$$

(b) Mean, $\bar{X} = 20 + \bar{v} = 20 + \frac{(21)}{200} = 20.105;$

$$m_1' = \frac{21}{200} = 0.105; \quad m_2' = \frac{1265}{200} = 6.325;$$

$$m_3' = \frac{-627}{200} = -3.135; \quad m_4' = \frac{14169}{200} = 70.845$$

$$m_2(\text{variance, } s^2) = m_2' - (m_1')^2 = 6.325 - (0.105)^2$$

$$= 6.314, \text{ so that } s = \sqrt{6.314} = 2.51.$$

$$m_3 = m_3' - 3m_1'm_2' + 2(m_1')^3$$

$$= -3.135 - 3(0.105)(6.325) + 2(0.105)^3 = -5.125;$$

Similarly, $m_4 = 82.58;$

$$b_1 = \frac{(-5.125)^2}{(6.314)^3} = 0.104 \text{ and } b_2 = \frac{82.58}{(6.314)^2} = 2.07$$

(i) A distribution with smaller C.V. will be more consistent.

$$\text{C.V. for (a)} = \frac{s}{\bar{x}} \times 100 = \frac{7.60}{8.16} \times 100 = 93.14\%$$

$$\text{C.V. for (b)} = \frac{2.51}{20.105} \times 100 = 12.48\%$$

Hence distribution (b) is more eonsistent.

(ii) A distribution having m_3 negative, will be negatively skewed. The distribution (b) has $m_3 = -5.125$, so it is negatively skewed.

(iii) A distribution with $b_1 = 0$ and $b_2 = 3$ will be mesokurtic. None of the distributions is mesokurtic.

4.50. (a) For a distribution to be mesokurtic, b_2 equals 3.

We are given that $m_4 = 24.3$. Therefore

$$\frac{m_4}{m_2^2} = b_3 = 3 \quad i.e. \quad \frac{243}{m_2^2} = 3$$

or $$m_2^2 = \frac{243}{3} = 81 \quad \text{or} \quad m_2 = 9$$

Hence the desired value of the standard deviation $= \sqrt{9} = 3$.

(b) First we calculate the moments about the mean. Thus

$$m_1 = 0;$$

$$m_2 = m_2' - (m_1')^2 = 17 - (-1.5)^2 = 14.75;$$

$$m_3 = m_3' - 3m_1'm_2' + 2(m_1')^3$$

$$= -30 - 3(-1.5)\,(17) + 2\,(-1.5)^3$$

$$= 76.5 - 36.75 = 39.75; \text{ and}$$

$$m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2\,(m_2') - 3(m_1')^4$$

$$= 108 - 4(-1.5)\,(-30) + 6(-1.5)^2\,(17) - 3(-1.5)^4$$

$$= 337.50 - 195.19 = 142.31$$

Hence $$b_1 = \frac{m_3^2}{m_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.49, \text{ and}$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{142.31}{(14.75)^2} = 0.65.$$

The distribution is platy-kurtic as $b_2 < 3$.

4.51. (a) b_2 in the first case $= \dfrac{m_4}{m_2^2} = \dfrac{230}{(9)^2} = 2.84 < 3$

b_2 in the seconc case $= \dfrac{m_4}{m_2^2} = \dfrac{780}{(16)^2} = 3.05 > 3$

Hence we may conclude that

(i) the second distribution is lepto-kurtic,

(ii) neither distribution is mesokurtic, and

(iii) the first distribution is platy-kurtic.

(b) We have $b_2 = \dfrac{m_4}{m_2^2} = \dfrac{m_4}{(25)^2}$

(i) For leptokurtic, b_2 must be greater than 3,

$i.e.$ $\dfrac{m_4}{625} > 3$ or $m_4 > 1875$.

(ii) For meso-kurtic, b_2 is equal to 3,

$i.e.$ $\dfrac{m_4}{625} = 3$ or $m_4 = 1875$

(iii) For platy-kurtic, b_2 must be less than 3,

$i.e.$ $\dfrac{m_4}{625} < 3$ or $m_4 < 1875$

4.54 Here, Mean (X) $= \bar{x} = 60$
S.D (X) = 8.944

i) Let Y = values of X increased by 20 points = X + 20
Then, Mean (Y) $= \bar{Y} = \bar{X} + 20 = 60 + 20 = 80$
and S.D (Y) = S.D (X+20) = S.D(X) = 8.944

ii) Let Z = Values of X increased by 25%
= X + 0.25 X = 1.25 X
Thus, Mean (Z) $= \bar{Z} = 1.25\,\bar{X} = 1.25(60) = 75$
and S.D (Z) = S.D (1.25X) = 1.25 S.D (X)
= 1.25 (8.944) = 11.1800

4.55 Given is,

$Mean = \overline{X} = 3133.33,\ Mode = 2804.35$

$S.D(X) = 796.70$

i) $Skewness = \dfrac{Mean - Mode}{S.D} = \dfrac{3133.33 - 2804.35}{796.70} = 0.413$

Distribution is positively skewed.

$\text{Coefficient of variation} = \dfrac{S.D}{Mean} \times 100$

$= \dfrac{796.70}{3133.33} \times 100 = 25.43\%$

ii) $Let\ Y = X + 500$

$Then\ Mean(Y) = \overline{Y} = \overline{X} + 500 = 3133.33 + 500 = 3633.33$

$and,\ S.D(Y) = S.D(X + 500) = S.D(X) = 796.70$

iii) $Let\ Z = X + 0.10X = 1.10X$

$Mean(Z) = \overline{Z} = 1.10\overline{X} = 1.10(3133.33) = 3446.6630$

$and,\ S.D(Z) = S.D(1.10X) = 1.10\ S.D(X)$

$= 1.10(796.70) = 876.3700$

●●●●●●●●●●

Chapter 5

INDEX NUMBERS

5.20. Calculation of the price relatives

Year	Retail Price (Rs)	Price Relatives (i) 1995 as base	Price Relatives (ii) 1998 as base
1995	14.95	100	95.5
1996	14.94	99.9	95.5
1997	15.10	101.0	96.5
1998	15.65	104.7	100
1999	16.28	108.9	104.0
2000	16.53	110.6	105.6

5.21. Calculation of Index Numbers.

Year	Production	Price relatives (i) 1954 as base	Price relatives (ii) Average of 1958, 59 and 60 as base
1954	282	100	51.90
1955	389	137.94	71.60
1956	438	155.32	80.61
1957	470	166.67	86.50
1958	511	181.21	94.05
1959	555	196.81	102.15
1960	564	200.00	103.80
1961	630	223.40	115.95
1962	662	234.75	121.84
1963	681	241.49	125.34

5.22. Computation of the simple aggregative Index Numbers.

Commodity	Average prices in Rs. Per unit					
	2001	2002	2003	2004	2005	2006
Gold	25.3	30.8	33.4	35.5	35.3	36.0
Wheat	17.3	14.5	4.9	5.7	17.1	11.6
Cotton	7.8	5.4	6.7	5.6	7.2	10.2
Total	50.4	50.7	45.0	46.8	59.6	57.8

The simple aggregative indices are calculated by the formula

$$P_{0n} = \frac{\Sigma p_n}{\Sigma p_0} \times 100,$$

where the symbols have their usual meaning. Thus the simple aggregative price index for

$$2002 = \frac{50.7}{50.4} \times 100 = 100.6,$$

$$2003 = \frac{45.0}{50.4} \times 100 = 89.3,$$

$$2004 = \frac{46.8}{50.4} \times 100 = 92.9,$$

$$2005 = \frac{59.6}{50.4} \times 100 = 118.3,$$

$$2006 = \frac{57.8}{50.4} \times 100 = 114.7,$$

Calculation of the simple Average of Relatives price indices for years 2002 to 2006:

Years	Price relatives			Sum of price relatives	Index No. as average of relatives
	Gold	Wheat	Cotton		
2001	100	100	100	300	100
2002	122	84	69	275	91.6
2003	132	28	86	246	82.1
2004	140	33	72	245	81.7
2005	140	100	92	332	110.7
2006	142	67	131	340	113.4

5.23. Calculation of the price Index Numbers for April 2002 and May 2002 with May 2001 as base, using

(i) Simple aggregative method

$$P_{0n} = \frac{\Sigma p_n}{\Sigma p_0} \times 100,$$

The simple aggregative price index for April 2002 is

$$P_{0n} = \frac{18.37 + 14.58 + 13.94 + 13.75}{17.50 + 14.58 + 14.67 + 17.50} \times 100 = 94.38,$$

and the simple aggregative price index for May 2002 is

$$P'_{02} = \frac{17.58 + 16.50 + 15.25 + 13.42}{17.50 + 14.58 + 14.67 + 17.50} \times 100 = 97.67$$

(ii) Simple average (mean) of price relatives:

Years	Price relatives				Sum of price relatives	Index No. as average of relatives
	Wheat	Barley	Jawar	Bajra		
May 2001	100	100	100	100	400	100
April 2002	105.0	100	95.0	78.6	378.6	94.65
May 2002	100.5	113.2	104.0	76.7	394.4	98.60

(iii) Geometric mean of price relatives:

Years	Price relatives				$\Sigma \log x$	$1/n\ \Sigma \log x$	Indices (G.M.)
	Wheat	Barley	Jawar	Bajra			
May 2001	100	100	100	100			
log	2.0000	2.0000	2.0000	2.0000	8.0000	2.0000	100
April 2002	105.0	100	95.0	78.6			
log	2.0212	2.000	1.9777	1.8954	7.8943	1.9738	94.14
May 2002	100.5	113.2	104.0	76.7			
log	2.0021	2.0539	2.0170	1.8848	7.9578	1.9894	97.59

5.24. Construction of chain index numbers for average prices.

Year	Link Relatives						Total	Average	Chain Indices
	Rice	Wheat	Linseed	Gur	Cotton	Tobacco			
1928	100	100	100	100	100	100	600	100	100
1929	105.5	73.3	114.3	115.9	87.4	98.8	595.2	99.2	99.2
1930	75.3	65.5	81.2	84.9	58.1	84.8	449.8	75.0	74.4
1931	70.7	75.0	64.6	67.6	76.7	80.0	434.7	72.4	53.9

Hence the chain indices are 100, 99.2, **74.4 and** 53.9.

5.25. Construction of chain indices for prices relatives.

Year	Link Relatives				Total of L.R.	Average	Chain Indices
	Sugar	Gur	Tea	Coffee			
1941	98	75	82	99	354	88.5	88.5
1942	102	109	90	101	402	100.5	88.9
1943	114	101	105	104	424	106.0	94.2
1944	96	101	114	91	402	100.5	94.7

Hence the chain indices are 88.5, 88.9, 94.2 and 94.7.

5.26. Construction of chain indices from price relatives.

Years	Price Relatives of Commodities			Total of relatives	(a) Average	(b) Chain Indices
	A	B	C			
2001	100	100	100	300	100	100
2002	105.0	97.0	121.0	323.0	107.7	107.7
2003	104.8	96.9	103.3	305.0	101.7	109.5
2004	104.5	106.4	104.0	314.9	105.0	115.0
2005	100.9	99.0	98.5	298.4	99.5	114.4
2006	103.5	106.1	101.6	311.2	103.7	118.6

5.27. (i) Calculation of Index Numbers, using the simple mean of price relatives.

Years	Price relatives of commodities				Total of relatives	Index no. as mean
	A	B	C	D		
2001	100	100	100	100	400	100
2002	125	120	87	75	407	101.8
2003	112	110	92	125	439	109.8
2004	125	120	108	150	503	125.8
2005	131	127	122	140	520	130.0

(ii) Calculation of Index Numbers, using the chain base method.

Year	Link relatives of commodities				Total of Relatives	Average	Chain Indices
	A	B	C	D			
2001	100	100	100	100	400	100	100
2002	125	120	87	75	407	101.8	101.8
2003	90	92	106	167	455	113.8	$\frac{101.8\times113.8}{100}=115.8$
2004	112	109	117	120	458	114.5	$\frac{115.8\times114.5}{100}=132.6$
2005	105	106	113	93	417	104.2	$\frac{132.6\times104.2}{100}=138.2$

(iii) The fixed base and the chain indices are generally not in close agreement. The reason being that the farther it is from the base year, the less reliable is the chain index likely to be.

5.28. Computation of weighted aggregative price index for 2007.

Groups	Base year		Current year		p_0q_0	p_1q_0
	p_0	q_0	p_1	q_1		
A	85	50	116	45	4250	5800
B	34	120	42	185	4080	5040
C	10	35	15	68	350	525
D	48	210	50	250	10080	10500
Total	---	---	--	--	18760	21865

Hence $p_{01}=\frac{\Sigma p_1q_0}{\Sigma p_0q_0}\times100=\frac{21865}{18760}\times100=116.6$

The weighted aggregative price index for 2007 on the basis of 2003 is 116.6. This means that the prices have increased by16.6%.

5.29. (a) Computation of the weighted price indices with base period quantities as weights.

Commodity	Base year		Current year				Price relatives	$p_0q_0 \times$
	p_0	q_0	p_1	q_1	p_0q_0	p_1q_0	p_1/p_0	(p_1/p_0)
Rice	3.50	71	3.15	80	248.50	223.65	0.900	223.65
Barley	2.00	107	1.80	138	214.00	192.60	0.900	192.60
Maize	2.60	62	1.75	57	161.20	108.50	0.673	108.49
Total	---	---	--	--	623.70	524.75	--	524.74

Weighted-aggregative price index, $P_{01} = \frac{\sum p_1 p_0}{\sum p_0 q_0} \times 100$

$$= \frac{524.75}{623.70} \times 100 = 84.14$$

Weighted Average of Relatives Price Index is

$$P_{01} = \frac{\sum\left(\frac{p_1}{p_0}\right) p_0 q_0}{\sum p_0 q_0} \times 100 = \frac{524.74}{623.70} \times 100 = 84.13$$

(b) Computation of weighted price indices with 2005 quantities as weights.

Commodity	Base year		Current year				Price relatives	$p_0q_1 \times$
	p_0	q_0	p_1	q_1	p_0q_1	p_1q_1	p_1/p_0	(p_1/p_0)
Rice	3.50	71	3.15	80	280.0	252.00	0.900	252.00
Barley	2.00	107	1.80	138	276.0	248.00	0.900	248.40
Maize	2.60	62	1.75	57	148.2	99.75	0.673	99.74
Total	---	---	--	--	704.2	600.15	--	600.14

∴ Weighted aggregative price Index is

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100 = \frac{600.15}{704.20} \times 100 = 85.22; \text{ and}$$

Weighted Average of Relatives Price Index is

5.30 Computation of weighted aggregative price index for 2007

Commodities	1997 Base year		2007 Current year					
	p_0	q_0	p_1	q_1	p_0q_0	p_1q_0	p_1q_1	p_0q_1
A	12	501	10	600	6012	5010	6000	7200
B	38	100	50	194	3800	5000	9700	7372
C	40	56	40	76	2240	2240	3040	3040
Total	--	--	--	--	12052	12250	18740	17612

(i) By Laspeyre's method:

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{12250}{12052} \times 100 = 101.6$$

(ii) By Paasche's method:

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100 = \frac{18740}{17612} \times 100 = 106.4$$

5.31. Computation of weighted price index for 2007

Group	Base year		Current year				Price Relatives	Weighted price relative
	p_0	q_0	p_1	q_1	p_1q_1	p_0q_1	p_1/p_0	p_0q_1 (p_1/p_0)
A	85	50	116	45	5220	3825	1.3647	5219.98
B	34	120	42	185	7770	6290	1.2353	7770.04
C	10	35	15	68	1020	680	1.5000	1020.00
D	48	210	50	250	12500	12000	1.0417	12500.40
Total	---	---	--	--	26510	22795	--	26510.42

(i) Weighted Aggregative Price Index by Paasche's method:

$$P_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100 = \frac{26510}{22795} \times 100 = 116.30$$

(ii) Weighted Average of Relatives Price Index by Paasche's method:

$$P_{01} = \frac{\sum\left(\frac{p_1}{p_0}\right) p_0 q_1}{\sum p_0 q_1} \times 100 = \frac{26510.42}{22795} \times 100 = 116.30$$

5.32. Construction of weighted Index numbers of prices for 2004 and 2005.

Commodity	2000 (base year)		2004		2005		Products		
	p_0	q_0	p_1	q_1	p_2	q_2	p_0q_0	p_1q_0	p_2q_0
A	70.50	270	80.65	276	85.00	290	19035.00	21775.50	2295.00
B	146.95	24	155.00	18	154.75	44	3526.80	3720.00	3714.00
C	25.50	130	32.50	121	30.50	137	3315.00	4225.00	3965.00
D	64.75	185	75.00	257	60.95	355	11978.75	13875.00	11275.75
Total	–	–	–	–	–	–	37855.55	43595.50	41904.75

Products Continued

	p_0q_1	p_1q_1	p_0q_2	p_2q_2
	19458.00	22259.4	20445.00	24650.00
	2645.10	2790.0	6465.80	6809.00
	3085.50	3932.5	34931.50	4178.50
	17288.25	20025.0	22986.25	21637.25
Σ	42476.65	49006.9	53390.55	57274.75

(i) Base year weighted indices:

Index for 2004

$$(p_{2004}) = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{43595.50}{37855.55} \times 100 = 115.16$$

Index for 2005

$$(p_{2005}) = \frac{\Sigma p_2 q_0}{\Sigma p_0 q_0} \times 100 = \frac{41904.75}{37855.55} \times 100 = 110.70$$

(ii) Current year weighted indices:

Index for 2004

$$(p_{2004}) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100 = \frac{49006.90}{42476.85} \times 100 = 115.37$$

Index for 2005

$$(p_{2005}) = \frac{\Sigma p_2 q_2}{\Sigma p_0 q_2} \times 100 = \frac{57274.75}{53390.55} \times 100 = 107.28$$

5.33. Computation of weighted index numbers.

Commodity	Base year		Current year					
	p_0	q_0	p_1	q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
A	7	70	5	49	490	350	343	245
B	5	27	7	28	135	189	140	196
C	10	35	9	29	350	315	290	261
D	9	50	4	42	450	200	378	168
E	3	16	10	25	48	160	75	250
Total	---	---	--	--	1473	1214	1226	1120

Fisher's index $= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$

$= \sqrt{\frac{1214}{1473} \times \frac{1120}{1226}} \times 100 = 86.8$; and

Marshall-Edgeworth index $= \frac{\sum p_1 (q_0 + q_1)}{\sum p_0 (q_0 + q_1)} \times 100$

$= \frac{\sum p_1 q_0 + \sum p_1 q_1}{\sum p_0 q_0 + \sum q_0 q_1} \times 100$

$= \frac{2334}{2699} \times 100 = 86.5.$

5.34. Construction of Fisher's Ideal Index Number.

Commodity	Base year		Current year					
	p_0	q_0	p_1	q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
Rice	9.3	100	4.5	90	930.0	450.0	837.0	405.0
Wheat	6.4	11	3.7	19	70.4	40.7	121.6	70.3
Jawar	5.1	5	2.7	3	25.5	13.5	15.3	8.1
Total	(ii) Taking 1964 as base:				1025.9	504.2	973.9	483.4
	p_1	q_1	p_0	q_0	p_1q_1	p_0q_1	p_1q_0	p_0q_0

(i) Fisher's index for 1964 $= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$

$= \sqrt{\frac{504.2}{1025.5} \times \frac{483.5}{973.9}} \times 100 = 49.4$

(ii) Fisher's index for 1957 $= \sqrt{\dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$= \sqrt{\dfrac{973.9}{483.4} \times \dfrac{1025.9}{504.2}} \times 100 = 202.5$

5.35. Construction of weighted index numbers.

Commodity	Base Year 2003		Current Years 2004		2005		p_0q_0	p_1q_0	p_2q_0
	p_0	q_0	p_1	q_1	p_2	q_2			
Wheat	9 35	3,974	8 12	3,862	8.78	3,930	37,156.90	32,268.88	34,891.72
Rice	11 25	973	11.7[illegible]	852	12 08	722	10,946.25	11,413.29	11,753.84
Gram	7.00	755	7.63	601	8.23	744	5,285.00	5,798.40	6,213.65
Total	–	–	–	–	–	–	53,388.15	49,480.57	52,859.21

Using Laspeyres' formula, we get

Index for 2004 $= \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \dfrac{49{,}480.57}{53{,}388.15} \times 100 = 92.68$

Index for 2005 $= \dfrac{\Sigma p_2 q_0}{\Sigma p_0 q_0} \times 100 = \dfrac{52{,}859.21}{53{,}388.15} \times 100 = 99.00$

5.36. Calculation of the weighted index numbers.

Diary Products	Base Year 1998		Current years 1999		2007	
	p_0	q_0	p_1	q_1	p_2	q_2
Milk	3.95	9,675	3.89	9,717	4.13	10,436
Butter	61.50	118	62.20	116	59.70	116
Cheese	34.80	78	35.40	78	38.90	83

The necessary prodeucts for the calculation of Laspeyre's, Paasche's, Fisher's and Marshall-Edgeworth's indices are given below:-

	p_0q_0	p_1q_0	p_1q_1	p_0q_1	p_2q_0	p_0q_2	p_2q_2
M	38,216.25	37,635 75	37,799.13	38,382.15	39,957.75	41,222.20	43,100.68
B	7,257 00	7,339.60	7,215.20	7,134.00	7,044.60	7,134.00	6,925.20
C	2,714 40	2,761 20	2,761.20	2,714.40	3,034 20	2,888.40	3,228.70
Σ	48,187 65	47,736.55	47,775.53	48,230 55	50,036.55	51,244.60	53,254.58

(i) Laspeyres'index for 1999 $= \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$= \dfrac{47{,}736.55}{48{,}187.65} \times 100 = 99.06$; and

Index for 2007 $= \dfrac{\Sigma p_2 q_0}{\Sigma p_0 q_0} \times 100$

$= \dfrac{50{,}036.55}{48{,}187.65} \times 100 = 103.84$

(ii) Paasche's index for 1999 $= \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$= \dfrac{47{,}775.53}{48{,}230.55} \times 100 = 99.06$; and

Index for 2007 $= \dfrac{\Sigma p_2 q_2}{\Sigma p_0 q_2} \times 100$

$= \dfrac{53{,}254.58}{51{,}244.60} \times 100 = 103.92$

(iii) Fisher's Ideal index for 1999 $= \sqrt{\dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$= \sqrt{\dfrac{47{,}736.65}{48{,}187.65} \times \dfrac{47{,}775.53}{48{,}230.55}} \times 100$

$= \sqrt{0.9906 \times .9906} \times 100 = 99.06;$

Index for 2007 $= \sqrt{\dfrac{\Sigma p_2 q_0}{\Sigma p_0 q_0} \times \dfrac{\Sigma p_2 q_2}{\Sigma p_0 q_2}} \times 100$

$= \sqrt{\dfrac{50{,}036.55}{48{,}187.65} \times \dfrac{53{,}254.58}{51{,}244.60}} \times 100$

$= \sqrt{1.0384 \times 1.0392} \times 100 = 103.88;$

(iv) Marshall-Edgeworth index for 1999 $= \dfrac{\Sigma p_1(q_0 + q_1)}{\Sigma p_0(q_0 + q_1)} \times 100$

$$= \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

$$= \frac{95,512.08}{96,418.20} \times 100 = 99.06; \text{ and}$$

Index for 2007 $$= \frac{\Sigma p_2 (q_0 + q_2)}{\Sigma p_0 (q_0 + q_2)} \times 100$$

$$= \frac{\Sigma p_2 q_0 + \Sigma p_2 q_2}{\Sigma p_0 q_0 + \Sigma p_0 q_2} \times 100$$

$$= \frac{103,291.13}{49,432.25} \times 100 = 103.88$$

(v) The necessary products for the calculation of Walsh indices are obtained as below:-

Diary Products	$\sqrt{q_0 q_1}$	$\sqrt{q_0 q_2}$	$p_0\sqrt{q_0 q_1}$	$p_1\sqrt{q_0 q_1}$	$p_0\sqrt{q_0 q_2}$	$p_2\sqrt{q_0 q_2}$
Milk	9.696	10.048	38,299 20	37,717 44	39,689 60	41,498.24
Butter	117	117	7,195 50	7,277 40	7,195 50	6,984 90
Cheese	78	80	2,714 40	2,761 20	2,784.00	3,112.00
Total	--	--	48,209 10	47,756 04	49,669 10	51,595 14

Thus Walsh index for 1999 $$= \frac{\Sigma p_1 \sqrt{q_0 q_1}}{\Sigma p_0 \sqrt{q_0 q_1}} \times 100$$

$$= \frac{47,756.04}{48,209.10} \times 100 = 99.06 \text{ ; and}$$

Index for 2007 $$= \frac{\Sigma p_1 \sqrt{q_0 q_2}}{\Sigma p_0 \sqrt{q_0 q_2}} \times 100$$

$$= \frac{51,595.14}{49,669.10} \times 100 = 103.88$$

(vi) The necessary calculations for Palgrave indices are given below:-

Diary Products	$\frac{p_1}{p_0}$	$\frac{p_2}{p_0}$	$p_1 q_1 \left(\frac{p_1}{p_0}\right)$	$p_2 q_2 \left(\frac{p_2}{p_0}\right)$
Milk	0.9848	1.0456	37224.58	45,066.07
Butter	1.0114	0.9707	7,292.45	6,722.29
Cheese	1.0172	1.1178	2,808.69	3,609.04
Total	--	--	47,325.72	55,397.40

Thus Palgrave index for 1999 $$= \frac{\Sigma p_1 q_1 \left(\frac{p_1}{p_0}\right)}{\Sigma p_1 q_1} \times 100$$

$$= \frac{47,325.72}{47,775.53} \times 100 = 99.06 \text{ ; and}$$

Index for 2007 $$= \frac{\Sigma p_2 q_2 \left(\frac{p_2}{p_0}\right)}{\Sigma p_2 q_2} \times 100$$

$$= \frac{55,397.40}{53,254.58} \times 100 = 104.02$$

5.37. (a) Rewriting the values in columns 2 and 3 in symbols, we get

Year (1)	Declared values (2)	Value on the basis of 1938 values (3)
1938	$\sum p_0 q_0$	$\sum p_0 q_0$
1939	$\sum p_1 q_1$	$\sum p_0 q_1$
1940	$\sum p_2 q_2$	$\sum p_0 q_2$
1941	$\sum p_3 q_3$	$\sum p_0 q_3$

(i) Index numbers of volumes are thus obtained by dividing the values in column 3 by $\sum p_0 q_0$, *i.e.* by Laspeyres's method.

Thus the volume index for 1938 $$= \frac{\sum p_0 q_0}{\sum p_0 q_0} \times 100$$

$$= \frac{858}{858} \times 100 = 100;$$

Volume index for 1939 $$= \frac{\sum p_0 q_1}{\sum p_0 q_0} \times 100$$

$$= \frac{832}{858} \times 100 = 97;$$

$$\text{Volume index for 1940} = \frac{\sum p_0q_2}{\sum p_0q_0} \times 100$$

$$= \frac{807}{858} \times 100 = 94; \text{ and}$$

$$\text{Volume index for 1941} = \frac{\sum p_0q_3}{\sum p_0q_0} \times 100$$

$$= \frac{704}{858} \times 100 = 82$$

(ii) Index numbers of average values (prices) are obtained by dividing the entry in column 2 for any year by corresponding entry in column 3., *i.e.* by Paasche's formula. Thus

$$\text{the average value index for 1938} = \frac{\sum p_0q_0}{\sum p_0q_0} \times 100$$

$$= \frac{858}{858} \times 100 = 100;$$

$$\text{the average value index for 1939} = \frac{\sum p_1q_1}{\sum p_0q_1} \times 100$$

$$= \frac{840}{832} \times 100 = 101;$$

$$\text{the average value index for 1940} = \frac{\sum p_2q_2}{\sum p_0q_2} \times 100$$

$$= \frac{1126}{807} \times 100 = 140; \text{ and}$$

$$\text{the average value index for 1941} = \frac{\sum p_3q_3}{\sum p_0q_3} \times 100$$

$$= \frac{1132}{704} \times 100 = 161.$$

(b) Computation of weighted price indices:

Cereals	1960	1961	1962	Weight	Price relatives			
	p_0	p_1	p_2	w	(p_1/p_0)	(p_2/p_0)	$(p_1/p_0)w$	$(p_2/p_0)w$
Wheat	110.70	122.85	129.00	12	1.1098	1.1653	31.3176	13.9836
Barley	99.50	124.55	120.60	8	1.2518	1.2111	10.0144	9.6888
Maize	60.25	74.90	76.25	5	1.2432	1.2656	6.2160	6.3280
Total	---	---	---	25	---	--	29.5480	30.0001

$$\text{Price index for 1961} = \frac{\sum \left(\frac{p_1}{p_0}\right) w}{\sum w} \times 100 = \frac{29.5480}{25} \times 100 = 118.19; \text{ and}$$

$$\text{Price index for 1962} = \frac{\sum \left(\frac{p_2}{p_0}\right) w}{\sum w} \times 100 = \frac{30.0004}{25} \times 100 = 120.00.$$

5.38. Computation of Fisher's price index number for 1976:

Commodity	Quantity (units)		Value (Rs.)	
	1961 (q_0)	1976 (q_1)	1961 (p_0q_0)	1976 (p_1q_1)
A	100	150	600	1200
B	80	100	400	700
C	60	72	180	432
D	30	33	450	363
Total	--	--	1630	2695
			$\sum p_0q_0$	$\sum p_1q_1$

To find other necessary products, we first compute the unit prices in both 1961 and 1976 by dividing the value figures by the corresponding quantity figures in both the years. Thus

Unit prices			
p_0	p_1	p_1p_0	p_0q_1
6	8	800	900
5	7	560	500
3	6	360	216
15	11	330	495
---	---	2050	2111
		$\sum p_1q_0$	$\sum p_0q_1$

Hence

$$\text{Fisher's price index for 1976} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{2050}{1630} \times \frac{2695}{2111}} \times 100$$

$$= \sqrt{1.2577 \times 1.2767} \times 100 = 126.72$$

5.39. Computation of quantity and price index numbers.

Commodity	Quantity (units)		Value (£)	
	1997 (q_0)	2007 (q_1)	1997	2007
1	100	150	500	900
2	80	100	320	500
3	60	72	150	360
4	30	33	360	297
Total	---	---	$1330 = \Sigma p_0 q_0$	$2057 = \Sigma p_1 q_1$

To find other necessary products, we first calculate the average or unit prices in both 1997 and 2007 by dividing the value figures by corresponding quantity figure in both the years.

Thus

Commodity	Unit price		$p_0 q_1$	$p_1 q_0$
	1997(p_0)	2007(p_1)		
1	5	6	750	600
2	4	5	400	400
3	2.5	5	180	300
4	12	9	396	270
Total	---	---	$1726 = \Sigma p_0 q_1$	$1570 = \Sigma p_1 q_0$

From these products, the following quantity and price indices (i.e. Laspeyres' type) can be obtained as follows:

The quantity index for 2007 with 1997 as base is

$$Q_{01} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \times 100 = \frac{1726}{1330} \times 100 = 129.8$$

The quantity index for 1997 with 2007 as base is

$$Q_{10} = \frac{\Sigma p_1 q_0}{\Sigma p_1 q_1} \times 100 = \frac{1570}{2075} \times 100 = 76.3$$

The price index for 2007 with 1997 as base is

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{1570}{1330} \times 100 = 118.0 \text{ and}$$

the price index for 1997 with 2007 as base is

$$P_{10} = \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times 100 = \frac{1726}{2057} \times 100 = 83.9$$

5.40. The necessary products are given below:

Com.	1955		1965		$p_1 q_0$	$p_0 q_0$	$\frac{p_1}{p_0}$	$\left(\frac{p_1}{p_0}\right) p_0 q_0$
	p_0	q_0	p_1	q_1				
A	10	501	12	600	6012	5010	1.20	6012
B	40	100	38	194	3800	4000	0.95	3800
C	50	76	40	56	3040	3800	0.80	3040
Σ	--	--	--	--	12852	12810	--	12852

Now the Weighted Aggregate price index for 1965, using base period quantities as weight, *i.e.* by Laspyres' formula is

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{12852}{12810} \times 100 = 100.33;$$

and the Weighted Average of relatives price index for 1965 by using Laspyres' method is

$$P_{01} = \frac{\Sigma (p_1/q_0)\, p_0 q_0}{\Sigma p_0 q_0} \times 100 = \frac{12852}{12810} \times 100 = 100.33.$$

5.41. Calculation of quantity index numbers.

Articles		Quantity 1938 (q_0)	Quantity 1943 (q_1)	Values (in 000£) 1938	Values (in 000£) 1943	(i) $\frac{q_1}{q_0} \times W$	(ii) $\frac{q_1}{q_0} \times W$
Cotton yarn (million 1b)	A	10.0	5.4	397	817	214.38	1512.96
	B	2.9	2.8	278	315	268.41	326.25
Cotton manufactures (thousand of yds)	A	35.3	69.3	841	2,854	1651.03	1453.77
	B	21.6	83.2	776	3,319	2989.04	861.66
	C	81.9	102.9	1,452	5,805	1821.31	4620.31
	D	68.8	90.2	1,028	6,381	1347.46	4867.01
Total		---	---	4,772	19,491	8291.63	13642.05

(i) Index for 1943 on 1938 as base $= \dfrac{\Sigma\left(\frac{q_1}{q_0}\right)W}{\Sigma W} \times 100$

$$= \frac{8291.63}{4772} \times 100 = 173.8$$

(ii) Index for 1938 on 1943 as base $= \dfrac{13642.05}{19491} \times 100 = 70.0$

5.42. Calculation of the cost of living index number.

Expenses on	Weight (W)	Price relative for 1928	I = price relative for 1929	$I \times W$
Food	35	100	97	3395
Rent	15	100	100	1500
Clothing	20	100	87	1740
Fuel	10	100	92	920
Misc.	20	100	113	2260
Total	100	--	--	9815

Hence cost of living index number for 1929 $= \dfrac{\Sigma I \times W}{\Sigma W}$

$$= \frac{9815}{100} = 98.15$$

As the cost of living index number for 1929 with 1928 = 100, is less than 100, we therefore conclude that prices in 1929 as compared with the prices in 1928 have fallen down.

5.43. (i) Construction of the consumer price index number, using the Aggregate Expenditure Method.

Commodity	Quantity consumed in 1939 (q_0)	Unit of price	Price in 1939 (p_0)	Price in 1940 (p_1)	p_0q_0	p_1q_0
Rice	240 kg	Rs. per kg.	5.75	6.00	1380	1440
Wheat	240 kg	Rs. per kg.	5.00	8.00	1200	1920
Gram	40 kg	Rs. per kg.	6.00	9.00	240	360
Sugar	40 kg	Rs. per kg.	20.00	15.00	800	600
Arhar	240 kg	Rs. per kg.	8.00	10.00	1920	2400
Ghee	4 kg	Rs. per kg.	2.00	1.50	8	6
Total:	---	---	---	---	5548	6726

$\therefore$ Consumer price index for 1940 $= \dfrac{\sum p_1q_0}{\sum p_0q_0} \times 100$

$$= \frac{6726}{5548} \times 100 = 121.23$$

(ii) Construction of the consumer price index number, using the Household Budget Method.

Commodity	Quantity consumed in 1939 (q_0)	Unit of price	Price in 1939 (p_0)	Price in 1940 (p_1)	Price relative $I = p_1/p_0 \times 100$	Weight W ($= p_0q_0$)	$W \times I$
Rice	240 kg	Rs. per kg.	5.75	6.00	104.3	1380	143934
Wheat	240 kg	Rs. per kg.	5.00	8.00	160.0	1200	192000
Gram	40 kg	Rs. per kg.	6.00	9.00	150.0	240	36000
Sugar	40 kg	Rs. per kg.	20.00	15.00	75.0	800	60000
Arhar	240 kg	Rs. per kg.	8.00	10.00	125.0	1290	240000
Ghee	4 kg	Rs. per kg	2.00	1.50	75.00	8	600
Total:	---	---	—	—	—	5548	672534

Hence the consumer price index number for 1940 $= \dfrac{\Sigma W \times 1}{\Sigma W}$

$$= \frac{672534}{5548} = 121.22$$

5.44. (i) Computation of consumer price index using the base year quantities as weights.

Article	Quantity	Price (Rs.) in		p_0q_0	p_1q_0
	2000 (q_0)	2000 p_0	2007 p_1		
Food	5 mds	18.00	26.50	90.00	132.50
Cloth	30 meters	2.60	2.80	78.00	84.00
Electricity	75 units	0.25	0.30	18.75	22.50
Rent	3 rooms	30.00	27.50	90.00	82.50
Miscellaneous	34 units	0.50	0.60	17.00	20.40
Total	--	--	--	293.75	341.90

Consumer price index for 2007 is

$$P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{341.90}{293.75} \times 100 = 116.39$$

(ii) Consumer price index number using the values in the base year as weights.

Article	Quantity 2000 (q_0)	Price in 2000 (p_0)	Price in 2007 (p_1)	Price Relative $I = p_1/p_0 \times 100$	Weight W ($= p_0q_0$)	$W \times I$
Food	5 mds	18.00	26.50	147.22	90.00	13249.80
Cloth	30 meters	2.60	2.80	107.69	78.00	8399.82
Electricity	75 units	0.25	0.30	120.00	18.75	2250.00
Rent	3 rooms	30.00	27.50	91.67	90.00	8250.30
Miscellaneous	34 units	0.50	0.60	120.00	17.00	2040.00
Total	---	---	---	---	293.75	34189.92

Hence the consumer price index number for 2000 is

$$P_{01} = \frac{\Sigma WI}{\Sigma W} = \frac{34189.92}{293.75} = 116.39$$

5.45. Computation of retail price index number.

Item	Unit	Purchase (units) q_0	Prices in July 2000 (p_0)	Prices on 1 Sept. 2003 (p_1)	p_0q_0	p_1q_0
Flour	Kilogram	18	1.90	2.30	34.20	41.40
Meat	Kilogram	2	22.00	28.00	44.00	56.00
Bread	200 gram	2	1.00	1.50	2.00	3.00
Tea	450 gram	4	8.25	10.35	33.00	41.40
Sugar	Kilogram	3	7.00	7.75	21.00	23.25
Milk	Liter	2.5	3.50	4.00	8.75	10.00
Butter	450 gram	2	12.30	15.00	24.60	30.00
Eggs	Dozen	1.5	6.50	10.50	9.75	15.75
Potatoes	Kilogram	10	2.60	3.20	26.00	32.00
Total	---	---	---	---	203.30	252.80

Hence the desired price index for 2003 is

$$P_{01} = \frac{\Sigma p_1 p_0}{\Sigma p_0 p_0} \times 100 = \frac{252.80}{203.30} \times 100 = 124.35$$

5.46 Computation of Fisher's quantity Index and Simple aggregative value index.

Quantity		Value		Price			
q_0	q_1	q_0p_0	q_1p_1	$p_0 = \frac{q_0p_0}{q_0}$	$p_1 = \frac{q_1p_1}{q_1}$	p_0q_1	q_0p_1
100	150	600	1200	$\frac{600}{100} = 6$	$\frac{1200}{150} = 8$	900	800
80	100	400	700	$\frac{400}{80} = 5$	$\frac{700}{100} = 7$	500	562
60	72	180	432	$\frac{180}{60} = 3$	$\frac{432}{72} = 6$	216	360
30	33	450	363	$\frac{450}{30} = 15$	$\frac{363}{33} = 11$	495	330
		Σq_0p_0 1630	Σq_1p_1 2695			Σp_0q_1 2111	Σq_0p_1 2050

i) Fisher's Quantity Index Number for 2006:

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} \times 100$$

$$= \sqrt{\frac{2111}{1630} \times \frac{2695}{2050}} \times 100 = 130.48$$

ii) Simple Aggregative value Index for 2006:

$$V_{01} = \frac{\Sigma V_1}{\Sigma V_0} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_0} \times 100 = \frac{2695}{1630} \times 100 = 165.34$$

5.47

i) Simple Aggregative Value Index

$$V_{01} = \frac{\Sigma V_1}{\Sigma V_0} \times 100$$

$$= \frac{3203}{4560} \times 100$$

$$= 70.24$$

ii) Index of Average values for each year by using Paasche's Formula is:

for 2003; $P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100 = \frac{950}{832} \times 100 = 114.18$

for 2004; $P_{02} = \frac{\Sigma p_2 q_2}{\Sigma p_0 q_2} \times 100 = \frac{1300}{807} \times 100 = 161.09$

for 2005; $P_{03} = \frac{\Sigma p_3 q_3}{\Sigma p_0 q_3} \times 100 = \frac{1450}{704} \times 100 = 205.97$

●●●●●●●●●

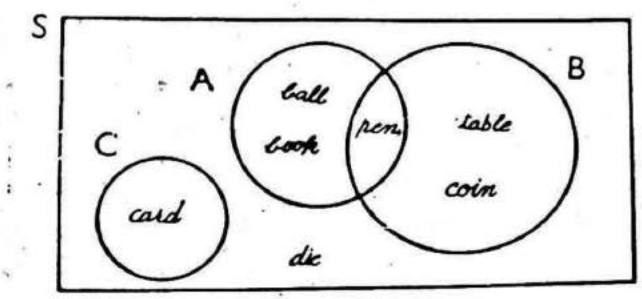
Chapter 6

PROBABILITY

6.1. Given S = {chair, student, pen}.

Proper subsets are {chair, student}, {chair, pen}, {student, pen}, {chair}, {student}, {pen}, ϕ.

6.2. The given sample space is illustrated by the following Venn Diagram:



6.3. (a) Given $A = \{1, 2\}$, $B = \{2, 3\}$ and $C = \{3\}$. Then

(i) $A \times A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$

(ii) $A \times B = \{(1, 2), (1, 3), (2, 2), (2, 3)\}$

(iii) $B \times A = \{(2, 1), (2, 2), (3, 1), (3, 2)\}$

(iv) Now

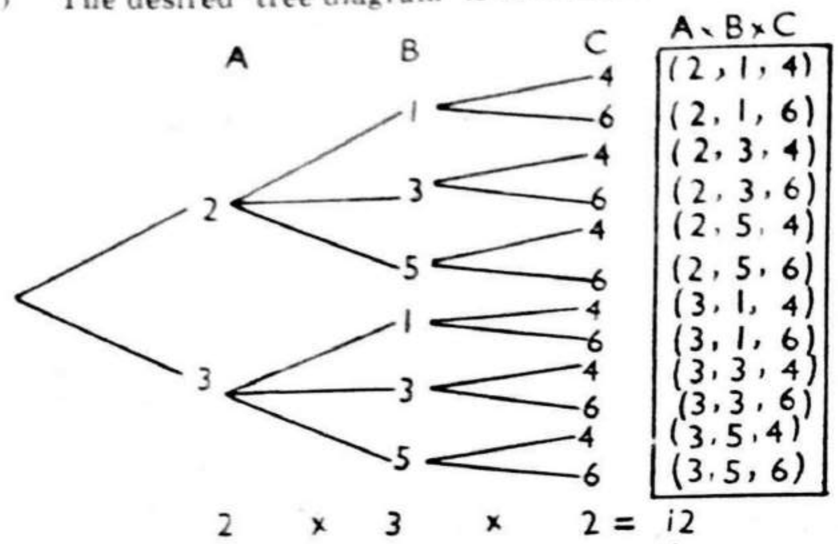
$A \times B = \{(1, 2), (1, 3), (2, 2), (2, 3)\}$, and

$B \times C = \{(2, 3), (3, 3)$. Therefore

$(A{\times}B) \cup (B{\times}C) = \{(1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$

(v) $(A \times B) \cap (B \times C) = \{(2, 3)\}$

(b) The desired "tree diagram" is constructed as below:



6.4. (b) $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

$A = \{2, 3, 4\}$, $B = \{3, 4, 5\}$ and $C = \{5, 6, 7\}$

Now $\bar{A}$ is the complement of A. It consists of the members in S which are not in A. Thus

$\bar{A} = \{1, 5, 6, 7, 8, 9, 10\}$

Similarly, $\bar{B} = \{1, 2, 6, 7, 8, 9, 10\}$

(i) $\bar{A} \cap B$ = The intersection of $\bar{A}$ and B consists of those elements which belong to both $\bar{A}$ and B.

Hence $\bar{A} \cap B = \{5\}$.

(ii) $\bar{A} \cup B$ = The union of $\bar{A}$ and B consists of those elements whcih belong to $\bar{A}$ or to B or to both.

Hence $\bar{A} \cup B = \{1, 3, 4, 5, 6, 7, 8, 9, 10\}$.

(iii) $\bar{A} \cap \bar{B}$ = The intersection of $\bar{A}$ and $\bar{B}$ consists of those elements belonging to $\bar{A}$ and $\bar{B}$

$= \{1, 6, 7, 8, 9, 10\}$

$\overline{\bar{A} \cap \bar{B}}$ = The complement of $\bar{A} \cap \bar{B}$ consists of those elements in S which are not in $\bar{A} \cap \bar{B}$

Hence $\overline{\bar{A} \cap \bar{B}} = \{2, 3, 4, 5\}$.

(iv) Now $B \cup C = \{3, 4, 5, 6, 7\}$, $A \cap (B \cup C) = \{3, 4\}$

$\therefore$ $\overline{A \cap (B \cup C)}$ = The complement of $A \cap (B \cup C)$

$= \{1, 2, 5, 6, 7, 8, 9, 10\}$.

6.5 Here $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{0, 2, 4, 6, 8\}$,

$B = \{1, 3, 5, 7, 9\}$, $C = \{2,3,4,5\}$ and $D = \{1,6,7\}$

(i) $A \cup C$ = The union of A and C consists of those elements which belong to A or to C or to both.

Hence $A \cup C = \{0, 2, 3, 4, 5, 6, 8\}$.

(ii) $A \cap B$ = The intersection of A and B consists of those elements which belong to both A and B.

Hence $A \cap B = \phi$

(iii) $\bar{C}$ = The complement of C consists of those elements in S which are not in C.

Hence $\bar{C} = \{0, 1, 6, 7, 8, 9\}$

(iv) $\bar{C} \cap D$ = The intersection of $\bar{C}$ and D consists of those elements which belong to both $\bar{C}$ and D.

$\therefore$ $\bar{C} \cap D = \{1, 6, 7\}$.

$(\bar{C} \cap D) \cup B$ = The union of $(\bar{C} \cap D)$ and B consists of those elements which belong to at least one of them.

Hence $(\bar{C} \cap D) \cup B = \{1, 3, 5, 6, 7, 9\}$

(v) $S \cap C$ = The intersection of S and C consists of those elements which belong to both S and C.

$\therefore$ $S \cap C = \{2, 3, 4, 5\}$

$\overline{S \cap C}$ = The complement of $S \cap C$ consists of elements in S which are not in C.

Hence $(\overline{S \cap C}) = \{0, 1, 6, 7, 8, 9\}$.

(vi) $A \cap C \cap \bar{D}$ = The intersection of A, C and $\bar{D}$ consists of elements which belong to A and to C and to $\bar{D}$.

Hence $A \cap C \cap \bar{D} = \{2, 4\}$.

6.6 (a) The sample space S is represented by the following array of 36 equally likely outcomes:

$$S = \begin{Bmatrix} (1,1) & (2,1) & (3,1) & (4,1) & (5,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{Bmatrix}$$

The two events are:

$A = \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}$

$B = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,6)\ (6,1), (6,2), (6,3), (6,4), (6,5)\}$.

(b) Using the sample space given in (a), we see that the two events A and B consists of the following elements:

$A = \{(1,2), (1,4), (1,6), (2,1), (2,3), (2,5), (3,2), (3,4), (3,6), (4,1), (4,3), (4,5), (5,2), (5,4), (5,6), (6,1), (6,3), (6,5)\}$; and

$B = \{(1,3), (2,3), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,3)\ (5,3), (6,3)\}$

The complements of A and B are

$\bar{A} = \{(1,1), (1,3), (1,5), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), (4,2), (4,4), (4,6), (5,1), (5,3), (5,5), (6,2), (6,4), (6,6)\}$; and

$\bar{B} = \{(1,1), (2,1), (4,1), (5,1), (6,1), (1,2), (2,2), (4,2), (5,2), (6,2), (1,4), (2,4), (4,4), (5,4), (6,4), (1,5), (2,5), (4,5), (5,5), (6,5), (1,6), (2,6), (4,6), (5,6), (6,6)\}$

Now $A \cup B$ = The union of A and B consists of those elements which belong to A or B or both.

$\therefore$ $A \cup B = \{(1,2), (1,4), (1,6), (2,1), (2,3), (2,5), (3,2), (3,4), (3,6), (4,1), (4,3), (4,5), (5,2), (5,4), (5,6), (6,1), (6,3), (6,5), (1,3), (3,1), (3,3), (3,5), (5,3)\}$

Thus $A \cup B$ contian 23 elements.

$A \cap B$ = The intersection of A and B consists of those elements which belong to both A and B. Thus

$A \cap B = \{(2,3), (3,2), (3,4), (4,3), (3,6), (6,3)\}$, *i.e.* $n(A \cap B) = 6$,

$A - B = A \cap \bar{B}$ = The set of all elements of A which do not belong to B.

$= \{(1,2), (1,4), (1,6), (2,1), (2,5), (4,1), (4,5), (5,2), (5,4), (5,6), (6,1), (6,5)\}$; *i.e.* $n(A-B) = 12$.

$(A \cap \bar{B}) \cup \bar{A}$ = The union of $(A \cap \bar{B})$ and $\bar{A}$ consists of those elements which belong to $(A \cap \bar{B})$ or $\bar{A}$ or both. Thus

$(A \cap \bar{B}) \cup \bar{A} = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,4), (2,5), (2,6), (3,1), (3,3), (3,5), (4,1), (4,2), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,4), (6,5), (6,6)\}$,

i.e. $(A \cap \bar{B}) \cup \bar{A}$ consists of 30 sample points.

6.7 (a) All the possible (i) combinations and (ii) permutations of 3 letters chosen from the four letters A, B, C and D are listed in the table below:

Combinations	Permutations
ABC	*ABC, ACB, BAC, BCA, CAB, CBA*
ABD	*ABD, ADB, BAD, BDA, DAB, DBA*
ACD	*ACD, ADC, CAD, CDA, DAC, DCA*
BCD	*BCD, BDC, CBD, CDB, DBC, DCB*

Enumeration gives a total of 4 combinations and 24 permutations: These results are also obtained by using the

(i) The total number of combinations is

$$\binom{n}{r} = \binom{4}{3} = \frac{4!}{3!\,(4-3)!} = 4, \text{ and}$$

(ii) the total number of permutations is

$$^{n}P_{r} = {}^{4}P_{3} = \frac{4!}{(4-3)!} = 4 \times 3 \times 2 = 24.$$

(b) (i) It is evident that the order in which 3 officers: *i.e.* 1 president, 1 vice-president and 1 secretary-treasurer are to be chosen is of significance. Thus there are 15 choices for the first officer, 14 choices for the second officer **and** 13 choices for the third officer. Hence, the number of sample points is **$15\times14\times13$**$=2730$. In other words, **the number of permutations is**

$$^{15}P_3 = \frac{15!}{(15-3)!} = \frac{15 \times 14 \times 13 \times 12!}{12!} = 2730$$

(ii) Since the order in which the three persons of the committee are chosen, is unimportant, it is therefore a problem involving combinations. Thus

$$\binom{15}{3} = \frac{15!}{(15-3)!\,3!} = \frac{15 \times 14 \times 13}{1 \times 2 \times 3} = 455.$$

6.8. The total number of possible bridge hands that can be selected from an ordinary deck of 52 cards is given by

$$\binom{52}{13} = \frac{52!}{(52-13)!\,13!} = \frac{52!}{39!\,13!} = 635{,}013{,}559{,}600$$

6.9. The total number of possible ways in which $n=10$ persons can be divided into 3 groups consisting of 5, 3 and 2 persons is given by

$$\binom{10}{5\ 3\ 2} = \frac{10!}{5!\,3!\,2!} = \frac{10\times9\times8\times7\times6\times5!}{5!\,3\times2\times1\times2\times1} = 2520$$

6.10. (b) The sample space consists of 4 elements a, b, c, d, i.e. S = {a, b, c, d}.

Then the possible subsets or events are

ϕ, $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$, $\{c, d\}$, $\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, $\{b, c, d\}$, $\{a, b, c, d\}$.

Thus we have $16=2^4$ subsets of events which include the empty set and the set itself.

(c) Let $S = \{e_1, e_2, e_3, \ldots, e_n\}$. Then all the possible subsets will consists of

(1) the subsets with no element, *i.e.* the empty set ϕ.

(2) the subsets consisting of exactly 1 elements such as $\{e_1\}, \{e_2\}, \ldots$

(3) the subsets consisting of exactly 2 elements such as $\{e_1, e_2\}, \{e_1, e_3\}, \ldots$

... ...

$(n+1)$, the set itself consisting of all n elements $\{e_1, e_2, \ldots, e_n\}$.

Now,

(1) the number of empty subsets is given by $\binom{n}{0}$;

(2) the number of subsets with exactly one element is given by $\binom{n}{1}$;

(3) the number of subsets with exactly 2 elements is given by $\binom{n}{2}$;

... ...

$(n+1)$ the number of subsets with all n elements is given by $\binom{n}{n}$.

Thus the number of all possible subsets of a set containing n elements is given by

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n}.$$

Again, $2^n = (1 + 1)^n = \binom{n}{0}(1)^n + \binom{n}{1}(1)^{n-1}(1) + \binom{n}{2}(1)^{n-2}(1)^2 + \ldots + \binom{n}{n}(1)^n.$

$$= \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n}$$

Hence the result.

6.12. (a) Let A, B and C denote the events that the stock price will go up, remain unchanged and go down respectively.

Then $P(A) = 0.60$, $P(B) = 0.38$ and $P(C) = 0.25$.

Since the three events are mutually exclusive and collectively exhaustive, therefore their sum should be equal to one. But

$$P(A) + P(B) + P(C) = 0.60 + 0.38 + 0.25 = 1.23,$$

which is greater than 1.

Hence the investment counsellor's claim is wrong.

(b) When two coins are tossed once, the sample space consists of 4 equally likely sample points, *i.e.*

$$S = \{HH, HT, TH, TT\}.$$

As the sample points are equally likely, a probability of $\frac{1}{4}$ is assigned tc each sample point. Thus

$P(HH) = 1/4$, $P(\text{one head and one tail}) = 2/4 = 1/2$, and

$P(2 \text{ tails}) = 1/4$.

Hence the given statement is wrong.

(c) The events "no accident", "one accident" and "two or more accidents" are mutually exclusive, therefore the sum of their probabilities should not exceed 1.

Now the sum of the given probabilities = 0.90 + 0.02 + 0.09 = 1.01, which exceeds 1.

Hence the statement is not correct.

(d) **As the events A, B and C are mutually exclusive, the sum of their probabilties should not exceed unity.**

Now $P(A) = \dfrac{1}{6}$,

$$\frac{2}{3}P(B) = \frac{1}{6} \text{ or } P(B) = \frac{1}{6} \times \frac{3}{2} = \frac{1}{4};$$

$$\frac{1}{4}P(C) = \frac{1}{6} \text{ or } P(C) = \frac{2}{3},$$

Adding, $P(A) + P(B) + P(C) = \dfrac{1}{6} + \dfrac{1}{4} + \dfrac{2}{3} = \dfrac{2 + 3 + 8}{12} = \dfrac{13}{12}$.

The sum turns out to be greater than 1, so the given statement is wrong.

6.13. (i) The sample space S contains 6 sample points, i.e.

$$S = \{1, 2, 3, 4, 5, 6\}$$

Let A denote the event that an odd number appears.

Then $A = \{1, 3, 5\}$

Hence $P(A) = \dfrac{\text{number of sample points in } A}{\text{number of sample points in } S} = \dfrac{n(A)}{n(S)} = 3/6 = 1/2.$

(ii) **The sample space S in a single toss of a pair of fair dice consists of 36 elements.**

Let A be the event that the sum 8 appears. Then

$$A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$

Hence $P(A) = \dfrac{\text{number of sample points in } A}{\text{number of sample points in } S} = \dfrac{5}{36}$.

(iii) **The sample space S consists of 8 elements.**

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

Let A be the event that at least one head appears. Then

$$A = \{HHH, HHT, HTH, THH, HTT, THT, TTH\}.$$

Hence $P(A) = \frac{7}{8}$.

(iv) **Let S = sample space and A = a king, ace, jack of clubs or queen of diamonds appear. Then**

S can occur in $\dbinom{52}{1} =$ 52 ways, the number of ways that a single card can be drawn from 52 cards.

A can occur in $\binom{10}{1} = 10$ ways, the number of ways that the card drawn is a king, ace, jack of clubs or queen of diamonds.

Hence $P(A) = \frac{10}{52} = \frac{5}{26}$

6.14. (b) The box contains 10 red, 30 white, 20 blue and 15 orange, i.e. 75 marbles in all.

One marble can be drawn in $\binom{75}{1} = 75$ ways

$\therefore$ (i) $P(\text{marble is orange or red}) = \frac{15+10}{75} = \frac{1}{3}$

(ii) $P(\text{marble is not-'red or blue'}) = \frac{45}{75} = \frac{3}{5}$

(iii) $P(\text{marble is not blue}) = \frac{55}{75} = \frac{11}{15}$

(iv) $P(\text{marble drawn is white}) = \frac{30}{75} = \frac{2}{5}$

(v) $P(\text{marble is red, white or blue}) = \frac{60}{75} = \frac{4}{5}$

6.15. (a) The sample space S is represented by the following array of 36 equally likely outcomes:

$$S = \begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{Bmatrix}$$

The various total number of dots that may turn up are, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Let X denote the total of dots that may turn up. Then

$$P(X=2) = \frac{\text{number of sample points making a total of 2}}{\text{number of sample points in } S} = 1/36;$$

$$P(X=3) = \frac{\text{number of sample points making a total of 3}}{\text{number of sample points in } S} = 2/36,$$

Similarly, we find that

$P(X=4) = 3/36$; $\quad P(X=5) = 4/36$;

$P(X=6) = 5/36$; $\quad P(X=7) = 6/36$;

$P(X=8) = 5/36$; $\quad P(X=9) = 4/36$;

$P(X=10) = 3/36$; $\quad P(X=11) = 4/36$;

$P(X=12) = 1/36$,

Again, $P(\text{dots will total at least 4}) = P(X \le 4) = \frac{33}{36} = \frac{11}{12}$

(b) Whenever two dice are thrown, the sample space S has 36 outcomes, which we assume, are equally likely and a probability of $\frac{1}{36}$ is attached with each outcome.

Let A be the event that a total of *more than* 7 occurs and B, the event that a total of *less than* 7 comes up. Then the event A has 15 outcomes, *i.e.*

$$A = \{(2,6), (3,6), (4,6), (5,6), (6,6), (3,5), (4,5), (5,5), (6,5), (4,4), (5,4), (6,4), (5,3), (6,3), (6,2)\}.$$

Thus $P(A) = \frac{15}{36} = \frac{5}{12}$.

Similarly, we see that the event B contains 15 sample points.

$\therefore P(B) = \frac{15}{36} = \frac{5}{12} = P(A)$.

Let C be the event that *exactly* 7 is thrown. Then A, B and C are mutually exclusive and collectively exhaustive events. Therefore

$$P(A) + P(B) + P(C) = 1$$

or $\quad P(C) = 1 - P(A) - P(B) = 1 - \frac{5}{12} - \frac{5}{12} = \frac{1}{6}$.

6.16. (a) S consists of 36 sample points.

Let $\quad A$ = (sum of the upper face numbers is odd), and

B = (at least one ace). Then

$$A = \begin{Bmatrix} (1,2) & (2,1) & (3,2) & (4,1) & (5,2) & (6,1) \\ (1,4) & (2,3) & (3,4) & (4,3) & (5,4) & (6,3) \\ (1,6) & (2,5) & (3,6) & (4,5) & (5,6) & (6,5) \end{Bmatrix}$$

$B = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (3,1), (4,1), (5,1), (6,1)\}$

$$\bar{B} = \begin{Bmatrix} (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{Bmatrix}$$

Now $A \cap B$ = The intersection of A and B consists of those points which belong to both A and B.

$\therefore \quad A \cap B = \{(1,2), (1,4), (1,6), (2,1), (4,1), (6,1)\}$

It thus contains 6 sample points.

And $A \cup B$ = The unon of A and B consists of those points which belong to A or B or both.

$$\therefore \quad A \cup B = \begin{Bmatrix} (1,1) & (1,5) & (2,5) & (4,1) & (5,1) & (6,1) \\ (1,2) & (1,6) & (3,1) & (4,3) & (5,2) & (6,3) \\ (1,3) & (2,1) & (3,2) & (4,5) & (5,4) & (6,5) \\ (1,4) & (2,3) & (3,4) & (3,6) & (5,6) & \end{Bmatrix}$$

i.e. $A \cup B$ contains 23 sample points.

Again $A \cap \bar{B}$ = The intersection of A and $\bar{B}$ consists of points which belong to both A and $\bar{B}$.

Thus $A \cap \bar{B} = \begin{Bmatrix} (2,3), (2,5), (3,2), (3,4), (3,6), (4,3) \\ (4,5), (5,2), (5,4), (5,6), (6,3), (6,5) \end{Bmatrix}$

i.e. $A \cap \bar{B}$ consists of 12 sample points.

Hence $P(A \cap B) = \frac{6}{36} = \frac{1}{6}$, $P(A \cup B) = \frac{23}{36}$, and

$$P(A \cup \bar{B}) = \frac{12}{36} = \frac{1}{3}.$$

(b) Let S = sample space, A = the sum shown is 8, and B = the two show the same number. Then S consists of 36 elements.

$A = \{(6,2), (5,3), (4,4), (3,5), (2,6)\}$

$B = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$

Hence $P(A) = \frac{5}{36}$; $P(B) = \frac{6}{36} = \frac{1}{6}$;

$P(A \cap B) = \frac{1}{36}$; as $A \cap B = \{(4,4)\}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= \frac{5}{36} + \frac{6}{36} - \frac{1}{36} = \frac{10}{36} = \frac{5}{18};$$

6.17. The possible outcomes, when the numbers on two discs are drawn without replacement, are $\binom{6}{2} = 15$, which are given below:

(1,2), (1,3), (1,4), (1,5), (1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)

Let k denote the sum of two numbers. Then

$P(k = 3) = \frac{1}{15}$; $P(k=4) = 1/15$; $P(k=5) = 2/15$;

$P(k = 6) = \frac{2}{15}$; $P(k=7) = 3/15$; $P(k=8) = 2/15$;

$P(k = 9) = \frac{2}{15}$; $P(k=10) = 1/15$; $P(k=11) = 1/15$;

6.18. The sample space S consists of $6^2 = 36$ sample points.

(i) Let A be the event that the product of the numbers on the dice is between 8 and 16. Then

$A = \{(4,2), (3,3), (2,4), (5,2), (4,3), (3,4), (2,5), (6,2), (5,3), (4,4), (3,5), (2,6)\}$, *i.e.* $n(A) = 12$ sample points.

$$\therefore \quad P(A) = \frac{n(A)}{n(S)} = \frac{12}{36} = \frac{1}{3}.$$

(ii) Let B be the event that the product of the numbers on the dice is divisible by 4. Then

$$B = \{(4,1),\ (1,4),\ (4,2),\ (2,4),\ (6,2),\ (4,4),\ (2,6),\ (4,5),\ (4,3),\ (2,2),\ (3,4),\ (5,4),\ (6,4),\ (4,6),\ (6,6)\},$$

i.e. $n(B) = 15$ sample points.

$$\therefore \quad P(B) = \frac{n(B)}{n(S)} = \frac{15}{36}.$$

6.19. (b) Let A be the event that *at least* one 6 occurs in 4 tosses of a fair die, and $\bar{A}$ be the event that no 6 occurs. Then

$$P(A) = 1 - P(\bar{A}) = 1 - \left(\frac{5}{6}\right)^4 = 1 - \frac{625}{1296} = 0.52$$

Let B be the event that *at least* one double six occurs in 24 tosses of two fair dice, and $\bar{B}$, the event that no double six occurs. Then

$$P(B) = 1 - P(\bar{B}) = 1 - \left(\frac{35}{36}\right)^{24} = 0.49$$

(c) The sample space S consists of 216 sample points.

The sample points in different events can be conveniently obtained by combining the faces of the third die with pertinent sums of the first two dice. The sums are given below:

Ist die

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

Now 9 can be made up by the following combinations:

3rd die	Sum of the 1st and 2nd die	No. of possibilities
1	8	5
2	7	6
3	6	5
4	5	4
5	4	3
6	3	2
Total		25

Hence $\quad P(\text{getting a nine}) = \dfrac{25}{216}.$

Similarly, 10 can be made up in 27 ways.

Thus $\quad P(\text{getting a 'ten'}) = \dfrac{27}{216} = \dfrac{1}{8}.$

6.20. (a) The sample space S contains $\binom{52}{2} = 1326$ sample points, the number of ways in which two cards can be drawn from 52 cards.

Let A be the event that one card drawn is a king and the other is a queen. Then

A contains $\binom{4}{1}\binom{4}{1} = 16$ sample points.

$$\therefore \quad P(A) = \frac{n(A)}{n(S)} = \frac{16}{1326} = \frac{8}{663}.$$

(b) $\quad S$ consists of $\binom{8}{5} = 56$ sample points.

A can choose one joker and 4 other cards in $\binom{1}{1}\binom{7}{4}$, *i.e.* 35 ways.

$$\therefore \quad P(A \text{ has the Joker}) = \frac{35}{56} = \frac{5}{8}.$$

6.21 (a) There are 10 good and 2 bad eggs in the refrigerator.

S can occur in $\binom{12}{4} = 495$ ways, the number of ways in which 4 eggs can be chosen from 12 eggs.

Let A denote the event that exactly one egg is bad and B denote the event that at least one egg is bad. Then

(i) A can occur in $\binom{2}{1}\binom{10}{3} = 240$ ways.

Hence $P(A) = \frac{240}{495} = \frac{16}{33}$.

(ii) $P(B) = 1 - P(\text{no egg is bad}) = 1 - P(\bar{B})$, where

$\bar{B}$ can occur in $\binom{10}{4} = 210$ ways.

$\therefore$ $P(B) = 1 - \frac{210}{495} = \frac{285}{495} = \frac{19}{33}$.

(b) The carton contains 3 bad eggs and 9 good eggs. S can occur in $\binom{12}{3} = 220$ ways, the number of ways in which 3 eggs can be chosen from 12 eggs.

(i) Let A denote the event that no bad egg (*i.e.* all good eggs) is chosen. Then $n(A) = \binom{9}{3} = 84$.

$\therefore$ $P(A) = \frac{84}{220} = 0.38$.

(ii) *At least* one bad egg is the complement of the event that no bad egg is chosen. Therefore,

$P(\text{at least one bad egg}) = 1 - P(\text{no bad egg})$

$= 1 - 0.38 = 0.62$

(iii) Let B be the event that *exactly* 2 bad eggs are chosen. Then $n(B) = \binom{3}{2}\binom{9}{1} = 27$.

$\therefore$ $P(B) = \frac{27}{220} = 0.12$.

6.22. (a) The sample space S for this experiment is

$S = \{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, and therefore $n(S) = 10$.

Let A be the event that the integer chosen is an even number. Then

$A = \{4, 6, 8, 10, 12\}$, *i.e.* $n(A) = 5$.

$\therefore$ $P(A) = \frac{n(A)}{n(S)} = \frac{5}{10} = \frac{1}{2}$

Let B be the event that the integer chosen is an even number and is divisible by 3. Then

$B = \{6, 12\}$, *i.e.* $n(B) = 2$.

$\therefore$ $P(B) = \frac{n(B)}{n(S)} = \frac{2}{10} = \frac{1}{5}$.

(b) The sample space S contains $\binom{20}{3}$, *i.e.* 1140 sample points, the number of ways in which 3 digits can be chosen from 20 digits.

(i) There are two mutually exclusive ways of getting an even sum, *i.e.* either 3 digits are even or 2 odd and 1 even. This can occur in $\binom{10}{3} + \binom{10}{2}\binom{10}{1}$, *i.e.* $120 + 450 = 570$ ways.

$\therefore$ $P(\text{Sum is even}) = \frac{570}{1140} = \frac{1}{2}$.

(ii) When all 3 digits are odd, the product will not be even; and 3 odd digits from 10 odd digits can be chosen in $\binom{10}{3}$, *i.e.* 120 ways.

Thus $P(\text{product is even}) = 1 - P(\text{product is not even})$

$= 1 - \frac{120}{1140} = 1 - \frac{2}{19} = \frac{17}{19}$.

(c) S consists of $\binom{13}{3}$ = 286 sample points.

(i) Let A be the event that all balls drawn are of different colours. Then A contains $\binom{4}{1}\binom{4}{1}\binom{5}{1}$, *i.e.* $4 \times 4 \times 5 = 80$ sample points.

$\therefore \quad P(A) = \frac{n(A)}{n(S)} = \frac{80}{286} = \frac{40}{143}.$

(ii) Let B be the event that all balls drawn are of same colour. Then B contains $\binom{4}{3} + \binom{4}{3} + \binom{5}{3}$, *i.e.* $4+4+10=18$ sample points.

$\therefore \quad P(B) = \frac{n(B)}{n(S)} = \frac{18}{286} = \frac{9}{143}.$

6.23 (a) We know that in one throw of a die, $P(6) = \frac{1}{6}$ and $P(\text{not } 6) = \frac{5}{6}$

(i) When 5 dice are thrown, then

$$P(\textit{at least} \text{ one six}) = 1 - P(\text{no 6's})$$
$$= 1 - \left(\frac{5}{6}\right)^5 = 0.598$$

(ii) When n dice are thrown, then

$$P(\textit{at least} \text{ one six}) = 1 - P(\text{no 6's}) = 1 - \left(\frac{5}{6}\right)^n.$$

(b) We need to find n such that

$$1 - \left(\frac{5}{6}\right)^n \geq 0.99, \text{ i.e. } \left(\frac{5}{6}\right)^n \leq 0.01$$

Taking logs of both sides, we get

$$n \log\left(\frac{5}{6}\right) \leq \log 0.01$$

Dividing both sides by $\log\left(\frac{5}{6}\right)$ and reversing the inequality sign as $\log\left(\frac{5}{6}\right)$ is negative, we have

$$n \geq \frac{\log(0.01)}{\log(5/6)} \; [\because \log 0.01 = \bar{2}.0000, \log(5/6) = \log(0.833) = \bar{1}.9206]$$

≥ 25.3 so least $n = 26$.

Hence 26 dice must be thrown so that the probability of obtaining at least one 6 is at least 0.99.

(c) $P(\text{target is hit at least once}) = 1 - P(\text{target is not hit})$

$$= 1 - (0.3)^n$$

Now we need to find n such that

$$1 - (0.3)^n \geq 0.995, \text{ i.e. } (0.3)^n \leq 0.005$$

Taking log of both sides, we obtain

$$n \log(0.3) \leq \log(0.005)$$

Dividing both sides by log (0.3) and reversing the inequality sign since log (0.3) is negative, we have

$$n \geq \frac{\log(0.005)}{\log(0.3)}$$
$$\geq \frac{\bar{3}.6990}{\bar{1}.4771} \geq 4.4$$

so least $n = 5$

Hence 5 missiles should be fired so that the probability that the target is hit *at least* once is greater than 0.995.

6.24. The sample space S contains $\binom{14}{6} = 3003$ sample points, which are all equally likely, exhaustive and mutually exclusive.

(i) Let A denote the event that 3 red and 3 not-red, *i.e.* black or white balls are drawn. Then

A contains $\binom{4}{3} \times \binom{10}{3}$, *i.e.* $4 \times 120 = 480$ sample points.

$\therefore \quad P(A) = \frac{n(A)}{n(S)} = \frac{480}{3003} = \frac{160}{1001}.$

(ii) Let B denote the event that *at least* two white balls are drawn. Then *at least* two white means 2, 3, 4 or 5 white balls. Let B_1, B_2, B_3 and B_4 denote the events that 2 white and 4 others, 3 white and 3 others, 4 white and 2 others, 5 white and 1 other ball respectively are drawn. Then

B_1 contains $\binom{5}{2}\binom{9}{4}$ = 1260 sample points;

B_2 contains $\binom{5}{3}\binom{9}{3}$ = 840 sample points;

B_3 contains $\binom{5}{4}\binom{9}{2}$ = 180 sample points; and

B_4 contains $\binom{5}{5}\binom{9}{1}$ = 9 sample points.

Now $B = B_1 \cup B_2 \cup B_3 \cup B_4$. Therefore

$$n(B) = n(B_1) + n(B_2) + n(B_3) + n(B_4)$$
$$= 1260 + 840 + 180 + 9 = 2289.$$

Hence $P(B) = \dfrac{n(B)}{n(S)} = \dfrac{2289}{3003} = \dfrac{109}{143}$.

6.25. (a) The sample space S contains $\binom{10}{3}$ = 120 sample points, the number of ways in which 3 people can be selected from 10 applicants.

(i) Let A represent the event that the three selected will be girls. Then A contains $\binom{6}{3}$ = 20 sample points, the number of ways in which 3 girls can be selected from 6 girls.

Therefore $P(A) = \dfrac{n(A)}{n(S)} = \dfrac{20}{120} = \dfrac{1}{6}$.

(ii) Let B denote the event that the three selected will be boys. Then B contains $\binom{4}{3}$ = 4 sample points.

Therefore $P(B) = \dfrac{n(B)}{n(S)} = \dfrac{4}{120} = \dfrac{1}{30}$.

(iii) *At least* one boy means, one, two or three boys. Let C denote the event that *at least* one boy is selected. Then

$$n(C) = \binom{4}{1}\binom{6}{2} + \binom{4}{2}\binom{6}{1} + \binom{4}{3}\binom{6}{0}$$
$$= 60 + 36 + 4 = 100 \text{ sample points.}$$

Hence $P(C) = \dfrac{n(C)}{n(S)} = \dfrac{100}{120} = \dfrac{5}{6}$.

(b) The sample space S contains $\binom{14}{5}$ = 2002 sample points.

Let A denote the event that more men are chosen than women. Then

$$n(A) = \binom{6}{3}\binom{8}{2} + \binom{6}{4}\binom{8}{1} + \binom{6}{5}$$
$$= (20 \times 28) + (15 \times 8) + 6 = 686$$

Hence $P(A) = \dfrac{686}{2002} = 0.34$.

6.26 The sample space S contains $\binom{10}{3}$ = 120 sample points

(i) Let A denote the event that orders are placed with the in-state suppliers only. Then A contains $\binom{4}{3}$ = 4 sample points.

$$\therefore \quad P(A) = \frac{n(A)}{n(S)} = \frac{4}{120} = \frac{1}{30}.$$

(ii) Let B denote the event that orders are placed with the out-of-state suppliers only. Then B cointains $\binom{6}{3}$ = 20 sample points

$$\therefore \quad P(B) = \frac{n(B)}{n(S)} = \frac{20}{120} = \frac{1}{6}.$$

(iii) *At least* one in-state supplier means one, two or three in-state suppliers. Let C be the event that orders are placed with *at least* one in-state supplier. Then C contains $\binom{4}{1}\binom{6}{2} + \binom{4}{2}\binom{6}{1} + \binom{4}{3}$ = 60 + 36 + 4 = 100 sample points.

Therefore $\quad P(C) = \dfrac{n(C)}{n(S)} = \dfrac{100}{120} = \dfrac{5}{6}$.

6.27. (b) Let A be the event that the person is a man, and B be the event that the person has brown eyes. Then we need P(A ∪ B).

Now $\quad P(A) = \dfrac{10}{30}$, $P(B) = \dfrac{15}{30}$ and $P(A \cap B) = \dfrac{5}{30}$.

Hence $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= \frac{10}{30} + \frac{15}{30} - \frac{5}{30} = \frac{20}{30} = \frac{2}{3}.$$

(c) Let W be the event that the person chosen be a woman, and G be the event that the person wears glasses. Then we need P(W ∪ G)

Now $P(W) = \frac{7}{20}$, $P(G) = \frac{6}{20}$ and $P(W \cap G) = \frac{4}{20}$

Hence $P(W \cup G) = P(W) + P(G) - P(W \cap G)$

$$= \frac{7}{20} + \frac{6}{20} - \frac{4}{20} = \frac{9}{20} = 0.45.$$

6.28. (b) The sample space in this case is S = {1, 2, 3, 4, ..., 50} and therefore n(S) = 50.

Let A represent the event that the integer selected is divisible by 6, B, the event that the integer chosen is divisible by 8, and $A \cap B$, the event that the integer chosen is divisible by both 6 and 8, *i.e.* by 24. Then we seek $P(A \cup B)$.

Now $n(A) = [\frac{50}{6}] = 8$, $n(B) = [\frac{50}{8}] = 6$, and

$n(A \cap B) = [\frac{50}{24}] = 2$, where $[x]$ stands for the highest integer in x.

Hence $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= \frac{8}{50} + \frac{6}{50} - \frac{2}{50} = \frac{12}{50} = \frac{6}{25}.$$

6.29. (c) The sample space S consists of $\binom{200}{1}$ = 200 sample points.

Let A be the event that the item chosen is a bolt and B, the event that the item chosen is rusted. Then we need $P(A \cup B)$.

Now A contains $\binom{50}{1} = 50$ sample points, and

B contains $\binom{100}{1} = 100$ sample points. Therefore

$$P(A) = \frac{n(A)}{n(S)} = \frac{50}{200}, \text{ and } P(B) = \frac{n(B)}{n(S)} = \frac{100}{200}$$

The events A and B are not mutually exclusive, as an item may be both rusted and a bolt. Therefore the event

$A \cap B$ contains $\binom{25}{1} = 25$ sample points and

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{25}{200}.$$

Hence $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= \frac{50}{200} + \frac{100}{200} - \frac{25}{200} = \frac{125}{200} = \frac{5}{8}.$$

6.30. (b) S consists of 6^2 = 36 sample points.

Let A represent the event that the sum of dots is equal to 7 and B, be the event that the sum of dots is equal to 11. Then we need $P(A \cup B)$.

Now $P(A) = \frac{n(A)}{n(S)} = \frac{6}{36} = \frac{1}{6}$, and

$$P(B) = \frac{n(A)}{n(S)} = \frac{2}{36} = \frac{1}{18}.$$

Since the two events A and B are mutually exclusive, therefore

$$P(A \cup B) = P(A) + P(B) = \frac{1}{6} + \frac{1}{18} = \frac{2}{9}.$$

(c) (i) Since A and B are mutually exclusive events, therefore

$$P(A \cup B) = P(A) + P(B) = 0.4 + 0.5 = 0.9.$$

(ii) Now $P(\bar{A}) = 1 - P(A) = 1 - 0.4 = 0.6$.

6.31. (a) Now $A \cap \bar{B}$, A–B or A – (A ∩ B) are the same sets.

$\therefore$ $(A \cap \bar{B}) \cup (B \cap \bar{A}) = [A - (A \cap B)] \cup [B - (B \cap A)]$; and

$P[(A \cap \bar{B}) \cup (B \cap \bar{A})] = P[A - (A \cap B)] + P[B - (B \cap A)]$

$= P(A) - P(A \cap B) + P(B) - P(A \cap B)$

$= P(A) + P(B) - 2P(A \cap B)$

(b) $P(A \cup B) = \frac{3}{4}, P(\bar{A}) = \frac{2}{3}$ and $P(A \cap B) = \frac{1}{4}$ (given)

$\therefore$ (i) $P(A) = 1 - P(\bar{A}) = 1 - \frac{2}{3} = \frac{1}{3}$

(ii) Now $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\therefore$ $P(B) = P(A \cup B) - P(A) + P(A \cap B)$

$$= \frac{3}{4} - \frac{1}{3} + \frac{1}{4} = \frac{9 - 4 + 3}{12} = \frac{8}{12} = \frac{2}{3}.$$

(iii) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

$$= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

(c) Given: $P(A) = \frac{1}{2}, P(A \cup B) = \frac{3}{4}$ and $P(\bar{B}) = \frac{5}{8}$.

Then $P(\bar{A}) = 1 - P(A) = \frac{1}{2}$, and $P(B) = 1 - P(\bar{B}) = \frac{3}{8}$.

$\therefore$ (i) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

or $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

$$= \frac{1}{2} + \frac{3}{8} - \frac{3}{4} = \frac{4 + 3 - 6}{8} = \frac{1}{8}$$

(ii) Using De Morgan's Law $(\overline{A \cup B}) = \bar{A} \cap \bar{B}$, we have

$$P(\bar{A} \cap \bar{B}) = P[(\overline{A \cup B})]$$

$$= 1 - P(A \cup B) = 1 - \frac{3}{4} = \frac{1}{4}.$$

(iii) Using De Morgan's Law $(\overline{A \cap B}) = \bar{A} \cup \bar{B}$, we have

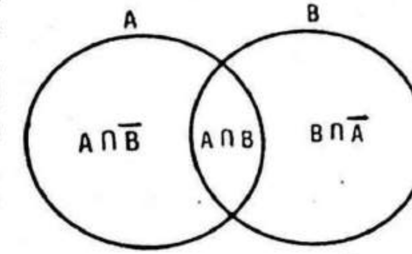
$$P(\bar{A} \cup \bar{B}) = P[(\overline{A \cap B})]$$

$$= 1 - P(A \cap B) = 1 - \frac{1}{8} = \frac{7}{8}.$$

(iv) $P(B \cap \bar{A}) = P(B - A \cap B) = P(B) - P(B \cap A)$

$$= \frac{3}{8} - \frac{1}{8} = \frac{1}{4}.$$

6.32. (i) Let A and B denote two sets of points with points in common represented by $A \cap B$. From the Venn diagram, we find that set A is composed of two disjoint sets $A \cap \bar{B}$ and $A \cap B$. Similarly, B is composed of $A \cap B$ and $B \cap \bar{A}$. The three mutually exclusive sets, i.e. $A \cap \bar{B}$, $A \cap B$ and $B \cap \bar{A}$ include points which are either in A or B. Thus the required probability is given by



$$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(B \cap \bar{A})$$

Now to obtain $A \cap \bar{B}$, the points in common represented by $A \cap B$ are removed from A. Thus

$$A \cap \bar{B} = A - A \cap B$$

$\therefore$ $P(A \cap \bar{B}) = P(A - A \cap B) = P(A) - P(A \cap B)$

Similarly, $P(B \cap \bar{A}) = P(B) - P(A \cap B)$

Substituting these values, we get

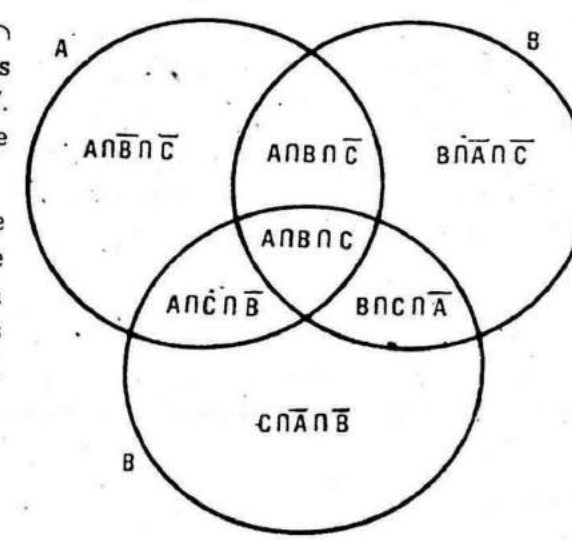
$$P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$$

$$= P(A) + P(B) - P(A \cap B).$$

(ii) Let A, B and C denote three sets of points having points in common as indicated in the Venn diagram below:

The symbol $A \cap \bar{B} \cap \bar{C}$ means the set of points in A but not in B nor in C. The other symbols have similar meanings.

The points which are either in A or B or C are included in the seven mutually exclusive sets (fig.) The required probability is given by



$$P(A \cup B \cup C) = P(A \cap \bar{B} \cap \bar{C}) + P(A \cap B \cap \bar{C}) + P(B \cap \bar{A} \cap \bar{C}) +$$
$$P(B \cap C \cap \bar{A}) + P(C \cap \bar{A} \cap \bar{B}) + P(A \cap C \cap \bar{B}) + P(A \cap B \cap C)$$

In order to obtain $A \cap \bar{B} \cap \bar{C}$, we remove points common to A and B, and to A and C, but in so doing we have removed points common to A, B and C twice.

Thus $A \cap \bar{B} \cap \bar{C} = A - A \cap B - A \cap C + A \cap B \cap C$.

$\therefore P(A \cap \bar{B} \cap \bar{C}) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$.

Similarly, we find that

$$P(A \cap B \cap \bar{C}) = P(A \cap B) - P(A \cap B \cap C)$$
$$P(B \cap \bar{A} \cap \bar{C}) = P(B) - P(B \cap C) - P(B \cap A) + P(B \cap C \cap A).$$
$$P(B \cap C \cap \bar{A}) = P(B \cap C) - P(A \cap B \cap C)$$
$$P(C \cap \bar{A} \cap \bar{B}) = P(C) - P(C \cap A) - P(C \cap B) + P(C \cap A \cap B)$$
$$P(A \cap C \cap \bar{B}) = P(A \cap C) - P(A \cap B \cap C)$$

Substituting these values, considering that $P(A \cap B) = P(B \cap A)$, etc., and simpifying, we get

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C).$$

6.33. (b) The sample space in this experiment is reduced by excluding the sample points containing the same numbers.

Since there are 6 sample points which contain the same numbers, therefore the *reduced* sample space consists of $36-6=30$ sample points.

(i) Let A denote the event that the sum of 6 occurs when the numbers appearing are different. Then

$A = \{(1, 5), (2, 4), (4, 2), (5, 1)\}$, and therefore

$$P(A) = \frac{4}{30} = \frac{2}{15}.$$

(ii) Let B be the event that the sum of 4 or less occurs when the numbers appearing are different. Then

$B = \{(3,1), (1,3), (2,1), (1,2)\}$ and thus

$$P(B) = \frac{4}{30} = \frac{2}{15}.$$

Alternative Method:

The sample space S consists of $6^2 = 36$ sample points.

(i) Let A be the event that the sum of 6 appears and B, the event that two different numbers appear on the dice. Then

$$P(B) = \frac{n(B)}{n(S)} = \frac{30}{36}, \text{ and } P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{4}{36}.$$

Hence $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{4}{36} \div \dfrac{30}{36} = \dfrac{4}{30} = \dfrac{2}{15}$.

(ii) Let C be the event that the sum of 4 or less appears. Then

$$C \cap B = \{(1,3), (3,1), (2,1), (1,2)\}, \text{ and } P(C \cap B) = \frac{4}{36}.$$

Hence $P(C/B) = \dfrac{P(C \cap B)}{P(B)} = \dfrac{4}{36} \div \dfrac{30}{36} = \dfrac{4}{30} = \dfrac{2}{15}$.

6.34 (a) Let A = (first tube is good) and B = (other tube is good). Then S can occur in $\binom{10}{2} = 45$ ways, the number of ways in which 2 tubes can be drawn from 10 tubes.

Now $A \cap B$ can occur in $\binom{6}{2} = 15$ ways, the number of ways in which 2 good tubes can be drawn from 6 good tubes.

$$\therefore \quad P(A \cap B) = \frac{15}{45} = \frac{1}{3}.$$

$$P(A) = \frac{6}{10} = \frac{3}{5}$$

Hence $P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{1/3}{3/5} = \dfrac{5}{9}$.

(b) Let A = (the sum of dots is odd) and B = (the sum of dots is 7). Then we seek $P(B/A)$.

S consists of $6^2 = 36$ sample points and A contains 18 sample points. Therefore

$$P(A) = \frac{18}{36} = \frac{1}{2}.$$

The event $A \cap B$ contains 6 sample points. Thus

$$P(A \cap B) = \frac{6}{36} = \frac{1}{6}.$$

Hence $P(B/A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{1/6}{1/2} = \dfrac{1}{3}.$

(c) Let A represent the event that the employee has accounting background and E be the event that the employee is an executive. Then

$$P(A) = 0.20 \text{ and } P(A \cap E) = 0.05$$

Hence the desired probability is

$$P(E/A) = \frac{P(A \cap E)}{P(A)} = \frac{0.05}{0.20} = 0.25.$$

6.35. (i) Let R_1, R_2 and R_3 denote the events that the first flower is red, the second flower is red and the third flower is red. Then we need $P(R_1 \cap R_2 \cap R_3)$. Since flowers are picked up at random one by one without replacement, therefore

$$P(R_1 \cap R_2 \cap R_3) = P(R_1)\, P(R_2/R_1)\, P(R_3/R_1 \cap R_2)$$
$$= \frac{10}{22} \times \frac{9}{21} \times \frac{8}{20} = \frac{6}{77} = 0.0779$$

(ii) The number of sequences in which the event 2 red and 2 white flowers in the first four picked up can occur, is given by

$$\frac{4!}{2!\,2!} = 6$$

Now $P(\text{2 red and 2 white in one sequence}) = \dfrac{10 \times 9 \times 12 \times 11}{22 \times 21 \times 20 \times 19} = \dfrac{9}{133}$

$\therefore$ Required probability $= \dfrac{9}{133} \times 6 = 0.406$

(iii) $P(\text{3rd is red/first 2 are white}) = \dfrac{P(\text{3rd red} \cap \text{first two white})}{P(\text{first 2 are white})}$

$$= \frac{\frac{12}{22} \times \frac{11}{21} \times \frac{10}{20}}{\frac{12}{22} \times \frac{11}{21}} = \frac{1}{2}.$$

6.36. (c) Given P(A) = 0.60, P(B) = 0.40 and P(A ∩ B) = 0.24. Then

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.24}{0.40} = 0.60;$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.60 + 0.40 - 0.24 = 0.76;$$

$$P(A/\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{0.60 - 0.24}{1 - 0.4} = \frac{0.36}{0.60} = 0.60;$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{0.24}{0.60} = 0.40; \text{ and}$$

$$P(\bar{B}) = 1 - P(B) = 1 - 0.40 = 0.60.$$

The events A and B are *independent* as

$$P(A/B) = 0.60 = P(A);\ P(B/A) = 0.40 = P(B) \text{ and}$$
$$P(A \cap B) = 0.24 = (0.60)\,(0.40) = P(A)\,P(B).$$

6.37. (b) (i) When A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B).$$

i.e. $\quad 0.7 = 0.4 + p$ or $p = 0.3$.

(ii) When A and B are independent, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= P(A) + P(B) - P(A) \cdot P(B)$$
$$= P(A) + P(B)\,[1 - P(A)]$$

or $\quad 0.7 = 0.4 + p(1 - 0.4)$

or $\quad 0.3 = 0.6p$ or $p = 0.5$.

(c) (i) When A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B)$$

i.e. $\quad 0.60 = 0.50 + P(B)$

$\therefore \quad P(B) = 0.10$.

(ii) When A and B are independent, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= P(A) + P(B) - P(A) \cdot P(B)$$

$$= P(A) + P(B)\,[1 - P(A)]$$

or $\quad 0.60 = 0.50 + P(B)\,[1 - 0.50]$

or $\quad 0.10 = 0.5\,P(B)$

or $\quad P(B) = 0.20$

(iii) $P(A/B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A) + P(B) - P(A \cup B)}{P(B)}$

or $\quad 0.40 = \dfrac{0.50 + P(B) - 0.60}{P(B)}$

or $\quad 0.40\,P(B) = P(B) - 0.10$

or $\quad 0.10 = P(B) - 0.40\,P(B)$

or $\quad 0.10 = 0.60\,P(B)$

or $\quad P(B) = 1/6 = 0.17$

6.38 (a) The statement is true.

(b) The statement is false. If A and B are independent, then

$$P(A/B) = P(A).$$

(c) The statement is true.

(d) The statement is false "Independent" does not mean that two events have equal probabilities.

6.39. (b) Given P(A and B will occur simultaneously), i.e. $P(A \cap B) = \frac{1}{8}$ and P(neither of them will occur), *i.e.* $P(\bar{A} \cap \bar{B}) = \frac{3}{8}$.

$\because$ A and B are independent, therefore we have

$$\frac{1}{8} = P(A \cap B) = P(A) \cdot P(B), \text{ and}$$

$$\frac{3}{8} = P(\bar{A} \cap \bar{B}) = [P(\bar{A}) \cdot P(\bar{B})]$$

$$= [1 - P(A)]\ [1 - P(B)]$$

$$= 1 - P(A) - P(B) + P(A)\,P(B)$$

or $\quad P(A) + P(B) = 1 - \dfrac{3}{8} + \dfrac{1}{8} = \dfrac{6}{8}$.

Now, we find two numbers whose sum is 6/8 and whose product is 1/8. Two such numbers are 1/2 and 1/4.

Hence $P(A) = 1/2$ and $P(B) = 1/4$ or $P(A) = \frac{1}{4}$ and $P(B) = \frac{1}{2}$

6.40 S consists of $6^2 = 36$ sample points.

Now $\quad E_1$ = the event that a 6 appears on at least one die, therefore

$E_1 = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,6), (6,5), (6,4), (6,3), (6,2), (6,1)\}$

i.e. $n(E_1) = 11$ sample points.

E_2 = the event that a 5 appears on exactly one die, therefore

$E_2 = \{(1,5),(2,5),(3,5),(4,5),(5,6),(6,5),(5,4),(5,3),(5,2),(5,1)\}$

It thus contains 10 sample points.

E_3 = the event that same number appears on both dice. Then

$E_3 = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$

i.e. $n(E_3) = 6$ sample points.

Now $E_1 \cap E_2 = \{(5,6), (6,5)\}$, $E_2 \cap E_3 = \{\ \}$, and $E_1 \cap E_3 = \{(6,6)\}$

Thus $n(E_1 \cap E_2) = 2$, $n(E_2 \cap E_3) = 0$ and $n(E_1 \cap E_3) = 1$.

We therefore have $P(E_1) = \frac{11}{36}$, $P(E_2) = \frac{10}{36}$, $P(E_3) = \frac{6}{36}$,

$$P(E_1 \cap E_2) = \frac{2}{36},\ P(E_2 \cap E_3) = 0,\ \text{and } P(E_1 \cap E_3) = \frac{1}{36}.$$

(i) As $\left(\frac{11}{36}\right)\left(\frac{10}{36}\right) \neq \frac{2}{36}$, *i.e.* $P(E_1)\,P(E_2) \neq P(E_1 \cap E_2)$,

$\therefore$ E_1 and E_2 are not independent.

(ii) As $\left(\frac{10}{36}\right)\left(\frac{6}{36}\right) \neq 0$, *i.e.* $P(E_2)\,P(E_3) \neq P(E_2 \cap E_3)$

$\therefore$ E_2 and E_3 are not independent.

(iii) As $\left(\frac{11}{36}\right)\left(\frac{6}{36}\right) \neq \frac{1}{36}$, *i.e.* $P(E_1)\,P(E_3) \neq P(E_1 \cap E_3)$

$\therefore$ E_1 and E_3 are not independent.

6.41 (a) $P(A \text{ cannot solve a problem}) = \frac{25}{100} = \frac{1}{4}$

$$P(B \text{ cannot solve a problem}) = \frac{30}{100} = \frac{3}{10}$$

$$\therefore\ P(\text{both } A \text{ and } B \text{ cannot solve a problem}) = \frac{1}{4} \times \frac{3}{10} = \frac{3}{40}$$

Hence $P(\text{either } A \text{ or } B \text{ can solve a problem}) = 1 - \frac{3}{40} = \frac{37}{40}$.

(b) Let A_1 be the event that the first card is a red ace, A_2 be the event that the second card is a ten or jack, and A_3 be the event that the third card is greater than 3 but less than 7. Then, as the cards are drawn in succession, without replacement, we have

$$P(A_1) = \frac{2}{52},\ P(A_2/A_1) = \frac{8}{51},\ P(A_3/A_1 \cap A_2) = \frac{12}{50}$$

$$\therefore\ P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2/A_1)\,P(A_3/A_1 \cap A_2)$$

$$= \frac{2}{52} \times \frac{8}{51} \times \frac{12}{50} = \frac{8}{5525} = 0.0014$$

6.42 Here $P(A) = \frac{5}{7}$ and $P(B) = \frac{7}{9}$.

(i) $P(\text{both of them will die}) = P(\bar{A} \cap \bar{B})$

$$= P(\bar{A}).P(\bar{B}) = \left(1 - \frac{5}{7}\right)\left(1 - \frac{7}{9}\right)$$

$$= \frac{2}{7} \times \frac{2}{9} = \frac{4}{63}.$$

(ii) $P(A \text{ will be alive and } B \text{ dead}) = P(A \cap \bar{B}) = P(A)\,.\,P(\bar{B})$

$$= \frac{5}{7}\left(1 - \frac{7}{9}\right) = \frac{10}{63}$$

(iii) $P(B \text{ will be alive and } A \text{ dead}) = P(\bar{A} \cap B) = P(\bar{A})\,.\,P(B)$

$$= \left(1 - \frac{5}{7}\right) \times \frac{7}{9} = \frac{2}{9}$$

(iv) $P(\text{both of them will be alive}) = P(A \cap B) = P(A)\,.\,P(B)$

$$= \frac{5}{7} \times \frac{7}{9} = \frac{5}{9}.$$

6.43 (a) Given

Bag	Red	Black	Total
1	3	5	8
2	4	7	11
Total	7	12	19

Now the probability of selecting the first bag is $\frac{1}{2}$, and if the first bag is selected, the probability that the ball drawn is red, is $\frac{3}{8}$.

Hence the probability that a red ball is drawn from the first bag is $\frac{1}{2} \times \frac{3}{8} = \frac{3}{16}$.

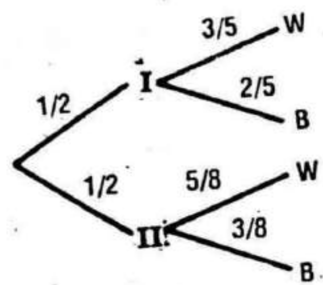
Similarly, the probability of drawing a red ball from the second bag is $\frac{1}{2} \times \frac{4}{11} = \frac{2}{11}$.

Hence the required probability $= \frac{3}{16} + \frac{2}{11}$ (the events are mutually exclusive)

$$= \frac{65}{176}.$$

(b) First we select one of the two urns and then we draw a ball which is either white (W) or black (B).

Let us construct the following tree diagram with respective probabilities to describe this process.



Now the probability of a white ball if urn I is selected is $\frac{1}{2} \times \frac{3}{5}$.

Similarly, the probability of a white ball from urn II, is $\frac{1}{2} \times \frac{5}{8}$

Hence the required probability $= \frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{5}{8}$

$$= \frac{3}{10} + \frac{5}{16} = \frac{49}{80}.$$

6.44. Let A = (the first drawing gives 3 white balls), B = (the second drawing gives 3 black balls) and S = Sample space. We seek P(A $\cap$ B).

S can occur in $\binom{13}{3} = 286$ ways, the number of ways in which 3 balls can be drawn from 13 balls.

A can occur in $\binom{5}{3} = 10$ ways, the number of ways in which 3 white balls can be drawn from 5 white balls.

B can occur in $\binom{8}{3} = 56$ ways.

$$\therefore \quad P(A) = \frac{10}{286}$$

To find the probability of B, when the white balls drawn in the first drawing are not being replaced, *i.e.* $P(B/A)$, the sample space S is reduced. Thus

S can occur in $\binom{10}{3} = 120$ ways.

$$\therefore \quad P(B/A) = \frac{56}{120}.$$

Hence $P(A \cap B) = P(A) \,.\, P(B/A) = \frac{10}{286} \times \frac{56}{120} = \frac{7}{429}.$

6.45 (b) Let S_1 denote the event that the spotted egg is chosen in the first draw, S_2, the event that the spotted egg is chosen in the second draw, and C_3, the event that a clear egg is chosen in the third draw.

Then we need $P(S_1 \cap S_2 \cap C_3)$ which by the multiplication law, may be written as

$$P(S_1 \cap S_2 \cap C_3) = P(S_1)\, P(S_2/S_1)\, P(C_3/S_1 \cap S_2).$$

Now $\quad P(S_1) = \frac{5}{30},$

$P(S_2/S_1) = \frac{4}{29}$, as after the occurrence of S_1, ther are 29 eggs left out of which 4 are spotted; and

$P(C_3/S_1 \cap S_2) = \frac{25}{28}$, as after the second draw, there are 28 eggs left out of which 25 are clear eggs.

Hence $P(S_1 \cap S_2 \cap C_3) = \frac{5}{30} \times \frac{4}{29} \times \frac{25}{28} = \frac{25}{1218}.$

6.46. Let A denote the event that 1 girl and 2 boys are selected. Then A can occur in any of the following three mutually exclusive ways.

A_1 = a girl from first group, a boy from second and a boy from third group.

A_2 = a boy from first group, a girl from second and a boy from 3rd group.

A_3 = a boy from 1st group, a boy from 2nd group and a girl from 3rd group.

Thus $P(A_1) = \frac{3}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{9}{32}$

$P(A_2) = \frac{1}{4} \times \frac{2}{4} \times \frac{3}{4} = \frac{3}{32}$

$P(A_3) = \frac{1}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{1}{32}$.

Hence $P(A) = P(A_1) + P(A_2) + P(A_3)$, ($A_1$, A_2, A_3 are mutually exclusive).

$$= \frac{9}{32} + \frac{3}{32} + \frac{1}{32} = \frac{13}{32}.$$

6.47. Let A = (only girls turn up for the party) and B = (2 girls and 1 boy turn up for the party). Then

(i) $P(A)$ = the probability of turning up of a girl from 1st family, a girl from 2nd family and a girl from 3rd family.

$$= \frac{2}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{32}.$$

(ii) B can occur in any of the following 3 mutually exclusive ways, where order corresponds to first, second and third family.

B_1 = girl – girl – boy;

B_2 = girl – boy – girl;

B_3 = boy – girl – girl.

Thus $P(B_1) = \frac{2}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{32}$,

$P(B_2) = \frac{2}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{9}{32}$,

$P(B_3) = \frac{2}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{32}$.

Hence $P(B) = P(B_1) + P(B_2) + P(B_3)$

$$= \frac{1}{32} + \frac{9}{32} + \frac{3}{32} = \frac{13}{32}.$$

6.48. Let A = (the book is favourably reviewed by the first reviewer), B = (the book is favourably reviewed by the second reviewer) and C = (the book is favourably reviewed by the third reviewer). Then

$$P(A) = \frac{3}{5},\ P(B) = \frac{4}{7} \text{ and } P(C) = \frac{2}{5}.$$

Since the events are independent, so $P(\bar{A}) = \frac{2}{5}$, $P(\bar{B}) = \frac{3}{7}$, and $P(\bar{C}) = \frac{3}{5}$.

A majority out of 3 reviewers will favour if two or three review the book favourably.

Let E be the event that only two favour. Then

$$E = (A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C).$$

In other words, if only two favour, then they would be either only the first and second, $A \cap B \cap \bar{C}$, or only the first and third, $A \cap \bar{B} \cap C$, or only the second and third, $\bar{A} \cap B \cap C$. Since these events are mutually exclusive, therefore

$$P(E) = P(A \cap B \cap \bar{C}) + P(A \cap \bar{B} \cap C) + P(\bar{A} \cap B \cap C).$$

$$= P(A).P(B).P(\bar{C}) + P(A).P(\bar{B})\,P(C) + P(\bar{A})\,P(B)\,P(C)$$

$$= \frac{3}{5}\cdot\frac{4}{7}\cdot\frac{3}{5} + \frac{3}{5}\cdot\frac{3}{7}\cdot\frac{2}{5} + \frac{2}{5}\cdot\frac{4}{7}\cdot\frac{2}{5}$$

$$= \frac{36}{175} + \frac{18}{175} + \frac{16}{175} = \frac{70}{175} = \frac{2}{5}.$$

In case all three also favour, we have the required probability as

$$\frac{2}{5} + P(A \cap B \cap C) = \frac{2}{5} + \frac{3}{5}\cdot\frac{4}{7}\cdot\frac{2}{5} = \frac{94}{174}.$$

6.49. (a) Here $P(A) = \frac{4}{5}$, P(B) = $\frac{3}{4}$, and P(C) = $\frac{2}{3}$

At least two shots mean two or more. Thus

$$P(A \text{ and } B \text{ hit but not } C) = \frac{4}{5} \times \frac{3}{4} \times \left(1 - \frac{2}{3}\right) = \frac{1}{5}$$

$$P(A \text{ and } C \text{ hit but not } B) = \frac{4}{5}\left(1 - \frac{3}{4}\right) \cdot \frac{2}{3} = \frac{2}{15}$$

$$P(B \text{ and } C \text{ hit but not } A) = \left(1 - \frac{4}{5}\right) \cdot \frac{3}{4} \times \frac{2}{3} = \frac{1}{10}$$

$$P(A, B \text{ and } C \text{ hit}) = \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} = \frac{2}{5}$$

Since these ways are mutually exclusive, therefore the required probability $= \frac{1}{5} + \frac{2}{15} + \frac{1}{10} + \frac{2}{5}$

$$= \frac{6 + 4 + 3 + 12}{30} = \frac{25}{30} = \frac{5}{6}.$$

(b) Given the probabilities of a wrong decision by each member as

$P(A) = 0.05$, $P(B) = 0.05$, and $P(C) = 0.10$,

Let each member vote independently, then

$P(\bar{A}) = 0.95$, $P(\bar{B}) = 0.95$ and $P(\bar{C}) = 0.90$.

Let E be the event that a wrong decision on the basis of a majority vote is made by the committee. Then

$$E = (A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C) \cup (A \cap B \cap C).$$

$$\therefore \quad P(E) = P(A \cap B \cap \bar{C}) + P(A \cap \bar{B} \cap C) + P(\bar{A} \cap B \cap C) + P(A \cap B \cap C)$$

$$= P(A) \,.\, P(B) \,.\, P(\bar{C}) + P(A)P(\bar{B})P(C) + P(\bar{A})P(B)P(C) + P(A)\, P(B)\, P(C)$$

$$= (0.05)\,(0.05)\,(0.90) + (0.05)\,(0.95)\,(0.10) + (0.95)\,(0.05)\,(0.10) + (0.05)\,(0.05)\,(0.10)$$

$$= 0.00225 + 0.00475 + 0.00475 + 0.00025 = 0.012$$

This indicates that the committee will be wrong in 1.2% of its decisions.

6.50. Let A = (first man hits the target), B = (second man hits the target), and C = (third man hits the target).

Then $\quad P(A) = \frac{1}{6}$, $P(B) = \frac{1}{4}$ and $P(C) = \frac{1}{3}$.

Also $\quad P(\bar{A}) = \frac{5}{6}$, $P(\bar{B}) = \frac{3}{4}$ and $P(\bar{C}) = \frac{2}{3}$ as the events are independent.

(i) Let E be the event that exactly one man hits the target.

Then $\quad E = (A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)$,

where $A \cap \bar{B} \cap \bar{C}$ implies that only the first man hits, $\bar{A} \cap B \cap \bar{C}$ means only the second man hits, etc. Since the three events are mutually exclusive, therefore

$$P(E) = P(A \cap \bar{B} \cap \bar{C}) + P(\bar{A} \cap B \cap \bar{C}) + P(\bar{A} \cap \bar{B} \cap C)$$

$$= P(A).P(\bar{B}).P(\bar{C}) + P(\bar{A}).P(B).P(\bar{C}) + P(\bar{A}).P(\bar{B}).P(C)$$

$$= \frac{1}{6} \cdot \frac{3}{4} \cdot \frac{2}{3} + \frac{5}{6} \cdot \frac{1}{4} \cdot \frac{2}{3} + \frac{5}{6} \cdot \frac{3}{4} \cdot \frac{1}{3}$$

$$= \frac{6}{72} + \frac{10}{72} + \frac{15}{72} = \frac{31}{72}.$$

(ii) Here we require the probability that the first man hit the target given that only one man hit the target, *i.e.* $P(A/E)$.

Now $\quad A \cap E = A \cap \bar{B} \cap \bar{C}$ is the event that fonly the first man hit the target. Therefore

$$P(A \cap E) = P(A \cap \bar{B} \cap \bar{C}) = P(A) \,.\, P(\bar{B}) \,.\, P(\bar{C})$$

$$= \frac{1}{6} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{12}.$$

Hence $P(A/E) = \frac{P(A \cap E)}{P(E)} = \frac{1/12}{31/72} = \frac{6}{31}$

6.51. Let A_1, A_2, A_3 represent the events that the first, the second and the third missile hit the target respectively. Then

$P(A_1) = 0.40$, $P(A_2) = 0.50$, $P(A_3) = 0.60$ and

$P(\bar{A}_1) = 0.60$, $P(\bar{A}_2) = 0.50$, $P(\bar{A}_3) = 0.40$

(i) P(all the missiles hit the target) $= P(A_1 \cap A_2 \cap A_3)$
$= P(A_1)\, P(A_2)\, P(A_3)$
$= (0.4)\,(0.5)\,(0.6) = 0.12$

(ii) P(*at least* one of the three hits the target)
$= 1 - P$ (no missile hits the target)
$= 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3)$
$= 1 - P(\bar{A}_1)\, P(\bar{A}_2)\, P(\bar{A}_3)$
$= 1 - (0.6)\,(0.5)\,(0.4) = 0.88.$

(iii) P(exactly one hits the target)
$= P(A_1 \cap \bar{A}_2 \cap \bar{A}_3) + P(\bar{A}_1 \cap A_2 \cap \bar{A}_3) + P(\bar{A}_1 \cap \bar{A}_2 \cap A_3)$
$= P(A_1)P(\bar{A}_2)P(\bar{A}_3) + P(\bar{A}_1)P(A_2) \times P(\bar{A}_3) + P(\bar{A}_1)\, P(\bar{A}_2)\, P(A_3)$
$= (0.40)(0.5)(0.4) + (0.6)(0.5)(0.4) + (0.6)(0.5)(0.6)$
$= 0.08 + 0.12 + 0.18 = 0.38.$

(iv) P(exactly two hit the target)
$= P(A_1 \cap A_2 \cap \bar{A}_3) + P(A_1 \cap \bar{A}_2 \cap A_3) + P(\bar{A}_1 \cap A_2 \cap A_3)$
$= P(A_1)P(A_2)P(\bar{A}_3) + P(A_1)P(\bar{A}_2)P(A_3) + P(\bar{A}_1)P(A_2)P(A_3)$
$= (0.4)\,(0.5)\,(0.4) + (0.4)\,(0.5)\,(0.6) + (0.6)\,(0.5)\,(0.6)$
$= 0.08 + 0.12 + 0.18 = 0.38.$

6.52. Given that $P(A \text{ wins any one game}) = \frac{6}{12} = \frac{1}{2}$,

$P(B \text{ wins any one game}) = \frac{4}{12} = \frac{1}{3}$, and

$P(\text{any one game ends in a tie}) = \frac{2}{12} = \frac{1}{6}$.

Let A_1, A_2, A_3 represent the events "A wins" in first, second and third game respectively;
B_1, B_2, B_3 represent the events "B wins" in first, second and third game respectively; and
T_1, T_2, T_3 represent the events "a tie occurs" in first, second and third game respectively. Then

(a) $P(A \text{ wins all 3 games}) = P(A_1 \cap A_2 \cap A_3)$

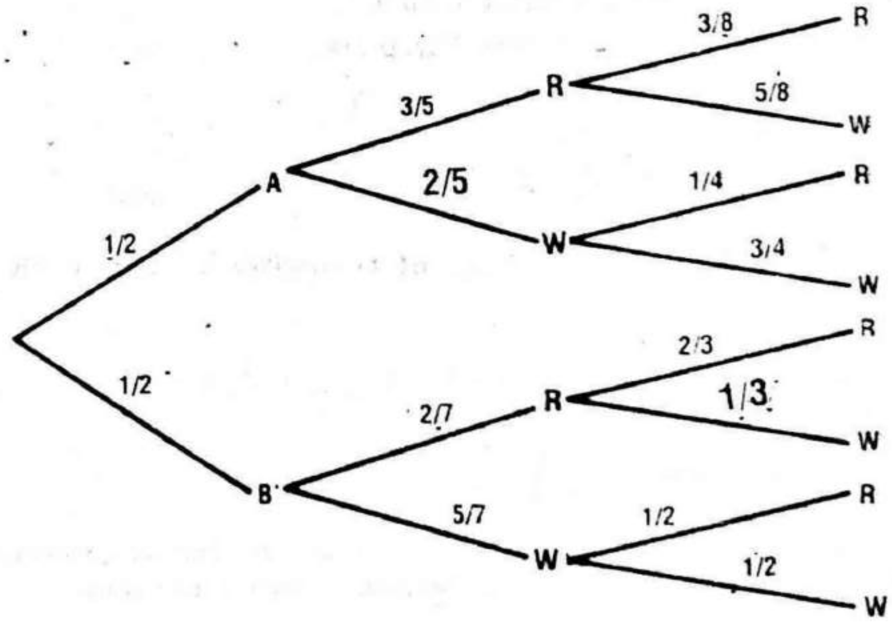
$= P(A_1).P(A_2).P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

(b) P(2 games end in a tie) $= P(T_1 \cap T_2 \cap \bar{T}_3) + P(T_1 \cap \bar{T}_2 \cap T_3) + P(\bar{T}_1 \cup T_2 \cap T_3)$; (2 in tie and third not)
$= P(T_1).P(T_2).P(\bar{T}_3) + P(T_1) + P(\bar{T}_2).P(T_3) + P(\bar{T}_1).P(T_2).P(T_3)$
$= \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} + \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{1}{6}$
$= \frac{5}{216} + \frac{5}{216} + \frac{5}{216} = \frac{15}{216} = \frac{5}{72}.$

(c) P (A & B wins alternately) $= P(A_1 \cap B_2 \cap A_3) + P(B_1 \cap A_2 \cap B_3)$
$= P(A_1) . P(B_2) . P(A_3) + P(B_1) . P(A_2) . P(B_3)$
$= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{12} + \frac{1}{18} = \frac{5}{36}.$

(d) $P(B \text{ wins at least one game}) = 1 - P(B \text{ wins no game})$
$= 1 - P(\bar{B}_1 \cap \bar{B}_2 \cap \bar{B}_3)$
$= 1 - P(\bar{B}_1) . P(\bar{B}_2) . P(\bar{B}_3)$
$= 1 - \left(\frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3}\right) = 1 - \frac{8}{27} = \frac{19}{27}$

6.53. The tree diagram of the process is constructed as below:



Suppose urn A is selected and a red ball (R) is drawn, and put into urn B, then urn B will contain 3 red balls (R) and 5 white balls (W).

Since there are four paths leading to two balls of the same colour, we get the required probability as

$$p = \left(\frac{1}{2}\cdot\frac{3}{5}\cdot\frac{3}{8}\right) + \left(\frac{1}{2}\cdot\frac{2}{5}\cdot\frac{3}{4}\right) + \left(\frac{1}{2}\cdot\frac{2}{7}\cdot\frac{2}{3}\right) + \left(\frac{1}{2}\cdot\frac{5}{7}\cdot\frac{1}{2}\right)$$

$$= \frac{9}{80} + \frac{3}{20} + \frac{2}{21} + \frac{5}{28} = \frac{901}{1680}$$

6.54. There are 6 persons (5 friends and 1 host) and not two persons will have the same birth day.

Now, each person can have any one of the 30 days (days of the month of April) as his birth day and assuming each day of the month is equally likely, we have $(30)^6$ ways in which the 6 persons can have their birth days.

As the 6 persons will have distinct birth days, the first person can have any one of the 30 days as his birth day, the second can have any one of the remaining 29 days as his birth day, the third person can have any one of the remaining 28 days as his birth day, and so on.

Thus there are $30 \times 29 \times 28 \times 27 \times 26 \times 25$ ways in which the 6 persons can have distinct birth day.

Hence the required probability, p, is

$$p = \frac{30 \times 29 \times 28 \times 27 \times 26 \times 25}{(30)^6}$$

$$= \frac{30}{30} \times \frac{29}{30} \times \frac{28}{30} \times \frac{27}{30} \times \frac{26}{30} \times \frac{25}{30} = \frac{2639}{4500} = 0.586$$

6.55. (a) The probability of throwing a head with a coin $= \frac{1}{2}$.

A can win in the first, third, fifth, ..., toss, while B can win in the second, fourth, sixth, ..., toss.

Thus $\quad P(\text{A's winning}) = \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^4 + \ldots$

($\because$ A's win in 3rd toss is associated with failure of both A and B once).

$$= \frac{\frac{1}{2}}{1-\left(\frac{1}{2}\right)^2} = \frac{2}{3}$$

$$P(\text{B's winning}) = \frac{1}{2} \times \frac{1}{2} + \left(\frac{1}{2}\right)^3 \times \frac{1}{2} + \left(\frac{1}{2}\right)^5 \times \frac{1}{2} + \ldots$$

$$= \frac{\frac{1}{2} \times \frac{1}{2}}{1-\left(\frac{1}{2}\right)^2} = \frac{1}{3}$$

Hence their respective chances of winning are $\frac{2}{3} : \frac{1}{3}$ or $2 : 1$.

(b) Let A, B and C represent the three persons.

The probability of throwing a head with a coin $= \frac{1}{2}$.

Now A can win in the first, fourth, seventh, ..., throws

$\therefore$ Chance of A's winning $= \frac{1}{2} + \left(\frac{1}{2}\right)^3 \times \frac{1}{2} + \left(\frac{1}{2}\right)^6 \times \frac{1}{2} + \ldots$

$$= \frac{\frac{1}{2}}{1-\left(\frac{1}{2}\right)^3} = \frac{4}{7}$$

B can win in the second, fifth, eighth, ..., throws

$\therefore$ Chance of B's winning $= \frac{1}{2} \times \frac{1}{2} + \left(\frac{1}{2}\right)^4 \times \frac{1}{2} + \left(\frac{1}{2}\right)^7 \times \frac{1}{2} + \ldots$

$$= \frac{\frac{1}{4}}{1-\left(\frac{1}{2}\right)^3} = \frac{2}{7}$$

C can win in the third, sixth, ninth, ..., throws

$\therefore$ Chance of C's winning $= \left(\frac{1}{2}\right)^2 \times \frac{1}{2} + \left(\frac{1}{2}\right)^5 \times \frac{1}{2} + \left(\frac{1}{2}\right)^8 \times \frac{1}{2} + \ldots$

$$= \frac{\left(\frac{1}{2}\right)^3}{1-\left(\frac{1}{2}\right)^3} = \frac{1}{7}$$

Hence their respective chances of winning are $\frac{4}{7}, \frac{2}{7}$ and $\frac{1}{7}$.

6.56. (a) S contains $2^6 = 64$ sample points.

Let A be the event that exactly 4 heads are obtained. Then the sample points corresponding to A are

$$\binom{6}{4} = 15$$

$$\therefore \quad P(A) = \frac{15}{64}.$$

(b) Here n = 16, $p = \frac{1}{2}$ and $q = \frac{1}{2}$

(i) $P(\text{exactly 8 heads are obtained}) = \binom{16}{8}\left(\frac{1}{2}\right)^{16}$

$$= \frac{16!}{8!\,(16-8)!}\cdot\frac{1}{65536}$$

$$= \frac{6435}{32768} = 0.196.$$

(ii) $P(\text{exactly 11 heads are obtained}) = \binom{16}{11}\left(\frac{1}{2}\right)^{16}$

$$= \frac{16!}{11!\,(16-11)!}\cdot\frac{1}{65536}$$

$$= \frac{273}{4096} = 0.067.$$

6.57. Let p denote the probability of passing an examination. Then $p = 0.4$ and $q = 1 - p = 0.6$. Here $n = 6$.

(i) $P(\text{2 candidates will pass}) = \binom{6}{2}(0.4)^2(0.6)^{6-2}$

$$= 15(0.4)^2(0.6)^4 = 0.31104 = 0.31.$$

(ii) $P(\text{5 candidates will pass}) = \binom{6}{5}(0.4)^5(0.6)^{6-5}$

$$= 6(0.4)^5(0.6) = 0.036864 = 0.04.$$

Now $P(\text{all candidates pass}) = p^6$, and

$P(\text{all candidates fail}) = q^6$.

But $p = 0.4$ and $q = 0.6$, therefore the probabilities are not equal.

The probability of all passing would be equal to the probability of all failing when $p = q = 0.5$.

6.58. (b) The probability of selecting an urn is $\frac{1}{3}$, i.e.

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

Let E be the event that the ball drawn is white. Then

$$P(E/A) = \frac{1}{3},\quad P(E/B) = \frac{2}{3} \text{ and } P(E/C) = \frac{2}{4}.$$

By Bayes' theorem, we have

$$P(C/E) = \frac{P(C)\,P(E/C)}{P(A)\,P(E/A) + P(B)\,P(E/B) + P(C)\,P(E/C)}$$

$$= \frac{\frac{1}{3}\cdot\frac{2}{4}}{\frac{1}{3}\cdot\frac{1}{3}+\frac{1}{3}\cdot\frac{2}{3}+\frac{1}{3}\cdot\frac{2}{4}} = \frac{\frac{2}{12}}{\frac{1}{9}+\frac{2}{9}+\frac{1}{6}} = \frac{\frac{2}{12}}{\frac{1}{2}} = \frac{1}{3}.$$

6.59. Let A_1 denote the event that the ideal coin is selected, A_2 denote the event that the 2nd coin is selected and A_3 denote the event that the 3rd coin is selected.

Then $\quad P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$

Again, let E denote the event that head appears both the times when the coin is tossed twice. Then

$P(E/A_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, (Probability of two heads, given that the ideal coin is tossed).

$P(E/A_2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$, and

$P(E/A_3) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$.

Hence by Bayes' theorem, the desired probaility is

$$P(A_1/E) = \frac{P(A_1)\, P(E/A_1)}{\sum P(A_i)\, P(E/A_i)}, \quad i = 1, 2, 3.$$

$$= \frac{\frac{1}{3} \cdot \frac{1}{4}}{\frac{1}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{4}{9}} = \frac{1/12}{29/108} = \frac{1}{12} \times \frac{108}{29} = \frac{9}{29}.$$

6.60. Let A denote the event that the student is taller than 6 feet, *W*, the event that the student chosen is a woman, and M, the event that the student is a man. Then a need P(W/A).

By Bayes' theorem, we get the desired probability as

$$P(W/A) = \frac{P(W)\, P(A/W)}{P(W)\, P(A/W) + P(M)\, P(A/M)}$$

$$= \frac{(0.60)\,(0.01)}{(0.60)\,(0.01) + (0.40)\,(0.04)} = \frac{0.006}{0.022} = \frac{3}{11}.$$

6.61. Let A, B **and** C denote the events that the cake is baked by cook A, cook B and cook C respectively. Then

$P(A) = 0.50$, $P(B) = 0.30$ and $P(C) = 0.20$.

Again, let E denote the event that the cake fails to rise. Then

$P(E/A) = 0.02$, $P(E/B) = 0.03$ and $P(E/C) = 0.05$.

We need $P(A/E)$.

Using Bayes' theorem, we get

$$P(A/E) = \frac{P(A)\, P(E/A)}{P(A)\, P(E/A) + P(B)\, P(E/B) + P(C)\, P(E/C)}$$

$$= \frac{(0.50)\,(0.02)}{(0.50)\,(0.02) + (0.30)\,(0.03) + (0.20)\,(0.05)}$$

$$= \frac{0.010}{0.029} = \frac{10}{29}.$$

Hence the desired proportion of "failures" caused by A

$$= \left(\frac{10}{29}\right) \times 100 = \frac{1000}{29}\,\%$$

6.62. Let H, M **and** L denote the events that the box chosen contains high quality, medium quality and low quality bulbs respectively. Then

$P(H) = 0.2$, $P(M) = 0.4$ and $P(L) = 0.4$.

Again, let S denote the event that the two bulbs tested are satisfactory. Then

$P(S/H) = 1.0$, $P(S/M) = (1.0 - 0.1)^2 = 0.81$, and

$P(S/L) = (1.0 - 0.2)^2 = 0.64$.

We need (i) $P(H/S)$, (ii) $P(M/S)$ and (iii) $P(L/S)$.

Using Bayes' theorem, we get

(i) $$P(H/S) = \frac{P(H)\, P(S/H)}{P(H)\, P(S/H) + P(M)\, P(S/M) + P(L)\, P(S/L)}$$

$$= \frac{(0.2)\,(1.0)}{(0.2 \times 1.0) + (0.4 \times 0.81) + (0.4 \times 0.64)}$$

$$= \frac{0.2}{0.2 + 0.324 + 0.256} = \frac{0.2}{0.78} = 0.256.$$

Similarly,

(ii) $P(M/S) = \frac{0.324}{0.78} = 0.415$, and

(iii) $P(L/S) = \frac{0.256}{0.78} = 0.328$.

6.63. The probabilities of three diseases A_1, A_2, A_3, **under the given conditions are**

$$P(A_1) = \frac{1}{2},\ P(A_2) = \frac{1}{6} \text{ and } P(A_3) = \frac{1}{3}.$$

Let E denote the test carried out to help the diagnosis. Then

$P(E/A_1) = \binom{5}{4}(0.1)^4(0.9)$ [Given disease A_1, the probability of 4 positive results out of 5 test]
$= 0.00045.$

Similarly,

$$P(E/A_2) = \binom{5}{4}(0.2)^4(0.9) = 0.00640, \text{ and}$$

$$P(E/A_3) = \binom{5}{4}(0.9)^4(0.1) = 0.32805$$

Hence by Bayes' theorem, the desired probabilities are

$$P(A_1/E) = \frac{P(A_1)\,P(E/A_1)}{\sum P(A_i)\,P(E/A_i)}, \text{ where } i = 1, 2, 3.$$

$$= \frac{\frac{1}{2} \times (0.00045)}{\frac{1}{2} \times (0.00045) + \frac{1}{6} \times (0.00640) + \frac{1}{3} \times (0.32805)}$$

$$= \frac{0.000225}{(0.000225) + (0.001067) + (0.109350)} = \frac{0.000225}{0.110642}$$

$$= 0.0020;$$

$P(A_2/E) = 0.0096$; and $P(A_3/E) = 0.9883$.

These results can be summarized as below:

Event	Prior Probabilities $P(A_i)$	Conditional Probabilities $P(E/A_i)$	Joint Probabilities $P(A_i \cap E)$	Posterior Probabilities $P(A_i/E)$
Disease A_1	1/2	0.00045	0.000225	0.0020
Disease A_2	1/6	0.00640	0.001067	0.0096
Disease A_3	1/3	0.32805	0.109350	0.9883
	1		0.110642	0.9999

The posterior probability $P(A_i/E)$ is, by Bayes' theorem,

$$P(A_i/E) = \frac{P(A_i \cap E)}{\sum P(A_i \cap E)}, \text{ where } P(A_i \cap E) = P(A_i)\,P(E/A_i).$$

●●●●●●●●●●

Chapter 7

RANDOM VARIABLES

7.1. (b) The sample space for this experiment is

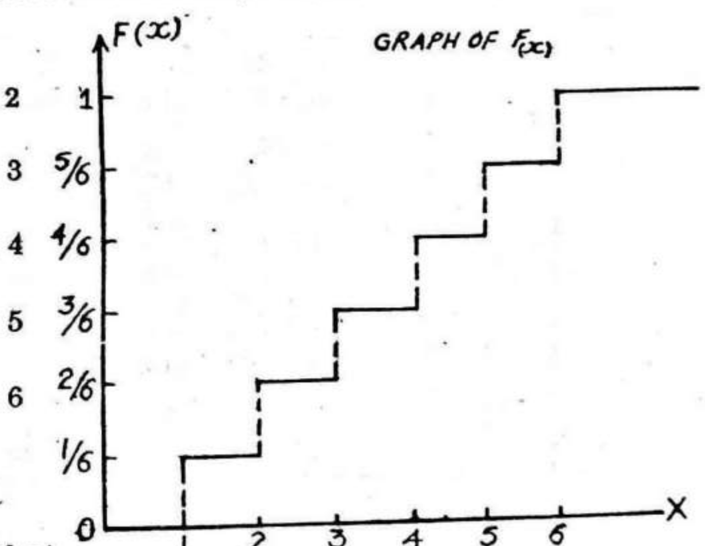
$$\mathbf{S = \{1, 2, 3, 4, 5, 6\}}$$

Let X be the r.v. that denote the number of points appearing. Then the values of x are 1, 2, 3, 4, 5 and 6. As all six faces are equally likely to appear, therefore

$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

Hence the desired distribution function, $F(x)$, is

$$F(x) = \begin{cases} 0, \text{ for } x < 1 \\ \frac{1}{6}, \text{ for } 1 \le x < 2 \\ \frac{2}{6}, \text{ for } 2 \le x < 3 \\ \frac{3}{6}, \text{ for } 3 \le x < 4 \\ \frac{4}{6}, \text{ for } 4 \le x < 5 \\ \frac{5}{6}, \text{ for } 5 \le x < 6 \\ 1, \text{ for } x \ge 1 \end{cases}$$



7.2 (b). Given that

x	-1	0	1	Total
$f(x)$	$3c$	$3c$	$6c$	$12c$

(i) Since the sum of probabilities in a *p.d.* should be equal to 1, therefore $12c = 1$ or $c = \frac{1}{12}$.

(ii) The probability distribution of $Y = 2X + 1$ is found as below:

x_i	-1	0	1
$f(x_i)$	$3c = \frac{3}{12}$	$\frac{3}{12}$	$\frac{6}{12}$
$y_i = 2x_i + 1$	-1	1	3

Hence the desired *p.d.* of $Y = 2X + 1$ is

y_i	-1	1	3
$f(y_i)$	3/12	3/12	6/12

(c) Let X be the random variable that denotes the number of aces in a bridge hand. Then the values that X can take, are 0, 1, 2, 3, and 4. The desired *p.d.* is given below:

x_i	$f(x_i)$
0	$\binom{4}{0}\binom{48}{13} \div \binom{52}{13} = \frac{6327}{20825} = 0.3038$
1	$\binom{4}{1}\binom{48}{12} \div \binom{52}{13} = \frac{9139}{20825} = 0.4389$
2	$\binom{4}{2}\binom{48}{11} \div \binom{52}{13} = \frac{4446}{20825} = 0.2135$
3	$\binom{4}{3}\binom{48}{10} \div \binom{52}{13} = \frac{858}{20825} = 0.0412$
4	$\binom{4}{4}\binom{48}{9} \div \binom{52}{13} = \frac{55}{20825} = 0.0026$
Total	1

7.3. (b) S contains $\binom{10}{4} = 210$ sample points.

Let X be the r.v. that denotes the number of red balls. Then the values of X are 0, 1, 2, 3 and 4, and their probabilities are

$$f(0) = P(X=0) = \frac{\binom{4}{0}\binom{6}{4}}{\binom{10}{4}} = \frac{15}{210}$$

$$f(1) = P(X=1) = \frac{\binom{4}{1}\binom{6}{3}}{\binom{10}{4}} = \frac{80}{210}$$

$$f(2) = P(X=2) = \frac{\binom{4}{2}\binom{6}{2}}{\binom{10}{4}} = \frac{90}{210}$$

$$f(3) = P(X=3) = \frac{\binom{4}{3}\binom{6}{1}}{\binom{10}{4}} = \frac{24}{210}$$

$$f(4) = P(X=4) = \frac{\binom{4}{4}\binom{6}{0}}{\binom{10}{4}} = \frac{1}{210}$$

Putting this information in a tabular form, we obtain the desired probability distribution of X as

No. of red balls: x_i	0	1	2	3	4
Probability: $f(x_i)$	$\frac{15}{210}$	$\frac{80}{210}$	$\frac{90}{210}$	$\frac{24}{210}$	$\frac{1}{210}$

7.4. (a) S consists of $\binom{15}{3} = 455$ sample points.

Let X be the r.v. that denotes the number of defective clock radios in the sample. Then the values of x are 0, 1, 2, 3 and their probabilities are

$$f(0) = P(X=0) = \binom{5}{0}\binom{10}{3} \div \binom{15}{3} = \frac{120}{455}$$

$$f(1) = P(X=1) = \binom{5}{1}\binom{10}{2} \div \binom{15}{3} = \frac{225}{455}$$

$$f(2) = P(X=2) = \binom{5}{2}\binom{10}{1} \div \binom{15}{3} = \frac{100}{455}$$

$$f(3) = P(X=3) = \binom{5}{3}\binom{10}{0} \div \binom{15}{3} = \frac{10}{455}$$

Hence the desired probability distribution of X is

No. of defectives: x_i	0	1	2	3	Total
Probability: $f(x_i)$	$\frac{120}{455}$	$\frac{225}{455}$	$\frac{100}{455}$	$\frac{10}{455}$	1

(b) S consists of $\binom{8}{3}$ = 56 sample points.

Let X be the r.v. that denotes the number of white balls, drawn from the bag. Then the values of x are 0, 1, 2, 3 and their probabilities are:

$$f(0) = P(X=0) = \binom{5}{0}\binom{3}{3} \div \binom{8}{3} = \frac{1}{56}$$

$$f(1) = P(X=1) = \binom{5}{1}\binom{3}{2} \div \binom{8}{3} = \frac{15}{56}$$

$$f(2) = P(X=2) = \binom{5}{2}\binom{3}{1} \div \binom{8}{3} = \frac{30}{56}$$

$$f(3) = P(X=3) = \binom{5}{3}\binom{3}{0} \div \binom{8}{3} = \frac{10}{56}$$

Putting this information in a tabular form, we obtain the desired probability distribution of X as

No. of white balls: x_i	0	1	2	3	Total
Probability: $f(x_i)$	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$	1

7.5. (b) The function f(x) will be a density function, if

(i) $f(x) \geq 0$, and (ii) $\int\limits_{-\infty}^{\infty} f(x)\, dx = 1$

The first condition is satisfied for every x in the given range, and the second coindition will be satisfied if

$$A \int\limits_{0}^{2} (4x - 2x^2)\, dx = 1.$$

$$i.e. \text{ if } \quad A \left\{ \left[\frac{4x^2}{2} \right]_0^2 - 2 \left[\frac{x^3}{3} \right]_0^2 \right\} = 1,$$

$$i.e. \text{ if } \quad A \left\{ 8 - 2\left(\frac{8}{3}\right) \right\} = 1,$$

$$i.e. \text{ if } \quad A \left\{ \frac{24 - 16}{3} \right\} = 1.$$

This gives $A = \frac{3}{8}$.

7.6. (b) Clearly $f(x) \geq 0$, and

$$\int\limits_{-\infty}^{\infty} f(x)\, dx = \int\limits_0^1 \frac{x}{2} dx + \frac{1}{4} \int\limits_1^2 (3-x) dx + \frac{1}{4} \int\limits_2^3 dx + \frac{1}{4} \int\limits_3^4 (4-x)\, dx$$

$$= \left[\frac{x^2}{4} \right]_0^1 + \frac{1}{4} \left[3x - \frac{x^2}{2} \right]_1^2 + \frac{1}{4} \left[x \right]_2^3 + \frac{1}{4} \left[4x - \frac{x^2}{2} \right]_3^4$$

$$= \frac{1}{4} + \frac{1}{4} \left[(6-2) - \left(3 - \frac{1}{2}\right) \right] + \frac{1}{4} [3 - 2] + \frac{1}{4} \left[\left(16 - \frac{16}{2}\right) - \left(12 - \frac{9}{2}\right) \right]$$

$$= \frac{1}{4} + \frac{1}{4}\left(\frac{3}{2}\right) + \frac{1}{4} + \frac{1}{4}\left(\frac{1}{2}\right)$$

$$= \frac{1}{4} + \frac{3}{2} + \frac{1}{4} + \frac{1}{3} = 1.$$

As $f(x)$ is a density function, therefore

$$P(X \geq 3) = \frac{1}{4} \int\limits_3^4 (4 - x)\, dx = \frac{1}{4} \left[4x - \frac{x^2}{2} \right]_3^4$$

$$= \frac{1}{4} \left[\left(16 - \frac{16}{2}\right) - \left(12 - \frac{9}{2}\right) \right] = \frac{1}{8};$$

$P(X=2) = 0$, because for a continuous r.v., the probability at any particular value is equal to zero.

$$P(|X| < 1.5) = P(-1.5 < X < 1.5)$$

$$= \int_{-1.5}^{0} 0.\, dx + \int_{0}^{1} \frac{x}{2}\, dx + \frac{1}{4} \int_{1}^{1.5} (3-x)\, dx$$

$$= \left[\frac{x^2}{4}\right]_0^1 + \frac{1}{4}\left[3x - \frac{x^2}{2}\right]_1^{3/2} = \frac{1}{4} + \frac{1}{4}\left[\left(\frac{9}{2} - \frac{9}{8}\right) - \left(3 - \frac{1}{2}\right)\right]$$

$$= \frac{1}{4} + \frac{1}{4}\left[\frac{27}{8} - \frac{5}{2}\right] = \frac{1}{4} + \frac{7}{32} = \frac{15}{32}, \text{ and}$$

$$P(1 < X < 3) = \frac{1}{4}\int_1^2 (3-x)\, dx + \frac{1}{4}\int_2^3 dx$$

$$= \frac{1}{4}\left[3x - \frac{x^2}{2}\right]_1^2 + \frac{1}{4}\Big[x\Big]_2^3$$

$$= \frac{1}{4}\left[\left(6 - \frac{4}{2}\right) - \left(3 - \frac{1}{2}\right)\right] + \frac{1}{4}[3 - 2]$$

$$= \frac{1}{4}\left(4 - \frac{5}{2}\right) + \frac{1}{4} = \frac{3}{8} + \frac{1}{4} = \frac{5}{8}.$$

7.7. (b) Given $f(x) = A\,(2 - x)\,(2 + x)$, if $0 \le x \le 2$

$= 0$, elsewhere

(i) To find the value of A, we must have that

$$\int_0^2 A(2 - x)\,(2 + x)\, dx = 1$$

Solving, $A \int_0^2 (4 - x^2)\, dx = 1$ or $A\left[4x - \frac{x^3}{3}\right]_1^2 = 1$ which

gives $A = \frac{3}{16}$.

(ii) $P\left(X = \frac{1}{2}\right) = 0$, as probability at a particular value is zero.

(iii) $P(X \le 1) = \frac{3}{16}\int_0^1 (4 - x^2)\, dx = \frac{3}{16}\left[4x - \frac{x^3}{3}\right]_0^1$

$$= \frac{3}{16}\left[\left(4 - \frac{1}{3}\right) - 0\right] = \frac{11}{16}.$$

(iv) $P(X \ge 2) = 0$, as outside the range 0 to 2, $f(x) = 0$.

(v) $P(1 \le X \le 2) = \frac{3}{16}\int_1^2 (4 - x^2)\, dx = \frac{3}{16}\left[4x - \frac{x^3}{3}\right]_1^2$

$$= \frac{3}{16}\left[\left(8 - \frac{8}{3}\right) - \left(4 - \frac{1}{3}\right)\right]$$

$$= \frac{3}{16}\left[\frac{16}{3} - \frac{11}{3}\right] = \frac{3}{16} \times \frac{5}{3} = \frac{5}{16}.$$

7.8. (i) The function f(x) will be a density function, if

$$f(x) \ge 0 \text{ for every } x, \text{ and } \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Now the first condition is obvious and the second condition will be satisfied if

$$\int_0^1 6x\,(1 - x)\, dx = 1,$$

i.e. $6\int_0^1 (x - x^2)\, dx = 1$ *i.e.* $6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1$

i.e. $6\left[\frac{1}{2} - \frac{1}{3}\right] = 1$, which is true.

Therefore $f(x)$ is a legitimate *p.d.f.*

(ii) The cumulative distribution function (*c.d.f.*) is obtained as

$$F(x) = \int_{-\infty}^{x} f(x)\, dx = 6\int_0^x x(1 - x)\, dx$$

$$= 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^x = 3x^2 - 2x^3$$

(iii) Now the probability in the interval $\frac{1}{3}$ to $\frac{2}{3}$ is

$$P(\tfrac{1}{3} < X < \tfrac{2}{3}) = \int_{1/3}^{2/3} f(x)\,dx$$

$$= 6\int_{1/3}^{2/3} (x - x^2)\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{1/3}^{2/3}$$

$$= 6\left[\left(\frac{1}{2}\cdot\frac{4}{9} - \frac{1}{3}\cdot\frac{8}{27}\right) - \left(\frac{1}{2}\cdot\frac{1}{9} - \frac{1}{3}\cdot\frac{1}{27}\right)\right]$$

$$= 6\left[\frac{10}{81} - \frac{7}{162}\right] = 6 \times \frac{13}{162} = \frac{13}{27};\ \text{and}$$

$$P(X \le \tfrac{1}{2} / \tfrac{1}{3} \le X \le \tfrac{2}{3}) = \frac{P(\frac{1}{3} \le X \le \frac{1}{2})}{P(\frac{1}{3} \le X \le \frac{2}{3})},\ \text{where}$$

$$P(\tfrac{1}{3} \le X \le \tfrac{1}{2}) = \int_{1/3}^{1/2} 6(x - x^2)\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{1/3}^{1/2}$$

$$= 6\left[\left(\frac{1}{2}\cdot\frac{1}{4} - \frac{1}{3}\cdot\frac{1}{8}\right) - \left(\frac{1}{2}\cdot\frac{1}{9} - \frac{1}{3}\cdot\frac{1}{27}\right)\right]$$

$$= 6\left[\frac{1}{12} - \frac{7}{162}\right] = 6 \times \frac{13}{324} = \frac{13}{54},$$

$$\therefore\ P(X \le \tfrac{1}{2} / \tfrac{1}{3} \le X \le \tfrac{2}{3}) = \left(\frac{13}{54}\right) \Big/ \left(\frac{13}{27}\right) = \frac{13}{54} \times \frac{27}{13} = \frac{1}{2}.$$

(iv) Now $\quad P(X < b) = 6\int_0^b x(1 - x)\,dx$

$$= 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^b = 3b^2 - 2b^3;\ \text{and}$$

$$P(X > b) = 6\int_b^1 x(1 - x)\,dx$$

$$= 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_b^1 = 6\left[\left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{b^2}{2} - \frac{b^3}{3}\right)\right]$$

$$= 1 - 3b^2 + 2b^3$$

We are given that $P(X < b) = 2P(X > b)$

i.e. $\quad 3b^2 - 2b^3 = 2(1 - 3b^2 + 2b^3)$

i.e. $\quad 6b^3 - 9b^2 = -2$

i.e. $\quad b^3 - \frac{3}{2}b^2 + \frac{1}{3} = 0.$

Solving this equation by the Newton-Raphson method, we find that $b = 0.6130$.

Note: The Newton-Raphson method and solution is given below:

Newton-Raphson Formula. Let x_r be an approximation to real root of a function $f(x) = 0$. Then a better approximation, x_{r+1} is given by

$$x_{r+1} = x_r - \frac{f(x_r)}{f'(x_r)}.$$

That is, if x_0 is the initial root, then

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)},$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)},\ \text{etc.}$$

are the first approximation, the second approximation, etc.

Here $\quad f(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{3}$, and $\quad$ (replacing b by x)

$$f'(x) = 3x^2 - 3x.$$

Thus $$x_{r+1} = x_r - \frac{x_r^3 - \frac{3}{2}x_r^2 + \frac{1}{3}}{3x_r^2 - 3x_r} = \frac{2x_r^3 - \frac{3}{2}x_r^2 - \frac{1}{3}}{3x_r^2 - 3x_r}$$

Let the approximate root be 0.5. Then for $r = 0$, and $x_0 = 0.5$, we have

$$x_1 = \frac{2(0.5)^3 - \frac{3}{2}(0.5)^2 - \frac{1}{3}}{3(0.5)^2 - 3(0.5)} = \frac{-0.458}{-0.75} = 0.6107$$

Putting $r = 1$ and $x_1 = 0.6107$, we find

$x_2 = 0.6130$ as a better approximation.

7.9. (i) Let A be the event that a tube will last less than 200 hours, and B be the event that the tube functions after 150 hours of service. We seek. P(A/B) which is given as

$$P(A/B)\ \frac{P(A \cap B)}{P(B)}$$

Now $$P(A \cap B) = \int_{150}^{200} \frac{100}{x^2}\,dx$$

$$= \left[\frac{-100}{x}\right]_{150}^{200} = \frac{100}{150} - \frac{100}{200} = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

$$P(B) = \int_{150}^{\infty} \frac{100}{x^2}\,dx = \left[\frac{-100}{x}\right]_{150}^{\infty} = \frac{100}{150} = \frac{2}{3}$$

Hence $P(A/B) = \frac{1}{6} \div \frac{2}{3} = \frac{1}{6} \times \frac{3}{2} = \frac{1}{4}$.

(ii) Probability that exactly one out of 3 such tubes will have to be replaced after 150 hours of service (i.e. one not functioning and 2 functioning) **is**

$$\binom{3}{1}\left(1 - \frac{1}{4}\right)\left(\frac{1}{4}\right)^2 = \frac{9}{64}.$$

(iii) The maximum no. n, of tubes is obtained by solving

$$\left(\frac{2}{3}\right)^n = 0.5, \text{ which gives } n = 2.$$

7.10. (b) The given joint p.d. of two r.v.'s is presented in the following table:

X \ Y	1	2	3	$P(X=x_i)$ $g(x)$
1	6/30	1/30	1/30	8/30
2	4/30	5/30	1/30	10/30
3	2/30	4/30	6/30	12/30
$P(Y=y_i):h(y)$	$\frac{12}{30}$	$\frac{10}{30}$	$\frac{8}{30}$	1

The marginal *p.d.* of X is obtained by adding over the rows and that of Y, by adding over the columns. Thus the two marginal distributions are:

x	1	2	3
$g(x)$	$\frac{8}{30}$	$\frac{10}{30}$	$\frac{12}{30}$

and

y	1	2	3
$h(y)$	$\frac{12}{30}$	$\frac{10}{30}$	$\frac{8}{30}$

By definition the conditional *p.d.* of $X = x_i$ given $Y = y_j$ is

$$f(x_i/y_j) = P(X=x_i/Y=y_j) = \frac{f(x_i, y_j)}{h(y_j)}, \text{ for } i = 1,2,3; j = 1, 2, 3.$$

Thus $$f(1/1) = \frac{f(1,1)}{h(1)} = \frac{6/30}{12/30} = \frac{1}{2};$$

$$f(2/1) = \frac{f(2,1)}{h(1)} = \frac{4/30}{12/30} = \frac{1}{3};$$

$$f(3/1) = \frac{f(3,1)}{h(1)} = \frac{2/30}{12/30} = \frac{1}{6}.$$

Hence the conditional *p.d.* of X given that $Y = 1$ is

x	1	2	3
$f(x/1)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Similarly the conditional distributions of X given that $y = 2, 3$ could be found.

The conditional *p.d.* of Y for given $X = x_i$ is

$$f(y_j/x_i) = P(Y=y_j/X=x_i) = \frac{f(x_i/y_j)}{g(x_i)}.$$

Thus $f(Y/2) = \dfrac{f(2, y)}{g(2)}$ for $y = 1, 2, 3$. That is

$$f(1/2) = \frac{f(2, 1)}{g(2)} = \frac{4/30}{10/30} = \frac{2}{5};$$

$$f(2/2) = \frac{f(2, 2)}{g(2)} = \frac{5/30}{10/30} = \frac{1}{2}; \text{ and}$$

$$f(3/2) = \frac{f(2, 3)}{g(2)} = \frac{1/30}{10/30} = \frac{1}{10}.$$

Hence the conditional *p.d.* of Y given that $X = 2$ is

y	1	2	3
$f(y/2)$	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{1}{10}$

Similarly the conditional distributions of Y given that $X=1$ and 3 could be found.

7.11. The marginal distributions g(x) and h(y) are obtained by adding over the columns and rows respectively in the following table:

Y \ X	1	2	3	$h(y)$
1	1/12	1/6	0	3/12
2	0	1/9	1/5	14/45
3	1/18	1/4	2/15	79/180
$g(x)$	$\frac{5}{36}$	$\frac{19}{36}$	$\frac{5}{15}$	1

Thus the marginal distributions are:

x	1	2	3
$g(x)$	$\frac{5}{36}$	$\frac{19}{36}$	$\frac{5}{15}$

and

y	1	2	3
$h(y)$	$\frac{3}{12}$	$\frac{14}{45}$	$\frac{79}{180}$

By definition, the conditional *p.d.* of $X = x_i$ given $Y = y_j$ is

$$f(x_i/y_j) = P(X=x_i/Y=y_j) = \frac{f(x_i, y_j)}{h(y_j)}$$

For $y = 1$, the conditional distribution of X is

$$f(1/1) = \frac{f(1, 1)}{h(1)} = \frac{1/12}{3/12} = \frac{1}{3};$$

$$f(2/1) = \frac{f(2, 1)}{h(1)} = \frac{1/6}{3/12} = \frac{2}{3}; \text{ and}$$

$$f(3/1) = \frac{f(3, 1)}{h(1)} = \frac{0}{3/12} = 0.$$

Hence the conditional *p.d.* of X given that $Y = 1$ is

x	1	2	3
$f(x/1)$	$\frac{1}{3}$	$\frac{2}{3}$	0

Similarly the conditional distributions of X could be found for given $y = 2, 3$ respectively.

The conditional *p.d.* of $Y = y_j$ for given $X = x_i$ are

$$f(y_j/x_i) = P(Y=y_j/X=x) = \frac{f(x,y)}{g(x)}$$

Thus for x = 3, the conditional distribution of Y is

$$f(1/3) = \frac{f(3, 1)}{g(3)} = \frac{0}{5/15} = 0;$$

$$f(2/3) = \frac{f(3, 2)}{g(3)} = \frac{1/5}{5/15} = \frac{3}{5}; \text{ and}$$

$$f(3/3) = \frac{f(3, 3)}{g(3)} = \frac{2/15}{5/15} = \frac{2}{5}$$

Hence the conditional distribution of Y given that $X = 3$ is

y	1	2	3
$f(y/3)$	0	$\frac{3}{5}$	$\frac{2}{5}$

Similarly for $x=1$ and $x=2$, conditional *p.d.* of y could be found.

(b) The r.v.'s X and Y will be independent, if

$$f(x, y) = g(x)\ h(y).$$

Now $\quad f(1,1) = \frac{1}{12}$ and $g(1) \cdot h(1) = \left(\frac{3}{12}\right)\left(\frac{5}{36}\right) = \frac{5}{144}$.

Since $f(1,1) \neq g(1) \cdot h(1)$, therefore X and Y are not indpendent.

7.12. (i) The marginal probability distribution for X is

$$g(x) = \sum_y f(x, y)$$
$$= \sum_{y=1}^{3} \frac{xy}{66} = \frac{x}{66} + \frac{2x}{66} + \frac{3x}{66} = \frac{x}{11}, \text{ for } x = 2,4,5;$$

and the marginal probability distribution for Y is

$$h(y) = \sum_x f(x, y)$$
$$= \sum_x \frac{xy}{66} = \frac{2y}{66} + \frac{4y}{66} + \frac{5y}{66} = \frac{y}{6}, \text{ for } x = 1,2,3.$$

Now $\quad \frac{xy}{66} = \frac{x}{11} \times \frac{y}{6}$.

i.e. $f(x,y) = g(x)h(y)$, therefore X and Y are independent.

(ii) The marginal *p.d.* for X is

$$g(x) = \sum_y f(x, y)$$
$$= \sum_{y=1}^{2} \frac{xy^2}{30} = \frac{x}{30} + \frac{4x}{30} = \frac{x}{6}, \text{ for } x = 1,2,3; \text{ and}$$

the marginal *p.d.* for Y is

$$h(y) = \sum_x f(x, y)$$
$$= \sum_{x=1}^{2} \frac{xy^2}{30} = \frac{y^2}{30} + \frac{2y^2}{30} + \frac{3y^2}{30} = \frac{y^2}{5}, \text{ for } y = 1,2.$$

Now $\quad \frac{xy^2}{30} = \frac{x}{6} \times \frac{y^2}{5}$, *i.e.* $f(x, y) = g(x)\ h(y)$

Hence X and Y are independent.

7.13. (b) Given the joint p.d.f. $f(x,y) = 3xy(x+y)$, for $0 \le x \le 1$, $0 \le y \le 1$.

The marginal *p.d.f.* of X is

$$g(x) = 3\int_0^1 xy\,(x + y)\, dy, \qquad 0 \le x \le 1,$$
$$= 3\int_0^1 (x^2y + xy^2)\, dy, \qquad 0 \le x \le 1,$$
$$= 3\left[\frac{x^2y^2}{2} + \frac{xy^3}{3}\right]_0^1 \qquad 0 \le x \le 1,$$
$$= 3\left[\frac{x^2}{2} + \frac{x}{3}\right] \qquad 0 \le x \le 1,$$
$$= \frac{3x^2}{2} + x, \qquad 0 \le x \le 1.$$

Similarly, the marginal *p.d.f.* of Y is

$$h(y) = 3\int_0^1 (x^2y + xy^2)\, dx, \qquad 0 \le y \le 1,$$
$$= 3\left[\frac{x^3y}{3} + \frac{x^2y^2}{2}\right]_0^1 \qquad 0 \le y \le 1,$$
$$= 3\left[\frac{y}{2} + \frac{y^2}{2}\right] \qquad 0 \le y \le 1,$$
$$= y + \frac{3}{2}y^2. \qquad 0 \le y \le 1.$$

The conditional *p.d.f.* of X given $Y = y$ is

$$f(x/y) = \frac{f(x, y)}{h(y)}, \quad \text{where } h(y) > 0$$
$$= \frac{3xy\,(x + y)}{y + \frac{3}{2}y^2} = \frac{3x\,(x + y)}{1 + \frac{3}{2}y},$$

and the conditional $p.d.f.$ of Y given $X = x$, is

$$f(y/x) = \frac{f(x, y)}{g(x)} = \frac{3xy\,(x + y)}{x\,(1 + \frac{3}{2}x)} = \frac{3y\,(x + y)}{1 + \frac{3}{2}x}$$

7.14. Given that

$$f(x, y) = x^2 + \frac{1}{3}xy, \qquad 0 \le x \le 1, \quad 0 \le y \le 2$$

(a) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = \int_0^1\int_0^2 \left(x^2 + \frac{1}{3}xy\right) dy\,dx$$

$$= \int_0^1 \left[x^2y + \frac{1}{6}xy^2\right]_0^2 dx$$

$$= \int_0^1 \left(2x^2 + \frac{2x}{3}\right) dx$$

$$= \left[\frac{2x^3}{3} + \frac{x^2}{3}\right]_0^1 = \frac{2}{3} + \frac{1}{3} = 1.$$

(b) (i) $P(X > \frac{1}{2})$ $$= \int_{1/2}^1\int_0^2 \left(x^2 + \frac{1}{3}xy\right) dy\,dx$$

$$= \int_{1/2}^1 \left[x^2y + \frac{xy^2}{6}\right]_0^2 dx = \int_{1/2}^1 \left(2x^2 + \frac{2x}{3}\right) dx$$

$$= \left[\frac{2x^3}{3} + \frac{x^2}{3}\right]_{1/2}^1 = \left(\frac{2}{3} + \frac{1}{3}\right) - \left(\frac{1}{12} + \frac{1}{12}\right)$$

$$= 1 - \frac{2}{12} = \frac{5}{6}.$$

(ii) $P(Y < X)$ $$= \int_0^1\int_0^x \left(x^2 + \frac{1}{3}xy\right) dy\,dx \quad (\because\ y<x)$$

$$= \int_0^1 \left[x^2y + \frac{1}{6}xy^2\right]_0^x dx = \int_0^1 \left(x^3 + \frac{x^3}{6}\right) dx$$

$$= \frac{7}{6}\int_0^1 x^3 dx = \frac{7}{6}\left[\frac{x^4}{4}\right]_0^1 = \frac{7}{6} \times \frac{1}{4} = \frac{7}{24}$$

(iii) $P(X + Y > 1)$ $= 1 - P(X + Y < 1)$

$$= 1 - \int_0^1\int_0^{1-x} \left(x^2 + \frac{1}{3}xy\right) dy\,dx$$

$$(\because\ x+y<1,\ \therefore\ y = 1 - x)$$

$$= 1 - \int_0^1 \left[x^2y + \frac{1}{6}xy^2\right]_0^{1-x} dx$$

$$= 1 - \int_0^1 \left\{x^2\,(1 - x) + \frac{1}{6}x\,(1 - x)^2\right\} dx$$

$$= 1 - \int_0^1 \frac{1}{6}\,(x + 4x^2 - 5x^3)\,dx$$

$$= 1 - \frac{1}{6}\left[\frac{x^2}{2} + \frac{4x^3}{3} - \frac{5x^4}{4}\right]_0^1$$

$$= 1 - \frac{1}{6}\left[\frac{1}{2} + \frac{4}{3} - \frac{5}{4}\right] = 1 - \frac{1}{6}\left(\frac{7}{12}\right)$$

$$= 1 - \frac{7}{72} = \frac{65}{72}.$$

(iv) Now $P(Y < \frac{1}{2} \mid X < \frac{1}{2})$ $= \dfrac{P(Y < \frac{1}{2} \text{ and } X < \frac{1}{2})}{P(X<\frac{1}{2})}$, and

$$P(Y < \tfrac{1}{2} \text{ and } X < \tfrac{1}{2}) = \int_0^{1/2}\int_0^{1/2} (x^2 + \tfrac{1}{3}xy)\, dy\, dx$$

$$= \int_0^{1/2} \left[x^2y + \frac{1}{6}xy^2\right]_0^{1/2} dx$$

$$= \int_0^{1/2} \left(\frac{x^2}{2} + \frac{x}{24}\right) dx$$

$$= \left[\frac{x^3}{6} + \frac{x^2}{48}\right]_0^{1/2} = \frac{1}{48} + \frac{1}{192} = \frac{5}{192}$$

$$\therefore P(Y<\tfrac{1}{2} \mid X < \tfrac{1}{2}) = \frac{5/192}{1/6} = \frac{5}{32}, \quad (\because P(X<\tfrac{1}{2}) = \tfrac{1}{6}).$$

7.15. (b) The marginal density functions are obtained as below:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad 0 < x < 1,$$

$$= \frac{2}{3}\int_0^1 (x + 2y)\, dy, \qquad 0 < x < 1,$$

$$= \frac{2}{3}\left[xy + \frac{2y^2}{2}\right]_0^1 = \frac{2}{3}[x + 1] \qquad 0 \le x < 1.$$

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad 0 < y < 1,$$

$$= \frac{2}{3}\int_0^1 (x + 2y)\, dx, \qquad 0 < y < 1,$$

$$= \frac{2}{3}\left[\frac{x^2}{2} + 2xy\right]_0^1, \qquad 0 < y < 1.$$

$$= \frac{2}{3}\left[\frac{1}{2} + 2y\right]_0^1, \qquad 0 < y < 1,$$

$$= \frac{1}{3} + \frac{2}{3}y \qquad 0 < y < 1.$$

The conditional *p.d.f.* of X given $Y = y$ is

$$f(x/y) = \frac{f(x, y)}{h(y)}, \quad \text{where } h(y) > 0.$$

$$= \frac{\frac{2}{3}(x + 2y)}{\frac{2}{3}[\frac{1}{2} + 2y]} = \frac{x + 2y}{\frac{1}{2} + 2y} = \frac{2x + 4y}{1 + 4y};$$

and the conditional *p.d.f.* of Y given $X = x$ is

$$f(y/x) = \frac{f(x, y)}{g(y)}, \quad \text{where } g(x) > 0$$

$$= \frac{\frac{2}{3}(x + 2y)}{\frac{2}{3}(x + 1)} = \frac{x + 2y}{x + 1}.$$

We know that $f(x/y) = \dfrac{2x + 4y}{1 + 4y}$, therefore

$$P(X<\tfrac{1}{2} / y=\tfrac{1}{2}) = \int_0^{1/2} f(x/\tfrac{1}{2})\, dx = \int_0^{1/2} \frac{2x + 4(\frac{1}{2})}{1 + 4(\frac{1}{2})} dx$$

$$= \int_0^{1/2} \frac{2x + 2}{3} . dx = \frac{2}{3}\left[\frac{x^2}{2} + x\right]_0^{1/2}$$

$$= \frac{2}{3}\left[\frac{1}{8} + \frac{1}{2}\right] = \frac{2}{3} \times \frac{5}{8} = \frac{5}{12}.$$

7.16. The marginal density functions are obtained below:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= 3\int_0^1 (x^2y + y^2x)\,dy = 3\left[\frac{x^2y^2}{2} + \frac{y^3x}{3}\right]_0^1$$

$$= \frac{3}{2}x^2 + x, \qquad 0 \le x \le 1.$$

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

$$= 3\int_0^1 (x^2y + y^2x)\,dx = 3\left[\frac{x^3y}{3} + \frac{y^2x^2}{2}\right]_0^1$$

$$= y + \frac{3}{2}y^2. \qquad 0 \le y \le 1.$$

The conditional distributions are

$$f(x/y) = \frac{f(x, y)}{h(y)} = \frac{3xy(x + y)}{y + (3/2)\,y^2} = \frac{6x(x + y)}{2 + 3y}, \text{ and}$$

$$f(y/x) = \frac{f(x, y)}{g(x)} = \frac{3xy(x + y)}{x + (3/2)\,x^2} = \frac{6y(x + y)}{2 + 3x}.$$

The conditional probability is obtained below:

$f(x/y) = \frac{f(x,y)}{h(y)}$

$$P\left[\tfrac{1}{2}\le x\le\tfrac{3}{4} \,\middle|\, \tfrac{1}{2}\le y\le\tfrac{2}{3}\right] = \frac{\int_{1/2}^{3/4}\int_{1/2}^{2/3}(3x^2y+3y^2x)\,dydx}{\int_{1/2}^{2/3}\left(y + \frac{3}{2}y^2\right)dy}$$

Now $$\int_{1/2}^{3/4}\int_{1/2}^{2/3}(3x^2y+3y^2x)\,dydx = \int_{1/2}^{3/4}\left[\frac{3x^2y^2}{2} + \frac{3y^2x}{3}\right]_{1/2}^{2/3}dx$$

$$= \int_{1/2}^{3/4}\left(\frac{7}{24}x^2 + \frac{37}{216}x\right)dx$$

$$= \left[\frac{7}{24}\cdot\frac{x^2}{3} + \frac{37}{216}\cdot\frac{x^2}{2}\right]_{1/2}^{3/4} = \frac{769}{13824}$$

and $$\int_{1/2}^{2/3}\left(y + \frac{3}{2}y^2\right)dy = \left[\frac{y^2}{2} + \frac{3y^3}{6}\right]_{1/2}^{2/3} = \frac{79}{432}$$

Hence the required conditional probability is

$$\frac{769}{13824} \times \frac{432}{79} = \frac{769}{2528} = 0.304.$$

7.17. The marginal distribution of X is obtained as

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$$

$$= 24\int_0^1 x^2y(1 - x)\,dy = 24\left[\frac{x^2y^2(1-x)}{2}\right]_0^1$$

$$= 12x^2\,(1 - x)$$

The conditional distribution of Y given X is

$$f(y/x) = \frac{f(x, y)}{g(x)} = \frac{24\,x^2y\,(1 - x)}{12x^2\,(1 - x)} = 2y$$

Also the marginal distribution of Y is obtained as

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

$$= 24\int_0^1 x^2y(1 - x)\,dx = 24\left[y\left(\frac{x^3}{3} - \frac{x^4}{4}\right)\right]_0^1 = 2y$$

Now obviously $f(x, y) = g(x)\,.\,h(y)$

Hence X and Y are independent.

7.19. (c) Let the r.v. X represent the sum of the spots on the dice. Then X has the following p.d.

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The expected value of the payment is

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \dots + 12 \times \frac{1}{36} = \frac{252}{36} = 7$$

7.20. (a) Let the r.v. X represent the number of women on the committee. Then X has the following p.d.

x	$f(x)$	$xf(x)$
0	$\binom{3}{0}\binom{5}{5} \div \binom{8}{5} = \frac{1}{56}$	0
1	$\binom{3}{1}\binom{5}{4} \div \binom{8}{5} = \frac{15}{56}$	$\frac{15}{56}$
2	$\binom{3}{2}\binom{5}{3} + \binom{8}{5} = \frac{30}{56}$	$\frac{60}{56}$
3	$\binom{3}{3}\binom{5}{2} \div \binom{8}{5} = \frac{10}{56}$	$\frac{30}{56}$
Σ	... 1	$\frac{105}{56}$

Hence the expected number of women on the committee is

$$E(X) = \sum xf(x) = \frac{105}{56}$$

(b) The mathematical expectation of a random variable X, i.e. E(X) is defined as

$$E(X) = \sum_i x_i\, f(x_i)$$

the summation extending over all possible values of X, provided that $\sum_i \mid x_i \mid f(x_i)$ converges.

Now $\quad x_i = 2, -2, \frac{8}{3}, -4, \dots, (-1)^{j+1}\frac{2^j}{j}, \dots$

$$f(x_i) = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots, \left(\frac{1}{2}\right)^j, \dots$$

$$\therefore \quad \sum_j |x_j|\, f(x_j) = \sum_j \frac{2^j}{j}\left(\frac{1}{2}\right)^j = \sum_{j=1}^{\infty} \frac{1}{j},$$

which is a divergent series, and hence $E(X)$ does not exist. It is however interesting to note that

$$\sum_j x_j\, f(x_j) = \sum_{j=1}^{\infty} (-1)^{j+1}\frac{1}{j}, \text{ converges.}$$

0.25 0.5

7.21. (b) As the variables X and Y are independent, therefore their joint distribution is

$Y \backslash X$	1	2	3	4	$f(y)$	$yf(y)$
0	1/32	1/32	1/32	1/32	1/8	0
1	3/32	3/32	3/32	3/32	3/8	3/8
2	3/32	3/32	3/32	3/32	3/8	6/8
3	1/32	1/32	1/32	1/32	1/8	3/8
$f(x)$	1/4	1/4	1/4	1/4	1	12/8
$xf(x)$	1/4	2/4	3/4	4/4	10/4	

Now $\quad E(X) = \sum xf(x) = \frac{10}{4} = 2.5$; and

$$E(Y) = \sum yf(y) = \frac{12}{4} = 1.5;$$

Again $\quad E(X+Y) = \sum(x + y)\, f(x, y)$

$$= \left(1\times\frac{1}{32}\right) + \left(2\times\frac{1}{32}\right) + \left(3\times\frac{1}{32}\right) + \left(4\times\frac{1}{32}\right)$$
$$+ \left(2\times\frac{3}{32}\right) + \left(3\times\frac{3}{32}\right) + \left(4\times\frac{3}{32}\right) + \left(5\times\frac{3}{32}\right)$$
$$+ \left(3\times\frac{3}{32}\right) + \left(4\times\frac{3}{32}\right) + \left(5\times\frac{3}{32}\right) + \left(6\times\frac{3}{32}\right)$$
$$+ \left(4\times\frac{1}{32}\right) + \left(5\times\frac{1}{32}\right) + \left(6\times\frac{1}{32}\right) + \left(7\times\frac{1}{32}\right)$$
$$= \frac{128}{32} = 4$$

Also $\quad E(X) + E(Y) = 2.5 + 1.5 = 4$

Hence $\quad E(X) + E(Y) = E(X + Y)$

7.22 (a) Computation of the expected values E(X) and E(X²).

x_i	$f(x_i)$	$x_i f(x_i)$	$x_i^2 f(x_i)$
0	$\binom{3}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^3 = \frac{27}{64}$	0	0
1	$\binom{3}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2 = \frac{27}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
2	$\binom{3}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1 = \frac{9}{64}$	$\frac{18}{64}$	$\frac{36}{64}$
3	$\binom{3}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^0 = \frac{1}{64}$	$\frac{3}{64}$	$\frac{9}{64}$
$\sum$	1	$\frac{48}{64}$	$\frac{72}{64}$

Hence $\quad E(X) = \sum x_i f(x_i) = \frac{48}{64} = \frac{3}{4}$, and

$$E(X^2) = \sum x_i^2 f(x_i) = \frac{72}{64} = 1\frac{1}{8}.$$

(b) (i) Computation of E(X) and Var(X).

x_i	$f(x_i)$	$x_i f(x_i)$	$x_i^2 f(x)$
−1	0.125	−0.125	0.125
0	0.500	0.000	0.000
1	0.200	0.200	0.200
2	0.050	0.100	0.200
3	0.125	0.375	1.125
Total	1.000	0.550	1.650

Thus $\quad E(X) = \sum x f(x) = 0.55$; and

$$Var(X) = E(X^2) - [E(X)]^2 = 1.650 - (0.55)^2 = 1.650 - 0.3025$$
$$= 1.3475 = 1.35$$

(ii) Computation of the p.d. of $Y = 2X + 1$, $E(Y)$ and $Var(Y)$

$y = (2x+1)$	$f(y)$	$yf(y)$	$y^2 f(y)$
−1	0.125	−0.125	0.125
1	0.500	0.500	0.500
3	0.200	0.600	1.800
5	0.050	0.250	1.250
7	0.125	0.875	6.125
Total	1.000	2.100	9.800

$\therefore \quad E(Y) = \sum y f(y) = 2.1$, and

$$Var(Y) = E(Y^2) - [E(Y)]^2 = 9.80 - (2.1)^2 = 9.80 - 4.41 = 5.39.$$

(iii) For the relationship between $E(X)$ and $E(Y)$ where $Y = 2X+1$, we have

$$E(Y) = 2.1$$
$$= 2(0.55) + 1 = 2E(X) + 1.$$

For the relationship between $Var(X)$ and $Var(Y)$, we have

$$Var(Y) = 5.39$$
$$= 4(1.3475) = (2)^2\, Var(X).$$

7.23 (a) The function f(x) will be a density function, if

$$A \int_0^1 x^3(1-x)\, dx = 1, \text{ i.e. if } A \int_0^1 (x^3 - x^4)\, dx = 1,$$

$$\text{i.e. if } A \left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = 1, \text{ i.e. if } A = 20.$$

$\therefore \quad f(x) = 20x^3(1-x)$ is a proper density function.

(b) The mean and variance are calculated below:

Now $\quad E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$

$$= 20 \int_0^1 x(x^3-x^4)\,dx = 20 \int_0^1 (x^4 - x^5)\,dx$$

$$= 20 \left[\frac{x^5}{5} - \frac{x^6}{6} \right]_0^1 = 20 \left[\frac{1}{5} - \frac{1}{6} \right] = \frac{2}{3},$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= 20 \int_0^1 x^2(x^3-x^4)\,dx = 20 \int_0^1 (x^5 - x^6)\,dx$$

$$= 20 \left[\frac{x^6}{6} - \frac{x^7}{7} \right]_0^1 = 20 \left[\frac{1}{6} - \frac{1}{7} \right] = \frac{10}{21}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{10}{21} - \left(\frac{2}{3}\right)^2 = \frac{2}{63}$$

(c) The distribution function is

$$F(x) = P(X \le x) = 20 \int_0^x x^3(1-x)\,dx = 20 \int_0^x (x^3 - x^4)\,dx$$

$$= 20 \left[\frac{x^4}{4} - \frac{x^5}{5} \right]_0^x = 20 \left[\frac{x^4}{4} - \frac{x^5}{5} \right]$$

$$= 5x^4 - 4x^5$$

Now $$P\left(\frac{1}{4} < X < \frac{1}{2}\right) = P\left(X < \frac{1}{2}\right) - P\left(X < \frac{1}{4}\right)$$

$$= F\left(\frac{1}{2}\right) - F\left(\frac{1}{4}\right)$$

$$= \left[5\left(\frac{1}{2}\right)^4 - 4\left(\frac{1}{2}\right)^5\right] - \left[5\left(\frac{1}{4}\right)^4 - 4\left(\frac{1}{4}\right)^5\right]$$

$$= \left(\frac{5}{16} - \frac{4}{32}\right) - \left(\frac{5}{256} - \frac{4}{1024}\right)$$

$$= \frac{6}{32} - \frac{16}{1024} = \frac{3}{16} - \frac{1}{64} = \frac{11}{64}$$

7.24. (a) $$E[X-E(X)]^2 = E[X^2 - 2XE(X) + \{E(X)\}^2]$$

$$= E(X^2) - 2E(X) \,.\, E(X) + [E(X)]^2$$

$$= E(X^2) - 2[E(X)]^2 + [E(X)]^2$$

$$= E(X^2) - [E(X)]^2.$$

(b) Given that $\text{Var}(X_1) = k$, $\text{Var}(X_2) = 2$, and $\text{Var}(3X_2 - X_1) = 25$.

Since X_1 and X_2 are independent r.v.'s, therefore

$$\text{Var}(3X_2 - X_1) = 9\,\text{Var}(X_2) + \text{Var}(X_1)$$

$$= 9(2) + k$$

Thus $9(2) + k = 25$

or $k = 25 - 18 = 7$.

(c) (i) The probability distribution of r.v. X is

x	1	2	3	4
$f(x)$	1/8	2/8	3/8	2/8

(ii) $$E(X) = \sum x\, f(x)$$

$$= \left(1 \times \frac{1}{8}\right) + \left(2 \times \frac{2}{8}\right) + \left(3 \times \frac{3}{8}\right) + \left(4 \times \frac{2}{8}\right)$$

$$= \frac{22}{8} = 2.75.$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2, \text{ where}$$

$$E(X^2) = \sum x^2 f(x)$$

$$= \left(1^2 \times \frac{1}{8}\right) + \left(2^2 \times \frac{2}{8}\right) + \left(3^2 \times \frac{3}{8}\right) + \left(4^2 \cdot \frac{2}{8}\right.$$

$$= \frac{68}{8} \qquad \frac{1}{8} + \frac{8}{8} + \frac{27}{8} + \frac{32}{8}$$

$$\therefore \quad \text{Var}(X) = \left(\frac{68}{8}\right) - \left(\frac{22}{8}\right)^2 = \frac{60}{64} = \frac{15}{16}. \qquad \frac{1+8+27+32}{8}$$

7.25. (a) Calculation of mean and variance.

x_i	$f(x_i) = \frac{6 - \lvert 7-x \rvert}{36}$	$x_i f(x_i)$	$x_i^2 f(x_i)$
2	1/36	2/36	4/36
3	2/36	6/36	18/36
4	3/36	12/36	48/36
5	4/36	20/36	100/36
6	5/36	30/36	180/36
7	6/36	42/36	294/36
8	5/36	40/36	320/36
9	4/36	36/36	324/36
10	3/36	30/36	300/36
11	2/36	22/36	242/36
12	1/36	12/36	144/36
Σ	1	252/36 = 7	1974/36

$\therefore$ Mean $= E(X) = \sum x_i f(x_i) = 7$, and

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$
$$= \sum x_i^2 f(x_i) - [\sum x_i f(x_i)]^2$$
$$= \frac{1974}{36} - (7)^2 = \frac{210}{36} = \frac{35}{6}.$$

(b) Here $x_i = 1, 2, 3, \ldots, n$

$$f(x_i) = \frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}$$

$\therefore$
$$E(X) = \sum x_i f(x)$$
$$= 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + 3 \cdot \frac{1}{n} + \ldots + n \cdot \frac{1}{n}$$
$$= \frac{1}{n}[1 + 2 + 3 + \ldots + n]$$
$$= \frac{1}{n}\left[\frac{n(n+1)}{2}\right] = \frac{n+1}{2}.$$

$$E(X^2) = 1^2 \cdot \frac{1}{n} + 2^2 \cdot \frac{1}{n} + 3^2 \cdot \frac{1}{n} + \ldots + n^2 \cdot \frac{1}{n}$$
$$= \frac{1}{n}[1^2 + 2^2 + 3^2 + \ldots + n^2]$$
$$= \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] = \frac{(n+1)(2n+1)}{6}$$

$\therefore$
$$\text{Var}(X) = E(X^2) - [E(X)]^2$$
$$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$
$$= \frac{2n^2 + 3n + 1}{6} - \frac{n^2 + 2n + 1}{4} = \frac{n^2 - 1}{12}$$

(c) Here $x_i = 0, 1, 2, 3, \ldots, n$

$$f(x_i) = k\binom{n}{0}, \ k\binom{n}{1}, \ k\binom{n}{2}, \ldots, k\binom{n}{n}.$$

where k is the constant of proportionality.

Since $\sum f(x_i) = 1$, therefore we first find the value of k, the constant of proportionality. The sum is

$$k\binom{n}{0} + k\binom{n}{1} + k\binom{n}{2} + \ldots + k\binom{n}{n} = 1$$

or
$$k\left[1 + n + \frac{n(n-1)}{2!} + \ldots + 1\right] = 1$$

or
$$k[1+1]^n = 1 \text{ or } k2^n = 1 \text{ or } k = \frac{1}{2^n}$$

Now
$$E(X) = \sum x_i f(x_i)$$
$$= \frac{1}{2^n}\left[0.1 + 1.n + 2.\frac{n(n-1)}{2!} + 3.\frac{n(n-1)(n-2)}{3!} + \ldots + n.1\right]$$
$$= \frac{n}{2^n}\left[1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \ldots + 1\right]$$
$$= \frac{n}{2^n}[1+1]^{n-1} = \frac{n}{2^n}(2)^{n-1} = \frac{n}{2}.$$

$$E(X^2) = \frac{1}{2^n}\left[0^2.1 - 1^2.n + 2^2.\frac{n(n-1)}{2!} + 3^2.\frac{n(n-1)(n-2)}{3!} + \ldots + n^2.1\right]$$

$$= \frac{1}{2^n}\left[n + 2.\frac{n(n-1)}{1!} + 3.\frac{n(n-1)(n-2)}{2!} + \ldots + n^2\right]$$

$$= \frac{n}{2^n}\left[\left\{1 + (n-1) + \frac{(n-1)(n-2)}{2!} + \ldots + 1\right\} + \left\{(n-1) + \frac{2(n-1)(n-2)}{2!} + \ldots + (n-1)\right\}\right]$$

$$= \frac{n}{2^n}\left[(1+1)^{n-1} + (n-1)(1+1)^{n-2}\right]$$

$$= \frac{n}{2^n}\left[2^{n-1} + (n-1)2^{n-2}\right] = \frac{n}{2} + \frac{n(n-1)}{2^2}$$

Hence $\quad \mathrm{Var}(X) = E(X^2) - [E(X)]^2$

$$= \frac{n}{2} + \frac{n(n-1)}{4} - \left(\frac{n}{2}\right)^2 = \frac{2n + n^2 - n - n^2}{4} = \frac{n}{4}.$$

7.26. (a) The probability of throwing a 6 with one die is $\frac{1}{6}$ and that of not throwing a 6 with one die is $\frac{5}{6}$.

As A has the first throw, therefore he can win in the first, third, fifth, ..., throws.

The probability that A wins in the first throw $= \frac{1}{6}$

The probability that A wins in the 3rd throw $= \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6}$, as

both A and B have failed once before A's 3rd throw.

The probability that A wins in the 5th throw $= \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6}$ and

so on. Since these cases are mutually exclusive, therefore the probability that A wins $= \frac{1}{6} + \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6} + \ldots$

$$= \frac{1/6}{1 - \left(\frac{5}{6}\right)^2} = \frac{1}{6} \times \frac{36}{11} = \frac{6}{11}.$$

The probability that B wins $= 1 - P(A \text{ wins})$

$$= 1 - \frac{6}{11} = \frac{5}{11}.$$

Hence $\quad A$'s expectation $= \frac{5}{11} \times$ Rs. 11 = Rs. 6, and

B's expectation $= \frac{5}{11} \times$ Rs. 11 = Rs. 5.

(b) The probability of getting a spade $= \frac{13}{52} = \frac{1}{4}$, and the probability of not getting a spade $= 1 - \frac{1}{4} = \frac{3}{4}$.

Now A can cut a spade in the first, fifth, ninth, ..., drawings, the respective probabilities of which are

$$\frac{1}{4},\ \left(\frac{3}{4}\right)^4 \times \frac{1}{4},\ \left(\frac{3}{4}\right)^8 \times \frac{1}{4}, \ldots$$

Thus the probability that A cuts a spade first

$$= \frac{1}{4} + \left(\frac{3}{4}\right)^4 \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^8 \cdot \frac{1}{4} + \ldots$$

$$= \frac{1}{4}\left[1 + \left(\frac{3}{4}\right)^4 + \left(\frac{3}{4}\right)^8 + \ldots\right]$$

$$= \frac{1/4}{1 - \left(\frac{3}{4}\right)^4} = \frac{1}{4} \times \frac{256}{175} = \frac{64}{175}$$

B can cut a spade in the second, sixth, tenth, ..., drawings, the respective probabilities of which are $\frac{3}{4} \times \frac{1}{4}, \left(\frac{3}{4}\right)^5 \times \frac{1}{4}, \left(\frac{3}{4}\right)^9 \times \frac{1}{4}, \ldots$

Therefore $P(B \text{ cuts a spade first}) = \frac{3}{4} \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^5 \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^9 \cdot \frac{1}{4} + \ldots$

$$= \frac{3}{4} \cdot \frac{1}{4} \left[1 + \left(\frac{3}{4}\right)^4 + \left(\frac{3}{4}\right)^8 + \ldots\right]$$

$$= \frac{\frac{3}{4} \times \frac{1}{4}}{1 - \left(\frac{3}{4}\right)^4} = \frac{3}{16} \times \frac{256}{175} = \frac{48}{175}$$

C can cut a spade in the third, seventh, eleventh, ..., drawings, the respective probabilities of which are

$$\left(\frac{3}{4}\right)^2 \cdot \frac{1}{4}, \left(\frac{3}{4}\right)^6 \cdot \frac{1}{4}, \left(\frac{3}{4}\right)^{10} \frac{1}{4}, \ldots \text{ Thus}$$

$$P(C \text{ cuts a spade first}) = \left(\frac{3}{4}\right)^2 \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^6 \cdot \frac{1}{4} + \left(\frac{3}{4}\right)^{10} \frac{1}{4} + \ldots$$

$$= \left(\frac{3}{4}\right)^2 \cdot \frac{1}{4} \left[1 + \left(\frac{3}{4}\right)^4 + \left(\frac{3}{4}\right)^8 + \ldots\right]$$

$$= \frac{\left(\frac{3}{4}\right)^2 \cdot \frac{1}{4}}{1 - \left(\frac{3}{4}\right)^4} = \frac{9}{64} \times \frac{256}{175} = \frac{36}{175}.$$

and $P(D \text{ cuts a spade first}) = 1 - \left(\frac{64}{175} + \frac{48}{175} + \frac{36}{175}\right)$

$$= 1 - \frac{148}{175} = \frac{27}{175}.$$

Hence expectation of $A = \frac{64}{175} \times £\ 175 = £\ 64$:

expectation of $B = \frac{48}{175} \times £\ 175 = £\ 48$:

expectation of $C = \frac{36}{175} \times £\ 175 = £\ 36$: and

expectation of $D = \frac{27}{175} \times £\ 175 = £\ 27$.

(c) The bag contians 2 white and 2 black balls.

Let A denote the event that A draws a white ball first. Then

$$P(A) = \binom{2}{1} \div \binom{4}{1} = \frac{1}{2}$$

If A fails to draw the white ball, then the bag will contain 2 white and 1 black ball, as the ball drawn by A is not replaced.

Let B denote the event that B draws a white ball after A fails to draw it. Then

$P(B) = P(B$ draws white ball/A has failed to draw white ball)

$$= \frac{\binom{2}{1}}{\binom{3}{1}} \times \frac{2}{4} = \frac{1}{3}$$

If both A and B have failed to draw white ball, then the bag will contain 2 white balls as the balls being drawn are not replaced.

Let C denote the event that C draws a white ball given that both A and B have failed to draw it. Then

$$P(C) = \frac{2}{2} \times \frac{2}{4} \times \frac{1}{3} = \frac{1}{6}$$

Hence expectation of $A = \frac{1}{2} \times$ Rs. 18 = Rs. 9.

expectation of $B = \frac{1}{3} \times$ Rs. 18 = Rs. 6, and

expectation of $C = \frac{1}{6} \times$ Rs. 18 = Rs. 3

7.27. (a) The total area under the curve is

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 x^2(1-x)\, dx$$

$$= \left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \ (= N, \text{ say})$$

Hence this is not a density function.

Now mean $= \frac{1}{N}\int x \cdot f(x)\,dx$

$$= 12\int_0^1 (x^3 - x^4)\,dx = 12\left[\frac{x^4}{4} - \frac{x^5}{4}\right]_0^1$$

$$= 12\left[\frac{1}{4} - \frac{1}{5}\right] = 12 \times \frac{1}{20} = \frac{3}{5} = 0.06.$$

$$E(X^2) = 12\left[\int_0^1 x^2 f(x)\,dx\right]$$

$$= 12\int_0^1 (x^4 - x^5)\,dx = 12\left[\frac{x^5}{5} - \frac{x^6}{6}\right]_0^1$$

$$= 12\left[\frac{1}{5} - \frac{1}{6}\right] = \frac{12}{30} = \frac{2}{5} = 0.4.$$

Thus σ^2 or $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 0.4 - (0.6)^2 = 0.04$, so that

$$\sigma = \sqrt{0.04} = 0.2.$$

(b) In order that $f(x)$ may be a density function, we should have $f(x) \geq 0$ for every x, and

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

The first condition is clearly satisfied and the second condition will be satisfied if

$$c\int_1^2 x\,dx = 1, \text{ i.e. if } c\left[\frac{x^2}{2}\right]_1^2 = 1,$$

i.e. if $c\left(\frac{3}{2}\right) = 1$ *i.e.* if $c = \frac{2}{3}$.

(ii) Mean $= E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \frac{2}{3}\int_1^2 x \cdot x\,dx$

$$= \frac{2}{3}\int_1^2 x^2 dx = \frac{2}{3}\left[\frac{x^3}{3}\right]_1^2$$

$$= \frac{2}{3}\left[\frac{8}{3} - \frac{1}{3}\right] = \frac{2}{3} \cdot \frac{7}{3} = \frac{14}{9}.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \frac{2}{3}\int_1^2 x^2 x\,dx = \frac{2}{3}\int_1^2 x^3 dx = \frac{2}{3}\left[\frac{x^4}{4}\right]_1^2$$

$$= \frac{2}{3}\left[\frac{16}{4} - \frac{1}{4}\right] = \frac{2}{3} \times \frac{15}{4} = \frac{5}{2}.$$

$\therefore$ $\mathrm{Var}(X) = E(X^2) - \{E(X)\}^2 = \frac{5}{2} - \left(\frac{14}{9}\right)^2 = \frac{5}{2} - \frac{196}{81} = \frac{13}{162}.$

and $S.D.(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{13}{162}}.$

(c) $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$ (by definition)

$$= \int_{-\infty}^{0} x \cdot 0\,dx + \frac{3}{8}\int_0^2 x(x-2)^2\,dx + \int_2^{\infty} x \cdot 0\,dx$$

$$= \frac{3}{8}\int_0^2 x(x^2 - 4x + 4)\,dx = \frac{3}{8}\int_0^2 (x^3 - 4x^2 + 4x)\,dx$$

$$= \frac{3}{8}\left[\frac{x^4}{4} - \frac{4x^3}{3} + \frac{4x^2}{2}\right]_0^2$$

$$= \frac{3}{8}\left[4 - \frac{32}{3} + 8\right] = \frac{3}{8}\left(\frac{4}{3}\right) = \frac{1}{2}, \text{ and}$$

$\sigma = \sqrt{E(X^2) - [E(X)]^2}$, where

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \frac{3}{8}\int_0^2 x^2(x-2)^2\, dx = \frac{3}{8}\int_0^2 (x^4-4x^3+4x^2)dx$$

$$= \frac{3}{8}\left[\frac{x^5}{5} - \frac{4x^4}{4} + \frac{4x^3}{3}\right]_0^2 = \frac{3}{8}\left[\frac{32}{5} - 16 + \frac{32}{3}\right]$$

$$= \frac{3}{8}\left(\frac{16}{15}\right) = \frac{2}{5}$$

$$\therefore \quad \sigma = \sqrt{\frac{2}{5} - \left(\frac{1}{2}\right)^2} = \sqrt{0.4 - 0.25}$$

$$= \sqrt{0.15} = 0.387$$

7.28. (a) The function f(x) will be a density function, if

$$k\int_0^2 x^3(1-x)\, dx = 1, \text{ i.e. if } k\int_0^1 (x^3 - x^4)\, dx = 1,$$

or $$k\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = 1, \text{ or } k\left[\frac{1}{4} - \frac{1}{5}\right] = 1, \text{ or } k = 20$$

$\therefore$ $f(x) = 20x^3(1-x)$ is a proper density function.

The mean, $E(X)$ and the Var(X) are determined as below:

$$E(X) = \int_{-\infty}^{\infty} xf(x)\, dx$$

$$= 20\int_0^1 x\,.\,x^3(1-x)\, dx = 20\left[\frac{x^5}{5} - \frac{x^6}{6}\right]_0^1$$

$$= 20\left[\frac{1}{5} - \frac{1}{6}\right] = \frac{20}{30} = \frac{2}{3},$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= 20\int_0^1 x^2\,.\,x^3(1-x)\, dx = 20\int_0^1 (x^5 - x^6)\, dx$$

$$= 20\left[\frac{x^6}{6} - \frac{x^7}{7}\right]_0^1 = 20\left(\frac{1}{42}\right) = \frac{10}{21}, \text{ and}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{10}{21} - \left(\frac{2}{3}\right)^2 = \frac{10}{21} - \frac{4}{9} = \frac{2}{63}.$$

(b) The function $f(x)$ will be a density function, if

$$\int_0^3 Ax(9-x^2)\, dx = 1$$

i.e. $$A\left[\frac{9x^2}{2} - \frac{x^4}{4}\right]_0^3 = 1.$$

i.e. $$A\left[\frac{81}{2} - \frac{81}{4}\right] = 1, \text{ which gives } A = \frac{4}{81}.$$

$$\therefore \quad f(x) = \frac{4}{81}x(9-x^2), \quad \text{for } 0 \le x \le 3$$

$$= 0, \quad \text{elsewhere}$$

Now Mean $$= E(X) = \frac{4}{81}\int_0^3 x\,.\,x(9-x^2)\, dx$$

$$= \frac{4}{81}\left[\frac{9x^3}{3} - \frac{x^5}{5}\right]_0^3$$

$$= \frac{4}{81}\left[81 - \frac{243}{5}\right] = \frac{4}{81} \times \frac{162}{5} = \frac{8}{5}$$

$$E(X^2) = \frac{4}{81}\int_0^3 x^2(9x - x^3)\,dx$$

$$= \frac{4}{81}\left[\frac{9x^4}{4} - \frac{x^6}{6}\right]_0^3 = \frac{4}{81}\left[\frac{729}{4} - \frac{729}{6}\right]$$

$$= \frac{4}{81} \times \frac{729}{12} = 3.$$

$$\therefore \quad S.D. = \sqrt{\Sigma(X^2) - [E(X)]^2} = \sqrt{3 - \left(\frac{8}{5}\right)^2}$$

$$= \sqrt{3 - 2.56} = 0.66$$

7.29. The total area must be unity.

$$\frac{1}{16}\int_{-3}^{-1}(3+x)^2\,dx + \frac{1}{16}\int_{-1}^{+1}(6-2x^2)\,dx + \frac{1}{16}\int_1^3 (3-x)^2\,dx$$

$$= \frac{1}{16}\left[\frac{(3+x)^3}{3}\right]_{-3}^{-1} + \frac{1}{16}\left[6x - \frac{2x^3}{3}\right]_{-1}^{+1} + \frac{1}{16}\left[\frac{-(3-x)^3}{3}\right]_1^3$$

$$= \frac{1}{16}\left[\frac{8}{3} + \frac{32}{3} + \frac{8}{3}\right] = 1.$$

$$\text{Now } \mu = \int_{-\infty}^{\infty} xf(x)\,dx$$

$$= \frac{1}{16}\int_{-3}^{-1} x(3+x)^2 dx + \frac{1}{16}\int_{-1}^{+1} x(6-2x^2)\,dx + \frac{1}{16}\int_1^3 x(3-x)^2 dx$$

$$= \frac{1}{16}\left[\int_{-3}^{-1}(9x+6x^2+x^3)dx + \int_{-1}^{1}(6x-2x^3)dx + \int_1^3(9x-6x^2+x^3)\,dx\right]$$

$$= \frac{1}{16}\left[\frac{9x^2}{2} + \frac{6x^3}{3} + \frac{x^4}{4}\right]_{-3}^{-1} + \frac{1}{16}\left[\frac{6x^2}{2} - \frac{2x^4}{4}\right]_{-1}^{+1} +$$

$$\frac{1}{16}\left[\frac{9x^2}{2} - \frac{6x^3}{3} + \frac{x^4}{4}\right]_1^3$$

$$= \frac{1}{16}\left[\left(\frac{9}{2} - 2 + \frac{1}{4}\right) - \left(\frac{81}{2} - 54 + \frac{81}{4}\right)\right] +$$

$$\frac{1}{16}\left[\left(3 - \frac{1}{2}\right) - \left(3 - \frac{1}{2}\right)\right] + \frac{1}{16}\left[\left(\frac{81}{2} - 54 + \frac{81}{4}\right) - \left(\frac{9}{2} - 2 + \frac{1}{4}\right)\right]$$

$$= \frac{1}{16}\left[\frac{11}{4} - \frac{189}{4}\right] + 0 + \frac{1}{16}\left[\frac{189}{4} - \frac{11}{4}\right] = 0$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - (\mu)^2$$

$$= \frac{1}{16}\int_{-3}^{-1} x^2(3+x)^2\,dx + \frac{1}{16}\int_{-1}^{+1} x^2(6-2x^2)\,dx +$$

$$\frac{1}{16}\int_1^3 x^2(3-x)^2\,dx \qquad (\because \ \mu = 0)$$

$$= \frac{1}{16}\int_{-3}^{-1}(9x^2+6x^3+x^4)\,dx + \frac{1}{16}\int_{-1}^{+1}(6x^2-2x^4)\,dx +$$

$$\frac{1}{16}\int_1^3(9x^2-6x^3+x^4)\,dx$$

$$= \frac{1}{16}\left[3x^3 + \frac{3x^4}{2} + \frac{x^5}{5}\right]_{-3}^{-1} + \frac{1}{16}\left[2x^3 - \frac{2x^5}{5}\right]_{-1}^{+1} +$$

$$\frac{1}{16}\left[3x^3 - \frac{3x^4}{2} + \frac{x^5}{5}\right]_1^3$$

$$= \frac{1}{16}\left[\left(-3 + \frac{3}{2} - \frac{1}{5}\right) - \left(-81 + \frac{243}{2} - \frac{243}{5}\right)\right] +$$

$$\frac{1}{16}\left[\left(2 - \frac{2}{5}\right) - \left(-2 + \frac{2}{5}\right)\right] +$$

$$\frac{1}{16}\left[\left(81 - \frac{243}{2} + \frac{243}{5}\right) - \left(3 - \frac{3}{2} + \frac{1}{5}\right)\right]$$

$$= \frac{1}{16}\left[\frac{81}{10} - \frac{17}{10}\right] + \frac{1}{16}\left[\frac{8}{5} + \frac{8}{5}\right] + \frac{1}{16}\left[\frac{81}{10} - \frac{17}{10}\right]$$

$$= \frac{2}{5} + \frac{1}{5} + \frac{2}{5} = 1$$

Hence $\sigma = \sqrt{1} = 1$

7.30. The function f(x) will be a proper density function, if

$$k \int_0^3 (6 + x - x^2)\, dx = 1,$$

i.e. if $k \left[6x + \frac{x^2}{2} - \frac{x^3}{3} \right]_0^3 = 1$, i.e. if $k \left[18 + \frac{9}{2} - \frac{27}{3}\right] = 1$,

i.e. if $k \left[\frac{27}{2} \right] = 1$, i.e. if $k = \frac{2}{27}$.

Now $\mu = E(X) = \int_{-\infty}^{\infty} x \,.\, f(x)\, dx$

$$= \frac{2}{27} \int_0^3 x \,.\, (6 + x - x^2)\, dx = \frac{2}{27} \int_0^3 (6x + x^2 - x^3)\, dx$$

$$= \frac{2}{27} \left[\frac{6x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \right]_0^3 = \frac{2}{27} \left[27 + 9 - \frac{81}{4}\right]$$

$$= \frac{2}{27} \left(\frac{63}{4}\right) = \frac{7}{6}.$$

For the mode of the distribution, $f'(x) = 0$ and $f''(x) < 0$.

Now $f(x) = \frac{2}{27}(6 + x - x^2)$

$\therefore \quad f'(x) = \frac{2}{27}(1 - 2x)$, and $f''(x) = \frac{-4}{27} < 0$.

Thus $f'(x) = 0$ gives $x = \frac{1}{2}$

Hence Mode $= \frac{1}{2}$.

Again $\text{Var}(X) = E(X^2) - [E(X)]^2$, where

$$E(X^2) = \frac{2}{27} \int_0^3 x^2(6 + x - x^2)\, dx = \frac{2}{27} \int_0^3 (6x^2 + x^3 - x^4)\, dx$$

$$= \frac{2}{27} \left[\frac{6x^3}{3} + \frac{x^4}{4} - \frac{x^5}{5} \right]_0^3 = \frac{2}{27} \left[54 + \frac{81}{4} - \frac{243}{5}\right]$$

$= \frac{2}{27} \left[\frac{513}{20} \right] = \frac{19}{10}$. Therefore

$$\sigma^2 = \frac{19}{10} - \left(\frac{7}{6}\right)^2 = \frac{19}{10} - \frac{49}{36} = \frac{97}{180}.$$

7.31. Since f(x) is a probability density function, therefore

$$k \int_2^5 (2 - x)(x - 5)\, dx = 1.$$

We find the value of k as follows:

$$k \int_2^5 (2 - x)(x - 5)\, dx = 1$$

or $\quad k \int_2^5 (-x^2 + 7x - 10)\, dx = 1$

i.e. $\quad k \left[-\frac{x^3}{3} + \frac{7x^2}{2} - 10x \right]_2^5 = 1$

i.e. $\quad k \left[\left(-\frac{125}{3} + \frac{175}{2} - 50\right) - \left(-\frac{8}{3} + 14 - 20\right) \right] = 1$

i.e. $\quad k \left[-\frac{25}{6} + \frac{26}{3} \right] = 1$, i.e. $k \left(\frac{9}{2}\right) = 1$ giving $k = 2/9$

Now the mean of X is

$$E(X) = \frac{2}{9}\int_2^5 x(2-x)(x-5)\,dx$$

$$= \frac{2}{9}\int_2^5 (-x^3 + 7x^2 - 10x)\,dx$$

$$= \frac{2}{9}\Big[-\frac{x^4}{4} + \frac{7x^3}{3} - \frac{10x^2}{2}\Big]_2^5$$

$$= \frac{2}{9}\Big[\Big(-\frac{625}{4} + \frac{875}{3} - \frac{250}{2}\Big) - \Big(-4 + \frac{56}{3} - 20\Big)\Big]$$

$$= \frac{2}{9}\Big[\frac{125}{12} + \frac{16}{3}\Big] = \frac{2}{9}\Big(\frac{189}{12}\Big) = \frac{7}{2}, \text{ and}$$

$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$, where

$$E(X^2) = \int_2^5 x^2 f(x)\,dx = \frac{2}{9}\int_2^5 x^2(2-x)(x-5)\,dx$$

$$= \frac{2}{9}\int_2^5 (-x^4 + 7x^3 - 10x^2)\,dx$$

$$= \frac{2}{9}\Big[-\frac{x^5}{5} + \frac{7x^4}{4} - \frac{10x^3}{3}\Big]_2^5 = \frac{2}{9}\Big(\frac{1143}{20}\Big) = \frac{127}{10}$$

$$\mathrm{Var}(X) = \frac{127}{10} - \Big(\frac{7}{2}\Big)^2 = \frac{9}{20}$$

For the <u>mode</u> of the distribution, $f'(x) = 0$ and $f''(x) < 0$.

Now $\quad f(x) = \frac{2}{9}(-x^2 + 7x - 10)$

$\therefore \quad f'(x) = \frac{2}{9}(-2x + 7)$, and

$$f''(x) = \frac{2}{9}(-2) = \frac{-4}{9}, \text{ which is negative.}$$

Thus $\quad f'(x) = 0$ gives $x = \frac{7}{2}$

Hence <u>mode</u> of the distribution is at $x = \frac{7}{2}$.

The <u>median</u>, a, is given by $\int_{-\infty}^{a} f(x)\,dx = \frac{1}{2}$. Thus

$$\frac{2}{9}\int_2^a (2-x)(x-5)\,dx = \frac{1}{2}$$

or $$\frac{2}{9}\Big[-\frac{x^3}{3} + \frac{7x^2}{2} - 10x\Big]_2^a = \frac{1}{2}$$

or $$\frac{2}{9}\Big[-\frac{a^3}{3} + \frac{7a^2}{2} - 10a + \frac{26}{3}\Big] = \frac{1}{2}$$

This equation reduces to

$$4a^3 - 42a^2 + 120a - 77 = 0.$$

Since the probability density function, which is parabolic, is symmetrical, so the equation has solution $a = 7/2$ and thus we may factorize the equation given above as

$$(2a-7)(2a^2 - 14a + 11) = 0$$

Now $\quad 2a - 7 = 0$ gives $a = \frac{7}{2}$ and

$$2a^2 - 14a + 11 = 0 \text{ gives } a = 0.902 \text{ and } 6.098$$

But both the values 0.902 and 6.098 are unacceptable since a must lie in the interval (2, 5). The median is therefore given by $a = 7/2$.

7.32. By definition, we have

$$\log G = E[\log X] = \int_{-\infty}^{\infty} \log x\, f(x)\, dx, \text{ if it exists.}$$

$$= 6\int_1^2 (\log x)(2-x)(x-1)\, dx$$

$$= 6\int_1^2 (\log x)(-x^2+3x-2)\, dx$$

$$= -6\int_1^2 x^2 \log x\, dx + 18\int_1^2 x\log x\, dx - 12\int_1^2 \log x\, dx$$

To evaluate $\int_1^2 x^2 \log x dx$, we put $u = \log x$ and $dv = x^2 dx$ so that $du = \frac{1}{x}$ and $v = \frac{x^3}{3}$. Thus integrating by parts, we have

$$\int_1^2 x^2 \log x\, dx = \left[\frac{x^3}{3}\log x\right]_1^2 - \int_1^2 \frac{x^3}{3}\cdot\frac{1}{x}\, dx,$$

$$\left[\because \int_0^t u\, dv = \left[uv\right]_0^t - \int_0^t v\, du\right]$$

$$= \left[\frac{8}{3}\log 2 - 0\right] - \frac{1}{3}\left[\frac{x^3}{3}\right]_1^2 = \frac{8}{3}\log 2 - \frac{7}{9};$$

$$\int_1^2 x\log x\, dx = \left[\frac{x^2}{2}\log x\right]_1^2 - \int_1^2 \frac{x^2}{2}\cdot\frac{1}{x}\, dx$$

(Integrating by parts)

$$= 2\log 2 - \frac{1}{2}\left[\frac{x^2}{2}\right]_1^2 = 2\log 2 - \frac{3}{4}$$

$$\int_1^2 x\log x\, dx = \left[x\log x\right]_1^2 - \int_1^2 x\cdot\frac{1}{x}\, dx$$

(Integrating by parts)

$$= 2\log 2 - \left[x\right]_1^2 = 2\log 2 - 1.$$

Thus $\log G = -6\left[\frac{8}{3}\log 2 - \frac{7}{9}\right] + 18\left[2\log 2 - \frac{3}{4}\right] - 12\,[2\log 2 - 1]$

$$= -4\log 2 + \frac{19}{6} = -\log 2^4 + \frac{19}{6}$$

or $\log G + \log 16 = \frac{19}{6}$ or $\log(16G) = \frac{19}{6}$

or $6\log(16G) = 19.$

7.33. A function f(x) is a proper density function, if $\int_{-\infty}^{\infty} f(x)\, dx = 1$. **Therefore**

$$k\int_0^1 x^2(1-x)^3\, dx = 1$$

Using the given relationship $\int_0^1 x^m(1-x)^n\, dx = \frac{m!n!}{(m+n+1)!}$, **we get**

$$k\left[\frac{2!\,3!}{(2+3+1)!}\right] = 1, \text{ or } k = 60.$$

Hence $f(x) = 60x^2(1-x)^3$, **where** $0 \le x \le 1$, **is a proper** *p.d.f.*

The skewness, in this case, is measured by μ_3, **which in symmetrical distribution is zero.**

Now $\mu_1' = E(X) = \int_{-\infty}^{\infty} x\cdot f(x)\, dx$

$$= 60\int_0^1 x\cdot x^2(1-x)^3\, dx = 60\int_0^1 x^3(1-x)^3\, dx$$

Using the relationship $\int_0^1 x^m(1-x)^n\,dx = \frac{m!\,n!}{(m+n+1)!}$, we get

$$\mu_1' = 60 \cdot \frac{3!\,3!}{(3+3+1)!} = 60\left(\frac{1}{140}\right) = \frac{3}{7}.$$

$$\mu_2' = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = 60\int_0^1 x^2 \cdot x^2(1-x)^3\,dx$$

$$= 60\int_0^1 x^4(1-x)^3\,dx = 60 \cdot \frac{4!\,3!}{(4+3+1)!} = \frac{3}{14}, \text{ and}$$

$$\mu_3' = E(X^3) = \int_{-\infty}^{\infty} x^3 f(x)\,dx = 60\int_0^1 x^5(1-x)^3\,dx$$

$$= 60\,\frac{5!\,3!}{(5+3+1)!} = \frac{5}{42}.$$

Hence $\mu_3 = \mu_3' - 3\mu_1'\,\mu_2' + 2\,\mu_1'^3$

$$= \frac{5}{42} - 3\left(\frac{3}{7}\right)\left(\frac{3}{14}\right) + 2\left(\frac{3}{7}\right)^3$$

$$= \frac{5}{42} - \frac{27}{98} + \frac{54}{343} = \frac{2}{2058} = 0.001.$$

7.34 (a) First of all, we find the value of k, which should be such as to make

$$\int_{-a}^{a} k\,dx = 1$$

or $\quad k\left[x\right]_{-a}^{a} = 1$ or $2ak = 1$ or $k = \frac{1}{2a}$

Now, the mean, μ, is given by

$$\mu = \int_{-a}^{a} x \cdot \frac{1}{2a}\,dx = \frac{1}{2a}\left[\frac{x^2}{2}\right]_{-a}^{a} = \frac{1}{2a}(0) = 0.$$

The variance, σ^2 or μ_2, is given by

$$\sigma^2 = \frac{1}{2a}\int_{-a}^{a} x^2\,dx - (\mu)^2 = \frac{1}{2a}\left[\frac{x^3}{3}\right]_{-a}^{a} \qquad (\because\ \mu = 0)$$

$$= \frac{1}{2a}\left[\frac{a^3}{3} + \frac{a^3}{3}\right] = \frac{a^2}{3}.$$

Again Mean Deviation about the mean is

$$M.D. = \int_{-\infty}^{\infty} |x-\mu|\,f(x)\,dx = \frac{1}{2a}\int_{-a}^{a} |x-\mu|\,dx$$

$$= \frac{1}{2a}\left\{\int_{-a}^{0} -x\,dx + \int_0^a x\,dx\right\} \qquad (\because\ \mu = 0)$$

$$= \frac{2}{2a}\int_0^a x\,dx = \frac{1}{a}\left[\frac{x^2}{2}\right]_0^a = \frac{a}{2}.$$

(b) Calculation of the mean moments and the mean deviation.

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x\,dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

Now, $\mu_1 = 0$;

$$\mu_2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \qquad \int_{-\infty}^{\infty} x^2 f_x\, d_x$$

$$= \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \int_0^1 \left(x^2 - x + \frac{1}{4}\right) dx$$

$$= \left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12};$$

$$\mu_3 = \int_{-\infty}^{\infty} (x-\mu)^3 f(x)\,dx$$

$$= \int_0^1 \left(x-\frac{1}{2}\right)^3 dx = \int_0^1 \left(x^3 - \frac{3x^2}{2} + \frac{3x}{4} - \frac{1}{8}\right) dx$$

$$= \left[\frac{x^4}{4} - \frac{x^3}{2} + \frac{3x^2}{8} - \frac{x}{8}\right]_0^1 = \frac{1}{4} - \frac{1}{2} + \frac{3}{8} - \frac{1}{8} = \frac{0}{8} = 0;$$

$$\mu_4 = \int_{-\infty}^{\infty} (x-\mu)^4 f(x)\,dx = \int_0^1 \left(x-\frac{1}{2}\right)^4 dx$$

$$= \int_0^1 \left(x^4 - 2x^3 + \frac{3x^2}{2} - \frac{x}{2} + \frac{1}{16}\right) dx$$

$$= \left[\frac{x^5}{5} - \frac{x^4}{2} + \frac{x^3}{2} - \frac{x^2}{4} + \frac{x}{16}\right]_0^1$$

$$= \frac{1}{5} - \frac{1}{2} + \frac{1}{2} - \frac{1}{4} + \frac{1}{16} = \frac{1}{80}.$$

The Mean Deviation about the mean is

$$M.D. = \int_{-\infty}^{\infty} |x-\mu|\, f(x)\,dx = \int_0^1 |x - 1/2|\,dx$$

$$= \int_0^{1/2} \left(\frac{1}{2} - x\right) dx + \int_{1/2}^1 \left(x - \frac{1}{2}\right) dx$$

$$= \left[\frac{x}{2} - \frac{x^2}{2}\right]_0^{1/2} + \left[\frac{x^2}{2} - \frac{x}{2}\right]_{1/2}^1$$

$$= \left(\frac{1}{4} - \frac{1}{8}\right) + \left\{\left(\frac{1}{2} - \frac{1}{2}\right) - \left\{\left(\frac{1}{8} - \frac{1}{4}\right)\right\}\right.$$

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

7.35. The total probability must be unity.

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{2}\int_0^1 x\,dx + \frac{1}{2}\int_1^2 dx + \frac{1}{2}\int_2^3 (3-x)\,dx$$

$$= \frac{1}{2}\left[\frac{x^2}{2}\right]_0^1 + \frac{1}{2}\left[x\right]_1^2 + \frac{1}{2}\left[\frac{-(3-x)^2}{2}\right]_2^3$$

$$= \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{2}(1) + \frac{1}{2}\left(\frac{1}{2}\right) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1.$$

Now $\mu = \int_{-\infty}^{\infty} f(x)\,dx$

$$= \frac{1}{2}\int_0^1 x^2\,dx + \frac{1}{2}\int_1^2 x^2\,dx + \frac{1}{2}\int_2^3 x(3-x)\,dx$$

$$= \frac{1}{2}\left[\frac{x^3}{3}\right]_0^1 + \frac{1}{2}\left[\frac{x^2}{2}\right]_1^2 + \frac{1}{2}\left[\frac{3x^2}{2} - \frac{x^3}{3}\right]_2^3$$

$$= \frac{1}{2}\left(\frac{1}{3}\right) + \frac{1}{2}\left[2 - \frac{1}{2}\right] + \frac{1}{2}\left[\left(\frac{27}{2} - \frac{27}{3}\right) - \left(6 - \frac{8}{3}\right)\right]$$

$$= \frac{1}{6} + \frac{3}{4} + \frac{7}{12} = \frac{18}{12} = \frac{3}{2}.$$

$$\sigma^2 = \mu_2' - (\mu)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - (\mu)^2$$

$$= \left\{\frac{1}{2}\int_0^1 x^3 dx + \frac{1}{2}\int_1^2 x^2 dx + \frac{1}{2}\int_2^3 x^2(3-x)\,dx\right\} - \left(\frac{3}{2}\right)^2$$

$$= \left\{\frac{1}{2}\left[\frac{x^4}{2}\right]_0^1 + \frac{1}{2}\left[\frac{x^3}{3}\right]_1^2 + \frac{1}{2}\left[x^3 - \frac{x^4}{4}\right]_2^3\right\} - \frac{9}{4}$$

$$= \left\{\frac{1}{2}\left(\frac{1}{4}\right) + \frac{1}{2}\left(\frac{8}{3} - \frac{1}{3}\right) + \frac{1}{2}\left(\frac{27}{4} - \frac{4}{1}\right)\right\} - \frac{9}{4}$$

$$= \left\{\frac{1}{8} + \frac{7}{6} + \frac{11}{8}\right\} - \frac{9}{4} = \frac{8}{3} - \frac{9}{4} = \frac{5}{12}.$$

In order to find β_2, the moment measure of Kurtosis, we first find the third and fourth moments. Thus

$$\mu_3' = \int_{-\infty}^{\infty} x^3 . f(x)\, dx$$

$$= \frac{1}{2}\int_0^1 x^4 dx + \frac{1}{2}\int_1^2 x^3 dx + \frac{1}{2}\int_2^3 x^3(3-x)\, dx$$

$$= \frac{1}{2}\left[\frac{x^5}{5}\right]_0^1 + \frac{1}{2}\left[\frac{x^4}{4}\right]_1^2 + \frac{1}{2}\left[\frac{3x^4}{4} - \frac{x^5}{5}\right]_2^3.$$

$$= \frac{1}{10} + \frac{15}{8} + \frac{1}{2}\left(\frac{243}{20} - \frac{28}{5}\right) = \frac{1}{10} + \frac{15}{8} + \frac{131}{40} = \frac{21}{4}$$

$$\mu_4' = \int_{-\infty}^{\infty} x^4 . f(x)\, dx$$

$$= \frac{1}{2}\int_0^1 x^5 dx + \frac{1}{2}\int_1^2 x^4 dx + \frac{1}{2}\int_2^3 x^4(3-x)\, dx$$

$$= \frac{1}{2}\left[\frac{x^6}{6}\right]_0^1 + \frac{1}{2}\left[\frac{x^5}{5}\right]_1^2 + \frac{1}{2}\left[\frac{3x^5}{5} - \frac{x^6}{6}\right]_2^3$$

$$= \frac{1}{2}\left(\frac{1}{6}\right) + \frac{1}{2}\left(\frac{31}{5}\right) + \frac{1}{2}\left(\frac{729}{30} - \frac{128}{15}\right)$$

$$= \frac{1}{12} + \frac{31}{10} + \frac{473}{60} = \frac{166}{15}.$$

Now $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$

$$= \frac{21}{4} - 3\left(\frac{3}{2}\right)\left(\frac{8}{3}\right) + 2\left(\frac{3}{2}\right)^3 = \frac{21}{4} - \frac{12}{1} + \frac{27}{4} = 0, \text{ and}$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2' - 3\mu_1'^4$$

$$= \frac{166}{15} - 4\left(\frac{3}{2}\right)\left(\frac{21}{4}\right) + 6\left(\frac{3}{2}\right)^2\left(\frac{8}{3}\right) - 3\left(\frac{3}{2}\right)^4$$

$$= \frac{166}{15} - \frac{63}{2} + \frac{36}{1} - \frac{243}{16} = \frac{91}{240}$$

Hence $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{91}{240} \times \frac{144}{25} = \frac{273}{125} = 2.184.$

7.36. Here the distribution is $f(x) = x^2(6-x)^2$ between $x=0$ and $x=6$. Therefore we have

$$\int_0^6 x^2(6-x)^2\, dx = \int_0^6 (x^4 - 12x^3 + 36x^2)\, dx$$

$$= \left[\frac{x^5}{5} - \frac{12x^4}{4} + \frac{36x^3}{3}\right]_0^6$$

$$= \frac{(6)^5}{5} - 3(6)^4 + 12(6)^3$$

$$= 216\left[\frac{36}{5} - 18 + 12\right] = 216 \times \frac{6}{5} = \frac{1296}{5}$$

To make the total frequency unity, we must multiply $f(x)$ by $\frac{5}{1296}$.

Now we calculate the moments about origin as:

$$\mu_1' = \int_{-\infty}^{\infty} x f(x)\, dx = \frac{5}{1296}\int_0^6 x^3(x^2 - 12x + 36)\, dx$$

$$= \frac{5}{1296}\left[\frac{x^6}{6} - \frac{12x^5}{5} + \frac{36x^4}{4}\right]_0^6$$

$$= \frac{5}{1296}\left[(6)^4\left\{\frac{36}{6} - \frac{72}{5} + 9\right\}\right] = \frac{5 \times 1296}{1296}\left\{\frac{3}{5}\right\} = 3;$$

$$\mu_2' = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \frac{5}{1296}\int_0^6 x^4(x^2 - 12x + 36)\, dx$$

$$= \frac{5}{1296}\left[\frac{x^7}{7} - 2x^6 + \frac{36x^5}{5}\right]_0^6$$

$$= \frac{5}{1296} \times (6)^5 \left[\frac{36}{7} - 12 + \frac{36}{5}\right] = 30\left[\frac{12}{35}\right] = \frac{72}{7};$$

$$\mu_3' = \int_{-\infty}^{\infty} x^3 f(x)\, dx = \frac{5}{1296}\int_0^6 x^5(x^2 - 12x + 36)\, dx$$

$$= \frac{5}{1296}\left[\frac{x^8}{8} - \frac{12x^7}{7} + 6x^6\right]_0^6$$

$$= \frac{5}{1296}\left[x^6\left(\frac{x^2}{8} - \frac{12x}{7} + 6\right)\right]_0^6$$

$$= \frac{5}{1296}\left[1296 \times 36\left(\frac{9}{2} - \frac{72}{7} + 6\right)\right] = 180 \times \frac{3}{14} = \frac{270}{7}$$

$$\mu_4' = \int_{-\infty}^{\infty} x^4 f(x)\, dx = \frac{5}{1296}\int_0^6 x^6(x^2 - 12x + 36)\, dx$$

$$= \frac{5}{1296}\left[\frac{x^9}{9} - \frac{12x^8}{8} + \frac{36x^7}{7}\right]_0^6$$

$$= \frac{5}{1296}\left[x^7\left(\frac{x^2}{9} - \frac{3x}{2} + \frac{36}{7}\right)\right]_0^6$$

$$= \frac{5}{1296}\left[(1296 \times 216)\left(4 - 9 + \frac{36}{7}\right)\right]$$

$$= 1080 \times \frac{1}{7} = \frac{1080}{7}.$$

Hence the moments about the mean are:

$\mu_1 = 0;$

$$\mu_2 = \mu_2' - (\mu_1')^2 = \frac{72}{7} - (3)^2 = \frac{9}{7} = 1.29;$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$

$$= \frac{270}{7} - 3(3)\left(\frac{72}{7}\right) + 2(3)^3 = \frac{270}{7} - \frac{648}{7} + 54 = 0;$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4$$

$$= \frac{1080}{7} - \frac{3240}{7} + \frac{3888}{7} - 243 = \frac{27}{7} = 3.86;$$

and the kurtosis of the distribution is

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{27}{7} \times \left(\frac{7}{9}\right)^2 = \frac{7}{3} = 2.33.$$

7.37. To compute the marginal p.d.'s and the correlation co-efficient, the values are arranged in the tabular form as below:

x f xf(x) x²f(x)

X \ Y	1	2	3	$g(x)$
1	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{9}{15}$
2	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{6}{15}$
$h(y)$	$\frac{3}{15}$	$\frac{5}{15}$	$\frac{7}{15}$	1

The correlation co-efficient, ρ, is computed as

$$\rho = \frac{E(XY) - E(X)\,E(Y)}{\sqrt{[E(X^2) - \{(EX)\}^2]\,[E(Y^2) - \{E(Y)\}^2]}}, \text{ where}$$

$$E(X) = \sum_{i=1}^{2} x_i g(x) = 1 \times \frac{9}{15} + 2 \times \frac{6}{15} = \frac{9}{15} + \frac{12}{15} = \frac{21}{15};$$

$$E(Y) = \sum_{j=1}^{3} y_j h(y_j) = 1 \times \frac{3}{15} + 2 \times \frac{5}{15} + 3 \times \frac{7}{15}$$

$$= \frac{3}{15} + \frac{10}{15} + \frac{21}{15} = \frac{34}{15};$$

$$E(X^2) = \sum_{i=1}^{2} x_i^2 g(x_i) = 1 \times \frac{9}{15} + 4 \times \frac{6}{15} = \frac{9}{15} + \frac{24}{15} = \frac{33}{15};$$

$$E(Y^2) = \sum_{j=1}^{3} y_j^2 h(y_j) = 1 \times \frac{3}{15} + 4 \times \frac{5}{15} + 9 \times \frac{7}{15}$$

$$= \frac{3}{15} + \frac{20}{15} + \frac{63}{15} = \frac{86}{15}; \text{ and}$$

$$E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j) = 1 \times \frac{2}{15} + 2 \times \frac{4}{15} + 3 \times \frac{3}{15}$$

$$+ 2 \times \frac{1}{15} + 4 \times \frac{1}{15} + 6 \times \frac{1}{15}$$

$$= \frac{2}{15} + \frac{8}{15} + \frac{9}{15} + \frac{2}{15} + \frac{4}{15} + \frac{24}{15} = \frac{49}{15}$$

$$\therefore \rho = \frac{\frac{49}{15} - \left(\frac{21}{15}\right)\left(\frac{34}{15}\right)}{\sqrt{\left\{\frac{33}{15} - \left(\frac{21}{15}\right)^2\right\}\left\{\frac{86}{15} - \left(\frac{34}{15}\right)^2\right\}}} = \frac{\frac{49}{15} - \left(\frac{714}{225}\right)}{\sqrt{\left\{\frac{33}{15} - \frac{441}{225}\right\}\left\{\frac{86}{15} - \frac{1156}{225}\right\}}}$$

$$= \frac{\frac{735 - 714}{225}}{\sqrt{\left\{\frac{495 - 441}{225}\right\}\left\{\frac{1290 - 1156}{225}\right\}}} = \frac{\frac{21}{225}}{\sqrt{\left(\frac{54}{225}\right)\left(\frac{134}{225}\right)}}$$

$$= \frac{21}{85.06} = 0.25$$

7.38. (b) Given $f(x, y) = 2 - x - y,\ 0 \le x \le 1,\ 0 \le y \le 1$
$= 0$, elsewhere.

Now $g(x) = \int_0^1 (2-x-y)\, dy = \left[2y - xy - \frac{y^2}{2}\right]_0^1 = \frac{3}{2} - x$; and

$$h(y) = \int_0^1 (2-x-y)\, dx = \left[2x - \frac{x^2}{2} - xy\right]_0^1 = \frac{3}{2} - y.$$

$$E(X) = \int_0^1 x\left(\frac{3}{2} - x\right) dx = \left[\frac{3x^2}{4} - \frac{x^3}{3}\right]_0^1 = \frac{5}{12};$$

$$E(Y) = \int_0^1 y\left(\frac{3}{2} - y\right) dy = \left[\frac{3y^2}{4} - \frac{y^3}{3}\right]_0^1 = \frac{5}{12};$$

$$E(X^2) = \int_0^1 x^2\left(\frac{3}{2} - x\right) dx = \left[\frac{3x^3}{6} - \frac{x^4}{4}\right]_0^1 = \frac{1}{4};$$

$$E(Y^2) = \int_0^1 y^2\left(\frac{3}{2} - y\right) dy = \left[\frac{3y^3}{6} - \frac{y^4}{4}\right]_0^1 = \frac{1}{4};$$

$$E(XY) = \int_0^1\int_0^1 xy(2-x-y)\, dx\, dy = \int_0^1 \left[\frac{2xy^2}{2} - \frac{x^2y^2}{2} - \frac{xy^3}{3}\right]_0^1 dx$$

$$= \int_0^1 \left(x - \frac{x^2}{2} - \frac{x}{3}\right) dx = \left[\frac{x^2}{2} - \frac{x^3}{6} - \frac{x^2}{6}\right]_0^1 = \frac{1}{6}.$$

$$\therefore \quad \rho = \frac{E(XY) - E(X)\,E(Y)}{\sqrt{[E(X^2) - \{E(X)\}^2]\,[E(Y^2) - \{E(Y)\}^2]}}$$

$$= \frac{\frac{1}{6} - \left(\frac{5}{12}\right)\left(\frac{5}{12}\right)}{\sqrt{\left[\frac{1}{4} - \left(\frac{5}{12}\right)^2\right]\left[\frac{1}{4} - \left(\frac{5}{12}\right)^2\right]}}$$

$$= \frac{\frac{1}{6} - \frac{25}{144}}{\frac{1}{4} - \frac{25}{144}} = \frac{-\frac{1}{144}}{\frac{11}{144}} = -\frac{1}{11} = -0.091.$$

7.41(a) Two fair dice are rolled, so the sample space contains 36 outcomes, *i.e.*,

$$n(S) = 36$$

Let X = Minimum of two numbers. To get the p.d. of X, we exclude 6 sample points having the same numbers such as (1, 1), (2, 2), ..., (6, 6). Thus there remains 30 sample points for our purpose.

Then the p.d. of X is

x_i	1	2	3	4	5	Total
$f(x_i)$	$\frac{10}{30}$	$\frac{8}{30}$	$\frac{6}{30}$	$\frac{4}{30}$	$\frac{2}{30}$	1

(b) No. of White Balls = 2

No. of Black Balls = 3

Total Balls = 5

Let us assume that Person A draws first, then

$$P(A) = 2/5,$$

$$P(B/\overline{A}) = \frac{3}{5}.\frac{2}{4} = \frac{3}{10}$$

$$P(C/\overline{A}\cap\overline{B}) = \frac{3}{5}.\frac{2}{4}.\frac{2}{3} = \frac{1}{5}$$

$$P(D/\overline{A}\cap\overline{B}\cap\overline{C}) = \frac{3}{5}.\frac{2}{4}.\frac{1}{3}.\frac{2}{2} = \frac{1}{10}$$

and their expected values are:

$$E(A) = 10x\frac{2}{5} = \text{Rs}.4$$

$$E(B/\overline{A}) = 10x\frac{3}{10} = \text{Rs}.3$$

$$E(C/\overline{A}\cap\overline{B}) = 10x\frac{1}{5} = \text{Rs}.2$$

$$E(D/\overline{A}\cap\overline{B}\cap\overline{C}) = 10x\frac{1}{10} = \text{Rs}.1$$

7.42 (b)

The given m.g.f of r.v X is,

$$M_x(t) = (1-3t)^{-4}$$

$$Mean\,(X) = \mu_1' = \left[\frac{d}{dt}\{M_x\,(t)\}\right]_{t=0}$$

$$= \left[12\,(1-3t)^{-5}\right]_{t=0}$$

$$= 12$$

also,

$$\mu_2' = \left[\frac{d^2}{dt^2}\{M_x\,(t)\}\right]_{t=0}$$

$$= \left[\frac{d}{dt}\left\{12\,(1-3t)^{-5}\right\}\right]_{t=0}$$

$$= \left[180\,(1-3t)^{-6}\right]_{t=0} = 180$$

thus,

$$variance = \mu_2' - \mu_1'^2$$

$$= 180 - (12)^2 = 180 - 144 = 36$$

7.42 (c)

Let X = Annual gross earnings of Pop Singer

Then, E (X) = 40,00,000 , S.D (X) = 8,00,000

Let Y = Singer's Manager earning. Then

Y = 0.15X

Now, E (Y) = E (0.15X) = 0.15E(X) = 0.15 (4000000)

= 600000

and,

S.D (Y) = S.D (0.15X)

= 0.15*S.D (X) = 0.15 (800000) = 12000

Chapter 8

DISCRETE PROBABILITY DISTRIBUTIONS

8.2. (b) The binomial probability distribution with n = 3 and p = 0.4 is

$$f(x) = \binom{3}{x}(0.4)^x(0.6)^{3-x}, \quad \text{for } x = 0, 1, 2, 3.$$

Now $P\left(X = \frac{3}{2}\right) = f\left(\frac{3}{2}\right) = 0$; because a random variable X with a binomial distribution takes only one of the integer values $0, 1, 2, \ldots, n$.

$$P(X=2) = \binom{3}{2}(0.4)^2(0.6)^{3-2} = 0.288;$$

$$P(X\le 2) = \sum_{x=0}^{2}\binom{3}{x}(0.4)^x(0.6)^{3-x}$$

$$= \binom{3}{0}(0.4)^0(0.6)^3 + \binom{3}{1}(0.4)^1(0.6)^2 + \binom{3}{2}(0.4)^2(0.6)^1$$

$$= 0.216 + 0.432 + 0.288 = 0.936;$$

$P(X=-2) = f(-2) = 0$; because a random variable X with a binomial distribution takes only one of the non-negative integer values $0, 1, 2, 3, \ldots, n$.

$$P(X\ge 2) = \sum_{x=2}^{3}\binom{3}{x}(0.4)^x(0.6)^{3-x}$$

$$= \binom{3}{2}(0.4)^2(0.6)^1 + \binom{3}{3}(0.4)^3(0.6)^0$$

$$= 0.288 + 0.064 = 0.352.$$

8.3. (a) We observe that

(i) there are two possible outcomes, *i.e.* we will get *or* will not get a "5 or 6";

(ii) the probability of getting a "5 or 6" in each trial is $p = \frac{2}{6}$,

(iii) the successive trials are independent, and

(iv) there ore 5 trials.

Therefore the binomial distribution with $n=5$ and $p=1/3$ is appropriate.

Let X denote the number of successes. Then

(i) $P(\text{no success}) = P(X=0) = \binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5$

$$= \frac{32}{243} = 0.1317;$$

(ii) $P(\text{at least 2 successes}) = P(X\ge 2) = 1 - P(X<2)$

$$= 1 - \sum_{x=0}^{1}\binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x}$$

$$= 1 - \left[\binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5 + \binom{5}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^{5-1}\right]$$

$$= 1 - \left[\frac{32}{243} + \frac{80}{243}\right] = \frac{131}{243} = 0.5391$$

(iii) $P(\text{at least 1 but not more than 3}) = P(1\le X\le 3)$

$$= \sum_{x=1}^{3}\binom{5}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{5-x}$$

$$= \binom{5}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^4 + \binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 + \binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2$$

$$= \frac{80}{243} + \frac{80}{243} + \frac{40}{243} = \frac{200}{243} = 0.8230.$$

(b) (i) Let X denote the number of successes. Then using the binomial distribution, we get

$$P(X=3) = \binom{8}{3}(0.4)^3(0.6)^{8-3}$$

$$= (56)(0.064)(0.07776) = 0.2787$$

(ii) In the binomial distribution with $n=6$, 2 failures imply four successes, *i.e.* P(2 failures in 6 trials) = P(4 successes in 6 trials). Thus

$$P(X=4) = \binom{6}{4}(0.6)^4(0.4)^2$$

$$= (15)(0.1296)(0.16) = 0.3110$$

(iii) Let X denote the number of successes. Then

$$P(X \le 2) = \sum_{x=0}^{2}\binom{9}{x}(0.4)^x(0.6)^{9-x}$$

$$= \binom{9}{0}(0.4)^0(0.6)^9 + \binom{9}{1}(0.4)(0.6)^8 + \binom{9}{2}(0.4)^2(0.6)^7$$

$$= 0.010078 + 0.060466 + 0.161248 = 0.2318$$

8.4. (a) Let X denote the number of heads when 6 coins are tossed. Then

(i) $P(X = \text{exactly 4 heads}) = \binom{6}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^2 = 15 \cdot \frac{1}{64} = \frac{15}{64}$.

(ii) $P(X = \text{not more than 4 heads}) = 1 - P(X>4)$

$$= 1 - \sum_{x=5}^{6}\binom{6}{x}\left(\frac{1}{2}\right)^6$$

$$= 1 - \left(\frac{6}{64} + \frac{1}{64}\right) = \frac{57}{64}$$

(b) Let X denote the number of heads in a single toss of 6 fair coins. Then

(i) $P(X \ge 2) = 1 - P(X < 2)$

$$= 1 - \sum_{x=0}^{1}\binom{6}{x}\left(\frac{1}{2}\right)^6 = 1 - \left(\frac{1}{64} + \frac{6}{64}\right) = \frac{57}{64}$$

(ii) $P(X<4) = \sum_{x=0}^{3}\binom{6}{x}\left(\frac{1}{2}\right)^6$

$$= \frac{1}{64} + \frac{6}{64} + \frac{15}{64} + \frac{20}{64} = \frac{42}{64} = \frac{21}{32}$$

8.5. (a) Let p denote the probability of getting caught copying in the exam. Then $p=0.2$, $q=0.8$ and $n=3$. Since the attempts are independent, therefore

$$P(\text{not getting caught}) = q^3 = (0.8)^3 = 0.512$$

(b) Let X denote the number of voters who prefer candidate A and p, the probability that a voter prefers A.

Then $X=7$, $p=0.60$ and $n=12$. Hence

$$P(X=7) = \binom{12}{7}(0.60)^7(0.40)^{12-7}$$

$$= (792)(0.0279936)(0.01024) = 0.2270$$

(c) Let X denote the number of patients recovering from a delicate heart operation, and p, the probability that a patient recovers. Then $p=0.9$, $n=7$ and $X=5$.

Thus $P(X=5) = \binom{7}{5}(0.9)^5(0.1)^2$

$$= (21)(0.59049)(0.01) = 0.1240$$

8.6 (a) Let X denote the number of workmen, and p, the probability that a workman catches the disease. Then

$$p = \frac{20}{100} = \frac{1}{5}, \text{ so that } q = 1 - \frac{1}{5} = \frac{4}{5}, \text{ and } n = 6.$$

(i) We seek the probability that "not more than 2" will catch the disease, *i.e.* $X \le 2$. Hence

$$P(X \le 2) = \sum_{x=0}^{2}\binom{6}{x}\left(\frac{1}{5}\right)^x\left(\frac{4}{5}\right)^{6-x}$$

$$= \binom{6}{0}\left(\frac{4}{5}\right)^6 + \binom{6}{1}\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^5 + \binom{6}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^4$$

$$= \frac{4096}{15625} + \frac{6144}{15625} + \frac{3840}{15625} = \frac{14080}{15625} = \frac{2816}{3125}$$

(ii) We seek the probability that 4 workmen, 5 workmen or 6 workmen will catch the disease. Hence

$$P(X \geq 4) = \sum_{x=4}^{6} \binom{6}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{6-x}$$

$$= \binom{6}{4}\left(\frac{1}{5}\right)^4 \left(\frac{4}{5}\right)^2 + \binom{6}{5}\left(\frac{1}{5}\right)^5 \left(\frac{4}{5}\right) + \left(\frac{1}{5}\right)^6$$

$$= \frac{240}{15625} + \frac{24}{15625} + \frac{1}{15625} = \frac{265}{15625} = \frac{53}{3125}$$

(b) Here $p = \frac{12}{30} = 0.4$ so that $q = 0.6$. Thus

(i) P(1st 3 days fine and the remaining 4 days wet)

$$= q^3p^4 = (0.6)^3\,(0.4)^4$$

$$= (0.216)\,(0.0256) = 0.0055.$$

(ii) $P(\text{rain falls on just three days}) = \binom{7}{3} p^3 q^4$

$$= 35\,(0.4)^3\,(0.6)^4$$

$$= 35\,(0.064)\,(0.1296)$$

$$= 0.2903$$

8.7. We observe that

(i) there are two possible outcomes, *i.e.* a man will be alive *or* will not be alive.

(ii) the probability of being alive for each man is 2/3,

(iii) the men will remain alive independently, and

(iv) there are 5 men.

Therefore the binomial distribution with $n = 5$ and $p = \frac{2}{3}$ is appropriate to get the desired probabilities.

Let X denote the number of people who will be alive in 30 years hence. Then we seek (i) $P(X=5)$, (ii) $(P \geq 3)$, (iii) $P(X=2)$ and (iv) $P(X \leq 1)$. Hence

(i) $P(X=5) = \binom{5}{5}\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^0 = \frac{32}{243} = 0.1317,$

(ii) $P(X \geq 3) = \sum_{x=3}^{5}\binom{5}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{5-x}$

$$= \binom{5}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 + \binom{5}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right) + \binom{5}{5}\left(\frac{2}{3}\right)^5$$

$$= \frac{80}{243} + \frac{80}{243} + \frac{32}{243} = \frac{192}{243} = 0.7901;$$

(iii) $P(X=2) = \binom{5}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3 = \frac{40}{243} = 0.1646$; and

(iv) $P(X \leq 1) = \sum_{x=0}^{1}\binom{5}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{5-x}$

$$= \binom{5}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^5 + \binom{5}{1}\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^4$$

$$= \frac{1}{243} + \frac{10}{243} = \frac{11}{243} = 0.0453.$$

8.8. (a) The probability of a correct answer is $\frac{1}{4}$ = 0.25, and n = 15.

Let X denote the number of correct answers. Then

$$P(5 \leq X \leq 10) = \sum_{x=5}^{10}\binom{15}{x}(0.25)^x\,(0.75)^{15-x} = \sum_{x=5}^{10} b(x;\,15, 0.25)$$

$$= \sum_{x=0}^{10} b(x;\,15, 0.25) - \sum_{x=0}^{4} b(x;\,15, 0.25)$$

$= 0.9999 - 0.6865$ (From binomial tables),

$= 0.3134$

(b) Let p denote the probability that a spot light is red when the commuter gets to it. Then $p = 0.2$, $q = 1 - 0.2 = 0.8$ and $n = 10$.

Let X be a *r.v.* that denotes the number of times, the commuter must stop for a red light on her way to work. Then X is $b(x;\ 10, 0.2)$. To evaluate $P(X=0)$ and $P(X \le 5)$, we do the following calculations.

Now $P(X=x) = \binom{10}{x}(0.2)^x (0.8)^{10-x}$, so

$$P(X=0) = \binom{10}{0}(0.2)^0(0.8)^{10} = 0.107374 \text{ and}$$

$$P(X \le 5) = \sum_{x=0}^{5}\binom{10}{x}(0.2)^x(0.8)^{10-x}$$

$$= (0.8)^{10} + \binom{10}{1}(0.2)(0.8)^9 + \binom{10}{2}(0.2)^2(0.8)^8 + \binom{10}{3}(0.2)^3(0.8)^7 + \binom{10}{4}(0.2)^4(0.8)^6 + \binom{10}{5}(0.2)^5(0.8)^5$$

$$= 0.107374 + 0.268345 + 0.301990 + 0.201327 + 0.088080 + 0.026424 = 0.9936.$$

8.9. (a) The successive terms of the binomial distribution 600 (0.3 + 0.7) ^{6} are calculated below:

X	Probability	$f(x)$	$600\,f(x)$
0	$q^6 = (0.3)^6$	$= 0.000729$	0.4374
1	${}^6C_1\, q^5 p = 6(0.3)^5(0.7)$	$= 0.010206$	6.1236
2	${}^6C_2\, q^4 p^2 = 15(0.3)^4(0.7)^2$	$= 0.059535$	35.7210
3	${}^6C_3\, q^3 p^3 = 20(0.3)^3(0.7)^3$	$= 0.185220$	111.1320
4	${}^6C_4\, q^2 p^4 = 15(0.3)^2(0.7)^4$	$= 0.324135$	194.4810
5	${}^6C_5\, q p^5 = 6(0.3)(0.7)^5$	$= 0.302526$	181.5156
6	$p^6 = (0.7)^6$	$= 0.117649$	70.5894

(b) Let p denote the probability of getting 4, 5 or 6 with one die.

Then $\quad p = \frac{3}{6} = \frac{1}{2}$, so that $q = \frac{1}{2}$

We seek the expected frequencies of getting 0, 1, 2, ..., 5 successes with 5 dice in tossing 96 times, which are the successive terms in the binomial expansion of

$$96\left(\frac{1}{2} + \frac{1}{2}\right)^5$$

Hence the expected frequencies are:

$$96\left(\frac{1}{2}\right)^5;\ 96\binom{5}{1}\left(\frac{1}{2}\right)^5;\ 96\binom{5}{2}\left(\frac{1}{2}\right)^5,\ 96\binom{5}{3}\left(\frac{1}{2}\right)^5,$$

$$96\binom{5}{4}\left(\frac{1}{2}\right)^5 \text{ and } 96\cdot\left(\frac{1}{2}\right)^5 \text{ or } 3, 15, 30, 30, 15 \text{ and } 3$$

8.10. (a) Let p denote the probability of getting a 5 or a 6 with one die.

Then $\; p = \frac{2}{6} = \frac{1}{3}$, so that $q = 1 - \frac{1}{3} = \frac{2}{3}$.

As the dice are in sets of 8, the binomial distribution is

$$N\left(\frac{2}{3} + \frac{1}{3}\right)^8$$

Thus the expected number of 3 successs

$$= N\cdot\binom{8}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^5 = N\cdot\frac{1792}{6561}$$

Hence the proportion of the sets in which 3 successes are expected

$$= N\cdot\frac{1792}{6561} \div N = \frac{1792}{6561} \times 100\% = 27.31\%$$

(b) Let p denote the probability of getting an even number.

Then the expected number of getting 5 even numbers in 10 throws = $N \cdot \binom{10}{5} p^5 q^5$, where N is the no. of sets and $q = 1-p$.

Similarly, the expected number of getting 4 even numbers in 10 throws = $N \cdot \binom{10}{4} p^4 q^6$.

Now, by the question, we have

$$N \cdot \binom{10}{5} p^5 q^5 = 2N \cdot \binom{10}{4} p^4 q^6$$

Or $\quad \frac{10!}{5!\,5!} p^5 q^5 = 2 \cdot \frac{10!}{4!\,6!} p^4 q^6$

Or $\quad \frac{p}{5} = \frac{q}{3}$ Or $3p = 5(1-p)$

Or $\quad p = \frac{5}{8}$ and hence $q = \frac{3}{8}$.

Then the binomial distributioin becomes

$$10{,}000 \left(\frac{3}{8} + \frac{5}{8}\right)^{10}$$

Hence the expected number of getting no even number

$$= 10{,}000 \left(\frac{3}{8}\right)^{10} = \frac{36905625}{67108864} = 1 \text{ approximately.}$$

8.11. Let p be the probability of getting a 6 with one die. Then $p = \frac{1}{6}$ and $q = 1 - p = \frac{5}{6}$.

Let X denote the number of *sixes* when 4 dice are thrown. Then the theoretical frequencies of 0, 1, 2, 3, and 4 sixes are the successive terms of the binomial distribution $108\left(\frac{5}{6} + \frac{1}{6}\right)^4$, which are calculated as follows:

x	Probability	$= f(x)$	$108 \times f(x)$
0	$q^4 = \left(\frac{5}{6}\right)^4$	$= \frac{625}{1296}$	52.08
1	$\binom{4}{1} q^3 p = 4 \cdot \left(\frac{5}{6}\right)^3 \left(\frac{1}{6}\right)$	$= \frac{500}{1296}$	41.67
2	$\binom{4}{2} q^2 p^2 = 6 \cdot \left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right)^2$	$= \frac{150}{1296}$	12.50
3	$\binom{4}{3} q p^3 = 4 \cdot \left(\frac{5}{6}\right) \left(\frac{1}{6}\right)^3$	$= \frac{20}{1296}$	1.67
4	$p^4 = \left(\frac{1}{6}\right)^4$	$= \frac{1}{1296}$	0.08

The mean number of sixes in a single throw = $np = 4 \times \frac{1}{6} = \frac{2}{3}$.

8.12. (a) Calculation of the mean and standard deviation of the binomial $(q+p)^3$.

x	$f(x)$	$x \cdot f(x)$	$x^2 \cdot f(x)$
0	q^3	0	0
1	$3q^2p$	$3q^2p$	$3q^2p$
2	$3qp^2$	$6qp^2$	$12qp^2$
3	p^3	$3p^3$	$9p^3$
Σ	1		

Now $\quad \sum x \cdot f(x) = 3q^2p + 6qp^2 + 3p^3$
$= 3p\,[q^2 + 2qp + p^2] = 3p(q+p)^2 = 3p.$

$\sum x^2 \cdot f(x) = 3q^2p + 12qp^2 + 9p^3$
$= 3p\,[q^2 + 4qp + 3p^2]$
$= 3p\,[(q^2 + 2qp + p^2) + (2qp + 2p^2)]$
$= 3p\,[(q+p)^2 + 2p\,(q+p)],$
$= 3p\,[1 + 2p] = 3p + 6p^2$

Hence $\mu = \sum x \cdot f(x) = 3p$, and $\qquad (n = \sum f(x) = 1)$

$$\sigma = \sqrt{\sum x^2 . f(x) - [\sum x . f(x)]^2}$$
$$= \sqrt{3p + 6p^2 - (3p)^2}$$
$$= \sqrt{3p - 3p^2} = \sqrt{3p\,(1-p)} = \sqrt{3pq}$$

(c) As the mean and variance of the binomial $(q+p)^n$ are np and npq, therefore np=3 and npq=2.

Dividing, we get

$$\frac{npq}{np} = \frac{2}{3} \text{ or } q = \frac{2}{3}.$$

$$\therefore \quad p = 1 - q = \frac{1}{3}, \text{ and } n\left(\frac{1}{3}\right) = 3, \text{ gives } n = 9.$$

Thus the binomial distribution is $\left(\frac{2}{3} + \frac{1}{3}\right)^9$.

Hence $P(X=7) = \binom{9}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^{9-7} = \frac{16}{2187} = 0.0073.$

8.13 (a) The mean and the standard deviation of the binomial $(q+p)^n$ are np and $\sqrt{npq}$. Then

$$np = 36 \text{ and } \sqrt{npq} = 4.8 \text{ or } npq = (4.8)^2$$

Dividing, we get

$$\frac{npq}{np} = \frac{(4.8)^2}{36} \text{ or } q = 0.64$$

$$\therefore \quad p = 1 - q = 0.36$$

Putting $p = 0.36$ in $np = 36$, we get

$$n = \frac{36}{0.36} = 100.$$

(b) If X is a binomial r.v. with parameters n and p, then mean = np and variance = npq.

Now Mean = 12.38 so $np = 12.38$, and

Variance = 8.64 so $npq = 8.64$.

Substituting for np in the second equation, we get

$$(12.38)\,q = 8.64 \text{ or } q = \frac{8.64}{12.38} = 0.6979$$

so $p = 1 - 0.6979 = 0.3021$, and

$n(0.3021) = 12.38$ givs $n = 41$.

(c) Let the binomial distribution be $(q+p)^n$; $q + p = 1$.

Then $np = 5$ and $\sqrt{npq} = 3$ or $npq = 9$

Dividing, we get

$$\frac{npq}{np} = \frac{9}{5} \text{ or } q = 1.8, \text{ which is greater than unity.}$$

Hence it is not possible to have a binomial distribution with mean = 5 and *s.d.* = 3.

8.17. (a) The binomial distribution of the r.v. X is given by $P(X=x) = \binom{25}{x}(0.2)^x(0.8)^{25-x}$.

Now $\mu = np = 25\,(0.2) = 5$, and

$\sigma = \sqrt{npq} = \sqrt{25\,(0.2)\,(0.8)} = 2.$

Hence $P(X<\mu-2\sigma) = P(X<1)$ $(\because \mu-2\sigma = 5-2(2) = 1)$

$= P(X=0)$

$= \binom{25}{0}(0.2)^0(0.8)^{25} = 0.00378.$

(b) To find the median and mode of the binomial $\binom{10}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{10-x}$, **we expand it. Therefore, expanding the binomial, we get**

x	0	1	2	3	4	5	6	7	8	9	10
$f(x)$	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$
F	$\frac{1}{1024}$	$\frac{11}{1024}$	$\frac{56}{1024}$	$\frac{176}{1024}$	$\frac{386}{1024}$	$\frac{638}{1024}$	$\frac{848}{1024}$	$\frac{968}{1024}$	$\frac{1013}{1024}$	$\frac{1023}{1024}$	1

Since the binomial distribution is a *discrete* distribution, therefore

Median = a value at or below which 50% probabilities lie

= 5; and

Mode = a value that corresponds to maximum probability

= 5.

8.18. (a) To find the probability of obtaining a head, we need to calculate the mean of the frequency distribution. Thus

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{(0)(12) + (1)(50) + (2)(151) + (3)(200) + (4)(87)}{500}$$

$$= \frac{1300}{500} = 2.6$$

Let X denote the number of heads obtained in 4 tosses. Then X is a binomial *r.v.* with $n=4$, and mean $= np = 4p$.

Now $4p = 2.6$ or $p = 0.65$

Thus the probability that the coin will show heads is 0.65.

(b) Now X is b(x; 4, 0.65) and therefore

$$P(X=x) = \binom{4}{x} (0.65)^x (0.35)^{4-x} \text{ for } x = 0,1,2,3,4.$$

The theoretical binomial probabilities and frequencies are computed as follows:

x	Probability	$= f(x)$	Frequency
0	$\binom{4}{0}(0.65)^0(0.35)^4$	= 0.015006	8
1	$\binom{4}{1}(0.65)^1(0.35)^3$	= 0.111475	56
2	$\binom{4}{2}(0.65)^2(0.35)^2$	= 0.310538	155
3	$\binom{4}{3}(0.65)^3(0.35)^1$	= 0.384475	192
4	$\binom{4}{4}(0.65)^4(0.35)^0$	= 0.178506	89

8.19. We first calculate the mean of the given distribution and equate it to np to find the value of p.

Now $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0+70+122+75+28+5+0}{200} = \frac{300}{200} = 1.5.$

Thus $6p = 1.5$ or $p = 0.25$ and hence $q = 0.75$.

The fitted binomial distribution is

$$b(x; 6, 0.25) = \binom{6}{x} (0.25)^x (0.75)^{6-x}$$

The theoretical binomial prbabilities and frequencies are computed as below:

x	Probability		Theoretical Frequency
0	$\binom{6}{0}(0.75)^6$	= 0.177978	35.60
1	$\binom{6}{1}(0.25)(0.75)^5$	= 0.355957	71.19
2	$\binom{6}{2}(0.25)^2(0.75)^4$	= 0.296631	59.33
3	$\binom{6}{3}(0.25)^3(0.75)^3$	= 0.131836	26.37
4	$\binom{6}{4}(0.25)^4(0.75)^2$	= 0.032959	6.59
5	$\binom{6}{5}(0.25)^5(0.75)$	= 0.004394	0.88
6	$\binom{6}{6}(0.25)^6$	= 0.000244	0.05
$\sum$		= 0.999999	200.01

8.20. Fitting of the Binomial Distribution.

We first calculate the mean of the given distribution and equate it to np to find the value of p.

Now $\bar{x} = \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{0 + 62 + 92 + 30 + 8}{150} = \dfrac{192}{150} = 1.28.$

Thus $4p = 1.28$ or $p = 0.32$ and hence $q = 0.68$.

The fitted binomial distribution is

$$b(x; 4, 0.32) = \binom{4}{x} (0.32)^x (0.68)^{4-x}$$

The expected frequencies are computed as below:

x	Probability		Expected frequency
0	$(0.68)^4$	$= 0.213814$	32
1	$\binom{4}{1} (0.32)(0.68)^3$	$= 0.402473$	60
2	$\binom{4}{2} (0.32)^2 (0.68)^2$	$= 0.284099$	43
3	$\binom{4}{3} (0.32)^3 (0.68)$	$= 0.089129$	13
4	$(0.32)^4$	$= 0.010486$	2
Σ		$= 1.000001$	150

8.21. Calculation of the mean and standard deviation.

x	f	fx	fx^2
0	0	0	0
1	1	1	1
2	3	6	12
3	8	24	72
4	16	64	256
5	28	140	700
6	18	108	648
7	13	91	637
8	9	72	576
9	4	36	324
10	0	0	0
Total	100	542	3226

$\therefore \quad \bar{x} = \dfrac{\sum fx}{\sum f} = \dfrac{542}{100} = 5.42$, and

$$s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2} = \sqrt{\frac{3226}{100} - \left(\frac{542}{100}\right)^2}$$

$$= \sqrt{32.26 - 29.3764} = \sqrt{2.8836} = 1.70.$$

Now putting the mean equal to np, we get

$10p = 5.42$ so that $p = 0.542$ and $q = 0.458$.

Thus the fitted binomial distribution is

$$b(x; 10, 0.542) = \binom{10}{x} (0.542)^x \cdot (0.458)^{10-x}$$

The expected frequencies are computed as below:

x	Probability		Expected frequency
0	$(0.458)^{10}$	$= 0.000406$	0.0
1	$\binom{10}{1} (0.542)(0.458)^9$	$= 0.004808$	0.5
2	$\binom{10}{2} (0.542)^2 (0.458)^8$	$= 0.0256$	2.6
3	$\binom{10}{3} (0.542)^3 (0.458)^7$	$= 0.0808$	8.1
4	$\binom{10}{4} (0.542)^4 (0.458)^6$	$= 0.1673$	16.7
5	$\binom{10}{5} (0.542)^5 (0.458)^5$	$= 0.2376$	23.8
6	$\binom{10}{6} (0.542)^6 (0.458)^4$	$= 0.2343$	23.4
7	$\binom{10}{7} (0.542)^7 (0.458)^3$	$= 0.1584$	15.8
8	$\binom{10}{8} (0.542)^8 (0.458)^2$	$= 0.0703$	7.0
9	$\binom{10}{9} (0.542)^9 (0.458)$	$= 0.01849$	1.8
10	$(0.542)^{10}$	$= 0.002188$	0.2
Σ		$= 1.000192$	99.9

Mean and variance of the expected distribution are

$$\mu = np = 10\,(0.542) = 5.42$$

$$\sigma = \sqrt{npq} = \sqrt{10(0.542)\,(0.458)}$$

$$= \sqrt{2.48236} = 1.57$$

8.22. For a symmetrical binomial distribution, we have

$$p = q = \frac{1}{2}$$

Then the symmetrical binomial distribution of degree n and observations, N, is $N\left(\frac{1}{2}+\frac{1}{2}\right)^n$.

Let T_r and T_{r+1} denote the rth and $(r+1)$th terms. Then

$$T_r = N\cdot\binom{n}{r-1}\left(\frac{1}{2}\right)^{r-1}\left(\frac{1}{2}\right)^{n-r+1}$$

$$= N\cdot\binom{n}{r-1}\left(\frac{1}{2}\right)^{n}, \text{ and}$$

$$T_{r+1} = N\cdot\binom{n}{r}\left(\frac{1}{2}\right)^{r}\left(\frac{1}{2}\right)^{n-r}$$

$$= N\cdot\binom{n}{r}\left(\frac{1}{2}\right)^{n}$$

Superposing in such a way that the rth term of one coincides with the $(r+1)$th term of the other, we get

$$T_r+T_{r+1} = N\cdot\binom{n}{r-1}\left(\frac{1}{2}\right)^{r} + N\cdot\binom{n}{r}\left(\frac{1}{2}\right)^{n}$$

$$= N\left[\binom{n}{r-1}+\binom{n}{r}\right]\left(\frac{1}{2}\right)^{n}$$

$$= N\binom{n+1}{r}+\left(\frac{1}{2}\right)^{n} = 2N\binom{n+1}{r}\left(\frac{1}{2}\right)^{n+1}$$

$$= 2N\cdot\binom{n+1}{r}\left(\frac{1}{2}\right)^{r}\left(\frac{1}{2}\right)^{n+1-r}$$

$= (r+1)$th term of the binomial distribution $2N\cdot\left(\frac{1}{2}+\frac{1}{2}\right)^{n+1}$, which is symmetrical binomial of degree $(n+1)$.

8.23. (b) Given that

$M_0(t) = \left(\frac{1}{4}+\frac{3}{4}e^t\right)^{12}$. Then

$$E(X) = \left[\frac{d}{dt}\left(\frac{1}{4}+\frac{3}{4}e^t\right)^{12}\right]_{t=0}$$

$$= \left[12\left(\frac{3}{4}e^t\right)\left(\frac{1}{4}+\frac{3}{4}e^t\right)^{11}\right]_{t=0} = 12\left(\frac{3}{4}\right)(1) = 9;$$

$$E(X^2) = \left[\frac{d^2}{dt^2}\left(\frac{1}{4}+\frac{3}{4}e^t\right)^{12}\right]_{t=0}$$

$$= \left[12\left(\frac{3}{4}e^t\right)\left(\frac{1}{4}+\frac{3}{4}e^t\right)^{11}\right]_{t=0} + \left[12\,(11)\left(\frac{3}{4}\right)^2 e^{2t}\left(\frac{1}{4}+\frac{3}{4}e^t\right)^{10}\right]_{t=0}$$

$$= 12\left(\frac{3}{4}\right)(1) + 12(11)\left(\frac{3}{4}\right)^2(1) = 9+\frac{297}{4} = \frac{333}{4}.$$

$$\therefore \quad \text{Var}(X) = E(X^2)-[E(X)]^2 = \frac{333}{4}-(9)^2 = \frac{333-324}{4} = \frac{9}{4} = 2.25.$$

Alternatively. Comparing $\left(\frac{1}{4}+\frac{3}{4}e^t\right)^{12}$ with $(q+pe^t)^n$, we find that $p=\frac{3}{4}$, $q=\frac{1}{4}$ and $n=12$.

Therefore $E(X) = np = 12\times\frac{3}{4} = 9$; and

$$\text{Var}(X) = npq = 12\times\frac{3}{4}\times\frac{1}{4} = \frac{9}{4} = 2.25.$$

Now $$P(X\geq 10) = \sum_{x=10}^{12}\binom{12}{x}\left(\frac{3}{4}\right)^x\left(\frac{1}{4}\right)^{12-x}$$

$$= \binom{12}{10}\left(\frac{3}{4}\right)^{10}\left(\frac{1}{4}\right)^2+\binom{12}{11}\left(\frac{3}{4}\right)^{11}\left(\frac{1}{4}\right)+\left(\frac{3}{4}\right)^{12}$$

$$= (66)\left(\frac{59049}{1048576}\right)\left(\frac{1}{16}\right) + 12\left(\frac{177147}{4194304}\right)\left(\frac{1}{4}\right) + \left(\frac{531441}{16777216}\right)$$

$$= 0.232293 + 0.126705 + 0.031676$$

$$= 0.3907.$$

8.24. The m.g.f. of the binomial distribution (q+p) ^{n} is

$$M_0(t) = (q + pe^t)^n, \text{ and}$$

the cumulant generating function is

$$\kappa(t) = n \log (q + pe^t), \text{ and}$$

therefore *rth* order cumulant is given by

$$\kappa_r = n \left[\frac{d^r}{dt^r} \log (q + pe^t)\right]_{t=0}$$

Differentiating with respect to p, we get

$$\frac{d\kappa_r}{dp} = n \left[\frac{d^r}{dt^r}\left(\frac{-1+e^t}{q+pe^t}\right)\right]_{t=0} \qquad \because q+pe^t = 1-p+pe^t,$$

Now $$pq\frac{d\kappa_r}{dp} = npq \left[\frac{d^r}{dt^r}\left(\frac{-1+e^t}{q+pe^t}\right)\right]_{t=0} \qquad \therefore \frac{d}{dp} = -1 + e^t$$

Again $$\kappa_{r+1} = n \left[\frac{d^{r+1}}{dt^{r+1}} \log (q + pe^t)\right]_{t=0}$$

$$= n \left[\frac{d^r}{dt^r} \cdot \frac{d}{dt} \log (q + pe^t)\right]_{t=0} = np \left[\frac{d^r}{dt^r}\left(\frac{e^t}{q+pe^t}\right)\right]_{t=0}$$

$$\therefore \quad \kappa_{r+1} - pq\frac{d\kappa_r}{dp} = np \left[\frac{d^r}{dt^r}\left(\frac{e^t}{q+pe^t} - \frac{q(-1+e^t)}{q+pe^t}\right)\right]_{t=0}$$

$$= np \left[\frac{d^r}{dt^r}\left(\frac{q+pe^t}{q+pe^t}\right)\right]_{t=0} = np\left[\frac{d^r}{dt^r}(1)\right]_{t=0} = 0,$$

so that $\kappa_{r+1} = pq\dfrac{d\kappa_r}{dp}$.

Putting $r = 1$, we get $\kappa_2 = pq\dfrac{d\kappa_1}{dp}$, where $\kappa_1 = np$

$$= npq$$

Putting $r = 2$, we get $\kappa_3 = pq\dfrac{d\kappa_2}{dp} = npq\dfrac{d}{dp}(qp)$

$$= npq(1 - 2p) = npq(q - p)$$

$$= n(p - 3p^2 + 2p^3)$$

Putting $r = 3$, we get $\kappa_4 = pq\dfrac{d\kappa_3}{dp} = npq\dfrac{d}{dp}(p - 3p^2 + 2p^3)$

$$= npq(1 - 6p + 6p^2)$$

$$= npq(1 - 6pq).$$

8.26. (b) Let X denote the number of white beads. Then the probability that X assumes a value x, is

$$P(X=x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \text{ for } x \text{ such that } 0 \le x \le n \text{ and } 0 \le x \le k.$$

where k = number of white beads, which is 4;

n = number of beads drawn, which is 5,

N = total number of beads in the bowl which is 11, and

x = 0,1,2,3, and 4. (Here possible values of x are equal to k)

Now we calculate the probabilities for various values of x as below:

$$P(X=0) = \binom{4}{0}\binom{11-4}{5-0} \div \binom{11}{5} = \frac{21}{462},$$

$$P(X=1) = \binom{4}{1}\binom{7}{4} \div \binom{11}{5} = \frac{140}{462},$$

$$P(X=2) = \binom{4}{2}\binom{7}{3} \div \binom{11}{5} = \frac{210}{462},$$

$$P(X=3) = \binom{4}{3}\binom{7}{2} \div \binom{11}{5} = \frac{84}{462},$$

$$P(X=4) = \binom{4}{4}\binom{7}{1} \div \binom{11}{5} = \frac{7}{462},$$

Hence the desired probability distribution is

x	0	1	2	3	4	Total
$P(X=x)$	$\frac{21}{462}$	$\frac{140}{462}$	$\frac{210}{462}$	$\frac{84}{462}$	$\frac{7}{462}$	1

Now $\sum xf(x) = (0)\left(\frac{21}{462}\right) + 1\left(\frac{140}{462}\right) + 2\left(\frac{210}{462}\right) + 3\left(\frac{84}{462}\right) + 4\left(\frac{7}{462}\right)$

$$= 0 + \frac{140}{462} + \frac{420}{462} + \frac{252}{462} + \frac{28}{462} = \frac{840}{462}, \text{ and}$$

$$\sum x^2 f(x) = 0\left(\frac{21}{462}\right) + 1^2\left(\frac{140}{462}\right) + 2^2\left(\frac{210}{462}\right) + 3^2\left(\frac{84}{462}\right) + 4^2\left(\frac{7}{462}\right)$$

$$= 0 + \frac{140}{462} + \frac{840}{462} + \frac{756}{462} + \frac{112}{462} = \frac{1848}{462}.$$

Hence Mean $= \sum x_i f(x_i) = \frac{840}{462} = 1.82$, and

$$\text{Variance} = \sum x_i^2 f(x_i) - \{\sum x_i f(x_i)\}^2 = \frac{1848}{462} - \left(\frac{840}{462}\right)^2$$

$$= 4 - 3.31 = 0.69.$$

The formulas for the mean and variance of the distribution are:

$$\mu = np \text{ and } \sigma^2 = npq\frac{N-n}{N-1}, \text{ where } p = \frac{k}{N} \text{ and } q = \frac{N-k}{N}$$

Here $p = \frac{4}{11}$ and $q = \frac{7}{11}$. Therefore

$$\mu = 5\left(\frac{4}{11}\right) = \frac{20}{11} = 1.82, \text{ and}$$

$$\sigma^2 = 5\left(\frac{4}{11}\right)\left(\frac{7}{11}\right)\frac{11-5}{11-1} = \frac{84}{121} = 0.69.$$

8.27. (a) Here N = 4 men + 2 women = 6 persons,

$n = 3$ persons to be selected,

$k = 4$, and $x = 0, 1, 2, 3$.

Hence the Hypergeometric distribution is

$$h(x; 6, 3, 4) = \frac{\binom{4}{x}\binom{2}{3-x}}{\binom{6}{3}}$$

When $x = 0$, $h(0; 6, 3, 4) = 0$, (the event is impossible)

When $x = 1$, the probability is

$$h(1; 6, 3, 4) = \frac{\binom{4}{1}\binom{2}{2}}{\binom{6}{3}} = \frac{4}{20}$$

When $x = 2$, the probability is

$$h(2, 6, 3, 4) = \frac{\binom{4}{2}\binom{2}{1}}{\binom{6}{3}} = \frac{12}{20}$$

When $x = 3$, the probability is

$$h(3; 6, 3, 4) = \frac{\binom{4}{3}\binom{2}{0}}{\binom{6}{3}} = \frac{4}{20}$$

(b) Here 6 bulbs can be selected in $\binom{8}{6}$ ways,

4 tulip bulbs can be selected in $\binom{4}{4}$ ways, and

2 daffodil bulbs can be selected in $\binom{4}{2}$ ways.

$\therefore$ the total number of ways to select 2 daffodil bulbs and 4 tulip bulbs is $\binom{4}{2}\binom{4}{4}$ ways.

Hence the required probability $= \frac{\binom{4}{2}\binom{4}{4}}{\binom{8}{6}} = \frac{4\times 3}{2} \times \frac{2}{8\times 7} = \frac{3}{14}$

Or Applying Hypergeometric distribution, we identify

$$N = 8,\ n = 6,\ k = 4,\ x = 2.$$

$$\therefore \quad h(2; 8, 6, 4) = \frac{\binom{4}{2}\binom{8-4}{6-2}}{\binom{8}{6}} = \frac{3}{14}.$$

8.28. Here **N = 10 cans to draw from,**

$n = 5$, the number of cans to be drawn.

$k = 5$ of the 10 cans are tomatoes.

Thus the Hypergeometric distribution is

$$h(x; 10, 5, 5) = \frac{\binom{5}{x}\binom{5}{5-x}}{\binom{10}{5}}.$$

Hence the probability that all contain tomatoes, *i.e.* $x = 5$ is

$$h(5; 10, 5, 5) = \frac{\binom{5}{5}\binom{5}{0}}{\binom{10}{5}} = 0.003968.$$

Let A denote the event that 3 or more cans contain tomatoes. Then

$$P(A) = \sum_{x=3}^{5}\binom{5}{x}\binom{5}{5-x} \div \binom{10}{5}$$

$$= \left[\binom{5}{3}\binom{5}{2} + \binom{5}{4}\binom{5}{1} + \binom{5}{5}\binom{5}{0}\right] \div \binom{10}{5}$$

$$= (100 + 25 + 1) \div 252 = 0.5$$

8.29. (a) Let X denote the number of income tax returns with illegitimate deductions. Then we have

$$N = 20,\ n = 6,\ k = 8 \text{ and } x = 3.$$

Hence using the hypergeometric distribution

$$h(x; 20, 6, 8) = \frac{\binom{8}{x}\binom{12}{6-x}}{\binom{20}{6}},$$ **we get the desired probability,**

i.e. $P(X = 3)$, as

$$h(3; 20, 6, 8) = \frac{\binom{8}{3}\binom{12}{3}}{\binom{20}{6}} = \frac{12320}{38760} = \frac{308}{969} = 0.3179.$$

(b) Let X denote the number of narcotic tablets. Then we have

$$k = 6,\ n = 3 \text{ and } N = 6 + 9 = 15.$$

The hypergeometric distribution is then

$$P(X = x) = h(x; 15, 3, 6) = \frac{\binom{6}{x}\binom{9}{3-x}}{\binom{15}{3}}.$$

The traveller will be arrested for illegal possession of narcotics if the customs official detects *at least* 1 narcotic tablet, *i.e.* if $x = 1, 2, 3$.

Hence the desired probability is

$$\sum_{x=1}^{3} h(x; 15,3,6) = \sum_{x=1}^{3}\frac{\binom{6}{x}\binom{9}{3-x}}{\binom{15}{3}}$$

$$= \frac{\binom{6}{1}\binom{9}{2}}{455} + \frac{\binom{6}{2}\binom{9}{1}}{455} + \frac{\binom{6}{3}\binom{9}{0}}{455}$$

$$= \frac{216}{455} + \frac{135}{455} + \frac{20}{455} = \frac{371}{455} = 0.82.$$

8.31. (a) When a sample is obtained under sampling without replacement, the *hypergeometric* distribution is used to find the desired probability. We are given that

$$N = 150,\ n = 4,\ x = 3 \text{ and } k = 20\% \text{ of } 150 = 30.$$

$$\therefore \quad P(X=3) = h(3;150,4,30) = \frac{\binom{30}{3}\binom{120}{1}}{\binom{150}{4}} = 0.0240$$

(b) When a sample is obtained under sampling with replacement from a finite population or sampling from infinite population, the *binomial* probability is appropriate to calculate the desired probability. We are given that $n = 4$, $x = 3$ and $p = 0.20$.

$$\therefore \quad P(X=3) = b(3;4,0.20) = \binom{4}{3}(0.20)^3(0.80) = 0.0256.$$

(c) There is a slight difference because the sample size is small.

8.32. (c) Since X is a Poisson r.v. with $\mu = 1.6$, therefore

$$P(X=0) = e^{-1.6}\frac{(1.6)^0}{0!} = 0.2019, \quad (\because e^{-1.6} = 0.2019)$$

$$P(X=1) = \frac{e^{-1.6}(1.6)}{1!} = 0.3230,$$

$$P(X=2) = \frac{e^{-1.6}(1.6)^2}{2!} = 0.2584, \text{ and}$$

$$\begin{aligned} P(X>2) &= 1 - P(X \le 2) \\ &= 1 - [P(X=0) + P(X=1) + P(X=2)] \\ &= 1 - 0.7833 = 0.2167. \end{aligned}$$

8.33. **(a) The desired probabilities for x = 0, 1, 2, 3 and 4 are computed as below:**

$$P(X=0) = \frac{e^{-3}.3^0}{0!} = e^{-3} = 0.0498,$$

$$P(X=1) = \frac{e^{-3}.3}{1!} = (0.0498)(3) = 0.1494,$$

$$P(X=2) = \frac{e^{-3}.3^2}{2!} = (0.1494)\left(\frac{3}{2}\right) = 0.2241,$$

$$P(X=3) = \frac{e^{-3}.3^3}{3!} = (0.2241)(1) = 0.2241, \text{ and}$$

$$P(X=4) = \frac{e^{-3}.3^4}{4!} = (0.2241)\left(\frac{3}{4}\right) = 0.1681,$$

Note that $e^{-3} = \dfrac{1}{(2.71828)^3} = 0.0498$

(b) Let X have the Poisson distribution with parameter μ. Then

$$P(X=x) = \frac{e^{-\mu}.\mu^x}{x!}, \; x = 0, 1, 2, \ldots, \infty$$

It is given that

$$P(X=1) = e^{-\mu}.\mu = 0.3, \text{ and}$$

$$P(X=2) = \frac{e^{-\mu}.\mu^2}{2!} = 0.2$$

Dividing the second equation by the first, we get

$$\frac{\mu}{2!} = \frac{0.2}{0.3}, \text{ or } \mu = 1.33$$

$$\therefore \quad P(X=0) = e^{-1.33} = 0.2644, \text{ and}$$

$$\begin{aligned} P(X=3) &= \frac{e^{-1.33}(1.33)^3}{3!} \\ &= \frac{(0.2644)(2.352637)}{6} = 0.1037 \end{aligned}$$

To evaluate $e^{-1.33}$, we let $y = e^{-1.33}$

$$\begin{aligned} \text{Then} \quad \log y &= -1.33 \log e = -1.33\,(0.4343) \\ &= -0.5776 = \bar{1}.4223. \end{aligned}$$

$$\therefore \quad y = \text{Antilog}\,(\bar{1}.4223) = 0.2644$$

(c) Let the Poisson distribution be

$$P(X=x) = \frac{\mu^x.e^{-\mu}}{x!} \text{ for } x = 0, 1, 2, \ldots$$

Given that $P(X=2) = 3P(X=4)$

i.e. $\frac{e^{-\mu} . \mu^2}{2!} = 3 . \frac{e^{-\mu} . \mu^4}{4!}$, i.e. $\frac{\mu^2}{2} = \frac{\mu^4}{8}$ giving $\mu = 2$.

(i) $P(X=0) = e^{-2} = 0.1353 = 0.135$

(ii) $P(X \le 3) = \sum_{x=0}^{3} \frac{e^{-2} (2)^x}{x!}$

$$= e^{-2} + \frac{e^{-2} . 2}{1!} + \frac{e^{-2} . 4}{2!} + \frac{e^{-2} . 8}{3!}$$

$$= e^{-2} \left(1 + 2 + 2 + \tfrac{4}{3}\right)$$

$$= 0.1353 \, (6.3333) = 0.857$$

8.34. (b) Here p = 0.03 and n = 500.

$\therefore$ the mean of the Poisson distribution is

$$\mu = np = 500 \, (0.03) = 15.$$

Let X denote the number of defective components.

The Poisson distribution then becomes

$$p(x; 15) = \frac{e^{-15} (15)^x}{x!}, \; x = 0, 1, 2, 3, \ldots$$

(i) $P(X \ge 3) = 1 - P(X < 3)$

$$= 1 - \left\{e^{-15} + e^{-15} (15) + \frac{e^{-15} (15)^2}{2!}\right\}$$

$$= 1 - e^{-15} \{1 + 15 + 112.5\}$$

$$= 1 - (0.00000031) (128.5) \; (\because e^{-15} = 0.00000031)$$

$$= 1 - 0.000060 = 0.999940$$

(ii) $P(X \ge 6) = 1 - P(X < 6)$

$$= 1 - 0.028298 = 0.971702.$$

Hence the probability that 2 successive boxes contain 6 or more defective $= (0.971702)^2 = 0.9442.$

8.35. (b) Here p, the probability of a defective tool = 0.1, so that q = 0.9 and n = 10.

(i) $P(\text{2 defective tools in 10}) = \binom{10}{2} (0.1)^2 (0.9)^9 = 0.1937$

(ii) Let μ be the mean of the Poisson distribution.

Then $\mu = np = 10(0.1) = 1.$

$\therefore$ $P(\text{2 defective tools in 10}) = \frac{\mu^2 . e^{-\mu}}{2!}$

$$= \frac{(1)^2 . e^{-1}}{2!} = \frac{1}{2e}$$

$$= \frac{1}{2(2.718)} = 0.1839$$

8.36. (a) Let X denote the number of claims. Then the Poisson distribution with $\mu = 0.05$, is

$$p(x; 0.05) = \frac{e^{-0.05} (0.05)^x}{x!}, \; x = 0, 1, 2, 3, \ldots$$

Therefore (i) $P(X=0) = e^{-0.05} = 0.9513$

(ii) $P(\text{1 or fewer claims}) = P(X \le 1) = e^{-0.05} + e^{-0.05} (0.05)$

$$= 0.9513 + (0.9513)(0.05)$$

$$= 0.9513 + 0.0476 = 0.9989.$$

(b) Let p be the chance of being killed in an accident.

Then the mean of the Poisson distribution is

$$\mu = np = 350 \times \frac{1}{1400} = \frac{1}{4} = 0.25$$

Let X denote the number of fatal accidents. Then

$P(\text{At least one fatal accident}) = P(X \ge 1)$

$$= 1 - P(X < 1)$$

$$= 1 - e^{-0.05} \quad (X=0)$$

$$= 1 - 0.779 = 0.221$$

8.37. (a) Let X be the r.v. the number of errors per page. Then X has a Poisson distribution with parameter $\mu = 2$.

So $P(X=x) = \frac{e^{-2}(2)^x}{x!}$

Now (i) $P(X \geq 4) = 1 - \sum_{x=0}^{3} \frac{e^{-2}(2)^x}{x!}$, where $e^{-2} = 0.1353$.

$$= 1 - e^{-2}[1 + 2 + 2 + 4/3]$$

$$= 1(0.1353)(6.3333) = 1 - 0.8569 = 0.1431, \text{ and}$$

(ii) $P(X=0) = e^{-2} = 0.1353$.

(b) Given that the number of demands for a car is distributed as a Poisson distribution with mean $\mu = 1.5$.

$\therefore P(X=x) = \frac{e^{-1.5}(1.5)^x}{x!}$, where x denotes the number of demands

Proportion of days on which neither car is used

= probability of there being no demand for the car

$$= \frac{e^{-1.5}(1.5)^0}{0!} = e^{-1.5} = 0.2231$$

Proportion of days on which some demand is refused

= probability for the number of demands to be more than two.

$$= 1 - P(X \leq 2), = 1 - \sum_{x=0}^{2} \frac{e^{-1.5}(1.5)^x}{x!}$$

$$= 1 - \left[e^{-1.5} + e^{-1.5}(1.5) + \frac{e^{-1.5}(1.5)^2}{2!}\right]$$

$$= 1 - (0.2231 + 0.3346 + 0.2510) = 1 - 0.8087 = 0.1913.$$

8.38. Let p be the probability of the product being defective. Then p = 5/100 = 0.05 and n = 100.

The mean of the Poisson distribution, $\mu = np = 0.05 \times 100 = 5$ and the Poisson distribution then becomes

$$p(x; 5) = \frac{e^{-5} \cdot (5)^x}{x!}, \quad x = 0, 1, 2, \ldots \text{ and } e^{-5} = 0.0067$$

We seek the probability that a box will fail to meet the guaranteed quality, *i.e.* the probability of there being more than 4 defective pins in a box of 100 pins. In other words, we are to find $P(X>4)$.

Thus $P(X>4) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^{4} \frac{e^{-5} \cdot (5)^x}{x!}$

$$= 1 - e^{-5}\left[1 + 5 + \frac{(5)^2}{2} + \frac{(5)^3}{6} + \frac{(5)^4}{24}\right]$$

$$= 1 - 0.0067\,[1 + 5 + 12.5 + 20.83 + 26.04]$$

$$= 1 - (0.0067)(65.37)$$

$$= 1 - 0.4380 = 0.5620.$$

8.39. (b) As the mean and variance of the Poisson distribution are equal, therefore $\sigma^2 = \mu = 1$; and the Poisson distribution will be

$$P(X=x) = \frac{e^{-1}(1)^x}{x!}.$$

Hence the desired probability is

$$P(X=2) = \frac{e^{-1}(1)^2}{2!} = \frac{e^{-1}}{2!} = \frac{1}{2e} = \frac{1}{2(2.7183)}$$

$$= \frac{1}{5.4366} = 0.1839.$$

8.40. (a) Let μ be the parameter of the Poisson distribution.

Then mean = μ and $\sigma = \sqrt{\mu}$, *i.e.* $\sigma = \sqrt{\text{mean}}$

In the given statement, mean = 5 and σ = 4

$\therefore \quad \sqrt{\text{mean}}$, *i.e.* $\sqrt{5} \neq \sigma$.

Hence the given statement is wrong.

(b) Let μ be the parameter of the Poisson distribution and N be the total number of observations. Then the Poisson distribution is

$$\frac{N \cdot e^{-\mu} \cdot \mu^x}{x!}, \; x = 0, 1, 2, \ldots, \infty$$

It is given that

$$N \cdot e^{-\mu} = 250, \text{ and } N \cdot e^{-\mu} \cdot \mu = 160.$$

Dividing the second equation by the first, we get

$$\mu = \frac{16}{25} = 0.64$$

$$\therefore \quad e^{-0.64} = 1 - (0.64) + \frac{(0.64)^2}{2} - \frac{(0.64)^3}{3!} + \frac{(0.64)^4}{4!} - \ldots$$

$$= 0.5273$$

The frequencies of the next two values are

$$N \cdot e^{-\mu} \cdot \frac{\mu^2}{2!} \text{ and } Ne^{-\mu} \cdot \frac{\mu^3}{3!}$$

Now $N \cdot e^{-\mu} \cdot \frac{\mu^2}{2!} = (Ne^{-\mu} \cdot \mu)\left(\frac{\mu}{2}\right) = 160 \times \frac{0.64}{2} = 51$, and

$$Ne^{-\mu} \cdot \frac{\mu^3}{3!} = (Ne^{-\mu} \cdot \frac{\mu^2}{2!})\left(\frac{\mu}{3}\right) = 51 \times \frac{0.64}{3} = 11.$$

8.44. Fitting of the Poisson distribution.

We first calculate the mean of the given distribution and equate it to μ, the parameter of the Poisson distribution $p(x; \mu)$ $= \frac{e^{-\mu} \cdot \mu^x}{x!}$, where x denotes the yeast cell counts.

Now $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0 + 128 + 74 + 54 + 12 + 5}{400} = \frac{273}{400} = 0.6825$

Therefore μ = 0.6825 and the fitted Poisson distribution is

$$p(x; 0.6852) = \frac{e^{-.6825} (0.6825)^x}{x!}, \; x = 0, 1, 2, 3, \ldots$$

The expected frequencies are obtained by multiplying the probabilities by 400. Thus the expected probabilities and frequencies are calculated as below:

x	Probability $p(x; 0.6825)$		Expected frequencies $400 \times p$
0	$e^{-0.6825}$	= 0.5054	202.16
1	$e^{-0.6825}$ (0.6825)	= 0.3449	137.96
2	$\frac{e^{-0.6825} (0.6825)^2}{2!}$	= 0.1177	47.08
3	$\frac{e^{-0.6825} (0.6825)^3}{3!}$	= 0.0268	10.72
4	$\frac{e^{-0.6825} (0.6825)^4}{4!}$	= 0.0046	1.84
5	$\frac{e^{-0.6825} (0.6825)^5}{5!}$	= 0.0006	0.24
Total		1.0000	400

8.45. (a) We first calculate the mean of the given frequency distribution and equate it to (i) μ and (ii) np.

Now $\bar{x} = \frac{\sum fx}{\sum f} = \frac{(0)(531) + (1)(354) + (2)(99) + (3)(15) + (4)(1)}{1000}$

$$= \frac{601}{1000} = 0.60$$

(i) Therefore μ = 0.60 and the fitted Poisson distribution is

$$P(X = x) = \frac{e^{-0.60} (0.60)^x}{x!} \text{ for } x = 0, 1, 2, 3, \ldots$$

(ii) **Equating to** np, *i.e.* $4p$, **we get** $p = \frac{0.60}{4} = 0.15$. Thus the **fitted binomial distribution is**

$$P(X = x) = \binom{4}{x} (0.15)^x (0.85)^{4-x}, \text{ for } x = 0, 1, 2, 3, 4.$$

The expected frequencies are obtained by multiplying the probabilities by 1000. The probabilities and the expected frequencies are given below:

x	(i) Poisson Distribution		(ii) Binomial Distribution	
	$P(X=x)$	Expected f	$P(X=x)$	Expected f
0	0.5488	549	0.5220	522
1	0.3293	329	0.3685	.368
2	0.0988	99	0.0975	98
3	0.0198	20	0.0115	12
4	0.0030	3	0.0005	0
>4	0.0003	0	--	--

(b) The mean of the given distribution is

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0 + 96 + 68 + 27 + 4 + 0}{440}$$

$$= \frac{195}{440} = 0.443 = \mu$$

Thus the fitted Poisson distribution is

$$p(x; 0.443) = \frac{e^{-0.443}\,(0.443)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

To find the value of $e^{-0.443}$, we let $y = e^{-0.443}$

Taking logs, we get

$$\log y = -0.443 \log e, \text{ where } e = 2.7183$$

$$= -0.443\,(0.4343) = -0.1924 = \bar{1}.8076$$

$$y = \text{Anti-log}\,(\bar{1}.8076) = 0.6421$$

(i) $P(\text{no accident}) = P(X=0) = e^{-0.443} = 0.6421$

(ii) $P(\text{more than one accident}) = P(X>1)$

$$= 1 - \sum_{x=0}^{1} \frac{e^{-0.443}\,(0.443)^x}{x!}$$

$$= 1 - [0.6421 + 0.2845]$$

$$= 1 - 0.9266 = 0.0734$$

8.46. Computation of the frequencies of the fitted Poisson distribution.

Here $\bar{x} = \dfrac{\sum f_i x_i}{\sum f_i} = \dfrac{0+90+84+36+36+15+6}{300} = \dfrac{267}{300} = 0.89$

Thus the fitted Poisson distribution is

$$p(x; 0.89) = \frac{e^{-0.89}\,(0.89)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

To find the value of $e^{-0.89}$, we put $y = e^{-0.89}$.

Taking logs, we get

$$\log y = -0.89 \log e = -0.89\,(0.4343) \quad (\text{where } e = 2.718)$$

$$= -0.3865 = \bar{1}.6135, \quad \therefore y = \text{Anti-log}\,(\bar{1}.6135) = 0.4107$$

Hence the expected frequencies are computed as below:

x	Probabilities	Expected f
0	$e^{-0.89} = 0.4107$	123.2
1	$e^{-0.89}\frac{(0.89)}{1!} = \frac{0.4107\times0.89}{1} = 0.3655$	109.6
2	$e^{-0.89}\frac{(0.89)^2}{2!} = \frac{0.3655\times0.89}{2} = 0.1626$	48.8
3	$e^{-0.89}\frac{(0.89)^3}{3!} = \frac{0.1626\times0.89}{3} = 0.0482$	14.5
4	$e^{-0.89}\frac{(0.89)^4}{4!} = \frac{0.0482\times0.89}{4} = 0.0107$	3.2
5	$e^{-0.89}\frac{(0.89)^5}{5!} = \frac{0.0107\times0.89}{5} = 0.0019$	0.6
6	$e^{-0.89}\frac{(0.89)^6}{6!} = \frac{0.0019\times0.89}{6} = 0.0003$	0.1

8.47. The mean number of accidents a day is

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0+113+128+63+28+15+6+7}{300} = \frac{360}{300} = 1.2$$

Thus the fitted Poisson distribution is

$$p(x;\,1.2) = \frac{e^{-1.2}\,(1.2)^x}{x!}.$$

To evaluate $e^{-1.2}$, we let $y = e^{-1.2}$

Then $\log y = -1.2 \log e$

$= -1.2\,(0.4343) = -0.52116 = \bar{1}.47884$

$\therefore \quad y = \text{Anti-log}\,(\bar{1}.47884) = 0.3011.$

The expected frequencies are calculated as below:

Accidents per day (x)	Probabilities $p(x;\,1.2)$		Expected f $300\,f(x)$
0	$e^{-1.2}$	$= 0.3011$	90.3
1	$e^{-1.2}\cdot\frac{(1.2)}{1!} = (0.3011)(1.2)$	$= 0.3613$	108.4
2	$e^{-1.2}\cdot\frac{(1.2)^2}{1!} = (0.3613)(0.6)$	$= 0.2168$	65.0
3	$e^{-1.2}\cdot\frac{(1.2)^3}{3!} = (0.2168)(0.4)$	$= 0.0867$	26.0
4	$e^{-1.2}\cdot\frac{(1.2)^4}{4!} = (0.0867)(0.3)$	$= 0.0260$	7.8
5	$e^{-1.2}\cdot\frac{(1.2)^5}{5!} = (0.0260)(0.24)$	$= 0.0062$	1.9
6	$e^{-1.2}\cdot\frac{(1.2)^6}{6!} = (0.0062)(0.2)$	$= 0.0012$	0.4
7+	$1 - \sum_{x=0}^{6} \frac{e^{-1.2}(1.2)^x}{x!}$	$= 0.0007$	0.2
Total		$= 1.0000$	300.0

And $P(4 \text{ or more accidents}) = 1 - P\,(\text{less than 4 accidents})$

$= 1 - (0.3011 + 0.3613 + 0.2168 + 0.0867)$

$= 1 - 0.9659 = 0.0341.$

8.48. (b) Let X denote the number of customers entering the shop in t units of time.

Then we have

$$\lambda = 30 \text{ persons per hour} = 30 \text{ persons per 60 minutes}$$

As 2 minutes is $\frac{2}{60} = \frac{1}{30}$ units of time, so $t = \frac{1}{30}$, and therefore the average number of persons per 2 minutes interval, *i.e.* $\lambda t = 30 \times \frac{1}{30} = 1.$

Hence, using the Poisson process formula

$$p(x;\,\lambda t) = \frac{e^{-\lambda t}\,(\lambda t)^x}{x!}, \text{ we get}$$

$$p(0;\,1) = \frac{e^{-1}\cdot(1)^0}{0!} = e^{-1} = \frac{1}{2.7183} = 0.3679.$$

8.49. (a) Taking 50 square feet as the unit of area, we have

$$\lambda = 1 \text{ flaw per 50 square feet.}$$

As 32 square feet are $\frac{32}{50} = 0.64$ units of area, so $t = 0.64$, and therefore the average number of flaws per 32 square feet, *i.e.* $\lambda t = 1 \times 0.64 = 0.64$. Assuming the flaws are a Poisson process, we have

$$P(\text{no flaw occurs in 32 square feet}) = \frac{e^{-0.64}(0.64)^0}{0!} = 0.5272$$

$$P(\text{atmost one flaw occurs in 32 square feet}) = \sum_{x=0}^{1} \frac{e^{-0.64}(0.64)^x}{x!}$$

$= e^{-0.64} + e^{-0.64}\,(0.64)$

$= 0.5272 + 0.3374$

$= 0.8646$

(b) Taking 12 hours from 9 p.m. until 9 a.m. the next morning as the unit of time, we have $\lambda = 3$ calls per 12 hours

As 6 hours from midnight to 6 a.m. are $\frac{6}{12} = \frac{1}{2}$ units of time, so $t = \frac{1}{2}$ and therefore the average number of calls per 6 hours, *i.e.* $\lambda t = 3 \times \frac{1}{2} = 1.5$.

As the arrivals of calls are assumed to be Poisson process, therefore we have

$$P(\text{doctor receives no call in 6 hours}) = \frac{e^{-1.5}(1.5)^0}{0!}$$
$$= e^{-1.5} = 0.2231$$

8.50. Taking 25 days as unit of time, we have

$$\lambda = 1 \text{ breakdown in 25 days.}$$

As 10 days is $\frac{10}{25} = 0.4$ unit of time, so $t = 0.4$ and therefore the average number of breakdowns per 10 days, *i.e.* $\lambda t = 1 \times 0.4 = 0.4$.

Hence, using the Poisson process formula $p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$, we have (i) $P(\text{exactly one breakdown in 10 days}) = e^{-0.4}\frac{(0.4)^1}{1!}$

$$= (0.6704)(0.4)$$
$$= 0.2682$$

(ii) $P(\text{more than one breakdown in 10 days})$

$$= 1 - P(0 \text{ or } 1 \text{ breakdown}) = 1 - \sum_{x=0}^{1} \frac{e^{-0.4}(0.4)^x}{x!}$$
$$= 1 - [0.6704 + 0.2682] = 1 - 0.9386 = 0.0614.$$

8.51. Taking 60 minutes, (10 to 11 A.M.) as unit of time, we have

$$\lambda = 300 \text{ cars passing in 60 minutes.}$$

As 1 minute is $\frac{1}{60}$ unit of time, so $t = \frac{1}{60}$ and therefore the average number of cars passing per minute, *i.e.* $\lambda t = 300 \times \frac{1}{60} = 5$

Hence, using the Poisson process formula $p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$, we have (i) $P(\text{not more than 4 cars in 1-minute}) = \sum_{x=0}^{4} \frac{e^{-5} \cdot (5)^x}{x!}$

$$= e^{-5}\left[1 + 5 + \frac{(5)^2}{2} + \frac{(5)^3}{6} + \frac{(5)^4}{24}\right]$$
$$= (0.0067)[1 + 5 + 12.5 + 20.83 + 26.04]$$
$$= (0.0067)(65.37) \quad (\because e^{-5} = 0.0067)$$
$$= 0.4380$$

(ii) $P(5 \text{ or more cars pass in 1-minute})$

$$= 1 - P(\text{less than 5 cars pass})$$
$$= 1 - \sum_{x=0}^{4} \frac{e^{-5} \cdot (5)^x}{x!} = 1 - 0.4380 = 0.5620.$$

8.53. (b) Let N be equal to number of eggs laid by an insect. Then N is assumed to have a Poisson distribution with mean m.

It is given that the probability of an egg developing is p.

If X denotes the number of eggs that develop, then for given $N = n$, the variate X has a binomial distribution based on n repetitions with parameter p. That is

$$P(X = k \mid N = n) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

Hence according to the theorem of total probability, we have

$$P(X = k) = \sum_{n=k}^{\infty} P(X = k \mid N = n) \cdot P(N = n)$$
$$= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \cdot \frac{e^{-m} \cdot m^n}{n!}$$

$$= \frac{e^{-m} p^k}{k!} \sum_{n=k}^{\infty} \frac{m^n q^{n-k}}{(n-k!)} = \frac{e^{-m} . p^k m^k}{k!} \sum_{n=k}^{\infty} \frac{(mq)^{n-k}}{(n-k)!}$$

$$= \frac{e^{-m} (mp)^k}{k!} \sum_{j=0}^{\infty} \frac{(mq)^j}{j!} \quad \text{where } j = n-k$$

$$= \frac{e^{-m} (mp)^k}{k!} . e^{mq} = \frac{e^{-m} (mp)^k . e^{m(1-p)}}{k!} = \frac{e^{-mp} . (mp)^k}{k!},$$

which is a Poisson distribution with mean mp.

8.54. (b) We observe that

(i) there are two possible outcomes, *i.e.* the swimmer will *or* will not cross the lake,

(ii) probability of success in swimming across the lake for cach swimmer is $p = 0.4$,

(iii) the swimmers will cross the lake independently, and

(iv) the experiment is repeated 10 times to get fourth success.

Therefore the negative binomial distribution with $k = 4$(4th success) and $x = 10$, is appropriate. Hence

$$b^*(10; 4, 0.4) = \binom{10-1}{4-1} (0.4)^4 (0.6)^{10-4}$$

$$= 84\,(0.4)^4 (0.6)^6 = 0.10033$$

8.55. (b) Here E(X) = 10, and $\sigma^2 = 9$.

Since mean is greater than variance, so X cannot have a negative binomial distribution.

(c) Here p, the probability of getting a head = $\frac{1}{2}$,

$k = 3$ (third success), and $x = 7$ (seventh trial).

$$\therefore \quad b^*(7; 3, \tfrac{1}{2}) = \binom{7-1}{3-1} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{7-3}$$

$$= 15 . \left(\frac{1}{2}\right)^7 = \frac{15}{128} = 0.1172.$$

8.56. (b) Here p, the probability of getting installed a black telephone = 0.3.

$k = 5$ (5th success), and $x = 10$ (10th telephone)

$$\therefore \quad b^*(10; 5, 0.3) = \binom{10-1}{5-1} (0.3)^5 (0.7)^{10-5}$$

$$= 126\,(0.3)^5 (0.7)^5 = 0.0515$$

8.57. (b) The Negative Binomial distribution gives the probability that the kth success occurs on the xth trial for x = k, k+1, k+2, ... Its probability function is

$$p(x) = \binom{x-1}{k-1} p^k q^{x-k}, \text{ where } x = k, k+1, k+2, \ldots$$

It has two parameters k and $p > 0$. To show that the probabilities add to 1, we proceed as below:

$$\sum p(x) = \sum_{x=k}^{\infty} \binom{x-1}{k-1} p^k q^{x-k}, \quad (x = k, k+1, k+2, \ldots)$$

Let $y = x - k$. Then

$$\text{Sum of prob} = \sum_{y=0}^{\infty} \binom{y+k-1}{k-1} p^k q^y, \quad (y = 0, 1, 2, \ldots)$$

$$= p^k \sum_{y=0}^{\infty} \binom{y+k-1}{k-1} q^y$$

$$= p^k \left[\binom{k-1}{k-1} q^0 + \binom{k}{k-1} q + \binom{k+1}{k-1} q^2 + \ldots\right]$$

$$= p^k \left[1 + kq + \frac{k(k+1)}{2!} q^2 + \ldots\right]$$

$$= p^k [1-q]^{-k} = p^k p^{-k} = 1.$$

The *m.g.f.* about the origin is

$$M_0(t) = E(e^{tY}) = p^k \sum_{y=0}^{\infty} \binom{y+k-1}{k-1} q^y . e^{ty}$$

$$= p^k \sum_{y=0}^{\infty} \binom{y+k-1}{k-1} (qe^t)^y = p^k(1 - qe^t)^{-k}$$

$$= \left[\frac{p}{1 - qe^t}\right]^k = \left[\frac{p}{p + q - qe^t}\right]^k$$

$$= \left[\frac{1}{1 - \frac{q}{p}(e^t - 1)}\right]^k$$

$$= [1 - \lambda (e^t - 1)]^{-k}, \quad \text{where } \lambda = q/p.$$

Taking logs, we have

$$\kappa(t) = \log M_0(t) = -k \log [1 - \lambda(e^t - 1)]$$

$$= k\left[\lambda(e^t-1) + \tfrac{1}{2}\lambda^2(e^t-1)^2 + \tfrac{1}{3}\lambda^3(e^t-1)^3 + \tfrac{1}{4}\lambda^4(e^t-1)^4 + \ldots\right]$$

$$= k\left[\lambda\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \ldots\right) + \tfrac{1}{2}\lambda^2 t^2\left(1 + \frac{t}{2!} + \frac{t^2}{3!} + \frac{t^3}{4!} + \ldots\right)^2 \right.$$

$$\left. + \tfrac{1}{3}\lambda^3 t^3\left(1 + \frac{t}{2!} + \ldots\right)^3 + \frac{1}{4}\lambda^4 t^4 + \ldots\right]$$

$$= k\left[\lambda t + (\lambda + \lambda^2)\frac{t^2}{2!} + (\lambda + 3\lambda^2 + 2\lambda^3)\frac{t^3}{3!} \right.$$

$$\left. + (\lambda + 7\lambda^2 + 12\lambda^3 + 6\lambda^4)\frac{t^4}{4!} + \ldots\right]$$

$$= k\left[\frac{q}{p}t + \frac{q}{p^2}\cdot\frac{t^2}{2!} + \frac{q(1+q)}{p^3}\cdot\frac{t^3}{3!} + \frac{q(1+4q+q^2)}{p^4}\cdot\frac{t^4}{4!} + \ldots\right]$$

dentifying co-efficients, we get

$$\kappa_1 = \frac{kq}{p}, \; \kappa_2 = \frac{kq}{p^2}, \; \kappa_3 = \frac{kq(1+q)}{p^3}, \; \text{and}$$

$$\kappa_4 = \frac{kq(1 + 4q + q^2)}{p^4}.$$

8.58. (c) Let X denote the number of trials required o get the occurrence of 1st head.

The head occurs on the third trial, *i.e.* the first success ccurs on the 3rd trial, therefore the geometric distribution with =1/2 and $x=3$ is used.

Hence, using the geometric distribution $P(X=x) = pq^{x-1}$, we get

$$P(X=3) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 = \frac{1}{8}.$$

8.59. (b) Sample space S associated with this experiment contains $\binom{52}{13}$ sample points.

Let A be the event that a bridge hand contains 5 spades, 2 hearts, 3 diamonds and 3 clubs.

Then A contains $\binom{13}{5}\cdot\binom{13}{2}\cdot\binom{13}{3}\cdot\binom{13}{3}$ sample points.

$$\text{Hence } P(A) = \frac{\binom{13}{5}\binom{13}{2}\binom{13}{3}\binom{13}{3}}{\binom{52}{13}} = \frac{(1287)\,(78)\,(286)\,(286)}{635{,}013{,}559{,}600}$$

$$= 0.0129.$$

This question actually relates to the extension of the hypergeometric distribution where the population is partitioned into 4 cells.

8.60. (a) Here $p_1 = P(\text{bulb is red}) = 0.50,$

$p_2 = P(\text{bulb is blue}) = 0.30,$

$p_3 = P(\text{bulb is green}) = 0.20,$ and

$x_1 = 2, x_2 = 1$ and $x_3 = 2.$

$\therefore P(\text{2 are red, 1 is green, 2 are blue}).$

$$= \frac{5!}{2!\,1!\,2!}(0.5)^2(0.3)(0.2)^2$$

$$= \frac{30 \times 3 \times 4}{4 \times 10 \times 100} = \frac{9}{100} = 0.09.$$

(b) Here $p_1 = P(\text{red marble is drawn}) = \frac{5}{10},$

$p_2 = P(\text{white marble is drawn}) = \frac{3}{10},$

$p_3 = P(\text{blue marble is drawn}) = \frac{2}{10}$, and

(i) $x_1 = 3, x_2 = 2$ and $x_3 = 1$.

Thus P(3 are red, 2 are white, 1 is blue).

$$= \frac{6!}{3!\,2!\,1!}\left(\frac{5}{10}\right)^3\left(\frac{3}{10}\right)^2\left(\frac{2}{10}\right)^1$$

$$= \frac{60 \times 1 \times 9 \times 2}{8 \times 100 \times 10} = \frac{27}{200} = 0.135.$$

(ii) $x_1 = 2, x_2 = 3$ and $x_3 = 1$.

Thus P(2 are red, 3 are white, 1 is blue).

$$= \frac{6!}{2!\,3!\,1!}\left(\frac{5}{10}\right)^2\left(\frac{3}{10}\right)^3\left(\frac{2}{10}\right)^1$$

$$= \frac{60 \times 27 \times 2}{4 \times 1000 \times 10} = \frac{81}{1000} = 0.081.$$

(iii) $x_1 = 2, x_2 = 2$ and $x_3 = 2$.

Thus P(2 of each colour appear)

$$= \frac{6!}{2!\,2!\,2!}\left(\frac{5}{10}\right)^2\left(\frac{3}{10}\right)^2\left(\frac{2}{10}\right)^2$$

$$= \frac{90 \times 9 \times 4}{4 \times 100 \times 100} = \frac{81}{1000} = 0.081.$$

8.61

Let X = No. of calls received

i) Here, $\mu = 4.9$

$$P(X = x) = \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-4.9}(4.9)^x}{x!}, x = 0,1,2,\ldots$$

$$P(x = 6) = \frac{e^{-4.9}(4.9)^6}{6!} = 0.1432$$

ii) Here, $\lambda = 4.9$ per half hour

and $t = 2, \lambda = 4.9 \times 2 = 9.8$ calls per hour

thus,

$$P(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-9.8}(9.8)^x}{x!}, \quad x = 0,1,2,\ldots$$

and

$$P(X = 6) = \frac{e^{-9.8}(9.8)^6}{6!} = 0.0682$$

(c) **See solution of Q. No. 8.60(*a*).**

8.62

Here $p = P(\text{Winning the prize}) = \frac{1}{10} = 0.10$

$q = 1 - 01 = 0.90, n = 7$

Let X = No. of days the child wins prizes,

then, Binomial p.d of r.v X is,

$$p(X = x) = {}^7c_x (0.1)^x (0.9)^{7-x}, \quad x = 0, 1, \ldots, 7$$

i) $P(X \geq 1) = 1 - P(X = 0)$

$= 1 - {}^7c_0 (0.1)^0 (0.9)^{7-0} = 1 - 0.4783 = 0.5217$

ii) P (First two days) = P(W). P(W). P(L). P(L). P(L). P(L). P(L)

$= (0.1)(0.1)(0.9)(0.9)(0.9)(0.9)(0.9)$

$= (0.1)^2 (0.9)^5 = 0.0059$

iii) $P(X = 7) = {}^7c_7 (0.1)^7 (0.9)^{7-7}$

$= 0.0000001$

●●●●●●●●●●

Chapter 9

CONTINUOUS PROBABILITY DISTRIBUTIONS

9.1. (b) The given rectangular distribution is

$$f(x) = 1, \ -\frac{1}{2} \le x \le \frac{1}{2},$$

$$= 0, \text{ elsewhere.}$$

The *m.g.f.* is

$$M_0(t) = E(e^{Xt})$$

$$= \int_{-1/2}^{1/2} e^{xt}\,dx = \left[\frac{1}{t}\cdot e^{xt}\right]_{-1/2}^{1/2}$$

$$= \frac{1}{t}\left[e^{t/2} - e^{-t/2}\right]$$

$$= \frac{1}{t}\left[\left(1 + \frac{t}{2} + \frac{1}{2!}\cdot\frac{t^2}{4} + \frac{1}{3!}\cdot\frac{t^3}{8} + \frac{1}{4!}\cdot\frac{t^4}{16} + \ldots\right) - \left(1 - \frac{t}{2} + \frac{1}{2!}\cdot\frac{t^2}{4} - \frac{1}{3!}\cdot\frac{t^3}{8} + \frac{1}{4!}\cdot\frac{t^4}{16} - \ldots\right)\right]$$

$$= \frac{2}{t}\left[\frac{t}{2} + \frac{1}{3!}\cdot\frac{t^3}{8} + \frac{1}{5!}\cdot\frac{t^5}{32} + \ldots\right]$$

$$= 1 + \frac{1}{3!}\cdot\frac{t^2}{4} + \frac{1}{5!}\cdot\frac{t^4}{16} + \ldots$$

Now $\mu_r' =$ co-efficient of $\frac{t^r}{r!}$ in the expansion of $M_0(t)$

$$\therefore \quad \mu_1' = 0, \quad \mu_2' = \frac{1}{12}, \quad \mu_3' = 0, \quad \mu_4' = \frac{1}{80}.$$

Hence the mean moments are,

$$\mu_1 = 0, \quad \mu_2 = \frac{1}{12}, \quad \mu_3 = 0 \text{ and } \mu_4 = \frac{1}{80}$$

9.2. (b) Here the mean, i.e. $\frac{1}{\lambda} = 12$, so that $\lambda = \frac{1}{12}$.

Then the probability that the length of life is equal to or greater than 18 months is given by

$$P(X \ge 18) = \int_{18}^{\infty}\left(\frac{1}{12}\right) e^{-x/12}\,dx$$

$$= \left[-e^{-x/12}\right]_{18}^{\infty} = e^{-1.5} = 0.2231$$

9.3. (a) The exponential distribution is given by

$$f(x) = \frac{1}{2}e^{-x/2}, \quad 0 \le x < \infty.$$

Now $\mu = E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_0^{\infty} x\cdot\frac{1}{2}\cdot e^{-x/2}\,dx$

$$= \left[-xe^{-x/2}\right]_0^{\infty} + \int_0^{\infty} e^{-x/2}\,dx \text{ (Integrating by parts)}$$

$$= 0 + \left[\frac{-e^{-x/2}}{1/2}\right]_0^{\infty} = 2$$

Again, $E(X^2) = \int_0^{\infty} x^2\cdot\frac{1}{2}e^{-x/2}\,dx$

$$= \left[-x^2e^{-x/2}\right]_0^{\infty} + \int_0^{\infty} 2x\cdot e^{-x/2}\,dx$$

(Integrating by parts)

$$= 0 + \frac{2E(X)}{1/2} = 4(2) = 8.$$

$$\therefore \text{Var}(X) = E(X^2) - [E(X)]^2 = 8 - (2)^2 = 4.$$

The *m.g.f.* is $M_0(t) = \int_0^{\infty} e^{tx}\left(\frac{1}{2}\right)e^{-x/2}\,dx$

$$= \frac{1}{2}\int_0^\infty e^{-(1/2-t)x}\,dx$$

$$= \frac{1}{2}\left[\frac{-e^{-(1/2-t)x}}{\frac{1}{2}-t}\right]_0^\infty = \frac{\frac{1}{2}}{\frac{1}{2}-t} = \frac{1}{1-2t}$$

$$P(X>3) = \int_3^\infty \frac{1}{2}e^{-x/2}\,dx$$

$$= \left[-e^{-x/2}\right]_3^\infty = e^{-1.5} = 0.2231.$$

Now $P(X>5 \mid X>2) = \dfrac{P(X>5)}{P(X>2)}$, where

$$P(X>5) = \int_5^\infty \left(\frac{1}{2}\right)e^{-x/2}\,dx$$

$$= \left[-e^{-x/2}\right]_5^\infty = e^{-2.5} = 0.0821, \text{ and}$$

$$P(X>2) = \int_2^\infty \left(\frac{1}{2}\right)e^{-x/2}\,dx$$

$$= \left[-e^{-x/2}\right]_2^\infty = e^{-1} = 0.3679.$$

Hence $P(X>5 \mid X>2) = \dfrac{P(X>5)}{P(X>2)} = \dfrac{0.0821}{0.3679} = 0.2232,$.

(b) Let the r.v.X denote the distance in kilometers travelled by customers. Then

$$P(X\le 1) = \frac{1}{5}\int_0^1 e^{-x/5}\,dx = \left[-e^{-x/5}\right]_0^1$$

$$= \left[-e^{-1/5}+1\right] = 1-0.8187 \quad (\because e^{-0.2}=0.8187)$$

$$= 0.1813,$$

$$P(X\ge 15) = \frac{1}{5}\int_{15}^\infty e^{-x/5}\,dx = \left[-e^{-x/5}\right]_{15}^\infty$$

$$= \left[0+e^{-15/5}\right] = e^{-3} = 0.0498$$

Thus the required proportions of customers are 18.13% and 4.98% respectively.

9.4. (a) The expected value of X is

$$E(X) = \int_0^\infty x\,.\,\alpha e^{-\alpha x}\,dx$$

Let $\alpha e^{-\alpha x}\,dx = dv$ and $x = u$ so that $v = -e^{-\alpha x}$ and $du = dx$. Then integrating by parts, we have

$$E(X) = \left[-xe^{-\alpha x}\right]_0^\infty + \int_0^\infty e^{-\alpha x}\,dx = \left[\frac{-e^{-\alpha x}}{\alpha}\right]_0^\infty = \frac{1}{\alpha}.$$

The expected value of X^2 is

$$E(X^2) = \int_0^\infty x^2\,.\,\alpha e^{-\alpha x}\,dx$$

$$= \left[-x^2e^{-\alpha x}\right]_0^\infty + 2\int_0^\infty xe^{-\alpha x}\,dx \text{ (Integrating by parts)}$$

$$= \frac{2}{\alpha}\left[-xe^{-\alpha x}\right]_0^\infty + \frac{2}{\alpha}\int_0^\infty e^{-\alpha x}\,dx = \frac{2}{\alpha^2}$$

$$\therefore \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2}{\alpha^2} - \left(\frac{1}{\alpha}\right)^2 = \frac{1}{\alpha^2}$$

Hence $\sigma = \dfrac{1}{\alpha}$.

(b) The m.g.f. of the distribution given is obtained as

$$M_0(t) = \frac{1}{\alpha}\int_0^\infty e^{tx}\,.\,e^{-x/\alpha}\,dx$$

$$= \frac{1}{\alpha}\int_0^\infty e^{-x(1/\alpha-t)}\,dx = \frac{1}{\alpha}\left[\frac{-e^{-x(1/\alpha-t)}}{\frac{1}{\alpha}-t}\right]_0^\infty$$

$$= \frac{1}{\alpha}\cdot\frac{\alpha}{1-\alpha t} = \frac{1}{1-\alpha t} = (1-\alpha t)^{-1}$$

$$= 1 + \alpha t + \alpha^2t^2 + \alpha^3t^3 + \alpha^4t^4 + \ldots$$

Now μ'_r is the co-efficient of $\frac{t^r}{r!}$, therfore we get

$$\mu'_1 = \alpha,\ \mu'_2 = 2\alpha^2,\ \mu'_3 = 6\alpha^3 \text{ and } \mu'_4 = 24\alpha^4.$$

Hence the moments about the mean are obtained below:

$$\mu_1 = 0;$$

$$\mu_2 = \mu'_2 - \mu'^2_1 = 2\alpha^2 - (\alpha)^2 = \alpha^2;$$

$$\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2\mu'^3_1 = 6\alpha^3 - 3\alpha(2\alpha^2) + 2\alpha^3 = 2\alpha^3, \text{ and}$$

$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1$$

$$= 24\alpha^4 - 4\alpha(6\alpha^3) + 6\alpha^2(2\alpha^2) - 3\alpha^4 = 9\alpha^4.$$

9.5. The m.g.f. is obtained first as below:

$$M_0(t) = \int_0^\infty e^{tx} \cdot xe^{-x}\, dx = \int_0^\infty x \cdot e^{-x(1-t)}\, dx$$

Let $u = x(1-t)$ so that $dx = \frac{du}{1-t}$. Then

$$M_0(t) = \int_0^\infty \frac{u}{1-t} e^{-u} \frac{du}{1-t} = \frac{1}{(1-t)^2} \int_0^\infty u \cdot e^{-u}\, du$$

$$= \frac{1}{(1-t)^2}\Gamma(2) \qquad \text{(by Gamma function)}$$

$$= (1-t)^{-2}, \quad t<1.$$

Thus $M_0(t) = 1 + 2t + 3t^2 + 4t^3 + 5t^4 + \ldots$

$$= 1 + 2(t) + 3 \times 2!\, \frac{t^2}{2!} + 4 \times 3!\, \frac{t^3}{3!} + 5 \times 4!\, \frac{t^4}{4!} + \ldots$$

Now μ'_r is the co-efficient of $\frac{t^r}{r!}$ in the expansion of $M_0(t)$.

Thus $\quad \mu'_1 = 2,\ \mu'_2 = 6,\ \mu'_3 = 24,\ \mu'_4 = 120$

Hence the moments about the mean are

$$\mu_1 = 0;\ \mu_2 = \mu'_2 - \mu'^2_1 = 2;$$

$$\mu_3 = \mu'_3 - 3\mu'_1\mu'_2 + 2\mu'^3_1 = 4; \text{ and}$$

$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1 = 24.$$

9.6. First of all, we find the value of y_0 which should be such as to make

$$\int_0^\infty y_0\, e^{-x/\sigma}\, dx = 1 \text{ or } y_0 \left[-\sigma\, e^{-x/\sigma}\right]_0^\infty = 1 \text{ or } y_0\sigma = 1$$

$\therefore \quad y_0 = \frac{1}{\sigma}$

Moments about the origin are obtained below;

$$\mu'_r = \frac{1}{\sigma}\int_0^\infty x^r\, e^{-x/\sigma}\, dx$$

Let $\quad u = \frac{x}{\sigma}$ so that $dx = \sigma du$. Then

$$\mu'_r = \sigma^r \int_0^\infty u^r\, e^{-u}\, du$$

$$= \sigma^r\, \Gamma(r+1) \qquad \text{(by Gamma function)}$$

$$= r!\, \sigma^r$$

$\therefore \quad \mu'_1 = \sigma,\ \mu'_2 = 2\sigma^2,\ \mu'_3 = 6\sigma^3$ and $\mu'_4 = 24\sigma^4$,

Now mean $= \mu'_1 = \sigma$, and

$$S.D. = \sqrt{\mu'_2 - \mu'^2_1} = \sqrt{2\sigma^2 - \sigma^2} = \sigma.$$

Moments about the mean are obtained as

$$\mu_1 = 0,\ \mu_2 = \sigma^2,\ \mu_3 = 2\sigma^3 \text{ and } \mu_4 = 9\sigma^4.$$

Hence $\beta_1 = \frac{\mu_3^2}{\mu_2^2} = \frac{(2\sigma^3)^2}{(\sigma^2)^3} = \frac{4\sigma^6}{\sigma^6} = 4$, and

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{9\sigma^4}{(\sigma^2)^2} = 9.$$

The first quartile, Q_1, is given by the relation

$$\int_0^{Q_1} \frac{1}{\sigma} e^{-x/\sigma}\, dx = \frac{1}{4} \text{ or } \frac{1}{\sigma}\left[-\sigma e^{-x/\sigma}\right]_0^{Q_1} = \frac{1}{4}$$

Or $-e^{-Q_1/\sigma} + 1 = \frac{1}{4}$ or $e^{-Q_1/\sigma} = \frac{3}{4}$ ... (A)

Also $\int_0^{Q_3} \frac{1}{\sigma} e^{-x/\sigma}\, dx = \frac{3}{4}$, giving

$$e^{-Q_3/\sigma} = \frac{1}{4} \qquad \text{... (B)}$$

Dividing (A) by (B), we get $e^{(Q_3 - Q_1)/\sigma} = 3$

Hence $Q_3 - Q_1 = \sigma \log_e 3$.

9.7. (a) $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \frac{1}{2}\int_{-\infty}^{\infty} x \cdot e^{-|x|}\, dx.$

$$= \frac{1}{2}\int_{-\infty}^{0} x e^{x}\, dx + \frac{1}{2}\int_{0}^{\infty} x e^{-x}\, dx$$

$$= -\frac{1}{2}\int_{0}^{\infty} x e^{-x}\, dx + \frac{1}{2}\int_{0}^{\infty} x e^{-x}\, dx$$

(changing x into $-x$)

$$= 0$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|}\, dx$$

$$= \frac{1}{2}\int_{-\infty}^{0} x^2 e^{x}\, dx + \frac{1}{2}\int_{0}^{\infty} x^2 e^{-x}\, dx$$

$$= \frac{1}{2}\int_{0}^{\infty} x^2 e^{-x}\, dx + \frac{1}{2}\int_{0}^{\infty} x^2 \cdot e^{-x}\, dx$$

$$= \int_{0}^{\infty} x^2 e^{-x}\, dx = \Gamma(3) = 2$$

$\therefore$ $\text{Var}(X) = E(X^2) - [E(X)]^2 = 2.$

(b) The total probability shall be unity, i.e.

$$\int_{-\infty}^{\infty} \frac{a\, dx}{e^{-x} + e^{x}} = 1$$

Or $\int_{-\infty}^{\infty} \frac{a\, e^{x}\, dx}{1 + e^{2x}} = 1$ (multiplying L.H.S. by $\frac{e^x}{e^x}$)

Now let $t = e^x$, then $dt = e^x\, dx$.

$\therefore$ $\int_{0}^{\infty} \frac{a\,dt}{1 + t^2} = 1$ Or $a\left[\tan^{-1} t\right]_0^{\infty} = 1$

Or $a\frac{\pi}{2} = 1$, or $a = \frac{2}{\pi}$

(ii) Now $P(X<1) = \frac{2}{\pi}\int_{-\infty}^{1} \frac{dx}{e^{-x} + e^{x}}$

$$= \frac{2}{\pi}\int_{-\infty}^{1} \frac{e^{x}}{1 + e^{2x}}\, dx \text{ [multiplying by } \frac{e^x}{e^x}\text{]}$$

$$= \frac{2}{\pi}\int_{0}^{e} \frac{dt}{1 + t^2} \qquad \text{(on putting } e^x = 1\text{)}$$

$$= \frac{2}{\pi}\left[\tan^{-1} t\right]_0^{e}$$

$$= \frac{2}{\pi}\tan^{-1} e = \frac{2}{\pi}\tan^{-1}(2.7183)$$

$$= \frac{2}{\pi}\left(\frac{\pi}{180}\times\frac{349}{5}\right) = 0.7755.$$

Hence the probability that two independent observations will take on values less than 1 = (0.7755) (0.7755) = 0.6014.

9.8. (b) The mode is that value of x for which

(i) $f'(x) = 0$ and (ii) $f''(x) < 0$.

Here $f(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}$.

Differentiating, we get

$$f'(x) = \frac{1}{\Gamma(m)}[(-e^{-x})x^{m-1} + (m-1)x^{m-2}e^{-x}]$$

$$= \frac{1}{\Gamma(m)} \cdot e^{-x}[(m-1)x^{m-2} - x^{m-1}]$$

Now $f'(x) = 0$ gives

$$e^{-x}[(m-1)x^{m-2} - x^{m-1}] = 0$$

or $e^{-m} \cdot x^{m-2}[(m-1) - x] = 0$ or $x = m-1$.

Differentiating $f'(x)$, we get

$$f''(x) = \frac{1}{\Gamma(m)}\left[\left\{-e^{-x}(m-1)x^{m-2} - x^{m-1}\right\} + e^{-x}\left\{(m-1)\times(m-2)x^{m-3} - (m-1)x^{m-2}\right\}\right]$$

$$= \frac{1}{\Gamma(m)}e^{-x}\left[(m-1)(m-2)x^{m-3} - (m-1)x^{m-2} - (m-1)x^{m-2} + x^{m-1}\right]$$

$$= \frac{1}{\Gamma(m)}e^{-x}x^{m-3}\left[x^2 - 2(m-1)x + (m-1)(m-2)\right]$$

Putting $x = m-1$ in $f''(x)$, we get

$$f''(m-1) = \frac{1}{\Gamma(m)}e^{1-m}(m-1)^{m-3}[(m-1)^2 - 2(m-1)^2 + (m-1)(m-2)]$$

$$= \frac{1}{\Gamma(m)}e^{1-m}(m-1)^{m-2}[-(m-1)+(m-2)]$$

$$= \frac{1}{\Gamma(m)}e^{1-m}(m-1)^{m-2}(-1)$$

which is negative for an integral value of m.

Hence mode = $m - 1$.

9.9. (a) (i) The rth moment about origin of $\gamma(m)$ variate is given by

$$\mu_r' = E(X^r) = \int_0^\infty x^r f(x)\,dx = \int_0^\infty x^r \cdot \frac{1}{\Gamma(m)}x^{m-1} \cdot e^{-x}dx$$

$$= \frac{1}{\Gamma(m)}\int_0^\infty x^{m+r-1} \cdot e^{-x}\,dx = \frac{1}{\Gamma(m)}\Gamma(m+r)$$

$$= \frac{(m+r-1)(m+r-2)\ldots(m+1)\,m\Gamma(m)}{\Gamma(m)}$$

$$= m(m+1)\ldots(m+r-2)(m+r-1).$$

(ii) The rth moment about origin using mgf is obtained as below:

The *m.g.f.* of $\gamma(m)$ variate is given by

$$M_0(t) = \int_0^\infty e^{tx} \cdot \frac{1}{\Gamma(m)}x^{m-1}e^{-x}dx = \frac{1}{\Gamma(m)}\int_0^\infty x^{m-1}e^{-x(1-t)}dx$$

Let $u = x(1-t)$. Then $dx = \frac{du}{1-t}$

$$\therefore \quad M_0(t) = \frac{1}{\Gamma(m)}\int_0^\infty \left(\frac{u}{1-t}\right)^{m-1} e^{-u}\frac{du}{1-t}$$

$$= \frac{1}{(1-t)^m} \cdot \frac{1}{\Gamma(m)} \int_0^\infty u^{m-1} e^{-u}\, du$$

$$= (1-t)^{-m}, \text{ provided that } |t| < 1.$$

Differentiating $M_0(t)$ r times *w.r.* to t and putting $t=0$, we get

$$\mu_r' = m(m+1) \dots (m+r-1)$$

(b) The m.g.f. of a $\gamma(m)$ variate with respect to origin is given by

$$M_0(t) = \int_0^\infty \frac{e^{tx}\, e^{-x} \,.\, x^{m-1}}{\Gamma(m)} dx = \frac{1}{\Gamma(m)} \int_0^\infty x^{m-1} \,.\, e^{-x(1-t)}\, dx$$

$$= \frac{1}{(1-t)^m} = (1-t)^{-m}, \text{ provided that } |t| < 1.$$

Similarly, the *m.g.f.* of a $\gamma(n)$ variate is

$$M_0(t) = (1-t)^{-n}$$

Now the *m.g.f.* of the sum is equal to the product of their *m.g.f*'s. That is

$$M_{X+Y}(t) = E[e^{t(X+Y)}], \text{ where}$$

$$X \sim \frac{1}{\Gamma(m)} \int_0^\infty x^{m-1} e^{-x}\, dx \text{ and } Y \sim \frac{1}{\Gamma(n)} \int_0^\infty y^{n-1} \,.\, e^{-y}\, dy$$

Thus $M_{X+Y}(t) = M_X(t) \,.\, M_Y(t)$

$$= (1-t)^{-m} (1-t)^{-n} = (1-t)^{-(m+n)}$$

But this is the *m.g.f.* of a $\gamma(m+n)$ variate.

Hence the result.

9.10. (b) The harmonic mean, H, of the Beta distribution is given by

$$\frac{1}{H} = \int_0^1 \frac{1}{B(m,n)} \,.\, x^{m-1} (1-x)^{n-1} \,.\, \frac{1}{x} dx$$

$$= \frac{1}{B(m,n)} \int_0^1 x^{m-2} (1-x)^{n-1}\, dx$$

$$= \frac{B(m-1,n)}{B(m,n)} = \frac{\Gamma(m-1)\,\Gamma(n)}{\Gamma(m+n-1)} \div \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}$$

$$= \frac{\Gamma(m-1)\,\Gamma(n)}{\Gamma(m+n-1)} \times \frac{(m+n-1)\,\Gamma(m+n-1)}{(m-1)\,\Gamma(m-1)\,\Gamma(n)} = \frac{m+n-1}{m-1}$$

Hence $H = \dfrac{m-1}{m+n-1}$.

9.11. (a) The joint probability density of X and Y is

$$f(x,y) = \frac{1}{\Gamma(m)\,\Gamma(n)} e^{-(x+y)} \,.\, x^{m-1} \,.\, y^{n-1}, \quad \begin{matrix} 0 \le x \le \infty \\ 0 \le y \le \infty \end{matrix}$$

Let $u = x + y$ and $v = \dfrac{x}{y}$, so that $x = \dfrac{uv}{1+v}$ and $y = \dfrac{u}{1+v}$.

Then the Jacobian of the transformation is

$$\frac{1}{J} = \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ \dfrac{1}{y} & \dfrac{-x}{y^2} \end{vmatrix} = -\frac{x+y}{y^2} = -\frac{(1+v)^2}{u}$$

$$\therefore \quad |J| = \frac{u}{(1+v)^2} \text{ and } dxdy = \frac{u}{(1+v)^2} dudv.$$

After substitution, the joint probability density for u and v is obtained as

$$g(u,v) = \frac{1}{\Gamma(m)\,\Gamma(n)} e^{-u} \left(\frac{uv}{1+v}\right)^{m-1} \cdot \left(\frac{u}{1+v}\right)^{n-1} \cdot \frac{u}{(1+v)^2}$$

$$= \frac{1}{\Gamma(m)\,\Gamma(n)} \,.\, e^{-u}\, u^{m+n-1} \frac{v^{m-1}}{(1+v)^{m+n}}$$

As x and y range from 0 to ∞, the range of u is from 0 to ∞ and of v is from 0 to ∞.

By integrating out *w.r.t.u.*, we get the probability density for v as follows:

$$h(v) = \int_0^\infty g(u, v)\, du = \frac{1}{\beta(m, n)} \cdot \frac{v^{m-1}}{(1+v)^{m+n}}$$

Thus the distribution of v is the Beta distribution of the second kind.

(b) Let X be a $\gamma(l)$ variate. Then

$$f(x) = \frac{1}{\Gamma(l)} x^{l-1} . e^{-x}, \qquad x > 0$$

Put $u = \sqrt{x}$, so that $du = \dfrac{dx}{2\sqrt{x}}$ or $dx = 2udu, \quad 0 < u < \infty$

Then the distribution of u,

$$g(u) = \frac{1}{\Gamma(l)} (u^2)^{l-1} . e^{-u^2} . 2udu$$

$$= \frac{2}{\Gamma(l)} u^{2l-1} . e^{-u^2} du$$

which is another form for Γ-function.

The mean value of u, the positive square root of x, is given by

$$E(u) = \int_0^\infty u\, g(u)\, du$$

$$= \frac{2}{\Gamma(l)} \int_0^\infty u.u^{2l-1}.e^{-u^2} du = \frac{2}{\Gamma(l)} \int_0^\infty e^{2l-1+1/2}.e^{-u^2} du$$

$$= \frac{\Gamma(l + 1/2)}{\Gamma(l)}$$

9.15. (a) The equation of the normal curve is

$$f(x) = ke^{-(x^2-6x+9)/24} = ke^{-(x-3)^2/24}$$

Since the area under the normal curve is unity, therefore

$$1 = k \int_{-\infty}^{\infty} e^{-(x-3)^2/24} dx$$

Let $z = \dfrac{x-3}{\sqrt{24}}$ so that $dx = \sqrt{24}\, dz$. Then

$$1 = \sqrt{24}\, k \int_{-\infty}^{\infty} e^{-z^2} dz = \sqrt{24}\, k.2 \int_0^{\infty} e^{-z^2} dz$$

Again let $z^2 = v$. Then $2z\, dz = dv$ or $dz = (1/2)\, v^{-1/2}\, dv$

Thus $\quad 1 = \sqrt{24}\, k \int_0^\infty e^{-v} e^{-1/2} dv = \sqrt{24}\, k \sqrt{\pi}$

$$\left(\because \int_0^\infty e^{-v} v^{-1/2} dv = \frac{\Gamma(1)}{(2)} = \sqrt{\pi}\right)$$

Hence $\quad k = \dfrac{1}{\sqrt{24\pi}}$.

The equation of the normal curve may be written as

$$f(x) = \frac{1}{\sqrt{12}\,\sqrt{2\pi}} e^{-(x-3)^2/2(\sqrt{12})^2}$$

Comparing with the general form of the normal curve, we find that

$$\mu = 3 \text{ and } \sigma = \sqrt{12} = 3.464.$$

9.16. The Quartile Deviation, Q, is found as

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-Q}^{\mu+Q} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{2}$$

Or $\quad \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^{Q/\sigma} e^{-z^2/2} dz = \dfrac{1}{4}$, where $z = \dfrac{x-\mu}{\sigma}$

From area tables, we find that $\dfrac{Q}{\sigma} = 0.6745$

Or $\quad Q = \dfrac{2}{3}\sigma$ (approximately)

$$\therefore \quad \frac{Q}{10} = \frac{\sigma}{15} \qquad \ldots (1)$$

The mean deviation from the mean, μ, is given by

$$M.D. = \int_{-\infty}^{\infty} |x-\mu| \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-z^2/2} dz, \text{ where } z = \frac{x-\mu}{\sigma}$$

$$= \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2} dz = \sigma\sqrt{\frac{2}{\pi}}$$

$$= 0.7979\,\sigma = \frac{4}{5}\sigma, \text{ approximately.}$$

$$\therefore \quad \frac{M.D.}{12} = \frac{\sigma}{15} \qquad \ldots (2)$$

Hence from (1) and (2), we find that

$$\frac{Q}{10} = \frac{M.D.}{12} = \frac{\sigma}{15}$$

Thus the Quartile Deviation, the Mean Deviation and the Standard Deviation are approximately in the ratio of 10:12:15.

9.18. (b) The Standardized Normal Variable is

$$Z = \frac{X-\mu}{\sigma}.$$

Now (i) $P(\mu \le X \le \mu + 1.54\sigma) = P\left(\frac{\mu-\mu}{\sigma} \le Z \le \frac{\mu+1.54\sigma-\mu}{\sigma}\right)$

$= P(0 \le Z \le 1.54)$

$= 0.4382$ (From area tables)

(ii) $P(\mu-1.73\sigma \le X \le \mu+0.56\sigma)$

$= P\left(\frac{\mu-1.73\sigma-\mu}{\sigma} \le Z \le \frac{\mu+0.56\sigma-\mu}{\sigma}\right)$

$= P(-1.73 \le Z \le 0.56)$

$= P(-1.73 \le Z \le 0) + P(0 \le Z \le 0.56)$

$= 0.4582 + 0.2123 = 0.6705$

Hence the required percentages of area between the given intervals are 43.82% and 67.05% respectively.

9.19 (a) The *r.v.X* is $N[0,(0.6)^2]$, so that *S.N.V.* is

$$Z = \frac{X-0}{0.6}$$

Now (i) $P(X>0) = P\left(\frac{X-0}{0.6} > \frac{0-0}{0.6}\right)$

$= P(Z>0) = 0.5$

(ii) $P(0.2<X<1.8) = P\left(\frac{0.2-0}{0.6} < \frac{X-0}{0.6} < \frac{1.8-0}{0.6}\right)$

$= P(0.33 < Z < 3.00)$

$= P(0 < Z < 3.00) - P(0<Z<0.33)$

$= 0.49865 - 0.1293$

$= 0.36935$ (From area tables)

(b) The S.N.V. is $Z = \frac{X-1}{3}$

(i) $P(3.43 \le X \le 6.19) = P\left(\frac{3.43-1}{3} \le \frac{X-1}{3} \le \frac{6.19-1}{3}\right)$

$= P(0.81 \le Z \le 1.73)$

$= P(0 \le Z \le 1.73) - P(0 \le Z \le 0.81)$

$= 0.4582 - 0.2910$ (From area tables)

$= 0.1672$

(ii) $P(-1.43 \le X \le 6.19) = P\left(\frac{-1.43-1}{3} \le \frac{X-1}{3} \le \frac{6.19-1}{3}\right)$

$= P(-0.81 \le Z \le 1.73)$

$= P(-0.81 \leq Z \leq 0) + P(0 \leq Z \leq 1.73)$

$= 0.2910 + 0.4582$ (From area tables)

$= 0.7492$

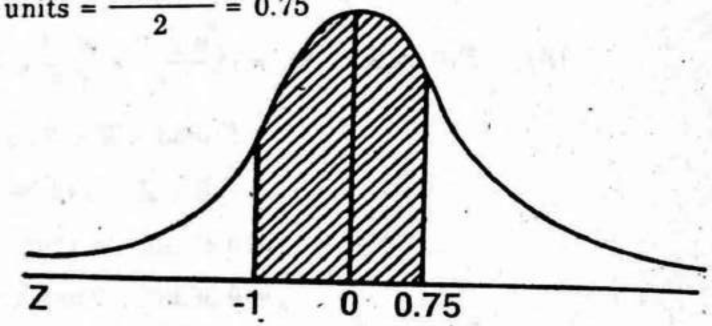
9.20. The normal distribution with $\mu = 12$ and $\sigma = 2$, is

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-[(x-12)/2]^2/2}$$

Thus the standardized normal variate is $Z = \dfrac{X-12}{2}$.

(a) 10 in standard units $= \dfrac{10-12}{2} = -1$,

13.5 in standard units $= \dfrac{13.5-12}{2} = 0.75$



Required area

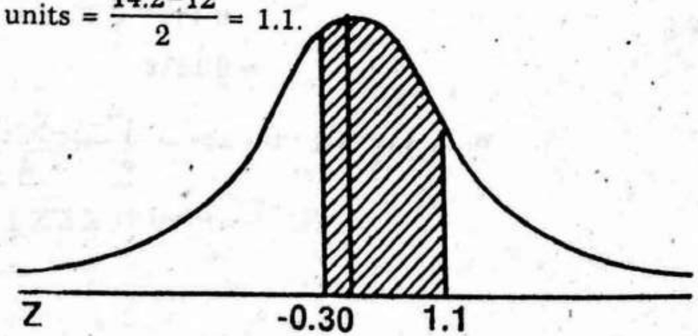
= (area between $z = -1$ and $z = 0.75$)

= (area between $z = -1$ and $z = 0$) + (area between $z = 0$ and $z = 0.75$)

$= 0.3413 + 0.2734 = 0.6147$ (From area tables)

(b) 11.4 in standard units $= \dfrac{11.4-12}{2} = -0.3$,

14.2 in standard units $= \dfrac{14.2-12}{2} = 1.1$.



$\therefore$ **Required area**

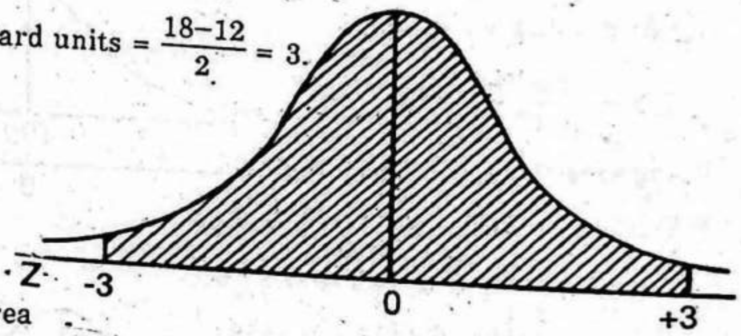
= (area between $z = -0.3$ and $z = 1.1$)

= (area between $z = -0.3$ and $z = 0$) + (area between $z = 0$ and $z = 1.1$)

$= 0.1179 + 0.3643 = 0.4822$ (From area tables)

(c) 6 in standard units $= \dfrac{6-12}{2} = -3$,

18 in standard units $= \dfrac{18-12}{2} = 3$.



Required area

= (area between $z = -3$ and $z = 3$)

= (area between $z = -3$ and $z = 0$) + (area between $z = 0$ and $z = 3$)

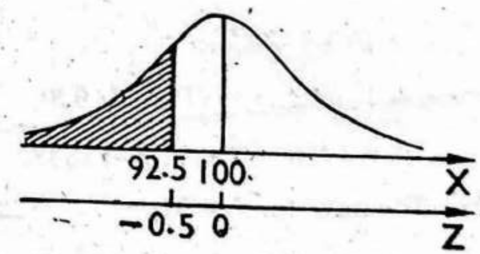
$= 0.49865 + 0.49865 = 0.9973$ (From area tables)

9.21. We draw the normal curve sketch showing x and z values, and the desired area for each part. With $\mu = 100$ and $\sigma = \sqrt{225} = 15$, we have

$$z = \frac{x-100}{15}$$

(a) At $x = 92.5$, we calculate

$$z = \frac{92.5-100}{15} = -0.5.$$



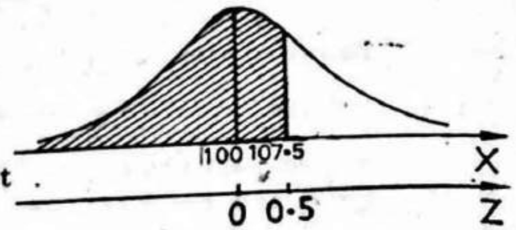
$P(X \leq 92.5) = P(Z \leq -0.5)$

$= 0.5 - P(-0.5 \leq Z \leq 0)$

$= 0.5 - 0.1915 = 0.3085$ (From area tables)

(b) At $x = 107.5$, we compute

$$z = \frac{107.5-100}{15} = 0.5.$$

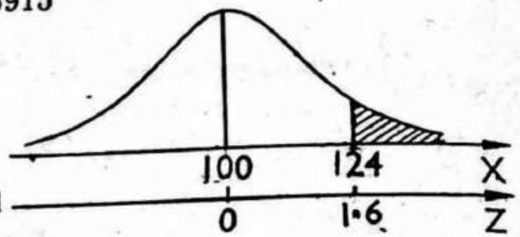


Using area table, we therefore get

$$\begin{aligned} P(X \leq 107.5) &= P(Z \leq 0.5) \\ &= P(-\infty \leq Z \leq 0) + P(0 \leq Z \leq 0.5) \\ &= 0.5 + 0.1915 = 0.6915 \end{aligned}$$

(c) At $x = 124$, we compute

$$z = \frac{124-100}{15} = 1.6.$$

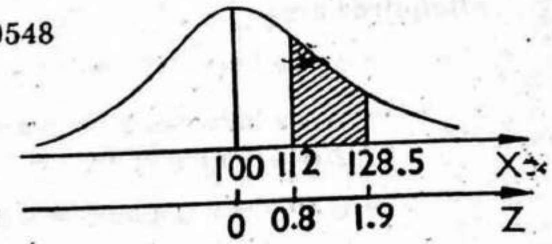


Using area table, we therefore find

$$\begin{aligned} P(X \geq 124) &= P(Z \geq 1.6) \\ &= P(0 \leq Z \leq \infty) - P(0 \leq Z \leq 1.6) \\ &= 0.5 - 0.4452 = 0.0548 \end{aligned}$$

(d) We have for $x = 112$,

$$z = \frac{112-100}{15} = 0.8, \text{ and}$$



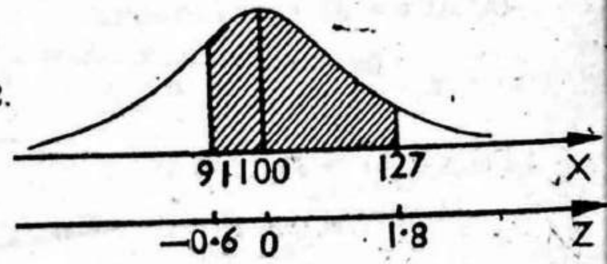
for $x = 128.5$, $z = \dfrac{128.5-100}{15} = 1.9.$

Therefore using area table, we get

$$\begin{aligned} &P(112 \leq X \leq 128.5) \\ &= P(0.8 \leq Z \leq 1.9) \\ &= P(0 \leq Z \leq 1.9) - P(0 \leq Z \leq 0.8) \\ &= 0.4713 - 0.2881 = 0.1832. \end{aligned}$$

(e) We have for $x = 91$,

$$z = \frac{91-100}{15} = -0.6, \text{ and}$$



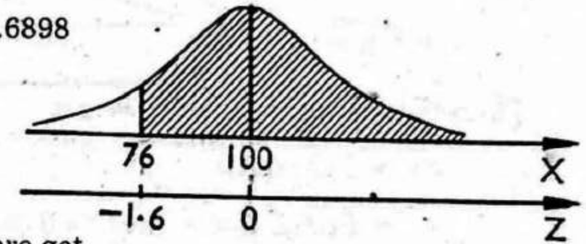
for $x = 127$, $z = \dfrac{127-100}{15} = 1.8.$

Using area table, we therefore get

$$\begin{aligned} &P(91 \leq X \leq 127) \\ &= P(-0.6 \leq Z \leq 1.8) \\ &= P(-0.6 \leq Z \leq 0) + P(0 \leq Z \leq 1.8) \\ &= 0.2257 + 0.4641 = 0.6898 \end{aligned}$$

(f) At $x = 76$, we calculate

$$z = \frac{76-100}{15} = -1.6.$$



Using area table, we therefore get

$$\begin{aligned} P(X \geq 76) &= P(Z \geq -1.6) \\ &= P(-0.6 \leq Z \leq 0) + P(0 \leq Z \leq \infty) \\ &= 0.4452 + 0.5 = 0.9452. \end{aligned}$$

9.22. (b) Here X is normally distributed with a mean 10 and variance 25. Further $Y = 5X + 10$, therefore

$$E(Y) = 5E(X) + 10 = 5(10) + 10 = 60, \text{ and}$$

$$\mathrm{Var}(Y) = 25\,\mathrm{Var}(X) = 25(25) = 625.$$

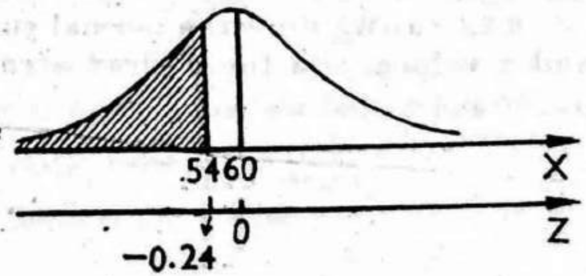
Hence Y is N (60, 625).

We draw the normal curve sketch showing x and z values and the desired area for each part. Wtih $\mu = 60$ and $\sigma = \sqrt{625} = 25$, we have

$$z = \frac{y-60}{25}$$

(i) At $y = 54$, we compute

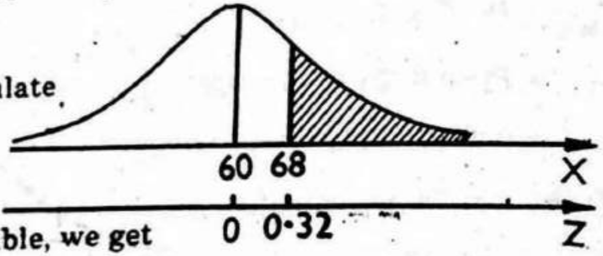
$$z = \frac{54-60}{25} = -0.24$$



Therefore using area table, we get

$$P(Y \le 54) = P(Z \le -0.24)$$
$$= 0.5 - P(-0.24 \le Z \le 0)$$
$$= 0.5 - 0.0948$$
$$= 0.4052.$$

(ii) At $y = 68$, we calculate

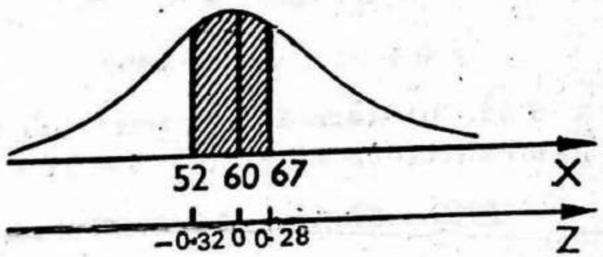


$$z = \frac{68-60}{25} = 0.32$$

Therefore using area table, we get

$$P(Y \ge 68) = P(Z \ge 0.32)$$
$$= P(0 \le Z < \infty) - P(0 \le Z < 0.32)$$
$$= 0.5 - 0.1255$$
$$= 0.3745.$$

(iii) We have for $y = 52$



$$z = \frac{52-60}{25} = -0.32,$$

and for $y = 67$.

$$z = \frac{67-60}{26} = 0.28.$$

Using area table, we therefore get

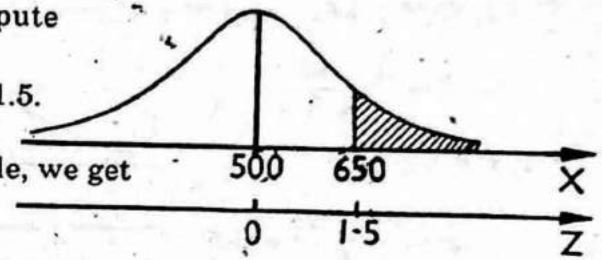
$$P(52 \le Y \le 67) = P(-0.32 \le Z \le 0.28))$$
$$= P(-0.32 \le Z \le 0) + P(0 \le Z \le 0.28)$$
$$= 0.1255 + 0.1103 = 0.2358.$$

9.23. (a) **We draw the normal curve sketch showing x and z values, and the desired area for each part.** With $\mu = 500$ and $\sigma = 100$, we have

$$Z = \frac{X - 500}{100}.$$

(i) We need the probability that a student will score over 650, *i.e.* $P(X > 650)$.

At $x = 650$, we compute



$$z = \frac{650 - 500}{100} = 1.5.$$

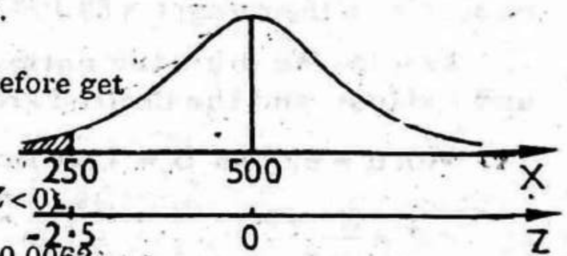
Hence using the area table, we get

$$P(X > 650) = P(Z > 1.5)$$
$$= 0.5 - P(0 \le Z \le 1.5)$$
$$= 0.5 - 0.4337 = 0.0663.$$

(ii) We need the probability that a student will score less than 250, *i.e.* $P(X < 250)$.

At $x = 250$, we compute

$$z = \frac{250 - 500}{100} = -2.5.$$

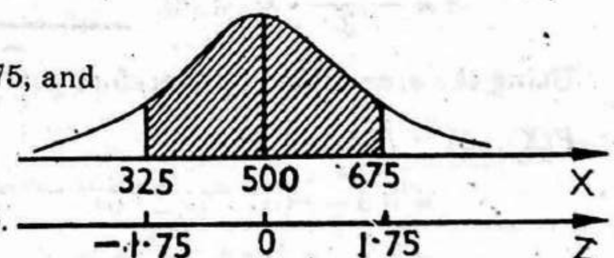


Using the area table, we therefore get

$$P(X < 250) = P(Z < -2.5)$$
$$= 0.5 - P(-2.5 < Z < 0),$$
$$= 0.5 - 0.4938 = 0.0062.$$

(iii) We need the probability that a student will score between 325 and 675, *i.e.* $P(325 < X < 675)$.

We have for $x = 325$,



$$z = \frac{325-500}{100} = -1.75, \text{ and}$$

for $x = 675$,

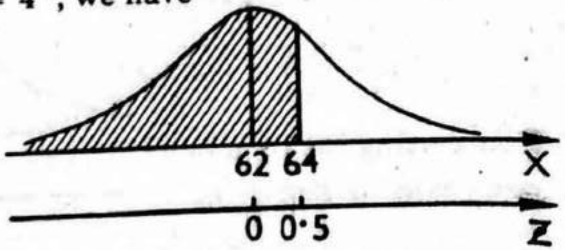
$$z = \frac{675-500}{100} = 1.75.$$

Hence using the area table, we get

$$P(325 < X < 675) = P(-1.75 < Z < 1.75)$$
$$= P(-1.75 < Z < 0) + P(0 < Z < 1.75)$$
$$= 0.4599 + 0.4599 = 0.9198.$$

(b) The percentage of boys who would be rejected on account of their height is equal to the area under the normal curve for $X < 64$. With $\mu = 62''$ and $\sigma = 4''$, we have

$$Z = \frac{X - 62}{4}$$



For $x = 64$,

$$z = \frac{64 - 62}{4} = 0.5.$$

Therefore using the area table, we obtain

$$P(X < 64) = P(Z < 0.5)$$
$$= 0.5 + P(0 < Z < 0.5)$$
$$= 0.5 + 0.1915 = 0.6915.$$

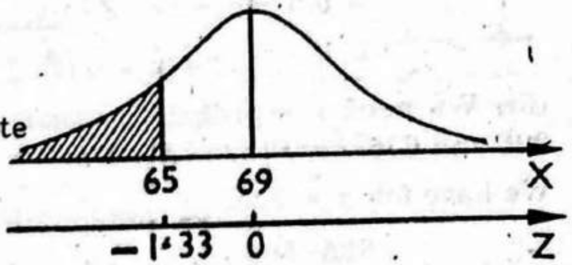
Hence the desired percentage of boys who would be rejected on account of their height is 69.15%.

9.24. (a) We draw the normal curve sketch showing x and z vlaues, and the desired area for each part.

With $\mu = 69$ and $\sigma = 3$, we have

$$Z = \frac{X - 69}{3}.$$

(i) At $x = 65$, we compute



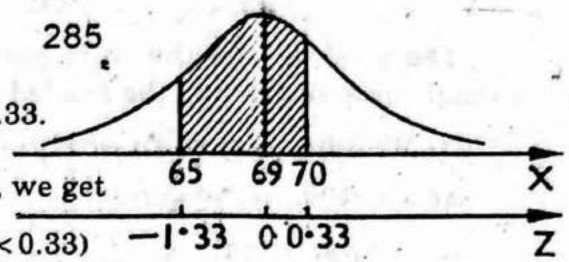
$$z = \frac{65 - 69}{3} = -1.33.$$

Using the area table, we therefore get

$$P(X < 65) = P(Z < -1.33)$$
$$= 0.5 - P(-1.33 < Z < 0)$$
$$= 0.5 - 0.4082 = 0.0918.$$

(ii) For $x = 65$,

$$z = \frac{65 - 69}{3} = -1.33, \text{ and}$$



for $x = 70$, $z = \frac{70 - 69}{3} = 0.33.$

Hence, using the area table, we get

$$P(65 \le X \le 70) = P(-1.33 < Z < 0.33)$$
$$= P(-1.33 < Z < 0) + P(0 < Z < 0.33)$$
$$= 0.4082 + 0.1293 = 0.5375$$

(b) Comparing $M_0(t) = e^{-6t + 32t^2}$ with the *m.g.f.* of $N(\mu, \sigma^2)$, we find that $\mu = -6$ and $\sigma^2 = 64$ or $\sigma = 8$.

To find the desired probabilities, we transform x values to z values, using $z = \frac{x - (-6)}{8}$. Therefore

at $x = -4$, we get $z = \frac{-4 + 6}{8} = 0.25$, and

for $x = 16$, $z = \frac{16 + 6}{8} = 2.75.$

Hence using the area table, we get

$$P(-4 \le X < 16) = P(0.25 \le Z < 2.75)$$
$$= P(0 \le Z \le 2.75) - P(0 \le Z \le 0.25)$$
$$= 0.4970 - 0.0987 = 0.3983.$$

Again, for $x = -10$, we have $z = \frac{-10 + 6}{8} = -0.5$, and

for $x = 0$, $z = \frac{0 + 6}{8} = 0.75.$

Using the area table, we therefore get

$$P(-10 < X \le 0) = P(-0.5 < Z \le 0.75)$$
$$= P(-0.5 < Z \le 0) + P(0 < Z \le 0.75)$$
$$= 0.1915 + 0.2734 = 0.4649.$$

9.25. With mean = 155 and standard deviation = 20, we have

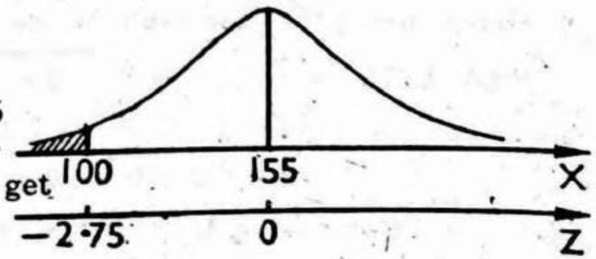
$$Z = \frac{X - 155}{20}$$

The x values and the corresponding z values are shown in normal curve sketch and the needed areas are shaded.

(i) Weights less than or equal to 100 pounds.

At $x = 100$, we compute

$$z = \frac{100 - 155}{20} = -2.75$$



Thus using the area table, we get

$$\begin{aligned} P(X \le 100) &= P(Z \le -2.75) \\ &= 0.5 - P(-2.75 \le Z \le 0) \\ &= 0.5 - 0.4970 = 0.0030 \end{aligned}$$

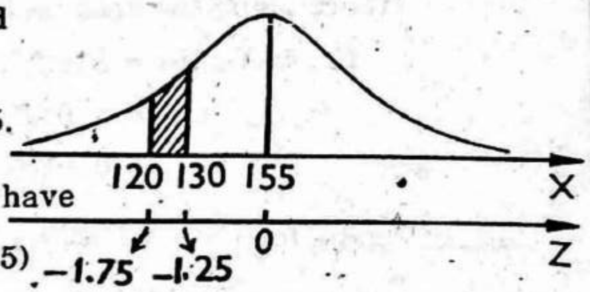
Therefore the number of students with weights less than or equal to 100 pounds is $2{,}000 \times 0.0030 = 6$.

(ii) Weights between 120 and 130 pounds.

At $x = 120$, we have

$$z = \frac{120 - 155}{20} = -1.75, \text{ and}$$

for $x = 130$, $z = \dfrac{130 - 155}{20} = -1.25$.



Using the area table, we therefore have

$$\begin{aligned} P(120 \le X \le 130) &= P(-1.75 \le Z \le -1.25) \\ &= P(-1.75 \le Z \le 0) - P(-1.25 \le Z \le 0) \\ &= 0.4599 - 0.3944 = 0.0655. \end{aligned}$$

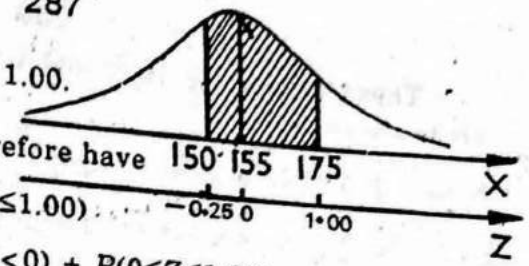
Thus the number of students with weights between 120 and 130 pounds is $2{,}000 \times 0.0655 = 131$.

(iii) Weights between 150 and 175 pounds.

At $x = 15$, we compute

$$z = \frac{150 - 155}{20} = -0.25, \text{ and}$$

for $x = 175$, $z = \dfrac{175 - 155}{20} = 1.00$.



Using the area table, we therefore have

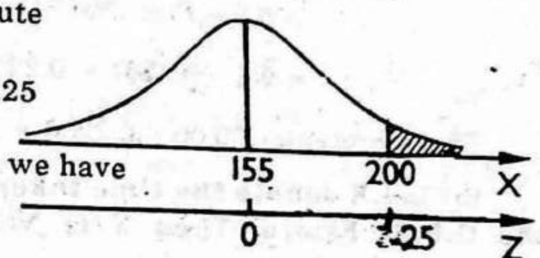
$$\begin{aligned} P(150 \le X \le 175) &= P(-0.25 \le Z \le 1.00) \\ &= P(-0.25 \le Z < 0) + P(0 \le Z \le 1.00) \\ &= 0.0987 + 0.3413 = 0.4400 \end{aligned}$$

Thus the number of students with weights between 150 and 175 pounds is $2{,}000 \times 0.4400 = 880$.

(iv) Weights greater than or equal to 200 pounds.

At $x = 200$, we compute

$$z = \frac{200 - 155}{20} = 2.25$$



Thus using the area table, we have

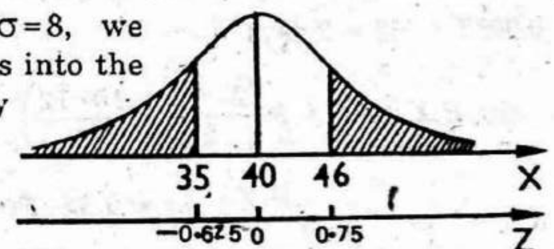
$$\begin{aligned} P(X \ge 200) &= P(Z \ge 2.25) \\ &= 0.5 - P(0 \le Z \le 2.25) \\ &= 0.5 - 0.4878 = 0.0122. \end{aligned}$$

Hence the number of students with weights greater than or equal to 200 pounds is $2{,}000 \times 0.0122 = 24$.

9.26. (a) The number of pairs of stockings that would need replacement is equal to the area under the normal curve for X≤35 and X≥46 days.

With $\mu = 40$ and $\sigma = 8$, we therefore change x values into the corresponding z values by

$$Z = \frac{X - 40}{8}$$



Thus at $x = 35$, $z = \dfrac{35 - 40}{8} = -0.625$.

and at $x = 46$, $z = \dfrac{46 - 40}{8} = 0.75$.

These values are indicated in the figure and the needed areas are shaded.

Now $P(X\leq 35) = P(Z\leq -0.625)$

$= 0.5 - P(-0.625\leq Z\leq 0)$

$= 0.5 - 0.2340 = 0.2660.$

This represents 100,000 × 0.2660 = 26,600 pairs of stockings.

Again $P(X\geq 46) = P(Z\geq 0.75)$

$= 0.5 - P(0\leq Z\leq 0.75)$

$= 0.5 - 0.2734 = 0.2266.$

This represents 100,00 × 0.2266 = 22,660 pairs of stockings.

(b) Let X denote the time taken to deliver the milk to the G.O.R. Estate. Then X is N(12, 2^2). The S.N.V. is $Z = \frac{X-12}{2}$.

(i) $P(X>17) = P\left(\frac{X-12}{2} > \frac{17-12}{2}\right)$

$= P(Z>2.5) = 0.5 - P(0<Z<2.5)$

$= 0.5 - 0.4938 = 0.0062$

$\therefore$ The number of days when he takes longer than 17 minutes $= 0.0062 \times 365 = 2$ days.

(ii) $P(X<10) = P\left(\frac{X-12}{2} < \frac{10-12}{2}\right)$

$= P(Z<-1) = 0.5 - P(-1<Z<0)$

$= 0.5 - 0.3413 = 0.1587$

$\therefore$ The number of days when he takes less than 10 minutes $= 0.1587 \times 365 = 58$ days.

(iii) $P(9<X<13) = P\left(\frac{9-12}{2} < \frac{X-12}{2} < \frac{13-12}{2}\right)$

$= P(-1.5 < Z < 0.5)$

$= P(-1.5 < Z < 0) + P(0<Z<0.5)$

$= 0.4332 + 0.1915 = 0.6247$

Thus the number of days when he takes between 9 and 13 minutes $= 0.6247 \times 365 = 228$ days.

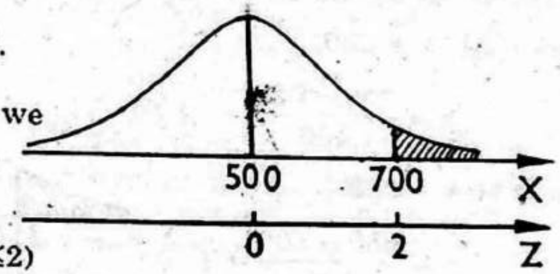
9.27. We draw the normal curve sketch showing x and z values, and the desired area for each part. With $\mu = 500$ and $\sigma = 100$, we have

$$Z = \frac{X-500}{100}$$

(a) (i) *Candidates receiving scores greater than 700.*

At $x = 700$, we compute

$$z = \frac{700-500}{100} = 2.$$



Using the area table, we therefore get

$P(X\geq 700) = P(Z>2)$

$= 0.5 - P(0\leq Z\leq 2)$

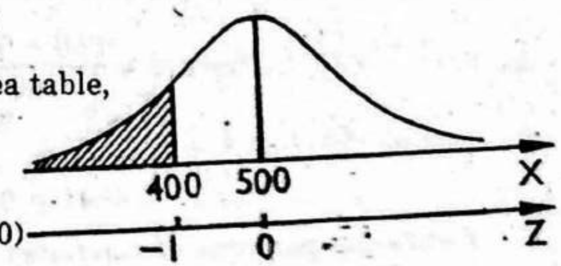
$= 0.5 - 0.4772 = 0.0228$

Hence the required %age = 2.28.

(ii) *Candidates receiving scores less than 400.*

At $x = 400$, we have

$$z = \frac{400-500}{100} = -1.$$



Therefore using the area table, we get

$P(X<400) = P(Z\leq -1)$

$= 0.5 - P(-1\leq Z\leq 0)$

$= 0.5 - 0.3413 = 0.1587$

Hence the desired %age = 15.87.

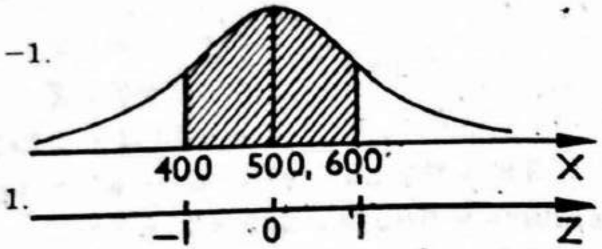
(iii) *Candidates receiving scores between 400 and 600.*

At $x = 400$, we have

$$z = \frac{400-500}{100} = -1.$$

and at $x = 600$,

$$z = \frac{600-500}{100} = 1.$$



Using the area table, we therefore get

$$P(400 \le X \le 600) = P(-1 \le Z \le 1)$$
$$= 2P(0 \le Z \le 1) = 2(0.3413) = 0.6826$$

Hence the desired %age = 68.26.

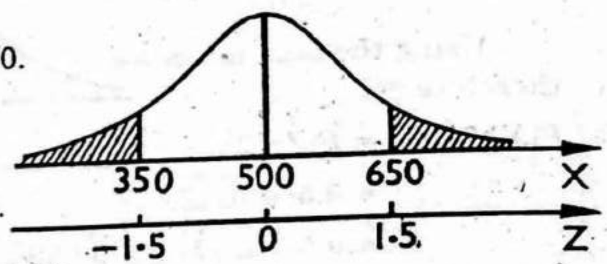
(iv) *Which differ from mean by more than 150.* This implies that the scores should be either less than 350 or more than 650.

At $x = 350$,

$$z = \frac{350-500}{100} = -1.50.$$

and at $x = 650$,

$$z = \frac{650-500}{100} = 1.50.$$



Now $P(X \le 350) = P(Z \le -1.5)$

$$= 0.5 - 0.4332 = 0.0668, \text{ and}$$

$$P(X \ge 650) = P(Z \le 1.5) = 0.5 - 0.4332 = 0.0668.$$

Hence the desired proportion = 0.0668 + 0.0668 = 0.1336 or 13.36%.

(b) For $x = 680$, we have $z = \dfrac{680-500}{100} = 1.8$. Therefore

$$P(X > 680) = P(Z > 1.8)$$
$$= 0.5 - 0.4641 = 0.0359.$$

Hence the per cent of candiates having higher scores than 680 is 3.59%.

9.28. The better route would be one which gives the smaller probability of the man's being late for the appointment.

Let X denote the time taken for the journey by the *route through the city centre*. Then with $\mu = 27$ and $\sigma = 5$, we have

$$Z = \frac{X-27}{5}. \text{ Thus}$$

(i) at $x = 28$, we compute $z = \dfrac{28-27}{5} = 0.2$, and

(ii) at $x = 32$, we calculate $z = \dfrac{32-27}{5} = 1.0$.

Therefore the probabilities of being late are

(i) $P(X \ge 28) = P(Z > 0.2)$

$$= 0.5 - P(0 \le Z \le 0.2) = 0.5 - 0.0793 = 0.4207, \text{ and}$$

(ii) $P(X \ge 32) = P(Z > 1.0)$

$$= 0.5 - P(0 \le Z \le 1.0) = 0.5 - 0.3413 = 0.1583.$$

Again, t Y denote the time taken for the journey by the *new route*. T..en with $\mu = 29$ and $\sigma = 2$, we have $Z = \dfrac{Y-29}{2}$. Thus

(i) at $y = 28$, we compute $z = \dfrac{28-29}{2} = -0.5$, and

(ii) at $y = 32$, we have $z = \dfrac{32-29}{2} = 1.5$.

Therefore the probabilities of being late are

(i) $P(Y \ge 28) = P(Z \ge -0.5)$

$$= 0.5 + P(-0.5 \le Z \le 0)$$
$$= 0.5 + 0.1915 = 0.6915, \text{ and}$$

(ii) $P(Y \ge 32) = P(Z \ge 1.5)$

$$= 0.5 - P(0 \le Z \le 1.5)$$
$$= 0.5 - 0.4332 = 0.0668.$$

Hence comparing the probabilities of being late, we find that the route through the city centre is better if the man has 28 minutes, but if he has 32 minutes, the new route is better.

9.29. (a) Given $P(X>a) = 0.5636$, indicating that a must be less than the mean, 100.

Now $\quad P\left(\frac{X-100}{4} > \frac{a-100}{4}\right) = 0.5636$

i.e. $\quad P\left(Z > \frac{a-100}{4}\right) = 0.5636$

Since a is less than 100, so $\frac{a-100}{4}$ is negative and can be expressed as $-\left(\frac{100-a}{4}\right)$.

Thus $\quad P\left(Z > -\frac{100-a}{4}\right) = 0.5636 = 0.0636 + 0.5$

$$= P(-z<0) + P(z>0)$$

Corresponding to an area of 0.0636, from area tables (under normal curve) we find that $z = 0.16$, so

$$\frac{100-a}{4} = 0.16 \text{ or } 100 - a = 0.64$$

which gives $a = 99.36$.

(b) For a normal distribution with mean μ and variance σ^2, $Q.D. = 0.6745\sigma$ (see property 10 one page 383 of the text).

Let x be the point P_{10} where P_{10} is a point at or below which 10% of the area lies. Then the area to the left of x is 0.1 and the area between μ and x is $0.5-0.1 = 0.4$.

Now $\quad (z/P = 0.4) = 1.28$, where $z = (x-\mu)/\sigma$

So $\quad \frac{x-\mu}{\sigma} = -1.28$, or $x = \mu - 1.28\sigma$ (x lies to the left of μ)

Thus $\quad P_{10} = \mu - 1.28\sigma$

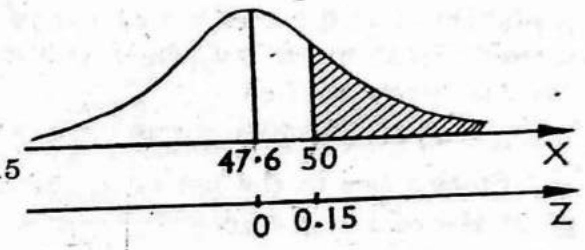
Similarly, we find that $P_{90} = \mu + 1.28\sigma$

Hence $\quad k = \frac{Q.D.}{P_{90} - P_{10}} = \frac{0.6745\sigma}{(\mu+1.28\sigma) - (\mu-1.28\sigma)}$

$$= \frac{0.6745\sigma}{2.56\sigma} = 0.263.$$

9.30. We draw the normal curve sketch showing x and z values, and the desired area for each part. With $\mu = 47.6$ and $\sigma = 16.2$, we have

$$Z = \frac{X - 47.6}{16.2}$$



(i) At $x = 50$, we compute

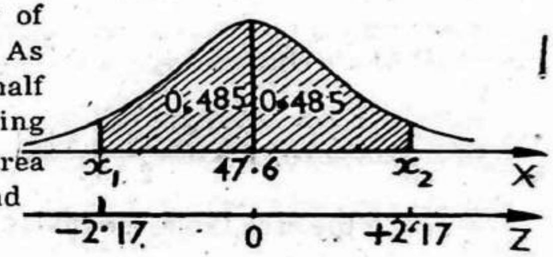
$$z = \frac{50 - 47.6}{16.2} = 0.15$$

Using the area table, we therefore get

$P(X>50) = P(Z>0.15)$

$\quad = 0.5 - P(0<Z<0.15)$

$\quad = 0.5 - 0.0596 = 0.4404.$



(ii) Let x_1 and x_2 be the two points between which the probability of an observation falling is 0.97. As the curve is symmetrical, so half of 0.97, *i.e.* 0.485 is the area lying on either side of μ. Using area table inversely, we therefore find

$$(|z| \; P = 0.485) = 2.17.$$

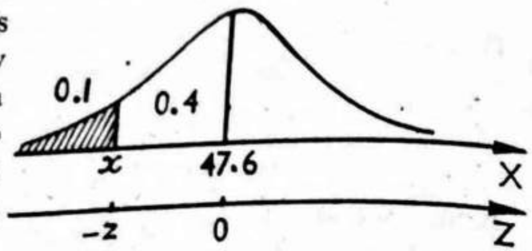
Since x_1 lies to the left of μ, therefore z is negative at this point. Hence $x_1 = \mu + z\sigma$

$$= 47.6 + (-2.17)(16.2) = 12.4.$$

Again, as x_2 lies to the right of μ, therefore z is positive at this point, and $x_2 = \mu + z\sigma$

$$= 47.6 + (2.17)(16.2) = 82.8.$$

(iii) Let x be the point P_{10}, where P_{10} is a point at or below which 10% of the area lies. Then the area to the left of x is 0.1 and area between μ and x is $0.5-0.1=0.4$



Looking at the area table, we find that a probability of 0.4 does not appear in the table, so we take the closest probability to 0.4, which is 0.3997. Thus, using the area table inversely, we find

$(z \mid P = 0.3997) = 1.28$

Since x lies to the left of μ, therefore z is negative at this point. Hence $x = \mu + z\sigma$

$= 47.6 + (-1.28)(16.2) = 26.86.$

Similarly, we find that $P_{30} = 39.11$ and $P_{99} = 85.3$.

9.31. (a) We first transform x values to z values, using

$$Z = \frac{X - 14.40}{2.50}$$

(i) At $x = 12$, we have $z = \dfrac{12 - 14.40}{2.50} = -0.96$, and

at $x = 16$, $z = \dfrac{16 - 14.40}{2.50} = 0.64.$

Using the area table, we therefore get

$$\begin{aligned} P(12 \le X \le 16) &= P(-0.96 \le Z \le 0.64) \\ &= P(-0.96 \le Z \le 0) + P(0 \le Z \le 0.64) \\ &= 0.3315 + 0.2389 = 0.5704. \end{aligned}$$

Hence the number of rods between 12 and 16 metres is

$1{,}000 \times 0.5704 = 570$ rods.

(ii) At $x = 15$, we compute $z = \dfrac{15 - 14.40}{2.50} = 0.24.$

Therefore using the area table, we get

$$\begin{aligned} P(X > 15) &= P(Z \ge 0.24) \\ &= 0.5 - P(0 \le Z \le 0.24) \\ &= 0.5 - 0.0948 = 0.4052. \end{aligned}$$

(b) With mena = 34.4 and s.d. = 16.6, we have

$$Z = \frac{X - 34.4}{16.6}$$

At $x = 30$, we compute $z = \dfrac{30 - 34.4}{16.6} = -0.27$, and

at $x = 50$, we have $z = \dfrac{60 - 34.4}{16.6} = 1.54$

Therefore using the area table, we obtain

$$\begin{aligned} P(30 \le X \le 60) &= P(-0.27 \le Z \le 1.54) \\ &= P(-0.27 \le Z \le 0) + P(0 \le Z \le 1.54) \\ &= 0.1064 + 0.4384 = 0.5446 \end{aligned}$$

Hence the number of students expected to obtain marks between 30 and 60 is $1{,}000 \times 0.5446 = 545$.

The central 70% of the candidates implies that 35% of the candidates will lie on either side of the mean, *i.e.* the area lying on either side of μ is 0.35.

Using area table inversely, we therefore find

$(z \mid P = 0.35) = 1.04$

$$\begin{aligned} \text{Thus the limits of marks} &= \mu \pm z\sigma \\ &= 34.4 \pm (1.04)(16.6) \\ &= 34.4 \pm 17.26 = 17.14,\ 51.68 \end{aligned}$$

Hence the desired limits of marks of the central 70% of the candidates are 17 and 52.

9.32. Let X be the r.v. the amount of drink in milliliters. Then X is N[200, $(15)^2$].

The S.N.V. is $Z = \frac{X-\mu}{\sigma} = \frac{X-200}{15}$

(a) Now $P(X > 240) = P\left(\frac{X-200}{15} > \frac{240-200}{15}\right)$

$= P(Z > 2.67) = 0.5 - P(0 < Z < 2.67)$

$= 0.5 - 0.4962 = 0.0038$

The required fraction is therefore 0.38%.

(b) $P(191 < X < 209) = P\left(\frac{191-200}{15} < \frac{X-200}{15} < \frac{209-200}{15}\right)$

$= P(-0.6 < Z < 0.6)$

$= P(-0.6 < Z < 0) + P(0 < Z < 0.6)$

$= 0.2257 + 0.2257 = 0.4514$

(b) A cup will likely overflow if it contains 230 or more than 230 milliliters.

$P(X \geq 230) = P\left(\frac{X-200}{15} \geq \frac{230-200}{15}\right)$

$= P(Z \geq 2) = 0.5 - 0.4772 = 0.0228$

Thus the required number of cups is

$1000 \times 0.0228 = 23$

(d) Let x be the value below which the smallest 25% of the drinks lie. Then the area to the left of x is 0.25. Looking the area tables inversely, we find that

$(z / P = 0.25) = 0.675$

$x = \mu + z\sigma = 200 + (-0.675)(15)$

$= 200 - 10.125 = 189.875$

9.33. (a) Let X be the r.v. the height of applicant to police force. Then X is normally distributed with $\mu = 170$ cm and $\sigma = 3.8$ cm.

The S.N.V. is $Z = \frac{X-170}{3.8}$

Let x denote the minimum acceptance height for the police force. Then the area to the left of x is 0.30 and the area between μ and x is $0.50 - 0.30 = 0.20$

Looking the are table inversely, we find that

$(z/P = 0.20) = 0.527$

As x lies to the left of μ, so z is negative at this point.

$\therefore \quad x = \mu - z(\sigma) = 170 - (0.527)(3.8) = 168$ cm

(b) Let X be the r.v. the life of motors in years. Then X is $N(10, (2)^2)$. The S.N.V. is $Z = \frac{X-10}{2}$.

Let x be the number of years, the manufacturer should offer as guaranteed period. Then the area between μ and x is $0.5-0.03 = 0.47$. Looking the area table inversely, we find that

$(z/P = 0.47) = 1.88$

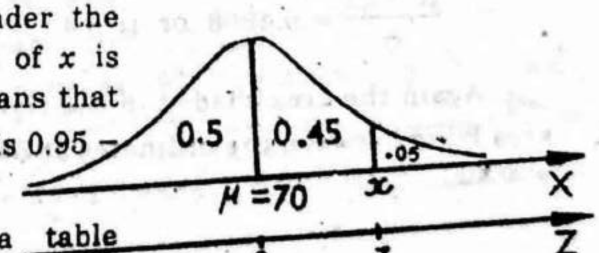
Since x lies to the left of μ, so z is negative at this point.

$\therefore \quad x = \mu - z\sigma = 10 - (1.88)(2)$

$= 10 - 3.76 = 6.24$ years.

9.34. Let x be the point corresponding to the height needed by 95% of men using the door.

Then the area under the normal curve to the left of x is 0.95 (see figure). This means that the area between μ and x is $0.95 - 0.50 = 0.45$.



Looking the area table inversely, we find $(z \mid P = 0.45) = 1.645$

With $\mu = 70$ and $\sigma = 3$, we have $z = \frac{x-70}{3}$ or $x = 70 + 3z$

As x lies to the right of μ, so z is positive at this point.

Thus $x = 70 + 3\,(1.645) = 70 + 4.935 = 75$ inches.

Since the architect wants to have at least a one-foot clearance, therefore he must make the doors 75 + 12 = 87 inches high.

9.35. (a) Let μ be the mean and σ the standard deviation of the normal distribution. Then the two quartiles are given by

$$Q_1 = \mu - 0.6745\sigma, \text{ and } Q_3 = \mu + 0.6745\sigma$$

Substituting the values, we get

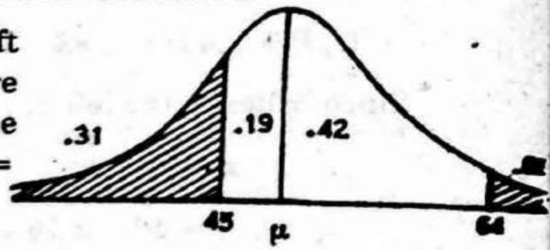
$$\mu - 0.6745\sigma = 8,$$

$$\mu + 0.6745\sigma = 17$$

Solving, we get $\mu = 12.5$ and $\sigma = 6.67$.

(b) Let μ and σ be the mean and the standard deviation respectively of the normal distribution.

The area shaded in the left hand tail is 0.31 (given), therefore the area lying between the ordinates at 45 and the mean = 0.5 − 0.31 = 0.19.



From area tables, we find that the value of z corresponding to an area of 0.19 is 0.4958. Therefore

$$\frac{\mu - 45}{\sigma} = 0.4958 \text{ or } \mu - 45 = 0.4958\sigma \quad \dots (A)$$

Again the area shaded in the right tail is 0.08 (given) so the area lying between the ordinates at 64 and the mean = 0.5 − 0.08 = 0.42.

The value of z corresponding to an area of 0.42 is 1.4053, implying that $\frac{64-\mu}{\sigma} = 1.4053$ or $64 - \mu = 1.4053\sigma \quad \dots (B)$

Solving the equations (A) and (B), we get

$\mu = 50$ approx. and $\sigma = 10$ approx.

(c) Now $P(X<89) = 0.90$ implies that the probability between μ and 89 is 0.90 − 0.50 = 0.40.

From area tables, we find that the value of Z corresponding to an area 0.40 is 1.28. Therefore

$$\frac{89-\mu}{\sigma} = 1.28 \text{ or } 89 - \mu = 1.28\sigma \quad \dots (1)$$

Again $P(X>94) = 0.05$ implies that the probability between μ and 94 is 0.50 − 0.05 = 0.45.

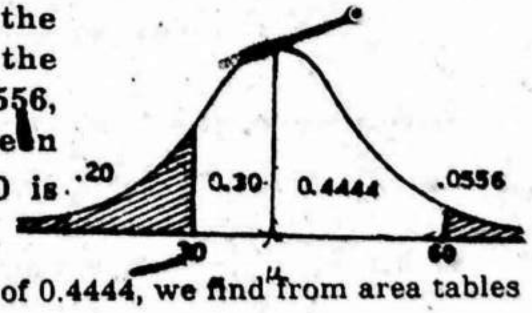
From area tables, we find that the value of Z corresponding to an area 0.45 is 1.645. Therefore

$$\frac{94-\mu}{\sigma} = 1.645 \text{ or } 94 - \mu = 1.645\sigma \quad 0\dots (2)$$

Solving the equations (1) and (2), we get

$\mu = 71.464$ and $\sigma = 13.70$.

9.36. Let the total marks be 100, and let μ and σ be the mean and the standard deviation. The number of students getting 60 or more marks is 25, that is $\frac{25}{450}$ = 0.0556 of the total number of students. In other words, the area to the right of the ordinate at x = 60 is 0.0556, therefore the area between the ordinates at μ and 60 is 0.5000−0.0556 = 0.4444.



Corresponding to an area of 0.4444, we find from area tables that the value of z is 1.59.

$$\therefore \quad \frac{60-\mu}{\sigma} = 1.59 \text{ or } 60-\mu = 1.59\sigma \qquad \dots (A)$$

The number of students who failed (got less than 30 marks) is 90, that is $\frac{90}{450} = 0.20$ of the total number of students. This means that the area to the left of the ordinate at $x=30$ is 0.20, that is the area between the ordinates at μ and 30 is 0.50–0.20=0.30. For this value of area, the value of z from area tables is 0.84.

$$\therefore \quad \frac{\mu-30}{\sigma} = 0.84 \text{ or } \mu-30 = 0.84\sigma \qquad \dots (B)$$

Solving the equatoins (A) and (B), we get

$\mu = 40.37$ and $\sigma = 12.35$.

9.37. Let X denote the height the boy can reach in various attempts. Then we are given

$$P(X \geq 1.85) = \frac{1}{5} = 0.2 \text{ and}$$

$$P(X \geq 1.70) = \frac{9}{10} = 0.9, \text{ which implies } P(X \leq 1.70) = 0.1.$$

Let μ and σ be the mean and standard deviation of the normal distribution of heights. Then $Z = \frac{X-\mu}{\sigma}$.

Looking the area table inversely, we find that $(z_1/P=0.3) = 0.84$ and $(z_2/P=0.4) = -1.28$ as it lies to the left of μ.

Thus $\frac{1.85-\mu}{\sigma} = 0.84$ and $\frac{1.70-\mu}{\sigma} = -1.28$.

Solving, we get $\mu = 1.7905\ m$ and $\sigma = 0.0708$.

Again, $(z/P = 0.5 - 0.001$, i.e. $0.499) = 3.08$,

so that $\frac{x-1.7905}{0.0708} = 3.08$, which gives

$x = 1.7905 + 0.2181 = 2.009$ m.

9.38. (b) Le X denote the number of deaths. Then the *p.d.* of X is

$$f(x) = \binom{500}{x}\left(\frac{20}{100}\right)^x\left(\frac{80}{100}\right)^{500-x}$$

and we need $P(70 \leq X \leq 80)$.

To use the normal curve area, the given values of X are to be adjusted for continuity. Therefore the interval of discrete values $70 \leq X \leq 80$ is replaced by the interval $69.5 \leq X \leq 80.5$.

With $\mu = np = 500 \times \frac{20}{100} = 100$, and

$$\sigma = \sqrt{npq} = \sqrt{500 \times \frac{20}{100} \times \frac{80}{100}} = 8.944,$$

we compute the z values.

At $x = 69.5$, we compute $z = \frac{69.5-100}{8.944} = -3.41$, and

at $x = 80.5$, we find that $z = \frac{80.5-100}{8.944} = -2.18$.

Hence, using the area table, we get

$$\begin{aligned} P(70 \leq X \leq 80) &= P(-3.41 \leq Z \leq -2.18) \\ &= P(-3.41 \leq Z \leq 0) - P(-2.18 \leq Z \leq 0) \\ &= 0.4997 - 0.4854 = 0.0143. \end{aligned}$$

9.39. (b) Let X denote the number of heads when a fair coin is tossed. Then the *p.d.* of X is

$$f(x) = \binom{200}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{200-x}$$

To use the normal curve area, the given values of X are to be adjusted for continuity. Therefore (i) the interval of discrete values $80 \leq X \leq 120$ becomes $79.5 \leq X \leq 120.5$, (ii) the value 90 starts at 89.5, and (iii) the discrete value 100 becomes the interval 99.5 to 100.5.

With $\mu = np = 200\times\frac{1}{2} = 100$, and $\sigma = \sqrt{npq} = \sqrt{200\times\frac{1}{2}\times\frac{1}{2}} = 7.07$, we compute the z values.

(i) At $x = 79.5$, we find $z = \dfrac{79.5-100}{7.07} = -2.90$, and

at $x = 120.5$, we find $z = \dfrac{120.5-100}{7.07} = 2.90$.

Using the area table, we therefore get

$$\begin{aligned} P(80\le X\le 120) &= P(-2.90\le Z\le 2.90) \\ &= P(-2.90\le Z\le 0) + P(0\le Z\le 2.90) \\ &= 0.4981 + 0.4981 = 0.9962. \end{aligned}$$

(ii) At $x = 89.5$, we compute $z = \dfrac{89.5-100}{7.07} = -1.49$.

Therefore using the area table, we find

$$\begin{aligned} P(X<90) &= P(Z\le -1.49) \\ &= 0.5-P(-1.49\le Z\le 0) = 0.5-0.4319 = 0.0681. \end{aligned}$$

(iii) At $x = 99.5$, we compute $z = \dfrac{99.5-100}{7.07} = -0.07$, and

at $x = 100.5$, we find $z = \dfrac{100.5-100}{7.07} = 0.07$.

Using the area table, we therefore get

$$\begin{aligned} P(X=100) &= P(-0.07\le Z\le 0.07) \\ &= P(-0.07\le Z\le 0) + P(0\le Z\le 0.07) \\ &= 0.0279 + 0.0279 = 0.0558. \end{aligned}$$

9.40. (a) Let X denote the number of heads appearing. Then the *p.d.* of X is

$$f(x) = \binom{200}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{200-x}$$

In order to use the normal curve area table, we adjust the given vlaues of X. Consequently (i) the interval of discrete values $105\le X\le 110$ is replaced by the interval $104.5\le X\le 110.5$, and (ii) the discrete value 95 starts at 94.5.

With $\mu = np = 200\times\frac{1}{2} = 100$, and $\sigma = \sqrt{200\times\frac{1}{2}\times\frac{1}{2}} = 7.07$, we compute z values.

(i) At $x = 104.5$, $z = \dfrac{104.5-100}{7.07} = 0.64$, and

at $x = 110.5$, $z = \dfrac{110.5-100}{7.07} = 1.49$.

Therefore using the area table, we get

$$\begin{aligned} P(105\le X\le 110) &= P(0.64\le Z\le 1.49) \\ &= P(0<Z\le 1.49) - P(0\le Z\le 0.64) \\ &= 0.4319 - 0.2382 = 0.1925. \end{aligned}$$

(ii) At $x = 94.5$, $z = \dfrac{94.5-100}{7.07} = -0.78$.

Hence using the area table, we get

$$\begin{aligned} P(X<95) &= P(Z<-0.78) = 0.5 - P(-0.78<Z<0) \\ &= 0.5 - 0.2823 = 0.2177. \end{aligned}$$

(b) **Let X denote the number of heads appearing.** Then the *p.d.* of X is

$$f(x) = \binom{15}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{15-x}$$

(i) We need $P(6\le X\le 9)$ by applying the binomial distribution. Thus

$$\begin{aligned} P(6\le X\le 9) &= \sum_{x=6}^{9}\binom{15}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{15-x} \\ &= \left(\frac{1}{2}\right)^{15}\left[\binom{15}{6}+\binom{15}{7}+\binom{15}{8}+\binom{15}{9}\right] \\ &= \frac{1}{32768}[5005 + 6435 + 6435 + 5005] \\ &= \frac{22880}{32768} = 0.6982 \end{aligned}$$

(ii) To use the normal curve area table, the given values of X are to be adjusted for continuity. Accordingly, the interval of discrete values $6 \le X \le 9$ is replaced by the interval $5.5 \le X \le 9.5$.

With $\mu = np = 15 \times \frac{1}{2} = 7.5$, and $\sigma = \sqrt{npq} = \sqrt{15 \times \frac{1}{2} \times \frac{1}{2}} = 1.94$, we compute z values. Thus

$$\text{at } x = 5.5,\ z = \frac{5.5 - 7.5}{1.94} = -1.03, \text{ and}$$

$$\text{at } x = 9.5,\ z = \frac{9.5 - 7.5}{1.94} = 1.03$$

Hence using the area table, we get

$$\begin{aligned} P(6 \le X \le 9) &= P(-1.03 \le Z \le 1.03) \\ &= P(-1.03 \le Z \le 0) + P(0 \le Z \le 1.03) \\ &= 0.3485 + 0.3485 = 0.6970. \end{aligned}$$

9.41. (a) Let the r.v.X be the number of incoming calls in one minute. Then, assuming that the number of incoming calls in one minute to have the Poisson distribution, we have

$$P(X=x) = \frac{e^{-5} \cdot (5)^x}{x!} \text{ for } x = 0, 1, 2, \ldots$$

If Y be the number of incoming calls in 20 minutes, then Y has the Poisson distribution with mean $= 5 \times 20 = 100$. We need $P(Y \le 102)$. To find the required probability, we use a normal approximatioin as the mean of the distribution is large. So Y is $N(100, 100)$ as the mean and variance of the Poisson distribution are equal.

$$\begin{aligned} \text{Hence } P(Y \le 102) &= P\left(\frac{(Y+1/2) - 100}{10} \le \frac{(102+1/2) - 100}{10}\right) \\ &= P(Z \le 0.25) = 0.5 + P(0 \le Z \le 0.25) \\ &= 0.5 + 0.0987 = 0.5987 \end{aligned}$$

(b) Given X is $b(x; 20, 0.4)$.

$$\begin{aligned} \text{Now} \quad P(6 \le X \le 10) &= \sum_{x=6}^{10} b(x; 20, 0.4) \\ &= \sum_{x=0}^{10} b(x; 20, 0.4) - \sum_{x=0}^{5} b(x; 20, 0.4) \\ &= 0.8724 - 0.1255 = 0.7469 \end{aligned}$$

Approximations using

(i) The Poisson distribution: We have $\mu = np = 20 \times 0.4 = 8$

$$\text{So} \quad P(X=x) = \frac{e^{-8} \cdot (8)^x}{x!}$$

$$\text{Now} \quad P(X=6) = \frac{e^{-8} \cdot (8)^6}{6!} = 0.122138$$

Using the recurrence formula $P(X=x) = \frac{\mu}{x} \cdot P(X=x-1)$, we get

$$P(X=7) = \frac{8}{7} P(X=6) = \frac{8}{7} \times 0.122138 = 0.139586$$

$$P(X=8) = \frac{8}{8}(0.139586) = 0.139586$$

$$P(X=9) = \frac{8}{9}(0.139586) = 0.124076$$

$$P(X=10) = \frac{8}{10} P(X=9) = 0.099261$$

$$\therefore \quad P(6 \le X \le 10) = P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10) = 0.6246.$$

This answer shows a poor approximation.

(ii) The normal distribution:

Here X is $N(np, npq)$, where

$$np = (20)(0.4) = 8 \text{ and } npq = (20)(0.4)(0.6) = 4.8$$

Adjusting the values of X for continuity, the interval of discrete values $6 \le X \le 10$ is replaced by the interval $5.5 \le X \le 10.5$. Therefore

$$P(5.5 \le X \le 10.5) = P\left(\frac{5.5-8}{2.19} \le \frac{X-8}{2.19} \le \frac{10.5-8}{2.19}\right)$$

$$= P(-1.14 \le Z \le 1.14)$$

$$= 0.3729 + 0.3729 = 0.7458.$$

9.42. (a) From ordinates table, we find that

(i) Ordinate at $z = 0.064$ is 0.3981;

(ii) Ordinate at $z = 1.27$ is 0.1781;

(iii) Ordinate at $z = -0.84$ = Ordinate at $z = 0.84 = 0.2803$,

and (iv) Ordinate at $z = -2.08 = 0.0459$

(b) As the bulk of the normal distribution lies between $\mu-3\sigma$ and $\mu+3\sigma$, so the range of classes would be $27.0 \pm 3(2.2)$, *i.e.* 20.4 to 33.6.

As range = 13.2, we use 7 classes with a common width $h = 2$. We then construct the classes as 20.0–22.0, 22.0–24.0, ..., 32.0–34.0. The necessary computations are shown below:

Classes	Upper class boundary	$z = \frac{ucb-\mu}{\sigma}$	$P(Z<z)$, $\Phi(z)$	Probab. $(\hat{p})$	Expected freq.
Upto 22.0	22.0	−2.27	0.0116	0.0116	2
22.0-24.0	24.0	−1.36	0.0869	0.0753	16
24.0-26.0	26.0	−0.45	0.3264	0.2395	50
26.0-28.0	28.0	+0.45	0.6736	0.3472	73
28.0-30.0	30.0	1.36	0.9131	0.2395	50
30.0-32.0	32.0	2.27	0.9884	0.0753	16
Over 32.0	∞	∞	1.0000	0.0116	2

9.43. (a) We first calculate the mean and the standard deviation of the given distribution.

Stature (x)	f	$D(=x-68)$	fD	fD^2
61	2	−7	−14	98
62	10	−6	−60	360
63	11	−5	−55	275
64	38	−4	−152	608
65	57	−3	−171	513
66	93	−2	−186	372
67	106	−1	−106	106
68	126	0	−744	0
69	109	1	109	109
70	87	2	174	348
71	75	3	225	675
72	23	4	92	368
73	9	5	45	225
74	4	6	24	144
Σ	750	--	+669 −75	4201

$$\bar{x} = a + \frac{\sum fD}{n} = 68 + \frac{(-75)}{750} = 67.9 \text{ inches, and}$$

$$s = \sqrt{\frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2} = \sqrt{\frac{4201}{750} - \left(\frac{-75}{750}\right)^2} = \sqrt{5.59} = 2.36,$$

Now $\bar{x} + s = 67.9 \pm 2.36 = 65.54,\ 70.26.$

The number of students having stature in this range

$= 93 + 106 + 126 + 109 + 87 = 521$

$\therefore$ Proportion $= \frac{521}{750} \times 100 = 69\%$

$\bar{x} + 2s = 67.9 \pm 2(2.36) = 63.18,\ 72.62$

The number of students having stature in this range

$= 38 + 57 + 93 + 106 + 126 + 109 + 87 + 75 + 23 = 714$

$\therefore$ Proportion $= \frac{714}{750} \times 100 = 95\%$

$\bar{x} + 3s = 67.9 \pm 3(2.36) = 60.82,\ 74.98$

The number of students having stature in this range = 750

$\therefore$ Proportion $= \frac{750}{750} \times 100 = 100\%$

In a normal distribution, the proportions (from the area tables) lying between these ranges are 68%, 95% and 99.7% respectively. Hence the given distribution can be considered to be reasonably normal.

(b) **Here the mean = 67.9, and s = 2.36. The necessary computations of the expected frequencies are given below:**

(x_i)	f	ucb	$z = \frac{ucb-\bar{x}}{s}$	$P(Z<z)$ $\Phi(z_i)$	Probab. (p)	Expected $f_i(np)$
61	2	Upto 61.5	−2.71	0.0034	0.0034	2.55
62	10	62.5	−2.29	0.0110	0.0076	5.70
63	11	63.5	−1.86	0.0314	0.0204	15.30
64	38	64.5	−1.44	0.0749	0.0435	32.62
65	57	65.5	−1.02	0.1539	0.0790	59.25
66	93	66.5	−0.59	0.2776	0.1237	92.78
67	106	67.5	−0.17	0.4325	0.1549	116.18
68	126	68.5	0.25	0.5987	0.1662	124.65
69	109	69.5	0.68	0.7518	0.1531	114.82
70	87	70.5	1.10	0.8643	0.1125	84.38
71	75	71.5	1.53	0.9370	0.0727	54.52
72	23	72.5	1.95	0.9744	0.0374	28.05
73	9	73.5	2.37	0.9911	0.0167	12.52
74	4	$+\infty$	∞	1.0000	0.0089	6.68
Σ	750	--	--	--	--	750

9.44. We first calculate the mean and standard deviation of this distribution.

Computation of mean and standard deviation

Classes	x_i	f	$u=(x-67)/5$	fu	fu^2
40–44	42	7	−5	−35	175
45–49	47	8	−4	−32	128
50–54	52	25	−3	−75	225
55–59	57	38	−2	−76	152
60–64	62	51	−1	−51	51
65–69	67	60	0	0	0
70–74	72	45	1	45	45
75–79	77	32	2	64	128
80–84	82	10	3	30	90
85–89	87	4	4	16	64
Total		280		−114	1058

Now, $\bar{x} = a + \frac{\sum f_i u_i}{\sum f} \times h$

$= 67 + \frac{(-114)}{280} \times 5 = 67 - 2.04 = 64.96$, and

$$s = h \times \sqrt{\frac{\sum fu^2}{\sum f} - \left(\frac{\sum fu}{\sum f}\right)^2}$$

$$= 5 \times \sqrt{\frac{1058}{280} - \left(\frac{-114}{280}\right)^2} = 5 \times \sqrt{3.7786 - 0.1681}$$

$$= 5 \times \sqrt{3.6105} = 5 \times (1.90) = 9.50$$

The necessary calculations of the expected frequencies of the fitted normal distribution are shown as follows:

Upper class boundary	$z = \frac{ucb-\bar{x}}{s}$	$P(Z<z)$ $\Phi(z_i)$	Probability (p)	Expected f (np)
Upto 39.5	−2.68	0.0037	0.0037	
44.5	−2.15	0.0158	0.0121	4.42
49.5	−1.63	0.0515	0.0357	10.00
54.5	−1.10	0.1357	0.842	23.58
59.5	−0.57	0.2843	0.1486	41.61
64.5	−0.05	0.4801	0.1958	54.82
69.5	+0.48	0.6844	0.2043	57.20
74.5	1.00	0.8413	0.1567	43.80
79.5	1.53	0.9370	0.0957	26.80
84.5	2.06	0.9803	0.0433	12.12
Over 84.5	+∞	1.0000	0.0197	5.52

And the necessary calculations for the heights of the ordinates appear below:

x_i	$z_i = \frac{x_i-\bar{x}}{s}$	$\phi(z_i)$	Ordinates $(nh/s)\,\phi(z_i)$
42	−2.42	0.0213	3.14
47	−1.89	0.0669	9.86
52	−1.36	0.1582	23.31
57	−0.84	0.2803	41.31
62	−0.31	0.3802	56.03
67	+0.21	0.3902	57.50
72	0.74	0.3034	44.71
77	1.27	0.1781	26.25
82	1.79	0.0804	11.85
87	2.32	0.0270	3.98

9.45. Let x denote the mid points. Then taking $u=(x-19)/2$, we find that $\sum fu=70$ and $\sum fu^2=2092$.

Hence, in terms of the units of original measurements, $\bar{x}=19.14$ and $s = 2.89$.

Computation of the Expected Frequencies

ucb	$z_i = \frac{ucb-\bar{x}}{s}$	$P(Z<z)$ $\phi(z_i)$	$\hat{p}$	Expected f $(n\hat{p})$
Upto 10	−3.16	0.0008	0.0008	6.8
12	−2.47	0.0068	0.0060	
14	−1.78	0.0375	0.0307	30.7
16	−1.09	0.1379	0.1004	100.4
18	−0.39	0.3483	0.2104	210.4
20	+0.30	0.6179	0.2696	269.6
22	0.99	0.8389	0.2210	221.0
24	1.68	0.9535	0.1146	114.6
26	2.37	0.9911	0.0376	37.6
∞	∞	1.0000	0.0089	8.9
Total	--	--	--	1000

●●●●●●●●●●

Chapter 10

SIMPLE REGRESSION AND CORRELATION

10.6. (a) The necessary calculations for determining the equation of the least squares regression line are shown below:

X	Y	X^2	Y^2	XY
20	5	400	25	100
11	15	121	225	165
15	14	225	196	210
10	17	100	289	170
17	8	289	64	136
19	9	361	81	171
92	68	1496	880	952

The estimated linear regression line of Y on X is

$$\hat{Y} = a + bX,$$

where a and b are the least squares estimates of the parameters α and β respectively and are given by

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \text{ and } a = \bar{Y} - b\bar{X}$$

Substituting the sums, we get

$$b = \frac{(6)(952) - (92)(68)}{(6)(1496) - (92)^2}$$

$$= \frac{5712 - 6256}{8976 - 8464} = \frac{-544}{512} = -1.0625$$

$$a = \frac{68}{6} - (-1.0625)\left(\frac{92}{6}\right) = 11.3333 + 16.2917 = 27.625$$

Thus the estimated regression line is

$$\hat{Y} = 27.625 - 1.0625X.$$

(b) The predicted values of Y are found by substituting the X values in the estimated equation. Thus for $X = 10$,

$$\hat{Y} = 27.625 - 1.0625\ (10) = 27.625 - 10.625 = 17.00$$

Similarly, the predicted values of Y for $X = 11, 15, 17, 19$ and 20 are 15.94, 11.69, 9.56, 7.44, 6.38 respectively.

(c) The standard error of estimate is given by

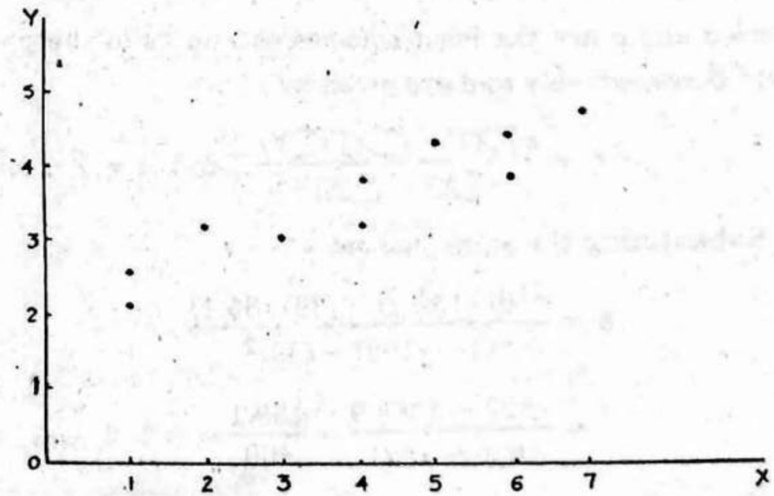
$$s_{Y.X} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}}$$

Now $$\sum(Y-\hat{Y})^2 = (5 - 6.38)^2 + (15 - 15.94)^2 + (14 - 11.69)^2 + (17-17.00)^2 + (8-9.56)^2 + (9-7.44)^2$$

$$= (-1.38)^2 + (-0.94)^2 + (2.31)^2 + 0 + (-1.56)^2 + (1.56)^2 = 12.9913$$

Hence $$s_{Y.X} = \sqrt{\frac{12.9913}{4}} = \sqrt{3.2478} = 1.80.$$

10.7. (a) The scatter diagram of the given data appears below:



(b) The necessary computations for the least-squares regression line and residuals are given in the following table:

X_i	Y_i	X_i^2	X_iY_i	$\hat{Y}_i = 2.00 + 0.387X_i$	Residuals $e_i = Y_i - \hat{Y}_i$
1	2.1	1	2.1	2.387	−0.287
1	2.5	1	2.5	2.387	0.113
2	3.1	4	6.2	2.774	0.326
3	3.0	9	9.0	3.161	−0.161
4	3.8	16	15.2	3.548	0.252
4	3.2	16	12.8	3.548	−0.348
5	4.3	25	21.5	3.935	0.365
6	3.9	36	23.4	4.322	−0.422
6	4.4	36	26.4	4.322	0.078
7	4.8	49	33.6	4.709	0.091
39	35.1	193	152.7	35.093	0.007

The estimated linear regression line of Y on X is

$$\hat{Y} = a + bX,$$

where a and b are the least-squares estimates of the parameters α and β respectively and are given by

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}, \text{ and } a = \bar{Y} - b\bar{X}$$

Substituting the sums, we get

$$b = \frac{(10)(152.7) - (39)(35.1)}{(10)(193) - (39)^2}$$

$$= \frac{1527 - 1368.9}{1930 - 1521} = \frac{158.1}{409} = 0.387, \text{ and}$$

$$a = 3.51 - (0.387)(3.9)$$

$$= 3.51 - 1.5093 = 2.00$$

Thus the estimated regression line is

$$\hat{Y} = 2.00 + 0.387X$$

(c) To compute the residuals, we first calculate the estimated values $\hat{Y}$ which appear in fifth column of the above table. The residuals are given in the last column and they add to 0.007. Theoretically the sum is zero and this small difference is due to rounding.

(d) When $X = 10$, the predicted value of Y is

$$\hat{Y} = 2.00 + 0.387(10) = 5.87$$

10.8. The estimated regression equation is $\hat{Y} = a + bX$, where

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}, \text{ and}$$

$$a = \bar{Y} - b\bar{X}.$$

(a) Now $\quad b = \dfrac{1000 - (10)(10)(20)}{2{,}000 - (10)(10)^2} = \dfrac{-1000}{1000} = -1$, and

$$a = 20 - (-1)(10) = 30.$$

Hence the desired estimated regression equation is

$$\hat{Y} = 30 - X.$$

(b) Here $\quad b = \dfrac{(32)(193640) - (528)(11720)}{(32)(11440) - (528)^2}$

$$= \frac{6196480 - 6188160}{366080 - 278784} = \frac{8320}{87296} = 0.095, \text{ and}$$

$$a = \frac{11720}{32} - (0.095)\left(\frac{528}{32}\right)$$

$$= 366.25 - (0.095)(16.5) = 364.68$$

Therefore the desired estimated regression equation is

$$\hat{Y} = 364.68 + 0.095X.$$

(c) Here $\quad b = \dfrac{(100)(1613) - (1239)(79)}{(100)(17322) - (1239)^2}$

$$= \frac{161300 - 97881}{1732200 - 1535121} = \frac{63419}{197079} = 0.32, \text{ and}$$

$$a = \bar{Y} - b\bar{X} = \frac{79}{100} - (0.32)\frac{1239}{100}$$

$$= 0.79 - 3.96 = -3.17$$

Here the desired estimated regression equation is

$$\hat{Y} = -3.17 + 0.32X$$

(d) Here $b = \dfrac{(10)(130628) - (1710)(760)}{(10)(293162) - (1710)^2}$

$$= \frac{1306280 - 1299600}{2931620 - 2924100} = \frac{6686}{7523} = 0.8887, \text{ and}$$

$$a = 76 - (0.8887)(171) = 76 - 151.97 = -75.97$$

Thus the desired estimated regression line is

$$\hat{Y} = -75.97 + 0.89X$$

(e) Here $b = \dfrac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2} = \dfrac{9871}{2800} = 3.53$, and

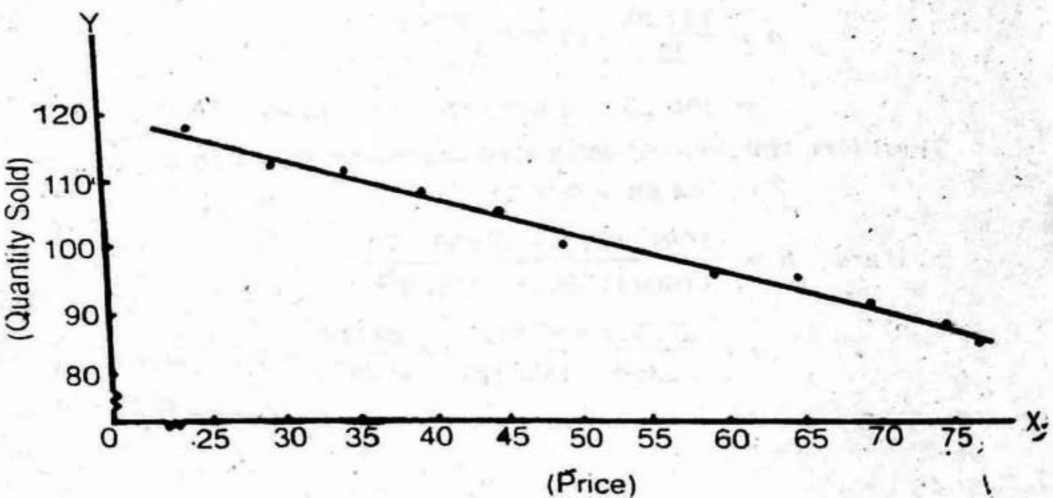
$$a = \bar{Y} - b\bar{X} = 237 - (3.53)(52)$$

$$= 237 - 183.56 = 53.44$$

Hence the desired estimated regression equation is

$$\hat{Y} = 53.44 + 3.53X.$$

10.9. (a) The scatter diagram of the given data appears below:



(b) The equation for the estimated regression line is

$$\hat{Y} = a + bX,$$

where a and b, the least-squares estimates of the parameters α and β, are given as

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} \text{ and } a = \bar{Y} - b\bar{X}.$$

The necessary computations are given in the table below:

Price (X)	Quantity (Y)	X^2	XY	$\hat{Y} = 130.62 - 0.57X$	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
25	118	625	2950	116.37	1.63	2.6569
45	105	2025	4725	104.97	0.03	0.0009
30	112	900	3360	113.52	−1.52	2.3104
50	100	2500	5000	102.12	−2.12	4.4944
35	111	1225	3885	110.67	0.33	0.1089
40	108	1600	4320	107.82	0.18	0.0324
65	95	4225	6175	93.57	1.43	2.0449
75	88	5625	6600	87.87	0.13	0.0169
70	91	4900	6370	90.72	0.28	0.0784
60	96	3600	5760	96.42	−0.42	0.1764
495	1024	27225	49145	1024.05	−0.05	11.9374

Substituting the sums, we get

$$b = \frac{(10)(49145) - (495)(1024)}{(10)(27225) - (495)^2} = \frac{491450-506880}{272250-245025}$$

$$= \frac{-15430}{27225} = -0.57, \text{ and}$$

$$a = \frac{1024}{10} - (-0.57)\left(\frac{495}{10}\right) = 102.4 + 28.215 = 130.62.$$

Hence the equation of the desired regression line is

$$\hat{Y} = 130.62 - 0.57X.$$

The estimated regression line is shown on the scatterdiagram.

(c) The standard deviation of regression (standard error of estimate) is given by

$$s_{y.x} = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{11.9374}{8}} = \sqrt{1.4922} = 1.22.$$

10.10. **The necessary computations are given in the following table:**

X	Y	XY	X^2	Y^2
3.2	6.5	20.80	10.24	42.25
2.7	5.3	14.31	7.29	28.09
4.5	8.6	38.70	20.25	73.96
1.0	1.2	1.20	1.00	1.44
2.0	4.2	8.40	4.00	17.64
1.7	2.9	4.93	2.89	8.41
0.6	1.1	0.66	0.36	1.21
1.9	3.0	5.70	3.61	9.00
17.6	32.8	94.70	49.64	182.00

(a) The estimated least-squares regression equation for Y values on X values is

$$\hat{Y} = a + bX, \text{ where}$$

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} = \frac{(8)(94.70) - (17.6)(32.8)}{(8)(49.64) - (17.6)^2}$$

$$= \frac{757.60 - 577.28}{397.12 - 309.76} = \frac{180.32}{87.36} = 2.064, \text{ and}$$

$$a = \bar{Y} - b\bar{X} = \frac{32.8}{8} - (2.064)\left(\frac{17.6}{8}\right)$$

$$= 4.1 - 4.5408 = -0.441$$

Hence the desired regression equation is $\hat{Y} = -0.441 + 2.064X$.

(b) The standard error of estimate, $s_{y.x}$ is

$$s_{y.x} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$= \sqrt{\frac{182.00 - (-0.441)(32.8) - (2.064)(94.70)}{8-2}}$$

$$= \sqrt{\frac{182.00 + 14.4648 - 195.4608}{6}}$$

$$= \sqrt{\frac{1.004}{6}} = \sqrt{0.1673} = 0.41.$$

(c) The estimated least-squares regression equation for X values on Y values is

$$\hat{X} = a_0 + b_0 Y, \text{ where}$$

$$b_0 = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum Y^2 - (\sum Y)^2} = \frac{(8)(94.70) - (17.6)(32.8)}{(8)(182.00) - (32.8)^2}$$

$$= \frac{180.32}{380.16} = 0.474, \text{ and}$$

$$a_0 = \bar{X} - b_0\bar{Y} = 2.2 - (0.474)(4.1) = 0.257$$

Thus the desired least-squares regression equation is

$$\hat{X} = 0.257 + 0.474Y.$$

(d) The standard error of estimate, $s_{x.y}$ is

$$s_{x.y} = \sqrt{\frac{\sum X^2 - a_0\sum X - b_0\sum XY}{n-2}}$$

$$= \sqrt{\frac{49.64 - (0.257)(17.6) - (0.474)(94.70)}{8-2}}$$

$$= \sqrt{\frac{49.64 - 4.5232 - 44.8878}{6}}$$

$$= \sqrt{\frac{0.2290}{6}} = \sqrt{0.0382} = 0.20.$$

10.11. (b) The computation of the co-efficient of determination.

Income (X) (000)	Expenditure (Y) (000)	X^2	Y^2	XY
10	7	100	49	70
20	21	400	441	420
30	23	900	529	690
40	34	1600	1156	1360
50	36	2500	1296	1800
60	53	3600	2809	3180
210	174	9100	6280	7520

The co-efficient of determination, r^2, is given by the ratio

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}} = \frac{\sum(\hat{Y}-\bar{Y})^2}{\sum(Y-\bar{Y})^2} = 1-\frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2}$$

Now $\sum(Y-\hat{Y})^2 = \sum Y^2 - a\sum Y - b\sum XY$,

where $b = \dfrac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2} = \dfrac{(6)(7520) - (210)(174)}{(6)(9100) - (210)^2}$

$$= \frac{45120 - 36540}{54600 - 44100} = \frac{8580}{10500} = 0.8171, \text{ and}$$

$$a = \frac{174}{6} - (0.8171)\left(\frac{210}{6}\right) = 29 - 28.5985 = 0.4015$$

$\therefore \quad \sum(Y-\hat{Y})^2 = 6280 - (0.4015)(174) - (0.8171)(7520)$

$= 6280 - 69.861 - 6144.592 = 65.547$, and

$\sum(Y-\bar{Y})^2 = \sum Y^2 - (\sum Y)^2/n = 6280 - (174)^2/6$

$= 6280 - 5046 = 1234$

Hence $r^2 = 1 - \dfrac{65.547}{1234} = 1 - 0.053 = 0.947$.

Alternative Method:

$$r^2 = \left[\frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}\right]^2$$

$$= \left[\frac{(6)(7520) - (210)(174)}{\sqrt{[(6)(9100) - (210)^2][(6)(6280) - (174)^2]}}\right]^2$$

$$= \left[\frac{8580}{\sqrt{(10500)(7404)}}\right]^2 = \left[\frac{8580}{8817.1423}\right]^2 = (0.9731)^2 = 0.947$$

This means that 94.7% of the variation in expenditure (Y) is related to the variation in income (X).

10.12. (b) For the data given in question 10.11 (b), we have

(i) Total variation $= \sum(Y-\bar{Y})^2 = \sum Y^2 - \dfrac{(\sum Y)^2}{n}$

$= 6280 - \dfrac{(174)^2}{6} = 1234.$

(ii) Unexplaiend variation $= \sum(Y-\hat{Y})^2 = \sum Y^2 - a\sum Y - b\sum XY$

$= 6280-(0.4015)(174)-(0.8171)(7520)$

$= 6280-69.861-6144.592 = 65.547$, and

(iii) Explaiend variation = Total variation–Unexplained variation

$= 1234 - 65.547 = 1168.453.$

10.14. (b) Calculation of the Co-efficient of Correlation.

X	Y	XY	X^2	Y^2
2	18	36	4	324
4	12	48	16	144
5	10	50	25	100
6	8	48	36	64
8	7	56	64	49
11	5	55	121	25
36	60	293	266	706

$$r_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left\{\sum X^2 - \frac{(\sum X)^2}{n}\right\}\left\{\sum Y^2 - \frac{(\sum Y)^2}{n}\right\}}}$$

$$= \frac{293 - \frac{(36)(60)}{6}}{\sqrt{\left\{266 - \frac{(36)^2}{6}\right\}\left\{706 - \frac{(60)^2}{6}\right\}}}$$

$$= \frac{-67}{\sqrt{(50)(106)}} = \frac{-67}{72.80} = -0.92$$

(c) Computation of the correlation co-efficient, when each X value is multiplied by 2 and 6 is added to each product, and when each Y value is multiplied by 3 and 15 is subtracted from each product, *i.e.* when $u = 2X+6$ and $v = 3Y-15$.

u	v	uv	u^2	v^2
10	39	390	100	1521
14	21	294	196	441
16	15	240	256	225
18	9	162	324	81
22	6	132	484	36
28	0	0	784	0
108	90	1218	2144	2304

$$r_{uv} = \frac{\sum uv - \frac{(\sum u)(\sum v)}{n}}{\sqrt{\left\{\sum u^2 - \frac{(\sum u)^2}{n}\right\}\left\{\sum v^2 - \frac{(\sum v)^2}{n}\right\}}}$$

$$= \frac{1218 - \frac{(108)(90)}{6}}{\sqrt{\left\{2144 - \frac{(108)^2}{6}\right\}\left\{2304 - \frac{(90)^2}{6}\right\}}}$$

$$= \frac{-402}{\sqrt{(200)(954)}} = \frac{-402}{436.81} = -0.92$$

We obtain the same result as in part (b), because the correlation co-efficient is independent of the origin and scale.

10.15. (b) Computation of the correlation coefficient.

X	Y	$u = 10X-70$	$v = 10Y-60$	u^2	v^2	uv
8.2	8.7	12	27	144	729	324
9.6	9.6	26	36	676	1296	936
7.0	6.9	0	9	0	81	0
9.4	8.5	24	25	576	625	600
10.9	11.3	39	53	1521	2809	2067
7.1	7.6	1	16	1	256	16
9.0	9.2	20	32	400	1024	640
6.6	6.3	−4	3	16	9	−12
8.4	8.4	14	24	196	576	336
10.5	12.4	35	64	1225	4096	2240
Total	--	167	289	4755	11501	7147

$$\therefore \quad r = \frac{\sum uv - \frac{(\sum u)(\sum v)}{n}}{\sqrt{\left\{\sum u^2 - \frac{(\sum u)^2}{n}\right\}\left\{\sum v^2 - \frac{(\sum v)^2}{n}\right\}}}$$

$$= \frac{7147 - \frac{(167)(289)}{10}}{\sqrt{\left\{4755 - \frac{(167)^2}{10}\right\}\left\{11501 - \frac{(289)^2}{10}\right\}}}$$

$$= \frac{2320.7}{\sqrt{(1966.1)(3148.9)}} = \frac{2320.7}{2488.18} = 0.93.$$

10.16. (a) Given $r_{XY} = 0.7$.

(i) Since the co-efficient of correlation is symmetric, therefore

$$r_{YX} = r_{XY} = 0.7.$$

(ii) Here $u = -2X$, $v = 3Y$ so that $\bar{u} = -2\bar{X}$ and $\bar{v} = 3\bar{Y}$. Also $u-\bar{u} = -2(X-\bar{X})$ and $v-\bar{v} = 3(Y-\bar{Y})$

$$\therefore \quad r_{uv} = \frac{\sum(u-\bar{u})(v-\bar{v})}{\sqrt{\sum(u-\bar{u})^2 \sum(v-\bar{v})^2}} = \frac{(-2)(3)\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{(4)(9)\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}}$$

$$= \frac{-6}{\sqrt{36}} r_{XY} = -r_{XY} = -0.7.$$

because the signs of the co-efficient of X and Y are different.

(b) Computation of the co-efficient of correlation for the given sample values.

$$r = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{[\sum X^2 - n\bar{X}^2][\sum Y^2 - n\bar{Y}^2]}} = \frac{404-(20)(2)(8)}{\sqrt{[180-20(2)^2][1424-20(8)^2]}}$$

$$= \frac{404-320}{\sqrt{(100)(144)}} = \frac{84}{120} = 0.70$$

10.17. (i) Computation of the correlation co-efficient.

$$\text{Now} \quad r = \frac{\sum XY - \dfrac{(\sum X)(\sum Y)}{n}}{\sqrt{\left\{\sum X^2 - \dfrac{(\sum X)^2}{n}\right\}\left\{\sum Y^2 - \dfrac{(\sum Y)^2}{n}\right\}}}$$

$$= \frac{482788 - \dfrac{(2433)(4245)}{23}}{\sqrt{\left[281019 - \dfrac{(2433)^2}{23}\right]\left[841786 - \dfrac{(4245)^2}{23}\right]}}$$

$$= \frac{33740.83}{\sqrt{(23649.92)(58306.66)}} = \frac{33740.83}{37134.185} = 0.91$$

(ii) The high value of correlation co-efficient indicates that there does exist a relationship between X and Y. Therefore the estimated least-squares line of regression is

$$\hat{Y} = a + bX$$

The two normal equations are

$$\sum Y = na + b\sum X;$$

$$\sum XY = a\sum X + b\sum X^2.$$

Substituting the sums in the normal equations, we get

$$4245 = 23a + 2433b$$

$$482788 = 2433a + 281019b$$

Solving them simultaneously, we obtain

$$a = 33.3 \text{ and } b = 1.43.$$

Hence $\hat{Y} = 33.3 + 1.43X$, is the desired estimated least-squares line of regression.

10.18. Calculation of the co-efficient of correlation between traffic density (X) and accident rate (Y).

	X	Y	X^2	Y^2	XY
	30	2	900	4	60
	35	4	1225	16	140
	40	5	1600	25	200
	45	5	2025	25	225
	50	8	2500	64	400
	60	15	3600	225	900
	70	24	4900	576	1680
	80	30	6400	900	2400
	90	32	8100	1024	2880
Σ	500	125	31250	2959	8885

$$\therefore \quad r = \frac{\sum XY - (\sum X)(\sum Y)/n}{\sqrt{\sum X^2 - (\sum X)^2/n}\sqrt{\sum Y^2 - (\sum Y)^2/n}}$$

$$= \frac{8885 - (500)(125)/9}{\sqrt{31250 - (500)^2/9}\sqrt{2959 - (125)^2/9}}$$

$$= \frac{8885 - 6944.44}{\sqrt{3472.22}\sqrt{1222.89}} = \frac{1940.56}{2060.62} = 0.94$$

The correlation between traffic density and accident rate is positive and is 0.94.

The coefficient of determination $= r^2$

$$= (0.94)^2 = 0.8836$$

A value of $r^2 = 0.8836$ indicates that 88.36% of the variability in Y is explained by its linear relationship with the variable X and 11.64% of the variation is due to chance or other factors.

10.19. Calculaton of the correlation co-efficient between the marks of Economics Paper (X) and Physics Paper (Y).

Roll No.	X	Y	D_x $(X-50)$	D_y $(Y-50)$	D_x^2	Dy^2	D_xD_y
1	36	62	−14	+12	196	144	−168
2	56	42	6	−8	36	64	−48
3	41	60	−9	10	81	100	−90
4	46	53	−4	3	16	9	−12
5	59	36	9	−14	81	196	−126
6	46	50	−4	0	16	0	0
7	65	42	15	−8	225	64	−120
8	31	66	−19	16	361	256	−304
9	68	44	18	−6	324	36	−108
10	41	58	−9	8	81	64	−72
11	70	65	20	15	400	225	300
12	36	71	−14	21	196	441	−294
Σ	595	649	−5	49	2013	1599	-1042

$$\therefore \quad r = \frac{\sum D_xD_y - (\sum D_x)(\sum D_y)/n}{\sqrt{\sum D_x^2 - (\sum D_x)^2/n}\sqrt{\sum D_y^2 - (\sum D_y)^2/n}}$$

$$= \frac{-1042 - (-5)(49)/12}{\sqrt{2013 - (-5)^2/12}\sqrt{1599 - (49)^2/12}}$$

$$= \frac{-1042 + 20.42}{\sqrt{(2010.92)(1398.92)}} = \frac{-1021.58}{1679.91} = -0.61.$$

There is negative correlation between the marks in Economics and marks in Physics.

10.20. Calculation of the co-efficient of correlation.

	X	Y	D_x $(X-8)$	D_y $(Y-17)$	D_x^2	Dy^2	D_xD_y
	3	25	−5	8	25	64	−40
	4	24	−4	7	16	49	−28
	5	20	−3	3	9	9	−9
	6	20	−2	3	4	9	−6
	7	19	−1	2	1	4	−2
	8	17	0	0	0	0	0
	9	16	1	−1	1	1	−1
	10	13	2	−4	4	16	−8
	11	10	3	−7	9	49	−21
	12	6	4	−11	16	121	−44
Σ	75	170	−5	0	85	322	−159

Here $\bar{X} = \frac{\sum X}{n} = \frac{75}{10} = 7.5, \quad \bar{Y} = \frac{\sum Y}{n} = \frac{170}{10} = 17,$

$$S_x = \sqrt{\frac{\sum D_x^2}{n} - \left(\frac{\sum D_x}{n}\right)^2}$$

$$= \sqrt{\frac{85}{10} - \left(\frac{-5}{10}\right)^2} = \sqrt{85 - 0.25} = 2.87$$

$$S_y = \sqrt{\frac{\sum D_y^2}{n} - \left(\frac{\sum D_y}{n}\right)^2}$$

$$= \sqrt{\frac{322}{10} - \left(\frac{0}{10}\right)^2} = \sqrt{32.2} = 5.67$$

$$\therefore \quad r = \frac{\frac{\sum D_x D_y}{n} - \frac{\sum D_x}{n} \cdot \frac{\sum D_y}{n}}{S_x S_y}$$

$$= \frac{-15.9 - (-0.5)\,(0)}{(2.87)\,(5.67)} = \frac{-15.9}{16.27} = -0.98.$$

The equation to the regression line of Y on X is

$$Y - \bar{Y} = r \frac{S_y}{S_x} (X - \bar{X})$$

or $$Y - 17 = -0.98 \left(\frac{5.67}{2.87}\right) (X - 7.5)$$

$$= -1.94\,(X - 7.5)$$

or $$Y = 17 - 1.94X + 14.55 = 31.55 - 1.94X$$

The equation to the regression line of X on Y is

$$X - \bar{X} = r \frac{S_x}{S_y} (Y - \bar{Y})$$

or $$X - 7.5 = -0.98 \left(\frac{2.87}{5.67}\right) (Y - 17)$$

$$= -0.50\,(Y - 17)$$

or $$X = 7.5 - 0.50Y + 8.5 = 16.0 - 0.50Y.$$

10.21 (a) Calculation of the correlation co-efficient.

	X	Y	D_x $(X-18)$	D_y $(Y-20)$	D_x^2	D_y^2	$D_x D_y$
	5	11	−13	−9	169	81	117
	12	16	−6	−4	36	16	24
	14	15	−4	−5	16	25	20
	16	20	−2	0	4	0	0
	18	17	0	−3	0	9	0
	21	19	3	−1	9	1	−3
	22	25	4	5	16	25	20
	23	24	5	4	25	16	20
	25	21	7	1	49	1	7
Σ	156	168	−6	−12	324	174	205

$$\therefore \quad r = \frac{\sum D_x D_y - (\sum D_x)(\sum D_y)/n}{\sqrt{\sum D_x^2 - (\sum D_x)^2/n}\,\sqrt{\sum D_y^2 - (\sum D_y)^2/n}}$$

$$= \frac{205 - (-6)\,(-12)/9}{\sqrt{324 - (-6)^2/9}\,\sqrt{174 - (-12)^2/9}}$$

$$= \frac{205 - 8}{\sqrt{(320)\,(158)}} = \frac{197}{224.85} = 0.876.$$

(b) Calculation of the co-efficient of correlation between persons employed (X) and cloth manufactured (Y).

	X	Y	D_x $(X-176)$	D_y $(Y-40)$	$D_x D_y$	D_x^2	D_y^2
	137	23	−39	−17	663	1521	289
	209	47	33	7	231	1089	49
	113	22	−43	−18	774	1849	324
	189	40	13	0	0	169	0
	176	39	0	− 1	0	0	1
	200	51	24	11	264	576	121
	219	49	43	9	387	1849	81
Σ	1243	271	31	−9	2319	7053	865

$$\therefore \quad r = \frac{\sum D_x D_y - (\sum D_x)(\sum D_y)/n}{\sqrt{\sum D_x^2 - (\sum D_x)^2/n}\,\sqrt{\sum D_y^2 - (\sum D_y)^2/n}}$$

$$= \frac{2319 - (31)\,(-9)/7}{\sqrt{7053 - (31)^2/7}\,\sqrt{865 - (-9)^2/7}}$$

$$= \frac{2319 + 39.86}{\sqrt{(6915.71)\,(853.43)}} = \frac{2358.86}{2429.42} = 0.97.$$

Thus the co-efficient of correlation between persons employed and cloth manufactured is 0.97.

10.22. To find relation between age and blindness, we first calculate the numbers of blind per lakh (100,000) and then correlate with the midpoints of age groups.

Age Group	Mid. value X	Blinds per lakh Y	D_x $(X-44.5)$	D_y $(Y-150)$	D_xD_y	D_x^2	D_y^2
0–9	4.5	55	−40	−95	3800	1600	9025
10–19	14.5	67	−30	−83	2490	900	6889
20–29	24.5	100	−20	−50	1000	400	2500
30–39	34.5	111	−10	−39	390	100	1521
40–49	44.5	150	0	0	0	0	0
50–59	54.5	200	10	50	500	100	2500
60–69	64.5	300	20	150	3000	400	22500
70–79	74.5	500	30	350	10500	900	122500
Total	--	--	−40	283	21680	4400	167435

$$\therefore \quad r = \frac{\sum D_xD_y - (\sum D_x)(\sum D_y)/n}{\sqrt{\sum D_x^2 - (\sum D_x)^2/n}\sqrt{\sum D_y^2 - (\sum D_y)^2/n}}$$

$$= \frac{21680 - (-40)(283)/8}{\sqrt{4400 - (-40)^2/8}\sqrt{167435 - (283)^2/8}}$$

$$= \frac{21680 + 1415}{\sqrt{(4200)(157424)}} = \frac{23095}{25713} = 0.898.$$

The co-efficient of correlation is positive and very high, implying that blindness increases with age.

10.23. The corrected sums are calculated first as below:

Corrected $\sum X = 125 - 6 - 8 + 8 + 6 = 125$

Corrected $\sum Y = 100 - 14 - 6 + 12 + 8 = 100$

Corrected $\sum X^2 = 650 - (6^2 + 8^2) + (8^2 + 6^2) = 650$

Corrected $\sum Y^2 = 640 - (14^2 + 6^2) + (12^2 + 8^2) = 436$

Corrected $\sum XY = 508 - (6 \times 14 + 8 \times 6) + (12 \times 8 + 6 \times 8)$

$= 508 - (84 + 48) + (96 + 48) = 520$

Here $$r = \frac{\frac{\sum XY}{n} - \left(\frac{\sum X}{n}\right)\left(\frac{\sum Y}{n}\right)}{\sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}\sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2}}$$

$$= \frac{\frac{520}{25} - \left(\frac{125}{25}\right)\left(\frac{100}{25}\right)}{\sqrt{\frac{650}{25} - \left(\frac{125}{25}\right)^2}\sqrt{\frac{436}{25} - \left(\frac{100}{25}\right)^2}}$$

$$= \frac{20.8 - 20}{\sqrt{(26-25)(17.44-16)}} = \frac{0.8}{1.2} = 0.67.$$

10.24. (a) Here $b_{yx} = -0.219$ and $b_{xy} = -0.785$

Since the regression coefficients aie negative, therefore r is given by

$$r = -\sqrt{b_{yx} \cdot b_{xy}} = -\sqrt{(-0.219)(-0.785)} = -0.415$$

(b) Here $b_{yx} = 0.648$ and $b_{xy} = 0.917$

Since the regression co-efficients are positive, therefore r is given by

$$r = +\sqrt{b_{yx} \cdot b_{xy}} = +\sqrt{(0.648)(0.917)} = +0.77$$

(c) Here $b_{yx} = 1.94$ and $b_{xy} = 0.15$

Since the regression co-efficients are positive, so r is given by

$$r = +\sqrt{b_{yx} \cdot b_{xy}} = +\sqrt{(1.94)(0.15)} = +0.54$$

(d) Here $b_{yx} = -1.96$, but the line of X on Y is given as $Y = 15.91 - 2.22X$ To determine b_{xy}, we should transform the above equatioin as follows:

$$-2.22X = Y - 15.91$$

or $$X = \frac{-Y}{2.22} + \frac{15.91}{2.22} = -0.45Y + 7.167$$

Thus $b_{xy} = -0.45$

Hence $r = -\sqrt{b_{yx} \times b_{xy}}$, as the regression coefficients are negative.

$$= -\sqrt{(-1.96)\,(-0.45)} = -0.94.$$

10.25. Taking $u = \dfrac{X-29.5}{10}$ and $v = \dfrac{Y-24.5}{10}$, the arithmetic is arranged in the table below:

Y_i \ X_j	u_j / v_i	9.5 (−2)	19.5 (−1)	29.5 (0)	39.5 (1)	49.5 (2)	f_i	$f_i v_i$	$f_i v_i^2$	$f_{ij}u_jv_i$
4.5	−2	3 (12)	1 (2)	--	--	--	4	−8	16	14
14.5	−1	12 (24)	8 (8)	14 (0)	1 (−1)	--	35	−35	35	31
24.5	0	2 (0)	13 (0)	40 (0)	12 (0)	3 (0)	70	0	0	0
34.5	1	---	3 (−3)	40 (0)	27 (27)	7 (14)	77	77	77	38
44.5	2	---	---	6 (0)	4 (8)	4 (16)	14	28	56	24
$f_{\cdot j}$		17	25	100	44	14	200	62	184	107
$f_{\cdot j}u_j$		−34	−25	0	44	28	13			
$f_{\cdot j}u_j^2$		68	25	0	44	56	193			
$f_{ij}u_jv_i$		36	7	0	34	30	107	←—— Check		

$$\therefore \quad r = \frac{\sum fuv - (\sum fu)(\sum fv)/n}{\sqrt{\left[\sum fu^2 - (\sum fu)^2/n\right]}\sqrt{\left[\sum fv^2 - (\sum fv)^2/n\right]}}$$

$$= \frac{107 - (13)\,(62)/200}{\left[\sqrt{193 - (13)^2/200}\right]\left[\sqrt{184 - (62)^2/200}\right]}$$

$$= \frac{107 - 4.03}{\sqrt{(192.155)\,(164.78)}} = \frac{102.97}{177.94} = 0.58.$$

The estimated equation to the regression line of Y on X is

$$Y - \bar{Y} = r\frac{s_y}{s_x}(X - \bar{X}),$$

where $\bar{Y} = b + \dfrac{\sum fv}{n} \times k = 24.5 + \dfrac{62 \times 10}{200} = 27.6;$

$$\bar{X} = a + \frac{\sum fu}{n} \times h = 29.5 + \frac{13 \times 10}{200} = 30.15;$$

$$s_x = \sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} \times h = \sqrt{\frac{193}{200} - \left(\frac{13}{200}\right)^2} \times 10 = 9.8; \text{ and}$$

$$s_y = \sqrt{\frac{\sum fv^2}{n} - \left(\frac{\sum fv}{n}\right)^2} \times k = \sqrt{\frac{184}{200} - \left(\frac{62}{200}\right)^2} \times 10 = 9.1.$$

Substituting the values, we get

$$Y - 27.6 = 0.58\left(\frac{9.1}{9.8}\right)(X - 30.15)$$

or $\quad Y - 27.6 = 0.54\,(X - 30.15)$

or $\quad Y = 0.54X - 16.28 + 27.6 = 11.32 + 0.54X,$

which is the desired regression equation of Y on X.

10.26. Let $u = \dfrac{X-130}{20}$ and $v = \dfrac{Y-63}{3}$, where X denotes weight in pounds and Y, the height in inches. The calculations are then arranged as in the following table:

Y_i	u / v	90 (−2)	110 (−1)	130 (0)	150 (1)	170 (2)	190 (3)	Total (f)	fv	fv^2	fuv
57	−2	--	--	--	--	1 (−4)	--	1	−2	4	−4
60	−1	8 (16)	21 (21)	8 (0)	1 (−1)	--	--	38	−38	38	36
63	0	3 (0)	50 (0)	57 (0)	12 (0)	--	2 (0)	124	0	0	0
66	1	1 (−2)	24 (−24)	54 (0)	19 (19)	3 (6)	--	101	101	101	−1
69	2	--	1 (−2)	8 (0)	5 (10)	3 (12)	--	17	34	68	20
72	3	--	--	2 (0)	--	--	1 (9)	3	9	27	9
Total (f)		12	96	129	37	7	3	284	104	238	60
fu		−24	−96	0	37	14	9	−60			
fu^2		48	96	0	37	28	27	236			
fuv		14	−5	0	28	14	9	60	←— Check		

$$\therefore \quad r = \frac{\sum fuv - (\sum fu)(\sum fv)/n}{\sqrt{\sum fu^2 - (\sum fu)^2/n}\,\sqrt{\sum fv^2 - (\sum fv)^2/n}}, \text{ where } n = \sum f$$

$$= \frac{60 - (-60)(104)/284}{\sqrt{236 - (-60)^2/284}\,\sqrt{238 - (104)^2/284}}$$

$$= \frac{60 + 21.97}{\sqrt{(223.32)(199.92)}} = \frac{81.97}{211.30} = 0.39$$

10.27 (b) Here

$$r = \frac{\sum fuv - (\sum fu)(\sum fv)/n}{\sqrt{\sum fu^2 - (\sum fu)^2/n}\,\sqrt{\sum fv^2 - (\sum fv)^2/n}}$$

$$= \frac{91 - (-4)(-53)/66}{\sqrt{109 - (-4)^2/66}\,\sqrt{115 - (-53)^2/66}}$$

$$= \frac{87.79}{\sqrt{(108.76)(72.56)}} = \frac{87.79}{88.83} = 0.9883$$

Now $u = \dfrac{x - 1250}{500}$ or $x = 1250 + 500\,u$

$\therefore \quad \bar{x} = 1250 + 500\left(\dfrac{-4}{66}\right) = 1250 - 30.30 = 1219.70,$

And $v = \dfrac{y - 500}{200}$ or $y = 500 + 200\,v$

$\therefore \quad \bar{y} = 500 + 200\left(\dfrac{-53}{66}\right) = 500 - 160.61 = 339.39$

$$s_x = h\sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2} = 500\sqrt{\frac{109}{66} - \left(\frac{-4}{66}\right)^2}$$

$$= 500\sqrt{1.6515 - 0.0037} = 500\,(1.2837) = 641.85,$$

$$s_y = k\sqrt{\frac{\sum fv^2}{n} - \left(\frac{\sum fv}{n}\right)^2} = 200\sqrt{\frac{115}{66} - \left(\frac{-53}{66}\right)^2}$$

$$= 200\sqrt{1.7424 - 0.6449} = 200\,(1.0476) = 209.52$$

Regression coefficients are:

$$b_{yx} = r \cdot \frac{s_y}{s_x} = 0.9883 \times \frac{209.52}{641.85} = 0.3226, \text{ and}$$

$$b_{xy} = r \cdot \frac{s_x}{s_y} = 0.9883 \times \frac{641.85}{209.52} = 3.0276.$$

Regression line of Y on X is $Y - \bar{Y} = b_{yx}(X - \bar{X})$,

i.e. $\quad Y - 339.39 = 0.3226\,(X - 1219.70)$

or $\quad Y = 0.3226X - 54.085$

Regression line of X on Y is $X - \bar{X} = b_{xy}(Y - \bar{Y})$

or $\quad X - 1219.70 = 3.0276\,(Y - 339.39)$ or $X = 3.0276Y + 192.163.$

10.28. Let $u = X_1 + X_2$ and $v = X_2 + X_3$. Then

$$r_{uv} = \frac{\sum(u - \bar{u})(v - \bar{v})}{\sqrt{\sum(u - \bar{u})^2 \sum(v - \bar{v})^2}}.$$

Now $\quad u - \bar{u} = (X_1 + X_2) - (\bar{X}_1 + \bar{X}_2), \quad (\because \bar{u} = \bar{X}_1 + \bar{X}_2)$

$\quad = (X_1 - \bar{X}_1) + (X_2 - \bar{X}_2),$ and

$\quad v - \bar{v} = (X_2 + X_3) - (\bar{X}_2 + \bar{X}_3), \quad (\because \bar{v} = \bar{X}_2 + \bar{X}_3)$

$\quad = (X_2 - \bar{X}_2) + (X_3 - \bar{X}_3).$

$\therefore \quad \sum(u - \bar{u})^2 = [(X_1 - \bar{X}_1) + (X_2 - \bar{X}_2)]^2$

$\quad = \sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2 + 2\sum(X_1 - \bar{X}_1)\sum(X_2 - \bar{X}_2)$

$\quad = nS^2 + nS^2 + 0 \quad (\because X_1$ and X_2 are uncorrelated)

$\quad = 2nS^2,$

$\therefore \quad \sum(v - \bar{v})^2 = \sum[(X_2 - \bar{X}_2) + (X_3 - \bar{X}_3)]^2$

$\quad = \sum(X_2 - \bar{X}_2)^2 + \sum(X_3 - \bar{X}_3)^2 + 2\sum(X_2 - \bar{X}_2)(X_3 - \bar{X}_3)$

$\quad = nS^2 + nS^2 + 0 \quad (\because X_2$ and X_3 are uncorrelated)

$\quad = 2nS^2,$ and

$\therefore \quad \sum(u - \bar{u})(v - \bar{v}) = \sum[(X_1 - \bar{X}_1) + (X_2 - \bar{X}_2)]\,[(X_2 - \bar{X}_2) + (X_3 - \bar{X}_3)]$

$\quad = \sum(X_1 - \bar{X}_1)(X_2 - \bar{X}_2) + \sum(X_1 - \bar{X}_1)(X_3 - \bar{X}_3)$

$\quad + \sum(X_2 - \bar{X}_2)^2 + \sum(X_2 - \bar{X}_2)(X_3 - \bar{X}_3).$

$\quad = 0 + 0 + nS^2 + 0 = nS^2.$

Substituting these values, we get

$$r_{uv} = \frac{nS^2}{\sqrt{(2nS^2)\,(2nS^2)}} = \frac{nS^2}{2nS^2} = \frac{1}{2}$$

10.29. (b) Calculation of the rank correlation co-efficient.

Here $d = 0, -8, 0, 0, 0, -1, 5, 2, 1, -1, -4, 3, -1, 2, -1, 3.$

$\therefore \quad \sum d^2 = 0 + 64 + 0 + 0 + 0 + 1 + 25 + 4 + 1 + 1 + 16 + 9 + 1 + 4 + 1 + 9 = 136$

Hence $r_s = 1 - \dfrac{6\sum d^2}{n(n^2-1)}$

$$= 1 - \frac{6 \times 136}{16(256-1)} = 1 - 0.2 = 0.8$$

10.30. (b) Calculations for the product moment co-efficient of correlation.

a	b	a^2	b^2	ab
7.4	8.5	54.76	72.25	62.90
9.0	6.1	81.00	37.21	54.90
11.0	2.4	121.00	5.76	26.40
2.5	6.7	6.25	44.89	16.75
4.6	12.6	21.16	158.76	57.96
6.5	3.3	42.25	10.89	21.45
41.0	39.6	326.42	329.76	258.36

$$r = \frac{n\sum ab - (\sum a)(\sum b)}{\sqrt{n\sum a^2 - (\sum a)^2}\sqrt{n\sum b^2 - (\sum b)^2}}$$

$$= \frac{(6)\,(258.36) - (41.0)\,(39.6)}{\sqrt{(6)\,(326.42) - (41.0)^2}\sqrt{(6)\,(329.76) - (39.6)^2}}$$

$$= \frac{1550.16 - 1623.60}{\sqrt{(277.52)\,(410.40)}} = \frac{-73.44}{337.48} = -0.22$$

Ranking the values and calculation of the rank correlation co-efficient.

a	b	Ranks a	Ranks b	$d(=a-b)$	d^2
7.4	8.5	3	2	1	1
9.0	6.1	2	4	−2	4
11.0	2.4	1	6	−5	25
2.5	6.7	6	3	3	9
4.6	12.6	5	1	4	16
6.5	3.3	4	5	−1	1
$\sum$	--	--	--	--	56

$\therefore \quad r_s = 1 - \dfrac{6\sum d^2}{n(n^2-1)}$

$$= 1 - \frac{6 \times 56}{6(36-1)} = 1 - 1.6 = -0.6.$$

10.31. (b) Calculation of the co-efficient of rank correlation.

Laboratory (X)	8	3	9	2	7	10	4	6	1	5	Total
Lecture (Y)	9	5	10	1	8	7	3	4	2	6	--
$d(=X-Y)$	−1	−2	−1	1	−1	3	1	2	−1	−1	--
d^2	1	4	1	1	1	9	1	4	1	1	24

$\therefore \quad r_s = 1 - \dfrac{6\sum d^2}{n(n^2-1)}$

$$= 1 - \frac{6 \times 24}{10(100-1)} = 1 - 0.14556 = 0.8545$$

10.32. Denoting the judges by 1, 2 and 3, the calculations are given below:

Judge 1	Judge 2	Judge 3	d_{12}	d_{12}^2	d_{23}	d_{23}^2	d_{13}	d_{13}^2
1	3	6	−2	4	−3	9	−5	25
6	5	4	1	1	1	1	2	4
5	8	9	−3	9	−1	1	−4	16
10	4	8	6	36	−4	16	2	4
3	7	1	−4	16	6	36	2	4
2	10	2	−8	64	8	64	0	0
4	2	3	2	4	−1	1	1	1
9	1	10	8	64	−9	81	−1	1
7	6	5	1	1	1	1	2	4
8	9	7	−1	1	2	4	1	1
Σ	--	--	--	200	--	214	--	60

$$\therefore \quad r_{s(12)} = 1 - \frac{6\sum d_{12}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 200}{10(100-1)} = 1 - 1.21 = -0.21:$$

$$r_{s(23)} = 1 - \frac{6\sum d_{23}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 214}{10 \times 99} = 1 - 1.30 = -0.30; \text{ and}$$

$$r_{s(13)} = 1 - \frac{6\sum d_{13}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 60}{10(100-1)} = 1 - 0.36 = +0.64$$

The co-efficient of rank correlation between the first Judge and the third Judge is positive, meaning that the Judges 1 and 3 have the nearest approach to the tastes in beauty.

10.33. Calculation of Spearman's rank correlation coefficients.

Entry	Judge X	Judge Y	Judge Z	d_{xy}	d_{xy}^2	d_{yz}	d_{yz}^2	d_{xz}	d_{xz}^2
A	5	1	6	4	16	−5	25	−1	1
B	2	7	4	−5	25	3	9	−2	4
C	6	6	9	0	0	−3	9	−3	9
D	8	10	8	−2	4	2	4	0	0
E	1	4	1	−3	9	3	9	0	0
F	7	5	2	2	4	3	9	5	25
G	4	3	3	1	1	0	0	1	1
H	9	8	10	1	1	−2	4	−1	1
K	3	2	5	1	1	−3	9	−2	4
L	10	9	7	1	1	2	4	3	9
Σ	--	--	--	--	62	--	82	--	54

Now $$r_{s(xy)} = 1 - \frac{6\sum d_{xy}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 62}{10(100-1)} = 1 - 0.38 = 0.62;$$

$$r_{s(yz)} = 1 - \frac{6\sum d_{yz}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 82}{10(100-1)} = 1 - 0.50 = 0.50;$$

$$r_{s(xz)} = 1 - \frac{6\sum d_{xz}^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 54}{10(100-1)} = 1 - 0.33 = 0.67.$$

The rank correlation co-efficient between Judges X and Z is the highest and is positive. We may therefore conclude that the pair of Judges (X, Z) has the nearest approach to common tastes.

10.34. (b) We observe that both the sets of ranks contain ties. The co-efficient of rank correlation is therefore calculated as below:

X	8	3	6.5	3	6.5	9	3	1	5	Total
Y	8	9	6.5	2.5	4	5	6.5	1	2.5	--
$d(=X-Y)$	0	−6	0	0.5	2.5	4	−3.5	0	2.5	0
d^2	0	36	0	0.25	6.25	16	12.25	0	6.25	77

For ties, we add the following quantity to $\sum d^2$

$$\frac{1}{12}(3^3-3)+\frac{1}{12}(2^3-2)+\frac{1}{12}(2^3-2)+\frac{1}{12}(2^3-2)$$

i.e. $2 + 0.5 + 0.5 + 0.5$ or 3.5

Hence $r_s = 1 - \dfrac{6(77+3.5)}{9(81-1)} = 1 - \dfrac{483}{720} = 1 - 0.67 = 0.33.$

10.35. Calculation of the co-efficient of concordance.

X	1	2	3	4	5	6	7	8	9	10
Y	7	10	4	1	6	8	9	5	2	3
Z	9	6	10	3	5	4	7	8	2	1
Total	17	18	17	8	16	18	23	21	13	14

Here $m = 3$ and $n = 10$.

$\therefore$ Mean $= \dfrac{m(n+1)}{2} = \dfrac{3(10+1)}{2} = 16.5$, and

$$S = (17-16.5)^2 + (18-16.5)^2 + \ldots + (13-16.5)^2 + (14-16.5)^2$$
$$= (0.5)^2 + (1.5)^2 + (0.5)^2 + (-8.5)^2 + (-0.5)^2 + (1.5)^2 + (6.5)^2 + (4.5)^2 + (-3.5)^2 + (-2.5)^2 = 158.50.$$

Hence the co-efficient of concordance, W, is

$$W = \frac{12\times S}{m^2(n^3-n)} = \frac{12\times 158.50}{9(1000-10)} = \frac{634}{2970} = 0.21.$$

●●●●●●●●●●

Chapter 11

MULTIPLE REGRESSION AND CORRELATION

11.2. The estimated multiple linear regression equation is

$$\hat{Y} = a + b_1X_1 + b_2X_2,$$

where a, b_1 and b_2 are the *least-squares estimates* of the parameters α, β_1 and β_2. The three normal equations are

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2,$$
$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2,$$
$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2.$$

The calculations needed to find a, b_1 and b_2 are shown below:

Y	X_1	X_2	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y
12	2	1	4	1	2	24	12
10	2	1	4	1	2	20	10
9	3	0	9	0	0	27	0
13	4	0	16	0	0	52	0
20	4	3	16	9	12	80	60
64	15	5	49	11	16	203	82

Substituting the sums in the normal equations, we get

$$5a + 15b_1 + 5b_2 = 64$$
$$15a + 49b_1 + 16b_2 = 203$$
$$5a + 16b_1 + 11b_2 = 82$$

Solving them simultaneously, we obtain

$$a = 3.88,\ b_1 = 2.09 \text{ and } b_2 = 2.65$$

which are the values of the desired least-squares estimates.

11.3. (a) The estimated multiple linear regression equation is

$$\hat{Y} = a + b_1X_1 + b_2X_2,$$

where a, b_1 and b_2 are the least-squares estimates of the parameters α, β_1 and β_2.

The three normal equation are

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2$$

$$\sum X_1Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2$$

$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2$$

The calculations needed to find a, b_1 and b_2 are shown below:

Y	X_1	X_2	X_1^2	X_2^2	X_1Y	X_2Y	X_1X_2
2	8	0	64	0	16	0	0
5	8	1	64	1	40	5	8
7	6	1	36	1	42	7	6
8	5	3	25	9	40	24	15
5	3	4	9	16	15	20	12
27	30	9	198	27	153	56	41

Substituting the sums in the normal equations, we get

$$5a + 30b_1 + 9b_2 = 27$$

$$30a + 198b_1 + 41b_2 = 153$$

$$9a + 41b_1 + 27b_2 = 56$$

Solving them simultaneously, we obtain

$$a = 4.49,\ b_1 = -0.04 \text{ and } b_2 = 0.64$$

Hence the desired estimated multiple linear regression is

$$\hat{Y} = 4.49 - 0.04X_1 + 0.64X_2$$

(b) The partial regression co-efficient b_1 measures the average change in Y for a unit change in X_1 while the effect of X_2 is held constant.

Similarly, b_2 measures the average change in Y for a unit change in X_2 while the effect of X_1 is held constant.

11.4. (a) The estimated multiple linear regression equation is

$$\hat{X}_1 = a + b_1X_2 + b_2X_3.$$

The three normal equations are

$$\sum X_1 = na + b_1\sum X_2 + b_2\sum X_3$$

$$\sum X_1X_2 = a\sum X_2 + b_1\sum X_2^2 + b_2\sum X_2X_3$$

$$\sum X_1X_3 = a\sum X_3 + b_1\sum X_2X_3 + b_2\sum X_3^2$$

The calculations needed to find a, b_1 and b_2 are shown below:

X_1	X_2	X_3	X_2^2	X_3^2	X_2X_3	X_1X_2	X_1X_3	X_1^2
1	1	2	1	4	2	1	2	1
4	8	8	64	64	64	32	32	16
1	3	1	9	1	3	3	1	1
3	5	7	25	49	35	15	21	9
2	6	4	36	16	24	12	8	4
4	10	6	100	36	60	40	24	16
16	33	28	235	170	188	103	88	47

Substituting the sums in the normal equations, we get

$$6a + 33b_1 + 28b_2 = 15$$

$$33a + 235b_1 + 188b_2 = 103$$

$$28a + 188b_1 + 170b_2 = 88$$

Solving them simultaneously, we obtain

$$a = 0.04,\ b_1 = 0.21 \text{ and } b_2 = 0.28$$

Hence the desired multiple linear regression is

$$\hat{X}_1 = 0.04 + 0.21X_2 + 0.28X_3$$

(b) The standard error of estimate, $s_{1.23}$ is obtained as below:

$$s_{1.23} = \sqrt{\frac{\sum(X_1-\hat{X}_1)^2}{n-3}} = \sqrt{\frac{\sum X_1^2 - a\sum X_1 - b_1\sum X_2X_1 - b_2\sum X_3X_1}{n-3}}$$

$$= \sqrt{\frac{47-(0.04)(15)-(0.21)(103)-(0.28)(88)}{6-3}}$$

$$= \sqrt{\frac{0.13}{3}} = \sqrt{0.0433} = 0.2082 = 0.21$$

(c) The co-efficient of multiple determination, $R^2{}_{1.23}$ is computed as

$$R^2_{1.23} = \frac{\sum(\hat{X}_1-\bar{X}_1)^2}{\sum(X_1-\bar{X}_1)^2}$$

$$= \frac{a\sum X_1 + b_1\sum X_1X_2 + b_2\sum X_1X_3 - \frac{(\sum X_1)^2}{n}}{\sum X_1^2 - \frac{(\sum X_1)^2}{n}}$$

$$= \frac{(0.04)(15)+(0.21)(103)+(0.28)(88) - (15)^2/6}{47 - \frac{(15)^2}{6}}$$

$$= \frac{9.37}{9.50} = 0.9863, \text{ and}$$

therefore the multiple correlation co-efficient $R_{1.23}$, is

$$R_{1.23} = \sqrt{0.9863} = 0.99.$$

11.5. (a) The estimated multiple regression of X_3 on X_1 and X_2 is

$$\hat{X}_3 = b_{3.12} + b_{31.2}X_1 + b_{32.1}X_2$$

where b's are the least squares estimates of population parameters.

The three normal equations are

$$\sum X_3 = nb_{3.12} + b_{31.2}\sum X_1 + b_{32.1}\sum X_2,$$

$$\sum X_1X_3 = b_{3.12}\sum X_1 + b_{31.2}\sum X_1^2 + b_{32.1}\sum X_1X_2,$$

$$\sum X_2X_3 = b_{3.12}\sum X_2 + b_{31.2}\sum X_1X_2 + b_{32.1}\sum X_2^2.$$

The calculations needed to find b's are shown in the table below:

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_2X_3	X_1X_3
3	16	90	9	256	8100	48	1440	270
5	10	72	25	100	5184	50	720	360
6	7	54	36	49	2916	42	378	324
8	4	42	64	16	1764	32	168	336
12	3	30	144	9	900	36	90	360
14	2	12	196	4	144	28	24	168
48	42	300	474	434	19008	236	2820	1818

Substituting the sums in the normal equations, we get

$$6b_{3.12} + 48b_{31.2} + 42b_{32.1} = 300$$
$$48b_{3.12} + 474b_{31.2} + 236b_{32.1} = 1818$$
$$42b_{3.12} + 236b_{31.2} + 434b_{32.1} = 2820$$

Solving them simultaneously, we obtain

$$b_{3.12} = 61.3993,\ b_{31.2} = -3.6461 \text{ and } b_{32.1} = 2.5385$$

Hence the desired estimated regression equation of X_3 on X_1 and X_2 is

$$\hat{X}_3 = 61.40 - 3.65X_1 + 2.54X_2.$$

(b) When $X_1 = 10$ and $X_2 = 6$, we get

$$X_3 = 61.40 - 3.65(10) + 2.54(6)$$
$$= 61.40 - 36.5 + 15.24 = 40.$$

(c) The multiple correlation co-efficient is the positive square root of $R^2{}_{3.12}$, where

$$R^2{}_{3.12} = \frac{\sum(\hat{X}_3-\bar{X}_3)^2}{\sum(X_3-\bar{X}_3)^2}$$

$$= \frac{b_{3.12}\sum X_3 + b_{31.2}\sum X_1X_3 + b_{32.1}\sum X_2X_3 - (\sum X_3)^2/n}{\sum X_3^2 - (\sum X_3)^2/n}$$

$$= \frac{(61.3993)(300) + (-3.6461)(1818) + (2.5385)(2820) - (300)^2/6}{19008 - (300)^2/6}$$

$$= \frac{18419.79 - 6628.6 + 7158.57 - 15000}{19008 - 15000}$$

$$= \frac{3949.75}{4008} = 0.9855$$

Hence $R_{3.12} = \sqrt{0.9855} = 0.99.$

Alternatively:

Now, $\bar{X}_1 = \frac{\sum X_1}{n} = \frac{48}{6} = 8;\ \bar{X}_2 = \frac{\sum X_2}{n} = \frac{42}{6} = 7;$

$$\bar{X}_3 = \frac{\sum X_3}{n} = \frac{300}{6} = 50.$$

$$S_1 = \sqrt{\frac{\sum X_1^2}{n} - \left(\frac{\sum X_1}{n}\right)^2} = \sqrt{\frac{474}{6} - \left(\frac{48}{6}\right)^2}$$

$$= \sqrt{79 - 64} = \sqrt{15} = 3.87;$$

$$S_2 = \sqrt{\frac{\sum X_2^2}{n} - \left(\frac{\sum X_2}{n}\right)^2} = \sqrt{\frac{434}{6} - \left(\frac{42}{6}\right)^2}$$

$$= \sqrt{72.33 - 49} = \sqrt{23.33} = 4.83;$$

$$S_3 = \sqrt{\frac{\sum X_3^2}{n} - \left(\frac{\sum X_3}{n}\right)^2} = \sqrt{\frac{19008}{6} - \left(\frac{300}{6}\right)^2}$$

$$= \sqrt{3168 - 2500} = \sqrt{668} = 25.84;$$

and $$r_{12} = \frac{\sum X_1X_2 - n\bar{X}_1\bar{X}_2}{\sqrt{[\sum X_1^2 - n\bar{X}_1^2]\,[\sum X_2^2 - n\bar{X}_2^2]}}$$

$$= \frac{236 - 6(8)(7)}{\sqrt{[474 - 6(64)]\,[434 - 6(49)]}} = \frac{236 - 336}{\sqrt{(90)(140)}}$$

$$= \frac{-100}{112.25} = -0.89;$$

$$r_{23} = \frac{\sum X_2X_3 - n\bar{X}_2\bar{X}_3}{\sqrt{[\sum X_2^2 - n\bar{X}_2^2]\,[\sum X_3^2 - n\bar{X}_3^2]}}$$

$$= \frac{2820 - 6(7)(50)}{\sqrt{[434 - 6(49)]\,[19008 - 6(2500)]}} = \frac{2820 - 2100}{\sqrt{(140)(4008)}}$$

$$= \frac{720}{749.08} = 0.96;$$

$$r_{31} = \frac{\sum X_1X_3 - n\bar{X}_1\bar{X}_3}{\sqrt{[\sum X_1^2 - n\bar{X}_1^2]\,[\sum X_3^2 - n\bar{X}_3^2]}}$$

$$= \frac{1818 - 2400}{\sqrt{(90)(4008)}} = \frac{-582}{600.6} = -0.97.$$

(a) The linear regression equation of x_3 on x_1 and x_2 is

$$\frac{x_3}{S_3} = \left(\frac{r_{32} - r_{13}r_{12}}{1 - r_{12}^2}\right)\frac{x_2}{S_2} + \left(\frac{r_{31} - r_{23}r_{12}}{1 - r_{12}^2}\right)\frac{x_1}{S_1},$$

where $x_1 = X_1 - \bar{X}_1$, $x_2 = X_2 - \bar{X}_2$ and $x_3 = X_3 - \bar{X}_3$.

Substituting the values, we get

$$\frac{x_3}{25.84} = \left(\frac{(0.96)-(-0.97)(-0.89)}{1-(-0.89)^2}\right)\frac{x_2}{4.83} + \left(\frac{-0.97-(0.96)(-0.89)}{1-(-0.89)^2}\right)\frac{x_1}{3.87}$$

Or $$\frac{x_3}{25.84} = \left(\frac{0.0967}{0.2079}\right)\frac{x_2}{4.83} + \left(\frac{-0.1156}{0.2079}\right)\frac{x_1}{3.87}$$

Or $x_3 = 2.49x_2 - 3.71x_1$

Or $(X_3 - 50) = 2.49\,(X_2 - 7) - 3.71\,(X_1 - 8)$

Or $X_3 = 62.25 - 3.71\,X_1 + 2.49X_2.$

(b) When $X_1 = 10$, and $X_2 = 6$, we get

$$X_3 = 62.25 - 3.71(10) + 2.49(6) = 40$$

$$\text{(c) } R^2_{3.12} = \frac{r^2_{31} + r^2_{32} - 2r_{12}r_{23}r_{31}}{1 - r^2_{12}}$$

$$= \frac{0.9409 + 0.9216 - 2(-0.89)(0.96)(-0.97)}{1 - 0.7921}$$

$$= \frac{1.8625 - 1.6575}{0.2079} = \frac{0.2050}{0.2079} = 0.9860$$

Hence $R_{3.12} = \sqrt{0.9860} = 0.993$.

11.6. (a) Calculations needed to find b's are shown in the table below:

Y	X_1	X_2	YX_1	YX_2	X_1X_2	X_1^2	X_2^2
57.5	78	2.75	4485.0	158.125	214.50	6084	7.5625
52.8	69	2.15	3643.2	113.520	148.35	4761	4.6225
61.3	77	4.41	4720.1	270.333	339.57	5929	19.4481
67.0	88	5.52	5896.0	369.840	485.76	7744	30.4704
53.5	67	3.21	3584.5	171.735	215.07	4489	10.3041
62.7	80	4.32	5016.0	270.864	345.60	6400	18.6624
56.2	74	2.31	4158.8	129.822	170.94	5476	5.3361
68.5	94	4.30	6439.0	294.550	404.20	8836	18.4900
69.2	102	3.71	7058.4	256.732	378.42	10404	13.7641
548.7	729	32.68	45001.0	2035.521	2702.41	60123	128.6602

Now $\bar{Y} = \frac{\sum Y}{n} = \frac{548.7}{9} = 60.9667,$

$$\bar{X}_1 = \frac{\sum X_1}{n} = \frac{729}{9} = 81,$$

$$\bar{X}_2 = \frac{\sum X_2}{n} = \frac{32.68}{9} = 3.6311,$$

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n} = 60123 - \frac{(729)^2}{9}$$

$$= 60123 - 59049 = 1074;$$

$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n} = 128.6602 - \frac{(32.68)^2}{9}$$

$$= 128.6602 - 118.6647 = 9.9955;$$

$$\sum x_1x_2 = \sum X_1X_2 - \frac{(\sum X_2)(\sum X_1)}{n} = 2702.41 - \frac{(729)(32.68)}{9}$$

$$= 2702.41 - 2647.08 = 55.33;$$

$$\sum x_1y = \sum X_1Y - \frac{(\sum X_1)(\sum Y)}{n} = 45001.0 - \frac{(729)(548.7)}{9}$$

$$= 45001.0 - 44444.7 = 556.3;$$

$$\sum x_2y = \sum X_2Y - \frac{(\sum X_2)(\sum Y)}{n} = 2035.521 - \frac{(32.68)(548.7)}{9}$$

$$= 2035.521 - 1992.3907 = 43.1303;$$

$$b_1 = \frac{(\sum x_1y)(\sum x_2^2) - (\sum x_2y)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

$$= \frac{(556.3)(9.9955) - (43.1303)(55.33)}{(1074)(9.9955) - (55.33)^2}$$

$$= \frac{5560.49665 - 2386.39950}{10735.167 - 3061.4089} = \frac{3174.09715}{7673.7581} = 0.4136$$

$$b_2 = \frac{(\sum x_2y)(\sum x_1^2) - (\sum x_1y)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

$$= \frac{(43.1303)(1074) - (556.3)(55.33)}{(1074)(9.9955) - (55.33)^2}$$

$$= \frac{46321.9422 - 30780.079}{7673.7581} = \frac{15541.8632}{7673.7581}$$

$$= 2.0253; \text{ and}$$

$$a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 = 60.9667 - (0.4136)(81) - (2.0253)(3.6311)$$

$$= 60.9667 - 33.5016 - 7.35406 = 20.111$$

Hence the fitted least squares regression is

$$\hat{Y} = 20.111 + 0.4136X_1 + 2.0253X_2$$

(b) When $X_1 = 75$ days and $X_2 = 3.15$ kg, then the predicted average length of infants is

$$\hat{Y} = 20.111 + (0.4136)(75) + (2.0253)(3.15)$$
$$= 20.111 + 31.02 + 6.3797 = 57.5 \text{ cm}$$

(c) Now $s_{Y.12} = \sqrt{\dfrac{\sum(Y-\hat{Y})^2}{n-3}}$, where

$$\sum(Y-\hat{Y})^2 = \sum Y^2 - a\sum Y - b_1\sum X_1Y - b_2\sum X_2Y$$
$$= 33773.65 - (20.111)(548.7) - (0.4136) \times (45001.0) - (2.0253)(2035.521)$$
$$= 33773.65 - 11034.9057 - 18612.4136 - 4122.54068$$
$$= 33773.65 - 33769.86 = 3.79$$

$$\therefore \quad s_{Y.12} = \sqrt{\frac{3.79}{6}} = \sqrt{0.6317} = 0.79$$

11.7. (b) Calculation of the multiple correlation coefficient $R_{1.23}$

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_2X_3	X_1X_3
4	2	8	16	4	64	8	16	32
3	5	10	9	25	100	15	50	30
2	3	13	4	9	169	6	39	26
4	2	15	16	4	225	8	30	60
6	1	17	36	1	289	6	17	102
7	4	16	49	16	256	28	64	[illegible]
8	5	20	64	25	400	40	100	160
34	22	99	194	84	1503	111	316	522

Now,
$$r_{12} = \frac{n\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_2^2 - (\sum X_2)^2}}$$
$$= \frac{7 \times 111 - (34)(22)}{\sqrt{7 \times 194 - (34)^2}\sqrt{7 \times 84 - (22)^2}}$$
$$= \frac{777 - 748}{\sqrt{(202)(104)}} = \frac{29}{145} = 0.2,$$

$$r_{13} = \frac{n\sum X_1X_3 - (\sum X_1)(\sum X_3)}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_3^2 - (\sum X_3)^2}}$$
$$= \frac{7 \times 522 - (34)(99)}{\sqrt{(202)}\sqrt{7 \times 1503 - (99)^2}}$$
$$= \frac{3654 - 3366}{\sqrt{(202)(72)}} = \frac{288}{381} = 0.76,$$

$$r_{23} = \frac{n\sum X_2X_3 - (\sum X_2)(\sum X_3)}{\sqrt{n\sum X_2^2 - (\sum X_2)^2}\sqrt{n\sum X_3^2 - (\sum X_3)^2}}$$
$$= \frac{7 \times 316 - (27)(99)}{\sqrt{(104)(720)}} = \frac{34}{274} = 0.12.$$

Hence
$$R_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{23}^2}$$
$$= \frac{(0.2)^2 + (0.76)^2 - 2(0.2)(0.12)(0.76)}{1 - 0.0144}$$
$$= \frac{0.6176 - 0.03648}{0.9856} = \frac{0.58112}{0.9856} = 0.59, \text{ so that}$$

$$R_{1.23} = \sqrt{0.59} = 0.768$$

11.8. (a) Calculation of the multiple correlations, etc.

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1X_2	X_2X_3	X_1X_3
32	3	2	1024	9	4	96	6	64
18	2	4	324	4	16	36	8	72
52	5	2	2704	25	4	260	10	104
16	1	5	256	1	25	16	5	80
42	4	3	1764	16	9	168	12	126
48	6	9	2304	36	81	288	54	432
208	21	25	8376	91	139	864	95	878

Now, $$r_{12} = \frac{n\sum X_1X_2 - (\sum X_1)(\sum X_2)}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_2^2 - (\sum X_2)^2}}$$

$$= \frac{6 \times 684 - (208)(21)}{\sqrt{6 \times 8376 - (208)^2}\sqrt{6 \times 91 - (21)^2}}$$

$$= \frac{5184 - 4368}{\sqrt{(6992)(105)}} = \frac{816}{857} = 0.952;$$

$$r_{13} = \frac{n\sum X_1X_3 - (\sum X_1)(\sum X_3)}{\sqrt{n\sum X_1^2 - (\sum X_1)^2}\sqrt{n\sum X_3^2 - (\sum X_3)^2}}$$

$$= \frac{6 \times 878 - (208)(25)}{\sqrt{6992}\sqrt{6 \times 139 - (25)^2}}$$

$$= \frac{5268 - 5200}{\sqrt{(6992)(209)}} = \frac{68}{1209} = 0.056;$$

$$r_{23} = \frac{n\sum X_2X_3 - (\sum X_2)(\sum X_3)}{\sqrt{n\sum X_2^2 - (\sum X_2)^2}\sqrt{n\sum X_3^2 - (\sum X_3)^2}}$$

$$= \frac{6 \times 95 - (21)(25)}{\sqrt{(105)(209)}} = \frac{45}{148} = 0.304.$$

Hence $$R_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{23}^2}$$

$$= \frac{(0.952)^2 + (0.056)^2 - 2(0.952)(0.304)(0.056)}{1 - (0.304)^2}$$

$$= \frac{0.877026}{0.907584} = 0.9663; \text{ so that}$$

$$R_{1.23} = \sqrt{0.9663} = 0.98;$$

$$R_{2.13}^2 = \frac{r_{21}^2 + r_{23}^2 - 2r_{12}r_{23}r_{31}}{1 - r_{13}^2}$$

$$= \frac{(0.952)^2 + (0.304)^2 - 2(0.952)(0.304)(0.056)}{1 - (0.056)^2}$$

$$= \frac{0.966306}{0.996864} = 0.9693; \text{ so that}$$

$$R_{2.13} = \sqrt{0.9693} = 0.98; \text{ and}$$

$$R_{3.12}^2 = \frac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{12}^2}$$

$$= \frac{(0.056)^2 + (0.304)^2 - 2(0.952)(0.304)(0.056)}{1 - (0.952)^2}$$

$$= \frac{0.063138}{0.093696} = 0.6739; \text{ so that}$$

$$R_{3.12} = \sqrt{0.6739} = 0.82.$$

(b) The estimated regression equation of x_3 on x_1 and x_2 (in deviation form) in terms of standard deviations and the linear correlation coefficients is

$$\frac{x_3}{S_3} = \left(\frac{r_{13} - r_{12}r_{23}}{1 - r_{12}^2}\right)\left(\frac{x_1}{S_1}\right) + \left(\frac{r_{23} - r_{12}r_{13}}{1 - r_{12}^2}\right)\left(\frac{x_2}{S_2}\right)$$

Substituting the values, we get

$$\frac{x_3}{1.5} = \left(\frac{0.4-(-0.2)(0.5)}{1-(-0.2)^2}\right)\left(\frac{x_1}{1.0}\right) + \left(\frac{0.5-(-0.2)(0.4)}{1-(-0.2)^2}\right)\left(\frac{x_2}{2.0}\right)$$

or $$\frac{x_3}{1.5} = \left(\frac{0.5}{0.96}\right)\left(\frac{x_1}{1.0}\right) + \left(\frac{0.58}{0.96}\right)\left(\frac{x_2}{2.0}\right)$$

or $x_3 = 0.78\,x_1 + 0.45\,x_2$

To obtain the regression equation in terms of original values, we replace x_1 by $X_1-\bar{X}_1$, x_2 by $X_2-\bar{X}_2$ and x_3 by $X_3-\bar{X}_3$. Thus

$$(X_3-12) = 0.78\,(X_1-20) + 0.45\,(X_2-36)$$

i.e. $$X_3 = 12 + 0.78X_1 - 15.60 + 0.45X_2 - 16.20$$

$$= -19.80 + 0.78X_1 + 0.45X_2$$

11.9. (b) To find the multiple correlation $R_{2.13}$, we first compute the simple correlations.

Now $b_{12} = 0.75$, $b_{13} = 0.58$, $b_{21} = 0.88$,

$b_{23} = 0.53$, $b_{31} = 1.68$ and $b_{32} = 1.30$. Therefore

$$r_{12} = \sqrt{b_{12}\times b_{21}} = \sqrt{(0.75)(0.88)} = 0.81,$$

$$r_{13} = \sqrt{b_{13}\times b_{31}} = \sqrt{(0.58)(1.68)} = 0.99, \text{ and}$$

$$r_{23} = \sqrt{b_{23}\times b_{32}} = \sqrt{(0.53)(1.30)} = 0.83.$$

Now $$R^2_{2.13} = \frac{r^2_{21} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1-r^2_{13}}$$

$$= \frac{(0.81)^2 + (0.83)^2 - 2(0.81)(0.99)(0.33)}{1-(0.99)^2}$$

$$= \frac{0.013846}{0.0199} = 0.6958$$

Hence $R_{2.13} = \sqrt{0.6958} = 0.83.$

low $$R^2_{2.13} = \frac{r^2_{21} + r^2_{23} - 2r_{12}r_{23}r_{31}}{1-r^2_{13}}$$

$$= \frac{(0.60)^2 + (0.65)^2 - 2(0.6)(0.7)(0.65)}{1-(0.70)^2}$$

$$= \frac{0.3600 + 0.4225 - 0.546}{1-0.49} = \frac{0.2365}{0.51} = 0.4637$$

Hence $R_{2.13} = \sqrt{0.4637} = 0.68.$

11.10. (a) The multiple correlation co-efficient, $R_{1.23}$ is given by

$$R^2_{1.23} = \frac{r^2_{21} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1-r^2_{23}}$$

$$= \frac{(0.8)^2 + (-0.7)^2 - 2(0.8)(-0.7)(-0.9)}{1-(-0.9)^2}$$

$$= \frac{0.122}{0.19} = 0.64, \text{ so that } R_{1.23} = \sqrt{0.64} = 0.80.$$

(b) To calculate the multiple correlatioin co-efficient and the partial correlation co-efficient, we first calculate the simple correlation co-efficients.

Thus $r_{12} = -\sqrt{b_{12}\times b_{21}} = -\sqrt{(-0.1)(-0.4)} = -0.20.$

$$r_{13} = +\sqrt{b_{13}\times b_{31}} = \sqrt{(0.27)(0.6)} = \sqrt{0.162} = 0.40, \text{ and}$$

$$r_{23} = +\sqrt{b_{23}\times b_{32}} = \sqrt{(0.67)(0.38)} = \sqrt{0.2546} = 0.50.$$

Now $$R^2_{2.13} = \frac{r^2_{21} + r^2_{23} - 2r_{12}r_{13}r_{23}}{1-r^2_{13}}$$

$$= \frac{(-0.2)^2+(0.5)^2-2(-0.2)(0.4)(0.5)}{1-(0.4)^2} = \frac{0.37}{0.84} = 0.44$$

Hence $R_{2.13} = \sqrt{0.44} = 0.66$

And $$r_{23.1} = \frac{r_{23} - r_{21}\,r_{31}}{\sqrt{1-r^2_{21}}\sqrt{1-r^2_{31}}} = \frac{0.5-(-0.2)(0.4)}{\sqrt{1-(-0.2)^2}\sqrt{1-(0.4)^2}}$$

$$= \frac{0.58}{\sqrt{(0.96)(0.84)}} = \frac{0.58}{0.90} = 0.64$$

11.11. (b) The estimated simple regression equation of X_1 on X_3 is

$$\hat{X}_1 = b_{1.3} + b_{13}X_3,$$

where $b_{13} = \dfrac{n\sum X_1X_3 - (\sum X_1)(\sum X_3)}{n\sum X_3^2 - (\sum X_3)^2}$ and $b_{1.3} = \bar{X}_1 - b_{13}\bar{X}_3$.

The estimated simple regression equation of X_2 on X_3 is

$$\hat{X}_2 = b_{2.3} + b_{23}X_3,$$

where $b_{23} = \dfrac{n\sum X_2X_3 - (\sum X_2)(\sum X_3)}{n\sum X_3^2 - (\sum X_3)^2}$ and $b_{2.3} = \bar{X}_2 - b_{23}\bar{X}_3$.

The computations needed to find the b's are given in the table below:

X_1	X_2	X_3	X_1X_3	X_2X_3	X_3^2
5	10	2	10	20	4
9	12	6	54	72	36
7	8	4	28	32	16
10	9	5	50	45	25
12	11	7	84	77	49
8	7	6	48	42	36
6	5	4	24	20	16
10	8	6	60	48	36
67	70	40	358	356	218

Now $\bar{X}_1 = \dfrac{\sum X_1}{n} = \dfrac{67}{8} = 8.375$, $\bar{X}_2 = \dfrac{\sum X_2}{n} = \dfrac{70}{8} = 8.75$, $\bar{X}_3 = 5$;

$$b_{13} = \frac{(8)(358) - (67)(40)}{(8)(218) - (40)^2} = \frac{184}{144} = 1.28,$$

$$b_{1.3} = 8.375 - (1.28)(5) = 1.975,$$

$$b_{23} = \frac{8(356) - (70)(40)}{8(218) - (40)^2} = \frac{48}{144} = 0.33, \text{ and}$$

$$b_{2.3} = 8.75 - (0.33)(5) = 7.10.$$

Hence the desired regression equations are

$$X_1 = 1.975 + 1.28X_3 \text{ and } X_2 = 7.10 + 0.33X_3$$

Next, we compute the residuals $X_{1.3} = X_1 - 1.975 - 1.28X_3$ and $X_{2.3} = X_2 - 7.10 - 0.33X_3$, and the simple correlation co-efficient between them. The necessary computations are given in the following table:

X_1	X_2	X_3	$X_{1.3}$	$X_{2.3}$	$X_{1.3}X_{2.3}$	$X_{1.3}^2$	$X_{2.3}^2$
5	10	2	+0.465	2.24	1.04160	0.2162	5.0176
9	12	6	−0.655	2.92	−1.91260	0.4290	8.5264
7	8	4	−0.095	−0.42	0.03990	0.0090	0.1764
10	9	5	1.625	0.25	0.40625	2.6406	0.0625
12	11	7	1.065	1.59	1.69335	1.1342	2.5281
8	7	6	−1.655	−2.08	3.44240	2.7390	4.3264
6	5	4	−1.095	−3.42	3.74490	1.1990	11.6964
10	8	6	0.345	−1.08	−0.37260	0.1190	1.1664
67	70	40	0	0	8.0832	8.4860	33.5002

Hence the co-efficient of correlation between $X_{1.3}$ and $X_{2.3}$, *i.e.* $r_{12.3}$ is obtained as

$$r_{12.3} = \frac{\sum X_{1.3}X_{2.3}}{\sqrt{\sum X_{1.3}^2 \sum X_{2.3}^2}} \quad (\because \sum X_{1.3} = \sum X_{2.3} = 0)$$

$$= \frac{8.0832}{\sqrt{(8.4860)(33.5002)}} = \frac{8.0832}{16.8607} = 0.48.$$

11.12. $$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.80-(-0.40)(-0.56)}{\sqrt{(1-0.16)(1-0.3136)}}$$

$$= \frac{0.80-0.224}{\sqrt{(0.84)(0.6864)}} = \frac{0.576}{0.7593} = 0.759,$$

$$r_{23.1} = \frac{r_{23} - r_{12}\,r_{13}}{\sqrt{(1-r_{12}^2)\,(1-r_{13}^2)}} = \frac{-0.56-(0.80)\,(-0.40)}{\sqrt{(1-0.64)\,(1-0.16)}}$$

$$= \frac{-0.56+0.32}{\sqrt{(0.36)\,(0.84)}} = \frac{-0.24}{0.55} = -0.436, \text{ and}$$

$$r_{13.2} = \frac{r_{13} - r_{12}\,r_{23}}{\sqrt{(1-r_{12}^2)\,(1-r_{23}^2)}} = \frac{-0.40-(0.80)\,(-0.56)}{\sqrt{(1-0.64)\,(1-0.3136)}}$$

$$= \frac{-0.40+0.448}{\sqrt{(0.36)\,(0.6864)}} = \frac{0.048}{0.497} = 0.097.$$

The linear regression equation of X_1 on X_2 and X_3 is

$$\frac{X_1 - \bar{X}_1}{S_1} = \left(\frac{r_{12}-r_{13}\,r_{23}}{1-r_{23}^2}\right)\frac{X_2 - \bar{X}_2}{S_2} + \left(\frac{r_{13}-r_{12}\,r_{23}}{1-r_{23}^2}\right)\frac{X_3-\bar{X}_3}{S_3}$$

Substituting the values, we get

$$\frac{X_1 - 28.02}{4.42} = \left(\frac{0.576}{0.6864}\right)\frac{X_2 - 4.91}{1.10} + \left(\frac{0.048}{0.6864}\right)\frac{X_3 - 594}{85}$$

Or $\quad X_1 - 28.02 = (3.37)\,(X_2 - 4.91) + (0.0038)\,(X_3 - 594)$

Or $\quad X_1 = 28.02 + 3.37X_2 - 16.5467 + 0.0038X_3 - 2.2572$

Or $\quad X_1 = 9.22 + 3.37X_2 + 0.0038X_3,$

which is the required regression equation for hay-crop on spring rainfall and accumulated temperature.

11.13. The regression equation for estimating marks obtained (X_1) is

$$\frac{X_1 - \bar{X}_1}{S_1} = \left(\frac{r_{12}-r_{13}\,r_{23}}{1-r_{23}^2}\right)\frac{X_2 - \bar{X}_2}{S_2} + \left(\frac{r_{13}-r_{12}\,r_{23}}{1-r_{23}^2}\right)\frac{X_3-\bar{X}_3}{S_3}$$

Substituting the values, we get

$$\frac{X_1 - 18.5}{11.2} = \left(\frac{0.60-(0.32)(0.35)}{1-0.1225}\right)\frac{X_2 - 100.6}{15.8} + \left(\frac{0.32-(0.60)\,(0.35)}{1-0.1225}\right)\frac{X_3 - 24}{6.0}$$

Or $\quad \dfrac{X_1 - 18.5}{11.2} = (0.5561)\,\dfrac{X_2 - 100.6}{15.8} + (0.1254)\,.\,\dfrac{X_3 - 24}{6.0}$

Or $\quad X_1 - 18.5 = 0.3942\,(X_2 - 100.6) + 0.2341\,(X_3 - 24)$

Or $\quad X_1 = 18.5 + 0.3942X_2 - 39.6565 + 0.2341X_3 - 5.6184$

$$= 0.39X_2 + 0.23X_3 - 26.77.$$

Partial correlaton co-efficients:

$$r_{12.3} = \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{(1-r_{13}^2)\,(1-r_{23}^2)}} = \frac{0.60-(0.32)\,(0.35)}{\sqrt{(1-0.1024)\,(1-0.1225)}}$$

$$= \frac{0.60-0.1120}{\sqrt{(0.8976)\,(0.8775)}} = \frac{0.4880}{0.8875} = 0.55,$$

$$r_{13.2} = \frac{r_{13} - r_{12}\,r_{23}}{\sqrt{(1-r_{12}^2)\,(1-r_{23}^2)}} = \frac{0.32-(0.60)\,(0.35)}{\sqrt{(1-0.36)\,(1-0.1225)}}$$

$$= \frac{0.32-0.21}{\sqrt{(0.64)\,(0.8775)}} = \frac{0.11}{0.75} = 0.15, \text{ and}$$

$$r_{23.1} = \frac{r_{23} - r_{21}\,r_{31}}{\sqrt{(1-r_{21}^2)\,(1-r_{31}^2)}} = \frac{0.35-(0.60)\,(0.32)}{\sqrt{(1-0.36)\,(1-0.1024)}}$$

$$= \frac{0.35-0.192}{\sqrt{(0.64)\,(0.8976)}} = \frac{0.158}{0.758} = 0.21.$$

Interpretation:

(i) $r_{12.3} = 0.55$ represents the correlation co-efficient between the marks obtained and general intelligence scores for college students having the same number of hours of study.

(ii) $r_{13.2} = 0.15$ represents the correlaton co-efficient between marks obtained and hours of studies for college students having the same general intelligence scores.

(iii) $r_{23.1} = 0.21$ represents the correlation co-efficient between general intelligence scores and hours of studies for college students having the same marks obtained.

11.14 (b) The equations of the three regression planes are

$$x_1 = b_{12.3}x_2 + b_{13.2}x_3,$$
$$x_2 = b_{21.3}x_1 + b_{23.1}x_3,$$
$$x_3 = b_{31.2}x_1 + b_{32.1}x_2.$$

Comparing, we find

$$b_{12.3} = 0.41, \quad b_{13.2} = 0.23,$$
$$b_{21.3} = 0.96, \quad b_{23.1} = -0.025,$$
$$b_{31.2} = 1.04, \quad b_{32.1} = -0.05.$$

Now, $r_{12.3} = \sqrt{b_{12.3} \times b_{21.3}} = \sqrt{(0.41)(0.96)} = \sqrt{0.3936} = 0.63;$

$r_{13.2} = \sqrt{b_{13.2} \times b_{31.2}} = \sqrt{(0.23)(1.04)} = \sqrt{0.2392} = 0.49$, and

$$r_{23.1} = -\sqrt{b_{23.1} \times b_{32.1}} = -\sqrt{(-0.025)(-0.05)}$$
$$= -\sqrt{0.00125} = -0.035, \text{ (both } b_{23.1} \text{ and } b_{32.1} \text{ are negative)}$$

Again, $$r_{12} = \frac{r_{12.3} + r_{13.2}\, r_{23.1}}{\sqrt{(1-r_{13.2}^2)(1-r_{23.1}^2)}} = \frac{0.63 + (0.49)(-0.035)}{\sqrt{1-(0.49)^2}\sqrt{1-(0.035)^2}}$$

$$= \frac{0.63 - 0.01715}{\sqrt{(0.7599)(0.998775)}} = \frac{0.61285}{0.8694} = 0.70,$$

$$r_{13} = \frac{r_{13.2} + r_{12.3}\, r_{23.1}}{\sqrt{(1-r_{12.3}^2)(1-r_{23.1}^2)}} = \frac{0.49 + (0.63)(-0.035)}{\sqrt{[1-(0.63)^2][1-(-0.035)^2]}}$$

$$= \frac{0.49 - 0.02205}{\sqrt{(0.6031)(0.998775)}} = \frac{0.46795}{0.7761} = 0.60,$$

$$r_{23} = \frac{r_{23.1} + r_{12.3}\, r_{13.2}}{\sqrt{(1-r_{12.3}^2)(1-r_{13.2}^2)}} = \frac{-0.035 + (0.63)(0.49)}{\sqrt{[1-(0.63)^2][1-(0.49)^2]}}$$

$$= \frac{-0.035 + 0.3087}{\sqrt{(0.60331)(0.7599)}} = \frac{0.2737}{0.6770} = 0.40.$$

11.15.(a)The regression equation of X_1 on X_2 and X_3 is

$$X_1 = a + b_{12.3} X_2 + b_{13.2} X_3.$$

Let $x_1 = X_1 - \bar{X}_1$, $x_2 = X_2 - \bar{X}_2$ and $x_3 = X_3 - \bar{X}_3$. Then the regression equation can be written more simply as

$$x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3.$$

The two normal equations are

$$\sum x_1 x_2 = b_{12.3} \sum x_2^2 + b_{13.2} \sum x_2 x_3$$
$$\sum x_1 x_3 = b_{12.3} \sum x_2 x_3 + b_{13.2} \sum x_3^2$$

Expressing the summations in terms of variances and zero order correlation co-efficients, we get

$$nS_1S_2r_{12} = b_{12.3}\, nS_2^2 + b_{13.2} nS_2S_3\, r_{23}$$
$$nS_1S_3r_{13} = b_{12.3}\, nS_2S_3\, r_{23} + b_{13.2}\, nS_3^2$$

Simplifying, we obtain

$$S_1\, r_{12} = b_{12.3}\, S_2 + b_{13.2}\, S_3\, r_{23},$$
$$S_1\, r_{13} = b_{12.3}\, S_2\, r_{23} + b_{13.2}\, S_3$$

Solving them simultaneously, we get

$$b_{12.3} = \frac{S_1}{S_2}\left(\frac{r_{12} - r_{13}\, r_{23}}{1 - r_{23}^2}\right), \quad b_{13.2} = \frac{S_1}{S_3}\left(\frac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\right)$$

Similarly from the regression equation

$x_3 = b_{32.1}\, x_2 + b_{31.2}\, x_1$, we find that

$$b_{32.1} = \frac{S_3}{S_2}\left(\frac{r_{23} - r_{13}\, r_{12}}{1 - r_{12}^2}\right), \text{ and } b_{31.2} = \frac{S_3}{S_1}\left(\frac{r_{13} - r_{23}\, r_{12}}{1 - r_{12}^2}\right)$$

Hence $b_{13.2} \times b_{31.2} = \dfrac{S_1}{S_3}\left(\dfrac{r_{13} - r_{12}\, r_{23}}{1 - r_{23}^2}\right) \cdot \dfrac{S_3}{S_1}\left(\dfrac{r_{13} - r_{23}\, r_{12}}{1 - r_{12}^2}\right)$

$$= \frac{(r_{13} - r_{12}\, r_{23})^2}{(1 - r_{12}^2)(1 - r_{23}^2)} = r_{13.2}^2 .$$

(b) (i) Here $r_{12.3} = \dfrac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$

$$= \frac{0.60 - (0.80)(-0.50)}{\sqrt{(1-0.64)(1 - 0.25)}} = \frac{0.60+0.40}{\sqrt{(0.36)(0.75)}}$$

$$= \frac{1.00}{0.52} = 1.92.$$

This is an impossible value of $r_{12.3}$ as it should be numerically less than 1. Hence it is not possible to get these correlations from a set of experimental data.

(ii) Here $r_{12.3} = \dfrac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$

$$= \frac{0.6 - (-0.4)(0.7)}{\sqrt{(1-0.16)(1 - 0.49)}} = \frac{0.6+0.28}{\sqrt{(0.84)(0.51)}}$$

$$= \frac{0.88}{0.654} = 1.34.$$

This is an impossible value of $r_{12.3}$ as the partial correlation co-efficient is to be numerically less than 1. Hence there are good reasons to suspect these values by noting that while r_{12} and r_{23} are both large and positive, r_{13} is fairly large and negative, contrary to what would be expected.

(iii) Here $r_{12.3} = \dfrac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$

$$= \frac{0.01 - (0.66)(-0.70)}{\sqrt{(1-(0.66)^2)(1-(-0.7)^2)}} = \frac{0.01+0.462}{\sqrt{(0.5644)(0.51)}}$$

$$= \frac{0.472}{0.5365} = 0.88,$$

$$r_{13.2} = \frac{r_{13} - r_{12}\, r_{32}}{\sqrt{(1 - r_{12}^2)(1 - r_{32}^2)}}$$

$$= \frac{0.66 - (0.01)(-0.70)}{\sqrt{(1-(.01)^2)(1-(0.7)^2)}} = \frac{0.66+0.007}{\sqrt{(0.9999)(0.51)}}$$

$$= \frac{0.667}{0.7141} = 0.93, \text{ and}$$

$$r_{23.1} = \frac{r_{23} - r_{21}\, r_{31}}{\sqrt{(1 - r_{21}^2)(1 - r_{31}^2)}}$$

$$= \frac{-0.70 - (0.01)(0.66)}{\sqrt{(1-(0.01)^2)(1-(0.66)^2)}}$$

$$= \frac{-0.70+0.0066}{\sqrt{(0.9999)(0.5644)}} = \frac{-0.7066}{0.7512} = -0.94.$$

It is possible to obtain the given correlations from a set of experimental data.

11.16. Substituting the given values in the formula for $r_{13.2}$, *i.e.*

$$r_{13.2} = \frac{r_{13} - r_{21}\, r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}, \text{ we get}$$

$$r_{13.2} = \frac{-0.641 - (0.370)(-0.736)}{\sqrt{(1 - (0.370)^2)(1 - (-0.736)^2)}}$$

$$= \frac{-0.641 + 0.27232}{\sqrt{(0.8631)(0.458304)}} = \frac{-0.36868}{0.62894} = -0.586.$$

Next, by definition, $r_{42} = \dfrac{\text{Cov}(X_4, X_2)}{\sqrt{\text{Var}(X_4)\,\text{Var}(X_2)}}$.

Since $X_4 = X_1 + X_2$, therefore $\bar{X}_4 = \bar{X}_1 + \bar{X}_2$ and

$$X_4 - \bar{X}_4 = (X_1 - \bar{X}_1) + (X_2 - \bar{X}_2).$$

Now
$$\text{Cov}(X_4, X_2) = \frac{1}{n}\sum(X_4 - \bar{X}_4)(X_2 - \bar{X}_2)$$
$$= \frac{1}{n}\sum[(X_1-\bar{X}_1)+(X_2-\bar{X}_2)](X_2-\bar{X}_2)$$
$$= \frac{1}{n}\sum[(X_1-\bar{X}_1)(X_2-\bar{X}_2)+(X_2-\bar{X}_2)^2]$$
$$= r_{12}S_1S_2 + S_2^2;$$

$$\text{Var}(X_4) = \frac{1}{n}\sum(X_4 - \bar{X}_4)^2$$
$$= \frac{1}{n}\sum[(X_1-\bar{X}_1)+(X_2-\bar{X}_2)]^2$$
$$= \frac{1}{n}\sum[(X_1-\bar{X}_1)^2+(X_2-\bar{X}_2)^2+2(X_1-\bar{X}_1)(X_2-\bar{X}_2)]$$
$$= S_1^2 + S_2^2 + 2r_{12}S_1S_2: \text{ and}$$

$$\text{Var}(X_2) = \frac{1}{n}\sum(X_2-\bar{X}_2)^2 = S_2^2$$

Substituting these values, we get

$$r_{42} = \frac{S_1S_2r_{12} + S_2^2}{\sqrt{(S_1^2 + S_2^2 + 2S_1S_2r_{12})(S_2^2)}}$$
$$= \frac{(1)(1.3)(0.370) + (1.3)^2}{\sqrt{[(1)^2 + (1.3)^2 + 2(1)(1.3)(0.370)][(1.3)^2]}}$$
$$= \frac{0.481 + 1.69}{\sqrt{(3.652)(1.69)}} = \frac{2.171}{2.484} = 0.874.$$

Similarly, we find that $r_{43} = -0.836$.

Now
$$r_{43.2} = \frac{r_{43} - r_{42}r_{23}}{\sqrt{(1 - r_{42}^2)(1 - r_{23}^2)}}$$
$$= \frac{-0.836 - (0.874)(-0.736)}{\sqrt{1-(0.874)^2}\sqrt{1-(-0.736)^2}}$$
$$= \frac{-0.836 + 0.643264}{\sqrt{(0.236124)(0.458304)}} = \frac{-0.192736}{0.328963} = -0.586.$$

Thus the two partial correlation co-efficients are equal.

11.17. (b) (i) We are given that $R_{1.23} = 1$

Squaring, $R_{1.23}^2 = 1$

Or
$$\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{31}}{1 - r_{23}^2} = 1$$

Or
$$r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{31} = 1 - r_{23}^2$$

Then
$$r_{12}^2 + r_{23}^2 - 2r_{12}r_{23}r_{31} = 1 - r_{13}^2$$

Dividing both sides by $1 - r_{13}^2$, we get

$$\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{23}r_{31}}{1 - r_{13}^2} = 1$$

Or $R_{2.13}^2 = 1$

Or $R_{2.13} = 1$ ($\because$ multiple correlation co-efficient is considered non-negative)

(ii) We have shown in (i) that $R_{2.13} = 1$

Squaring, $R_{2.13}^2 = 1$

Or
$$\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{23}r_{31}}{1 - r_{13}^2} = 1$$

Or $r_{12}^2 + r_{23}^2 - 2r_{12}\,r_{23}\,r_{31} = 1 - r_{13}^2$

Or $r_{13}^2 + r_{23}^2 - 2r_{12}\,r_{23}\,r_{31} = 1 - r_{12}^2$

Or $\dfrac{r_{13}^2 + r_{23}^2 - 2r_{12}\,r_{23}\,r_{31}}{1 - r_{12}^2} = 1$

Or $R_{3.12}^2 = 1$ or $R_{3.12} = 1$

(c) **Now** $R_{1.23} = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}\,r_{13}\,r_{23}}{1 - r_{23}^2}}$

$R_{1.23} = 0$, if and only if

$r_{12}^2 + r_{13}^2 - 2r_{12}\,r_{13}\,r_{23} = 0,$

i.e. if $r_{12}^2 + r_{13}^2 = 2r_{12}\,r_{13}\,r_{23}$

Again $R_{2.13} = \sqrt{\dfrac{r_{12}^2 + r_{23}^2 - 2r_{12}\,r_{13}\,r_{23}}{1 - r_{13}^2}}$

$$= \sqrt{\frac{r_{12}^2 + r_{23}^2 - (r_{12}^2 + r_{13}^2)}{1 - r_{13}^2}} = \sqrt{\frac{r_{23}^2 - r_{13}^2}{1 - r_{23}^2}}$$

which is not necessarily zero.

(d) **Given** $r_{12} = r_{23} = r_{13} = r \neq 1$

Now $R_{1.23}^2 = \dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}$

$$= \frac{r^2 + r^2 - 2r.r.r}{1 - r^2} = \frac{2r^2(1-r)}{1 - r^2} = \frac{2r^2}{1+r}$$

$\therefore$ $R_{1.23} = \dfrac{r\sqrt{2}}{\sqrt{1+r}}$

Similarly, we find that

$$R_{2.13} = \frac{r\sqrt{2}}{\sqrt{1+r}} = R_{3.12}.$$

Hence the result.

11.18. (a) $r_{12.3} = \dfrac{r_{12} - r_{13}\,r_{23}}{\sqrt{(1-r_{13}^2)\,(1 - r_{23}^2)}} = 0$, if and only if

$r_{12} - r_{13}r_{23} = 0,$

i.e. if $r_{12} = r_{13}r_{23}$

But $r_{12} = 0,$

$\therefore$ $r_{13}r_{23} = 0$; if either $r_{13} = 0$ or $r_{23} = 0$.

(b) The mean of $aX_1 + bX_2 + cX_3 = 0$ is $a\bar{X}_1 + b\bar{X}_2 + c\bar{X}_3 = 0$. Substracting, we get

$a(X_1 - \bar{X}_1) + b(X_2 - \bar{X}_2) + c(X_3 - \bar{X}_3) = 0,$

i.e. $ax_1 + bx_2 + cx_3 = 0$, whcih can be written as

$x_3 = \left(-\dfrac{a}{c}\right)x_1 + \left(-\dfrac{b}{c}\right)x_2$, where x_1 and x_2 are independent and measured from their respective means.

Squaring and summing, we get

$$\sum x_3^2 = \frac{a^2}{c^2}\sum x_1^2 + \frac{b^2}{c^2}\sum x_2^2. \quad \text{(product term vanishes)}$$

$$= \frac{a^2}{c^2} \cdot nS_1^2 + \frac{b^2}{c^2} \cdot nS_2^2$$

Also $\sum x_1x_3 = \left(-\dfrac{a}{c}\right)\sum x_1^2 + \left(-\dfrac{b}{c}\right)\sum x_1x_2$

$$= \left(-\frac{a}{c}\right)nS_1^2 \qquad (\textstyle\sum x_1x_2 = 0)$$

Thus $r_{13} = \dfrac{\sum x_1x_3}{\sqrt{\sum x_1^2 \sum x_3^2}} = \dfrac{-\frac{a}{c}nS_1^2}{\sqrt{(nS_1^2)\left(\frac{a^2}{c^2}nS_1^2 + \frac{b^2}{c^2}nS_2^2\right)}}$

$$= \frac{-aS_1}{\sqrt{a^2S_1^2 + b^2S_2^2}}.$$

Similarly, we find that $r_{23} = \dfrac{-bS_2}{\sqrt{a^2S_1^2 + b^2S_2^2}}$ and $r_{12} = 0$

Now, $$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$= \frac{\dfrac{-aS_1}{\sqrt{a^2S_1^2 + b^2S_2^2}} - 0}{\sqrt{(1-0)\left(1 - \dfrac{b^2S_2^2}{a^2S_1^2 + b^2S_2^2}\right)}} = \frac{-aS_1}{+\sqrt{a^2S_1^2}} = -1;$$

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

$$= \frac{0 - \left(\dfrac{-aS_1}{\sqrt{a^2S_1^2 + b^2S_2^2}}\right)\left(\dfrac{-bS_2}{\sqrt{a^2S_1^2 + b^2S_2^2}}\right)}{\sqrt{\left(1 - \dfrac{a^2S_1^2}{a^2S_1^2 + b^2S_2^2}\right)\left(1 - \dfrac{b^2S_2^2}{a^2S_1^2 + b^2S_2^2}\right)}}$$

$$= \frac{-ab\,S_1S_2}{+\sqrt{a^2b^2S_1^2S_2^2}} = -1, \text{ and}$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}}$$

$$= \frac{\dfrac{-bS_2}{\sqrt{a^2S_1^2 + b^2S_2^2}} - 0}{\sqrt{\left(1 - \dfrac{b^2S_2^2}{a^2S_1^2 + b^2S_2^2}\right)(1-0)}} = \frac{-b\,S_2}{+\sqrt{b^2S_2^2}} = -1.$$

Hence $r_{13.2} = r_{12.3} = r_{23.1} = -1$

11.20. The estimated least-squares equation of the quadratic regression is

$$\hat{Y} = a + b_1X + b_2X^2,$$

where a, b_1 and b_2 are the estimates of the parameters.

The three normal equations are

$$\sum Y = na + b_1\sum X + b_2\sum X^2,$$
$$\sum XY = a\sum X + b_1\sum X^2 + b_2\sum X^3,$$
$$\sum X^2Y = a\sum X^2 + b_1\sum X^3 + b_2\sum X^4.$$

The computations needed to find a, b_1 and b_2 are shown below:

X	Y	X^2	X^3	X^4	XY	X^2Y	Y^2
−2	0.4	4	−8	16	−0.8	1.6	0.16
−1	1.3	1	−1	1	−1.3	1.3	1.69
0	2.2	0	0	0	0	0	4.84
1	2.5	1	1	1	2.5	2.5	6.25
2	3.0	4	8	16	6.0	12.0	9.00
0	9.4	10	0	34	6.4	17.4	21.94

Substituting the sums in the normal equations, we get

$$5a + 10b_2 = 9.4$$
$$10b_1 = 6.4$$
$$10a + 34b_2 = 17.4$$

Solving them, we obtain

$$a = 2.08,\ b_1 = 0.64 \text{ and } b_2 = -0.1.$$

Hence the equation of the desired quadratic regression is

$$\hat{Y} = 2.08 + 0.64X - 0.1X^2.$$

The standard error of estimate in this case is

$$s_{y.x.x^2} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 3}}$$

$$= \sqrt{\frac{\sum Y^2 - a\sum Y - b_1\sum XY - b_2\sum X^2Y}{n-3}}$$

$$= \sqrt{\frac{21.94-(2.08)(9.4)-(0.64)(6.4)-(-0.7)(17.4)}{5-3}}$$

$$= \sqrt{\frac{0.032}{2}} = \sqrt{0.016} = 0.13$$

●●●●●●●●●●

Chapter 12

CURVE FITTING BY LEAST SQUARES

12.1. (c) Let the equation of the straight line to be fitted to the data, be

$Y = a + bX$, where a and b are to be determined.

The two normal equations for determining a and b are

$$\sum Y = na + b\sum X,\ \sum XY = a\sum X + b\sum X^2.$$

The necessary calculations are shown in the following table:

	X	Y	XY	X^2	$Y=a+bX$	$D=Y-Y'$
	1	2	2	1	3.18	−1.18
	2	6	12	4	4.84	1.16
	3	7	21	9	6.50	0.50
	4	8	32	16	8.16	−0.16
	5	10	50	25	9.82	0.18
	6	11	66	36	11.48	−0.48
$\sum$	21	44	183	91	--	0

Substituting these totals in the normal equations, we get

$6a + 21b = 44$ and $21a + 91b = 183$

Solving, we obtain $a = 1.52$ and $b = 1.66$

Hence the equation of the desired straight line is

$$Y = 1.52 + 1.66X.$$

The values of deviations D_i are given in the last column of the above table.

12.2. (c) Let the equation of the straight line to be fitted to the given data, be

$Y = a + bX$, where a and b are to be evaluated.

The two normal equations are

$\sum Y = na + b\sum X, \quad \sum XY = a\sum X + b\sum X^2.$

The computations involved are shown in the table below:

	X	Y	XY	X^2	Y'
	0	5	0	0	7.20
	1	11	11	1	8.48
	2	8	16	4	9.76
	3	14	42	9	11.04
	4	10	40	16	12.32
	5	16	80	25	13.60
	6	12	72	36	14.88
	7	20	140	49	16.16
	8	15	120	64	17.44
$\sum$	36	111	521	204	--

Putting these values in the normal equtioins, we get

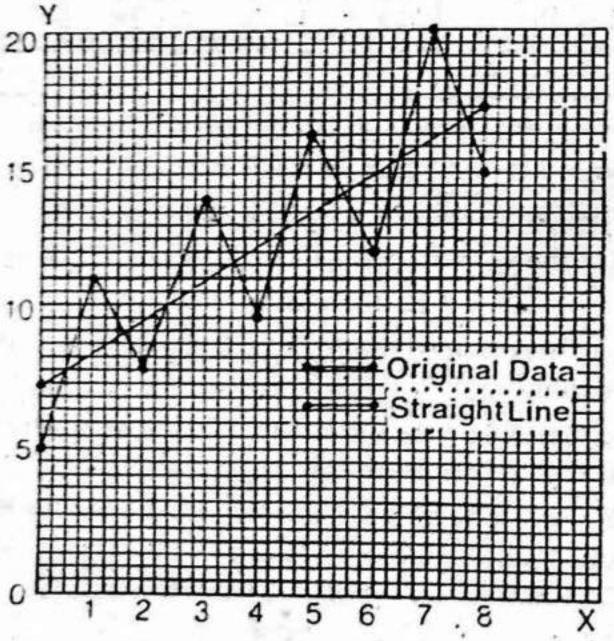
$$9a + 36b = 111, \; 36a + 204b = 521.$$

Solving them simultaneously, we obtain $a = 7.2$ and $b = 1.28$.

Hence the equation of the required straight line is $Y = 7.2 + 1.28X$.

The calculated values of $Y(=Y')$ from this equation are shown in the last column of the table.

Graph



12.3. (b) Let the equation to the least squares line be

$Y = a + bX$, where a and b are to be determined

The two normal equations are

$$\sum Y = na + b\sum X,$$

$$\sum XY = a\sum X + b\sum X^2.$$

The computations involved are shown below:

	Year (X)	Output (Y)	XY	X^2	Y'	$Y-Y'$	$(Y-Y')^2$
	1	1	1	1	1.67	−0.67	0.4489
	2	3	6	4	2.17	0.83	0.6889
	3	2	6	9	2.67	−0.67	0.4489
	4	4	16	16	3.17	0.83	0.6889
	5	3	15	25	3.67	−0.67	0.4489
	6	5	30	36	4.17	0.83	0.6889
	7	4	28	49	4.67	−0.67	0.4489
	8	6	48	64	5.17	0.83	0.6889
	9	5	45	81	5.67	−0.67	0.4489
$\sum$	45	33	195	285	33.03	--	5.0001

Putting these values in the normal equations, we get

$$9a + 45b = 33$$

$$45a + 285b = 195$$

Solving them simultaneously, we get $a = 1.17$ and $b = 0.50$

Hence the required least squares line is $Y = 1.17 + 0.50X$

Now, we find the values of $Y(Y')$ from this line by substituting the values of X. These values appear in the above table in column 5 and other calculations are also shown there.

The sum of the squared deviations $= \sum(Y-Y')^2 = 5.0001$

12.4. (b) Let the equation of the straight line be

$Y = a + bX$, where a and b are to be determined.

The normal equations are $\sum Y = na + b\sum X$ and $\sum XY = a\sum X + b\sum X^2$.

The necessary computations are shown in the following table:

	X	Y	XY	X^2
	0	12	0	0
	5	15	75	25
	10	17	170	100
	15	22	330	225
	20	24	480	400
	25	30	750	625
$\sum$	75	120	1805	1375

The normal equations then become

$$6a + 75b = 120 \text{ and } 75a + 1375b = 1805$$

Solving them simultaneously, we get $a = 11.25$ and $b = 0.70$.

Hence the required equation of the straight line is $Y = 11.25 + 0.70X$.

12.5. (a) Let the least squares line be Y=a+bX, where

$$b = \frac{n\sum XY - \sum X\sum Y}{n\sum X^2 - (\sum X)^2} = \frac{(\sum XY)/n - \bar{X}\,\bar{Y}}{(\sum X^2)/n - \bar{X}^2},$$

$$= \frac{(404)/20 - (2)(8)}{(180)/20 - (2)^2} = \frac{4.2}{5} = 0.84, \text{ and}$$

$$a = \bar{Y} - b\bar{X} = 8 - (0.84)(2) = 6.32.$$

Hence the desired least squares line is

$$Y = 6.32 + 0.84\,X.$$

(b) The normal equations in respect of the straight line $\dot{Y}$=a+bX are

$$\sum Y = na + b\sum X,$$

$$\sum XY = a\sum X + b\sum X^2.$$

And the normal equations when the straight line is $X = c + dY$, are

$$\sum X = nc + d\sum Y,$$

$$\sum XY = c\sum Y + d\sum Y^2.$$

The computations involved are shown in the following table:

	X	Y	XY	X^2	Y^2
	1	8	8	1	64
	2	9	18	4	81
	3	13	39	9	169
	4	18	72	16	324
	5	27	135	25	729
$\sum$	15	75	272	55	1367

Straight line $Y = a + bX$

Substituting these values, we get

$$5a + 15b = 75$$
$$15a + 55b = 272$$

Solving them simultaneously, we obtain

$$a = 0.9 \text{ and } b = 4.7$$

Hence the required equation is

$$Y = 0.9 + 4.7X$$

When $X = 6$, then

$$Y = 0.9 + 4.7(6) = 29.1$$

Straight line $X = c + dY$

Substituting these values, we get

$$5c + 75d = 15$$
$$75c + 1367d = 272$$

Solving them simultaneously, we obtain

$$c = 0.09 \text{ and } d = 0.19$$

Hence the required equation is

$$X = 0.09 + 0.19Y$$

When $X = 6$, then

$$6 = 0.09 + 0.19Y$$

or $Y = 31.1$

12.6(c) As $\sum X = 0 = \sum X^3$, the normal equations reduce to

$$\sum Y = na + c\sum X^2,\ \sum XY = b\sum X^2,\ \sum X^2Y = a\sum X^2 + c\sum X^4.$$

The arithmetic involved in computing the necessary summations is shown in the following table:

X	Y	X^2	X^4	XY	X^2Y
−2	−5	4	16	10	−20
−1	−2	1	1	2	−2
0	1	0	0	0	0
1	2	1	1	2	2
2	1	4	16	2	4
0	−3	10	34	16	−16

Substituting these values, we get

$$5a + 10c = -3$$

$$10b = 16$$

$$10a + 34c = -16$$

Solving them, we get

$$a = 0.83,\ b = 1.60 \text{ and } c = -0.71$$

Hence the equation of the fitted parabola is

$$Y = 0.83 + 1.60X - 0.71X^2$$

12.7. (b) Let the equation of the second degree parabola be

$$Y = a + bX + cX^2.$$

Then the normal equations are

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

Now $\quad \sum X = n(\bar{X}) = (5)(2) = 10$, and

$$\sum Y = n(\bar{Y}) = (5)(15) = 75.$$

Substituting the values in the normal equations, we get

$$5a + 10b + 30c = 75 \quad \text{... A}$$

$$10a + 30b + 100c = 242 \quad \text{... B}$$

$$30a + 100b + 354c = 850 \quad \text{... C}$$

Now 2A: $\quad 10a + 20b + 60c = 150 \quad \text{... D}$

D−B: $\quad 10b + 40c = 92 \quad \text{... E}$

Again 6A: $\quad 30a + 60b + 180c = 450 \quad \text{... F}$

C−F: $\quad 40b + 174c = 400 \quad \text{... G}$

Again 4E: $\quad 40b + 160c = 368 \quad \text{... H}$

H−G: $\quad 14c = 32$ or $c = 2.286$

Putting c = 2.286 in E, we get

$$10b + 40(2.286) = 92, \text{ which gives}$$

$$b = \frac{92 - 91.44}{10} = 0.056.$$

Now, putting $b = 0.056$ and $c = 2.286$ in A, we obtain

$$5a + 10(0.056) + 30(2.286) = 75$$

$$\therefore \quad 5a = 75 - 0.56 - 68.58$$

$$= 5.86, \text{ giving } a = 1.172$$

Hence the fitted second degree parabola is

$$Y = 1.172 + 0.056X + 2.286X^2.$$

(c) Let the equation of the parabola of the second degree fitting the data be $Y = a + bX + cX^2$, where a, b and c are to be determined. The three normal equations are

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

The necessary computations are given in the table below:

X	Y	X^2	X^3	X^4	XY	X^2Y
0	1	0	0	0	0	0
1	5	1	1	1	5	5
2	10	4	8	16	20	40
3	22	9	27	81	66	198
4	38	16	64	256	152	608
10	76	30	100	354	243	851

Substituting these values in the three normal equations, we get

$$5a + 10b + 30c = 76$$
$$10a + 30b + 100c = 243$$
$$30a + 100b + 354c = 851$$

Solving them simultaneously, we obtain

$$a = 1.428,\ b = 0.244 \text{ and } c = 2.214$$

Hence the equation of the required second degree parabola is

$$Y = 1.428 + 0.244X + 2.214X^2$$

12.8. The arithmetic may be simplified by taking $u = \frac{X-2.5}{0.5}$.

The equation of the second degree parabola then becomes

$$Y = a + bu + cu^2.$$

Since $\sum u = 0 = \sum u^3$, the normal equations are therefore reduced to

$$\sum Y = na + c\sum u^2,$$
$$\sum uY = b\sum u^2,$$
$$\sum u^2Y = a\sum u^2 + c\sum u^4.$$

The necessary calculations are shown in the following table:

X	Y	$u = \frac{X-2.5}{0.5}$	u^2	u^4	uY	u^2Y
1.0	1.1	–3	9	81	–3.3	9.9
1.5	1.3	–2	4	16	–2.6	5.2
2.0	1.6	–1	1	1	–1.6	1.6
2.5	2.0	0	0	0	0	0
3.0	2.7	1	1	1	2.7	2.7
3.5	3.4	2	4	16	6.8	13.6
4.0	4.1	3	9	81	12.3	36.9
Σ	16.2	0	28	196	14.3	69.9

Substituting these values in the normal equations, we get

$$7a + 28c = 16.2$$
$$28b = 14.3$$
$$28a + 196c = 69.9$$

Solving, we get

$$a = 2.071,\ b = 0.511 \text{ and } c = 0.061, \text{ giving}$$
$$Y = 2.071 + 0.511u + 0.061u^2$$

Whence, by writing $\frac{X-2.5}{0.5}$ for u, we get

$$Y = 2.071 + 0.511\left(\frac{X-2.5}{0.5}\right) + 0.061\left(\frac{X-2.5}{0.5}\right)^2$$
$$= 2.071 + 0.511\,(2X-5) + 0.061\,(2X-5)^2$$
$$= 1.04 - 0.20X + 0.24X^2.$$

which is the required equation of the second degree parabola.

12.9. Taking $u = \frac{X-24}{4}$, we find that $\sum u = 0 = \sum u^3$.

Then the equation of the second degree parabola of Y on u is $Y = a+bu+cu^2$, and the normal equations become

$$\sum Y = na + c\sum u^2,\ \sum uY = b\sum u^2$$
$$\sum u^2Y = a\sum u^2 + c\sum u^4$$

The calculations involved are shown in the following table:

X	$u = \frac{X-24}{4}$	Y	u^2	u^4	uY	u^2Y
8	–4	2.4	16	256	–9.6	38.4
12	–3	4.8	9	81	–14.4	43.2
16	–2	8.3	4	16	–16.6	33.2
20	–1	9.5	1	1	–9.5	9.5
24	0	11.2	0	0	–50.1	0
28	1	24.3	1	1	24.3	24.3
32	2	22.2	4	16	44.4	88.8
36	3	21.2	9	81	63.6	190.8
40	4	25.4	16	256	101.6	406.4
Σ	0	129.3	60	708	233.9 183.8	834.6

Substituting these values, we get

$$9a + 60c = 129.3, \quad 60b = 183.8,$$
$$60a + 708c = 834.6$$

Solving these equations, we obtain

$a = 14.96$, $b = 3.063$ and $c = -0.089$, giving

$Y = 14.96 + 3.063\,u - 0.089\,u^2$.

Whence by writing $\frac{X-24}{4}$ for u, we get

$$Y = 14.96 + 3.063\left(\frac{X-24}{4}\right) - 0.089\left(\frac{X-24}{4}\right)^2$$

or $\quad Y = 14.96 + 3.063(0.25X - 6) - 0.089\,(0.25X-6)^2$

$\quad = 14.96 + 0.766X - 18.378 - 0.0056X^2 + 0.267X - 3.204$

$\quad = -6.622 + 1.033X - 0.0056X^2$,

which is the required equation of the second degree parabola.

12.10. Since $u = X - 3$ (given), so we find that $\sum u = 0 = \sum u^3$.

The normal equations are reduced to

$\sum v = na + c\sum u^2$, where $v = (Y-1650)/50$,

$\sum uv = b\sum u^2$, $\sum u^2 v = a\sum u^2 + c\sum u^4$

The necessary calculations are arranged in the following table:

X	$u = X-3$	Y	$v = \frac{Y-1650}{50}$	u^2	u^4	uv	u^2v
1	−2	1250	−8	4	16	16	−32
2	−1	1400	−5	1	1	5	−5
3	0	1650	0	0	0	0	0
4	1	1950	6	1	1	6	6
5	2	2300	13	4	16	26	52
$\sum$	---	---	6	10	34	53	21

Substituting these values in the normal equations, we get

$$5a + 10c = 6,$$
$$10b = 53,$$
$$10a + 34c = 21$$

Solving them, we find $a = -0.086$, $b = 5.30$ and $c = 0.643$.

Hence the equation of the required parabolic curve of v on u is

$$v = -0.086 + 5.30u + 0.643\,u^2.$$

Whence, by writing $X-3$ for u and $\frac{Y-1650}{50}$ for v, we get

$$\frac{Y-1650}{50} = -0.086 + 5.30(X-3) + 0.643\,(X-3)^2$$

Simplifying, we have

$$Y = 1140 + 72.14X + 32.14X^2$$

as the required parabolic curve of Y on X.

12.11. The arithmetic is simplified by taking $u = 2\left(\frac{X-37.5}{5}\right)$ as the number of pairs of values is even. As $\sum u = 0 = \sum u^3$, the normal equations become regarding the

(i) **Straight line:**

$\sum Y = na$ and $\sum uY = b\sum u^2$;

(ii) **Second degree parabola:**

$\sum Y = na + c\sum u^2$

$\sum uY = b\sum u^2$

$\sum u^2 Y = a\sum u^2 + c\sum u^4$

The necessary computations are give in the following table:

X	Y	u	u^2	u^4	uY	u^2Y
20	240	−7	49	2401	−1680	11760
25	315	−5	25	625	−1575	7875
30	403	−3	9	81	−1209	3627
35	450	−1	1	1	−450	450
40	488	1	1	1	+488	488
45	520	3	9	81	1560	4680
50	525	5	25	625	2625	13125
55	532	7	49	2401	3724	26068
$\sum$	3473	0	168	6216	3483	68073

Substituting these values in the normal equations, we get

(i) **Straight line:**

$$8a = 3473 \text{ and } 168b = 3483 \text{ giving}$$

$$a = 434.125 \text{ and } b = 20.73$$

Thus $Y = 434.125 + 20.73\,u$

The straight line of Y on X is obtained by writing $\frac{2(X-37.5)}{5}$ for u. Then

$$Y = 434.125 + 20.73\left[\frac{2(X-37.5)}{5}\right]$$

$$= 434.125 + 20.73\,(0.4X - 15)$$

$$= 123.175 + 8.292X$$

which is the equation of the required straight line.

(ii) **Second degree parabola:**

$$8a + 168c = 3473$$

$$168b = 3483$$

$$168a + 6216c = 68073.$$

Solving them simultaneously, we find that

$$a = 472.093,\ b = 20.73 \text{ and } c = -1.808, \text{ giving}$$

$$Y = 472.093 + 20.73\,u - 1.808\,u^2.$$

Whence, by writing $2\frac{(X-37.5)}{5}$, *i.e.* $(0.4X-15)$ for u, we get

$$Y = 472.093 + 20.73\,(0.4X-15) - 1.808\,(0.4X-15)^2$$

$$= -245.657 + 29.988X - 0.289\,X^2,$$

which is the equation of the required second degree parabola of Y on X.

The computations needed to find the sums of squares of residuals are given below:

X	Y	X^2	XY	Y^2	X^2Y
20	240	400	4800	57600	96000
25	315	625	7875	99225	196875
30	403	900	12090	162409	362700
35	450	1225	15750	202500	551250
40	488	1600	19520	238144	780800
45	520	2025	23400	270400	1053000
50	525	2500	26250	275625	1312500
55	532	3025	29260	283024	1609300
$\sum$	3473	12300	138945	1588927	5962425

The sum of squres of residuals in case of straight line is

$$S = \sum(Y-a-bX)^2 = \sum Y^2 - a\sum Y - b\sum XY$$

$$= 1588927 - (123.175)\,(3473) - (8.292)\,(138945)$$

$$= 1588927 - 427786.77 - 1152131.9 = 9008.4;$$

and in case of 2nd degree parabola, is

$$S = \sum(Y-a-bX-cX^2)^2 = \sum Y^2 - a\sum Y - b\sum XY - c\sum X^2Y$$

$$= 1588927-(-245.657)\,(3473) - (29.988)\,(138945) - (-0.289)\,(5962425)$$

$$= 1588927 + 853166.76 - 4166682.06 + 172929.5$$

$$= 340.6.$$

12.12. Taking $u = X-2$, we find that $\sum u = 0 = \sum u^3 = \sum u^5$. Then the normal equations in case of

(i) *Straight line*, reduce to $\sum Y = na$ and $\sum uY = b\sum u^2$;

(ii) *Quadratic parabola*, reduce to

$$\sum Y = na + c\sum u^2,\ \sum uY = b\sum u^2,\ \sum u^2Y = a\sum u^2 + c\sum u^4;$$

(iii) *Cubic parabola,* reduce to

$$\sum Y = na + c\sum u^2, \quad \sum uY = b\sum u^2 + d\sum u^4,$$
$$\sum u^2Y = a\sum u^2 + c\sum u^4, \quad \sum u^3Y = b\sum u^4 + d\sum u^6.$$

The necessary computations are given in the following table:

u	Y	u^2	u^4	u^6	uY	u^2Y	u^3Y
−2	1.0	4	16	64	−2.0	4.0	−8.0
−1	1.8	1	1	1	−1.8	1.8	−1.8
0	1.3	0	0	0	0	0	0
1	2.5	1	1	1	2.5	2.5	2.5
2	6.3	4	16	64	12.6	25.2	50.4
Σ	12.9	10	34	130	11.3	33.5	43.1

Substituting these values in the normal equations and solving them, we get

(i) **Straight line:**

$$5a = 12.9, \text{ and } 10b = 11.3$$

or $a = 2.58$ and $b = 1.13$.

Thus $Y = 2.58 + 1.13u$.

But the straight line of Y on X is obtained by writing $X-2$ for u. Therefore

$$Y = 2.58 + 1.13\,(X-2)$$
$$= 0.32 + 1.13X,$$

which is the equation of the desired straight line.

(ii) **Parabola of second degree:**

$$5a + 10c = 12.9$$
$$10b = 11.3$$
$$10a + 34c = 33.5$$

Solving them simultaneously, we find that

$a = 1.48$, $b = 1.13$ and $c = 0.55$.

Then $Y = 1.48 + 1.13u + 0.55u^2$.

Writing $X-2$ for u, we get the quadratic parabola as

$$Y = 1.48 + 1.13\,(X-2) + 0.55\,(X-2)^2$$
$$= 1.42 - 1.07X + 0.55X^2$$

(iii) **Cubic parabola:**

$$5a + 10c = 12.9, \quad 10b + 34d = 11.3,$$
$$10a + 34c = 33.5, \quad 34b + 130d = 43.1.$$

Solving them simultaneously, we find that

$a = 1.48$, $b = 0.025$, $c = 0.55$ and $d = 0.325$

$\therefore$ $Y = 1.48 + 0.025u + 0.55u^2 + 0.325u^3$

Replacing u by $X-2$, we get the desired parabola of third degree

$$Y = 1.48 + 0.025\,(X-2) + 0.55\,(X-2)^2 + 0.325(X-2)^3$$
$$= 1.03 + 1.725X - 1.40X^2 + 0.325X^3.$$

Next, we calculate the expected values $(\hat{Y})$ and the sums of squares of residuals as below:

X	Y	Straight line		Quadratic parabola		Cubic parabola	
		$\hat{Y} = 0.32 + 1.13X$	$(Y-\hat{Y})^2$	$\hat{Y} = 1.42 - 1.07X + 0.55X^2$	$(Y-\hat{Y})^2$	$\hat{Y} = 1.03 + 1.725X - 1.40X^2 + 0.325X^3$	$(Y-\hat{Y})^2$
0	1.0	0.32	0.4624	1.42	0.1764	1.03	0.0009
1	1.8	1.45	0.1225	0.90	0.8100	1.68	0.0144
2	1.3	2.58	1.6384	1.48	0.0324	1.48	0.0324
3	2.5	3.71	1.4641	3.16	0.4356	2.38	0.0144
4	6.3	4.84	2.1316	5.94	0.1296	6.33	0.0009
Σ	12.9	12.9	5.8190	12.9	1.5840	12.9	0.0630

Alternatively. The computations needed to find the sums of squares of residuals are given as follows:

X	Y	X^2	XY	Y^2	X^2Y	X^3Y
0	1.0	0	0	1.00	0	0
1	1.8	1	1.8	3.24	1.8	1.8
2	1.3	4	2.6	1.69	5.2	10.4
3	2.5	9	7.5	6.25	22.5	67.5
4	6.3	16	25.2	39.69	100.8	403.2
$\sum$	12.9	30	37.1	51.87	130.3	482.9

(i) *Straight line*:

$S = \sum(Y-a-bX)^2 = \sum Y^2 - a\sum Y - b\sum XY$

$= 51.87 - (0.32)(12.9) - (1.13)(37.1)$

$= 51.87 - 4.128 - 41.932 = 5.82$

(ii) *Quadratic parabola*:

$S = \sum(Y-a-bX-cX^2)^2 = \sum Y^2 - a\sum Y - b\sum XY - c\sum X^2Y$

$= 51.87 - (1.42)(12.9) - (-1.07)(37.1) - (0.55)(130.3)$

$= 51.87 - 18.318 + 39.697 - 71.665 = 1.584.$

(iii) *Cubic parabola*:

$S = \sum(Y-a-bX-cX^2-dX^3)^2$

$= \sum Y^2 - a\sum Y - b\sum XY - c\sum X^2Y - d\sum X^3Y$

$= 51.87 - (1.03)(12.9) - (1.725)(37.1) - (-1.4) \times (130.3) - (0.325)(482.9)$

$= 51.87 - 13.287 - 63.9975 + 182.42 - 156.9425$

$= 234.29 - 234.227 = 0.063,$

12.13. (b) To find a suitable curve,, we construct a difference table as below:

X	Y	ΔY	$\Delta^2 Y$
0	10		
		7	
1	17		4
		11	
2	28		4
		15	
3	43		4
		19	
4	62		

As the second order differences are constant, therefore the suitable curve is the second degree parabola, *i.e.* $Y = a + bX + cX^2$. **Then the three normal equations are**

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4.$$

The necessary computations are given below:

X	Y	X^2	X^3	X^4	XY	X^2Y
0	10	0	0	0	0	0
1	17	1	1	1	17	17
2	28	4	8	16	56	112
3	43	9	27	81	129	387
4	62	16	64	256	248	992
10	160	30	100	354	450	1508

Substituting these values in the normal equations, we get

$$5a + 10b + 30c = 160$$

$$10a + 30b + 100c = 450$$

$$30a + 100b + 354c = 1508$$

Solving them simultaneously, we find that

$$a = 10, \quad b = 5 \text{ and } c = 2$$

Hence the desired second degree parabolic curve is

$$Y = 10 + 5X + 2X^2.$$

12.14. (b) The given curve Y = ab^X **may be written as**

$\log Y = \log a + X \log b$

or $\quad Y' = A + BX'$ **(Form of** *a* **st. line)**

where $Y' = \log Y$, $A = \log a$ **and** $B = \log b$.

The normal equations are:

$$\sum Y' = nA + B\sum X$$

$$\sum XY' = A\sum X + B\sum X^2.$$

The calculations involved are shown in the following table.

	X	Y	X^2	$Y' = \log Y$	XY
	0	73	0	1.8633	0
	1	91	1	1.9590	1.9590
	2	112	4	2.0492	4.0984
	3	131	9	2.1173	6.3519
	4	162	16	2.2095	8.8380
Σ	10	–	30	10.1983	21.2473

Substituting these values in the normal equations, we get

$$5A + 10B = 10.1983$$

$$10A + 30B = 21.2473$$

Solving them, we find that $A = 1.8695$ and $B = 0.08507$

Thus a = antilog of A

= antilog (1.8695) = 74.04, and

b = antilog of B

= antilog (0.08507) = 1.216

Hence the eqution of the required curve is

$$Y = 74.04\,(1.22)^X$$

Next, we estimate the values of Y when $X=5$ and 6. The equation is $Y=74.04(1.22)^X$. Putting $X=5$ and 6, we get

$$Y_5 = 74.04\,(1.22)^5 = 200.11; \text{ and}$$

$$Y_6 = 74.04\,(1.22)^6 = 244.13.$$

12.15. The equaton of the curve $Y = ab^X$ may be written as

$$\log Y = \log a + X \log b$$

or $\quad Y' = A + BX,$ (form of a st. line)

where $Y' = \log Y$, $A = \log a$ and $B = \log b$

The two normal equations are

$$\sum Y' = nA + B\sum X$$

$$\sum XY' = A\sum X + B\sum X^2.$$

The necessary calculations are given in the table below:

	X	Y	X^2	$Y' = \log Y$	XY'
	0	32	0	1.5051	0
	1	47	1	1.6721	1.6721
	2	65	4	1.8129	3.6258
	3	92	9	1.9638	5.8914
	4	132	16	2.1206	8.4824
	5	190	25	2.2788	11.3940
	6	275	36	2.4393	14.6358
Σ	21	--	91	13.7926	45.7015

Substituting these values in the normal equations, we get

$$7A + 21B = 13.7926, \quad 21A + 91B = 45.7015$$

Solving them, we find that $A = 1.5072$ and $B = 0.1544$

Thus $\quad a$ = Antilog of A

= Antilog (1.5072) = 32.15, and

b = Antilog of B = Antilog (0.1544) = 1.427

Hence the least-squares curve is of the form

$$Y = 32.15\,(1.427)^X$$

Estimated value of Y when $X = 7$, is

$$Y = 32.15\,(1.427)^7 = 387.40$$

12.16. We may write the exponential equation in the logarithmic form as

$$\log Y = \log a + X \log b$$

i.e. $\quad Y' = A + BX,$ (Form of a straight line)

where $Y' = \log Y$, $A = \log a$ and $B = \log b$.

The two normal equations are

$$\sum Y' = nA + B\sum X, \quad \sum XY' = A\sum X + B\sum X^2$$

The necessary calculatons are arranged in the following table:

	Day (X)	Height (Y)	X^2	$Y' = \log Y$	XY'
	0	0.75	0	−0.1249	0
	1	1.20	1	0.0792	0.0792
	2	1.75	4	0.2430	0.4860
	3	2.50	9	0.3979	1.1937
	4	3.45	16	0.5378	2.1512
	5	4.70	25	0.6721	3.3605
	6	6.20	36	0.7924	4.7544
	7	8.25	49	0.9165	6.4155
	8	11.50	64	1.0607	8.4856
Σ	36	--	204	4.5747	26.9261

Substituting these values, we get

$$9A + 36B = 4.5747$$

$$36A + 204B = 26.9261$$

Solving them simultaneously, we get

$$A = -0.0669 \text{ and } B = 0.1438$$

Whence, we find that

$$a = \text{Antilog of } A = \text{Antilog } (-0.0669)$$

$$= \text{Antilog } (\bar{1}.9331) = 0.8572, \text{ and}$$

$$b = \text{Antilog of } B$$

$$= \text{Antilog } (0.1438) = 1.393$$

Hence the equation of the required exponential curve is

$$Y = 0.86\,(1.39)^X.$$

12.17. The equation of the curve $Y = ab^X$ may be written as

$$\log Y = \log a + X \log b$$

or $\quad Y' = A + BX,$

where $Y' = \log Y$, $A = \log a$ and $B = \log b$.

The two normal equations are

$$\sum Y' = nA + B\sum X$$

$$\sum XY' = A\sum X + B\sum X^2$$

The necessary calculatons are given in the table below:

	X	Y	X^2	$Y' = \log Y$	XY'
	1	10.0	1	1.0000	1.0000
	2	12.2	4	1.0864	2.1728
	3	14.5	9	1.1614	3.4842
	4	17.3	16	1.2380	4.9520
	5	21.0	25	1.3222	6.6110
	6	25.0	36	1.3979	8.3874
	7	29.0	49	1.4624	10.2368
Σ	28	---	140	8.6683	36.8442

Substituting these values in the normal equations, we get

$$7A + 28B = 8.6683$$

$$28A + 140B = 36.8442$$

Solving them, we find that $A = 0.9283$ and $B = 0.0775$

Thus $\quad a = \text{Antilog of } A$

$$= \text{Antilog } (0.9283) = 8.478, \text{ and}$$

$$b = \text{Antilog of } B = \text{Antilog } (0.0775) = 1.195.$$

Hence the equation of the desired curve is

$$Y = 8.478\,(1.195)^X$$

12.18. The equation of the curve $Y = ab^X$ may be written as

$$\log Y = \log a + X \log b$$

or $\quad Y' = A + BX$

where $Y' = \log Y$, $A = \log a$ and $B = \log b$.

The two normal equations are

$$\sum Y' = nA + B\sum X$$
$$\sum XY' = A\sum X + B\sum X^2$$

The following table contains the necessary computations:

X (years)	Y	X^2	$Y' = \log Y$	XY'
1	304	1	2.4829	2.4829
2	341	4	2.5328	5.0656
3	393	9	2.5944	7.7832
4	457	16	2.6599	10.6396
5	548	25	2.7388	13.6940
6	670	36	2.8261	16.9566
7	882	49	2.9455	20.6185
28	--	140	18.7804	77.2404

Substituting these values in the normal equations, we get

$$7A + 28B = 18.7804$$
$$28A + 140B = 77.2404$$

Solving them, we find that $A = 2.380$ and $B = 0.076$

Thus a = Antilog of A = Antilog (2.380) = 239

b = Antilog of B = Antilog of (0.076) = 1.19

Hence the least squares curve is of the form

$$Y = 239\,(1.19)^X$$

The expected enrolment 5 years from now ($X = 12$) is

$$Y_{12} = 239\,(1.19)^{12} = 1954$$

12.19. (a) Considering $\sum X^3 = 582$ redundant, we may fit to the given values the curve $Y = ab^X$, which in terms of logs becomes

$$\log Y = \log a + X \log b$$

or $\log Y = A + BX \quad (A = \log a,\ B = \log b)$

The two normal equations then are

$$\sum \log Y = nA + B\sum X,$$
$$\sum X \log Y = A\sum X + B\sum X^2$$

Substituting the values, we get

$$8A + 16B = 23$$
$$16A + 204B = 104$$

Solving, we obtain $A = 2.20$ and $B = 0.3372$

Whence $a = 158.5$ and $b = 2.17$

Hence the curve is $Y = 158.5\,(2.17)^X$

(b) The equation of the curve $Y = a + b\sqrt{X}$ may be written as

$$Y = a + bX', \text{ where } X' = \sqrt{X}$$

The two normal equations are

$$\sum Y = na + b\sum X'$$
$$\sum X'Y = a\sum X' + b\sum X'^2$$

The necessary calculations are given in the table below:

X	Y	$X' = \sqrt{X}$	X'^2	$X'Y$
1.20	6.33	1.0954	1.20	6.9339
2.50	8.03	1.5811	2.50	12.6962
3.40	8.95	1.8439	3.40	16.5029
4.70	10.09	2.1679	4.70	21.8741
5.30	10.56	2.3022	5.30	24.3112
$\sum$	43.96	8.9905	17.10	82.3183

Substituting these values in the normal equations, we get

$$5a + 8.9905\,b = 43.96,$$
$$8.9905\,a + 17.10\,b = 82.3183.$$

Solving them, we find that $a = 2.50$ and $b = 3.50$

Hence the least-squares curve is of the form

$$Y = 2.50 + 3.50\sqrt{X}$$

12.20. The equatoin of the cure $Y = aX^b$ may be written as

$$\log Y = \log a + b \log X$$

or $$Y' = A + bX',$$

where $Y' = \log Y$, $A = \log a$ and $X' = \log X$.

The two normal equation are

$$\sum Y' = nA + b\sum X'$$

$$\sum X'Y' = A\sum X' + b\sum X'^2$$

The necessary calculations are shown in the following table:

X	Y	$X' = \log X$	X'^2	$X' = \log X$	$X'Y'$
1	1200	0	0	3.0792	0
2	900	0.3010	0.090601	2.9542	0.889214
3	600	0.4771	0.227624	2.7782	1.325479
4	200	0.6021	0.362524	2.3010	1.385432
5	110	0.6990	0.488601	2.0414	1.426939
6	50	0.7782	0.605595	1.6990	1.322162
$\sum$	--	2.8574	1.774945	14.8530	6.349226

Substituting these values in the normal equations, we get

$$6A + 2.8574\,b = 14.8530$$

$$2.8574\,A + 1.774945\,b = 6.349226$$

Solving them, we find that $A = 3.3083$ and $b = -1.7488$

Thus a = Antilog of A

= Antilog of (3.3083) = 2033.0

Hence the equation of the desired curve is

$$Y = 2033\,(X)^{-1.7488}$$

12.21. The equation of the curve $Y = aX^b$ may be written as

$$\log Y = \log a + b \log X,$$

i.e. $Y' = A + bX'$, **(a straight line)**

where $Y' = \log Y$, $A = \log a$ and $X' = \log X$

Then the two normal equations are

$$\sum Y' = nA + b\sum X'$$

$$\sum X'Y' = A\sum X' + b\sum X'^2$$

The necessary calculations are given in the table below:

X	Y	$X' = \log X$	X'^2	$Y' = \log Y$	$X'Y'$
50	108	1.6990	2.886601	2.0334	3.454747
100	53	2.0000	4.000000	1.7243	3.448600
250	24	2.3979	5.749924	1.3802	3.309582
500	9	2.6990	7.284601	0.9542	2.575386
1000	5	3.0000	9.000000	0.6990	2.097000
$\sum$	--	11.7959	28.921126	6.7911	14.885315

Substituting these values in the normal equations, we get

$$5A + 11.7959\,b = 6.7911,$$

$$11.7959\,A + 28.921126\,b = 14.885315$$

Solving them, we find that $A = 3.81177$ and $b = -1.04$

Thus a = Antilog of A

= Antilog (3.81177) = 6483.1

Hence the equation of the required curve is

$$Y = 6483.1\,(X)^{-1.04}$$

To estimate the value of Y when $X = 400$, we use the equation $Y' = A + bX'$

Therefore $Y' = 3.81177 + (-1.04)(2.6021)$

$= 1.1056$

Hence taking Antilog, we get $Y = 12.76$

12.22. We may write the given relation in logs as

$$\log Y = \log a + n \log X$$

or $Y' = A + nX'$; **(a straight line)** where $Y' = \log Y$, $A = \log a$ and $X' = \log X$

The two normal equations are

$$\sum Y' = kA + n\sum X' \quad (k = \text{no. of values})$$

$$\sum X'Y' = A\sum X' + n\sum X'^2$$

The computations involved are shown in the following table:

X	Y	$X' = \log X$	X'^2	$Y' = \log Y$	$X'Y'$
0.5	3.4	−0.3010	0.0906	0.5315	−0.15998
1.5	7.0	0.1761	0.0310	0.8451	+0.14882
2.5	12.8	0.3979	0.1582	1.1072	0.44055
5.0	29.8	0.6990	0.4886	1.4742	1.03047
10.0	68.2	1.0000	1.0000	1.8338	1.83380
$\sum$	--	1.9720	1.7684	5.7918	3.29366

Substituting these values in the normal equations, we get

$$5A + 1.9720\,n = 5.7918,$$

$$1.9720\,A + 1.7684\,n = 3.29366.$$

Solving them simultaneously, we find that

$$A = 0.7561 \text{ and } n = 1.02$$

Now $\quad a = \text{Antilog of } A$

$$= \text{Antilog}\,(0.7561) = 5.703$$

Hence $a = 5.703$ and $n = 1.02$ are the required values.

12.23. The equation of the curve $v = ae^{bt}$ may be written as

$$\log v = \log a + (b \log e)\,t$$

or $\quad y = A + Bt,$

where $y = \log v$, $A = \log a$ and $B = b \log e$

Then the two normal equations are

$$\sum y = nA + B\sum t$$

$$\sum yt = A\sum t + B\sum t^2$$

The necessary calculations are given in the table below:

t	v	$y = \log v$	yt	t^2
0.5	9.1	0.9590	0.47950	0.25
0.8	8.5	0.9294	0.74352	0.64
1.4	7.5	0.8751	1.22514	1.96
2.0	6.7	0.8261	1.65220	4.00
2.5	6.1	0.7853	1.96325	6.25
7.2	--	4.3749	6.06361	13.10

Substituting these values in the normal equations, we get

$$5A + 7.2B = 4.3749$$

$$7.2A + 13.10B = 6.06361.$$

Solving them, we find that $A = 0.9995$ and $B = -0.0865$

Thus $\quad a = \text{Antilog of } A$

$$= \text{Antilog}\,(0.9995) = 9.998$$

$$b = B/\log e \text{ where } \log e = 0.4343$$

$$= -0.0865/0.4343 = -0.1992 = -0.20$$

Hence the equation of the curve to be fitted tc the data is

$$v = 9.998\ e^{-0.2t}$$

$$= 10\,e^{-0.2t}$$

12.24.(a) The equation of the curve $Y = ce^{bX}$ may be written as

$$\log Y = \log a + (b \log e)\,X$$

or $\quad Y' = A + BX,$

where $Y' = \log Y$, $A = \log a$ and $B = b \log e$

Then the two normal equations are

$$\sum Y' = nA + B\sum X$$

$$\sum XY' = A\sum X + B\sum X^2$$

The computations needed to fit the curve are given in the table below:

X	Y	$Y' = \log Y$	XY'	X^2
1	27	1.4314	1.4314	1
2	73	1.8633	3.7266	4
3	200	2.3010	6.9030	9
4	545	2.7364	10.9456	16
5	1484	3.1715	15.8575	25
15	--	11.5036	38.8641	55

Substituting these values in the normal equations, we get

$$5A + 15B = 11.5036$$

$$15A + 55B = 38.8641$$

Solving them, we find that $A = 0.99473$ and $B = 0.43533$

Therefore $Y = 0.99473 + 0.43533X$

Now $\quad a =$ Antilog of A

$= \text{Antilog}\ (0.99473) = 9.88$, and

$0.4343b = 0.43533 \quad (\because \log_{10}e = 0.4343)$

or $\quad b = \dfrac{0.43533}{0.4343} = 1.002$

Hence the equation of the curve fitted to the data is

$$Y = 9.88\ e^{1.002X}$$

(b) The values of Y from the approximating line for various values of X are obtained below:

X	$Y' = 0.99473 + 0.43533X$	$\hat{Y} =$ Antilog Y'	$Y - \hat{Y}$
1	1.43006	26.92	0.08
2	1.86539	73.34	−0.34
3	2.30072	199.81	0.19
4	2.73605	544.56	0.44
5	3.17138	1484.30	−0.30
Σ	--	--	0.07

The deviations of the estimated values of Y from the corresponding observed values add to 0.07. This slight difference is due to rounding off.

12.25. The Pareto curve $n = AX^{-\alpha}$ may be written as

$$\log n = \log A - \alpha \log X$$

or $\quad Y = a + bX'$, (a straight line)

where $Y = \log n$, $a = \log A$, $b = -\alpha$ and $X' = \log X$

The two normal equations are

$$\sum Y = Na + b\sum X'$$

$$\sum X'Y = a\sum X' + b\sum X'^2$$

The computations involved are shown in the following table:

Income (X)	$X' = \log X$	X'^2	No. (n)	$Y = \log n$	$X'Y$
150	2.1761	4:7354	14,000,000	7.1461	15.55063
500	2.6990	7.2846	825,000	5.9165	15.96763
1,000	3.0000	9.0000	173,000	5.2380	15.71400
2,000	3.3010	10.8966	35,500	4.5502	15.02021
Σ	11.1761	31.9166	---	22.8508	62.25247

Substituting these values, we get

$$4a + 11.1761b = 22.8508$$

$$11.1761a + 31.9166b = 62.25247$$

Multiplying the first equation by 11.1761 and the second by 4, we get

$$44.7044a + 124.9052b = 255.38283$$

$$44.7044a + 127.6664b = 249.00988$$

Subtraction gives

$$-2.7612b = 6.37295$$

$\therefore \quad b = -2.31$

Whence, we find that $\alpha = -b = 2.31$.

Putting $b = -2.31$ in the first normal equation and simplifying, we get $a = 12.1669$.

Thus A = Antilog of a

= Antilog (12.1669) = 1,469,000,000,000.

Hence the result.

12.26. Taking logs of both sides of the equation $pv^{\gamma}=c$, we obtain

$$\log p + \gamma \log v = \log c$$

or $\log v = \frac{1}{\gamma} \log c - \frac{1}{\gamma} \log p$, as p is to be taken as independent variable.

Calling $\log v = y$, and $\log p = x$, we may write it as

$$y = a + bx, \text{ where } a = \frac{1}{\gamma}\log c \text{ and } b = -\frac{1}{\gamma}$$

Then the two normal equations are

$$\sum y = na + b\sum x$$

$$\sum xy = a\sum x + b\sum x^2, \text{ giving}$$

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}; \text{ and}$$

$$a = \frac{(\sum x^2)(\sum y) - (\sum x)(\sum xy)}{n\sum x^2 - (\sum x)^2}.$$

The computations involved are shown in the following table:

p	v	$x(=\log p)$	$y(=\log v)$	xy	x^2
0.5	1.62	−0.3010	0.2095	−0.06306	0.09060
1.0	1.00	0.0000	0.0000	0	0
1.5	0.75	0.1761	−0.1249	−0.02079	0.03101
2.0	0.62	0.3010	−0.2076	−0.06249	0.09060
2.5	0.52	0.3979	−0.2840	−0.11400	0.15832
3.0	0.46	0.4771	−0.3372	−0.16088	0.22762
Σ	--	1.0511	−0.7442	−0.42122	0.59815

Substitution gives

$$b = \frac{(6)(-0.42122) - (-0.7442)(1.0511)}{(6)(0.59815) - (1.0511)^2}$$

$$= \frac{-2.52732 + 0.78223}{3.58890 - 1.1048112} = \frac{-1.74509}{2.4840888} = -0.702509$$

Thus $-\frac{1}{\gamma} = -0.702509$

or $\gamma = \frac{1}{0.702509} = 1.42.$

12.27. (b) The equation of the curve $\frac{1}{Y} = a + bX$ may be written as

$$Y' = a + bX, \text{ where } Y' = \frac{1}{Y}.$$

The two normal equations are

$$\sum Y' = na + b\sum X$$

$$\sum XY' = a\sum X + b\sum X^2$$

The necessary calculations are given in the table below:

	X	Y	X^2	$Y'=1/Y$	XY'
	0	10	0	0.100	0
	1	8	1	0.125	0.125
	4	5	16	0.200	0.800
	6	4	36	0.250	1.500
	12	2.5	144	0.400	4.800
	16	2	256	0.500	8.000
Σ	39	---	453	1.575	15.225

Substituting these values in the normal equations, we get

$$6a + 39b = 1.575$$

$$39a + 453b = 15.225$$

Solving them, we find that $a = 0.10$ and $b = 0.025$

Hence the required equation of the reciprocal curve is

$$\frac{1}{Y} = 0.10 + 0.025\,X$$

12.28. (a) The given linear equation is

$$Y = a + bX_1 + cX_2,$$

where a, b and c are to be determined.

The method of least-squres calls for the selection of those values a, b and c which make the sum of squares of deviations, S, a minimum. In other words, we have to minimize $S = \sum(Y-a-bX_1-cX_2)^2$.

For a minimum value, $\frac{\partial S}{\partial a}, \frac{\partial S}{\partial b}$ and $\frac{\partial S}{\partial c}$ must be zero.

Thus
$$\frac{\partial S}{\partial a} = 2\sum (Y-a-bX_1-cX_2)(-1) = 0,$$

$$\frac{\partial S}{\partial b} = 2\sum (Y-a-bX_1-cX_2)(-X_1) = 0, \text{ and}$$

$$\frac{\partial S}{\partial c} = 2\sum (Y-a-bX_1-cX_2)(-X_2) = 0,$$

which on simplification become

$$\sum Y = na + b\sum X_1 + c\sum X_2$$

$$\sum X_1Y = a\sum X_1 + b\sum X_1^2 + c\sum X_1X_2$$

$$\sum X_2Y = a\sum X_2 + b\sum X_1X_2 + c\sum X_2^2$$

These are the required normal equations.

(b) The normal equations of $Y = a + bX_1 + cX_2$ are

$$\sum Y = na + b\sum X_1 + c\sum X_2$$

$$\sum X_1Y = a\sum X_1 + b\sum X_1^2 + c\sum X_1X_2$$

$$\sum X_2Y = a\sum X_2 + b\sum X_1X_2 + c\sum X_2^2$$

The necessary calculations are given in the table below:

	Y	X_1	X_2	X_1Y	X_2Y	X_1^2	X_2^2	X_1X_2
	2	8	0	16	0	64	0	0
	5	8	1	40	5	64	1	8
	7	6	1	42	7	36	1	6
	8	5	3	40	24	25	9	15
	5	3	4	15	20	9	16	12
$\sum$	27	30	9	153	56	198	27	41

Substituting these values in the normal equations, we get

$$5a + 30b + 9c = 27$$

$$30a + 198b + 41c = 153$$

$$9a + 41b + 27c = 56$$

Solving them, we find that $a = 4.49$, $b = -0.04$ and $c = 0.64$.

Hence the required equation is $Y = 4.49 - 0.04X_1 + 0.64X_2$.

12.32. (a) According to the principle of least squares, the sum to be minimized is given by

$$S = (X+7Y-17)^2 + (2X-Y-0)^2 + (3X-2Y+1)^2$$

The two normal equations are obtained as

$$\frac{\partial S}{\partial X} = 2(X+7Y-17)(1)+2(2X-Y-0)(2)+2(3X-2Y+1)(3) = 0,$$

$$\frac{\partial S}{\partial Y} = 2(X+7Y-17)(7)+2(2X-Y-0)(-1)+2(3X-2Y+1)(-2) = 0.$$

Simplifying, we get

$$14X - Y = 14, \text{ and } -X + 54Y = 121$$

Multiplying the first equation by 1 and second by 14, we have

$$14X - Y = 14, \text{ and } -14X + 756Y = 1694$$

Adding, we get

$$755Y = 1708 \text{ or } Y = 2.262$$

Putting $Y = 2.262$ in the first normal equation, we have

$$14X - 2.262 = 14 \text{ or } 14X = 14 + 2.262$$

$$\therefore \quad X = \frac{16.262}{14} = 1.162$$

Hence the required solution is $X = 1.162$ and $Y = 2.262$

(b) The normal equation for X is obtianed by multiplying each equatioin by the co-efficient of X in it and adding them together as below:

$$4X + 2Y = 9.6$$
$$9X - 6Y = -6.3$$
$$+X - 3Y = -6.3$$
$$\underline{9X + 6Y = 24.0}$$
$$23X - Y = 21.0 \quad \ldots (A)$$

Similarly, the normal equation for Y is obtained as below:

$$2X + Y = 4.8$$
$$-6X + 4Y = 4.2$$
$$-3X + 9Y = 18.9$$
$$\underline{6X + 4Y = 16.0}$$
$$-X + 18Y = 43.9 \quad \ldots (B)$$

Now $18(A)$: $\quad 414X - 18Y = 378 \quad \ldots (C)$

$\therefore (B) + (C)$: $\quad 413X = 421.9$ or $X = 1.02$

Putting $X = 1.02$ in (B), we get

$$-1.02 + 18Y = 43.9 \text{ or } Y = \frac{43.9 + 1.02}{18} = 2.50$$

Hence the required solution is $X = 1.02$ and $Y = 2.50$.

Substituting $X = 1.02$ and $Y = 2.50$ in the given set of equations, we get the residuals as

$$e_1 = -0.26,\ e_2 = 0.16,\ e_3 = 0.18,\ e_4 = 0.06$$

$\therefore$ Sum of squres of residuals is

$$\sum e_i^2 = (-0.26)^2 + (0.16)^2 + (0.18)^2 + (0.06)^2 = 0.1292$$

12.33. (a) Normal equation for X is

$$X + Y = 3.01$$
$$4X - 2Y = 0.06$$
$$X + 3Y = 7.02$$
$$\underline{9X + 3Y = 14.91}$$
$$15X + 5Y = 25.00 \quad \ldots (A)$$

Normal equation for Y is

$$X + Y = 3.01$$
$$-2X + Y = -0.03$$
$$3X + 9Y = 21.06$$
$$\underline{3X + Y = 4.97}$$
$$5X + 12Y = 29.01 \quad \ldots (B)$$

Now $3(B)$: $\quad 15X + 36Y = 87.03 \quad \ldots (C)$

$(C)-(A)$: $\quad 31Y = 62.03$

$\therefore \quad Y = 2.001$

Putting this value of Y in (A) and simplifying, we get

$$X = 0.9997$$

(b) Normal equation for X is

$$4X + 2Y = 8$$
$$X + 2Y = 5.02$$
$$9X - 3Y = 30.06$$
$$\underline{9X + 6Y = 2.91}$$
$$23X + 7Y = 45.99 \quad \ldots (A)$$

Normal equation for Y is

$$2X + Y = 4$$
$$2X + 4Y = 10.04$$
$$-3X + Y = -10.02$$
$$\underline{6X + 4Y = 1.94}$$
$$7X + 10Y = 5.96 \quad \ldots (B)$$

Now $10(A)$: $\quad 230X + 70Y = 459.9 \quad \ldots (C)$

$7(B)$: $\quad \underline{49X + 70Y = 41.72} \quad \ldots (D)$

$(C)-(D)$: $\quad 181X = 418.18$

$$\therefore \quad X = \frac{418.18}{181} = 2.31$$

Substituting $X = 2.31$ in (B), we get

$$7(2.31) + 10Y = 5.96$$

or $\quad 10Y = 5.96 - 16.17$

or $\quad Y = \dfrac{-10.21}{10} = -1.02$

Hence the best possible values are

$$X = 2.31 \text{ and } Y = -1.02$$

12.34. Normal equation for X is obtained below:

$$\begin{aligned} X + 2Y + Z &= 1 \\ X - Y - 2Z &= -3 \\ 4X + 2Y + 2Z &= 8 \\ \underline{16X + 8Y - 20Z} &\underline{= -28} \\ 22X + 11Y - 19Z &= -22 \quad \ldots (A) \end{aligned}$$

Normal equation for Y is obtained below:

$$\begin{aligned} 2X + 4Y + 2Z &= 2 \\ -X + Y + 2Z &= 3 \\ 2X + Y + Z &= 4 \\ \underline{8X + 4Y - 10Z} &\underline{= -14} \\ 11X + 10Y - 5Z &= -5 \quad \ldots (B) \end{aligned}$$

Normal equation for Z is obtained as below:

$$\begin{aligned} X + 2Y + Z &= 1 \\ -2X + 2Y + 4Z &= 6 \\ 2X + Y + Z &= 4 \\ \underline{-20X - 10Y + 25Z} &\underline{= 35} \\ -19X - 5Y + 31Z &= 46 \quad \ldots (C) \end{aligned}$$

Now $2(B)$: $\quad 22X + 20Y - 10Z = -10 \quad \ldots (D)$

$(D)-(A)$: $\quad 9Y + 9Z = 12$

or $\quad 3Y + 3Z = 4 \quad \ldots (E)$

Again $19(B)$: $\quad 209X + 190Y - 95Z = -95 \quad \ldots (F)$

$11\,(C)$: $\quad \underline{-209X - 55Y + 341Z = 506} \quad \ldots (G)$

$(F)+(G)$: $\quad 135Y + 246Z = 411 \quad \ldots (H)$

Solving equations (E) and (H) simultaneously, we get

$$Y = -0.75 \text{ and } Z = 2.08$$

Putting these two values in (A), we get

$$22X + 11(-0.75) - 19(2.08) = -22$$

or $\quad 22X = -22 + 8.25 + 39.52$

or $\quad X = \dfrac{25.77}{22} = 1.17$

Hence the required solution is

$$X = 1.17,\ Y = -0.75, \text{ and } Z = 2.08$$

12.35. Normal equation for X is:

$$\begin{aligned} X - Y + 2Z &= 3 \\ 9X + 6Y - 15Z &= 15 \\ 16X + 4Y + 16Z &= 84 \\ \underline{X - 3Y - 3Z} &\underline{= -14} \\ 27X + 6Y &= 88 \quad \ldots (A) \end{aligned}$$

Normal equation for Y is:

$$\begin{aligned} -X + Y - 2Z &= -3 \\ 6X + 4Y - 10Z &= 10 \\ 4X + Y + 4Z &= 21 \\ \underline{-3X + 9Y + 9Z} &\underline{= 22} \\ 6X + 15Y + Z &= 70 \quad \ldots (B) \end{aligned}$$

Normal equation for Z is:

$$
\begin{aligned}
2X - 2Y + 4Z &= 6 \\
-15X - 10Y + 25Z &= -25 \\
16X + 4Y + 16Z &= 84 \\
\underline{-3X + 9Y + 9Z} &\underline{= 42} \\
Y + 54Z &= 107 \qquad \dots (C)
\end{aligned}
$$

Solving these equations by determinants, we get

$$\frac{X}{\begin{vmatrix} 88 & 6 & 0 \\ 70 & 15 & 1 \\ 107 & 1 & 54 \end{vmatrix}} = \frac{Y}{\begin{vmatrix} 27 & 88 & 0 \\ 6 & 70 & 1 \\ 0 & 107 & 54 \end{vmatrix}} = \frac{Z}{\begin{vmatrix} 27 & 6 & 88 \\ 6 & 15 & 70 \\ 0 & 1 & 107 \end{vmatrix}} = \frac{1}{\begin{vmatrix} 27 & 6 & 0 \\ 6 & 15 & 1 \\ 0 & 1 & 54 \end{vmatrix}}$$

$$\frac{X}{49154} = \frac{Y}{70659} = \frac{Z}{38121} = \frac{1}{19899}$$

$$\therefore \quad X = \frac{49154}{19899} = 2.47;$$

$$Y = \frac{70659}{19899} = 3.55; \text{ and}$$

$$Z = \frac{38121}{19899} = 1.92.$$

Hence the most plausible values are

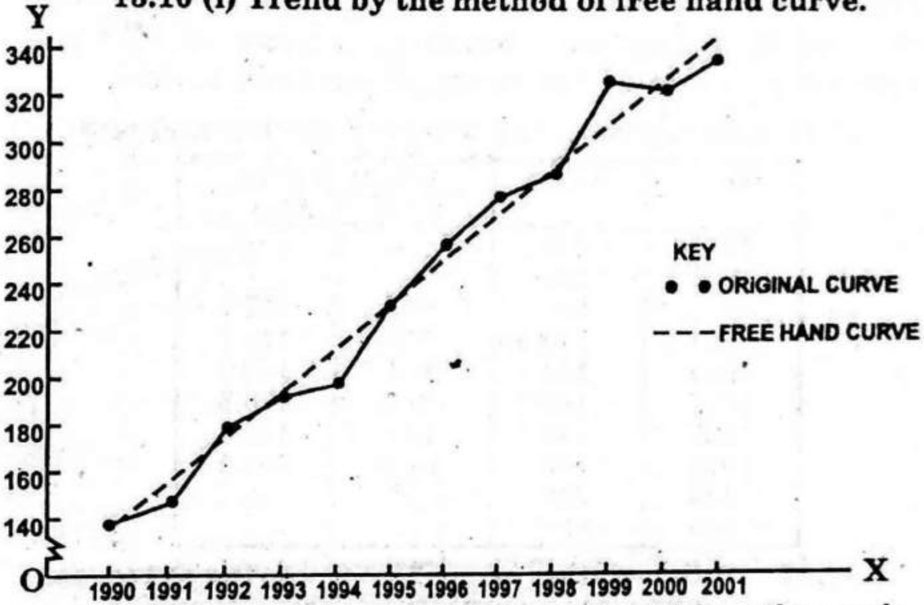
$$X = 2.47,\ Y = 3.55 \text{ and } Z = 1.92.$$

●●●●●●●●●

Chapter 13

TIME SERIES ANALYSIS

13.10 (i) Trend by the method of free hand curve.



In this case, the trend values can be read from the graph.

(ii) Trend by the method of Semi-Averages.

Year	Wages	Total	Averages	Trend values
1990	140			132.5
1991	148			152.7
1992	180	1098	1098÷6 = 183	172.9
1993	195			193.1
1994	200			213.3
1995	235			233.5
1996	260			253.7
1997	280			273.9
1998	290	1825	1825÷6 = 304	294.1
1999	330			314.3
2000	325			334.5
2001	340			354.7

To compute the trend values, we should know the increase or decrease per year. It follows from the two values of the semi-averages that there has been an increase of 304 − 183 = 121 in 6 years (from the middle of 1992 and 1993 to the middle of 1998 and 1999) or 121 ÷ 6 = 20.2 per year. Thus the trend value for 1993 is $183 + \left(\frac{1}{2}\right)(20.2) = 193.1$ and for 1994 is 193.1 + 20.2 = 213.3. The trend values obtained by taking into account an increase of 20.2 per year, are shown in the last column of the above table.

13.11. Calculation of the five-year moving averages.

Year	Value	5-Year Moving Total	5-Year Moving Average
1921	102	--	--
1922	108	--	--
1923	130	638	127.6
1924	140	716	143.2
1925	158	804	160.8
1926	180	884	176.8
1927	196	964	192.8
1928	210	1036	207.2
1929	220	--	--
1930	230	--	--

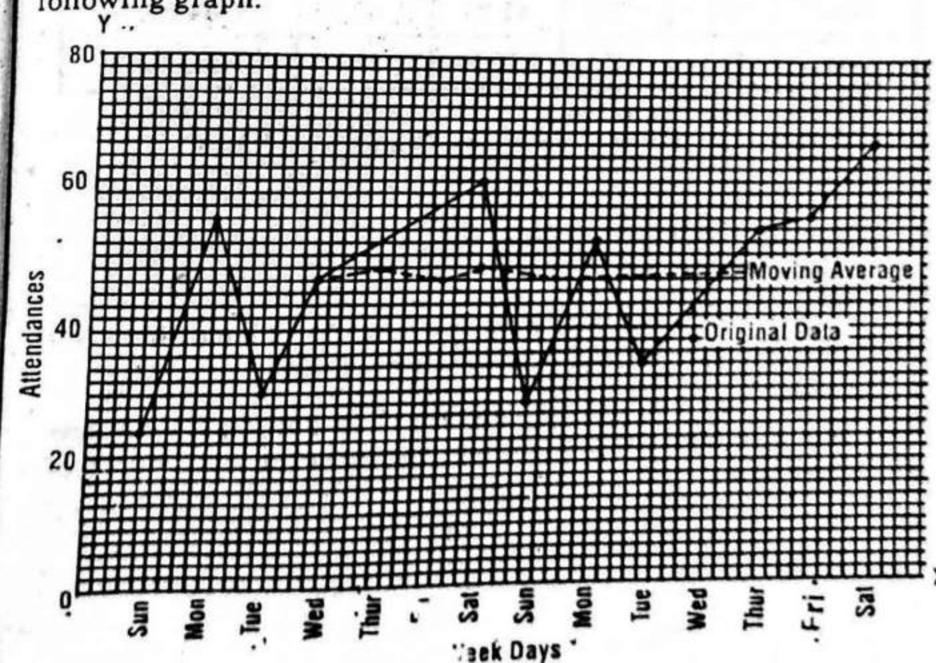
13.12. (ii) Calculation of the 5-year moving averages.

Year	Number	5-Year Moving Total	5-Year Moving Average
1951	170	--	--
1952	210	--	--
1953	188	749	149.8
1954	98	710	142.0
1955	83	705	141.0
1956	131	699	139.8
1957	205	691	138.2
1958	182	700	140.0
1959	90	710	142.0
1960	92	688	137.6
1961	141	641	128.2
1962	183	631	126.2
1963	135	599	119.8
1964	80	565	113.0
1965	60	522	104.4
1966	107	511	102.2
1967	140	---	---
1968	124	---	---

13.3 (a) Calculation of the 7-day moving average for attendances.

Week and Day	Attendances	7-Day Moving Total	7-Day Moving Average
1. Sun.	24	--	--
Mon.	55	--	--
Tues.	29	--	--
Wed.	48	324	46.3
Thur.	52	327	46.7
Fri.	55	324	46.3
Sat.	61	327	46.7
2. Sun.	27	322	46.0
Mon.	52	323	46.1
Tues.	32	324	46.3
Wed.	43	328	46.9
Thur.	53	--	--
Fri.	56	--	--
Sat.	65	--	--

The attendances and moving averages are plotted on the following graph.



(b) The linear trend to be fitted is $\hat{Y} = a + bX$. The arithmetic can be arranged as in the table below:

Week & Day	Y	X	X^2	XY	Trend Values ($\hat{Y} = 46.57 + 1.28X$)
1. Sun.	24	−6.5	42.25	−156.0	38.25
Mon.	55	−5.5	30.25	−302.5	39.53
Tues.	29	−4.5	20.25	−130.5	40.81
Wed.	48	−3.5	12.25	−168.0	42.09
Thur.	52	−2.5	6.25	−130.0	43.37
Fri.	55	−1.5	2.25	−82.5	44.65
Sat.	61	−0.5	0.25	−30.5	45.93
2. Sun.	27	+0.5	0.25	13.5	47.21
Mon.	52	1.5	2.25	78.0	48.49
Tues.	32	2.5	6.25	80.0	49.77
Wed.	43	3.5	12.25	150.5	51.05
Thur.	53	4.5	20.25	238.5	52.33
Fri.	56	5.5	30.25	308.0	53.61
Sat.	65	6.5	42.25	422.5	54.89
Total	652	0	227.50	219.0	651.98

Now $a = \frac{\sum Y}{n} = \frac{652}{14} = 46.57$, and $(\because \sum X = 0)$

$$b = \frac{\sum XY}{\sum X^2} = \frac{291.0}{227.5} = 1.28$$

$\therefore$ $\hat{Y} = 46.57 + 1.28X$, with origin at the midpoint between Saturday of first week and Sunday of second week.

The trend values are determined by putting the various values of X in the equation and they appear in the last column of the above table.

13.14 Calculation of the 7-year moving averages.

Year	Index	7-Year Moving Total	7-Year Moving Average
1980	187	---	---
1981	161	---	---
1982	149	---	---
1983	142	1026	146.6
1984	125	966	138.0
1985	129	935	133.6
1986	133	915	130.7
1987	137	902	128.9
1988	130	907	129.6
1989	129	914	130.6
1990	129	933	133.3
1991	130	977	139.6
1992	136	1016	145.1
1993	152	1105	157.9
1994	171	1234	176.3
1995	169	1383	197.6
1996	218	1542	220.3
1997	258	1704	243.4
1998	279	---	---
1999	295	---	---
2000	314	---	---

13.15.(a) Calculation of the Nine-Yer moving average.

Year	Value	9-Year Moving Total	9-Year Moving Average
1	8	--	--
2	7	--	--
3	5	--	--
4	2	--	--
5	4	62	6.9
6	9	60	6.7
7	10	57	6.3
8	9	59	6.6
9	8	68	7.6
10	6	77	8.6
11	4	79	8.8
12	7	78	8.7
13	11	77	8.6
14	13	74	8.2
15	11	78	8.7
16	9	87	9.7
17	8	95	10.6
18	5	96	10.7
19	10	93	10.3
20	13	90	10.0
21	15	87	9.7
22	12	90	10.0
23	10	97	10.8
24	8	103	11.4
25	6	--	--
26	11	--	--
27	12	--	--
28	16	--	--

(b) Computation of 4-month centred moving average.

Month (1)	Value (2)	4-Month moving totals (3)	4-Month *centred* moving Totals (4)	4-Month Centred moving average (Col. 4÷8)
1	23		--	--
2	26		--	--
		107		
3	28		222	27.75
		115		
4	30		239	29.88
		124		
5	31		257	32.12
		133		
6	35		268	33.50
		135		
7	37		273	34.12
		138		
8	32		279	34.88
		141		
9	34		--	--
10	38		--	--

13.16. Calculaton of the centred moving averages of 4-Quarters.

Year and Quarter (1)	Index Number (2)	4-Quarters moving totals (3)	4-Quarters *centred* moving Totals (4)	4-Quarters Centred moving average (Col. 4÷8)
1951-I	86		--	--
II	80		--	--
		333		
III	83		665	83.12
		332		
IV	84		664	83.00
		332		
1952-I	85		661	82.62
		329		
II	80		652	81.50
		323		
III	80		638	79.15
		315		
IV	78		630	78.75
		315		
1953-I	77		631	78.89
		316		
II	80		634	79.25
		318		
III	81		641	80.12
		323		
IV	80		647	80.88
		324		
1954-I	82		650	81.25
		326		
II	81		654	81.75
		328		
III	83		657	82.12
		329		
IV	82		661	82.62
		332		
1955-I	83		666	83.25
		334		
II	84		672	84.00
		338		
III	85		--	--
IV	86		--	--

13.17. Calculation of the 4-quarter centered moving averages.

Year and Quarter (1)	Data (2)	4-Quarters moving totals (3)	4-Quarters centered moving totals (4)	4-Quarters centered moving average (Col. 4 ÷ 8)
2000-I	102		--	--
II	71		--	--
		318		
III	47		659	82.38
IV	98		717	89.62
		376		
2001-I	125		778	97.25
		402		
II	106		937	117.12
		535		
III	73		1226	153.25
		691		
IV	231		1505	188.12
		814		
2002-I	281		1764	220.50
		950		
II	229		2157	269.62
		1207		
III	209		2617	327.12
		1410		
IV	488		3038	379.75
		1628		
2003-I	484		3504	438.00
		1876		
II	447		4230	528.75
		2354		
III	457		--	--
IV	966		--	--

13.18.(a) Calculation of the 2-year centered moving averages.

Year (1)	Visitors (2)	2-Year moving totals (3)	2-Year centred moving totals (4)	2-Year centred moving average (Col. 4 ÷ 4) (5)
1995	31			
		80		
1996	49		203	50.75
		123		
1997	74		259	64.75
		136		
1998	62		263	65.75
		127		
1999	65		265	66.25
		138		
2000	73		281	70.25
		143		
2001	70		297	74.25
		154		
2002	84		324	81.00
		170		
2003	86		335	83.75
		165		
2004	79			

Calculation of the 3-years weighted moving average with weights 1, 2, 1 respectively.

Year (1)	Visitors (2)	3-Year weighted moving totals (3)	3-Year weighted moving average (Col. 3 ÷ 4) (4)
1995	31	--	--
1996	49	203	50.75
1997	74	259	64.75
1998	62	263	65.75
1999	65	265	66.25
2000	73	281	70.25
2001	70	297	74.25
2002	84	324	81.00
2003	86	335	83.75
2004	79	--	--

The calculations shown above indicate that the 2-year centred moving average is equivalent to a 3-year weighted moving average with weights 1, 2, 1 respectively.

(b) Calculation of the 3-year weighted moving average with weights, 1, 4, 1 respectively.

Year (1)	Visitors (2)	3-Year weightedm oving totals (3)	3-Year weighted moving average (Col. 3 ÷ 4) (4)
1995	31	--	--
1996	49	301	50.17
1997	74	407	67.83
1998	62	387	64.50
1999	65	395	65.83
2000	73	427	71.17
2001	70	437	72.83
2002	84	492	82.00
2003	86	507	84.50
2004	79	--	--

13.19. Trend by using the method of

(i) Semi-averages **(ii) 3-year moving averages**

Year	Values	Totals	Average	Trend Values	Values	3-Year Moving Total	3-Year Moving Average
1968	2			3.8	2	--	--
1969	4			4.6	4	12	4.0
		20	5				
1970	6			5.4	6	18	6.0
1971	8			6.2	8	21	7.0
1972	7			7.0	7	21	7.0
1973	6			7.8	6	21	7.0
1974	8			8.6	8	24	8.0
		36	9				
1975	10			9.4	10	30	10.0
1976	12			10.2	12	--	--

(iii) Least-squares for fitting a straight line.

Year	Values Y	X	X^2	XY	Trend Values
1968	2	−4	16	−8	3
1969	4	−3	9	−12	4
1970	6	−2	4	−12	5
1971	7	−1	1	−8	6
1972	7	0	0	−40	7
1973	6	1	1	6	8
1974	8	2	4	16	9
1975	10	3	9	30	10
1976	12	4	16	48	11
Total	63	0	60	+100 / + 60	63

Let the equation of the straight line be

$$\hat{Y}_t = a + bX.$$

Then the two normal equations reduce to

$$\sum X = na$$

$$\sum XY = b\sum X^2$$

Substituting the values, we get

$$a = \frac{\sum Y}{n} = \frac{63}{9} = 7, \text{ and}$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{60}{60} = 1$$

Hence the required equation is $\hat{Y}_t = 7 + X$ with origin at 1972. The trend values are obtained by substituting the values of X in the fitted equation. These values appear in the last column of the above table. We prefer the least-squares trend as it is the better fit.

13.20. (b) Determination of the trend by using the method of

(i) 3-year moving averages (ii) Least-squares for fitting a st. line

Year	Value	3-year moving Total	3-year moving Average	Value Y	X	X^2	XY	Trend value
1970	12	--	--	12	−4	16	−48	18.09
1971	23	72	24.0	23	−3	9	−69	24.04
1972	37	108	36.0	37	−2	4	−74	29.99
1973	48	126	42.0	48	−1	1	−48	35.94
1974	41	126	42.0	41	0	0	−239	41.89
1975	37	126	42.0	37	1	1	37	47.84
1976	48	146	48.7	48	2	4	96	53.79
1977	61	179	59.7	61	3	9	183	59.74
1978	70	--	---	70	4	16	280	65.69
Total				377	0	60	596 +357	377.01

Let the equation of the straight line be

$$\hat{Y}_t = a + bX$$

Then the two normal equations reduce to

$$\sum Y = na \text{ and } \sum XY = b\sum X^2.$$

Substituting the values, we get

$$a = \frac{\sum Y}{n} = \frac{377}{9} = 41.89, \text{ and } b = \frac{\sum XY}{\sum X^2} = \frac{357}{60} = 5.95$$

Hence the equation of the least-squares trend is $\hat{Y}_t = 41.89 + 5.95X$, with origin at 1974 and units of X are 1 year. The trend values are computed by substituting the values of X in the equation and are shown in the last column.

We prefer the *least-squares* trend as it is the better fit.

13.21. (a) Fitting a linear trend.

Let the equation of the linear trend be $\hat{Y}_t = a + bX$.
Then the two normal equations for determining a and b are

$$\sum Y = na + b\sum X \quad \text{and} \quad \sum XY = a\sum X + b\sum X^2$$

Since the number of years in the data is odd, we assign X=0 to the middle year 1997, X=1,2 to the successive years and X=-1, -2 to the preceding years. The normal equations then reduce to

$$\sum Y = na \quad \text{and} \quad \sum XY = b\sum X^2$$

The arithmetic can be arranged as in the table below:

Year	X	Index (Y)	XY	X^2	Trend Values
1995	−2	125	−250	4	124.4
1996	−1	114	−114	1	112.0
1997	0	99	0	0	99.6
1998	1	80	80	1	87.1
1999	2	80	160	4	74.7
Σ	---	498	−124	10	497.8

Substituting, we get

$$a = \frac{\sum Y}{n} = \frac{498}{5} = 99.6, \qquad \textit{and}$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{-124}{10} = -12.4$$

Hence the equation of the desired linear trend is

$\hat{Y}_t = 99.6 - 12.4X$ with origin at 1997 and units of X are 1 year. The trend values are shown in the last column of the above table.

$$a = \frac{\sum Y}{n} = \frac{438.9}{11} = 39.9, \text{ and} \qquad (\because n = 11)$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{-84.4}{110} = -0.7673.$$

Hence the required equation is

$Y = 39.9 - 0.7673X$, with origin at 1953.

Putting $X = -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$ in the equation, we obtain the trend values as

43.74, 42.97, 42.20, 41.43, 40.67, 39.90, 39.13, 38.37, 37.60, 36.83, 36.06.

13.22. Fitting of a straight line trend and a parabolic trend.

As the number of years in the data is odd, we can assign $X=0$ to the middle year 1980, $X=1,2,3$ to the succeeding years and $X=-1,-2,-3$ to the preceding years.

(i) Let the equation of the straight line be

$$Y_t = a + bX.$$

Then the two normal equations reduce to

$$\sum Y = na \text{ and } \sum XY = b\sum X^2.$$

(ii) Let the equation of the quadratic parabola be $Y_t = a + bX + cX^2$.

Then the three normal equations reduce to

$$\sum Y = na + c\sum X^2$$

$$\sum XY = b\sum X^2$$

$$\sum X^2 Y = a\sum X^2 + c\sum X^4$$

The arithmetic can be arranged as in the table below:

Year	Y	X	X^2	X^4	XY	X^2Y	Y^2
1977	88	−3	9	81	−264	792	7744
1978	101	−2	4	16	−202	404	10201
1979	105	−1	1	1	−105	105	11025
1980	91	0	0	0	0	0	8281
1981	113	1	1	1	113	113	12769
1982	120	2	4	16	240	480	14400
1983	132	3	9	81	396	1188	17424
$\sum$	750	0	28	196	178	3082	81844

Substituting the values in the normal equations of the straight line, we get

$$a = \frac{\sum Y}{n} = \frac{750}{7} = 107.14, \text{ and}$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{178}{28} = 6.38$$

Hence the equation of the straight line is

$$\hat{Y} = 107.14 + 6.38X,$$

with origin at 1980 and units of X are 1 year.

The estimate of the profits in 1985 is obtained by putting $X=5$ in the fitted equation. Thus

$$\hat{Y}_t = 107.14 + 6.38(5) = 139.04$$

Substituting the values in the normal equations of the quadratic parabola, we get

$$b = \frac{\sum XY}{\sum X^2} = \frac{178}{28} = 6.36; \text{ and}$$

$$7a + 28c = 750$$

$$28a + 196c = 3082$$

Solving them simultaneously, we get

$$a = 103.24, \text{ and } c = 0.976$$

Hence the equation of the desired parabola trend is

$$\hat{Y}_t = 103.24 + 6.36X + 0.796X^2,$$

with origin at 1980 and units of X are 1 year.

(iii) In order to determine which is the better fitting trend, we calculate the sum of squares of residuals in both cases. In case of straight line trend, the sum of squares of residuals is

$$\sum(Y - \hat{Y}_t)^2 = \sum Y^2 - a\sum Y - b\sum XY$$

$$= 81844 - (107.14)(850) - (6.38)(178)$$

$$= 81844 - 80355 - 1135.64 = 353.36$$

The sum of squares of residuals in case of parabolic trend is

$$\sum(Y-\hat{Y}_t)^2 = \sum Y^2 - a\sum Y - b\sum XY - c\sum X^2Y$$
$$= 81844 - (103.24)(750) - (6.38)(178) - (0.976)(3.082)$$
$$= 81844 - 77430 - 1135.64 - 3008.03$$
$$= 270.33$$

The parabolic trend is the better fit as it has the smaller sum of squares of residuals.

13.23. Let the equation of the second degree parabola be

$$Y_t = a + bX + cX^2 .$$

Then the normal equations for determining a, b and c are:

$$\sum Y = na + b\sum X + c\sum X^2$$
$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$
$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

The arithmetic can be arranged as in the table below:

Year	Production Y	X	X^2	X^3	X^4	XY	X^2Y
1970–71	136	0	0	0	0	0	0
1971–72	162	1	1	1	1	162	162
1972–73	187	2	4	8	16	374	748
1973–74	225	3	9	27	81	675	2025
1974–75	272	4	16	64	256	1088	4352
1975–76	277	5	25	125	625	1385	6925
1976–77	322	6	36	216	1296	1932	11592
Σ	1581	21	91	441	2275	5616	25804

Substituting these sums, we get

$$7a + 21b + 91c = 1581$$
$$21a + 91b + 441c = 5616$$
$$91a + 441b + 2275c = 25804$$

Solving these equations simultaneously, we get

$$a = 133.09,\ b = 30\ 25 \text{ and } c = 0.155$$

Hence the equation of the desi ed quadratic parabola is

$$Y_t = 133.09 + 30.25X + 0.155X^2$$

Estimated production for 1978–79 is obtained by putting $X=8$ in the fitted equation. Therefore

$$\hat{Y}_t = 133.09 + 30.25(8) + 0.155(8)^2 = 385$$

13.24. The equation of the parabola of second order is

$$Y_t = a + bX + cX^2$$

Since the number of years is odd, we therefore assign $X = 0$ to the middle year, i.e. 2000 and $X = -1, -2, \ldots$ to the preceding years and $X = 1,2, \ldots$ to the successive years.

The normal equations then become

$$\Sigma Y = na + c\Sigma X^2$$
$$\Sigma XY = b\Sigma X^2$$
$$\Sigma X^2Y = a\Sigma X^2 + c\Sigma X^4$$

The arithmetic can be arranged as in the table below:

Year	Production Y	X	X^2	X^4	XY	X^2Y	Trend values
1995	17	–5	25	625	–85	425	18.52
1996	20	–4	16	256	–80	320	18.85
1997	19	–3	9	81	–57	171	20.32
1998	26	–2	4	16	–52	104	22.93
1999	24	–1	1	1	–24	24	26.68
2000	40	0	0	0	0	0	31.57
2001	35	1	1	1	35	35	37.60
2002	35	2	4	16	70	140	44.77
2003	51	3	9	81	153	459	53.08
2004	74	4	16	256	296	1184	62.53
2005	69	5	25	625	345	1725	73.12
Σ	410	0	110	1958	601	4587	409.97

Substituting these summations in the normal equations, we get

$$11a + 110c = 410$$
$$110b = 601$$
$$110a + 1958c = 4587$$

Solving them simultaneously, we obtain

$$a = 31.57, b = 5.46 \text{ and } c = 0.57$$

Hence the equation of the required parabola is

$$Y_t = 31.57 + 5.46X + 0.57X^2$$

with origin at 2000 and units of X are 1 year.

The trend values are obtained by putting the different values of X in the above equation, and they are giver. in the last column in the table above.

13.25. The number of years is even and they are equispaced with interval of one year. We therefore assign X = 0 to the midpoint of the years 2004 and 2005, X = –1, –3, –5, –7 to the preceding years and X = 1,3,5,7 to the successive years (half year units), so that

$$\Sigma X = 0 = \Sigma X^3$$

The equation of the parabola fitting the data is

$$Y_t = a_0 + a_1X + a_2X^2 \quad \text{(t is coded)}$$

The normal equations for determining a_0, a_1 and a_2 then reduce to

$$\Sigma Y = na_0 + a_2\Sigma X^2$$

$$\Sigma XY = a_1\Sigma X^2$$

$$\Sigma X^2Y = a_0\Sigma X^2 + a_2\Sigma X^4$$

The arithmetic involved in computation is arranged in the following table:

Year	Coded year (X)	Values (Y)	X^2	X^4	XY	X^2Y	Trend values
2001	–7	273.7	49	2401	–1915.9	13411.3	273.96
2002	–5	293.5	25	625	–1467.5	7337.5	292.90
2003	–3	315.0	9	81	–945.0	2835.0	314.32
2004	–1	336.8	1	1	–336.8	336.8	338.22
2005	+1	364.4	1	1	364.4	364.4	364.60
2006	3	394.8	9	81	1184.4	3553.2	393.46
2007	5	424.2	25	625	2121.0	10605.0	424.80
2008	7	458.7	49	2401	3210.9	22476.3	458.62
Σ	0	2861.1	168	6216	2215.5	60919.5	2860.88

Substituting these values in the normal equations, we get

$$8a_0 + 168a_2 = 2861.1$$
$$168a_1 = 2215.5$$
$$168a_0 + 6216a_2 = 6091.5$$

Solving these equations, we obtain

$$a_0 = 351.1023, a_1 = 13.1875 \text{ and } a_2 = 0.3112$$

Hence the equation of the required parabola is

$$\hat{Y}_t = 351.10 + 13.19X + 0.31X^2$$

with origin at the midpoint of 2004 and 2005 (i.e. January 1, 2005) and X is measured in units of a half year. The trend values appear in the last column.

13.26. Let the equation of the quadratic parabola be

$$\hat{Y}_t = a + bX + cX^2$$

The normal equations for determining *a*, *b* and *c* are:

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

The arithmetic can be arranged as in the table below:

Year	Y	X	X^2	X^3	X^4	XY	X^2Y
1924	187	0	0	0	0	0	0
1927	142	3	9	27	81	426	1278
1930	133	6	36	216	1296	798	4788
1933	129	9	81	729	6561	1161	10449
1936	136	12	144	1728	20736	1632	19584
1939	169	15	225	3375	50625	2535	38025
1942	279	18	324	5832	104976	5022	90396
$\sum$	1175	63	819	11907	184275	11574	164520

Substituting these sums, we get

$$7a + 63b + 819c = 1175 \quad \text{... (A)}$$

$$63a + 819b + 11907c = 11574 \quad \text{... (B)}$$

$$819a + 11907b + 184275c = 164520 \quad \text{... (C)}$$

Now 9(A): $63a + 567b + 7371c = 10575$ ... (D)

(C)-(D): $252b + 4536c = 999$ ... (E)

13(B): $819a + 10647b + 154791c = 150462$ ... (F)

(C)-(F): $1260b + 29484c = 14058$ ... (G)

5(E): $1260b + 22680c = 4995$ ... (H)

(G)-(H): $6804c = 9063$

$\therefore$ $c = 1.332$

Putting $c = 1.332$ in (E), we get

$$252b + 4536\,(1.332) = 999$$

or $252b = 999 - 6041.952$

or $b = \dfrac{999 - 6041.952}{252} = -20.012$

Putting $b = -20.012$ and $c = 1.332$ in (A), we get $a = 192$

Hence the required equation of the quadratic parabola is

$$\hat{Y}_t = 192 - 20.012X + 1.332\,X^2,$$

with origin at 1924.

To estimate the value of the index for 1935, we put $X = 11$ in the above equation and get

$$\hat{Y} = 192 - 20.012\,(11) + (1.332)\,(11)^2$$

$$= 192 - 220.132 + 161.172 = 133.$$

13.27. The given curve $Y = ab^X$ may be written as

$$\log Y = \log a + X \log b$$

or $Y' = A + BX$, (Form of a st. line)

where $Y' = \log Y$, $A = \log a$ and $B = \log b$.

The two normal equations are

$$\sum Y' = nA + B\sum X$$

$$\sum XY' = A\sum X + B\sum X^2$$

The calculations involved are shown in the following table:

Year	X	Y	X^2	$Y' = \log Y$	XY'
1911	0	5.38	0	0.7308	0
1921	1	7.22	1	0.8585	0.8585
1931	2	9.64	4	0.9841	1.9682
1941	3	12.70	9	1.1038	3.3114
1951	4	17.80	16	1.2504	5.0016
1961	5	24.02	25	1.3806	6.9030
1971	6	31.34	36	1.4961	8.9766
$\sum$	21	--	91	7.8043	27.0193

Substituting these values in the normal equations, we get

$$7A + 21B = 7.8043$$

$$21A + 91B = 27.0193$$

Solving them, we find that $A = 0.7285$ and $B = 0.1288$.

Thus a = antilog of A

= antilog (0.7285) = 5.35, and

b = antilog of B

= antilog (0.1288) = 1.345.

Hence the equation of the required curve is

$\hat{Y} = 5.35\ (1.345)^X$, where origin is at 1911 and X is measured in ten yearly intervals.

In order to forecast the populaton for the year 1991, we put $X=8$ in the fitted equation. Thus $Y_{1991} = 5.35\ (1.345)^8 = 57.30$

13.28. (b) To compute the seasonal indices, we first find the 4-quarter centred moving averages.

Year and Quarter (1)	Y (2)	4-quarter moving totals (3)	4-quarter *centred* moving Totals (4)	4-quarter *centred* moving Averages (Col 4÷8)	$\frac{\text{Data}}{\text{Trend}} \times 100$
	TSCI			*TC*	*SI*
1949 – 1	105		--	--	--
2	77		--	--	--
		345			
3	68		692	86.5	78.6
		347			
4	95		700	87.5	108.6
		353			
1950–1	107		712	89.0	120.2
		359			
2	83		729	91.1	91.1
		370			
3	74		750	93.8	78.9
		380			
4	106		776	97.0	109.3
		396			
1951–1	117		804	100.5	116.4
		408			
2	99		822	102.8	96.3
		414			
3	86		--	--	--
4	112		--	--	--

The percentages are put in the following table in order to obtain the seasonal indices:

Year	Quarters 1	2	3	4	
1949	--	--	78.6	108.6	
1950	120.2	91.1	78.9	109.3	
1951	116.4	96.3	--	--	
Total	236.6	187.4	157.5	217.9	
Mean (S.I.)	118.30	93.70	78.75	108.95	Total 399.7

The sum of these mean percentages is 399.7 which is close to the desired 400, so that no adjustment is necessary. Hence the desired seasonal indices are 118.30, 93.70, 78.75 and 108.95.

13.29. To compute the seasonal indices, we first find the 4-quarter centred moving averages.

Year and Quarter (1)	Prices Y (2)	4-quarter moving Total (3)	4-quarter *centred* moving Totals (4)	4-quarter *centred* moving Averages (Col 4÷8)	$\frac{\text{Data}}{\text{Trend}} \times 100$
	TSCI			*TC*	*SI*
1961 – I	122		--	--	--
II	125		--	--	--
		482			
III	118		961	120.1	98.3
		479			
IV	117		947	118.4	98.8
		468			
1962–I	119		932	116.5	102.1
		464			
II	114		920	115.0	99.1
		456			
III	114		898	112.2	101.6
		442			
IV	109		869	108.6	100.4
		427			
1963–I	105		833	104.1	100.9
		406			
II	99		792	99.0	100.0
		386			
III	93		753	94.1	98.8
		367			
IV	89		715	89.4	99.6
		348			
1964–I	86		686	85.8	100.2
		338			
II	80		671	83.9	95.4
		333			
III	83	--	--	--	--
IV	84	--	--	--	--

The percentages are put in the following table in order to obtain the seasonal indices:

Year	Quarters				
	I	II	III	IV	
1961	--	--	98.3	98.8	
1962	102.1	99.1	101.6	100.4	
1963	100.9	100.0	98.8	99.6	
1964	120.2	95.4	--	--	Total
Total	303.2	294.5	298.7	298.8	398.5
Main	101.1	98.2	99.6	99.6	400.1
Seasonal Index	101.5	98.6	100	100	

13.30. (a) To compute the seasonal indices by the percentage of annual average method, we first compute the annua-average as below:

Year	Quarters				Annual or Yearly	
	I	II	III	IV	Total	Average
1996	118	87	47	83	335	83.75
1997	94	73	41	68	276	69.00
1998	73	61	36	56	226	56.50

We next divide each of the quarterly observations by the corresponding annual-average and express the result as a percentage. The percentage so obtained are given below:

Year	Quarters				Total
	I	II	III	IV	
1996	140.90	103.88	56.12	99.10	
1997	136.23	105.80	59.42	98.55	
1998	129.20	107.96	63.72	99.12	
Total	406.33	317.64	179.26	296.77	
Mean	135.44	105.88	59.75	98.92	399.99

The sum of the mean percentages is 399.99 which is close to the desired total of 400, therefore no adjustment is necessary. Hence the desired seasonal indices are 135.44, 105.88, 59.75 and 98.92.

(b) To compute the indices of seasonal variations using ratio-to-moving average method, we first obtain the 4-quarter centred moving averages.

Year and Quarter (1)	Sales Y (2)	4-quarter moving Total (3)	4-quarter centred moving Totals (4)	4-quarter centred moving Average (Col 4÷8)	$\frac{\text{Data}}{\text{Trend}}\times 100$
	$TSCI$			TC	SI
1996 – I	118				
II	87		--	--	--
		335	--	--	--
III	47				
			646	80.8	
		311			
IV	83				58.2
			608	76.0	
		297			109.2
1997–I	94				
			588	73.5	
		291			127.9
II	73				
			567	70.9	
		276			103.0
III	41				
			531	66.4	61.7
		255			
IV	68				
			498	62.2	109.3
		243			
1998–I	73				
			481	60.1	121.5
		238			
II	61				
			464	58.0	105.2
		226			
III	36	--	--	--	--
IV	56	--	--	--	--

The percentages are put in the following table in order to obtain the seasonal indices:

Year	Quarters				
	I	II	III	IV	
1996	--	--	58.2	109.2	
1997	127.9	103.0	61.7	109.3	
1998	121.5	105.2	--	--	
Total	249.4	208.2	119.9	218.5	Total
Mean	124.7	104.1	60.0	109.2	398.0
Seasonal Index	125.3	104.6	60.3	109.7	399.9

Deseasonalization of Data for 1997

13.31. To estimate the seasonal indices, we first compute the trend by means of centred moving averages.

Year and Quarter (1)	Sales Y (2)	4-quarters moving Total (3)	4-quarters *centred* moving Totals (4)	4-quarters *centered moving* Averages (Col. 4 ÷ 8)	$\frac{\text{Data}}{\text{Trend}} \times 100$
1994–I	48		--	--	--
II	52	151	--	--	--
III	16	153	304	38.00	42.1
IV	35	147	300	37.50	93.3
1995–I	50	153	300	37.50	133.3
II	46	158	311	38.88	118.3
III	22	176	334	41.75	52.7
IV	40	164	340	42.50	94.1
1996–I	68	168	332	41.50	163.9
II	34	163	331	41.38	82.2
III	26	188	351	43.88	59.3
IV	35	210	398	49.75	70.4
1997–I	93	200	410	51.25	181.5
II	56	210	410	51.25	109.3
III	16		411	51.38	31.1
IV	45	206	407	50.88	88.4
1998–I	84	219	425	53.12	158.1
II	61	222	441	55.12	110.7
III	29		--	--	--
IV	48		--	--	--

The percentages are put in the following table in order to obtain the seasonal indices:

Year	Quarters I	II	III	IV	
1994	--	--	42.1	93.3	
1995	133.3	118.3	52.7	94.1	
1996	163.9	82.2	59.3	70.4	
1997	181.5	109.3	31.1	88.4	
1998	158.1	110.7	--	--	
Total	636.8	420.5	185.2	346.2	Total
Mean	159.2	105.1	46.3	86.6	397.2
Seasonal Index	160.3	105.8	46.6	87.2	399.9

To forecast sales for each quarter of 1999, first find the projected trend values for these quarters by extending the moving averages (graphically) to these quarters, then multiply these projected trend values by the corresponding seasonal indices and divide each product by 100.

13.32. Let the equation of the linear trend be $Y_t = a + bX$. Since the number of quarters in the observed series is even, therefore the middle point of the two middle quarters is taken as X = 0.

Year	Y	X	X^2	XY	Trend $\hat{Y}_t = 103.56 - 1.60X$
1961-I	122	−15	225	−1830	127.56
II	125	−13	169	−1625	124.36
III	118	−11	121	−1298	121.16
IV	117	−9	81	−1053	117.96
1962-I	119	−7	49	−833	114.76
II	114	−5	25	−570	111.56
III	114	−3	9	−342	108.36
IV	109	−1	1	−109	105.16
1963-I	105	1	1	105	101.96
II	99	3	9	297	98.76
III	93	5	25	465	95.56
IV	89	7	49	623	92.36
1964-I	86	9	81	774	89.16
II	80	11	121	880	85.96
III	83	13	169	1079	82.76
IV	84	15	225	1260	79.56
Σ	1657	0	1360	−2177	1656.96

Since $\sum X = 0$, the normal equations take the form

$$\sum Y = na \text{ and } \sum XY = b\sum X^2$$

Substituting the values, we obtain

$$a = \frac{\sum Y}{n} = \frac{1657}{16} = 103.56, \text{ and}$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{-2177}{1360} = -1.60$$

Thus the equation of the linear trend is

$$\hat{Y}_t = 103.56 - 1.60X,$$

with origin at the middle of IV quarter 1962 and I quarter 1963, and X is measured in half quarter units. The trend values are computed by substituting the values of X in the least-squares equation and they appear in the last column of the above table.

Next, we divide each observation in the original data by the corresponding trend value and multiply it by 100. The percentages so obtained are arranged by quarters as shown in the following table in order to compute the indices of seasonal variation.

Computation of Seasonal Indices

Year	Quarters				
	I	II	III	IV	
1961	95.64	100.51	97.39	99.19	
1962	103.69	102.19	*105.20	103.65	
1963	102.98	100.24	97.32	96.36	
1964	96.46	*93.07	100.29	105.58	
Total	398.77	302.94	295.00	404.78	Total
Mean	99.69	100.98	98.33	101.20	400.20
Seasonal Indices	99.64	100.93	98.28	101.15	400

* Discard these extreme relatives before computing totals

13.33. Let the equation of the linear trend be $Y_t = a + bX$. Since the number of quarters in the observed series is even, therefore the middle point of the two middle quarters is taken as X=0.

Year and quarter	Y	X	XY	X^2	$\hat{Y}_t = 109.42 + 3.81X$
1982-1	70	−11	−770	121	67.51
2	81	−9	−729	81	75.13
3	89	−7	−623	49	82.75
4	115	−5	−575	25	90.37
1983-1	75	−3	−225	9	97.99
2	93	−1	−93	1	105.61
3	108	1	108	1	113.23
4	152	3	456	9	120.85
1984-1	80	5	400	25	128.47
2	105	7	735	49	136.09
3	150	9	1350	81	143.71
4	195	11	2145	121	151.33
$\sum$	1313	0	2179	572	----

Since $\sum X = 0$, the normal equations take the form

$$\sum Y = na \text{ and } \sum XY = b\sum X^2.$$

Substituting the values, we obtain

$$a = \frac{\sum Y}{n} = \frac{1313}{12} = 109.42, \text{ and}$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{2179}{572} = 3.81$$

Thus the trend line is $\hat{Y}_t = 109.42 + 3.81X$, with origin at the middle of 2nd and 3rd quarters of 1983, and X is measured in half quarter units. The trend values are computed by substituting the values of X in the least-squares equation and they appear in the last column of the above table.

Next, we divide each observation in the original data by the corresponding trend value and multiply it by 100. The percentages so obtained are arranged by quarters as shown in the following table in order to compute the indices of seasonal variation.

Computation of Seasonal Indices

Year	Quarters				
	Summer	Autumn	Winter	Spring	
1982	*103.69	*107.81	107.55	127.25	
1983	76.54	88.06	*95.38	125.78	
1984	62.27	77.15	104.38	128.86	
Total	138.81	165.21	211.93	381.89	Total
Mean	69.40	82.60	105.96	127.30	385.26
Seasonal Indices	72.06	85.76	110.02	132.18	400.02

* Discard these extreme relatives before computing totals.

$$\text{Deseasonalized data} = \frac{\text{Period's original value}}{\text{Period's seasonal index}} \times 100$$

$$= \frac{TSCI}{S} \times 100 = TCI \times 100$$

Thus the deseasonalized data for 1984 are shown below:

Deseasonalization of the 1984 values

Quarter	Y	Seasonal Index	Diseasonalized Data
Summer	80	72.06	111.02
Autumn	105	85.76	122.43
Winter	150	110.02	136.34
Spring	195	132.18	147.53

13.34. Let the equation of the linear trend b $Y_t = a + bX$. Since the number of quarters in the observe series is even, therefore the middle of the two middle quarters is taken as X = 0.

Year	Y	X	X^2	XY	Trend values $\hat{Y}_t = 114 + 0.247X$
2000–I	107	–15	225	–1605	110.295
II	115	–13	169	–1495	110.789
III	103	–11	121	–1133	111.283
IV	98	–9	81	–882	111.777
2001–I	118	–7	49	–826	112.271
II	122	–5	25	–610	112.765
III	115	–3	9	–345	113.259
IV	104	–1	1	–104	113.753
2002–I	126	+1	1	+126	114.247
II	129	3	9	387	114.741
III	118	5	25	590	115.235
IV	107	7	49	749	115.729
2003–I	121	9	81	1089	116.223
II	122	11	121	1342	116.717
III	116	13	169	1508	117.211
IV	103	15	225	1545	117.705
Total	1824	0	1360	336	1824

Since $\Sigma X = 0$, the normal equations are reduced to

$$\Sigma Y = na \text{ and } \Sigma XY = b\Sigma X^2$$

Substituting the values, we get

$$a = \frac{1824}{16} = 114 \text{ and } b = \frac{336}{1360} = 0.247$$

Thus the trend line is $Y_t = 114 + 0.247X$, with origin at the middle of IV quarter 2001 and I quarter 2002, and X is measured in half quarter units. The trend values are computed by substituting the values of X in the least squares equation and they are shown in the last column of the table. Next, we divide each observation in the original data by the corresponding trend value and multiply it by 100. The percentages thus obtained are arranged in the following table in order to compute the indices of seasonal variation.

Computation of Seasonal Indices

Year	Quarters I	II	III	IV	
2000	97.01	103.80	92.56	87.67	
2001	105.10	108.19	101.54	91.43	
2002	110.29	112.43	102.40	92.46	
2003	104.11	104.53	98.97	87.51	
Total	416.51	428.95	395.47	359.07	Total
Mean (S.I.)	104.13	107.24	98.87	89.77	400

13.35. (a) Let the equation of the linear trend be $Y_t = a + bX$. Since the number of quarters in the observed series is even, therefore the middle point of the two middle quarters is taken as X = 0.

Year and quarter	Y	X	XY	X^2	Trend $\hat{Y}_t = 94.08 + 0.83X$
1949-1	105	−11	−1155	121	84.95
2	77	−9	−693	81	86.61
3	68	−7	−476	49	88.27
4	95	−5	−475	25	89.93
1950-1	107	−3	−321	9	91.59
2	83	−1	−83	1	93.25
3	74	1	74	1	94.91
4	106	3	318	9	96.57
1951-1	117	5	585	25	98.23
2	99	7	693	49	99.89
3	86	9	774	81	101.55
4	112	11	1232	121	103.21
Σ	1129	0	473	572	1128.96

34

Since $\sum X = 0$, the normal equations take the form

$$\sum Y = na \quad \text{and} \quad \sum XY = b\sum X^2.$$

Substituting the values, we obtain

$$a = \frac{\sum Y}{n} = \frac{1129}{12} = 94.08, \text{ and}$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{473}{572} = 0.83.$$

Thus the trend line is $\hat{Y}_t = 94.08 + 0.83X$, with the origin at the middle of 2nd quarter 1950 and 3rd quarter 1950, and X is measured in half quarter units. The trend values are computed by substituting the values of X in the least-squares equation and they appear in the last column of the above table.

(b) (i) Now we divide each observation in the original data by the corresponding trend value and multiply it by 100. The percentages so obtained are arranged by quarters as shown in the following table in order to compute the indices of seasonal variation.

Computation of Seasonal Indices

Year	Quarters 1	2	3	4	
1949	123.60	88.90	77.04	105.64	
1950	116.82	89.01	77.97	109.76	
1951	119.11	*99.11	*84.69	108.52	
Total	359.53	177.91	155.01	323.92	Total
Mean	119.84	88.96	77.50	107.97	394.27
Seasonal Indices	121.58	90.25	78.62	109.54	399.99

*** Discard these extreme relatives before computing totals.**

(ii) Deseasonalization of data

$$\text{Deseasonalized data} = \frac{\text{Period's original value}}{\text{Period's seasonal index}} \times 100$$

$$= \frac{TSCI}{S} \times 100 = TCI \times 100$$

Thus the deseasonalized data are shown below:

Deseasonalization of data

Year	Quarters			
	1	2	3	4
1949	86.36	85.32	86.49	86.73
1950	88.01	91.97	94.12	96.77
1951	96.23	109.70	109.39	102.25

13.36. Expressing the data for each quarter as a percentage of the data for the previous quarter, we get the link relatives as below:

Year	Quarter			
	1	2	3	4
1970	---	111.6	103.2	85.3
1971	108.2	110.3	111.4	78.2
1972	104.3	118.3	105.6	78.7
1973	108.5	118.0	107.3	77.2
Median	108.2	114.8	106.4	78.4

Next, we calculate the chain relatives for the four quarters, taking the value of the first quarter equal to 100%. The chain relatives are:

Quarter	1	2	3	4	1
Chain Relative	100	114.8	122.1	95.7	103.5

Continuing the process, we find that the chain relative for the first quarter works out to be 103.5, which as a matter of fact, ought to have been 100. The increase of (103.5–100), *i.e.* 3.5 is due to secular trend present in the data. Adjusting for the trend, we subtract one-fourth of 3.5 from the second quarter figure, two-fourth from the third quarter figure and three-fourth from

13.37. (b) We fit a straight line $Y_t = a+bX$, by the method of least-squares to the yearly averages, which are assumed to correspond to the midpoint of each year.

Year	Yearly Total	Yearly average (Y)	X	X^2	XY
2001	156	39.00	–1	1	–39.00
2002	180	45.00	0	0	0
2003	279	69.75	+1	1	69.75
Total	---	153.75	0	2	30.75

The two normal equations for determining a and b are

$\Sigma Y = na$ and $\Sigma XY = b\Sigma X^2$

Substituting, we get $a = 51.25$ and $b = 15.38$.

Thus the trend line is $Y_t = 51.25 + 15.38X$.

The value of b = 15.38 indicates that Y values increase by Y = 15.38 after every year or ¼ of (15.38) = 3.84 after every quarter. Assuming that the given quarterly data correspond to the middle of the quarter, we calculate the trend values as below.

When X = 0, which corresponds to July 1, 2002 Y = 51.25. But we need the value of Y, a half quarter later. This is the trend value corresponding to third quarter of 2002

The quarterly values found by this trend line are shown in the following table.

Year	Quarters			
	1	2	3	4
2001	30.13	33.97	37.81	41.65
2002	45.49	49.33	53.17	57.01
2003	60.85	64.69	68.53	72.37

Dividing each of the actual values by the corresponding trend value and expressing the result as a percentage, we get

Year	Quarters				
	1	2	3	4	
2001	139.40	132.47	87.28	86.43	
2002	65.95	91.22	75.23	114.02	
2003	83.81	106.66	94.85	129.89	Total
Total	289.16	330.35	257.36	330.34	1207.21
Mean	96.39	110.12	85.79	110.11	402.41
Seasonal Indices	95.81	109.46	85.28	109.45	400.03

To forecast the sales for each quarter of 2004, we first get the projected trend values for these quarters, by means of the fitted least-squares equation. The projected values are then multiplied by the corresponding seasonal indices and the product divided by 100. The desired forecasts are given below:

Quarter of 2004	Projected Trend Value	Seasonal Index	Forecast
I	76.21	95.81	73
II	80.05	109.46	88
III	83.89	85.28	72
IV	87.73	109.45	96

13.38. **(b) Computations of the cyclical irregular movements and cyclical relatives are shown in the table below. To remove the irregular variations, a three-quarter moving average has been thought appropriate.**

Year & quarter (1)	Y values *TSCI* (2)	Trend values *T* (3)	Seasonal Index (S)% (4)	Trend · Seasonal TS ÷ 100 (5)	Cyclical Irregular %s CI (%) (6)	3-quarter moving total (7)	Cyclical relative C (%) (8)
1961-I	122	127.78	101.5	129.70	94.06	..	..
II	125	124.55	98.6	122.81	101.78	293.10	97.70
III	118	121.32	100	121.32	97.26	298.12	99.37
IV	117	118.09	100	118.09	99.08	298.42	99.47
1962-I	119	114.86	101.5	116.58	102.08	304.73	101.58
II	114	111.63	98.6	110.07	103.57	310.82	103.61
III	114	108.40	100	108.40	105.17	312.38	104.13
IV	109	105.17	100	105.17	103.64	310.27	103.42
1963-I	105	101.96	101.5	103.49	101.46	306.82	102.27
II	99	98.71	98.6	97.33	101.72	300.58	100.19
III	93	95.48	100	95.48	97.40	295.60	98.53
IV	89	92.25	100	92.25	96.48	299.05	96.35
1964-I	86	89.02	101.5	90.36	95.17	286.22	95.41
II	80	85.79	98.6	84.59	94.57	290.27	96.76
III	83	82.56	100	82.56	100.53	300.99	100.33
IV	84	79.33	100	79.33	105.89	...	...

13.39. (a) Let the equation of the linear trend by $y_t = a + bX$. Since the number of months in the observed series is even, therefore the middle point of the two middle months is taken as $X = 0$.

Year and months		Y	X	XY	X^2	Trend Values $Y = 2.9688 + 0.0113X$
2000	Jan	6.2	−47	−291.4	2209	2.4377
	Feb	1.8	−45	−81	2025	2.4603
	Mar	0.9	−43	−38.7	1849	2.4829
	Apr	1.4	−41	−57.4	1681	2.5055
	May	3.2	−39	−124.8	1521	2.5281
	Jun	2.3	−37	−85.1	1369	2.5507
	Jul	2.2	−35	−77	1225	2.5733
	Aug	3.2	−33	−105.6	1089	2.5959
	Sep	3.4	−31	−105.4	961	2.6185
	Oct	3.4	−29	−98.6	841	2.6411
	Nov	2.7	−27	−72.9	729	2.6637
	Dec	2.1	−25	−52.5	625	2.6863
2001	Jan	3.3	−23	−75.9	529	2.7089
	Feb	1.7	−21	−35.7	441	2.7315
	Mar	0.5	−19	−9.5	361	2.7541
	Apr	2.2	−17	−37.4	289	2.7767
	May	1.5	−15	−22.5	225	2.7993
	Jun	2.5	−13	−32.5	169	2.8219
	Jul	2.8	−11	−30.8	121	2.8445
	Aug	3.2	−9	−28.8	81	2.8671
	Sep	4.2	−7	−29.4	49	2.8897
	Oct	4.5	−5	−22.5	25	2.9123
	Nov	6.1	−3	−18.3	9	2.9349
	Dec	2.8	−1	−2.8	1	2.9575
2002	Jan	3.4	1	3.4	1	2.9801
	Feb	3.1	3	9.3	9	3.0027
	Mar	1.3	5	6.5	25	3.0253
	Apr	1.7	7	11.9	49	3.0479
	May	3.2	9	28.8	81	3.0705
	Jun	3.2	11	35.2	121	3.0931
	Jul	2.6	13	33.8	169	3.1157
	Aug	2.8	15	42	225	3.1383

Year and months		Y	X	XY	X^2	Trend Values $\hat{Y} = 2.9688 + 0.0113X$
	Sep	2.5	17	42.5	289	3.1609
	Oct	4.1	19	77.9	361	3.1835
	Nov	0.8	21	16.8	441	3.2061
	Dec	4.1	23	94.3	529	3.2287
2003	Jan	3.4	25	85	625	3.2513
	Feb	3.4	27	91.8	729	3.2739
	Mar	1.5	29	43.5	841	3.2965
	Apr	1.8	31	55.8	961	3.3191
	May	3.0	33	99	1089	3.3417
	Jun	3.4	35	119	1225	3.3643
	Jul	3.1	37	114.7	1369	3.3869
	Aug	5.4	39	210.6	1521	3.4095
	Sep	4.9	41	200.9	1681	3.4321
	Oct	1.5	43	64.5	1849	3.4547
	Nov	6.2	45	279	2025	3.4773
	Dec	4.0	47	188	2209	3.4999
	Σ	142.5	0.0	417.7	36848.0	142.5

Since $\Sigma X = 0$, the normal equations take the form

$$\Sigma Y = na \text{ and } \Sigma XY = b\Sigma X^2$$

Substituting the values, we obtain

$$a = \frac{\Sigma Y}{n} = \frac{142.5}{48} = 2.9688$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = 0.0113$$

Thus, the trend line is $\hat{Y}_t = 2.9688 + 0.0113$, with the origin at the middle of Dec 2001 and Jan 2002 and X is measured in half month units. The trend values are computed by substituting the values of X in the least square equation and they appear in the last column of the above table.

13.40. The calculations needed to compute the first serial correlation co-efficient r_1 are shown below:

Y_t	Y_{t+1}	$Y_t-\bar{Y}$	$Y_{t+1}-\bar{Y}$	$(Y_t-\bar{Y})(Y_{t+1}-\bar{Y})$	$(Y_t-\bar{Y})^2$
65	64	1.55	0.55	0.8525	2.4025
64	63	0.55	−0.45	−0.2475	0.3025
63	61	−0.45	−2.45	1.1025	0.2025
61	60	−2.45	−3.45	8.4525	6.0025
60	58	−3.45	−5.45	18.8025	11.9025
58	63	−5.45	−0.45	2.4525	29.7025
63	64	−0.45	0.55	−0.2475	0.2025
64	62	0.55	−1.45	−0.7975	0.3025
62	64	−1.45	0.55	−0.7975	2.1025
64	63	0.55	−0.45	−0.2475	0.3025
63	63	−0.45	−0.45	0.2025	0.2025
63	62	−0.45	−1.45	0.6525	0.2025
62	60	−1.45	−3.45	5.0025	2.1025
60	62	−3.45	−1.45	5.0025	11.9025
62	64	−1.45	0.55	−0.7975	2.1025
64	66	0.55	2.55	1.4025	0.3025
66	68	2.55	4.55	11.6025	6.5025
68	68	4.55	4.55	20.7025	20.7025
68	69	4.55	5.55	25.2525	20.7025
69	--	5.55	--	−3.135 + 75.5275	30.8025
1269	--	0	--	72.3925	148.9500

Now $\bar{Y} = \sum_{t=1}^{n} \frac{Y_t}{n} = \frac{1269}{20} = 63.45$, and

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t+1} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} = \frac{72.3925}{148.9500} = 0.49$$

Calculation of the coefficient of auto-correlation of lag 2

Y_t	Y_{t+2}	$Y_t-\bar{Y}$	$Y_{t+2}-\bar{Y}$	$(Y_t-\bar{Y})(Y_{t+2}-\bar{Y})$
65	63	1.55	−0.45	−0.6975
64	61	0.55	−2.45	−1.3475
63	60	−0.45	−3.45	1.5525
61	58	−2.45	−5.45	13.3525
60	63	−3.45	−0.45	1.5525
58	64	−5.45	0.55	−2.9975
63	62	−0.45	−1.45	0.6525
64	64	0.55	0.55	0.3025
62	63	−1.45	−0.45	0.6525
64	63	0.55	−0.45	−0.2475
63	62	−0.45	−1.45	0.6525
63	60	−0.45	−3.45	1.5525
62	62	−1.45	−1.45	2.1025
60	64	−3.45	0.55	−1.8975
62	66	−1.45	2.55	−3.6975
64	68	0.55	4.55	2.5025
66	68	2.55	4.55	11.6025
68	69	4.55	5.55	25.2525
68	--	4.55	--	+61.7300
69	--	5.55	--	−10.8850
1269	--	0	--	+50.8450

$$r_2 = \frac{\sum_{t=1}^{n-2} (Y_t - \bar{Y})(Y_{t+2} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} = \frac{50.845}{148.950} = +0.34$$

●●●●●●●●●●