Paper supplied by The Titaghur Paper Mills Company Ltd. on Control Rate

S. CHAND & CO.

ASAF ALI ROAD	—	New Delhi
FOUNTAIN	—	Delhi
MAI HIRAN GATE	—	Jullundur City
LAL BAGH	—	Lucknow

Price : Rs. 12·50

Published by G. S. Sharma, for S. Chand & Co., Fountain, Delhi and Printed by S. P. Dhawan at The Central Electric Press, 80-D, Kamla Nagar, Delhi-6.

PREFACE

This work covers the practical problems in Statistics for B.Sc., B. Com., M. Com., M. A. (Eco.), M. Sc., P. C. S., I. A. S. and other competitive examinations. Problems in general, have been selected from Papers of different Universities, competitive examinations and from authoritative text books. These have been fully solved. In solving these problems, we have tried to use as many different methods as possible.

We do not claim absolute originality, as we have taken extensive help and matter from different sources. We have tried our best to quote the sources for such material, as far as possible. We thank all the authors, publishers and copyright holders of the books from which such matter has been taken.

We are highly thankful to Dr. M. Ray, D. Sc., F. N. I., Dr. Balak Ram, Ph. D., Dr. J. P. Agarwal, Ph. D., Prof. H. S. Sharma, M. A., Dr. Raghu Raj Singh, Ph. D., Prof. S. S. Seth, M. Sc., and Prof. D. N. Agarwal, M. Sc., for giving us encouragement and valuable suggestions.

We are also thankful to Sri. Shyam Lal ji and Sri. Raj Kumar Seth of M/s. S. Chand & Co., Delhi, for bringing out the publication in time. Our thanks are also due to Messrs Central Electric Press, Delhi, for printing the book in very short time.

We shall regard our labour well rewarded if this book satisfies those for whom it has been intended. Suggestions for improvements will be thankfully acknowledged. Omissions and printing mistakes, if any, are regretted.

AGRA COLLEGE, AGRA — AUTHORS

1st January, 1959

CHAPTER I

"AVERAGES"

An average is defined as 'a single expression representative of the whole distribution'; or, it may be defined as, "an average is a single simple expression in which the net result of a complex group of large numbers is concentrated". Important forms of averages are the Mean, Mode, Median, Quartiles, Deciles, Percentiles etc.

[1] MEAN.—represented by (M)

(*a*) *Individual Series.*—(*i*) Main Method $(M)=\frac{\Sigma x}{n}$,

where x=size of items

n=no. of items.

(*ii*) Short-cut-Method.—An assumed mean (A) is taken and from it the deviations $(x-A)$ are calculated with proper algebrical signs. Then it is summed up and Mean calculated. $M=A+\frac{\Sigma\xi}{n}$

$\xi=(x-A)$.

Problem 1.—Compute the arithmetic mean of the following:

Serial No.:	1	2	3	4	5	6	7	8	9	10
Marks:	20	25	15	10	10	30	35	20	40	45

Solution:

Serial No.	Marks (x)	$\xi=(x-A)$
1	20	−10
2	25	−5
3	15	−15
4	10	−20
5	10	−20
6	30	0
7	35	+5
8	20	−10
9	40	+10
10	45	+15
$n=10$	$\Sigma x=250$	$\Sigma\xi=-50$

Let the assumed mean (A)=30

M by main Method $=\frac{\Sigma x}{n}$

$=\frac{250}{10}$

$=25$ marks

M by Short-cut-Method

$M=A+\frac{\Sigma\xi}{n}$

$=30+\left[\frac{-50}{10}\right]$

$=30-5$

$=25$ marks.

M represents Mean.

(*b*) *Discrete Series.*—(*i*) Main Method :—$\frac{\Sigma xf}{n}$ where f=frequency.

[NOTE :—Deviation $(x-A)$ can also be shown by other letters other than ξ ; *e.g.*, d_a etc.]

(*d*) There is yet another method of computation of the Mean—Short-cut with class-interval units. The only difference between Short-cut-Method for continuous series and this method is that instead of using actual deviations of the mid-points from the guessed mean, the deviations are written in terms of class-intervals.

Problem 4.—Quoted from *Elementary Statistical Methods*—

Hourly Earning	Mid-Point Cents	Number of Workers	M′ (ass. mean)=55	
			Deviation	fd′
	a	b	c	d
27·5 and under 32·5 Cents.	30	120	− 5	− 600
32·5 ,, ,, 37.5 ,,	35	152	− 4	− 608
37·5 ,, ,, 42·5 ,,	40	170	− 3	− 510
42·5 ,, ,, 47·5 ,,	45	214	− 2	− 428
47·5 ,, ,, 52·5 ,,	50	410	− 1	− 410
52·5 ,, ,, 57·5 ,,	55	429	0	0
57·5 ,, ,, 62·5 ,,	60	568	+ 1	+ 568
62·5 ,, ,, 67·5 ,,	65	650	+ 2	+1300
67·5 ,, ,, 72·5 ,,	70	795	+ 3	+2385
72·5 ,, ,, 77·5 ,,	75	915	+ 4	+3660
77·5 ,, ,, 82·5 ,,	80	745	+ 5	+3725
82·5 ,, ,, 87·5 ,,	85	530	+ 6	+3180
87·5 ,, ,, 92·5 ,,	90	259	+ 7	+1813
92·5 ,, ,, 97·5 ,,	95	152	+ 8	+1216
97·5 ,, ,, 102·5 ,,	100	107	+ 9	+ 963
102·5 ,, ,, 107·5 ,,	105	50	+10	+ 500
107·5 ,, ,, 112·5 ,,	110	25	+11	+ 275
		N=6291		+ 17,029

Neiswanger, p. 260. Source of original table ; as quoted in the above referred book : "Earnings in the Manufacture of Rubber Products, May, 1940". Monthly Labour Review, U.S. Dept. of Labor, Washington, D.C., June, 1941 p. 1503.

$$C=\frac{\Sigma\,(fd')}{N}\cdot i \quad = \frac{+17029}{6291}.05 \quad = \cdot 1353$$

$$M=M'+C=0{\cdot}55+{\cdot}1353=\$0{\cdot}6853 \text{ or } 68{\cdot}5 \text{ Cents.}$$

(*e*) *Weighted Mean* (M_w).—For the Calculation of weighted mean, we follow the same process as that in Discrete and Continuous Series. The only difference being (f) replaced by (w) where w=weights.

(*i*) $n=\Sigma w$ in place of Σf.

(*ii*) xw or $w\xi$ in place of

xf or $f\xi$

Otherwise, there is no difference.

[2] Mode (M_0).—"Mode is the value of that item in a variable which occurs most frequently."

(a) *Individual Series.*—Mode is located by inspection in the individual series. That item which repeats the maximum number of times is said to be modal.

Problem 5.—Size of the item Frequency.

Size of the item	Frequency	
1	5	
2	10	Mode (M_0) is located
3	25	at 3 with 25 as max.
4	14	frequency.

(b) *Discrete Series.*—In discrete series, Mode (M_0) is located by grouping. The process is : in all, we make six columns, including frequency column.

(*i*) In the first Column (freq. col.) mode is located by inspection item containing maximum frequency.

(*ii*) In the second column, we group frequencies in two's, starting from the top. Their totals are found and items containing maximum frequency located.

(*iii*) In the third column we do like Col. (*ii*) except that we start grouping by leaving the first frequency.

(*iv*) We group, in this column, frequencies in three's—starting from the first frequency. Totals calculated and the items containing maximum frequency located.

(*v*) Same as in Col. (*iv*), starting grouping in three's from the second frequency.

(*vi*) Same as in Col. (*iv*), starting grouping in three's from the third frequency. Finally, we find out the size of item which repeats largest number of times. On this item, Mode is located

Problem 6.—Compute the Mode.

Size of item :	2	3	4	5	6	7	8	9	10	11	12	13
Frequency :	3	8	10	12	16	14	10	8	17	5	4	1

(*Nagpur, B. Com.*, 1945)

Solution :

Size of item	Frequency (I)	(II)	(III)	(IV)	(V)	(VI)	No. of times cont. max. freq.
2	3	11		21			×
3	8		18		30		×
4	10	22				38	1
5	12		28	42			3
6	16	30			40		5
7	14		24			32	3
8	10	18		35			1
9	8		25		30		×
10	**17**	22				26	1
11	5		9	10			×
12	4	5					×
13	1						×

∴ Mode is located at 6 (frequency $=16$).

(c) *Continuous series.*—Mode in the continuous series is computed by the formula :

$$(i) \quad M_0 = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} (l_2 - l_1)$$

where M_0 = Mode.

l_1 and l_2 = lower and upper limit of the modal group.

f_1 = frequency of the modal group.

f_0 = frequency of the next lower group than the modal group.

f_2 = frequency of the next higher group than the modal group.

Another formula for computing Mode of grouped data is

$$(ii) \quad M_0 = l_1 + \frac{f_2}{f_2 - f_0} (l_2 - l_1)$$

$$(iii) \quad M_0 = L + \frac{N_2}{N_1 + N_2} \times I$$

$$(iv) \quad M_0 = 3Md - 2M$$

where L = lower limit of the class-interval.

N_2 = frequency of the next higher group.

N_1 = frequency of the next lower group.

I = Magnitude of the class-interval.

Out of these formulas, however, the first is regarded as the best, and in general, followed.

Problem 7.—

Marks	No. of students.
2—4	20
4—6	40
6—8	30
8—10	10

(*Practical Statistics*—Ziaudin, p. 31)

Solution :

By (i) formula.

$$M_0 = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} (l_2 - l_1).$$

$$= 4 + \frac{40 - 20}{80 - 20 - 30} (6 - 4).$$

$$= 4 + \frac{20}{30} \times 2$$

$$= 4 + 1{\cdot}33$$

$$= 5{\cdot}33 \text{ approx.}$$

by (*ii* and *iii*) methods $M_0 = L + \frac{N_2}{N_1+N_2} \times I$

$$= 4 + \frac{30}{20+30} \times 2$$

$$= 5{\cdot}2 \text{ approx.}$$

by (*iv*) method. $M_0 = 3Md - 2M$

$$= \left(3 \times \frac{11}{2}\right) - \left(2 \times \frac{28}{5}\right)$$

$$= \frac{33}{2} - \frac{56}{5}$$

$$= 5{\cdot}3 \text{ approx.}$$

Md being $\frac{11}{2}$ and M being $5\frac{3}{5}$.

[3] MEDIAN.—"Median is the value of that item in a series which divides the series into two equal parts, one part consisting of all values less, and the other all values greater than it." (*Ghosh and Chaudhry—Statistics*, p. 99)

(*a*) *Individual Series.*—For calculating median of an individual series, it is essential that the series is placed in an arranged order. It may either be ascending, descending or ascending-cum-descending order. However, it is always safe and better to array it in ascending or descending order. Then, *Md* is calculated as :

$$Md = \text{Size of the } \left(\frac{n+1}{2}\right)^{th} \text{ item.}$$

Sometimes it so happens that the value of the median comes in the fraction, say it is equal to 5·5th item. Then it is located as :

$$\frac{\text{Size of the 5th item} + \text{size of the 6th item}}{2}$$

i.e., average is taken of the two values—the Vth and the VIth values. This gives the value of the Median.

(*b*) *Discrete Series.*—(*i*) There is no need of putting it in any arrayed form. The reason is, in this case, we have to find out the cumulative frequency—which automatically places the series in an ascending order.

(*ii*) Cumulative frequency is known.

(*iii*) Value of $\left(\frac{n+1}{2}\right)^{th}$ item is known.

(*iv*) *Md* is located at the size of the item in whose cumulative frequency this value of $\left(\frac{n+1}{2}\right)^{th}$ item falls.

If it is not just equal to any value of the *cmf*, then the *cmf* next higher to the value of $\left(\frac{n+1}{2}\right)^{th}$ item should be considered.

(*c*) *Continuous Series.*—(*i*) Cumulative frequency is calculated.

(*ii*) *m* is calculated as the size of $\left(\frac{n+1}{2}\right)^{th}$ item ; following the method mentioned in (*b*) above.

(*iii*) On the basis of the value of *m* the median group is located.

(*iv*) Finally, median is calculated following this formula of interpolating median in a grouped data :

$$(a)\quad Md = l_1 + \left\{\frac{l_2 - l_1}{f_1}(m - c)\right\}$$

where, Md = Median

l_1 and l_2 = lower and upper limit of the median group,

f_1 = frequency of the median group.

m = size of $\left(\frac{n+1}{2}\right)^{th}$ item.

c = *cmf* of the next lower group than the median group.

$$(b)\quad Md = l + \frac{i}{f}\left(\frac{n}{2} - c\right)$$

where n = total frequency.

l = lower limit of the median group.

i = class interval.

c = cumulative frequency of the group preceding the median group.

$$(c)\quad Md = L + \frac{\frac{N}{2} - F}{f} \times i$$

where L = lower limit of the median group.

f = frequency of the median group.

i = class interval.

F = total of all frequencies before median group.

N = Total of frequencies.

There are a few more formulas for the computation of Md in grouped data, but the first one (*a*) is regarded as the simplest and the best.

[NOTE.—Many writers calculate the value of Md in individual and discrete series as size of $\frac{n}{2}^{th}$ item and not $\left(\frac{n+1}{2}\right)^{th}$ item.]

[4] QUARTILES, DECILES, PERCENTILES ETC.

(a) They are also calculated as the Median. Difference is for the quartiles, $(n+1)$ is divided by 4 in place of 2 ; for Deciles it is divided by 10 ; for Percentiles it is divided by 100.

(b) In Individual series, the series is to be arrayed.

(c) In Discrete series, the cumulative frequency is to be calculated.

(d) The value given is used for multiplication to the value of $\left(\frac{n+1}{x}\right)$ where $x=4$ for quartiles, 10 for deciles and so on.

e.g. (i) We have to compute Q_3. Now 3 is the power

$$=\left(\frac{3(n+1)}{4}\right).$$

(ii) $D_7 \quad = \frac{7(n+1)}{10}$.

(iii) $O_6 \quad = \frac{6(n+1)}{8}$

(iv) $P_{46} \quad = \frac{46(n+1)}{100}$

and so on.

(e) for continuous series, find the value of that particular item, as required, say, $d_7 = \frac{7(n+1)}{10}$; $o_6 = \frac{6(n+1)}{8}$; $p_{46} = \frac{46(n+1)}{100}$

(f) Then interpolate as : (using first formula)

$$P_{47} = l_1 + \left\{ \frac{l_2 - l_1}{f_1} (p_{47} - c) \right\}$$

where p_{47} = size of the $\frac{47(n+1)}{100}^{th}$ item.

$$O_6 = l_1 + \left\{ \frac{l_2 - l_1}{f_1} \left(o_6 - c \right) \right\}$$

$$Q_3 = l_1 + \left\{ \frac{l_2 - l_1}{f_1} (q_3 - c) \right\}$$

that is, in the small bracket $(m-c)$, we replace m by the particular unit whose value is to be known say q_3 for finding out Q_3.

[5] GEOMETRIC MEAN.—(Quoted – slightly changed—from *Elementary Statistical Methods* by Neiswanger, pp. 287—289).

The geometric mean (g) is the nth root of the product of n numbers. The formula in natural numbers is :

$$g = \sqrt[n]{a_1 \cdot a_2 \cdot a_3 \ldots a_n}$$

If, for example, only two values, such as 4 and 9, are used, the geometric average is the square root of the product of these numbers: $4 \times 9 = 36$ and the square root of $36 = 6$.

When a geometric average of a series of numbers is decided, the computation is best undertaken through the use of logarithms. The formula then becomes : $g = \text{Antilog}\ \frac{\log a_1 + \log a_2 + \log a_3 + \ldots \log a_n}{n}$

OR $$\log g = \frac{\log a_1 + \log a_2 + \log a_3 + \ldots\ldots \log a_n}{n}$$

where a_1, a_2, a_3 are the digits of whose g is to be calculated.
n = total No. of items in the series.

To compute the geometric mean (g) of 4 and 9, the logarithms of 4 and 9 (·60206 and ·95424) are secured from log. table. These logs are then added and divided by 2 (here n being = 2).

$$\log g = \frac{\cdot 60206 + \cdot 95424}{2}$$

$$\log g = \cdot 77815$$

$g = 6$ where g = geometric mean.

When the data are in frequency-distribution form, the formula for the geometric mean in logarithmic terms becomes :

$$\log g = \frac{f_1 \log a_1 + f_2 \log a_2 + f_3 \log a_3 + \ldots f_n \log a_n}{\Sigma f}$$

$a_1, a_2, a_3 \ldots a_n$ represent the mid-points of classifications ; and $f_1, f_2, f_3 \ldots f_n$ symbolize the class frequencies associated with each mid-point. $\Sigma f = n$.

Problem 8. — (This alone from *Practical Statistics* by Ziauddin. p. 34).

Class-group	Mid-point (x)	Frequency (f)	log of (x)	$f \log x$
2—4	3	20	·4772	9·542
4—6	5	40	·6990	27·66
6—8	7	30	·8451	25·353
8—10	9	10	·9542	9·542
		$\Sigma f = n = 100$		$\Sigma f \log x = 72{\cdot}397$

x = mid-points. $n = \Sigma f$

$$\log g = \frac{\Sigma(f \log x)}{n}$$

$$= \frac{72 \cdot 739}{100} = \cdot 72397$$

Therefore, $g = 5 \cdot 279$ where g = geometric mean.

In case we have to compute weighted geometric mean (g'), we multiply the digits by weights (w) in place of (f)—other things as mentioned above.

[**6**] HARMONIC MEAN—The harmonic mean is the reciprocal of the arithmetic average of the reciprocals of the values ; symbolically :

$$h = \frac{n}{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \ldots\ldots \frac{1}{n}}$$

where $a, b, c, \ldots n$ represent the values of the items. n = total no. of items, h = harmonic mean.

It can also be expressed as :

$$h = \text{reciprocal } \frac{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \ldots\ldots \frac{1}{n}}{n}$$

$$\text{OR} \quad h = \frac{n}{\Sigma f. \frac{1}{x}} \quad \text{OR} \quad \frac{1}{h} = \frac{\Sigma f. \frac{1}{x}}{\Sigma f}$$

where x = size of the items ; f = frequency.

✓**Problem 9.**—(from *Practical Statistics*—Ziauddin, pp. 34—35).

(*a*) Calculate the harmonic mean of ·5 and ·25.

(*i*) $h = \frac{2}{\frac{1}{\cdot 5} + \frac{1}{\cdot 25}} = \frac{2}{2+4} = \cdot 33.$ [$n = 2.$]

(*ii*) $\frac{1}{h} = \frac{\frac{1}{\cdot 5} + \frac{1}{\cdot 25}}{2} = \frac{2+4}{2}$ or $\frac{1}{h} = \frac{6}{2}$ $\therefore h = \frac{2}{6}$

= ·33 approximately.

(*b*)

Class group	Mid value (x)	Reciprocals $\left(\frac{1}{x}\right)$	Frequency f	$f \cdot \frac{1}{x}$
2— 4	3	$\frac{1}{3}$ = ·333	20	6·66
4— 6	5	$\frac{1}{5}$ = ·2	40	8·00
6— 8	7	$\frac{1}{7}$ = ·14	30	4·29
8—10	9	$\frac{1}{9}$ = ·111	10	1·11
			$\Sigma f = 100$	$\Sigma f \cdot \frac{1}{x} = 20{\cdot}06$

(*i*) harmonic mean $(h) = \frac{100}{20{\cdot}06} = 4{\cdot}98$

$$\left(h = \frac{\Sigma f = n}{\Sigma f \cdot \frac{1}{x}} \right)$$

(*ii*) harmonic mean $(h) = \frac{1}{h} = \frac{\Sigma f \cdot \frac{1}{x}}{n} = \frac{20{\cdot}06}{100}$

$$\therefore h = \frac{100}{20{\cdot}06} = 4{\cdot}98.$$

weighted harmonic mean is calculated by multiplying the size of the items by weights (w) in place of (f) - other procedure is common.

[7] Quadratic Mean, Moving Average, Progressive Average, Composite Average and Death Rates. [Quoted from *Elements of Statistics* by Profs. D. K. Sakhwalkar and M.P. Singh, pp. 149—152.]

Quadratic Mean.—Quadratic mean is an arithmetic average obtained by extracting the root of the sum of square of item values comprising a series. It is also known as "Root Mean Square" and calculated by the formula :

$$Q_m = \sqrt{\frac{a^2 + b^2 + c^2 + \ldots\ldots n^2}{n}}$$

Q_m = Quadratic mean.

$a^2, b^2, c^2, \ldots n^2$ = square of various item values in a series.

n = number of items.

Moving Average.—It is a form of arithmetic average computed to obtain a new series, by dropping off the earliest item-value and taking in its place the succeeding trading period and repeating the same process year after year. It may be on 3 yearly, 5 yearly, 7 yearly etc. basis. Symbolically—

(*i*) 3 yearly— $\frac{a+b+c}{(n=3)}$, $\frac{b+c+d}{3}$, $\frac{c+d+e}{3}$, $\frac{d+e+f}{3}$ ... etc.

(*ii*) 5 yearly— $\frac{a+b+c+d+e}{(n=5)}$, $\frac{b+c+d+e+f}{5}$, $\frac{c+d+e+f+g}{5}$ ... etc.

PROGRESSIVE AVERAGE.— A progressive average is a cumulative average used occasionally during the 'early years of the life of a business.' This is computed by taking all the figures available in each succeeding year. Thus, the average for the different years will be :

$$\frac{a+b}{2}\,;\ \frac{a+b+c}{3}\,;\ \frac{a+b+c+d}{4}$$ ... etc. where a, b, c, d, etc. represent the amounts of trading profits in different years.

COMPOSITE AVERAGE.—A composite average is an arithmetic mean computed by taking out an average of various averages. Thus, if the monthly sales of a certain business house are represented by $x_1, x_2, x_3 \ldots x_{12}$ the average monthly sales for that year would be

$$\frac{x_1+x_2+x_3\ldots+x_{12}}{12}$$

and, if monthly sales of different years are available, monthly average sale for various years will be

$$\frac{a+b+c\ldots+n}{n}$$ [n = no. of items.]

where $a, b, c, \ldots n$ represent yearly averages of monthly sales for different years. This is known as a composite average.

DEATH RATES.—

Problem 10.—

Age group	Local Population (Town A)			Standard Population (Town B)		
(Years)	Population	Death	Death Rate per 1000	Population	Death	Death Rate per 1000
Under 10	20,000	600	30	12,000	372	31
10—20	12,000	240	20	30,000	660	22
20—40	50,000	1250	25	62,000	1,612	26
40—60	30,000	1050	35	15,000	525	35
Above 60	10,000	500	50	3,000	180	60
	1,22,000	3,640	29·9	1,22,000	3,349	26·6

General Death Rate of Town A—

$$=\frac{1}{122000}[(20000\times 30)+(12000\times 20)+(50000\times 25)+(30000\times 35)+(10000\times 50)]$$

$=29{\cdot}9$

General Death Rate of Town B—

$$\frac{1}{122000}[(12000\times 31)+(30000\times 22)+(62000\times 26)+(15000\times 35)+(3000\times 60)]$$

$=26{\cdot}6.$

Standardized Death Rate of Town A—

[Standardized death rate is found by taking the population of one town, say, town B, as standard and the death rates of another town, say, Town A local.]

$$=\frac{1}{122000}[(12000\times 30)+(30000\times 20)+(62000\times 25)+(15000\times 35)+(3000\times 60)]$$

$=26{\cdot}1.$

Problem 11.—According to the census of 1941, following are the populationfigures, in thousands, of first 10 cities: Find the Median and quartiles.

2488, 1490, 777, 733, 522, 672, 591, 407, 387, 391.

Solution

For finding out the median, the series is re-distributed in either ascending or descending order. Here, it is placed in the descending order.

Cities	*Population in thousand*	*Cities*	*Population in thousand*
1	2488	6	591
2	1490	7	522
3	777	8	407
4	733	9	391
5	672	10	387

$n=10$ $\quad$ $\text{Md}=\frac{(n+1)}{2}^{th}$ item. where Md=median

n=number of items.

$$=\left(\frac{10+1}{2}\right)^{th} \text{ item,}$$

$$=\frac{11}{2}=5{\cdot}5^{th} \text{ item.}$$

$$=\frac{5th \text{ item}+6^{th} \text{ item}}{2}$$

$$=\frac{672+591}{2}$$

$$=\frac{1263}{2}=631{\cdot}5 \text{ thousands}$$

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item,} \qquad \text{where } Q_1 = \text{First quartile.}$$

$$= \frac{11}{4}^{th} \text{ item,}$$

$$= 777 \text{ thousands approximately.}$$

$$\left(\frac{11}{4} \text{ is more near to 3 than 2}\right):$$

$$Q_3 = \frac{3(n+1)}{4}^{th} \text{ item.}$$

$$= \frac{33}{4}^{th} \text{ item} = 8^{th} \text{ item approximately.}$$

$$= 407 \text{ thousands approximately.}$$

Problem 12.—The following table gives the marks obtained by a batch of 25 students in a certain class-test in Economics and politics.

Roll No. of the students	*Economics*	*Politics*	*Roll No. of the students*	*Economics*	*Politics*
1	29	36	12	35	33
2	65	30	13	46	80
3	33	38	14	47	44
4	45	39	15	60	85
5	51	64	16	30	20
6	72	50	17	32	32
7	48	46	18	52	25
8	33	15	19	54	55
9	42	42	20	56	28
10	25	10	21	58	53
11	28	72	22	49	35
			23	38	40
			24	40	62
			25	46	58

In which subject is the level of knowledge of the students, as revealed from the above figures, higher ? Give reasons.

(*M.A., Allahabad, 1937*)

Solution :

To find the level of knowledge and to ascertain as to in which subject the knowledge of students is higher, we have o compute the median of both the subjects and on the basis of that the result would be known. For finding out the median, the series is to be re-distributed either in the ascending or descending order. Here, it is arrayed in the descending order.

Roll No. of students	*Economics*	*Politics*	*Roll No. of Students*	*Economics*	*Politics*
1	72	85	12	46	42
2	65	80	13	46	40
3	60	72	14	45	39
4	58	64	15	42	38
5	56	62	16	40	36
6	54	58	17	38	35
7	52	55	18	35	33
8	51	53	19	33	32
9	49	50	20	33	30
10	48	46	21	32	28
11	47	44	22	30	25
			23	29	20
			24	28	15
			25	25	10

Median for Economics marks :

$Md_{(e)}=\left(\frac{n+1}{2}\right)^{th}$ item. $Md_{(e)}$=Median for Economics Marks.

n=number of items.

$=\left(\frac{25+1}{2}\right)^{th}$ item$=13^{th}$ item.

=46 marks.

Median for Politics marks : $Md_{(p)}=\left(\frac{n+1}{2}\right)^{th}$ item.

$Md_{(p)}$=Median for Politics marks.

$=\left(\frac{25+1}{2}\right)^{th.}$ item.

=13*th.* item.

=40 marks.

Thus, we arrive at a conclusion that standard of knowledge in Economics is higher.

Problem 13.—Find the Mode of the following series :

Size	*Frequency*	*Size*	*Frequency*
5	48	13	52
6	52	14	41
7	56	15	57
8	60	16	63
9	63	17	52
10	57	18	48
11	55	19	40
12	50		

(Allahabad, B. Com. 1943)

Solution :

Mode of the series is to be located by the grouping method.

Size	Frequency (I)	(II)	(III)	(IV)	(V)	(VI)
5	48	100		156		
6	52		108		168	
7	56	116				**179**
8	60		**123**	**180**		
9	**63**	**120**			**175**	
10	57		112			162
11	55	105		157		
12	50		102		143	
13	52	93				150
14	41		98	161		
15	57	**120**			172	
16	**63**		115			163
17	52	100		140		
18	48		88			
19	40					

Columns	Size of items containing maximum frequency								
	(7)	(8)	(9)	(10)	(11)	15)	(16)	(17)	(18)
1			9				16	—	—
2			9	10		15	16	—	—
3		8	9					—	—
4		8	9	10				—	—
5			9	10	11			—	—
6	7	8	9					—	—
Number of times	1	3	6	3	1	1	2		

Mode is located at 9, *i.e.* consisting of 63 frequency.

Problem 14.—Compute the weighted geometric average of relative prices of the following commodities for the year 1939 (base year 1938-Price 100) :

Commodity	*Relative Price*	*Weight* (Value produced in 1938)
1. Corn	128·8	1385
2. Cotton	62·4	819
3. Hay	117·7	842
4. Wheat	99·0	561

Commodity	*Relative Price*	*Weight (Value produced in 1938)*
Oats	130·9	408
Potatoes	143·5	194
Sugar	125·6	142
Barley	150·2	100
Tobacco	101·1	103
Rye	116·2	25
Rice	117·5	17
Oil Seeds	78·7	29

(*B.Com., Alld., 1943*)

Solution :

Commodity	Relative Price	Log. for Relative Price	Weight	Weight x Log. for R. Price
Corn	128·8	2·4594=2·5	1385	3462·5
Cotton	62·4	1·7952=1·8	819	1474·2
Hay	117·7	2·2480=2·5	842	2105·0
Wheat	99·0	1·9956=2·0	561	1122·0
Oats	130·9	2·4900=2·5	408	1020·0
Potatoes	143·5	2·6385=2·6	194	504·4
Sugar	125·6	2·4082=2·4	142	340·8
Barley	150·2	2·7007=2·7	100	270·0
Tobacco	101·1	2·0043=2·0	103	206·0
Rye	116·2	2·2095=2·2	25	55·0
Rice	117·5	2·2430=2·3	17	39·1
Oil Seed	78.7	1·8960=1·9	29	55·1
			$\Sigma W=4625$	$\Sigma Wl=10654·1$

Where ΣW=summation of weights

ΣWl= „ „ weight *x* log of relative price.

$$g'=\text{antilog } \frac{10654·1}{4625}$$

$$=\text{antilog } 2·3 \text{ approximately.}$$

$$=169·8 \text{ approximately.}$$

where g'=weighted geometric mean.

Problem 15.—The following table gives the results of certain examinations of three universities in the year 1936. Which is the best university ?

University Examination	Percentage Results in the University		
	A	B	C
M.A.	80	75	70
M.Sc.	70	70	60
B.A.	65	80	70
B.Sc.	60	70	80
B.Com.	75	65	75

(*Cal. M.A., 1937*)

Solution :

This question can be solved in two ways. Either individual median can be calculated of these respective universities ; and on the basis of it, the best university can be known, or weighted (weights to be ascertained by ourselves) arithmetic average is to be calculated and then the best university is known :

University Examination	Percentage Results in the University					
	A	Weights W_1	B	Weights W_2	C	Weights W_3
M.A.	80	10	75	10	70	10
M.Sc.	70	5	70	5	60	5
B.A.	65	35	80	35	70	35
B.Sc	60	30	70	30	80	30
B.Com.	75	20	65	20	75	20
		$\Sigma W_1 = n = 100$		$\Sigma W_2 = 100$		$\Sigma W_3 = 100$

Average Result for University A :

Examination	% of marks (X)	weights (W_1)	$x.w_1$
M.A.	80	10	800
M.Sc.	70	5	350
B.A.	65	35	2275
B.Sc.	60	30	1800
B.Com.	75	20	1500
		$\Sigma W_1 = n = 100$	$\Sigma XW_1 = 6725$

∴ weighted average for University $A = \frac{6725}{100} = 67 \cdot 25$. Thus, we ascertain that the students of 'A' University secured, on an average 67·25% marks.

Average Result for University B :

Examination	*% of marks* (X)	*weights* (W_2)	$X.W_2$
M.A.	75	10	750
M.Sc.	70	5	350
B.A.	80	35	2800
B.Sc.	70	30	2100
B.Com.	65	20	1300
		$\Sigma W_2 = 100$	$\Sigma XW_2 = 7300$

weighted arithmetic average for University $B = \frac{7300}{100}$

$= 73 \cdot 0$

On an average, the students of 'B' University secured 73% marks.

Average Result for University C :

Examination	*% of marks*(X)	*weight* (W_3)	$X.W_3$
M.A.	70	10	700
M.Sc.	60	5	300
B.A.	70	35	2450
B.Sc.	80	30	2400
B.Com.	75	20	1500
		$\Sigma W_3 = n = 100$	$\Sigma XW_3 = 7350$

weighted arithmetic average for the results of University C $= \frac{7350}{100}$ $= 73{\cdot}5$.

Thus, we find that the students of University C secured, on an average, 73·5% marks.

Therefore, our inference is that University C is the best.

Problem 16.—Locate the Mode of the following :

Size of item	*Frequency*	*Size of item*	*Frequency*
4	2	9	14
5	5	10	14
6	8	11	15
7	9	12	11
8	12	13	13

Solution :

Mode is to be located by the grouping method.

Size of item (X)	I Frequency	II	III	V	V	VI
4	2	7		15		
5	5		13		22	
6	8	17				29
7	9		21	35		
8	12	26			40	
9	14		28			43
10	14	29		40		
11	**15**		26		39	
12	11	24				
13	13					

Size of items with max. frequency in col.

I	—	11,	11 = 4 times.
II	—	10, 11	10 = 5 times.
III	—	9, 10	9 = 3 times.
IV	—	10, 11, 12	8 = 1 time.
V	—	8, 9, 10	12 = 1 time.
VI	—	9, 10, 11	

Therefore, our conclusion is that mode is located at 10.

Problem 17.—The following table gives the number of persons with different incomes in the U.S.A. during the year 1929. Calculate the average income per head.

Income in thousands of dollars	*No. of persons in lakhs*	*Income in thousands of dollars*	*No. of persons in lakhs*
under 1	13	10—25	27
1—2	90	25—50	6
2—3	81	50—100	2
3—5	117	100—1000	2
5—10	66		

(*Lucknow, B. Com., 1939*)

Solution :

For finding out the average income per head, we have to compute simple arithmetic average. Arithmetic average by short-cut method :

Income in thousands o dollars	Mid-value (x)	No. of persons in lakhs (f)	(ξ)	$f\xi$
under 1	·5	13	−17	−221
1— 2	1·5	90	−16	−14400
2— 3	2·5	80	−15	−1200
3— 5	4·0	117	−13·5	−1589·5
5— 10	7·5	66	−10·0	−660
10— 25	17·5	27	0	0
25— 50	37·5	6	+20	+120
50— 100	75·0	2	+57·5	+115
100—1000	550·0	2	+532·5	+1065
		$\Sigma f = n = 403$	$\xi = (X - A)$	$\Sigma f = -3810{\cdot}5$

$$M = A + \frac{\Sigma f\xi}{n} \qquad \text{Let the ass. mean } (A) = 17{\cdot}5$$

$$= 17{\cdot}5 + \frac{-3810{\cdot}5}{403}$$

$$= \frac{7052{\cdot}5 - 3810{\cdot}5}{403} = \frac{3242}{403} = 8{\cdot}044 \text{ approx.}$$

Therefore, the average income per head is 8·044 thousand dollars.

Problem 18.—The marks obtained by students of classes A and B are given below. Give as much information as you can regarding the composition of the classes in respect of intelligence :

Marks obtained	*No. of students in class A*	*No. of students in class B*
5—10	1	5
10—15	10	6
15—20	20	15
20—25	8	10
25—30	6	5
30—35	3	4
35—40	1	2
40—45	0	2

(*Agra, B.Com.*, 1939)

Solution :

(*a*) highest marks secured in class A is 35—40 by only one student.

(*b*) highest marks obtained in class B is 40—45 by two students.

(*c*) Arithmetic average for marks obtained in class A and B.

Marks obtained	Mid-value (x)	Class A		Class B	
		No. of students in class A (f_1)	$x.f_1$	No. of students in class B (f_2)	$x.f_2$
5—10	7·5	1	7·5	5	37·5
10—15	12·5	10	125·0	6	75·0
15—20	17·5	20	350·0	15	262·5
20—25	22·5	8	120·0	10	225·0
25—30	27·5	6	165·0	5	137·5
30—35	32·5	3	97·5	4	130·0
35—40	37·5	1	37·5	2	75·0
40—45	42·5	0	0	2	85·0
		$\Sigma f_1 = n_1 = 49$	$\Sigma xf_1 = 902·5$	$\Sigma f_2 = n_2 = 49$	$\Sigma xf_2 = 1027·5$

Arith. average for the marks obtained by the students of class A

$$A = \frac{\Sigma xf_1}{n_1}$$

$$= \frac{902·5}{49} = 18·4 \text{ approximately.}$$

Arith. average for the marks obtained by the students of class B.

$$B = \frac{\Sigma xf_2}{n_2}$$

$$= \frac{1027·5}{49}$$

$$= 20·9 \text{ approximately.}$$

Thus, on the basis of this comparison, we can conclude that the average marks obtained by class B is more than that obtained by class A, that is, class B is more intelligent.

(*d*) The comparative intelligence can also be known by finding out the median of class A and B, and also by weighted arithmetic average.

✓**Problem 19.**—Calculate the (*i*) unweighted mean of the prices in column III and (*ii*) the mean obtained by weighting each price by the quantity consumed, and explain why they differ as they do :

I	II	III
Articles of Food	*Quantity consumed*	*Price in Rs. per md.*
Flour	11·5 mds.	5·8
Ghee	5·6 ,,	58·4
Sugar	·28 ,,	8·2
Potato	·16 ,,	2·5
Oil	·35 ,,	20·0

(*Calcutta, M.A., 1937*)

Solution :

(A) Unweighted mean of column III—

Articles of Food	*Prices in Rs.* (x)	
Flour	5·8	
Ghee	58·4	$M = \frac{\Sigma x}{n}$
Sugar	8·2	$= \frac{94·9}{5}$
Potato	2·5	$= 18·9$ approx.
Oil	20·0	
$n = 5$	$\Sigma x = 94·9$	

(B) Weighted mean—weighting with the quantity consumed (short-cut method)

Commodity	Price in Rs. (X)	Weight (W)	ξ A=10	Wξ
Flour	5·8	11·5	−4·2	−48·3
Ghee	58·4	5·6	+48·4	+271·04
Sugar	8·2	·28	−1·8	−·504
Potato	2·5	·16	−7·5	−1·2
Oil	20·0	·35	+10	+3·5
		$\Sigma W = n = 17·89$	$\xi = (X - A)$	$\Sigma W\xi = 221·036$

$M^1 = A + \frac{\Sigma W\xi}{n}$ $\quad$ M^1 = weighted mean.

$= 10 + \frac{221 \cdot 036}{17 \cdot 89}$ $\quad$ $n = \Sigma W$.
A = assumed mean = 10.

$= \frac{178 \cdot 9 + 221 \cdot 036}{17 \cdot 89}$

$= \frac{399 \cdot 9}{17 \cdot 9}$ approximately.

$= 22 \cdot 3$ approximately.

√**Problem 20.**—The following data relate to sizes of shoes sold at a store during a given week. Find the average size by the short-cut method :

Size of shoes	4·5	5	5·5	6	6·5	7	7·5	8	8·5	9	9·5	10	10·5	11·0
No. of pairs	1	2	4	5	15	30	60	95	82	75	44	25	15	4

(*Calcutta, M.A., 1936*)

Solution :

Size of shoes (x)	No. of pairs (f)	ξ	fξ
4·5	1	−3·5	−3·5
5	2	−3	−6·0
5·5	4	−2·5	−10.0
6	5	−2	−10·0
6·5	15	−1·5	−22·5
7	30	−1	−30·0
7·5	60	−·5	−30 0
8	95	0	0
8·5	82	+·5	+41·0
9	75	+1	+75·0
9·5	44	+1·5	+66·0
10	25	+2	+50·0
10·5	15	+2·5	+37·5
11	4	+3	+12·0
	Σf = n = 457	ξ = (x—A) A = 8	Σfξ = 169·5

$M = A + \frac{\Sigma f\xi}{n}$ where M = arithmetic average.

$= 8 + \frac{169 \cdot 5}{457} = 8 + \cdot 3$

$= 8 \cdot 3$ approximately.

Problem 21.—Find out the average rates of (*a*) motion in the case of a person who rides the first mile @ 10 miles an hour, the next mile @ 8 miles an hour, and the third mile @ 6 miles an hour. (*b*) Increase in population which in the first decade has increased 20%, in the next 25% and in the third 44%.

(*Agra, B.Com., 1938*)

Solution :

For (*a*) part we have to compute harmonic mean and for (*b*) Geometric mean.

(*a*) Harmonic mean $(h)=\dfrac{n}{\dfrac{1}{a}+\dfrac{1}{b}+\cdots\cdots\dfrac{1}{x}}$

where n = number of items ; a, b, c, x etc. the values of respective items.

Putting the given values we get

$$h=\frac{3}{\frac{1}{10}+\frac{1}{8}+\frac{1}{6}}$$

$$=\frac{3}{\frac{12+15+20}{120}}$$

$$=\frac{360}{47} \qquad = 7\frac{31}{47} \qquad \textbf{Ans.}$$

(*b*) Geometric mean $(g)=\text{Antilog}\left[\dfrac{\log a+\log b+\cdots\log n}{n}\right]$

where g = geometric mean ; $\log a$, $\log b$ etc. are the logarithmic value of the respective sizes of the items a, b, n, etc. and n = total number of items.

$$g=\text{Antilog}\,\frac{\log 20+\log 25+\log 44}{3}$$

$$=\text{Antilog}\,\frac{1\cdot3010+1\cdot3979+1\cdot6435}{3}$$

$$=\text{Antilog}\,\frac{4\cdot3424}{3}$$

$$=\text{Antilog}\ 1\cdot4474$$

$=27\cdot54$ approximately. **Ans.**

Problem 22.—Following is the frequency distribution of yield of cane in tons per acre. Calculate the mean.

Class intervals	*Frequency*	*Class intervals*	*Frequency*
35—	7	60—	42
40—	8	65—	42
45—	12	70—	15
50—	26	75—	17
55—	32	80—85	9

(*Raj. M.A., 1950*)

Solution :

We have to compute the mean—simple arithmetic. Short-cut method is being applied. As the difference in the group-value is 5 the class intervals are computed first and then mean calculated.

Class intervals	Mid value (x)	Frequency(f)	$\xi(x-A)$	$f\xi$
35—40	37·5	7	−25	−175
40—45	42·5	8	−20	−160
45—50	47·5	12	−15	−180
50—55	52·5	26	−10	−260
55—60	57·5	32	− 5	−160
60—65	62·5	42	0	0
65—70	67·5	42	+ 5	+210
70—75	72·5	15	+10	+150
75—80	77·5	17	+15	+255
80—85	82·5	9	+20	+180
Let the assumed average (A) =62·5		$\Sigma f=n=210$	$\xi=(x-A)$	$\Sigma f\xi=-140$

$$M=A+\frac{\Sigma f\xi}{n}$$

$$=62\cdot5+\left[\frac{-140}{210}\right]$$

$$=62\cdot5+[-\cdot66]$$

$$=61\cdot84$$ **Ans.**

Problem 23.—Calculate the values of (*a*) the mode, (*b*) the median, and (*c*) the two quartiles of the following :

Wages in Rs.	*No. of workmen*	*Wages in Rs.*	*No. of workmen*
20—	8	25—	25
21—	10	26—	16
22—	11	27—	9
23—	16	28—29	6
24—	20		

(*Raj., M.A., 1950*)

Solution :

Wages in Rs.	No. of workmen (f)	Cumulative Frequency (cmf)
20—21	8	8
21—22	10	18
22—23	11	29
23—24	16	45
24—25	20	65
25—26	25	90
26—27	16	106
27—28	9	115
28—29	6	121

(*a*) THE MODE : 25—26 with 25 frequency is the modal group.

$$M_o = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1).$$

where M_o = mode.

l_1 and l_2 = lower and upper limits of the modal group.

f_1 = frequency of modal group.

f_0 = frequency of next lower group.

f_2 = frequency of next higher group.

$$= 25 + \frac{25-20}{50-20-16}(26-25)$$

$$= 25 + \frac{5}{14} \times 1$$

$= 25 + \cdot 35$ approximately.

$= 25 \cdot 35$ approx. **Ans.**

(*b*) THE MEDIAN—

$$Md = l_1 + \left\{ \frac{l_2 - l_1}{f_1}(m - c) \right\}$$

where Md = median

l_1 and l_2 = lower and upper limits of median group.

f_1 = frequency of median group.

m = size of $\left(\frac{n+1}{2}\right)^{th}$ item.

c = cumulative frequency of the next lower group than the median group.

$$Md = 24 + \left\{ \frac{25-24}{20}(61-45) \right\}$$

$$=24+\left\{\frac{1}{20}\times 16\right\}$$

$$=24+\cdot 8$$

$$=24\cdot 8 \text{ approx.} \quad \textbf{Ans.}$$

(c). THE QUARTILES —

$$Q_1=l_1+\left\{\frac{l_2-l_1}{f_1}(q_1-c)\right\}$$

where Q_1 = first quartile and q_1 = size of $\left(\frac{n+1}{4}\right)^{th}$ item.

Others as in (b).

$$Q_1=23+\left\{\frac{24-23}{16}\ (30-29)\right\}$$

$$=23+\frac{1}{16}$$

$$=23\cdot 06 \text{ approx.} \quad \textbf{Ans.}$$

$$Q_3=l_1+\left\{\frac{l_2-l_1}{f_1}(q_3-c)\right\}$$

where Q_3 = third quartile and q_3 = size of $\frac{3(n+1)^{th}}{4}$ item.

$$Q_3=26+\left\{\frac{27-26}{16}\ (91-90)\right\}$$

$$=26+\frac{1}{16}$$

$$=26\cdot 06 \text{ approximately.} \quad \textbf{Ans.}$$

Problem 24.—The data below gives the frequency distribution of the weights of 629 Indian soldiers. Calculate the average weight of a soldier and its standard error :—

Weights in lbs.	*Frequency*	*Weights in lbs.*	*Frequency*
105—	15	171—	35
116—	43	182—	16
127—	138	193—	5
138—	162	204—	3
149—	129	216—229	1
160—	82		

Solution :

For the first part simple arithmetic average is to be computed ; for the second part, we have to find the value of $\sigma_m=\frac{\sigma}{\sqrt{n}}$.

Weights in Lbs.	Mid-value (x)	Frequency (f)	(ξ)	$f\xi$	ξ^2	$f\xi^2$
105—116	110·5	15	—55	— 825	3025	45375
116—127	121·5	43	—44	—1892	1936	83248
127—138	132·5	138	—33	—4554	1189	150282
138—149	143·5	162	—22	—3564	484	78408
149—160	154·5	129	—11	—1419	121	14809
160—171	165·5	82	0		0	0
171—182	176·5	35	+11	+ 385	121	4235
182—193	187·5	16	+22	+ 352	484	7744
193—204	198·5	5	+33	+ 165	1089	5445
204—216	209·5	3	+44	+ 132	1936	5808
216—229	221·5	1	+55	+ 55	3025	3025
		$\Sigma = n = 629$	$\xi = (X-A)$	$\Sigma f\xi = -11165$		$\Sigma f\xi^2 = 398379$

Let the assumed average (A) = 165·5

$$\text{Arithmetic Mean (M)} = A + \frac{\Sigma f\xi}{n}$$

$$= 165{\cdot}5 + \left[\frac{-11165}{629}\right]$$

$$= 165{\cdot}5 - 17{\cdot}7 \text{ approx.}$$

$$= 147{\cdot}8 \text{ approximately.}$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma f\xi^2}{n}}$$

$$= \sqrt{\frac{398379}{629}}$$

$$= \sqrt{633{\cdot}35}$$

$$= 25{\cdot}16 \text{ approximately.}$$

$$\text{Standard Error of the Mean } (\sigma_m) = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{25{\cdot}16}{\sqrt{629}}$$

$$= \frac{25{\cdot}16}{25{\cdot}07} \text{ approx.}$$

$$= 1{\cdot}003 \text{ approximately.}$$

Problem 25.—Below are given the marks obtained by a batch of 20 students in a certain class-test in English and Hindi :—

Roll No. of students	*Marks in English*	*Marks in Hindi*	*Roll No. of students*	*Marks in English*	*Marks in Hindi*
1	53	58	11	25	10
2	54	55	12	42	42
3	52	25	13	33	15
4	32	32	14	48	46
5	30	26	15	72	50
6	60	85	16	51	64
7	47	44	17	45	39
8	46	80	18	33	38
9	35	33	19	65	30
10	28	72	20	29	36

In which subject is the level of knowledge of the students higher ?

(*Punjab, Sept., M.A., 1951*)

Solution :

To find out as to in which subject the level of knowledge is higher, we have to compute the median of both the series and then compare them. Knowledge will be higher in that subject, the median value of which is higher.

The series being an individual one, the series (marks obtained) are to be arrayed and the median is to be computed :—

No. of students	Marks obtained in English (arrayed in ascending order)	Marks obtained in Hindi (arranged in ascending order)
1	25	10
2	28	15
3	29	25
4	30	26
5	32	30
6	33	32
7	33	33
8	35	36
9	42	38
10	45	39
11	46	42
12	47	44
13	48	46
14	51	50
15	52	55
16	53	58
17	54	64
18	60	72
19	65	80
20	72	85

M*d* (for marks in English) $= \text{size of} \left(\frac{n+1}{2}\right)^{th}$ item.

$= ,, \;\; ,, \;\; \frac{21}{2}^{th}$ item.

$=$ 10.5th item.

$=\dfrac{\text{size of 10th item}+\text{size of 11th item}}{2}$

$=\dfrac{45+46}{2}$

$=45{\cdot}5$ approximately.

M*d* (for marks in Hindi) $=\text{size of } \dfrac{(n+1)}{2}^{th}$ item,

$=$,, $\frac{21}{2}^{th}$ item,

$=$,, $10{\cdot}5^{th}$ item,

$=\dfrac{\text{size of } 10^{th} \text{ item}+\text{size of } 11^{th} \text{ item}}{2}$

$=\dfrac{39+42}{2}$

$=40{\cdot}5$ approximately.

∴ knowledge in English is higher.

Problem 26.—Compute the Mode and Median of the following series. Account for their difference :

Size of item	*Frequency*	*Size of item*	*Frequency*
2	3	8	10
3	8	9	8
4	10	10	17
5	12	11	5
6	16	12	4
7	14	13	1

(*Nagpur, B. Com. 1945*)

Solution :

(*a*) Mode.—

Size of item	Frequency (I)	(II)	(III)	(IV)	(V)	(VI)
2	3	11		21		
3	8		18		30	
4	10	22				**38**
5	12		**28**	42		
6	16	**30**			**40**	
7	14		24			32
8	10	8		35		
9	8		25		30	
10	**17**	22				26
11	5		9	10		
12	4	5				
13	1					

Size of the items containing maximum frequency :—

Col. I —10.

Col. II —6,7.

Col. III—5,6

Col. IV—5,6,7

Col. V—6,7,8

Col. VI—4,5,6

∴ Mode is located at the size of 6^{th} item=16(frequency).

(*b*) Median.—

Size of the item	Frequency	*cmf*
2	3	3
3	8	11
4	10	21
5	12	33
6	16	49
7	14	63
8	10	73
9	8	81
10	17	98
11	5	103
12	4	107
13	1	108

$$Md = \text{Size of} \left(\frac{n+1}{2}\right)^{th} \text{item}$$

$= \text{,,} \ \frac{109}{2}^{th}$ item,

$=$,, $54{\cdot}5^{th}$ item.

$=$ 14 approximately.

or Md = Size of the $\left(\frac{n+1}{2}\right)$th item = $\left(\frac{13+1}{2}\right) = \frac{14}{2} = 7$

Then Median is located at the size of the 7th item is (14)

Problem 27.—The following table gives the population of males at different age groups of the U.K. and India at the time of the census of 1931.

Age group	*U.K.* (*lakhs*)	*India* (*lakhs*)
0—5	18	214
5—10	19	258
10—15	20	222
15—20	18	157
20—25	16	145
25—30	14	161
30—40	27	257
40—50	25	184
50—60	19	120
Above 60	17	100

Compare the average age of males in the two countries and account for the difference, if any.

(Alld., B.Com., 1936 ; Luck., B.Com., 1941).

Solution :

Age group	U.K.				India	
	lakhs (f)	Mid-value (x)	ξ (x—A)	$f\xi$	lakhs (f)	$f\xi$
0— 5	18	2·5	—25	—450	214	—5350
5—10	19	7·5	—20	—380	258	—5160
10—15	20	12·5	—15	—300	222	—3330
15—20	18	17·5	—10	—180	157	—1570
20—25	16	22·5	—5	—80	145	—725
25—30	14	27·5	0	0	161	0
30—40	27	35	+7·5	+202·5	257	+1927·5
40—50	25	45	+17·5	+437·5	184	+3220
50—60	19	55	+27·5	+522·5	120	+3300
Above 60	17	65	+37·5	+637·5	100	+3750
	n=193		A=20·5	$\Sigma f\xi$= +410	n=1818	$\Sigma f\xi$= —3937·5

Average age of males in U.K. :

$$M = A + \frac{\Sigma f\xi}{n}$$

$$= 27{\cdot}5 + \left[\frac{410}{193}\right] \qquad \text{where } A = 27{\cdot}5 \text{ years.}$$

$$= 27{\cdot}5 + 2{\cdot}12 \text{ years.}$$

$$= 29{\cdot}62 \text{ years approx.}$$

Average age of males in India :

$$M = A + \frac{\Sigma f\xi}{n}$$

$$= 27{\cdot}5 + \left[\frac{-3937{\cdot}5}{1818}\right] \text{ years.}$$

$$= 27{\cdot}5 - 2{\cdot}19 \text{ years.}$$

$$= 25{\cdot}31 \text{ years approximately.}$$

Thus, we find that the average age of males in U.K. is higher than that of India. Mainly, the reasons may be classified under two heads :—

(*i*) Due to climatological factor,

(*ii*) Due to better economic status and higher standard of living prevailing in U.K.

Problem 28.—The following table gives the marks obtained by a batch of 30 B.Com. students in a class-test in statistics. (Marks 100).

Roll No.	*Marks obtained*	*Roll No.*	*Marks obtained*
1	33	16	24
2	32	17	33
3	55	18	42
4	47	19	38
5	21	20	45
6	50	21	26
7	27	22	33
8	12	23	44
9	68	24	48
10	49	25	52
11	40	26	30
12	17	27	58
13	44	28	37
14	48	29	38
15	62	30	35

Find the values of the Mode, the Median and the Quartiles.

(*Alld., B.Com., 1938*).

Solution :

Roll No.	Marks obtained (I)	(II)	(III)	(IV)	(V)	(VI)
1	33	65		120		
2	32		87		134	
3	55	102				123
4	47		68	118		
5	21	71			98	
6	50		77			89
7	27	39		107		
8	12		80		129	
9	**68**	**117**				**157**
10	49		89	106		
11	40	57			101	
12	17		61			109
13	44	92		**154**		
14	48		**110**		134	
15	62	86				119
16	24		57	79		
17	33	75			113	
18	42		80			127
19	38	83		109		
20	45		71		104	
21	26	59				103
22	33		77	125		
23	44	92			144	
24	48		100			130
25	52	82		140		
26	30		88		125	
27	58	95				133
28	37		75	110		
29	38	73				
30	35					

I Col—9
II Col—9, 10
III Col—14, 15
IV Col—13, 14, 15
V Col—23, 24, 25
VI Col—9, 10, 11

∴ Mode is located at 9 **Ans.**

Median and Quartiles (series arranged in ascending order) :

Roll No.	Marks obtained	Roll No.	Marks obtained	Roll No.	Marks
1	12	11	33	21	47
2	17	12	35	22	48
3	21	13	37	23	48
4	24	14	38	24	49
5	26	15	38	25	50
6	27	16	40	26	52
7	30	17	42	27	55
8	32	18	44	28	58
9	33	19	44	29	62
10	33	20	45	30	68

$$Md = \text{Size of } \left(\frac{n+1}{2}\right)^{th} \text{item.}$$

$$= \quad ,, \quad \frac{31}{2}^{th} \text{ item.}$$

$$= \quad ,, \quad 15{\cdot}5^{th} \text{ item.}$$

$$= \frac{\text{size of } 15^{th} \text{ item} + \text{size of } 16^{th} \text{ item}}{2}$$

$$= \frac{38+40}{2}$$

$$= 39 \text{ approx.}$$

$$Q_1. \quad = \text{Size of } \left(\frac{n+1}{4}\right)^{th} \text{item.}$$

$$= \quad ,, \quad \frac{31}{4}^{th} \text{ item.}$$

$$= \quad ,, \quad 7{\cdot}75^{th} \text{ item.}$$

$$= \frac{\text{size of } 7^{th} \text{ item} + \text{size of } 8^{th} \text{ item}}{2}$$

$$= \frac{30+32}{2}$$

$$= 31 \text{ approx.}$$

$Q_3 = \text{size of } \frac{3(n+1)}{4}^{th} \text{ item.}$

$= \;\; ,, \;\; \frac{93}{4}^{th} \text{ item.}$

$=$23rd item approx.

$=$48 approximately.

✓**Problem 29.** Following is the frequency distribution of the yield of cane in tons per acre :

Class-intervals	*Frequency*	*Class-intervals*	*Frequency*
35	7	60	42
40	8	65	42
45	12	70	15
50	26	75	17
55	32	80—85	9

Calculate the Mean. (*Raj., M.A., 1949*)

Solution :

Class-interval	Mid-value (x)	Frequency (f)	ξ ($x-A$)	$f\xi$
35—40	37·5	7	−20	−140
40—45	42·5	8	−15	−120
45—50	47·5	12	−10	−120
50—55	52·5	26	− 5	−130
55—60	57·5	32	0	0
60—65	62·5	42	+ 5	+210
65—70	67·5	42	+10	+420
70—75	72·5	15	+15	+225
75—80	77·5	17	+20	+340
80—85	82·5	9	+25	+225
		$n=210$	A=57·5	$\Sigma f\xi=+910$

$$\text{Arithmetic Mean (M)} = A + \frac{\Sigma f \xi}{n}$$

$$= 57{\cdot}5 + \frac{910}{210}$$

$$= 57{\cdot}5 + 4{\cdot}33$$

$$= 61{\cdot}83 \text{ approximately.}$$

✓**Problem 30.**—Calculate the Mean of the following by short-cut method :—

Size of the item :	20	19	18	17	16	15	14	13	12	11
Frequency :	1	2	4	8	11	10	7	4	2	1

(*Practical Statistics*—Zia-ud-din, Ch. I. Q. 9)

Solution :

Size of the item (x)	Frequency (f)	ξ (x—A)	$f\xi$
20	1	+9	+ 9
19	2	+8	+16
18	4	+7	+28
17	8	+6	+48
16	11	+5	+55
15	10	+4	+40
14	7	+3	+21
13	4	+2	+ 8
12	2	+1	+ 2
11	1	0	0
	$n=50$	A=11	$\Sigma f\xi=227$

$$\text{Arithmetic Mean (M)} = A + \frac{\Sigma f \xi}{n}$$

$$= 11 + \frac{227}{50}$$

$$= 11 + 4{\cdot}54$$

$$= 15{\cdot}54 \text{ approximately.}$$

✓**Problem 31.**—The following table gives the annual birth and death rates in the U.S.A. during the period 1931 to 1945 :

Year :	1931	1932	1933	1934	1935	1936	1937	1938
Birth Rate :	18·0	17·4	16·6	17·2	16·9	16·7	17·1	16·6
Death Rate :	11·1	10·9	10·7	11·1	10·9	11·6	11·3	10·6
Year :	1939	1940	1941	1942	1943	1944	1945	
Birth Rate :	17·3	17·9	18·9	20·9	21·5	20·2	19·6	
Death Rate :	10·6	10·7	10·5	10·4	10·9	10·6	10·6	

Calculate the arithmetic average, the median and the mode of birth and death rates respectively. Which of these averages represents the series the best ?

(Raj., M.A. (P), 1954)

Solution :

Year	Birth Rate (x)	Death Rate (x_1)
1931	18·0	11·1
1932	17·4	10·9
1933	16·6	10·7
1934	17·2	11·1
1935	16·9	10·9
1936	16·7	11·6
1937	17·1	11·3
1938	16·6	10·6
1939	17·3	10·6
1940	17·9	10·7
1941	18·9	10·5
1942	20·9	10·4
1943	21·5	10·9
1944	20·2	10·6
1945	19·6	10·6
$n=15$	$\Sigma x=272·8$	$\Sigma x_1=162·5$

Arith. average of birth rate

$$=\frac{\Sigma x}{n}=\frac{272·8}{15}=18·18 \text{ approx.}$$

Arith. average of death rate

$$=\frac{\Sigma x_1}{n}=\frac{162.5}{15}=10·82 \text{ approx.}$$

Mode (by inspection)

(*a*) for birth rate $=21·5$ approx.

(*b*) for death rate $=11·6$ approx.

M*d* (for birth rate) = size of $\left(\frac{n+1}{2}\right)^{th}$ item.

= size of 8^{th} item.

= 16·6 approx.

M*d* (for death rate) = size of $\left(\frac{n+1}{2}\right)^{th}$ item.

= ,, ,, $\left(\frac{15+1}{2}\right)^{th}$ item.

= size of 8^{th} item.

= 10·6 approximately.

		Birth Rate	Death Rate
M	=	18·18	10·82
M*o*	=	21·5	11·6
M*d*	=	16·6	10·6

A comparative study would indicate, beyond doubt, that in this case the mean gives the best result because it takes into consideration all individual items. In mode we lay undue emphasis only on one item in the series ; which may be abnormal. Median takes into consideration only the median (centre) value ; and it is not much effected by the extreme variations ; which may actually influence the results to a very considerable extent. Thus, mean gives the most satisfactory result in this case as it takes into consideration all values.

Problem 32.—A class of 16 boys and girls was given a common intelligence test and the following marks were obtained :—

S. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Marks (boys)	15	35	43	46	48	48	49	50	55	56	60	64	71	75	80	85
Marks (girls)	10	30	45	52	55	58	61	61	63	69	70	72	74	75	75	90

(*a*) Calculate the arithmetic mean and the median of the marks obtained by boys and girls separately. (*b*) On the basis of the above figures would you assert that girls are more intelligent than boys ?

(*Raj., M.A., 1954*)

Solution:

S. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$n=16$
Marks (boys) (x)	15	35	43	46	48	48	49	50	55	56	60	64	71	75	80	85	$\Sigma x=880$
Marks (girls) (x_1)	10	30	45	52	55	58	61	61	63	69	70	72	74	75	75	90	$\Sigma x_1=960$

$$\text{Mean for boys' marks} = \frac{\Sigma x}{n} = \frac{880}{16}$$

$$= 55 \text{ marks.}$$

$$\text{Mean for girls' marks} = \frac{\Sigma x_1}{n} = \frac{960}{16}$$

$$= 60 \text{ marks.}$$

$$\text{Median (for boys' marks)} = \text{Size of } \frac{(n+1)^{th}}{2} \text{ item.}$$

$$= \text{,,} \frac{17}{2}^{th} \text{ item} = 8{\cdot}5^{th.} \text{ item}$$

$$= \frac{\text{Size of } 8^{th} \text{ item} + \text{Size of } 9^{th} \text{ item}}{2}$$

$$= \frac{50+55}{2} = \frac{105}{2}$$

$$= 52{\cdot}5 \text{ marks.}$$

$$\text{Median (for girls' marks)} = \text{Size of } \frac{(n+1)}{2}^{th} \text{ item.}$$

$$= \text{,,} \frac{17^{th}}{2} \text{ item.}$$

= Size of 8.5^{th} item.

$$= \frac{\text{Size of } 8th \text{ item} + \text{Size of } 9th \text{ item}}{2}$$

$$= \frac{61+63}{2} = \frac{124}{2}$$

= 62 marks.

(*b*) On the basis of the values obtained by the mean and median, we assert that in this particular case of 16 boys and 16 girls the average intelligence of girls are higher.

Problem 33. Obtain the Mean, Median and Mode of the following distribution :—

Marks	*Frequency*	*Marks*	*Frequency*
10—25	6	55—70	26
25—40	20	70—85	3
40—55	44	85—100	1

Solution :

Marks	Mid-value (x)	Frequency (f)	Cm f	ξ (x—A)	$f\,\xi$
10— 25	17·5	6	6	—30·0	—180
25— 40	32·5	20	26	—15·0	—300
40— 55	47·5	44	70	0	0
55— 70	62·5	26	96	+15·0	+390
70— 85	77·5	3	99	+30·0	+ 90
85—100	92·5	1	100	+45·0	+ 45
		n=100		A=47.5	$\Sigma f\,\xi$=+45

Let the assumed mean (A)=47·5

$$M = A + \frac{\Sigma f\,\xi}{n}$$

$$= 47·5 + \frac{45}{100}$$

$$= 47·5 + ·45$$

$$= 47·95 \text{ approx.}$$

$$M_o = l_1 + \frac{f_1 - f_o}{2f_1 - f_o - f_2}(l_2 - l_1)$$

$$= 40 + \frac{44-20}{88-20-26}(55-40)$$

$$= 40 + \frac{24}{42} \times 15$$

$$= 40 + 8·57 \text{ approx.}$$

$$= 48·57 \text{ approx.}$$

$$\text{Md} = l_1 + \left\{ \frac{l_2 - l_1}{f_1} (m-c) \right\}$$

$$\left[\text{where } m = \text{Size of} \left(\frac{n+1}{2} \right)^{th} \text{ item.} \right.$$

$$= \text{,,} \ \frac{101^{th}}{2} \text{ item.}$$

$$= 50{\cdot}5^{th} \text{ item. (40—55 is the Md. group)} \Big]$$

$$= 40 + \left\{ \frac{55-40}{44} (50{\cdot}5 - 26) \right\}$$

$$= 40 + \frac{15}{44} \times 24{\cdot}5$$

$$= 40 + 8{\cdot}3 \text{ approx.}$$

$$= 48{\cdot}3 \text{ approximately.}$$

$$= 40 + \left\{ \frac{15}{44} \times 24{\cdot}5 \right\}$$

$$= 40 + 8.35 \text{ approx.}$$

$$= 48{\cdot}35 \text{ approx.}$$

Problem 34.—The following is the distribution of wages per thousand employees in a certain factory :

Daily wages in as.	2	4	6	8	10	12	14	16	18	20	22	24	Total
No. of employees	3	13	43	102	175	220	204	139	69	25	6	1	1000

Calculate the modal and median wages, and explain why there is a difference between the two.

(*Alld. ,M.A., 1940*)

Solution :

[A] Mode.—

Daily wages in as.	No. of empl. I (*f*)	II	III	IV	V	VI
2	3	16		59		
4	13		56		158	
6	43	145				320
8	102		277	497		
10	175	**395**			599	
12	**220**		**424**			**563**
14	204	343		412		
16	139		208		233	
18	69	94				100
20	25		31	32		
22	6	7				
24	1					

Col. I—12,
Col. II—10,12
Col. III—12,14
Col. IV—8,10,12
Col. V—10,12,14
Col. VI—12,14,16

∴ Mode is located at 12 (*e.g.* the largest number of employees are paid 12 as. as daily wages)

[B] MEDIAN.—

Daily wages in as.	2	4	6	8	10	12	14	16	18	20	22	24
No. of Employees (f)	3	13	43	102	175	220	204	139	69	25	6	1
Cmf	3	16	59	161	336	556	760	899	968	993	999	1000 $n=1000$

$Md=$ size of $\left(\frac{n+1}{2}\right)^{th}$ item $=$ size of $\left(\frac{1000+1}{2}\right)^{th}$ item.

$=$ size of $500{\cdot}5^{th}$ item. Thus, it is located at the 556 *cmf.* group.

$=12$ annas approximately.

There is no significant difference between these two values.

Probelm 35.—From the following table find the mean, median and modal ages of married women at first child-births and calculate also the coefficient of variation :

Age at the brith of first child :

13 14 15 16 17 18 19 20 21 22 23 24 25

Number of married women :

37 162 343 390 256 433 161 355 65 85 49 46 40

(*Raj, M.A., 1955*)

Solution :

Age at the brith of first child (x)	No. of married women (f)	Cmf	ξ ($x-A$)	$f\xi$
13	37	37	−6	−122
14	162	199	−5	−810
15	343	542	−4	−1372
16	390	932	−3	−1170
17	256	1188	−2	− 512
18	433	1621	−1	− 433
19	161	1782	0	0
20	355	2137	+1	+ 355
21	65	2202	+2	+ 130
22	85	2287	+3	+ 255
23	49	2336	+4	+ 196
24	46	2382	+5	+ 230
25	40	2422	+6	+ 240
	$n=2422$		$A=19$	$\Sigma f\xi=-3013$

Let the assumed mean (A) $=19$

(*i*) Mean (M) $=A+\frac{\Sigma f\xi}{n}$

$$=19+\left[\frac{-3013}{2422}\right]$$

$$=19-1\cdot2 \text{ approx.}$$

$$=17\cdot8 \text{ years approx.}$$

(*ii*) Mode (M*o*) = 18 years by inspection (highest frequency = 433)

(*iii*) Median (M*d*) = size of $\left(\frac{n+1}{2}\right)^{th}$ item

$$= \text{,, ,, } \frac{2423}{2} \text{ ,, ,,}$$

$$= \text{,, ,, } 1211\cdot5 \text{ ,, ,,}$$

$$=18 \text{ years approx.}$$

To compute the co-efficient of variation, we have to find out the standard deviation (σ).

Then, Co-efficient of variation $=\frac{\sigma}{M}\times100$

where M = arithmetic mean.

Problem 36. Find out the average height of a clerk in a certain office from the following figures: What is the median height and how far does it differ from the mode?

Height	*Frequency*	*Weight*	*Frequency*
5′ 6″	1	5′ 11″	1
5′ 7″	2	6′ 0″	2
5′ 8″	4	6′ 1″	1
5′ 9″	3	6′ 2″	1
5′ 10″	2	6′ 3″	1

(*Agra, M.A., 1937*)

Solution:

(*i*) Mode.—

Height in inches	Freq. (I)	II	III	IV	V	VI
66	1	3		7		
67	2		6		9	
68	4	7				9
69	3		5	6		
70	2	3			5	
71	1		3			4
72	2	3		4		
73	1		2		3	
74	1	2				
75	1					

Col. I—68,
Col. II—68,69
Col. III—67,68
Col. IV—66,67,68
Col. V—67,68,69
Col. VI—68,69,70

∴ mode is located at 68″ or 5′ 8″ having highest frequency (4).

(*ii*) MEDIAN.—

Height in inches	66	67	68	69	70	71	72	73	74	75
Frequency	1	2	4	3	2	1	2	1	1	
Cmf	1	3	7	10	12	13	15	16	17	18

$n=18$

$Md=$ size of $\left(\frac{n+1}{2}\right)^{th}$ item. $=$ size of $\frac{19}{2}^{th}$ item.

$=$ size of $9{\cdot}5^{th}$ item. $=69''$ approximately.

Or

$9{\cdot}5^{th}$ item $=3+\frac{1}{2}(4-3)=3+\frac{1}{2}$

$=3{\cdot}5^{th}$ frequency group.

$=$ between the heights of $5'\ 8''$ and $5'\ 9''$.

$=5'\ 9''$ approximately.

✓**Problem 37.**—The following is the age distribution of candidates appearing at the Matriculation and Intermediate Arts Examinations of the Patna University in 1937 :

Age in years	12—	13—	14—	15—	16—	17—	18—	19—	20—	21—	22—	Total
Matriculation	5	48	189	303	522	980	981	794	515	474	—	4811
Intermediate	—	—	—	5	45	87	127	150	155	127	175	871

Compare the median and modal ages of the Matriculation candidates with those of I.A. candidates.

(*Patna, M.A., 1940*)
(B.com., Alld., 1947)

Solution :

Age in years	Matriculation		Intermediate	
	No of candidates	*Cmf*	No. of candidates	*Cmf*
12—13	5	5	—	—
13—14	48	53	—	—
14—15	189	242	—	—
15—16	303	545	5	5
16—17	522	1067	45	50
17—18	980	2047	187	137
18—19	981	3028	127	264
19—20	794	3822	150	414
20—21	515	4337	155	569
21—22	474	4811	127	696
22—23	—	4811	175	871

m for matriculation=size of $\left(\frac{n+1}{2}\right)^{th}$ item.

$$= \text{,,} \quad \text{,,} \quad \frac{4812}{2}^{th} \text{ item} = 2406th \text{ item.}$$

$$\text{Md (for Matriculation)} = l_1 + \left\{\frac{l_2 - l_1}{f_1}(m-c)\right\}$$

[l_1, l_2 are the lower and upper limits of the median group ; f_1=freq. of the median group ; m=size of $\left(\frac{n+1}{2}\right)^{th}$ item ; c=$cmf.$ of the next lower group]

$$= 18 + \left\{\frac{19-18}{981}(2406-980)\right\}$$

$$= 18 + \left\{\frac{1}{981} \times 1426\right\} = 18 + 1{\cdot}4$$

$$= 19{\cdot}4 \text{ years approximately.}$$

$$\text{Md (for Intermediate)} = l_1 + \left\{\frac{l_2 - l_1}{f_1}(m-c)\right\} \quad m = \text{size of } \left(\frac{n+1}{2}\right)^{th} \text{ item.}$$

$$= 20 + \left\{\frac{21-20}{155}(436-414)\right\} = \frac{872^{th}}{2} \text{ item.}$$

$$= 20 + \left\{\frac{1}{155} \times 22\right\} = 20 + {\cdot}14 \quad = 436^{th} \text{ item.}$$

$$= 20{\cdot}14 \text{ years approx.}$$

M_0 (by location) (*i*) For Intermediate— age group 17—18 years.
(*ii*) For Matriculation—age-group 18—19 years.

Problem 38.—Amend the following table, and locate the median from the amended table.

Sizes	*Frequency*	*Sizes*	*Frequency*
10—15	10	30—35	28
15—17·5	15	35—40	30
17·5—20	17	45—up	40
22—30	25		

(*Alld., B. Com., 1942*)

Solution :

Amended size	*Frequency*	*Cumulative frequency*
10—20	(10+15+17)=42	42
20—30	25	67
30—40	58	125
40—50	40	165

$$m = \text{size of } \left(\frac{n+1}{2}\right)^{th} \text{ item.}$$

$$= \text{,, ,, } \left(\frac{165+1}{2}\right) \text{ ,, } = \frac{166}{2} = 83^{rd} \text{ item.}$$

$$Md = l_1 + \left\{ \frac{l_2 - l_1}{f_1} (m - c) \right\}$$

[l_1 and l_2=lower and upper limits of the median group ; f_1= frequency of the median group ; m=size of $\left(\frac{n+1}{2}\right)^{th}$ item. c=cumulative frequency of the next lower group].

$$= 30 + \left\{ \frac{40-30}{58} (83-67) \right\}$$

$$= 30 + \left\{ \frac{10}{58} \times 16 \right\}$$

$$= 30 + 2{\cdot}76$$

$$= 32{\cdot}76 \text{ approximately.}$$

Problem 39.—Find the mode and Median from the the following table :

Marks	*Students*	*Marks*	*Students*
0—10	2	40—50	35
10—20	18	50—60	20
20—30	30	60—70	6
30—40	45	70—80	3

(*Agra, B.Com., 1941*)

Solution :

[**A**] Mode.—

Marks (x)	Frequency (f)	Cumulative frequency (cmf)
0—10	2	2
10—20	18	20
20—30	30	50
30—40	45	95
40—50	35	130
50—60	20	150
60—70	6	156
70—80	3	159

By inspection, we find that marks-group 30—40, with 45 (f) is the modal group. Then, by interpolation,

$$M_0 = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} (l_2 - l_1)$$

$$= 30 + \frac{45 - 30}{90 - 30 - 35} (40 - 30)$$

$$=30+\frac{15}{25}\times 10$$
$$=30+6$$
$$=36 \text{ approx.}$$

[**B**] MEDIAN.

$$m=\text{size of } \left(\frac{n+1}{2}\right)^{th} \text{ item.}$$
$$= \text{,, ,, } \frac{160}{2} \text{ ,, ,, } =80^{th} \text{ item.}$$

and median group = 30 − 40 marks group.

$$\text{Md} = l_1+\left\{\frac{l_2-l_1}{f_1}(m-c)\right\}$$
$$=30+\left\{\frac{40-30}{45}(80-50)\right\}$$
$$=30+\left\{\frac{10}{45}\times 30\right\}$$
$$=30+6{\cdot}66 \text{ approx.}$$
$$=36{\cdot}66 \text{ approx.}$$

Problem 40. The monthly incomes of 10 families in rupees in a certain locality are given below :

Family :	A	B	C	D	E	F	G	H	I	J
Income :	85	70	10	75	500	8	42	250	40	36

Calculate the Mean, the Geometric Mean and the Harmonic Mean. Which of the above three averages represents the figures best ?

(*Agra, B.Com., 1945*)

Solution :

Family	A	B	C	D	E	F	G	H	I	J	$n=10$
Income (x)	85	70	10	75	500	8	42	250	40	36	$\Sigma x=$ 1116
Logarithms	1·9294	1·8451	1·0000	1·8751	2·6990	·9031	1·6232	2·3979	1·6021	1.5563	
Reciprocals	·1176	·1429	·0000	·1333	·2000	·1250	·2381	·4000	·2500	·2778	

(i) $M = \frac{\Sigma x}{n} = \frac{1116}{10} = 111{\cdot}6$ approx.

(ii) $G_m = \text{Antilog}\left[\frac{\log a + \log b + \ldots\ldots \log_n}{n}\right]$

$= \text{Antilog}\left[\frac{1{\cdot}9294+1{\cdot}8451+1+1{\cdot}8751+2{\cdot}6990+{\cdot}9031+1{\cdot}6232+2{\cdot}3979+1{\cdot}6021+1{\cdot}5563}{10}\right]$

$= \text{Antilog}\frac{17{\cdot}2312}{10}$

$= \text{Antilog } 1{\cdot}72312$

$= \text{Antilog } 5284$

$= 52{\cdot}84$ approx.

(iii) $H_m = \frac{n}{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\ldots\ldots\frac{1}{n}}$

$= \frac{10}{{\cdot}1176+{\cdot}1429+{\cdot}1333+{\cdot}2+{\cdot}125+{\cdot}2381+{\cdot}4+{\cdot}25+{\cdot}2778}$

$= \frac{10}{1{\cdot}8847}$ — approx.

$= 5{\cdot}3$ approx.

$= 18{\cdot}87$ approx.

Geometric mean represents the figures best.

Problem 41.—From the following figures given below, find the Mode, Median and Quartiles :

Age :	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
No. of persons :	50	70	100	180	150	120	70	59

(*Agra, B.Com., 1949*)

Solution :

Age	20—25	25—30	30—35	35—40	40—45	45—50	50—55	55—60
No. of persons (*f*)	50	70	100	180	150	120	70	59
cmf	50	120	220	400	550	670	740	799

(A) Mode—By inspection we find that 35—40 is the modal group.

$$M_0 = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= 35 + \frac{180-100}{360-100-150} = 40-35$$

$=35+3{\cdot}63$ approx.

$=38{\cdot}63$ approx.

(B) Median (Md) $=l_1+\left\{\frac{l_2-l_1}{f_1}(m-c)\right\}$

where $m=$ size of $\left(\frac{n+1}{2}\right)^{th}$ item.

$=\frac{780}{2}=390^{th}$. item.

$=35+\left\{\frac{40-35}{180}(390-220)\right\}$

$=35+4{\cdot}72$

$=39{\cdot}72$ approx.

(C) $Q_1=l_1+\left\{\frac{l_2-l_1}{f_1}(q_1-c)\right\}$

where $q_1=$ size of $\left(\frac{n+1}{4}\right)^{th}$ item.

$=\frac{780}{4}=195^{th}$ item.

thus, 30—35 is the Q_1 group.

$=30+\left\{\frac{35-30}{100}(195-120)\right\}$

$=30+3{\cdot}75$ approx.

$=33{\cdot}75$ approx.

(D) $Q_3=l_1+\left\{\frac{l_2-l_1}{f_1}(q_3-c)\right\}$

where $q_3=$ size of $\frac{3(n+1)}{4}^{th}$ item.

$=585^{th}$ item.

= located in the 45-50 group.

$=45+\left\{\frac{50-45}{120}(585-550)\right\}$

$=45+1{\cdot}45$ approx.

$=46{\cdot}45$ approx.

Problem 42.—From the table given below, find the Mean and the Mode

Marks	1—5	6—10	11—15	16—20	21—25	26—30	31—35	36—40	41—45
No. of candidates	7	10	16	32	24	18	10	5	1

(*Agra, B.Com.*, 1951)

Solution :

Marks	Mid-value (x)	Frequency (f)	$\xi=(x-A)$	$f\xi$
1—5	3	7	−20	−140
6—10	8	10	−15	−150
11—15	13	16	−10	−160
16—20	18	32	−5	−160
21—25	23	24	0	0
26—30	28	18	+5	+90
31—35	33	10	+10	+100
36—40	38	5	+15	+75
41—45	43	1	+20	+20
		$n=123$	A=23	$\Sigma f\xi=-325$

Let the assumed mean (A)=23.

$$M=A+\frac{\Sigma f\xi}{n}$$

$$=23+\left[\frac{-325}{123}\right]$$

$$=23-2{\cdot}64 \text{ approx.}$$

$$=20{\cdot}36 \text{ approx.}$$

$$Md=l_1+\frac{f_1-f_0}{2f_1-f_0-f_2}(l_2-l_1)$$

$$=16+\frac{32-16}{64-16-24}(20-16)$$

$$=16+\frac{16}{24}\times 4$$

$$=16+2{\cdot}66$$

$$=18{\cdot}66 \text{ approx.}$$

Problem 43.—Refer to the following frequency table and find out the Mean, Median and Mode.

Class interval (score)	*Frequencies*	*Class interval (score)*	*Frequencies*
130—134	1	75—79	39
125—129	2	70—74	29
120—124	4	65—69	23
115—119	7	60—64	17
110—114	9	55—59	11
105—109	10	50—54	8
100—104	15	45—49	6
95—99	28	40—44	3
85—94	41	35—39	2
80—84	72	30—34	2

(*L.T. 1945*)

Solution :

Class interval	Mid-value (x)	Freq. (f)	$C\,m\,f$	ξ ($x-$A)	$f\xi$	Class interval	Mid-value (x)	(f)	$C\,m\,f$	ξ	$f\xi$
130—134	132	1	1	−45	−45	75—79	77	39	228	+5	+195
125—129	127	2	3	−40	−120	70—74	72	29	257	+10	+290
120—124	122	4	7	−35	−245	65—69	67	23	280	+15	+345
115—119	117	7	14	−30	−420	60—64	62	17	297	+20	+340
110—114	112	9	23	−25	−575	55—59	57	11	308	+25	+275
105—109	107	10	33	−20	−660	50—54	52	8	316	+30	+240
100—104	102	15	48	−15	−720	45—49	47	6	322	+35	+210
95—99	97	28	76	−10	−760	40—44	42	3	325	+40	+120
85—94	87	41	117	−5	−585	35—39	37	2	327	+45	+90
80—84	82	72	189	0	0	30—34	32	2	329	+50	+100
									$n=329$		$\Sigma f\xi=-1925$

Let the assumed mean (A) = 82 $\qquad n=\Sigma f = 329.$

$$M = A + \frac{\Sigma f\xi}{n}$$

$$= 82 + \left[\frac{-1925}{329}\right]$$

$$= 82 - 5{\cdot}85$$

$$= 76{\cdot}15 \text{ approximately.}$$

$$Md = l_1 + \left\{\frac{l_2 - l_1}{f_1}(m-c)\right\}$$

[where m = size of $\left(\frac{n+1}{2}\right)^{th}$ item

= ,, 165^{th}. item.

and 80—84 is the median group]

$$=80+\left\{\frac{84-80}{72}\ (165-117)\right\}$$

$=80+2{\cdot}66$ approx.

$=82{\cdot}66$ approximately.

$$M_o=l_1+\frac{f_1-f_o}{f_1-f_o-f_2}\ (l_2-l_1)$$

$$=80+\frac{72-41}{144-41-39}(84-80)$$

[where 80—84 is the modal group

$$=80+\frac{31}{64}\times 4$$

$=80+1{\cdot}93$ approx.

$=81{\cdot}93$ approximately.

✓**Problem 44.**—The frequency distribution of cost of production of gur in rupees per maund for different holdings in two districts are given below. Find the average cost in each district.

Cost in Rs. per md.	2—3	3—4	4—5	5—6	6—7	7—8	8—9	9—10	10—11	11—12	12—13
District A	9	32	37	21	13	7	5	2	1	2	1
District B	1	10	34	23	21	14	10	9	5	2	1

(*I.C.S. 1939*)

Solution :

Cost in Rs. per md.	Mid-value (x)	District A (f)	ξ	fξ	District B (f)	ξ	fξ
2— 3	2·5	9	—5·0	— 45	1	—5	— 5
3— 4	3·5	32	—4·0	—128	10	—4	— 40
4— 5	4·5	37	—3·0	—111	34	—3	—102
5— 6	5·5	21	—2·0	— 42	23	—2	— 46
6— 7	6.5	13	—1.0	— 13	21	—1	— 21
7— 8	7.5	7	0	0	14	0	0
8— 9	8·5	5	+1·0	+ 5	10	+1	+10
9—10	9·5	2	+2·0	+ 4	9	+2	+18
10—11	10·5	1	+3·0	+ 3	5	+3	+15
11—12	11.5	2	+4·0	+ 8	2	+4	+ 8
12—13	12·5	1	+5·0	+ 5	1	+5	+ 5
		n=130		Σfξ = —314	n=130		Σfξ = —158

Let the assumed mean (A) = 7·5

$$\text{Mean for District A} = A + \frac{\Sigma f \xi}{n}$$

$$= 7{\cdot}5 + \left[\frac{-314}{130}\right] \quad = 7{\cdot}5 - 2{\cdot}41 \text{ approx.}$$

$$= 5{\cdot}09 \text{ approx.}$$

$$\text{Mean for District B} = A + \frac{\Sigma f \xi}{n}$$

$$= 7{\cdot}5 + \left[\frac{-158}{130}\right] \quad = 7{\cdot}5 - 1{\cdot}21 \text{ approx.}$$

$$= 6{\cdot}29 \text{ approx.}$$

Problem 45.—The following is the frequency distribution of a random sample of 509 employees by weekly earnings. Calculate the average weekly earnings and its sampling error :

Weekly earnings ...	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40—42
No. of Employees ...	3	6	10	15	24	42	75	90	79	55	36	26	19	13	9	7

(*I.A.S. 1948*)

Solution :

On the basis of the last group (40—42), we have first to place the earnings in grouped form with 2 as the class-interval.

Weekly Earnings	Mid-Value (x)	No. of Employees (f)	$\xi = (x - A)$	$f\xi$
10—12	11	3	—16	— 48
12—14	13	6	—14	— 84
14—16	15	10	—12	—120
16—18	17	15	—10	—150
18—20	19	24	— 8	—192
20—22	21	42	— 6	—252
22—24	23	75	— 4	—300
24—26	25	90	— 2	—180
26—28	27	79	0	0
28—30	29	55	+ 2	+110
30—32	31	36	+ 4	+144
32—34	33	26	+ 6	+156
34—36	35	19	+ 8	+152
36—38	37	13	+10	+130
38—40	39	9	+12	+108
40—42	41	7	+14	+98
		n=509		$\Sigma f\xi$ = —428

Let the assumed mean (A) = 27

$$M = A + \frac{\Sigma f \xi}{n}$$

$$= 27 + \left[\frac{-428}{509}\right]$$

$$= 27 - \cdot 84 \text{ approx.}$$

$$= 26 \cdot 14 \text{ approx.}$$

Absolute error (AE) $= a - e$

where a = actual value
e = estimated value.

Relative Error (R) $= \frac{AE}{e}$ or $\frac{a-e}{e}$

The relative error when expressed by way of a percentage is known as the Percentage Error.

The relative error is positive if $a > e$ and negative if $a < e$.

Problem 46.—The marks (out of a maximum of 100) obtained by candidates in an examination are shown in the following frequency table. Calculate the Mode and the Arithmetic Average.

Marks	17·5—22·5	22·5—27·5	27·5—32·5	32·5—37·5	37·5—42·5	42·5—47·5	47·5—52·5	52·5—57·5	57·5—62·5	62·5—67·5	67·5—72·5
No. of Candidates	2	8	33	80	170	243	213	145	67	35	4

(*Agra, B.Com. Part. II, 1954*)

Solution :

Marks	Mid-value (x)	No. of Candidates (f)	ξ (X—A)	$f\xi$
17·5—22·5	20	2	−25	− 50
22·5—27·5	25	8	−20	− 160
27·5—32·5	30	33	−15	− 495
32·5—37·5	35	80	−10	− 800
37·5—42·5	40	170	− 5	− 850
42·5—47·5	45	243	0	0
47·5—52·5	50	213	+ 5	+1065
52·5—57·5	55	145	+10	+1450
57·5—62·5	60	67	+15	+1005
62·5—67·5	65	35	+20	+ 700
67·5—72·5	70	4	+25	+ 100
		n=1000		Σf = +1965

Let the assured mean (A) = 45

$$M = A + \frac{\Sigma f\xi}{n}$$

$$= 45 + \frac{1965}{1000}$$

$$= 45 + 1{\cdot}965 \text{ approx.}$$

$$= 46{\cdot}965 \text{ approx.}$$

$$M_0 = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= 42{\cdot}5 + \frac{243 - 170}{486 - 170 - 213}(47{\cdot}5 - 42{\cdot}5)$$

[where 42·5 − 47·5 is the modal group]

$$= 42{\cdot}5 + \frac{73}{103} \times 5$$

$$= 42{\cdot}5 + 3{\cdot}54 \text{ marks approx.}$$

$$= 46{\cdot}04 \text{ marks approximately.}$$

Problem 47.—Determine Mode and the Median from the following figures :

25, 15, 23, 40, 27, 25, 23, 25, 20

(*Agra, M.Com. 1954*)

Solution :

Putting the figures in ascending order, we get

Serial No.	*Frequency*	*Serial No.*	*Frequency*
1	15	6	25
2	20	7	25
3	23	8	27
4	23	9	40
5	25		

$$Md = \text{size of } \left(\frac{n+1}{2}\right)^{th} \text{ item.}$$

$$= \text{,, ,, } \frac{10}{2}^{th} \text{ item } (n \text{ being equal to } 9)$$

$$= \text{,, } 5^{th} \text{ item.}$$

$$= 25.$$

$$M_0 = 25$$

[Mode in an individual series is located by inspection only. In this case, we find that the figure 25 repeats the largest number of times. Thus, we interpolate that Mode is located at 25].

Problem 48.—Calculate the mean, Q_1, Q_3, D_7, D_5, D_4, H_5, O_3, O_7, O_5, P_{15}, P_{10}, P_{20}, P_{40}, P_{60}, P_{90} from the following :

Earnings :	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Employees :	3	6	10	15	24	42	75	90	79	55	36	26	19	13	8

Solution :

Earnings (x)	Employees (f)	*cmf*	$\xi=(x-A)$	$f\xi$
9	3	3	−7	−21
10	6	9	−6	−36
11	10	19	−5	−50
12	15	34	−4	−60
13	24	58	−3	−72
14	42	100	−2	−84
15	75	175	−1	−75
16	90	265	0	0
17	79	344	+1	+79
18	55	399	+2	+110
19	36	435	+3	+108
20	26	461	+4	+104
21	19	480	+5	+95
22	13	493	+6	+78
23	8	501	+7	+56
	$n=501$			$\Sigma f\xi=+232$

Let the assumed mean (A)=16

$$M=A+\frac{\Sigma f\xi}{n}$$

$$=16+\frac{232}{501}=16+\cdot46 \qquad =16\cdot46 \text{ approx.}$$

$$Q_1=\text{size of }\left(\frac{n+1}{4}\right)^{th}\text{ item}$$

$$= \text{,, ,, } \frac{502}{4} \text{ ,, ,, } \qquad =125\cdot5^{th}\text{ item.}$$

$=15$ approx.

$$Q_3=\text{size of }3\left(\frac{n+1}{4}\right)^{th}\text{ item}$$

$$= \text{ ,, } 3\times\frac{502}{4}{}^{th}\text{ item} \qquad =376\cdot5^{th}\text{ item.}$$

$=18$ approximately.

$$D_7=\text{size of }\frac{7(n+1)th}{10}\text{ item}$$

$$=7\times\frac{502}{10}{}^{th}\text{ item} \qquad =351\cdot4^{th}\text{ item.}$$

$=18$ approximately.

D_5=size of $\frac{5(n+1)}{10}^{th}$ item $=\left(\frac{n+1}{2}\right)^{th}$ item.

$=251^{th}$ item.

=16 approximately.

D_4=size of $\frac{4(n+1)}{10}^{th}$ item.

= ,, ,, $4\times\frac{502}{10}^{th}$ item. $=4\times50{\cdot}2^{th}$ item.

$=200{\cdot}8^{th}$ item.

=16 approximately.

H_5=size of $\frac{5(n+1)}{6}^{th}$ item.

= ,, $5\times\frac{502}{6}^{th}$ item. $=418{\cdot}3^{th}$ item.

=19 approximately.

O_3=size of $\frac{3(n+1)}{8}^{th}$ item. $=3\times\frac{502}{8}^{th}$ item.

$=188{\cdot}2^{th}$ item.

=16 approximately

O_7=size of $\frac{7(n+1)}{8}^{th}$ item.

= ,, 7×502^{th} item. $=439{\cdot}2^{th}$ item.

=20 approximately.

O_5=size of $\frac{5(n+1)}{8}^{th}$ item. $=\frac{2510th}{8}$ item.

$=313{\cdot}7^{th}$ item.

=17 approximately.

P_{15}=size of $\frac{15(n+1)}{100}^{th}$ item. $=15\times\frac{502}{100}^{th}$ item.

$=75{\cdot}3^{th}$ item.

=14 approximately.

P_{10}=size of $\frac{10(n+1)}{100}^{th}$ item. $=50{\cdot}2^{th}$ item,

=13 approximately.

$$P_{20} = \text{size of } \frac{20(n+1)}{100}^{th} \text{ item.}$$

$= 100{\cdot}4$ *th* item.

$= 15$ approx.

$$P_{40} = \text{size of } \frac{40(n+1)^{th}}{100} \text{ item.}$$

$= 200{\cdot}8$ th item.

$= 16$ approximately.

$$P_{60} = \text{size of } \frac{60(n+1)^{th}}{100} \text{ item.}$$

$= 301{\cdot}2$ th item.

$= 17$ approximately.

$$P_{90} = \text{size of } \frac{90(n+1)^{th}}{100} \text{ item.}$$

$= 451{\cdot}8$ th item.

$= 20$ approximately.

[Q=Quartile (4) ; D=Decile (10) ; H=Hexile (6) ; O=Octile (8) ; P=Percentile (100)].

Problem 49.—Amend the following table and locate the median from the amended table :

Sizes	*Frequencies*
10—15	10
15—17·5	15
17·5—20	17
22—30	25
30—35	28
35—40	30
45—Onwards	40

(Alld. B.Com., 1942)

Solution :

The series can be amended in different forms. One of these methods being its re-writing 10 as class-interval. To keep consistency the last unit, 45—onwards is taken within 40—50 group.

Sizes (x)	*Frequencies* (f)	*Cumulative freq.* (cmf)
10—20	(10+15+17)=42	42
20—30	25	67
30—40	58	125
40—50	40	165

$$Md = l_1 + \left\{ \frac{l_2 - l_1}{f_1} (m - c) \right\}$$

$\left[\text{where } m = \text{size of } \left(\frac{n+1}{2}\right)^{th} \text{ item.}\right.$

$= \quad ,, \quad \frac{166^{th}}{2}$ item.

$= 83$rd item.

$=$ it is located in the 30—40 group]

$= 30 + \left\{\frac{40-30}{58}\ (83-67)\right\}$

$= 30 + \frac{10}{58} \times 16$

$= 30 + 2{\cdot}75$ approx.

$= 32{\cdot}75$ approximately.

✓ **Problem 50.**—Which of the two places for which the mortality data are given below would you describe as more healthy than the other? ✓

Age in years	Town x (Standard)		Town y (Local)	
	Population	Deaths	Population	Deaths
Under 10	15,000	375	10,000	300
10—30	50,000	250	52,000	312
30—70	1,20,000	840	1,26,000	1,008
Above 70	15,000	975	12,000	840

(*Statistics—Ghosh and Choudhry*, p. 181)

Solution :

Age in years	Town x (Standard)			Town y (Local)		
	Population	Deaths	Deaths per 1000	Population	Deaths	Deaths per 1000
Under 10	15,000	375	25	10,000	300	30
10—30	50,000	250	5	52,000	312	6
30—70	1,20,000	840	7	1,26,000	1,008	8
Above 70	15,000	975	65	12,000	840	70
	n=2,00,000	2440		n=2,00,000	2460	

Hence A healthier.

(A) Crude death Rate of Town x :—

$= \frac{1}{200000}\ [(15000 \times 25) + (50000 \times 5) + (120000 \times 7) + (15000 \times 65)]$

For More accurate working, we find death 'rate'.

$= \frac{1}{200000} \times 2440000$

$= \frac{61}{5}$

$= 12{\cdot}2$ deaths per 1000 approx.

(B) Crude death Rate of Town y :—

$= \frac{1}{200000} [(10000 \times 30) + (52000 \times 6) + (126000 \times 8) + (12000 \times 70)]$

$= \frac{1}{200000} \times 2460000$

$= \frac{123}{10}$

$= 12{\cdot}3$ deaths per thousand approximately.

Therefore, the population of Town x is healthier.

Problem 51.—Compute the Crude and Standardized death rates in the following, and state if local population has higher or lower death rate.

Age group in years	Standard Population		Local Population	
	Population	Deaths	Population	Deaths
Under 5	6,000	150	2,500	63
5—15	10,000	20	12,500	25
15—65	12,500	50	20,000	80
Above 65	4,000	160	5,000	200

(*Statistics—Ghosh and Choudhry*, p. 181)

Solution :

Age group in years	Standard Population			Local Population		
	Population	Deaths	Deaths per 1000	Population	Deaths	Deaths per 1000
Under 5	6,000	150	25	2,500	63	25·2
5—15	10,000	20	2	12,500	25	2
15—65	12,500	50	4	20,000	80	4
Above 65	4,000	160	40	5,000	200	40
Total	32,500			40,000		

[A] Crude death rate of standard population :—

$$=\frac{1}{32500}\left[(6000\times25)+(1000\times2)+(12500+4)+(4000\times40)\right]$$

$$=\frac{1}{32500}\times380000$$

$$=\frac{152}{13}$$

$=11{\cdot}69$ deaths per 1000 approx.

[B.] Crude death rate of local population :—

$$=\frac{1}{40000}\left[(2500\times25{\cdot}2)+(12500\times2)+(20000\times4)+(5000\times40)\right]$$

$$=\frac{1}{40000}\times368000$$

$$=\frac{46}{5}$$

$=9{\cdot}2$ deaths per 1000 approximately.

[C] Standard death rate of local population :—

$$=\frac{1}{32500}\left[(6000\times25{\cdot}2)+(10000\times2)+(12500\times4)+(4000\times40)\right]$$

$$=\frac{1}{32500}\times451200$$

$$=\frac{4512}{325}$$

$=13{\cdot}8$ deaths per thousand approx.

[D] Standard death rate of standard population :—

$$=\frac{1}{40000}\left[(2500\times25)+(12500\times2)+(20000\times4)+(5000\times40)\right]$$

$$=\frac{1}{40000}\times367500$$

$$=\frac{147}{16}$$

$=9{\cdot}18$ deaths per thousand approx.

Therefore, standard population is healthier.

Problem 52.—Which of the two places for which the mortality data are given below would you describe as more healthy than another ?

Age	Town A		Town B	
	Numbers	Deaths	Numbers	Deaths
20—40 yrs.	5,10,000	2,050	2,95,000	1,000
40—60 ,,	2,45,000	2,550	3,40,000	2,750
Over 60 ,,	20,000	1,200	50,000	3,000
Total	7,75,000	5,800	6,85,000	6,750

(*Elementary Statistics—Dubey and Agarwal*, p. 114)

Solution :

Age	Town A			Town B		
	Numbers	Deaths	Death per 1000	Numbers	Deaths	Deaths per 1000
20—40yrs.	5,10,000	2,050	5	2,95,000	1,000	3
40—60 ,,	2,45,000	2,550	10	3,40,000	2,750	8
Over 60 ,,	20,000	1,200	60	50,000	3,000	60
Total	7,75,000	5,800		6,85,000	6,750	

[A] Crude death rate of Town A

$$=\frac{1}{775000}[(510000\times5)+(245000\times10)+(20000\times60)]$$

$$=\frac{1}{775000}\times6200000$$

$=8$ approx. per thousand.

[B] Crude death rate of Town B

$$=\frac{1}{685000}[(295000\times3)+(340000\times8)+(50000\times60)]$$

$$=\frac{1}{685000}\times6610000$$

$=9{\cdot}6$ deaths per thousand approx.

Therefore, A is healthier than B.

Problem 53.—The following table gives the monthly income of 24 families in a certain locality :

Serial No. of the family :	1	2	3	4	5	6	7	8	9	10	11	12
Monthly income in Rs. :	60	400	86	95	100	150	110	74	90	92	280	180
S. No. of the family :	13	14	15	16	17	18	19	20	21	22	23	24
Monthly income in Rs. :	96	98	104	75	80	94	100	75	600	82	200	84

Calculate the Arithmetic average, the median and the mode.

(*P.C.S., 1955*)

Solution :

Writing the series in ascending order :

S. No. :	1	2	3	4	5	6	7	8	9	10	11	12
Income :	60	74	75	75	80	82	84	86	90	92	94	95
S. No. :	13	14	15	16	17	18	19	20	21	22	23	24
Income :	96	98	100	100	104	110	150	180	200	280	400	600

Arithmetic Mean (M)

$= \frac{\Sigma x}{n} = \frac{3405}{24} =$ Rs. 141·8 approximately.

Median (Md)

$=$ Size of $\frac{(n+1)}{2}^{th}$ item

$=$,, ,, $\frac{(24+1)}{2}^{th}$ item

$=$,, ,, $\frac{25}{2}^{th}$ item

$=$,, ,, $12\cdot5^{th}$ item

$= \frac{\text{Size of } 12^{th} \text{ item} + \text{Size of } 13^{th} \text{ item}}{2}$

$=$ Rupees $\frac{95+96}{2}$ approximately

$=$ Rs. $\frac{191}{2}$

$=$ Rs. 95·8 approximately.

Mode (M_o)

$=$ By inspection is located at the highest freq. point.

Problem 54 —The following distribution below gives the cost of production of sugarcane in differenth oldings. Obtain the Arithmetic Mean.

Cost	*Frequency*	*Cost*	*Frequency*
2—6	1	18—	52
6—	9	22—	36
10—	21	26—	19
14—	47	30—34	3

(*I.A.S., 1941*)

Solution :

Cost	Mid-value (x)	Frequency (f)	ξ (x—A)	$f\xi$
2—6	4	1	−16	− 16
6—10	8	9	−12	−108
10—14	12	21	− 8	−168
14—18	16	47	− 4	−188
18—22	20	52	0	0
22—26	24	36	+ 4	+144
26—30	28	19	+ 8	+152
30—34	32	3	+12	+ 36
		$n=188$		$\Sigma f\xi=-148$

Let the assumed mean (A) = 20

$$\text{Arithmetic Mean (M)} = A + \frac{\Sigma f\xi}{n}$$

$$= 20 + \left[\frac{-148}{188}\right]$$

$$= 20 - \cdot 78 \text{ approx.}$$

= Rs. 19·22 approximately.

Problem 55.—Calculate the values of (*a*) the mode (*b*) the median and (*c*) the two quartiles from the following :

Wages in Rs.	*No. of workmen*	*Wages in Rs.*	*No of workmen*
20	8	24	20
21	10	25	25
22	11	26	15
23	16	27	9
		28—29	6

(*I.A.S., 1950*)

Solution :

Wages in Rs,	Mid-Value (x)	No. of workmen (f)	Cumulative Frequency (cmf)
20—21	20·5	8	8
21—22	21·5	10	18
22—23	22·5	11	29
23—24	23·5	16	45
24—25	24.5	20	65
25—26	25·5	25	90
26—27	26.5	15	105
27—28	27·5	9	114
28—29	28·5	6	120

$$\text{Mode } (M_0) = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}(l_2 - l_1)$$

$$= 25 + \frac{25-20}{50-20-15}(26-25)$$

$$= 25 + \frac{5}{15} = 25 + \cdot333 = 25\cdot333 \text{ approximately.}$$

$$\text{Median } (Md) = l_1 + \left\{ \frac{l_2 - l_1}{f_1}(m - c) \right\} \text{ where } m = \frac{120+1}{2} = 60\cdot5$$

$$= 60 \text{ approx.}$$

$$= 24 + \frac{1}{20}\ (60-45) = 24 + \frac{15}{20} = 24 + \cdot75$$

$$= 24\cdot75 \text{ approximately.}$$

$$Q_1 = l_1 + \left\{ \frac{l_2 - l_1}{f_1}\ (q_1 - c) \right\}$$

$$\text{where } q_1 = \frac{121}{4} = 30 \text{ approximately.}$$

$$= 23 + \frac{1}{16}\ (30-29)$$

$$= 23 + \frac{1}{16} \text{ approximately.}$$

$$= 23 + \cdot0625 \text{ approximately.}$$

$$= 23\cdot0625 \text{ approximately.}$$

$$Q_3 = l_1 + \left\{ \frac{l_2 - l_1}{f_1}(q_3 - c) \right\}$$

$$\left[\text{Where } q_3 = \frac{121 \times 3}{4} = \frac{363}{4} = 90 \text{ approx.}\right]$$

$$= 25 + \frac{1}{25}(90-65)$$

$$= 25 + \frac{25}{25}$$

$$= 26 \text{ approximately.}$$

Geometric and Harmonic Means

Article :—(A) Geometric mean G is given by :

(i) $$G = [X_1\ X_2\ X_3 \ldots\ldots X_n]^{\frac{1}{n}}$$

$$\text{or } \log G = \frac{1}{n}\ \log X_1 + \log X_2 + \ldots\ldots \log X_n]$$

$$\therefore\quad G = \text{Antilog}\,\frac{1}{n}\left[\log X_1 + \log X_2 + \ldots + \log X_n\right]$$

$$= \text{Anti log}\,\frac{1}{n}\,\Sigma \log X$$

(ii) $$G = \left[X_1^{f_1} \cdot X_2^{f_2} \cdot X_3^{f_3} \ldots X_n^{f_n}\right]^{\frac{1}{\Sigma f}}$$

$$\text{or } \log G = \frac{1}{\Sigma f}\left[f_1 \log X_1 + f_2 \log X_2 + \ldots + f_n \log X_n\right]$$

$$\therefore \quad G = \text{Antilog } \frac{1}{\Sigma f}\left[f_1 \log X_1 + f_2 \log X_2 + \ldots + f_n \log X_n\right]$$

$$\therefore \; G = \text{Antilog } \frac{1}{\Sigma f} \cdot \Sigma f \log X$$

(B) The Harmonic mean, H is given by

(i) $$\frac{1}{H} = \frac{1}{n}\left[\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_n}\right]$$
$$= \frac{1}{n}\, \Sigma \frac{1}{x}$$

(ii) $$\frac{1}{H} = \frac{1}{\Sigma f}\left[\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots\ldots + \frac{f_n}{x_n}\right]$$
$$= \frac{1}{\Sigma f}\, \Sigma \frac{f}{x}.$$

✓**Problem 56.** Compute the weighted geometric average of Relative prices of the following commodities for the year 1939 (Base year 1938—price 100) :

Commodity	*Relative price*	*Weight (value produced in 1938)*
Corn	128·8	1385
Cotton	62·4	819
Hay	117·7	842
Wheat	99·0	561
Oats	130·9	408
Potatoes	143·5	194
Sugar	125·6	142
Barley	150·2	100
Tobacco	101·1	103
Rye	116·2	25
Rice	117·5	17
Oilseeds	78·7	29

How does it differ from the unweighted geometric mean, and why ?

(B.com., Alld, 1943)

Solution :

Relative Price X	Weight (f)	log X	f log X
128·8	1385	2·1106	2923·1810
62·4	819	1·7952	1470·2688
117·7	842	2·0119	1764·5398
99·0	561	1·9956	1119·5316
130·9	408	2·1173	863·8584
143·5	194	2·1553	418·1282
125·6	142	2·1004	298·2568
150·2	100	2·1761	217·6100
101·1	103	2·0043	206·4429
116·2	25	2·0645	51·6125
117·5	17	2·0719	35·2223
78·7	29	1·8960	54·9840
TOTAL	$\Sigma f=4625$	$\Sigma \log X=24·5581$	9423·6263

(*i*) Weighted geometric mean.

$$G=\text{Antilog } \frac{\Sigma f \log X}{\Sigma f}$$

$$=\text{Antilog } \frac{9423·6263}{4625}$$

$$=\text{Antilog } 2·0375$$

$$=109·14$$

(*ii*) Simple geometric average.

$$G=\text{Antilog } \frac{\Sigma \log X}{n}$$

$$=\text{Antilog } \frac{24·5581}{12}$$

$$=\text{Antilog } 2·0465$$

$$=111·3$$

Problem 57. Find the Harmonic mean for the following distribution.

Class	*Frequency*
40—50	19
50—60	25
60—70	36
70—80	72
80—90	51
90—100	43

Solution :

Class	Mid value X	Frequency (f)	$\frac{1}{X}$	$\frac{f}{X}$
40—50	45	19	·0222	·4218
50—60	55	25	·0182	·4550
60—70	65	36	·0154	·5544
70—80	75	72	·0133	·9576
80—90	85	51	·0118	·6018
90—100	95	43	·0105	·4515
	TOTAL	$\Sigma f = 246$	TOTAL	$\Sigma \frac{f}{X} = 3{\cdot}4421$

Harmonic mean H is given by

$$\frac{1}{H} = \frac{\Sigma \frac{f}{X}}{\Sigma f}$$

$$= \frac{3{\cdot}4421}{246}$$

$$\therefore H = \frac{246}{3{\cdot}4421}$$

$$= 71{\cdot}4$$

✓**Problem 58.** The annual income of 14 families are given below in Rupees:

180, 250, 490, 120, 1400, 7000, 1050, 150, 360, 100, 80, 200, 500, 240.

Calculate the geometric and Harmonic means.

Solution :

Annual income in Rupees X	Log X	$\frac{1}{X}$
180	2.2553	·00555
250	2·3979	·00400
490	2·6902	·00204
120	2·0792	·00833
1,400	3·1461	·00071
7,000	3·8451	·00014
1,050	3·0212	·00095
150	2·1761	·00666
360	2·5563	·00279
100	2·0000	·00100
80	1·9031	·01250
200	2·3010	·00500
500	2.6990	·00200
240	2.3802	·00417
TOTAL	Σ logX=35·4507	$\Sigma \frac{1}{X}$ =·05584

Geometric mean G is given by

$$G = \text{Antilog}\ \frac{\Sigma \log X}{n}$$

$$= \text{Antilog}\ \frac{35{\cdot}4507}{14}$$

$$= \text{Antilog}\ 2{\cdot}5322$$

$$= 340{\cdot}6$$

Harmonic mean H is given by

$$\frac{1}{H} = \frac{\Sigma \frac{1}{X}}{n}$$

$$= \frac{{\cdot}05584}{14}$$

$$= {\cdot}00399$$

$$\therefore\ H = \frac{1}{{\cdot}00399}$$

$$= 250{\cdot}6$$

$\therefore$ Geometric mean $= 340{\cdot}6$

Harmonic Mean $= 250{\cdot}6$

Problem 59. Calculate the geometric and harmonic means for the following distribution.

Weight of boys in lbs.	Number of boys
100—104	24
105—109	30
110—114	45
115—119	65
120—124	72
125—129	84
130—134	124
135—139	58
140—144	22
145—149	32
150—154	8
TOTAL	564

Soultion :

Weight of boys in lbs.	Mid-Value (X)	Number of boys (f)	log X	f log X	$\frac{f}{X}$
100—104	102	24	2·0086	48·2064	·2353
105—109	107	30	2·0294	60·8820	·2804
110—114	112	45	2·0492	92·2140	·4018
115—119	117	65	2·0682	134·4330	·5555
120—124	122	72	2·0804	149·7888	·5901
125—129	127	84	2·1038	176·7192	·6614
130—134	132	124	2·1206	262·9544	·9393
135—139	137	58	2·1376	123·9808	·4233
140—144	142	22	2·1523	47·3506	·1549
145—149	147	32	2·1673	69·3536	·2176
150—154	52	8	2·1818	17·4544	·0526
	TOTAL	$\Sigma f = 564$	TOTAL	1183·3372	4·5122

$$\therefore G = \text{Antilog} \frac{\Sigma f \log X}{\Sigma f}$$

$$= \text{Antilog} \frac{1183 \cdot 3372}{564}$$

$$= \text{Antilog } 2 \cdot 0981$$

$$= 125 \cdot 3$$

$$\frac{1}{H} = \frac{\Sigma \frac{f}{X}}{\Sigma f}$$

$$= \frac{4.5122}{564}$$

$$\therefore H = \frac{564}{4 \cdot 5122}$$

$$= 124 \cdot 9 \text{ approx.}$$

CHAPTER II

(DISPERSION)

"Two distributions of statistical data may be symmetrical and have common means, medians and modes, and identical frequencies in the modal class. Yet, with these points in common they may differ widely in the scatter of their values about the measures of central tendency"*.

"Dispersion or scatter variation or variability is relative to any typical value and is a measure of the extent to which the individual itmes vary. If the scatter about the measure of central tendency is very large, it is of little use as a typical value...... Measures of dispersion are also called, Averages of the Second Order."†

Various measures of Dispersion are—

(i) *The Range*

(ii) *Quartile Deviation or Semi-interquartile Range.*

(iii) *Mean Deviation.*

(iv) *Standard Deviation.*

The Range.

The range is the difference between the maximum and minimum item values in the series.

Example.

Problem 60. Find the Range of the following series .

30, 35, 40, 70, 92, 54, 68.

Range $=(M_a-M_i)$

$=92-30$ $\qquad$ M_a = Maximum item value

$=62$ $\qquad$ M_i = Minimum item value.

Quartile Deviation or Semi-Interquartile Range.

"When the items have been arranged according to their magnitude, the inter-quartile range is the range within which the middle half of the values fall. The extreme values do not influence the interquartile range as they do the range measured by the absolute difference between the highest and lowest values in a series."‡

The customary formula which is generally applied in the computation of Quartile deviation or interquartile range is :

$$Q=\frac{Q_3-Q_1}{2}$$ where Q = Quartile Deviation

---

* *Elementary Statistical Methods*—Neiswanger. p. 312.

† *Practical Statistics*—M. Zia-ud-din. p. 70.

‡ *Elementary Statistical Methods* Neiswanger. p. 315.

Q_1=First Quartile
Q_3=Third Quartile.

and, the Quartile Co-efficient is calculated by the formula :

$$\text{Quartile Co-efficient}=\frac{\dfrac{Q_3-Q_1}{2}}{\dfrac{Q_3+Q_1}{2}}$$

$$=\frac{Q_3-Q_1}{Q_3+Q_1}$$

Example.

Problem 61. Calculate the semi inter-quartile range and quartile co-efficient of :

Age	No. of Members	Cumulative frequency
20	3	3
30	61	64
40	132	196
50	153	349
60	140	489
70	51	540
80	2	542
	542	

(*Nag., B.Com., 1942*)

Q_1=age of $\left(\frac{n+1}{4}\right)^{th}$ item.

=age of 136^{th} item.

=50

Q_3=age of $\left(\frac{3n+1}{4}\right)^{th}$ item.

Q_3=age of 408^{th} item.

=60.

$\therefore$ Semi-interquartile range$=\frac{Q_3-Q_1}{2}$

$=\frac{60-50}{2}$

=5. **Ans.** [Q. D.=Quartile Deviation.]

$\therefore$ Co-eff. of Q. D. $=\frac{Q_3-Q_1}{Q_3+Q_1}=\frac{60-50}{60+50}=\frac{10}{110}$

=·0909 approximately.

Mean Deviation or Average Deviation.

Mean or average deviation is the 'average of the deviations of the items from the Median, Mean or Mode.' *All the deviation in this calculation is taken positively* and represented symbolically as $|d|$.

Symbolically,

Mean Deviation (δ_m) from the Mean $= \frac{\Sigma|d_m|}{n}$

Mean Coefficient Dispersion ($C\delta_m$) from the Mean $= \frac{\frac{\Sigma|d_m|}{n}}{M}$

Mean Deviation from the Median (δ_{md}) $= \frac{\Sigma|d_{md}|}{n}$

n = number of items.
M = Mean
$|d|$ = deviations omitting negative signs.

Mean Coefficient Dispersion from the Median ($C\delta_{md}$)

$= \frac{\delta_{md}}{M_d}$ M_d = Median.

Mean Deviation from the Mode (δ_{m0}) $= \frac{\Sigma|d_{m0}|}{n}$, M_0 = Mode

Mean Co-efficient Dispersion from the Mode

$(C\delta_{m0}) = \frac{\delta_{m0}}{M_0}$

Mean Deviation for grouped data and discrete

series $= \frac{\Sigma f\,|d|}{n}$

Example.

Problem 62. Calculate (A) Mean Deviation and (B) Mean Co-efficient Dispersion from the Median.

Earnings	f	$\lvert d_{md}\rvert$*
$25·00	1	0·00
$22·00	1	3·00
$30·00	1	5·00
$18·00	1	7.00
$28·00	1	3·00
$123·00	5	18·00

$M_d = 25{\cdot}00$

$\delta_{md} = \frac{\Sigma|d_{md}|}{n}$

$= \frac{18}{5}$

$= 3{\cdot}60$

$C\delta_{md} = \frac{\delta_{md}}{md} = \frac{3{\cdot}60}{25{\cdot}00}$

$= {\cdot}144$ **Ans.**

If the series is a discrete one, or a grouped data, then $|d|$ is multiplied by the f and, finally, $\Sigma f\,|d|$ is calculated. On the basis of it the value of mean deviation or Mean Coefficient Dispersion is calculated:

* *Elementary Statistical Method*—Neiswanger. p. 324.

Example.

Problem 63. Calculate the Mean Deviation from the Median for the following : (*Practical Statistics*—Zia-ud-din. p. 71)

Class-interval	Frequencies	Class-interval	Frequencies
2—4	3	6—8	2
4—6	4	8—10	1

Solution :

$$\text{Median} = 4 + \tfrac{2}{4}\left(\tfrac{10}{2} - 3\right) = 5$$

Class-intervals	Central Values	$\lvert d_{md} \rvert$	f	$f\lvert d_{md} \rvert$
2—4	3	2	3	6
4—6	5	0	4	0
6—8	7	2	2	4
8—10	9	4	1	4
			$n=10$	$\Sigma f\lvert d_{md}\rvert = 14$

$$\text{Mean Deviation } (\delta_{md}) = \frac{\Sigma f \lvert d_{md} \rvert}{n} = \frac{14}{10} = 1{\cdot}4.$$

∴ Mean Co-efficient Dispersion from the Median $(C\delta_{md}) = \dfrac{\delta_{md}}{M_d} = \dfrac{1{\cdot}4}{5} = {\cdot}28$

Mean Deviation by Short-cut Method.*

Example.

Problem 64. Computation of Mean Deviation from Assumed Mean.

Size X	Frequency (f)	Dev. from Assumed Mean d_a	$\lvert d_a \rvert$	$f\lvert d_a \rvert$	fd_a (Signs considered)
4	2	−6	6	12	−12
6	1	−4	4	4	− 4
8	3	−2	2	6	− 6
10	6	0	0	0	0
12	4	+2	2	8	+ 8
14	3	+4	4	12	+12
16	1	+6	6	6	+ 6
Assumed Mean (A)=10	$n=20$			$\Sigma f\lvert d_a \rvert = 48$	$\Sigma fd_a = 4$

* *Elements of Statistics* by Profs. D. K. Sakhwalkar and M.P. Singh. pp. 205—206. (Symbols etc. slightly changed).

Actual Mean (M) $=10+\frac{4}{20}=10 \cdot 2$

$\therefore$ Mean Deviation from (δ_k) $=\frac{\Sigma f|d_a|}{n}=\frac{48}{20}=2 \cdot 4$

Difference between the two means $=(10 \cdot 2-10)=\cdot 2$

Difference between the No. of frequencies below and above Mean $=(12-8)=4$

$\therefore$ the correction to the Mean Deviation from Assumed Mean (Total Error) $=\cdot 2\times 4=\cdot 8$

Mean Deviation from actual

$$\text{Mean } (\delta = \frac{\Sigma f|d_a|+\text{Total Error}}{n}$$

$$=\frac{48}{20}=2 \cdot 44$$

This method is called the short-cut method used for the calculation of mean deviation. Its formula is $(\delta_a)=\frac{\Sigma|d_a|+\text{Total Error}}{n}$

Standard Division.

(i) The standard deviation is always calculated from the Arithmetic Mean. The customary formula of it for un-grouped data is :

Standard Deviation $(\sigma)=\sqrt{\frac{\Sigma d^2}{n}}$ d= deviations of the items from the mean.

n= number of items.

For grouped data, we have to find out fd^2 and on the basis of it, the formula becomes :

$$\sigma=\sqrt{\frac{\Sigma fd^2}{n}}$$

Many authors take ξ for d indicating deviations from the Mean.

(ii) *Standard Deviation by short-cut method.*—"The short-cut method avoids the labour of finding the Arithmetic Mean. This consists in taking convenient Provisional Mean or Working Mean and employing the formula. *It is customary to take a Provisional Mean corresponding to the greatest frequency*".

$$\sigma=\sqrt{\frac{\Sigma f(D)^2}{n}-\left(\frac{\Sigma fD}{n}\right)^2}$$

Where D= deviations of the central values from the Provisional Mean.

As usual, the D^2 is not to be multiplied by f if it is an individual series. In that case, the formula becomes :

$$\sigma = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2}$$

(iii) It can also be expressed in the following forms :

(a) $\sigma = h\sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2}$

where h=class-difference

$$u = \frac{X - A}{h}$$

A=Assured Mean and X=size of the item.

(b) $\sigma = \sqrt{\frac{\Sigma f\xi^2}{n} - \left(\frac{\Sigma f\xi}{n}\right)^2}$

Where $\xi = X - A$.

(iv) *Variance* — The square of the Standard Deviation is called Variance, and generally represented by σ^2.

Relative Measures of Dispersion.

*To relate the measure of dispersion to its average and to convert it to percentage form, the standard deviation is divided by Arithmetic Mean. This measure is known as *Co-efficient of variation* and is given by.

$$C.\ V. = \frac{100\ \sigma}{M}$$ C. V.=Co-efficient of variation

M=Mean

σ=Standard Deviation.

and is generally used for comparison of consistency of two or more quantities.

The ratio $\frac{\sigma}{\text{Mean}}$ is called the *Co-efficient of Standard Deviation.*

Some other comparative co-efficients of dispersion are given as follows: —

Quartile Co-efficient of Dispersion $= \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$

Mean Co-efficient of Dispersion $= \frac{\text{Mean Deviation} \times 100}{\text{Median (or Arith. Mean, if used)}}$.

Problem 65. Find out the standard deviation for the following frequency distribution :—

Variable	5	15	25	35	45	55	65	75
Frequency	3	7	9	23	15	8	6	4.

* *Practical Statistics* by Zia-ud-din. p. 74.

Solution :

Take A=35, h=10. (class diff. 15—5=10) (A=Assumed means)

X	f	$u=\frac{X-A}{h}$	fu	fu^2
5	3	—3	— 9	27
15	7	—2	—14	28
25	9	—1	— 9	9
35	23	0	0	0
45	15	1	15	15
55	8	2	16	32
65	6	3	18	54
75	4	4	16	64
TOTAL	$n=75$		$\Sigma fu=33$	$\Sigma fu^2=229$

∴ Standard deviation $\sigma=h\sqrt{\frac{\Sigma fu^2}{\Sigma f}-\left(\frac{\Sigma fu}{\Sigma f}\right)^2}$

$$=10\sqrt{\frac{229}{75}-\left(\frac{33}{75}\right)^2}$$

$$=\tfrac{10}{75}\sqrt{229\times75-33\times33}$$

$$=\frac{10\times126\cdot8}{75}$$

$$=\frac{1268}{75}$$

$$=17 \text{ appr.}$$

Problem 66. Find the standard deviation for the following frequency distribution.

Variable	2	3	4	5	9	10	12	13	15
Frequency	25	37	44	59	68	43	31	23	12.

Solution :

ξ=X—A where A=9 (assumed average)

X	f	ξ (X—a)	$f\xi$	$f\xi^2$
2	25	—7	—175	1225
3	37	—6	—222	1332
4	44	—5	—220	1100
5	59	—4	—236	944
9	68	0	0	0
10	43	1	43	43
12	31	3	93	279
13	23	4	92	368
15	12	6	72	432
TOTAL	342		—553	5723

Substituting the values from table we have

$$\sigma = \sqrt{\frac{\Sigma f \xi^2}{n} - \left(\frac{\Sigma f \xi}{n}\right)^2}$$

$$\text{or } \sigma^2 = \frac{5723}{342} - \left(\frac{-553}{342}\right)^2$$

$$\text{or } \sigma = 3.76.$$

Problem 67. Find out standard deviation and co-efficient of skewness for the following distribution :

Variable	*Frequency*
0—5	2
5—10	5
10—15	7
15—20	13
20—25	21
25—30	16
30—35	8
35—40	3

Solution :

In the calculation of the mean and the standard deviation we take for each class its mid-value as the value of X. The calculations for finding them are shown in the following table:

X	f	$U = \frac{X - 22\cdot5}{5}$	fu	fu^2
2·5	2	−4	−8	32
7·5	5	−3	−15	45
12·5	7	−2	−14	28
17·5	13	−1	−13	13
22·5	21	0	0	0
27·5	16	1	16	16
32·5	8	2	16	32
37·5	3	3	9	27
TOTAL	75		−9	193

Here A=22·5, (assumed average), h=5 (class-difference)

$$\text{Mean} = 22\cdot5 - \frac{9}{75} \times 5 \left(\text{Formula, } M = A + h \frac{\Sigma fu}{n}\right)$$

$$= 21\cdot9 \text{ appr.}$$

$$\text{and } \sigma = h\sqrt{\frac{\Sigma f u^2}{n} - \left(\frac{\Sigma fu}{n}\right)^2}$$

$$= 5 \times \sqrt{\frac{193}{75} - \left(\frac{-9}{75}\right)^2}$$

$$= 8\cdot0 \text{ appr.}$$

For mode referring the original table

$$M_0 = 20 + \frac{16}{16+13} \times 5$$

$$= 22\cdot76.$$

$$\therefore \ Sk = \frac{M - M_0}{6}$$

$$= \frac{21 \cdot 9 - 22 \cdot 76}{8}$$

$$= -\cdot 1 \text{ appr.}$$

Problem 68. Find the mean yield of paddy and the standard deviation for the distribution of the results of 3061 crop-cutting experiments shown in the following table :—

Yield of paddy per acre in lbs.	*No. of Experiments*
0 — 400	236
401 — 800	481
801 — 1200	604
1201 — 1600	576
1601 — 2000	419
2001 — 2400	333
2401 — 2800	217
2801 — 3200	87
3201 — 3600	64
3601 — 4000	23
4001 — 4400	14
4401 — 4800	6
4801 — 5200	1
TOTAL	3061

(*B. Com. Bombay, 1945*)

Solution :

We take mid value of each class as X., A=1800·5, h=400.
(assumed average)(class-interval)

X	f	$U = \frac{X-A}{h}$	fu	fu^2
200·5	236	−4	− 944	3776
600·5	481	−3	−1443	4329
1000·5	604	−2	−1208	2416
1400·5	576	−1	− 576	576
1800·5	419	0	0	0
2200·4	333	1	333	333
2600·5	217	2	434	868
3000·5	87	3	261	783
3400·5	64	4	256	1024
3800·5	23	5	115	575
4200·5	14	6	84	504
4600·5	6	7	42	294
5000·5	1	8	8	64
TOTAL	3061		−2638	15542

$$\text{Mean yield of paddy} = A + h\frac{\Sigma f u}{n}$$

$$= 1800{\cdot}5 - \frac{2638 \times 400}{3061}$$

$$= 1455{\cdot}8 \text{ lbs. per acre.}$$

Standard deviation of the yield paddy for 3061 crop cutting experiments.

$$\sigma = h \sqrt{\frac{\Sigma f u^2}{n} - \left(\frac{\Sigma f u}{n}\right)^2}$$

$$\text{or } \sigma = 400 \times \sqrt{\frac{15542}{3061} - \left(\frac{-2638}{3061}\right)^2}$$

$$= \sqrt{4{\cdot}34} \times 400$$

$$= 833{\cdot}2 \text{ lbs.}$$

Problem 69. The fluctuations in the rates of Kohinoor and Tata Deferred on the 7^{th} march are given below. Find out which of the two shares shows greater variability.

Kohinoor—618, 619, 616, 623, 620, 624, 622, 625, 622, 625, 626, 625.

Tata Deferred—$2152\frac{1}{2}$, $2132\frac{1}{2}$, $2134\frac{1}{4}$, $2132\frac{1}{2}$, 2145, $2142\frac{1}{2}$, $2146\frac{1}{4}$, 2130, $2146\frac{1}{4}$, $2142\frac{1}{2}$, 2150, 2135, $2152\frac{1}{2}$.

(*Bombay, 1945*)

Solution :

Calculation of standard deviation.

Kohinoor			Tata Deferred		
Prices of the security X	A=(620) ξ=X−A	ξ^2	Prices of the security	ξ=X−A =X−2142·5	ξ^2
618	−2	4	2152·5	10	100
619	−1	1	2132·5	−10	100
616	−4	16	2134·25	−8·25	68·06
623	3	9	2132·5	−10	100
620	0	0	2145	2·5	6·25
624	4	16	2142·5	0	0
622	2	4	2146·25	3·75	14·06
625	5	25	2130	−12·5	156·25
622	2	4	2146·25	3·75	14·06
625	5	25	2142·5	0	0
626	6	36	2150	7·5	56·25
625	5	25	2135	−7·5	56·25
			2152·5	10	100
n=12 TOTAL	25	165	n=13	−10·75	771·18

Kohinoor—

Arithmetic average or $M = A + \frac{\Sigma\xi}{n}$

$$= 620 + \frac{25}{12}$$

$$= 622 \cdot 1$$

Standard deviation $\sigma = \sqrt{\frac{\Sigma\xi^2}{n} - \left(\frac{\Sigma\xi}{n}\right)^2}$

$$= \sqrt{\frac{165}{12} - \left(\frac{25}{12}\right)^2}$$

$$= \sqrt{13 \cdot 75 - 4 \cdot 5}$$

$$= 3 \cdot 07.$$

Coefficient of variation $= \frac{\sigma}{M} \times 100$

$$= \frac{3 \cdot 07}{622 \cdot 1} \times 100$$

$$= \cdot 493$$

Tata Deferred—

$$M = A + \frac{\Sigma\xi}{n}$$

$$= 2142 \cdot 5 + \frac{-10 \cdot 75}{13}$$

$$= 2141 \cdot 67$$

$$\sigma = \sqrt{\frac{771 \cdot 8}{13} - \left(\frac{-10 \cdot 75}{13}\right)^2}$$

$$= \sqrt{58 \cdot 64}$$

$$= 7 \cdot 66$$

Co-efficient of variation $= \frac{\sigma}{M} \times 100$

$$= \frac{7 \cdot 66}{2141 \cdot 67} \times 100$$

$$= \cdot 358.$$

On comparison of the co-efficients of variation, it is obvious that Kohinoor shows greater variability.

Problem 70.—The following table gives the frequency distribution of area under wheat in a sample of 282 villages in Meerut District during 1936-37. Calculate :—

(*a*) The standard deviation, and

(*b*) The semi-inter-quartile range of the distribution.

Bighas under wheat	*Frequency*	*Bighas under wheat*	*Frequency*
0—100	3	1100—	14
100—	7	1200—	14
200—	10	1300—	16
300—	17	1400—	8
400—	33	1500—	8
500—	29	1600—	6
600—	27	1700—	5
700—	21	1800—	2
800—	23	1900—2000	1
900—	20		
1000—	18		
contd.	contd.		

(*I. A. S., 1948*)

Solution : We take mid-value of each class as X.

let A=850, h=100 (class interval)
(assumed average)

Bighas under wheat	x	f	Cumulative frequency	$u=\frac{X-A}{h}$	fu	fu^2
0—100	50	3	3	—8	—24	192
100—	150	7	10	—7	—49	343
200—	250	10	20	—6	—60	360
300—	350	17	37	—5	—85	425
400—	450	33	70	—4	—132	528
500—	550	29	99	—3	—87	261
600—	650	27	126	—2	—54	108
700—	750	21	147	—1	—21	21
800—	850	23	170	0	0	0
900—	950	20	190	1	20	20
1000—	1050	18	208	2	36	72
1100—	1150	14	222	3	42	126
1200—	1250	14	236	4	56	224
1300—	1350	16	252	5	80	400
1400—	1450	8	260	6	48	288
1500—	1550	8	268	7	56	392
1600—	1650	6	274	8	48	384
1700—	1750	5	279	9	45	405
1800—	1850	2	281	10	20	200
1900—2000	1950	1	282	11	11	121
	TOTAL	282		TOTAL	—50	4870

Standard deviation $\sigma = h \cdot \sqrt{\frac{\Sigma f u^2}{n} - \left(\frac{\Sigma f u}{n}\right)^2}$

$$= 100\sqrt{\frac{4870}{282} - \left(\frac{-50}{282}\right)^2}$$

$$=100\sqrt{17{\cdot}27-{\cdot}03}$$

$$=100\times\sqrt{17{\cdot}24}$$

$$=415 \text{ approx.}$$

(*b*) For semi-inter-quartile range.

First quartile

$$Q_1=L+\frac{\frac{1}{4}n-F}{f}\times i$$

$$=500+\frac{70{\cdot}5-70}{29}\times 100$$

$$=500+\frac{50}{29}$$

$$=501{\cdot}72.$$

$$Q_3=L+\frac{\frac{3}{4}n-F}{5}\times i$$

$$=1100+\frac{211{\cdot}5-208}{14}\times 100$$

$$=1100+\frac{350}{14}$$

$$=1125.$$

$$\therefore \text{ Semi-inter-quartile range}=\frac{Q_3-Q_1}{2}$$

$$=\frac{1125-501{\cdot}72}{2}$$

$$=311{\cdot}64$$

$$\text{Coefficient of the S. I. Range}=\frac{Q_3-Q_1}{Q_3+Q_1}$$

$$=\frac{623{\cdot}28}{1626{\cdot}72}$$

$$={\cdot}38$$

Problem 71.—Goals scored by two teams A and B in a football season were as follows :—

No. of goals scored in a Match	No. of Matches A	No. of Matches B
0	27	17
1	9	9
2	8	6
3	5	5
4	4	3

By calculating the coefficient of variation in each case, find which team may be considered the more consistent.

(*I.A.S., 1954*)

Solution :

$$M = \frac{0+1+2+3+4}{5} = 2 \text{ ; M = arithmetic average.}$$

X	X—M	team A f	team A $f(X-M)^2$	team B f	team B $f(X-M)^2$
0	−2	27	108	17	68
1	−1	9	9	9	9
2	0	8	0	6	0
3	1	5	5	5	5
4	2	4	16	3	12
		53	138	40	94

$$\text{For team A } \sigma = \sqrt{\frac{\Sigma f(X-M)^2}{n}}$$

$$= \sqrt{\frac{138}{53}} = 1{\cdot}6$$

$$\therefore \text{ Coefficient of variation} = \frac{\sigma}{M} \times 100$$

$$= \frac{1{\cdot}6}{2} \times 100 = 80$$

$$\text{For team B } \sigma = \sqrt{\frac{94}{40}}$$

$$= 1{\cdot}53$$

$$\therefore \text{ Coefficient of variation} = \frac{\sigma}{M} \times 100$$

$$= \frac{1{\cdot}53 \times 100}{2}$$

$$= 76{\cdot}5$$

Hence the team **B** is more consistent.

Problem 72. The following table gives the yield of paddy in maunds per acre based on crop cutting experiments in a certain area during 1940-41.

Yields in mds. per acre	*Frequency*	*Yield in mds. per acre*	*Frequency*
0	4	24	128
3	4	27	73
6	32	30	50
9	81	33	13
12	135	36	12
15	198	39	5
18	210	42	1
21	144		
continued	continued		

Calculate the arithmetic mean, the standard deviation and median of the distribution.

(*I.A.S.*, *1949*)

Solution : Let X represents yield in mds. per acre.

X	f	Cumulative frequency	$u=\frac{X-18}{3}$	fu	fu^2
0	4	4	−6	−24	144
3	4	8	−5	−20	100
6	32	40	−4	−128	512
9	81	121	−3	−243	729
12	135	256	−2	−270	540
15	198	454	−1	−198	198
18	210	664	0	0	0
21	144	808	1	144	144
24	128	936	2	256	512
27	73	1009	3	219	657
30	50	1059	4	200	800
33	13	1072	5	65	325
36	12	1084	6	72	432
39	5	1089	7	35	245
42	1	1090	8	8	64
Total	1090			$\Sigma fu=116$	$\Sigma fu^2=5402$

Arithmetic mean $M=A+h\,\frac{\Sigma fu}{\Sigma f}$

$$=18+3\times\frac{116}{1090}$$

$$=21{\cdot}19.$$

Standard deviation $\sigma=h\,.\sqrt{\frac{\Sigma fu^2}{\Sigma f}-\left(\frac{\Sigma fu}{\Sigma f}\right)^2}$

$$=3\times\sqrt{\frac{5402}{1090}-\left(\frac{116}{1090}\right)^2}$$

$$=5{\cdot}85.$$

Clearly in this case 'Median' is the measure of

$$\frac{\frac{n}{2}^{th}\text{ item}+\left(\frac{n}{2}+1\right)^{th}\text{ item}}{2}$$ where n is total frequency

i.e., of $\frac{545^{th}\text{ item}+546^{th}\text{ item}}{2}$

which is $\frac{18+18}{2}=18.$

Hence $M_d=18.$

✓**Problem 73.**—The following tables give the frequency distribution of expenditure on food per family per month among working class families in two localities. Find the arithmetic average and the standard deviation of the expenditure at both places.

Range of expenditure in Rs. per month		*No. of families*	
		Place A	*Place B*
Rs.	3— 6	28	39
,,	6— 9	292	284
,,	9—12	389	401
,,	12—15	212	202
,,	15—18	59	48
,,	18—21	18	21
,,	21—24	2	5

(*P.C.S., 1941*)

Solution :

Mid-Value X	$u=\frac{x-10\cdot5}{3}$	Place A			Place B		
		f	fu	fu^2	f	fu	fu^2
4·5	−2	28	−56	112	39	−78	156
7·5	−1	292	−292	292	284	−284	284
10·5	0	389	0	0	401	0	0
13·5	1	212	212	212	202	202	202
16·5	2	59	118	236	48	96	192
19·5	3	18	54	162	21	63	189
22·5	4	2	8	32	5	20	80
		1000	44	1046	1000	19	1103

Place A.

$$\text{Arithmetic average}=A+h\,\frac{\Sigma fu}{\Sigma f}$$

$$=10\cdot5+\frac{3\times44}{1000}$$

$$=10\cdot632.$$

$$\text{Standard deviation}=h\sqrt{\frac{\Sigma fu^2}{\Sigma f}-\left(\frac{\Sigma fu}{\Sigma f}\right)^2}$$

$$=3\times\sqrt{\frac{1046}{1000}-\left(\frac{44}{1000}\right)^2}$$

$$=3\times\sqrt{1\cdot046-\cdot0174}$$

$$=3\times\sqrt{1\cdot0286}$$

$$=3\cdot03.$$

Place B.

$$\text{Arithmetic average}=10\cdot5+\frac{3\times19}{1000}$$

$$=10\cdot557.$$

$$\text{Standard deviation} = 3 \times \sqrt{\frac{1103}{1000} - \left(\frac{19}{1000}\right)^2}$$

$$= 3 \times \sqrt{1 \cdot 103 - \cdot 000361}$$

$$= 3 \times \sqrt{1 \cdot 102639}$$

$$= 3 \cdot 15$$

Problem 74.—The values of the arithmetic mean and standard deviation of the following frequency distribution of a continuous variable derived by using both arbitrary working origin and scale are 135·3 lbs. and 9·6 lbs. respectively. Determine the actual class intervals. (I. A. S. 1952)

f	u	fu	fu^2
2	−4	−8	32
5	−3	−15	45
8	−2	−16	32
18	−1	−18	18
22	0	0	0
13	1	13	13
8	2	16	32
4	3	12	36
80		−16	208

Solution :

If $u = \frac{X - A}{h}$, then

$$\text{Mean} = A + h \frac{\Sigma f u}{\Sigma f} = 135 \cdot 3,$$

and $\sigma^2 = h^2 \frac{\Sigma f u^2}{\Sigma f} - h^2 \left(\frac{\Sigma f u}{\Sigma f}\right)^2 = (9 \cdot 6)^2$

Substituting the values from the table we get

$$A - \frac{16h}{80} = 135 \cdot 3$$

$$\frac{208h^2}{80} - \frac{16^2 h^2}{80^2} = (9 \cdot 6)^2$$

which give $A = 136 \cdot 5$ and $h = 6$

If x_1 and x_2 are class limits for the class where the arbitrary origin lies, we have

$$\frac{x_1 + x_2}{2} = 136 \cdot 5 \text{ and } x_2 - x_1 = h = 6$$

$$\therefore \quad x_1 = 133 \cdot 5, \; x_2 = 139 \cdot 5$$

Thus this class is 133·5—139·5. The class interval is 6 and hence the classes are

$$(109{\cdot}5-115{\cdot}5), \ldots\ldots\ldots, (151{\cdot}5-157{\cdot}5)$$

Problem 75.—Calculate the standard deviation of the following two series, which shows greater deviation.

Series A	*Series B*	*Series A*	*Series B*
192	83	260	126
288	87	348	126
236	93	291	101
229	109	330	102
184	124	243	108
—continued	—continued		

(*R.C.S.*, 1938)

Solution :

Series A			*Series B*		
Size of the item X	Deviation from arith. average X−M	(X−M)²	X	X−M	(X−M)²
192	−68·1	4638	83	−22·9	524·4
288	27·9	778·4	87	−18·9	357·2
236	−24·1	580·8	93	−12·9	166·4
229	−31·1	967·2	109	3·1	9·61
184	−76·1	5791	124	18·1	327·6
260	−·1	·01	126	20·1	404·0
348	87·9	7726	126	20·1	404·0
291	30·9	954·8	101	−4·9	24·01
330	69·9	4886	102	−3·9	15·21
243	−17·1	292·4	108	·1	4·41
ΣX=2601 n=10		26614·61	ΣX=1059		2236·84

For series A,

$$\text{Arithmetic average } M = \frac{\Sigma X}{n} = 260{\cdot}1$$

For series B,

$$\text{Arithmetic average } M = \frac{1059}{10} = 105{\cdot}9$$

For series A,

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma (X-M)^2}{n}}$$

$$= \sqrt{\frac{26614{\cdot}61}{10}}$$

$$= 51{\cdot}6.$$

∴ Standard coefficient of deviation

$$=\frac{\sigma}{M}$$

$$=\frac{51{\cdot}6}{260{\cdot}1}$$

$$=\cdot198$$

For series B,

$$\text{Standard deviation } \sigma=\sqrt{\frac{\Sigma(X-M)^2}{n}}$$

$$=\sqrt{\frac{2236{\cdot}84}{10}}$$

$$=14{\cdot}96.$$

Standard coefficient of variation

$$=\frac{\sigma}{M}$$

$$=\frac{14{\cdot}96}{105{\cdot}9}$$

$$=\cdot141$$

On comparison of standard coefficient of deviation, it is obvious that series B shows greater deviation.

Problem 76.—The index numbers of prices of cotton and coal shares in 1942 were as under :—

Month	*Index number of prices of cotton shares*	*Index number of prices of coal shares*
January	188	131
February	178	13
March	173	130
April	164	129
May	172	129
June	183	120
July	184	127
August	185	127
September	211	130
October	217	137
November	232	140
December	240	142

which of the two shares do you consider more variable in price.

(*M. A. Agra, 1944*)

Solution :

Cotton Shares			Coal Shares		
Index No. of prices X	Deviation rom mean 194) X—M	$(X-M)^2$	Index No. of price's X	X—M M=131	$(X-M)^2$
188	−6	36	131	0	0
178	−16	256	130	−1	1
173	−21	441	130	−1	1
164	−30	900	129	−2	4
172	−22	484	129	−2	4
183	−11	121	120	−11	121
184	−10	100	127	−4	16
185	−9	81	127	−4	16
211	17	289	130	−1	1
217	23	529	137	6	36
232	38	1444	140	9	81
240	46	2116	142	11	121
$\Sigma X=2327$ $n=12$		6767	$\Sigma X=1581$ $n=12$		402

Cotton Shares.

Arithmetic average, $M=\frac{2327}{12}=194.$

Standard deviation, $\sigma=\sqrt{\frac{6767}{12}}=23{\cdot}8.$

Coefficient of variation,

$$=\frac{\sigma}{M}\times 100$$

$$=\frac{23{\cdot}8}{194}\times 100$$

$$=12{\cdot}27.$$

Coal Shares.

$$M=\frac{1581}{12}=131$$

$$\sigma=\sqrt{\frac{402}{12}}=5{\cdot}79$$

Coefficient of variation $=\frac{5{\cdot}79}{131}\times 100$

$$=4{\cdot}42.$$

From these datas we see that cotton shares are more variable in prices than the coal shares.

Problem 77. Calculate the standard deviation of the following data with regard to 2298 families in the U.K.

No. of persons in the family	*No. of families*
1	166
2	552
3	580
4	433
5	268
6	148
7	77
8	41
9	20
10	8
11	5
12	1
TOTAL	2298

(*M.A. Alld*, *1942*)

Solution :

No. of persons in family x	No. of families f	$\xi = x - A$	$f\xi$	$f\xi^2$
1	165	−3	− 495	1485
2	552	−2	−1104	2208
3	580	−1	− 580	580
4	433	0	0	0
5	268	1	268	268
6	148	2	296	592
7	77	3	231	693
8	41	4	164	656
9	20	5	100	500
10	8	6	48	284
11	5	7	35	245
12	1	8	8	64
TOTAL	2298		−1029	7575

$$\text{Standard deviation} = \sqrt{\frac{\Sigma f \xi^2}{\Sigma f} - \left(\frac{\Sigma f \xi}{f}\right)^2}$$

$$= \sqrt{\frac{7575}{2298} - \left(\frac{-1029}{2298}\right)^2}$$

$$= 1{\cdot}76.$$

Problem 78.—A collar manufacturer is considering the production of a new style of collar to attract young man. The following statistics of neck circumferences are available based upon measurements of a typical group of college students :—

Mid-Value (Inches)	No. of students	Mid-Value (Inches)	No. of students
12·5	4	15·0	29
13·0	19	15·5	18
13·5	30	16·0	1
14·0	63	16·5	1
14·5	66		

Compute the Standard Deviation and use the criterion ($M \pm 3\sigma$), where σ is standard deviation and M is arithmetic mean, to determine the largest and smallest sizes of collars he should make in order to meet the needs of practically all his customers, bearing in mind that collars are worn, on average, $\frac{3}{4}$ inches larger than neck size.

(*B. Com. Raj.*, *1949*)

Solution :

Neck circumference in inches x	No. of students f	$u=\frac{x-14\cdot5}{\cdot5}$	fu	fu^2
12·5	4	−4	−16	64
13	19	−3	−57	171
13·5	30	−2	−60	120
14	63	−1	−63	63
14·5	66	0	0	0
15	29	1	29	29
15·5	18	2	36	72
16	1	3	3	9
16·5	1	4	4	16
TOTAL	231		−124	544

Mean of the neck-circumferences

$$M=14\cdot5+\frac{-124}{231}\times\cdot5$$
$$=14\cdot5-\cdot26$$
$$=14\cdot24 \text{ inches}$$

Standard deviation of the neck-circumferences

$$\sigma=\cdot5\times\sqrt{\frac{544}{231}-\left(\frac{-124}{231}\right)^2}$$
$$=\cdot5\times\sqrt{2\cdot35-\cdot28}$$
$$=\cdot72 \text{ inches,}$$

Using the criterion $M\pm3\sigma$,

(*i*) The smallest size of the collars $=M-3\sigma+\frac{3}{4}$ inches.

$$=14\cdot24-3\times(7\cdot2)+\cdot75 \text{ inches}$$
$$=12\cdot83 \text{ inches.}$$

(*ii*) The largest size of the collars $= M + 3\,\sigma + {\cdot}75$ inches

$$= 14{\cdot}24 + 3({\cdot}72) + {\cdot}75$$

$$= 17{\cdot}15 \text{ inches.}$$

Problem 79. The following table shows the number of workers in two factories whose weekly earnings are given in column (1). Determine the mean values of weekly earnings and standard deviations in both factories.

Range of weekly earnings	Number of workers	
	Factory A	Factory B
Rs.		
4 — 6	74	71
6 — 8	376	379
8 — 10	304	303
10 — 12	110	112
12 — 14	18	18
14 — 16	0	1
16 — 18	9	3
18 — 20	9	9
20 — 22	0	4

(*M.A. Cal.*, *1936*)

Solution :

Let X represents the mid-value of the range of weekly earnings.

X	$u = \frac{X-9}{2}$	Factory A			Factory B		
		f	fu	fu^2	f	fu	fu^2
5	−2	74	−148	296	71	−142	284
7	−1	376	−376	376	379	−379	379
9	0	304	0	0	303	0	0
11	1	110	110	110	112	112	112
13	2	18	36	72	18	36	72
15	3	0	0	0	11	3	9
17	4	9	36	144	3	12	48
19	5	9	45	225	9	45	225
21	6	0	0	0	4	24	204
TOTAL		900	−297	1223	900	−289	1273

Factory A.

Mean value of weekly earning $M = A + h\,\frac{\Sigma fu}{\Sigma f}$

$$= 9 + 2 \times \frac{-297}{900}$$

$$= 8{\cdot}34 \text{ rupees.}$$

Standard deviation of weekly earnings

$$\sigma = h\sqrt{\frac{\Sigma f u^2}{\Sigma f} - \left(\frac{\Sigma f u}{\Sigma f}\right)^2}$$

$$= 2 \times \sqrt{\frac{1223}{900} - \left(\frac{-297}{900}\right)^2}$$

$$= 2{\cdot}236 \text{ rupees.}$$

Factory B.

$$M = 9 + 2 \times \frac{-289}{900}$$

$$= 8{\cdot}36 \text{ rupees.}$$

$$\sigma = 2 \times \sqrt{\frac{1273}{900} - \left(\frac{-289}{900}\right)^2}$$

$$= 2{\cdot}29 \text{ rupees.}$$

√**Problem 80.**—Calculate standard deviation and semi-interquartile range from the following table giving the age distribution of 542 members of the House of commons.

Age	*No. of Members*
20	3
30	61
40	132
50	153
60	140
70	51
80	2

(*B. Com., Nag., 1942*)

Solution :

$A = 50,\ h = 10$

x	f	Cumulative frequency	$u = \frac{X - A}{h}$	fu	fu^2
20	3	3	−3	−9	27
30	61	64	−2	−122	244
40	132	196	−1	−132	132
50	153	349	0	0	0
60	140	489	1	140	140
70	51	540	2	102	204
80	2	542	3	6	18
TOTAL	542			−15	765

Standard deviation of ages $\sigma = h\sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2}$

$$\text{or} \quad \sigma = 10 \times \sqrt{\frac{765}{542} - \left(\frac{15}{542}\right)}$$

$$= 11 \cdot 6 \text{ appr.}$$

For semi-inter-quartile range

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item's age}$$

$$= 136^{th} \text{ item's age}$$

$$= 50$$

$$Q_3 = 408^{th} \text{ item's age}$$

$$= 60$$

Semi-inter-quartile range

$$= \frac{Q_3 - Q_1}{2} = \frac{60 - 50}{2} = 5 \quad \textbf{Ans.}$$

Problem 81. In any two samples, where the variates n_1 and n_2 are measured in same units,

$n^1 = 36$ (summation) $\Sigma x^2_1 = 49428$
$n_2 = 49$ (do.) $\Sigma x^2_2 = 71258$

compute the values of the S.Ds of the two samples.

What additional information is required to calculate the coefficient of variation of the above two samples?

(*B.Com. Luck., 1943*)

Solution :

Standard deviation of the first sample

$$= \sqrt{\frac{\Sigma x_1^2}{n_1}} = \sqrt{\frac{49428}{36}} = 37 \cdot 01$$

Standard deviation of the second sample

$$= \sqrt{\frac{\Sigma x^2_2}{n_2}} = \sqrt{\frac{71258}{49}} = 38 \cdot 08.$$

Since coefficient of variation $= \frac{\sigma}{M} \times 100$

where σ is S.D. and M is arithmetic average.

Hence arithmetic averages M_1 and M_2 are more required to calculate coefficients of variation.

Problem 82.—Compile a table, showing the frequencies with which words of different numbers of letters occur in the extract reproduced below (omitting punctuation marks) treating as the variable the number letters in each word, and obtain the mean, median, and the coefficient of variation of the distribution :—

Success in the examination confers no absolute right to appointment, unless Government is satisfied, after such enquiry as may be considered necessary, that the candidate is suitable in all respects for appointment to the public service.

(I.A.S., 1947)

Solution :

The frequency distribution is as follows :

Number of letters (x)	*Number of words i.e. frequency* (f)	*Number of letters* (x)	*Number of words i.e. frequency* (f)
2	9	7	4
3	6	8	3
4	2	9	3
5	2	10	2
6	2	11	3

Calculation of the mean, median and S.D.

A=6 (assumed mean)

x	f	$\xi = x - A$	ξf	$f\xi^2$	Cumulative frequency cf
2	9	−4	−36	144	9
3	9	−3	−18	54	15
4	2	−2	−4	8	17
5	2	−1	−2	4	19
6	2	0	0	0	21
7	4	1	4	4	25
8	3	2	6	12	28
9	3	3	9	27	31
10	2	4	8	32	33
11	3	5	15	75	36
TOTAL	36	TOTAL	−18	360	

Mean M $= A + \dfrac{\Sigma f\xi}{\Sigma f}$

$= 6 - \dfrac{18}{36} = 5{\cdot}5$

Median M_d = the size of the $\left(\dfrac{36+1}{2}\right)$ *i.e.* $18{\cdot}5^{th}$ item.

$=5$

Standard deviation $\sigma = \sqrt{\dfrac{\Sigma f\xi^2}{\Sigma f} - \left(\dfrac{\Sigma f\xi}{\Sigma f}\right)^2}$

$= \sqrt{\dfrac{360}{36} - \left(\dfrac{-18}{36}\right)^2} = 3{\cdot}12$

Coefficient of variation $= \dfrac{\sigma}{M} \times 100 = \dfrac{3{\cdot}12}{5{\cdot}5} \times 100 = 56{\cdot}7$

Problem 83. The following table gives the number of finished articles turned out per day by different number of workers in a factory. Find the mean value and standard deviation of the daily output of finished articles, and explain the significance of standard deviation.

No. of articles (*X*)	*No. of workers* (*f*)	*No. of articles* (*X*)	*No. of workers* (*f*)
18	3	23	17
19	7	24	13
20	11	25	8
21	14	26	9
22	18	27	4
Continued	Continued		

(*B.Com., Cal., 1937*)

Solution :

A=22 (assumed average)

X	f	ξ=X−A	$f\xi$	$f\xi^2$
18	3	−4	−12	48
19	7	−3	−21	63
20	11	−2	−22	44
21	14	−1	−14	14
22	18	0	0	0
23	17	1	17	17
24	13	2	26	52
25	8	3	24	72
26	9	4	36	144
27	4	5	20	100
	104		54	554

$$\text{Mean M} = \text{A} + \frac{\Sigma f\xi}{\Sigma f}$$

$$= 22 + \frac{54}{104}$$

$$= 22{\cdot}519$$

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma f\xi^2}{\Sigma f} - \left(\frac{\Sigma f\xi^2}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{554}{104} - \left(\frac{54}{104}\right)^2}$$

$$= 2{\cdot}2.$$

Problem 84. From the following figures find the standard deviation and the coefficient of variation :—

Marks	*No. of persons*	*Marks*	*No. of persons*
0—10	5	40—50	30
10—20	10	50—60	20
20—30	20	60—70	10
30—40	40	70—80	4

(*B. Com., (S) Agra, 1948*)

Solution :

Taking X to represent the mid-values of marks column $h = 10$ (class difference), $A = 35$ (assumed average).

X	No. of persons f	$u = \frac{X-A}{h}$	fu	fu^2
5	5	−3	−15	45
15	10	−2	−20	40
25	20	−1	−20	20
35	40	0	0	0
45	30	1	30	30
55	20	2	40	80
65	10	3	30	90
75	4	4	16	64
	139		61	369

$$\text{Mean } M = 35 + 10 \times \frac{61}{139} \qquad \left(= A + h\frac{\Sigma f u}{\Sigma f} \right)$$

$$= 39{\cdot}5$$

$$\text{Standard deviation } \sigma = h\sqrt{\frac{\Sigma f u^2}{\Sigma f} - \left(\frac{\Sigma f u}{\Sigma f}\right)^2}$$

$$= 10 \times \sqrt{\frac{369}{139} - \left(\frac{61}{139}\right)^2}$$

$$= 10 \times \sqrt{2{\cdot}65 - {\cdot}20}$$

$$= 15{\cdot}6$$

$$\text{Coefficient of variation} = \frac{\sigma}{M} \times 100.$$

$$= \frac{15{\cdot}5}{39{\cdot}5} \times 100$$

$$= 39{\cdot}2$$

Problem 85. Show that in a discrete series if the deviations, x, from the mean are small so that $\left(\frac{x}{M}\right)^3$ and higher powers of $\frac{x}{M}$ can be neglected, the following relations are found to hold approximately between the arithmetic mean, A, the geometric mean, G, and the harmonic mean, H. (σ being the standard deviation and $\frac{\sigma 3}{M^3}$ can also be neglected). [*Statistics*—J. C. Chaturwedi]

(i) $G = M\left(1 - \frac{\sigma^2}{2M^2}\right)$

(ii) $M^2 - G^2 = \sigma^2$

(iii) $H = M\left(1 - \frac{\sigma^2}{M^2}\right)$

(iv) $MH = G^2$ (v) $M - 2G + H = 0$

Solution :

If X is the variable value, we have, $X - M = x$, so that $X = M + x$

(i) Now if n is the total frequency

$$\log G = \frac{1}{n} \Sigma f \log (x + M)$$

$$= \frac{1}{n}\left[\Sigma f \log M + \Sigma f \log\left(1 + \frac{x}{M}\right)\right]$$

$$= \log M + \frac{1}{n} \Sigma f \left(\frac{x}{M} - \frac{x^2}{2M^2}\right)$$

$\left(\frac{x}{M}\right)^3$ and higher powers are neglected.

but $\Sigma f x = 0$ and $\frac{1}{n} \Sigma f x^2 = \sigma^2$

$$\therefore \log G = \log M - \frac{\sigma^2}{2M^2}$$

$$\therefore \quad G = M e^{-\frac{\sigma^2}{2M^2}} \quad \dots \quad \dots \quad (1)$$

$$= M\left(1 - \frac{\sigma^2}{2M^2}\right)$$

(ii) From (1) on squaring

$$G^2 = M^2 e^{-\frac{\sigma^2}{M^2}}$$

$$= M^2\left(1 - \frac{\sigma^2}{M^2}\right)$$

$$\therefore M^2 - G^2 = \sigma^2$$

(iii) now

$$\frac{1}{H} = \frac{1}{n} \Sigma f. \frac{1}{x + M}$$

$$=\frac{1}{n}\Sigma\frac{f}{M}\left(1+\frac{x}{M}\right)^{-1}$$

or $$\frac{1}{H}=\frac{1}{nM}\Sigma f\left(1-\frac{x}{M}+\frac{x^2}{M^2}\right)$$

$\left[\text{since}\left(\frac{x}{M}\right)^3\text{ etc., can be neglected.}\right]$

Also $\Sigma f x=0$ and $\frac{1}{n}\Sigma f x^2=\sigma^2$

$$\therefore\quad\frac{1}{H}=\frac{1}{M}\left(1+\frac{\sigma^2}{M^2}\right)$$

$$\therefore\quad H=M\left(1+\frac{\sigma^2}{M^2}\right)^{-1}$$

or $$H=M\left(1-\frac{\sigma^2}{M^2}\right)$$ neglecting higher powers ... ... (2)

(*iv*) From relation (2)

$$MH=(M^2-\sigma^2)$$

or $MH=G^2$ by relation (*ii*) above

Problem 86. An analysis of the monthly wages paid to workers in two firms A and B, belonging to the same industry, gives the following results :—

	Firm A	*Firm B*
Number of wage-earners	586	648
Average monthly wage	Rs. 52·5	Rs. 47·5
Variance of the distribution of wage	100	121

(*a*) which firm, A or B, pays out the larger amount as monthly wages ?

(*b*) In which firm, A or B, is there greater variability in individual wages ?

(*c*) what are the measures of—

(*i*) average monthly wage, and

(*ii*) the variability in individual wages, of all the workers in the two firms, A and B, taken together.

[*I.A.S. etc., exams., 1951*]

Solution :

(*a*) *Firm A.*

The average monthly wage per worker =Rs. 52·5

Total number of workers = 586

Total monthly wages paid to workers = $52 \cdot 5 \times 586$

=Rs. 30763.

Firm B.

The total monthly wages paid to workers=Rs. $47 \cdot 5 \times 648$

=Rs. 30780

From these results it is clear that firm B pays out a larger amount.

(*b*) *Firm A.*

Variance of the distribution of wages is 100

∴ Standard deviation $\sigma_a = \sqrt{100} = 10$

Arithmetic average $M_a = 52 \cdot 5$

∴ Coefficient of the standard deviation of

the distribution of wages$= \frac{\sigma_a}{M_a} = \frac{10}{52 \cdot 5}$

$= \cdot 19$

Firm B.

Standard deviation $\sigma_b = \sqrt{121} = 11$

Coefficient of the standard deviation of the distribution

of wages$= \frac{\sigma_b}{M_b} = \frac{11}{47 \cdot 5}$

$= \cdot 23.$

Hence there is greater variability in individual wages in Firm B.

(*c*) (*i*) The average monthly wage of all the workers in both the factories A and B would be obtained by the following formula :

$$M_{ab} = \frac{n_a M_a + n_b M_b}{n_a + n_b} \text{ where}$$

n_a and n_b stands for the number of items in the two firms A and B respectively M_a and M_b for the mean of the two firms A and B respectively and M_{ab} for the combined mean of the two firms.

$$\therefore \quad M_{ab} = \frac{586 \times 52 \cdot 5 + 648 \times 47 \cdot 5}{586 + 648}$$

$$= \text{Rs. } 49 \cdot 9$$

(*ii*) The combined standard deviation of the two series would be given by the following formula —

$$\sigma_{ab} = \sqrt{\frac{n_a\sigma_a^2 + n_a(M_a - M_{ab})^2 + n_b\sigma_b^2 + n_b(M_b - M_{ab})^2}{n_a + n_b}}$$

where suffixes *a* and *b* represent for firms A and B respectively and *n*, M, σ have their usual meaning (*i.e.* frequency, mean and standard deviation).

$$\therefore\ \sigma_{ab} = \sqrt{\frac{586\times100 + 586(52{\cdot}5 - 49{\cdot}9)^2 + 648\times121 + 648(47{\cdot}5 - 49{\cdot}9)^2}{586 + 648}}$$

$$= \sqrt{\frac{58600 + 3984{\cdot}6 + 78408 + 3693{\cdot}6}{1234}}$$

$$= \sqrt{\frac{144686{\cdot}2}{1234}}$$

$$= \sqrt{117{\cdot}2}$$

$$= 10{\cdot}8.$$

Problem 87. In any two series, where ξ_1 and ξ_2 represent the deviation from an assumed average, 100,

$n_1 = 150$ $\Sigma\xi_1 = 100$ $\Sigma\xi_1^2 = 245320$

$n_2 = 200$ $\Sigma\xi_2 = 250$ $\Sigma\xi_2^2 = 43850$

A = 100

Calculate the coefficient of variation for the two series.

Solution :

Series I.

Arithmetic average $M_1 = A + \frac{\Sigma\xi_1}{n_1}$

or $M_1 = 100 + \frac{180}{150}$

$= 101{\cdot}2$

Standard deviation $\sigma_1 = \sqrt{\frac{\Sigma\xi_1^2}{n_1} - \left(\frac{\Sigma\xi}{n_1}\right)^2}$

or $\sigma_1 = \sqrt{\frac{245320 - 150(1{\cdot}2)^2}{150}}$

$= 40{\cdot}4$

Coefficient of variation $= \frac{\sigma_1}{M_1} \times 100$

$= \frac{40{\cdot}4}{101{\cdot}2} \times 100$

$= 39{\cdot}9$

Series 2 :

$$M_2 = A + \frac{\Sigma\xi_2}{n_2}$$

$$= 100 + \frac{250}{200}$$

$$= 101 \cdot 25$$

$$\sigma_2 = \sqrt{\frac{\Sigma\xi_2^2}{n_2} - \left(\frac{\Sigma\xi_2}{n_2}\right)^2}$$

$$= \sqrt{\frac{43850}{200} - \left(\frac{250}{200}\right)^2}$$

$$= 14 \cdot 75$$

$$\text{Coefficient of variation} = \frac{\sigma_2}{M_2} \times 100$$

$$= \frac{14 \cdot 75}{101 \cdot 25} \times 100$$

$$= 14 \cdot 6.$$

√**Problem 88.** From the figures given below compare the variability of population of Banaras and Allahabad (in thousands.)

	Allahabad	Banaras
1881	160	218
1891	175	223
1901	172	213
1911	172	204
1921	157	198
1931	184	205
1941	261	263

(*M.A. Agra, 1948*)

Solution :

Year	Population of Allahabad in thousands X_1	Deviation from Arithmetic mean of Allahabad X_1—M (183)	$(X_1-M)^2$	Population of Banaras in thousands X_2	Deviation from arithmetic mean of Banaras X_2—M(218)	$(X_2-M)^2$
1881	160	—23	529	218	0	0
1891	175	— 8	64	223	5	25
1901	172	—11	121	213	— 5	25
1911	172	—11	121	204	—14	196
1921	157	—26	676	198	—20	400
1931	184	1	1	205	—13	169
1941	261	78	6084	263	45	202
N=7	1281	1524	7596			2840

Arithmetic average of Allahabad is 183 thousands
Arithmetic average of Banaras is 218 thousands

Standard Deviation of Allahabad, $\sigma_1 = \sqrt{\frac{\Sigma(X_1 - M)^2}{N}}$

$$\sigma_1 = \sqrt{\frac{7596}{7}} = 32 \cdot 9 \text{ thousands.}$$

Coefficient of variation of Allahabad $= \frac{\sigma_1}{M}$

$$= \frac{32 \cdot 9}{183} = \cdot 18$$

Standard Deviation of Banaras, $\sigma_2 = \sqrt{\frac{2840}{7}}$

$$= 20 \cdot 1 \text{ thousands}$$

Coefficient of variation of Banaras

$$= \frac{20 \cdot 1}{218}$$

$$= \cdot 09.$$

Therefore, the Population of Allahabad is more variable than the population of Banaras.

Problem 89. Calculate standard Deviation of the following :—

Variable :	Above 0	10	20	30	40	50	60	70	80
Frequency :	150	140	100	80	80	70	30	14	0

(*Raj., M.A., 1956*)

Solution :

Marks (X)	Frequency (f)	$u = \frac{X - A}{h}$	fu	fu^2
0	150	−4	0	0
10	140	−3	−420	1260
20	100	−2	−200	400
30	80	−1	−80	80
40	80	0	0	0
50	70	+1	+70	70
60	30	+2	+60	120
70	14	+3	+42	126
80	0	+4	0	0
	$n = 664$		$\Sigma fu = -528$	$\Sigma fu^2 = 2056$

Let the assumed mean (A) = 40 $h=10$

$$\text{Standard Deviation } (\sigma) = h\sqrt{\frac{\Sigma f u^2}{\Sigma f} - \left(\frac{\Sigma f u}{\Sigma f}\right)^2}$$

$$= 10\sqrt{\frac{2056}{664} - \left(\frac{-528}{664}\right)^2}$$

$$= \frac{10}{664}\sqrt{2056\times 664-(-528\times -528)}$$

$$= \frac{10}{664}\sqrt{(2056\times 664)-(278784)}$$

$$= \frac{10}{664}\sqrt{1092376}$$

$$= \frac{10\times 1078}{664} \text{ approximately.}$$

$$= \frac{10780}{664} \text{ approx.}$$

$$= 16{\cdot}2 \text{ approx.}$$

Problem 90. From the following table calculate the coefficient of variation :

Age at the birth of First child	: 13	14	15	16	17	18	19	20	21	22	23	24	25
No. of married women	: 37	162	343	390	256	433	161	355	65	85	49	46	40

(Raj., M.A., 1955)

Solution :

Age at the birth of First child (X)	No. of married women (f)	Xf	ξ (X—M)	ξ^2	$f\xi^2$
13	37	481	−5	25	925
14	162	2168	−4	16	2592
15	343	5145	−3	9	3087
16	390	6240	−2	4	1560
17	256	4352	−1	1	256
18	433	7794	0	0	0
19	161	3059	+1	1	161
20	355	7100	+2	4	1420
21	65	1365	+3	9	585
22	85	1870	+4	16	1360
23	49	1127	+5	25	1225
24	46	1104	+6	36	1656
25	40	1000	+7	49	1960
	$n=2422$	ΣXf =42805			$\Sigma f\xi^2$ =16787

$$= \frac{\Sigma Xf}{n} = \frac{42805}{2422} = 17{\cdot}6 = 18 \text{ approx.}$$

Standard Deviation $(\sigma) = \sqrt{\frac{\Sigma f \xi^2}{n}}$

$$= \sqrt{\frac{16787}{2422}}$$

$$= \sqrt{6{\cdot}9}$$

$$= 2{\cdot}6 \text{ approx.}$$

Therefore, coefficient

of variation $= \frac{\sigma}{M} \times 100$

$$= \frac{2{\cdot}6}{17{\cdot}6} \times 100$$

$$= \frac{260}{17{\cdot}6} \text{ approx.}$$

$$= 14{\cdot}7 \text{ approximately.}$$

Problem 91. Calculate the coefficient of variation of the following monthly incomes of twenty families given below in rupees :—

2,000 ; 35 ; 400 ; 15 ; 40 ; 1500 ; 300 ; 6 ; 90 ; 250 ; 20 ; 12 ; 450 ; 10 ; 150 ; 8 ; 25 ; 30 ; 1200 ; 60.

(*Alld., B. Com., 1941*)

Solution :

Serial No. of Families	Income (X)	Serial No. of Families	Income (X)
1	2000	11	20
2	35	12	12
3	400	13	450
4	15	14	10
5	40	15	150
6	1500	16	8
7	300	17	25
8	6	18	30
9	90	19	1200
10	250	20	60
			$\Sigma X = 6601$

Range = (2000 − 60) rupees.

= 1940 rupees.

$$M = \frac{6601}{20} \left[= \frac{\Sigma X}{n} \text{ where } \Sigma X = 6601, n = 20 \right]$$

= 330 rupees approximately.

∴ Coefficient of variation by Range method and with the arithmetic average $=\frac{R}{n}\times 100$

[R=Range
$n=\Sigma X$

$=\frac{1940}{6601}\times 100$

$=29{\cdot}3$ approximately.

Problem 92. Find the Arithmetic Average, the First Moment of Dispersion and the Standard Deviation from the data in the following series :

Size of item	*Frequency*	*Size of item*	*Frequency*
3—4	3	7—8	85
4—5	7	8—9	32
5—6	22	9—10	8
6—7	60		

(*Alld., B.Com. 1942*)

Solution :

Size	Mid-value (X)	Frequency (f)	Xf	ξ (± ignored)	ξ^2	$f\xi$	$f\xi^2$
3— 4	3.5	3	10·5	3·0	9	9	27
4— 5	4·5	7	31·5	2·0	4	14	28
5— 6	5·5	22	121·0	1·0	1	22	22
6— 7	6·7	60	390·0	0	0	0	0
7— 8	7·5	85	637·5	1·0	1	85	85
8— 9	8.5	32	272·0	2·0	4	64	128
9—10	9·5	8	76·0	3·0	9	24	72
		n=217	1438·5			$\Sigma f\xi$ =218	$\Sigma f\xi^2$= 362

$$M=\frac{\Sigma Xf}{n}=\frac{1438{\cdot}5}{217}=6{\cdot}5 \text{ approx.}$$

$$\text{Mean Deviation}=\frac{\Sigma f\xi}{n}$$

$$=\frac{218}{6{\cdot}5}$$

Standard Deviation $=1{\cdot}004$ approx.

$$=\sqrt{\frac{\Sigma f\xi^2}{n}}$$

$$=\sqrt{\frac{362}{217}}$$

$$=\sqrt{1{\cdot}66}$$

$=1{\cdot}2$ approximately.

Problem 93. Calculate the Mean Deviation from the following data :—

Difference in age between husband and wife in a particular Community.

Difference in years	*Frequency*	*Difference in years*	*Frequency*
0— 5	449	20—25	109
5—10	705	25—30	52
10—15	507	30—35	16
15—20	281	35—40	4

(*Bom., B.Com., 1936*)

Solution :

Age-group	Mid-value (X)	Frequency (f)	Xf	ξ (X$-M$)	$f\xi$
0— 5	2·5	449	1122·5	8	3592
5—10	7·5	705	5287·5	3	2115
10—15	12·5	507	6337·5	2	1014
15—20	17·5	281	4917·5	7	1967
20—25	22·5	109	2452·5	12	1308
25—30	27·5	52	1430·0	17	884
30—35	32·5	16	520·0	22	352
35—40	37·5	4	150·0	27	108
		n=2129	Σxf= 22217·5	± Signs ignored	$\Sigma f\xi$= 113.0

$$M=\frac{\Sigma Xf}{n}$$

$$=\frac{22217·5}{2129}$$

$$=10·5 \text{ approximately.}$$

$$\text{Mean Deviation from M } (S_m)=\frac{\Sigma f\xi}{n}$$

$$=\frac{11340}{2129}$$

$$=5·32 \text{ approximately.}$$

Problem 94. Calculate the Standard Deviation from the following data :—

Size of item	*Frequency*
6	3
7	6
8	9
9	13
10	8
11	5
12	4
	48

(*Bom., B.Com., 1936*
Nagpur, B.Com., 1944)

Solution :

Size of item (X)	Frequency (f)	Xf	ξ (X−M)	ξ^2	$f\xi^2$
6	3	18	−3	9	27
7	6	42	−2	4	24
8	9	72	−1	1	9
9	13	117	0	0	0
10	8	80	+1	1	8
11	5	55	+2	4	20
12	4	48	+3	9	36
	n=48	Σxf =432	M=9		$\Sigma f\xi^2$ =124

$$M = \frac{\Sigma Xf}{n}$$

$$= \frac{432}{48}$$

$$= 9$$

Standard Deviation

$$= \sqrt{\frac{\Sigma f\xi^2}{n}}$$

$$= \sqrt{\frac{124}{48}}$$

= 1·6 approximately.

Problem 95. The following table gives the values of imports of certain commodities into India in Lakhs of Rupees :—

	1908	1909	1910	1911	1912	1913
Cotton goods	3220	3282	3754	4120	5180	6054
Woollen goods	238	158	243	279	240	306
Boots and shoes	39	57	46	55	65	74

Calculate a coefficient to determine which of the above imports is most variable from year to year.

(*Elementary Statistics*—Dubey and Agarwal, p. 260]

Solution :

(Mean Deviation and coefficient of Mean Deviation.)

	Cotton Goods		Woollen Goods		Boots and Shoes	
Year	Imports in lakhs (X)	ξ	Imports in lakhs (X_1)	ξ_1	Imports in lakhs (X_2)	ξ_2
1908	3220	1048	238	6	39	17
1909	3282	986	158	86	57	1
1910	3754	514	243	1	46	10
1911	4120	148	279	35	55	1
1912	5180	912	240	4	65	9
1913	6054	1786	306	62	74	18
$n=6$	$\Sigma X=25610$	$\Sigma\xi=5394$	$\Sigma X_1=1464$	$\Sigma\xi_1=194$	$\Sigma X_2=336$	$\Sigma\xi_2=56$

(A) Cotton Goods—

(*i*) $M=\frac{\Sigma X}{n}=\frac{25610}{6}=4268$ approximately.

(*ii*) Mean Deviation $=\frac{\Sigma|\xi|}{n}$

$=\frac{5394}{6}$

$=899$ approximately.

(*iii*) Coeff. of M.D. $=\frac{899}{4268}$

$=\cdot 21$ approximately.

(B) Woollen Goods—

(*i*) $M=\frac{\Sigma X_1}{n}=\frac{1464}{6}=244$ approx.

(*ii*) Mean Deviation $=\frac{194}{6}=32{\cdot}3\dot{3}$ approximately.

(*iii*) Coefficient of M.D. $=\frac{32}{244}$

$=\cdot 13$ approximately.

(C) Boots and Shoes—

(*i*) $M=\frac{336}{6}=56$ approximately.

(*ii*) Mean Deviation $=\frac{56}{6}=9{\cdot}3\dot{3}$ approximately.

(*iii*) Coeff. of M. D. $=\frac{9}{56}$ approx.

$=\cdot 16$ approximately.

The import of cotton have highest variation.

Problem 96. The following table gives the number of the passengers carried and the amount received by a certain Motor Bus Company during the years 1925—1931 :

Years	*Receipts in Rs.*	*Passengers*
1925	1354	50,010
1926	2780	61,060
1927	3011	70,005
1928	3020	70,110
1929	3541	82,001
1930	4150	91,000
1931	5000	1,00,000

Find out one measure of dispersion and state whether the variation in receipts is greater than that in passengers.

(*Elementary Statistics*—Dubey and Agarwal, p. 261)

Solution :

Years	Receipts		Passengers	
	in Rs. (X_1)	$\xi_1=(X_1-M_1)$	Nos. (X_2)	$\xi_2=(X_2-M_2)$
1925	1354	1911	50010	24874
1926	2780	485	61060	13824
1927	3011	254	70005	4879
1928	3020	245	70110	4774
1929	3541	276	82001	7117
1930	4150	885	91000	16116
1931	5000	1735	1,00,000	25116
$n=7$	$\Sigma X_1=22856$	$\Sigma\xi_1=5791$	$\Sigma X_2=524186$	$\Sigma\xi_2=96700$

Arithmetic Average for Receipts $=\frac{\Sigma X_1}{n}$

$=\frac{22856}{7}$

$=3265\cdot1$ rupees approximately

$=3265$ rupees approximately.

Mean Deviation of Receipts $=\frac{\Sigma\xi_1}{n}$

$=\frac{5791}{7}=827\cdot2$ approximately.

Arith. average

for Passengers $=\frac{\Sigma X_2}{n}$

$$= \frac{524186}{7} = 74883{\cdot}71 \text{ approx.}$$

$$= 74884 \text{ approximately.}$$

Mean Deviation ,, $= \frac{96700}{7}$

$$= 13814{\cdot}2 \text{ approximately.}$$

Thus, we find that the variation of passengers are greater than the receipts.

Problem 97. Find the Average deviation from the Mean for the following :—

Class	*Frequency*	*Class*	*Frequency*
0— 6	8	18—24	9
6—12	10	24—30	5
12— 18	12		

(*Practical Statistics*—Dr. J.C. Chaturvedi, p. 104)

Solution :

Class	Mid-value (X)	Frequency (f)	X f	ξ (X—M)	$f\,\xi$
0— 6	3	8	24	11	88
6—12	9	10	90	5	50
12—18	15	12	180	1	12
18—24	21	9	189	7	63
24—30	27	5	135	13	65
		n=44	ΣXf=618	± signs ignored	$\Sigma f\xi$ =278

Arithmetic Average (M) $= \frac{\Sigma Xf}{n}$

$$= \frac{618}{44}$$

$$= 14{\cdot}04 \text{ approx.}$$

$$= 14 \text{ approximately.}$$

Mean Deviation from (M) $= \frac{\Sigma f\xi}{n}$

$$= \frac{278}{44}$$

$$= 6{\cdot}31 \text{ approximately.}$$

∴ Coefficient of Mean

Deviation $= \frac{\delta_m}{M}$ where δ_m = Mean Deviation from M

$= \frac{6 \cdot 31}{14}$

$= \cdot 45$ approximately.

Problem 98.—Find the Mean Deviation from the Median of the following :

Value of variable	*Frequency*	*Value of variable*	*Frequency*
6	4	30	15
12	7	36	10
18	9	42	5
24	18		

(*Practical Statistics*—Dr. J.C. Chaturvedi : p. 104.)

Solution :

Value of variable (X)	Frequency (f)	$C\ mf$	ξ ($X - M_d$)	$f\xi$
6	4	4	18	72
12	7	11	12	84
18	9	20	6	54
24	18	38	0	0
30	15	53	6	90
36	10	63	12	120
42	5	68	18	90
	$n = 68$		± omitted	$\Sigma f\xi = 510$

Md = Size of $\left(\frac{n+1}{2}\right)^{th}$ item.

= ,, $\frac{69}{2}^{th}$ item.

= ,, $34 \cdot 5^{th}$ item.

= 24 approximately.

Mean Deviation from the

Md (δ_{md}) $= \frac{\Sigma f\xi}{n}$

$= \frac{510}{68}$

$= 7 \cdot 5$ approximately.

∴ Coefficient of Mean

Deviation from $Md = \frac{\delta_{md}}{Md}$

$= \frac{7{\cdot}5}{24}$

$=$ ·31 approximately.

Problem 99.—Find the Range and the Range coefficient of the following :

Marks	*Candidates*	*Marks*	*Candidates*
10	26	15	310
11	201	16	80
12	673	17	13
13	100	18	1
14	739		

(*Practical Statistics*—
Dr. J.C. Chaturvedi : p. 105.)

Solution :

Range $= (M_a - M_i)$ where M_a = maximum value
M_i = minimum value.

$= (18 - 10)$

$= 8$

∴ Range Coefficient

$= \frac{(M_a - M_i)}{(M_a + M_i)}$

$= \frac{(18-10)}{(18+10)}$

$= \frac{8}{28}$

$=$ ·28 approx.

Problem 100.—Find the Range and the Range coefficient of the following :

Age in years	*Frequency*
5—10	10
10—15	15
15—20	20
20—25	5

Solution :

Age in years	*Mid-value (X)*	*Frequency (f)*
5—10	7·5	10
10—15	12·5	15
15—20	17·5	20
20—25	22·5	5

Range $= (22{\cdot}5 - 7{\cdot}5)$
$= 15{\cdot}0$

Coefficient from Range

$$= \frac{22{\cdot}5 - 7{\cdot}5}{22{\cdot}5 + 7{\cdot}5}$$

$$= \frac{15{\cdot}0}{30{\cdot}0}$$

$= {\cdot}5$ approx.

Problem 101.—Find the Range and Range Coefficient of the following :

Value of variable	*Frequency*	*Value of variable*	*Frequency*
2	3	8	47
4	10	10	17
6	25		

Solution :

Range $= (M_a - M_i)$
$= (10 - 2)$
$= 8$

Coeff. $= \frac{M_a - M_i}{M_a + M_i}$

$$= \frac{10 - 2}{10 + 2}$$

$= \frac{8}{12} = {\cdot}66$ approx.

Problem 102.—Find the Range and coefficient of the following :

Height in inches	*Frequency*	*Height in inches*	*Frequency*
66	10	70	15
67	20	71	12
68	25	72	4
69	30		

Solution :

Range $= (M_a - M_i)$
$= (72 - 66)$
$= 6$ inches

Coeff. of Range Disp. $= \left(\frac{M_a - M_i}{M_a + M_i}\right)$

$$= \frac{72 - 66}{72 + 66}$$

$$= \frac{6}{138}$$

$= {\cdot}043$ approximately.

Problem 103.—Calculate the Mean Deviation and Coefficient of Deviation of the following :

Consumption in K.W. Hours	*No. of users.*
0 but less than 10	10
10 but less than 20	25
20 but less than 30	30
30 but less than 40	20
40 but less than 50	15
TOTAL	100

(*Elementary Statistical Methods*—Neiswanger, p. 365.)

Solution :

Cons in K.W.H.	Mid-value (X)	Frequency (f)	Xf	ξ (M—x)	$f\xi$
0—10	5	10	50	21	210
10—20	15	25	375	11	275
20—30	25	30	750	1	30
30—40	35	20	700	9	180
40—50	45	15	675	19	285
		$n=100$	$\Sigma Xf=2550$	$\pm$ omitted	$\Sigma f\xi = 980$

$$M = \frac{\Sigma X f}{n}$$

$$= \frac{2550}{100}$$

$= 25{\cdot}5$ approximately

$= 26$ approximately.

Mean Deviation from Mean

$$(\delta_m) = \frac{\Sigma f \xi}{n}$$

$$= \frac{980}{100}$$

$= 9{\cdot}80$ approximately.

∴ Coefficient of Mean

Deviation from Mean $= \dfrac{\delta_m}{M}$

$$= \frac{9{\cdot}8}{26}$$

$= {\cdot}37$ approximately.

Problem 104.—From the following frequency distribution of the size of sales of tickets, calculate the Mean Deviation (and its coefficient)

Sales in Dollars	*No. of sales*
0— 1·99	2
2— 3·99	10
4— 5·99	26
6— 7·99	32
8— 9·99	8
10—11·99	2
TOTAL	80

(*Elementary Statistical Methods*—Neiswanger p. 365.)

Solution :

Sales in $	Mid-Value (X)	Freq. (f)	Xf	ξ (X−M)	$f\xi$
0—1·99	1	2	2	5	10
2—3·99	3	10	30	3	30
4—5·99	5	26	130	1	26
6—7·99	7	32	224	1	32
8—9·99	9	8	72	3	24
10—11·99	10	2	22	5	10
		n = 80	$\Sigma Xf = 480$	± Omitted	$\Sigma f\xi = 132$

$$M = \frac{\Sigma Xf}{n}$$

$$= \frac{480}{80}$$

$$= \$\ 6 \text{ approx.}$$

$$\text{Mean Deviation from Mean}(\delta m) = \frac{\Sigma f\xi}{n}$$

$$= \frac{132}{80}$$

$$= \$\ 1{\cdot}65 \text{ approx.}$$

$$\therefore \text{Coefficient of Mean Deviation} = \frac{\delta m}{M}$$

$$= \frac{1{\cdot}65}{6}$$

$$= {\cdot}27 \text{ approximately.}$$

Problem 105.—Calculate Quartile Deviation and Coefficient of quartile deviation from the following :

Height in inches	*Frequencies*	*Height in inches*	*Frequencies*
50	10	54	14
51	12	55	18
52	15	56	6
53	10		

Solution :

Height in inches (X)	Frequencies (f)	*cmf*
50	10	12
51	12	22
52	15	37
53	10	47
54	14	61
55	18	79
56	6	85
	$n = 85$	

$$Q_1 = \text{Size of } \left(\frac{n+1}{4}\right)^{th} \text{ item.}$$

$$= \text{,, ,, } \frac{86^{th}}{4} \text{ item.}$$

$$= 21{\cdot}5^{th} \text{ item.}$$

$$= 51$$

$$Q_3 = \text{Size of } \frac{3(n+1)^{th}}{4} \text{ item.}$$

$$= \text{,, ,, } \frac{258}{4} = 64{\cdot}5^{th} \text{ item.}$$

$$= 55.$$

Quartile Deviation (Semi-interquartile range) $= \dfrac{Q_3 - Q_1}{2}$

$$= \frac{55-51}{2}$$

$$= \frac{4}{2} = 2 \text{ approximately.}$$

Coefficient of Quartile Deviation

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{55-51}{55+51}$$

$$= \frac{4}{106}$$

$$= {\cdot}037 \text{ approximately.}$$

Problem 106. Calculate Quartile Deviation and Co-efficient of semi-interquartile range of the following :—

Months	*Sales in Rs.*	*Months*	*Sales in Rs.*
Jan.	55	July	120
Feb.	60	Aug.	130
March	70	Sept.	145
April	70	Oct.	145
May	90	Nov.	155
June	110	Dec.	170

Solution :

Months	*Sales in Rs.*	*Months*	*Sales in Rs.*
Jan.	55	July	120
Feb.	60	Aug.	130
March	70	Sept.	145
April	70	Oct.	145
May	90	Nov.	155
June	110	Dec.	170

$$Q_1 = \text{Size of the } \left(\frac{n+1}{4}\right)^{th} \text{ item.}$$

$$= \text{,, ,, } \frac{13}{4}^{th} \text{ item} = 3{\cdot}25^{th} \text{ item.}$$

$= \text{Rs. } 70$ approximately

$$Q_3 = \text{Size of the } \frac{3(n+1)^{th}}{4} \text{ item.}$$

$$= \text{,, ,, } \frac{39^{th}}{4} \text{ item} = 9{\cdot}75^{th} \text{ item.}$$

$= \text{Rs. } 145$ approximately.

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{145-70}{2} = \frac{85}{2} = \text{Rs. } 42{\cdot}5 \text{ approximately.}$$

Coefficient of Semi Inter quartile range

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{145-70}{145+70}$$

$$= \frac{85}{215}$$

$= {\cdot}39$ approximately.

✓**Problem 107.**—Calculate the Quartile Deviation and Coefficient of Quartile Deviation of the following :—

Monthly wages in Rs.	*No. of wage earners*	*Monthly wages in Rs.*	*No of wage earners*
12·5—17·5	2	32·5—37·5	3
17·5—22·5	22	37·5—42·5	4
22·5—27·5	19	42·5—47·5	6
27·5—32·5	14	47·5—52·5	1
		52·5—57·5	1

(*Practical Statistics*—Ziauddin, p. 77.)

Solution :

Monthly Wages in Rs.	Mid-value (X)	Frequency (f)	*cmf*
12·5—17·5	15	2	2
17·5—22·5	20	22	24
22·5—27·5	25	19	43
27·5—32·5	30	14	57
32·5—37·5	35	3	60
37·5—42·5	40	4	64
42·5—47·5	45	6	70
47·5—52·5	50	1	71
52·5—57·5	55	1	72
		$n=72$	

q_1 $=$Size of $\left(\frac{n+1}{4}\right)^{th}$ item$=\frac{73}{4}=18{\cdot}25^{th}$ item.

q_3 $=$Size of $\frac{3(n+1)^{th}}{4}$ item$=\frac{219}{4}=54{\cdot}75^{th}$ item.

Q_1 $=l_1+\frac{i}{f}\left(\frac{n}{4}-c\right)$ where l_1=lower limit of Q_1 group.

$=17{\cdot}5+\frac{5}{22}(18{\cdot}25-2)$ i=class-difference

$=17{\cdot}5+\frac{5}{22}\times 16{\cdot}25$ f=freq. of the Q_1 group.

$=17{\cdot}5+3{\cdot}69$ approx. $\frac{n}{4}$=value of q_1

$=21{\cdot}19$ approximately c=cmf. of next lower group

Q_3 $=l_1+\frac{i}{f}\left(\frac{3n}{4}-c\right)$ where $\left[\frac{3n}{4}\right]$ in this case is the value of q_3

$$=27{\cdot}5+\frac{5}{14}\ (54{\cdot}75-43)$$

$$=27{\cdot}5\ +\frac{5}{14}\times 11{\cdot}75$$

$$=27{\cdot}5+4{\cdot}19=31{\cdot}69 \text{ approximately.}$$

Quartile Deviation $=\frac{31{\cdot}69-21{\cdot}19}{2}=\frac{10{\cdot}5}{2}=5{\cdot}25$ approx.

Coefficient of Quartile Deviation

$$=\frac{31{\cdot}69-21{\cdot}19}{31{\cdot}69+21{\cdot}19}=\frac{10{\cdot}5}{52{\cdot}88}={\cdot}19 \text{ approx.}$$

CHAPTER III

SKEWNESS AND KURTOSIS

Skewness is a lack of symmetry. For an asymmetric distribution the range is usually greater on one side of the mode than on the other. The curve is said to have a longer tail on one side than on the other."*

"Skewness is a term for the degree of distortion from symmetry. When a distribution is symmetrical, the values of the Mean, Median and Mode coincide......

"A large number of frequency distribution occurring in practice fall into four types :

(*i*) Symmetrical

(*ii*) Moderately skewed

(*iii*) Extremely skewed, or J-shaped

(*iv*) The 'U' shaped."†

Generally the following formulae are used to calculate the coefficient of skewness (j)

(*i*) Coefficient of Skewness $(j) = \dfrac{\text{Mean}-\text{Mode}}{\text{St. Deviation}}$

(*ii*) Co-efficient of Skewness $(j) = \dfrac{3(\text{Mean}-\text{Median})}{\sigma}$

(*iii*) $(j) = \dfrac{Q_3+Q_1-2\ \text{Median}}{Q_3-Q_1}$

Problem 108.—Find out coefficient of dispersion and a coefficient of skewness from the following table giving wages of 230 persons and explain their significance.

Wages Rs.	*No. of persons*	*Wages* Rs.	*No. of persons*
70—80	12	110—120	50
80—90	18	120—130	45
90—100	35	130—140	20
100—110	42	140—150	8

(*Agra, B.Com.*, 1940)

* *Practical Statistics*—Dr. J. C. Chaturvedi, p. 89.

† *Practical Statistics*—Zia-ud-din, pp. 69-70.

Solution :

Calculation of standard deviation

Wages in Rs.	mid-value X	No. of persons (f)	$u=\frac{X-A}{h}$ A=105 (assumed mean) h=10 (class-interval).	fu	fu^2
70— 80	75	12	—3	—36	108
80— 90	85	18	—2	—36	72
90—100	95	35	—1	—35	35
100—110	105	42	0	0	0
110—120	115	50	1	50	50
120—130	125	45	2	90	180
130—140	135	20	3	60	180
140—150	145	8	4	32	128
TOTAL		230		125	753

Arithmetic average, $M=A+h\frac{\Sigma fu}{\Sigma f}$

$$=105+10\times\frac{125}{230}$$

$$=110 \cdot 4 \text{ rupees.}$$

Standard deviation $=h\sqrt{\frac{\Sigma fu^2}{\Sigma f}-\left(\frac{\Sigma fu}{\Sigma f}\right)^2}$

$$=10\times\sqrt{\frac{753}{230}-\left(\frac{125}{230}\right)^2}$$

$$=17 \cdot 3 \text{ rupees.}$$

The location of modal group :

In this question when we study the series, we find that it has only one mode. Because, the series consists of ascending/descending order with one maximum frequency (50) as the turning point. By inspection we locate mode at this maximum point which lies in the interval (110—120). Now the exact point within the class-interval at which mode is, is given by the formula :—

$$M_0=L_1+\frac{f_1}{f_{-1}+f_1}\times I$$

where M_0 stands for mode in the modal class, L_1 stands for the lower limit of the modal class, f_{-1} stands for the frequency in the next preceding group (here $f_{-1}=42$), f_1 stands for the frequency in the next succeeding group (here $f_1=45$) and I stands for the class-interval.

$$\therefore \quad M_0=110+\frac{45}{42+45}\times 10$$

$$=115 \cdot 17$$

Now coefficient of skewness by Karl Pearson's formula is given by, $j=\frac{M-M_0}{\sigma}$ where j is coeff. of skewness.

$$\therefore \quad j=\frac{110{\cdot}4-115{\cdot}17}{17{\cdot}3}$$

$$=-{\cdot}275 \quad \text{Ans.}$$

Problem 109.—Find out the mean wage and a coefficient of skewness for the following :

Wages in Rs.	*No. of men*	*Wages in Rs.*	*No. of men*
4·5	35	8·5	125
5·5	40	9·5	87
6·5	48	10·5	43
7·5	100	11·5	22

Solution :

Calculation of standard Deviation.

Wages in Rs. X	No. of mem (f)	Deviation from the assumed average (7 5) ξ=X—A	$f\xi$	$f\xi^2$
4·5	35	—3	—105	315
5·5	40	—2	— 80	160
6·5	48	—1	— 48	48
7·5	100	0	0	0
8·5	125	1	125	125
9·5	87	2	174	348
10·5	43	3	129	387
11·5	22	4	88	352
TOTAL	500	TOTAL	283	1735

Mean wage, $M=A+\frac{\Sigma f\xi}{\Sigma f}$

$$=7{\cdot}5+\frac{283}{500}$$

$$=8{\cdot}66 \text{ rupees.}$$

Standard deviation, $\sigma=\sqrt{\frac{\Sigma f\xi^2}{\Sigma f}-\left(\frac{\Sigma f\xi}{\Sigma f}\right)^2}$

$$=\sqrt{\frac{1735}{500}-\left(\frac{283}{500}\right)^2}$$

$$=1{\cdot}77$$

Location of mode : Since the series consists of ascending/descending order with one maximum frequency (125) as the turning point, hence mode is located at this point and is equal to 8·5.

$$\therefore \qquad M_0 = 8{\cdot}5 \text{ rupees.}$$

$$\text{Coefficient of Skewness} = \frac{M - M_0}{\sigma}$$

$$= \frac{8{\cdot}069 - 8{\cdot}5}{1{\cdot}77}$$

$$= -{\cdot}245 \quad \textbf{Ans.}$$

Problem 110.—Compute the coefficient of skewness and variation for the following frequency distribution of wages:

Weekly wages in Rs.	*No. of men*
4·5—12·5	4
12·5—20·5	24
20·5—28·5	21
28·5—36·5	18
36·5—44·5	5
44·5—52·5	3
52·5—60·5	5
60·5—68·5	8
68·5—76·5	2
	90

Solution :

Calculation of standard deviation. $A = 32{\cdot}5$, $h = 8$

Weekly wages in Rs.	Mid-value (X)	No. of men f	$u = \frac{x - A}{h}$ A=assumed mean, h=class interval	fu	fu^2
4·5—12 5	8·5	4	—3	—12	36
12·5—20·5	16·5	24	—2	—48	96
20 5—28·5	24·5	21	—1	—21	21
28·5—36·5	32·5	18	0	0	0
36·5—44·5	40·5	5	1	5	5
44·5—52·5	48.5	3	2	6	12
52·5—60.5	56·5	5	3	15	45
60·5—68·5	64·5	8	4	32	128
68·5—76 5	72·5	2	5	10	50
		90		—13	393

$$\text{Mean wage } M = A + h\frac{\Sigma f u}{\Sigma f}$$

$$= 32{\cdot}5 + 8 \times \frac{-13}{90}$$

$$= 30{\cdot}35 \text{ rupees.}$$

Standard Deviation, $\sigma = h\sqrt{\frac{\Sigma f u^2}{\Sigma f} - \left(\frac{\Sigma f u}{\Sigma f}\right)^2}$

or $\sigma = 8\sqrt{\frac{393}{90} - \left(\frac{-13}{90}\right)^2}$

$= 13{\cdot}92$

Location of modal group : Since the series is multimodal as there are more than one point of maximum frequencies (24 and 8) hence the modal group will be located by the method of grouping.

Weekly wages in Rs.	No. of men (frequency) (1)	(2)	(3)	(4)	(5)	(6)	No. of times a group contains maximum frequency
4·5—12·5	4	28		**49**			1
12·5—20·5	**24**		**45**		**63**		4
20·5—28·5	21	**39**				**44**	5
28·5—36·5	18		23	26			3
36·5—44·5	5	8			13		1
44·5—52·5	3		8			16	0
52·5—60·5	5	13		15			0
60·5—68·5	**8**		10				1
68·5—76·5	2						0

Maximum frequencies in each column is under-lined. From above it is quite clear that the modal-class is (20·5—28·5).

By interpolation, $M_0 = L_1 + \frac{f_1}{f_{-1} + f_1} \times \bar{I}$

$= 20{\cdot}5 + \frac{18}{24 + 18} \times 8$

$= 23{\cdot}93.$

Coeff. of skewness $j = \frac{M - M_0}{\sigma} = \frac{30{\cdot}35 - 23{\cdot}93}{13{\cdot}92} = {\cdot}46$

Coeff. of variation $= \frac{\sigma}{M} \times 100 = \frac{13{\cdot}92 \times 100}{30{\cdot}35} = 45{\cdot}8$ **Ans.**

Problem 111.—Find the coefficient of skewness of the two groups given below and point out which distribution is more skew ?

Marks	*Group A*	*Group B*
55—58	12	20
58—61	17	22
61—64	23	25
64—67	18	13
67—70	11	7

(*M.A., Agra, 1954*)

Solution :

Calculation of Standard Deviation.

Marks	Mid-Value (X)	Step deviation $u=\frac{X-62\cdot5}{3}$	Group A			Group B		
			f	fu	fu^2	f	fu	fu^2
55—58	56·5	—2	12	—24	48	20	—40	80
58—61	59·5	—1	17	—17	17	22	—22	22
61—64	62·5	0	23	0	0	25	0	0
64—67	65·5	1	18	18	18	13	13	13
67—70	68·5	2	11	22	44	7	14	28
		TOTAL	81	—1	127	87	—35	143

Group A. By inspection modal class is (61—64)

$$\therefore \quad M_0=61+\frac{18}{17+18}\times 3$$

$$=62\cdot54$$

Mean, M $=62\cdot5+\frac{-1}{81}\times 3=62\cdot463$

Standard deviation σ $=3\times\sqrt{\frac{127}{81}-\left(\frac{-1}{81}\right)^2}$

$=3\cdot75$

Coefficient of skewness $=\frac{62\cdot46-62\cdot54}{3\cdot75}$

$=-\cdot02$ **Ans.**

Group B. By inspection modal-class is 61—64

$$\therefore \quad M_0=61+\frac{13}{22+13}\times 3$$

$$=62\cdot1$$

Mean, M $=62\cdot5+\frac{-35}{87}\times 3=61\cdot3$

Standard deviation σ $=3\times\sqrt{\frac{143}{87}-\left(\frac{-35}{87}\right)^2}$

$=3\cdot6$

Coefficient of skewness $\sigma=\frac{61\cdot3-62\cdot1}{3\cdot6}$

$=-\cdot22$ **Ans.**

Hence group B is more skewed.

✓**Problem 112.**—Find the mean, mode, standard deviation and a coefficient of skewness for the following :

Years under	10,	20,	30,	40,	50,	60.
No. of persons	15,	32,	51,	78,	97,	109.

(*P.C.S., 1952*)

Solution :

Calculation of standard deviation.

Years	Mid-Value X	No. of persons f	Step deviation $u=\frac{X-35}{10}$	fu	fu^2
0—10	5	15	—3	—45	135
10—20	15	17	—2	—34	68
20—30	25	19	—1	—19	19
30—40	35	27	0	0	0
40—50	45	19	1	19	19
50—60	55	12	2	24	48
	TOTAL	109	TOTAL	—55	289

Mean $M=35+10\times\frac{-55}{109}$

$=29{\cdot}95$

Standard Deviation $\sigma=10\times\sqrt{\frac{289}{109}-\left(\frac{-55}{109}\right)^2}$

$=15{\cdot}49$

The modal-group is by inspection 30—40.

$\therefore \quad M_o=30+\frac{19}{19+19}\times 10$

$=35$

Co-efficient of skewness $=\frac{29{\cdot}95-35}{15{\cdot}49}$

$=\frac{-5{\cdot}05}{15{\cdot}49}$

$=-{\cdot}32$ **Ans.**

✓**Problem 113.**—Explain the meaning and significance of skewness. Which of the following two distributions is more skew ?

Distribution of weekly index Nos. of cost of living in Bombay 1942		Distribution of weekly index Nos. of cost of living in Bombay 1943	
Index No.	No. of weeks	Index No.	No. of weeks
140—150	5	200—210	10
150—160	10	210—220	10
160—170	20	220—230	10
170—180	9	230—240	8
180—190	6	240—250	7
190—200	2	250—260	7

(*M. Com. Agra, 1950*)

Solution :

For the year 1942.

Calculation of Standard Deviation. A=165, h=10

Index No.	Mid-value X	No. of weeks f	Step deviation $u=\frac{X-A}{h}$	fu	fu^2
140—150	145	5	−2	−10	20
150—160	155	10	−1	−10	10
160—170	165	20	0	0	0
170—180	175	9	1	9	9
180—190	185	6	2	12	24
190—200	195	2	3	6	18
	TOTAL	52		7	81

$$\text{Mean, } M = A + h\,\frac{\Sigma fu}{\Sigma f}$$

$$= 165 + 10 \times \frac{7}{52}$$

$$= 166{\cdot}34$$

Standard deviation,

$$\sigma = h.\sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2}$$

$$= 10 \times \sqrt{\frac{81}{52} - \left(\frac{7}{52}\right)^2}$$

$$= 12{\cdot}3$$

For the year 1943.

Calculation of Standard Deviation

A=225 h=10

Index No.	Mid-value X	No. of weeks f	Step deviation $u=\frac{X-A}{h}$	fu	fu^2
200—210	205	10	−2	−20	40
210—220	215	10	−1	−10	10
220—230	225	10	0	0	0
230—240	235	8	1	8	8
240—250	245	7	2	14	28
250—260	255	7	3	21	63
	TOTAL	52	TOTAL	13	149

$$\text{Mean, } M = 225 + 10 \times \frac{13}{52}$$

$$= 227{\cdot}5$$

Standard Deviation

$$\sigma = 10 \times \sqrt{\frac{149}{52} - \left(\frac{13}{52}\right)^2}$$

$$= 16{\cdot}7$$

Location of modal class.

For the year 1942 by inspection the mode lies in the class-interval (160—170).

For the year 1943 the mode is located by grouping.

Index No. Class	Frequency (1)	(2)	(3)	(4)	(5)	(6)	No. of times a group contained maximum frequency
200—210	10	**20**		**30**			3
210—220	10		**20**				5
220—230	10	18			**28**	**25**	5
230—240	8		15				2
240—250	7	14		22			1
250—260	7						0

Hence the mode is not clearly defined.

Therefore median is to be determined to get coefficient of skewness.

Determination of median (M*d*)

1942 Class	*f*	Cumulative frequency	1943 Class	*f*	Cumulative frequency
140—150	5	5	200—210	10	10
150—160	10	15	210—220	10	**20**
160—170	20	35	220—230	10	30
170—180	9	44	230—240	8	38
180—190	6	50	240—250	7	45
190—200	2	52	250—260	**7**	52

Median is computed by the use of the following formula.

$$Md = L_1 + \frac{\frac{N}{2} - F}{f} \times I$$

where Md is median, L is lower limit of median class, f is frequency of median class, I is class interval of median class, F is the total of all frequencies before the median class and N is the total of frequencies.

$$\therefore \text{ For 1942} \qquad Md = 160 + \frac{26-15}{20} \times 10$$

$$= 165{\cdot}5$$

$$\text{For 1943} \qquad Md = 220 + \frac{26-20}{10} \times 10$$

$$= 226.$$

For 1942. Coeff. of skewness

$$j = \frac{3(M - Md)}{\sigma}$$

$$= \frac{3(166{\cdot}3 - 165{\cdot}5)}{12{\cdot}3}$$

$$= {\cdot}19 \text{ } \mathbf{Ans.}$$

For 1943. Coeff. of skewness

$$j = \frac{3(M - Md)}{\sigma}$$

$$= \frac{3(227{\cdot}5 - 226)}{16{\cdot}7}$$

$$= {\cdot}27 \text{ } \mathbf{Ans.}$$

Thus, the coeff. of skewness is higher in the year 1943.

✓**Problem 114.**—Find the coefficient of skewness from the following data :

Heights of school boys at age 5.

Height in inches	No.	Height in inches	No.	Height in inches	No.	Height in inches	No.
28	1	34	43	40	1670	46	89
29	0	35	59	41	1614	47	27
30	1	36	166	42	1541	48	19
31	3	37	344	43	1028	49	4
32	8	38	740	44	567	50	4
33	13	39	1167	45	233	51	1
continued	continued	continued	continued	continued	continued		

(*M.A., Bombay, 1935*)

Solution :

Calculation of standard-deviation.

Height in inches X	No. of boys f	Deviation from assumed average (40) ξ	$f\xi$	$f\xi^2$
28	1	—12	—12	144
29	0	—11	0	0
30	1	—10	—10	100
31	3	— 9	—27	243
32	8	— 8	—64	512
33	13	— 7	—91	637
34	40	— 6	—240	1440
35	59	— 5	—295	1475
36	166	— 4	—664	2656
37	344	— 3	—1032	3096
38	740	— 2	—1480	2960
39	1167	— 1	—1167	1167
40	1670	0	0	0
41	1614	1	1614	1614
42	1541	2	3082	6164
43	1028	3	3084	9252
44	576	4	2268	9072
45	233	5	1165	5825
46	89	6	534	3204
47	27	7	189	1323
48	19	8	152	1216
49	4	9	36	324
50	4	10	40	400
51	1	11	11	121
TOTAL	9339		7093	52945

Mean height, $M = 40 + \frac{7093}{9339} = 40{\cdot}75$

Standard deviation—

$$\sigma = \sqrt{\frac{52945}{9339} - \left(\frac{7093}{9339}\right)^2}$$
$$= 2{\cdot}26 \text{ inch.}$$

Location of mode by grouping.

Height in inches x	No. of frequencies (1)	(2)	(3)	(4)	(5)	(6)	No. of times an item faces maximum frequency
28	1	1					0
29	0		1	2			0
30	1	4			4		0
31	3		11			12	0
32	8	21		24			0
33	13		53		61		0
34	40	99				112	0
35	59		225	265			0
36	166	510			569		0
37	344		1084			1250	0
38	740	1907		2251			1
39	1167		2837		3577		4
40	**1670**	**3284**				**4451**	5
41	1614		**3155**	**4825**			3
42	1541	2569			**4183**		1
43	1028		1595			3136	0
44	567	800		1828			0
45	233		322		889		0
46	89	116				349	0
47	27		46	135			0
48	19	23			50		0
49	4		8			27	0
50	4	5		9			0
51	1						0

Hence mode is 41 inches.

Coefficient of skewness

$$j=\frac{M-M_0}{\sigma}$$

$$=\frac{40{\cdot}75-41}{2{\cdot}26}$$

$$= -{\cdot}106. \textbf{ Ans.}$$

Problem 115.—Calculate coefficient of skewness from the following data :

Variable : Marks above	0	10	20	30	40	50	60	70	80
Frequency :	150	140	100	80	80	70	30	14	0

(Raj., M.A., 1959)

Solution :

Marks (X)	Frequency (f)	Xf	$u=\frac{X-A}{h}$	u^2	fu	fu^2
0	150	0	−4	16	0	0
10	140	1400	−3	9	−420	1260
20	100	2000	−2	4	−200	400
30	80	2400	−1	1	− 80	80
40	80	3200	0	0	0	0
50	70	3500	+1	1	+ 70	70
60	30	1800	+2	4	+ 60	120
70	14	980	+3	9	+ 42	126
80	0	0	+4	16	0	0
	$n=664$	$\Sigma Xf=$ 15280			$\Sigma fu=$ −528	$\Sigma fu^2=$ 2056

Let the assumed mean (A) = 40 $h=10$ (class interval)

Arith. Average (M) $=\frac{\Sigma Xf}{n}$

$=\frac{15280}{664}$

= 23 approximately.

Standard Deviation

$$(\sigma)=h\sqrt{\frac{\Sigma fu^2}{\Sigma f}-\left(\frac{\Sigma fu}{\Sigma f}\right)^2}$$

$$=10\sqrt{\frac{2056}{664}-\left(\frac{-528}{664}\right)^2}$$

$$=\frac{10}{664}\sqrt{(2056\times664)-(-528\times-528)}$$

$$=\frac{10}{664}\sqrt{1092376}$$

$$=\frac{10\times1078}{664}\text{ approx.}$$

= 16·2 approx.

M_0 by inspection is found to be located at 10 with freq. 140 because at 0 it cannot be

∴ Coefficient of skewness

$$(j)=\frac{M-M_0}{\sigma}$$

$$=\frac{23-10}{16{\cdot}2}\text{ approx.}$$

$$=\frac{13}{16{\cdot}2}\text{ approx.}$$

=·8 approx. **Ans.**

✓Problem 116.—From the following table calculate the coefficient of Skewness.

Age at the birth of first child :
13 14 15 16 17 18 19 20 21 22 23 24 25
No. of married women :
37 162 343 390 256 433 161 355 65 85 49 46 40
(*Data from Raj, M.A., 1955*)

Solution :

Age at the birth of first child (X)	No. of married women (*f*)	X*f*	ξ (X—M)	ξ²	*f*ξ²
13	37	481	—5	25	925
14	162	2168	—4	16	2592
15	343	5145	—3	9	3087
16	390	6240	—2	4	1560
17	256	4352	—1	1	256
18	433	7794	0	0	0
19	161	3059	+1	1	161
20	355	7100	+2	4	1420
21	65	1365	+3	9	585
22	85	1870	+4	16	1360
23	49	1127	+5	25	1225
24	46	1104	+6	36	1656
25	40	1000	+7	49	1960
	n=2422	ΣX*f*= 42805			Σ*f*ξ² =16787

$$M=\frac{\Sigma Xf}{n}=\frac{42805}{2422}=17{\cdot}6=18 \text{ approx for finding } \xi$$

$$\text{Standard Deviation } (\sigma)=\sqrt{\frac{\Sigma f.\xi^2}{n}}$$

$$=\sqrt{\frac{16787}{2422}}$$

$$=\sqrt{6{\cdot}9}$$

$$=2{\cdot}6 \text{ approx.}$$

M_0 by inspection=18.

$$\therefore \text{ Coefficient of Skewness } (j)=\frac{M-M_0}{\sigma}$$

$$=\frac{17{\cdot}6-18}{2{\cdot}6}$$

$$=\frac{-{\cdot}4}{2{\cdot}6}$$

$$=-{\cdot}15 \text{ approx.}$$ **Ans.**

Problem 117.—Calculate the coefficient of skewness of the following series :

Size of item	*Frequency*	*Size of item*	*Frequency*
2	3	8	10
3	8	9	8
4	10	10	17
5	12	11	5
6	16	12	4
7	14	13	1

(*Statistics : Theory and Practice*—Ghosh and Choudhry, p. 174.)

Solution :

Size of Item (X)	Frequency (f)	fX	cmf	$\lvert d_{md}\rvert$	$\lvert fd_{md}\rvert$
2	3	6	2	5	15
3	8	24	11	4	32
4	10	40	21	3	30
5	12	60	33	2	24
6	16	96	49	1	16
7	14	98	63	0	0
8	10	80	73	1	10
9	8	72	81	2	16
10	17	170	98	3	51
11	5	55	103	4	20
12	4	48	107	5	20
13	1	13	108	6	6
	$n=108$	$\Sigma fX=762$			$\Sigma\lvert f^{d}_{md}\rvert=240$

Mean (M) $=\frac{\Sigma fX}{n}=\frac{762}{108}=7{\cdot}05$ approximately.

Median (Md) $=$size of $\left(\frac{n+1}{2}\right)^{th}$ item.

$=$,, ,, $\left(\frac{108+1}{2}\right)^{th}$ item.

$=$,, ,, $54{\cdot}5^{th}$. item.

$=7$ approximately.

Mean Deviation from the median $(\delta Md)=\frac{\Sigma\lvert fd_{md}\rvert}{n}$

$=\frac{240}{108}$

$=2{\cdot}2$ approximately.

$$\text{Coefficient of skewness } (j) = \frac{M - Md}{\delta Md}$$

$$= \frac{7{\cdot}05 - 7}{2{\cdot}2}$$

$$= {\cdot}022 \text{ approximately}$$

Problem 118.—Find out the Standard Deviation and a coefficient of skewness for the following distribution :

Variable	*Frequency*	*Variable*	*Frequency*
0— 5	2	20—25	21
5—10	5	25—30	16
10—15	7	30—35	8
15—20	13	35—40	3

(*Mathematical Statistics*—Dr. J.C. Chaturvedi, p 150.)

Solution :

In the calculation of the Mean and the Standard Deviation we take for each class its mid-value as the value of X.

Variable	Mid-Value (X)	Frequency (f)	ξ	Or ξ can be written as	$f\xi$	$f\xi^2$
0— 5	2·5	2	−20	−4×5	−8×5	32×25
5—10	7·5	5	−15	−3×5	−15×5	45×25
10—15	12·5	7	−10	−2×5	−14×5	28×25
15—20	17·5	13	−5	−1×5	−13×5	13×25
20—25	22·5	21	0	0	0	0
25—30	27·5	16	5	1×5	16×5	16×25
30—35	32·5	8	10	2×5	16×5	32×25
35—40	37·5	3	15	3×5	9×5	27×25
		$n=75$			$\Sigma\xi=$ [−9×5]	$\Sigma f\xi^2=$ [193×25]

Mean $= 22 - \frac{45}{75} = 21{\cdot}9$ approximately.

and $\sigma^2 = \frac{193 \times 25}{75} - \left(\frac{-9 \times 5}{75}\right)^2$

$\therefore$ $\sigma = 8{\cdot}0$ approximately.

For Mode

$$M_0 = L + \frac{f_1}{f_1 + f_2} \times i$$

where $L = 20$
$f_1 = 16$
$f_2 = 13$
$i = 5$

$$= 20 + \frac{16}{13 + 16} \times 5$$

$$= 22{\cdot}76 \text{ approximately.}$$

$\therefore$ Co-eff. of skewness $(j)=\dfrac{M-M_0}{\sigma}$

$$=\frac{21\cdot9-22\cdot76}{8}$$

$=-\cdot1$ approximately. **Ans.**

✓**Problem 119.**—Find the Mean, Mode, Standard Deviation and a Coefficient of skewness for the following :

Years Under	10	20	30	40	50	60
No. of persons	15	32	51	78	97	109

(*Mathematical Statistics*—Dr. J. C. Chaturvedi, p. 167.)

Solution :

Years	Mid-Value (X)	Freq. (f)	Cumulative Frequency (cmf)	ξ (X−A) A=35	$f\xi$	$f\xi^2$
0—10	5	15	15	—30	—45×10	135×100
10—20	15	17	32	—20	—34×10	68×100
20—30	25	19	51	—10	−19×10	19×100
30—40	35	27	78	0	0	0
40—50	45	19	97	10	19×10	19×100
50—60	55	12	109	20	24×10	48×100
			$n=109$		−55×10	289×100

$$\text{Mean}=A+\frac{\Sigma f\xi}{n}=35+\frac{-35\times10}{109}=29{\cdot}96 \text{ app.}$$

$$\text{Mode}=L+\frac{f_1}{f_1+f_2}\times i \quad (\text{where } L=30, f_1=19, f_2=19, i=10)$$

$$=30+\frac{19}{19+19}\times10=35. \text{ app.}$$

$$\sigma^2=\frac{\Sigma f\xi^2}{\Sigma f}-\left(\frac{\Sigma f\xi}{n}\right)^2$$

$=265\cdot13-25\cdot4$

$=239\cdot73$ appr.

$\therefore\ \sigma$ $=15\cdot47$ appr.

Co-efficient of skewness $(j)=\dfrac{M-M_0}{\sigma}$

$$=\frac{29\cdot96-35}{15\cdot47}$$

$=-\cdot3$ approx. **Ans.**

Problem 120. Find Mean, Mode, Standard Deviation and Co-efficient of skewness for the following :

Class	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
Frequency	2	5	7	13	21	16	8	3

(*Mathemetical Statistics*—Dr. J. C. Chaturvedi, p. 168.)

Solution :

Class	Frequency (f)	Mid-Value (X)	ξ A=22·5	$f\xi$	$f\xi^2$
0—5	2	2·5	—20	—40	800
5—10	5	7·5	—15	—75	1125
10—15	7	12·5	—10	—70	700
15—20	13	17·5	— 5	—65	325
20—25	21	22·5	0	0	0
25—30	16	27·5	5	80	400
30—35	8	32·5	10	80	800
35—40	3	37·5	15	45	675
	n=75			$\Sigma f\xi=-45$	$\Sigma f\xi^2=4825$

Mean $=A+\frac{\Sigma f\xi}{n}=22·5+\frac{-45}{75}$

$=22·5-·6=21·9$ approx.

M_o $=L+\frac{f_1}{f_1+f_2}\times i=20+\frac{16}{16+13}\times 5$

$=20+\frac{16}{29}\times 5=20+2·7=22·7$ approx.

σ^2 $=\frac{\Sigma f\xi^2}{n}-\left(\frac{\Sigma f\xi}{n}\right)^2$

$=\frac{4825}{75}-\left(\frac{-45}{75}\right)^2$

$=64·3-·36=63·94$

$\therefore\ \sigma$ $=\sqrt{63·94}=8$ approx.

$\therefore$ Coefficient of skewness $=\frac{M-M_o}{\sigma}$

$=\frac{21·9-22·7}{8}$

$=-·1$ approximately. **Ans.**

Problem 121.—Find the Lower and Upper Quartiles Median and a Coefficient of skewness of the following :

Variable	*No. of students*	*Variable*	*No. of students*
0—10	15	40—50	12
10—20	20	50—60	31
20—30	25	60—70	71
30—40	24	70—80	52

(From two solved problems done *by Dr. J. C. Chaturvedi in 'Practical Statistics'* pp. 74 and 101.)

Solution :

Variable	Mid-Value	Frequency	Cumulative frequency
0—10	5	15	15
10—20	15	20	35
20—30	25	25	60
30—40	35	24	84
40—50	45	12	96
50—60	55	31	127
60—70	65	71	198
70—80	75	52	250

Q_1 $=30+\frac{62{\cdot}5-60}{24}\times 10$. Because $\frac{N}{4}=62{\cdot}5$

$=31{\cdot}04$ app.

Q_3 $=60+\frac{187{\cdot}5-127}{71}\times 10$. Because $\frac{3N}{4}=187{\cdot}5$

$=68{\cdot}52$

M*d* $=50+\frac{125-96}{31}\times 10$

$=59{\cdot}3.$

Coefficient of skewness $(j)=\frac{Q_3+Q_1-2Md}{Q_3-Q_1}$

$=\frac{68{\cdot}52+31{\cdot}04-118{\cdot}6}{68{\cdot}52-31{\cdot}04}$

$=-{\cdot}5$ Approx. **Ans.**

CHAPTER IV

CONTINUOUS VARIABLE

Probability density function : If the probability that the variable x lies in the interval α_1 to α_2 is given by

$$\int_{\alpha_1}^{\alpha_2} f(x)\,dx$$

when the total area under the curve given by $y=f(x)$ is unity, then $y=f(x)$ is called the probability density function of x. The interval can be finite or infinite and the function $f(x)$ is non negative.

Definitions :—

For a probability density function $y=f(x)$ in between the limits $-\infty$ to $+\infty$

(*i*) The mode is the value of x for which the frequency y is maximum *i.e.*, where

$$\frac{dy}{dx}=0 \text{ and } \frac{d^2y}{dx^2} \text{ is negative}$$

(*ii*) The Median is the value of x which divides the total frequency into two equal parts *i.e.*, if x is the value of the median, it is given by

$$\int_{-\infty}^{x} f(x)dx=\int_{x}^{\infty} f(x)dx=\tfrac{1}{2} \quad \int_{-\infty}^{\infty} f(x)dx=\tfrac{1}{2}$$

(*iii*) The Mean is obtained by finding the x—coordinate of the centre of gravity of the area between the curve and the x—axis *i.e.*, by

$$M=\int_{-\infty}^{\infty} xf(x)\,dx$$

(*iv*) The Geometric Mean G given by the relation

$$\log G=\int_{-\infty}^{\infty} f(x)\log x\,dx$$

(*v*) The Harmonic Mean H is given by the relation

$$\frac{1}{H}=\int_{-\infty}^{\infty}\frac{1}{x}f(x)\,dx$$

(*vi*) The value of $\mu'n$ (n^{th} moment about zero) is given by

$$\mu'n=\int_{-\infty}^{\infty} x^n f(x)\,dx \qquad \text{Since } \xi=x-0=x$$

If $\mu'n$ is the moment about any arbitrary point a, its value is given by

$$\mu'n = \int_{-\infty}^{\infty} (x-a)^n f(x)\, dx.$$

Note :—(*i*) Since arbitrary origin A is zero the first moment μ'_1 will clearly be equal to the mean of x.

Since μ'_1 is the mean, the moments about the mean are given by the relation.

$$\mu n = \int_{-\infty}^{\infty} (x-\mu'_1)^n f(x) dx.$$

$\therefore$ Standard deviation $\sigma = \sqrt{\mu_2}$

and $\mu_2 = \mu'_2 - (\mu'_1)^2$

(*ii*) If mean coincides with mode then the distribution is symmetrical.

Problem 122.—Find the harmonic mean for the distribution given by

$$df = \frac{1}{\beta(m,n)} (1-x)^{m-1} x^{n-1} dx \text{ for } o \leqslant x \leqslant 1$$

(*P.C.S., 1952*)
(*M.Sc. Agra, 1952, 1956*)

Solution :

The total frequency $f = \frac{1}{\beta(m,n)} \int_0^1 (1-x)^{m-1} x^{n-1} dx$

$$= \frac{1}{\beta(m,n)} \cdot \beta(m,n)$$

$$= 1.$$

$\therefore$ the function is probability density function.

The harmonic mean, H is given by

$$\frac{1}{H} = \frac{1}{\beta(m,n)} \int_0^1 \frac{1}{x} (1-x)^{m-1} x^{n-1} dx$$

$$= \frac{1}{\beta(m,n)} \int_0^1 (1-x)^{m-1} x^{n-2} dx$$

$$=\frac{\beta(m,n-1)}{\beta(m,n)}$$

$$=\frac{\Gamma m\ \Gamma(n-1)}{\Gamma(m+n-1)}\cdot\frac{\Gamma(m+n)}{\Gamma m\ \Gamma n}$$

$$=\frac{m+n-1}{n-1}$$

$$\therefore \quad H=\frac{n-1}{m+n-1}.$$

Problem 123.—Find the arithmetic mean and standard deviation for the distribution given by

$$df=\frac{1}{\beta(m,n)}\ (1-x)^{m-1}\ .\ x^{n-1}\ dx$$

for $o \leqslant x \leqslant 1$

(M.Sc. Agra, 1954, 1956)

Solution :

$$\text{Total frequency } f=\frac{1}{\beta(m,\ n)}\int_0^1 (1-x)^{m-1}\ .\ x^{n-1}\ dx$$

$$=1$$

The arithmetic mean, M, is given by μ'_1

$$\therefore \quad M=\frac{1}{\beta(m,n)}\int_0^1 x.\ (1-x)^{m-1}\ x^{n-1}\ .\ dx$$

$$=\frac{1}{\beta(m,n)}\int_0^1 (1-x)^{m-1}\ x^n\ dx$$

$$=\frac{\beta(m,n+1)}{\beta(m,n)}$$

$$=\frac{\Gamma m\ \Gamma(n+1)}{\Gamma(m+n+1)}\cdot\frac{\Gamma(m+n)}{\Gamma m\ \Gamma n}$$

$$=\frac{n}{m+n}.$$

Since standard deviation $\sigma=\sqrt{\mu'_2-(\mu'_1)^2}$

$$\text{where} \quad \mu'_2=\frac{1}{\beta(m,n)}\int_0^1 x^2.\ (1-x)^{m-1}\ x^{n-1}\ dx$$

[Taking the origin at 0]

$$= \frac{1}{\beta(m,n)} \int_0^1 (1-x)^{m-1} x^{n+1} dx$$

$$= \frac{\beta(m,n+2)}{\beta(m,n)}$$

$$= \frac{\Gamma m \, \Gamma n+2}{\Gamma(m+n+2)} \cdot \frac{\Gamma(m+n)}{\Gamma m \, \Gamma n}$$

$$= \frac{n(n+1)}{(m+n+1)(m+n)}$$

Also $\mu'_1 = \frac{n}{m+n}$

$\therefore$

$$\sigma = \sqrt{\mu_2} = \sqrt{\mu'_2 - (\mu'_1)^2}$$

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

$$= \frac{1}{(m+n)} \sqrt{\frac{n(m+n)(n+1) - n^2(m+n+1)}{m+n+1}}$$

$$= \frac{1}{(m+n)} \sqrt{\frac{mn}{m+n+1}}$$

Problem 124. Find the geometric mean for the distribution

$$df = \frac{1}{\beta(m,n)} (1-x)^{m-1} x^{n-1} dx \quad o \leqslant x \leqslant 1$$

Solution :

Total frequency $f = \frac{1}{\beta(m,n)} \int_0^1 (1-x)^{m-1} x^{n-1} dx$

$= 1$

$\therefore$ The geometric mean G is given by

$$\log G = \frac{1}{\beta(m,n)} \int_0^1 (1-x)^{m-1} x^{n-1} \log x \, dx$$

Now $\int_0^1 (1-x)^{m-1} x^{n-1} dx = \beta(m,n)$

differeniating both sides with respect to n we have

$$\int_0^1 (1-x)^{m-1} x^{n-1} \log x \, dx = \frac{\partial}{\partial n} \beta(m,n)$$

$$\therefore \quad \log G = \frac{1}{\beta(m,n)} \frac{\partial}{\partial n} \beta(m,n)$$

$$= \frac{\partial}{\partial n} \log \beta(m,n)$$

$$= \frac{\partial}{\partial n} \log \frac{\Gamma m \; \Gamma n}{\Gamma(m+n)}$$

$$= \frac{\partial}{\partial n}\left[\log \Gamma m + \log \Gamma n - \log \Gamma (m+n)\right]$$

Problem 125.—Show that for the rectangular population

$$df = dx, \qquad 0 \leqslant x \leqslant 1$$

μ'_1 (about the origin)$=\frac{1}{2}$,

and $\mu_2=\frac{1}{12}$

and mean deviation$=\frac{1}{4}$

(*I.A.S., 1950*)

Solution :

$$f \text{(total frequency)} = \int_0^1 dx = 1$$

Hence the function is probability density function ;

now taking the assumed average zero

$$\mu'_1 = \int_0^1 x\, dx$$

$$= \left(\frac{x^2}{2}\right)_0^1$$

$$= \tfrac{1}{2}$$

and

$$\mu'_2 = \int_0^1 x^2\, dx$$

$$= \left(\frac{x^3}{3}\right)_0^1$$

$$= \tfrac{1}{3}$$

$$\therefore \quad \mu_2 = \mu'_2 - (\mu'_1)^2$$

$$= \tfrac{1}{3} - \tfrac{1}{4}$$

$$= \tfrac{1}{12}$$

Since μ_1' is the mean as assumed mean is zero

$$\therefore \quad \text{Mean deviation} = \int_0^1 |(x-\mu'_1)|\, dx \qquad \text{[taking +ive sign of deviation)}$$

$$= \tfrac{1}{4}$$

$$\therefore \quad \mu_1' = \tfrac{1}{2},\ \mu_2 = \tfrac{1}{12},\ \text{mean deviation} = \tfrac{1}{4}$$

Problem 126.—Find the mode and the median for the frequency curve $y=\frac{1}{2}\sin x$, for $0 \leqslant x \leqslant \pi$

(*M.Sc. Agra 1951*)

Solution :

Clearly the total frequency $= \int_0^{\pi} \frac{1}{2} \sin x \, dx = 1$ and hence the function is a probability density function.

The mode is given by

$$\frac{dy}{dx} = 0 = \frac{1}{2} \cos x$$

$$\therefore \quad x = \frac{\pi}{2}$$

$$\text{Also } \left(\frac{d^2y}{dx^2}\right) = -\frac{1}{2} \sin x$$

$$\left(\frac{d^2y}{dx^2}\right)_{x=\frac{\pi}{2}} = -\frac{1}{2}\left[\text{ negative when } x = \frac{\pi}{2}\right]$$

$\therefore$ Mode is given by $x = \frac{\pi}{2}$

The median is given by x if

$$\int_0^x \frac{1}{2} \sin x \, dx = \frac{1}{2}\int_0^{\pi} \frac{1}{2} \sin x \, dx$$

or $\quad (\cos x - 1) = \frac{1}{2}[-2]$

or $\quad \cos x = 0$

$$\therefore \quad x = \frac{\pi}{2}$$

$\therefore$ Median is also given by $x = \frac{\pi}{2}$

Problem 127.—What is frequency curve ? A frequency function in the range (—3, 3) is defined by

$$y = \tfrac{1}{16}(3+x)^2, \quad -3 \angle x \angle -1$$
$$y = \tfrac{1}{16}(6-2x^2), \quad -1 \angle x \angle 1$$
$$\text{and } y = \tfrac{1}{16}(3-x)^2, \quad 1 \angle x \angle 3$$

Find the mean and the standard deviation of the distribution.

(*M.Sc. Agra 1948*)

Solution :

Since total frequency $= \int_{-3}^{-1} \frac{1}{16}(3+x)^2 \, dx + \int_{-1}^{1} \frac{1}{16}(6-2x^2) \, dx + \int_{1}^{3} \frac{1}{16}(3-x)^2 dx$

$$= \left[\tfrac{1}{16}\frac{(3+x)^3}{3}\right]_{-3}^{-1}+\left[\tfrac{1}{16}[6x-\tfrac{2}{3}x^3]\right]_{-1}^{1}+\left[-\tfrac{1}{16}\frac{(3-x)^3}{3}\right]_{1}^{3}$$

$$= \frac{1}{6}+\frac{2}{3}+\frac{1}{6}$$

$$=1$$

Hence the function is probability density function.

$\therefore$ The mean M is given by

$$M=\int_{-3}^{-1}\tfrac{1}{16}(3+x)^2 x\;dx+\int_{-1}^{1}\tfrac{1}{16}(6-2x^2)\,xdx+\int_{1}^{3}\tfrac{1}{16}(3-x)^2.\,xdx$$

$$=\tfrac{1}{16}\left[\frac{9x^2}{2}+\frac{6x^3}{3}+\frac{x^4}{4}\right]_{-3}^{-1}+0+\tfrac{1}{16}\left[\frac{9x^2}{2}-\frac{6x^3}{3}+\frac{x^4}{4}\right]_{1}^{3}$$

$$=\tfrac{1}{16}\left[\frac{9}{2}-2+\frac{1}{4}-\frac{81}{2}+54-\frac{81}{4}\right]+\tfrac{1}{16}\left[\frac{81}{2}-54+\frac{81}{4}-\frac{9}{2}+2-\tfrac{1}{4}\right]$$

$$=0$$

$\therefore\quad \mu_1'=0 \qquad$ as $\qquad M=\mu_1'$

and $\quad \mu_2=\mu_2'-\mu_1'^2=\mu_2'$

$$\therefore\quad \mu_2=\int_{-3}^{-1}\tfrac{1}{16}(3+x)^2\,x^2dx+\tfrac{1}{16}\int_{-1}^{1}(6-2x^2).\,x^2dx+\tfrac{1}{16}\int_{1}^{3}(3-x)^2.\,x^2dx$$

$$=\tfrac{1}{16}\left[3x^3+\tfrac{3}{2}x^4+\tfrac{1}{5}x^5\right]_{-3}^{-1}+\tfrac{1}{16}\left[2x^3-\tfrac{2}{5}x^5\right]_{-1}^{1}+\tfrac{1}{16}\left[3x^3-\tfrac{3}{2}x^4+\tfrac{1}{5}x^5\right]_{1}^{3}$$

$$\mu_2=\frac{1}{16}\left[-3+\frac{3}{2}-\frac{1}{5}+81-\frac{243}{2}+\frac{243}{5}\right]+\frac{1}{16}\left[4-\frac{4}{5}\right]+\frac{1}{16}\left[81-3-\frac{243}{2}+\frac{3}{2}+\frac{243}{5}-\frac{1}{5}\right]$$

$$=\frac{1}{16}\left[-84+100\right]$$

$$=1$$

$\therefore\quad$ Standard deviation $\sigma=\sqrt{\mu_2}=1$

$\therefore\quad$ Mean$=0$

Standard deviation$=1$

Problem 128.—Show that for the distribution

$$df = y_0 e^{-\frac{x^2}{2}} x^{n-1} dx \qquad 0 \leqslant x \leqslant \infty$$

$$\mu_1' \text{ (about the origin)} = \sqrt{2}\, \Gamma\left(\frac{n+1}{2}\right) \Big/ \Gamma\frac{n}{2} \text{ and } \mu'_2 = n$$

(M.Sc. Agra 1953)

Solution :

$$\text{Total frequency } f = \int_0^\infty y_0\, e^{-\frac{x^2}{2}} x^{n-1} dx$$

$$= y_0 .\ 2^{\frac{n-2}{2}}\, \Gamma\frac{n}{2}$$

$\therefore$ taking $y_0 = \dfrac{1}{2^{\frac{n-2}{2}}\,\Gamma\frac{n}{2}}$ the function becomes probability density function.

then
$$= \frac{1}{2^{\frac{n-2}{2}}\Gamma\frac{n}{2}} \int_0^\infty e^{-\frac{x^2}{2}} x^{n+r-1} dx$$

$$= \frac{2^{\frac{r}{2}}\, \Gamma\left(\frac{n+r}{2}\right)}{\Gamma\frac{n}{2}}$$

$$\text{Putting } r=1\ \mu_1' = \sqrt{2}\, \frac{\Gamma\frac{n+1}{2}}{\Gamma\frac{n}{2}}$$

$$\text{Putting } r=2\ \mu'_2 = 2.\ \frac{\Gamma\frac{n+2}{2}}{\Gamma\frac{n}{2}}$$

$$= n.$$

Problem 129.—Obtain the values of the first four moments for the distribution given by $dF = \text{constant} \times e^{-x/\sigma} dx.$ $(0 \leqslant x \leqslant \infty).$

(I.A.S. Part I, 1949)

Solution :

Total frequency $F = \text{constant} \times \int_0^\infty e^{-x/\sigma} dx$

$= \text{constant} \times \sigma$

$\therefore$ If we take constant $= \frac{1}{\sigma}$ the function becomes

$dF = \frac{1}{\sigma} e^{-x/\sigma} dx$ which is probability density function.

Now $\mu'_r = \frac{1}{\sigma} \int_0^\infty x^r e^{-x/\sigma} dx$

$= \lfloor r . \sigma^r$

$\therefore$ $\mu'_1 = \sigma$, $\mu'_2 = 2\sigma^2$, $\mu'_3 = 6\sigma^3$ and $\mu'_4 = 24\sigma^4$

now $\mu_1 = 0$,

$\mu_2 = \mu'_2 - \mu'^2_1$

$= \sigma^2$

$\mu_3 = \mu'_3 - 3\mu_2 \mu'_1 - \mu'^3_1$

$= 6\sigma^3 - 3\sigma^3 - \sigma^3$

$= 2\sigma^3$

and $\mu_4 = \mu'_4 - 4\mu_3 \mu_1' - 6\mu_2 \mu'^2_1 - \mu'^4_1$

$= 24\sigma^4 - 8\sigma^4 - 6\sigma^4 - \sigma^4$

$= 9\sigma^4$

$\therefore$ $\mu'_1 = \sigma$; $\mu'_2 = 2\sigma^2$; $\mu'_3 = 6\sigma^3$; $\mu'_4 = 24\sigma_4$.

and $\mu_1 = 0$; $\mu_2 = \sigma^2$; $\mu_3 = 2\sigma^3$; $\mu_4 = 9\sigma^4$

Problem 130.—Obtain the values of the mean, standard deviation and interquartile range for the distribution given by

$dF = \text{constant} \times e^{-x/\sigma}$ $(0 \leqslant x \leqslant \infty)$

(*I.A.S., Part II. 1949*)

Solution :

As in the previous question

$\mu_1' = \sigma$

$\mu_2 = \sigma^2$

$\therefore$ Mean $= \sigma$

and standard deviation $= \sqrt{\mu_2} = \sigma$

Problem 131. In a continuous distribution, whose probability density function is given by $f(x) = 3x(2-x)/4$, the variable ranges from 0 to 2, show that the distribution is symmetrical, with mean $x=1$, and variance 1/5. Show that the second and third moments about $x=0$ are 6/5 and 8/5 respectively; and verify that

$$\mu_3=0.$$

(J. C.)

Solution :

$$\text{Mean}=\mu'_1=\int_0^2 x\cdot\frac{3x(2-x)}{4}\,dx$$
$$=\frac{3}{4}\int_0^2 (2x^2-x^3)\,dx$$
$$=1$$

Mode is the value of x for which

$$\frac{dy}{dx}=\frac{3}{4}\,[\,2-2x\,]=0$$

or $\quad x=1$

$$\frac{d^2y}{dx^2}=-\frac{3}{2}$$

$\therefore\quad$ Mode $=1$

Hence the function is symmetrical as Mean and Mode are equal.

$$\mu'_2=\int_0^2 \frac{3}{4}\,x^3(2-x)\,dx$$
$$=\frac{3}{4}\left[\frac{x^4}{2}-\frac{x^5}{5}\right]_0^2$$
$$=\frac{6}{5}$$

$$\mu'_3=\int_0^2 \frac{3}{4}\,x^4(2-x)dx$$
$$=\frac{3}{4}\left[\frac{2x^5}{5}-\frac{x^6}{6}\right]_0^2$$
$$=\frac{8}{5}$$

$\therefore$ Variance $=\mu_2=\mu'_2-\mu'^2_1$

$$=\frac{6}{5}-1$$
$$=\frac{1}{5}$$

Now

$$\mu_3=\mu_3'-3\mu_2\,\mu'_1-\mu'^3_1$$
$$=\frac{8}{5}-3\times\frac{1}{5}\times 1-1$$
$$=0$$

CHAPTER V

INTERPOLATION

The following methods are generally used for interpolating values of a given variable for different cases :—

(1) The Algebraic Method :—

Let corresponding to $X_1, X_2, \ldots X_n$ values of X the values of Y be $Y_1, Y_2, \ldots Y_n$. Next, assume that $x_1, x_2, \ldots x_n$ are the deviations of the values of X from any assumed value A (*i.e.*, $x_r = X_r - A$ for $r=1, 2, \ldots n$) and $y_1, y_2, \ldots y_n$ are deviations of the values of Y from assumed value B (*i.e.*, $y_r = Y_r - B$). Then, to the above values of x and y we fit a curve—

$$y = a + bx + cx^2 + \ldots\ldots\ldots + kx^{n-1} \qquad \ldots(1)$$

where $a, b, \ldots\ldots k$ are constants.

By substituting from above the values of x, y we get the equations connecting the constants of the form :

$$\begin{aligned} y_1 &= a + bx_1 + cx^2_1 + \ldots\ldots + kx_1^{n-1} \\ y_2 &= a + bx_2 + cx_2^2 + \ldots\ldots + kx_n^{n-1} \\ &\ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ &\ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ yn &= a + bx_n + cx^2_n + \ldots + kx_n^{n-1} \end{aligned}$$

Solving these equations, by the process of elimination, we obtain the constants $a, b, c \ldots k$.

Thus, we get the curve—

$$Y - B = a + b(X-A) + c\ (X-A)^2 + \ldots\ldots + k(X-A)^{n-1} \quad \ldots(2)$$

passing through the points (X_1, Y_1) ; $(X_2, Y_2) \cdots (Xn, Yn)$. Now to find any value of Y corresponding to given value of X we substitute for X in (2) and get the intepolated value of Y.

(2) Lagrange's Method :—

Suppose corresponding to values $X_1, X_2, \ldots Xn$ of X the values of Y be $Y_1, Y_2, \ldots Yn$ and it is required to find the value of Y corresponding to value x of X. Then the required value of Y is given by the following formula.

$$Yx = \frac{(x-X_2)(x-X_3)\ldots(x-X_n)}{(X_1-X_2)(X_1-X_3)\ldots(X_1-X_n)} Y_1 + \frac{(x-X_1)(x-X_3)\ldots(x-X_n)}{(X_2-X_1)(X_2-X_3)\ldots(X_2-X_n)}$$

$$Y_2 + \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots + \frac{(x-X_1)(x-X_2)\ldots(x-X_{n-1})}{(X_n-X_1)(X_n-X_2)\ldots(X_n-X_{n-1})} Y_n.$$

(3) Newton's Binomial Method :—

Let X be the independent variable and Y be the dependent variable and the values be given as in the following table :

X	Y
a	Y_0
$a+h$	Y_1
$a+2h$	Y_2
$a+3h$	Y_3
⋮	⋮
$a+nh$	Y_n

(*i*) When out of these $Y_0, Y_1, Y_2 \ldots Y_n$ values of Y any one value of Y is missing we obtain it by the formula :

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3} Y_{n-3} + \ldots = 0$$

The values of $Y_0, Y_1 \ldots$ etc. are substituted in this formula and the missing value of Y is obtained.

(*ii*) When out of these $Y_0, Y_1, Y_2, \ldots Y_n$ values of Y any two values of Y are missing we obtain these by solving the following two equations :

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3} Y_{n-3} + \ldots = 0 \qquad \ldots(1)$$

and $$Y_{n-1} - (n-1)Y_{n-2} + \frac{(n-1)(n-2)}{1\times 2} Y_{n-3} - \frac{(n-1)(n-2)(n-3)}{1\times 2\times 3} Y_{n-4} + \ldots = 0 \ldots(2)$$

By substituting the given values of $Y_0, Y_1, \ldots$ etc. of Y in (1) and (2) we get two linear equations which are solved to get the missing values.

(4) Newton's Method of Advancing Differences :—

Let X be the independent variable and Y be the dependent variable and the values be given as in the following table.

X	Y
a	Y_0
$a+h$	Y_1
$a+2h$	Y_2
$a+3h$	Y_3
⋮	⋮
$a+nh$	Y_n

It is required to find the value of Y corresponding to any value x of X where x is not equal to a, $a+h$, $a+2h$ etc. for then the method given under (3) will be followed.

First find a parameter $t\left(=\frac{x-a}{h}\right)$ by dividing the difference of the first given value of X and the required value by the difference between any two consecutive values of X.

Next form a difference table as given below.

X	Y	Differences				
		1st	2nd	3rd	4th	5th
a	Y_0	$\triangle Y_0$	$\triangle^2 Y_0$	$\triangle^3 Y_0$	$\triangle^4 Y_0$	$\triangle^5 Y_0$
$a+h$	Y_1	$\triangle Y_1$	$\triangle^2 Y_1$	$\triangle^3 Y_1$	$\triangle^4 Y_1$	
$a+2h$	Y_2	$\triangle Y_2$	$\triangle^2 Y_2$	$\triangle^3 Y_2$		
$a+3h$	Y_3	$\triangle Y_3$	$\triangle^2 Y_3$			
$a+4h$	Y_4	$\triangle Y_4$				
$a+5h$	Y_5					

where $\triangle Y_0 = Y_1 - Y_0$, $\triangle Y_1 = Y_2 - Y_1$ etc.

and $\triangle^2 Y_0 = \triangle Y_1 - \triangle Y_0$, etc. $\triangle^2 Y_1 = \triangle Y_2 - \triangle Y_1$ etc., etc.

The powers of $\triangle$ indicate the order of the difference in the above table.

Now the required value of Y is obtained by the help of the following formula :

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1 \times 2} \cdot (\triangle^2 Y_0) + \frac{t(t-1)(t-2)}{1 \times 2 \times 3} \cdot (\triangle^3 Y_0) + \ldots$$

where $t = \frac{x-a}{h}$ and Y_0, $\triangle Y_0$, $\triangle^2 Y_0$ etc. are the uppermost value of the second, third ...columns in the difference table.

(5) **Newton's method for Divided Differences** :

This method is used in case of unequal intervals. Let Y_a, Y_b, Y_c, Y_d, etc. be the values of Y corresponding to the values a, b, c, d,... etc. of X ; then we have the divided differences table as given below.

X	Y	Divided differences			
		1st	2nd	3rd	...
a	Y_a	$\Delta_1 = \frac{Y_b - Y_a}{b-a}$	$\Delta^2_1 = \frac{\Delta_2 - \Delta_1}{c-a}$	$\Delta^3_1 = \frac{\Delta^2_2 - \Delta^2_1}{d-a}$	
b	Y_b	$\Delta_2 = \frac{Y_c - Y_b}{c-b}$	$\Delta^2_2 = \frac{\Delta_3 - \Delta_2}{d-b}$		
c	Y_c	$\Delta_3 = \frac{Y_d - Y_c}{d-c}$			
d	Y_d				
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...

The value of Y corresponding to the value x of X is given by the Newton's formula for divided difference which is

$$Y_x = Y_a + (x-a) \Delta_1 + (x-a)(x-b) \Delta^2_1 + (x-a)(x-b)(x-c) \Delta_1^3 + \ldots$$

where Δ_1, Δ_1^2, Δ_1^3,...etc. are the divided differences.

Note : 1. In case of equal intervals Newtons Binomial method and Newton Advancing Differences method are used.

2. In case of unequal intervals Algabric method, Lagranges method and Newtons Divided Differences methods are used.

Problem 132.—The population of a town increases according to the compound interest law. In 1890 and 1940 it was 19,500 and 34,670 respectively. Use it to estimate the population in 1929 and 1945.

(*Alld. M.Com.*, *1945*)

Solution :

X	1890	1940
Y	19,500	34,670

Let us fit the curve

$$Y = a + bX$$

to the above data. Substituting for X and Y we get

$19500 = a + 1890\, b$ I

$34670 = a + 1940\, b$ II

Subtracting I from II we have

$15170 = 50\, b$

$\therefore \quad b = 303{\cdot}4$

$\therefore \quad a = -553926$

$\therefore \quad Y = -553926 + 303{\cdot}4\, b$

Therefore for X=1926 the value of Y is given by

$Y = -553926 + 303{\cdot}4 \times 1926$

or $\quad Y = 30422{\cdot}0$

Again for X =1945 the value of Y is given by

$Y = -553926 + 1945 \times 303{\cdot}4$

or $\quad Y = 36187$

Therefore estimated population in 1926 and 1946 was 30422·4 and 36187

Problem 133.—Find by algebraic method of interpolation, using all the information given, the likely number for 1950 from the following table of index numbers of production of a certain article in India.

Year	1948	1949	1950	1951	1952
Index Number	100	107	...	157	212

(*P.C.S. 1953*)

Solution :

Let X denote the year and Y the index number

then				
X	1948	1949	1951	1952
Y	100	107	157	212

put $x = X - 1951$
$y = Y - 157$

∴				
x	-3	-2	0	1
y	-57	-50	0	55

Now let us fit to the above data a curve

$$y = a + bx + cx^2 + dx^3 \qquad \ldots(1)$$

Substituting for x and y we get

$$-57 = a - 3b + 9c - 27d \qquad \ldots(2)$$
$$-50 = a - 2b + 4c - 8d \qquad \ldots(3)$$
$$0 = a + 0 + 0 + 0 \qquad \ldots(4)$$
$$55 = a + b + c + d \qquad \ldots(5)$$

From (4) we get $a = 0$

$$\therefore \quad -19 = -b + 3c - 9d \qquad \ldots(6)$$
$$-25 = -b + 2c - 4d \qquad \ldots(7)$$
and $$55 = b + c + d \qquad \ldots(8)$$

Adding (6) and (8) $\quad 36 = 4c - 8d \qquad \ldots(9)$

Adding (7) and (8) $\quad 30 = 3c - 3d \qquad \ldots(10)$

From (9) and (10) $\quad c = 11,\ d = 1$

$\therefore \quad b = 43$

Hence the curve is

$$y = 43 + 11x^2 + x^3$$

or $$Y - 157 = 43(X - 1951) + 11(X - 1951)^2 + (X - 1951)^3 \qquad \ldots II$$

Now to get value of Y corresponding to X=1950 we substitute for X is II

$$\therefore \quad Y - 157 = 43(1950 - 1951) + 11(1950 - 1951)^2 + (1950 - 1951)^3$$

or $$Y = 157 - 43 + 11 - 1$$
$$= 124.$$

∴ The likely number for 1950 is 124.

Problem 134.—The following values are given in a table :

X	1	2	3	4	5
Y	216000	226981	...	250047	262144

Using any suitable algebraic method, find the value of Y for X=3.

(*I.A.S., 1953*)

Solution :

Let $x=X-4$
$y=Y-250047$

∴				
x	-3	-2	0	1
y	-34047	-23066	0	12097

Now let us fit to the above data a curve

$$y=a+bx+cx^2+dx^3 \quad \ldots(1)$$

Since (0, 0) is a point on the curve ∴ $a=0$

Now substituting values of x and y in (1) we get

$$-34047=-3b+9c-27d \quad \ldots(2)$$
$$-23066=-2b+4c-8d \quad \ldots(3)$$
$$12097=b+c+d \quad \ldots(4)$$

or from (2) $-11349=-b+3c-9d$ ...(5)

from (3) $-11533=-b+2c-4d$ ...(6)

from (4) & (6) adding

$$564=3c-3d$$

or $188=c-d$ ...(7)

from (4) & (5) adding

$$748=4c-8d$$

or $187=c-2d$ ...(8)

From (7) & (8) $d=1,\ c=189$

∴ $b=11907$

∴ the curve is $y=11907x+189x^2+x^3$

or $Y-250047=11907(X-4)+189(X-4)^2+(X-4)^3$.

The value of Y for X=3 is given by,

$$Y-250047=-11907+189-1$$

or $Y=250236-11908$
$=238328$

∴ Value of Y for X=3 is 238328.

Problem 135.—The observed values of a function are respectively 168, 120, 72 and 63 at the four positions 3, 7, 9 and 10 of the independent variable. What is the best estimate you can give for the value of the function at the position 6 of the independent variable.

(*I.A.S., 1951 ; M.Sc. Agra, 1957*)

Solution :

X	3	7	9	10
Y	168	120	72	63

Let $x = X - 7$
$y = Y - 120$

$\therefore$ x	-4	0	2	3
y	48	0	-48	-57

Now let us fit to the above data a curve

$$y = a + bx + cx^2 + dx^3$$

Since (0, 0) is a point on the curve $\therefore$ $a = 0$

Substituting for x and y we get

$$48 = -4b + 16c - 64d \quad \ldots(1)$$
$$-48 = 2b + 4c + 8d \quad \ldots(2)$$
$$-57 = 3b + 9c + 27d \quad \ldots(3)$$

or
$$12 = -b + 4c - 16d \quad \ldots(4)$$
$$-24 = b + 2c + 4d \quad \ldots(5)$$
$$-19 = b + 3c + 9d \quad \ldots(6)$$

from (4) & (5)
$$-12 = 6c - 12d$$
or
$$-2 = c - 2d \quad \ldots(7)$$
From (4) and (5) $-7 = 7c - 7d$

or
$$-1 = c - d \quad \ldots(8)$$
From (7) and (8) $c = 0,\ d = 1$

$\therefore$ $b = -28$

$\therefore$ the curve is $y = -28x + x^3$

or $Y - 120 = -28(X-7) + (X-7)^3$

Value of Y for X=6 is given by

$$Y - 120 = -28(6-7) + (6-7)^3$$
or
$$Y = 120 + 28 - 1$$
$$= 147$$

This question can also be done by Lagrange's method for unequal intervals.

Problem 136.—If lx represents the number living at age X in a life table, find, as accurately as the data will permit, lx for values of X=35, 42 and 47 ; given

$$l_{20} = 512,\ l_{30} = 439,$$
$$l_{40} = 346,\ l_{50} = 243\ ;$$

(*I.A.S., 1948*)

Solution :

Let Y denote l_x

∴	X	20	30	40	50
	Y	512	439	346	243

Let
$$x = X - 40$$
$$y = Y - 346$$

∴	x	-20	-10	0	10
	y	166	93	0	-103

Now let us fit to the above data a curve
$$y = a + bx + cx^2 + dx^3$$

Since curve passes through (0,0) ∴ $a = 0$

Substituting for x and y we get

$$166 = -20\,b + 400\,c - 8000\,d \quad \ldots (1)$$
$$93 = -10\,b + 100\,c - 1000\,d \quad \ldots (2)$$
$$-103 = 10\,b + 100\,c + 1000\,d \quad \ldots (3)$$

Add (2) and (3)
$$-10 = 200\,c$$
∴ $$c = -\cdot 05$$

From (1) $b + 400\,d = -9\cdot 3$

From (2) $b + 100\,d = -9\cdot 8$

∴ $d = \cdot 0016$, $b = -9\cdot 9666$

∴ Curve is
$$y = -9\cdot 966\,x - \cdot 05x^2 + \cdot 0016\,x^3$$
or
$$Y - 346 = -9\cdot 966(X-40) - \cdot 05(X-40)^2 + \cdot 0016(X-40)^3$$

∴ for $X = 35$, $Y = 346 + 49\cdot 83 - 1\cdot 25 - \cdot 2$
$= 394\cdot 38$

for $X = 42$, $Y = 346 - 19\cdot 932 - \cdot 2 + \cdot 0128$
$= 325\cdot 8808$

for $X = 47$, $Y = 346 - 69\cdot 762 - 2\cdot 45 + \cdot 5488$
$= 274\cdot 3368$

∴ $l_{35} = 394\cdot 38$, $l_{42} = 325\cdot 8808$
and $l_{47} = 274\cdot 3368$

Problem 137.—From the following life table, calculate the number living at ages 25, 35 and 47.

Age (in years)	20	30	40	50
Number living	51	44	35	24

(*M.A. Alld.*, 1952)

Solution :

X	20	30	40	50
Y	51	44	35	24

put $x = X - 40$

$y = Y - 35$

∴ x	-20	-10	0	10
y	16	9	0	-11

Now let us fit to the above data a curve

$$y = a + bx + cx^2 + dx^3$$

Since curve passes through (0, 0), $\therefore\ a=0$

Substituting for x and y we get

$$16 = -20\,b + 400\,c - 8000d \quad \dots (1)$$
$$9 = -10\,b + 100\,c - 1000\,d \quad \dots (2)$$
$$-11 = 10\,b + 100\,c + 1000\,d \quad \dots (3)$$

Add (2) & (3) $\therefore\ 200\,c = -2$

$\therefore\ c = -\cdot 01$

$\therefore$ From (1) $16 = -20\,b - 4 - 8000\,d$

or $-1 = b + 400\,d$

From (2) $-1 = b + 100\,d$

$\therefore\ b = -1,\ d = 0$

$\therefore$ Curve is

$$y = -x - \cdot 01\,x^2$$

or $Y - 35 = -(X-40) - \cdot 01(X-40)^2$

$\therefore$ For $X = 25$

$Y_{25} - 35 = -(25-40) - \cdot 01\ (25-40)^2$

or $Y_{25} = 35 + 15 - 2\cdot 25$

or $Y_{25} = 47\cdot 75$ ... I

For $X = 35$

$Y_{35} - 35 = -(35-40) - \cdot 01\ (35-40)^2$

or $Y_{35} = 35 + 5 - \cdot 25$

or $Y_{35} = 39\cdot 75$ ... II

or $X = 47$

$Y_{47} - 35 = -(47-40) - \cdot 01\ (47-40)^2$

or $Y_{47} = 35 - 7 - \cdot 49$

or $Y_{47} = 35 - 7 - \cdot 49$

or $Y_{47} = 27\cdot 51$ ... III

Hence the number of living at age 25, 35 and 47 are respectively 47·75, 39·75 and 27·51.

Problem 138.—Determine by Lagrange's formula the percentage number of criminals under 35 years.

Age	% number of criminals
under 25 years	52
,, 30 ,,	67·3
,, 40 ,,	84·1
,, 50 ,,	94·4

(*M.A. Agra, 1934*)

Solution :

X	25	30	40	50
	a	*b*	*c*	*d*
Y	52	67·3	84·1	94·4
	Y_1	Y_2	Y_3	Y_4

Lagrange's formula is

$$Y_x=\frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)}Y_1+\frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)}Y_2$$
$$+\frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)}Y_3+\frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}Y_4$$

$$\therefore\ Y_{35}=\frac{(35-30)(35-40)(35-50)}{(25-30)(25-40)(25-50)}\times 52$$
$$+\frac{(35-25)(35-40)(35-50)}{(30-25)(30-40)(30-50)}\times 67\cdot 3$$
$$+\frac{(35-25)(35-30)(35-50)}{(40-25)(40-30)(40-50)}\times 84\cdot 1$$
$$+\frac{(35-25)(35-30)(35-40)}{(50-25)(50-30)(50-40)}\times 94\cdot 4$$
$$=\frac{(5)(-5)(-15)}{(-5)(-15)(-25)}\times 52+\frac{(10)(-5)(-15)}{(5)(-10)(-20)}\times 67\cdot 3$$
$$+\frac{(10)(5)(-15)}{(15)(10)(-10)}\times 84\cdot 1+\frac{(10)(5)(-5)}{(+25)(+20)(10)}\times 94\cdot 4.$$
$$=-\frac{52}{5}+\frac{3}{4}\times 67\cdot 3+\frac{84\cdot 1}{2}-\frac{94\cdot 4}{20}$$
$$=-10\cdot 4+50\cdot 475+42\cdot 05-4\cdot 72$$
$$=77\cdot 405$$

$\therefore$ Criminals under 35 years are 77·405.

Problem 139.—If l_x represents the number living at age x in a life table

X	10	20	30	40
l_x	512	439	346	243,

estimate the numbers living at age 22 and 25.

(*I.A.S., Part II, 1949*)

Solution :

X	10	20	30	40
	a	b	c	d
Y	512	439	346	243
	Y_1	Y_2	Y_3	Y_4

Lagrange's formula is

$$Y_x=\frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)}Y_1+\frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)}Y_2$$
$$+\frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)}Y_3+\frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}Y_4.$$

$$\therefore \quad Y_{22}=\frac{(22-20)(22-30)(22-40)}{(10-20)(10-30)(10-40)}\times 512$$
$$+\frac{(22-10)(22-30)(22-40)}{(20-10)(20-30)(20-40)}\times 439$$
$$+\frac{(22-10)(22-20)(22-40)}{(30-10)(30-20)(30-40)}\times 346$$
$$+\frac{(22-10)(22-20)(22-30)}{(40-10)(40-20)(40-30)}\times 243$$

or
$$Y_{22}=\frac{(2)(-8)(-18)}{(-10)(-20)(-30)}\times 512+\frac{(12)(-8)(-18)}{(10)(-10)(-20)}\times 439$$
$$+\frac{(12)(2)(-18)}{(20)(10)(-10)}\times 346+\frac{(12)(2)(-8)}{(30)(20)(10)}\times 243$$
$$=-\frac{2\times 8\times 18\times 512}{6000}+\frac{12\times 8\times 18\times 439}{2000}$$
$$+\frac{12\times 2\times 18\times 346}{2000}-\frac{12\times 2\times 8\times 243}{6000}$$
$$=-\frac{48\times 512}{1000}+\frac{48\times 18\times 439}{1000}+\frac{12\times 18\times 346}{1000}$$
$$-\frac{32\times 243}{1000}$$
$$=-24{\cdot}576+379{\cdot}296+74{\cdot}736-7{\cdot}776$$
$$=421{\cdot}65$$

$$Y_{25}=\frac{(25-20)(25-30)(25-40)}{(10-20)(10-30)(10-40)}\times 512$$
$$+\frac{(25-10)(25-30)(25-40)}{(20-10)(20-30)(20-40)}\times 439$$
$$+\frac{(25-10)(25-20)(25-40)}{(30-10)(30-20)(30-40)}\times 346$$
$$+\frac{(25-10)(25-20)(25-30)}{(40-10)(40-20)(40-30)}\times 243$$

$$=\frac{(5)(-5)(-15)}{(-10)(-20)(-30)}\times 512+\frac{(15)(-5)(-15)}{(10)(-10)(-20)}\times 439$$

$$+\frac{(15)(5)(-15)}{(+20)(+10)(-10)}\times 346$$

$$+\frac{(15)(5)(-5)}{(30)(20)(10)}\times 243$$

$$=-\frac{25\times 15\times 512}{6000}+\frac{225\times 5\times 439}{2000}+\frac{225\times 5\times 346}{2000}$$

$$-\frac{25\times 15\times 243}{6000}$$

$$=-\frac{25\times 15(512+243)}{6000}+\frac{225\times 5(439+346)}{2000}$$

$$=-\frac{25\times 5\times 755}{2000}+\frac{225\times 5\times 785}{2000}$$

$$=\frac{25\times 5(-755+9\times 785)}{2000}$$

$$=\frac{-755+7065}{16}$$

$$=\frac{6310}{16}$$

$$=394{\cdot}375$$

$\therefore \quad l_{22}=421{\cdot}65$

$l_{25}=394{\cdot}375.$

Problem 140.—The following table gives the number of income tax assessees in the U.P. :

Income not exceeding	No. of assessees
Rs. 2500	7166
Rs. 3000	10576
Rs. 5000	17200
Rs. 7500	20505
Rs. 10000	21975

Estimate the number of assessees with income not exceeding Rs. 4000.

(*M.A. Allahabad, 1944*)

Solution :

	a	*b*	*c*	*d*	*e*
X	2500	3000	5000	7500	10000
Y	7166	10576	17200	20505	21975
	Y_1	Y_2	Y_3	Y_4	Y_5

Lagranges formula is

$$Y_x=\frac{(x-b)(x-c)(x-d)(x-e)}{(a-b)(a-c)(a-d)(a-e)}\,Y_1+\frac{(x-a)(x-c)(x-d)(x-e)}{(b-a)(b-c)(b-d)(b-e)}Y^2$$

$$+\frac{(x-a)(x-b)(x-d)(x-e)}{(c-a)(c-b)(c-d)(c-e)}Y_3+\frac{(x-a)(x-b)(x-c)(x-e)}{(d-a)(d-b)(d-c)(d-e)}Y_4$$

$$+\frac{(x-a)(x-b)(x-c)(x-d)}{(e-a)(e-b)(e-c)(e-d)}Y_5$$

$\therefore \quad Y_{4000}$

$$=\frac{(4000-3000)(4000-5000)(4000-7500)(4000-10000)}{(2500-3000)(2500-5000)(2500-7500)(2500-10000)}\times 7166$$

$$+\frac{(4000-2500)(4000-5000)(4000-7500)(4000-10000)}{(3000-2500)(3000-5000)(3000-7500)(3000-10000)}\times 10576$$

$$+\frac{(4000-2500)(4000-3000)(4000-7500)(4000-10000)}{(5000-2500)(5000-3000)(5000-7500)(5000-10000)}\times 17200$$

$$+\frac{(4000-2500)(4000-3000)(4000-5000)(4000-10000)}{(7500-2500)(7500-3000)(7500-5000)(7500-10000)}\times 20505$$

$$+\frac{(4000-2500)(4000-3000)(4000-5000)(4000-7500)}{(10000-2500)(10000-3000)(10000-5000)(10000-7500)}\times 21975$$

Cancelling the two zeros from each bracket we get

$$Y_{4000}=\frac{(40-30)(40-50)(40-75)(40-100)}{(25-30)(25-50)(25-75)(25-100)}\times 7166$$

$$+\frac{(40-25)(40-50)(40-75)(40-100)}{(30-25)(30-50)(30-75)(30-100)}\times 10576$$

$$+\frac{(40-25)(40-30)(40-75)(40-100)}{(50-25)(50-30)(50-75)(50-100)}\times 17200$$

$$+\frac{(40-25)(40-30)(40-50)(40-100)}{(75-25)(75-30)(75-50)(75-100)}\times 20505$$

$$+\frac{(40-25)(40-30)(40-50)(40-75)}{(100-25)(100-30)(100-50)(100-75)}\times 21975$$

or Y_{4000}

$$=\frac{(+10)(-10)(-35)(-60)}{(-5)(-25)(-50)(-75)}\times 7166+\frac{(15)(-10)(-35)(-60)}{(+5)(-20)(-45)(-70)}\times 10576$$

$$+\frac{(15)(10)(-35)(-60)}{(25)(20)(-25)(-50)}\times 17200+\frac{(15)(10)(-10)(-60)}{(50)(45)(25)(-25)}\times 20505$$

$$+\frac{(15)(10)(-10)(-35)}{(75)(70)(50)(25)}\times 21975$$

or $\quad Y_{4000}=-\dfrac{56\times 7166}{125}+10576+\dfrac{63\times 17200}{125}-\dfrac{8\times 20505}{125}$

$$+\frac{21975}{125}$$

$$=-3210{\cdot}37+10576+8668{\cdot}8-1312{\cdot}32+175{\cdot}8$$

$$=-4522{\cdot}69+19420{\cdot}6$$

$$=14897{\cdot}91$$

Thus the number of persons with the income not exceeding Rs. 4000 is 14897·91

or 14898 appr.

Problem 141.—Determine by Lagrange's Method the percentage number of patients over 40 years.

Age over years	% number of patients
30	148
35	96
45	68
55	34

Solution :

X	30	35	45	55
Y	148	96	68	34

$$Y_x = \frac{(x-35)(x-45)(x-55)}{(30-35)(30-45)(30-55)} \times 148 + \frac{(x-30)(x-45)(x-55)}{(35-30)(35-45)(35-55)} \times 96$$

$$+ \frac{(x-30)(x-35)(x-55)}{(45-30)(45-35)(45-55)} \times 68 + \frac{(x-30)(x-35)(x-45)}{(55-30)(55-35)(55-45)} \times 34$$

$$\therefore \quad Y_{40} = \frac{(40-35)(40-45)(40-55)}{(-5)(-15)(-25)} \times 148 + \frac{(40-30)(40-45)(40-55)}{(5)(-10)(-20)} \times 96$$

$$+ \frac{(40-30)(40-35)(40-45)}{(15)(10)(-10)} \times 68 + \frac{(40-30)(40-35)(40-45)}{(25)(20)(10)} \times 34$$

$$= -\frac{(5 \times 5 \times 15)}{5 \times 15 \times 25} \times 148 + \frac{10 \times 5 \times 15}{5 \times 10 \times 20} \times 96 + \frac{10 \times 5 \times 5}{15 \times 10 \times 10} \times 68 - \frac{10 \times 5 \times 5}{25 \times 20 \times 10} \times 34$$

$$= -\frac{148}{5} + \frac{3 \times 96}{4} + \frac{68}{6} - \frac{34}{25}$$

$$= -29{\cdot}6 + 72 + 11{\cdot}33 - 1{\cdot}36$$

$$= 83{\cdot}33 - 30{\cdot}96$$

$$= 52{\cdot}37$$

$$\therefore \quad Y_{40} = 52{\cdot}37.$$

Problem 142.—In the following table h is the height above sea level and p the barometric pressure. Calculate p when $h=5280$.

$h=0$	4763	6252	10593
$p=27$	25	23	20

(*M.A. Aligarh, 1942*)

Solution :

X	0	4763	6252	10593
	a	b	c	d
Y	27	25	23	20
	Y_1	Y_2	Y_3	Y_4

Lagrange's formula is

$$Y_x = \frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)}Y_1 + \frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)}Y_2$$

$$+\frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)}Y_3 + \frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}Y_4$$

$$\therefore\quad Y_{5280} = \frac{(5280-4763)(5280-6252)(5280-10593)}{(0-4763)(0-6252)(0-10593)}\times 27$$

$$+\frac{(5280-0)(5280-6252)(5280-10593)}{(4763-0)(4763-6252)(4763-10593)}\times 25$$

$$+\frac{(5280-0)(5280-4763)(5280-10593)}{(6252-0)(6252-4763)(6252-10593)}\times 23$$

$$+\frac{(5280-0)(5280-4763)(5280-6252)}{(10593-0)(10593-4763)(10593-6252)}\times 20$$

$$\text{or}\quad Y_{5280} = -\frac{517\times 972\times 5313\times 27}{4763\times 6252\times 10593}$$

$$+\frac{5280\times 972\times 5313\times 25}{4763\times 1489\times 5830}$$

$$+\frac{5280\times 517\times 5313\times 23}{6252\times 1489\times 4341}$$

$$-\frac{5280\times 517\times 972\times 20}{10593\times 5830\times 4341}$$

$$Y_{5280} = -\frac{47\times 243\times 483}{433\times 521\times 107}+\frac{1200\times 972\times 483}{433\times 1489\times 53}$$

$$+\frac{440\times 517\times 5313\times 23}{521\times 1489\times 4341}-\frac{320\times 47\times 324}{107\times 53\times 4341}$$

$\therefore$ By the method of logarithms

$$Y_{5280}^{*} = -\cdot 2393+16\cdot 48+7\cdot 883-\cdot 189$$

$$=24\cdot 363-\cdot 4283$$

$$=23\cdot 9347$$

$\therefore$ Barometric pressure corresponding to $h=5280$ is $23\cdot 9347$

Problem 143.—Extrapolate the population of a town for 1946 from the following data about its population during the previous four censuses.

Census year	*Population in thousands*
1911	473
1921	468
1931	454
1941	484

(*M.Com. Raj*, *1952*)

Solution :

X	1911	1921	1931	1941
	a	b	c	d
Y	473	468	454	484
	Y_1	Y_2	Y_3	Y_4

Lagranges formula is

$$Y_x = \frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)}Y_1 + \frac{(x-a)(x-c)(x-d)}{(b-a)(b-c)(b-d)}Y_2 + \frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)}Y_3 + \frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}Y_4.$$

$$\therefore \quad Y_x = \frac{(x-1921)(x-1931)(x-1941)}{(1911-1921)(1911-1931)(1911-1941)} \times 473 + \frac{(x-1911)(x-1931)(x-1941)}{(1921-1911)(1921-1931)(1921-1941)} \times 468 + \frac{(x-1911)(x-1921)(x-1941)}{(1931-1911)(1931-1921)(1931-1941)} \times 454 + \frac{(x-1911)(x-1921)(x-1931)}{(1941-1911)(1941-1921)(1941-1931)} \times 484$$

$$\therefore \quad Y_{1946} = -\frac{(1946-1921)(1946-1931)(1946-1941)}{10\times20\times30} \times 473 + \frac{(1946-1911)(1946-1931)(1946-1941)}{10\times10\times20} \times 468 - \frac{(1946-1911)(1946-1921)(1946-1941)}{20\times10\times10} \times 454 + \frac{(1946-1911)(1946-1921)(1946-1931)}{30\times20\times10} \times 484$$

$$\text{or} \quad Y_{1946} = -\frac{25\times15\times5}{10\times20\times30}\times473 + \frac{35\times15\times5}{10\times10\times20}\times468 - \frac{35\times25\times5}{20\times10\times10}\times454 + \frac{35\times25\times15}{30\times20\times10}\times484$$

$$\text{or} \quad Y_{1946} = -147{\cdot}8125 + 614{\cdot}25 - 993{\cdot}125 + 1058{\cdot}75$$

$$\therefore \quad Y_{1946} = 532{\cdot}0625$$

Hence population of 1946 is extrapolated as 532 thousands approximately.

Problem 144.—From the table given below, interpolate the number under 30 years. Use Lagrange's Formula :

Ages (years)	*Population occupied per* 10,000
10—15	193·5
15—20	880
20—25	933
25—35	1636
35—45	1201
45—55	830

(*Alld. M.A.*, 1951)

Solution :

Ages under (years)	X	15	20	25	35	45	55
		a	*b*	*c*	*d*	*e*	*f*
Population occupied		Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
per 10,000	Y	193·5	1073·5	2006·5	3642·5	4863·5	5693·5

Lagranges Formula is

$$Y_x=\frac{(x-b)(x-c)(x-d)(x-e)(x-f)}{(a-b)(a-c)(a-d)(a-e)(a-f)}Y_1$$
$$+\frac{(x-a)(x-c)(x-d)(x-e)(x-f)}{(b-a)(b-c)(b-d)(b-e)(b-f)}Y_2$$
$$+\frac{(x-a)(x-b)(x-d)(x-e)(x-f)}{(c-a)(c-b)(c-d)(c-e)(c-f)}Y_3$$
$$+\frac{(x-a)(x-b)(x-c)(x-e)(x-f)}{(d-a)(d-b)(d-c)(d-e)(d-f)}Y_4$$
$$+\frac{(x-a)(x-b)(x-c)(x-d)(x-f)}{(e-a)(e-b)(e-c)(e-d)(e-f)}Y_5$$
$$+\frac{(x-a)(x-b)(x-c)(x-d)(x-e)}{(f-a)(f-b)(f-c)(f-d)(f-e)}Y_6$$

$$\therefore\quad Y_{30}=\frac{(30-20)(30-25)(30-35)(30-45)(30-55)}{(15-20)(15-25)(15-35)(15-45)(15-55)}\times 193{\cdot}5$$
$$+\frac{(30-15)(30-25)(30-35)(30-45)(30-50)}{(20-15)(20-25)(20-35)(20-45)(20-55)}\times 1073{\cdot}5$$
$$+\frac{(30-15)(30-20)(30-35)(30-45)(30-55)}{(25-15)(25-20)(25-35)(25-45)(25-55)}\times 2006{\cdot}5$$
$$+\frac{(30-15)(30-20)(30-25)(30-45)(30-55)}{(35-15)(35-20)(35-25)(35-45)(35-55)}\times 3642{\cdot}5$$
$$+\frac{(30-15)(30-20)(30-25)(30-35)(30-55)}{(45-15)(45-20)(45-25)(45-35)(45-55)}\times 4863{\cdot}5$$

$$+\frac{(30-15)(30-20)(30-25)(30-35)(30-45)}{(55-15)(55-20)(55-25)(55-35)(55-45)}\times 5693\cdot5$$

$$\text{or}\quad Y_{30}=\frac{10\times5\times5\times15\times25}{5\times10\times20\times30\times40}\times193\cdot5-\frac{15\times5\times5\times15\times25}{5\times5\times15\times25\times35}\times1073\cdot5$$

$$+\frac{15\times10\times5\times15\times25}{10\times5\times10\times20\times30}\times2006\cdot5+\frac{15\times10\times5\times15\times25}{20\times15\times10\times10\times20}\times3642\cdot5$$

$$-\frac{15\times10\times5\times5\times25}{30\times25\times20\times10\times10}\times4863\cdot5+\frac{15\times10\times5\times5\times15}{40\times35\times30\times20\times10}\times5693\cdot5$$

$$\text{or}\quad Y_{30}=\frac{5\times193.5}{64}-\frac{3\times1073\cdot5}{7}+\frac{15\times2006.5}{16}+\frac{15\times3642\cdot5}{32}-\frac{4863\cdot5}{16}+\frac{3\times5693\cdot5}{448}$$

$$\text{or}\quad Y_{30}=\frac{967\cdot5}{64}-\frac{3220\cdot5}{7}+\frac{30097\cdot5}{16}+\frac{54637\cdot5}{32}-\frac{4863\cdot5}{16}+\frac{17080\cdot5}{448}$$

$$\text{or}\quad Y_{30}=15\cdot1-460\cdot07+1881\cdot1+1707\cdot4-303\cdot9+38\cdot1$$

$$\text{or}\quad Y_{30}=3641\cdot7-763\cdot97$$

$$\therefore\quad Y_{30}=2877\cdot73$$

Therefore the numbers under 30 years Ages are 2877·73 or 2878. appr.

Problem 145.—It is required to find the missing value in the following table. Establish any suitable formula for interpolation and find the missing value.

Serial No.	*Value*	*Serial No.*	*Value*
1	6·4577	6	1·7849
2	3·4531	7	1·6874
3	2·5604	8	1·6177
4	2·1521	9	1·5646
5	....	10	1·5232

(*I. C. S., 1940*)

Solution :

X	1	2	3	4	5	6	7	8	9	10
Y	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
	6·4577	3·4531	2·5604	2·1521	...	1·7849	1·6874	1·6177	1·5646	1·5232

Newton's formula for one missing value is

$$Y_n-nY_{n-1}+\frac{n(n-1)}{1\times2}Y_{n-2}-\frac{n(n-1)(n-2)}{1\times2\times3}Y_{n-3}+\ldots=0$$

$$\because\quad Y_9-9Y_8+\frac{9\times8}{1\times2}Y_7-\frac{9\times8\times7}{1\times2\times3}Y_6+\frac{9\times8\times7\times6}{1\times2\times3\times4}Y_5$$

$$-\frac{9\times8\times7\times6\times5}{1\times2\times3\times4\times5}Y_4+\frac{9\times8\times7\times6\times5\times4}{1\times2\times3\times4\times5\times6}Y_3$$

$$-\frac{9\times8\times7\times6\times5\times4\times3}{1\times2\times3\times4\times5\times6\times7}Y_2+\frac{9\times8\times7\times6\times5\times4\times3\times2}{1\times2\times3\times4\times5\times6\times7\times8}Y_1$$

$$-\frac{9\times8\times7\times6\times5\times4\times3\times2\times1}{1\times2\times3\times4\times5\times6\times7\times8\times9}Y_0=0$$

or $Y_9-9Y_8+36Y_7-84Y_6+126Y_5-126Y_4+84Y_3-36Y_2+9Y_1-Y_0=0$

or $(Y_9-Y_0)-9(Y_8-Y_1)+36(Y_7-Y_2)-84(Y_6-Y_3)+126(Y_5-Y_4)=0$

or $(1{\cdot}5232-6{\cdot}4577)-9(1{\cdot}5646-3{\cdot}4531)+36(1{\cdot}6177-2{\cdot}5604)-84(1{\cdot}6874-2{\cdot}1521)+126\ (1{\cdot}7849-Y_4)=0.$

or $-4{\cdot}9345+9\times1{\cdot}8885-36\times{\cdot}9427+84\times{\cdot}4647+126(1{\cdot}7849-Y_4)=0$

or $-126(1{\cdot}7849-Y_4)=-4{\cdot}9345+16{\cdot}9965-33{\cdot}9372+39{\cdot}0348$

or $126\ Y_4=126\times1{\cdot}7849-38{\cdot}8717+56{\cdot}0313$

or $126\ Y_4=242{\cdot}057$

$\therefore\quad Y_4=1{\cdot}921$

Therefore missing value is 1·921.

Problem 146.—Interpolate the missing figure in the following table with the help of a suitable formula :

1911	1912	1913	1914	1915	1916	1917
1331	1728	2197	...	3375	4096	4913

(*M.A. Delhi, 1953*)

Solution :

X	1911	1912	1913	1914	1915	1916	1917
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y	1331	1728	2197	...	3375	4096	4913

Newtons formula for one missing value is

$$Y_n-nY_{n-1}+\frac{n(n-1)}{1\times2}Y_{n-2}-\frac{n(n-1)(n-2)}{1\times2\times3}Y_{n-3}+\ldots=0$$

$$Y_6-6Y_5+\frac{6\times5}{1\times2}Y_4-\frac{6\times5\times4}{1\times2\times3}Y_3+\frac{6\times5\times4\times3}{1\times2\times3\times4}Y_2$$

$$-\frac{6\times5\times4\times3\times2}{1\times2\times3\times4\times5}Y_1+\frac{6\times5\times4\times3\times2\times1}{1\times2\times3\times4\times5\times6}Y_0=0$$

or $Y_6-6Y_5+15Y_4-20Y_3+15Y_2-6_1Y+Y_0=0$

or $(Y_6+Y_0)-6(Y_5+Y_1)+15(Y_4+Y_2)-20Y_3=0.$

or $(4913+1331)-6(4096+1728)+15(3375+2197)-20Y_3=0$

or $20Y_3=6244-6\times5824+15\times5572$

or $Y_3=2744$

$\therefore$ Missing value is 2744.

Problem 147.—The annual sales of a concern are given below :

Years	1915	1920	1925	1930	1935
Sales of cloth in Lacs of yards	125	163	204	238	282

Assuming the conditions of the market to be the same, estimate the sales for the year 1940.

(*M.A. Patna, 1941*)

Solution :

X	1915	1920	1925	1930	1935	1940
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5
Y	125	163	204	238	282	...

Newtons formula for one missing value is

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times2\times3} Y_{n-3} + \ldots = 0$$

$$\therefore \quad Y_5 - 5Y_4 + \frac{5\times4}{1\times2} Y_3 - \frac{5\times4\times3}{1\times2\times3} Y_2 + \frac{5\times4\times3\times2}{1\times2\times3\times4} Y_1 - \frac{5\times4\times3\times2\times1}{1\times2\times3\times4\times5} Y_0 = 0$$

or $Y_5 - 5Y_4 + 10Y_3 - 10Y_2 + 5Y_1 - Y_0 = 0$

or $Y_5 - 5(Y_4 + Y_1) + 10(Y_3 - Y_2) - Y_0 = 0$

or $Y_5 - 5(282 - 163) + 10(238 - 204) - 125 = 0$

or $Y_5 - 5\times119 + 10\times34 - 125 = 0$

or $Y_5 - 595 + 340 - 125 = 0$

$\therefore \quad Y_5 = 380$

$\because$ The missing value is 380.

Problem 148.—Estimate the annual sales of cloth for 1935 from the following records of a cloth dealer.

Year	1920	1925	1930	1935	1940
Sale of cloth in Lakhs of yards	250	285	328	...	444

(*M.Com., Allahabad, 1944*)

Solution :

Newtons formula for one missing value is :

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times2\times3} Y_{n-3} + \ldots = 0$$

X	1920	1925	1930	1935	1940
	Y_0	Y_1	Y_2	Y_3	Y_4
Y	250	285	328	...	444

$$\therefore Y_4 - 4Y_3 + \frac{4\times3}{1\times2} Y_2 - \frac{4\times3\times2}{1\times2\times3} Y_1 + \frac{4\times3\times2\times1}{1\times2\times3\times4} Y_0 = 0$$

or $Y_4-4Y_3+6Y_2-4Y_1+Y_0=0$

or $444-4Y_3+6\times328-4\times285+250=0$

$\therefore$ $4Y_3=1522$

$\therefore$ $Y_3=380{\cdot}5$

$\therefore$ The missing value is 380·5

Problem 149.—Give an estimate of the population of Bengal in 1911 and in 1941 from the following figures.

Year	1881	1891	1901	1911	1921	1931
Population in Lakhs	363	391	421	445	467	501

(B.Com. Agra, 1940)

Solution :

X	1881	1891	1901	1911	1921	1931
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5
Y	363	391	421	...	467	501

Newtons formula for one missing value is :

$$Y_n-nY_{n-1}+\frac{n(n-1)}{1\times2}Y_{n-2}-\frac{n(n-1)(n-2)}{1\times2\times3}Y_{n-3}+\ldots=0.$$

$\therefore$ $Y_5-5Y_4+10Y_3-10Y_2+5Y_1-Y_0=0$

or $(Y_5-Y_0)-5(Y_4-Y_1)+10Y_3-10Y_2=0$

or $(501-363)-5(467-391)+10Y_3-10\times421=0$

or $10Y_3=-138+5\times76+4210$

or $10Y_3=4452$

$\therefore$ $Y_3=445{\cdot}2$

$\therefore$ Population in 1911 is 445·2 lakhs.

Now to Extrapolate population for 1941 we have the table as below.

X	1881	1891	1901	1911	1921	1931	1941
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y	363	391	421	445·2	467	501	...

Applying Newtons formula for $n=6$ we have

$$Y_6-6Y_5+15Y_4-20Y_3+15Y_2-6Y_1+Y_0=0$$

or $Y_6-6(Y_5+Y_1)+15(Y_4+Y_2)-20Y_3+Y_0=0$

or $Y_6-6\times892+15\times888-20\times445{\cdot}2+363=0$

or $Y_6-5352+13320-8904+363=0$

or $Y_6-14256+13683=0.$

or $Y_6-573=0$

$\therefore$ $Y_6=573$

∴ Population for 1941 should be 573 lakhs.

Problem 150.—The following table gives the population of Indore at the time of the last six censuses.

Year	1881	1891	1901	1911	1921	1931
Population	75401	82984	86686	44947	93091	127327

Estimate the population for 1941.

(B.Com. Agra, 1944)

Solution :

In the given data of population there are sudden jumps during 1911 and 1931. Hence the value for 1941 cannot be extrapolated directly. We will first interpolate the value for 1911 and then for 1931 and with the help of these interpolated datas we will determine population for 1941.

(*i*) **To determine the value for 1911.**

X	1881	1891	1901	1911	1921
	Y_0	Y_1	Y_2	Y_3	Y_4
Y	75401	82984	86686	...	93091

Newton's formula for one missing value is :

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3} Y_{n-3} + \ldots = 0.$$

∴ $Y_4 - 4Y_3 + 6Y_2 - 4Y_1 + Y_0 = 0$

or $93091 - 4Y_3 + 6\times 86686 - 4\times 82984 + 75401 = 0$

or $Y_3 = 89168.$

(*ii*) **To determine the value for 1931.**

X	1881	1891	1901	1911	1921	1931
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5
Y	75401	82984	86686	89168	93091	...

Putting $n=5$ in Newtons formula we have

$$Y_5 - 5Y_4 + 10Y_3 - 10Y_2 + 5Y_1 - Y_0 = 0$$

or $Y_5 - 5\times 93091 + 10\times 89168 - 10\times 86686 + 5\times 82984 - 75401 = 0.$

or $Y_5 - 465455 + 891680 - 866860 + 414920 - 75401 = 0$

or $Y_5 - 101116 = 0$

∴ $Y_5 = 101116.$

(*iii*) **To determine the value for 1941.**

X	1881	1891	1901	1911	1921	1931	1941
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y	75401	82984	86686	89168	93091	101116	...

Putting $n=6$ in Newtons formula we have

$$Y_6 - 6\,Y_5 + 15\,Y_4 - 20\,Y_3 + 15\,Y_2 + 6Y_1 - Y_0 = 0$$

or $Y_6-6\times101126+15\times93091-20-89168+15\times86686$
$-6\times82984+75401=0.$

or $Y_6-606696+1396365-1783360+1300290-497904$
$+75401=0$

or $Y_6-115904=0$

or $Y_6=115904$

Hence the extrapolated population for 1941 is 115904.

Problem 151.—The age of mother and average number of children born per mother are given in the following table. Find by any method of interpolation the average number of children born per mother aged 30–34.

Age of Mother	*Number of Children*	*Age of Mother*	*Number of Children*
15—19	·7	30—34	...
20—24	2·1	35—39	5·7
25—29	3·5	40—44	5·8

(*P. C. S., 1943 ; M.Com., Alld., 1946*)

Solution :

X	Y		X	Y	
15—19	·7	Y_0	30—34	......	Y_3
20—24	2·1	Y_1	35—39	5·7	Y_4
25—29	3·5	Y_2	40—44	5·8	Y_5

Newton's method for one missing value is

$$Y_n-nY_{n-1}+\frac{n(n-1)}{1\times2}Y_{n-2}-\frac{n(n-1)(n-2)}{1\times2\times3}Y_{n-3}+\ldots\ldots\ldots=0$$

∴ putting $n=5$ we have

$$Y_5-5Y_4+10Y_3-10Y_2+5Y_1-Y_0=0$$

∴ $5\cdot8-5\times5\cdot7+10Y_3-10\times3\cdot5+5\times2\cdot1-\cdot7=0$

or $5\cdot8-28\cdot5+10Y_3-35+10\cdot5-\cdot7=0$

or $10Y_3-47\cdot9=0$

∴ $Y_3=4\cdot79$

∴ Expected number of children born per mother aged 30—34 is 4·79.

Problem 152.—Estimate the outflow of gold from India in 1937-38 from the data given below :—

Average value of net exports of gold coin and Bullion.

Year	Rs.
1933—34	895632418
1934—35	525374607
1935—36	373558955
1936—37	278461129
1938—39	232602068

(*M. Com. Alld., 1948*)

Solution :

X	Y	
1933—34	895632418	Y_0
1934—35	525374607	Y_1
1935—36	373558955	Y_2
1936—37	278461129	Y_3
1937—38	....	Y_4
1938—39	232602068	Y_5

Newton's formula for one missing value is :

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2}Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3}Y_{n-3} + \ldots\ldots = 0$$

For $n=5$

$$Y_5 - 5Y_4 + 10Y_3 - 10Y_2 + 5Y_1 - Y_0 = 0$$

or $232602068 - 5Y_4 + 10\times 278461129 - 10\times 373558955 + 5\times 525374607 - 895632418 = 0$

or $-5Y_4 + 5644086393 - 4631221968 = 0$

or $-5Y_4 + 1012864425 = 0$

$\therefore \quad Y_4 = 202572885.$

Hence the estimated value of outflow of gold in 1937-38 is 202572885.

Note : The result, at the first look, seems to be wrong. But, in reality, it is obvious due to the fact that the values of upper items are considerably higher than lower values ; that results in this type of somewhat unexpected result.

Problem 153.—Estimate the value of imports in the year 1913 from the data given below.

Year	Value of Imports in Rupees
1910	39202000
1911	26510000
1912	26163000
1913	....
1914	33755000
1915	32987000
1916	27431000

(*M.A. Alld., 1950*)

Solution :

X	Y (Rupees in thousands)	
1910	39202	Y_0
1911	26510	Y_1
1912	26163	Y_2
1913	....	Y_3
1914	33755	Y_4
1915	32987	Y_5
1916	27431	Y_6

Newton's formula for one missing value is :

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2}Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3}Y_{n-3} + \ldots\ldots = 0$$

For $n=6$.

$$Y_6 - 6Y_5 + 15Y_4 - 20Y_3 + 15Y_2 - 6Y_1 + Y_0 = 0$$

$$27431 - 6\times 32987 + 15\times 33755 - 20Y_3 + 15\times 26163 - 6\times 26510 + 39202 = 0$$

or $\quad 965403 - 356982 - 20Y_3 = 0$

$\therefore \quad 20Y_3 = 608421$

$\quad Y_3 = 30421{\cdot}05$

$\therefore$ Value of Import in Rupees is 30421050 in the year 1913.

Problem 154.—Below are given weighted index numbers of cost of living of labourers in an industrial centre in India. Interpolate to find out the missing index number for 1933 to the nearest integer, using all the figures :

Year	1930	1931	1932	1934	1935
Index No.	173	149	145	131	141

(*M.A. Agra, 1938*)

Solution :

X	1930	1931	1932	1933	1934	1935
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5
Y	173	149	145	...	131	141

Applying Newton's formula for $n=5$

$$Y_5 - 5Y_4 + 10Y_3 - 10Y_2 + 5Y_1 - Y_0 = 0.$$

$$141 - 5\times 131 + 10Y_3 - 10\times 145 + 5\times 149 - 173 = 0$$

or $\quad 10Y_3 + 886 - 2278 = 0$

or $\quad 10Y_3 = 1392$

$\therefore \quad Y_3 = 139{\cdot}2$

Thus the index number for cost of living for 1933 is 139·2

Problem 155.—Obtain the estimate of the missing figures below :

Value of X	2·0	2·1	2·2	2·3	2·4	2·5	2·6
Value of Y	·135	...	·111	·110	...	·082	·074

(*I.A.S.*)

Solution :

X	2·0	2·1	2·2	2·3	2·4	2·5	2·6
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
Y	·135	?	·111	·110	?	·082	·074

The two missing values are obtained by solving the equations

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times} Y_{n-3} + \ldots = 0 \ldots (1)$$

and $Y_{n-1} - (n-1)Y_{n-2} + \frac{(n-1)(n-2)}{1\times 2} Y_{n-3} - \frac{(n-1)(n-2)(n-3)}{1\times 2\times 3}$

$$Y_{n-4} + \ldots = 0 \ldots\ldots (2)$$

For $n=6$ these equations become

$$Y_6 - 6Y_5 + 15Y_4 - 20Y_3 + 15Y_2 - 6Y_1 + Y_0 = 0$$

and $$Y_5 - 5Y_4 + 10Y_3 - 10Y_2 + 5Y_1 - Y_0 = 0$$

these give

$$\cdot 074 - 6\times \cdot 082 + 15Y_4 - 20\times \cdot 110 + 15\times \cdot 111 - 6Y_1 + \cdot 135 = 0$$

and $$\cdot 082 - 5Y_4 + 10\times \cdot 110 - 10\times \cdot 111 + 5Y_1 - \cdot 135 = 0$$

or $$15Y_4 - 6Y_1 = \cdot 818$$

or $$-5Y_4 + 5Y_1 = \cdot 063$$

or $$15Y_4 - 6Y_1 = \cdot 818$$

$$-15Y_4 + 15Y_1 = \cdot 189$$

Adding $$+9Y_1 = 1\cdot 007$$

$\therefore$ $$Y_1 = \cdot 1119$$

$\therefore$ $$Y_4 = \cdot 0993$$

$\therefore$ for values 2·1 and 2·4 of X the values of Y are ·1119 and ·0993 respectively.

Problem 156.—Interpolate the missing figures in the following table of rice cultivation :

Year	Acres in million	Year	Acres in millions
1911	76·6	1916	......
1912	78·7	1917	80·6
1913	......	1918	77·6
1914	77·7	1919	78·7
1915	78·7		

(*B.Com. Agra, 1937 ; M.A. Agra, 1950*)

Solution :

X	Y		X	Y	
1911	76·6	Y_0	1916	......	Y_5
1912	78·7	Y_1	1917	80·6	Y_6
1913	......	Y_2	1918	77·6	Y_7
1914	77·7	Y_3	1919	78·7	Y_8
1915	78·7	Y_4			

Two missing values are obtained by the equations

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2} Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3} Y_{n-3} + \ldots\ldots\ldots = 0$$

and $Y_{n-1}-(n-1)Y_{n-2}$

$$+\frac{(n-1)(n-2)}{1\times 2}Y_{n-3}-\frac{(n-1)(n-2)(n-3)}{1\times 2\times 3}Y_{n-4}+\ldots\ldots\ldots\ldots=0$$

for $\qquad n=8$

the equations become

$$Y_8-8Y_7+28Y_6-56Y_5+70Y_4-56Y_3+28Y_2-8Y_1-Y_0=0$$

and $\qquad Y_7-7Y_6+21Y_5-35Y_4+35Y_3-21Y_2+7Y_1-Y_0=0$

or $\qquad 78{\cdot}7-8\times 77{\cdot}6+28\times 80{\cdot}6-56Y_5+70\times 78{\cdot}7-56\times 77{\cdot}7$
$\qquad\qquad +28Y_2-8\times 78{\cdot}7+76{\cdot}6=0$

and $\qquad 77{\cdot}6-7\times 80{\cdot}6+21Y_5-35\times 78{\cdot}7+35\times 77{\cdot}7-21Y_2+7$
$\qquad\qquad \times 78{\cdot}7-76{\cdot}6=0$

or $\qquad -56Y_5+28Y_2=-2319{\cdot}5$

and $\qquad 21Y_5-21Y_2=49{\cdot}3$

or $\qquad 2Y_5-Y_2=82{\cdot}84$

and $\qquad Y_5-Y_2=2{\cdot}35$

$\therefore \qquad Y_5=80{\cdot}49$

$\qquad\qquad Y_2=78{\cdot}14$

Therefore in the year 1913 and 1916 the areas under cultivation were 78·14 and 80·49 million acres.

Problem 157.—With the help of any formula of extrapolation estimate the population of Jaipur in 1961:

Year	1891	1901	1911	1921	1931	1941	1951
Population (in thousands)	165	167	143	126	150	175	292

(Raj. M.A. Old course, 1955)

Solution :

X	1891	1901	1911	1921	1931	1941	1951	1961
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
Y	165	167	143	126	150	175	292	?

Newtons formula for one missing value is

$$Y_n-nY_{n-1}+\frac{n(n-1)}{1\times 2}Y_{n-2}-\frac{n(n-1)(n-2)}{1\times 2\times 3}Y_{n-3}+\ldots=0$$

$$Y_7-7Y_6+21Y_5-35Y_4+35Y_3-21Y_2+7Y_1-Y_0=0$$

$\therefore \quad Y_7-7\times 292+21\times 175-35\times 150+35\times 126-21\times 143+7$
$\qquad\qquad \times 167-165=0$

or $\quad Y_7-2044+21\times 32-35\times 24+1169-165=0$

or $\quad Y_7-2044+672-840+1169-165=0$

or $\quad Y_7-3049+1841=0$

or $\quad Y_7-1208=0$

$\therefore \qquad Y_7 = 1208.$

Therefore extrapolated value of population in 1961 is 1208 thousands. The value comes out to be very high which is due to the fact that the differences in the values are not regular.

Problem 158.—From the following table of membership of a certain club, calculate the probable membership for 1943, assuming a linear law :

Year	1931	1932	1933	1934	1935	1936	1937	1938
Membership	508	524	539	556	574	590	604	619
Year	1939	1940	1941	1942				
Membership	638	655	671	690				

Is your conclusion valid ? If not, why ?

(*Andhra B.Com.*, *1943*)

Solution :

X	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940
Y	508	524	539	556	574	590	604	619	638	655
	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
X	1941	1942	1943							
Y	671	690	?							
	Y_{10}	Y_{11}	Y_{12}							

Applying Newtons method for single missing value we have

$$Y_n - nY_{n-1} + \frac{n(n-1)}{1\times 2}Y_{n-2} - \frac{n(n-1)(n-2)}{1\times 2\times 3}Y_{n-3} + \ldots = 0$$

$$\therefore \; Y_{12} - 12Y_{11} + 66Y_{10} - 220Y_9 + 495Y_8 - 792Y_7 + 924Y_6 - 792Y_5 + 495Y_4 - 220Y_3 + 66Y_2 - 12Y_1 + Y_0 = 0$$

or $Y_{12} - 12\times 690 + 66\times 671 - 220\times 655 + 495\times 638 - 792\times 619 + 924\times 604 - 792\times 590 + 495\times 574 - 220\times 556 + 66\times 539 - 12\times 524 + 508 = 0$

or $Y_{12} - 12\times 1214 + 66\times 1210 - 220\times 1211 + 495\times 1212 - 792\times 1209 + 924\times 604 + 508 = 0$

or $Y_{12} - 14568 + 79860 - 266420 + 599940 - 957528 + 558096 + 508 = 0$

or $Y_{12} - 1238516 + 1238404 = 0$

or $Y_{12} - 112 = 0$

$\therefore \quad Y_{12} = 112.$

Therefore in the year 1943 membership comes out to be 112 ; which is not valid.

Because the differences in the values of Y is not regular.

Problem 159.—Find out by interpolation from the following data the number of workers earning Rs. 24 or more but less than Rs. 25.

Earning less than	*Number of workers*
Rs.	
20	296
25	599
30	804
35	918
40	966

(*U.P.C.S., 1948*)

Solution :

X	20	25	30	35	40
Y	296	599	804	918	966.

From the above values of X it is clear that the difference between any two consecutive values of X is constant and hence the value of Y for X=24 (which lies between 20 and 25) will be determined by Newtons Advancing Differences Method. The formula for which is

$$Y=Y_0+t\,(\Delta Y_0)+\frac{t\,(t-1)}{1\times 2}\,(\Delta^2 Y_0)+\frac{t\,(t-1)\,(t-2)}{1\times 2\times 3}(\Delta^3 Y_0)+\ldots\ldots$$

where $t=\dfrac{\text{Value of X for which Y is to be determined}-\text{First value of X}}{\text{Difference between any two consecutive values of X}}$

$$=\frac{24-20}{5}=\cdot 8$$

and $\Delta_0 Y$, $\Delta^2 Y_0$ etc. are determined by difference table as below :

Earning less than in Rs. X	No. of workers Y	Differences			
		1st	2nd	3rd	4th
20	296 Y_0	303 ΔY_0	-98 $\Delta^2 Y_0$	7 $\Delta^3 Y_0$	18 $\Delta^4 Y_0$
25	599 Y_1	205 ΔY_1	-91 $\Delta^2 Y_1$	25 $\Delta^3 Y_1$	
30	804 Y_2	114 ΔY_2	-66 $\Delta^2 Y_2$		
35	918 Y_3	48 ΔY_3			
40	966 Y_4				

$$\therefore\quad Y=296+\cdot 8\times 303+\frac{\cdot 8(\cdot 8-1)}{1\times 2}(-98)+\frac{\cdot 8(\cdot 8-1)(\cdot 8-2)}{1\times 2\times 3}\times 7$$

$$+\frac{\cdot 8(\cdot 8-1)(\cdot 8-2)(\cdot 8-3)}{1\times 2\times 3\times 4}\times 18$$

$$=296+242\cdot 4+\frac{\cdot 8\times\cdot 2}{2}\times 98+\frac{\cdot 8\times\cdot 2\times 1\cdot 2}{6}\times 7$$

$$-\frac{\cdot 8\times\cdot 2\times 1\cdot 2\times 2\cdot 2}{24}\times 18$$

$$=538\cdot 4+7\cdot 84+\cdot 224-\cdot 03168$$

$$=546\cdot 432$$

∴ Earning less than Rs. 24 is of 546·432 workers

and Earning less than Rs. 25 is of 599 workers.

∴ Workers earning Rs. 24 or more but less than Rs. 25 is =52·57.

Problem 160 —Given

sin 45°=·7071,
sin 50°=·7660,
sin 55°=·8192,
sin 60°=·8660,

find sin 52°, by using any method of interpolation.

(*I.A.S., 1955*)

Solution :

X	45	50	55	60
Y	·7071	·7660	·8192	·8660

The difference between any two consecutive values of X is constant and the value of Y for X=52 lies between 50 and 55 that is within the interval. Hence applying the Newtons Advancing Differences method we have the formula

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots=0$$

where

$$t=\frac{\text{value of X for which Y is to be determined.}-\text{first value of X}}{\text{Difference between any two consecutive values of X}}$$

$$=\frac{52-45}{5}=1{\cdot}4$$

and $\triangle Y_0$, $\triangle^2 Y_0$ etc. are determined from the difference table given below :

X	Y	Differences		
		1st	2nd	3rd
45	Y_0 ·7071			
50	Y_1 ·7660	·0589 $\triangle Y_0$	— ·0057 $\triangle^2 Y_0$	
55	Y_2 ·8192	·0532 $\triangle Y_1$	— ·0064 $\triangle^2 Y_1$	— ·0007 $\triangle^3 Y_0$
60	Y_3 ·8660	·0468 $\triangle Y_2$		

$$\therefore \quad Y=(\cdot 7071)+(1{\cdot}4\times\cdot 0589)-\left(\frac{1{\cdot}4\times\cdot 4}{1\times 2}\times\cdot 0057\right)+\left(\frac{1{\cdot}4\times\cdot 4\times\cdot 6}{1\times 2\times 3}\times\cdot 0007\right)$$

$$=\cdot 7071+\cdot 08246-\cdot 001546+\cdot 0000392$$

$$=\cdot 7880532$$

∴ sin 52°=·7880532.

Problem 161.—The following table gives the census population of an Indian State in 1901, 1911, 1921 and 1931.

Estimate the population of State in 1924 making your method clear.

Year	1901	1911	1921	1931
Population (in thousands)	2797	2935	3047	3354

(*P.C.S., 1939*)

Solution :

X	1901	1911	1921	1931
Y	2797	2935	3047	3354

Applying the Newtons Advancing Differences method we have

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1\times 2}(\triangle^2 Y_0) + \frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0) + \ldots\ldots = 0$$

$$t = \frac{1924 - 1901}{10} = 2{\cdot}3$$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Year	Population in thousands	Differences		
X	Y	1st	2nd	3rd
1901	2797 Y_0	138 $\triangle Y_0$	-26 $\triangle^2 Y_0$	221 $\triangle^3 Y_0$
1911	2935 Y_1	112 $\triangle Y_1$	195 $\triangle^2 Y_1$	
1921	3047 Y_2	307 $\triangle Y_2$		
1931	3354 Y_3			

$$\therefore \quad Y = 2797 + 2{\cdot}3 \times 138 - \frac{2{\cdot}3 \times 1{\cdot}3}{1 \times 2} \times 26 + \frac{2{\cdot}3 \times 1{\cdot}3 \times {\cdot}3}{1 \times 2 \times 3} \times 221$$

$$= 2797 + 317{\cdot}4 - 38{\cdot}87 + 32{\cdot}5395$$

$$= 3108{\cdot}0695$$

$\therefore$ Population in the year 1924 comes out to be 3108 thousands approximately.

Problem 162.—The following are the marks obtained by 492 candidates in a certain examination :

Not more than 40 marks	212	candidates
,, ,, ,, 45 ,,	296	,,
,, ,, ,, 50 ,,	368	,,
,, ,, ,, 55 ,,	429	,,
,, ,, ,, 60 ,,	460	,,
,, ,, ,, 65 ,,	481	,,
,, ,, ,, 70 ,,	490	,,
,, ,, ,, 75 ,,	492	,,

Find out the number of candidates who secured more than 42 but not more than 45 marks.

(*M.A. Cal.*, *1935* ; *M.Com. Alld.*, *1947*)

Solution :

To determine the number of candidates who secured more than 42 marks :—

X	40	45	50	55	60	65	70	75
Y	212	296	368	429	460	481	490	492

Applying the Newtons Advancing Differences method we have

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1\times 2}(\triangle^2 Y_0) + \frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0) + \ldots\ldots$$

where $t = \frac{42-40}{5} = {\cdot}4$

for $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the difference table as below :

Not more than X marks	No. of Candidates Y	Differences						
		1st	2nd	3rd	4th	5th	6th	7th
40	$212 Y_0$	$84 \triangle Y_0$	$-12 \triangle^2 Y_0$	$1 \triangle^3 Y_0$	$-20 \triangle^4 Y_0$	$59 \triangle^5 Y_0$	$-120 \triangle^6 Y_0$	$210 \triangle^7 Y_0$
45	$296 Y_1$	$72 \triangle Y_1$	$-11 \triangle^2 Y_1$	$-19 \triangle^3 Y_1$	$39 \triangle^4 Y_1$	$-61 \triangle^5 Y_1$	$90 \triangle^6 Y_1$	
50	$368 Y_2$	$61 \triangle Y_2$	$-30 \triangle^2 Y_2$	$20 \triangle^3 Y_2$	$-22 \triangle^4 Y_2$	$29 \triangle^5 Y_2$		
55	$429 Y_3$	$31 \triangle Y_3$	$-10 \triangle^2 Y_3$	$-2 \triangle^3 Y_3$	$7 \triangle^4 Y_3$			
60	$460 Y_4$	$21 \triangle Y_4$	$-12 \triangle^2 Y_4$	$5 \triangle^3 Y_4$				
65	$481 Y_5$	$9 \triangle Y_5$	$-7 \triangle^2 Y_5$					
70	$490 Y_6$	$2 \triangle Y_6$						
75	$492 Y_7$							

$$\therefore\ Y = 212 + {\cdot}4\times 84 + \frac{{\cdot}4\times{\cdot}6}{1\times 2}\times 12 + \frac{{\cdot}4\times{\cdot}6\times 1{\cdot}6}{1\times 2\times 3}\times 1 + \frac{{\cdot}4\times{\cdot}6\times 1{\cdot}6\times 2{\cdot}6}{1\times 2\times 3\times 4}\times 20$$

$$+ \frac{{\cdot}4\times{\cdot}6\times 1{\cdot}6\times 2{\cdot}6\times 3{\cdot}6}{1\times 2\times 3\times 4\times 5}\times 59 + \frac{{\cdot}4\times{\cdot}6\times 1{\cdot}6\times 2{\cdot}6\times 3{\cdot}6\times 4{\cdot}6}{1\times 2\times 3\times 4\times 5\times 6}\times 120$$

$$+ \frac{{\cdot}4\times{\cdot}6\times 1{\cdot}6\times 2{\cdot}6\times 3{\cdot}6\times 4{\cdot}6\times 5{\cdot}6}{1\times 2\times 3\times 4\times 5\times 6\times 7}\times 210$$

$$= 212 + 33{\cdot}6 + 1{\cdot}44 + {\cdot}064 + {\cdot}832 + 1{\cdot}716 + 2{\cdot}755 + 3{\cdot}858$$

$$= 256{\cdot}256 = 256 \text{ Appr.}$$

Thus number of candidates getting not more than 42 is 256.

and the number of candidates getting not more than 45 is 296.

$\therefore$ No. of candidates securing more than 42 but not more than 45 marks is $296 - 256 = 40$.

Problem 163.—Using any method of interpolation, find the value of Y corresponding to X=21·64 from the following table :

X	20	21	22	23	24	25
Y	2727	3345	4041	4732	5496	6300

(*I.A.S.*)

Solution :

X	20	21	22	23	24	25
Y	2727	3345	4041	4732	5496	6300

Applying the Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times2\times3}(\triangle^3 Y_0)+\ldots\ldots$$

where
$$t=\frac{21{\cdot}64-20}{1}=1{\cdot}64$$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the difference table as below :

X	Y	Differences 1st	2nd	3rd	4th	5th
20	2727 Y_0	618 $\triangle Y_0$	78 $\triangle^2 Y_0$	−83 $\triangle^3 Y_0$	161 $\triangle^4 Y_0$	−272 $\triangle^5 Y_0$
21	3345 Y_1	696 $\triangle Y_1$	−5 $\triangle^2 Y_1$	78 $\triangle^3 Y_1$	−111 $\triangle^4 Y_1$	
22	4041 Y_2	691 $\triangle Y_2$	73 $\triangle^2 Y_2$	−33 $\triangle^3 Y_2$		
23	4732 Y_3	764 $\triangle Y_3$	40 $\triangle^2 Y_3$			
24	5496 Y_4	804 $\triangle Y_4$				
25	6300 Y_5					

$$\therefore\quad Y=2727+1{\cdot}64\times618+\frac{1{\cdot}64\times{\cdot}64}{1\times2}\times78+\frac{1{\cdot}64\times{\cdot}64\times{\cdot}36}{1\times2\times3}\times83$$

$$+\frac{1{\cdot}64\times{\cdot}64\times{\cdot}36\times1{\cdot}36}{1\times2\times3\times4}\times161+\frac{1{\cdot}64\times{\cdot}64\times{\cdot}36\times1{\cdot}36\times2{\cdot}36}{1\times2\times3\times4\times5}\times272$$

$$=2727+1013{\cdot}52+87{\cdot}7344+11{\cdot}205+7{\cdot}3899+5{\cdot}8752$$

$$=3852{\cdot}7245.$$

Problem 164.—Obtain with the help of any suitable interpolation formula the value of Y for X=6·3 in the following table, and compare the result with the value obtained by graphical interpolation. Discuss briefly the relative accuracy and convenience of the two methods.

X	4	5	6	7	8	9
Y	20·256	20·625	21·296	22·401	24·096	26·561

(*I.A.S.*)

Solution :

X	4	5	6	7	8	9
Y	20·256	20·625	21·296	22·401	24·096	26·561

Applying Newtons Advancing Differences method we have

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1\times 2}(\triangle^2 Y_0) + \frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0) + \ldots\ldots$$

where $= \frac{6\cdot 3-4}{1} = 2\cdot 3$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

X	Y	Differences 1st	2nd	3rd	4th	5th
4	20·256 Y_0	·369 $\triangle Y_0$	·302 $\triangle^2 Y_0$	·132 $\triangle^3 Y_0$	·024 $\triangle^4 Y_0$	0 $\triangle^5 Y_0$
5	20·625 Y_1	·671 $\triangle Y_1$	·434 $\triangle^2 Y_1$	·156 $\triangle^3 Y_1$	·024 $\triangle^4 Y_1$	
6	21·296 Y_2	1·105 $\triangle Y_2$	·590 $\triangle^2 Y_2$	·180 $\triangle^3 Y_2$		
7	22·401 Y_3	1·695 $\triangle Y_3$	·770 $\triangle^2 Y_3$			
8	24·096 Y_4	2·465 $\triangle Y_4$				
9	26·561 Y_5					

$$\therefore\quad Y = 20\cdot 256 + 2\cdot 3\times \cdot 369 + \frac{2\cdot 3\times 1\cdot 3}{1\times 2}\times \cdot 302 + \frac{2\cdot 3\times 1\cdot 3\times \cdot 3}{1\times 2\times 3}\times \cdot 132$$

$$-\frac{2\cdot 3\times 1\cdot 3\times \cdot 3\times \cdot 7}{1\times 2\times 3\times 4}\times \cdot 024 + \frac{2\cdot 3\times 1\cdot 3\times \cdot 3\times \cdot 7\times 1\cdot 7}{1\times 2\times 3\times 4\times 5}\times 0$$

$$= 20\cdot 256 + \cdot 8487 + \cdot 4515 + \cdot 0197 - \cdot 00063$$

$$= 21\cdot 575.$$

Problem 165.—The following are the marks abtained by 492 candidates in a certain examination :

Not more than 40 marks	210 candidates
,, ,, 45 ,,	253 ,,
,, ,, 50 ,,	307 ,,
,, ,, 55 ,,	381 ,,
,, ,, 60 ,,	413 ,,
,, ,, 65 ,,	492 ,,

Find out the number of candidates (*a*) who secured more than 48 but not more than 50 marks. (*b*) less than 48 but not less than 45 marks.

(*P.C.S., 1954*)

Solution :

To determine both the results the value for not more than 48 marks is to be interpolated.

∴						
X	40	45	50	55	60	65
Y	210	253	307	381	413	492

Applying Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots\ldots\ldots$$

where $t=\frac{48-40}{5}=1{\cdot}6$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Not more than X marks	No. of candidates Y	Differences 1st	2nd	3rd	4th	5th
40	210 Y_0					
		43 $\triangle Y_0$				
45	253 Y_1		11 $\triangle^2 Y_0$			
		54 $\triangle Y_1$		9 $\triangle^3 Y_0$		
50	307 Y_2		20 $\triangle^2 Y_1$		−71 $\triangle^4 Y_0$	
		74 $\triangle Y_2$		−62 $\triangle^3 Y_1$		222 $\triangle^5 Y_0$
55	381 Y_3		−42 $\triangle^2 Y_2$		151 $\triangle^4 Y_1$	
		32 $\triangle Y_3$		89 $\triangle^3 Y_2$		
60	413 Y_4		47 $\triangle^2 Y_3$			
		79 $\triangle Y_4$				
65	492 Y_5					

$$\therefore\quad Y=210+1{\cdot}6\times 43+\frac{1{\cdot}6\times{\cdot}6}{1\times 2}\times 11-\frac{1{\cdot}6\times{\cdot}6\times{\cdot}4}{1\times 2\times 3}\times 9$$

$$-\frac{1{\cdot}6\times{\cdot}6\times{\cdot}4\times 1{\cdot}4}{1\times 2\times 3\times 4}\times 71-\frac{1{\cdot}6\times{\cdot}6\times{\cdot}4\times 1{\cdot}4\times 2{\cdot}4}{1\times 2\times 3\times 4\times 5}\times 222$$

or $Y=210+68{\cdot}8+5{\cdot}28-{\cdot}576-1{\cdot}59-2{\cdot}387$
$=279{\cdot}53=280$ Appr.

$\therefore$ No. of candidates securing not more than 48 marks is 280

$\therefore$ (*a*) No. of candidates who secured more than 48 but not more than 50 marks is $=307-280=27$

(*b*) No. of candidates who secured less than 48 but not less than 45 marks is $=280-253=27$.

Problem 166.—Estimate, by the method of interpolation, the expectation of life at age 22 from the following data, stating the assumptions underlying the formula used by you.

Age	10	15	20	25	30	35
Expectation of life (in years)	35·4	32·2	29·1	26·0	23·1	20·4

(I.A.S., 1949)

Solution :

X	10	15	20	25	30	35
Y	35·4	32·2	29·1	26·0	23·1	20·4

Applying the Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots\ldots$$

where $t=\frac{22-10}{5}=2{\cdot}4$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Age X	Expectation of life in years Y	Differences 1st	2nd	3rd	4th	5th
10	35·4 Y_0	$-3·2$ $\triangle Y_0$	·1 $\triangle^2 Y_0$	$-·1$ $\triangle^3 Y_0$	·3 $\triangle^4 Y_0$	$-·5$ $\triangle^5 Y_0$
15	32·2 Y_1	$-3·1$ $\triangle Y_1$	0 $\triangle^2 Y_1$	·2 $\triangle^3 Y_1$	$-·2$ $\triangle^4 Y_1$	
20	29·1 Y_2	$-3·1$ $\triangle Y_2$	·2 $\triangle^2 Y_2$	0 $\triangle^3 Y_2$		
25	26·0 Y_3	$-2·9$ $\triangle Y_3$	·2 $\triangle^2 Y_3$			
30	23·1 Y_4	$-2·7$ $\triangle Y_4$				
35	20·4 Y_5					

$$\therefore \quad Y = 35·4 - 2·4 \times 3·2 + \frac{2·4 \times 1·4}{1 \times 2} \times ·1 - \frac{2·4 \times 1·4 \times ·4}{1 \times 2 \times 3 \times 4} \times ·1$$

$$- \frac{2·4 \times 1·4 \times ·4 \times ·6}{1 \times 2 \times 3 \times 4} \times ·3 - \frac{2·4 \times 1·4 \times ·4 \times ·6 \times 1·6}{1 \times 2 \times 3 \times 4 \times 5} \times ·5$$

$$= 35·4 - 7·68 + ·168 - ·0224 - ·01008 - ·005376$$

$$= 27·85.$$

Problem 167.—The following table gives the quantities of a certain brand of tea sold at prices noted against each :

Price of tea per pound (in Rs.)	Quantity sold in thousand pounds
1—4 —0	97·5
1—8 —0	79·7
1—12—0	65·9
2—0 —0	56·0
2—4 —0	49·9

Estimate the probable sale when the price is 1—6—0 per pound.

(*Patna, M.A., 1944*)

Solution :

Price of tea per pound in As.	X	20	24	28	32	36
Quantity sold in thousand pounds	Y	97·5	79·7	65·9	56·0	49·9

Applying Newtons method for Advancing Differences we have

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1 \times 2}(\triangle^2 Y_0) + \frac{t(t-1)(t-2)}{1 \times 2 \times 3}(\triangle^3 Y_0) + \ldots\ldots$$

where $t = \frac{22-20}{4} = ·5$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table os below :

X	Y	Differences			
		1st	2nd	3rd	4th
20	97·5 Y_0				
24	79·7 Y_1	−17·8 $\triangle Y_0$	4 $\triangle^2 Y_0$	·1 $\triangle^3 Y_0$	−·4 $\triangle^4 Y_0$
28	65·9 Y_2	−13·8 $\triangle Y_1$	4·1 $\triangle^2 Y_1$	−·3 $\triangle^3 Y_1$	
32	56·0 Y_3	− 9·9 $\triangle Y_2$	3·8 $\triangle^2 Y_2$		
36	49·9 Y_4	− 6·1 $\triangle Y_3$			

$$Y = 97{\cdot}5 - {\cdot}5 \times 17{\cdot}8 - \frac{{\cdot}5 \times {\cdot}5}{1 \times 2} \times 4 + \frac{{\cdot}5 \times {\cdot}5 \times 1{\cdot}5}{1 \times 2 \times 3} \times {\cdot}1 + \frac{{\cdot}5 \times {\cdot}5 \times 1{\cdot}5 \times 2{\cdot}5}{1 \times 2 \times 3 \times 4} \times {\cdot}4$$

$$= 97{\cdot}5 - 8{\cdot}9 - {\cdot}5 + {\cdot}0042 + {\cdot}0156$$

$$= 97{\cdot}5198 - 9{\cdot}4$$

$$= 88{\cdot}1198$$

∴ Estimated sale is 88·12 thousand pounds when the price is Rs. 1—6—0 per pound.

Problem 168.—Find an interpolated figure for population of 1896 from the following table :—

Year	1881	1891	1901	1911
Population	25974	29003	32528	36070

(*Nagpur, B.Com., 1942*)

Solution :

X	1881	1891	1901	1911
Y	25974	29003	32528	36070

Applying Newton's method of Advancing Differences we have

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1 \times 2}(\triangle^2 Y_0) + \frac{t(t-1)\ (t-2)}{1 \times 2 \times 3}(\triangle^3 Y_0) + \ldots$$

where $t = \dfrac{1896 - 1881}{10} = 1{\cdot}5$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below:

X	Y	Differences		
		1st	2nd	3rd
1881	25974 Y_0	3029 $\triangle Y_0$	496 $\triangle^2 Y_0$	−479 $\triangle^3 Y_0$
1891	29003 Y_1	3525 $\triangle Y_1$	17 $\triangle^2 Y_1$	
1901	32528 Y_2	3542		
1911	36070 Y_3			

$$\therefore\ Y = 25974 + 1{\cdot}5 \times 3029 + \frac{1{\cdot}5 \times {\cdot}5}{1 \times 2} \times 496 + \frac{1{\cdot}5 \times {\cdot}5 \times {\cdot}5}{1 \times 2 \times 3} \times 479$$

$$= 25974 + 4543{\cdot}5 + 186 + 29{\cdot}94$$

$$= 30733{\cdot}44$$

Therefore interpolated figure for population of 1896 is 30733·44

Problem 169.—The following table gives the different premiums at different ages in a life Assurance Company.

Age	25	30	35	40	45	50
Premium in Rs.	23	26	30	35	42	51

Find the premium at age 28.

(*Nagpur, B.Com., 1944*)

Solution :

X	25	30	35	40	45	50
Y	23	26	30	35	42	51

Newtons formula for Advancing Differences method is

$$Y = Y_0 + t(\Delta Y_0) + \frac{t(t-1)}{1 \times 2}(\Delta^2 Y_0) + \frac{t(t-1)(t-2)}{1 \times 2 \times 3}(\Delta^3 Y_0) + \ldots$$

where $t = \frac{28-25}{5} = {\cdot}6$

For ΔY_0, $\Delta^2 Y_0$ etc. we have the Difference table as below :

X	Y	Differences 1st	2nd	3rd	4th	5th
25	23 Y_0	3 ΔY_0	1 $\Delta^2 Y_0$	0 $\Delta^3 Y_0$	1 $\Delta^4 Y_0$	$-2 \Delta^5 Y_0$
30	26 Y_1	4 ΔY_1	1 $\Delta^2 Y_1$	1 $\Delta^3 Y_1$	$-1 \Delta^4 Y_1$	
35	30 Y_2	5 ΔY_2	2 $\Delta^2 Y_2$	0 $\Delta^3 Y_2$		
40	35 Y_3	7 ΔY_3	2 $\Delta^2 Y_3$			
45	42 Y_4	9 ΔY_4				
50	51 Y_5					

$$Y = 23 + {\cdot}6 \times 3 - \frac{{\cdot}6 \times {\cdot}4}{1 \times 2} \times 1 + 0 - \frac{{\cdot}6 \times {\cdot}4 \times 1{\cdot}4 \times 2{\cdot}4}{1 \times 2 \times 3 \times 4} \times 1 - \frac{{\cdot}6 \times {\cdot}4 \times 1{\cdot}4 \times 2{\cdot}4 \times 3{\cdot}4}{1 \times 2 \times 3 \times 4 \times 5} \times 2$$

$$= 23 + 1{\cdot}8 - {\cdot}12 + 0 - {\cdot}0336 - {\cdot}0457$$

$$= 24{\cdot}8 - {\cdot}1993$$

$$= 24{\cdot}6007$$

Therefore premium at age 28 is 24·6 rupees.

Problem 170.—Find the annual premiums for ages 34 and 36 nearer birth day respectively :

Age nearer birth day	*Annual premium in Rs.*
25	27 14 0
30	31 13 0
35	37 0 0
40	43 13 0
45	52 12 0

(*Andhra, B. Com., 1944*)

Solution :

Age	X	25	30	35	40	45
Annual premium in As.	Y	446	509	592	701	844

Newtons formula for Advancing Differences method is

$$Y=Y_0+t(\Delta Y_0)+\frac{t(t-1)}{1\times 2}(\Delta^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\Delta^3 Y_0)+\ldots$$

where (*i*) $t=\frac{34-25}{5}=1\cdot 8$

(*ii*) $t=\frac{36-25}{5}=2\cdot 2$

For ΔY_0 $\Delta^2 Y_0$ etc. we have the difference table as given below :

X	Y		Differences 1st	2nd	3rd	4th
25	446	Y_0	63 ΔY_0	20 $\Delta^2 Y_0$	6 $\Delta^3 Y_0$	2 $\Delta^4 Y_0$
30	509	Y_1	83 ΔY_1	26 $\Delta^2 Y_1$	8 $\Delta^3 Y_1$	
35	592	Y_2	109 ΔY_2	34 $\Delta^2 Y_2$		
40	701	Y_3	143 ΔY_3			
45	844	Y_4				

$\therefore$ (*i*) $Y_{34}=446+1\cdot 8\times 63+\frac{1\cdot 8\times\cdot 8}{1\times 2}\times 20-\frac{1\cdot 8\times\cdot 8\times\cdot 2}{1\times 2\times 3}\times 6$

$+\frac{1\cdot 8\times\cdot 8\times\cdot 2\times 1\cdot 2}{1\times 2\times 3\times 4}\times 2$

$=573\cdot 54$ Annas $=$ Rs. 35-13-6

(*ii*) $Y_{36}=446+2\cdot 2\times 63+\frac{2\cdot 2\times 1\cdot 2}{1\times 2}\times 20+\frac{2\cdot 2\times 1\cdot 2\times\cdot 2}{1\times 2\times 3}\times 6$

$-\frac{2\cdot 2\times 1\cdot 2\times\cdot 2\times\cdot 8}{1\times 2\times 3\times 4}\times 2$

$=611\cdot 49$ Annas $=$ Rs. 38-3-6.

Therefore premiums for ages 34 and 36 are Rs. 35-13-6 and Rs. 38-3-6 respectively.

Problem 171.—Use Newtons formula to estimate the population in 1925 of a place having the following record :

Year :	1891	1901	1911	1921	1931
Population in thousands	46	66	11	93	101

(*Raj. M. A., 1949*)

Solution :

X	1891	1901	1911	1921	1931
Y	46	66	81	93	101

Newtons formula for Advancing Differences is

$$Y=Y_0+t(\Delta Y_0)+\frac{t(t-1)}{1\times 2}(\Delta^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\Delta^3 Y_0)+\ldots\ldots$$

where $t=\frac{1925-1891}{10}$

$=3{\cdot}4$

ΔY_0, $\Delta^2 Y_0$ etc. are given by the following Difference table.

Year X	Population in thousands Y	Differences 1st	2nd	3rd	4th
1891	46 Y_0	20 ΔY_0	-5 $\Delta^2 Y_0$	2 $\Delta^3 Y_0$	-3 $\Delta^4 Y_0$
1901	66 Y_1	15 ΔY_1	-3 $\Delta^2 Y_1$	-1 $\Delta^3 Y_1$	
1911	81 Y_2	12 ΔY_2	-4 $\Delta^2 Y_2$		
1921	93 Y_3	8 ΔY_3			
1931	101 Y_4				

$$\therefore \quad Y=46+3{\cdot}4\times 20+\frac{3{\cdot}4\times 2{\cdot}4}{1\times 2}\times(-5)+\frac{3{\cdot}4\times 2{\cdot}4\times 1{\cdot}4}{1\times 2\times 3}\times 2$$

$$+\frac{3{\cdot}4\times 2{\cdot}4\times 1{\cdot}4\times {\cdot}4}{1\times 2\times 3\times 4}\times(-3)$$

$$=46+68-20{\cdot}4+3{\cdot}808-{\cdot}5712$$

$$=96{\cdot}8368$$

$\therefore$ Population in 1925 is found to be 96·8368 thousands.

Problem 172.—From the following table, find the number of students who obtained less than 45 marks :

Marks	*No. of students*
30—40	31
40—50	42
50—60	51
60—70	35
70—80	31

(*Alld. M.Com., 1952*)

Solution :

Marks more than 30 but less than	X	40	50	60	70	80
No. of Students	Y	31	73	124	159	190

Applying Newtons formula for Advancing Differences we have

$$Y = Y_0 + t(\Delta Y_0) + \frac{t(t-1)}{1\times 2}(\Delta^2 Y_0) + \frac{t(t-1)(t-2)}{1\times 2\times 3}(\Delta^3 Y_0) + \ldots$$

where $t = \frac{45-40}{10} = \cdot 5$

For ΔY_0, $\Delta^2 Y_0$ etc. we have the Difference table as below :

Marks more than 30 but less than X	No. of Students Y	Differences 1st	2nd	3rd	4th
40	31 Y_0	42 ΔY_0	9 $\Delta^2 Y_0$	-25 $\Delta^3 Y_0$	47 $\Delta^4 Y_0$
50	73 Y_1	51 ΔY_1	-16 $\Delta^2 Y_1$	12 $\Delta^3 Y_1$	
60	124 Y_2	35 ΔY_2	-4 $\Delta^2 Y_3$		
70	159 Y_3	31 ΔY_3			
80	190 Y_4				

$$\therefore \quad Y = 31 + \cdot 5\times 42 - \frac{\cdot 5\times \cdot 5}{1\times 2}\times 9 - \frac{\cdot 5\times \cdot 5\times 1\cdot 5}{1\times 2\times 3}\times 25$$
$$- \frac{\cdot 5\times \cdot 5\times 1\cdot 5\times 2\cdot 5}{1\times 2\times 3\times 4}\times 47$$

$= 31 + 21 - 1\cdot 125 - 1\cdot 562 - 1\cdot 836$
$= 52 - 4\cdot 523$
$= 47\cdot 477.$

Therefore number of students getting less than 45 marks are 48 nearly.

Problem 173.—The following figures refer to the working of a Tramway.

Rate per unit (pice)	*Number of passengers*
5	30,000
4·5	40,000
4	60,000
3·5	100,000
3	150,000

Estimate the probable number of passengers if the rate be 4·2.
(*M.Com., Agra, 1949*)

Solution :

Rate per unit (pice)	X	5	4·5	4	3·5	3
No. of passengers (in thousand)	Y	30	40	60	100	150

Applying Newtons formula for Advancing Differences method we have

$$Y=Y_0+t(\Delta Y_0)+\frac{t(t-1)}{1\times 2}(\Delta^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\Delta^3 Y_0)+\ldots\ldots$$

where $t=\frac{4\cdot 2-5}{-\cdot 5}=+1\cdot 6$

For ΔY_0, $\Delta^2 Y_0$ etc. we have the Difference table as below :

Rate per unit (in pice) X	No. of passengers (in thousands) Y	Differences 1st	2nd	3rd	4th
5	30 Y_0				
4·5	40 Y_1	10 ΔY_0	10 $\Delta^2 Y_0$		
4	60 Y_2	20 ΔY_1	20 $\Delta^2 Y_1$	10 $\Delta^3 Y_0$	−20 $\Delta^4 Y_0$
3·5	100 Y_3	40 ΔY_2	10 $\Delta^2 Y_2$	−10 $\Delta^3 Y_1$	
3	150 Y_4	50 ΔY_3			

$$\therefore \quad Y=30+1\cdot 6\times 10+\frac{1\cdot 6\times \cdot 6}{1\times 2}\times 10-\frac{1\cdot 6\times \cdot 6\times \cdot 4}{1\times 2\times 3}\times 10$$

$$-\frac{1\cdot 6\times \cdot 6\times \cdot 4\times 1\cdot 4}{1\times 2\times 3\times 4}\times 20$$

$$=30+16+4\cdot 8-\cdot 64-\cdot 448$$

$$=50\cdot 8-1\cdot 088$$

$$=49\cdot 722 \text{ thousands.}$$

Therefore probable number of passengers if the rate be 4·2 are 49722.

Problem 174.—Estimate the expectation of life at the age of 16 years using the following data :

Age in years	10	15	20	25	30	35
Expectation of life (in years)	35·4	32·3	29·2	26·0	23·2	20·4

(*P.C.S., 1951*)

Solution :

X	10	15	20	25	30	35
Y	35·4	32·3	29·2	26	23·2	20·4

Applying Newtons Advancing Differences method we have

$$Y=Y_0+t(\Delta Y_0)+\frac{t(t-1)}{1\times 2}(\Delta^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\Delta^3 Y_0)+\ldots\ldots$$

where $t=\frac{16-10}{5}=1\cdot 2$

For $\triangle Y_0$, $\Delta^2 Y_0$ etc. we have the Difference table as below :

Age in Years X	Expectation of life in years Y	Differences 1st	2nd	3rd	4th	5th
10	35·4 Y_0	$-3·1$ ΔY_0	0 $\Delta^2 Y_0$	$-·1$ $\Delta^3 Y_0$	·6 $\Delta^4 Y_0$	$-1·5$ $\Delta^5 Y_0$
15	32·3 Y_1	$-3·1$ ΔY_1	$-·1$ $\Delta^2 Y_1$	·5 $\Delta^3 Y_1$	$-·9$ $\Delta^4 Y_1$	
20	29·2 Y_2	$-3·2$ ΔY_2	$+·4$ $\Delta^2 Y_2$	$-·4$ $\Delta^3 Y_2$		
25	26·0 Y_3	$-2·8$ ΔY_3	0 $\Delta^2 Y_3$			
30	23·2 Y_4	$-2·8$ ΔY_4				
35	20·4 Y_5					

$$\therefore \quad Y = 35·4 - 1·2 \times 3·1 + 0 + \frac{1·2 \times ·2 \times ·8}{1 \times 2 \times 3} \times ·1 + \frac{1·2 \times ·2 \times ·8 \times 1·8}{1 \times 2 \times 3 \times 4} \times ·6 + \frac{1·2 \times ·2 \times ·8 \times 1·8 \times 2·8}{1 \times 2 \times 3 \times 4 \times 5} \times 1·5$$

$$= 35·4 - 3·72 + ·0032 + ·00864 + ·012$$

$$= 31·7.$$

Problem 177.—What do you understand by interpolation ? What are its uses ? Describe some method of interpolation known to you, and apply the method in calculating the expectation of life at age 24, using the following data :

Age	10	15	20	25	30	35
Expectation in life	35·45	32·20	29·06	26·03	23·11	20·41

(P.C.S., B.Com. Nagpur, 1943)

Solution:

X	10	15	20	25	30	35
Y	35·45	32·20	29·06	26·03	23·11	20·41

Applying Newtons Advancing Differences method we have

$$Y = Y_0 + t\,(\triangle Y_0) + \frac{t\,(t-1)}{1 \times 2}\,(\triangle^2 Y_0) + \frac{t\,(t-1)(t-2)}{1 \times 2 \times 3}\,(\Delta^3 Y_0) + \ldots\ldots$$

where $\qquad t = \frac{24-10}{5} = 2·8$

For $\triangle Y_0$, $\Delta^2 Y_0$ etc. we have the Difference table as below :

Age X	Expectation in life Y	Differences 1st	2nd	3rd	4th	5th
10	35·45 Y_0	$-3·25$ ΔY_0	·11 $\Delta^2 Y_0$	0 $\Delta^3 Y_0$	0 $\Delta^4 Y_0$	·11 $\Delta^5 Y_0$
15	32·20 Y_1	$-3·14$ ΔY_1	·11 $\Delta^2 Y_1$	0 $\Delta^3 Y_1$	·11 $\Delta^4 Y_1$	
20	29·06 Y_2	$-3·03$ ΔY_2	·11 $\Delta^2 Y_2$	.11 $\Delta^3 Y_2$		
25	26·03 Y_3	$-2·92$ ΔY_3	·22 $\Delta^2 Y_3$			
30	23·11 Y_4	$-2·70$ ΔY_4				
35	20·41 Y_5					

$$\therefore \quad Y = 35{\cdot}45 - 2{\cdot}8 \times 3{\cdot}25 + \frac{2{\cdot}8 \times 1{\cdot}8}{1 \times 2} \times {\cdot}11 + 0 + 0 + \frac{2{\cdot}8 \times 1{\cdot}8 \times {\cdot}8 \times {\cdot}2 \times 1{\cdot}2}{1 \times 2 \times 3 \times 4 \times 5} \times {\cdot}11$$

$$= 35{\cdot}45 - 9{\cdot}1 + {\cdot}277 + {\cdot}004$$

$$= 26{\cdot}63.$$

Problem 176.—State Newtons formula for interpolation for equal intervals and the assumptions underlying it. Use it to find the annual net premium at age 25 from the table given below :

Age	20	24	28	32
Annual net premium.	·01427	·01581	·01772	·01996

(*I.A.S., 1950 ; Agra M.A. 1956*)

Solution :

X	20	24	28	32
Y	·01427	·01581	·01772	·01996

Applying the Newtons Advancing Differences method we have

$$Y = Y_0 + t\,(\triangle Y_0) + \frac{t\,(t-1)}{1 \times 2}\,(\triangle^2 Y_0) + \frac{t\,(t-1)(t-2)}{1 \times 2 \times 3}\,(\triangle^3 Y_0) + \ldots\ldots$$

where $$t = \frac{25-20}{4} = 1{\cdot}25$$

For $\triangle Y_0$, $\triangle Y_0$ etc. we have the Difference table as below :

Age X	Annual net premium Y	Differences 1st	2nd	3rd
20	·01427 Y_0	·00154 $\triangle Y_0$	·00037 $\triangle^2 Y_0$	—·00004 $\triangle^3 Y_0$
24	·01581 Y_1	·00191 $\triangle Y_1$	·00033 $\triangle^2 Y_1$	
28	·01772 Y_2	·00224 $\triangle Y_2$		
32	·01996 Y_3			

$$\therefore \quad Y = {\cdot}01427 + 1{\cdot}25 \times {\cdot}00154 + \frac{1{\cdot}25 \times {\cdot}25}{1 \times 2} \times {\cdot}00037 + \frac{1{\cdot}25 \times {\cdot}25 \times {\cdot}75}{1 \times 2 \times 3} \times {\cdot}00004$$

$$= {\cdot}01427 + {\cdot}001925 + {\cdot}000058 + {\cdot}0000016$$

$$= {\cdot}0162546.$$

Problem 177.—Estimate the population in 1925 of a place having the following record :

Year	1891	1901	1911	1921	1931
Population in thousand	46	66	81	93	101

(*M.A Raj., 1950*)

Solution :

X	1891	1901	1911	1921	1931
Y	46	66	81	93	101

Applying the Newtons Advancing Differences method we have

$$Y = Y_0 + t(\Delta Y_0) + \frac{t(t-1)}{1\times 2}(\Delta^2 Y_0) + \frac{t(t-1)(t-2)}{1\times 2\times 3}(\Delta^3 Y_0) + \ldots\ldots$$

where $t = \frac{1925-1891}{10} = 3{\cdot}4$

For ΔY_0, $\Delta^2 Y_0$ etc. we have the Difference table as below :

Year X	Population in thousands Y	Differences 1st	2nd	3rd	4th
1891	46 Y_0	20 ΔY_0	-5 $\Delta^2 Y_0$	2 $\Delta^3 Y_0$	-9 $\Delta^4 Y_0$
1901	66 Y_1	15 ΔY_1	-3 $\Delta^2 Y_1$	-7 $\Delta^3 Y_1$	
1911	81 Y_2	12 ΔY_2	-4 $\Delta^2 Y_2$		
1921	93 Y_3	8 ΔY_3			
1931	101 Y_4				

$$\therefore \quad Y = 46 + 3{\cdot}4\times 20 - \frac{3{\cdot}4\times 2{\cdot}4}{1\times 2}\times 5 + \frac{3{\cdot}4\times 2\,4\times 1{\cdot}4}{1\times 2\times 3}\times 2 - \frac{3{\cdot}4\times 2{\cdot}4\times 1{\cdot}4\times{\cdot}4}{1\times 2\times 3\times 4}\times 9$$

$$= 46 + 68 - 30{\cdot}4 + 3{\cdot}808 - 1{\cdot}714$$

$$= 95{\cdot}694.$$

Problem 178.—The following are the annual premiums in a certain life Insurance Company for a policy of Rs. 500 payable at the death with an agreed bonus :

Age Next birthday	Annual Premium Rs. As.
25	24—10
30	27—11
35	31— 9
40	36— 6
45	42— 5

Calculate the premium at age 36.

(*M.Com. Luck., 1942*)

Solution :

X	25	30	35	40	45
Y	394 As.	443 As.	505 As.	582 As.	677 As.

Applying the Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots\ldots$$

where $t=\frac{36-25}{5}=2{\cdot}2$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Age Next Birthday X	Annual Premium in As. Y	Differences 1st	2nd	3rd	4th
25	394 Y_0				
30	443 Y_1	49 $\triangle Y_0$	13 $\triangle^2 Y_0$	2 $\triangle^2 Y_0$	1 $\triangle^3 Y_0$
35	505 Y_2	62 $\triangle Y_1$	15 $\triangle^2 Y_1$	3 $\triangle^2 Y_1$	
40	582 Y_3	77 $\triangle Y_2$	18 $\triangle^2 Y_2$		
45	677 Y_4	95 $\triangle Y_3$			

$$\therefore\quad Y=394+2{\cdot}2\times 49+\frac{2{\cdot}2\times 1{\cdot}2}{1\times 2}\times 13+\frac{2{\cdot}2\times 1{\cdot}2\times{\cdot}2}{1\times 2\times 3}\times 2-\frac{2{\cdot}2\times 1{\cdot}2\times{\cdot}2\times{\cdot}8}{1\times 2\times 3\times 4}\times 1$$

$$=394+107{\cdot}8+17{\cdot}16+{\cdot}176-{\cdot}0176$$

$$=519{\cdot}118 \text{ Annas.}$$

Thus the premium at age 36 is Rs. 32 As. 7.

Problem 179.—From the following data, estimate the number of persons earning wages between 60 and 70 rupees :

Wages in rupees	No. of persons in thousands
Below 40	250
40— 60	120
60— 80	100
80—100	70
100—120	50

(*M.Com. Agra, 1951*)

Solution :

Wages in Rs. below	X	40	60	80	100	120
No. of persons in thousands	Y	250	370	470	540	590

Applying Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots\ldots$$

where $t=\frac{70-40}{20}=1{\cdot}5$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Wages in rupees X	No. of persons in thousands Y	Differences 1st	2nd	3rd	4th
Below 40	250 Y_0	120 ΔY_0	-20 $\Delta^2 Y_0$	-10 $\Delta^3 Y_0$	20 $\Delta^4 Y_0$
,, 60	370 Y_1	100 ΔY_1	-30 $\Delta^2 Y_1$	10 $\Delta^3 Y_1$	
,, 80	470 Y_2	70 ΔY_2	-20 $\Delta^2 Y_2$		
,, 100	540 Y_3	50 ΔY_3			
,, 120	590 Y_4				

$$\therefore \quad Y = 250 + 1{\cdot}5 \times 120 - \frac{1{\cdot}5 \times {\cdot}5}{1 \times 2} \times 20 + \frac{1{\cdot}5 \times {\cdot}5 \times {\cdot}5}{1 \times 2 \times 3} \times 10 + \frac{1{\cdot}5 \times {\cdot}5 \times {\cdot}5 \times 1{\cdot}5}{1 \times 2 \times 3 \times 4} \times 20$$

$$= 250 + 180 - 7{\cdot}5 + {\cdot}625 + {\cdot}46875$$

$$= 423{\cdot}6 \text{ thousands.}$$

Thus 423·6 thousand persons earn wages below 70 rupees.

$\therefore$ No. of persons earning wages between 60 and 70 rupees are $= 423{\cdot}6 - 370 = 53{\cdot}6$ thousands.

Problem 180.—The following table gives the quantities of a certain brand of tea demanded at prices noted against each. Estimate the probable demand when the price is Rs. 1—14—0 a pound.

Price of tea per pound Rs. As.	Quantity demanded in thousand pounds
1— 4	82·5
1— 8	70·8
1—12	63·1
2— 0	55·0
2— 4	48·9

(*M.A. Al'd.*, 1942 ; *Raj. M.A.*, 1956)

Solution :

Rs. 1/14/- = 30 annas.

Price of per pound in annas.	X	20	24	28	32	36
Quantity demanded in thousand pounds	Y	82·5	70·8	63·1	55·0	48·9

Applying Newtons Advancing differences method we have

$$Y = Y_0 + t(\Delta Y_0) + \frac{t(t-1)}{1 \times 2}(\Delta^2 Y_0) + \frac{t(t-1)(t-2)}{1 \times 2 \times 3}(\Delta^3 Y_0) + \ldots\ldots\ldots$$

where $\qquad t = \dfrac{30-20}{4} = 2{\cdot}5$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Price of tea per lb. in annas X	Quantity demanded in thousand lbs. Y	Differences 1st	2nd	3rd	4th
20	82·5 Y_0	−11·7 $\triangle Y_0$	4 $\triangle^2 Y_0$	−4·4 $\triangle^3 Y_0$	6·8 $\triangle^4 Y_0$
24	70·8 Y_1	−7·7 $\triangle Y_1$	−·4 $\triangle^2 Y_1$	2·4 $\triangle^3 Y_1$	
28	63·1 Y_2	−8·1 $\triangle Y_2$	2 $\triangle^2 Y_2$		
32	55·0 Y_3	−6·1 $\triangle Y_3$			
36	48·9 Y_4				

$$\therefore \quad Y = 82{\cdot}5 - 2{\cdot}5 \times 11{\cdot}7 + \frac{2{\cdot}5 \times 1{\cdot}5}{1 \times 2} \times 4 - \frac{2{\cdot}5 \times 1{\cdot}5 \times {\cdot}5}{1 \times 2 \times 3} \times 4{\cdot}4 - \frac{2{\cdot}5 \times 1{\cdot}5 \times {\cdot}5 \times {\cdot}5}{1 \times 2 \times 3 \times 4} \times 6{\cdot}8$$

$$= 82{\cdot}5 - 29{\cdot}25 + 7{\cdot}5 - 1{\cdot}375 - {\cdot}2656$$

$= 59{\cdot}11$ thousand pounds.

Thus the estimated demand of tea at a price of Rs. 1/14/- is 59·11.

Problem 181.—The length of the day was 12 hours on March 19th ; 14 hours on April 18th, and 15 hours 40 minutes on May 18th. Required an approximate value of (*a*) the length of the day on May 3rd (*b*) the mean length of the day during the period, March 19th to May 18th. (*I.A.S., 1947*)

Solution :

(*a*) Since the number of days between 19th March and 18th April are 30 and also between 18th April and 18th May number of days are 30 ; hence this is the case of equal intervals and Newtons formula for Advancing Differences is applicable.

Also 15 hours 40 minutes = 15·66 hours = 15·7 hours approx.

∴ X	March 19th	April 18th	May 18th
Y	12	14	15·7

It is required to find the duration of day on May 3rd. Formula is

$$Y = Y_0 + t(\triangle Y_0) + \frac{t(t-1)}{1 \times 2}(\triangle^2 Y_0) + \frac{t(t-1)(t-2)}{1 \times 2 \times 3}(\triangle^3 Y_0) + \ldots\ldots$$

where
$$t = \frac{\text{May 3rd} - \text{March 19th}}{\text{April 18th} - \text{March 19th}}$$

$$= \frac{45 \text{ days}}{30 \text{ days}} = 1{\cdot}5$$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Date of observation X	Length of day in hours	Differences 1st	2nd
March 19th April 18th May 18th	12 Y_0 14 Y_1 15·7 Y_2	2 ΔY_0 1·7 ΔY_1	$-·3$ $\Delta^2 Y_0$

$$\therefore \quad Y = 12 + 1·5 \times 2 - \frac{1·5 \times ·5}{1 \times 2} \times ·3$$

$$\therefore \quad Y = 12 + 3 - ·1125$$

$= 14·88$ hours appr.

$= 14$ hours 53 minutes approximately.

$\therefore$ Length of the day on May 3rd was 14 hours 53 min. approximately.

(b) Mean length of the day $= \frac{12 + 14 + 15·7}{3}$ hours

$= \frac{41·7}{3}$ hours

$= 13·9$ hours.

Hence mean length of day is 13·9 hours.

Problem 182.—The Gross profits of the Buland Sugar Co., Ltd. are given below :

Years	*Gross profits in Lakhs of Rupees*
1935—36	4·86
1937—38	12·64
1939—40	13·68
1941—42	16·65
1943—44	23·29

Make an estimate for 1942-43 and 1944-45. (*B.Com. Raj.*, *1949*)

Solution :

X	Y
1935—36	4·86
1937—38	12·64
1939—40	13·68
1941—42	16·65
1943—44	23·29

Applying the Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots$$

where $t=\dfrac{\text{year of interpolation}-\text{year of origin}}{\text{Time distance between adjoining years}}$

$\therefore$ t for 1942-43

$$=\frac{(1942-43)-(1935-36)}{(1937-38)-(1935-36)}=\frac{7}{2}=3{\cdot}5$$

t for 1944-45

$$=\frac{(1944-45)-(1935-36)}{(1937-38)-(1935-36)}=\frac{9}{2}=4{\cdot}5$$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Years X	Gross Profit in Lakhs of Rs. Y	Differences			
		1st	2nd	3rd	4th
1935-36	4·86 Y_0				
1937-38	12·64 Y_1	7·78 $\triangle Y_0$			
1939-40	13·68 Y_2	1·04 $\triangle Y_1$	−6·74 $\triangle^2 Y_0$		
1941-42	16·65 Y_3	2·97 $\triangle Y_2$	1·93 $\triangle^2 Y_1$	8·67 $\triangle^3 Y_0$	
1943-44	23·29 Y_4	6·64 $\triangle Y_3$	3·67 $\triangle^2 Y_2$	1·74 $\triangle^3 Y_1$	−6·93 $\triangle^4 Y_0$

(*i*) Estimate of the profit for the year 1942-43.

$$Y=4{\cdot}86+3{\cdot}5\times 7{\cdot}78-\frac{3{\cdot}5\times 2{\cdot}5}{1\times 2}\times 6{\cdot}74+\frac{3{\cdot}5\times 2{\cdot}5\times 1{\cdot}5}{1\times 2\times 3}\times 8{\cdot}67$$
$$-\frac{3{\cdot}5\times 2{\cdot}5\times 1{\cdot}5\times {\cdot}5}{1\times 2\times 3\times 4}\times 6{\cdot}93$$

$=4{\cdot}86+27{\cdot}23-29{\cdot}488+18{\cdot}956-1{\cdot}895$

$=19{\cdot}663$ lakhs of Rs.

(*ii*) Estimate of the profit for the year 1944-45

$$Y=4{\cdot}86+4{\cdot}5\times 7{\cdot}78-\frac{4{\cdot}5\times 3{\cdot}5}{1\times 2}\times 6{\cdot}74$$
$$+\frac{4{\cdot}5\times 3{\cdot}5\times 2{\cdot}5}{1\times 2\times 3}\times 8{\cdot}67-\frac{4{\cdot}5\times 3{\cdot}5\times 2{\cdot}5\times 1{\cdot}5}{1\times 2\times 3\times 4}\times 6{\cdot}93$$

$=4{\cdot}86+35{\cdot}01-53{\cdot}073+56{\cdot}897-17{\cdot}054$

$=26{\cdot}64$ lakhs of Rs.

Thus the gross profit in (1942-43) and (1944-45) are 19·663 lakhs of Rs. and 26·64 lakhs of Rs. respectively

Problem 183.—Estimate the probable number of incomes between Rs. 20 and Rs. 25 from the following data :

Income in Rs.	*No. of incomes*
under Rs. 10	20
Rs. 10 to Rs. 20	45
Rs. 20 to Rs. 30	115
Rs. 30 to Rs. 40	210
Rs. 40 to Rs. 50	115

(*M.Sc. Agra, 1955 ; M.A. Raj., 1955 ; P.C.S.*)

Solution :

Income in Rs. under	X	10	20	30	40	50
No. of incomes	Y	20	65	180	390	505

Applying Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots$$

where $t=\dfrac{25-10}{10}=1{\cdot}5$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Income in Rs. X	No. of incomes Y	Differences 1st	2nd	3rd	4th
under 10	20 Y_0	45 $\triangle Y_0$	70 $\triangle^2 Y_0$	25 $\triangle^3 Y_0$	-215 $\triangle^4 Y_0$
under 20	65 Y_1	115 $\triangle Y_1$	95 $\triangle^2 Y_1$	-190 $\triangle^3 Y_1$	
under 30	180 Y_2	210 $\triangle Y_2$	-95 $\triangle^2 Y_2$		
under 40	390 Y_3	115 $\triangle Y_3$			
under 50	505 Y_4				

$$\therefore\ Y=20+1{\cdot}5\times 45+\frac{1{\cdot}5\times{\cdot}5}{1\times 2}\times 70-\frac{1{\cdot}5\times{\cdot}5\times{\cdot}5}{1\times 2\times 3}\times 25-\frac{1{\cdot}5\times{\cdot}5\times{\cdot}5\times 1{\cdot}5}{1\times 2\times 3\times 4}\times 215$$

$$=20+67{\cdot}5+26{\cdot}25-1{\cdot}5625-5{\cdot}039$$

$$=117{\cdot}15=117$$

$\therefore$ No. of incomes under Rs. 25 is 117

$\therefore$ No. of incomes between Rs. 20 to Rs. 25 is

117−65=52.

Problem 184.—The following are the numbers of deaths in four successive ten year age-group. Calculate the number of deaths at 45-50 and 50-55.

Age-group	*Deaths*
25—	13229
35—	18139
45—	24225
55—	31496

(*P.C.S., 1952 ; M.Sc. Agra, 1956*)

Solution :

Exact Age X	35	45	55	65
Sum of deaths from 25 to Y age stated	13229	31368	55593	87089

Applying the Newtons Advancing Differences method we have

$$Y=Y_0+t(\triangle Y_0)+\frac{t(t-1)}{1\times 2}(\triangle^2 Y_0)+\frac{t(t-1)(t-2)}{1\times 2\times 3}(\triangle^3 Y_0)+\ldots\ldots$$

where $t=\frac{50-35}{10}=\frac{15}{10}=1{\cdot}5$

For $\triangle Y_0$, $\triangle^2 Y_0$ etc. we have the Difference table as below :

Exact Age X	Sum of Deaths from 25 to Age stated Y	Differences 1st	2nd	3rd
35	13229 Y_0			
45	31368 Y_1	18139 $\triangle Y_0$	6086 $\triangle^2 Y_0$	1185 $\triangle^3 Y_0$
55	55593 Y_2	24225 $\triangle Y_1$	7271 $\triangle^2 Y_1$	
65	87089 Y_3	31496 $\triangle Y_2$		

$$\therefore \quad Y=13229+1{\cdot}5\times 18139+\frac{1{\cdot}5\times{\cdot}5}{1\times 2}\times 6086-\frac{1{\cdot}5\times{\cdot}5\times{\cdot}5}{1\times 2\times 3}\times 1185$$

$$=42645{\cdot}7=42646 \text{ appr.}$$

Thus number of deaths from 25 to 50 are 42646, whereas number of Deaths from 25 to 45 and from 25 to 55 are respectively 31368 and 55593.

$\therefore$ No. of Deaths between 45—50 are $=42646-31368=11278$

and No. of Deaths between 50—55 are $=55593-42646=12947$

Problem 185.—Given $\log_{10}654=2{\cdot}8156$; $\log_{10}658=2{\cdot}8182$
$\log_{10}659=2{\cdot}8189$; $\log_{10}661=2{\cdot}8202$.

Find $\log_{10}656$ using two different interpolation formulae available for observations at unequal intervals, say, Lagrange's formula and the formula for divided differences.

(I.A.S., 1956)

Solution :

(*i*) **Lagrange's formula** :—

	a	*b*	*c*	*d*
X	654	658	659	661
	Y_1	Y_2	Y_3	Y_4
Y	2·8156	2·8182	2·8189	2·8202

Lagrange's formula is

$$Y_x=\frac{(x-b)(x-c)(x-d)}{(a-b)(a-c)(a-d)}Y_1+\frac{(x-a)(x-c)(x-a)}{(b-a)(b-c)(b-d)}Y_2$$
$$+\frac{(x-a)(x-b)(x-d)}{(c-a)(c-b)(c-d)}Y_3+\frac{(x-a)(x-b)(x-c)}{(d-a)(d-b)(d-c)}Y_4$$

$$Y_{656}=\frac{(656-658)(656-659)(656-661)}{(654-658)(654-659)(654-661)}\times 2{\cdot}8156$$
$$+\frac{(656-654)(656-659)(656-661)}{(658-654)(658-659)(658-661)}\times 2{\cdot}8182$$
$$+\frac{(656-654)(656-658)(656-661)}{(659-654)(659-658)(659-661)}\times 2{\cdot}8189$$
$$+\frac{(656-654)(656-658)(656-659)}{(661-654)(661-658)(661-659)}\times 2{\cdot}8202$$

$$=\frac{(-2)(-3)(-5)}{(-4)(-5)(-7)}\times 2{\cdot}8156$$
$$+\frac{(2)(-3)(-5)}{(4)(-1)(-3)}\times 2{\cdot}8182$$
$$+\frac{(2)(-2)(-5)}{(5)(1)(-2)}\times 2{\cdot}8189$$
$$+\frac{(2)(-2)(-3)}{(7)(3)(2)}\times 2{\cdot}8202$$

or $$Y_{656}=\frac{3\times 2{\cdot}8156}{14}+\frac{5\times 2{\cdot}8182}{2}-2\times 2{\cdot}8189+\frac{2\times 2{\cdot}8202}{7}$$
$$=\cdot 6033+7{\cdot}0455-5{\cdot}6378+\cdot 8058$$
$$=2{\cdot}8168$$

$\therefore\quad \log_{10}656=2{\cdot}8168.$

(*ii*) **By the formula for divided differences** :—

X	654	658	659	661
Y	2·8156	2·8182	2·8189	2·8202

The difference table is as given below :

X	Y	Divided Differences		
		1st	2nd	3rd
654	2·8156			
		$\Delta_1=\frac{2\cdot8182-2\cdot8156}{658-654}=\cdot00065$		
			$\Delta^2_1=\frac{\cdot0007-\cdot00065}{659-654}=\cdot00001$	
658	2·8182			
		$\Delta_2=\frac{2\cdot8189-2\cdot8182}{659-658}=\cdot0007$		$\Delta^3_1=\frac{-\cdot000027}{661-654}=-\cdot0000038$
			$\Delta^2_2=\frac{\cdot00065-\cdot0007}{661-658}=-\cdot000017$	
659	2·8189			
		$\Delta_3=\frac{2\cdot8202-2\cdot8189}{661-659}=\cdot00065$		
661	2·8202			

Hence $\Delta_1=\cdot00065$, $\Delta^2_1=\cdot00001$, $\Delta^3_1=-\cdot0000038$

Formula for divided differences is

$$Y_{656}=2\cdot8156+(656-654)\times\cdot00065$$
$$+(656-654)(656-658)\times\cdot00001$$
$$+(656-654)(656-658)(656-659)\times(-\cdot0000038)$$
$$=2\cdot8156+2\times\cdot00065+(2)(-2)(\cdot00001)+(2)(-2)(-3)(-\cdot0000038)$$
$$=2\cdot8156+\cdot0013-\cdot00004-\cdot0000456$$
$$=2\cdot8169-\cdot0000856$$
$$=2\cdot8168144$$

Hence $\log_{10}656=2\cdot8168144$.

Problem 186.—By means of divided differences, find the value of Y_{19} from the following table :

X	11	17	21	23	31
Y_x	14646	83526	194426	279846	923526

Solution :

X	11	17	21	23	31
Y	14646	83526	194486	279846	923526.

The difference table is as given below :

X	Y	Divided Differences 1st	2nd	3rd	4th
11	14646				
		$\Delta_1=\frac{68880}{6}=11480$			
17	83526		$\Delta^2_1=\frac{16260}{10}=1626$		
		$\Delta_2=\frac{110960}{4}=27740$		$\Delta^3_1=\frac{864}{12}=72$	
21	194486		$\Delta^2_2=\frac{14940}{6}=2490$		$\Delta^4_1=\frac{20}{20}=1$
		$\Delta_3=\frac{85360}{2}=42680$		$\Delta^3_2=\frac{1288}{14}=92$	
23	279846		$\Delta^2_3=\frac{37780}{10}=3778$		
		$\Delta_4=\frac{643680}{8}=80460$			
31	923526				

$\therefore \quad \Delta_1=11480,\ \Delta^2_1=1626,\ \Delta^3_1=72,\ \Delta^4_1=1$

$$\therefore \quad Y_{19}=14646+(19-11)\times 11480+(19-11)(19-17)\times 1626+(19-11)(19-17)(19-21)\times 72+(19-11)(19-17)(19-21)(19-23)\times 1$$

$$=14646+8\times 11480+16\times 1626+(-32)\times 72+128$$
$$=14646+91840+26016-2304+128$$
$$=130326.$$

Problem 187.—Given

X	40	42	44	45
Y	43833	46568	49431	50912

Use divided differences to find Y for X=43.

Solution :

X	40	42	44	45
Y	43833	465.8	49431	50912

The divided differences are obtained from the table given below:

X	Y	Divided Differences 1st	2nd	3rd
40	43833			
		$\Delta_1=\frac{2735}{2}=1367{\cdot}5$		
42	46568		$\Delta^2_1=\frac{64}{4}=16$	
		$\Delta_2=\frac{2863}{2}=1431{\cdot}5$		$\Delta^3_1=\frac{{\cdot}5}{5}={\cdot}1$
44	49431		$\Delta^2_2=\frac{49{\cdot}5}{3}=16{\cdot}5$	
		$\Delta_3=\frac{1481}{1}=1481$		
45	50912			

$\therefore \quad \Delta_1 = 1367{\cdot}5, \ \Delta^2_1 = 16, \ \Delta^3_1 = {\cdot}1$

By the formula for divided differences

$$Y_{43} = 43833 + (43-40) \times 1367{\cdot}5 +)(43-40)(43-42) \times 16$$
$$+ (43-40)(43-42)(43-44) \times {\cdot}1$$
$$= 43833 + 4102{\cdot}5 + 48 - {\cdot}3.$$
$$= 47983{\cdot}2.$$

Problem 188.—Given

$$Y_{-30} = 30 \ ; \ Y_{-13} = 34 \ ; \ Y_3 = 38 \ ; \ Y_{18} = 42$$

Find Y_0

Solution :

X	−30	−13	3	18
Y	30	34	38	42

The difference table is as given below :

X	Y	Divided Differences 1st	2nd	3rd
−30	30			
		$\Delta_1 = \frac{4}{-13-(-30)} = {\cdot}235$		
−13	34		$\Delta^2_1 = \frac{{\cdot}015}{33} = {\cdot}00046$	
		$\Delta_2 = \frac{4}{16} = {\cdot}25$		$\Delta^3_1 = \frac{{\cdot}00005}{48} = {\cdot}000001$
3	38		$\Delta^2_2 = \frac{{\cdot}016}{31} = {\cdot}00051$	
		$\Delta_3 = \frac{4}{15} = {\cdot}266$		
18	42			

$\Delta_1 = {\cdot}235, \ \Delta^2_1 = {\cdot}00046, \ \Delta^3_1 = {\cdot}000001$

$$Y_0 = 30 + (0+30) \times {\cdot}235 + 30 \times 13 \times {\cdot}00046 - 30 \times 13 \times 3 \times {\cdot}000001$$
$$= 30 + 7{\cdot}05 + {\cdot}1794 - {\cdot}00117$$
$$= 37{\cdot}2294 - {\cdot}00117$$
$$= 37{\cdot}22823.$$

CHAPTER VI

CORRELATION

According to Profs. Croxton and Cowden, "Correlation, also called Co-variation, is the casual relationship existing between any two variables depicting separate characters, the connection being that of direct cause and effect or of mutually reactive causes or the like, the Co-efficient of Correlation being its numerical measurements".

In the words of W. I. King, "Correlation means that between two series or groups of data there exits some casual connection."

When two series or groups of data are compared, the one that is taken as the standard of measure is called 'Subject', and the other which is compared is known as 'Relative'.

Croxton and Cowden has mentioned in their book 'Applied General Statistics' that the theory of correlation involves three kinds of measurements :

(*a*) An estimating equation which describes the functional relationship between the two variables.

(*b*) A measure of the amount of variation of the actual values of the dependent variable from their estimated or computed values. It is called the 'scatter', or **standard error** of estimate and is calculated as :

$$\sigma y_s = \sqrt{\frac{\Sigma(y-y_e)^2}{N}}$$

$y-y_e$ = deviation from the estimated value.

N = Total number of observations.

σy_s = Standard error of estimate.

(*c*) A measure of the degree of relationship, or correlation (r) between the variables, independent of the units or terms in which they were originally expressed.

Some of the common formulas that are used to calculate Correlation are as follows :

(1) $$r = \frac{\Sigma xy}{n.\sigma_x.\sigma y}$$

r = Co-efficient of Correlation.

n = number of observations.

σx = Standard Deviation of X series.

σ_y = Standard Deviation of Y series.

[If Σxy is positive r=positive ; if Σxy is negative, r=negative]

(2) $r = \dfrac{\Sigma xy}{\sqrt{\Sigma x^2 . \Sigma y^2}}$

(3) $r = \dfrac{\Sigma \xi\eta - nC_xC_y}{n\,\sigma_x.\sigma_y}$

A=assumed mean of X series.
B=assumed mean of Y series.
ξ=deviation of X from A
η=deviation of Y from B
C_x=true mean—assumed mean of X series. $[M_x - A]$
C_y= ,, ,, — ,, ,, ,, Y ,, $[M_y - B]$

(4) $r = \dfrac{\Sigma xy - n\,(a_1 - x)(a_2 - y)}{n\,\sigma_x.\sigma_y}$

a_1=true average of X series.
a_2=true average of Y series.
x=assumed mean of X series.
y=assumed mean of Y series.

(5) $r = \dfrac{\Sigma xy - n\left[\dfrac{\Sigma d_x}{n}\right]\left[\dfrac{\Sigma d_y}{n}\right]}{\sqrt{\Sigma d_x^2 - \left[\dfrac{\Sigma d_x}{n}\right]}\sqrt{\Sigma d_y^2 - \left[\dfrac{\Sigma d_y^2}{n}\right]}}$

d_x=deviations of X series.
d_y=deviations of Y series.

The above formula (5) is also written as :

(6) $r = \dfrac{\Sigma \xi\eta - \dfrac{\Sigma\xi.\Sigma\eta}{n}}{\sqrt{\left[\Sigma\xi^2 - \dfrac{(\Sigma\xi)^2}{n}\right]\left[\Sigma\eta^2 - \dfrac{(\Sigma\eta)^2}{n}\right]}}$

where ξ=deviations of X series.
η=deviations of Y series.

(7) Correlation from table ;

$$r = \frac{\xi f\,\xi\eta - \dfrac{\xi f\,\xi . \Sigma f\,\eta}{n}}{\sqrt{\left[\Sigma f\,\xi^2 - \dfrac{(\Sigma f\,\xi)^2}{n}\right]\left[\Sigma\,f\,\eta^2 - \dfrac{(\Sigma f\,\eta)^2}{n}\right]}}$$

(8) $r = \dfrac{\Sigma f\,uv - \dfrac{\Sigma f\,u\,\Sigma f\,v}{\Sigma f}}{\sqrt{\Sigma f\,u^2 - \dfrac{(\Sigma fu^2)}{\Sigma f}}\sqrt{\Sigma fv^2 - \dfrac{(\Sigma f\,v)^2}{\Sigma f}}}$

(using step-deviation method)

$u = X -$ assumed mean

$v = Y -$ assumed mean

$\Sigma f = n$

(9) Rank Correlation

$$\rho = 1 - \frac{6\Sigma d^2}{n^3 - n}$$

$\rho =$ rank correlation

$d =$ rank differences

$n =$ number of Observations.

[The highest value is ranked 1, second highest is to be ranked 2 and so on for both the X and Y series.]

(10) Co-efficient of Concurrent deviations.

$$r_{cd} = \pm \sqrt{\pm \left[\frac{2c - n}{n} \right]}$$

$r_{cd} =$ Co-efficient of concurrent deviations.

$c =$ the number of concurrent deviations.

$n =$ total number of observations.

In this calculation, the differences are indicated only by + or − signs. Then the total number of concurrent deviations for the X and Y series is counted (c) and then the coefficient of concurrent deviation is calculated.

Note :—(*i*) It should be remembered that the value of r must fall within the range of 0 to ± 1. *If the value crosses 1, then the calculation is wrong.*

(*ii*) Correlation is also known by graph. If the graphs of the subject and the relative move in sympathy, there is + correlation, if they move in opposite directions, then negative (−) correlation, if one curve cuts the other, then there is no correlation at all.

(*iii*) For finding out the value of Σxy for the tabular data (Formula 7 and 8) the following points may be taken into cosideration :

(*a*) Mid-values of X and Y series can be known to make the calculations easier.

(*b*) Products of ξ and η should be placed on one corner of the value table, [400 / 6 / 2400→] say 400 in this example. Then

400	
	6
	2400→

multiply 400 with 6 which gives the value = 2400.

Then count the figures of the extreme corner (2400) for the whole table and place it in the column $x.y$ or $f\xi.f\eta$.

Sum it up, which will give the value of Σxy or $\Sigma f\xi\eta$ or $\Sigma f uv$.*

(iv) Except the formula number 9 and 10, which are meant for special cases, any of the other formulas may be used for calculating co-efficient of correlation.

Problem 189.—The following table gives the index numbers of the whloesale prices in India and the Bombay cost of living index numbers. Calculate the coefficient of correlation :

	Index Numbers of wholesale prices	*Bombay Cost of living index numbers*
April 1953	385	342
May 1953	398	346
June 1953	406	353
July 1953	408	355
August 1953	410	358
Sept. 1953	404	353
Oct. 1953	394	350
Nov. 1953	390	344
Dec. 1953	390	342
Jan. 1954	399	346
Feb. 1954	395	331
March 1954	394	332

(*M.A., Agra, 1955*)

Solution :

X	Y	ξ (X−404)	η (Y−350)	ξ^2	η^2	$\xi\eta$
385	342	−19	−8	361	64	152
398	346	−6	−4	36	16	24
406	353	+2	+3	4	9	6
408	355	+4	+5	16	25	20
410	358	+6	+8	36	64	48
404	353	0	+3	0	9	0
394	350	−10	0	100	0	0
390	344	−14	−6	196	36	84
390	342	−14	−8	196	64	112
399	346	−5	−4	25	16	20
395	331	−9	−19	81	361	171
394	332	−10	−18	100	324	180
		−75	−48	1151	988	817

Let the assumed mean in X and Y series be 404 and 350.

ξ = X − assumed mean (404)

η = Y − assumed mean (350)

Co efficient of correlation $= r$

$$r=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\Sigma\eta}{n}}{\sqrt{\left[\Sigma\xi^2-\frac{(\Sigma\xi)^2}{n}\right]\left[\Sigma\eta^2-\frac{(\Sigma\eta)^2}{n}\right]}}$$

* This method would be helpful for those who have not sufficient mathematical background.

$$= \frac{817 - \frac{(-75)(-48)}{12}}{\sqrt{\left[1151 - \frac{(-75)^2}{12}\right]\left[988 - \frac{(48)^2}{12}\right]}}$$

$$= \frac{817 - 300}{\sqrt{[1151 - 469][988 - 192]}}$$

$$= \frac{517}{\sqrt{682 \times 796}}$$

$$= \frac{517}{26 \cdot 1 \times 28 \cdot 2}$$

$$= \frac{517}{736 \cdot 02}$$

$$= + \cdot 70$$

This shows a high degree of positive correlation in the changes in wholesale prices and cost of living index numbers.

Problem 190.—The following table gives the index numbers of wholesale prices of rice and wheat for the year 1955-56. Calculate the Co-efficient of correlation between the index numbers of wholesale prices of rice and wheat.

Month		*Index Numbers of wholesale Prices*	
		Rice	*Wheat*
April	1955	410	400
May	1955	405	350
June	1955	410	365
July	1955	455	415
Aug.	1955	490	420
Sept.	1955	510	410
Oct.	1955	490	430
Nov.	1955	475	470
Dec.	1955	465	505
Jan.	1956	450	530
Feb.	1956	470	525
March	1956	505	545

(*M. A., Agra, 1957*)

Solution :

X	Y	d_x (X−461)	d_x^2	d_y	d_y^2	$d_x d_y$ (xy)
410	400	−51	2601	−47	2209	+2397
405	350	−56	3136	−97	9409	+5432
410	365	−51	2601	−82	6724	+4182
455	415	− 6	36	−32	1024	+ 192
490	420	+29	841	−27	729	− 783
510	410	+49	2401	−37	1369	−1813
490	430	+29	841	−17	289	− 493
475	470	+14	196	+23	529	+ 322
465	505	+ 4	16	+58	3364	+ 232
450	530	−11	121	+83	6889	− 913
470	525	+ 9	81	+78	6084	+ 702
505	545	+44	1936	+98	9604	+4312
$\Sigma X =$ 5535	$\Sigma Y =$ 5365		$\Sigma d_x^2 =$ 14807		$\Sigma d_y^2 =$ 48223	$\Sigma xy =$ +13769

Arithmetic Average for X series $= \frac{\Sigma X}{n}$

$= \frac{5535}{12}$

$= 461 \cdot 25$

$= 461$ approximately.

Arithmetic Average for Y series $= \frac{\Sigma Y}{n}$

$= \frac{5365}{12}$

$= 447$ approximately

Standard Deviation of X series (σ_x) $= \sqrt{\frac{\Sigma dx^2}{n}}$

$= \sqrt{\frac{14807}{12}}$

$= \sqrt{1234}$ approximately.

$= 35$ approximately.

Standard Deviation of Y series (σ_y) $= \sqrt{\frac{\Sigma d_y^2}{n}}$

$= \sqrt{\frac{48223}{12}}$

$= \sqrt{4019}$ approximately.

$= 63$ approximately.

Co-efficient of Correlation (r) $= \frac{\Sigma xy}{n.\sigma x.\sigma_y}$

$= \frac{+13769}{12 \times 35 \times 63}$ approx.

$= \frac{+13769}{26460}$ approximately.

$= + \cdot 52$ approximately.

Thus, we find that there is a positive correlation of a moderate high degree between the index numbers of wholesale prices of rice and wheat.

√**Problem 191.**—Calculate the Co-efficient of correlation between the ages of husbands and wives according to the data given :

Ages of Husbands	*Ages of Wives*	*Ages of Husbands*	*Ages of Wives*
33	18	30	29
27	20	31	27
28	22	33	29
28	27	35	28
29	21	36	29

(*M A. Agra, 1950*)

Solution :

Subject (X) ages of husbands			Relative (Y) ages of wives			Product of deviations of husbands' & wives' ages
Ages of husbands in years	Deviation from average (29 years)	Square of deviation	Ages of wife in years	Deviations from average (21 years)	Square of deviations	
m_1	x	x^2	m_2	y	y^2	(xy)
33	+4	16	18	−3	9	−12
27	−2	4	20	−1	1	+2
28	−1	1	22	+1	1	−1
28	−1	1	27	+6	36	−6
29	0	0	21	0	0	0
30	+1	1	29	+8	64	+8
31	+2	4	27	+6	36	+12
33	+4	16	29	+8	64	+32
35	+6	36	28	+7	49	+42
36	+7	49	29	+8	64	+56
$\Sigma m_1=310$ $n=10$	$a_1=29$ years	$\Sigma x^2=128$	$\Sigma m_2=250$	$a_2=21$ yrs	$\Sigma y^2=324$	$\Sigma xy= +133$

Standard Deviation of X series $(\sigma_1)=\sqrt{\frac{\Sigma x^2}{n}}$

$$=\sqrt{\frac{128}{10}}$$

$$=\sqrt{12\cdot8}$$

$=3\cdot5$ years approximately.

Standard Deviation of Y series $(\sigma_2)=\sqrt{\frac{\Sigma y^2}{n}}$

$$=\sqrt{\frac{324}{10}}$$

$$=\sqrt{32\cdot4}$$

$=5\cdot6$ years approximately

Co-efficient of correlation (r) $= \dfrac{\Sigma xy}{n \cdot \sigma_1 \cdot \sigma_2}$

$= \dfrac{133}{10 \times 3{\cdot}5 \times 5{\cdot}6}$

$= \dfrac{133}{196}$

$= +{\cdot}6$ approximately.

$+{\cdot}6$ indicates a moderately high positive degree of correlation in the ages of husbands and wives.

Problem 192.—Ten students got the following percentage of marks in Principles of Economics and Statistics :

Student :	1	2	3	4	5	6	7	8	9	10
Marks in Economics :	78	36	98	25	75	82	90	62	65	39
Marks in Statistics :	84	51	91	60	68	62	86	58	53	47

Calculate the co-efficient of correlation.

(*M.A. Agra, 1951*)

Solution :

Let the two subjects be denoted by X and Y. Then putting them in a tabular form, we have—

X	Y	ξ	η	$\xi\eta$	ξ^2	η^2
78	84	4	−16	−64	16	256
36	51	46	17	782	2116	289
98	91	−16	−23	368	256	529
25	60	57	8	456	3249	64
75	68	7	0	0	49	0
82	62	0	6	0	0	36
90	86	−8	−18	144	64	324
62	58	20	10	200	400	100
65	53	17	15	255	289	225
39	47	43	21	903	1849	441
Total (n)		+170	+20	+3044	8288	2264

Applying the method of deviation Correlation, let

$\xi = 82 - X$

$\eta = 68 - Y$

Hence, applying the short-cut method of the calculation of correlation—

$$r = \frac{\Sigma \xi \eta - \dfrac{\Sigma \xi \, \Sigma \eta}{n}}{\sqrt{\left[\Sigma\xi^2 - \dfrac{(\Sigma\xi)^2}{n} \right] \left[\Sigma\eta^2 - \dfrac{(\Sigma\eta)^2}{n} \right]}}$$

$$= \frac{3044 - \dfrac{170 \times 20}{10}}{\sqrt{\left[8288 - \dfrac{(170)^2}{10}\right]\left[2264 - \dfrac{(20)^2}{10}\right]}}$$

$$= \frac{3044 - 340}{\sqrt{\left[8288 - \dfrac{28900}{10}\right]\left[2264 - \dfrac{400}{10}\right]}}$$

$$= \frac{2704}{\sqrt{[5398][2224]}}$$

$$= \frac{2704}{37{\cdot}4 \times 47{\cdot}1}$$

$$= \frac{2704}{3457{\cdot}14}$$

$$= \frac{2704}{3457} \text{ approximately.}$$

$= {\cdot}78$ approx. **Ans.**

✓**Problem 193.**—Calculate the co-efficient of correlation for the following ages of husband and wife :

Husband's age	23	27	28	29	30	31	33	35	36	39
Wife's age	18	22	23	24	25	26	28	29	30	32

(M.A. Agra, 1952)

Solution :

S. No. of observations	Ages of Husbands (X series)			Ages of Wives (Y series)			Product of deviations of X and Y series
	Age of Husbands	Deviations from assumed mean ($x_1=30$)	Square of the deviations	Ages of wives	Deviation from assumed average ($x_2=24$)	Square of the deviations	
(No.)	(years)	d_x	dx^2	(years)	d_y	d_y^2	xy
1	23	−7	49	18	−6	36	+42
2	27	−3	9	22	−2	4	+ 6
3	28	−2	4	23	−1	1	+ 2
4	29	−1	1	24	0	0	0
5	30	0	0	25	+1	1	0
6	31	+1	1	26	+2	4	+ 2
7	33	+3	9	28	+4	16	+12
8	35	+5	25	29	+5	25	+25
9	36	+6	36	30	+6	36	+36
10	39	+9	81	32	+8	64	+72
$n=10$	311		$\Sigma dx^2=215$	257		$\Sigma dy^2=187$	$\Sigma xy=+197$

Arithmetic Average of X series $= \frac{311}{10}$

$= 31{\cdot}1\ (a_1)$

and assumed average of X series $= 30 (x_1)$

Arithmetic average of Y series $= \frac{257}{10}$

$= 25{\cdot}7\ (a_2)$

and assumed average of Y series $= 24\ (x_2)$

Standard Deviation of X series $= (\sigma_x) \sqrt{\frac{\Sigma d_x^2}{n}}$

$= \sqrt{\frac{215}{10}}$

$= \sqrt{21{\cdot}5}$

$= 4{\cdot}5$ approximately.

Standard Deviation of Y series $= (\sigma_y) = \sqrt{\frac{\Sigma d_y^2}{n}}$

$= \sqrt{\frac{187}{10}}$

$= \sqrt{18{\cdot}7}$

$= 4{\cdot}3$ approximately.

Co-efficient of correlation $(r) = \frac{\Sigma xy - n(a_1 - x_1)(a_2 - x_2)}{n \cdot \sigma x \cdot \sigma y}$

$= \frac{197 - 10(31{\cdot}1 - 30)(25{\cdot}7 - 24)}{10 \times 4{\cdot}5 \times 4{\cdot}3}$

$= \frac{197 - 10(1{\cdot}1)(1{\cdot}7)}{10 \times 4{\cdot}5 \times 4{\cdot}3}$

$= \frac{197 - 12{\cdot}7}{193{\cdot}5}$

$= \frac{184{\cdot}3}{193{\cdot}5}$

$= +{\cdot}95$ approximately.

This indicates a very high degree of positive correlation between the ages of husbands and wives.

✓**Problem 194.**—$\sigma_1 = 4{\cdot}5$ and $\sigma_2 = 3{\cdot}6$ are the standard deviations of two groups $x_1, x_2, x_3 \ldots\ldots x_n$ and $y_1, y_2, y_3 \ldots\ldots y_n$ and $\Sigma xy =$

4800, $n=1000$. Calculate the co-efficient of correlation between the above two groups and also give the probable error of the co-efficient.

(Statistics—Ghosh and Choudhry. p. 431)

Solution :

$$\sigma_1=4{\cdot}5$$
$$\sigma_2=3{\cdot}6$$
$$\Sigma xy=4800$$
$$n=1000$$

$\therefore$ Coefficient of Correlation $(r)=\dfrac{\Sigma xy}{n\,\sigma_1\,\sigma_2}$

$$=\frac{4800}{1000\times4{\cdot}5\times3{\cdot}6}$$

$$=\frac{4800}{16200}$$

$$=+{\cdot}296 \text{ approximately.}$$

Probable Error $={\cdot}6745\dfrac{1-r^2}{\sqrt{n}}$

$$={\cdot}6745\frac{1-({\cdot}296)^2}{\sqrt{1000}}$$

$$={\cdot}6745\times\frac{{\cdot}912384}{10}$$

$$=\frac{{\cdot}6154030080}{10}$$

$={\cdot}0615403008$ approximately. **Ans.**

Problem 195.—Calculate r from the following table, and indicate its probable error :

	Net area sown in lakhs of acres	*No. of ploughs in lakhs*
U.P.	359	52
Madras	310	44
Bombay	285	12
Punjab	275	24
B. & O.	257	35
C.P.	245	16
Bengal	240	46
Assam	64	11
Sind	48	3
N.W.F.P.	23	2
Average for India	211	25

(P.C.S., 1943)

Solution :

Provinces	Net area sown in lakhs of acres			No. of ploughs in lakhs			Product
	X	$\xi=X-211$	ξ^2	Y	$\eta=Y-25$	η^2	$\xi\eta$
U.P.	359	148	21904	52	27	729	3996
Madras	310	99	9801	44	19	361	1881
Bombay	285	74	5476	12	−13	169	−962
Punjab	275	64	4096	24	−1	1	−64
B. & O.	257	46	2116	35	10	100	460
C.P.	245	34	1156	16	−9	81	−306
Bengal	240	29	841	46	21	441	609
Assam	64	−147	21609	11	−14	196	2058
Sind	48	−163	26569	3	−22	484	3586
N.W.F.P.	23	−188	35344	2	−23	529	4324
$n=10$		$\Sigma\xi=-4$	$\Sigma\xi^2=128912$		$\Sigma\eta=-5$	$\Sigma\eta^2=3091$	$\Sigma\xi\eta=15582$

$\therefore \quad \Sigma\xi=-4,\ \Sigma\xi^2=128912,\ \Sigma\eta=-5,\ \Sigma\eta^2=3091$

$\Sigma\xi\eta=15582 \qquad n=10$

$$r=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi\,.\,\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{15582-\dfrac{-4\times-5}{10}}{\sqrt{128912-\dfrac{4^2}{10}}\sqrt{3091-\dfrac{5^2}{10}}}$$

$$=\frac{15580}{\sqrt{128910{\cdot}4}\sqrt{3088{\cdot}5}}$$

$$=\frac{15580}{\sqrt{398139770{\cdot}4}}$$

$$=\frac{15580}{19953{\cdot}4}$$

$$={\cdot}78$$

Probable error (P. E.) $={\cdot}6745\dfrac{1-r^2}{\sqrt{n}}$

$$={\cdot}6745\ \frac{1-({\cdot}78)^2}{\sqrt{10}}$$

$$={\cdot}6745\ \frac{1-({\cdot}6084)}{3{\cdot}162}$$

$$={\cdot}0835$$

$\therefore \quad r={\cdot}78$ and P.E. $={\cdot}0835$.

Problem 196.—The following table gives the results of the Matriculation Examination held in 1936 :

Age of Candidates	13—	14—	15—	16—	17—	18—	19—	20—	21—22
Percentage of Failures	39·2	40·6	43·4	34·2	36·6	39·2	48·9	47·1	54·5

Calculate the coefficient of correlation and estimate its probable error. From your results can you definitely assert that failure is correlated with age ? *(P.C.S., 1940)*

Solution :

Age of Candidates				Percentage of Failures			Product
Class-interval	Mid-value X	$\xi = X - 17{\cdot}5$	ξ^2	Y	$\eta = Y - 40{\cdot}6$	η^2	$\xi\eta$
13—14	13·5	−4	16	39·2	−1·4	1·96	+5·6
14—15	14·5	−3	9	40·6	0	0	0
15—16	15·5	−2	4	43·4	2·8	7·84	−5·6
16—17	16·5	−1	1	34·2	−6·4	40·96	6·4
17—18	17·5	0	0	36·6	−4·0	16·00	0
18—19	18·5	1	1	39·2	−1·4	1·96	−1·4
19—20	19·5	2	4	48·9	8·3	68·89	16·6
20—21	20·5	3	9	47·1	6·5	42·25	19·5
21—22	21·5	4	16	54·5	13·9	193·21	55·6
TOTAL	157·5	0	60	383·7	18·3	373·07	96·7

$$\Sigma\xi = 0, \quad \Sigma\xi^2 = 60$$

$$\Sigma\eta = 18{\cdot}3, \quad \Sigma\eta^2 = 373{\cdot}07$$

$$n = 9, \quad \Sigma\xi\eta = 96{\cdot}7$$

$$r = \frac{\Sigma\xi\eta - \dfrac{\Sigma\xi \cdot \Sigma\eta}{n}}{\sqrt{\Sigma\xi^2 - \dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2 - \dfrac{(\Sigma\eta)^2}{n}}}$$

$$= \frac{96{\cdot}7 - \dfrac{0 \times 18{\cdot}3}{9}}{\sqrt{60 - \dfrac{0}{9}}\sqrt{373{\cdot}07 - \dfrac{(18{\cdot}3)^2}{9}}}$$

$$= \frac{96{\cdot}7}{\sqrt{60}\sqrt{335{\cdot}86}}$$

$$= \frac{96{\cdot}7}{\sqrt{20151{\cdot}6}}$$

$$= \frac{96{\cdot}7}{141{\cdot}9}$$

$$= {\cdot}681$$

$$\text{Probable error (P.E.)} = \cdot 6745 \frac{1-r^2}{\sqrt{n}}$$

$$= \cdot 6745 \frac{1-(\cdot 68)^2}{\sqrt{9}}$$

$$= \cdot 6745 \frac{1-\cdot 4625}{3}$$

$$= \cdot 6745 \times \cdot 1792$$

$$= \cdot 1208$$

$$6 \text{ P.E.} = \cdot 7296$$

here we see that $r < \cdot 7296$ hence it cannot be said that failure is correlated with age.

Problem 197.—Psychological tests of intelligence and of arithmetical ability were applied to 10 children. Here is a record of ungrouped data showing intelligence and arithmetic ratios. Calculate r.

Child	A	B	C	D	E	F	G	H	I	J
I.R.	105	104	102	101	100	99	98	96	93	92
A.R.	101	103	100	98	95	96	104	92	97	94

(*M.Sc. Agra, 1950*)
(*M.Sc. Agra, 1954*)

Solution :

Child	Intelligence ratio			Arithmetic ratio			Product
	X	$x = X - M_x = X - 99$	x^2	Y	$y = Y - M_y = Y - 98$	y^2	xy
A	105	6	36	101	3	9	18
B	104	5	25	103	5	25	25
C	102	3	9	100	2	4	6
D	101	2	4	98	0	0	0
E	100	1	1	95	−3	9	−3
F	99	0	0	96	−2	4	0
G	98	−1	1	104	6	36	−6
H	96	−3	9	92	−6	36	18
I	93	−6	36	97	−1	1	6
J	92	−7	49	94	−5	25	35
$n = 10$	990	0	170	980	0	149	99

$$M_x = \frac{990}{10} = 99. \quad M_y = \frac{980}{10} = 98$$

$$\Sigma x^2 = 170. \ \Sigma y^2 = 149. \ \Sigma xy = 99, \ n = 10$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2} \ \sqrt{\Sigma y^2}}$$

$$= \frac{99}{\sqrt{170} \ \sqrt{149}}$$

$$= \frac{99}{159 \cdot 1}$$

$$= \cdot 622.$$

Problem 198.—From the following data calculate the coefficient of correlation and find its probable error :

Mean annual Birth Rate per 1000 of population	35·3	33·5	31·4	30·5	29·3	28·2	26·3	23·6	20·1	19·9	16·7
Mean annual Death Rate per 1000 of population	20·8	19·4	18·9	18·7	17·7	16·0	14·7	14·3	14·4	12·2	12·1

(*M A , Raj, 1955*)

Solution :

Mean annual Birth Rate per 1000 of population			Mean annual Death Rate per 100 of population			Product $\xi\eta$
X	ξ=X−26·8	ξ^2	Y	η=Y—16·3	η^2	
35·3	8·5	72·25	20·8	4·5	20·25	38·25
33·5	6·7	44·89	19·4	3·1	9·61	20·77
31·4	4·6	21·16	18·9	2·6	6·76	11·96
30·5	3·7	13·69	18·7	2·4	5·76	8·88
29·3	2·5	6·25	17·7	1·4	1·96	3·50
28·2	1·4	1·96	16·0	— ·3	·09	·42
26·3	— ·5	·25	14·7	—1·6	2·56	·80
23·6	—3·2	10·24	14·3	—2·0	4·00	6·40
20·1	—6·7	44·89	14·4	—1·9	3·61	12·73
19·9	—6·9	47·61	12·2	—4·1	16·81	28·29
16·7	—10·1	102·01	12·1	—4·2	17·64	42·42
294·8	0	365·20	179·2	— ·1	89·05	174·42

$M_x = 26 \cdot 8$, $M_y = 16 \cdot 29$ (nearly).

$n = 11$, $\Sigma\xi = 0$, $\Sigma\xi^2 = 365 \cdot 2$. $\Sigma\eta = - \cdot 1$, $\Sigma\eta^2 = 89 \cdot 05$, $\Sigma\xi\eta = 174 \cdot 42$

$$r = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \cdot \Sigma\eta}{n}}{\sqrt{\Sigma\xi^2 - \frac{(\Sigma\xi)^2}{n}} \ \sqrt{\Sigma\eta^2 - \frac{(\Sigma\eta)^2}{n}}}.$$

$$=\frac{174{\cdot}42-\dfrac{0\times-{\cdot}1}{11}}{\sqrt{365{\cdot}2-\dfrac{0}{11}}\;\sqrt{89{\cdot}05-\dfrac{({\cdot}1)^2}{11}}}$$

$$=\frac{174{\cdot}42}{\sqrt{365{\cdot}2}\;\sqrt{89{\cdot}05}}$$

$$=\frac{174{\cdot}42}{\sqrt{325209{\cdot}60}}$$

$$=\frac{174{\cdot}42}{570{\cdot}2}$$

$$={\cdot}306$$

$$\text{P.E.}={\cdot}6745\,\frac{1-({\cdot}306)^2}{\sqrt{11}}$$

$$={\cdot}6745\times\frac{{\cdot}91}{3{\cdot}316}$$

$$=\frac{{\cdot}613795}{3{\cdot}316}$$

$$={\cdot}185.$$

✓**Problem 199.**—The following table gives data regarding the size of holdings in Uttar Pradesh during the year ending on 30th June 1945 :

Size of holdings in acres	*Total number of persons in thousands*	*Total area in thousands of acres.*
0—1	4639	2481
1—3	3635	6734
3—5	1695	6608
5—10	1563	10822
10—15	392	4752
15—20	167	2836
20—25	70	1570
above 25	115	5310
	12276	41113

Calculate the coefficient of correlation between the total area of holding and the total number of persons.

(*M.A Raj., 1953*)

Solution :

Size of holdings in acres	Total number of persons in thousands			Total area in thousands of acres			Product
	X	ξ=X-1534		Y	η=Y−5146		
0–1	4639	3105	9641025	2481	−2665	7102225	−8264825
1–3	3635	2101	4414401	6734	1588	2521744	3336388
3–5	1695	161	25921	6608	1462	2137444	235382
5–10	1568	29	841	10822	5676	32216976	164604
10–15	392	−1142	1304164	4752	− 394	155236	449948
15–20	167	−1367	1868689	2836	−2310	5336100	3157770
20–25	70	−1464	2143290	1570	−3576	12787776	5235264
over 25	115	−1419	2013561	5310	164	26896	−232716
n = 8		4	21411892		−55	62284397	4081815

$n=8,\ \Sigma\xi=4,\ \Sigma\xi^2=21411892 \qquad \Sigma\xi\eta=4081815$

$\Sigma\eta=-55,\ \Sigma\eta^2=62284397$

$$r=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{4081815-\dfrac{4\times-55}{8}}{\sqrt{21411892-\dfrac{4^2}{8}}\sqrt{62284397-\dfrac{5^2}{8}}}$$

$$=\frac{4081842\cdot5}{\sqrt{21411890}\ \sqrt{62284393\cdot9}}$$

$$=\frac{4081842\cdot5}{\sqrt{1333626590903471}}$$

$$=\frac{4081842\cdot5}{36518852\cdot5}$$

$=\cdot11$

$\therefore\quad r=\cdot11.$

Problem 200.—Calculate the coefficient of correlation of the short-time oscillations for the following indices of supply and price of a commodity:

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Index of Supply	186	168	156	166	178	186	154	136	154	192	210	216	204	194	234
Index of Price	222	228	232	216	200	196	208	214	204	192	185	184	190	196	182

(M. A. Raj., 1951)

Solution :

Year	Index of Supply			Index of price			Product
	X	$\xi = X - 182$	ξ^2	Y	$\eta = Y - 203$	η^2	$\xi\eta$
1	186	4	16	222	19	361	76
2	168	—14	196	228	25	625	—350
3	156	—26	676	232	29	841	—754
4	166	—16	256	216	13	169	—208
5	178	— 4	16	200	— 3	9	12
6	186	4	16	196	— 7	47	—28
7	154	—28	784	208	5	25	—140
8	136	—46	2116	214	11	121	—566
9	154	—28	784	204	1	1	— 28
10	192	10	100	192	—11	121	—110
11	210	28	784	185	—18	324	—504
12	216	34	1156	184	—19	361	—646
13	204	22	484	190	—13	169	—286
14	194	12	144	196	— 7	49	— 84
15	234	52	2704	182	—21	441	—1092
$n=15$	2734	4	10232	3049	4	3664	—4648

$n=15,\quad \Sigma\xi=4,\quad \Sigma\eta=4,\quad \Sigma\xi^2=10232,\quad \Sigma\eta^2=3664,\quad \Sigma\eta\xi=-4648$

$$=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\cdot\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\frac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\frac{(\Sigma\eta^2)}{n}}}$$

$$r=\frac{-4648-\frac{4\times4}{15}}{\sqrt{10232-\frac{4^2}{15}}\sqrt{3664-\frac{4^2}{15}}}$$

$$=\frac{-4649\cdot07}{\sqrt{10230\cdot93}\ \sqrt{3662\cdot93}}$$

$$=\frac{-4649\cdot07}{\sqrt{37475180\cdot4249}}$$

$$=\frac{-4649\cdot07}{6121\cdot69}$$

$$=-\cdot759$$

Problem 201.—The following table gives the population of two towns at the time of the last seven censuses :

Year	1881	1891	1901	1911	1921	1931	1941
Agra (The figures are in thousands)	160	169	188	154	164	205	284
Kanpur (The figures are in thousands)	155	194	203	179	216	243	287

Calculate the coefficient of correlation between the population of Agra and Kanpur. *(M.A. Raj., 1948)*

Solution :

Year	Population of Agra (in thousands)			Population of Kanpur (in thousands)			Product
	X	$\xi=X-189$	ξ^2	Y	$\eta=Y-211$	η^2	$\xi\eta$
1881	160	−29	841	155	−56	3136	1624
1891	169	−20	400	194	−17	289	340
1901	188	−1	1	203	−8	64	8
1911	154	−35	1225	179	−32	1024	1120
1921	164	−25	625	216	5	25	−125
1931	205	16	256	243	32	1024	512
1941	284	95	9025	287	76	5776	7220
$n=7$	1324	1	12373	1477	0	11338	10699

$$n=7,\ \Sigma\xi=1,\ \Sigma\xi^2=12373,\ \Sigma\eta=0,\ \Sigma\eta^2=11338$$

$$\Sigma\xi\eta=10699$$

$$r=\frac{\Sigma\xi\eta-\frac{\Sigma\xi.\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\frac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\frac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{10699-\frac{1\times 0}{7}}{\sqrt{12373-\frac{1}{7}}\sqrt{11338-0}}$$

$$=\frac{10699}{\sqrt{12372{\cdot}86}\sqrt{11338}}$$

$$=\frac{10699}{\sqrt{140283486{\cdot}68}}$$

$$=\frac{10699}{11844\cdot13}$$

$$=\cdot903.$$

Problem 202.—The following table gives the annual birth and death rates in U.S.A. during the period 1931 to 1945 :

Year	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945
Birth Rate	18·0	17·4	16·6	17·2	16·9	16·7	17·1	16·6	17·3	10·9	18·9	20·9	21·5	20·2	19·6
Death Rate	11·1	10·9	10·7	11·1	10·9	11·6	11·3	10·6	10·6	10·7	10·5	10·4	10·9	10·6	10·6

Calculate the coefficient of correlation* between birth rates and death rates.

(*M.A. Raj.*, 1954)

Solution :

Year	Birth Rate			Death Rate			Product
	X	$\xi=X-17\cdot7$	ξ^2	Y	$\eta=Y-10\cdot6$	η^2	$\xi\eta$
1931	18·0	·3	·09	11·1	·5	·29	·15
1932	17·4	− ·3	·09	10·9	·3	·09	− ·09
1933	16·6	−1·1	1·21	10·7	·1	·01	− ·11
1934	17·2	− ·5	·25	11·1	·5	·25	− ·25
1935	16·9	− ·8	·64	10·9	·3	·09	− ·24
1936	16·7	−1·0	1·00	11·6	1·0	1·00	−1·00
1937	17·1	− ·6	·36	11·3	·7	·49	− ·42
1938	16·6	−1·1	1·21	10·6	0	0	0
1939	17·3	− ·4	·16	10·6	0	0	0
1940	10·9	−6·8	46·24	10·7	·1	·01	− ·68
1941	18·9	1·2	1·44	10·5	−·1	·01	− ·12
1942	20·9	3·2	10·24	10·4	−·2	·04	− ·64
1943	21·5	3·8	14·44	10·9	·3	·09	1·14
1944	20·2	2·5	6·25	10·6	0	0	0
1945	19·6	1·9	3·61	10·6	0	0	0
$n=15$	265·8	·3	87·23	162·5	3·5	2·37	−2·26

$n=15,\ \Sigma\xi=\cdot3,\ \Sigma\xi^2=87\cdot23,\ \Sigma\eta=3\cdot5,\ \Sigma\eta^2=2\cdot37$

$$\Sigma\xi\eta=-2\cdot26$$

$$r=\frac{\Sigma\xi n-\dfrac{\Sigma\xi\cdot\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$= \frac{-2{\cdot}26 - \frac{{\cdot}3 \times 3{\cdot}5}{15}}{\sqrt{87{\cdot}23 - \frac{({\cdot}3)^2}{15}}\sqrt{2{\cdot}37 - \frac{(3{\cdot}5)^2}{15}}}$$

$$= \frac{-2{\cdot}33}{\sqrt{87{\cdot}17}\sqrt{1{\cdot}55}}$$

$$= \frac{-2{\cdot}33}{\sqrt{135\ 1135}}$$

$$= \frac{-2{\cdot}33}{11{\cdot}62}$$

$$= -{\cdot}2.$$

Problem 203.—From the following table calculate the coefficient of correlation between the value of the exports of raw cotton and the value of the imports of cotton manufactured goods. Calculate the standard error of the coefficient of correlation also.

Year	1913-14	1917-18	1919-20	1921-22	1923-24	1929-30	1931-32
Export of raw cotton (In crores of Rs.)	42	44	58	55	89	98	66
Import of manufactured cotton goods (In crores of Rs.)	56	49	53	58	65	76	58

(*M.A. Cal., 1937*) ; (*M.A. Raj., 1956*); (*B.Com. Nagpur, 1944*)

Solution :

Year	Export of raw cotton (in crores of Rs.)			Import of manufactured cotton goods			Product $\xi\eta$
	X	$\xi = X - 64$	ξ^2	Y	$\eta = Y - 59$	η^2	
1913-14	42	−22	484	56	−3	9	66
1917-18	44	−20	400	49	−10	100	200
1919-20	58	−6	36	53	−6	36	36
1921-22	55	−9	81	58	−1	1	9
1923-24	89	25	625	65	6	36	150
1929-30	98	34	1156	76	17	289	578
1931-32	66	2	4	58	−1	1	−2
$n=7$	452	4	2786	415	2	472	1037

$$n=7,\ \Sigma\xi=4,\ \Sigma\xi^2=2786,\ \Sigma\eta=2,\ \Sigma\eta^2=472,\ \Sigma\xi\eta=1037$$

$$r = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \cdot \Sigma\eta}{n}}{\sqrt{\Sigma\xi^2 - \frac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2 - \frac{(\Sigma\eta)^2}{n}}}$$

$$= \frac{1037 - \frac{4\times 2}{7}}{\sqrt{2786 - \frac{4^2}{7}}\sqrt{472 - \frac{2^2}{7}}}$$

$$= \frac{1037 - 1{\cdot}14}{\sqrt{2783{\cdot}72}\sqrt{471{\cdot}43}}$$

$$= \frac{1035{\cdot}86}{\sqrt{1312329{\cdot}1196}}$$

$$= \frac{1035{\cdot}86}{1145{\cdot}56}$$

$$= {\cdot}904$$

$$\text{Standard error} = \frac{1 - r^2}{\sqrt{n}}$$

$$= \frac{1 - ({\cdot}904)^2}{\sqrt{7}}$$

$$= \frac{1 - ({\cdot}817216)}{2{\cdot}645}$$

$$= \frac{{\cdot}182784}{2{\cdot}645}$$

$$= {\cdot}069$$

$$\therefore \quad r = {\cdot}904$$

And Standard Error $= {\cdot}069$.

Problem 204.—From the following table, find out how far the fluctuations in prices correspond to the amount of money in circulation in India.

Year	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
Rs. and Notes in circulation in crores	248	256	248	266	297	338	407	463	411	393
Index Numbers of prices (1873=100)	137	143	147	152	184	196	225	276	281	260

(B. Com. Agra, 1937)

Solution :

Year	Rs. and Notes in circulation in crores			Index Number of Prices (1873=100)			Product
	X	$\xi = X - 332$	ξ^2	Y	$\eta = Y - 200$	η^2	
1912	248	−84	7056	137	−63	3969	5292
1913	256	−76	5776	143	−57	3249	4332
1914	248	−84	7056	147	−53	2809	4452
1915	266	−66	4356	152	−48	2304	3168
1916	297	−35	1225	184	−16	256	560
1917	338	6	36	196	− 4	16	−24
1918	407	75	5625	225	25	625	1875
1919	463	131	17161	276	76	5776	9956
1920	411	79	6241	281	81	6561	6399
1921	393	61	3721	260	60	3600	3660
$n=10$	3327	7	58253	2001	1	29165	39670

$n=10$, $\Sigma\xi=7$, $\Sigma\xi^2=58253$, $\Sigma\eta=1$, $\Sigma\eta^2=29165$, $\Sigma\xi\eta=39670$

$$r=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\,.\,\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\frac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\frac{(\Sigma\eta)^2}{n}}}$$

$$r=\frac{39670-\frac{7\times1}{10}}{\sqrt{58253-\frac{7^2}{10}}\;\sqrt{29165-\frac{1}{10}}}$$

$$=\frac{39669\cdot3}{\sqrt{58248\cdot1}\;\sqrt{29164\cdot9}}$$

$$=\frac{39669\cdot3}{\sqrt{1698800011\cdot69}}$$

$$=\frac{39669\cdot3}{41216\cdot5}$$

$$=\cdot96$$

$$\text{P.E.}=\cdot6745\,\frac{1-r^2}{\sqrt{n}}$$

$$=\cdot6745\,\frac{1-(\cdot96)^2}{\sqrt{10}}$$

$$=\cdot6745\,\frac{1-\cdot9216}{3\cdot162}$$

$$=\cdot6745\times\cdot0248$$

$$=\cdot01672760$$

$$6\ \text{P.E.}=\cdot1003656$$

since $r > 6$ P.E.

hence the fluctuations in prices correspond to the amount of money in circulation to a very great extent.

Problem 205.—The following table gives the birth rates of a few countries of the world during the year 1931 :—

Country	Egypt	Canada	U.S.A.	India	Japan	Germany	France	I.F.S.	U.K.	Russia	Australia	Newzeland	Palestine	Sweden	Norway
Birth-rate	44	24	19	33	32	16	18	20	16	40	20	18	53	15	17
Death-rate	27	11	12	24	19	11	16	14	12	18	9	8	23	12	11

Find r between the Birth rate and Death rate figures.

(B. Com. Luck. 1938)

Solution :

Country	Birth-Rate			Death-rate			Product
	X	$\xi =$ X−25	ξ^2	Y	$\eta =$ Y−15	η^2	
Egypt	44	19	361	27	12	144	228
Canada	24	−1	1	11	−4	16	4
U.S.A.	19	−6	36	12	−3	9	18
India	33	8	64	24	9	81	72
Japan	32	7	49	19	4	16	28
Germany	16	−9	81	11	−4	16	36
France	18	−7	49	16	1	1	−7
I.F.S.	20	−5	25	14	−1	1	5
U.K.	16	−9	81	12	−3	9	27
Russia	40	15	225	18	3	9	45
Australia	20	−5	25	9	−6	36	30
Newzeland	18	−7	49	8	−7	49	49
Palestine	53	28	784	23	8	64	224
Sweden	15	−10	100	12	−3	9	30
Norway	17	−8	64	11	−4	16	32
$n=15$	385	10	1994	227	2	476	821

$$n = 15,\ \Sigma\xi = 10,\ \Sigma\xi^2 = 1994,\ \Sigma\eta = 2,\ \Sigma\eta^2 = 476$$

$$\Sigma\xi\eta = 821$$

$$r = \frac{\Sigma\xi\eta - \dfrac{\Sigma\xi \cdot \Sigma\eta}{n}}{\sqrt{\Sigma\xi^2 - \dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2 - \dfrac{(\Sigma\eta)^2}{n}}}$$

$$= \frac{821 - \dfrac{10 \times 2}{15}}{\sqrt{1994 - \dfrac{10^2}{15}}\sqrt{476 - \dfrac{2^2}{15}}}$$

$$= \frac{821 - 1{\cdot}33}{\sqrt{1994 - 6{\cdot}6}\ \sqrt{476 - {\cdot}27}}$$

$$= \frac{819{\cdot}67}{\sqrt{1987{\cdot}4 \times 475{\cdot}73}}$$

$$= \frac{819{\cdot}67}{\sqrt{945465{\cdot}802}}$$

$$= \frac{819{\cdot}67}{972{\cdot}35}$$

$$= {\cdot}842$$

Problem 206.—Calculate the coefficient of correlation between the cost of living and the weekly wage rates from the following data :—

Date	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932
Cost of living Index	151	110	102	101	103	100	100	96	95	95	87	84	81
Index of weekly wage Rates	155	120	99	98	101	101	102	100	99	99	98	96	94

(*M.A., Alld., 1937*)

Solution :

Date	Cost of living Index			Index of weekly wage Rates			Product
	X	$\xi = X-100$	ξ^2	Y	$\eta = Y-104$	η^2	$\xi\eta$
1920	151	51	2601	155	51	2601	2601
1921	110	10	100	120	16	256	160
1922	102	2	4	99	−5	25	−10
1923	101	1	1	98	−6	36	−6
1924	103	3	9	101	−3	9	−9
1925	100	0	0	101	−3	9	0
1926	100	0	0	102	−2	4	0
1927	96	−4	16	100	−4	16	16
1928	95	−5	25	99	−5	25	25
1929	95	−5	25	99	−5	25	25
1930	87	−13	169	98	−6	36	78
1931	84	−16	256	96	−8	64	128
1932	81	−19	361	94	−10	100	190
$n=13$	1305	5	3567	1362	10	3206	3198

$n=13,\ \Sigma\xi=5,\ \Sigma\xi^2=3567,\ \Sigma\eta=10,\ \Sigma\eta^2=3206$

$\Sigma\xi\eta=3198$

$$r=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi\,.\,\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{3198-\dfrac{5\times10}{13}}{\sqrt{3567-\dfrac{5^2}{13}}\sqrt{3206-\dfrac{10^2}{13}}}$$

$$=\frac{3198-3{\cdot}846}{\sqrt{3565{\cdot}08}\sqrt{3198{\cdot}3}}$$

$$=\frac{3194{\cdot}154}{\sqrt{11402195{\cdot}364}}$$

$$=\frac{3194{\cdot}154}{3376{\cdot}7}$$

$= {\cdot}945.$

Problem 207.—The following data gives the index numbers of Industrial production of Great Britain and the number of registered unemployed persons in the same country during the years 1924-31:

Year	*Industrial production (Index Number)*	*Number of Registered unemployed (Hundred thousand)*
1924	100	11·3
1925	102	12·4
1926	104	14·0
1927	107	11·1
1928	105	12·3
1929	112	12·2
1930	103	19·1
1931	94	26·4

Calculate the coefficient of correlation between production and the number of unemployed. (*B. Com. Luck., 1944*)

Solution :

Year	Industrial Production (Index Number)			Number of Registered unemployed (Hundred thousand)			Product
	X	$\xi=X-103$	ξ^2	Y	$\eta=Y-14\cdot8$	η^2	$\xi\eta$
1924	100	−3	9	11·3	−3·5	12·25	+10·5
1925	102	−1	1	12·4	−2·4	5·76	2·4
1926	104	1	1	14·0	− ·8	·64	− ·8
1927	107	4	16	11·1	−3·7	13·69	−14·8
1928	105	2	4	12·3	−2·5	6·25	− 5·0
1929	112	9	81	12·2	−2·6	6·76	−23·4
1930	103	0	0	19·1	4·3	18·49	0
1931	94	−9	81	26·4	11·6	134·56	−104·4
$n=8$	827	3	193	118·8	·4	198·40	−135·5

$n=8$, $\Sigma\xi=3$, $\Sigma\xi^2=193$, $\Sigma\eta=\cdot4$, $\Sigma\eta^2=198\cdot4$

$$\Sigma\xi\eta=-135\cdot5$$

$$r=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi\cdot\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{-135\cdot5-\dfrac{3\times\cdot4}{8}}{\sqrt{193-\dfrac{3^2}{8}}\sqrt{198\cdot4-\dfrac{(\cdot4)^2}{8}}}$$

$$=\frac{-135\cdot65}{\sqrt{191\cdot875}\sqrt{198\cdot38}}$$

$$= \frac{-135\cdot65}{\sqrt{38064\cdot1625}}$$

$$= \frac{-135\cdot65}{195\cdot1}$$

$$= -\cdot695$$

$$\therefore r = -\cdot695.$$

Problem 208.—The following table gives the wholesale price Index numbers for Calcutta and Karachi for the period 1927-1941 :

Year	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
Calcutta Index Numbers (Base : July, 1914)	148	145	141	116	96	91	87	89	91	91	102	95	108	120	139
Karachi Index Numbers (Base : July, 1914)	137	137	135	108	95	99	97	96	99	102	108	104	108	116	120

(*a*) Calculate the coefficient of correlation between the above two series and state what it indicates.

(*b*) Point out whether the Calcutta indices are more variable than the Karachi ones.

(*B. Com. Alld.*, *1944*)

Solution :

Year	Calcutta Index Number			Karachi Index Number			Product $\xi\eta$
	X	ξ=X−110	ξ^2	Y	η=Y−110	η^2	
1927	148	38	1444	137	27	729	1026
1928	145	35	1225	137	27	729	945
1929	141	31	961	133	23	529	713
1930	116	6	36	108	−2	4	−12
1931	96	−14	196	95	−15	225	210
1932	91	−19	361	99	−11	121	209
1933	87	−23	529	97	−13	169	299
1934	89	−21	441	96	−14	196	294
1935	91	−19	361	99	−11	121	209
1936	91	−19	361	102	−8	64	152
1937	102	−8	64	108	−2	4	16
1938	95	−15	225	104	−6	36	90
1939	108	−2	4	108	−2	4	4
1940	120	10	100	116	6	36	60
1941	139	29	841	120	10	100	290
n=15	1659	9	7149	1659	9	3067	4505

$n=15,\ \Sigma\xi=9,\ \Sigma\xi^2=7149,\ \Sigma\eta=9,\ \Sigma\eta^2=3067$

$$\Sigma\xi\eta=4505$$

$$r=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi.\xi\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\ \sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{4505-\dfrac{9\times9}{15}}{\sqrt{7149-\dfrac{9^2}{15}}\ \sqrt{3067-\dfrac{9^2}{15}}} \quad \text{— ①}$$

$$=\frac{4505-5{\cdot}4}{\sqrt{7149-5{\cdot}4}\ \sqrt{3067-5{\cdot}4}}$$

$$=\frac{4499{\cdot}6}{\sqrt{7143.6\times3061{\cdot}6}}$$

$$=\frac{4499{\cdot}6}{\sqrt{21870845{\cdot}76}}$$

$$=\frac{4499{\cdot}6}{4676{\cdot}6}$$

$$={\cdot}962$$

$$\text{Probable error}={\cdot}6745\frac{1-r^2}{\sqrt{n}}$$

$$={\cdot}6745\,\frac{1-({\cdot}962)^2}{\sqrt{15}}$$

$$={\cdot}6745\,\frac{1-{\cdot}925}{3{\cdot}872}$$

$$={\cdot}6745\,\frac{{\cdot}075}{3{\cdot}872}$$

$$=\frac{{\cdot}0505875}{3{\cdot}872}$$

$$={\cdot}013.$$

Since $r>6$ P.E, hence, Calcutta Index Number is correlated to Karachi Index Number to a very great extent.

Since standard deviation of Calcutta Index number is greater than standard deviation for Karachi, hence Calcutta indices are more variable than the Karachi ones.

from the denominator of ①

Problem 209.—The index numbers of prices of all commodities in Bombay and in Calcutta were as under :

Month	May 1942	June 1942	July 1942	Aug. 1942	Sep. 1942	Oct. 1942	Nov. 1942	Dec. 1942	Jan. 1943	Feb. 1943
Index Number of Commodity prices in Calcutta	169	182	182	192	198	209	227	238	250	253
Index Number of Commodity prices in Bombay	204	222	225	228	229	233	249	266	255	255

Do you think prices in Bombay and in Calcutta are correlated ?

(*M.A. Agra, 1944*)

Solution :

Month	Index number of commodity prices in Calcutta			Index number of commodity prices in Bombay			Product $\xi\eta$
	X	$\xi=$ X−210	ξ^2	Y	$\eta=$ Y−230	η^2	
May 1942	169	−41	1681	204	−26	676	1066
June 1942	182	−28	784	222	−8	64	224
July 1942	182	−28	784	225	−5	25	140
Aug. 1942	192	−18	324	228	−2	4	36
Sept. 1942	198	−12	144	229	−1	1	12
Oct. 1942	209	−1	1	233	3	9	−3
Nov. 1942	227	17	289	249	19	361	323
Dec. 1942	238	28	784	266	36	1296	1008
Jan. 1943	250	40	1600	255	25	625	1000
Feb. 1943	253	43	1849	255	25	625	1075
$n=10$	2100	0	8240	2366	66	3686	4881

$\therefore$ $n=10$, $\Sigma\xi=0$, $\Sigma\xi^2=8240$, $\Sigma\eta=66$, $\Sigma\eta^2=3686$

$\Sigma\xi\eta=4881$

$$r=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi\cdot\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\dfrac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\dfrac{(\Sigma\eta)^2}{n}}}$$

$$=\frac{4881-0}{\sqrt{8240}\sqrt{3686-\dfrac{66^2}{10}}}$$

$$=\frac{4881}{\sqrt{8240}\sqrt{3686-435\cdot6}}$$

$$= \frac{4881}{\sqrt{8240}\ \sqrt{3250 \cdot 4}}$$

$$= \frac{4881}{\sqrt{26783296}}$$

$$= \frac{4881}{5175 \cdot 2}$$

$$= \cdot 943$$

Problem 210.—Calculate the coefficient of correlation from the following table, and interpret it.

Year	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934
Average daily No. of labourers (in thousands)	368	384	385	361	347	384	395	403	400	385
Lakhs of Bales consumed by mills	22	21	24	20	22	26	26	29	28	27

(*B. Com. Agra, 1941*)

Solution :

Year	Av. daily No. of labourers (in thousands)			Lakhs of Bales consumed by mills			Product
	X	$\xi =$ X—381	ξ^2	Y	$\eta =$ Y—24	η^2	$\xi\eta$
1925	368	—13	169	22	—2	4	26
1926	384	3	9	21	—3	9	— 9
1927	385	4	16	24	0	0	0
1928	361	—20	400	20	—4	16	80
1929	347	—34	1156	22	—2	4	68
1930	384	3	9	26	2	4	6
1931	395	14	196	26	2	4	28
1932	403	22	484	29	5	25	110
1933	400	19	361	28	4	16	76
1934	385	4	16	27	3	9	12
$n=10$	3812	2	2816	245	5	91	397

$$n=10,\ \Sigma\xi=2,\ \Sigma\xi^2=2816,\ \Sigma\eta=5,\ \Sigma\eta^2=91$$

$$\Sigma\xi\eta=397$$

$$r=\frac{\Sigma\xi\eta-\frac{\Sigma\xi.\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2-\frac{(\Sigma\xi)^2}{n}}\sqrt{\Sigma\eta^2-\frac{(\xi\eta)^2}{n}}}$$

$$= \frac{397 - \frac{2 \times 5}{10}}{\sqrt{2816 - \frac{4}{10}}\sqrt{91 - \frac{25}{10}}}$$

$$= \frac{396}{\sqrt{2815 \cdot 6}\ \sqrt{88 \cdot 5}}$$

$$= \frac{396}{\sqrt{249180 \cdot 6}}$$

$$= \frac{396}{499 \cdot 18}$$

$$= \cdot 793$$

Since the coefficient of correlation is positive and very great hence, this shows that increase in the number of labourers follows with the increase in consumption of bales and *vice versa*.

Problem 211.—Explain correlation and calculate co-efficient of correlation between ages of husbands and wives in the following : *(M.A., Agra, 1953)*

Age of husband	Age of wives in years					Total
	10—20	20—30	30—40	40—50	50—60	
15—25	6	3				9
25—35	3	16	10			29
35—45		10	15	7		32
45—55			7	10	4	21
55—65				4	5	9
Total	9	29	32	21	9	100

Solution :

$$\text{Let } \xi = \frac{X - \text{assumed mean}}{i}$$

when i is the difference between two intervals

$$\eta = \frac{Y - \text{assumed mean}}{i}$$

Let the assumed mean of the X series = 40

" " " " Y " = 35

Ages of Husbands X↓ / Ages of Wives Y→	10 to 20 (15)	20 to 30 (25)	30 to 40 (35)	40 to 50 (45)	50 to 60 (55)	Total of f↓	ξ	ξ²	fξ	fξ²
15 to 25 (20)	[4] 6 24	[2] 3 6				9 [30]	-2	4	-18	36
25 to 35 (30)	[2] 3 6	[1] 16 16	[0] 10 0			29 [22]	-1	1	-29	29
35 to 45 (40)		[0] 10 0	[0] 15 0	[0] 7 0		32 (0)	0	0	0	0
45 to 55 (50)			[0] 7 0	[1] 10 10	[2] 4 8	21 [18]	+1	1	21	21
55 to 65 (60)				[2] 4 8	[4] 5 20	9 [28]	+2	4	18	36
Total of f→	9 [30]	29 [22]	32 [0]	21 [18]	9 [28]	100 [98]	0	10	-8	122
η	-2	-1	0	1	2	0				
η²	4	1	0	1	4	10				
fη	-18	-29	0	21	18	-8				
fη²	36	29	0	21	36	122				

No need

$$r = \frac{\Sigma f \xi \eta - \frac{\Sigma f \xi \, \Sigma f \eta}{n}}{\sqrt{\left[\Sigma f \xi^2 - \frac{(\Sigma f \xi)^2}{n}\right]\left[\Sigma f \eta^2 - \frac{(\Sigma f \eta)^2}{n}\right]}}$$

$$\Sigma f \xi \eta = 30 + 22 + 18 + 28 = 98$$

$$\therefore r = \frac{98 - \frac{(-8)\times(-8)}{100}}{\sqrt{\left[122 - \frac{64}{100}\right]\left[122 - \frac{64}{100}\right]}}$$

$$= \frac{98 - \frac{64}{100}}{122 - \frac{64}{100}}$$

$$= \frac{98 - \cdot 64}{122 - \cdot 64}$$

$$= \frac{97 \cdot 36}{121 \cdot 36}$$

$$= \frac{9736}{12136}$$

$$= + \cdot 802$$

Thus, we find that there is a high degree of positive correlation between the ages of husbands and wives.

Problem 212.—Define Pearson's coefficient of correlation, and establish a formula for it taking into account deviations from assumed means. Find its value for the following table :—

X \ Y	0—5	5—10	10—15	15—20	20—25
0— 4	1	2			2
4— 8		4	5	8	
8—12		1	3	4	
12—16		2		2	1

(*M.Sc. Agra, 1951*)

Solution :

Using the step-deviation method we have

X \ Y	0-5	5-10	10-15	15-20	20-25	TOTAL f	u	uf	fu²
0-4	[3] 1	[2] 2			[-1] 2	5	-1	-5	5
4-8		[0] 4	[0] 5	[0] 8		17	0	0	0
8-12		[-2] 1	[-1] 3	[0] 4		8	1	8	8
12-16		[-4] 2		[0] 2	[2] 1	5	2	10	20
TOTAL f	1	9	8	14	3	35	TOTAL	13	33
v	-3	-2	-1	0	1	TOTAL			
fv	-3	-18	-8	0	3	-26			
fv²	3	36	8	0	3	50			

$$u = \frac{X-(4\text{-}8)}{4}$$

$$v = \frac{Y-(15\text{-}20)}{5}$$

SMALL SQUARES REPRESENTS PRODUCT uv

$\Sigma fu=13,\ \Sigma fu^2=33,\ \Sigma fv=-26,\ \Sigma fv^2=50,\ \Sigma f=35$

$$\Sigma fuv=(3\times 1)+(2\times 2)+(-1\times 2)+(-2\times 1)+(-1\times 3)+(-4\times 2)+(2\times 1)$$

$$=3+4-2-2-3-8+2=-6$$

$$r=\frac{\Sigma fuv-\frac{\Sigma fu\ \Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2-\frac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2-\frac{(\Sigma fv)^2}{\Sigma f}}}$$

$$=\frac{-6-\frac{13\times -26}{35}}{\sqrt{33-\frac{13^2}{35}}\sqrt{50-\frac{26^2}{35}}}$$

$$=\frac{3{\cdot}65}{\sqrt{865{\cdot}126}}=\frac{3{\cdot}65}{29{\cdot}4}={\cdot}12$$

$\therefore\quad r={\cdot}12$

Problem 213.—Calculate the value of the coefficient of correlation for the following data :—

X/Y	16-18	18-20	20-22	22-24
10-20	2	1	1	
20-30	3	2	3	2
30-40	3	4	5	6
40-50	2	2	3	4
50-60		1	2	2
60-70		1	2	1

(*P.C.S. 1952 ; M.A. Raj. 1952 ; M.A. Aligarh 1941*)

Solution :

Using the step-deviation Method for Grouped datas we have the table as below :

X \ Y	16-18	18-20	20-22	22-24	TOTAL f	STEP DEVIATION u	fu	fu^2
10-20	[4] 2	[2] 1	[0] 1	[-2]	4	-2	-8	16
20-30	[2] 3	[1] 2	[0] 3	[-1] 2	10	-1	-10	10
30-40	[0] 3	[0] 4	[0] 5	[0] 6	18	0	0	0
40-50	[-2] 2	[-1] 2	[0] 3	[1] 4	11	1	11	11
50-60	[-4]	[-2] 1	[0] 2	[2] 2	5	2	10	20
60-70	[-6]	[-3] 1	[0] 2	[3] 1	4	3	12	36
TOTAL f	10	11	16	15	52	TOTAL	15	93
STEP DEVIATION v	-2	-1	0	1	TOTAL			
fv	-20	-11	0	15	-16			
fv^2	40	11	0	15	66			

$$u = \frac{X-(30-40)}{10}$$

$$v = \frac{Y-(20-22)}{2}$$

SMALL SQUARES REPRESENT PRODUCT uv

$\therefore$ from the table we have

$\Sigma fu = 15, \quad \Sigma fu^2 = 93$

$\Sigma fv = -16, \quad \Sigma fv^2 = 66 \quad$ and $\Sigma f = 52.$

$$\Sigma fuv = (4\times 2)+(2\times 1)+(0\times 1)+(2\times 3)+(1\times 2)+(0\times 3)$$
$$+(-1\times 2)+(0\times 3)+(0\times 4)+(0\times 5)+(0\times 6)$$
$$+(-2\times 2)+(-1\times 2)+(0\times 2)+(2\times 2)+(-3\times 1)$$
$$+(0\times 2)+(3\times 1)$$
$$=8+2+0+6+2+0-2+0+0+0+0-4-2+0+4-3$$
$$+0+3$$
$$=14$$

The Pearson's Coefficient of Correlation is given by the formula :

$$r=\frac{\Sigma fuv-\frac{\Sigma fu\Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2-\frac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2-\frac{(\Sigma fv)^2}{\Sigma f}}}$$

$$=\frac{14-\frac{15\times-16}{52}}{\sqrt{93-\frac{15^2}{52}}\sqrt{66-\frac{16^2}{52}}}$$

$$=\frac{14+4{\cdot}6}{\sqrt{93-4{\cdot}3}\sqrt{66-4{\cdot}9}}$$

$$=\frac{18{\cdot}6}{\sqrt{88{\cdot}7}\sqrt{61{\cdot}1}}$$

$$=\frac{18{\cdot}6}{\sqrt{5419{\cdot}57}}$$

$$=\frac{18{\cdot}6}{73{\cdot}6}$$

$$={\cdot}25$$

Hence, Coefficient of correlation is $+{\cdot}25$.

Problem 214.—Prove that the coefficient of correlation lies between +1 and −1.

Calculate the same for the following table :

X \ Y	17·5	22·5	27·5	32·5	37·5	42·5	47·5	52·5	57·5
17·5	11	62	19	3	1				
22·5	6	220	190	34	6	2			
27·5		46	165	59	13	5	2	1	
32·5		3	25	33	14	6	3	2	
37·5		1	3	8	9	6	4	3	1
42·5				1	3	5	4	3	2
47·5					1	2	3	3	2
52·5							1	2	2

(P.C.S. 1954)

Solution :

Using the step-deviation method for Gro uped datas we have

X \ Y	17·5	22·5	27·5	32·5	37·5	42·5	47·5	52·5	57·5	TOTAL f	u	fu	fu^2
17·5	[6] 11	[4] 62	[2] 19	[0] 3	[-2] 1					96	-2	-192	384
22·5	[3] 6	[2] 220	[1] 190	[0] 34	[-1] 6	[-2] 2				458	-1	-458	458
27·5		[0] 46	[0] 165	[0] 59	[0] 13	[0] 5	[0] 2	[0] 1		291	0	0	0
32·5		[-2] 3	[-1] 25	[0] 33	[1] 14	[2] 6	[3] 3	[4] 2		86	1	86	86
37·5		[-4] 1	[-2] 3	[0] 8	[1] 9	[4] 6	[6] 4	[8] 3	[10] 1	35	2	70	140
42·5				[0] 1	[3] 3	[6] 5	[9] 4	[12] 3	[15] 2	18	3	54	162
47·5					[4] 1	[8] 2	[12] 3	[16] 3	[20] 2	11	4	44	176
52·5							[15] 1	[20] 2	[25] 2	5	5	25	125
TOTAL f	17	332	402	138	47	26	17	14	7	1000	TOTAL	-371	1531
v	-3	-2	-1	0	1	2	3	4	5	TOTAL			
fv	-51	-664	-402	0	47	52	51	56	35	-876			
fv^2	153	1328	402	0	47	104	153	224	175	2586			

$u = \dfrac{X-27\cdot5}{5}$

$v = \dfrac{Y-32\cdot5}{5}$

SMALL SQUARES REPRESENT PRODUCT uv

$$\therefore \quad \Sigma fu = -371, \quad \Sigma fu^2 = 1531$$
$$\Sigma fv = -876, \quad \Sigma fv^2 = 2586 \qquad \Sigma f = 1000$$

$$\begin{aligned}\Sigma fuv = &(6\times11)+(4\times62)+(2\times19)+(-2\times1)+(3\times6)\\&+(2\times220)+(1\times190)+(-1\times6)+(-2\times2)+(-2\times3)\\&+(-1\times25)+(1\times14)+(2\times6)+(3\times3)+(4\times2)\\&+(-4\times1)+(-2\times3)+(1\times9)+(4\times6)+(6\times4)\\&+(8\times3)+(10\times1)+(3\times3)+(6\times5)+(9\times4)+(12\times3)\\&+(15\times2)+(4\times1)+(8\times2)+(12\times3)+(16\times3)\\&+(20\times2)+(15\times1)+(20\times2)+(25\times2)\end{aligned}$$

$$\begin{aligned}\text{or } \Sigma fuv = &66+248+38-2+18+440+190-6-4-6-25+14\\&+12+9+8-4-6+9+24+24+24+10+9+30\\&+36+36+30+4+16+36+48+40+15+40+50\\=&1471\end{aligned}$$

The Pearson's Coefficient of Correlation is given by the formula

$$r=\frac{\Sigma fuv-\frac{(\Sigma fu)(\Sigma fv)}{\Sigma f}}{\sqrt{\Sigma fu^2-\frac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2-\frac{(\Sigma fv)^2}{\Sigma f}}}$$

$$=\frac{1471-\frac{-371\times-876}{1000}}{\sqrt{1531-\frac{371^2}{1000}}\sqrt{2586-\frac{876^2}{1000}}}$$

$$=\frac{1471-324{\cdot}996}{\sqrt{1393{\cdot}359}\sqrt{1818{\cdot}624}}$$

$$=\frac{1146{\cdot}004}{\sqrt{2533996{\cdot}118016}}$$

$$=\frac{1146{\cdot}004}{1591{\cdot}8}$$

$$={\cdot}72$$

∴ Pearson's Coefficient of Correlation is ·72.

Problem 215.—The following table gives according to age the frequency of marks obtained by 100 students in an intelligence test :

Marks \ Age in years	18	19	20	21	Total
10—20	4	2	2		8
20—30	5	4	6	4	19
30—40	6	8	10	11	35
40—50	4	4	6	8	22
50—60		2	4	4	10
60—70		2	3	1	6
TOTAL	19	22	31	28	100

Calculate the coefficient of correlation between age and intelligence. (*P.C.S. 1955*)

Solution :

Using the step deviation method we have :

X \ Y	18	19	20	21	TOTAL f	u	fu	fu^2
10-20	[4] 4	[2] 2	[0] 2		8	-2	-16	32
20-30	[2] 5	[1] 4	[0] 6	[-1] 4	19	-1	-19	19
30-40	[0] 6	[0] 8	[0] 10	[0] 11	35	0	0	0
40-50	[-2] 4	[-1] 4	[0] 6	[1] 8	22	1	22	22
50-60		[-2] 2	[0] 4	[2] 4	10	2	20	40
60-70		[-3] 2	[0] 3	[3] 1	6	3	18	54
TOTAL f	19	22	31	28	100	TOTAL	25	167
v	-2	-1	0	1	TOTAL			
fv	-38	-22	0	28	-32			
fv^2	76	22	0	28	126			

$u = \frac{X-20}{1}$

$v = \frac{Y-(30-40)}{10}$

SMALL SQUARES REPRESENT PRODUCT uv

$\Sigma fu = 25, \quad \Sigma fu^2 = 167$

$\Sigma fv = -32, \quad \Sigma fv^2 = 126 \qquad \Sigma f = 100$

$$\Sigma fuv = (4\times4)+(2\times2)+(2\times5)+(1\times4)+(-1\times4)+(-2\times4)+(-1\times4)+(1\times8)+(-2\times2)+(2\times4)+(-3\times2)+(3\times1)$$

$$=16+4+10+4-4-8-4+8-4+8-6+3$$

$$=27$$

$$r = \frac{\Sigma f u v - \dfrac{\Sigma f u \, \Sigma f v}{\Sigma f}}{\sqrt{\Sigma f u^2 - \dfrac{(\Sigma f u)^2}{\Sigma f}} \sqrt{\Sigma f v^2 - \dfrac{(\Sigma f v)^2}{\Sigma f}}}$$

$$= \frac{27 - \dfrac{25 \times -32}{100}}{\sqrt{167 - \dfrac{25^2}{100}} \sqrt{126 - \dfrac{32^2}{100}}}$$

$$= \frac{35}{\sqrt{160{\cdot}75} \sqrt{115{\cdot}76}}$$

$$= \frac{35}{136{\cdot}9}$$

$$= {\cdot}25$$

∴ Coefficient of correlation is ·25

Problem 216.—Show that r, the coefficient of correlation, cannot numerically exceed unity.

Calculate from the data reproduced below pertaining to 66 selected villages in Meerut District, the value of r between 'total cultivable area' and the 'area under wheat.'

Total Cultivable Area (In Bighas)

Area under Wheat (In Bighas)	0—	500—	1000—	1500—	2000—2500	Total
0—	12	6	...	...	...	18
200—	2	18	4	2	1	27
400—	...	4	7	3	...	14
600—	...	1	...	2	1	4
800—1000	...	...	...	1	2	3
Total	14	28	11	8	4	66

(*I.A.S.* 1949)

Solution :

Using the step-deviation method for grouped series we have the table as below :

X \ Y	0–	500–	1000–	1500–	2000–2500	Total f	u	fu	fu²
0–	[1] 12	[0] 6				18	–1	–18	18
200–	[0] 2	[0] 18	[0] 4	[0] 2	[0] 1	27	0	0	0
400–		[0] 4	[1] 7	[2] 3		14	1	14	14
600–		[0] 1		[4] 2	[6] 1	4	2	8	16
800–1000				[6] 1	[9] 2	3	3	9	27
Total f	14	28	11	8	4	66	Total	13	75
v	–1	0	1	2	3	Total			
fv	–14	0	11	16	12	25			
fv²	14	0	11	32	36	93			

$u=\dfrac{X-(200-)}{200}$

$v=\dfrac{Y-(500-)}{500}$

SMALL SQUARES REPRESENT PRODUCT uv

$\therefore\ \Sigma fu=13,\quad \Sigma fu^2=75$

$\Sigma fv=25,\quad \Sigma fv^2=93 \qquad \Sigma f=66$

$$r=\frac{\Sigma fuv-\dfrac{\Sigma fu\,\Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2-\dfrac{(\Sigma fu)^2}{\Sigma f}}\ \sqrt{\Sigma fv^2-\dfrac{(\Sigma fv)^2}{\Sigma f}}}$$

$\Sigma fuv=(1\times12)+(1\times7)+(2\times3)+(4\times2)+(6\times1)+(6\times1)$
$+(9\times2)$

$=12+7+6+8+6+6+18$

$=63$

$$r=\frac{63-\dfrac{13\times25}{66}}{\sqrt{75-\dfrac{13^2}{66}}\ \sqrt{93-\dfrac{25^2}{66}}}$$

$$=\frac{58\cdot08}{\sqrt{72\cdot44}\ \sqrt{3\cdot53}}$$

$$= \frac{58 \cdot 08}{\sqrt{6050 \cdot 9132}}$$

$$= \frac{58 \cdot 08}{77 \cdot 78} = \cdot 74$$

$\therefore r \quad = \cdot 74$

Problem 217.—The correlation table given below shows the ages of husband and wife for 53 married couples living together on the census night of 1941. Calculate the co-coefficient of correlation between the age of husband and that of his wife.

Age of husband	*Age of wife* 15—25	25—35	35—45	45—55	55—65	65—75	*Total*
15—25	1	1	...	...	...	...	2
25—35	2	12	1	...	...	...	15
35—45	...	4	10	1	...	...	15
45—55	...	...	3	6	1	...	10
55—65	...	...	...	2	4	2	8
65—75	...	...	...	...	1	2	3
Total	3	17	14	9	6	4	53

(*I.A.S. 1950*)

Solution :

Using the step-deviation method for grouped datas we have the following table :

X AGE OF HUSBAND \ AGE OF WIFE Y	15-25	25-35	35-45	45-55	55-65	65-75	TOTAL f	u	fu	fu^2
15-25	[4] 1	[2] 1					2	−2	−4	8
25-35	[2] 2	[1] 12	[0] 1				15	−1	−15	15
35-45		[0] 4	[0] 10	[0] 1			15	0	0	0
45-55			[0] 3	[1] 6	[2] 1		10	1	10	10
55-65				[2] 2	[4] 4	[6] 2	8	2	16	32
65-75					[6] 1	[9] 2	3	3	9	27
TOTAL f	3	17	14	9	6	4	53	TOTAL	16	92
v	−2	−1	0	1	2	3	TOTAL			
fv	−6	−17	0	9	12	12	10			
fv^2	12	17	0	9	24	36	98			

$u = \frac{X-(35\text{-}45)}{10}$

$v = \frac{Y-(35\text{-}45)}{10}$

SMALL SQUARES REPRESENT PRODUCT uv

$\Sigma f u=16,\ \Sigma f u^2=92,\ \Sigma f v=10,\ \Sigma f v^2=98,\ \Sigma f=53$

$$\Sigma f u v=(4\times 1)+(2\times 1)+(2\times 2)+(1\times 12)+(1\times 6)+(2\times 1)+(2\times 2)+(4\times 4)+(6\times 2)+(6\times 1)+(9\times 2)$$

$$=4+2+4+12+6+2+4+16+12+6+18$$

$$=86$$

$$r=\frac{\Sigma f u v-\frac{\Sigma f u\,\Sigma f v}{\Sigma f}}{\sqrt{\Sigma f u^2-\frac{(\Sigma f u)^2}{\Sigma f}}\ \sqrt{\Sigma f v^2-\frac{(\Sigma f v)^2}{\Sigma f}}}$$

$$=\frac{86-\frac{16\times 10}{53}}{\sqrt{92-\frac{16^2}{53}}\ \sqrt{98-\frac{10^2}{53}}}$$

$$=\frac{82{\cdot}98}{\sqrt{87{\cdot}17}\ \sqrt{96{\cdot}1}}$$

$$=\frac{82{\cdot}98}{\sqrt{8377{\cdot}04}}$$

$$=\frac{82{\cdot}98}{91{\cdot}5}$$

$$=\cdot 906$$

$$\therefore\ r=\cdot 906$$

✓**Problem 218.**—The correlation table given below shows for each of 78 towns (1) measures of the amount of over-crowding present in a given year and (2) the infant mortality rate in the same year. Calculate the coefficient of correlation between over-crowding and infant mortality rate :

Infant mortality rate	*Percentage of population in families living more than two persons per room*					
	1·5—	4·5—	7·5—	10·5—	13·5—	16·5—19·5
36—	5	...	...	...	...	...
46—	9	1	...	...	...	...
56—	10	4	1	...	...	1
66—	4	7	5	2	...	...
76—	2	5	4	1	1	...
86—	...	2	2	3	...	1
96—	...	1	2	2	1	1
106—115	...	1	...	1	...	...

(*I.A.S. 1949 (II)*)

Solution :

Using the step deviation method for grouped datas we have :

X \ Y	1·5-	4·5-	7·5-	10·5-	13·5-	16·5-19·5	TOTAL f	u	fu	fu^2
36-	[4] 5						5	-4	-20	80
46-	[3] 9	[0] 1					10	-3	-30	90
56-	[2] 10	[0] 4	[-2] 1			[-8] 1	16	-2	-32	64
66-	[1] 4	[0] 7	[-1] 5	[-2] 2			18	-1	-18	18
76-	[0] 2	[0] 5	[0] 4	[0] 1	[0] 1		13	0	0	0
86-		[0] 2	[1] 2	[2] 3		[4] 1	8	1	8	8
96-		[0] 1	[2] 2	[4] 2	[6] 1	[8] 1	7	2	14	28
106-115		[0] 1		[6] 1			2	3	6	18
TOTAL f	30	21	14	9	2	3	79	TOTAL	-72	306
v	-1	0	1	2	3	4	TOTAL			
fv	-30	0	14	18	6	12	20			
fv^2	30	0	14	36	18	48	146			

$u = \dfrac{X-(76-)}{10}$

$v = \dfrac{Y-(4\cdot5-)}{3}$

SMALL SQUARES REPRESENT PRODUCT uv

$\Sigma f u = -72$, $\Sigma f u^2 = 306$, $\Sigma f v = 20$, $\Sigma f v^2 = 146$, $\Sigma f = 79$

$$\Sigma f u v = (4\times5)+(3\times9)+(2\times10)+(-2\times1)+(-8\times1)+(1\times4)$$
$$+(-1\times5)+(-2\times2)+(1\times2)+(2\times3)+(4\times1)+(2\times2)$$
$$+(4\times2)+(6\times1)+(8\times1)+(6\times1)$$
$$=20+27+20-2-8+4-5-4+2+6+4+4+8+6+8+6$$
$$=96$$

$$\therefore \quad r = \frac{\Sigma f u v - \dfrac{\Sigma f u \, \Sigma f v}{\Sigma f}}{\sqrt{\Sigma f u^2 - \dfrac{(\Sigma f u)^2}{\Sigma f}} \sqrt{\Sigma f v^2 - \dfrac{(\Sigma f v)^2}{\Sigma f}}}$$

$$= \frac{96 - \frac{-72 \times 20}{79}}{\sqrt{306 - \frac{72^2}{79}} \sqrt{146 - \frac{20^2}{79}}}$$

$$= \frac{114 \cdot 2}{\sqrt{240 \cdot 4} \sqrt{140 \cdot 94}}$$

$$= \frac{114 \cdot 2}{\sqrt{33881 \cdot 976}}$$

$$= \frac{114 \cdot 2}{154 \cdot 5}$$

$$= \cdot 73$$

$\therefore r = \cdot 739$

Problem 219.—Define Pearson's coefficient of correlation and deduce the suitable formula for calculating it from a correlation table. Calculate its value for the following table :—

x/y	16—18	18—20	20—22	22—24
10—20	2	1	1	
20—30	3	2	3	2
30—40	3	4	5	6
40—50	2	2	3	4
50—60		1	2	2
60—70		1	2	1

(*M.Sc. Agra, 1956*)

Solution :

Applying the step-deviation method for grouped data we have :—

X \ Y	16-18	18-20	20-22	22-24	TOTAL f	STEP DEVIATION u	fu	fu^2
10-20	[4] 2	[2] 1	[0] 1		4	-2	-8	16
20-30	[2] 3	[1] 2	[0] 3	[-1] 2	10	-1	-10	10
30-40	[0] 3	[0] 4	[0] 5	[0] 6	18	0	0	0
40-50	[-2] 2	[-1] 2	[0] 3	[1] 4	11	1	11	11
50-60		[-2] 1	[0] 2	[2] 2	5	2	10	20
60-70		[-3] 1	[0] 2	[3] 1	4	3	12	36
TOTAL f	10	11	16	15	52	TOTAL	15	93
STEP DEVIATION v	-2	-1	0	1	TOTAL			
fv	-20	-11	0	15	-16			
fv^2	40	11	0	15	66			

$u = \dfrac{X-(30-40)}{10}$

$v = \dfrac{Y-(20-22)}{2}$

SMALL SQUARES REPRESENTS THE PRODUCT uv

$\Sigma fu = 15,\ \Sigma fu^2 = 93,\ \Sigma fv = -16,\ \Sigma fv^2 = 66,\ \Sigma f = 52$

$$\Sigma fuv = (4\times2)+(2\times1)+(2\times3)+(1\times2)+(-1\times2)+(-2\times2)$$
$$+(-1\times2)+(1\times4)+(-2\times1)+(2\times2)+(-3\times1)$$
$$+(3\times1)$$

$$=8+2+6+2-2-4-2+4-2+4-3+3$$
$$=16$$

$$r=\frac{\Sigma fuv-\dfrac{\Sigma fu\ \Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2-\dfrac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2-\dfrac{(\Sigma fv)^2}{\Sigma f}}}$$

$$=\frac{16-\dfrac{15\times-16}{52}}{\sqrt{93-\dfrac{15^2}{52}}\sqrt{66-\dfrac{16^2}{52}}}$$

$$=\frac{20\cdot6}{\sqrt{5419\cdot57}}$$

$$=\frac{20\cdot6}{73\cdot6}$$

$$=\cdot28$$

$\therefore \quad r=\cdot28.$

✓Problem 220.—Find the value of Pearson's coefficient of correlation, r, for the following table :

Y \ X	94·5	96·5	98·5	100·5	102·5	104·5	106·5	108·5	110·5
29·5			4	3		4	1		1
59·5	1	3	6	18	6	9	2	3	1
89·5	7	3	16	16	4	4	1		1
119·5	5	9	10	9	2		1	2	
149·5	3	5	8	1		1			
179·5	4	2	3	1					
209·5	4	4		1					
239·5	1	1							

(*M.Sc. Agra 1953 ; M.Sc. Agra, 1955*)

Solution :

Using the step-deviation method we have the table :

Y \ X	94·5	96·5	98·5	100·5	102·5	104·5	106·5	108·5	110·5	TOTAL f	v	fv	fv^2
29·5		[4] 4	[2] 3		[-2] 4	[-4] 1			[-10] 1	13	-2	-26	52
59·5	[3] 1	[2] 3	[1] 6	[0] 18	[-1] 6	[-2] 9	[-3] 2	[-4] 3	[-5] 1	49	-1	-49	49
89·5	[0] 7	[0] 3	[0] 16	[0] 16	[0] 4	[0] 4	[0] 1		[0] 1	52	0	0	0
119·5	[-3] 5	[-2] 9	[-1] 10	[0] 9	[1] 2		[3] 1	[4] 2		38	1	38	38
149·5	[-6] 3	[-4] 5	[-2] 8	[0] 1		[4] 1				18	2	36	72
179·5	[-9] 4	[-6] 2	[-3] 3	[0] 1						10	3	30	90
209·5	[-12] 4	[-8] 4		[0] 1						9	4	36	144
239·5	[-15] 1	[-10] 1								2	5	10	50
TOTAL f	25	31	46	46	16	15	4	5	3	191	TOTAL	75	495
u	-3	-2	-1	0	1	2	3	4	5	TOTAL			
fu	-75	-62	-46	0	16	30	12	20	15	-90			
fu^2	225	124	46	0	16	60	36	80	75	662			

$v = \dfrac{Y - 89 \cdot 5}{30}$

$u = \dfrac{X - 100 \cdot 5}{2}$

$\Sigma fu = -90,\ \Sigma fu^2 = 662,\ \Sigma fv = 75,\ \Sigma fv^2 = 495,\ \Sigma f = 191$

$$\begin{aligned}\Sigma fuv = &(4\times4)+(2\times3)+(-2\times4)+(-4\times1)+(-10\times1)+(3\times1)\\ &+(2\times3)+(1\times6)+(-1\times6)+(-2\times9)+(-3\times2)\\ &+(-4\times3)+(-5\times1)+(-3\times5)+(-2\times9)+(-1\times10)\\ &+(1\times2)+(3\times1)+(4\times2)+(-6\times3)+(-4\times5)+(-2\times8)\\ &+(4\times1)+(-9\times4)+(-6\times2)+(-3\times3)+(-12\times4)\\ &+(-8\times4)+(-15\times1)+(-10\times1)\end{aligned}$$

$$\begin{aligned}&= 16+6-8-4-10+3+6+6-6-18-6-12-5-15-18\\ &\quad -10+2+3+8-18-20-16+4-36-12-9-48-32\\ &\quad -15-10\end{aligned}$$

$$= -274$$

$$r = \frac{\Sigma fuv - \dfrac{\Sigma fu\ \Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2 - \dfrac{(\Sigma fu)^2}{f}}\sqrt{\Sigma fv^2 - \dfrac{(\Sigma fv)^2}{\Sigma f}}}$$

$$= \frac{-274 - \frac{-90 \times 75}{191}}{\sqrt{662 - \frac{90^2}{191}}\sqrt{495 - \frac{75^2}{191}}}$$

$$= \frac{-274 + 35{\cdot}3}{\sqrt{619{\cdot}6}\,\sqrt{465{\cdot}55}}$$

$$= \frac{-238{\cdot}7}{\sqrt{288454{\cdot}78}}$$

$$= \frac{-238{\cdot}7}{537{\cdot}08}$$

$$= -{\cdot}44$$

$$\therefore\ r = -{\cdot}44.$$

✓**Problem 221.**—Calculate the co-efficient of correlation between the monthly income and the expenditure on food from the correlation table given below for certain 43 working class families in Kanpur. Arrange your numerical work in a tabular form and explain it where necessary.

Monthly income

Monthly expenditure on food in Rs.	24	28	32	36	40	44	48	52	Total
12	1	1							2
15		2		4					6
18		3	2	5		1			11
21			3	3	4	2			12
24				1	3	4			8
27							2	1	3
30								1	1
Total	1	6	5	13	7	7	[illegible]	2	43

(*M. A. Raj. 1949*)

Solution :

Applying the step-deviation method for grouped datas we have :—

X \ Y	24	28	32	36	40	44	48	52	TOTAL f	u	fu	fu^2
12	[9] 1	[6] 1							2	-3	-6	18
15		[4] 2		[0] 4					6	-2	-12	24
18		[2] 3	[1] 2	[0] 5		[-2] 1			11	-1	-11	11
21			[0] 3	[0] 3	[0] 4	[0] 2			12	0	0	0
24				[0] 1	[1] 3	[2] 4			8	1	8	8
27							[6] 2	[8] 1	3	2	6	12
30								[12] 1	1	3	3	9
TOTAL f	1	6	5	15	7	7	2	2	43	TOTAL	-12	82
v	-3	-2	-1	0	1	2	3	4	TOTAL			
fv	-3	-12	-5	0	7	14	6	8	15			
fv^2	9	24	5	0	7	28	18	32	123			

$u = \frac{X-21}{3}$

$v = \frac{Y-36}{4}$

$$\Sigma fu = -12,\ \Sigma fu^2 = 82,\ \Sigma fv = 15,\ \Sigma fv^2 = 123,\ \Sigma f = 43$$

$$\Sigma fuv = (9\times1)+(6\times1)+(4\times2)+(2\times3)+(1\times2)+(-2\times1)+(1\times3)+(2\times4)+(6\times2)+(8\times1)+(12\times1)$$

$$= 9+6+8+6+2-2+3+8+12+8+12$$

$$= 72.$$

$$r = \frac{\Sigma fuv - \frac{\Sigma fu . \Sigma f.v}{\Sigma f}}{\sqrt{\Sigma fu^2 - \frac{(\Sigma fu)^2}{\Sigma f}}\ \sqrt{\Sigma fv^2 - \frac{(\Sigma fv)^2}{\Sigma f}}}$$

$$= \frac{72 - \frac{15\times-12}{43}}{\sqrt{82 - \frac{12^2}{43}}\sqrt{123 - \frac{15^2}{43}}} = \frac{76{\cdot}19}{\sqrt{9270{\cdot}86}}$$

$$= \frac{76{\cdot}19}{96{\cdot}2} = {\cdot}79$$

$\therefore r = {\cdot}79.$

✓**Problem 222.**—The following table gives the frequency, according to age-groups, of marks obtained by 75 students in an intelligence test :—

Test marks	Age in Years					TOTAL
	19	20	21	22	23	
0 to 20	4	4	2	1	1	12
20 to 40	3	5	4	2	2	16
40 to 60	3	6	8	5	3	25
60 to 80		4	6	8	4	22
TOTAL	10	19	20	16	10	75

Calculate the coefficient of correlation between age and intelligence. (*M.A. Raj.*, 1955)

Solution :

X \ Y	19	20	21	22	23	TOTAL f	u	fu	fu²
0-20	[4] 4	[2] 4	[0] 2	[-2] 1	[-4] 1	12	-2	-24	48
20-40	[2] 3	[1] 5	[0] 4	[-1] 2	[-2] 2	16	-1	-16	16
40-60	[0] 3	[0] 6	[0] 8	[0] 5	[0] 3	25	0	0	0
60-80		[-1] 4	[0] 6	[1] 8	[2] 4	22	1	22	22
TOTAL f	10	19	20	16	10	75	TOTAL	-18	86
v	-2	-1	0	1	2	TOTAL			
fv	-20	-19	0	16	20	-3			
fv²	40	19	0	16	40	115			

$$u = \frac{X-(40\text{-}60)}{20}$$

$$v = Y - 21$$

$\Sigma fu = -18,\quad \Sigma fu^2 = 86,\quad \Sigma fu = -3,\quad \Sigma fv^2 = 115,\quad \Sigma f = 75$

$$\Sigma fuv = (4\times4)+(2\times4)+(-2\times1)+(-4\times1)+(2\times3)+(1\times5)+(-1\times2)+(-2\times2)+(-1\times4)+(1\times8)+(2\times4)$$

$$=16+8-2-4+6+5-2-4-4+8+8$$

$$=35$$

$$r=\frac{\Sigma fuv-\dfrac{\Sigma fu\,\Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2-\dfrac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2-\dfrac{(\Sigma fv)^2}{\Sigma f}}}$$

$$=\frac{35-\dfrac{-18\times-3}{75}}{\sqrt{86-\dfrac{18^2}{75}}\sqrt{115-\dfrac{3^2}{75}}}$$

$$=\frac{35-\cdot72}{\sqrt{81\cdot68}\quad\sqrt{144\cdot88}}$$

$$=\frac{34\cdot28}{\sqrt{11833\cdot7984}}$$

$$=\frac{34\cdot28}{108\cdot78}$$

$$=\cdot31$$

$\therefore\ r=\cdot31.$

Problem 223.—Prove that Pearson's, coefficient of correlation cannot numerically exceed unity. How far is r a true measure of correlation between two variables? Calculate the coefficient of correlation between the ages of husbands and wives from the following table and find its probable error.

Ages of husbands

Ages of wives	20—30	30—40	40—50	50—60	60—70	TOTAL
15—25	5	9	3			17
25—35		10	25	2		37
35—45			12	2		15
45—55			4	16	5	25
55—65				4	2	6
TOTAL	5	20	44	24	7	100

(*P.C.S. 1938*)

Solution :

X \ Y	20-30	30-40	40-50	50-60	60-70	TOTAL f	u	fu	fu²
15-25	[2] 5	[1] 9	[0] 3			17	-1	-17	17
25-35		[0] 10	[0] 25	[0] 2		37	0	0	0
35-45			[0] 1	[1] 12	[2] 2	15	1	15	15
45-55			[0] 4	[2] 16	[4] 5	25	2	50	100
55-65				[3] 4	[6] 2	6	3	18	54
TOTAL f	5	20	44	24	7	100	TOTAL	66	186
v	-2	-1	0	1	2	TOTAL			
fv	-10	-20	0	24	14	8			
fv²	20	20	0	24	28	92			

$$u=\frac{X-(25\text{-}35)}{10}$$

$$v=\frac{Y-(40\text{-}50)}{10}$$

$\Sigma fu=66,\ \Sigma fu^2=186,\ \Sigma fv=8,\ \Sigma fv^2=92,\ \Sigma f=100$

$$\Sigma fuv=(2\times5)+(1\times9)+(1\times12)+(2\times2)+(2\times16)+(4\times5)+(3\times4)+(6\times2)$$

$$=10+9+12+4+32+20+12+12$$

$$=\mathbf{111}$$

$$r=\frac{\Sigma fuv-\frac{\Sigma fu\ \Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2-\frac{(\Sigma fu)^2}{\Sigma f}}\ \sqrt{\Sigma fv^2-\frac{(\Sigma fv)^2}{\Sigma f}}}$$

$$=\frac{111-\frac{66\times8}{100}}{\sqrt{186-\frac{66^2}{100}}\sqrt{92-\frac{8^2}{100}}}$$

$$=\frac{111-5{\cdot}28}{\sqrt{3186-4{\cdot}56}\ \sqrt{92-{\cdot}64}}$$

$$= \frac{105{\cdot}72}{\sqrt{142{\cdot}44 \times 91{\cdot}36}}$$

$$= \frac{105{\cdot}72}{\sqrt{13013{\cdot}3184}}$$

$$= \frac{105{\cdot}72}{114{\cdot}07}$$

$$= {\cdot}92$$

$$\therefore \quad r = {\cdot}92$$

Probable Error of the coefficient of correlation (P.E.) is given by the formula

$$\text{P.E.} = {\cdot}6745 \; \frac{1 - r^2}{\sqrt{\Sigma f}}$$

$$= {\cdot}6745 \; \frac{1 - ({\cdot}92)^2}{\sqrt{100}}$$

$$= {\cdot}6745 \times \frac{1 - {\cdot}8464}{10}$$

$$= {\cdot}6745 \times {\cdot}01536$$

$$= {\cdot}0104 \text{ (appr.)}$$

Hence probable error $= {\cdot}0104$ (appr.)

Problem 224.—The following table gives the frequency according to age groups of marks obtained by 67 students in an intelligence test :—

	Age in years				
Test Marks	18	19	20	21	Total
200—250	4	4	2	1	11
250—300	3	5	4	2	14
300—350	2	6	8	5	21
350—400	1	4	6	10	21
TOTAL	10	19	20	18	67

Is there any relationship between age and intelligence ?

(*B. Com. Agra, 1942*)

Solution :

X \ Y	18	19	20	21	TOTAL f	u	fu	fu²
200–250	[4] 4	[2] 4	[0] 2	[−2] 1	11	−2	−22	44
250–300	[2] 3	[1] 5	[0] 4	[−1] 2	14	−1	−14	14
300–350	[0] 2	[0] 6	[0] 8	[0] 5	21	0	0	0
350–400	[−2] 1	[−1] 4	[0] 6	[1] 10	21	1	21	21
TOTAL f	10	19	20	18	67	TOTAL	−15	79
v	−2	−1	0	1	TOTAL			
fv	−20	−19	0	18	−21			
fv²	40	19	0	18	77			

$u = \frac{X-(300\text{-}350)}{50}$

$v = Y - 20$

$\Sigma fu = -15,\ \Sigma fu^2 = 79,\ \Sigma fv = -21,\ \Sigma fv^2 = 77,\ \Sigma f = 67$

$$\Sigma fuv = (4\times4)+(2\times4)+(-2\times1)+(2\times3)+(1\times5)+(-1\times2)+(-2\times1)+(-1\times4)+(1\times10)$$

$$=16+8-2+6+5-2-2-4+10$$

$$=35$$

$$r = \frac{\Sigma fuv - \frac{\Sigma fu\,\Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2 - \frac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2 - \frac{(\Sigma fv)^2}{\Sigma f}}}$$

$$= \frac{35 - \frac{-15\times-21}{67}}{\sqrt{79-\frac{15^2}{67}}\sqrt{77-\frac{21^2}{67}}}$$

$$= \frac{30\cdot3}{\sqrt{75\cdot65}\ \sqrt{70\cdot42}}$$

$$= \frac{30\cdot3}{\sqrt{5327\cdot27}}$$

$$= \frac{30\cdot3}{72\cdot9}$$

$$= \cdot41$$

$$\text{Probable error (P.E.)} = \cdot6745\ \frac{1-r^2}{\sqrt{\Sigma f}}$$

$$= \cdot6745\ \frac{1-(\cdot41)^2}{\sqrt{67}}$$

$$= \cdot6745\ \frac{\cdot82}{8\cdot18}$$

$$= \cdot06745 \text{ (appr.)}$$

Since $\cdot41 > 6 \times \cdot06745$

i.e., $r > 6$ P.E.

Hence the correlation is significant *i.e.*, its presence is strongly suggested.

Problem 225.—The following table gives the number of students having different heights and weights.

Height in Inches	Weight in pounds					Total
	80—90	90—100	100—110	110—120	120—130	
50—55	1	3	7	5	2	18
55—60	2	4	10	7	4	27
60—65	1	5	12	10	7	35
65—70		3	8	6	3	20
Total	4	15	37	28	16	100

Do you find any relation between height and weight?

(B. Com. All., 1940)

Solution :

Y / X	80-90	90-100	100-110	110-120	120-130	TOTAL f	u	fu	fu^2
50-55	[4] 1	[2] 3	[0] 7	[-2] 5	[-4] 2	18	-2	-36	72
55-60	[2] 2	[1] 4	[0] 10	[-1] 7	[-2] 4	27	-1	-27	27
60-65	[0] 1	[0] 5	[0] 12	[0] 10	[0] 7	35	0	0	0
65-70		[-1] 3	[0] 8	[1] 6	[2] 3	20	1	20	20
TOTAL f	4	15	37	28	16	100	TOTAL	-43	119
v	-2	-1	0	1	2	TOTAL			
fv	-8	-15	0	28	32	37			
fv^2	16	15	0	28	64	123			

$$u = \frac{X-(60\text{-}65)}{5}$$

$$v = \frac{Y-(100\text{-}110)}{10}$$

$$\Sigma f = 100,\ \Sigma fu = -43,\ \Sigma fu^2 = 119,\ \Sigma fv = 37,\ \Sigma fv^2 = 123.$$

$$\Sigma fuv = (4\times1)+(2\times3)+(-2\times5)+(-4\times2)+(2\times2)+(1\times4)$$
$$+(-1\times7)+(-2\times4)+(-1\times3)+(1\times6)+(2\times3)$$

$$= 4+6-10-8+4+4-7-8-3+6+6$$

$$= -6$$

$$r = \frac{\Sigma fuv - \dfrac{\Sigma fu\ \Sigma fv}{\Sigma f}}{\sqrt{\Sigma fu^2 - \dfrac{(\Sigma fu)^2}{\Sigma f}}\sqrt{\Sigma fv^2 - \dfrac{(\Sigma fv)^2}{\Sigma f}}}$$

$$= \frac{-6 - \dfrac{-43\times37}{100}}{\sqrt{119 - \dfrac{43^2}{100}}\sqrt{123 - \dfrac{37^2}{100}}}$$

$$= \frac{9{\cdot}91}{\sqrt{101{\cdot}51}\sqrt{109{\cdot}31}}$$

$$= \frac{9{\cdot}91}{\sqrt{11096{\cdot}0581}}$$

$$= \frac{9{\cdot}91}{105{\cdot}33}$$

$$= {\cdot}094$$

$$\text{P.E.} = {\cdot}6745\ \frac{1-({\cdot}094)^2}{\sqrt{100}}$$

$$= {\cdot}6745\ \frac{1-{\cdot}0088}{10}$$

$$= {\cdot}6745 \times {\cdot}09912$$

$$= {\cdot}067$$

Since r is not greater than 6 P.E. hence there is no relationship between the heights and weights. Also we see that $r < {\cdot}3$ hence the fact is further proved.

Problem 226.—Calculate r between production of Pig Iron (percentage of trend, 1897—1913) and Industrial Production (percentage of trend, 1897—1913) for the following table :—

Pig Iron Production

Industrial Production	50—60	60—70	70—80	80—90	90—100	100—110	110—120	120—130	TOTAL
120—130								15	15
110—120						6	34	1	41
100—110					5	51	6		62
90—100				3	33	1			37
80—90			2	24	3				29
70—80			7	2					9
60—70		2	1						3
50—60	6	2							8
TOTAL	6	4	10	29	41	58	40	16	204

(*M.A. Cal.*, *1936*)

Solution :

X \ Y	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	TOTAL f	u	fu	fu²
120-130								[6] 15	15	2	30	60
110-120						[1] 6	[2] 34	[3] 1	41	1	41	41
100-110					[0] 5	[0] 51	[0] 6		62	0	0	0
90-100				[1] 3	[0] 33	[-1] 1			37	-1	-37	37
80-90			[4] 2	[2] 24	[0] 3				29	-2	-58	116
70-80			[6] 7	[3] 2					9	-3	-27	81
60-70		[12] 2	[8] 1						3	-4	-12	48
50-60	[20] 6	[15] 2							8	-5	-40	200
TOTAL f	6	4	10	29	41	58	40	16	204	TOTAL	-103	583
v	-4	-3	-2	-1	0	1	2	3	TOTAL			
fv	-24	-12	-20	-29	0	58	80	48	101			
fv²	96	36	40	29	0	58	160	144	563			

$$u=\frac{X-(100\text{-}110)}{10} \qquad v=\frac{Y-(90\text{-}100)}{10}$$

$\Sigma f u=-103,\ \Sigma f u^2=583,\ \Sigma f v=101,\ \Sigma f v^2=563,\ \Sigma f=204$

$$\begin{aligned}\Sigma f uv &= (6\times15)+(1\times6)+(2\times34)+(3\times1)+(1\times3)+(-1\times1)\\ &\quad+(4\times2)+(2\times24)+(6\times7)+(3\times2)+(12\times2)+(8\times1)\\ &\quad+(20\times6)+(15\times2)\\ &=90+6+68+3+3-1+8+48+42+6+24+8+120+30\\ &=455\end{aligned}$$

$$r=\frac{\Sigma f u v-\dfrac{\Sigma f u\,\Sigma f v}{\Sigma f}}{\sqrt{\Sigma f u^2-\dfrac{(\Sigma f u)^2}{\Sigma f}}\ \sqrt{\Sigma f v^2-\dfrac{(\Sigma f v)^2}{\Sigma f}}}$$

$$r=\frac{455-\dfrac{-103\times101}{204}}{\sqrt{583-\dfrac{103^2}{204}}\ \sqrt{563-\dfrac{101^2}{204}}}$$

$$=\frac{455+50\cdot99}{\sqrt{531}\ \sqrt{513}}$$

$$=\frac{506}{\sqrt{272403}}$$

$$= \frac{506}{521 \cdot 9}$$
$$= \cdot 96$$
$$\therefore \quad r = \cdot 96$$

Problem 227.—Calculate the co-efficient of concurrent deviations from the following data :

n or number of pairs of observations = 47
c or number of concurrent deviations = 16

[*Statistics—Ghosh and Choudhary*, pp. 439].

Solution :

Co-efficient of concurrent deviation (r)

$$= \pm \sqrt{\pm \frac{(2c-n)}{n}}$$
$$= \pm \sqrt{\pm \frac{(32-47)}{47}}$$
$$= -\sqrt{-\left(-\frac{15}{47}\right)}$$
$$= -\sqrt{-(-\cdot 31)}$$
$$= -\cdot 55 \text{ Ans.}$$

Problem 228.—Calculate the coefficient of correlation from the following data :

Year	*Supply*	*Price*
1921	80	146
1922	82	140
1923	86	130
1924	91	117
1925	83	133
1926	85	127
1927	89	115
1928	96	95
1929	93	100

(*M. Com. Alld., 1943*)

Solution :

Calculation of coefficient of correlation by the method of concurrent deviations—

Year	Supply (X)		Price (Y)	
	Supply	Deviation signs from the preceding year	Price	Deviation signs from the preceding year
1921	80		146	
1922	82	+	140	−
1923	86	+	130	−
1924	91	+	117	−
1925	83	−	133	+
1926	85	+	127	−
1927	89	+	115	−
1928	96	+	95	−
1929	93	−	100	+

Coefficient of Concurent deviation $(r) = \pm\sqrt{\pm\frac{(2c-n)}{n}}$

where c = concurent deviation = 0

n = total number of observations.

$$= \pm\sqrt{\pm\frac{(0-9)}{9}}$$

$$= -\sqrt{-\frac{(-9)}{9}}$$

$$= -\sqrt{-\frac{(-9)}{9}}$$

$$= -\sqrt{+1} \text{ approximately}$$

$$= -1 \text{ approximately.}$$ **Ans.**

✓**Problem 229.**—Calculate r from the following table :

Year	Average No. of daily labourers (000's)	Lakhs of bales consumed by mills
1925	368	22
1926	384	21
1927	385	24
1928	361	20
1929	347	22
1930	384	26
1931	395	26
1932	403	29
1933	400	28
1934	385	27

(*B. Com., Agra, 1941*)

Solution :

Year	Average daily No. labourers (000)	Deviations over the preceding year	Lakhs of bales consumed by mills	Deviation over the preceding year
1925	368		22	
1926	384	+	21	−
1927	385	+	24	+
1928	361	−	20	−
1929	347	−	22	+
1930	384	+	26	+
1931	395	+	26	
1932	403	+	29	+
1933	400	−	28	−
1934	385	−	27	−

Co-efficient of concurrent deviation $(r) = \pm\sqrt{\pm\frac{(2c-n)}{n}}$

$$=\pm\sqrt{\pm\frac{(12-10)}{10}},$$ where c = Concurrent deviations = 6

n — total number of observations = 10

$$=\pm\sqrt{\pm\frac{(2)}{10}}$$

$$=+\sqrt{+\tfrac{1}{5}}$$

$$=\sqrt{\cdot 2}$$

$=\sqrt{\cdot 2}$ approximately. **Ans.**

Problem 230.—Ten students got the following percentage of marks in Principles of Economics and Statistics :—

Students	1	2	3	4	5	6	7	8	9	10
Marks in Economics	78	36	98	25	75	82	90	62	65	39
Marks in Statistics	84	51	91	60	68	62	86	58	53	47

Calculate Rank Correlation Co-efficient.

(*M.A., Agra, 1951*)

Solution :

In this case, the total number of observations (n) = 10.

Rank Correlation is calculated by the formula $\rho = 1-\frac{6\Sigma d^2}{n^3-n}$

$$\Sigma d^2=[(78-84)+(36-51)+(98-91)+(25-60)+(75-68) + (82-62)+(90-86)+(62-58)+(65-53)+(39-47)]^2$$

$$=[-6-15+7-35+7+20+4+4+12-8]^2$$

$$=[-64+54]^2$$

$$=[-10]^2$$

$$=100$$

Putting the value of Σd^2 in the formula we get

$$\rho=1-\frac{6\times 100}{1000-10}$$ where ρ = Rank Correlation.

$$=1-\frac{600}{990}$$

$$=1-\frac{20}{33}$$

$$=\frac{33-20}{33}$$

$$=\frac{13}{33}$$

$= +\cdot 39$ approximately.

Problem 231.—Ten competitors in a beauty contest are ranked by three judges in the following orders :

First Judge :	1	6	5	10	3	2	4	9	7	8
Second Judge :	3	5	8	4	7	10	2	1	6	9
Third Judge :	6	4	9	8	1	2	3	10	5	7

Use the Rank Correlation Coefficient to discuss which pair of judges have the nearest approach to common tastes in beauty.

(M.A., Alld. 1952)

Solution :

In this question, we have to calculate three sets of rank correlation coefficient with the following groups :

(*a*) First and Second Judge. (*b*) First and Third Judge. (*c*) Second and Third Judge.

First Judge's Opinion		Second Judge's Opinion		Third Judge's Opinion	
X	Ranks	Y	Ranks	Z	Ranks
1	10	3	8	6	5
6	5	5	6	4	7
5	6	8	3	9	2
10	1	4	7	8	3
3	8	7	4	1	10
2	9	10	1	2	9
4	7	2	9	3	8
9	2	1	10	10	1
7	4	6	5	5	6
8	3	9	2	7	4

[A] Renk Coefficient Correlation between the opinions of first and second judges :

$$\Sigma d^2 = [(10-8)+(5-6)+(6-3)+(1-7)+(8-4)+(9-1)+(7-9)+(2-10)+(4-5)+(3-2)]^2$$

$$=[-2-1+3-6+4+8-2-8-1+1]^2$$

$$=[16-22]^2$$

$$=[-6]^2$$

$$=36$$

$$\rho = 1-\frac{6\Sigma d^2}{n^3-n}$$ where ρ = Coefficient of rank correlation.

$$=1-\frac{6\times 36}{1000-10}$$

$$=1-\frac{216}{990}$$

$$=\frac{990-216}{990}$$

$=\cdot 7$ approximately.

[B] Rank Coefficient Correlation between the opinion of first and third judges :

$$\Sigma d^2=[(10-5)+(5-7)+(6-2)+(1-3)+(8-10)+(9-9)+(7-8)+(2-1)+(4-6)+(3-4)]^2$$

$$=[5-2+4-2-2+0-1+1-2-1]^2$$

$$=[10-10]^2$$

$=0$ (There is no difference in their approach)

[C] Rank Coefficient Correlation between the opinion of Second and third judges

$$\Sigma d^2=[(8-5)+(6-7)+(3-2)+(7-3)+(4-10)+(1-9)+(9-8)+(10-1)+(5-6)+(2-4)]^2$$

$$=[3-1+1+4-6-8+1+9-1-2]^2$$

$$=[18-18]^2$$

$=0$ (There is no difference in their approach)

Thus, we conclude that the [B] and [C] pairs of judges have the least disparity in their approach to beauty.

First and Second Judge

d = [illegible]

Σd^2 = 4 + 1 + 9 + 36 + 16 + 64 + [illegible] + 64 + 1 + 1 = 200

$\rho = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$

$= 1 - \frac{6 \times 200}{10(100-1)}$

$= 1 - \frac{40}{33} = -\frac{7}{33} = -\cdot 23$

First & third $\rho = +0\cdot 64$

Second & third $\rho = -0\cdot 20$

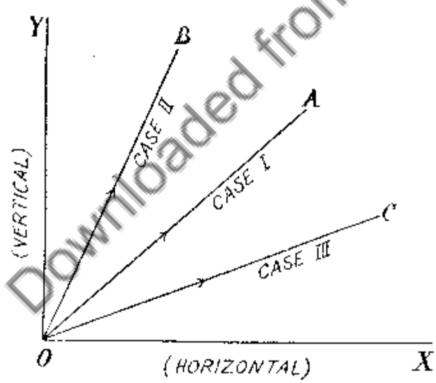
Ans. from Yule & Kendall Book

CHAPTER VII

REGRESSION

There may exist almost perfect correlation between the two series, but the proportional movements in the two may be very different. "In many cases, a measure of this proportional variation can be usefully employed and proportional variation for both series having been obtained, comparison of the two by means of a ratio gives us the ratio of variation". This line of variation may be regular, or irregular. As the economic forces are always dynamic, so the line of variation is almost always irregular. This line of variation may be plotted on a graph, known as the 'Galton Graph', or it may be calculated mathematically.

"If in the Galton graph both the subject and relative change by equal percentages, then the ratio of variation is equal to unity (one) and the line drawn through the plotted points will be a line at an angle of 45° to the horizontal (case I). When the relative shows a tendency to change less than the subject, the line will be at an angle less than 45° to the vertical (case II). If the relative changes move proportionally than the subject, the line will be at an angle less than 45° to the horizontal (case III). *This line is known as the Regression Line.* The nearer this regression line approaches the vertical, the slighter the degree of correlation."



∠AOX=45° to horizontal
∠YOB=less than 45° to vertical
∠COX=less than 45° to horizontal

Equations of the Lines of Regression.—

The equation of the line of regression of Y on X is

$y - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x - \bar{x})$ and the equation of the line of regression of X on Y is: $x - \bar{x} = \frac{r\sigma_x}{\sigma_y}(y - \bar{y})$

Where $\bar{x}$ and $\bar{y}$ denote the means of X and Y series.

$$y = \frac{r\sigma_y}{\sigma_x} \cdot x \text{ and } x = \frac{r\sigma_x}{\sigma_y} \cdot y.$$

It is also expressed as :

If X is the independent variable and Y is the dependent variable that is for given value of X the value of Y is to be determined, then Regression coefficient of Y on X is

$$= r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi\Sigma\eta}{n}}{\Sigma\xi^2 - n\left(\frac{\Sigma\xi}{n}\right)^2}$$

Where r is coefficient of correlation ; σ_y, σ_x are standard deviations ; x, y are deviations of the values of X,Y from their respective means M_x and N_y ; and ξ, η are deviations of X, Y from assumed means ; n is number of items.

Also Regression equation of Y on X is

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}(X - M_x)$$

If Y is the independent variable and X is the dependent variable then Regression coefficient of X on Y is

$$= r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi\,\Sigma\eta}{n}}{\Sigma\eta^2 - n\left(\frac{\Sigma\eta}{n}\right)^2}$$

and Regression equation of X on Y is

$$X - M_x = r\frac{\sigma_x}{\sigma_y}(Y - M_y)$$

√**Problem 232.**—Heights of fathers and sons are given in inches :—

Heights of father	65	66	67	67	68	69	71	73
Height of son	67	68	64	68	72	70	69	70

Form the two lines of regression and calculate the expected average height of the son when the height of the father is 67·5 inches.

(*M. Sc. Agra, 1957*)

Solution :— $n=8$

Height of father X	Height of son Y	Deviation of X from assumed mean 68 ξ	Deviation of Y from assumed mean 68 η	$\xi\eta$	ξ^2	η^2
65	67	−3	−1	3	9	1
66	68	−2	0	0	4	0
67	64	−1	−4	4	1	16
67	68	−1	0	0	1	0
68	72	0	4	0	0	16
69	70	1	2	2	1	4
71	69	3	1	3	9	1
73	70	5	2	10	25	4
		$\Sigma\xi=2$	$\Sigma\eta=4$	$\Sigma\xi\eta=22$	$\Sigma\xi^2=50$	$\Sigma\eta^2=42$

$$\text{mean } M_x = 68+\frac{2}{8}=68{\cdot}25$$

$$\text{mean } M_y = 68+\frac{4}{8}=68{\cdot}5$$

$$\text{Now } r\frac{\sigma_y}{\sigma_x}=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\,\Sigma\eta}{n}}{\Sigma\xi^2-n\left(\frac{\Sigma\xi}{n}\right)^2}$$

$$=\frac{22-\frac{2\times 4}{8}}{50-8\left(\frac{2}{8}\right)^2}$$

$$=\frac{21}{49{\cdot}5}={\cdot}424$$

∴ Equation to regression line of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

or $\quad Y-68{\cdot}5={\cdot}424\,(X-68{\cdot}25)$ ... ... (1)

$$\text{Again} \quad r\frac{\sigma_x}{\sigma_y}=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\Sigma\eta}{n}}{\Sigma\eta^2-n\left(\frac{\Sigma\eta}{n}\right)^2}$$

$$=\frac{22-\frac{2\times4}{8}}{42-8\left(\frac{4}{8}\right)^2}$$

$$=\frac{21}{40}=\cdot 525$$

Equation to regression line of X on Y is

$$X-M_x=r\ \frac{\sigma_x}{\sigma_y}(Y-M_y)$$

or $X-68{\cdot}25=\cdot525\ (Y-68{\cdot}5)$ ...(II)

Next to find the height of son (*i.e.* Y) for given height 67·5 of father (*i.e.* for X = 67·5).

Since X is given and Y is to be determined therefore equation (*i*) will be used. Therefore expected height of son is given by

$$Y-68{\cdot}5=\cdot424\ (67{\cdot}5-68{\cdot}25)$$

or $Y\ =68{\cdot}5-\cdot424\times\cdot75$

$Y\ =68{\cdot}18$ inches.

✓**Problem 233.**—A study of wheat prices at Hapur and Karanchi yields the following data :—

	Hapur	*Karanchi*
	Rs.	Rs.
Average price per maund	2·463	2·797
Standard deviation	·326	·207
Coefficient of Correlation between two prices + ·774		

Estimate from the above data the most likely price of wheat (1) at Hapur corresponding to the price Rs. 2·334 at Karanchi (2) at Karanchi corresponding to price of Rs. 3·052 at Hapur.

(*P.C.S. 1938*)
(*Alld. M.Com. 1950*)
(*Agra, M.A., 1953*)

Solution :

(1) To get the price at Hapur we must take it to be the dependent variable Y in the equation.

$$Y-M_y=r\ \frac{\sigma_y}{\sigma_x}(X-M_x)$$

Substituting the given values we get

$$Y-2{\cdot}463=\cdot774\times\frac{\cdot326}{\cdot207}(2{\cdot}334-2{\cdot}797)$$

or $Y\ =1{\cdot}899$, app.

(2) To get the price at Karanchi we must take it to be the dependent variable X in the regression equation

$$X-M_x=r\frac{\sigma_x}{\sigma_y}\ (Y-M_y)$$

Substituting the given values we get

$$X-2{\cdot}797={\cdot}774\frac{{\cdot}207}{{\cdot}326}\ (3{\cdot}052-2{\cdot}463)$$

or $$X=3{\cdot}086.$$

Problem 234.—Explain the concepts of 'correlation' and 'regression' and compare their usefulness.

Calculate the coefficient of correlation and obtain the lines of regression for the following data :—

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

Obtain an estimate of Y which should correspond on the average to X=6·2

(*I.A.S. 1955*)

Solution :

Calculation of coefficient of correlation (r).

X	Y	Deviation of X from mean 5 X	Deviation of Y from mean 12 Y	xy	x^2	y^2
1	9	−4	−3	12	16	9
2	8	−3	−4	12	9	16
3	10	−2	−2	4	4	4
4	12	−1	0	0	1	0
5	11	0	−1	0	0	1
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	12	9	16
9	15	4	3	12	16	9
ΣX=45	ΣY=108			Σxy=57	Σx²=60	Σy²=60

$$\text{mean } M_x=\frac{45}{9}=5$$

$$\text{mean } M_y=\frac{108}{9}=12$$

$$\text{Coefficient of correlation } r=\frac{\Sigma xy}{\sqrt{\Sigma x^2\,\Sigma y^2}}$$

$$=\frac{57}{\sqrt{60\times 60}}$$

$$={\cdot}95$$

Now $$r\,\frac{\sigma_y}{\sigma_x}=\frac{\Sigma xy}{\Sigma x^2}=\frac{57}{60}=\cdot 95$$

and $$r\,\frac{\sigma_x}{\sigma y}=\frac{\Sigma xy}{\Sigma y^2}=\frac{57}{60}=\cdot 95$$

The two lines of regression are

$$Y-M_y=r\,\frac{\sigma_y}{\sigma_x}\,(X-M_x)$$

and $$X-M_x=r\,\frac{\sigma_x}{\sigma_y}\,(Y-M_y)$$

or $$Y-12=\cdot 95(X-5) \quad \ldots \quad (1)$$

and $$X-5=\cdot 95(Y-12) \quad \ldots \quad (2)$$

Now to find Y for given $X=6\cdot 2$. Since Y is dependent variable therefore using the (1) equation,

$$Y-12=\cdot 95(6\cdot 2-5)$$

or $$Y=13\cdot 14$$

✓**Problem 235.**—The following marks have been obtained by a class of students in statistics (out of 100).

Paper I	80	45	55	56	58	60	65	68	70	75	85
Paper II	82	56	50	48	60	62	64	65	70	74	90

Compute the coefficient of correlation for the above data. Find the lines of regression and examine the relationship.

(I.A.S. 1945)

(Agra B.Sc. 1961)

Solution :

Calculation of coefficient of correlation (r).

Paper I (X)	Paper II (Y)	Deviation of X from ass. mean (ξ^{65})	Deviation of Y from ass. mean (η^{65})	$\xi\eta$	ξ^2	η^2
80	82	15	17	255	225	289
45	56	−20	−9	180	400	81
55	50	−10	−15	150	100	225
56	48	−9	−17	153	81	289
58	60	−7	−5	35	49	25
60	62	−5	−3	15	25	9
65	64	0	−1	0	0	1
68	65	3	0	0	9	0
70	70	5	5	25	25	25
75	74	10	9	90	100	81
85	90	20	25	500	400	625
		$\Sigma\xi=2$	$\Sigma\eta=6$	$\Sigma\xi\eta=1403$	$\Sigma\xi^2=1414$	$\Sigma\eta^2=1650$

$\therefore$ mean $M_x = 65 + \frac{2}{11} = 65{\cdot}18$ appr.

mean $M_y = 65 + \frac{6}{11} = 65{\cdot}54$ appr.

Coefficient of correlation (r)

$$= \frac{\Sigma\xi\eta - \frac{\Sigma\xi\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2 - n\left(\frac{\Sigma\xi}{n}\right)^2}\sqrt{\Sigma\eta^2 - n\left(\frac{\Sigma\eta}{n}\right)^2}}$$

or

$$r = \frac{1403 - \frac{2\times 6}{11}}{\sqrt{1414 - 11\left(\frac{2}{11}\right)^2}\sqrt{1651 - 11\left(\frac{6}{11}\right)^2}}$$

$$= \frac{1401{\cdot}91}{\sqrt{1414 - {\cdot}36}\sqrt{1651 - 3{\cdot}27}}$$

$$= \frac{1401{\cdot}91}{\sqrt{1413{\cdot}64 \times 1647{\cdot}73}}$$

$$= \frac{1401{\cdot}91}{1526{\cdot}23}$$

$$= {\cdot}91$$

Now
$$r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi\Sigma\eta}{n}}{\Sigma\xi^2 - n\left(\frac{\Sigma\xi}{n}\right)^2}$$

$$= \frac{1403 - \frac{2\times 6}{11}}{1441 - 11\left(\frac{2}{11}\right)^2}$$

$$= \frac{1401{\cdot}91}{1413{\cdot}64}$$

$$= {\cdot}99$$

Again
$$r\frac{\delta_x}{\delta_y} = \frac{\Sigma\xi\mu - \frac{\Sigma\xi\ \Sigma\eta}{n}}{\Sigma\eta^2 - n\left(\frac{\Sigma\eta}{n}\right)^2}$$

$$= \frac{1403 - \frac{2\times 6}{11}}{1651 - 11\left(\frac{6}{11}\right)^2}$$

$$=\frac{1401{\cdot}91}{1647{\cdot}73}$$

$$={\cdot}85$$

Hence the lines of regression are

$$Y-M_y=r\frac{\sigma_x}{\sigma_x}(X-M_x)$$

and $$X-M_x=r\frac{\sigma_x}{\sigma_y}(Y-M_y)$$

or $$Y-65{\cdot}54={\cdot}99(X-65{\cdot}18)$$

and $$X-65{\cdot}18={\cdot}85(Y-66{\cdot}54)$$

√**Problem 236.**—Explain the terms (*a*) correlation coefficient and (*b*) regression equations. Why should there be two reqression equations ? Write down the two regression equations that may be associated with the following pairs of values :—

(X) :	152	114	138	154	144	153	141	117	156
(Y) :	193	300	414	594	676	549	320	483	481

(*I.A S., 1951*)

Solution :

X	Y	Deviations from ass. mean 140 of X (ξ)	Deviations of Y from ass. mean 445 (η)	$\xi\eta$	ξ^2	η^2
152	193	12	−252	−3024	144	63504
114	300	−26	−145	3770	676	21025
138	414	−2	−41	82	4	1681
154	594	14	149	2086	196	22201
144	676	4	231	924	16	53361
153	549	13	104	1352	169	10816
141	320	1	−125	−125	1	15625
117	483	−23	38	874	529	1444
156	481	16	36	576	256	1296
		$\Sigma\xi=9$	$\Sigma\eta=-5$	$\Sigma\xi\eta=6515$	$\Sigma\xi^2=1991$	$\Sigma\eta^2=190953$

$$\text{mean } M_x=140+\frac{9}{9}=141$$

$$\text{mean } M_y=445+\frac{-5}{9}=444{\cdot}45$$

$$r\frac{\sigma_x}{\sigma_h}=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\;\Sigma\eta}{n}}{\Sigma\eta^2-n\left(\frac{\Sigma\eta}{n}\right)^2}$$

$$=\frac{6615-\frac{9\times -5}{9}}{190953-9\left(\frac{-5}{9}\right)^2}$$

$$=\frac{6520}{190950\cdot23}$$

$$=\cdot034$$

$$r\frac{\sigma_x}{\sigma_x}=\frac{\Sigma\xi\eta-\frac{\Sigma\xi\,\Sigma\eta}{n}}{\Sigma\xi^2-n\left(\frac{\Sigma\xi}{n}\right)^2}$$

$$=\frac{6520}{1991-9\left(\frac{9}{9}\right)^2}$$

$$=\frac{6520}{1982}$$

$$=3\cdot29 \text{ appro.}$$

∴ The two regression lines are

$$X-141=\cdot034\,(Y-444\cdot45)$$

and $$Y-444\cdot45=3\cdot29\,(X-141).$$

Problem 237. Calculate the regression equations from the following data :

Age of Husband :	18	19	20	21	22	23	24	25	26	27
Age of Wife :	17	17	18	18	18	19	19	20	21	22

(*M. Com., Alld., 1951*)

Solution :

Age of Husband X	Age of Wife Y	Deviation of X from ass. average 22 ξ	Deviation of Y from ass. average 18 η	ξ^2	η^2	$\xi\eta$
18	17	−4	−1	16	1	4
19	17	−3	−1	9	1	3
20	18	−2	0	4	0	0
21	18	−1	0	1	0	0
22	18	0	0	0	0	0
23	19	1	1	1	1	1
24	19	2	1	4	1	2
25	20	3	2	9	4	6
26	21	4	3	16	9	12
27	22	5	4	25	16	20
		$\Sigma\xi=5$	$\Sigma\eta=9$	$\Sigma\xi^2=85$	$\Sigma\eta^2=33$	$\Sigma\xi\eta=48$

Mean $M_x=22+\frac{5}{10}=22\cdot5$

$$\text{Mean } M_y = 18 + \frac{9}{10} = 18{\cdot}9$$

$$\therefore \quad r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \,.\, \Sigma\eta}{n}}{\Sigma\eta^2 - n\left(\frac{\Sigma\eta}{n}\right)^2}$$

$$= \frac{48 - \frac{5\times 9}{10}}{33 - \frac{81}{10}}$$

$$= \frac{48 - 4{\cdot}5}{33 - 8{\cdot}1}$$

$$= \frac{43{\cdot}5}{24{\cdot}9}$$

$$= 1{\cdot}74$$

$$\text{Again } r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \,.\, \Sigma\eta}{n}}{\Sigma\xi^2 - n\left(\frac{\Sigma\xi}{n}\right)^2}$$

$$\text{or} \quad r\frac{\sigma_y}{\sigma_x} = \frac{43{\cdot}5}{33 - \frac{25}{10}}$$

$$= \frac{43{\cdot}5}{30{\cdot}5}$$

$$= 1{\cdot}42$$

Therefore the regression equations are

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}(X - M_x)$$

$$\text{and } X - M_x = r\frac{\sigma_x}{\sigma_y}(Y - M_y)$$

$$\text{or } Y - 18{\cdot}9 = 1{\cdot}42\,(X - 22{\cdot}5) \quad \ldots \quad \ldots \quad \ldots \quad \text{(I)}$$

$$\text{and } X - 22{\cdot}5 = 1{\cdot}74\,(Y - 18{\cdot}9) \quad \ldots \quad \ldots \quad \ldots \quad \text{(II)}$$

Problem 238.—Find the equations of regression in its simplest form of the following :

Ages of Husbands and Wives at marriage.

Husband's Age :	23	27	28	28	29	30	31	33	35	36
Wife's Age :	18	20	22	27	21	29	27	29	28	29

(*Punjab, M.A., 1950*)

Solution :

Husband's Age X	Wife's Age Y	Deviation of X from ass. average 29 ξ	Deviation of Y from ass. average 25 η	ξ^2	η^2	$\xi\eta$
23	18	−6	−7	36	49	42
27	20	−2	−5	4	25	10
28	22	−1	−3	1	9	3
28	27	−1	2	1	4	−2
29	21	0	−4	0	16	0
30	29	1	4	1	16	4
31	27	2	2	4	4	4
33	29	4	4	16	16	16
35	28	6	3	36	9	18
36	29	7	4	49	16	28
		$\Sigma\xi=10$	$\Sigma\eta=0$	$\Sigma\xi^2=148$	$\Sigma\pi^2=164$	$\Sigma\xi\eta=123$

$$\text{Mean } M_x = 29 + \frac{10}{10} = 30$$

$$\text{Mean } M_y = 25 + \frac{0}{10} = 25$$

$$r\frac{\sigma_y}{\sigma_x} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \,.\, \Sigma\eta}{n}}{\Sigma\xi^2 - n\left(\frac{\Sigma\xi}{n}\right)^2}$$

$$= \frac{123-0}{148 - \frac{100}{10}}$$

$$= \frac{123}{138}$$

$$= \cdot 9 \text{ appr.}$$

Again
$$r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \,.\, \Sigma\eta}{n}}{\Sigma\eta^2 - n\left(\frac{\Sigma\eta}{n}\right)^2}$$

$$= \frac{123-0}{164-0}$$

$$= \cdot 75$$

∴ The Regression equations are

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}(X - M_x)$$

$$\text{and } X - M_x = r\frac{\sigma_x}{\sigma_y}(Y - M_y)$$

or $\quad Y-25=\cdot 9\ (X-30)$ ... ... (I)

$X-30=\cdot 75\ (Y-25)$ ... ... (II)

Problem 239.—In the following table are recorded data showing the test scores made by salesmen on an intelligence test and their weekly sales.

Salesmen	1	2	3	4	5	6	7	8	9	10
Test Scores	40	70	50	60	80	50	90	40	60	60
Sales (000)	2·5	6·0	4·5	5·0	4·5	2·0	5·5	3·0	4·5	3·0

Calculate the regression line of sales on test score, and estimate the most probable weekly sales volume if salesman makes a score of 70. *(I. A. S., 1948)*

Solution :

Test Scores X	Sales Y	Deviation of X from ass. average 60 (ξ)	Deviation of Y from ass. average 4 5 (η)	$\xi\eta$	ξ^2	η^2
40	2·5	—20	—2	40	400	4
70	6·0	10	1·5	15	100	2·25
50	4·5	—10	0	0	100	0
60	5·0	0	·5	0	0	·25
80	4·5	20	0	0	400	0
50	2·0	—10	—2·5	25	100	6·26
90	5 5	30	1	30	900	1
40	3 0	—20	—1·5	30	400	2·25
60	4·5	0	0	0	0	0
60	3·0	0	—1·5	0	0	2·25
		$\Sigma\xi=0$	$\Sigma\eta=-4\cdot 5$	$\Sigma\xi\eta=140$	$\Sigma\xi^2=2400$	$\Sigma\eta^2=18.25$

$$\text{Mean } M_x = 60 + \frac{0}{10} = 60$$

$$M_y = 4\cdot 5 + \frac{-4\cdot 5}{10} = 4\cdot 05$$

Regression line of sales on score that is of Y on X is

$$Y-M_y = r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

or $$Y-4\cdot 05=\frac{\Sigma\xi\eta-\dfrac{\Sigma\xi\,.\,\Sigma\eta}{n}}{\Sigma\xi^2-n\left(\dfrac{\Sigma\xi}{n}\right)^2}(X-60)$$

or $$Y-4\cdot 05=\frac{140-\dfrac{0\times(-4\cdot 5)}{10}}{2400-10\left(\dfrac{0}{10}\right)^2}(X-60)$$

or $$Y-4\cdot05=\frac{140}{2400}\ (X-60)$$

$\therefore$ for X = 70

$$Y-4\cdot05=\frac{140}{2400}\ (70-60)$$

or $$Y=4\cdot05+\frac{1400}{2400}$$

$$=4\cdot633 \text{ appr.}$$

$\therefore$ Most probable weekly sales volume if salesman makes a score of 70 is 4·633.

✓**Problem 240.**—The following data are given for marks in English and Maths. in the S. L. C. Examination of the U. P. in a certain year :

Mean Marks in English	=39·5
,, ,, Maths.	=47·6
r between marks in English and Maths.	=·42
Standard deviation of marks in English	=10·8
,, ,, in Maths.	=16·9

Form the two lines of regression and explain why there are two equations of regression. Calculate the expected average marks in Maths. of candidates who received 50 marks in English.

(U.P.C.S., 1941)
Agra B.Sc. 1960

Solution :

If X and Y represents the marks in English and Maths. respectively then

Give $M_x=39\cdot5,\ M_y=47\cdot6,\ \sigma_x=10\cdot8,\ \sigma_y=16\cdot9$
$r=\cdot42$

The regression line of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}\ (X-M_x)$$

or $$Y-47\cdot6=\cdot42\ \frac{16\cdot9}{10\cdot8}\ (X-39\cdot5)$$

or $$Y-47\cdot6=\cdot657\ (X-39\cdot5) \quad \ldots \quad \ldots \quad (I)$$

The regression line of X on Y is

$$X-M_x=r\ \frac{\sigma_x}{\sigma_y}\ (Y-M_y)$$

or $X-39{\cdot}5= {\cdot}42\frac{10{\cdot}8}{16{\cdot}9}(Y-47{\cdot}6)$

or $X-39{\cdot}5= {\cdot}268\ (Y-47{\cdot}6)$ ... ... (II)

For X=50 using formula I we have

$Y-47{\cdot}6= {\cdot}657\ (50-39{\cdot}5)$

or $Y=47{\cdot}6+6{\cdot}9$

$=54{\cdot}5.$

∴ Expected average marks in Maths. of candidates who secured 50 marks in English are=54·5

Problem 241.—The following data are given for marks in subjects A and B in a certain examination :

Mean Marks in A	=39·5
Mean Marks in B	=47·5
Standard Deviation of Marks in A	=10·8
Standard Deviation of marks in B	=16·8
Coefficient of Correlation between A and B	=+ ·42

Draw the two Lines of Regression and explain why there are two equations of regression.

Also give the expectations of marks in subject B for candidates who secured 50 marks in subject A.

(M. A., Pb. Sept., 1951)
(M. Com., Alld., 1946)
or (Agra B.Sc. 1960)

Solution :

Let X denotes subject A and Y subject B.

∴ $M_x=39{\cdot}5,\ M_y=47{\cdot}5,\ \sigma_x=10{\cdot}8,\ \sigma_y=16{\cdot}8,\ r={\cdot}42$

Two Lines of Regression are

$$Y-M_y = r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

and $$X-M_x = r\frac{\sigma_x}{\sigma_y}(X-M_y)$$

or $Y-47{\cdot}5= {\cdot}42\frac{16{\cdot}8}{10{\cdot}8}(X-39{\cdot}5)$

and $X-39{\cdot}5= {\cdot}42\frac{10{\cdot}8}{16{\cdot}8}(Y-47{\cdot}5)$

or $Y-47{\cdot}5= {\cdot}65\ (X-39{\cdot}5)$ ... ... (I)

and $X-39{\cdot}5= {\cdot}27\ (Y-47{\cdot}5)$ ... ... (II)

Hence equations I and II are two Lines of Regression. Therefore expected marks of B for 50 marks of A is given by

$$Y-47{\cdot}5= {\cdot}65\ (50-39{\cdot}5)$$

or $Y-47{\cdot}5=6{\cdot}825$

or $Y=54{\cdot}325.$

Problem 242.—

(*a*) Given

	X Series	Y Series
Mean	18	100
Standard Deviation	14	20

Coefficient of correlation between X and Y series is +·8. Find out the most probable value of Y if X is 70 and most probable value of X if Y is 90.

(*b*) If the regression coefficients are ·8 and ·6, what would be the value of the coefficient of correlation.

(*Alld., M.A., 1952*)

Solution :

(*a*) $M_x=18, \quad M_y=100$

$\sigma_x=14, \quad \sigma_y=20$

$r=\cdot 8$

Regression equation of X on Y is

$$X-M_x=r\frac{\sigma_x}{\sigma_y}(Y-M_y)$$

or $X-18=\cdot 8\frac{14}{20}(Y-100)$

or $X-18=\cdot 56\ (Y-100)$ ...I

Regression equation of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

or $Y-100=\cdot 8\times\frac{20}{14}(X-M_x)$

or $Y-100=1{\cdot}14\ (X-18)$ ...II

Value of Y if X is 70 is from II

$Y-100=1{\cdot}14\ (70-18)$

or $Y=159{\cdot}28$ ...III

Value of X if Y is 90 is

$X-18=\cdot 56\ (90-100)$

or $X=12{\cdot}4$ ...IV

(*b*) $r\frac{\sigma_x}{\sigma_y}=\cdot 8, \quad r\frac{\sigma_y}{\sigma_x}=\cdot 6$

$\therefore \qquad r^2 = \cdot 48$

$\therefore \qquad r = \cdot 69$

Problem 243.—Given the following data, calculate the expected value of Y when X is 12.

	X	Y
Average	7·6	14·8
Standard Deviation	3·6	2·5

$$r = \cdot 99$$

(*Alld. M. Com., 1945*)
(*Raj., M.A., 1951*)

Solution :

$$M_x = 7.6, \qquad M_y = 14{\cdot}8$$
$$\sigma_x = 3{\cdot}6, \qquad \sigma_y = 2{\cdot}5$$
$$r = \cdot 99$$

Regression equation of Y on X is

$$Y - M_y = r \frac{\sigma_y}{\sigma_u} (X - M_x)$$

or $$Y - 14{\cdot}8 = \cdot 99 \frac{2{\cdot}5}{3{\cdot}6} (X - 7{\cdot}6)$$

or $$Y - 14{\cdot}8 = \cdot 688 \, (X - 7{\cdot}6) \qquad \ldots I$$

expected value of Y when X is 12 is given by

$$Y - 14{\cdot}8 = \cdot 688 \, (12 - 7{\cdot}6)$$

or $$Y - 14{\cdot}8 = \cdot 688 \times 4{\cdot}4$$

or $$Y = 14{\cdot}8 + 3{\cdot}027$$
$$= 17{\cdot}827$$

Problem 244.—Given the following data, find what will be the probable yield when the rainfall is 29 inches :—

Mean rainfall = 25 inches
Mean production = 40 units per acre.

Coefficient of correlation between rainfall and production = ·8
Standard deviation of rainfall = 3 inches
Standard deviation of production = 6 units per acre
Explain your method.

(*M. Com., Agra, 1947*)

Solution :

Let X denotes rainfall in inches and Y production in units per acre. Then

$$M_x = 25 \text{ inches}, \qquad M_y = 40 \text{ units per acre}$$
$$\sigma_x = 3, \qquad \sigma_y = 6, \qquad r = \cdot 8$$

Regression equation of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

or $$Y-40=\cdot 8\ \frac{6}{3}\ (X-25)$$

or $$Y-40=1\cdot 6\ (X-25)$$

Therefore probable yield when rainfall is 29 is given by

$$Y-40=1\cdot 6\ (29-25)$$

or $$Y=46\cdot 4$$

∴ probable yield is 46·4 units per acre.

✓**Problem 245.**—Calculate regression equations from the following data :

Mean height = 50·07 inches St. Dev. of height = 5·26 inches
Mean Age = 9·98 years St. Dev. of age = 2·59 years.

$$r=+\cdot 898$$

(*Raj., M.A., 1950*)

Solution :

If X represents height and Y age then

$M_x=50\cdot 07$, $\sigma_x=5\cdot 26$
$M_y=9\cdot 98$, $\sigma_y=2\cdot 59$
$r=\cdot 898$

Regression equation of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}\ (X-M_x)$$

or $$Y-9\cdot 98=\frac{\cdot 898\times 2\cdot 59}{5\cdot 26}\ (X-50\cdot 07)$$

or $$Y-9\cdot 98=\cdot 442\ (X-50\cdot 07) \quad \ldots I$$

Regression equation of X on Y is

$$X-M_x=r\frac{\sigma_x}{\sigma_y}\ (Y-M_y)$$

or $$X-50\cdot 07=\frac{\cdot 898\times 5\cdot 26}{2\cdot 59}\ (Y-9\cdot 98)$$

or $$X-50\cdot 07=1\cdot 797\ (Y-9\cdot 98) \quad \ldots II$$

Therefore the two regression equations are

$$Y-9\cdot 98=\cdot 442\ (X-50\cdot 07)$$

and $$X-50\cdot 07=1\cdot 797\ (Y-9\cdot 98)$$

Problem 246.—Explain what is meant by a scatter diagram and line of regression. Why should there be in general two lines of regression for each bivariate distribution ?

You are given the following results for the heights (X) and weights (Y) of 1000 policemen of U.P. ;

$$M_x=68 \text{ inches}, \quad M_y=150 \text{ lbs.}, \quad \sigma_x=2\cdot5 \text{ inches}, \quad \sigma_y=20 \text{ lbs.} \quad r=\cdot6$$

Estimate from the above data (*a*) the height of a particular policeman whose weight is 200 lbs., (*b*) the weight of particular policemen who is 5 ft. tall.

(*P. C. S. 1953*)

Solution :

Given $M_x=68$ inches, $M_y=150$ lbs., $\sigma_x=2\cdot50$ inches, $\sigma_y=20$ lbs. $r=\cdot6$

The regression line of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

or $$Y-150=\cdot6\ \frac{20}{2\cdot5}(X-68)$$

or $$Y-150=4\cdot8(X-68) \quad \text{.......I}$$

The regression line of X on Y is

$$X-M_x=r\frac{\sigma_x}{\sigma_y}(Y-M_y)$$

or $$X-68=\cdot6\ \frac{2\cdot5}{20}(Y-150)$$

or $$X-68=\cdot75(Y-150) \quad \text{.......II}$$

(*a*) to find the height of a particular policeman whose weight is 200 lbs.

$Y=200$

$$X-68=\cdot75(200-150)$$

or $$X=105\cdot5 \text{ inches.}$$

(*b*) weight of particular policeman the height of whom is 5 ft.

$X=60 \quad \therefore \; Y-150=4\cdot8(60-68)$

or $\; Y=112\cdot4$ lbs.

$\therefore$ Height of a particular policeman whose weight is 200 lbs. is 105·5 inches and weight of a particular policeman whose height is 60 inches is 112·4 lbs.

✓**Problem 247.**—Illustrate the methodology of computing the correlation coefficient and the equation of the line of regression of Y on X by using the following data :—

X	2	6	4	7	5
Y	8	8	5	6	2

Incorporate checks on the arithmetic in your work. Show that the coefficient of correlation lies between -1 and $+1$.

(*M.Sc. Agra, 1948*)

Solution :

X	Y	Deviation of X from ass. average 4 ξ	Deviation of Y from ass. average 5 η	$\xi\eta$	ξ^2	η^2
2	8	-2	3	-6	4	9
6	8	2	3	6	4	9
4	5	0	0	0	0	0
7	6	3	1	3	9	1
5	2	1	-3	-3	1	9
		$\Sigma\xi=4$	$\Sigma\eta=4$	$\Sigma\xi\eta=0$	$\Sigma\xi^2=18$	$\Sigma\eta^2=28$

$$r = \frac{\Sigma\xi\eta - \dfrac{\Sigma\xi.\Sigma\eta}{n}}{\sqrt{\Sigma\xi^2 - n\left(\dfrac{\Sigma\xi}{n}\right)^2}\ \sqrt{\Sigma\eta^2 - n\left(\dfrac{\Sigma\eta}{n}\right)^2}}$$

$$= \frac{0 - \dfrac{4\times 4}{5}}{\sqrt{18-5\left(\dfrac{4}{5}\right)^2}\ \sqrt{28-5\left(\dfrac{4}{5}\right)^2}}$$

$$= \frac{-3{\cdot}2}{\sqrt{14{\cdot}8\times 24{\cdot}8}}$$

$$= \frac{-3{\cdot}2}{19{\cdot}15}$$

$$= -{\cdot}167.$$

$$M_x = 4+\tfrac{4}{5} = 4{\cdot}8$$

$$M_y = 5+\tfrac{4}{5} = 5{\cdot}8$$

The equation of the line of regression of Y on X is

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}(X - M_x)$$

$$\text{Now} \quad r\,\frac{\sigma_y}{\sigma_x} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi\Sigma\eta}{n}}{\Sigma\xi^2 - n\left(\frac{\Sigma\S}{n}\right)^2}$$

$$= \frac{0 - \frac{4\times 4}{5}}{18 - 5\left(\frac{4}{5}\right)^2}$$

$$= \frac{-3{\cdot}2}{14{\cdot}8}$$

$$= -{\cdot}21$$

∴ Regression line of Y on X is

$$Y - 5{\cdot}8 = -{\cdot}21\,(X - 4{\cdot}8)$$

Problem 248.—In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible :—

Variance of X=9

Regression Equations—

$$8X - 10Y + 66 = 0$$

$$40X - 18Y = 214.$$

What were (*a*) the mean values of X and Y (*b*) the standard deviation of Y, and (*c*) the coefficient of correlation between X and Y.

(*I.A.S. 1947*)

Solution :

(*a*) Calculation of the mean values of X and Y.

$$8X - 10Y = -66 \quad \ldots \quad \text{I}$$

$$40X - 18Y = 214 \quad \ldots \quad \text{II}$$

On solving these equations the values of X and Y so obtained gives M_x and M_y.

multiplying I by 5 we get

$$40X - 50Y = -330$$

$$40X - 18Y = 214$$

subtracting

$$-32Y = -544$$

$$\therefore \quad Y = 17$$

substituting the value of Y in I

$$8X - 170 = -66$$

or

$$X = 13$$

$$\therefore \quad M_x = 13 \text{ and } M_y = 17$$

(c) Calculation of the coefficient of correlation between X and Y

$$40X-18Y=214 \quad \ldots \quad (i)$$
$$8X-10Y=-66 \quad \ldots \quad (ii)$$

From (i) $X = \cdot 45Y + 5\cdot 35$

∴ Regression coefficient of X on Y is

$$r\frac{\sigma_x}{\sigma_y} = \cdot 45 \quad \ldots$$

From (ii) $10Y = 66 + 8X$

or $Y = \cdot 8X + 6.6$

∴ Regression coefficient of Y on X is

$$r\frac{\sigma_y}{\sigma_x} = \cdot 8$$

$$\therefore \quad r = \sqrt{r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y}}$$
$$= \sqrt{\cdot 45 \times \cdot 8}$$
$$= \cdot 6$$

If from (1) we determine Regression coefficient of Y on X and from (ii) Regression coefficient of X on Y then the value of r will come out to be greater than 1 which is impossible. Hence the Regression coefficients of X on Y and of Y on X will be determined from (i) and (ii) respectively.

(b) Calculation of standard deviation of Y

Variance of X that is $\sigma_x^2 = 9$

$$\therefore \sigma_x = 3$$

$$r\frac{\sigma_x}{\sigma_y} = \cdot 45$$

$$\therefore \sigma_y = \frac{r\sigma_x}{\cdot 45}$$
$$= \frac{\cdot 6 \times 3}{\cdot 45}$$
$$= 4$$

∴ (a) Mean values of X and Y are 13 and 17 respectively.

(b) Standard deviation of Y is 4.

(c) Coefficient of correlation between X and Y is ·6

Problem 249.—Two lines of regression are given by $x+2y-5=0$ and $2x+3y-8=0$ and $\sigma_x^2=12$. Calculate the value of $\bar{x}$, $\bar{y}$, σ_y^2 and r.

(*All., M.A., 1952*)

Solution :

$$2x+3y-8=0$$
$$x+2y-5=0$$

Solving these equation for x and y we have

$$x=1 \text{ and } y=2$$

$\therefore \quad \bar{x}=1, \quad \bar{y}=2$

The above equations can be written as

$$x=-\tfrac{3}{2}y+4$$

$$\text{and } y=-\tfrac{1}{2}x+\tfrac{5}{2}$$

$$\therefore \; r\frac{\sigma_x}{\sigma_y}=-\tfrac{3}{2}$$

$$\text{and } r\frac{\sigma_y}{\sigma_x}=-\tfrac{1}{2}$$

multiplying $r^2=\frac{3}{4}=\cdot75$

$\therefore \quad r=\cdot86$

Dividing we have $\dfrac{\sigma_x^2}{\sigma_y^2}=3$

$\therefore \; \sigma y^2=4$

Hence $\bar{x}=1, \; \bar{y}=2, \; r=\cdot86, \; \sigma_y^2=4$

Problem 250.—What are regression lines ? Why are there in general two regression lines ?

If α is the acute angle between the two regression lines in the case of two variables X and Y, show that

$$\tan \alpha = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}$$

where r, σ_x and σ_y have their usual meaning.

(*M.Sc., Agra 1955*)

Solution :

If θ_1, θ_2 are the angles which the two regression lines makes with the axis of x then

$$\tan \theta_1 = r\frac{\sigma_y}{\sigma_x}$$

$$\cot \theta_2 = r\frac{\sigma_x}{\sigma_y}$$

$$\therefore \quad \tan \theta_2 = \frac{\sigma_y}{r\sigma_x}$$

$$\text{Now } \tan \alpha = \tan(\theta_2 - \theta_1)$$

$$= \frac{\tan \theta_2 - \tan \theta_1}{1 + \tan \theta_2 \tan \theta_1}$$

$$= \frac{\frac{\sigma_y}{r\sigma_x} - \frac{r\sigma_y}{\sigma_x}}{1 + \frac{\sigma_y \cdot r\sigma_y}{r\sigma_x \cdot \sigma_x}}$$

$$= \frac{\frac{\sigma_y(1-r^2)}{\sigma_x}}{\frac{r(\sigma_x^2+\sigma_y^2)}{\sigma_x^2}}$$

$$= \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}$$

Problem 251.—The following statistical coefficients were deduced in the course of an examination of the relationship between yield of wheat and the amount of rainfall :—

	Yield in lb. per acre	*Annual rainfall in inches*
Mean	985	12·8
Standard deviation	70·1	1·6
r between yield and rainfall	+·52	

from the above data, calculate (*a*) the most likely yield of wheat per acre when the annual rainfall is 9·2 inches, (*b*) the probable annual rainfall for yield of 1400 lb. per acre.

(*M.A., Agra 1938*)

Solution :

If X represent the yield per acre and Y the annual rainfall then the two regression lines are

$$Y - M_y = r\,\frac{\sigma_y}{\sigma_x}\,(X - M_x)$$

and $$X - M_x = r\,\frac{\sigma_x}{\sigma_y}\,(Y - M_y)$$

or $$Y - 12{\cdot}8 = {\cdot}52\,\frac{1{\cdot}6}{70{\cdot}1}\,(X - 985)$$

and $$X - 985 = {\cdot}52\,\frac{70{\cdot}1}{1{\cdot}6}\,(Y - 12{\cdot}8)$$

or $$Y - 12{\cdot}8 = {\cdot}01(X - 985) \qquad \ldots \quad (i)$$

and $$X - 985 = 22{\cdot}8(Y - 12{\cdot}8) \qquad \ldots \quad (ii)$$

(*a*) The yield of wheat per acre (X) when annual rainfall (Y) is 9·2 inches is given by (*ii*)

$\therefore \quad X-985=22{\cdot}8(9{\cdot}2-12{\cdot}8)$

or $\quad X=902{\cdot}92$ lbs.

$\therefore$ yield of wheat per acre when annual rainfall is 9·2 is 902·92 lbs.

(*b*) Annual rainfall for yield of 1400 lbs. is given by (*i*)

$Y-12{\cdot}8={\cdot}01(1400-985)$

or $\quad Y=16{\cdot}95$ inches

$\therefore$ Annual rainfall for yield of 1400 lbs is 16·95 inches

Problem 252.—Find the most likely price in Bombay corresponding to the price of Rs. 70 at Calcutta from the following data : —

	Calcutta	*Bombay*
Average price	65	67
Standard deviation	2·5	3·5

Coefficient of correlation is ·8 between the two prices of the commodities in the two towns.

(*M. Com., Agra, 1951 ; M.A., Agra, 1954*)

Solution :

If X and Y represents the prices at Calcutta and Bombay respectively, then regression line of Y on X is

$$Y-M_y=r\frac{\sigma_y}{\sigma_x}(X-M_x)$$

or $\quad Y-67={\cdot}8\,\dfrac{3{\cdot}5}{2{\cdot}5}(X-65)$

or $\quad Y-67=1{\cdot}12(X-65)$

$\therefore$ price at Bombay corresponding to price Rs. 70 at Calcutta is given by

$$Y-67=1{\cdot}12(70-65)$$
$$Y=67+5{\cdot}6$$

or $\quad Y=72{\cdot}6$ Rs. approximately.

Hence the price at Bombay corresponding to price Rs. 70 at Calcutta is Rs. 72·6 approximately.

Problem 253.—Vital statistics of the U.P. (in thousands)

Years	*Fevers*	*Respiratory diseases*	*Dysentery and diarrhoea*	*Others*	*Total*
1931	1025	37	16	228	1306
1932	853	34	13	176	1076
1933	698	35	12	160	905
1934	970	47	18	260	1295

Find out r of the deaths from the fevers and total deaths given above. Calculate standard error of this coefficient and the line of regression of the deaths from fevers on total deaths.

(*M.A., Agra, 1937*)

Solution :

Deaths from the fevers X	Total deaths Y	Deviation of X from ass. average 885 ξ	Deviation of Y from ass. average 1145 η	$\xi\eta$	2	η^2
1025	1306	140	161	22540	19600	25921
853	1076	−32	−69	2208	1024	4761
698	905	−157	−240	44880	84969	57600
970	1295	85	150	12750	7225	22500
		$\Sigma\xi=6$	$\Sigma\eta=2$	$\Sigma\xi\eta=82378$	$\Sigma\xi^2=62818$	$\Sigma\eta^2=110782$

$$r=\frac{\Sigma\xi\eta-n\left(\frac{\Sigma\xi}{\eta}\right)\left(\frac{\Sigma\eta}{n}\right)}{\sqrt{\Sigma\xi^2-n\left(\frac{\Sigma\xi}{n}\right)^2}\sqrt{\Sigma\eta^2-n\left(\frac{\Sigma\eta}{n}\right)^2}}$$

$$=\frac{82378-4\left(\frac{6}{4}\right)\left(\frac{2}{4}\right)}{\sqrt{62818-4\left(\frac{6}{4}\right)^2}\sqrt{110782-4\left(\frac{2}{4}\right)^2}}$$

$$=\frac{82378-3}{\sqrt{62818-9}\sqrt{110782-1}}$$

$$=\frac{82375}{\sqrt{62809\times110781}}$$

$$=\frac{82375}{83414\cdot9}$$

$$=\cdot987$$

Standard Error of the coefficient of correlation.

$$\text{S.E.}=\frac{1-r^2}{\sqrt{n}}$$

$$=\frac{1-(\cdot987)^2}{\sqrt{4}}$$

$$=\frac{1-\cdot974}{2}$$

$$=\cdot013$$

mean of the deaths from fever $M_x = 885 + \frac{6}{4}$

$= 886 \cdot 5$ deaths

mean of the total deaths M_y $= 1145 + \frac{2}{4}$

$= 1145 \cdot 5$ deaths

$$r\frac{\sigma_x}{\sigma_y} = \frac{\Sigma\xi\eta - \frac{\Sigma\xi \cdot \Sigma\eta}{n}}{\Sigma\eta^2 - n\left(\frac{\Sigma\eta}{n}\right)}$$

$$= \frac{82378 - \frac{6 \cdot 2}{4}}{110782 - 4\left(\frac{2}{4}\right)^2}$$

$$= \frac{82375}{110782 - 1}$$

$$= \frac{82375}{110781}$$

$$= \cdot 74$$

∴ Regression line of X on Y is

$$X - M_x = r\frac{\sigma_x}{\sigma_y}\ (Y - M_y)$$

or $X - 886 \cdot 5 = \cdot 74(Y - 1145 \cdot 5)$

CHAPTER VIII

GRAPHS

The use of graphs are becoming more and more in the study of statistics. Graph simplifies complexity of numerical statement. It represents the tendency, and a mere look at it saves the time and energy of going through complex data.

In the construction of graphs, the following points may be kept in view :

(1) A complete and accurate *title* should be given to the graph constructed.

(2) The constant factor should be shown on the OX axis and the variable factor on the OY axis.

(3) *Fixation of scale.* It is perhaps the most important phase of the construction of a graph.

(*i*) The scale should be so adjusted that the graph is neither too small, nor too large.

(*ii*) Find out the total amount of space available on the OX and OY respectively, and the total amounts to be shown on them. Divide the total amount by the space available, which will help the fixation of base, as :

Space available on the graph (say OX axis) = 6″

Total amount to be shown on this axis = 6000 units.

$\therefore$ 1″ on OX axis will represent $\frac{6000}{6} = 1000$ units.

The same process is to be followed for the fixation of scale in the OY axis.

[*It is to be remembered, however, that it is not essential that it should always be a wholenumber. It may be in fraction.*]

(*iii*) If there is a long difference between the value of O and the starting value, then a *false base line* may be drawn, leaving a few lines from the OX axis and drawing a line parallel to OX axis.

This line will indicate the minimum value :—

Year	*Exports in crores of Rupees*
1950	540
1951	610
1952	725

In this case, the point (on OY axis) where from the false base line is drawn, will give the value of 540 crores.

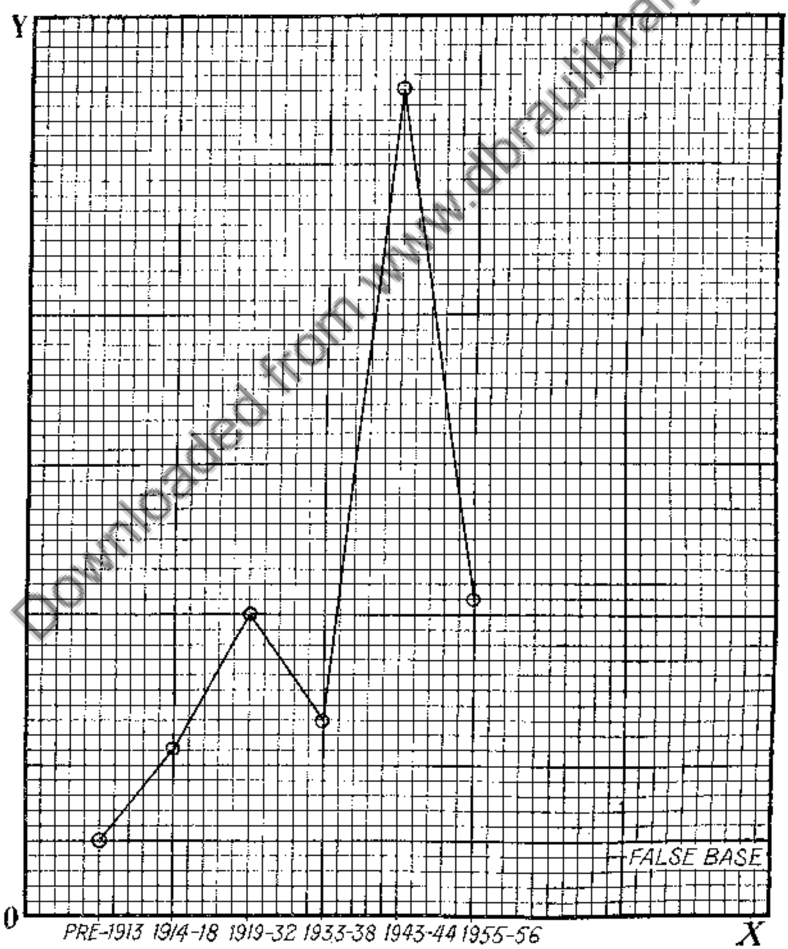
Problem 254.—Draw a graph to represent the following data :

Years	*Manufacture of shoes (pairs) at Agra per day*
Pre—1913	500
1914—18	6,000
1919—32	15,000
1933—38	8,000
1943—44	50,000
1955—56	16,000

Solution :

(*i*) Take years on OX axis.
(*ii*) Show pairs of shoes on OY axis.
(*ii*) Draw a false base line.

Scale :—
OX axis$=\frac{5}{10}''$ → one different year to be represented.
OY axis$=\frac{1}{10}''$ → 1000 pairs.



Problem 255.—Represent the following on a graph paper on natural Scale. Figures indicates employment of labourers in Tanneries of U.P.

Year	*Labourers Engaged in tanneries of U.P.*
1929	745
1939	1679
1943	3800
1953	2496
1954	——

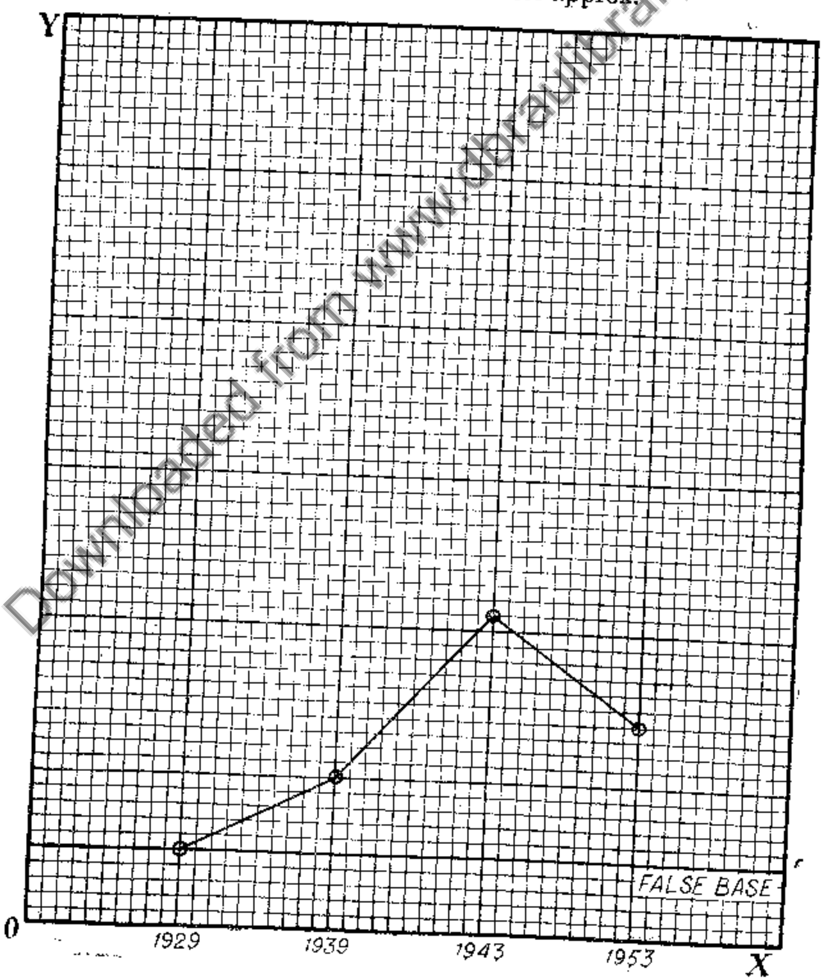
Solution :

Take years on OX axis and labourers engaged in tanneries on OY axis. Draw a false base line.

Scale :—

OX axis → 1″=one representable year.

OY axis → $\frac{1}{10}$″=186 labourers approx.



Problem 256.—Draw a graph showing the number of workers engaged in leather goods factories of U.P. (organised sector) over the mentioned years.

Year	*Labourers engaged in leather goods factories*
1929	2301
1939	2765
1943	9062
1953	3640
1954	——

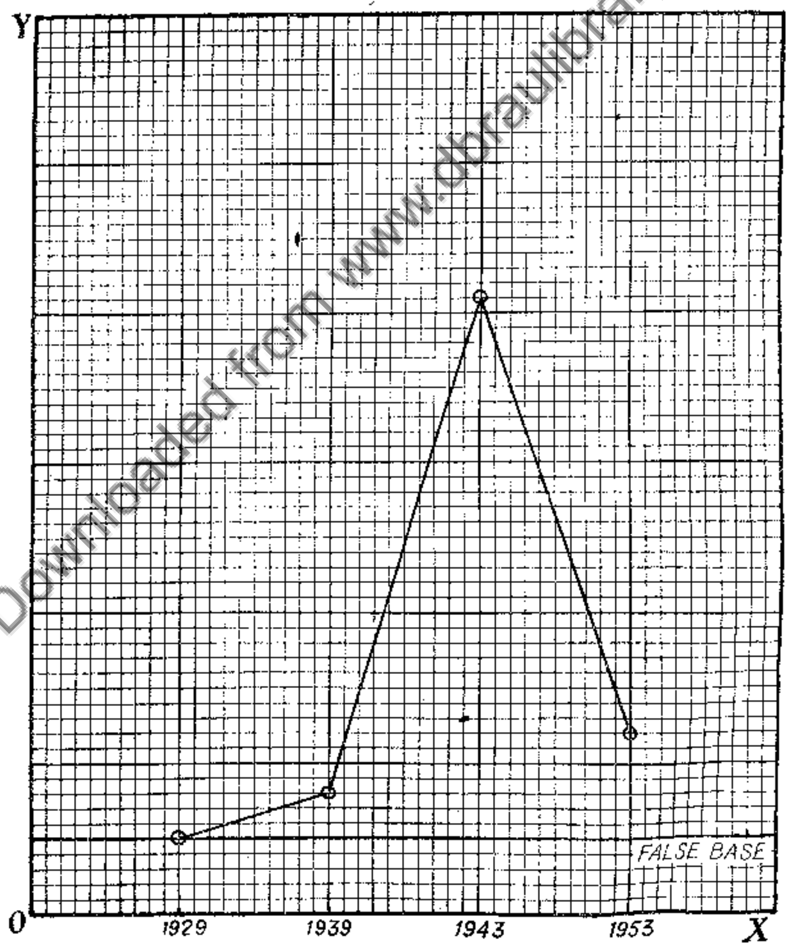
Solution :

Take years on OX axis and labourers engaged on OY axis. As usual, draw a false base line to save space.

Scale :—

OX axis → $1'' = 1$ representable year.

OY axis → $\frac{1}{10}'' = 186$ labourers approximately.



Problem 257.—Please draw a graph to show the yearly income of a man over different years, as mentioned below.

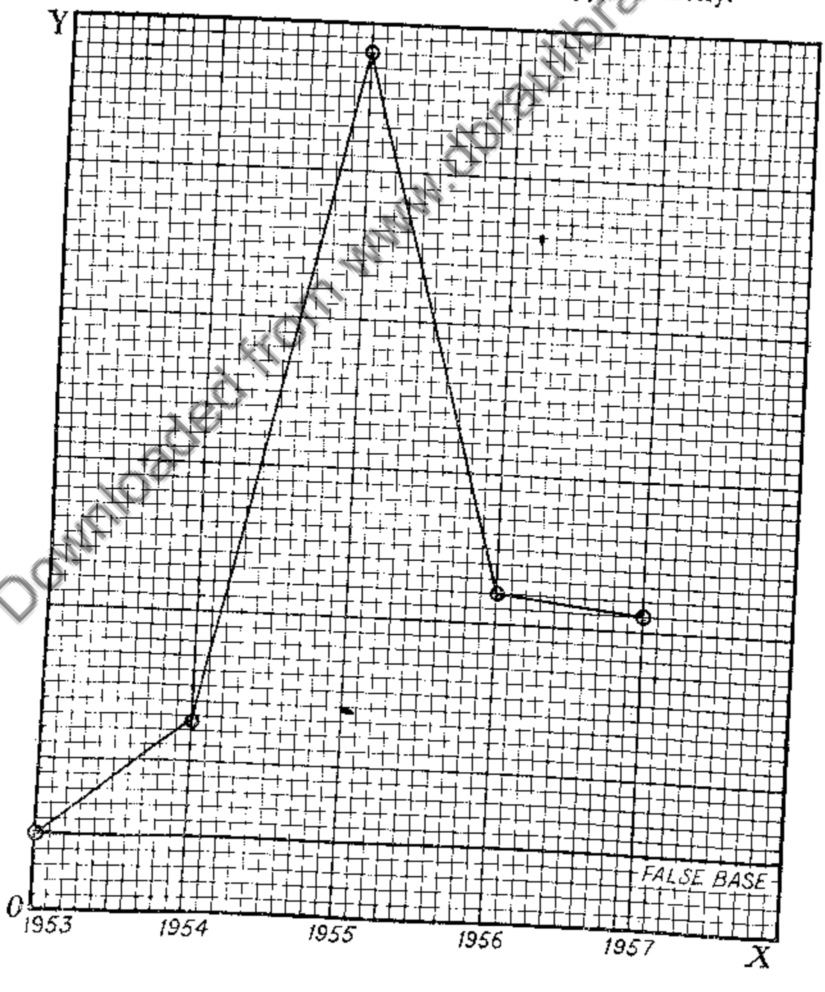
Year	*Income in Rs.*
1953	3046
1954	4444
1955	12682
1956	6184
1957	6096

Solution :

Draw the graph with years on the OX axis and income on OY axis.

Scale :—

OX axis → 1″=1 year.
OY axis → $\frac{1}{10}$″=186 rupees approximately.



Problem 258.—Represent graphically the percentage trend of the use of goats for milk purposes in U.P.

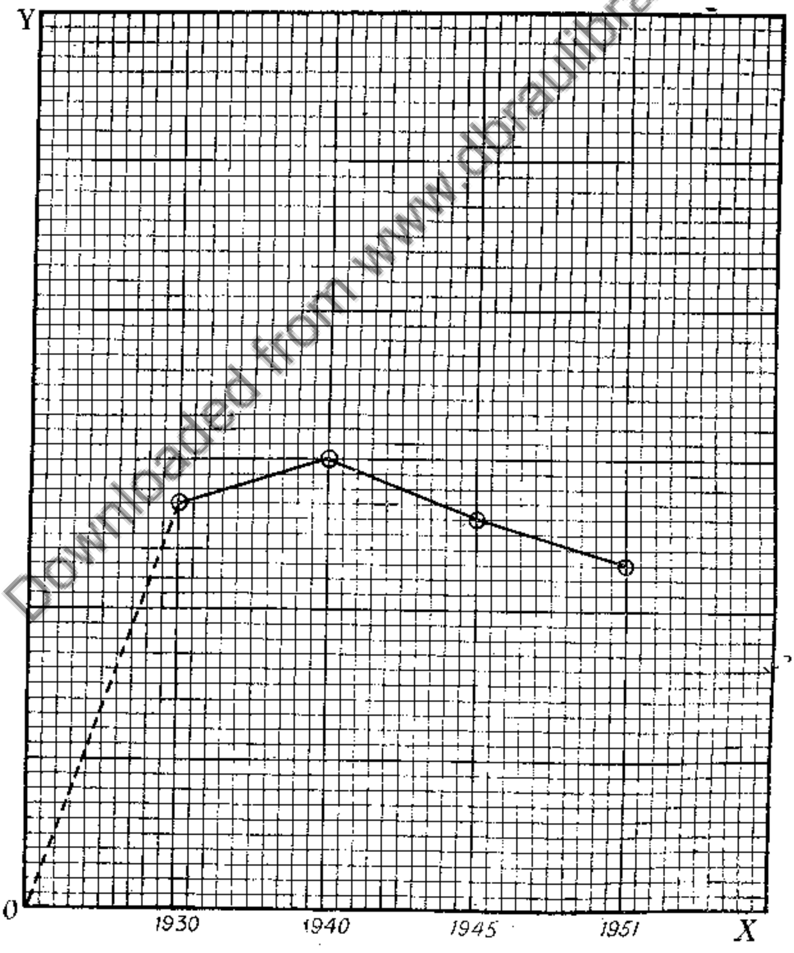
Year	*Percentage of goats used for milk purposes in U. P.*
1930	27
1940	30
1945	26
1951	23

Solution :

Represent years on OX axis and percentage of goats used for milk on OY axis.

Scale :—

OX axis → 1″ = Representable year.
OY axis → $\frac{1}{10}$″ = 1 percent.



Problem 259.—Represent graphically the use of Cow (percentage) in Bengal for milk purposes, over the mentioned years :

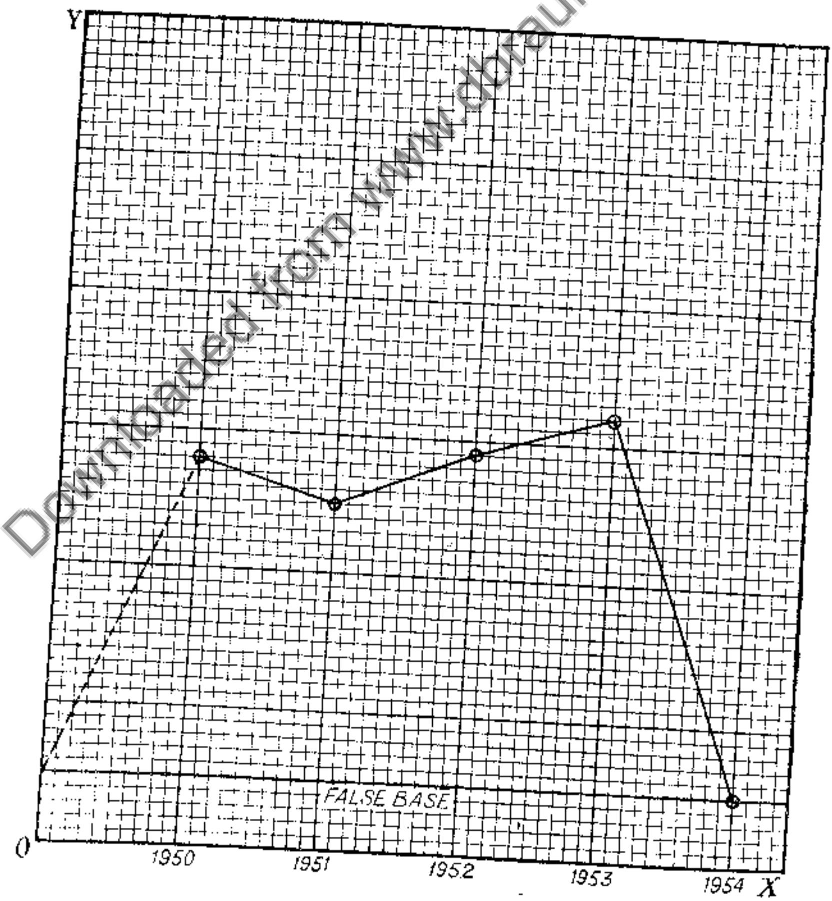
Year	*Percentage of Cow used for milk purposes*
1950	73
1951	70
1952	74
1953	77
1954	50

Solution :

Take years on OX axis and percentage of Cow used for milk purposes on OY axis. Draw a false base.

Scale :—

OX axis → 1″ = one mentioned year.
OY axis → $\frac{1}{10}$″ = one perent.



Problem 260.—Draw a graph to show the following data on a graph paper :

Year	*Workers Engaged*
1913	1,000
1918	12,000
1932	25,000
1938	10,000
1944	60,000
1956	20,000

Solution :

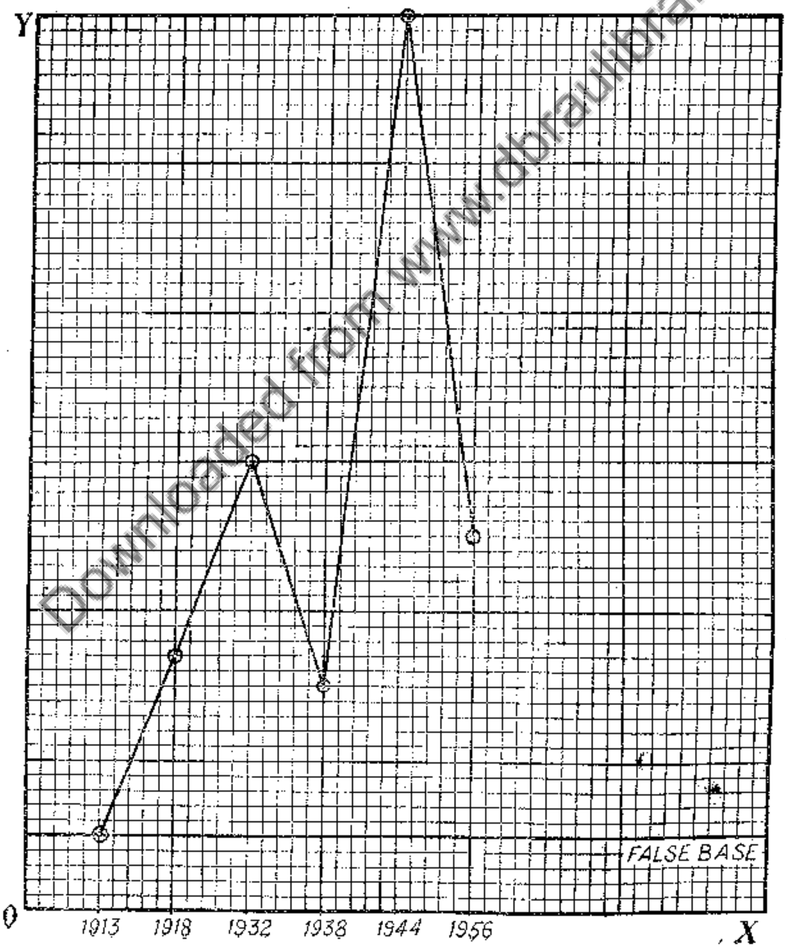
(*a*) Show years on OX axis and workers engaged on OY axis.

(*b*) Draw a false base line.

Scale :—

OX axis → $\frac{5}{10}''$=one mentionable year.

OY axis → $\frac{1}{10}''$=1000 workers.



Problem 261.—Represent graphically the growth of bacteria per c.c. on different degrees of temperature from the following data :

Temperature	*Count of bacteria per C.C.*
50°C	1,10,00,00,000
30°C	87,90,00,000
20°C	21,80,00,000
10°C	1,08,20,000

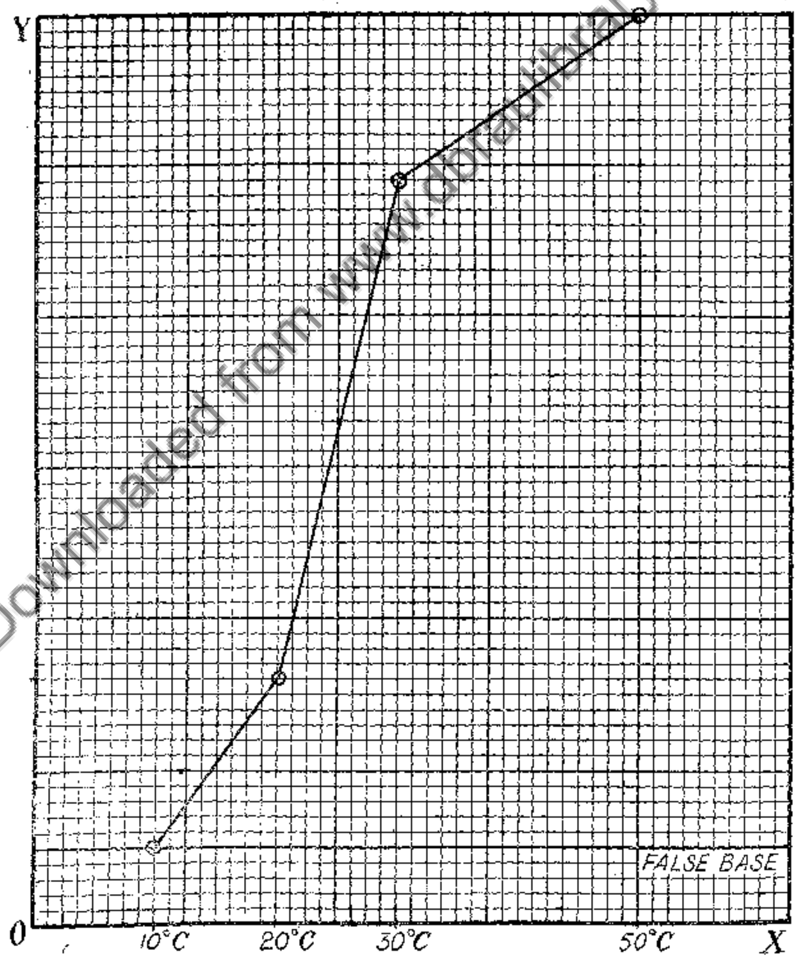
Solution :

Represent temperature on OX axis and count of bacteria per C.C. on OY axis. Draw a false base line.

Scale :—

OX axis→$\frac{8}{10}''$ = 10°C

OY axis→$\frac{1}{10}''$ = App. 200 crores of bacteria per c.c.



Problem 262.—Represent graphically the following data on natural scale.

Hours of incubation	*Count of bacteria for c.c. on broth at 20°C with salt*
0	10,600
1	11,200
2	11,500
3	11,300
5	10,800
24	60,100

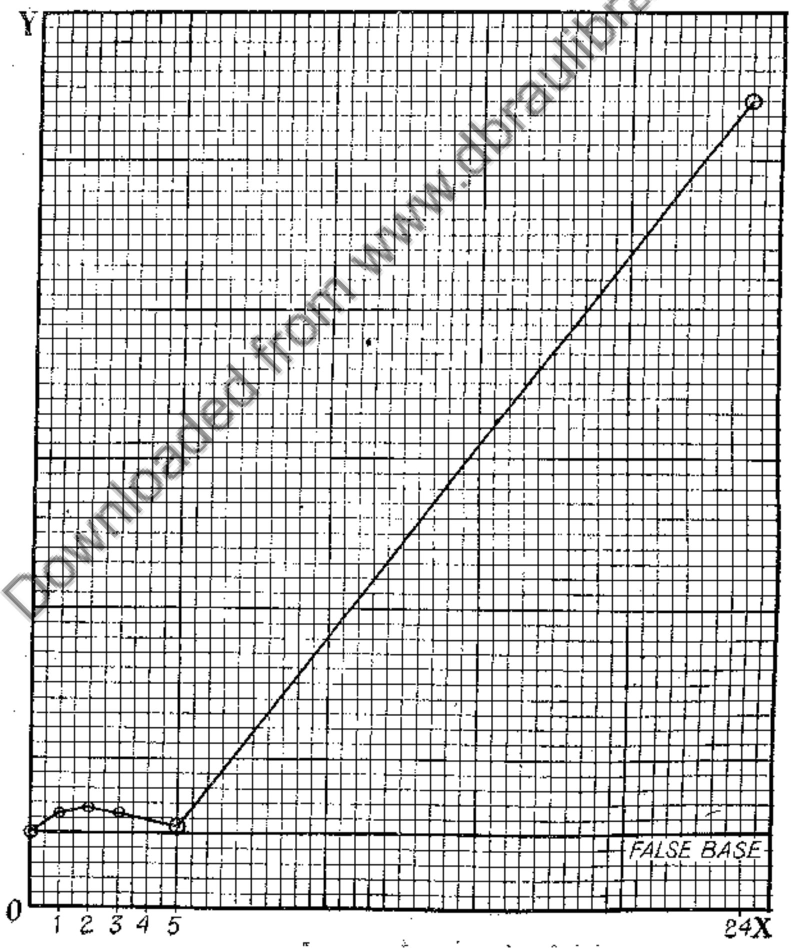
Solution :

Represent hours of incubation on OX axis and count of bacteria per c c. on broth at 20°C with salt on OY axis. Also draw a base line (false).

Scale :—

OX axis → $\frac{2}{10}''$ = 1 hr.

OY axis → $\frac{1}{10}''$ = 1000 app. Count of bacteria.



Problem 263.—Represent graphically the following :

Year	*(Balance of trade Rs. Crores)*
1950	+20
1951	+32
1952	−14
1953	−20
1954	+18
1955	− 6

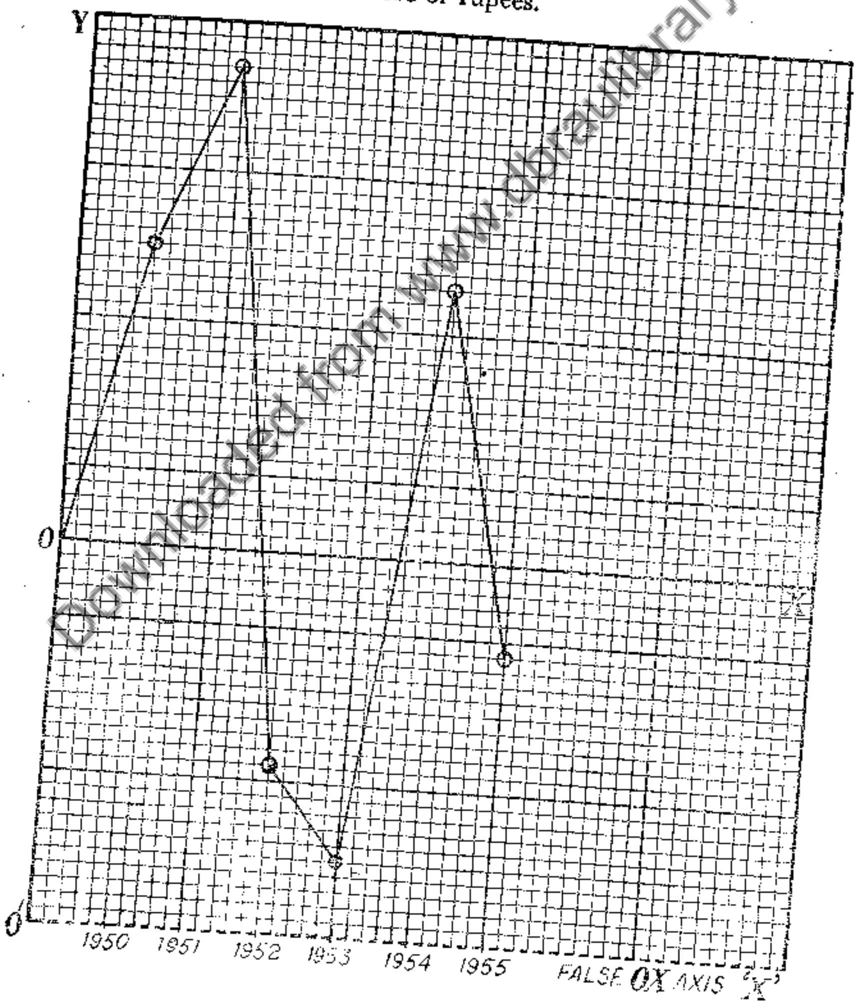
Solution :

Represent year on OX axis and balance of trade on OY axis. Positive numbers are to be plotted on the upper side of OX axis and negative figures below OX axis.

Scale :—

OX axis→$\frac{5}{10}$" = one year.

OY axis→$\frac{1}{10}$" = 1 crore of rupees.



CHAPTER IX

ASSOCIATION

THEORY OF ATTRIBUTES

I. Notation and Terminology.

(A) Attributes :—

Attribute means the quality of the observed object. When we collect data with regard to definite qualities and place them in one group, it is said to be classification according to attribute.

Classification of attributes may be in one group, which is also known as the simple classification of attributes, technically known as 'Dichotomy' (cutting into two). It can also be placed in more than one group, which is known as multiple attribute classification, as for example, in the Census Report, there are many classifications according to attributes as age, education, employment, civil conditions, etc.

(B) Notations and Terminology :—

For theoretical purposes it is necessary to have some simple notation for the classes formed and for the numbers of observations assigned to each.

The capitals A, B, C, ... will be used to denote the several attributes. An object or individual possessing the attribute A will be termed simply A. Their totals are represented by putting the letter within brackets as (A) will indicate the total frequency possessing the attribute A.

Generally, Greek letters α, β, γ, ... or small letters *a*, *b*, *c*, ... are used to denote the negative attributes. For example intelligent students are represented by A and not intelligent students by α or *a*. The frequency of these negative attributes are also represented by putting them within brackets.

In the computation of association problems often we have to put different attributes together. They are represented as follows :—

Say, intelligence is represented by A
wealthy persons by B

then intelligent and wealthy persons will be represented by AB and their frequency by (AB). In the same way we can denote intelligent and not wealthy persons by Aβ or A*b* and their frequencies by (Aβ) or (A*b*).

Similarly if there are three attributes to be represented they will be depicted as ABC and their frequency by (ABC).

(C) Order of classes and class-frequencies :—

If a class contains p attributes, it will be spoken as a class of p^{th} order and the frequency as a frequency of the p^{th} order. For example A, AB, ABC, ... etc., are regarded as attributes of the first, second, third ... etc., orders respectively.

(D) Ultimate class-frequencies :—*

Any class frequency can always be expressed in terms of class-frequencies of higher order. For the whole number of observations must clearly be equal to the number of A's added to the number of α's ; that is

$$N=(A)+(\alpha)$$ where N=universe or whole number of observations.

Similarly, the number of A's is equal to the number of A's which are B's added to the number of A's which are b's, *i.e.*

$$(A)=(AB)+(A\beta)$$

Similarly

$$(AB)=(ABC)+(AB\gamma)$$

and so on

It follows at once from the results we have just given that every class-frequency can be expressed in terms of the frequencies of the highest order, *i.e.* of order n. For any frequency can be analysed into higher frequencies, and process need only stop when we have reached the frequencies of highest order. For example, with three attributes,

$$(A)=(AB)+(A\beta)$$
$$=(ABC)+(AB\gamma)+(A\beta C)+(A\beta\gamma)$$

The classes specified by n attributes, *i.e.*, those of the highest order, are termed the *ultimate class frequencies.*

2. Consistence of data

Any class-frequencies which have been or might have been observed within one and the same universe may be said to be consistent with one another. They conform with one another, and do not in any way conflict.

Condition of consistence :—

The necessary and sufficient condition for the consistence of a set of independent class-frequencies relating to a particular universe is that no ultimate class-frequency which may be calculated from them is negative.

Thus to find out if a number of given frequencies are consistent or inconsistent, all we have to do is to find out all the ultimate frequencies. If any ultimate frequency is negative, the data are inconsistent. If no ultimate frequency is negative, the data are consistent.

*Theory of Statistics by Yule and Candall page 15.

Consistence of positive class-frequencies :

Generally the datas are given in terms of class frequencies hence the conditions of consistency are given in terms of these frequencies by expanding the ultimate frequencies and applying the condition that each expansion is not less than zero. We will consider the cases of one, two and three attributes in turn.

1. If only one attribute be noted the conditions are

$$(A) \geqslant 0$$

and $(\alpha) \geqslant 0$ *i. e.* $(A) \leqslant N$.

2. If two attributes are noted there are four ultimate frequencies (AB), $(A\beta)$, (αB) and $(\alpha\beta)$. The following conditions are obtained by expanding these ultimate frequencies and applying the condition that each of these ultimate frequencies are positive. Conditions are

(*a*) $(AB) \geqslant 0$

(*b*) $(AB) \geqslant (A)+(B)-N$

(*c*) $(AB) \leqslant (A)$

(*d*) $(AB) \leqslant (B)$

3. Now consider the case of three attributes. The number of ultimate class-frequencies are eight thus giving eight conditions.

(*a*) $(ABC) \geqslant 0$

(*b*) $(ABC) \geqslant (AB)+(AC)-(A)$ [From condition $(A\beta\gamma) \geqslant 0$.

(*c*) $(ABC) \geqslant (AB)+(BC)-(B)$ [From condition $(\alpha B\gamma) \geqslant 0$.

(*d*) $(ABC) \geqslant (AC)+(BC)-(C)$ [From condition $(\alpha\beta C) \geqslant 0$.

(*e*) $(ABC) \leqslant (AB)$ [From condition $(AB\gamma) \geqslant 0$.

(*f*) $(ABC) \leqslant (BC)$ [From condition $(A\beta C) \geqslant 0$.

(*g*) $(ABC) \leqslant (AC)$ [From condition $(\alpha BC) \geqslant 0$.

(*h*) $(ABC) \leqslant (AB)+(AC)+(BC)-(A)-(B)-(C)+N$ [From condition $(\alpha\beta\gamma) \geqslant 0$.

In these conditions (*a*) to (*d*) gives the minor limits and from (*e*) to (*h*) the major limits. It is clear that none of the major limits should be less than none of the minor limits.

Comparing (*a*) with (*h*), (*b*) with (*g*), (*c*) with (*f*) and (*d*) with (*e*) we obtain the following new conditions :

$$(AB)+(AC)+(BC) \geqslant (A)+(B)+(C)-N \quad \text{..1}$$

$$(AB)+(AC)-(BC) \leqslant (A) \quad \text{...2}$$

$$(AB)-(AC)+(BC) \leqslant (B) \quad \text{...3}$$

$$-(AB)+(AC)+(BC) \leqslant (C) \quad \text{...4}$$

These (1) to (4) conditions are used to determine the consistency of datas. If any two frequencies of second order are given from these we can find the upper and lower limits of the third frequency of second order.

These conditions may be applied to the examination of incomplete data. For they may allow us to assign limits of the unknown class-frequency.

3. Association of attributes

Association means some sort of relationship between two or more attributes or objects. Association may be complete either in the form of positive or negative. It may also be partial or, in some extreme cases this association is practically nil, which is technically known as 'Independence'.

(A) Independence :

In some of the observed cases, we find that the actual observation is just equal to the expectation. It is known as the independence of attributes, *i.e.*, there does not exist any association between the two or more observed attributes. In such a case, if A and B are two attributes, we get the same proposition of A's amongst B's as amongst β's.

$$\text{Expectation of } (AB)=\frac{(A)\times(B)}{N}$$

$$\text{Expectation of } (\alpha\beta)=\frac{(\alpha)\times(\beta)}{N}$$

But
$$(\alpha)=N-(A)$$
$$(\beta)=N-(B)$$

Attributes A and B may be regarded independent, if we find that the actual observation is just equal to expectation.

(B) Partial Association :

Partial association means that the attributes A and B are associated to each other in any way. It may be positive or negative. But, in no case, it should either be perfectly associated or disassociated.

$$\text{If } (AB) > \frac{(A)\times(B)}{N}$$

$$\text{or } (AB) < \frac{(A)\times(B)}{N}$$

we can conclude that the two attributes, A and B are associated to each other in any form. In the cases of Partial association, often we find that they are mutually associated due to the presence of a third attribute, C.

(C) Association and Disassociation :

If we proceed to study with two attributes A and B, and find that they "appear together in a larger number of cases than is to be expected if they are independent", they will be termed as associated.

Association of attributes, in general, is divided into two parts :

(*i*) **Positive Association :**

If the two attributes A and B are so related that

$$(AB) > \frac{(A) \times (B)}{N},$$

it will be known as positive association.

(*ii*) **Negative Association :**

If the two attributes A and B are so related that

$$(AB) < \frac{(A) \times (B)}{N}$$

it will be known as negative association.

[*Note* : It should be remembered that negative association is different from independence because in Independence $(AB) = \frac{(A) \times (B)}{N}$ or $(\alpha\beta) = \frac{(\alpha) \times (\beta)}{N}$ but in negative asscciation $(AB) < \frac{(A) \times (B)}{N}$]

(**D**) **Co-efficient of Association :**

Yule has given the following formula as the simplest to calculate the Co-efficient of association.

$$Q = \frac{(AB)\ (\alpha\beta) - (A\beta)\ (\alpha B)}{(AB)\ (\alpha\beta) + (A\beta)\ (\alpha B)}$$

where Q = Co-efficient of association.

The tabular form, given below, elaborately simplifies the finding out of the values of (AB), (αβ), (αB), (Aβ), (N), (α), (β), (A), (B) etc.

	A	α	Total
B	(AB)	(αB)	[B]
β	(Aβ)	(αβ)	[β]
Total	[A]	[α]	[N]

If there are three attributes, we will form a table with four divisions each, horizontally and vertically. [In the above case, there are three divisions.]

(1) Notation and Terminology

Problem 264.—The following are the number of boys observed with certain classes of defects amongst a number of school-children. A denotes development defects; B nerve signs; C low nutrition.

$$(ABC)=149 \qquad (\alpha BC)=204$$
$$(AB\gamma)=738 \qquad (\alpha B\gamma)=1762$$
$$(A\beta C)=225 \qquad (\alpha\beta C)=171$$
$$(A\beta\gamma)=1196 \qquad (\alpha\beta\gamma)=21842$$

Find the frequencies of the positive classes.

(Yule and Candall)

Solution :

$$N=(ABC)+(AB\gamma)+(\alpha BC)+(\alpha B\gamma)+(A\beta C)+(A\beta\gamma)+(\alpha\beta C)+(\alpha\beta\gamma)$$
$$=149+738+204+1762+225+1196+171+21842$$
$$=26287.$$

$$(A)=(AB)+(A\beta)=(ABC)+(AB\gamma)+(A\beta C)+(A\beta\gamma)$$
$$=149+738+225+1196.$$
$$=2308.$$

$$(B)=(AB)+(\alpha B)=(ABC)+(AB\gamma)+(\alpha BC)+(\alpha B\gamma)$$
$$=149+738+204+1762$$
$$=2853$$

$$(C)=(AC)+(\alpha C)=(ABC)+(A\beta C)+(\alpha BC)+(\alpha\beta C)$$
$$=149+225+204+171$$
$$=749$$

$$(AB)=(ABC)+(AB\gamma)=149+738=887$$
$$(BC)=(ABC)+(\alpha BC)=149+204=353$$
$$(CA)=(ABC)+(A\beta C)=149+225=374$$
$$(ABC)=149.$$

Problem 265.—Given the following positive class-frequencies, find all the ultimate-class frequencies.

$$N=23713$$
$$(A)=1618 \qquad (AB)=587$$
$$(B)=2015 \qquad (BC)=428$$
$$(C)=770 \qquad (AC)=335$$
$$(ABC)=156.$$

(Yule and Candall)

Solution :

$$(AB)=(AB\gamma)+(ABC)$$
$$\therefore\ (AB\gamma)=(AB)-(ABC)$$
$$=587-156=431$$

$(A\beta C) = (AC) - (ABC)$
$= 335 - 156 = 179$
$(\alpha BC) = (BC) - (ABC)$
$= 428 - 156 = 272$
$(\alpha\beta C) = \alpha\beta C \cdot N$
$= (1 - A)(1 - B) C . N$
$= (C - AC - BC + ABC) . N$
$= (C) - (AC) - (BC) + (ABC)$
$= 770 - 335 - 428 + 156.$
$= 163.$
$(A\beta\gamma) = A\beta\gamma . N = A(1 - B)(1 - C) \cdot N$
$= (A - AB - AC + ABC) N$
$= (A) - (AB) - (AC) + (ABC)$
$= 1618 - 587 - 335 + 156$
$= 852.$

Similarly $(\alpha B\gamma) = (B) - (AB) - (BC) + (ABC)$
$= 2015 - 587 - 428 + 156$
$= 1156.$
$(\alpha\beta\gamma) = N - (A) - (B) - (C) + (AB) + (BC) + (CA) - (ABC)$
$= 23713 - 1618 - 2015 - 770 + 587 + 428 + 335 - 156$
$= 20504.$

Problem 266.—Show that for n attributes A, B, C,.........M, $(ABC......M) \geqslant [(A) + (B) + (C) + ... + (M)] - (n - 1) N.$

where N is the total frequency.

(M.Sc. Agra, 1949, 1955, 1957)
(P.C.S., 1954)

Solution :

We have $(\alpha\beta) = (\alpha) - (\alpha B)$
$= N - (A) - (B) + (AB)$

Since no frequency can be negative,

Hence $(AB) \geqslant (A) + (B) - N$ ....(1)

Thus we can write

$$(AT) \geqslant (A) + (T) - N$$

Putting T=BC, we have

$(ABC) \geqslant (A) + (BC) - N$
$\geqslant (A) + (B) + (C) - 2N$

Now assume that it is true for m values A, B,..............., S ; so that

$(ABC.........S) \geqslant (A) + (B) + ... + (S) - (m - 1) N$

Now putting S=RX we have

$$(ABC\ldots\ldots RX) \geqslant (A)+(B)+\ldots\ldots+(RX)-(m-1)\,N$$
$$\geqslant (A)+(B)+\ldots\ldots+(R)+(X)-m\,N$$

Thus if theorem is true for $n=m$ it is found that it will be true for $n=m+1$. But it is true for $m=2$ hence for $m=3$ and hence for 4, 5,......and so on in general.

Problem 267.—At an examination at which 600 candidates appeared, boys outnumbered girls by 16 per cent. Also those passing the examination exceed in number those failing by 310. The number of successful boys choosing Science subjects was 300 while among the girls offering Arts subjects there were 25 failures. Altogether only 135 offered Arts and 33 among them failed. Boys failing in the examination numbered 18. Obtain all the class-frequencies.

(I. A.S., 1953)

Solution :

Let the Boys and Girls be denoted by A and α, Passing and Failing in examination by B and β, offering Science and Arts by C and γ respectively, we have the following data in the order of question.

$$N=600$$

$$(A)-(\alpha)=16\times\frac{600}{100}=96 \quad (1)$$
$$(B)-(\beta)=310 \quad (2)$$
$$(ABC)=300 \quad (3)$$
$$(\alpha\beta\gamma)=25 \quad (4)$$
$$(\gamma)=135 \quad (5)$$
$$(\beta\gamma)=33 \quad (6)$$
$$(A\beta)=18 \quad (7)$$

$$N=(A)+(\alpha)=(B)+(\beta)=(C)+(\gamma)$$

From (1) $(A)=348$, $(\alpha)=252$

From (2) $(B)=455$, $(\beta)=145$

From (5) $(C)=465.$

From (7) $(AB)=(A)-(A\beta)=348-18=330$

From (6) $(B\gamma)=(\gamma)-(\beta\gamma)$

$=135-33=102.$

$(BC)=(B)-(B\gamma)$

$=455-102$

$=353.$

$$(A\beta\gamma)=(\beta\gamma)-(\alpha\beta\gamma)=33-25=8$$
$$(A\beta C)=(A\beta)-(A\beta\gamma)=18-8=10$$
$$(AC)=(A\beta C)+(ABC)=10+300=310$$

$$(A\gamma)=(A)-(AC)=348-310=38$$
$$(\alpha C)=(C)-(AC)=465-310=155$$
$$(\alpha\gamma)=(\alpha)-(\alpha C)=252-155=97$$

In the same manner the other class frequencies may be obtained.

Problem 268.—100 children took three examinations. 40 passed the first, 39 passed the second and 48 passed the third. 10 passed all three, 21 failed all three, 9 passed the first two and failed the third, 19 failed the first two and passed the third. Find how many children passed at least two examinations.

(P.C.S., 1952)

Solution :

Let A, B, C denote the passing of first, second and third examination respectively, then

$$N=100$$
$$(A)=40$$
$$(B)=39$$
$$(C)=48$$
$$(ABC)=10$$
$$(\alpha\beta\gamma)=21$$
$$(AB\gamma)=9$$
$$(\alpha\beta C)=19$$

To find the children which passed at least two examinations *i.e.* to find the value of

$$(AB\gamma)+(A\beta C)+(\alpha BC)+(ABC)$$

Now
$$\begin{aligned}(ABC)+(AB\gamma)+(A\beta C)+(\alpha BC) &=(ABC)+(AB)-(ABC)+(AC)-(ABC)\\&\quad+(BC)-(ABC)\\&=(AB)+(BC)+(CA)-2(ABC)\end{aligned}$$

But
$$(\alpha\beta\gamma)=N-(A)-(B)-(C)+(AB)+(BC)+(CA)-(ABC)$$

or
$$21=100-40-39-48+(AB)+(BC)+(CA)-10$$

$$\therefore \quad (AB)+(BC)+(CA)=58$$

and
$$\begin{aligned}(AB)+(BC)+(CA)-2(ABC)&=58-20\\&=38\end{aligned}$$

Hence the number of children passing at least two examinations is 38.

Problem 269.—In a very hotly fought battle 70 per cent at least of the combatants lost an eye, 75 per cent at least lost an ear,

80 per cent at least lost an arm and 85 per cent at least lost a le
How many at least must have lost all four.

(*P.C.S. 195*
(*M. Com., All., 194*

Solution :

Let $N = 100$

Let A, B, C, D denote respectively lossing an eye, an ear, an arm and a leg. Then from the given datas :—

$$(A) \geqslant 70$$
$$(B) \geqslant 75$$
$$(C) \geqslant 80$$
$$(D) \geqslant 85$$

Now those persons which lost all the four will be denoted by (ABCD).

$$\text{Now} \quad (ABCD) = (A) + (B) + (C) + (D) - 3N$$
$$= 70 + 75 + 80 + 85 - 300$$
$$= 10$$

Hence those who lost all the four are at least 10%.

Problem 270.—Given that $(A) = (B) = (C) = \frac{1}{2}N$, and $(ABC) = (\alpha\beta\gamma)$,

Show that $2(ABC) = (AB) + (BC) + (CA) - \frac{1}{2}N.$

(*M.Sc., Agra ; 1949*)

Solution :

$$(\alpha\beta\gamma) = (1-A)\ (1-B)\ (1-C).\ N$$
$$= N - (A) - (B) - (C) + (AB) + (BC) + (CA) - (ABC)$$

or $$2(ABC) = N - \frac{N}{2} - \frac{N}{2} - \frac{N}{2} + (AB) + (BC) + (CA)$$

or $$2(ABC) = (AB) + (BC) + (CA) - \frac{1}{2}N.$$

Problem 271.—Measurements are made on a thousand husbands and a thousand wives. If the measurements of the husbands exceed the measurements of the wives is 800 cases for one measurement, in 700 cases for another, and in 660 cases for both measurement. In how many cases will both the measurements on the wife exceed the measurement on the husband ?

(*Yule and Candall*)

Solution :

Let A denote husband exceeding wife in first measurement ; β denote husband exceeding wife in second measurement. Then data are as given below.

$$N = 1000 \text{ (say)}$$

$$(A) = 800,\ (B) = 700$$

$$(AB) = 660$$

to find $\alpha\ \beta$.

$$(\alpha\beta) = N - (A) - (B) + (AB)$$
$$= 1000 - 800 - 700 + 660$$
$$= 160$$

Hence in 160 cases both the measurements on wife will exceed both the measurements on husband.

Problem 272.—In a war between white and Red forces there are more Red soldiers than white ; there are more armed whites than unarmed Reds ; there are fewer armed Reds with ammunition than unarmed whites without ammunition. Show that there are more armed Reds without ammunition than unarmed whites with ammunition.

(*Yule and Candali*)

Solution :

Let white soldier be denoted by A, armed by B, and possessed with ammunition by C respectively, we have from given data :

$$(\alpha) > (A) \qquad \ldots 1$$

$$(AB) > (\alpha\beta) \qquad \ldots 2$$

$$(A\beta\gamma) > (\alpha BC) \qquad \ldots 3$$

Now we have to prove $(\alpha B\gamma) > (A\beta C)$.

From (1), $(\alpha\beta) + (\alpha B) > (AB) + (A\beta)$

or $(\alpha\beta) + (\alpha B) > (\alpha\beta) + (A\beta)$ from 2

or $(\alpha B) > (A\beta)$

or $(\alpha B\gamma) + (\alpha BC) > (A\beta\gamma) + (A\beta C)$

or $(\alpha B\gamma) + (\alpha BC) > (\alpha BC) + (A\beta C)$ from 3

or $(\alpha B\gamma) >$ [redacted]

Problem 273.—Show that if A occurs in a larger proportion of the cases where B is than where B is not, then B will occur in a larger proportion of the cases where A is than where A is not, *i.e.* given $\frac{(AB)}{(B)} > \frac{(A\beta)}{(\beta)}$, show that $\frac{(AB)}{A} > \frac{(\alpha B)}{(\alpha)}$.

(*M.Sc. Agra, 1948*)
(*M.Sc. Agra, 1950*)

Solution :

$$\frac{(AB)}{(B)} > \frac{(A\beta)}{(\beta)}$$

or $(AB)\ (\beta) > (A\beta)\ (B)$

or $(AB)\ [N-(B)] > [(A)-(AB)]\ (B)$

or $(AB)\ .\ N > (A)\ (B)$

or $\frac{(AB)}{(B)} > \frac{(A)}{N}$

or $\frac{(AB)}{(B)-(AB)} > \frac{(A)}{N-(A)}$

or $\frac{(AB)}{(\alpha B)} > \frac{(A)}{(\alpha)}$

or $\frac{(AB)}{(A)} > \frac{(\alpha B)}{(\alpha)}$

Problem 274.—In a free vote in the House of Commons, 600 members voted. 300 Government members representing English constituencies (including Welsh) voted in favour of the motion. 25 opposition members representing Scottish constituencies voted against the motion. The Government majority among those who voted was 96. 135 of the members voting represented Scottish constituencies. 18 Government members voted against the motion. 102 Scottish members voted in favour of the motion. The motion was carried by 310 votes. Analyse the voting according to the nationality of the constituencies and party.

(Yule and Candall).

Solution :

Devoting the Government members by A, voting for the motion by B and English members by C respectively we have from the given data

$$N=600$$
$$(ABC)=300$$
$$(\alpha\beta\gamma)=25$$
$$(A)-(\alpha)=96$$
$$(\gamma)=135$$
$$(A\beta)=18$$
$$(B\gamma)=102$$
$$(B)-(\beta)=310$$

we wish to find the ultimate class frequencies. This is the case of three attributes hence there will be 8 ultimate class frequencies in which two are already known.

$$N = (A) + (\alpha) = 600$$
$$(A) - (\alpha) = 96$$
$$\therefore \quad (A) = 348$$
$$(\alpha) = 252$$

Similarly $(B) = 455$
$(\beta) = 145$
$(C) = N - (\gamma) = 465$

We have thus found all the first-order frequencies.

Again, $(AB) = (A) - (A\beta)$
$= 330$
$(BC) = (B) - (B\gamma)$
$= 353.$

Also $(\alpha\beta\gamma) = N - (A) - (B) - (C) + (AB) + (BC) + (CA) - (ABC)$

or $25 = 600 - 348 - 455 - 465 + 330 + 353 + (CA) - 300$

$\therefore \quad (CA) = 310$
$(AB\gamma) = (AB) - (ABC) = 30$
$(\alpha BC) = (BC) - (ABC) = 53$
$(A\beta C) = (AC) - (ABC) = 10$
$(\beta\gamma) = (\gamma) - (B\gamma) = 33$
$(A\beta\gamma) = (\beta\gamma) - (\alpha\beta\gamma) = 8$
$(\alpha\beta) = 127$
$(\alpha\beta C) = (\alpha\beta) - (\alpha\beta\gamma) = 102.$

N = the sum of ultimate class frequencies,

$\therefore \quad (\alpha B\gamma) = 72.$

✓**Problem 275.**—A market investigator returns the following data : Of 1000 people consulted, 811 liked chocolates, 752 liked toffee and 418 liked boiled sweets ; 570 liked chocolates and toffee, 356 liked chocolates and boiled sweets and 348 liked toffee and boiled sweets ; 297 liked all three. Show that this information as it stands must be incorrect.

(M.Sc. Agra, 1956)

Solution :

Let A, B, C denote the liking of chocolates, toffee and boiled sweets respectively then data are :

$$N = 1000$$
$$(A) = 811$$

$$(B)=752$$
$$(C)=418$$
$$(AB)=570$$
$$(AC)=356$$
$$(BC)=348$$
$$(ABC)=297$$

To show that one of the ultimate class frequency will be negative for incorrect information.

Now $(\alpha\beta\gamma)=N-(A)-(B)-(C)+(AB)+(BC)+(AC)-(ABC)$
$=1000-811-752-418+570+356+348-297$
$=-4$

since this ultimate class frequency is negative hence the information is not correct as there should not be any negative ultimate class frequency.

2. Consistence of Data

Problem 276.—If a report gives the following frequencies as actually observed, show that there must be a misprint or mistake of some sort, and that possibly the misprint consists in dropping of a 1 before the 85 given as the frequency (BC) :—

$$N=1000,\ (A)=510,\ (B)=490,\ (C)=427,\ (AB)=189,$$
$$(AC)=140 \text{ and } (BC)=85$$

(*I.A.S.*, *1949*)

Solution :

From the condition for consistency we have

$$(AB)+(BC)+(AC) \geqslant (A)+(B)+(C)-N$$
or $$(BC) \geqslant (A)+(B)+(C)-(AB)-(AC)-N$$
$$\geqslant 510+490+427-189-140-1000$$
$$\geqslant 98.$$

But 85 is less than 98 and hence 85 cannot be correct value of (BC). If (BC) is taken as 185, we find that all the remaining conditions are satisfied since

$(AB)+(AC)-(BC)=144$, which is less than (A)
$(BC)+(AC)-(AB)=136$, which is less than (C) and
$(AB)+(BC)-(AC)=234$, which is less than (B).

Problem 277.—If $(A)=50$, $(B)=60$, $(C)=50$, $(A\beta)=5$, $(A\gamma)=20$ and $N=100$, find the greatest and least possible values of (BC).

(*M.Sc. Agra*, *1952*)

Solution :

$$(AB)=(A)-(A\beta)=45$$
$$(AC)=(A)-(A\gamma)=30.$$

The following four conditions for consistence of data will give the limits of (BC)

$$(AB)+(BC)+(AC)\geqslant(A)+(B)+(C)-N \quad ...1$$
$$(AB)+(AC)-(BC)\leqslant(A) \quad ...2$$
$$(AB)-(AC)+(BC)\leqslant(B) \quad ...3$$
$$-(AB)+(AC)+(BC)\leqslant(C) \quad ...4$$

From (1) $(BC)\geqslant 50+60+50-100-45-30$
$\geqslant -15$

From (2) $(BC)\geqslant 45+30-50$
$\geqslant 25$

From (3) $(BC)\leqslant 60+30-45$
$\leqslant 45$

From (4) $(BC)\leqslant 50+45-30$
$\leqslant 65$

Since 65 is greater than 45 and 25 is greater than -15 hence the value of (BC) lies between 25 and 45 ; *i.e.*

$$25\leqslant(BC)\leqslant 45.$$

Problem 278.—If (A)=50, (B)=60, (C)=80, (AB)=35, (AC)=45 and (BC)=42, find the greatest and least possible values of (ABC).

(*M.Sc. Agra, 1949*)

Solution :

The greatest value of (ABC) is given by the relations

$$(ABC)\leqslant(AB)$$
$$(ABC)\leqslant(BC)$$
$$(ABC)\leqslant(AC)$$
$$(ABC)\leqslant(AB)+(BC)+(AC)-(A)-(B)-(C)+N$$

$\therefore$ $(ABC)\leqslant 35$
$(ABC)\leqslant 45$
$(ABC)\leqslant 42$

$\therefore$ $(ABC)\leqslant 35$ hence greatest value of (ABC) is 35.

The least value of (ABC) is given by the relations

$$(ABC)\geqslant 0$$
$$(ABC)\geqslant(AB)+(AC)-(A)$$

$(ABC) \geqslant (AC)+(BC)-(C)$
$(ABC) \geqslant (AB)+(BC)-(B)$

$\therefore \quad (ABC) \geqslant 0$
$(ABC) \geqslant 35+45-50 \geqslant 30$
$(ABC) \geqslant 45+42-80 \geqslant 7$
$(ABC) \geqslant 35+42-60 \geqslant 17$

Thus the least value of (ABC) is 30

Hence $30 \leqslant (ABC) \leqslant 35$.

Problem 279.—50 per cents of the imports of barley into a country come from the Dominions ; 80 per cent of the total imports go to brewing ; 75 per cent of the imports are grown in the Northern hemisphere ; 80 per cent. of Northern-grown barley goes to brewing ; 100 per cent. of foreign Southern-grown barley goes to stock feeding. Show that the foreign Northern-grown barley which goes to brewing cannot be less than 30 per cent nor more than 60 per cent. of the total imports. [It is assumed that brewing and stock feeding are the only two uses to which imported barley is put.]

(M.Sc. Agra, 1954)

Solution :

Denoting Imports from Dominions by A, going to brewing by B, and grown in Northern Hemisphere by C we have,

$N = 100$ (say)

then $(A)=50$, $(B)=80$, $(C)=75$,

$$(BC) = \frac{80}{100} \times 75 = 60$$

and $(\alpha\gamma) = (\alpha\beta\gamma)$ {as 100 per cent of foreign southern-grown barley goes to stock feeding}

To find the limits of (αBC).

Let $(\alpha BC) = x$

then $(ABC) = (BC) - (\alpha BC)$
$= 60 - x$

Also $(\alpha\beta\gamma) = N-(A)-(B)-(C)+(AB)+(BC)$
$+(CA)-(ABC)$

and $(\alpha\gamma) = N-(A)-(C)+(AC)$.

Comparing we get $(AB)+(BC) = (ABC)+(B)$

or $(AB) = 80 - x$

Using these values we get

$(A\beta) = (A)-(AB)$
$= x - 30$

and $(\alpha\beta) = N - (A) - (B) + (AB)$

$= 50 - x$

Since these frequencies cannot be negative hence

$$30 \leqslant x \leqslant 50$$

Hence the result.

Problem 280.—The following summary appears in a report on a survey covering, 1000 fields. Scrutinize the numbers, and point out if there be any mistake, or misprint in them :—

Manured fields 510	
Irrigated fields 490	
Fields growing improved varieties	427
Fields both irrigated and manured	189
Fields both manured and growing improved varieties	140
Fields both irrigated and growing improved varieties	85

(*I.A.S., 1949*)

Solution :

Denoting manured fields by A, irrigated fields by B, and fields growing improved varieties by C we have

$$N = 1000.$$

$(A) = 510, \quad (AB) = 189$

$(B) = 490, \quad (AC) = 140$

$(C) = 427, \quad (BC) = 85.$

Applying the condition for consistence of data we have

$(AB) + (AC) + (BC) \geqslant (A) + (B) + (C) - N$

or $189 + 140 + (BC) \geqslant 510 + 490 + 427 - 1000$

or $(BC) \geqslant 98.$

but given value of (BC) is 85 which is less hence there is mistake or misprint in the data.

Problem 281.—Given that $(A) = (B) = (C) = \frac{N}{2}$ and 80 per cent. of the A's are B's, 75 per cent. of A's are C's, find the limits of the percentage of B's that are C's.

Solution :

Let $(A) = (B) = (C) = \frac{N}{2} = 100$

then $(AB) = 80$

$(AC) = 75$

To find the limits of (BC).

we have

$$(AB)+(AC)+(BC) \geqslant (A)+(B)+(C)-N \quad \ldots\ldots(1)$$
$$(AB)+(AC)-(BC) \leqslant (A) \quad \ldots\ldots(2)$$
$$(AB)-(AC)+(BC) \leqslant (B) \quad \ldots\ldots(3)$$
$$-(AB)+(AC)+(BC) \leqslant (C) \quad \ldots\ldots(4)$$

From (1) $(BC) \geqslant -55$

From (2) $(BC) \geqslant 55$

From (3) $(BC) \leqslant 95$

From (4) $(BC) \leqslant 105$

Hence $55 \leqslant (BC) \leqslant 95$

Hence the value of B's that can be C's lies between 55 per cent and 95 per cent.

Problem 282.—Among the adult population of a certain town 50 per cent of the population are male, 60 per cent are wage earners and 50 per cent are 45 years of age or over. 10 per cent of the males are not wage-earners and 40 per cent of the males are under 45 years. Can we infer anything about what percentage of the population of 45 or over are wage-earners ?

Solution :

Denoting the attributes male, wage earner, and 45 years old or more by A, B, C respectively we have the following data :

$$N=100 \text{ (say)}$$

$(A)=50,\quad (A\beta)=\frac{10}{100}\times 50=5$

$(B)=60,\quad (A\gamma)=\frac{40\times 50}{100}=20$

$(C)=50$

We have to find the limits of (BC).

$$(AB)=(A)-(A\beta)=45$$
$$(AC)=(A)-(A\gamma)=30$$

Applying the conditions for consistence of data we have

$$(AB)+(BC)+(AC) \geqslant (A)+(B)+(C)-N \quad \ldots\ldots(1)$$
$$(AB)+(AC)-(BC) \leqslant (A) \quad \ldots\ldots(2)$$
$$(AB)-(AC)+(BC) \leqslant (B) \quad \ldots\ldots(3)$$
$$-(AB)+(AC)+(BC) \leqslant (C) \quad \ldots\ldots(4)$$

From (1) $(BC) \geqslant -15$

From (2) $(BC) \geqslant 25$

From (3) $(BC) \leqslant 45$

From (4) $(BC) \leqslant 65$

$\therefore \quad 25 \leqslant (BC) \leqslant 45.$

Since (C)=50, we have as two limits of (BC), 50%, and 90% of (C) ; *i.e.*, the percentage of the population of 45 years old or more who are wage earners lies between 50 and 90 per cent.

Problem 283.—A penny is tossed three times and the results, heads and tails, noted. The process is continued until there are 100 sets of threes. If 69 cases heads fell first, 49 cases heads fell second, and 53 cases heads fell third. In 33 cases heads fell both first and second, and in 21 cases heads fell both second and third. Show that there must have been atleast 5 occasions on which heads fell three times, and that there could not have been more than 15 occasions on which tails fell three times, though there need not have been any.

Solution :

Let A, B, C denote the heads felling first, second and third times respectively ; then according to given data :

$$(A)=69,\ (B)=49,\ (C)=53$$
$$(AB)=33,\ (BC)=21,\ N=100$$

to prove $(ABC) \geqslant 5$

and $(\alpha\beta\gamma) \leqslant 15$

From the conditions for consistence of data we have

$$\begin{aligned}(ABC) &\geqslant (BC)+(AB)-(B)\\ &\geqslant 21+33-49\\ &\geqslant 5\end{aligned}$$

Also

$$\begin{aligned}(\alpha\beta\gamma) &= N-(A)-(B)-(C)+(AB)+(BC)+(AC)-(ABC)\\ &= N-(A)-(B)-(C)+(AB)+(BC)+(AC\beta)\\ &= -17+(A\beta C)\end{aligned}$$

Now $(\beta C)=(C)-(BC)$

$=32$

And $(AC\beta)$ is sub-class of (βC) hence

$$(A\beta C) \leqslant 32$$

$$\therefore \quad (\alpha\beta\gamma) \leqslant -17+32$$
$$\leqslant 15$$

Problem 284.—The following is a summary of the statistical features of a census of ration cards :—

Item No.	*Category*	*Total number of cards belonging to the category*
1.	The whole of the census	1000
2.	Permanent resident	510
3.	Males	490
4.	Consumers of rice	427
5.	Permanent male resident	189
6.	Consumer of rice among permanent residents	140
7.	Males consuming rice	97

Show that the entry against item No. 7 is inconsistent with entries against all the previous items, namely, 1, 2, 3, 4, 5, and 6 taken together.

(*I.A.S., 1947*)

Solution :

Denoting Permanent residents by A, Males by B, and consumers of rice by C we have

$$N = 1000$$
$$(A) = 510$$
$$(B) = 490$$
$$(C) = 427$$
$$(AB) = 189$$
$$(AC) = 140$$
$$(BC) = 97$$

From the condition of consistency we have

$$(AB)+(BC)+(AC) \geqslant (A)+(B)+(C)-N$$

or
$$(BC) \geqslant (A)+(B)+(C)-N-(AB)-(AC)$$
$$\geqslant 510+490+427-1000-189-140$$
$$\geqslant 98$$

But 97 is less than this hence entry against item 7 is not consistent.

Problem 285.—Given that $(A)=(B)=(C)=\frac{1}{2}N$, and that $\frac{(AB)}{N}=\frac{(AC)}{N}=p$, find what must be greatest and least values of p in order that we may infer that $\frac{(BC)}{N}$ exceeds any given value, say, q.

(*Yule and Candall*)

Solution :

$$\frac{(BC)}{N} > q,$$

we have

$$(AB)+(AC)+(BC) \geqslant (A)+(B)+(C)-N \quad \ldots \quad (1)$$
$$(AB)+(AC)-(BC) \leqslant (A) \quad \ldots \quad (2)$$
$$(AB)-(AC)+(BC) \leqslant (B) \quad \ldots \quad (3)$$
$$-(AB)+(AC)+(BC) \leqslant (C) \quad \ldots \quad (4)$$

From (1) $$\frac{(BC)}{N} \geqslant \frac{1}{2}-2p \quad \ldots \quad (5)$$

From (2) $\quad \frac{(BC)}{N} \geqslant 2p - \frac{1}{2} \quad \ldots \quad (6)$

From (3) & (4) $\quad \frac{(BC)}{N} \leqslant \frac{1}{2} \quad \ldots \quad (7)$

Since $\frac{(BC)}{N}$ should be greater than q

hence from (5) $\quad \frac{1}{2} - 2p \geqslant q$

or $\quad p \leqslant \frac{1}{4} - \frac{1}{2} q \quad (8)$

and from (6) $\quad 2p - \frac{1}{2} \geqslant q$

or $\quad p \geqslant \frac{1}{2} q + \frac{1}{4} \quad \ldots \quad (9)$

Since $\frac{(BC)}{N} < \frac{(B)}{N}$

Hence $p < \frac{1}{2} \quad \ldots \quad (10)$

Also, p cannot be negative hence lower limit of p is zero.

i.e. $\quad p > 0 \quad \ldots \quad (11)$

Combining (8) with (11) we have

$$0 \leqslant p \leqslant \tfrac{1}{4}(1-2q)$$

Combining (9) with (10) we have

$$\tfrac{1}{4}(1+2q) \leqslant p \leqslant \tfrac{1}{2}$$

Problem 286.—Show that if $\frac{(A)}{N} = x$, $\frac{(B)}{N} = 2x$, $\frac{(C)}{N} = 3x$ and $\frac{(AB)}{N} = \frac{(AC)}{N} = \frac{(BC)}{N} = y$

the value of neither x nor y can exceed $\frac{1}{4}$.

(Yule and Candall)

Solution :

The condition of consistence are

$$\left.\begin{aligned} (BC) &\geqslant (B)+(C)-N \\ (AC) &\geqslant (C)+(A)-N \\ (AB) &\geqslant (A)+(B)-N \end{aligned}\right\} \quad (1)$$

Also $\quad \left.\begin{aligned} (BC) &\leqslant (B) \\ (AC) &\leqslant (C) \\ (AB) &\leqslant (A) \end{aligned}\right\} (i)$ $\quad$ [since (AB), (BC), (CA) are sub-classes of (A), (B), (C) respectively.]

From (1) $\quad y \geqslant 5x - 1$

$$y \geqslant 4x - 1$$

$$y \geqslant 3x - 1$$

From (2) $y \leqslant 2x$

$y \leqslant 3x$

$y \leqslant x$

$\therefore \quad 5x-1 \leqslant y \leqslant x$

This shows that

$x \geqslant 5x-1$

or $4x \leqslant 1$

or $x \leqslant \frac{1}{4}$

Hence x and y cannot exceed $\frac{1}{4}$

Association of Attributes

Problem 287.— Given the following ultimate class-frequencies, find the frequencies of the positive and negative classes and the whole number of observation, N :

$(AB)=200$; $(Ab)=100$
$(aB)=160$; $(ab)=80$.

(Statistic—Ghosh and Choudhry. pp. 506)

Solution :

To find N, A, B, a and b.

$N=(AB)+(ab)+(aB)+(Ab)$
$=200+100+160+80$
$=540.$

$A=(AB)+(Ab)$
$=200+100$
$=~~280~~$ 300
$B=(AB)+(Ba)$
$=200+160=360$ } +attributes

$a=N-(A) \quad =540-280=260$
$b=N-(B) \quad =540-360=180$ | —attributes.

Problem 288.—Find the association between eye-colour of husband and eye-colour of wife from the following data :—

Husbands with light eyes and wives with light eyes	=1236
Husbands with light eyes and wives with not light eyes	= 856
Husbands with not light eyes and wives with light eyes	= 528
Husbands with not light eyes and wives with not light eyes	= 476

(Statistics—Theory and Practice—By Ghosh and Choudhry pp. 506)

Solution :

Let husbands with light eyes be represented by $=A$
husbands with not light eyes $=a$
wives with light eyes $=B$
wives with not light eyes $=b$.

Thus, we get :— AB=1236 ; Ab=856 ; aB=528 ; ab=476

$$\text{Co-efficient of Association (Q)} = \frac{(AB)\ (ab)-(Ab)\ (aB)}{(AB)\ (ab)+(Ab)\ (aB)}$$

$$=\frac{[1236\times476]-856\times528]}{[1236\times476]+[856\times528]}$$

$$=\frac{588336-66768}{588336+66768}=\frac{521568}{655104}$$

$$=+\cdot8$$

That indicates a high degree of positive association between eye colour of husband and eye colour of wife.

Problem 289.—Investigate the association between darkness of eye-colour in father and son from the following data :

Fathers with dark eyes and sons with not dark eyes =237
Fathers with dark eyes and sons with dark eyes =150
Fathers with not dark eyes and sons with dark eyes =267
Fathers with not dark eyes and sons with not dark eyes=2346
(*Statistics—Theory and Practice—Ghosh & Choudhry. pp. 455*)

Solution :

Let Fathers with dark eyes=A
Fathers with not dark eyes=a
Sons with dark eyes =B
Sons with not dark eyes =b

Ab=237
AB=150
aB=267
ab=2346

$$\text{Co-efficient of association (Q)} = \frac{(AB)\ (ab)-(Ab)\ (aB)}{(AB)\ (ab)+(Ab)\ (aB)}$$

$$=\frac{[150\times2346]-[237\times267]}{[150\times2346]+[237\times267]}$$

$$=\frac{351900-63279}{351900+63279}$$

$$=\frac{288621}{415179}$$

$$=+\cdot69$$

Thus, we find that there is a positive association of high degree in the eye-colours of fathers and sons.

Problem 290.—Can vaccination be regarded as a preventive measure for small-pox from the data given below ?

'Of 1482 persons in a locality exposed to small pox, 368 in all were attacked'.

'Of 1482 persons, 343 had been vaccinated and of these only 35 were attacked'.

(*Alld., M.Com., 1944*).

Solution :

Let attacked $=A$

not attacked $=a$

Vaccinated $=B$

not vaccinated $=b$

	A	a	
B	AB 35	aB 308	[B] 343
b	Ab 368	ab 771	[b] 1139
	[A] 403	[a] 1079	[N] 1482

Not vaccinated and attacked $=(Ab)=368$

Vaccinated and attacked $=(AB)=35$

N $=1482$

$$\text{The Co-efficient of association } (Q)=\frac{(AB)\,(ab)-(aB)\,(Ab)}{(AB)\,(ab)+(aB)\,(Ab)}$$

$$=\frac{[35\times771]-[308\times368]}{[35\times771]+[308\times368]}$$

$$=\frac{26985-113344}{26985+113344}$$

$$=-\frac{86359}{140329}$$

$$=-\cdot615$$

Thus, we find that vaccination and small pox are in a high degree of negative association.

∴ Vaccination is a preventive of small pox of a high degree. **Ans.**

Problem 291.—From the figures in the following table, compare the association between literacy and unemployment in rural and urban areas.

	Urban	*Rural*
Total adult males	25 lakhs	200 lakhs
Literate males	10 ,,	40 ,,
Unemployed males	5 ,,	12 ,,
Literate and unemployed males	3 ,,	4 ,,

(*Alld., M.A., 1937, Patna, M.A., 1943*)

Solution :

Let literate males be represented by A
non-literate males by a
Employed males by B
Unemployed males by b
Total by N

(*i*) **Urban :**

	A	a	N	
B	(AB) 7	(aB) 13	(B) 20	
b	(Ab) 3	(ab) 2	(b) 5	in Lakhs
N	(A) 10	(a) 15	(N) 25	

Co-efficient of Association (Q) $= \frac{(AB)\ (ab) - (aB)\ (Ab)}{(AB)\ (ab) + (aB)\ (Ab)}$

$= \frac{(7 \times 2) - (13 \times 3)}{(7 \times 2) + (13 \times 3)}$

$= \frac{14 - 59}{14 + 39}$

$= \frac{-25}{39} = -\cdot 47$ approximately.

Thus, we find that there is negative association between literacy and unemployment.

(*ii*) **Rural :**

	A	a	N	
B	(AB) 36	(aB) 152	(B) 188	
b	(Ab) 4	(ab) 8	(b) 12	in lakhs
N	(A) 40	(a) 160	(N) 200	

$$\text{Coefficient of association } (Q) = \frac{(AB)\ (ab) - (aB)\ (Ab)}{(AB)\ (ab) + (aB)\ (Ab)}$$

$$= \frac{(36 \times 8) - (152 \times 4)}{(36 \times 8) + (152 \times 4)}$$

$$= \frac{288 - 608}{288 + 608}$$

$$= \frac{-320}{896} = -{\cdot}35 \text{ approximately.}$$

Thus, we conclude that (*i*) In both the urban and rural areas, there is negative association between literacy and unemployment.

(*ii*) In urban area the literate persons are less unemployed than in rural area.

Problem 292.—In the course of anti-malarial work quinine was administered to 606 adults out of a total population of 3540. The incidence of malarial fever is shown below. Discuss the preventine value of quinine.

	Fever	No-fever	Total
Quinine	19	587	606
No-quinine	193	2741	2934
Total	212	3328	3540

(*Calcutta, M.A., 1935*)

Solution :

Let fever be represented by A
No-fever by *a*
Quinine by B
No-quinine by *b*
Total by N

	A	*a*	N
B	(AB) 19	(*a*B) 587	(B) 606
b	(A*b*) 193	(*ab*) 2741	(*b*) 2934
N	(A) 212	(*a*) 3328	(N) 3540

$$\text{Co-efficient of association (Q)} = \frac{(AB)\ (ab) - (aB)\ (Ab)}{(AB)\ (ab) + (aB)\ (Ab)}$$

$$= \frac{(19 \times 2741) - (587 \times 193)}{(19 \times 2741) + (587 \times 193)}$$

$$= \frac{52079 - 113291}{52079 + 113291}$$

$$= \frac{-61212}{165370}$$

$$= -\cdot 37 \text{ approximately.}$$

Thus, we conclude (*i*) Quinine and malaria are negatively associated.

(*ii*) Quinine is preventive of fever equivalent to ·37 degree.

Problem 293.—The male population of the U.P. is 250 lakhs. The number of literate males is 20 lakhs, and total number of male criminals is 26 thousands. The number of literate male criminals is 2 thousands. Do you find any association between literacy and criminality ?

(*Agra, M.A., 1943*)
(*Agra, M.Sc., 1951*)

Solution :

Male Population (N) = 250 lakhs.

Literate males (A) = 20 lakhs.

Male criminals (B) = 26 thousands = 13/50 lakhs.

Literate male criminals (AB) = 2 thousands = 1/50 lakh.

∴ a (*i.e.* non-literate males) = N − (A).
= 250 − 20 lakhs.
= 230 lakhs.

∴ b (*i.e.* Male non-criminals) = N − (B)
= 250 − 13/50 lakhs.
= 249·74 lakhs.

AB = 1/50 lakhs.

ab = (230 × 249·74) lakhs.

aB = (230 × 13/50) lakhs.

Ab = (20 × 249·74) lakhs.

$$\therefore \quad \text{Co-efficient of association (Q)} = \frac{(AB)\ (ab) - (aB)\ (Ab)}{(AB)\ (ab) + (aB)\ (Ab)}$$

$$= \frac{[\frac{1}{50} \times (230 \times 249\cdot 74)] - [(230 \times \frac{13}{50})(20 \times 249\cdot 74)]}{[\frac{1}{50} \times (230 \times 249\cdot 74)] + [(230 \times \frac{13}{50})(20 \times 249\cdot 74)]}$$

$$=\frac{1148\cdot804-298589\cdot04}{1148\cdot804+298589\cdot04}$$

$$=\frac{-297440\cdot236}{299737\cdot844}$$

$$=-\cdot9 \text{ approximately.}$$

Thus, we find that (*i*) There is a very high degree of negative association between literacy and criminality.

(*ii*) Literacy checks criminality to a very large extent.

Problem 294.—Investigate if there is any association between extravagance in father and son from the following :

Extravagant sons with extravagant fathers	400
Miser sons with extravagant fathers	150
Extravagant sons with miser fathers	170
Miser sons with miser fathers	1150

Solution :

Let extravagant sons be represented by A
miser sons by *a*
extravagant fathers by B
miser fathers by *b*
Total by N

	A	*a*	N
B	(AB) 400	(*a*B) 150	(B) 550
b	(A*b*) 170	(*ab*) 1150	(*b*) 1320
N	(A) 570	(*a*) 1300	(N) 1870

∴ Co-efficient of association $(Q)=\frac{(AB)(ab)-(aB)(Ab)}{(AB)(ab)+(aB)(Ab)}$

$$=\frac{[400\times1150]-[150\times170]}{[400\times1150]+[150\times170]}$$

$$=\frac{460000-25500}{460000+25500}$$

$$=\frac{434500}{485500}$$

$$=+\cdot89 \text{ approximately.}$$

Thus, we find that there is a positive association of high degree in the extravagance between father and son. That is, in general, an extravagant father have an extravagant son.

Problem 295.—In an antimalarial campaign in a certain area, quinine was administered to 812 persons out of a total population of 3248. The number of fever cases is shown below :

Treatment	Fever	No Fever
Quinine	20	792
No-Quinine	220	2216

Discuss the usefulness of quinine in checking malaria.

(*Agra, M.A., 1953*)
(*P.C.S., 1941*)

Solution :

Let Fever be represented by A
No-fever by α
Quinine by B
No-quinine by β

Treatment	Fever A	No Fever α	Total
Quinine B	(AB) 20	(αB) 792	812
No-Quinine β	(Aβ) 220	($\alpha\beta$) 2216	2436

Yule co-efficient of association (Q) $= \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$

$$= \frac{[20 \times 2216] - [220 \times 792]}{[20 \times 2216] + [220 \times 792]}$$

$$= \frac{44320 - 174240}{44320 + 174240}$$

$$= \frac{-129920}{218560}$$

$$= -\cdot 58$$

The above shows that the quinine treatment and attack from fever are negatively associated.

Thus, it may be concluded that quinine is effective in checking Malaria.

Problem 296.—A census revealed the following figures of the blind and the insane in two age-groups in a certain population.

	Age-group 15-25 years	*Age group over 75 years*
Total Population	2,70,000	1,60,200
Number of blind	1,000	2,000
Number of Insane	6,000	1,000
Number of insane among the blind	19	9

(*a*) Obtain a measure of the association between blindness and insanity in each of the two age groups.

(*b*) Do you consider that blindness and insanity are associated or disassociated with each other in the two age-groups, or more in one age-group than in the other?

(M.A., Agra, 1955)
(P.C.S., 1948)
(Raj., M.A., 1954)

Solution :

(*i*) [**Age between 15 to 25 years**]

Let Blind be expressed by A
not blind by *a*
Insane by B
Sane by *b*

Total Population = N

	Blind A	Not Blind *a*	N
Insane B	[AB] 19	[*a*B] 5981	[B] 6000
Sane *b*	[A*b*] 981	[*ab*] 263019	[*b*] 264000
N	[A] 1000	[*a*] 269000	[N] 270000

$$\text{Co-efficient of association } (Q) = \frac{(AB)\ (ab) - (Ab)\ (aB)}{(AB)\ (ab) + (Ab)\ (aB)}$$

$$= \frac{[19 \times 263019] - [5981 \times 981]}{[19 \times 263019] + [5981 \times 981]}$$

$$= \frac{4997361 - 5867361}{4997361 + 5867361}$$

$$= \frac{-870000}{10864722}$$
$$= -\cdot 08 \text{ approximately.}$$

(ii) [**Age over 75 years**]

	Blind A	Not Blind a	N
Insane B	(AB) 9	(aB) 991	[B] 1000
Sane b	(Ab) 1991	(ab) 157209	[b] 159200
	[A] 2000	[a] 158200	[N] 160200

Co-efficient of association (Q) $= \frac{(AB)\ (ab) - (Ab)\ (aB)}{(AB)\ (ab) + (Ab)\ (aB)}$

$$= \frac{[9 \times 157209] - [991 \times 1991]}{[9 \times 157209] + [991 \times 1991]}$$
$$= \frac{1414881 - 1973081}{1414881 + 1973081}$$
$$= \frac{-558200}{3387962}$$
$$= -\cdot 16 \text{ approximately.}$$

Thus, we find that there is negative association between blindness and insanity in both the cases. The degree of negative association is less in the age-group of 15-25 and is more in the over 75 age-group.

Solution of (b) part :

If $AB < \frac{A \times B}{N}$, then A and B are negatively associated, or simply disassociated. Our example is of this type. Over 75 years age group is more disassociated.

Problem 297.—Calculate the coefficient of association between extravagance in fathers and sons from the following data :—

Extravagant fathers with extravagant sons	327
Extravagant fathers with miserly sons	545
Miserly fathers with extravagant sons	741
Miserly fathers with miserly sons	235

Solution :

Let extravagant fathers be represented by A
miserly fathers by *a*
extravagant sons by B
miserly sons by *b*

	Extravagant fathers A	Miserly fathers *a*	
Extravagant sons B	[AB] 327	[*a*B] 741	[B] 1068
Miserly sons *b*	[A*b*] 545	(*ab*) 235	[*b*] 780
	[A] 872	(*a*) 976	1848

Applying Yule's formula to find Co-efficient of Association (Q), we find—

$$Q = \frac{(AB)\ (ab) - (aB)\ (Ab)}{(AB)\ (ab) + (aB)\ (Ab)}$$

$$= \frac{[327 \times 235] - [741 \times 545]}{[327 \times 235] + [741 \times 545]}$$

$$= \frac{76845 - 403845}{76845 + 403845}$$

$$= \frac{-327000}{480690}$$

$$= -\cdot 67 \text{ approximately}$$

Hence, we conclude that there is disassociation between extravagance in fathers and sons.

Problem 298.—Calculate the coefficient of association between extravagance in fathers and sons from the following data :

Extravagant fathers with extravagant sons	=	237
Extravagant fathers with miserly sons	=	545
Miserly fathers with extravagant sons	=	741
Miserly fathers with miserly sons	=	235

(*Luck., M. Com., 1947*)

Solution :

Let extravagant fathers be represented by A
miserly fathers by *a*
Extravagant sons by B
miserly sons by *b*

	A	*a*	
B	(AB) 237	(*a*B) 741	[B] 978
b	(A*b*) 545	(*ab*) 235	[*b*] 780
	[A] 782	[*a*] 976	[N] 1758

Applying Yule's formula to find co-efficient of association (Q), we find :—

$$Q = \frac{(AB)\ (ab) - (aB)\ (Ab)}{(AB)\ (ab) + (aB)\ (Ab)}$$

$$= \frac{[237 \times 235] - [741 \times 545]}{[237 \times 235] + [741 \times 545]}$$

$$= \frac{55695 - 403845}{55695 + 403845}$$

$$= \frac{-348150}{459540}$$

$$= -\cdot 757$$

This indicates a high degree of negative association between the extravagance of father and son.

Problem 299.—1660 candidates appeared for a competitive examination, 422 were successful. 256 had attended a coaching class and of these 150 came out successful. Estimate the utility of the coaching class.

(*Agra, M. Com., 1947*)

Solution :

Let the total candidates be represented by [N] (1660)
Successful candidates by [A] 422)
Unsuccessful candidates by [*a*] (1238)
Candidates attending coaching class by B (422)
Candidates attending coaching class and successful (AB) (150)
Candidates attending coaching class but unsuccessful (αB) (106)
Candidates not attending coaching class by *b* (1238)

	A	a	
B	(AB) 150	(aB) 106	(B) 256
b	(Ab) 272	(ab) 1132	(b) 1 4
	(A) 422	(a) 1238	[N] 1660

$$\text{Coefficient of association (Q)} = \frac{(AB)(ab)-(aB)(Ab)}{(AB)(ab)+(aB)(Ab)}$$

$$= \frac{[150 \times 1132]-[106 \times 272]}{[150 \times 1132]+[106 \times 272]}$$

$$= \frac{169800-28832}{169800+28832}$$

$$= \frac{140968}{198632}$$

$$= +\cdot 709 \text{ approximately.}$$

This indicates that there is a very high degree of positive association between coaching and success in the competitive examination. That is, coaching helps in success in the competitive examination to a great extent.

Problem 300.—In an experiment on immunization of cattle from tuberculosis, the following results were obtained :—

	Died or affected	Unaffected
Inoculated	12	26
Not inoculated	16	6

Examine the effect of vaccine in controlling susceptibility to tuberculosis.

(I.A.S., 1948)

Solution :

Let Inoculated be represented by A
Not inoculated by a
Died or affected by B
Unaffected by b.

Inoculated and died (AB)=12
Inoculated and not affected (Ab)=26
Not inoculated and died (aB)=16
Not inoculated and unaffected (ab)=6

Yule's coefficient of association is given by :—

$$Q=\frac{(AB)(ab)-(Ab)(aB)}{(AB)(ab)+(Ab)(aB)}$$

$$=\frac{[12\times6]-[26\times16]}{[12\times6]+[26\times16]}$$

$$=\frac{72-416}{72+416}$$

$$=\frac{-344}{488}$$

$$=-\cdot7 \text{ approximately.}$$

Thus, it indicates that there is negative association between vaccine and susceptibility to tuberculosis. That is, vaccine prevents the attack of tuberculosis to a very considerable extent.

Problem 301.—The following table is reproduced from a memoir written by Karl Pearson.

		Eye colour in Sons	
		not light	Light
	Not light	230	148
Eye colour in fathers :			
	Light	151	471

Discuss whether the colour of the son's eye is associated with that of the father.

(I.A.S., 1949)

Solution :

Let light eye colour of father be represented by A
Not light colour of father's eye by a
Light colour of son's eye by B
Not light colour of son's eye by b.

Not light eye colour of father and son $(ab)=230$
Not light eye colour of father and light of son $(aB)=148$
Light of father and light of son $(AB)=471$
Light of father and not light of son $(Ab)=151$

$$\text{Co-efficient of association } (Q)=\frac{(AB)(ab)-(Ab)(aB)}{(AB)(ab)+(Ab)(aB)}$$

$$=\frac{[471\times151]-[151\times148]}{[471\times151]+[151\times148]}$$

$$=\frac{71121-22348}{71121+22348}$$

$$=\frac{48773}{93469}$$

$$=+\cdot5 \text{ approximately.}$$

This indicates that there is a fairly high degree of positive association between the eye colour of father and son.

Problem 302.—Can inoculation be regarded as a preventive measure for cholera from the data given below :

> "Of 593 men in a locality exposed to cholera, 147 were attacked.
>
> Of 593 men, 137 had been inoculated and of these only 14 were attacked".

(P.C.S., 1951)

Solution :

Let total number of observations be represented by N (593)
People inoculated be represented by B (137)
People not inoculated by b (593 − 137 = 456)
People attacked by A (147)
People not attacked by a (593 − 147 = 446)
Inoculated and attacked by AB (14)
Inoculated and not attacked by Ab (137 — 14 = 123).

	A	a	
B	(AB) 14	(aB) 123	[B] 137
(b)	(Ab) 123	(ab) 333	[b] 456
	[A] 147	[a] 446	[N] 593

$$\text{Co-efficient of association } (Q) = \frac{(AB)(ab) - (Ab)(aB)}{(AB)(ab) + (Ab)(aB)}$$

$$= \frac{[14 \times 333] - [123 \times 123]}{[14 \times 333] + [123 \times 123]}$$

$$= \frac{4662 - 15129}{4662 + 15129}$$

$$= \frac{-10467}{19791}$$

$$= -\cdot 5 \text{ approximately.}$$

Thus, we find that inoculation and attack of cholera are negatively associated. That is, inoculation checks the attack of cholera.

Problem 303.—Calculate the co-efficient of association between intelligence in father and son from the following data :

Intelligent fathers with intelligent sons	248
Intelligent fathers with dull sons	81
Dull fathers with intelligent sons	92
Dull fathers with dull sons	579

(*Raj, M.A.*, *1948*)

Solution :

Let intelligent fathers be represented by A
Dull fathers by *a*
Intelligent sons by B
Dull sons by *b*.

	A	*a*	
B	(AB) 248	(*a*B) 92	[B] 340
	(A*b*) 81	(*ab*) 579	[*b*] 660
	[A] 329	[*a*] 671	[N] 1000

Yule's co-efficient of association (Q) $= \frac{(AB)(ab) - (Ab)(aB)}{(AB)(ab) + (Ab)(aB)}$

$$= \frac{[248 \times 579] - [81 \times 92]}{[248 \times 579] + [81 \times 92]}$$

$$= \frac{143592 - 7452}{143592 + 7452}$$

$$= \frac{136140}{151044}$$

$$= + .9$$

This indicates a high degree of positive association between the intelligence of father and son. That is, the intelligence or dullness of father effects the intelligence or dullness of son to a very considerable extent.

Problem 304.—In an anti-malaria campaign in a village quinine was administered to 600 adults out of a total population of 3000. The number of fever cases is shown below :—

	Fever	No-fever	Total
Quinine	19	581	600
No-quinine	186	2214	2400
Total	205	2795	3000

(*Raj., M.A., 1950*)

Solution :

Let administration of Quinine be represented by A
No-quinine by a
Fever by B
No-fever by b.

	A	a	
B	(AB) 19	(aB) 186	[B] 205
b	(Ab) 581	(ab) 2214	[b] 2795
	[A] 600	[a] 2400	[N] 3000

$$\text{Co-efficient of association (Q)} = \frac{(AB)(ab)-(Ab)(aB)}{(AB)(ab)+(Ab)(aB)}$$

$$= \frac{[19 \times 2214]-[581 \times 186]}{[19 \times 2214]+[581 \times 186]}$$

$$= \frac{42066-108066}{42066+108066}$$

$$= \frac{-66000}{150132}$$

$$= -\cdot 43 \text{ approx.}$$

Thus, we conclude that quinine and malaria are negatively associated. That is, quinine prevents malaria attack.

Problem 305.—From the following figures, calculate the co-efficients of association between unemployment and literacy in rural and urban areas separately :

	Urban	*Rural*
Total adult males	50 lakhs	400 lakhs
Literate Males	20 ,,	80 ,,
Unemployed Males	5 ,,	12 ,,
Literate and unemployed males	3 ,,	4 ,,

(*Raj., M.A., Final, Old Course, 1955*)

Solution :

[A] **Urban—**

Let Literate males be represented by A
Illiterate by a
Employed by B
Unemployed by b.

	A	a	
B	(AB) 17	(aB) 28	[B] 45
b	(Ab) 3	(ab) 2	[b] 5
	[A] 20	[a] 30	[N] 50

All figures in lakhs.

$$\text{Co-efficient of association (Q)} = \frac{(AB)(ab)-(Ab)(aB)}{(AB)(ab)+(Ab)(aB)}$$
$$= \frac{[17\times 2]-[3\times 28]}{[17\times 2]+[3\times 28]}$$
$$= \frac{34-84}{34+84}$$
$$= \frac{-50}{118}$$
$$= -\cdot 42 \text{ approx.}$$

Thus, we find that literacy and un-employment are negatively associated. That is, educated people are more employed.

[B] **Rural—**

	A	a	
B	(AB) 76	(aB) 312	[B] 388
b	(Ab) 4	(ab) 8	[b] 12
	[A] 80	[a] 320	[N] 400

Figures in lakhs.

$$\text{Co-efficient of association (Q)} = \frac{(AB)(ab)-(Ab)(aB)}{(AB)(ab)+(Ab)(aB)}$$

$$= \frac{[76\times 8]-[4\times 312]}{[76\times 8]+[4\times 312]}$$

$$= \frac{608-1248}{608+1248}$$

$$= \frac{-640}{1856}$$

$$= -\cdot 34 \text{ approx.}$$

In the rural area the literacy and un-employment are negatively associated to the extent of $-\cdot 34$ approximately.

Problem 306.—Find the co-efficient of association between the type of college training and success in teaching from the following table :—

Instruction	*Successful*	*Unsuccessful*	*Total*
Teacher's College	58	42	100
University	49	51	100
Total	107	93	200

(Raj., M.A., 1956)

Solution :

Let teacher's College be represented by A
University by *a*
Successful by B
Unsuccessful by *b*

	A	*a*	
B	(B) 58	(*a*B) 49	[B] 107
b	(A*b*) 42	(*b*) 51	[*b*] 93
	[A] 100	[A] 100	[N] 200

$$\text{Co-efficient of association (Q)} = \frac{(AB)\ (ab)-(Ab)\ (aB)}{(AB)\ (ab)+(Ab)\ (aB)}$$

$$= \frac{[58\times 51]-[42\times 49]}{[58\times 51]+[42\times 49]}$$

$$=\frac{2958-2058}{2958+2058}$$

$$=\frac{900}{5016}$$

$$=+\cdot 17 \text{ approximately.}$$

Problem 307.—Find out the coefficient of association between the use of contraceptives and infant mortality from the following figures :

	Children born	*Infantile deaths*	*Survivors*
Users of Contraceptives	1000	250	750
Non-users of Contraceptives	1000	140	860

Solution :

Let the users of contraceptives be represented by A
Non-user of contraceptives by *a*
Infantile deaths by B
Non-infantile deaths (survivors) by *b*

	A	*a*	
B	(AB) 250	[B] 140	[B] 390
b	(A*b*) 750	(*ab*) 860	[*b*] 1610
	[A] 1000	[A] 1000	[N] 2000

$$\text{Co-efficient of association (Q)}=\frac{(AB)\ (ab)-(Ab)\ (aB)}{(AB)\ (ab)+(Ab)\ (aB)}$$

$$=\frac{[250\times 860]-[750\times 140]}{[250\times 860]+[750\times 140]}$$

$$=\frac{215000-105000}{21500+105000}$$

$$=\frac{110000}{320000}$$

$$=\frac{11}{32}$$

$$=+\cdot 34 \text{ approx.}$$

Thus, we find that the use of contraceptive and infant mortality are positively associated.

Problem 308.—Find if there is any association between inoculation and attack of typhoid from the following :—

	Attacked	*Not attacked*
Inoculated	12	674
Not-inoculated	47	1122
Total	59	1796

Solution :

Let inoculated be represented by A
Not-inoculated by a
Attacked by B
Not-attacked by b

	A	a	
B	(AB) 12	(aB) 47	[B] 59
b	(Ab) 674	(ab) 1122	[b] 1796
	[A] 686	[a] 1169	[N] 1855

$$\text{Co-efficient of association } (Q) = \frac{(AB)\ (ab) - (Ab)\ (aB)}{(AB)\ (ab) + (Ab)\ (aB)}$$

$$= \frac{[12 \times 1122] - [674 \times 47]}{[12 \times 1122] \times [674 \times 47]}$$

$$= \frac{13464 - 31678}{13464 + 31678}$$

$$= \frac{-18214}{45142}$$

$$= -\cdot 4 \text{ approx.}$$

Thus, we conclude that inoculation and the attack of typhoid are negatively associated. That is, inoculation check the attack of typhoid to some extent.

Problem 309.—Do you find any association between the tempers of brothers & sisters from the following data :

Good natured brothers and good natured sisters	=	1230
Good natured brothers and sullen sisters	=	850
Sullen brothers and good natured sisters	=	530
Sullen brothers and sullen sisters	=	980

(*Agra, M.Sc., 1957*)

Solution :

Let Good natured brothers be represented by A
Sullen brothers by α
Good natured sisters be represented by B
Sullen sisters by b

Thus,

$$AB=1230$$
$$Ab=850$$
$$aB=530$$
$$ab=980$$

$$\text{Co-efficient of association } (Q)=\frac{(AB)\ (ab)-(Ab)\ (aB)}{(AB)\ (ab)+(Ab)\ (aB)}$$
$$=\frac{[1230\times 980]-[850\times 530]}{[1230\times 980]+[850\times 530]}$$
$$=\frac{1205400-450500}{1205400+450500}$$
$$=\frac{754900}{1655900}$$
$$=\cdot 45$$

Thus, we find that there is positive association between the natures of brothers and sisters. That is, in general, good natured brothers have good natured sisters and *vice versa*.

✓**Problem 310.**—From the following, find whether blindness and baldness are associated :—

Total population	=	16264000
number of bald headed	=	24441
number of blind	=	7623
number of bald-headed blind	=	221

(*M.Sc., Agra, 1956*)

Solution :

N represents the total population
A=Attributes for bald-headed
B=Attributes for Blinds

Hence $N=16264000$

$(A)=24441$

$(B)=7623$

$(AB)=221$

$$\text{Now } (AB)_0=\frac{(A)\times(B)}{N}$$
$$=\frac{24441\times 7623}{16264000}$$
$$=\frac{186313743}{16264000}$$
$$=11\cdot 4$$

Since $(AB)>(AB)_0$ hence there is positive association.

Problem 311.—If in a collection of houses actually invaded by small pox, 70 per cent of the inhabitants are attacked and 85 per cent have been vaccinated, what is the lowest percentage of the vaccinated that must have been attacked.

(*I.A.S.*, 1955)

Solution :

$$\text{Let} \quad N = 100$$

Denoting the inhabitants invaded by small pox by A and vaccinated by B we have

$$(A) = 70$$
$$(B) = 85$$

To find the minimum value to (AB)

$$\begin{aligned} \text{We have } (AB) &\geqslant (A)+(B)-N \\ &\geqslant 70+85-100 \\ &\geqslant 55 \end{aligned}$$

∴ lowest value of (BC) is 55

But the inhabitants vaccinated are 85,

i.e. $(B) = 85$

hence lowest percentage of inhabitants vaccinated which have been attacked

$$= \frac{55}{85} \times 100$$
$$= 64{\cdot}7 \text{ per cent}$$

Problem 312.—Investigate the association between darkness of eye colour in father and son from the following data :—

Combination	*Frequency*
Fathers with dark eyes and sons with dark eyes	50
Fathers with dark eyes and sons with not dark eyes	79
Fathers with not dark eyes and sons with dark eyes	89
Fathers with not dark eyes and sons with not dark eyes	782

What would have been the frequency of fathers with dark eyes and sons with not dark eyes ; for the same total number, had there been complete independence ?

(*I.A.S.*, 1954)

Solution :

(A) Let fathers with dark eyes be represented by **A**

,,	,, not dark ,,	,, ,, ,,	α
Sons	,, dark eyes	,, ,, ,,	B
,,	,, not dark eyes	,, ,, ,,	β

	A	α	
B	(AB) 50	(αB) 89	[B] 139
β	(Aβ) 79	($\alpha\beta$) 782	[β] 861
	[A] 129	[α] 871	[N] 1000

Yule's Coefficient of Association (Q)

$$= \frac{(AB)\,(\alpha\beta) - (A\beta)\,(\alpha B)}{(AB)\,(\alpha\beta) + (A\beta)\,(\alpha B)}$$

$$= \frac{[50 \times 782] - [79 \times 89]}{[50 \times 782] \times [79 \times 89]}$$

$$= \frac{39100 - 7031}{39100 + 7031}$$

$$= \frac{32069}{38931}$$

$= + \cdot 82$ approximately.

Thus, we find that there is positive association of a high degree between the eye colour of father and son.

(B) Taking the attributes independent, the frequency of fathers with dark eyes and sons with not dark eyes for the same total number $= (A\beta)_0 = \frac{(A)\,(\beta)}{N}$

$$= \frac{129 \times 861}{1000}$$

$$= \frac{111069}{1000}$$

$$= 111{\cdot}069$$

$= 111$ approximately.

Problem 313.—Show that if $(AB)_1, (\alpha B)_1, (A\beta)_1, (\alpha\beta)_1,$
$(AB)_2, (\alpha B)_2, (A\beta)_2, (\alpha\beta)_2$

be two aggregates corresponding to the same values of (A), (B), (α) and (β),

$$(AB)_1-(AB)_2=(\alpha B)_2-(\alpha B)_1=(A\beta)_2-(A\beta)_1=(\alpha\beta)_1-(\alpha\beta)_2$$

(M.Sc., Agra, 1944)

Solution :

$$(A)=(AB)_1+(A\beta)_1$$

Also $$(A)=(AB)_2+(A\beta)_2$$

$\therefore$ $$(AB)_1+(A\beta)_1=(AB)_2+(A\beta)_2$$

or $$(AB)_1-(AB)_2=(A\beta)_2-(A\beta)_1$$

Similarly the other parts can be proved.

Problem 314.—Adopting the usual notation prove that

$$\delta=\frac{(B)(\beta)}{N}\left\{\frac{(AB)}{(B)}-\frac{(A\beta)}{(\beta)}\right\}.$$

(M.Sc., Agra, 1948)

Solution :

$$\begin{aligned}
\text{R.H.S.} &= \frac{(B)(\beta)}{N}\left\{\frac{(AB)}{(B)}-\frac{(A\beta)}{(\beta)}\right\}\\
&= \frac{(B)(\beta)}{N}\left[\frac{(AB)(\beta)-(A\beta)(B)}{(B)(\beta)}\right]\\
&= \frac{1}{N}\Big[(AB)[N-(B)]-[(A)-(AB)](B)\Big]\\
&= \frac{1}{N}\Big[(AB)\,N-(A)(B)\Big]\\
&= (AB)-\frac{(A)(B)}{N}\\
&= (AB)-(AB)_0\\
&= \delta.
\end{aligned}$$

Note :—In the similar manner we can prove the result

$$\delta=\frac{(A)(\alpha)}{N}\left[\frac{(AB)}{(A)}-\frac{(\alpha B)}{(\alpha)}\right].$$

Problem 315.—Show that if $\delta=(AB)-(AB)_0$;

$$(AB)^2+(\alpha\beta)^2-(\alpha B)^2-(A\beta)^2=[(A)-(\alpha)][(B)-(\beta)]+2N\delta.$$

[M.Sc., Agra, 1946, 1948, 1954].

Solution :

$$\text{R.H.S.} = [(A)-(\alpha)][(B)-(\beta)]+2N\delta.$$

$$=[(AB)+(A\beta)-(\alpha B)-(\alpha\beta)][(AB)+(\alpha B)-(A\beta)-(\alpha\beta)] + 2N\left[(AB)-\frac{(A)(B)}{N}\right].$$

$$=[(AB)^2-(A\beta)^2-(\alpha B)^2+(\alpha\beta)^2+(AB)(\alpha B)-(A\beta)(AB) -(AB)(\alpha\beta)+(A\beta)(AB)+(A\beta)(\alpha B)-(A\beta)(\alpha\beta) -(\alpha B)(AB)+(\alpha B)(A\beta)-(\alpha B)(\alpha\beta)-(\alpha\beta)(AB) -(\alpha\beta)(\alpha B)+(\alpha\beta)(A\beta)]+2[N(AB)-(A)(B)]$$

$$=(AB)^2-(A\beta)^2-(\alpha B)^2+(\alpha\beta)^2+2(A\beta)(\alpha B)-2(AB)(\alpha\beta) +2[(AB)+(A\beta)+(\alpha B)+(\alpha\beta)](AB) -2[(AB)+(A\beta)][(AB)+(\alpha B)]$$

$$=(AB)^2-(A\beta)^2-(\alpha B)^2+(\alpha\beta)^2.$$

$$=\text{L.H.S.}$$

Hence proved.

CHAPTER X

INDEX NUMBERS

"An Index Number is a number which indicates the level of a certain phenomena at any given date in comparison with the level of the same phenomena at some standard date".

Index numbers are mostly expressed in the form of percentages, compared with the base year index, which is always equal to 100.

According to Prof. Secrist, "Index Numbers are a series of numbers by which changes in the magnitudes of a phenomenon are measured from time to time, or from place to place........"

Construction of Index Numbers

Index numbers are constructed in many ways. Some of the important types and formula are given below :

(*i*) Simplest Method.

Commodity—A.

Base year Price=10/-

Base year Index=100

Current year Price=15/-

∴ Current year Index Number

$$=\frac{\text{Current year Price}}{\text{Base year Price}}\times 100$$

$$=\frac{15}{10}\times 100$$

$$=150$$

(*ii*) Chain Base and Link Index Numbers.

(Refer to Problems 318, 327........)

(*iii*) Weighted Index Numbers.

$$\text{Index Numbers for current year}=\frac{\Sigma IV}{\Sigma V}$$

Where I=Price Relative

V=Value (weight).

(*iv*) Weighted Aggregative Method.

$$\text{Index Number for Current year}=\frac{\Sigma p_1q_1}{\Sigma p_0q_0}\times 100$$

Where, p_1=current year price

q_1=current year quantity

p_0=base year price

q_0=base year quantity.

(*v*) Fisher's Ideal Formula

$$\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

Values as in (*iv*) above. To calculate current year index it is multiplied by 100.

(*vi*) Time Reversal Test.

$$P_{01} \times P_{10} = 1$$

Where P_{01} = Price change for the current year on the base year.

P_{10} = Reverse of P_{01}

$$\text{Again, } P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_1}}$$

(*vii*) Factor Reversal Test.

$$P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$\text{Again, } P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}}$$

$$= \sqrt{\frac{\Sigma p_1 q_1 \times \Sigma p_1 q_1}{\Sigma p_0 q_0 \times \Sigma p_0 q_0}}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

(*viii*) $P_{01} = \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$ — Laspeyres formula.

(*ix*) $P_{01} = \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$ — Paasche formula.

(*x*) $P_{01} = \frac{1}{2}\left[\dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} + \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}\right]$ —Arithmetic cross of Laspeyres and Paasche formulae.

(*xi*) $P_{01} = \sqrt{\dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$ — Geometric cross of (*viii*) & (*ix*).

(*xii*) $P_{01} = \dfrac{\Sigma (q_0 + q_1) p_1}{\Sigma (q_0 + q_1) p_0}$ — Arithmetically cross-weight aggregation formula.

(*xiii*) $P_{01} = \dfrac{\Sigma \sqrt{q_0 \cdot q_1} p_1}{\Sigma \sqrt{q_0 \cdot q_1} p_0}$ — Geometrically cross-weight aggregation formula.

(*xiv*) $P_{01} = \dfrac{\Sigma p_1 q a}{\Sigma p_0 q a}$ [qa = quantity for the year to which weights relate.]

There are approximately 135 formulae to calculate index numbers. Different formulae are used in the construction of different types of indices.

Problem 316.—Use the following data of industrial production in India to compare the annual fluctuations in Indian industrial activity by the chain base method :—

Index numbers of Industrial Production in India.

Year	Index No.	Year	Index No.
1919—20	120	1926—27	149
—21	122	—28	156
—22	116	—29	137
—23	120	—30	162
—24	120	—31	149
—25	137	—32	160
—26	136	—33	160

(*Luck., M.Com., 1943*)

Solution :

For the construction of Index Number by Chain Link Method, we assume the Relative for 1919-20 as 100 and then construct the other index.

Year Base: 1919-20=100	Index No. $\left[\frac{p_1}{p_0}\times 100\right]$	Year	Index No. $\left[\frac{p_1}{p_0}\times 100\right]$
1919—20	100	1926—27	109·6
1920—21	101 66	1927—28	104·7
1921—22	95·08	1928—29	87·8
1922—23	103·4	1929—30	118·2
1923—24	100	1930—31	91·9
1924—25	114·16	1931—32	107·4
1925—26	99·27	1932—33	100

Problem 317.—The following are the group index numbers and the group weights of an average working class family's budget. Construct the cost of living index number by assigning the given weights.

Groups	Index Nos.	Weight
Food	352	48
Fuel and Lighting	220	10
Clothing	230	8
Rent	160	12
Miscellaneous	190	15

(*I.A.S., 1950*)

Solution :

Construction of cost of living index by Family Budget Method :—

Groups	Index No. (I)	Weights (V)	(IV)
Food	352	48	16896
Fuel and Light	220	10	2200
Clothing	230	8	1840
Rent	160	12	1920
Misc.	190	15	2850
		ΣV=93	ΣIV=25706

$$\text{Index Number for the current year} = \frac{\Sigma IV}{\Sigma V}$$

$$= \frac{25706}{93}$$

$$= 276{\cdot}408 \text{ approx.}$$

Problem 318.—An enquiry into the budgets of the middle class families in a city in England gave the following information :—

Expenses on	Food 35%	Rent 15%	Clothing 20%	Fuel 10%	Misc 20%
Prices (1928)	£ 150	£ 30	£ 75	£ 25	£ 40
Prices (1929)	£ 145	£ 30	£ 65	£ 23	£ 45

What changes in cost of living figures of 1929 as compared with that of 1928 are seen. (*Luck., B. Com., 1944*)

Solution :

Index Numbers are constructed here based on fixed base system and regarding 1928 as the base year.

Articles	Prices (1928)	Index (1928)	Prices (1929)	Index (1929) $\left[\frac{p_1}{p_0} \times 100\right]$ (I)	Expenses % (V)	(IV)
Food	£ 150	100	£ 145	97 approx.	35	3395
Rent	£ 30	100	£ 30	100	15	1500
Clothing	£ 75	100	£ 65	87 ,,	20	1740
Fuel	£ 25	100	£ 23	92	10	920
Misc	£ 40	100	£ 45	113 ,,	20	2260
					ΣV=100	ΣIV=9815

$$\text{Index Number for the current year (1929)} = \frac{\Sigma IV}{\Sigma V}$$

$$= \frac{9815}{100}$$

$$= 98{\cdot}15 \text{ approx.}$$

Therefore, we find that as compared to the year 1928, the prices have been gone down in 1929

Problem 319.—Given the following data, what index numbers would you use for purposes of comparison ? Give reasons.

Year	Rice		Wheat		Jowar	
	Price	Quantity	Price	Quantity	Price	Quantity
1927	9·3	100	6·4	11	5·1	5
1934	4·5	90	3·7	10	2·7	3

Prices and quantities are given in arbitrary units.

(M.A., Cal., 1937)

Solution :

Commodity	Base year		Current year		p_0q_0	p_1q_0	p_0q_1	p_1q_1
	Price	Qty.	Price	Qty.				
Rice	9·3	100	4·5	90	930·0	450	837·0	405·0
Wheat	6·4	11	3·7	10	70·4	40·7	64·0	37·0
Jowar	5·1	5	2·7	3	25·5	13·5	15·3	8·1
					$\Sigma p_0q_0 = 1025{\cdot}9$	$\Sigma p_1q_0 = 504{\cdot}2$	$\Sigma p_0q_1 = 916{\cdot}3$	$\Sigma p_1q_1 = 450{\cdot}1$

For making the computations sufficiently easier, we take these in the form of approximates as : $\Sigma p_0q_0 = 1026$; $\Sigma p_1q_0 = 504$; $\Sigma p_0q_1 = 916$ and $\Sigma p_1 q_1 = 450$.

$$\text{Index No. for the current year} = 100 \times \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$$

$$= 100 \times \sqrt{\frac{504}{102} \times \frac{450}{916}}$$

$$= 100 \sqrt{\frac{504}{1026} \times \frac{450}{916}}$$

$$= 100 \times {\cdot}492 \text{ app.}$$

$$= 49{\cdot}2 \text{ approximately.}$$

Problem 320.—Construct Index Numbers for the year 1904 on the basis of the year 1902 of the following :

Year	Article I		Article II		Article III	
	Price	Quantity	Price	Quantity	Price	Quantity
1902	5	10	8	6	6	3
1904	4	12	7	7	5	4

(Agra, M. Com., 1947)

Solution :

Year 1902 is considered as base (100).

Let p_0 represent the price for the base year (1902)

q_0 ,, ,, quantity ,, ,, ,, ,,

p_1 ,, ,, price ,, ,, current year (1904)

q_1 ,, ,, quantity ,, ,, ,, ,,

Article	1902		1904		$p_0\,q_0$	$p_0\,q_1$	$p_1\,q_0$	$p_1\,q_1$
	Price (p_0)	Quantity (q_0)	Price (p_1)	Quantity (q_1)				
I	5	10	4	12	50	60	40	48
II	8	6	7	7	48	56	42	49
III	6	3	5	4	18	24	15	20
				Total	$\Sigma p_0q_0 = 116$	$\Sigma p_0\,q_1 = 140$	$\Sigma p_1\,q_0 = 97$	$\Sigma p_1q_1 = 117$

Index Number for 1904 by Fisher's Ideal Formula :

$$= 100 \times \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$$

$$= 100 \times \sqrt{\frac{97}{116} \times \frac{117}{140}}$$

$$= 100 \times \sqrt{\cdot 83 \times \cdot 83} \text{ approx.}$$

$$= 100 \times \cdot 83 \text{ approx.}$$

$= 83{\cdot}0$ approximately. **Ans.**

Problem 321. Construct with the help of data given below Fisher's Ideal Index, and show how it satisfies the factor reversal test.

Commodity	Estimated total produce in 000 tons in District Saran		Harvest Price per maund in District Saran	
	1931-32	1932-33	1931-32	1932-33
Winter Rice	71	26	3—8—0	3—2—0
Barley	107	83	2—0—0	1—14—0
Maize	62	48	2—9—0	1—12—0

(*Patna, M.A., 1942*)

Solution :

Commodity	Base year (1931-32)		Current Year (1932-33)		(Base year Price) × (Base year quantity)	(Base year Price) × (current year quantity)	(Current year Price) × (base year quantity)	(Current year price) × (Current year quantity)
	Price in nP. per md.	Quantity	Price in nP. per md.	Quantity				
	p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
Winter Rice	350	71	312	26	24850	9100	22152	8112
Barley	200	107	187	83	21400	16600	20009	15521
Maize	256	62	175	48	15872	12288	10850	8400
					$\Sigma p_0q_0 = 62122$	$\Sigma p_0q_1 = 37988$	$\Sigma p_1q_0 = 53011$	$\Sigma p_1q_1 = 32033$

Fisher's Ideal Index Number for the

$$\text{current year} = 100 \times \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}}$$

$$= 100 \times \sqrt{\frac{53011}{62122} \times \frac{32033}{37988}}$$

$$= 100 \times \sqrt{\frac{1698101363}{2359890536}}$$

$$= 100 \times \cdot 84 \text{ approximately.}$$

$$= 84.$$

If $P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$, then it will satisfy the condition of Factor Reversal Test, where

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \text{ and}$$

$$Q_{01} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}}$$

$$\text{Now, } P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}}$$

$$= \sqrt{\frac{53011}{62122} \times \frac{32033}{37988} \times \frac{37988}{62122} \times \frac{32033}{53011}}$$

$$= \sqrt{\frac{32033 \times 32033}{62122 \times 62122}}$$

$$= \frac{32033}{62122}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence proved that it satisfies the Factor Reversal Test.

Problem 322.—Construct approximate index number to discuss the fluctuation in the export of raw cotton from India for the period 1930-31 to 1935-36, using the average of the period 1926—30 as base.

Year	*Quantity of Raw Cotton in thousand tons*	*Value of Raw Cotton in Rs. lakhs*
1926—30 (base)	609	5941
1930-31	701	4633
1931-32	423	2345
1932-33	365	2037
1933-34	504	2753
1934-35	623	3495
1935-36	607	3377

(*I.C.S., 1939*)

Solution :

For 1926-30 (as base), index number is 100. On the basis of this, we have to construct index numbers :

Year	Raw Cotton			
	Quantity (1000 tons)		Value (Rs. lakhs)	
	Quantity	Index Numbers	Value	Index Numbers
		1926—30=100		1926—30=100
1926—30 (Average)	609	100	5941	100
1930—31	701	$\left[\frac{701}{609} \times 100 = 115{\cdot}1\right]$	4633	$\frac{4633}{5941} \times 100 = 78{\cdot}1$
1931—32	423	$\frac{423}{609} \times 100 = 69{\cdot}4$	2345	$\frac{2345}{5941} \times 100$ =40 approx.
1932—33	365	60 approx.	2037	34 ,,
1933—34	504	83 ,,	2753	46 ,,
1934—35	623	102 ,,	3495	59 ,,
1935—36	607	98 ,,	3377	57 ,,

Problem 323.—Show with the help of the following data that the factor reversal test is satisfied by Fisher's Ideal Formula for Index Number Construction :

Commodity	*Base year Price (Rs.)*	*Base year Quantity (Mds.)*	*Current year Price (Rs.)*	*Current year Quantity (Mds.)*
A	6	50	10	56
B	2	100	2	120
C	4	60	6	60
D	10	30	12	24
E	8	40	12	36

(*Alld. M. Com.*, *1946* ;
Punjab, M.A. 1951 ; *(Sept.)*
Delhi, B. Com., *1953*)

Solution :

Commodity	Base year Price (Rs.) p_0	Base year Quantity (Mds.) q_0	Current year Price (Rs.) p_1	Current year Quantity (Mds.) q_1	p_0q_0	p_0q_1	p_1q_0	p_1q_1
A	6	50	10	56	300	336	500	560
B	2	100	2	120	200	240	200	240
C	4	60	6	60	240	240	360	360
D	10	30	12	24	300	240	360	288
	8	40	12	36	320	288	480	432
					$\Sigma p_0q_0=1360$	$\Sigma p_0q_1=1344$	$\Sigma p_1q_0=1900$	$\Sigma p_1q_1=1880$

Fisher's Ideal Formula is $\sqrt{\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \dfrac{\Sigma p_1q_1}{\Sigma p_0q_1}}$

The condition of Factor Reversal Test is said to be satisfied if

$$P_{01} \times Q_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0}, \quad \text{where}$$

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \quad \text{and}$$

$$Q_{01} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_1q_0}}$$

Therefore, $P_{01} \times Q_{01} = \sqrt{\dfrac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \dfrac{\Sigma p_1q_1}{\Sigma p_0q_1} \times \dfrac{\Sigma p_0q_1}{\Sigma p_0q_0} \times \dfrac{\Sigma p_1q_1}{\Sigma p_1q_0}}$

$$= \sqrt{\frac{1900}{1360} \times \frac{1880}{1344} \times \frac{1344}{1360} \times \frac{1880}{1900}}$$

$$= \sqrt{\frac{1880 \times 1880}{1360 \times 1360}}$$

$$= \frac{1880}{1360}$$

$$= \frac{\Sigma p_1q_1}{\Sigma p_0q_0} \qquad \text{Hence proved.}$$

To find out the Index Number, by Fisher's Formula, for the current year, we have to multiply it by 100., *i.e.*,

$$100 \times \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}}.$$

Problem 324.—The following table gives the average wholesale prices of the commodities A, B, C, D, E during the years 1940 to 1943. Find out the index number by reference to 1940 as base.

Commodities	*Average Wholesale Prices in the years*			
	1940	*1941*	*1942*	*1943*
A	30	32	36	40
B	20	22	25	30
C	40	44	46	50
D	45	50	52	55
E	55	60	61	65

Solution :

(Base year : 1940=100)

Commodities	Percentage or Relative with 1940 as base			
	1940	1941	1942	1943
A	100	$\frac{32}{30} \times 100$ =106 app.	$\frac{36}{30} \times 100$ =120	$\frac{40}{30} \times 100$ 133 app.
B	100	110 ,,	125 app.	150 ,,
C	100	110 ,,	115 ,,	125 ,,
D	100	111 ,,	116 ,,	122 ,,
E	100	109 ,,	110 ,,	118 ,,
Total of Relatives	500	546	586	648
Index Number as average of Relatives	100	109	117	129

Problem 325.—From the fixed base index numbers given below, prepare chain base index numbers.

1940	1941	1942	1943	1944
267	275	280	290	320

Solution :

(1940 is taken as base=100)

To construct chain base index numbers from the index numbers given, based on fixed base system, we have to proceed as such :

Years	Fixed base Index Numbers	Fixed base index numbers changed to chain based index numbers	Chain base Index Numbers
1940	267		100
1941	275	$\frac{275}{267}\times 100$	106 app.
1942	280	$\frac{280}{275}\times 100$	101 ,,
1943	290	$\frac{290}{280}\times 100$	103 ,,
1944	320	$\frac{320}{290}\times 100$	110·4 ,,

Problem 328.—From the chain base index numbers given below, prepare fixed base index numbers.

1900	1901	1902	1903	1904
80	110	120	90	140

(*Figures Arbitrary*)

Solution :

1900 is base year=100

Years	Chain Base Index Numbers	Chain Base index numbers chained to 1900 as base	Fixed based index numbers
1900	80		80
1901	110	$\frac{80}{100}\times 110$	88
1902	120	$\frac{80}{100}\times\frac{110}{100}\times 120$	105 app.
1903	90	$\frac{80}{100}\times\frac{110}{100}\times\frac{120}{10}\times 90$	95 ,,
1904	140	$\frac{80}{100}\times\frac{110}{100}\times\frac{120}{100}\times\frac{90}{100}\times 140$	133 ,,

CHAPTER XI

THEORETICAL DISTRIBUTION

In this chapter we will deal with the following three theoretical distributions :—

1. Binomial Distribution.
2. Poisson's Distribution.
3. Normal Distribution.

I. Binomial Distribution

Let there be an event, the chance of being its success is p and the chance of its failure is q in one trial such that $p+q=1$. Let the event be tried n times and the results tabulated to form one set in which there can be only once a success, two successes......n successes. We assume that events in the series of trials are independent *i.e.*, the chances p and q are the same for each event and remain constants throughout the trials. Let the set of n trials be repeated N times where N is very large. In these N sets of n trials there will be a few cases in which there is no success and a few cases in which there is one success, two successes and so on. If we classify these cases in order of success, we will get a frequency distribution of the form

No. of successes	frequency
0	x_0
1	x_1
2	x_2
⋮	⋮
n	x_n

Now, on the assumption of independence of successive events the values of $x_0, x_1, x_2, \ldots\ldots x_n$ can be determined theoretically also and thus nature of the distribution may also be determined.

The probability that in a particular set there are exactly r successes is ${}^nC_r p^r q^{n-r}$. Hence in all the N sets, the frequency (number of cases) of r successes are $N.{}^nC_r p^r q^{n-r}$. So that our distribution becomes

No. of successes	Frequency
0	Nq^n
1	$N{}^nC_1 q^{n-1}p$
2	$N{}^nC_2 q^{n-2}p^2$
⋮	⋮
n	$N.{}^nC_n p^n$.

Hence for N sets of n trials, the frequencies of 0, 1, 2,.........n successes are given by the successive terms of the binomial expansion of N $(q+p)^n$ *i.e.*, $N(q^n+{}^nC_1q^{n-1}p+{}^nC_2q^{n-2}p^2+\ldots\ldots+{}^nC_np^n)$. This is called Binomial Theorem.

(2) Poisson's Distribution

The Poisson distribution is in a sense a particular limiting form of the binomial distribution in which the proportion p of successes is very small. We may suppose our number n large enough to render np itself appreciable though p is small ; and we are thus led to consider the limiting form of the binomial as $p \to o$ subject to the condition that np remains finite, and equal to m, say.

Under these conditions the term

$$^nC_rp^rq^{n-r}=\frac{\lfloor n}{\lfloor n-r.\lfloor r}\cdot\left(\frac{m}{n}\right)^r\left[1-\frac{m}{n}\right]^{n-r}$$

$$=\frac{m^r}{\lfloor r}\left(1-\frac{m}{n}\right)^n\cdot\frac{\lfloor n}{\lfloor n-r.\,n^r\cdot\left(1-\frac{m}{n}\right)^r}$$

$$\text{Now }\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^n=e^{-m}$$

and by Stirling's formula for $\lfloor n$ *viz.* $\lim_{n\to\infty}\lfloor n=\sqrt{2\pi n}.\,n^n.\,e^{-n}$ we have

$$\lim_{n\to\infty}\frac{\lfloor n}{\lfloor n-r.\,n^r.\left(1-\frac{m}{n}\right)^r}$$

$$=\lim_{n\to\infty}\frac{\sqrt{2\pi n}\,.\,n^ne^{-n}}{\sqrt{2\pi(n-r)}\,(n-r)^{n-r}.\,e^{-(n-r)}.\,n^r.\left(1-\frac{m}{n}\right)^r}$$

$$=\lim_{n\to\infty}\frac{1}{e^r\left(1-\frac{r}{n}\right)^{n-r+\frac{1}{2}}.\left(1-\frac{m}{n}\right)^r}$$

$$=\lim_{n\to\infty}\frac{1}{e^r.\left(1-\frac{r}{n}\right)^n}\cdot\frac{\left(1-\frac{r}{n}\right)^{r-\frac{1}{2}}}{\left(1-\frac{m}{n}\right)^r}$$

$$=\frac{1}{e^r.\,e^{-r}}\cdot1=1$$

$$\text{Hence }\lim_{n\to\infty}{}^nC_rp^rq^{n-r}=\frac{m^r}{\lfloor r}e^{-m}$$

Hence the successive terms of the binomial become

$$e^{-m},\ e^{-m}\cdot\frac{m}{|\underline{1}},\ e^{-m}\frac{m^2}{|\underline{2}},\ \ldots\ldots,\ e^{-m}\frac{m^r}{|\underline{r}},\ \ldots\ldots\ldots$$

and the limit of $(q+p)^n$ is

$$e^{-m}\left(1+\frac{m}{|\underline{1}}+\frac{m^r}{|\underline{2}}+\ldots\ldots+\frac{m^r}{|\underline{r}}+\ldots\ldots\right).$$

This is called the Poisson distribution.

Note :—If we take, in place of p, q very small tending to zero such that nq is finite we will get the same limiting distribution.

3. Normal Distribution

The normal distribution is a continuous distribution which may be derived from the binomial distribution in the following manner. In the binomial distribution $N(q+p)^n$ the frequency of r successes is given by

$$f_r = N\ {}^nC_r p^r q^{n-r}$$

and frequency of $r+1$ successes by

$$f_{r+1} = N\ {}^nC_{r+1}\ p^{r+1} q^{n-r-1}$$

Now $\qquad f_{r+1} \gtrless f_r$

if $\qquad \dfrac{N\ {}^nC_{r+1}\ p^{r+1}\ q^{n-r-1}}{N\ {}^nC_r p^r q^{n-r}} \gtrless 1$

or $\qquad \dfrac{n-r}{r+1}\cdot\dfrac{p}{q} \gtrless 1$

or $\qquad r(p+q) \lessgtr np-q$

or $\qquad r \lessgtr np-q$

Let np be a whole number then the maximum frequency is

$$f_{np} = N\ {}^nC_{np}\ p^{np}\cdot q^{n-np}$$

$$= N\cdot\frac{|\underline{n}}{|\underline{np}\cdot|\underline{nq}}\cdot p^{np}\ q^{nq}$$

$\left\{\begin{array}{l}\text{since if } r=np-1 \\ \quad f_{r+1} > f_r \\ \quad \text{if } r=np \\ \quad f_{r+1} < f_r \\ \therefore\ f_{np} \text{ is highest}\end{array}\right.$

The frequency of $np+x$ successes is

$$f_{np+x} = N\ {}^nC_{np+x}\cdot p^{np+x}\cdot q^{n-pn-x}$$

$$= N\cdot\frac{|\underline{n}}{|\underline{np+x}\cdot|\underline{nq-x}}\cdot p^{np+x}\ q^{nq-x}$$

Hence $\qquad \dfrac{f_{np+x}}{f_{np}} = \dfrac{|\underline{np}\cdot|\underline{nq}}{|\underline{np+x}\ |\underline{nq-x}}\cdot p^x\ q^{-x}$

Since n is large enough tending to infinity hence applying Stirling's approximation we have

$$\frac{f_{np+x}}{f_{np}}=\frac{\sqrt{2\pi\, np}\,(np)^{np}\,e^{-np}\cdot\sqrt{2\pi nq}\,(nq)^{nq}\,e^{-nq}\cdot p^{x}.q^{-x}}{\sqrt{2\pi(np+x)}\cdot(np+x)^{np+x}\cdot e^{-(np+x)}\cdot\sqrt{2\pi(nq-x)}\cdot(nq-x)^{nq-x}\cdot e^{-(nq-x)}}$$

$$=\frac{1}{\left(1+\frac{x}{np}\right)^{np+x+\frac{1}{2}}\left(1-\frac{x}{nq}\right)^{nq-x+\frac{1}{2}}}$$

$$\therefore\quad \log\frac{f_{np+x}}{f_{np}}=-(np+x+\tfrac{1}{2})\log\left(1+\frac{x}{np}\right)-(nq-x+\tfrac{1}{2})\log\left(1-\frac{x}{nq}\right)$$

$$\text{or}\quad \log\frac{f_{np+x}}{f_{np}}=-(np+x+\tfrac{1}{2})\left[\frac{x}{np}-\frac{x^2}{2n^2p^2}+\frac{x^3}{3n^3p^3}\cdots\cdots\right]$$

$$-(nq-x+\tfrac{1}{2})\left[-\frac{x}{nq}-\frac{x^2}{2n^2q^2}-\frac{x^3}{3n^3q^3}\cdots\cdots\right]$$

$$=-\frac{x^2}{2npq}-\frac{q-p}{2npq}\,x+\text{ terms of order }\frac{1}{n^2}\text{ and higher.}$$

Since n is large and $q-p$ is very small hence taking further approximation we have

$$\log\frac{f_{np+x}}{f_{np}}=-\frac{x^2}{2npq}$$

$$\text{or}\ \log\frac{f_{np+x}}{f_{np}}=-\frac{x^2}{2\sigma^2}$$

[where σ is the standard deviation of binomial.]

$$\text{or}\ f_{np+x}=f_{np}\,e^{-\frac{x^2}{2\sigma^2}}$$

This can be put in the form

$$y_x=y_0\,e^{-\frac{x^2}{2\sigma^2}}$$

Where y_x is the frequency of $np+x$ successes [or else of x successes if the origin is transformed to np successes.].

The curve $y=y_0\,e^{-\frac{x^2}{2\sigma^2}}$ is called the Normal curve.

Generally the total frequency is taken to be $+1$ and therefore the value of y_0 can be determined by putting $\int y=1$

$$\text{or} \int_{-\infty}^{\infty} y_0 \, e^{-\frac{x^2}{2\sigma^2}} \, dx = 1$$

$$\text{or} \quad y_0 \cdot \sigma\sqrt{2\pi} = 1$$

$$\therefore \quad y_0 = \frac{1}{\sqrt{2\pi}}$$

Therefore the standard form of the normal curve is

$$\therefore \quad y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

If the total frequency is N, the corresponding standard form of normal curve is

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

If the origin is changed to the point $(a, 0)$, this equation becomes,

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Examples.

Problem 327.—With the usual notation, in the binomial distribution $N(q+p)^n$, compute the value of the mean, the standard deviation, coefficients β_1 and β_2.

(M.Sc. Agra, 1950, 1951, 1952, 1954)

Solution :

In the Binomial Distribution $N(q+p)^n$ the frequencies of the 0, 1, 2, ...etc. successes are given by the successive terms of the expansion.

$$N\,[q^n + {}^nC_1\, q^{n-1} p + \ldots\ldots\ldots + {}^nC_r q^{n-r} p^r + \ldots\ldots\ldots + p^n].$$

Now, $\quad \beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

$\quad\quad\quad \beta_2 = \dfrac{\mu_4}{\mu_2^2}$

Mean $= \mu_1 = \mu'_1$ [about the origin at 0 successes)

Standard Deviation $= \sqrt{\mu_2}$

To determine μ_2, μ_3 and μ_4 we have to determine first μ_2', μ_3', and μ_4' *i.e.* moments about the arbitrarily chosen origin.

Now take an arbitrary origin at 0 successes. The successive values of the deviations ξ are 0, 1, 2, ...r, ...n and hence

$$\mu_1' = \frac{\Sigma f\xi}{N}$$

$$= \sum_{r=0}^{n} {}^nC_r q^{n-r} . p^r . r \qquad \left[\text{here } \xi = r,\ f = {}^nC_r q^{n-r} . p^r\right]$$

$$= \Sigma \frac{np(n-1)(n-2) \ldots (n-r+1)}{\lfloor r-1} q^{n-r} . p^{r-1}$$

$$= np \, \Sigma \, {}^{n-1}C_{r-1} q^{n-r} . p^{r-1}$$

$$= np \, (q+p)^{n-1}$$

$$= np$$

$\therefore$ Mean M $= np$ [Since arbitrary origin is zero]

Similarly $\mu'_2 = \frac{\Sigma f \xi^2}{N}$

or $\mu'_2 = \sum_{r=0}^{n} {}^nC_r \, q^{n-r} . p^r . r^2$

$$= \Sigma \, {}^nC_r q^{n-r} p^r [r(r-1)+r]$$

$$= \Sigma \frac{n(n-1) p^2 .(n-2)(n-3) \ldots (n-r+1)}{\lfloor r-2} q^{n-r} . p^{r-2}$$

$$+ \Sigma \frac{np (n-1)(n-2) \ldots (n-r+1)}{\lfloor r-1} q^{n-r} . p^{r-1}$$

$$= n(n-1)p^2 \, \Sigma \, {}^{n-2}C_{r-2} \, q^{n-r} . p^{r-2} + np \, \Sigma \, {}^{n-1}C_{r-1} \, q^{n-r} . p^{r-1}$$

$$= n(n-1) \, p^2 . (q+p)^{n-2} + np \, (q+p)^{n-1}$$

$$= n(n-1)p^2 + np$$

$$\therefore \mu'_2 = n(n-1)p^2 + np$$

Again

$$\mu'_3 = \frac{\Sigma f \xi^3}{N}$$

$$= \sum_{r=0}^{n} {}^nC_r \, q^{n-r} , p^r , r^3$$

$$= \sum_{0}^{n} {}^{n}C_r \, q^{n-r} \cdot p^r \, [r(r-1)(r-2)+3r(r-1)+r]$$

$$= n(n-1)(n-2)p^3+3n(n-1)p^2+np$$

$$\mu'_4 = \frac{\Sigma f \xi^4}{N}$$

$$= \Sigma \, {}^{n}C_r \, q^{n-r} \cdot p^r \cdot r^4$$

$$= \Sigma \, {}^{n}C_r q^{n-r} \cdot p^r . [r(r-1)(r-2)(r-3)+6r(r-1)(r-2) + 7r(r-1)+r]$$

$$= n(n-1)(n-2)(n-3)p^4+6n((n-1)(n-2)p^3 + 7n(n-1)p^2+np$$

Therefore $\mu_1 = 0$

$$\mu_2 = \mu'_2-(\mu'_1)^2$$

$$= n(n-1)p^2+np-n^2p^2$$

$$= np(1-p)$$

$$= npq$$

$\therefore$ Standard Deviation $\sigma = \sqrt{\mu_2} = \sqrt{npq}$

$$\mu_3 = \mu'_3-3\mu'_2\mu'_1+3\mu'_1 \cdot \mu'^2_1-\mu'^3_1$$

$$= \mu'_3-3\mu_2'\mu_1'+2\mu'^3_1$$

$$= n(n-1)(n-2)p^3+3n(n-1)p^2+np-3[n(n-1)p^2+np].np + 2n^3p^3$$

$$= 2np^3-3np^2+np$$

$$= -2np^2(1-p)+np(1-p)$$

$$= npq-2nqp^2$$

$$= npq[1-2p]$$

$\therefore \mu_3 = npq[q-p]$

$$\mu_4 = \mu'_4-4\mu'_3\mu'_1+6\mu'_2 \cdot \mu'^2_1-4\mu'_1 \cdot \mu'^3_1+\mu'^4_1$$

$$= \mu'_4-4\mu'_3\mu'_1+6\mu_2'\mu'^2_1-3\mu'^4_1$$

$$= n(n-1)(n-2)(n-3)p^4+6n(n-1)(n-2)p^3+7n(n-1)p^2 + np$$

$$-4[n(n-1)(n-2)p^3+3n(n-1)p^2+np].np+6[n(n-1)p^2 + np] \cdot n^2p^2-3n^4p^4$$

$$= np^4[n^3-6n^2+11n-6-4n^3+12n^2-8n+6n^3-6n^2-3n^3] + 6np^3[n^2-3n+2-2n^2+2n+n^2]+np^2[7n-7-4n]+np$$

$$= np^4[3n-6]+6np^3[-n+2]+np^2[3n-7]+np$$

$$= 3n^2p^2[p^2-2p+1]+6np^2[-p^2+2p-1]+np(1-p)$$

$$= 3n^2p^2(1-p)^2-6np^2(1-p)^2+npq$$

$$= 3n^2p^2q^2-6np^2q^2+npq$$

$$= 3n^2p^2q^2+npq\,(1-6pq)$$

Using these values we get

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq}$$

Also $\gamma_1 = \sqrt{\beta_1} = \dfrac{q-p}{\sqrt{npq}}$; $\gamma_2 = \beta_2 - 3 = \dfrac{1-6pq}{npq}$.

Problem 328.—If a coin is tossed N times when N is very large even number, show that the probability of exactly $\frac{N}{2} - p$ heads and $\frac{N}{2} + p$ tails is approximately

$$\left(\frac{2}{\pi N}\right)^{\frac{1}{2}} e^{-\frac{2p^2}{N}}$$

(*M.Sc., Agra, 1945, 1947, 1951, 1954, 1958*)

Solution :

In tossing a coin once the probability of getting head is $\frac{1}{2}$ and probability of getting tail is also $\frac{1}{2}$. Hence the probability of 0, 1, 2, ... heads in tossing N times will be given by successive terms of binomial

$$(\tfrac{1}{2}+\tfrac{1}{2})\ N$$

Let $N=2m$; and let f_m and f_{m-p} denote the probability of m and $m-p$ heads respectively. Then

$$f_m = {}^{2m}C_m (\tfrac{1}{2})^m . (\tfrac{1}{2})^m$$

and

$$f_{m-p} = {}^{2m}C_{m-p} (\tfrac{1}{2})^{m-p} (\tfrac{1}{2})^{m+p}$$

$$\therefore \quad \frac{f_m}{f_{m-p}} = \frac{{}^{2m}C_m (\frac{1}{2})^{2m}}{{}^{2m}C_{m-p} . (\frac{1}{2})^{2m}}$$

or

$$\frac{f_m}{f_{m-p}} = \frac{\underline{|m-p}}{\underline{|m}} . \frac{\underline{|m+p}}{\underline{|m}} .$$

Since m is very large hence applying Stirlings approximation we have

$$\frac{f_m}{f_{m-p}} = \frac{e^{-(m-p)} . (m-p)^{m-p+\frac{1}{2}} . \sqrt{2\pi} . e^{-(m+p)} (m+p)^{m+p+\frac{1}{2}} . \sqrt{2\pi}}{e^{-m} . m^{m+\frac{1}{2}} . \sqrt{2\pi} . e^{-m} m^{m+\frac{1}{2}} . \sqrt{2\pi}}$$

or

$$\frac{fm}{f_{m-p}} = \frac{[(m-p)\ (m+p)]^{m+\frac{1}{2}} . \left[\frac{m+p}{m-p}\right]^p}{m^{2m+1}}$$

or $$\frac{f_m}{f_{m-p}} = \left[\left(1-\frac{p}{m}\right)\left(1+\frac{p}{m}\right)\right]^{m+\frac{1}{2}} \left\{\frac{1+\frac{p}{m}}{1-\frac{p}{m}}\right\}^p$$

or $$\log \frac{f_m}{f_{m-p}} = \left(m+\frac{1}{2}\right)\left[\log\left(1-\frac{p}{m}\right)\left(1+\frac{p}{m}\right)\right] + p \log \frac{1+\frac{p}{m}}{1-\frac{p}{m}}$$

or $$\log \frac{f_m}{f_{m-p}} = (m+\tfrac{1}{2})\left[2\left(-\frac{p^2}{2m^2} - \frac{p^4}{4m^4} \cdots\right]\right. + p\left[2\left(\frac{p}{m} + \frac{p^3}{3m^3} + \cdots\right)\right]$$

or $$\log \frac{f_m}{f_{m-p}} = \left(-\frac{p^2}{m} + \frac{2p^2}{m}\right) - \frac{p^2}{2m^2} + \text{higher powers of } \frac{1}{m}$$

or $$\log \frac{f_m}{f_{m-p}} = \frac{p^2}{m} \text{ neglecting higher powers of } \frac{1}{m}$$

$$\therefore \quad \frac{f_m}{f_{m-p}} = e^{\frac{p^2}{m}}$$

or $$f_{m-p} = f_m \cdot e^{-\frac{p^2}{m}}$$

Now $$f_m = (\tfrac{1}{2})^{2m} \cdot \frac{\lfloor 2m}{\lfloor m \, \lfloor m}$$

Using stirlings approximation we have

$$f_m = (\tfrac{1}{2})^{2m} \frac{e^{-2m}\,(2m)^{2m+\frac{1}{2}}\,\sqrt{2\pi}}{[e^{-m}.(m)^{m+\frac{1}{2}}.\sqrt{2\pi}]^2}$$

$$= \frac{1}{\sqrt{\pi m}}$$

$$\therefore \quad f_{m-p} = \frac{1}{\sqrt{\pi m}} \cdot e^{-\frac{p^2}{m}}$$

or $$f_{\frac{N}{2}-p} = \left(\frac{2}{\pi N}\right)^{\frac{1}{2}} \cdot e^{-\frac{2p^2}{N}}$$

where $f_{\frac{N}{2}-p}$ denotes probability of exactly $\frac{N}{2}-p$ heads and $\frac{N}{2}+p$ tails.

Problem 329.—Derive the normal distribution as a limiting case of binomial distribution when $p=q$.

(*M.Sc*, *Agra*, *1948*, *1954*; *I.A.S.*, *1956*)

Solution :

Let the binomial distribution be $N(p+q)^n$. Since $p=q=\frac{1}{2}$, therefore the distribution is symmetrical. Let us assume that n is an even integer say $2k$ where k is large integer. There is no loss of generality in assuming this as n tends to infinity ultimately. The binomial distribution may be written as $N(\frac{1}{2}+\frac{1}{2})^{2k}$. The frequencies of r and $r+1$ successes are given by

$$f_r={}^{2k}C_r(\tfrac{1}{2})^{2k}$$

and $$f_{r+1}={}^{2k}C_{r+1}(\tfrac{1}{2})^{2k}$$

now $$f_{r+1} > \text{ or } < f_r$$

according as

$$\frac{{}^{2k}C_{r+1}(\frac{1}{2})^{2k}}{{}^{2k}C_r(\frac{1}{2})^{2k}} > \text{ or } < 1$$

or $$\frac{2k-r}{r+1} > \text{ or } < 1$$

or $$r < \text{ or } > k-\tfrac{1}{2}$$

If $r=k-1$ $\quad f_{r+1} > f_r$

If $r=k$ $\quad f_r > f_{r-1}$

hence f_k gives the maximum frequency let us assume it y_0.

$$\therefore \quad y_0={}^{2k}C_k(\tfrac{1}{2})^{2k}$$

The frequency of $k+x$ successes is given by

$$y_x={}^{2k}C_{k+x}(\tfrac{1}{2})^{2k}$$

$$\therefore \quad \frac{y_x}{y_0}=\frac{{}^{2k}C_{k+x}(\frac{1}{2})^{2k}}{{}^{2k}C_k(\frac{1}{2})^{2k}}$$

$$=\frac{|\underline{k}}{|\underline{k+x}}\cdot\frac{|\underline{k}}{|\underline{k-x}}$$

$$=\frac{e^{-2k}.k^{2k+1}2\pi}{e^{-k-x}(k+x)^{k+x+\frac{1}{2}}.e^{-k+x}(k-x)^{k-x+\frac{1}{2}}2\pi}$$

(using Stirlings approximation)

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

or $$\log\frac{y_x}{y_0} = -\left(k+x+\frac{1}{2}\right)\log\left(1+\frac{x}{k}\right) - \left(k-x+\frac{1}{2}\right)\log\left(1-\frac{x}{k}\right)$$

$$= -\frac{x^2}{k} \text{ neglecting higher powers of } \frac{1}{k}$$

$$= -\frac{2x^2}{n}$$

$\therefore$ $$y_x = y_0 \, e^{-\frac{2x^2}{n}}$$

or $$y_x = y_0 \, e^{-\frac{x^2}{2\sigma^2}} \qquad \left[\text{since } \sigma^2 = npq = \frac{1}{4}n\right.$$

This is normal distribution.

Problem 330.—Show that if two symmetrical binomial distributions of degree n (the same number of observations) are so superposed that the r^{th} term of the one coincides with the $(r+1)^{th}$ term of the other, the distribution formed by adding superposed terms is a symmetrical binomial of degree $(n+1)$.

(M.Sc. Agra, 1949, 1953, 1957)

Solution :

For symmetrical distribution $p=q=\frac{1}{2}$.

Let the binomial distribution be $N\left(\frac{1}{2}+\frac{1}{2}\right)^n$

or $$N = N\left[{}^nC_0\left(\tfrac{1}{2}\right)^n + {}^nC_1\left(\tfrac{1}{2}\right)^{n-1}\left(\tfrac{1}{2}\right) + \ldots\ldots + {}^nC_{r-1}\left(\tfrac{1}{2}\right)^{n-r+1}\left(\tfrac{1}{2}\right)^{r-1} + {}^nC_r\left(\tfrac{1}{2}\right)^{n-r}\left(\tfrac{1}{2}\right)^r + \ldots\ldots {}^nC_n\left(\tfrac{1}{2}\right)^n\right]$$

or $$N = N\left(\tfrac{1}{2}\right)^n\left[{}^nC_0 + {}^nC_1 + \ldots\ldots + {}^nC_{r-1} + {}^nC_r + \ldots {}^nC_n\right]$$

The other symmetrical distribution with same number of observations is

$$N = N\left(\tfrac{1}{2}\right)^n\left[{}^nC_0 + {}^nC_1 + \ldots\ldots + {}^nC_{r-1} + {}^nC_r + \ldots\ldots {}^nC_n\right]$$

Superposing the first distribution over second as required in the question we have

$$2N = N\left(\tfrac{1}{2}\right)^n\left[{}^nC_0 + ({}^nC_0 + {}^nC_1) + \ldots\ldots + ({}^nC_{r-1} + {}^nC_r) + ({}^nC_r + {}^nC_{r+1}) + \ldots\ldots + ({}^nC_{n-1} + {}^nC_n) + {}^nC_n\right]$$

or $$2N = N\left(\tfrac{1}{2}\right)^n\left[{}^{n+1}C_0 + {}^{n+1}C_1 + {}^{n+1}C_2 + \ldots + {}^{n+1}C_r + \ldots {}^{n+1}C_{n+1}\right]$$

or $$N = N\left(\tfrac{1}{2}\right)^{n+1}\left[{}^{n+1}C_0 + {}^{n+1}C_1 + {}^{n+1}C_2 + \ldots + {}^{n+1}C_r + \ldots + {}^{n+1}C_{n+1}\right]$$

or $N=N(\frac{1}{2}+\frac{1}{2})^{n+1}$

Which is a symmetrical binomial distribution of degree $(n+1)$.

Problem 331.—Show that if np be a whole number, the mean of the binomial distribution coincides with the greatest term.

Solution :

Let the binomial distribution be $N(q+p)^n$ then we know that mean$=np$

Also as shown in the article for normal distribution the maximum frequency is also for np successes. Hence the result.

Problem 332.—A perfect cubic die is thrown a large number of times in sets of 8. The occurrence of a 5 or a 6 is called a success. In what proportion of the sets would you expect 3 successes?

(*M. Sc., Agra. 1948*)

Solution :

The chance of getting 5 or 6 with one die is $\frac{2}{6}$ or $\frac{1}{3}$.

$\therefore \quad p=\frac{1}{3},\ q=\frac{2}{3}$.

Since dies are in sets of 8, hence the binomial distribution is $N(\frac{2}{3}+\frac{1}{3})^8$.

$\therefore$ Frequency of getting three successes$=N\,{}^8C_3\,(\frac{2}{3})^5\,(\frac{1}{3})^3$

$\therefore$ Proportion of the sets in which three successes are expected

$$=\frac{N\,{}^8C_3(\frac{2}{3})^5.(\frac{1}{3})^3}{N}$$

$$=\frac{8\times7\times6\times2^5}{3\times2\times3^8}$$

$$=\frac{1792}{6561}$$

$$=\frac{1792\times100}{6561}\ \text{per cent}$$

$$=27{\cdot}31\ \text{per cent.}$$

Problem 333.—A perfect coin is tossed 1000 times in sets of 12. (*a*) In how many cases should we expect to get 8 heads and 4 tails. (*b*) In how many cases should we expect to get 8 heads at least.

Solution :

Here $p=q=\frac{1}{2}$

Hence the numbers of successes 0, 1, 2,...12 are the terms in $1000\,(\frac{1}{2}+\frac{1}{2})^{12}$

i.e. $1000[(\frac{1}{2})^{12}+{}^{12}C_1(\frac{1}{2})^{11}.(\frac{1}{2})+{}^{12}C_2(\frac{1}{2})^{10}.(\frac{1}{2})^2+\ldots\ldots+(\frac{1}{2})^{12}]$.

(*a*) The term giving 8 successes and 4 failures is

$$=1000.\,{}^{12}C_8\left(\tfrac{1}{2}\right)^8.\left(\tfrac{1}{2}\right)^4$$

$$=1000.\,{}^{12}C_4\left(\tfrac{1}{2}\right)^{12}$$

$$=1000\,.\,\frac{12\times11\times10\times9}{4\times3\times2\times2^{12}}$$

$$=1000\,.\,\frac{495}{4096}$$

$$=\frac{495000}{4096}$$

$$=120{\cdot}849$$

$$=121 \text{ approximately.}$$

(*b*) Here we require the sum of terms with 8, 9, 10, 11 and 12 successes. Hence our expected number is

$$=1000\times\left(\tfrac{1}{2}\right)^{12}\left[{}^{12}C_8+{}^{12}C_9+{}^{12}C_{10}+{}^{12}C_{11}+{}^{12}C_{12}\right]$$

$$=\frac{1000}{4096}\left[\,{}^{12}C_4+{}^{12}C_3+{}^{12}C_2+{}^{12}C_1+1\,\right]$$

$$=\frac{1000}{4096}\left[\frac{12\times11\times10\times9}{4\times3\times2}+\frac{12\times11\times10}{3\times2}+\frac{12\times11}{2}+12+1\right]$$

$$=\frac{1000}{4096}\left[495+220+66+12+1\right]$$

$$=\frac{1000\times794}{4096}$$

$$=\frac{794000}{4096}$$

$$=193{\cdot}847$$

$$=194 \text{ approximately.}$$

Problem 334.—The following data show the results of throwing 12 dice 26306 times, a throw of 5 or 6 reckoned a success.

No. of successes	0	1	2	3	4	5	6	7	8	9	10 & over
Frequency	185	1149	3265	5475	6114	5194	3067	1331	403	105	18

Find the expected frequencies and compare the actual mean and standard deviation with those of the expected distribution.

Solution :

This is a binomial distribution in which $p=\frac{2}{6}=\frac{1}{3}$, $q=\frac{2}{3}$, $n=12$ and $N=26306$.

The expected frequencies for 0, 1, 2,......successes are given by the successive terms of expansion $26306(\frac{2}{3}+\frac{1}{3})^{12}$ which are obtained as below :

Successes	Frequency
0	$26306\,(\frac{2}{3})^{12}=203$
1	$26306\,{}^{12}C_1(\frac{2}{3})^{11}(\frac{1}{3})=1217$
2	$26306\,{}^{12}C_2(\frac{2}{3})^{10}(\frac{1}{3})^2=3345$
3	$26306\,{}^{12}C_3(\frac{2}{3})^9(\frac{1}{3})^3=5576$
4	$26306\,{}^{12}C_4(\frac{2}{3})^8(\frac{1}{3})^4=6273$
5	$26306\,{}^{12}C_5(\frac{2}{3})^7(\frac{1}{3})^5=5018$
6	$26306\,{}^{12}C_6(\frac{2}{3})^6(\frac{1}{3})^6=2927$
7	$26306\,{}^{12}C_7(\frac{2}{3})^5(\frac{1}{3})^7=1254$
8	$26306\,{}^{12}C_8(\frac{2}{3})^4(\frac{1}{3})^8=392$
9	$26306\,{}^{12}C_9(\frac{2}{3})^3(\frac{1}{3})^9=87$
10 & over	$26306\,{}^{12}C_{10}(\frac{2}{3})^2(\frac{1}{3})^{10}+\ldots+\ldots=14$

$$\text{The expected value of mean}=np=12\times\tfrac{1}{3}=4$$

$$\text{Standard deviation}=\sqrt{npq}=\sqrt{12\times\tfrac{2}{3}\times\tfrac{1}{3}}=1{\cdot}63$$

Let us now determine the value of the constants from the actual figures obtained by the experiment.

x	f	$\xi=x-4$	$f\xi$	$f\xi^2$
0	185	−4	−740	2960
1	1149	−3	−3447	10341
2	3265	−2	−6530	13060
3	5475	−1	−5475	5475
4	6114	0	0	0
5	5194	1	5194	5194
6	3067	2	6134	12268
7	1331	3	3993	11979
8	403	4	1612	6448
9	105	5	525	2625
10 and over	18	6	108	648
	26306		1374	70998

$$\text{Actual mean}=4+\frac{1374}{26306}=4{\cdot}05$$

Actual standard deviation $=\sqrt{\frac{\Sigma f\xi^2}{\Sigma f}-\left(\frac{\Sigma f\xi}{\Sigma f}\right)^2}$

$$\sqrt{\frac{70998}{26306}-(\cdot 05)^2}$$

$$=\sqrt{2\cdot 6989-\cdot 0025}$$

$$=\sqrt{2\cdot 6964}$$

$$=1\cdot 64.$$

Problem 335.—In the previous problem, find the equation of normal curve, which has the same mean, standard deviation and total frequency as the observed distribution.

Solution :

$$\text{Mean}=4\cdot 05$$

$$\text{Standard deviation } \sigma=1\cdot 64$$

The equation to the normal curve with origion at the mean is given by

$$y=y_0\, e^{-\frac{x^2}{2\sigma^2}}$$

where y_0 is to be determined.

Now $$2\int_0^{\infty} y_0\, e^{-\frac{x^2}{2\sigma^2}}\,dx=26306$$

Or $$2y_0\int_0^{\infty} e^{-t^2}\sqrt{2}\,\sigma dt=26306$$

[If $x=\sqrt{2}\,\sigma t$]

Or $$2y_0\frac{\sqrt{\pi}}{2}\sqrt{2}\,\sigma=26306$$

$\therefore$ $$y_0=\frac{26306}{\sqrt{2}\,\sigma\sqrt{\pi}}$$

$$=\frac{26306}{1\cdot 64\sqrt{2\pi}}$$

Hence equation to normal curve with origin at mean is

$$y=\frac{26306}{1\cdot 64\sqrt{2\pi}}\, e^{-\frac{x^2}{2(1\cdot 64)^2}}$$

The equation to normal curve with origin at zero success is

$$y=\frac{26306}{1\cdot64\sqrt{2\pi}}\, e^{-\frac{(x-4\cdot05)^2}{2(1\cdot64)^2}}$$

Problem 336.—Assuming that half the population are consumers of chocolate, so that the chance of an individual being a consumer is $\frac{1}{2}$, and assuming that 100 investigators each take ten individuals to see whether they are consumers, how many investigators would you expect to report that three people or less were consumers.

(*Yule & Kandall 10.4 Page 194*)

Solution :

$p=q=\frac{1}{2}$, $n=10$, $N=100$

This being a binomial distribution $100\ (\frac{1}{2}+\frac{1}{2})^{10}$; the number of investigators to report that no person is consumer $=100.\ (\frac{1}{2})^{10}$

$$=\frac{100}{1024}$$

the number of investigators to report that one person is consumer $=100\ {}^{10}C_1\ (\frac{1}{2})^9.\ (\frac{1}{2})=\frac{1000}{1024}$ the number of investigators to report that two persons are consumers $=100\ {}^{10}C_2\ (\frac{1}{2})^{10}=\frac{4500}{1024}$ the number of investigators to report that three persons are consumers $=$ $100\ {}^{10}C_3\ (\frac{1}{2})^{10}=\frac{12000}{1024}$

$\therefore$ Total investigators required

$$=\frac{100+1000+4500+12000}{1024}$$

$$=\frac{17600}{1024}$$

$=17$ (because the persons cannot be in fraction.)

Problem 337.—An irregular six-faced die is thrown, and the expectation that in 10 throws it will give five even numbers is twice the expectation that it will give four even numbers. How many times in 10,000 sets of 10 throws would you expect it to give no even numbers.

(*Yule & Kandall 10.5 Page 194*)

Solution :

If p is the expectation of getting an even number, then

$${}^{10}C_5\, p^5q^5=2\,{}^{10}C_4\, p^4q^6$$

$$\text{or}\quad \frac{\lfloor 10}{\lfloor 5\ \lfloor 5}\, p^5q^5=\frac{\lfloor 10}{\lfloor 4\ \lfloor 6}\, p^4q^6$$

$$\text{or} \quad \frac{p}{5} = \frac{q}{3} = \frac{p+q}{8} = \frac{1}{8}$$

$$\therefore \quad p = \frac{5}{8}$$

$$q = \frac{3}{8}.$$

Hence the required number of times

$$= 10000 \times {}^{10}C_0 \left(\tfrac{5}{8}\right)^0 \left(\tfrac{3}{8}\right)^{10}$$

$$= \frac{10000 \times 3^{10}}{8^{10}}$$

$$= \frac{36905625}{67108864}$$

$$= 1 \text{ appr.}$$

Problem 338.—What are the conditions under which a Poissonian distribution can be expected? Calculate the mean, the standard deviation, and the third and fourth moments about the mean of such a distribution. Also prove that $M\delta\gamma_1\gamma_2 = 1$, where the symbols have their usual meaning.

(M. Sc., Agra, 1944, 1945, 1950, 1953, 1956)
(I.A.S., 1949)

Solution :

Poissonian distribution is expected when one of the chances say p or q becomes very small and n is sufficiently large so that np or nq remains finite but not necessarily large.

The respective frequencies for 0, 1, 2,........ successes of the Poisson's distribution are given by

$$Ne^{-m} \; ; \; Ne^{-m}.\, m \; ; \; Ne^{-m}.\, \frac{m^2}{\lfloor 2} \ldots\ldots\ldots ; \; Ne^{-m} \frac{m^r}{\lfloor r} \ldots\ldots\ldots \text{etc.}$$

Let us assume the origin to be located at the first point of the distribution, then the values of the deviation ξ are given by 0, 1, 2, ...,r,......and we atonce obtain

$$\mu_1' = \sum_0^{\infty} e^{-m}.\, r \frac{m^r}{\lfloor r} \qquad \left(\mu_1' = \frac{\Sigma f \xi}{N} \right)$$

$$= me^{-m} \Sigma \frac{m^{r-1}}{\lfloor r-1}$$

$$= m\, e^{-m}.\, e^m \qquad \left[\text{since } \Sigma \frac{m^{r-1}}{\lfloor r-1} = e^m \right]$$

$$= m$$

Hence m is the mean of the distribution

$$\mu_2' = \frac{\Sigma f\xi^2}{N}$$

$$= \Sigma\, e^{-m}\, \frac{m^r}{\underline{|r}}\,.\, r^2$$

$$= e^{-m}\, \Sigma\, m^r.\, \frac{[r(r-1)+r]}{\underline{|r}}$$

$$= e^{-m}\, m^2\, \Sigma \frac{m^{r-2}}{\underline{|r-2}} + e^{-m}.\, m\, \Sigma \frac{m^{r-1}}{\underline{|r-1}}$$

$$= m^2 + m$$

$$\mu_3' = \frac{\Sigma f\xi^3}{N}$$

$$= \Sigma\, e^{-m}.\, \frac{m^r}{\underline{|r}}\,.\, r^3$$

$$= e^{-m}\, \Sigma\, m^r\, \frac{[r(r-1)(r-2)+3r(r-1)+r]}{\underline{|r}}$$

$$= e^{-m}\, m^3\, \Sigma \frac{m^{r-3}}{\underline{|r-3}} + e^{-m}.\, 3m^2\, \Sigma \frac{m^{r-2}}{\underline{|r-2}} + e^{-m}.\, m\, \Sigma \frac{m^{r-1}}{\underline{|r-1}}$$

$$= m^3 + 3m^2 + m$$

$$\mu_4' = \frac{\Sigma f\xi^4}{N}$$

$$= \Sigma\, e^{-m}.\, \frac{m^r}{\underline{|r}}\,.\, r^4$$

$$= \Sigma\, e^{-m}\, m^r\, \frac{[r(r-1)(r-2)(r-3)+6r(r-1)(r-2)+7r(r-1)+r]}{\underline{|r}}$$

$$= e^{-m}.\, m^4\, \Sigma \frac{m^{r-4}}{\underline{|r-4}} + e^{-m} 6m^3 \Sigma \frac{m^{r-3}}{\underline{|r-3}} + e^{-m}.\, 7m^2\, \Sigma \frac{m^{r-2}}{\underline{|r-2}} + e^{-m}.\, m\, \Sigma \frac{m^{r-1}}{\underline{|r-1}}$$

$$= m^4 + 6m^3 + 7m^2 + m.$$

If d is the mean then we know that the n^{th} moment about the mean is given by the formula

$$\mu_n = \mu'_n - {}^nC_1\, \mu'_{n-1} d + {}^nC_2\, \mu'_{n-2} d^2 - \ldots\ldots\ldots\ldots$$

but $d = m$

$$\therefore \mu_1 = \mu_1' - \mu'_0.m = m - m = 0$$

now $\mu_2 = \mu_2' - {}^2C_1 \mu_1'.\, m + {}^2C_2\, \mu_0'.\, m^2$

$$= m^2 + m - 2m^2 + m^2$$

$$= m$$

Standard deviation $=\sqrt{\mu}=\sqrt{m}$

$$\mu_3=\mu_3'-{}^3C_1\,\mu_2'.\,m+{}^3C_2\mu_1'.\,m^2-{}^3C_3.\,\mu_0'.\,m^3$$
$$=m^3+3m^2+m-3\,(m^2+m).\,m+3.\,m^3-m^3$$
$$=m$$

$$\mu_4=\mu'_4-{}^4C_1\mu'_3.m+{}^4C_2\mu_2'm^2-{}^4C_3\mu'_1m^3+{}^4C_4\mu'_1m^4$$
$$=m^4+6m^3+7m^2+m-4(m^3+6m^2+m).m+6(m^2+m).m^2$$
$$-4m^4+m^4$$
$$=3m^2+m$$

Now $\beta_1=\dfrac{\mu_3^2}{\mu_2^3}=\dfrac{m^2}{m^3}=\dfrac{1}{m}$

$$\beta_2=\frac{\mu_4^2}{\mu_2^2}=\frac{3m^2+m}{m^2}=3+\frac{1}{m}$$

$$\gamma_1=\sqrt{\beta_1}=\frac{1}{\sqrt{m}}$$

$$\gamma_2=\beta_2-3=\frac{1}{m}$$

Now to prove $M\,\sigma\,\gamma_1\gamma_2=1$

$$\text{L.H.S.}=m.\,\sqrt{m}.\,\frac{1}{\sqrt{m}}.\,\frac{1}{m}$$
$$=1$$
$$=\text{R.H.S.}$$

Hence mean $=m$

Standard deviation $=\sqrt{m}$

Third moment about the mean $=m$

Fourth moment about the mean $=3m^2+m$.

Problem 339.—(*a*) Obtain the two β. coefficients for the Poisson distribution.

(*b*) Letters were received in an office on each of 100 days. Assuming the following data to form a random sample from a Poisson distribution, find the expected frequencies, correct to the nearest unit, taking $e^{-4}=\cdot0183$

Number of letters	0	1	2	3	4	5	6	7	8	9	10
Frequency	1	4	15	22	21	20	8	6	2	0	1

(*M.Sc. Agra, 1948*)

Solution :

(*a*) For β-coefficients see prevoius problem.

(*b*) total frequency N=100.

$$\text{mean } m=\frac{\Sigma f x}{N}$$

$$=\frac{0+4+30+66+84+100+48+42+16+0+10}{100}$$

$$=4$$

The expected frequencies are given by N $e^{-m}\cdot\frac{m^r}{|\underline{r}}$ or

$100\,e^{-4}\,\frac{4^r}{|\underline{r}}$ for different values of r.

Now $r=0$, $100\,e^{-4}$. $=100\times{\cdot}0183=1{\cdot}83$

$r=1,\ 100\,e^{-4}.\frac{4}{|\underline{1}}=1{\cdot}83\times 4 \quad =7{\cdot}32$

$r=2,\ 100\,e^{-4}.\frac{4^2}{|\underline{2}}=1{\cdot}83\times 8 \quad =14{\cdot}64$

$r=3,\ 100\,e^{-4}.\frac{4^3}{|\underline{3}}=1{\cdot}83\times\frac{32}{3} \quad =19{\cdot}52$

$r=4,\ 100\,e^{-4}.\frac{4^4}{|\underline{4}}=1{\cdot}83\times\frac{32}{3} \quad =19{\cdot}52$

$r=5,\ 100\,e^{-4}.\frac{4^5}{|\underline{5}}=1{\cdot}83\times\frac{128}{15}=15{\cdot}616$

$r=6,\ 100\,e^{-4}.\frac{4^6}{|\underline{6}}=1{\cdot}83\times\frac{256}{45}=10{\cdot}41$

$r=7,\ 100\,e^{-4}.\frac{4^7}{|\underline{7}}=1{\cdot}83\times\frac{4^7}{|\underline{7}}=5{\cdot}949$

$r=8,\ 100\,e^{-4}.\frac{4^8}{|\underline{8}}=1{\cdot}83\times\frac{4^8}{|\underline{8}}=2{\cdot}9745$

$r=9,\ 100\,e^{-4}.\frac{4^9}{|\underline{9}}=1{\cdot}83\times\frac{4^9}{|\underline{9}}=1{\cdot}322$

$r=10,\ 100\,e^{-4}.\frac{4^{10}}{|\underline{10}}=1{\cdot}83\times\frac{4^{10}}{|\underline{10}}={\cdot}529$

Hence we have approximately

Number of letters	0	1	2	3	4	5	6	7	8	9	10
expected frequency	2	7	15	20	20	16	10	6	3	1	1

✓ **Problem 340.**—Fit a Poissons distribution to the set of observations :

Deaths	0	1	2	3	4
Frequency	122	60	15	2	1

and calculate the theoretical frequencies.

(*M.Sc. Agra, 1949, 1954, 1957*)

Solution :

$$N=122+60+15+2+1=200$$

$$\text{Mean } m=\frac{0+60+30+6+4}{200}$$

$$=\frac{1}{2}=\cdot 5$$

Now $$e^{-\cdot 5}=1-(\cdot 5)+\frac{(\cdot 5)^2}{\underline{|2}}-\frac{(\cdot 5)^3}{\underline{|3}}+\ldots\ldots\ldots\ldots$$

$$=1-\cdot 5+\cdot 125-\cdot 0208+\ldots\ldots\ldots\ldots$$

$$=\cdot 61 \text{ nearly.}$$

The theoretical frequency for r deaths is

$$Ne^{-m}.\frac{m^r}{\underline{|r}}=200\times\cdot 61\frac{(\cdot 5)^r}{\underline{|r}}$$

$$=122\frac{(\cdot 5)^r}{\underline{|r}}$$

which gives frequency for $r=0, 1, 2\ldots\ldots$as 122, 61, 15, 2 and 0 respectively.

Problem 341.—In 1000 consecutive issues of the 'Utopian Seven Daily Chronicle' the deaths of centenarians were recorded, the number x having frequency f according to the table.

x	0	1	2	3	4	5	6	7	8
f	229	325	257	119	50	17	2	1	0

Show that the distribution is roughly Poissonian by calculating its mean, and then the frequencies in the Poissonian distribution with the same mean and the same total frequency of 1000. Also calculate the variance of the given distribution and compare it with the mean. [Given $e^{-1\cdot 5}=\cdot 2231$ appr.] (*M. Sc., Agra, 1951*)

Solution :

$$N=1000$$

$$\text{Mean } m=\frac{\Sigma f\xi}{N}$$

$$=\frac{0+325+514+357+200+85+12+7+0}{1000}$$

$$=\frac{1500}{1000}=1\cdot 5$$

now $e^{-1\cdot 5}=\cdot 2231$

The theoretical frequency of Poissonian distribution for r successes is given by N $e^{-m}.\frac{m^r}{\underline{|r}}$. Giving r the values 0, 1, 2,......etc. we get the theoretical frequencies as below.

x	Expected frequency by Poissonian distribution	
0	$Ne^{-m} = 1000 \times \cdot 2231$	$= 223 \cdot 1$
1	$Ne^{-m}.m = 223 \cdot 1 \times 1 \cdot 5$	$= 334 \cdot 7$
2	$Ne^{-m} \frac{m^2}{\lfloor 2} = 223 \cdot 1 \times \frac{(1 \cdot 5)^2}{\lfloor 2}$	$= 251$
3	$Ne^{-m} . \frac{m^3}{\lfloor 3} = 223 \cdot 1 \times \frac{(1 \cdot 5)^3}{\lfloor 3}$	$= 125 \cdot 5$
4	$Ne^{-m} . \frac{m^4}{\lfloor 4} = 223 \cdot 1 \times \frac{(1 \cdot 5)^4}{\lfloor 4}$	$= 47 \cdot 1$
5	$Ne^{-m} . \frac{m^5}{\lfloor 5} = 223 \cdot 1 \times \frac{(1 \cdot 5)^5}{\lfloor 5}$	$= 14 \cdot 1$
6	$Ne^{-m} . \frac{m^6}{\lfloor 6} = 223 \cdot 1 \times \frac{(1 \cdot 5)^6}{\lfloor 6}$	$= 3 \cdot 5$
7	$Ne^{-m} . \frac{m^7}{\lfloor 7} = 223 \cdot 1 \times \frac{(1 \cdot 5)^7}{\lfloor 7}$	$= \cdot 8$
8	$Ne^{-m} . \frac{m^8}{\lfloor 8} = 223 \cdot 1 \times \frac{(1 \cdot 5)^8}{\lfloor 8}$	$= \cdot 2$

These two frequencies are roughly equal hence they represent the Poissonian distribution.

Now $$\sigma^2 = \frac{\Sigma f x^2}{\Sigma f} - d^2$$

$$= \frac{0+325+1028+1071+800+425+72+49+0}{1000} - (1 \cdot 5)^2$$

$$= \frac{3770}{1000} - (1 \cdot 5)^2$$

$$= 3 \cdot 77 - 2 \cdot 25$$

$$= 1 \cdot 52$$

$\therefore$ mean $= 1 \cdot 5$

variance $= 1 \cdot 52$

Problem 342.—In a certain factory turning razor blades, there is a small chance $\frac{1}{500}$ for any blade to be defective. The blades are supplied in packets of 10. Use Poisson's distribution to calculate the approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.

(*M.Sc., Agra, 1955*)

Solution :

$$N=10{,}000$$

$$m=nq \text{ approximately}$$

$$=10 \times \frac{1}{500}$$

$$=\cdot 02$$

$$e^{-m}=e^{-\cdot 02}=1-\cdot 02+\frac{(\cdot 02)^2}{\lfloor 2}$$

$$=\cdot 9802 \text{ appr.}$$

The respective frequencies are given by

$$Ne^{-m},\ Ne^{-m}.m \text{ and } Ne^{-m}\times\frac{m^2}{\lfloor 2}.$$

Substituting these values we get 0, 1 and 2 defective blades in nearly 9802, 196 and 2 packets respectively.

Problem 343.—Find the mean and standard deviation for the table of deaths of women over 85 years old recorded in a three-year period.

No. of deaths recorded in a day	0	1	2	3	4	5	6	7
No. of days	364	376	218	89	33	13	2	1

Find the expected number of days with one death recorded for the Poisson series fitted to the data.

(*M.Sc., Agra 1958*)

Solution :

$$N=1096 \text{ (Total frequency)}$$

Calculation of mean and standard deviation,

(Number of deaths) per day x	*Frequency* f	fx	fx^2
0	364	0	0
1	376	376	376
2	218	436	872
3	89	267	801
4	33	132	528
5	13	65	325
6	2	12	72
7	1	7	49
		1295	3023

$$\therefore \qquad m = \frac{1295}{1096} = 1{\cdot}18$$

$$\sigma^2 = \frac{\Sigma f x^2}{\Sigma f} - d^2$$

$$= \frac{3023}{1096} - (1{\cdot}18)^2$$

$$= 2{\cdot}758 - 1{\cdot}392$$

$$= 1{\cdot}366$$

$$\sigma = 1{\cdot}17 \text{ appr.}$$

$$e^{-m} = e^{-1{\cdot}18} = 1 - 1{\cdot}18 + \frac{(1{\cdot}18)^2}{\lfloor 2} - \frac{(1{\cdot}18)^3}{\lfloor 3} + \cdots$$

$$= {\cdot}307$$

Expected number of days with one death per day are $Ne^{-m}.m$

$$= 1096 \times {\cdot}307 \times 1{\cdot}18$$

$$= 397{\cdot}1$$

Problem 344.—In 1000 extensive sets of trials for an event of small probability, the frequency f of the number x of successes proved to be

x	0	1	2	3	4	5	6	7
f	305	365	210	80	28	9	2	1

Assuming it to be a Poissonian distribution calculate its mean, variance and expected frequencies for the Poissonian distribution with same mean. (*Weatherburn, Page 61*)

Solution :

Calculation of mean and variance.

x	f	fx	fx^2
0	305	0	0
1	365	365	365
2	210	420	840
3	80	240	720
4	28	112	448
5	9	45	225
6	2	12	72
7	1	7	49
	$\Sigma f = 1000$	$\Sigma fx = 1201$	$\Sigma fx^2 = 2719$

Mean $\quad m = \frac{\Sigma fx}{\Sigma f} = \frac{1201}{1000} = 1{\cdot}2$ approx

$$\sigma^2 = \frac{\Sigma fx^2}{\Sigma f} - d^2$$

$$= \frac{2719}{1000} - (1 \cdot 2)^2$$
$$= 2 \cdot 719 - 1 \cdot 44$$
$$= 1 \cdot 279$$

$$e^{-m} = e^{-1 \cdot 2} = 1 - 1 \cdot 2 + \frac{(1 \cdot 2)^2}{|\underline{2}} - \frac{(1 \cdot 2)^3}{|\underline{3}} + \ldots$$
$$= \cdot 3012$$

Expected frequencies are given below

x	Expected frequency (f)
0	$Ne^{-m} = \cdot 3012 \times 1000 = 301 \cdot 2$
1	$Ne^{-m}m = 301 \cdot 2 \times 1 \cdot 2 = 361 \cdot 4$
2	$Ne^{-m} \frac{m^2}{|\underline{2}} = 301 \cdot 2 \times \frac{(1 \cdot 2)^2}{|\underline{2}} = 216 \cdot 8$
3	$Ne^{-m} \frac{m^3}{|\underline{3}} = 301 \cdot 2 \times \frac{(1 \cdot 2)^3}{|\underline{3}} = 86 \cdot 7$
4	$Ne^{-m} \cdot \frac{m^4}{|\underline{4}} = 301 \cdot 2 \times \frac{(1 \cdot 2)^4}{|\underline{4}} = 26$
5	$Ne^{-m} \cdot \frac{m^5}{|\underline{5}} = 301 \cdot 2 \times \frac{(1 \cdot 2)^5}{|\underline{5}} = 6 \cdot 2$
6	$Ne^{-m} \cdot \frac{m^6}{|\underline{6}} = 301 \cdot 2 \times \frac{(1 \cdot 2)^6}{|\underline{6}} = 1 \cdot 2$
7	$Ne^{-m} \cdot \frac{m^7}{|\underline{7}} = 301 \cdot 2 \times \frac{(1 \cdot 2)^7}{|\underline{7}} = \cdot 2$

Problem 345.—Obtain the first four moments of the Normal distribution and hence the values of β_1 and β_2. *(I.A.S. 1948; M.Sc., Agra, 1948, 1953)*

Solution :

The equation to the normal curve with origin at the mean is

$$y = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$$

We have the n^{th} moment about the origin as

$$\mu'_n = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} x^n \, dx$$

(*i*) When n is odd this integral vainshes. Hence in a normal distribution all moments of odd order about the origin are zero, *i.e.* $\mu'_1 = \mu'_3 = \mu'_5 = \ldots = \mu'_{2r+1} = 0$

(ii) When n is even say $2r$

$$\text{then } \mu'_{2r} = 2\int_0^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} x^{2r}\, dx$$

$$= \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty e^{-t} \sigma^{2r} (2t)^r \frac{\sigma dt}{\sqrt{2t}} \qquad \left[\text{put } \frac{x^2}{2\sigma^2} = t \right]$$

$$= \frac{2^{r+\frac{1}{2}} \sigma^{2r+1}}{\sigma\sqrt{2\pi}} \int_0^\infty e^{-t} t^{r-\frac{1}{2}}\, dt$$

$$= \frac{2^r \sigma^{2r}}{\sqrt{\pi}} \cdot \Gamma\left(r+\tfrac{1}{2}\right)$$

$$= \frac{2^{\frac{n}{2}} \sigma^n}{\sqrt{\pi}} \cdot \Gamma\left(\frac{n+1}{2}\right)$$

putting $n=2$, $\therefore\ \mu_2' = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2}\sqrt{\pi} = \sigma^2$

,, $n=4$, $\mu_4' = \frac{2^2\sigma^4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right) = \frac{4\sigma^4}{\sqrt{\pi}} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} = 3\sigma^4$

also $\mu_0' = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1$

Since the mean is at the origin itself therefore all the μ' are μ.

Hence $\mu_2 = \mu_2' = \sigma^2$

$\mu_4 = \mu_4' = 3\sigma^4$

and $\mu_1 = \mu_3 = \mu_5 = \ldots\ldots = 0$

now $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3$

$\gamma_1 = \sqrt{\beta_1} = 0$

$\gamma_2 = \beta_2 - 3 = 0$ *i.e.* the normal curve has zero kurtosis.

Problem 346.—If two normal universes have the same total frequeney but the standard deviation of one is k times that of the other, show that the maximum frequency of the first is $\frac{1}{k}$ that of the other.

(*I.A.S. 1947 ;*
P.C.S. 1952 ;
M.Sc. Agra, 1947, 1951)

Solution :

	First universe	*Second universe*
Total frequency	N	N
Standard deviation	Kσ	σ

For first normal distribution is

$$y=y_0\, e^{-\frac{x^2}{2k^2\sigma^2}}$$

for which y_0 is the maximum frequency.

For second normal distribution is

$$y=y'_0\, e^{-\frac{x'^2}{2\sigma^2}}$$

for which maximum frequnecy is y_0'

Since the total frequency for both the distributions are same hence

$$\int_{-\infty}^{\infty} y_0\, e^{-\frac{x^2}{2k^2\sigma^2}}\, dx = \int_{-\infty}^{\infty} y_0'\, e^{-\frac{x'^2}{2\sigma^2}}\, dx'.$$

$$\text{or} \quad 2y_0\int_0^{\infty} e^{-\frac{x^2}{2k^2\sigma^2}}\, dx = 2y'_0 \int_0^{\infty} e^{-\frac{x'^2}{2\sigma^2}}\, dx'.$$

In L.H.S. put $\frac{x^2}{2k^2\sigma^2}=t^2 \quad \therefore \quad dx=\sqrt{2}\, k\sigma\, dt.$

In R.H.S. put $\frac{x'^2}{2\sigma^2}=t'^2 \quad \therefore \quad dx'=\sqrt{2}\, \sigma\, dt'$

$$\therefore \quad 2y_0\sqrt{2}\, k\sigma\int_0^{\infty} e^{-t^2}\, dt = 2y_0'\sqrt{2}\, \sigma\int_0^{\infty} e^{-t'^2}\, dt'.$$

or $$y_0\, k=y'_0$$

or max. frequency of first $= \frac{1}{k}\times$ maximum frequency of second.

Problem 347.—Find graphically or otherwise the point of inflection of the normal curve, and show that it occurs at a distance σ from the mean ordinate.

(*Yule & Kandall 10·7 Page 194*)

Solution :

Let the normal curve be

$$y=y_0\, e^{-\frac{x^2}{2\sigma^2}}$$

We know that at the point of inflection

$$\frac{d^2y}{dx^2}=0.$$

Differentiating $\frac{dy}{dx}=-\frac{y_0}{\sigma^2}\,e^{-\frac{x^2}{2\sigma^2}}\cdot x$

Again differentiating

$$\frac{d^2y}{dx^2}=+\frac{y_0}{\sigma^4}\,x^2\cdot e^{-\frac{x^2}{2\sigma^2}}-\frac{y_0}{\sigma^2}\,e^{-\frac{x^2}{2\sigma^2}}$$

$$=-y_0\,e^{-\frac{x^2}{2\sigma^2}}\cdot\frac{1}{\sigma^2}\left[1-\frac{x^2}{\sigma^2}\right]$$

∴ For the point of inflection

$$1-\frac{x^2}{\sigma^2}=0$$

$$\text{or}\quad x=\pm\sigma$$

∴ the point of inflection occurs at a distance σ from the mean ordinate.

The value of the ordinate at point of inflection is given by

$$y_\sigma=y_0\,e^{-\frac{\sigma^2}{2\sigma^2}}$$

$$=y_0\,e^{-\frac{1}{2}}\,.$$

Problem 348.—5000 candidates appeared in a certain examination paper carrying a maximum of 100 marks. It was found that the marks were normally distributed with a mean 39·5 and with standard deviation 12·5. Determine approximately the number of students who secured a first class, for which a minimum of 60 marks is necessary. You may use the table given below.

The proportion, A, of the whole area of the normal curve lying to the left of the ordinate at the deviation $\frac{x}{\sigma}$.

$\frac{x}{\sigma}$	1·5	1·6	1·7	1·8
A	·93319	·94520	·95543	·96407

Solution :

$$\frac{x}{\sigma}=\frac{60-39\cdot5}{12\cdot5}=1\cdot64 \text{ as } x \text{ is deviation from mean.}$$

Interpolation of the value of A for $\frac{x}{\sigma}=1\cdot64$

$\frac{x}{6}$	A	Differences		
		$\triangle$	$\triangle^2$	$\triangle^3$
1·5	·93319			
		·01201		
1·6	·94520		−·00178	
		·01023		−·00019
1·7	·95543		−·00159	
		·00864		
1·8	·96407			

$$t=\frac{x-a}{h}=\frac{1\cdot64-1\cdot5}{\cdot1}=\frac{\cdot14}{\cdot1}=1\cdot4$$

$$y_t=y_0+t\triangle y_0+\frac{t(t-1)}{\lfloor 2}\triangle^2 y_0+\frac{t(t-1)(t-2)}{\lfloor 3}\triangle^3 y_0$$

$$=\cdot93319+1\cdot4\times\cdot01201-\frac{1\cdot4\times\cdot4}{2}\times\cdot00178$$

$$+\frac{1\cdot4\times\cdot4\times\cdot6}{6}\times\cdot00019$$

$$=\cdot93319+\cdot016814-\cdot0004984-\ldots\ldots$$

$$=\cdot95 \text{ appr.}$$

Thus number of students getting 60 or more than 60 marks is only 5%.

Problem 349.—If skulls are classified as A, B, C according as the length-breadth index is under 75, between 75 and 80, or over 80 ; find approximately (assuming that the distribution is normal) the mean and standard deviation of a series in which A are 58%, B are 38% and C are 4%, being given that if

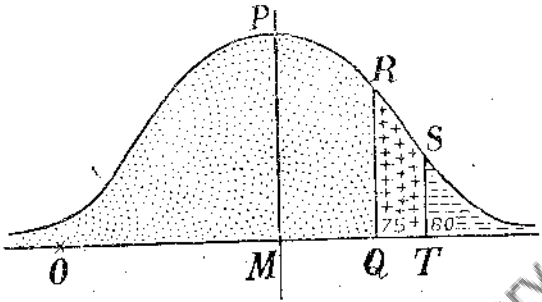
$$f(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}\,dt$$

then $f(\cdot20)=\cdot08$ and $f(1\cdot75)=\cdot46$

(P.C.S., 1954 ; M.Sc., Agra, 1949, 1955)

Solution :

Let m be the mean and σ the standard deviation of the distribution. Since the total frequency is taken to be one ; hence fre-



quency of skull A, whose length and breadth index is under 75 is ·58 ; the frequency of skull B, whose index lie between 75 and 80 is ·38 ; and the frequency of skull C, whose index is over 80 is ·04. Therefore the total area to the left of ordinate RQ is ·58, area between ordinates RQ and ST is ·38, and area to the right of ordinate ST is ·04.

Therefore the area between origin and $x(=75-m)$ *i.e.*, area PRQM is $= \cdot58 - \cdot5 = \cdot08$.

i.e., area corresponding to $t = \dfrac{75-m}{\sigma}$ is ·08.

But $f(\cdot20) = \cdot08$

Hence $\dfrac{75-m}{\sigma} = \cdot20$ ... ... (1)

Again, the area between origin and $x(=80-m)$ *i.e.*, area PSTM is

$$= \cdot08 + \cdot38 = \cdot46$$

i.e, area corresponding to $t = \dfrac{80-m}{\sigma}$ is ·46.

But $f(1\cdot75) = \cdot46$

Hence $\dfrac{80-m}{\sigma} = \cdot46$ ... (2)

From (1) $\cdot2\sigma + m = 75$

From (2) $\cdot46\,\sigma + m = 80$

Solving these two we have

$$m = 74{\cdot}3$$

$$\sigma = 3{\cdot}23$$

Problem 350.—If the probability that a deviation lies between x and $-x$ is given by

$$\phi(hx) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2}\, h\,dx$$

where $\phi(\ ) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-y^2}\, dy$

show that $\phi(y) = 1 - \frac{e^{-y^2}}{y\sqrt{\pi}} \left[1 - \frac{1}{2y^2} + \frac{1{\cdot}3}{(2y^2)^2} - \frac{1{\cdot}3{\cdot}5}{(2y^2)^3} + \ldots\ldots \right]$

(M.Sc., Agra, 1945, 1957)

Solution

Since $\sqrt{\pi} = 2\int_0^\infty e^{-y^2}\, dy$

we have $\sqrt{\pi} - \sqrt{\pi}\,\phi(y) = 2\int_0^\infty e^{-y^2}\, dy - 2\int_0^y e^{-y^2}\, dy$

$$= 2\int_y^\infty e^{-y^2}\, dy$$

$$= \int_t^\infty e^{-t} \,.\, t^{-\frac{1}{2}}\, dt \text{ if } y^2 = t$$

Integrating by parts we have

$$\sqrt{\pi} - \sqrt{\pi}\,\phi(y) = \left[-e^{-t} \,.\, t^{-\frac{1}{2}} \right]_t^\infty - \tfrac{1}{2}\int_t^\infty e^{-t} .\, t^{-\frac{3}{2}}\, dt$$

$$= e^{-t} \,.\, t^{-\frac{1}{2}} - \tfrac{1}{2}\int_t^\infty e^{-t} \,.\, t^{-\frac{3}{2}}\, dt$$

Successive integration by parts give the result

$$\sqrt{\pi}-\sqrt{\pi}\ \phi(y)=e^{-t}.\,t^{-\frac{1}{2}}-\tfrac{1}{2}\,e^{-t}.\,t^{-\frac{3}{2}}+\frac{1.3}{2.2}e^{-t}\,t^{-\frac{5}{2}}-\frac{1\cdot3\cdot5}{2\cdot2\cdot2}\,e^{-t}.\,t^{-\frac{7}{2}}+\ldots\ldots$$

or $\sqrt{\pi}-\sqrt{\pi}\,\phi\ (y)=e^{-y^2}.\,y^{-1}\left[1-\frac{1}{2y^2}+\frac{1\cdot3}{2\cdot2}\cdot\frac{1}{y^4}-\ \ldots\ldots\right]$

$$\therefore\ \phi(y)=1-\frac{e^{-y^2}}{y\sqrt{\pi}}\left[1-\frac{1}{2y^2}+\frac{1\cdot3}{(2y^2)^2}\ -\ldots\ldots\right]$$

Problem 351.—If the probability that a deviation lies between n and $-x$ is given by

$$\phi\ (hx)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-h^2x^2}hdx$$

$$\text{Where}\ \phi\ (y)=\frac{2}{\sqrt{\pi}}\int_0^y e^{-y^2}dy$$

Show that

$$\phi\ (y)=\frac{2}{\sqrt{\pi}}e^{-y^2}y\left[1+\frac{1}{3}.\ (2y^2)+\frac{1}{3\cdot5}\ (2y^2)^2+\frac{1}{3\cdot5\cdot7}\ (2y^2)^3+\ldots\ldots\ldots\ldots\right]$$

(M. Sc., Agra, 1946, 1955)

Solution :

$$\text{Let}\ \int_0^y e^{-y^2}\ dy=e^{-y^2}\cdot t$$

Differentiating we get

$$e^{-y^2}=\ e^{-y^2}\frac{dt}{dy}-2yte^{-y^2}$$

or $\dfrac{dt}{dy}-2\,yt=1\ldots\ldots$ ...(1)

as $\phi\ (y)$ is a function of y, $e^{-y^2}t$ is also a function of y hence put

$t=y+a\,y^3+\ldots\ldots\ldots$ ...(2)

[Since lower limit is zero, t will have no constant term ; also no even powers of y will be present since we have taken probability from $-y$ to y which is

$$=\int_{-y}^{y}\ldots\ldots\ldots\ldots\quad =2\int_{0}^{y} \text{ or else } 0 \Big].$$

From (1) and (2) we get

$$(1+3ay^2+5by^4+\ldots\ldots)-2y\,(y+ay^3+\ldots\ldots)-1=0$$

Equating coefficients of different powers of y on the two sides. we have

$$3a-2=0,\ 5b-2a=0,\ 7c-2b=0\ldots\ldots\ldots$$

$$\therefore\quad t=y+\frac{2}{3}\,y^3+\frac{2^2}{3.5}\,y^5+\ldots\ldots\ldots$$

$$\therefore\quad \phi\ (y)=\frac{2}{\sqrt{\pi}}\,e^{-y^2}.y\left[1+\frac{1}{3}\,(2y^2)+\frac{1}{3\cdot 5}\,(2y^2)^2+\ldots\ldots\right]$$

CHAPTER XII

CHI-SQUARE (X^2) DISTRIBUTION

Chi-square distribution was discovered by Helmert in 1875, and rediscovered independently by Karl Pearson in 1900 who applied it as a test for 'goodness of fit'.

The Chi-square distribution is used :—

(1) Test of goodness of fit

(2) For sampling purposes

(3) For finding Association and relationship between attributes.

Definition :—If f is the observed frequency and f_t is the theoretical frequency (or expected frequency), then the Chi-square is defined as

$$X^2 = \Sigma \frac{(f - f_t)^2}{f_t}$$

$$\text{or} \quad X^2 = \Sigma \frac{f^2}{f_t} - \Sigma f \qquad (\text{since } \Sigma f = \Sigma f_t)$$

If X^2 is zero all the $(f - f_t)$ are zero, hence the actual frequencies coincides with the expected frequencies. As the value of X^2 increases the correspondence becomes poorer and poorer.

Determination of theoretical frequency in a contingency table :—

The expected frequency in each cell is obtained on the null hypothesis *i.e.*, on the assumption that the two attributes are independent *i.e.*, not associated.

Let the classification of the A be p-fold and that of B's q-fold as shown below :—

	A_1	A_2	....	A_r	....	A_p	Total
B_1	(A_1B_1)						(B_1)
B_2							(B_2)
⋮							⋮
B_s				(A_rB_s)			(B_s)
⋮							
B_q						(A_pB_q)	(B_q)
Total	(A_1)	(A_2)	....	(A_r)		(A_p)	N

The values (A_1), (A_2).........etc. and (B_1), (B_2)...etc. and total frequency N is same as that of observed frequencies.

Now
$$A_rB_s = \frac{(A_r) \times (B_s)}{N}$$

where (A_rB_s) gives the expected frequency corresponding to the cell common to the rth column and sth row. Similarly other expected frequencies can be obtained.

Degrees of freedom.—If the table consists of p columns and q rows then number of cells are pq and degrees of freedom $=(p-1)(q-1)$.

The calculated value of X^2 is compared with the table value at either 5% level of significance or at 1% level of significance for the given degrees of freedom. If the calculated value is greater than the table value our null hypothesis is not correct and if less the hypothesis is correct. If the value is approximately equal the difference may be due to sampling fluctuations.

Note. The theoretical frequencies in case of 2×2 table can be determined very easily as below :

Let the attributes be A and B then table is

	A	α	Total
B	(AB)	$(B\alpha)$	(B)
β	$(A\beta)$	$(\alpha\beta)$	(β)
Total	(A	(α)	N

$$(AB) = \frac{(A) \times (B)}{N}$$

$$(B\alpha) = (B) - (AB)$$

$$(A\beta) = (A) - (AB)$$

$$(\alpha\beta) = (\alpha) - (\alpha B)$$

Thus we have to calculate only (AB) and others can be obtained directly by subtraction method.

Co-efficient of Contingency

The co-efficient of contingency is given by the relation,

$$C = \sqrt{\frac{X^2}{N + X^2}}, \text{ where N is total frequency.}$$

The value of C lies between 0 and 1.

Problem 352.—The table given below shows the data obtained during an epidemic of cholera.

	Attacked	*Not-attacked*	*Total*
Inoculated	31	469	500
Not Inoculated	185	1315	1500
	216	1784	2000

Test the effectiveness of inoculation in preventing the attack of cholera.

[Five percent value of X^2 for one degree of freedom is 3·84].

(*I.A.S*, *1941*)

Solution :

Considering the two attributes as Independent we have the following theoretical frequencies.

Theoretical frequency for Inoculated Attacked

$$= \frac{500}{2000} \times 216 = 54$$

From this the theoretical frequency table is as given below:

	Attacked	*Not-attacked*	*Total*
Inoculated	54	446	500
Not-inoculated	162	1338	1500
	216	1784	2000

Substituting the observed and theoretical frequencies in the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$=\frac{(31-54)^2}{54}+\frac{(469-446)^2}{446}+\frac{(185-162)^2}{162}+\frac{(1315-1338)^2}{1338}$$

$$=\frac{(23)^2}{54}+\frac{(23)^2}{446}+\frac{(23)^2}{162}+\frac{(23)^2}{1338}$$

$$=(23)^2\left[\frac{1}{54}+\frac{1}{446}+\frac{1}{162}+\frac{1}{1338}\right]$$

$$=(529)\ [\cdot0185+\cdot0022+\cdot0062+\cdot0007]$$

$$=529\times\cdot0276$$

$$=14\cdot6$$

The number of degrees of freedom

$$=(2-1)\ (2-1)$$

$$=1$$

But the value of X^2 for 1 degree of freedom at 5 Percent is 3·841, which is less than the calculated value of X^2. This shows that Hypothesis is not correct, and absence of attack from cholera and inoculation are associated.

Problem 353.—The following table is Published in a memoir written by Karl Pearson :

		Eye colour in Sons		
Eye colour		*Not light*	*Light*	*Total*
in	not light	230	148	378
Fathers	light	151	471	622
		381	619	1000

Test whether the colour of Son's eyes is associated with that of the Father's (you may use the fact that 5% value of chi-square for 1 degree of freedom is 3·84.)

(*I.A.S.*, *1949*)

Solution :

Considering that Eye colour of Fathers and Sons are not associated, *i.e.*, they are independent.

Theoretical frequency for not light eye colour of father and not light colour of son.

$$=\frac{378}{1000}\times 381=144$$

The theoretical frequency table will be as below :

		Eye colour in Sons.		
		Not light	*Light*	*Total*
Eye colour in fathers	not light	144	234	378
	light	237	385	622
		381	619	1000

Substituting the observed and the theoretical frequencies in the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$=\frac{(230-144)^2}{144}+\frac{(148-234)^2}{234}+\frac{(151-237)^2}{237}+\frac{(471-385)^2}{385}$$

$$=\frac{(86)^2}{144}+\frac{(86)^2}{234}+\frac{(86)^2}{237}+\frac{(86)^2}{385}$$

$$=(86)^2\left[\frac{1}{144}+\frac{1}{234}+\frac{1}{237}+\frac{1}{385}\right]$$

$$=(86)^2\,[\cdot 0069+\cdot 0042+\cdot 0042+\cdot 0026]$$

$$=7396\times\cdot 0179$$

$$=131\cdot 64$$

The number of degrees of freedom is

$$=(2-1)(2-1)$$

$$=1$$

The value of X^2 at 5% level of significance for 1 degree of freedom is 3·841. The calculated value is much more, which leads us to the conclusion that the hypothesis is wrong, and the colour of Sons eyes is Associated with Father's eyes.

Problem 354.—In an experiment of immunization of cattle from tuberculosis, the following results were obtanied :—

	Died or affected	*Unaffected*	*Total*
Inoculated	12	26	38
Not inoculated	16	6	22
	28	32	60

Examine the effect of vaccine in controlling susceptibility to tuberculosis.

(*I.A.S. 1948*)
IAS. 1963

Considering that the two attributes are not associated, they are Independent. Hence the theoretical frequency for Inoculated and died will be

$$=\frac{38}{60}\times 28=\frac{266}{15}$$
$$=18$$

The theoretical frequency table will be as below :

	Died or affected	*Unaffected*	*Total*
Inoculated	18	20	38
Not inoculated	10	12	22
	28	32	60

Substituting the observed and the theoretical frequencies in the Formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$
$$=\frac{(12-18)^2}{18}+\frac{(26-20)^2}{20}+\frac{(16-10)^2}{10}+\frac{(6-12)^2}{12}$$
$$=\frac{(6)^2}{18}+\frac{(6)^2}{20}+\frac{(6)^2}{10}+\frac{(6)^2}{12}$$
$$=(6)^2\left[\frac{1}{18}+\frac{1}{20}+\frac{1}{10}+\frac{1}{12}\right]$$
$$X^2=36[\cdot 055+\cdot 05+\cdot 01+\cdot 083]$$
$$=36\,[0\cdot 198]$$
$$=7\cdot 128$$

Degrees of freedom is
$$=(2-1)(2-1)$$
$$=1.$$

The value of X^2 for 1 degree of freedom is 3·841. The calculated value is greater than this. Hence the Hypothesis is wrong. Thus it is established that the vaccine is effective in controlling susceptibility to tuberculosis.

Problem 355.—Genetic theory States that childern having one Parent of blood type M and the other of blood type N will always be one of the three types M, MN, N, and that the proportions of these types will be on average as 1 : 2 : 1. A report states that out of 300 childern having one M Parent and one N Parent, 30% were found to be type M, 45% type MN and remainder type N. Test the hypothesis by X^2 test.

Solution :

The observed frequencies of type M are 30% and of type MN 45% and the remainder is of type N. The observed frequencies are

Children of type M	90
Children of type MN	135
Children of type N	75

By the Genetic theory they are in the ratio 1 : 2 : 1 Therefore

$$\frac{M}{1}=\frac{MN}{2}=\frac{N}{1}=\frac{M+MN+N}{4}$$

$$\frac{M}{1}=\frac{MN}{2}=\frac{N}{1}=\frac{300}{4}$$

Children of type M	=75
Children of type MN	=150
Children of type N	=75

Substituting the observed and the theoretical frequencies in the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$X^2=\frac{(90-75)^2}{75}+\frac{(135-150)^2}{150}+\frac{(75-75)^2}{75}$$

$$=\frac{(15)^2}{75}+\frac{(15)^2}{150}+\frac{0}{75}$$

$$=\frac{225}{75}+\frac{225}{150}+\frac{0}{75}$$

$$=3+1{\cdot}5+0$$

$$=4{\cdot}5$$

The number of degrees of freedom is 3−1 or 2. The value of X^2 for 2 degrees of freedom at 5% level is 5·991. The calculated value is less than this figure. The hypothesis holds ground so far as the X^2 test is concerned.

√ **Problem 356.**—In an experiment on the immunization of goats from anthrax the following results were obtained. Derive your inference on the efficacy of the vaccine.

	Died of anthrax	*Survived*	*Total*
Inoculated with vaccine	2	10	12
Not inoculated	6	6	12
	8	16	24

(*I.A.S., 1943*)
IAS 1963

Solution :

Considering the two attributes as independent we have the following theoretical frequencies.

Theoretical frequency for Inoculated with vaccine and died of anthrax $= \frac{12}{24} \times 8$

$= 4.$

From this the theoretical frequency table is as given below.

	Died of anthrax	*Survived*	*Total*
Inoculated with vaccine	4	8	12
Not inoculated	4	8	12
	8	16	24

Substituting the observed and theoretical frequencies in the formula

$$X^2 = \Sigma \frac{(f - f_t)^2}{f_t}$$

$$= \frac{(2-4)^2}{4} + \frac{(10-8)^2}{8} + \frac{(6-4)^2}{4} + \frac{(6-8)^2}{8}$$

$$= 1 + \tfrac{1}{2} + 1 + \tfrac{1}{2}$$

$$= 3.$$

The number of degrees of freedom

$= (2-1)(2-1)$

$= 1$

But the value of X^2 for 1 degree of freedom at 5 percent is 3·841 which is greater than calculated value for X^2 hence hypothesis is correct that is survival is not associated with inoculation of vaccine.

Problem 357.—The following table gives the results of a series of controlled experiments. Discuss whether the treatment may be considered to have any positive effect.

	Positive	*No effect*	*Negative*	*Total*
Treatment	9	2	1	12
Control	3	6	3	12
	12	8	4	24

(*I.A.S., 1944*)

Solution :

Let the two attributes experiment and effect are independent. Theoretical frequency for treatment and positive effect

$= \frac{12}{24} \times 12 = 6.$

Theoretical frequency for treatment and no effect is

$= \frac{12}{24} \times 8 = 4.$

From this the theoretical frequency table is as given below :

	Positive	*No effect*	*Negative*	*Total*
Treatment	6	4	2	12
Control	6	4	2	12
	12	8	4	24

Substituting the observed and theoretical frequencies in the formula

$$X^2 = \Sigma \frac{(f - f_t)^2}{f_t}$$

$$= \frac{(9-6)^2}{6} + \frac{(2-4)^2}{4} + \frac{(1-2)^2}{2} + \frac{(3-6)^2}{6} + \frac{(6-4)^2}{4} + \frac{(3-2)^2}{2}$$

$$= \tfrac{3}{2} + 1 + \tfrac{1}{2} + \tfrac{3}{2} + 1 + \tfrac{1}{2}$$

$$= 6.$$

Degree of freedom $= (3-1)(2-1)$

$= 2.$

But the value of X^2 for 2 degrees of freedom at 5% level of significant is 5·99 which is less than calculated value of X^2. Therefore the hypothesis is not correct that is the experiments and effect are associated or the treatment has positive effect.

Problem 358.—From the following table showing the number of plants having certain characters, test the hypothesis that the flower colour is independent of flatness of leaf.

	Flat Leaves	*Curled leaves*	*Total*
White flowers	99	36	135
Red flowers	20	5	25
	119	41	160

You may use the following table giving the value of X^2 for one degree of freedom, for different values of P.

P	·99	·95	·90	·50	·10	·05	·01
X^2	·000157	·00393	·0158	·455	2·706	3·841	6·635

(*I.A.S., 1946*)
(*Punjab Certificate in Statistics 1946*)
(*M. Sc. Agra, 1957*)

Solution :

Considering that the flower colour and flatness of leaf are not associated.

The theoretical frequency for white flowers and flat leaves

$$= \frac{135}{160} \times 119$$

$$= 100{\cdot}4$$

From this the theoretical frequency table is as given below :

	Flat leaves	*Curled leaves*	*Total*
White flowers	100·4	34·6	135
Red flowers	18·6	6·4	25
	119	41	160

Substituting the observed and theoretical frequencies in the formula

$$X^2 = \Sigma \frac{(f - f_t)^2}{f_t}$$

$$= \frac{(99-100{\cdot}4)^2}{100{\cdot}4} + \frac{(36-34{\cdot}6)^2}{34{\cdot}6} + \frac{(20-18{\cdot}6)^2}{18{\cdot}6} + \frac{(5-6{\cdot}4)^2}{6{\cdot}4}$$

$$= \frac{(1{\cdot}4)^2}{100{\cdot}4} + \frac{(1{\cdot}4)^2}{34{\cdot}6} + \frac{(1{\cdot}4)^2}{18{\cdot}6} + \frac{(1{\cdot}4)^2}{6{\cdot}4}$$

$$= \frac{1{\cdot}96}{100{\cdot}4} + \frac{1{\cdot}96}{34{\cdot}6} + \frac{1{\cdot}96}{18{\cdot}6} + \frac{1{\cdot}96}{6{\cdot}4}$$

$$= {\cdot}0195 + {\cdot}0566 + {\cdot}1053 + {\cdot}3062$$

$$= {\cdot}4876$$

The number of degrees of freedom

$$= (2-1)(2-1)$$

$$= 1.$$

The value of X^2 for 1 degree of freedom at 5% level of significance is 3·841 which is greater than calculated value of X^2 hence hypothesis is correct and the flower colour is independent of the flatness of the leaf.

Problem 359. The normal rate of infection for a certain disease in cattle is known to be 50%. In an experiment with seven animals injected with a new vaccine it was found that none of the animals caught infection. Can the evidence be regarded as conclusive (at the 1% level of significance) to prove the value of the new vaccine. Use the tabulated value of X^2 (for $g=1$) $=6{\cdot}64$ (at 1% level of significance).

(I.A.S., 1942)

Solution :

Since the normal rate of infection is known to be 50% and 7 animals are injected hence theoretical frequencies and observed frequencies are as below.

	Infection	*Not infection*
f	**0**	**7**
f_t	7/2	7/2

Using these datas we have X^2 by the formula

$$X^2 = \Sigma \frac{(f - f_t)^2}{f_t}$$

$$=\frac{\left(0-\frac{7}{2}\right)^2}{\frac{7}{2}}+\frac{\left(7-\frac{7}{2}\right)^2}{\frac{7}{2}}$$

$$=\frac{7}{2}+\frac{7}{2}$$

$$=7$$

Number of degrees of freedom

$$=(2-1)$$

$$=1$$

But the value of X^2 for one degree of freedom at 1% level of significance is 6·64 which is less than calculated value. Hence the evidence may be regarded as conclusive.

Problem 360.—The normal rate of infection in a certain disease in animals is known to be 40 percent. In an experiment with 6 animals injected with a new vaccine it was observed that none of the animals caught infection. Calculate the probability of the observed result and hence or otherwise explain whether the evidence can be regarded as exclusive 1 percent at the level of significance to prove the value of the vaccine.

(I.A.S. 1949 Part II)

Solution :

Since the normal rate of infection in animals is known to be 40 percent and total animals experimented are 6 in which none of the animals caught disease, hence the frequency distributions are as below.

	Infection	*Not infection*
f	0	6
f_t	$6\times\frac{40}{100}=2{\cdot}4$	3·6

Using these datas we have X^2 by the formula

$$X^2=\Sigma\frac{(f-f_t)^3}{f_t}$$

$$=\frac{(0-2{\cdot}4)^2}{2{\cdot}4}+\frac{(6-3{\cdot}6)^2}{3{\cdot}6}$$

$$=2{\cdot}4+\frac{(2{\cdot}4)^2}{3{\cdot}6}$$

$$=2{\cdot}4+1{\cdot}6$$

$$=4$$

Now for one degree of freedom the value of probability for $X^2=4$ is from the standard tables $=\cdot04550$ which is less than $P=\cdot05$ hence the hypothesis is not correct that is infection is effected by vaccination in animals which proves the value of vaccine.

Problem 361.—The following data are observed for hydrides of Datura :—

Flowers	*violet,*	*fruits*	*prickly*	...	47
,,	,,	,,	*smooth*	...	12
,,	*white*	,,	*prickly*	...	21
,,	,,	,,	*smooth*	...	3

using X^2 test, find the association between colour of flowers and character of fruit ; given that

$v=1$ | $P=\cdot 402$ for $X^2=\cdot 7$
$P=\cdot 399$ for $X^2=\cdot 71$

(*M. Sc. Agra 1956*)

Solution :

The observed results may the tabulated as below :

	Fruits prickly	*Fruits smooth*	*Total*
Flowers violet	47	12	59
Flowers white	21	3	24
	68	15	83

Considering that there is no association between the colour of flowers and character of fruits.

Theoretical frequency for violet flowers and prickly fruit is

$$=\frac{59}{83}\times 68=48\cdot 34$$

From this the theoretical frequency table is as given below :—

	Fruits prickly	*Fruits smooth*	*Total*
Flowers violet	48·34	10·66	59
Flowers white	19·66	4·34	24
	68	15	83

Substituting the observed and theoretical frequencies in the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$X^2=\frac{(47-48\cdot 34)^2}{48\cdot 34}+\frac{(21-19\cdot 66)^2}{19\cdot 66}+\frac{(12-10\cdot 66)^2}{10\cdot 66}+\frac{(3-4\cdot 34)^2}{4\cdot 34}$$

$$=\frac{(1\cdot 34)^2}{48\cdot 34}+\frac{(1\cdot 34)^2}{19\cdot 66}+\frac{(1\cdot 34)^2}{10\cdot 66}+\frac{(1\cdot 34)^2}{4\cdot 34}$$

$$=\frac{1\cdot 7956}{48\cdot 34}+\frac{1\cdot 7956}{19\cdot 66}+\frac{1\cdot 7956}{10\cdot 66}+\frac{1\cdot 7956}{4\cdot 34}$$

$$=\cdot 0371+\cdot 0913+\cdot 1684+\cdot 4137$$

$$=\cdot 7105$$

Degrees of freedon$=(2-1)(2-1)=1$

The value of X^2 for 1 degree of freedom at 40·2% level of significance is ·7 and at 39·9% level of significance is ·71. Now the value which is calculated is slightly greater. Hence this is due to sampling fluctuation.

Problem 362.—The following table shows the result of inoculation against cholera :

	Not-attacked	*Attacked*
Inoculated	431	5
Not-inoculated	291	9

Is there any significant association between inoculation and attack ? Given that

$$\nu=1 \begin{cases} P=\cdot074 \text{ for } X^2=3\cdot2 \\ P=\cdot069 \text{ for } X^2=3\cdot3 \end{cases}$$

(*M.Sc. Agra, 1958*)

Solution :

The observed frequencies may be tabulated as below :

	Not-attacked	*Attacked*	*Total*
Inoculated	431	5	436
Not-inoculated	291	9	300
	722	14	736

Considering that there is no association between the inoculation and attack we have the theoretical frequency for Inoculated and Not attacked $=\frac{436}{736}\times 722=427\cdot7$

From this the theoretical frequencies are obtained as below :

	Not attacked	*Attacked*	*Total*
Inoculated	427·7	8·3	436
Not-inoculated	294·3	5·7	300
	722	14	736

The value of X^2 is obtained from the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$\therefore X^2=\frac{(431-427\cdot7)^2}{427\cdot7}+\frac{(5-8\cdot3)^2}{8\cdot3}+\frac{(291-294\cdot3)^2}{294\cdot3}+\frac{(9-5\cdot7)^2}{5\cdot7}$$

$$=\frac{(3\cdot3)^2}{427\cdot7}+\frac{(3\cdot3)^2}{8\cdot3}+\frac{(3\cdot3)^2}{294\cdot3}+\frac{(3\cdot3)^2}{5\cdot7}$$

$$= \frac{10\cdot89}{427\cdot7} + \frac{10\cdot89}{8\cdot3} + \frac{10\cdot89}{294\cdot3} + \frac{10\cdot89}{5\cdot7}$$
$$= \cdot025 + 1\cdot312 + \cdot037 + 1\cdot905$$
$$= 3\cdot279$$

No. of degrees of freedom $=(2-1)\ (2-1)=1$

Since $X^2=3\cdot2$ corresponds to $P=\cdot074$

and $X^2=3\cdot3$ corresponds to $P=\cdot069$

Hence $X^2=3\cdot279$ corresponds to $P=\cdot0706$ appr.

The X^2 test shows that the hypothesis is incorrect and hence there is association between inoculation and attack.

Problem 363. In experiment on pea-breeding, Mendel obtained the following frequencies of seeds ; 315 round and yellow 101 wrinkled and yellow ; 108 round and green ; 32 wrinkled and green, Total, 556.

Theory predicts that the frequencies should be in the proportions 9 : 3 : 3 : 1.

Find the X^2 and examine the correspondence between theory and experiment.

For $r=3$ [$X^2=\cdot35$ for $P=\cdot95$, $X^2=\cdot115$ for $P=\cdot99$; and $X^2=\cdot6$ for $P=\cdot9$].

(Agra M.Sc., 1952)

Solution :

The observed frequencies are

A	Round and yellow	315
B	Wrinkled and yellow	101
C	Round and green	108
D	Wrinkled and green	32

The theoretical frequencies are obtained as below

$$\frac{A}{9}=\frac{B}{3}=\frac{C}{3}=\frac{D}{1}=\frac{A+B+C+D}{16}=\frac{556}{16}$$

∴ A Round and yellow = 312·75

B Wrinkled and yellow = 104·25

C Round and green = 104·25

D Wrinkled and green = 34·75

Now $$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

or $$X^2=\frac{(315-312\cdot75)^2}{312\cdot75}+\frac{(101-104\cdot25)^2}{104\cdot25}+\frac{(108-104\cdot25)^2}{104\cdot25}+\frac{(32-34\cdot75)^2}{34\cdot75}$$

$$= \frac{(2 \cdot 25)^2}{312 \cdot 75} + \frac{(3 \cdot 25)^2}{104 \cdot 25} + \frac{(3 \cdot 75)^2}{104 \cdot 25} + \frac{(2 \cdot 75)^2}{34 \cdot 75}$$

$$= \frac{5 \cdot 0625}{312 \cdot 75} + \frac{10 \cdot 5625}{104 \cdot 25} + \frac{14 \cdot 0625}{104 \cdot 25} + \frac{7 \cdot 5625}{34 \cdot 75}$$

$$= \cdot 0161 + \cdot 1013 + \cdot 1348 + \cdot 2176$$

$$= \cdot 4698 = \cdot 47 \text{ appr.}$$

The number of degrees of freedom$=4-1=3$

The corresponding value of P is ·943, hence there is very high degree of correspondence between theory and practice.

Problem 364.—Suppose that, in a public opinion survey answers to the questions :—

(*a*) Do you drink ?

(*b*) Are you in favour of local option on sale of liquor ?

were as tabulated below :

Question (b)	Question (a) *Yes*	*No*	*Total*
Yes	56	31	87
No	18	6	24
Total	74	37	111

Can you infer that opinion on local option is dependent on whether or not an individual drinks ?

Values of X^2 on levels of significance P

Degrees of freedom	P= ·05	P=·10
1	3·84	2·71
2	5·99	4·60
3	7·82	6·25

(*P. C. S. 1953*)

Solution :

Let there be no association between local option and an individual drink.

The theoretical frequency corresponding to observe frequency 56 is$=\frac{87}{111} \times 74 = 58$ (appr.)

Hence the theoretical frequencies may be tabulated as below :

Question (b)	Question (a) *Yes*	*No*	*Total*
Yes	58	29	87
No	16	8	24
Total	74	37	111

From this the value of X^2 is obtained by the formula

$$X^2 = \Sigma \frac{(f-f_t)^2}{f_t}$$

$$\therefore \quad X^2 = \frac{(56-58)^2}{58} + \frac{(31-29)^2}{30} + \frac{(18-16)^2}{16} + \frac{(6-8)^2}{8}$$

$$= \frac{4}{58} + \frac{4}{30} + \frac{4}{16} + \frac{4}{8}$$

$$= \cdot 0689 + \cdot 1333 + \cdot 25 + \cdot 5$$

$$= \cdot 9522.$$

Degrees of freedom $= (2-1)(2-1) = 1$

Value of X^2 for 1 degree of freedom at 5% level of significance is 3·84 which is greater than calculated value hence hypothesis is correct that is the local option is not dependent on individual drinks.

Problem 365.—What is the X^2 test of goodness of fit ? What cautions are necessary in using this test ?

Find the value of Chi-square for the following table :—

Class	A	B	C	D	E
Observed frequency	8	29	44	15	4
Theoretical frequency	7	24	38	24	7

(M.Sc. Agra 1951, 1955)

Solution :

Since some of the frequencies are less than 10, we regroup the data as in the following table.

Class	A and B	C	D and E
Observed frequency	37	44	19
Theoretical frequency	31	38	31

Using the formula

$$X^2 = \Sigma \frac{(f-f_t)^2}{f_t}$$

Or

$$X^2 = \frac{(37-31)^2}{31} + \frac{(44-38)^2}{38} + \frac{(19-31)^2}{31}$$

$$=\frac{6^2}{31}+\frac{6^2}{38}+\frac{12^2}{31}$$

$$=1{\cdot}16+{\cdot}95+4{\cdot}64$$

$$=6{\cdot}75 \text{ appr.}$$

Problem 366.—A painstaking experimeter rolled 12 dice 26,306 times, observing at each throw the number of dice recording a 5 or a 6. Here is a table of his results :

No. of successes	0	1	2	3	4	5	6	7	8	9	10 and over
Observed frequency	185	1149	3265	5475	6114	5194	3067	1331	403	105	18

obtain the corresponding theoretical frequencies and hence calculate X^2.

(M.Sc. Agra, 1944; 1950)
(P.C.S. 1954)

Solution :

If the dice are unbiassed, the chance of getting a 5 or a 6 with one die is $\frac{1}{3}$. Hence the theoretical frequencies of 0, 1, 2, 3,......etc. successes in throwing 12 dice 26,306 times are obtained by the successive terms of the binomial expansion $26306\left(\frac{2}{3}+\frac{1}{3}\right)^{12}$ which are tabulated as below :

No. of Successes	0	1	2	3	4	5	6	7	8	9	10 and over
Theoretical frequencies	203	1217	3345	5576	6273	5018	2927	1254	392	87	14

From these values X^2 is obtained by the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$=\frac{(185-203)^2}{203}+\frac{(1149-1217)^2}{1217}+\frac{(3265-3345)^2}{3345}$$

$$+\frac{(5475-5576)^2}{5576}+\frac{(6114-6273)^2}{6273}+\frac{(5194-5018)^2}{5018}$$

$$+\frac{(3067-2927)^2}{2927}+\frac{(1331-1254)^2}{1254}+\frac{(403-392)^2}{392}$$

$$+\frac{(105-87)^2}{87}+\frac{(18-14)^2}{14}$$

$$=\frac{(18)^2}{203}+\frac{(68)^2}{1217}+\frac{(80)^2}{3345}+\frac{(101)^2}{5576}+\frac{(159)^2}{6273}+\frac{(176)^2}{5018}$$

$$+\frac{(140)^2}{2927}+\frac{(77)^2}{2254}+\frac{(11)^2}{392}+\frac{(18)^2}{87}+\frac{(4)^2}{14}$$

$=1{\cdot}596+3{\cdot}800+1{\cdot}913+1{\cdot}829+4{\cdot}030+6{\cdot}173+6{\cdot}696+4{\cdot}728$
$+{\cdot}309+3{\cdot}724+1{\cdot}143$

$=35{\cdot}941$

∴ The value of $X^2=35{\cdot}941$.

Problem 367.—Five coins are tossed 3200 times and the following results are obtained.

No. of heads	0	1	2	3	4	5
Frequency	80	570	1100	900	500	50

Test the hypothesis that the coins are unbiassed. You can make the use of the following data in drawing your conclusion.

Degrees of freedom	1	2	3	4	5	6
X^2 value at 5% level of significance	3·841	5·991	7·815	9·488	11·070	12·592

(*Alld. M. Com., 1952*)

Solution :

If the coins are unbiassed, the chance of getting a head from one coin is $\frac{1}{2}$. Hence the theoretical frequencies of 0, 1, 2,......etc. heads in throwing 5 coins 3200 times are obtained by the successive terms of the binomial expansion $3200\,(\frac{1}{2}+\frac{1}{2})^5$ which are tabulated as below :—

No. of heads	0	1	2	3	4	5
Expected frequency	100	500	1000	1000	500	100

From these values X^2 is obtained by the formula

$$X^2=\Sigma\,\frac{(f-f_t)^2}{f_t}$$

$$\text{or } X^2=\frac{(80-100)^2}{100}+\frac{(570-500)^2}{500}+\frac{(1100-1000)^2}{1000}+\frac{(900-1000)^2}{1000}+\frac{(500-500)^2}{500}+\frac{(50-100)^2}{100}$$

$$=4+9{\cdot}8+10+10+0+25$$

$$=58{\cdot}8.$$

Number of degrees of freedom $=(6-1)=5$.

The value of X^2 for 5 degrees of freedom at 5% level of significant is 11·070 which is very less than calculated value hence our hypothesis is not correct *i.e.,* the coins are not unbiassed but biassed.

Problem 368.—The following table gives the results of a dice throwing experiment :—

12 Dice thrown 4096 times, a throw of 6 reckoned a success.

No. of successes	0	1	2	3	4	5	6	7 and over
Frequency	447	1145	1181	796	380	115	24	8

Find X^2 on the hypothesis that the dice were unbiassed and hence show that the data are consistent with the hypothesis so far as the X^2 test is concerned.

(*Theory of Statistics, Yule and Kendall*)

Solution :

On the hypothesis that the dice were unbiassed, the chance of getting a 6 with one die is $\frac{1}{6}$. Hence the expected frequencies of 0, 1, 2,.........successes in throwing 12 dice 4096 times are obtained by the successive terms of the binomial expansion $4096(\frac{5}{6}+\frac{1}{6})^{12}$, which are tabulated as below.

No. of successes	0	1	2	3	4	5	6	7 and over
Expected frequency	359	1102	1212	808	365	116	27	8

Since no frequency should be less than 10 for the application of X^2 test, we get on regrouping the datas :—

Observed frequency	447	1145	1181	796	380	115	32
Theoretical frequency	459	1102	1212	808	364	116	35

From these value X^2 is obtained by the formula

$$X^2 = \Sigma \frac{(f-f_t)^2}{f_t}$$

$$\text{or } X^2 = \frac{(447-459)^2}{459} + \frac{(1145-1102)^2}{1102} + \frac{(1181-1212)^2}{1212} + \frac{(796-808)^2}{808}$$

$$+ \frac{(380-364)^2}{364} + \frac{(115-116)^2}{116} + \frac{(32-35)^2}{35}$$

$$= \frac{12^2}{459} + \frac{43^2}{1102} + \frac{31^2}{1212} + \frac{12^2}{808} + \frac{16^2}{364} + \frac{1}{116} + \frac{3^2}{35}$$

$$= \cdot31 + 1\cdot68 + \cdot79 + \cdot18 + \cdot71 + \cdot01 + \cdot26$$

$$= 3\cdot94.$$

No. of degrees of freedom $= 8-1=7$

The value of X^2 for 7 degrees of freedom at 5% level of significance is 14·067 hence the calculated value is very less than this. From this we come to the conclusion that our hypothesis is correct and the datas are consistent with the hypothesis.

Problem 369.—Five dice were thrown 192 times, and the number of times 4, 5 or 6 was thrown were as follows :—

No. of dice showing 4, 5 or 6	5	4	3	2	1	0
Frequency	6	46	70	48	20	2

Calculate the value of X^2.

(*Agra M.Sc., 1945*)

Solution :

On the hypothesis that the dice are unbiassed, the chance of getting 4, 5 or 6 with one die is $\frac{3}{6}$ or $\frac{1}{2}$. Hence the expected frequencies of getting 5, 4, 3,.........successes with 5 dice in throwing 192 times are obtained by successive terms of binomial expansion 192 $(\frac{1}{2}+\frac{1}{2})^5$, which are tabulated as below.

No, of dice showing 4, 5 or 6	5	4	3	2	1	0
Frequency	6	30	60	60	30	6

Since for the application of x^2 test no frequency should be less than 10 hence on regrouping the frequencies we have

Observed frequency	52	70	48	22
Theoretical frequency	36	60	60	36

From this the value of X^2 is obtained from the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$\text{or } X^2=\frac{(52-36)^2}{36}+\frac{(70-60)^2}{60}+\frac{(48-60)^2}{60}+\frac{(22-36)^2}{36}$$

$$=\frac{256}{36}+\frac{100}{60}+\frac{144}{60}+\frac{196}{36}$$

$$=7{\cdot}11+1{\cdot}66+2{\cdot}4+5{\cdot}44$$

$$=16{\cdot}61.$$

Problem 370.—Five dice were thrown 96 times, and the number of times 4, 5 or 6 were thrown was :—

No. of dice showing 4, 5 or 6	5	4	3	2	1	0
Frequency	7	19	35	24	8	3

Calculate the value of X^2 upto 2 places of decimals.

(*M. Sc. Agra, 1946*)

Solution :

Considering that the dice are unbiassed, the chance of getting 4, 5 or 6 with one dic is $\frac{3}{6}$ or $\frac{1}{2}$. Hence the expected frequencies of getting 5, 4, 3,...successes with 5 dice in throwing 96 times are obtained by successive terms of the binomial expansion 96 $(\frac{1}{2}+\frac{1}{2})^5$, which are tabulated as below.

No. of dice showing 4, 5 or 6.	5	4	3	2	1	0
Frequency	3	15	30	30	15	3

Since no frequency should be less than 10 in the application of X^2 test, hence on regrouping the frequencies we have

Observed frequency	26	35	24	11
Theoretical frequency	18	30	30	18

From this the value of X^2 is obtained by the formula

$$X^2=\Sigma\frac{(f-f_t)^2}{f_t}$$

$$\text{or } X^2 = \frac{(26-18)^2}{18} + \frac{(35-30)^2}{30} + \frac{(24-30)^2}{30} + \frac{(11-18)^2}{18}$$

$$= \frac{64}{18} + \frac{25}{30} + \frac{36}{30} + \frac{49}{18}$$

$$= 3{\cdot}55 + {\cdot}83 + 1{\cdot}2 + 2{\cdot}72$$

$$= 8{\cdot}3.$$

Problem 371.—The number of males in each of 106 eight-pig litres was found, and they are given by the following frequency distribution :—

Number of males per litre	0	1	2	3	4	5	6	7	8	Total
Frequency	0	5	9	22	25	26	14	4	1	106

Assuming that the probability of an animal being male or female is even (*i.e.* $p=q=\frac{1}{2}$), and the frequency distribution follows the binomial law, calculate the expected frequencies of the nine classes. Find also the value of X^2 to test the goodness of fit.

(*Punjab M.A.* (*Maths.*), *1946.*)

Solution :

Taking the hypothesis as correct that is the probability of an animal being male or female is even, the chance of getting a male animal is $\frac{1}{2}$. Hence the frequencies of getting 8 males, 7 males,...... etc. in 106 eight-pig litres is obtained by the successive terms of binomial expansion $106\left(\frac{1}{2}+\frac{1}{2}\right)^8$ which are tabulated as below :—

No. of males per litre	0	1	2	3	4	5	6	7	8	Total
Frequency	·4	3·3	11·6	23·2	29	23·2	11·6	3·3	·4	106

Since no frequency should be less than 10 in calculating X^2 hence regrouping these frequencies we have

Observed frequency	14	22	25	26	19
Theoretical frequency	15·3	23 2	29	23·2	15·3

From these X^2 is obtained by the formula

$$X^2 = \Sigma \frac{(f-f_t)^2}{f_t}$$

$$\text{r } X^2 = \frac{(14-15{\cdot}3)^2}{15{\cdot}3} + \frac{(22-23{\cdot}2)^2}{23{\cdot}2} + \frac{(25-29)^2}{29} + \frac{(26-23{\cdot}2)^2}{23{\cdot}2} + \frac{(19-15{\cdot}3)^2}{15{\cdot}3}$$

$$= \frac{(1{\cdot}3)^2}{15{\cdot}3} + \frac{(1{\cdot}2)^2}{23{\cdot}2} + \frac{(4)^2}{29} + \frac{(2{\cdot}8)^2}{23{\cdot}2} + \frac{(3{\cdot}7)^2}{15{\cdot}3}$$

$$= \frac{1{\cdot}69}{15{\cdot}3} + \frac{1{\cdot}44}{23{\cdot}2} + \frac{16}{29} + \frac{7{\cdot}84}{23{\cdot}2} + \frac{13{\cdot}69}{15{\cdot}3}$$

$$= {\cdot}11 + {\cdot}062 + {\cdot}551 + {\cdot}338 + {\cdot}895$$

$$= 1{\cdot}956$$

No. of degrees of freedom = (5 − 1) = 4.

The value of X^2 for 4 degrees of freedom at 5% level of significance is 9·488 which is very large than calculated value hence our hypothesis is correct or the fit is good.

Problem 372 —200 digits were chosen at random from a set of tables. The frequencies of the digits were :—

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	18	19	23	21	16	25	22	20	21	15

Find the value of X^2 and give the significance of this value.

(*M. Sc. Agra, 1947*)

Solution :

If we take the hypothesis that the digits were distributed in equal number in the tables, the expected frequencies of the digits would be :

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	20	20	20	20	20	20	20	20	20	20

From these values of observed and expected frequencies we have X^2 from the formula

$$X^2 = \Sigma \frac{(f-f_t)^2}{f_t}$$

$$\text{or } X^2 = \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(15-20)^2}{20}$$

$$= \frac{4}{20} + \frac{1}{20} + \frac{9}{20} + \frac{1}{20} + \frac{16}{20} + \frac{25}{20} + \frac{4}{20} + \frac{0}{20} + \frac{1}{20} + \frac{25}{20}$$

$$= \frac{4+1+9+1+16+25+4+0+1+25}{20}$$

$$= \frac{86}{20}$$

$$= 4{\cdot}3.$$

The number of degrees of freedom = (10 − 1) = 9.

The value of X^2 for 9 degrees of freedom at 5% level of significance is 16·919 which is very large in comparison to calculated value. Hence our hypothesis is correct.

Problem 373.—Show that in a 2×2 contingency table wherein the frequencies are $\frac{a|b}{c|d}$, X^2 calculated from the 'independence' frequencies is

$$\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)}.$$

(M.Sc. Agra, 1948, 1954)

Solution :

Let the frequencies be tabulated as below

	B	β	
A	a	b	$a+b$
α	c	d	$c+d$
	$a+c$	$b+d$	$a+b+c+d$

Expected frequency for AB is $\frac{(a+b)(a+c)}{(a+b+c+d)}$

" " Aβ is $\frac{(a+b)(b+d)}{(a+b+c+d)}$

" " αB is $\frac{(c+d)(a+c)}{(a+b+c+d)}$

" " αβ is $\frac{(c+d)(b+d)}{(a+b+c+d)}$

Now the value of X^2 is given by

$$X^2 = \Sigma \frac{(f-f_t)^2}{f_t}$$

$$\text{or } X^2 = \frac{\left[a - \frac{(a+b)(a+c)}{(a+b+c+d)}\right]^2}{\frac{(a+b)(a+c)}{(a+b+c+d)}} + \frac{\left[b - \frac{(a+b)(b+d)}{(a+b+c+d)}\right]^2}{\frac{(a+b)(b+d)}{(a+b+c+d)}}$$

$$+ \frac{\left[c - \frac{(c+d)(a+c)}{(a+b+c+d)}\right]^2}{\frac{(c+d)(a+c)}{(a+b+c+d)}} + \frac{\left[d - \frac{(c+d)(b+d)}{(a+b+c+d)}\right]^2}{\frac{(c+d)(b+d)}{(a+b+c+d)}}$$

$$= \frac{[ad-bc]^2}{(a+b+c+d)(a+b)(a+c)} + \frac{[bc-ad]^2}{(a+b+c+d)(b+d)(a+b)}$$

$$+ \frac{[bc-ad]^2}{(a+b+c+d)(c+d)(a+c)} + \frac{[ad-bc]^2}{(a+b+c+d)(c+d)(b+d)}$$

$$=\frac{(ad-bc)^2}{(a+b+c+d)}\left[\frac{1}{(a+b)(a+c)}+\frac{1}{(b+d)(a+b)}+\frac{1}{(c+d)(a+c)}+\frac{1}{(c+d)(b+d)}\right]$$

$$=\frac{(ad-bc)^2}{(a+b+c+d)}\left[\frac{(b+d)(c+d)+(a+c)(c+d)+(a+b)(b+d)+(a+b)(a+c)}{(a+b)(a+c)(b+d)(c+d)}\right]$$

$$=\frac{(ad-bc)^2}{(a+b+c+d)}\left[\frac{(c+d)[b+d+a+c]+(a+b)[a+c+b+d]}{(a+b)(a+c)(b+d)(c+d)}\right]$$

$$=\frac{(ad-bc)^2}{(a+b+c+d)}\left[\frac{(a+b+c+d)(a+b+c+d)}{(a+b)(a+c)(b+d)(c+d)}\right]$$

$$=\frac{(a+b+c+d)(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

Problem 374.—Prove that for $2\times n$ table,

$$X^2=\Sigma\frac{\left[N_1N_2\left(\frac{\mu_{1r}}{N_1}-\frac{\mu_{2r}}{N_2}\right)^2\right]}{\mu_{1r}+\mu_{2r}}$$

where μ_{1r}, μ_{2r} are the two frequencies in the r^{th} column and N_1, N_2 are the marginal sums of the two rows.

(M. Sc. Agra, 1949, 1953)

Solution :

Let the frequencies be tabulated as below

	A_1	A_2	......	A_r	A_{r+1}	......	A_n	Total
B_1	$\mu_{1,1}$	$\mu_{1,2}$		$\mu_{1,r}$	$\mu_{1,r+1}$		$\mu_{1,n}$	N_1
B_2	$\mu_{2,1}$	$\mu_{2,2}$		$\mu_{2,r}$	$\mu_{2,r+1}$		$\mu_{2,n}$	N_2
Total	$\mu_{1,1}+\mu_{2,1}$	$\mu_{1,2}+\mu_{2,2}$		$\mu_{1,r}+\mu_{2,r}$	$\mu_{1,r+1}+\mu_{2,r+1}$		$\mu_{1,n}+\mu_{2,n}$	N_1+N_2

Expected frequency for $A_1\ B_1$ is $=\frac{N_1}{N_1+N_2}(\mu_{1,1}+\mu_{2,1})$

" " " $A_1\ B_2$ is $=\frac{N_2}{N_1+N_2}(\mu_{1,1}+\mu_{2,1})$

... ...

... ...

$$\text{Expected frequency for } A_r\,B_1 = \frac{1}{N_1+N_2}(\mu_{1,r}+\mu_{2,r})$$

$$\text{,, \quad ,, \quad ,, } A_r\,B_2 = \frac{N_2}{N_1+N_2}(\mu_{1,r}+\mu_{2,r})$$

... ...

... ...

Now $$X^2 = \Sigma\frac{(f-f_t)^2}{f_t}$$

or
$$X^2 = \left\{ \frac{\left[\mu_{1,1} - \frac{N_1}{N_1+N_2}(\mu_{1,1}+\mu_{2,1})\right]^2}{\frac{N_1(\mu_{1,1}+\mu_{2,1})}{N_1+N_2}} + \frac{\left[\mu_{2,1} - \frac{N_2}{N_1+N_2}(\mu_{1,1}+\mu_{2,1})\right]^2}{\frac{N_2}{N_1+N_2}(\mu_{1,1}+\mu_{2,1})} \right\}$$

$$+ \ldots \ldots \ldots \ldots \ldots + \ldots \ldots \ldots \ldots$$

$$+ \left\{ \frac{\left[\mu_{1,r} - \frac{N_1}{N_1+N_2}(\mu_{1,1}+\mu_{2,r})\right]^2}{\frac{N_1(\mu_{1,r}+\mu_{2,r})}{N_1+N_2}} + \frac{\left[\mu_{2,r} - \frac{N_2}{N_1+N_2}(\mu_{1,r}+\mu_{2,r})\right]^2}{\frac{N_2}{N_1+N_2}(\mu_{1,r}+\mu_{2,r})} \right\}$$

$$+ \ldots \ldots \ldots \ldots \ldots + \ldots \ldots \ldots \ldots$$

$$= \left\{ \frac{[N_2\,\mu_{1,1} - N_1\,\mu_{2,1}]^2}{N_1\,(N_1+N_2)(\mu_{1,1}+\mu_{2,1})} + \frac{[N_1\,\mu_{2,1} + N_2\,\mu_{1,1}]^2}{N_2\,(N_1+N_2)(\mu_{1,1}+\mu_{2,1})} \right\}$$

$$+ \ldots \ldots \ldots \ldots \ldots + \ldots \ldots \ldots \ldots$$

$$= \left\{ \frac{(N_2\,\mu_{1,1} - N_1\,\mu_{1,2})^2}{N_1\,N_2\,(\mu_{1,1}+\mu_{2,1})} \right\} + \ldots \ldots \ldots \ldots + \ldots \ldots \ldots$$

$$= \left\{ \frac{N_1 N_2 \left(\frac{\mu_{1,1}}{N_1} - \frac{\mu_{1,2}}{N_2}\right)^2}{(\mu_{1,1}+\mu_{2,1})} \right\} + \ldots \ldots \ldots \ldots \ldots \ldots$$

$$= \Sigma \frac{N_1\,N_2 \left(\frac{\mu_{1\cdot1}}{N_1} - \frac{\mu_{1,2}}{N_2}\right)^2}{(\mu_{1,1}+\mu_{2,1})}$$

Problem 375.—The following table shows the association, among 1000 school boys, between their general ability and their mathematical ability. Calculate the coefficient of contingency between the two.

General ability

Math. ability		*Good*	*Fair*	*Poor*	*Total*
	Good	44	22	4	70
	Fair	265	257	178	700
	Poor	41	91	98	230
	Total	350	370	280	1000

(*Punjab M.A.* (*Math.*), *1945*)

Solution :

Assuming that the two attributes the General ability and the Mathematical ability are not associated, we have the following theoretical frequencies :—

General ability

Maths. ability.		*Good*	*Fair*	*Poor*	*Total*
	Good	24·5	25·9	19·6	70
	Fair	245	259	196	700
	Poor	80·5	75·1	64·4	230
	Total	350	370	280	1000

Now $X^2=\Sigma\frac{(f-f_t)^2}{f_t}$

or $X^2=\frac{(44-24\cdot5)^2}{24\cdot5}+\frac{(22-25\cdot9)^2}{25\cdot9}+\frac{(4-19\cdot6)^2}{19\cdot6}+\frac{(265-245)^2}{245}$

$+\frac{(257-259)^2}{259}+\frac{(178-196)^2}{196}+\frac{(41-80\cdot5)^2}{80\cdot5}+\frac{(91-75\cdot1)^2}{75\cdot1}$

$+\frac{(98-64\cdot4)^2}{64\cdot4}$

$=\frac{(19\cdot5)^2}{24\cdot5}+\frac{(3\cdot9)^2}{25\cdot9}+\frac{(15\cdot6)^2}{19\cdot6}+\frac{(20)^2}{245}+\frac{2^2}{259}+\frac{(18)^2}{196}+\frac{(39\cdot5)^2}{80\cdot5}$

$+\frac{(15\cdot9)^2}{75\cdot1}+\frac{(33\cdot6)^2}{64\cdot4}$

$=15\cdot52+\cdot59+12.42+1\cdot63+\cdot02+1\cdot65+19\cdot37+3\cdot37+17\cdot53$

$=72\cdot1$

Now the coefficient of contingency is obtained by the Pearson's formula

$$C=\sqrt{\frac{X^2}{N+X^2}}$$

$$=\sqrt{\frac{72\cdot1}{1000+72\cdot1}}$$

$$=\sqrt{\frac{72\cdot1}{1072\cdot1}}$$

$$=\sqrt{\cdot0672}$$

$$=\cdot26$$

Hence the coefficient of contingency is ·26

Problem 376.—Discuss the resemblance of stature of parent and offspring from the following :—

Parent

Offspring	Very tall	Tall	Medium	Short	Total
Very tall	20	30	20	2	72
Tall	14	125	85	12	236
Medium	3	140	165	125	433
Short	3	37	68	151	259
Total	40	332	338	290	1000

(I.C.S., 1936)

Solution :

Assuming that the two attributes the stature of parent and the stature of offspring are independent we have the theoretical frequencies as below :—

Parent

Offspring	Very tall	Tall	Medium	Short	Total
Very tall	2·9	23·9	24·3	20·9	72
Tall	9·4	78·4	79·8	68·4	236
Medium	17·3	143·8	146·4	125·5	433
Short	10·4	85·9	87·5	75·2	259
Total	40	332	338	290	1000

Now $X^2=\Sigma\frac{(f-f_t)^2}{f_t}$

or $X^2=\frac{(20-2\cdot9)^2}{2\cdot9}+\frac{(30-23\cdot9)^2}{23\cdot9}+\frac{(20-24\cdot3)^2}{24\cdot3}+\frac{(2-20\cdot9)^2}{20\cdot9}$

$+\frac{(14-9\cdot4)^2}{9\cdot4}+\frac{(125-78\cdot4)^2}{78\cdot4}+\frac{(85-79\cdot8)^2}{79\cdot8}+\frac{(12-68\cdot4)^2}{68\cdot4}$

$+\frac{(3-17\cdot3)^2}{17\cdot3}+\frac{(140-143\cdot8)^2}{143\cdot8}+\frac{(165-146\cdot4)^2}{146\cdot4}+\frac{(125-125\cdot5)^2}{125\cdot5}$

$+\frac{(3-10\cdot4)^2}{10\cdot4}+\frac{(37-85\cdot9)^2}{85\cdot9}+\frac{(68-87\cdot5)^2}{87\cdot5}+\frac{(151-75\cdot2)^2}{75\cdot2}$

— Jagu mamc —

$$= \frac{(17\cdot1)^2}{2\cdot9} + \frac{(6\cdot1)^2}{23\cdot9} + \frac{(4\cdot3)^2}{24\cdot3} + \frac{(18\cdot9)^2}{20\cdot9} + \frac{(4\cdot6)^2}{9\cdot4} + \frac{(46\cdot6)^2}{78\cdot4}$$

$$+ \frac{(5\cdot2)^2}{79\cdot8} + \frac{(56\cdot4)^2}{68\cdot4} + \frac{(14\cdot3)^2}{17\cdot3} + \frac{(3\cdot8)^2}{143\cdot8} + \frac{(8\cdot6)^2}{146\cdot4}$$

$$+ \frac{(\cdot5)^2}{125\cdot5} + \frac{(7\cdot4)^2}{10\cdot4} + \frac{(48\cdot9)^2}{85\cdot9} + \frac{(8\cdot5)^2}{87\cdot5} + \frac{(75\cdot8)^2}{75\cdot2}$$

$$= \frac{292\cdot41}{2\cdot9} + \frac{37\cdot21}{23\cdot9} + \frac{18\cdot49}{24\cdot3} + \frac{357\cdot21}{20\cdot9} + \frac{21\cdot16}{9\cdot4} + \frac{2171\cdot56}{78\cdot4}$$

$$+ \frac{27\cdot04}{79\cdot8} + \frac{3180\cdot96}{68\cdot4} + \frac{204\cdot49}{17\cdot3} + \frac{14\cdot44}{143\cdot8} + \frac{73\cdot96}{146\cdot4}$$

$$+ \frac{\cdot25}{125\cdot5} + \frac{54\cdot76}{10\cdot4} + \frac{2391\cdot21}{85\cdot9} + \frac{72\cdot25}{87\cdot5} + \frac{5745\cdot64}{75\cdot2}$$

$$= 100\cdot8 + 1\cdot55 + \cdot76 + 17\cdot09 + 2\cdot25 + 27\cdot69 + \cdot34 + 46\cdot5$$
$$+ 11\cdot82 + \cdot10 + \cdot51 + \cdot00 + 5\cdot26 + 27\cdot83 + \cdot83 + 76\cdot40$$

$$= 319\cdot73$$

Now the coefficient of contingency is obtained by the Pearson's formula

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

$$= \sqrt{\frac{319\cdot73}{1000 + 319\cdot73}}$$

$$= \sqrt{\frac{319\cdot73}{1319\cdot73}}$$

$$= \sqrt{\cdot2317}$$

$$= \cdot48$$

The value of C thus obtained shows that there is association between the stature of parent and stature of offspring.

$$= \frac{67 \cdot 8 - 66}{3 \cdot 011} \sqrt{10}$$

$$= \frac{1 \cdot 8 \times 3 \cdot 162}{3 \cdot 011}$$

$$= 1 \cdot 89.$$

Interpolation of value of P corresponding to $t = 1 \cdot 89$

$t = 1 \cdot 9$	$P = \cdot 955$
$t = 1 \cdot 8$	$P = \cdot 947$
difference $= \cdot 1$	difference $= \cdot 008$

For $\cdot 1$ difference in t the difference in P is $\cdot 008$

$\therefore$ For $\cdot 09$ „ „ $\frac{\cdot 008 \times \cdot 09}{\cdot 1}$

$= \cdot 0072$

$\therefore$ For $t = 1 \cdot 89$, $P = \cdot 9542 = \cdot 954$ appr.

The value of students P comes out to be $\cdot 954$, thus the chance of getting a value of t greater than observed is $1 - \cdot 954$ or $\cdot 046$. The probability of getting t greater in absolute value is $2 \times \cdot 046$ or $\cdot 092$, which is greater than $\cdot 05$. This indicates that the value of t is not significant and hence the mean height in universe is 66 inches.

✓**Problem 378.**—The nine items of a sample had the following values :—

45, 47, 50, 52, 48, 47, 49, 53, 51.

Does the mean of the nine items differ significantly from the assumed population mean of 47·5 ? Given that

$$\nu = 8 \begin{cases} P = \cdot 945 \text{ for } t = 1 \cdot 8 \\ P = \cdot 953 \text{ for } t = 1 \cdot 9 \end{cases}$$

(M. Sc., Agra, 1958)

Solution :

Calculation of the mean and standard deviation.

No. of item	Sample value x	Deviation of x from mean 49·1	Square of deviation $(x - \bar{x})^2$
1	45	−4·1	16·81
2	47	−2·1	4·41
3	50	·9	·81
4	52	2·9	8·41
5	48	−1·1	1·21
6	47	−2·1	4·41
7	49	− ·1	·01
8	53	3·9	15·21
9	51	1·9	3·61
Total	$\Sigma x = 442$		$\Sigma(x - \bar{x})^2 = 54 \cdot 89$

$$\text{Sample mean } \bar{x} = \frac{\Sigma x}{n} = \frac{442}{9} = 49 \cdot 1 \text{ appr.}$$

$$\text{Standard deviation } \sigma_s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{54 \cdot 89}{8}}$$

$$= \sqrt{6 \cdot 86}$$

$$= 2 \cdot 6 \text{ appr.}$$

On the assumption that the mean for the universe is 47·5,

$$t = \frac{\bar{x} - M}{\sigma_s} \sqrt{n}$$

$$= \frac{49 \cdot 1 - 47 \cdot 5}{2 \cdot 6} \sqrt{9}$$

$$= \frac{1 \cdot 6 \times 3}{2 \cdot 6}$$

$$= 1 \cdot 85 \text{ appr.}$$

The interpolated value of **P** for $t = 1 \cdot 85$ comes out to be ·949, thus the chance of getting a value of t greater than observed is 1 — ·949 or ·051. The probability of getting t greater in absolute value is 2 × ·051 or ·102, which is greater than ·05. This indicates that the value of t is not significant and hence the mean in universe is 47·5.

Problem 379.—Find the student's t for the following variate values in a sample of eight :—

$$-4, -2, -2, 0, 2, 2, 3, 3.$$

taking the mean of the universe to be zero.

(*M.Sc., Agra, 1948*)

Solution :—

Calculation of mean and standard deviation.

No. of item	Variate value x	Deviation from mean $(x - \bar{x})$	Square of deviation $(x - \bar{x})^2$
1	−4	−4·25	18·0625
2	−2	−2·25	5·0625
3	−2	−2·25	5·0625
4	0	−·25	·0625
5	2	1·75	3·0625
6	2	1·75	3·0625
7	3	2·75	7·5625
8	3	2·75	7·5625
Total	$\Sigma x = 2$		49·5000

$$\text{Sample mean } \bar{x} = \frac{\Sigma x}{n} = \frac{2}{8} = \cdot 25$$

$$\text{Standard deviation } \sigma_s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{49 \cdot 5}{7}}$$

$$= \sqrt{7 \cdot 071428}$$

$$= 2 \cdot 487$$

On the assumption that the mean of the universe is 0, we get

$$t = \frac{\bar{x} - M}{\sigma_s}\sqrt{n}$$

$$= \frac{\cdot 25 - 0}{2 \cdot 487}\sqrt{8}$$

$$= \cdot 29 \text{ appr.}$$

Note :—To discuss the significance of this value of t, we find P from the table for 7 degrees of freedom. The value comes out nearly ·6, thus the chance of getting a value of t greater than observed is 1—·6 or ·4. The probability of getting t greater in absolute value is 2×·4 or ·8, which is greater than ·05. This indicates that the value of t is not significant and hence the mean of universe can be taken zero.

✓**Problem 380.**—Ten individuals are chosen at random from a population and their heights are found to be, in inches, 63, 63, 64, 65, 66, 69, 69, 70, 70, 71. Discuss the suggestion that the mean height in the universe is 65 inches, given that for 9 degrees of freedom the value of student's t at 5 per cent level of significance is 2·262. (*M.Sc., Agra, 1957*)

Solution :

Calculation of mean and standard deviation.

No. of individual	Heights of individual x	Deviation from mean height $(x-\bar{x})$	Square of deviation $(x-\bar{x})^2$
1	63	−4	16
2	63	−4	16
3	64	−3	9
4	65	−2	4
5	66	−1	1
6	69	2	4
7	69	2	4
8	70	3	9
9	70	3	9
10	71	4	16
Total	$\Sigma x = 670$		$\Sigma(x-\bar{x})^2 = 88$

Sample mean $\bar{x} = \frac{\Sigma x}{n} = \frac{670}{10} = 67$ inches

Standard deviation $\sigma_s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

$= \sqrt{\frac{88}{9}}$

$= \sqrt{9 \cdot 7778}$

$= 3 \cdot 13$ inches

On the assumption that the mean height of the universe is 65 inches, we have

$$t = \frac{\bar{x} - M}{\sigma_s} \sqrt{n}$$

$$= \frac{67 - 65}{3 \cdot 13} \sqrt{10}$$

$$= \frac{2 \times 3 \cdot 162}{3 \cdot 13}$$

$$= 2 \cdot 02$$

The calculated value of t is less than the table value, hence the value of t is not significant showing that mean height in the universe is 65 inches.

✓**Problem 381.**—Find student's t for the following variate values in a sample of 10 ; −6, −4, −3, −2, −2, 0, 1, 1, 3, 5, taking m to be zero, and find from the tables the probability of getting a value of t as great or greater on random sampling from a normal universe.

(Yule and Kandall 23·1 Page 460)

Solution :

Calculation of mean and standard deviation.

No. of item	Variate value x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$
1	−6	−5·3	28·09
2	−4	−3·3	10·89
3	−3	−2·3	5·29
4	−2	−1·3	1·69
5	−2	−1·3	1·69
6	0	·7	·49
7	1	1·7	2·89
8	1	1·7	2·89
9	3	3·7	13·69
10	5	5·7	32·49
Total	$\Sigma x = -7$		$\Sigma(x-\bar{x})^2 = 100 \cdot 10$

Sample mean $\bar{x} = \frac{\Sigma x}{n} = \frac{-7}{10} = -\cdot 7$

Standard deviation $\sigma_s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

$$= \sqrt{\frac{100\cdot 1}{9}}$$

$$= \sqrt{11\cdot 1222}$$

$$= 3\cdot 34$$

On the assumption that the mean of the universe is zero, we have

$$t = \frac{\bar{x} - M}{\sigma_s}\sqrt{n}$$

$$= \frac{-\cdot 7 - 0}{3\cdot 34}\sqrt{10}$$

$$= -\frac{\cdot 7 \times 3\cdot 162}{3\cdot 34}$$

$$= -\cdot 664$$

ν (degree of freedom) $= 10 - 1 = 9$

The value of P for 9 degrees of freedom corresponding to value $t = \cdot 664$ is $\cdot 738$, thus the chance of getting a value of t greater than observed is $1 - \cdot 738$ or $\cdot 262$. The probability of getting t greater in absolute value is $2 \times \cdot 262$ or $\cdot 524$.

Problem 382.—Why should there be different formula for testing significance of difference in means when the samples are (*a*) small and (*b*) large?

The yields of two types, 'Type 17' and 'Type 51', of grams in pounds per acre at 6 replications are given below. What comments would you make on the differences in the mean yields? You may assume that if there be 5 degrees of freedom and $P = \cdot 2$, t is approximately $1\cdot 476$.

Replication	*Yields in pounds (Type 17)*	*Yields in pounds (Type 51)*
1	20·50	24·86
2	24·60	26·39
3	23·06	28·19
4	29·98	30·75
5	30·37	29·97
6	23·83	22·04

(*I.A.S., 1951*)

Solution :

Calculation of the mean and standard deviation of the differences in yields of two types of gram.

No. of Replication	Yields in Pounds Type 17	Yields in Pounds Type 51	Difference in yields x	Deviation from mean value $x-\bar{x}$	Square of deviation $(x-\bar{x})$
1	20·50	24·86	4·36	2·72	7·3984
2	24·60	26·39	1·79	·15	·0225
3	23·06	28·19	5·13	3·49	12·1801
4	29·98	30·75	·77	— ·87	·7569
5	30·37	29·97	—·40	—2·04	4·1616
6	23·83	22·04	—1·79	3·43	11·7649
Total			$\Sigma x=9{\cdot}86$		36·2844

Sample mean of the difference of the yields is

$$\bar{x}=\frac{\Sigma x}{n}=\frac{9{\cdot}86}{6}=1{\cdot}64 \text{ pounds appr.}$$

Standard deviation of the difference of the yields

$$\sigma_s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$=\sqrt{\frac{36{\cdot}2844}{5}}$$

$$=\sqrt{7{\cdot}2569}$$

$$=2{\cdot}69 \text{ appr.}$$

On the assumption that the differences in the types has got no effect on yields *i.e.*, the true mean of difference to be zero, we have

$$t=\frac{\bar{x}-M}{\sigma_s}\sqrt{n}$$

$$=\frac{1{\cdot}64-0}{2{\cdot}69}\sqrt{6}$$

$$=\frac{1{\cdot}64\times 2{\cdot}449}{2{\cdot}69}$$

$$=\frac{4{\cdot}01636}{2{\cdot}69}$$

$$=1{\cdot}493$$

The number of degrees of freedom$=6-1=5$

For 5 degrees of freedom P$=\cdot 2$ for $t=1{\cdot}476$ indicates that the value of t at 2 per cent level of significance is 1·476, which is less than the calculated value showing that our hypothesis is not correct.

Hence it seems that at this level of significance there is a difference in yields corresponding to difference in the type of gram.

Problem 383.—Nine patients, to whom a certain drink was administered registered the following increments in blood pressure: 7, 3, −1, 4, −3, 5, 6, −4, 1. Show that the data do not indicate that the drink was responsible for these increments.

(*Weatherburn, Page 206*)

Solution :

Calculation of mean and standard deviation of increments in blood pressure.

No. of patient	Increments in blood pressure x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$
1	7	5	25
2	3	1	1
3	−1	−3	9
4	4	2	4
5	−3	−5	25
6	5	3	9
7	6	4	16
8	−4	−6	36
9	1	−1	1
Total	$\Sigma x=18$		$\Sigma(x-\bar{x})^2=126$

Sample mean of the increments in blood pressure

$$\bar{x}=\frac{\Sigma x}{n}=\frac{18}{9}=2$$

Standard deviation of the increments in blood pressure

$$\sigma_s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$=\sqrt{\frac{126}{8}}$$

$$=\sqrt{15{\cdot}75}$$

$$=3{\cdot}96$$

On the null hypothesis the above values are regarded as a random sample from a normal population whose mean is zero, we have

$$t=\frac{\bar{x}-M}{\sigma_s}\sqrt{n}$$

$$= \frac{2-0}{3 \cdot 96} \sqrt{9}$$

$$= \frac{6}{3 \cdot 96}$$

$$= 1 \cdot 515$$

Degrees of freedom $= 9 - 1 = 8$

The value of P for 8 degrees of freedom corresponding to $t = 1 \cdot 515$ is ·915 appr., thus the chance of getting a value of t greater than observed is $1 - \cdot 915$ or ·085. The probability of getting t greater in absolute value is $2 \times \cdot 085$ or ·17, which is greater than ·05. This indicates that the value of t is not significant and hence the drink does not effect the increment in blood pressure.

Problem 384.—In testing the superiority of Leaks drill over the Ordinary drill, plots in the form of long strips were cultivated, two adjacent strips being allotted at random to Leak's drill and the Ordinary. For ten such pairs of plots the values of the excess of the weight of grain from the plot treated by Leaks drill over that obtained by use of the Ordinary drill were 2·4, 1·0, ·7, 0, 1·1, 1·6, 1·1, −·4, ·1 and ·7. Show that the data furnish strong evidence of the superiority of Leaks drill.

(*Weatherburn, Page 206*)

Solution :

Calculation of mean and standard deviation of the excess of the weight of grain.

No. of plots	Excess of weight of grain x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$
1	2·4	1·57	2·4649
2	1	·17	·0289
3	·7	−·13	·0169
4	0	−·83	·6889
5	1·1	·27	·0729
6	1·6	·77	·5929
7	1·1	·27	·0729
8	−·4	−1·23	1·5129
9	·1	−·73	·5329
10	·7	−·13	·0169
Total	$\Sigma x = 8 \cdot 3$		6·0010

Sample mean of the excess weight of grain is

$$\bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{8 \cdot 3}{10} = \cdot 83$$

Standard deviation of the excess of weight of grains.

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{6\cdot0010}{9}}$$

$$= \sqrt{\cdot666778}$$

$$= \cdot816$$

On the null hypothesis that is the type of drill do not effect the production of grain we have the true mean of excess of weights of grain to be zero, we get

$$t = \frac{\bar{x} - M}{\sigma_s}\sqrt{n}$$

$$= \frac{\cdot83 - 0}{\cdot816}\sqrt{10}$$

$$= \frac{\cdot83 \times 3\cdot162}{\cdot816}$$

$$= 3\cdot22$$

Degrees of freedom $= 10 - 1 = 9$

The corresponding value of P is ·995, thus the chance of getting a value of t greater than observed is $1 - \cdot995$ or ·005. The probability of getting t greater in absolute value is $2 \times \cdot005$ or ·01. This value belongs to about the 1% level of significance giving strong evidence against the null hypothesis. Hence Leaks drill is superior to Ordinary drill.

Problem 385.—The table signifies additional hours of sleep gained by 10 patients in an experiment with a sleeping drug :—

Patient	1	2	3	4	5	6	7	8	9	10
Hours gained	·7	1·1	·2	1·2	·1	3·4	3·7	·8	1·8	2

Assuming that the hours of sleep is a normally distributed variable, calculate t for the above table.

(*M.Sc., Agra, 1953*)

Solution :

Calculation of mean and standard deviation.

Patient No.	Hours gained x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$
1	·	$-\cdot8$	$\cdot64$
2	$1\cdot1$	$-\cdot4$	$\cdot16$
3	$\cdot2$	$-1\cdot3$	$1\cdot69$
4	$1\cdot2$	$-\cdot3$	$\cdot09$
5	$\cdot1$	$-1\cdot4$	$1\cdot96$
6	$3\cdot4$	$1\cdot9$	$3\cdot61$
7	$3\cdot7$	$2\cdot2$	$4\cdot84$
8	$\cdot8$	$-\cdot7$	$\cdot49$
9	$1\cdot8$	$\cdot3$	$\cdot09$
10	2	$\cdot5$	$\cdot25$
Total	$\Sigma x=15$		$13\cdot82$

Mean of the additional hours of sleep, $\bar{x}=\dfrac{\Sigma x}{n}=\dfrac{15}{10}=1\cdot5$

Standard deviation of additional hours of sleeep

$$\sigma_s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$=\sqrt{\frac{13\cdot82}{9}}$$

$$=\sqrt{1\cdot535556}$$

$$=1\cdot239$$

then $\quad t=\dfrac{\bar{x}-M}{\sigma_s}\sqrt{n}\quad$ [On null hypothesis *i.e.* zero mean]

$$=\frac{1\cdot5-0}{1\cdot239}\sqrt{10}$$

$$=\frac{4\cdot743}{1\cdot239}$$

$$=3\cdot82$$

Problem 386.—A random sample of nine from the men of a large city gave a mean height of 68 inches ; and the unbiassed estimate s^2 of the population variance found from the sample was 4·5 in.². Are these data consistent with the assumption of a mean height of 68·5 in. for the men of the city ?

Solution :

Given

$$\bar{x} = 68 \text{ in.}$$

$$s = \sqrt{4 \cdot 5} = 2 \cdot 12$$

$$\nu = 9 - 1 = 8$$

Therefore on the assumption that M=68·5 we have

$$t = \frac{\bar{x} - M}{s} \sqrt{n}$$

$$= \frac{68 - 68 \cdot 5}{2 \cdot 12} \sqrt{9}$$

$$= \frac{1 \cdot 5}{2 \cdot 12}$$

$$= \cdot 707$$

The value of P corresponding to $t = \cdot 707$ for 8 degrees of freedom is from the table ·748, thus the chance of getting a value of t greater than observed is 1 — ·748 or ·252. The probability of getting t greater in absolute value is 2 × ·252 or ·504, which is greater than ·05. This indicates that the value of t is not significant hence the mean height of the men of city is 68·5 inches.

Problem 387.—A random sample of 16 values from a normal population showed a mean of 41·5 in. and a sum of squares of deviations from this mean equal to 135 in^2. Show that the assumption of a mean of 43·5 in. for the population is not reasonable.

Solution :

Given

$$\bar{x} = 41 \cdot 5 \text{ in.}$$

$$\Sigma(x - \bar{x})^2 = 135 \qquad n = 16$$

We have

$$\sigma_s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{13}{15}}$$

$$= 3$$

Therefore on the assumption that M = 43·5 in.

We have

$$t = \frac{\bar{x} - M}{\sigma_s} \sqrt{n}$$

$$= \frac{41 \cdot 5 - 43 \cdot 5}{3} \sqrt{16}$$

$$= -\frac{8}{3}$$

$$= -2 \cdot 667$$

Degrees of freedom $= 16 - 1 = 15$

The value of P corresponding to this value of t for 15 degrees of freedom is ·991 appr., thus the chance of getting a value of t greater than observed is 1—·991 or ·009. The probability of getting t greater in absolute value is $2 \times \cdot 009$ or ·018, which is less than ·05 ; hence the value of t is significant showing that assumption is not correct and mean cannot be 43·5 in.

Problem 388.—In a Test Examination given to two groups of students the marks obtained were as follows :

First group :— 18, 20, 36, 50, 49, 36, 34, 49, 41.
Second group :—29, 28, 26, 35, 30, 44, 46.

Examine the significance of difference between the arithmetic averages of the marks secured by the students of the above two groups. *(P.C.S. 1951)*

Solution :

Calculation of the mean and the standard deviation of the samples for first group and second group.

First group			Second group		
Marks obtained x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$	Marks obtained y	Deviation from mean $y-\bar{y}$	Square of deviation $(y-\bar{y})$
18	—19	361	29	—5	25
20	—17	289	28	—6	36
36	— 1	1	26	—8	64
50	13	169	35	1	1
49	12	144	30	—4	16
36	— 1	1	44	10	100
34	— 3	9	46	12	144
49	12	144			
41	4	16			
$\Sigma x = 333$		$\Sigma(x-\bar{x})^2 = 1134$	$\Sigma y = 238$		$\Sigma(y-\bar{y})^2 = 386$

Mean of first group $\bar{x} = \frac{\Sigma x}{n_1} = \frac{333}{9} = 37$

Mean of second group $\bar{y} = \frac{\Sigma y}{n_2} = \frac{238}{7} = 34$

Standard deviation of the combined groups is

$$s = \sqrt{\frac{\Sigma (x-\bar{x})^2 + \Sigma (y-\bar{y})^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{1134+386}{9+7-2}}$$

$$= \sqrt{\frac{1520}{14}}$$

$$= \sqrt{108 \cdot 57}$$

$$= 10 \cdot 4$$

Applying these values in the formula

$$t = \frac{\bar{x}-\bar{y}}{s} \sqrt{\frac{n_1 \ n_2}{n_1+n_2}}$$

$$= \frac{37-34}{10 \cdot 4} \sqrt{\frac{9 \times 7}{9+7}}$$

$$= \frac{3 \times 7 \cdot 93}{10 \cdot 4 \times 4}$$

$$= \cdot 57$$

The degrees of freedom $= 9+7-2 = 14$

The value of P corresponding to this value of t for 14 degrees of freedom is ·737, thus the chance of getting a value of t greater than observed is $1 - \cdot 737$ or ·263. The probability of getting t greater in absolute value is $2 \times \cdot 263$ or ·526, which is greater than ·05. Hence the value of t is not significant and the two groups of students are of same standard.

Problem 389.—Two independent samples of 8 and 7 items respectively had the following values :—

Sample 1.	9,	11,	13,	11,	15,	9,	12,	14.
Sample 2.	10,	12,	10,	14,	9,	8,	10.	

Is the difference between the means of the samples significant ? Given that

$$\nu = 13 \left[\begin{array}{l} P = \cdot 874 \text{ for } t = 1 \cdot 2 \\ P = \cdot 892 \text{ for } t = 1 \cdot 3 \end{array} \right.$$

(*M. Sc., Agra, 1958*)

Solution :

Determination of mean and standard deviation of samples.

Sample 1			Sample 2		
Size of item x	Deviation from mean $(x-\bar{x})$	Square of deviation $(x-\bar{x})^2$	Size of item y	Deviation from mean $(y-\bar{y})$	Square of deviation $(y-\bar{y})^2$
9	—2·75	7·5625	10	—·43	·1849
11	—·75	·5625	12	1·57	2·4649
13	—1·25	1·5625	10	—·43	·1849
11	—·75	·5625	14	3·57	12·7449
15	3·25	10·5625	9	—1·43	2·0449
9	—2·75	7·5625	8	—2·43	5·9049
12	·25	·0625	10	—·43	·1849
14	2·25	5·0625			
$\Sigma x=94$		$\Sigma(x-\bar{x})^2=33.5$	$\Sigma y=73$		$\Sigma(y-\bar{y})^2=23{\cdot}4143$

Mean of sample 1 is $\bar{x}=\frac{\Sigma x}{n_1}=\frac{94}{8}=11{\cdot}75$

Mean of sample 2 is $\bar{y}=\frac{\Sigma y}{n_2}=\frac{73}{7}=10{\cdot}43$

Standard deviation of the combined items is

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{33{\cdot}5+23{\cdot}4143}{8+7-2}}$$

$$=\sqrt{\frac{56{\cdot}9143}{13}}$$

$$=\sqrt{4{\cdot}3780}$$

$$=2{\cdot}09$$

From these values we h ve

$$t=\frac{\bar{x}-\bar{y}}{s}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$=\frac{11{\cdot}75-10{\cdot}43}{2{\cdot}09}\sqrt{\frac{8\times 7}{8+7}}$$

$$=\frac{1{\cdot}32\times 1{\cdot}93}{2{\cdot}90}$$

$$=\frac{2 \cdot 5476}{2 \cdot 09}$$

$$=1 \cdot 21$$

The degrees of freedom $=8+7-2=13$

The value of P corresponding to $t=1 \cdot 21$ for 13 degrees of freedom is ·875, thus the chance of getting a value of t greater than observed is $1-\cdot 875$ or ·125. The probability of getting t greater in absolute value is $2\times\cdot 125$ or ·25, which is greater than ·05 hence the value of t is not significant and hence the difference between the means of two samples is not significant.

Problem 390.—The heights of six randomly chosen sailors are, in inches: 63, 65, 68, 69, 71 and 72. Those of ten randomly chosen soldiers are ; 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Discuss the light that these data throw on the suggestion that soldiers are, on the average, taller than sailors ; given that

$$\nu=14 \begin{cases} P=\cdot 539 & \text{for } t=\cdot 1 \\ P=\cdot 527 & \text{for } t=\cdot 08 \end{cases}$$

(M. Sc., Agra, 1956)

Solution :

Calculation of mean and standard deviation.

Soldiers			Sailors		
Height x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$	Height y	Deviation from mean $y-\bar{y}$	Square of deviation $(y-\bar{y})^2$
61	—6·8	46·24	63	—5	25
62	—5·8	33·64	65	—3	9
65	—2·8	7·84	68	0	0
66	—1·8	3·24	69	1	1
69	1·2	1·44	71	3	9
69	1·2	1·44	72	4	16
70	2·2	4·84			
71	3·2	10·24			
72	4·2	17·64			
73	5·2	27·04			
678		153·60	408		60

Mean of the heights of soldiers is $\bar{x}=\frac{\Sigma x}{n_1}=\frac{678}{10}=67 \cdot 8$

Mean of the heights of sailors is $\bar{y}=\frac{\Sigma y}{n_2}=\frac{408}{6}=68$

Standard deviation of the combination is

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{153 \cdot 60 + 60}{10+6-2}}$$

$$= \sqrt{\frac{213 \cdot 60}{14}}$$

$$= \sqrt{15 \cdot 2571}$$

$$= 3 \cdot 9.$$

The value of t is determined by the formula

$$t = \frac{\bar{x} - \bar{y}}{s} \sqrt{\frac{n_1 \, n_2}{n_1+n_2}}$$

$$= \frac{67 \cdot 8 - 68}{3 \cdot 9} \sqrt{\frac{10 \times 6}{10+6}}$$

$$= -\frac{\cdot 2 \times 7 \cdot 74}{3 \cdot 9 \times 4}$$

$$= - \cdot 099$$

Degrees of freedom $= 10+6-2 = 14$

The corresponding value of **P** is ·538, thus the chance of getting a value of t greater than observed is $1 - \cdot 538$ or ·462. The probability of getting t greater in absolute value is $2 \times \cdot 462$ or ·924, which is much greater than ·05, hence the value of t is not significant and hence the suggestion that soldiers are taller than sailors is wrong.

Problem 391.—What is Fisher's t—test for small samples? Explain giving suitable formula, how this test can be applied for testing the significance of the difference between two sample means. Hence calculate the value of t in the case of two characters A and B whose corresponding values are given below :—

A :	16, 10, 8, 9, 9, 8
B :	8, 4, 5, 9, 12, 4

(P. C. S., 1952)
(M. Sc., Agra, 1945, 1946, 1951)

Solution :

Calculation of the mean and standard deviation.

Character A			Character B		
Size of item x	Deviation from mean $x-\bar{x}$	Square of deviation $(x-\bar{x})^2$	Size of item y	Deviation from mean $y-\bar{y}$	Square of deviation $(y-\bar{y})^2$
16	6	36	8	1	1
10	0	0	4	−3	9
8	−2	4	5	−2	4
9	−1	1	9	2	4
9	−1	1	12	5	25
8	−2	4	4	−3	9
60		46	42		52

Mean of items of character A is $\bar{x} = \frac{\Sigma x}{n_1} = \frac{60}{6} = 10$

Mean of items of character B is $\bar{y} = \frac{\Sigma y}{n_2} = \frac{42}{6} = 7$

Standard deviation of combination is given by

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{46+52}{6+6-2}}$$

$$= \sqrt{9 \cdot 8}$$

$$= 3 \cdot 13.$$

From these data the value of t is obtained by the formula

$$t = \frac{\bar{x}-\bar{y}}{s}\sqrt{\frac{n_1 \; n_2}{n_1+n_2}}$$

$$= \frac{10-7}{3 \cdot 13}\sqrt{\frac{6\times 6}{6+6}}$$

$$= \frac{3\times 1 \cdot 732}{3 \cdot 13}$$

$$= 1 \cdot 66$$

Degrees of freedom $= 6+6-2 = 10$.

Note :—The value of P corresponding to value 1·66 of t for 10 degrees of freedom is ·936, thus the chance of getting a value of t

greater than observed is 1 − ·936 or ·064. The probability of getting t greater in absolute value is $2 \times \cdot 064$ or ·128, which is greater than ·05, hence the value of t is not significant and the two characters comes from the same universe.

Problem 392. – For a random sample of 10 pigs, fed on diet A, the increases in weight in pounds in a certain period were

10, 6, 16, 17, 13, 12, 8, 14, 15, 9.

For another random sample of 12 pigs, fed on diet B, the increases in the same period were 7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17.

Test whether diets A and B differ significantly as regards their effects on increase on weight. You may use the following extract from statistieal table :–

Degree of freedom	*Value of 't' significant at 5% level of probability*
19	2·09
20	2·09
21	2·08
22	2·07
23	2·07

(*I.A.S. 1954*)

Solution :

Calculation of the mean and standard deviation.

On diet A			On diet B		
Increase in weight in pounds x	Deviation from mean $(x-\bar{x})$	Square of deviation $(x-\bar{x})^2$	Increase in weight in pounds y	Deviation from mean $y-\bar{y}$	Square of deviation $(y-\bar{y})^2$
10	−2	4	7	−8	64
6	−6	36	13	−2	4
16	4	16	22	7	49
17	5	25	15	0	0
13	1	1	12	−3	9
12	0	0	14	−1	1
8	−4	16	18	3	9
14	2	4	8	−7	49
15	3	9	21	6	36
9	3	9	23	8	64
			10	−5	25
			17	2	4
$\Sigma x=120$		120	$\Sigma y=180$		314

Mean of increases in weights on diet A is $\bar{x} = \frac{\Sigma x}{n_1} = \frac{120}{10}$

$= 12$ pounds

Mean of increases in weights on diet B is $\bar{y} = \frac{\Sigma y}{n_2} = \frac{180}{12}$

$= 15$ pounds

Standard deviation of the combination

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{120+314}{10+12-2}}$$

$$= \sqrt{21 \cdot 7}$$

$$= 4 \cdot 65$$

From these values we have the value of t by the formula

$$t = \frac{\bar{x}-\bar{y}}{S} \sqrt{\frac{n_1 \; n_2}{n_1+n_2}}$$

$$= \frac{12-15}{4 \cdot 65} \sqrt{\frac{12 \times 10}{12+10}}$$

$$= -\frac{3 \times 2 \cdot 33}{4 \cdot 65}$$

$$= -1 \cdot 5$$

Degrees of freedom $= 10+12-2 = 20$.

The value of t for 20 degrees of freedom at 5% level of significance is 2·09, which is greater than calculated value. Hence the difference is not significant.

Problem 393.—Show how you would use Students' t-test and Fisher's z-test to decide whether the two sets of observations

17, 27, 18, 25, 27, 29, 27, 23, 17

and 16, 16, 20, 16, 20, 17, 15, 21

indicate samples drawn from the same universe.

(*Agra M.Sc., 1949*)

Solution :

Calculation of means and standard deviations.

First Set			Second Set		
Size of item x	Deviation from mean $(x-\bar{x})$	Square of Deviation $(x-\bar{x})^2$	Size of item y	Deviation from mean $y-\bar{y}$	Square of deviation $(y-\bar{y})^2$
17	—6·33	40·0689	16	—1·63	2·6569
27	3·67	13·4689	16	—1·63	2·6569
18	—5·33	28·4089	20	2·37	5·6169
25	1·67	2·7889	16	—1·63	2·6569
27	3·67	13·4689	20	2·37	5·6169
29	5·67	32·1489	17	—·63	·3969
27	3·67	13·4689	15	—2·63	6·9169
23	—·33	·1089	21	3·37	11·3569
17	—6·33	40·0689			
210		184·0001	141		37·8752

Mean of the first set $\bar{x}=\frac{\Sigma x}{n_1}=\frac{210}{9}$ = 23·33 appr.

Mean of the second set $\bar{y}=\frac{\Sigma y}{n_2}=\frac{141}{8}$ = 17·63 appr.

Standard deviation of first set

$$\sigma x=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n_1-1}}.$$

$$\therefore \qquad \sigma^2 x=\frac{184}{8}=23$$

Similarly $$\sigma^2 y=\frac{37\cdot8752}{7}=5\cdot41 \text{ appr.}$$

Standard deviation of combination is given by

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{n_1+n_2-2}}$$

$$=\sqrt{\frac{184\cdot0001+37\cdot8752}{15}}$$

$$=\sqrt{14\cdot7917}$$

$$=3\cdot84$$

Students' t-Test

$$t=\frac{\bar{x}-\bar{y}}{s}\sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$=\frac{23{\cdot}33-17{\cdot}63}{3{\cdot}84}\sqrt{\frac{9\times 8}{9+8}}$$

$$=\frac{5{\cdot}7\times 2{\cdot}06}{3{\cdot}84}$$

$$=3{\cdot}05.$$

The degrees of freedom $=9+8-2=15$

We find the value of P from the tables corresponding to this value of t for 15 degrees of freedom. If the value of P is such that t is significant, it will be indicated that the samples do not come from the same universe. [The value of P is ·995, thus the probability that t is greater in absolute value is $2\times(1-{\cdot}995)$ or ·01 ; which is less than ·05 showing that t is significant.

Fishers' Z-test

$$z=\tfrac{1}{2}\log_e\frac{23}{5{\cdot}41}$$

$$=1{\cdot}1512\log_{10}\frac{23}{5{\cdot}41}$$

$$=1{\cdot}1512\log_{10}4{\cdot}251$$

$$=1{\cdot}1512\times{\cdot}6248.$$

$$={\cdot}7193.$$

The degrees of freedom are 8 and 7 respectively. We can find out for these degrees of freedom the value of z_0 from the tables for 1 and 5 per cent points of significance. Comparing ·7193 with these values we can decide if this value is significant or insignificant after which we can draw a suitable conclusion.

Problem 394.—The heights of the ten men of a random sample from an unknown population gave a mean of 69 in., and a sum of square of deviations from the mean equal to 42 in^2. Another random sample of nine men gave a mean height 68 in., and sum of squares of deviations from the mean equal to 36 in^2. Apply the t test to the hypothesis that the two samples are from the same population.

Solution :

$$\bar{x}=69\ ;\ \Sigma(x-\bar{x})^2=42\ ;\ n_1=10$$

$$\bar{y}=68\ ;\ \Sigma(y-\bar{y})^2=36\ ;\ n_2=9$$

The standard error $s=\sqrt{\frac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{n_1+n_2-2}}$

$$=\sqrt{\frac{42+36}{17}}$$

$$=2{\cdot}14.$$

The value of t is obtained by the formula

$$t=\frac{\bar{x}-\bar{y}}{s}\sqrt{\frac{n_1n_2}{n_1+n_2}}$$

$$=\frac{69-68}{2{\cdot}14}\sqrt{\frac{90}{19}}$$

$$=1{\cdot}01.$$

For 17 degrees of freedom this value is not at all significant. Hence the two samples are from same population.

Problem 395.—A random sample of 13 pairs of values from a normal population gave correlation co-efficient of ·6. Is this significant of the existence of correlation in the universe.

Solution :

$$r=\tfrac{6}{10}, \qquad n=13$$

we have

$$t=\frac{\sqrt{n-2}}{\sqrt{1-r^2}}r$$

$$=\frac{\sqrt{11}}{\sqrt{1-\frac{36}{100}}}\times\frac{6}{10}$$

$$=\sqrt{11}\cdot\tfrac{6}{8}$$

$$=\frac{3{\cdot}317\times3}{4}$$

$$=2{\cdot}488 \text{ appr.}$$

Degrees of freedom $=13-2=11$.

The value of P corresponding to this value of t for 11 degrees of freedom is ·984, thus the chance of getting a value of t greater in absolute value is $2\times(1-{\cdot}984)$ or ·032, which is less than ·05. Hence t is significant. Thus our hypothesis that the variables in the normal population are uncorrelated is not correct.

Problem 396.—Find the least value of r, in a sample of 27 pairs from a normal population, that is significant at the 5% level.

(*Weatherburn Page 193*)

Solution :

Degrees of freedom $=27-2=25$

Value of t at 5% level of significance for 25 degrees of freedom is $t=2{\cdot}06$.

Hence for r to be significant we must have

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} > 2{\cdot}06$$

or $$\frac{5r}{\sqrt{1-r^2}} > 2{\cdot}06$$

or $$\frac{r}{\sqrt{1-r^2}} > {\cdot}412$$

or $$r^2 > (1-r^2)\times{\cdot}1697$$

or $$1{\cdot}1697\, r^2 > {\cdot}1697$$

or $$r^2 > {\cdot}145$$

$\therefore$ $$|r| > {\cdot}38$$

Problem 397.—Twelve pictures submitted in a competition were ranked by two judges with results as shown in the table below :—

TABLE

Pictures	A	B	C	D	E	F	G	H	I	J	K	L
Rank assigned by first judge	5	9	6	7	1	3	4	12	2	11	10	8
Rank assigned by second judge	5	8	9	11	3	1	2	10	4	12	7	6

Calculate ρ. Is there a lack of independence in these rankings ?

(Assume that on the hypothesis of independence of two sets of n rankings, $t=\rho\sqrt{\dfrac{n-2}{1-\rho^2}}$ follows the t-distribution with $n-2$ degrees of freedom).

Given that

Degrees of freedom	10	11	12
Value of (t) significant at 5 per cent level of probability	2·23	2·20	2·18

(I.A.S. 1956)

Solution :

Spearman coefficient of rank correlation is

$$p=1-\frac{6\Sigma d^2}{n(n^2-1)}$$

Where n is the number of individuals ranked and d is the difference in the ranks assigned to the same individual.

$$\Sigma d^2=0+1+9+16+4+4+4+4+4+1+9+4$$
$$=60$$

$$\therefore \quad p=1-\frac{6\times 60}{12\,(12^2-1)}$$
$$=1-\frac{30}{143}$$
$$=1-{\cdot}209$$
$$={\cdot}791$$

Assuming that the two sets of ranking are independent we have

$$t=p\sqrt{\frac{n-2}{1-p^2}}$$
$$={\cdot}791\sqrt{\frac{10}{1-({\cdot}791)^2}}$$
$$={\cdot}791\times 5{\cdot}17$$
$$=4{\cdot}089$$

Degrees of freedom $=12-2=10$

Value of t for 10 degrees of freedom at 5% level of significance is 2·23 which is less than calculated value, hence the ranking is significant.

Problem 398.—In a random sample of 28 pairs of values from a bivariate normal population, the correlation was found to be ·7. Is this value consistent with the assumption that the correlation in the population is ·5 ?

(*Weatherburn Page 202*)

Solution :

$$r={\cdot}7,\ \rho={\cdot}5,\ n=28$$

$$Z=\tfrac{1}{2}\log_e\frac{1+r}{1-r}$$

$$=1{\cdot}1513\times\log_{10}\frac{1+\frac{7}{10}}{1-\frac{7}{10}}$$

$$=1{\cdot}1513\times\log_{10}\frac{17}{3}$$
$$=1{\cdot}1513\times\log_{10}5{\cdot}666$$
$$=1{\cdot}1513\times{\cdot}7533$$
$$={\cdot}867$$

$$\zeta=\tfrac{1}{2}\log_e\frac{1+\rho}{1-\rho}$$
$$=1{\cdot}1513\times\log_{10}3$$
$$=1{\cdot}1513\times{\cdot}4771$$
$$={\cdot}549$$
$$\therefore Z-\zeta={\cdot}867-{\cdot}549$$
$$={\cdot}318$$

The standard error of Z is $\frac{1}{\sqrt{25}}={\cdot}2$

Since $(Z-\zeta)$ is considerably less than twice the standard error, its value is not significant. So far as this test goes, the correlation in the population might very well be ·5.

Problem 399.—The first of two samples consists of 23 pairs, and gives a correlation of ·5 ; while the second of 28 pairs, has a correlation of ·8. Are these values significantly different.

Solution :

On the hypothesis that they are from the same normal population, the standard error of the difference of the Z's is

$$\varepsilon=\sqrt{\frac{1}{20}+\frac{1}{25}}$$
$$=\sqrt{{\cdot}09}$$
$$={\cdot}3$$

Now $$Z_1=\tfrac{1}{2}\log_e\frac{1+r_1}{1-r_1}$$
$$=1{\cdot}1513\times\log_{10}3$$
$$={\cdot}549$$
$$Z_2=1{\cdot}1513\log_{10}\frac{1+r_2}{1-r_2}$$
$$=1{\cdot}1513\times{\cdot}9542$$
$$=1{\cdot}098$$

$$\therefore\ |Z_1-Z_2|={\cdot}549$$

Which is little less than 2ε, and is therefore not quite significant at the 5% level. The hypothesis is not discredited.

CHAPTER XIV

PROBABILITY

Definition.—If an event can happen in 'a' ways and fail in 'b' ways, and all these ways are equally likely to occur, then the probability of its happening is $\frac{a}{a+b}$ and the probability of its failing is $\frac{b}{a+b}$.

Thus, if it be known that, in the long run, out of every 41 children born, there are 21 boys and 20 girls, the probability of any particular birth being that of a boy is $\frac{21}{41}$.

Note 1.—It follows at once from the definition of probability that if p be the probability that any event should occur, $1-p$ will be the probability of its failing to occur.

Note 2.—If the odds are in favour of the events as a to b (or b to a against it) then chance of its happening $=\frac{a}{a+b}$.

1. Mutually Exclusive Events :

Definition.—Events are called exclusive, if, when one of them does, no other one can happen. For example if we throw an ordinary cubical die. Obviously any of the six faces may be uppermost when the die comes to rest. Total number of possible ways is six for a single throw. The different cases are mutually exclusive, since no two faces can be uppermost at the same time.

Theorem.—If an event can happen in more than one way, all ways being mutually exclusive, the probability of its happening at all is the sum of the probability of its happenings in the several ways. *i.e.,* $p=p_1+p_2+\ldots+p_n$

2. Independent Events :

Definition.—Events are said to be independent, if the occurrence of one does not affect the occurrence of any of the others. We can define it in the following manner also.

Two events are said to be mutually independent when the probability for either is the same whether the other happen or not.

Theorem.—The probability that two independent events should both happen is the product of the separate probabilities of their happenings. Thus the probability of the occurrence of two independent events whose respective probabilities are p_1 and p_2 is $p_1\times p_2$.

Cor. I. If p_1 and p_2 be the probabilities of two independent events, the chance that they will both fail is $(1-p_1)(1-p_2)$, the chance that the first happens and the second fails is $p_1\ (1-p_2)$, and the chance that the second happens and the first fails is $(1-p_1)p_2$.

Cor. II. If $p_1, p_2, \ldots\ldots p_n$ be the probabilities of n independent events, then the probability that they all happen will be $p_1 \times p_2 \ldots\ldots p_n$ and that they all fail $(1-p_1)(1-p_2)\ldots\ldots\ldots(1-p_n)$.

3. Dependent Events :

If two events are not independent, but the probability of the second is different when the first happens from what it is when the first fails, the theorems for independent events will still hold good provided that p_2 is the probability that the second event happens when the first is known to have happened.

4. Use of Binomial Expansion :

When the probability of the happening of an event in one trial is known, the probability of its happening exactly once, twice, three times......etc. in n trials can be at once written with the help of binomial expansion.

For, if p be the probability of the happening of the event, the probability of its failing is $1-p=q$. Then the probability of the event happening r times exactly in n trials is ${}^nC_r p^r q^{n-r}$. Thus, if $(p+q)^n$ be expanded by binomial theorem, the successive terms will be the probability of the happening of the event exactly n times, $n-1$ times, $n-2$ times, etc. in n trials.

Cor. I. The probability of the event happening at least r times in n trials is

$$p^n + {}^nC_1\, p^{n-1}q + {}^nC_2\, p^{n-2}q^2 + \ldots\ldots + {}^nC_r\, p^r q^{n-r}.$$

Cor. II. To find the most probable number of successes and failures in n trials it is only necessary to find the greatest term in the expansion of $(p+q)^n$.

5. Use of Multinomial Expansion :

If a die has r faces marked from 1 to r, such n dice are thrown, then the chance of getting a sum S is

$$p = \frac{\text{coefficient of } x^s \text{ in the expansion of } (x+x^2+\ldots\ldots+x^r)^n}{r^n}$$

6. Expectation :

The value of a given chance of obtaining a given sum of money is called the expectation.

Thus if $\frac{a}{a+b}$ is the chance of obtaining a sum of money M, then the expectation is $\frac{a}{a+b}$ M.

7. Inverse Probability :

If $P_1, P_2, \ldots\ldots P_n$ be the probabilities of the existence of n causes, which are mutually exclusive and are such that a certain event must have followed from one of them, and let $p_1, p_2, \ldots p_n$ be the respective probabilities that when one of the causes $P_1, P_2, \ldots\ldots P_n$ exists it will be followed by the event in question ; then on any occasion when the event is known to have occurred the probability of the r^{th} cause is

$$P = \frac{P_r\, p_r}{P_1 p_1 + P_2 p_2 + \ldots\ldots P_n p_n}.$$

Examples

Problem 402.—6 cards are chosen at random from a pack of 52 ; what is the probability that 3 will be black and 3 red ?

Solution :

6 cards can be selected from 52 cards in $^{52}C_6$ ways *i. e.*, $\frac{|52}{|6\ |46}$ ways.

Now black cards are 26 and out of these three can be selected in $^{26}C_3$ ways ; similarly 3 red cards can be selected in $^{26}C_3$ ways.

$$\text{Now required chance} = \frac{\text{No. of favourable ways}}{\text{total No. of ways}}$$

$$= \frac{^{26}C_3 \times {^{26}C_3}}{^{52}C_6}$$

$$= \frac{|26 \times |26 \times |6 \times |46}{(|3\ |23)^2\ |52}.$$

Problem 401.—One of a pack of 52 cards having been removed, from the remainder of the pack, two cards are drawn and found to be spades. Find the chance that the missing card is a spade.

Solution :

The missing card may be, and is equally likely to be, any one of the pack except the two drawn, *i.e.*, any one of 50 cards. Hence total number of ways are 50. Also there are 11 remaining cards of spade from which the missing card may be. Hence number of favourable ways are 11. Therefore required chance is $\frac{11}{50}$.

Problem 402.—From a set of 19 cards, numbered 1, 2, 3...... ..., 19, one is, drawn at random. Find the chance that its number is divisible by 3 or 7.

Solution :

Total number of ways in which 1 card may be selected out of 19 are $^{19}C_1$ or 19.

There are 6 numbers (*i.e.*, 3, 6, 9, 12, 15, 18) which are divisible by 3 and two numbers (*i.e.* 7 and 14) which are divisible by 7. Hence out of 8 cards we have to select the 1 card so that its number may be divisible by 3 or 7.

$\therefore$ Number of favourable ways $= {}^8C_1 = 8$

$\therefore$ required chance $= \dfrac{8}{19}$.

Problem 403.—From a pack of 52 cards two are drawn at random. Find the chance that one is a king and the other a queen.

(*Agra, M. Sc., 1957*)

Solution :

Two cards can be drawn out of 52 in $^{52}C_2$ ways.

A king can be drawn in $^4C_1$ ways or 4 ways.

A queen can be drawn in $^4C_1$ ways or 4 ways.

$\therefore$ A king and a queen can be drawn in 4×4 ways.

$$\therefore \quad \text{required chance} = \frac{4 \times 4}{^{52}C_2} = \frac{4 \times 4}{26 \times 51}.$$

Problem 404.—If we draw 4 cards out of a pack, what is the chance that they each belong to a different suit ?

Solution :

4 cards out of 52 can be selected in $^{52}C_4$ ways, hence total number of ways are $^{52}C_4$.

There are 4 suits each containing 13 cards, hence number of ways in which one card can be selected from 1 suit is $^{13}C_1$ or 13.

Similarly for other three cards of different suits number of ways are 13 in each case.

Hence total favourable ways $= 13 \times 13 \times 13 \times 13$

$$\therefore \quad \text{required chance} = \frac{13 \times 13 \times 13 \times 13}{^{52}C_4}$$

$$= \frac{13^4 \times 4 \times 3 \times 2 \times 1}{52 \times 51 \times 50 \times 49}.$$

$$= \frac{13^3}{17 \times 25 \times 49}.$$

Problem 405.—Find the chance of drawing a king, a queen and a knave in that order from a pack of 52 cards in three consecutive draws, the cards drawn not being replaced.

Solution :

The chance of drawing a king is $p_1 = \frac{4}{52}$

The chance of drawing a queen is $p_2 = \frac{4}{51}$

The chance of drawing a knave is $p_3 = \frac{4}{50}$

Since these all dependent events

$\therefore$ required chance $= p_1 p_2 p_3 = \frac{4^3}{52 \times 51 \times 50}$.

Problem 406.—What is the probability that a specified player will get a hand containing 13 cards of one suit at a single deal at a game of bridge.

Solution :

There are 52 cards and 13 can be chosen from them in ${}^{52}C_{13}$ ways of these ways only four will contain cards of one suit.

Hence required probability $= \frac{4}{{}^{52}C_{13}}$

$$= \frac{4 \times 13! \times 39!}{52!}$$

Problem 407.—A pack of cards is separated into four packets, *viz.*, 13 hearts, 13 spades, 13 clubs, 13 diamonds. What is the chance of drawing the ace of clubs ?

Solution :

The chance of drawing a card from the packet of clubs is $\frac{1}{4}$.

Then the chance of drawing ace of clubs from the packet of clubs is $\frac{1}{13}$.

Hence required chance is $\frac{1}{4} \times \frac{1}{13}$

$$= \frac{1}{52}.$$

Problem 408.—What is the probability of getting 9 cards of the same suit in one hand at a game of bridge ?

(I. A. S., 1951)

Solution :

Since each player should have 13 cards hence total number of ways of selected 13 cards out of 52 cards is $^{52}C_{13}$.

There are four suits and each suit contains 13 cards hence number of ways of selected 9 cards of same suit is $^{4}C_1 \times {}^{13}C_9$.

After neglecting this vary suit there remains only 39 cards hence the remaining 4 cards he can select in $^{39}C_4$ ways.

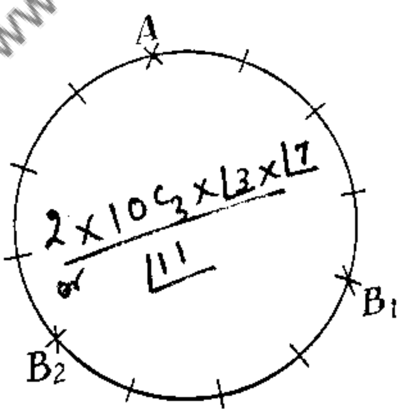
Hence favourable ways of selecting only 9 cards of same suit are $^{4}C_1 \times {}^{13}C_9 \times {}^{39}C_4$.

$$\therefore \text{ Required chance} = \frac{^{4}C_1 \times {}^{13}C_9 \times {}^{39}C_4}{^{52}C_{13}}$$

Problem 409.—A and B stand in a ring with ten other persons. If the arrangement of 12 persons is at random, find the chance that there are exactly three persons between A and B.

(M. Sc. Agra, 1951)

Solution :



Let A occupy one of the 12 seats ; then B can occupy either seat B_1 or seat B_2. Hence there are only 2 favourable ways for B to occupy a seat such that there are exactly 3 persons between A and B. In all B can occupy any of 11 seats hence only 11 ways.

Hence the required chance

$$= \frac{2}{11}.$$

Problem 410.—An experiment succeeds twice as often as it fails. Find the chance that in the next six trials there will be at least four successes.

(M. Sc. Agra, 1950)

Solution :

Chance of getting a success in one trial is $p = \frac{2}{3}$

$\therefore \quad q = \frac{1}{3}$.

Therefore with the help of binomial expression $(\frac{2}{3}+\frac{1}{3})^6$ we have the chance of at least 4 successes in 6 trials is

$$=(\tfrac{2}{3})^6+{}^6C_1(\tfrac{2}{3})^5(\tfrac{1}{3})+{}^6C_2(\tfrac{2}{3})^4(\tfrac{1}{3})^2$$

$$=\frac{64+192+240}{729}=\frac{496}{729}.$$

Problem 411.—Find the chance of throwing an odd number with an ordinary die.

Solution :

There are three odd numbers (*i.e.* 1, 3, 5) hence in 3 ways we can get an odd number. But to select one number out of 6 we have total 6 ways.

Hence required chance $=\frac{3}{6}=\frac{1}{2}$.

Problem 412.—Find the probability of throwing one 6 at least in six throws with a die.

Solution :

The probability of not throwing 6 is $\frac{5}{6}$ in each throw. Hence the probability of not throwing a 6 in six throws is $(\frac{5}{6})^6$; and therefore the probability of throwing one six at least is

$$1-\left(\frac{5}{6}\right)^6.$$

Problem 413.—Find the chance of throwing 10 with 4 dice.

Solution :

The whole number of different throws is 6^4, for any one of six numbers can be exposed on each die; also the number of ways of throwing 10 in the coefficient of x^{10} in $(x+x^2+x^3+x^4+x^5+x^6)^4$, for this coefficient gives the number of ways in which 10 can be made up by the addition of four of the numbers 1, 2, 3, 4, 5, 6, repetitions being allowed.

Now the coefficient of x^{10} in $(x+x^2+x^3+x^4+x^5+x^6)^4$

$=$coefficient of x^{10} in $x^4\left(\frac{1-x^6}{1-x}\right)^4$

$=$coefficient of x^{10} in $x^4\,[1-4x^6+6x^{12}\ldots\ldots][1+4x+10x^2+\ldots\ldots+\frac{4\times5\times6\times7\times8\times9}{1\times2\times3\times4\times5\times6}x^6+\ldots\ldots]$

$=$coefficient of x^{10} in $x^4\,[1-4x^6+6x^{12}\ldots][1+4x+\ldots+84x^6+\ldots]$

$=84-4=80$

Hence the required chance $=\frac{80}{6^4}=\frac{5}{81}$.

Problem 414.—What is the chance of two aces at least being thrown in 3 throws with two dice.

Solution :

This is the same as throwing 6 times with one die, and then at any one throw, the chance of one ace is $\frac{1}{6}$, and of no ace is $\frac{5}{6}$.

$\therefore$ The chance that no ace is thrown in 6 times is $\left(\frac{5}{6}\right)^6$;

„ „ only one ace „ is $6\left(\frac{5}{6}\right)^5 \times \frac{1}{6}$;

$\therefore$ The chance that 2 aces are not thrown $= \frac{5^6 + 5^5 \times 6}{6^6}$

$$= \frac{11 \times 5^5}{6^6}$$

$\therefore$ The chance that two aces at least are thrown is

$$1 - \frac{11 \times 5^5}{6^6}.$$

Problem 415.—2 dice are thrown ; what is the probability that the sum shown will be 7 or 11.

Solution :

The sum 7 can be shown in 6 different ways (*i.e.* 1+6, 2+5, 3+4, 4+3, 5+2, 6+1) and the sum 11 can be shown in two ways (*i.e.* 5+6, 6+5) hence total favourable ways=8

Total number of ways $= 6^2 = 36$

Required chance $= \frac{8}{36} = \frac{2}{9}$.

Problem 416.—A die is thrown 12 times ; what is the probability that the face 4, will appear just twice.

Solution :

The chance of getting face marked 4 once with one die in one trial is $\frac{1}{6}$ and not getting is $\frac{5}{6}$. Hence with the help of binomial expansion $\left(\frac{1}{6} + \frac{5}{6}\right)^{12}$ the chance of getting the face marked 4 twice in 12 trials $= {}^{12}C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{10}$

$$= \frac{12 \times 11}{2 \times 1}\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^{10}$$

$$= \frac{11 \times 5^{10}}{6^{11}}.$$

Problem 417.—Compare the chances of throwing 4 with one die, 8 with two dice and 12 with three dice.

Solution :

(*a*) 4 with one die can be thrown in one way only. There are 6 ways in all in which the die can be thrown. Hence the chance of getting 4 with one die $=\frac{1}{6}$

(*b*) 8 with two dice can be made of

6, 2 which can be achieved in $|\underline{2}$ ways or 2 ways

5, 3 which can be achieved in $|\underline{2}$ ways or 2 ways

4, 4 which can be achieved in 1 way

hence total ways of getting 8 with 2 dice are 5 and total number of ways in which two dice can be thrown are $6^2=36$

Hence chance of getting 8 with two dice $=\frac{5}{36}$.

(*c*) 12 with three dice can be made up of

6, 5, 1 which can be achieved in $|\underline{3}$ or 6 ways

6, 4, 2 „ „ „ $|\underline{3}$ or 6 ways

6, 3, 3 „ „ „ $\frac{|\underline{3}}{|\underline{2}}$ or 3 ways

5, 5, 2 „ „ „ $\frac{|\underline{3}}{|\underline{2}}$ or 3 ways

5, 4, 3 „ „ „ $|\underline{3}$ or 6 ways

4, 4, 4 „ „ „ 1 way

∴ Total ways of getting 12 with three dice are

$$6+6+3+3+6+1=25$$

∴ Chance of getting 12 is $\frac{25}{6^3}=\frac{25}{216}$

∴ Ratio of chances is $\frac{1}{6}:\frac{5}{36}:\frac{25}{216}$

or $36 : 30 : 25.$

Problem 418.—How many throws with a single die must a man have, in order that his chance of throwing an ace may be $\frac{1}{2}$?

Solution :

Let x be the required number of throws. His chance of not throwing an ace at one throw is $\frac{5}{6}$. Therefore his chance of not throwing an ace at all in the x throws is $\left(\frac{5}{6}\right)^x$.

$\therefore$ his chance of throwing an ace is $1-\left(\frac{5}{6}\right)^x$

$$\therefore \quad 1-\left(\frac{5}{6}\right)^x=\frac{1}{2}$$

$$\text{or} \quad \left(\frac{5}{6}\right)^x=\frac{1}{2}$$

$$\therefore \quad x=\frac{\log 2}{\log 6-\log 5}$$

$$=3{\cdot}8 \text{ (from log tables)}$$

This shows that there is no exact number of throws in which his chance will amount to $\frac{1}{2}$, but, if he throws 4 times, his chance will be a little greater than $\frac{1}{2}$.

Problem 419.—Find the chance of throwing 7 at least with two dice in a single throw.

Solution :

7 can be thrown with two dice in 6 ways hence chance of getting 7 with two dice is $\frac{6}{36}$

8 can be thrown with two dice in 5 ways hence chance of getting 8 is $\frac{5}{36}$

9 can be thrown in 4 ways hence chance of getting 9 is $\frac{4}{36}$

similarly chance of getting 10 is $\frac{3}{36}$.

chance of getting 11 is $\frac{2}{36}$

chance of getting 12 is $\frac{1}{36}$

Since these events are mutually exclusive therefore the cha[nce] of throwing 7 at least with two dice in one throw

$$=\frac{6}{36}+\frac{5}{36}+\frac{4}{36}+\frac{3}{36}+\frac{2}{36}+\frac{1}{36}$$

$$=\frac{21}{36}=\frac{7}{12}.$$

Problem 420.—Find the chance of throwing 15 with th[ree] dice.

Solution :

15 with three dice can be made up of

6, 6, 3 which can be achieved in $\frac{\lfloor 3}{\lfloor 2}$ or 3 ways

6, 5, 4 which can be achived in $\lfloor 3$ or 6 ways

5, 5, 5 which can be achieved in 1 way

∴ Total ways of getting 15 with three dice $=3+6+1=10$

∴ required chance $=\frac{10}{6\times6\times6}=\frac{5}{108}.$

Problem 421.—A throws 3 coins, B throws two ; what is th[e] chance that A will throw a greater number of heads than B.

Solution :

Since A is not to throw as many or more heads, but actually a greater number, this can be done in three mutually exclusive ways. We give them, with their chances.

(*i*) A throws three heads chance is $\left(\frac{1}{2}\right)^3=\frac{1}{8}$

(*ii*) A throws 2 heads, B does not chance is

$${}^3C_1\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)\times\left(1-\frac{1}{4}\right)=\frac{9}{32}$$

(*iii*) A throws 1 head, B throws 2 tails $={}^3C_2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2\times\frac{1}{4}$

$$=\frac{3}{32}$$

In case of A we took help of binomial expansion $\left(\frac{1}{2}+\frac{1}{2}\right)^3$ and in case of B the expansion $\left(\frac{1}{2}+\frac{1}{2}\right)^2$

∴ Total Probability $=\frac{1}{8}+\frac{9}{32}+\frac{3}{32}=\frac{16}{32}=\frac{1}{2}.$

✓**Problem 422.**—In two bags there are to be put altogether 2 red and 10 white balls, neither bag being empty. How must the balls be divided so as to give to a person who draws one ball from either bag, (1) the least chance and (2) the greatest chance of drawing a red ball.

Solution :

The least chance is when one bag contains only one white ball, and the greatest chance is when one bag contains only one red ball, the chances being $\frac{1}{11}$ and $\frac{6}{11}$ respectively.

✓**Problem 423.**—A card is drawn at random from a pack and replaced, then a second drawing is made, and so on. How many drawings must be made in order to have a chance of $\frac{1}{2}$ that the ace of spades shall appear at least once.

Solution :

It is assumed that the cards are properly suffled after each drawing. The different drawings are, thus, independent events, with the same probabilities each time. The chance that the ace of spades will never appear in n drawings is $\left(\frac{51}{52}\right)^n$. Therefore the chance that ace of spades will appear at least once in n trials is

$$1-\left(\frac{51}{52}\right)^n$$

$$\therefore \quad 1-\left(\frac{51}{52}\right)^n=\frac{1}{2}$$

$$\left(\frac{51}{52}\right)^n=\frac{1}{2}$$

$$\therefore \quad n=\frac{\log 2}{\log 52-\log 51}$$

$$=36.$$

✓**Problem 424.**—Find the chance of drawing a red ball from a bag which contains 5 white and 7 red balls.

Solution :

Out of total 12 balls one ball can be taken out in 12 ways; but there are only 7 red balls hence number of favourable ways are 7. Therefore the required probability $=\frac{7}{12}$.

✓**Problem 425.**—Two balls are to be drawn from a bag containing 7 red and 11 white balls, find the chance that they will both be white.

Solution :

Out of total 18 balls two can be drawn in $18C_2$ ways. Total white balls are 11 hence favourable ways are ${}^{11}C_2$.

$$\therefore \text{ the required chance} = \frac{{}^{11}C_2}{{}^{18}C_2}$$

$$= \frac{11\times 10}{18\times 17}$$

$$= \frac{55}{133}$$

Problem 426.—Find the chance of drawing 2 white balls in succession from a bag containing 11 red and 13 white balls, the balls drawn not being replaced.

Solution :

The chance of drawing a white ball first time is $\frac{13}{24}$; and having drawn a white ball the first time, there will be 11 red balls and 12 white balls left, and therefore the chance of drawing a white ball the second time will be $\frac{12}{23}$. These events are dependent hence the probability of drawing white balls in succession will be $\frac{13}{24}\times\frac{12}{23}$

$$= \frac{13}{46}.$$

Problem 427.—There are two bags, one of which contains 5 red and 7 white balls and the other 3 red and 12 white balls, and a ball is to be drawn (from one or other) of the two bags ; find the chance of drawing a red ball.

Solution :

The chance of choosing the first bag is $\frac{1}{2}$, and if the first bag be chosen the chance of drawing a red ball from it is $\frac{5}{12}$; hence the chance of drawing a red ball from the first bag is $\frac{1}{2}\times\frac{5}{12}=\frac{5}{24}$. Similarly the chance of drawing a red ball from the second bag is $\frac{1}{2}\times\frac{3}{15}=\frac{1}{10}$. Hence, as these events are mutually exclusive, the chance required is $\frac{5}{24}+\frac{1}{10}=\frac{37}{120}$.

Problem 428.—If a bag contains 3 white and 7 black balls, and a ball be drawn and always replaced, show that the most likely result in 10 trials is to draw 3 white and 7 black balls.

Solution :

The chance of a white occurring at any one drawing is $\frac{3}{10}$.

The chance of a black occurring at any one drawing is $\frac{7}{10}$.

Hence we have to find the greatest term in the expansion of $(3+7)^{10}$.

Now $T_{r+1} \gtrless T_r$ in expansion of $(x+a)^n$,

accordingly as $\frac{T_{r+1}}{T_r} \gtrless 1$

Or $\frac{n-r+1}{r} \cdot \frac{a}{x} \gtrless 1$

Or $r \lessgtr \frac{(n+1)a}{x+a}$

In this case $r \lessgtr \frac{(10+1)7}{7+3}$

$r \lessgtr \frac{77}{10}$

∴ 8th term is the greatest term, and to this term corresponds the chance of a combination in which 7 black and 3 white balls occur.

Problem 429.—There are 5 white and 7 red balls in a bag, what is the chance that a white ball is drawn and then a red, the first ball not being put back ?

Solution :

The chance of drawing a white ball first is $\frac{5}{12}$.

The chance of drawing a red ball afterwards is $\frac{7}{11}$.

These events are dependent, hence the total probability is $\frac{5}{12} \times \frac{7}{11} = \frac{35}{132}$.

Problem 430.—A bag contains 3 red, 4 black and 2 white balls. What is the chance of drawing a red and a white balls, each ball being put back after it is drawn.

1s. = 12d. 1 crown = 5s.

Solution :

Case I—If first ball drawn is white and second is red.

The chance is $\frac{2}{9} \times \frac{3}{9} = \frac{2}{27}$

Case II—If first ball drawn is red and second white

The chance is $\frac{3}{9} \times \frac{2}{9} = \frac{2}{27}$

These two ways are mutually exclusive, hence required probability is $\frac{2}{27} + \frac{2}{27} = \frac{4}{27}$.

Problem 431.—A bag contains 5 white balls and 7 black balls. Find the expectation of a man who is allowed to draw a ball from the bag and who is to receive one shilling if he draws a black ball, and a crown if he draws a white one.

Solution :

The chance of drawing a black ball is $\frac{7}{12}$, and therefore the expectation from drawing a black ball is 7*d*. The chance of drawing a white ball is $\frac{5}{12}$; and therefore the expectation from drawing a white ball is 2*s*. 1*d*. Hence, as these events are exclusive, the whole expectation is 2*s*. 8*d*.

Problem 432.—There are three urns, of which, one contains 4 white and 2 red, another 5 white and 1 red, the third 3 white and 4 red balls; and there is no general reason why one urn should be selected more than another, or one ball from an urn more than any other ball from that urn. Find the chance of a red ball being drawn.

Solution :

Since all the urns are equally likely to be drawn from, therefore of any number of drawings one-third will be made from each of the urns.

The chance that red ball is drawn from first urn is

$$\frac{2}{6} \times \frac{1}{3} = \frac{1}{9}$$

The chance that red ball is drawn from second urn is

$$\frac{1}{6} \times \frac{1}{3} = \frac{1}{18}$$

The chance that red ball is drawn from third urn is

$$\frac{4}{7} \times \frac{1}{3} = \frac{4}{21}$$

$\therefore$ required chance $= \frac{1}{9} + \frac{1}{18} + \frac{4}{21}$

$= \frac{45}{126} = \frac{5}{14}$.

Problem 433.—A bag contains 6 white and 9 black balls. Two drawings of four balls are made such that (*a*) the balls are replaced before the second trial (*b*) the balls are not replaced before the second trial.

Find the probability that the first drawing will give 4 white and the second 4 black balls in each case. (*I.A.S., 1945*)

Solution :

(*a*) The chance of drawing 4 white balls in first drawing

$= \frac{^6C_4}{^{15}C_4}$

The chance of drawing 4 black balls in second drawings $= \frac{^9C_4}{^{15}C_4}$ (since balls are replaced). These two events are independent.

Hence requird chance $= \frac{^6C_4}{^{15}C_4} \times \frac{^9C_4}{^{15}C_4}$

$= \frac{^6C_4 \times {^9C_4}}{(^{15}C_4)^2}$

(*b*) The chance of drawing 4 white balls in first drawing $= \frac{^6C_4}{^{15}C_4}$

Since balls are not replaced hence only 11 balls remain in the bag. Hence chance of drawing 4 black balls in second drawing

$= \frac{^9C_4}{^{11}C_4}$.

These are dependent events

Hence required chance $= \frac{^6C_4 \times {^9C_4}}{^{15}C_4 \times {^{11}C_4}}$.

Problem 434.—A and B draw from a log containing 2 white and 3 black balls, the ball being put back, after each drawing, until a white is drawn, what are their respective chances of drawing a white.

Solution :

The chance that A draws a white at the first time is $\frac{2}{5}$.

The chance that B has a drawing at all is the chance that A draws a black at the first time ; *i.e.* $\frac{3}{5}$.

The chance that having a drawing, B will draw a white, is $\frac{2}{5}$; therefore the chance of his drawing a white at the second drawing is $\frac{3}{5}\times\frac{2}{5}$.

The chance that A has a second drawing is the chance that A and B both draws black balls, which is $\left(\frac{3}{5}\right)^2$; therefore the chance that he draw a white at his second drawing is $\left(\frac{3}{5}\right)^2\times\frac{2}{5}$.

Hence A's various chances of drawing a white are

$$\frac{2}{5},\ \left(\frac{3}{5}\right)^2\times\frac{2}{5},\ \left(\frac{3}{5}\right)^4\times\frac{2}{5};\ \ldots\ldots\text{etc.}$$

and his drawing a white at any one time excludes the possibility of his doing so at any other ; his chance of a white ball is

$$\frac{2}{5}+\left(\frac{3}{5}\right)^2\times\frac{2}{5}+\left(\frac{3}{5}\right)^4\times\frac{2}{5}+\ldots\ldots\infty$$

$$=\frac{2}{5}\left[1+\left(\frac{3}{5}\right)^2+\left(\frac{3}{5}\right)^2+\ldots\ldots\infty\right]$$

$$=\frac{2}{5}\times\frac{1}{1-\left(\frac{3}{5}\right)^2}$$

$$=\frac{5}{8}.$$

Similarly B's chance is $\frac{3}{5}\times\frac{2}{5}+\left(\frac{3}{5}\right)^3\times\frac{2}{5}+\ldots\ldots\infty$

$$=\frac{3}{5}\times\frac{2}{5}\left[1+\left(\frac{3}{5}\right)^2+\left(\frac{3}{5}\right)^4+\ldots\ldots\infty\right]$$

$$=\frac{3}{5}\times\frac{2}{5}\times\frac{1}{1-\left(\frac{3}{5}\right)^2}$$

$$=\frac{3}{8}$$

Problem 435.—In the previous problem, what would be their respective chances, if a ball when drawn is not put back ?

Solution :

The chance of A drawing a white ball at first is $\frac{2}{5}$.

If B gets a drawing, there will be 2 white and 2 black balls, hence his chance then will be $\frac{1}{2}$;

$\therefore$ B's chance of a white at his first drawing is $\frac{3}{5} \times \frac{1}{2}$;

If A gets another drawing, there will be 2 white and 1 black, hence his chance then will be $\frac{2}{3}$. But his chance of getting another drawing is $\frac{3}{5} \cdot \frac{1}{2}$;

$\therefore$ A's chance of a white at his second drawing is $\frac{1}{5}$.

Similarly B's chance of a white at second drawing is

$$\frac{3}{5} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{10}.$$

Therefore A's chance of a white at either drawing is

$$\frac{2}{5} + \frac{1}{5} = \frac{3}{5}$$

(∵ only 2 w balls, ∴ only two draws).

and B's chance of a white at either drawing is

$$\frac{3}{10} + \frac{1}{10} = \frac{2}{5}.$$

Problem 436.—A and B throw with one die for a stake of Rs. 44/- which is to be won by the player who first throws 2. If A has the first throw, what are their respective expectations.

(*M. Sc., Agra, 1954, 1958*)

Solution :

A can win in the first, third, fifth......throw while B can win in second, fourth......throw.

The chance that A throws 2 in the first throw is $\frac{1}{6}$

The chance that B has a throw is $\left(1 - \frac{1}{6}\right)$ or $\frac{5}{6}$

The chance that having a throw B throws 2 is $\frac{1}{6}$

$\therefore$ The chance that B throws 2 at the second throw is $\frac{1}{6} \times \frac{5}{6}$

The chance that A has a second chance is $\left(\frac{5}{6}\right)^2$ therefore the chance that A wins in third throw is $\frac{1}{6} \times \left(\frac{5}{6}\right)^2$

$$\text{Hence A's chance} = \frac{1}{6} + \frac{1}{6} \times \left(\frac{5}{6}\right)^2 + \frac{1}{6} \times \left(\frac{5}{6}\right)^4 + \ldots\ldots$$

$$= \frac{\frac{1}{6}}{1-\left(\frac{5}{6}\right)^2}$$

$$= \frac{6}{11}$$

$$\text{B's chance} = \frac{5}{6} \times \frac{1}{6} + \left(\frac{5}{6}\right)^3 \times \frac{1}{6} + \left(\frac{5}{6}\right)^5 \times \frac{1}{6} + \ldots$$

$$= \frac{\frac{5}{6} \times \frac{1}{6}}{1-\left(\frac{5}{6}\right)^2}$$

$$= \frac{5}{11}$$

$$\therefore \quad \text{Expectation of A} = \frac{6}{11} \times 44 = 24/\text{-}$$

$$\text{Expectation of B} = \frac{5}{11} \times 44 = 20/\text{-}$$

Problem 437.—Out of $2n+1$ tickets consecutively numbered, three are drawn at random. Find the chance that the numbers on them are in arithmetical progression.

(M. Sc., Agra, 1950)

Solution :

Suppose the first selected number is 1, then the possible groups are :

1, 2, 3 ; 1, 3, 5 ; 1, 4, 7 ; 1, 5, 9 ; ....1, $n+1$, $2n+1$; giving n possible ways.

Similar grouping will show that when the lowest number selected are 2, 3, 4, 5,......., $2n-2$, $2n-1$, the number of favourable ways (*i.e.* such groups) are respectively $n-1$, $n-1$, $n-2$, $n-2$,...... 1, 1.

$\therefore$ Total number of favourable ways

$$= 2[1+2+3+\ldots\ldots+(n-1)]+n$$

$$= 2\frac{n-1}{2} \times n + n$$

$$= n^2$$

Total number of ways $= {}^{2n+1}C_3$

$$= \frac{(2n+1)\ 2n\ (2n-1)}{\lfloor 3}$$

$\therefore$ Required chance

$$= \frac{n^2 \times \lfloor 3}{(2n+1)\ 2n\ (2n-1)}$$

$$= \frac{3n}{4n^2-1}.$$

Problem 438.—What is the chance that a leap year, selected at random, will contain 53 Sundays?

(*M.Sc., Agra, 1955*)

Solution :

A leap year consists of 52 complete weeks and 2 days over. These two days can be,

(1) Monday & Tuesday
(2) Tuesday & Wednesday
(3) Wednesday & Thursday
(4) Thursday & Friday
(5) Friday & Saturday
(6) Saturday & Sunday
(7) Sunday & Monday

of these 7 likely cases only 2 are favourable (*i.e.* 6th and 7th cases)

Hence required chance $= \frac{2}{7}$

Problem 439.—Three groups of children contain respectively 3 girls and 1 boy; 2 girls and 2 boys; 1 girl and 3 boys. One child is selected at random from each group. Show that the chance that the three selected consist of 1 girl and 2 boys is $\frac{13}{32}$.

(*M.Sc., Agra, 1955*)

Solution :

This can be done in following three mutually exclusive ways: —

(1) girl from first, boy from second, boy from third chance $= \frac{3}{4} \times \frac{1}{2} \times \frac{3}{4}$

(2) boy from first, girl from second, boy from third chance $= \frac{1}{4} \times \frac{1}{2} \times \frac{3}{4}$

(3) boy from first, boy from second, girl from third; chance $= \frac{1}{4} \times \frac{1}{2} \times \frac{1}{4}$

$$\therefore \quad \text{Total chance} = \frac{9}{32} + \frac{3}{32} + \frac{1}{32}$$
$$= \frac{13}{32}.$$

√**Problem 440.**—Eight mice are selected at random from a large number and then divided into two groups of four each—group A and group B. Each mouse in group A is given a dose 'a' of a certain poison which is expected to kill one in four. Each mouse in group B is given a dose '*b*' of another poison which is expected to kill one in two. Show that, nevertheless, there may be fewer deaths in group B than in group A and find the probability of the happening.

(*I.A.S. 1952*)

Solution :

Favourable cases are

	No. of deaths in B	*No. of deaths in A*
(1)	0	1, 2, 3 or 4
(2)	1	2, 3 or 4
(3)	2	3 or 4
(4)	3	4

chances of 0, 1, 2, ... deaths in groups A and B are given respectively by the different terms of the binomial expansion $\left(\frac{3}{4}+\frac{1}{4}\right)^4$ and $\left(\frac{1}{2}+\frac{1}{2}\right)^4$

$$\therefore \quad \text{Chance for (1) case} = \left(\frac{1}{2}\right)^4 \times \left[{}^4C_1 \left(\frac{3}{4}\right)^3 \times \frac{1}{4} + {}^4C_2\left(\frac{3}{4}\right)^2 \times \left(\frac{1}{4}\right)^2 + {}^4C_3\left(\frac{3}{4}\right) \times \left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^4 \right]$$
$$= \frac{1}{16}\left[\frac{108}{256} + \frac{54}{256} + \frac{12}{256} + \frac{1}{256}\right]$$
$$= \frac{175}{16 \times 256}$$

$$\text{Chance for (2) case} = {}^4C_1 \left(\frac{1}{2}\right)^3 \times \frac{1}{2} \times \left[{}^4C_2\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 + {}^4C_3\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^4 \right]$$
$$= \frac{4}{16}\left[\frac{54}{256} + \frac{12}{256} + \frac{1}{256}\right]$$
$$= \frac{268}{16 \times 256}$$

Chance for (3) case $= {}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2\times\left[\,{}^4C_3\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)\right]$

$$= \frac{6}{16}\left[\frac{12}{256}+\frac{1}{256}\right]$$

$$= \frac{78}{16\times 256}$$

Chance for (4) case $= {}^4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)\left[\left(\frac{1}{4}\right)^4\right]$

$$= \frac{4}{16\times 256}$$

These ways are mutually exclusive hence required chance

$$= \frac{175+268+78+4}{16\times 256}$$

$$= \frac{525}{16\times 256}$$

$$= \frac{525}{4096}.$$

Problem 441.—A and B cast each with a pair of ordinary dice for a stake of £61. A wins if he throws 6 before B throw 7, and B wins if he throws 7 before A throws 6. Find their respective expectations if they cast alternately, A beginning.

(*M.Sc. Agra, 1945*)

Solution :

A can throw 6 (4+2 or 2+4 ; 5+1 or 1+5 ; 3+3) with two dice in 5 different ways hence his chance of throwing 6 in first trial is $\frac{5}{36}$

If A fails B will get chance. Hence B's chance of getting a throw is $1-\frac{5}{36}$ or $\frac{31}{36}$.

If B gets a chance he can throw 7 with two dice (4+3 or 3+4; 5+2 or 2+5 ; 6+1 or 1+6 ;) in 6 different ways.

Hence B's chance of winning if he gets a chance at all is

$$\frac{6}{36} \text{ or } \frac{1}{6}$$

Hence B's chance of winning is $\frac{31}{36}\times\frac{1}{6}$

If B also fails A will get again chance hence A's chance of getting second trial is $\frac{31}{36} \times \frac{5}{6}$. Therefore the chance that A wins in second trial is $\frac{5}{36} \times \frac{31}{36} \times \frac{5}{6}$.

If A again fails B will get a chance hence we can prove in similar way that B's chance of winning is $\left(\frac{31}{36}\right)^2 \times \frac{5}{6} \times \frac{1}{6}$ and so on.

$$\therefore \quad \text{Total Chance of A} = \frac{5}{36} + \frac{5}{36} \times \frac{31}{36} \times \frac{5}{6} + \frac{5}{36} \times \left(\frac{31}{36}\right)^2 \times \left(\frac{5}{6}\right)^2 + \ldots$$

$$= \frac{\frac{5}{36}}{1 - \frac{31}{36} \times \frac{5}{6}}$$

$$= \frac{5 \times 6}{61}$$

$$= \frac{30}{61}$$

$$\text{Total chance of B} = \frac{31}{36} \times \frac{1}{6} + \left(\frac{31}{36}\right)^2 \times \frac{5}{6} \times \frac{1}{6} + \left(\frac{31}{36}\right)^3 \times \left(\frac{5}{6}\right)^2 \times \frac{1}{6} + \ldots\ldots$$

$$= \frac{\frac{31}{36} \times \frac{1}{6}}{1 - \frac{31}{36} \times \frac{5}{6}}$$

$$= \frac{31}{61}$$

$$\therefore \quad \text{Expectation of A} = \frac{30}{61} \times 61 = 30£$$

$$\text{Expectation of B} = \frac{31}{61} \times 61 = 31£$$

Problem 442.—The odds against a certain event are 5 to 2 and the odds in favour of another (independent) event are 6 to 5; find the chance that one at least of the events will happen.

(M.Sc. Agra, 1945)

Solution :

Let p_1 be the probability that first event will happen, then $p_1=\frac{2}{7}$; hence $q_1=1-\frac{2}{7}=\frac{5}{7}$.

Similarly if p_2 is the chance that the second event happens then $p_2=\frac{6}{11}$; $q_2=\frac{5}{11}$.

The probability that none of the events happen is

$$q_1q_2=\frac{5}{7}\times\frac{5}{11}=\frac{25}{77}$$

Hence the probability that at least one event happens

$$=1-\frac{25}{77}=\frac{52}{77}.$$

Problem 443.—A, B, C, D cut a pack of cards successively in the order mentioned. If the person who cuts a spade first receives £ 175, what are their expectations.

(M.Sc., Agra, 1944)

Solution :

The chance of cutting a spade $=\frac{13}{52}=\frac{1}{4}$

A will get first, fifth, ninth,.........trials

Hence chance of $A=\frac{1}{4}+\frac{1}{4}\left(\frac{3}{4}\right)^4+\frac{1}{4}\left(\frac{3}{4}\right)^8+\ldots\ldots$

$$=\frac{\frac{1}{4}}{1-\left(\frac{3}{4}\right)^4}=\frac{64}{175}$$

B will get second, sixth, tenth,......trials

$\therefore$ chance of $B=\frac{3}{4}\times\frac{1}{4}+\left(\frac{3}{4}\right)^5\times\frac{1}{4}+\ldots\ldots$

$$=\frac{\frac{3}{4}\times\frac{1}{4}}{1-\left(\frac{3}{4}\right)^4}=\frac{48}{175}$$

C will get third, seventh......trials

$\therefore$ chance of $C=\left(\frac{3}{4}\right)^2\times\frac{1}{4}+\left(\frac{3}{4}\right)\times\frac{1}{4}+\ldots\ldots$

$$=\frac{\left(\frac{3}{4}\right)^2\times\frac{1}{4}}{1-\left(\frac{3}{4}\right)^4}=\frac{36}{175}$$

$\therefore$ D's chance $= 1 - \frac{64}{175} - \frac{48}{175} - \frac{36}{175}$

$= \frac{27}{175}$

$\therefore$ Expectation of A $= \frac{64}{175} \times 175 = 64$ £

Expectation of B $= \frac{48}{175} \times 175 = 48$ £

Expectation of C $= \frac{36}{175} \times 175 = 36$ £

Expectation of D $= \frac{27}{175} \times 175 = 27$ £

Problem 444.—Peter and Paul play a game with two dice. Peter plays first by throwing the dice together. If the total number of points is a prime number other than 2 he wins outright ; if it is even he throws again under the same conditions ; in other cases the throw passes to Paul, who throws under the same conditions. What is the probability of Peter's winning ?

(M.Sc., Agra, 1947)
(From Kandall Vol. I, Page 169)

Solution :

It is to be assumed that the probabilities of throwing any number 1 to 6 with either die are equal. The possible throws are 2, 3, 4,.........12 and the number of ways in which they can occur are :—

Total points	2	3	4	5	6	7	8	9	10	11	12
No. of ways	1	2	3	4	5	6	5	4	3	2	1

Thus the probability

(1) of throwing a prime other than 2 (*i.e.* 3, 5, 7, 11) is $\frac{14}{36}$,

(2) of throwing an even number (*i.e.* 2, 4, 6, 8, 10, 12) is $\frac{18}{36}$,

(3) of throwing neither *i.e.* throwing 9 is $\frac{4}{36}$.

These three events are mutually exclusive. Let P be the probability of Peters winning. Now if Peter throws a prime other than 2 he wins outright, and the probability of his doing so is thus $\frac{14}{36}$; if he throws an even number he throws again, and his probability of winning in this case is $\frac{18}{36}$ P ; if he throws neither, the throw passes to Paul, whose chance is then P, so that Peter's chance of winning is $\frac{4}{36}(1-P)$.

Thus we have

$$P=\frac{14}{36}+\frac{18}{36}P+\frac{4}{36}(1-P)$$

$$\text{giving } P=\frac{18}{22}=\frac{9}{11}.$$

Problem 445.—Eight letters to each of which corresponds an envelope, are placed in the envelopes at random. What is the probability that all letters are not placed in the right envelopes?

(M.Sc., Agra, 1947)

Solution

Eight letters in eight envelopes can be put in $|\underline{8}$ ways in which there will be only one way for posting the letters correctly.

Hence required chance $=\frac{1}{|\underline{8}}$.

Problem 446.—Goddard, the captain of West Indies cricket team is reported to have observed the rule of calling "heads" every time the toss was made during the five matches of the last test-series with the Indian team, what is the probability of his winning the toss in all the five matches?

How will the probability be affected if he had made a rule of tossing a coin privately to decide whether to call 'heads' or 'tails' on each occasion.

(I. A. S., 1950)

Solution :

The chance of winning the toss in one match is $\frac{1}{2}$.

Since the result of one match does not effect on the tossing of other hence these events are independent. Therefore the chance of winning the toss in all the five matches $=\left(\frac{1}{2}\right)^5$

$$=\frac{1}{32}$$

These events are mutually independent hence tossing privately will not effect the result.

Problem 447.—p is the probability that a man aged x will die in a year. Find the probability that out of 5 men A, B, C, D, and L each aged x, A will die in the year and be the first to die.

(I. A. S., 1954)

Solution :

p is the probability that a man die in a year. Hence the chance of failure $=1-p$.

The chance that no man die out of 5 men A, B, C, D, L is $(1-p)^5$.

Therefore chance that at least one man die out of 5 is $1-(1-p)^5$.

Thus the probability that A is first to die $= \frac{1}{5}\,[1-(1-p)^5]$.

Problem 448.—A, B and C in order toss a coin. The first one to throw a head wins. What are their respective chances of winning. Assume that the game may continue indefinitely.

(I.A.S., 1955)

Solution :

The chance of throwing head $= \frac{1}{2}$

A can win in first, fourth, seventh,......trials
B can win in second, fifth, eighth,......trials
C can win in third, sixth, ninth,......trials.

$$\therefore \quad \text{chance of A} = \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2}\right)^3 + \frac{1}{2}\left(\frac{1}{2}\right)^6 + \cdots\cdots$$

$$= \frac{\frac{1}{2}}{1-\left(\frac{1}{2}\right)^3} = \frac{4}{7}$$

$$\text{chance of B} = \frac{1}{2}\cdot\left(\frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{2}\right)^4 + \frac{1}{2}\left(\frac{1}{2}\right)^7 + \cdots\cdots$$

$$= \frac{\frac{1}{4}}{1-\left(\frac{1}{2}\right)^3} = \frac{2}{7}$$

$$\text{chance of C} = \frac{1}{2}\cdot\left(\frac{1}{2}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right)^5 + \cdots\cdots$$

$$= \frac{\frac{1}{8}}{1-\left(\frac{1}{2}\right)^3} = \frac{1}{7}.$$

Problem 449.—If 4 whole numbers taken at random are multiplied together, show that the chance that the last digit in the product is 1, 3, 7 or 9 is $\frac{16}{625}$.

(M. Sc., Agra, 1948)

Solution :

Method I. The whole numbers can end with any of the 10 digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

The whole numbers having last digit an even number *i.e.* 2, 4, 6, 8 when multiplied together will end with the even numbers. Similarly the whole numbers, ending with 0 or 5 will not give 1, 3, 7 or 9 when multiplied together.

Hence we conclude that only those whole numbers which end with 1, 3, 7 or 9, when multiplied together will give the last digits 1, 3, 7 or 9.

Hence out of 10 digits the end place can be occupied by any of these 4 digits. The chance of which for one whole number is $\frac{4}{10}$ or $\frac{2}{5}$. Similar is case with other three whole numbers also

Hence the required chance $=\left(\frac{2}{5}\right)^4=\frac{16}{625}$.

Method II. Last digit can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

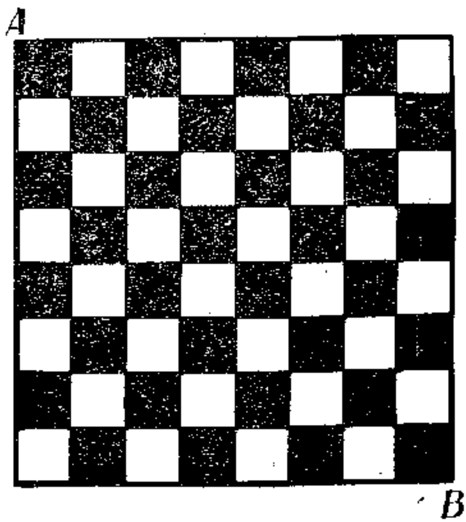
Chance that any of the four numbers is divisible by 2 or 5 is $\frac{6}{10}$ or $\frac{3}{5}$. Hence the chance that it is not divisible by 2 or 5 is $\frac{2}{5}$. Chance that all the four numbers are not divisible by 2 or 5 is $\left(\frac{2}{5}\right)^4$. This is the chance that the lastdig it in the product will not be 0, 5, 2, 4, 6, 8 that is the required chance.

Problem 450.—If three square are chosen at random on a chess board, find the chance that they should be in a diagonal line.

(M. Sc., Agra, 1946)

Solution :

There are 64 squares in a chess-board. The total number of ways in which 3 squares can be chosen out of 64 squares $={}^{64}C_3$.



Now we have to find the chance that they should be in a diagonal line.

Consider the diagonals parallel to diagonal AB. Diagonal AB contains 8 squares. On both sides of AB the number of squares in successive diagonals are 7, 6, 5, 4, 3, 2 and 1. Similar is case for diagonals perpendicular to AB.

Therefore the favourable ways *i.e.* the ways in which 3 squares are diagonally

$=2[{}^8C_3+2({}^7C_3+{}^6C_3+{}^5C_3+{}^4C_3+{}^3C_3)]$

$=2[56+2(35+20+10+4+1)]$

$=392$

$$\therefore \text{ required chance} = \frac{392}{^{64}C_3} = \frac{392 \times 3 \times 2}{64 \times 63 \times 62} = \frac{7}{744}.$$

Problem 451.—If m things are distributed among 'a' men and 'b' women. Show that the chance that the number of things received by men is odd is

$$\frac{1}{2} \cdot \frac{(b+a)^m - (b-a)^m}{(b+a)^m}$$

(*M.Sc., Agra, 1953*)

(B.Sc., 1957)

Solution :

Chance of getting the thing by men is $\frac{a}{a+b}$

Chance of getting the thing by women is $\frac{b}{a+b}$

If 'a' men get only one thing out of m and other things go to women then the chance for men is

$$^mC_1\left(\frac{a}{a+b}\right)\left(\frac{b}{a+b}\right)^{m-1}$$

If 'a' men get three things out of m then chance is

$$^mC_3\left(\frac{a}{a+b}\right)^3\left(\frac{b}{a+b}\right)^{m-3}$$

If 'a' men get five things out of m then chance is

$$^mC_5\left(\frac{a}{a+b}\right)^5\left(\frac{b}{a+b}\right)^{m-5}$$

and so on

Hence the chance that the number of things received by men is odd

$$= {^mC_1}\left(\frac{a}{a+b}\right)\left(\frac{b}{a+b}\right)^{m-1} + {^mC_3}\left(\frac{a}{a+b}\right)^3\left(\frac{b}{a+b}\right)^{m-3} + {^mC_5}\left(\frac{a}{a+b}\right)^5\left(\frac{b}{a+b}\right)^{m-5} + \ldots\ldots$$

$$= \frac{1}{(a+b)^m}\left[{^mC_1}ab^{m-1} + {^mC_3}a^3b^{m-3} + {^mC_5}a^5b^{m-5} + \ldots\ldots\right]$$

$$= \frac{1}{(a+b)^m}\left[\frac{1}{2}\left(b+a\right)^m - \frac{1}{2}\left(b-a\right)^m\right]$$

$$= \frac{1}{2} \cdot \frac{(b+a)^m - (b-a)^m}{(a+b)^m}.$$

Problem 452.—An ordinary six-sided die is thrown 4 times. What are the probabilities of obtaining 4, 3, 2, 1, 0 'aces' ?

(M.Sc., Agra, 1958)

Solution :

Chance of obtaining an ace with one die is $\frac{1}{6}$; hence chances of obtaining 4, 3, 2, 1 and 0 'aces' with 4 dice are successive terms in the binomial expansion $\left(\frac{1}{6}+\frac{5}{6}\right)^4$. The result is tabulated as below :

No. of 'aces'	4,	3,	2,	1,	0
Probability	$\left(\frac{1}{6}\right)^4$,	$4\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)$,	$6\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$,	$4\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3$,	$\left(\frac{5}{6}\right)^4$.

Problem 453.—Two players A and B want respectively m and n points of winning a set of games ; their chances of winning a single game are p and q respectively, where the sum of p and q is unity ; the stake is to belong to the player who first makes up his set ; determine the probabilities in favour of each player.

(Hall and Knight Algebra, page 388)

Solution :

Suppose that A wins in exactly $m+r$ games ; to do this he must win the last game and $m-1$ out of the preceding $m+r-1$ games. The chance of this is ${}^{m+r-1}C_{m-1}\, p^{m-1}q^r p$ or ${}^{m+r-1}C_{m-1}\, p^m q^r$.

Now the set will necessarily be decided in $m+n-1$ games, and A may win his m games in exactly m games, or $m+1$ games,..., or $m+n-1$ games ; therefore we shall obtain the chance that A wins the set by giving to r the values 0, 1, 2,......$n-1$ in the expression ${}^{m+r-1}C_{m-1}\,p^m q^r$.

Thus A's chance

$$=p^m\left[1+mq+\frac{m(m+1)}{\lfloor 2}q^2+\ldots\ldots+\frac{\lfloor m+n-2}{\lfloor m-1\,\lfloor n-1}q^{n-1}\right];$$

Similarly B's chance

$$=q^n\left[1+np+\frac{n(n+1)}{2}p^2+\ldots\ldots+\frac{\lfloor m+n-2}{\lfloor m-1\,\lfloor n-1}p^{m-1}\right].$$

Problem 454.—Two players of equal skill A and B, are playing a set of games ; they leave off playing when A wants 3 points and B wants 2 points. If the stake is £ 16, what share ought each to take.

(M.Sc., Agra, 1948)

Solution :

Since the players A and B are of equal skill, the chance of each winning is $\frac{1}{2}$. Now the game must be decided in four games out of which A must win the last game and two of the remaining three games.

Hence probability of A's winning

$$= {}^3C_1\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)^3$$

$$= \frac{3}{16}+\frac{1}{8}=\frac{5}{16}$$

Hence A's share $= \frac{5}{16}\times 16 = 5$ £

$\therefore$ B's share $= 11$ £.

Problem 455.—A problem in statistics is given to three students whose chances of solving it are $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$. What is the probability that the problem will be solved.

(M.Sc., Agra, 1957)

Solution :

Probability that first fails to solve the problem $=1-\frac{1}{2}=\frac{1}{2}$

" " second " " " $=1-\frac{1}{3}=\frac{2}{3}$

" " third " " " $=1-\frac{1}{4}=\frac{3}{4}$

$\therefore$ The probability that none could solve the problem *i.e.* problem is not solved $=\frac{1}{2}\times\frac{2}{3}\times\frac{3}{4}=\frac{1}{4}$.

Hence the probability that the problem is solved

$$=1-\frac{1}{4}=\frac{3}{4}.$$

Problem 456.—Suppose a white ball has been drawn from one of the three bags,

first containing 3 white and 5 red balls
the second containing 2 white and 7 red balls
the third containing 5 white and 9 red balls
What is the chance that it was drawn from the third bag.

Solution :

Method I.—Let these be n drawings, of those $\frac{n}{3}$ will be made from the first bag, and of these $\frac{3}{8}\times\frac{n}{3}$ will give a white ball.

Similarly $\frac{2}{9}\times\frac{n}{3}$ will be the number of occasions, on which a white ball be drawn from the second bag; and $\frac{5}{14}\times\frac{n}{3}$ will be the number of occasions, on which a white is drawn from the third bag.

Hence out of every $\left(\frac{3}{8}+\frac{2}{9}+\frac{5}{14}\right)\times\frac{n}{3}$ occasions on which a white is drawn, it will be from the third bag on $\frac{5}{14}\times\frac{n}{3}$ occasions.

Hence the chance that, a white ball being drawn, it comes from the third bag is

$$p=\frac{\frac{5}{14}\times\frac{n}{3}}{\left(\frac{3}{8}\times\frac{2}{9}+\frac{5}{14}\right)\times\frac{n}{3}}$$

$$=\frac{180}{481}.$$

Method II.—By article 7 we have

$$P_1=P_2=P_3=\frac{1}{3}$$

$$p_1=\frac{3}{8},\ p_2=\frac{2}{9},\ p_3=\frac{5}{14}$$

$$\therefore \quad \text{required } p=\frac{\frac{5}{14}\times\frac{1}{3}}{\frac{1}{3}\times\frac{3}{8}+\frac{1}{3}\times\frac{2}{9}+\frac{1}{3}\times\frac{5}{14}}$$

$$=\frac{180}{481}.$$

Problem 457.—A bag contains three balls, each of which may be either white or black; and a white ball is drawn. What is the chance that the bag contained 2 white and 1 black?

Solution :

This problem may be interpreted in two different ways, which we shall discuss separately.

I. If each ball is equally likely to be white or black, by taking the terms in the expansion of $\left(\frac{1}{2}+\frac{1}{2}\right)^3$, we find the probabilities in three cases :—

(*a*) All the three balls are white probability is $\left(\frac{1}{2}\right)^3$ or $\frac{1}{8}$

(*b*) 2 balls are white probability is ${}^3C_1\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)$ or $\frac{3}{8}$

(*c*) 1 ball is white probability is ${}^3C_2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2$ or $\frac{3}{8}$

$$\therefore \quad P_1=\frac{1}{8}\ ;\quad P_2=\frac{3}{8}\ ;\quad P_3=\frac{3}{8}$$

$$p_1=1\ ;\quad p_2=\frac{2}{3}\ ;\ p_3=\frac{1}{3}$$

$$\therefore \quad \text{required chance} = \frac{\frac{3}{8}\times\frac{2}{3}}{\frac{1}{8}\times 1+\frac{3}{8}\times\frac{2}{3}+\frac{3}{8}\times\frac{1}{3}}$$

$$=\frac{1}{2}.$$

II. The bag may either contain (1) all white, or (2) 2 white and 1 black, or (3) 1 white and 2 black, and there is no general a priori reason for supposing one assortment to exist rather than another ; we therefore assume them all three to be equally likely.

$$\therefore \quad P_1=P_2=P_3=\frac{1}{3}$$

$$p_1=1,\ \ p_2=\frac{2}{3},\ p_3=\frac{1}{3}$$

$$\therefore \quad \text{required chance} = \frac{\frac{1}{3}\times\frac{2}{3}}{\frac{1}{3}\times 1+\frac{1}{3}\times\frac{2}{3}+\frac{1}{3}\times\frac{1}{3}}$$

$$=\frac{1}{3}.$$

Problem 458.—A bag contains 5 coins which are known to be either sovereign or shillings. Two coins are drawn and are seen to be a sovereign and a shilling. If these be replaced and two again be drawn, find the chance that they will be a sovereign and a

Solution :

We can have four different kinds of bags,

(1) Containing 4 sovereign and 1 shilling,
(2) „ 3 „ 2 „ ,
(3) „ 2 „ 3 „ ;
(4) „ 1 „ 4 „ .

We assume that all the bags are a priori equally likely.

Now the chance that a sovereign comes at any drawing, from (1) is $\frac{4}{10}$, from (2) is $\frac{6}{10}$, from (3) is $\frac{6}{10}$, and from (4) is $\frac{4}{10}$.

Hence the chance that the bag is of first kind is

$$\frac{\frac{1}{4}\times\frac{4}{10}}{\frac{1}{4}\times\left(\frac{4}{10}+\frac{6}{10}+\frac{6}{10}+\frac{4}{10}\right)}=\frac{1}{5},$$

of the second kind $\frac{3}{10}$, of the third kind $\frac{3}{10}$, of the fourth $\frac{1}{5}$.

Hence the chance that the bag is of first kind, and that a sovereign and a shilling are drawn again $=\frac{1}{5}\times\frac{4}{10}=\frac{4}{50}$

Similarly for second kind chance $=\frac{3}{10}\times\frac{6}{10}=\frac{9}{50}$

„ „ third „ „ $=\frac{3}{10}\times\frac{6}{10}=\frac{9}{50}$

„ „ fourth „ „ $=\frac{1}{5}\times\frac{4}{10}=\frac{4}{50}$

These events are also exclusive. Hence the required chance

$$\text{is}=\frac{4}{50}+\frac{9}{50}+\frac{9}{50}+\frac{4}{50}$$
$$=\frac{13}{25}.$$

✓**Problem 459.**—A ball has been drawn at random from a bag containing 99 black balls and 1 white ball ; and a man whose statements are accurate 9 times out of 10 asserts that the white ball was drawn. Find the chance that the white ball was really drawn.

Solution :

The probability that the white ball will really be drawn in any case is $\frac{1}{100}$, and therefore the probability that the man will truly assert that the white ball is drawn is $\frac{1}{100}\times\frac{9}{10}$.

The probability that the white ball will not be drawn is $\frac{99}{100}$, and therefore the probability that the man will falsely assert that the white ball is drawn is $\frac{99}{100} \times \frac{1}{10}$.

Hence the required probability

$$= \frac{\frac{1}{100} \times \frac{9}{10}}{\frac{1}{100} \times \frac{9}{10} + \frac{99}{100} \times \frac{1}{10}}$$

$$= \frac{1}{12}.$$

Problem 460. – A speaks the truth three times out of four, and B five times out of six; and they agree in stating that a white ball has been drawn from a bag which was known to contain 1 white and 9 black balls. Find the chance that the white ball was really drawn.

Solution :

The probability that the white ball will be drawn in any case is $\frac{1}{10}$, and therefore the probability that A and B will agree in truly asserting that a white ball is drawn is $\frac{1}{10} \times \frac{3}{4} \times \frac{5}{6}$.

The probability that a black ball will really be drawn in any case is $\frac{9}{10}$; and therefore the probability that A and B will agree in falsely asserting that a white ball is drawn is $\frac{9}{10} \times \frac{1}{4} \times \frac{1}{6}$,

Hence the required probability

$$= \frac{\frac{1}{10} \times \frac{3}{4} \times \frac{5}{6}}{\frac{1}{10} \times \frac{3}{4} \times \frac{5}{6} + \frac{9}{10} \times \frac{1}{4} \times \frac{1}{6}}$$

$$= \frac{5}{8}.$$

Problem 461. – A speaks truth three times out of four, and B five times out of six; and they agree in stating that a white ball has been drawn from a bag which was known to contain 10 balls all of different colours, white being one. What is the chance that a white ball was really drawn?

Solution :

The probability that the white ball will really be drawn in any case is $\frac{1}{10}$, and therefore the probability that A and B will agree in truly asserting that the white ball is drawn is

$$\frac{1}{10} \times \frac{3}{4} \times \frac{5}{6} = \frac{1}{16}.$$

The probability that the white ball will not be drawn in any case is $\frac{9}{10}$. The probability that A will make a wrong statement is $\frac{1}{4}$, hence, as there are nine ways of making a wrong statement which may all be supposed to be equally likely, the chance that A will wrongly assert that a white ball is drawn is $\frac{1}{4} \times \frac{1}{9}$. Therefore the chance that A and B will agree in falsely asserting that a white ball is drawn $= \frac{9}{10} \times \frac{1}{4 \times 9} \times \frac{1}{6 \times 9} = \frac{1}{2160}$.

Hence required probability is

$$= \frac{\frac{1}{16}}{\frac{1}{16} + \frac{1}{2160}}$$

$$= \frac{135}{136}.$$

Q. There are 4 bags each containing 6 W & 3 B b... & 3 bags each containing 2 W & 4 B balls. A B ball is drawn, what is the chance that it come from 1st group.

Soln.

$$\text{Ans} = \frac{\frac{4}{7} \times \frac{3}{9}}{\frac{4}{7} \times \frac{3}{9} + \frac{3}{7} \times \frac{4}{6}}$$

CHAPTER XV

CURVE FITTING

Method of Least Squares:

Let us suppose that we have n pairs of values $X_1\ Y_1, \ldots\ldots, x_n\ y_n$, and that we wish to represent them by an equation of the type

$$y = a_0 + a_1 x + a_2 x^2 + \ldots\ldots + a_p x^p \ldots\ldots (1)$$

Now we have to determine the values of the constants $a_0, a_1, \ldots, a_p$ in terms of the given values X, Y, so as to get the best possible fit.

Substituting $X_1, X_2, \ldots X_n$ for x in equation (1) we get

$$y_1' = a_0 + a_1 X_1 + a_2 X_1^2 + \ldots\ldots + a_p X_1^p$$
$$y_2' = a_0 + a_1 X_2 + a_2 X_2^2 + \ldots + a_p X_2^p$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y'_n = a_0 + a_1 X_n + a_2 X^2_n + \ldots\ldots\ldots + a_p X_n^p$$

Now define a quantity U such that

$$U = \Sigma (Y_r - y_r')^2 \text{ where } r = 1, 2, \ldots\ldots n.$$

The values of $Y_r - y'_r$ for different values of r are called Residuals. The best fit is that for which U is minimum.

But $U = \Sigma (Y_r - a_0 - a_1 X_r - a_2 X^2_r \ldots\ldots - a_p X_r^p)^2$

The condition that U may be minimum gives

$$\frac{\partial U}{\partial a_0} = \frac{\partial U}{\partial a_1} = \frac{\partial U}{\partial a_2} = \ldots\ldots\ldots = \frac{\partial U}{\partial a_p} = 0$$

Which gives on simplification

$$\Sigma Y_r = na_0 + a_1 \Sigma X_r + a_2 \Sigma X^2_r + \ldots\ldots\ldots$$
$$\Sigma X_r Y_r = a_0 \Sigma X_r + a_1 \Sigma X^2_r + a_2 \Sigma X^3_r + \ldots\ldots$$
$$\Sigma X^2_r Y_r = a_0 \Sigma X^2_r + a_1 \Sigma X^3_r + a_2 \Sigma X^4_r + \ldots\ldots$$
$$\Sigma X^3_r Y_r = a_0 \Sigma X^3_r + a_1 \Sigma X^4_r + a_2 \Sigma X^5_r + \ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Writing these equation in general forms as

$$\Sigma Y = na_0 + a_1 \Sigma X + a_2 \Sigma X^2 + \ldots\ldots\ldots\ldots + a_p \Sigma X^p$$
$$\Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 + a_2 \Sigma X^3 + \ldots + a_p \Sigma X^{p+1}$$

$$\Sigma X^2Y = a_0 \Sigma X^2 + a_1 \Sigma X^3 + a_2 \Sigma X^4 + \ldots + a_p \Sigma X^{p+2}$$
$$\Sigma X^3Y = a_0 \Sigma X^3 + a_1 \Sigma X^4 + a_2 \Sigma X^5 + \ldots + a_p \Sigma X^{p+3}$$

....

....

....

These equations are known as Normal Equations and are $p+1$ in number. They can be solved like simultaneous equations to give the $p+1$ constants.

Cor. I. If we have to fit a curve of first degree *viz.* $y = a_0 + a_1 x$ then normal equations are $\Sigma Y = na_0 + a_1 \Sigma X$

$$\Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2$$

These two equations will give the constants a_0 and a_1.

Cor. II. If we have to fit a curve of second degree *viz.* $y = a_0 + a_1 x + a_2 x^2$ then normal equations are

$$\Sigma Y = na_0 + a_1 \Sigma X + a_2 \Sigma X^2$$
$$\Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 + a_2 \Sigma X^3$$
$$\Sigma X^2Y = a_0 \Sigma X^2 + a_1 \Sigma X^3 + a_2 \Sigma X^4$$

These three equations will give the constants a_0, a_1 and a_2.

Examples

Problems 462.—Fit a straight line to the following data regarding X as the independent variable.

x	0	1	2	3	4
y	1	1·8	3·3	4·5	6·3

(*M. Sc., Agra, 1949*)

Solution :

Let the straight line be

$$y = a_0 + a_1 x$$

then normal equations are

$$\Sigma y = na_0 + a_1 \Sigma x$$
$$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2$$

here $n = 5$

Calculation of Σx, Σy, Σxy, and Σx^2.

x	y	xy	x^2
0	1	0	0
1	1·8	1·8	1
2	3·3	6·6	4
3	4·5	13·5	9
4	6·3	25·2	16
$\Sigma x = 10$	$\Sigma y = 16·9$	$\Sigma xy = 47·1$	$\Sigma x^2 = 30$

$\therefore$ normal equations become

$$16{\cdot}9 = 5a_0 + 10a_1$$
$$47{\cdot}1 = 10a_0 + 30a_1$$

Solving $a_0 = {\cdot}72$, $a_1 = 1{\cdot}33$.

hence the required line is

$$y = {\cdot}72 + 1{\cdot}33x.$$

Problem 463.—Show that the line of fit to the following data is given by $y = {\cdot}7x + 11{\cdot}28$.

x	0	5	10	15	20	25
y	12	15	17	22	24	30

Solution :

We see that the values of x are equispaced so that $x_1 = 0$, $x_n = 25$ and $h = 5$

$$\therefore \quad M_x = \frac{x_1 + x_n}{2} = \frac{25}{2},$$

$$\text{and } t = \frac{x - M_x}{h} = \frac{x - 12{\cdot}5}{5}$$

Let the line be

$$y = a_0 + a_1 t$$

hence normal equations are

$$\Sigma y = na_0 + a_1 \Sigma t$$
$$\Sigma ty = a_0 \Sigma t + a_1 \Sigma t^2$$

Calculation of Σy, Σty, Σt and Σt^2.

x	t	y	ty	t^2
0	$-2{\cdot}5$	12	-30	$6{\cdot}25$
5	$-1{\cdot}5$	15	$-22{\cdot}5$	$2{\cdot}25$
10	$-{\cdot}5$	17	$-8{\cdot}5$	${\cdot}25$
15	${\cdot}5$	22	11	${\cdot}25$
20	$1{\cdot}5$	24	36	$2{\cdot}25$
25	$2{\cdot}5$	30	75	$6{\cdot}25$
	$\Sigma t = 0$	$\Sigma y = 120$	$\Sigma ty = 61$	$\Sigma t^2 = 17{\cdot}5$

$\therefore$ Normal equations become

$$120 = 6a_0 + 0 \qquad \therefore \quad a_0 = 20$$
$$61 = 0 + 17{\cdot}5a_1 \qquad \therefore \quad a_1 = 3{\cdot}48$$

$\therefore$ line is

$$y = 20 + 3{\cdot}48\, t$$

or $$y = 20 + 3{\cdot}48\, \frac{x - 12{\cdot}5}{5}$$

or $$y = {\cdot}7x + 11{\cdot}28.$$

Problem 464.—Fit a second degree parabola to the following data taking x as independent variable.

x	0	1	2	3	4
y	1	1·8	1·3	2·5	6·3

(*M.Sc. Agra, 1956*)

Solution :

Let the parabolic curve of fit be

$$y = a_0 + a_1 x + a_2 x^2$$

then normal equations are

$$\Sigma y = na_0 + a_1 \Sigma x + a_2 \Sigma x^2$$
$$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x^3$$
$$\Sigma x^2 y = a_0 \Sigma x^2 + a_1 \Sigma x^3 + a_2 \Sigma x^4.$$

The values of Σx etc. are calculated as below.

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	1·8	1·8	1	1·8	1	1
2	1·3	2·6	4	5·2	8	16
3	2·5	7·5	9	22·5	27	81
4	6·3	25·2	16	100·8	64	256
$\Sigma x = 10$	$\Sigma y = 12·9$	37·1	30	130·3	100	354

Substituting these values we get the normal equations as

$$12·9 = 5a_0 + 10a_1 + 30a_2$$
$$37·1 = 10a_0 + 30a_1 + 100a_2$$
$$130·3 = 30a_0 + 100a_1 + 354a_2$$

Solving we get

$$a_0 = 1·42,\ a_1 = -1·07,\ a_2 = ·55$$

∴ parabolic curve of fit is

$$y = 1·42 - 1·07\,x + ·55x^2$$

Problem 465.—Fit a second degree parabolic to the following data :

x	0	1	2	3	4
y	1	5	10	22	38

(*M. Sc. Agra, 1958*)

Solution :

Let the parabolic curve of fit be

$$y=a_0+a_1x+a_2x^2$$

Then normal equations are

$$\Sigma y=na_0+a_1\Sigma x+a_2\Sigma x^2$$
$$\Sigma xy=a_0\Sigma x+a_1\Sigma x^2+a_2\Sigma x^3$$
$$\Sigma x^2y=a_0\Sigma x^2+a_1\Sigma x^3+a_2\Sigma x^4$$

Calculation of Σx etc.

x	y	xy	x^2	x^2y	x^3	x^4
0	1	0	0	0	0	0
1	5	5	1	5	1	1
2	10	20	4	40	8	16
3	22	66	9	198	27	81
4	38	152	16	608	64	256
10	76	243	30	851	100	354

∴ Normal equations become

$$76=6a_0+10a_1+30a_2$$
$$243=10a_0+30a_1+100a_2$$
$$851=30a_0+100a_1+354a_2$$

Solving we get

$$a_0=\cdot 75,\ a_1=\cdot 83,\ a_2=2\cdot 106$$

Hence the required curve of fit is

$$y=\cdot 75+\cdot 83x+2\cdot 106x^2$$

Problem 466.—Fit a parabolic curve of regression of y on x to the seven pairs of values.

$x\rightarrow$	1·0	1·5	2·0	2·5	3·0	3·5	4·0
$y\rightarrow$	1·1	1·3	1·6	2·0	2·7	3·4	4·1

(*M.Sc. Agra, 1954*)

Solution :

I Method : Let the parabolic curve of fit be

$$y=a_0+a_1x+a_2x^2$$

Then normal equations are

$$\Sigma y=na_0+a_1\Sigma x+a_2\Sigma x^2$$
$$\Sigma xy=a_0\Sigma x+a_1\Sigma x^2+a_2\Sigma x^3$$
$$\Sigma x^2y=a_0\Sigma x^2+a_1\Sigma x^3+a_2\Sigma x^4$$

Calculation of Σx, Σx^2,......etc.

x	y	xy	x^2	x^2y	x^3	x^4
1·0	1·1	1·1	1·0	1·1	1·0	1·0
1·5	1·3	1·95	2·25	2·925	3·375	5·0625
2·0	1·6	3·2	4·0	6·4	8·0	16·0
2·5	2·0	5·0	6·25	12·5	15·625	39·0625
3·0	2·7	8·1	9·0	24·3	27·0	81·0
3·5	3·4	5·9	12·25	41·65	42·875	150·0625
4·0	4·1	16·4	16·0	65·6	64·0	256·0
17·5	16·2	41·65	50·75	154·475	161·875	548·1875

Hence normal equations become

$$16{\cdot}2 = 7a_0 + 17{\cdot}5a_1 + 50{\cdot}75a_2$$
$$41{\cdot}65 = 17{\cdot}5a_0 + 50{\cdot}75a_1 + 161{\cdot}875a_2$$
$$154{\cdot}475 = 50{\cdot}75a_0 + 161{\cdot}875a_1 + 548{\cdot}1875a_2$$

Solving we get

$$a_0 = 1{\cdot}04, \qquad a_1 = -{\cdot}2, \qquad a_2 = {\cdot}24$$

$\therefore$ Curve of fit is

$$y = 1{\cdot}04 - {\cdot}2x + {\cdot}24x^2$$

II Method :

Put $$u = \frac{x - 2{\cdot}5}{{\cdot}5} = 2x - 5$$

and let the curve of fit be

$$y = a_0 + a_1u + a_2u^2$$

Then normal equations are

$$\Sigma y = na_0 + a_1\Sigma u + a_2\Sigma u^2$$
$$\Sigma uy = a_0\Sigma u + a_1\Sigma u^2 + a_2\Sigma u^3$$
$$\Sigma u^2y = a_0\Sigma u^2 + a_1\Sigma u^3 + a_2\Sigma u^4$$

Calculation of Σu etc.

x	u	y	uy	u^2	u^2y	u^3	u^4
1·0	−3	1·1	−3·3	9	9·9	−27	81
1·5	−2	1·3	−2·6	4	5·2	−8	16
2·0	−1	1·6	−1·6	1	1·6	−1	1
2·5	0	2·0	0	0	0	0	0
3·0	1	2·7	2·7	1	2·7	1	1
3·5	2	3·4	6·8	4	13·6	8	16
4·0	3	4·1	12·3	9	36·9	27	81
	$\Sigma u=0$	16·2	14·3	28	69·9	0	196

The normal equations become

$$16{\cdot}2=7a_0+28a_2$$
$$14{\cdot}3=28a_1$$
$$69{\cdot}9=28a_0+196a_2$$

Solving $a_0=2{\cdot}07,\quad a_1={\cdot}511,\quad a_2={\cdot}061$

Hence curve is

$$y=2{\cdot}07+{\cdot}511(2x-5)+{\cdot}061(2x-5)^2$$

or $y=1{\cdot}04-{\cdot}2x+{\cdot}24x^2.$

Problem 467.—Fit a second degree parabola to the following data taking x as the independent variable :-

x ⟶	1	2	3	4	5	6	7	8	9
y ⟶	2	6	7	8	10	11	11	10	9

(M.Sc. Agra, 1953)

Solution :

Put $u=x-5,\quad v=y-8$

and let the parabolic curve of fit be

$$v=a_0+a_1u+a_2u^2$$

Normal equations are

$$\Sigma v=na_0+a_1\Sigma u+a_2\Sigma u^2$$
$$\Sigma uv=a_0\Sigma u+a_1\Sigma u^2+a_2\Sigma u^3$$
$$\Sigma u^2v=a_0\Sigma u^2+a_1\Sigma u^3+a_2\Sigma u^4$$

Calculation of Σu etc.

x	y	u	v	uv	u^2	u^2v	u^3	u^4
1	2	−4	−6	24	16	−96	−64	256
2	6	−3	−2	6	9	−18	−27	81
3	7	−2	−1	2	4	−4	−8	16
4	8	−1	0	0	1	0	−1	1
5	10	0	2	0	0	0	0	0
6	11	1	3	3	1	3	1	1
7	11	2	3	6	4	12	8	16
8	10	3	2	6	9	18	27	81
9	9	4	1	4	16	16	64	256
		0	2	51	60	−69	0	708

Hence the normal equations become

$$2=9a_0+60a_2$$
$$51=60a_1$$
$$-69=60a_0+708a_2$$

Solving $a_0=2{\cdot}002,\quad a_1={\cdot}85,\quad a_2=-{\cdot}267$

$\therefore$ Required curve is

$$y-8=2{\cdot}002+{\cdot}85(x-5)-{\cdot}267(x-5)^2$$

or $$y=10{\cdot}002+{\cdot}85(x-5)-{\cdot}267(x-5)^2.$$

Problem 468.—The profits, £100y, of a certain company in the xth year of its life are given by :

x	1	2	3	4	5
y	25	28	33	39	46

Taking $u=x-3$ and $v=y-33$, find the parabolic regression of v on u in the form

$$v=a+bu+cu^2$$

(*M.Sc. Agra, 1952*)

Solution :

Since curve is $v=a+bu+cu^2$

The normal equations are

$$\Sigma v=na+b\Sigma u+c\Sigma u^2$$
$$\Sigma uv=a\Sigma u+b\Sigma u^2+c\Sigma u^3$$
$$\Sigma u^2v=a\Sigma u^2+b\Sigma u^3+c\Sigma u^4$$

Calculation of Σu etc.

x	y	u	v	uv	u^2	u^2v	u^3	u^4
1	25	−2	−8	16	4	−32	−8	16
2	28	−1	−5	5	1	−5	−1	1
3	33	0	0	0	0	0	0	0
4	39	1	6	6	1	6	1	1
5	46	2	13	26	4	52	8	16
		0	6	53	10	21	0	34

Hence normal equations become

$$6=5a+10c$$
$$53=10b$$
$$21=10a+34c$$

$\therefore\quad a={\cdot}086,\quad b=5{\cdot}3,\quad c={\cdot}643$

$\therefore$ Required curve is

$$v=-{\cdot}086+5{\cdot}3u+{\cdot}643u^2$$

CHAPTER XVI

THE THEORY OF SAMPLING (LARGE SAMPLES)

Random Sampling :

The selection of an individual from a Parent universe is random when each member of the universe has the same chance of being chosen. It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately, but by some mechanical process so that every unit has equal probability of entering in sample.

Simple Sampling :

By simple sampling we mean Random Sampling in which each event has the same chance of success, and in which chances of success of different events are independent, whether previous trial has been made or not.

The theory of sampling can be studied under two heads (1) *The sampling of Attributes*, and (2) *The sampling of Variables.*

(1) The Sampling of Attributes

In the sampling of attributes we are concerned only with the possession or non-possession of some specified attribute or characteristic by the individual selected in sampling. For instance, in sampling from births we may be concerned only whether the child is male or female. The choosing of an individual in sampling may be called an 'event' or a 'trial', and the possession of the specified attribute by the individual selected a 'success' and the non-possession a 'failure'.

Suppose now that we take N samples with n events in each. The chance of success of each event is p and of the failure $q=1-p$; then the frequencies of samples with 0, 1, 2,... successes are the terms in the expansion $N(q+p)^n$.

The expected value, or the mean value, of the number of successes is therefore np; the variance is npq, and the standard deviation (or the standard error) of the number of successes is $\sqrt{npq}$.

The standard error of the proportion of successes is

$$=\sqrt{\frac{pq}{n}}.$$

Note. If the observed number (or proportion) of successes differs from expected number (or proportion) of successes by more than 3 times the standard error of number of successes (or proportion of successes) then the hypothesis is not correct and the deviation between the two values is not due to fluctuation of random sampling.

Comparison of Large Samples :

Let two samples give proportion of As' as p_1 and p_2, the numbers in the two samples being n_1 and n_2. We have to find if the difference $p_1 - p_2$ is significant of a real difference between two populations with respect to the given attributes. On the hypothesis that the populations are similar in this respect we can combine the samples to give the common value of the proportion of As' in the population by the formula

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

If E_1, E_2 are the standard errors in the two samples then

$$E_1^2 = \frac{p_0 q_0}{n_1} \text{ and } E_2^2 = \frac{p_0 q_0}{n_2}$$

Let E be the standard error of the parent universe then

$$E^2 = E_1^2 + E_2^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

If $p_1 - p_2 > 3E$ then it is not due to fluctuations of the simple sampling, but is due to some other reason.

If the proportion of As' are not the same in the material from which the two samples are drawn, but p_1 and p_2 are the true values of the proportions, the standard errors of sampling in the two cases

$$\text{are} \quad E_1 = \frac{p_1 q_1}{n_1}, \quad E_2 = \frac{p_2 q_2}{n_2}$$

$$\text{then } E^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

If the difference $p_1 - p_2 < 3E$, it may vanish on taking fresh samples because this difference may be due to the fluctuations of the simple sampling.

(2) Sampling of Variables

We will now consider the sampling of values of a variable, such as height, age, etc. Each member of the population of individuals provides a value of the variable, and we thus have a population of values of the variable, and the frequency distribution determined by it.

Sampling Distribution. If we take a number of samples from a universe and calculate some statistic, say mean or standard deviation, of each sample, we shall get a series of different values, one for each sample. If the number of samples is reasonably large, these values may be grouped to give a frequency distribution. As the value of the number of samples becomes larger this distribution approaches the ideal or the normal form of a continuous curve. Such a distribution is called a sampling distribution.

The standard errors of some important parameters when the parent universe is normal are given below :—

S.E. of Mean $= \frac{\sigma}{\sqrt{n}}$

,, ,, Standard deviation $= \frac{\sigma}{\sqrt{2n}}$

,, ,, Variance $= \sigma^2 \sqrt{\frac{2}{n}}$

,, ,, of coefficient of correlation $= \frac{1-r^2}{\sqrt{n}}$

,, ,, of coefficient of variation (V) $= \frac{V}{\sqrt{2n}} \sqrt{1+\frac{2V^2}{10}}$

Comparison of Means of two large samples :*

Given two independent simple samples of n_1 and n_2 members respectively, we may wish to examine whether the difference of their means may be accounted for by fluctuations of sampling, the two samples being regarded as drawn from the same population of S.D.σ. The Standard error's of the means of samples of n_1 and n_2 members from this population are $\frac{\sigma}{\sqrt{n_1}}$ and $\frac{\sigma}{\sqrt{n_2}}$ respectively. Hence, on the assumption that the samples are independent and drawn from this population, the S.E.E. of the difference of their means is given by

$$E^2=\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

The sampling distribution of this difference has zero for its mean, and is approximately normal if n_1 and n_2 are large. Consequently, if the observed difference of the means exceeds 3E, it can hardly be ascribed to fluctuations of sampling ; and our assumption that the samples were drawn from the same population is almost certainly incorrect.

If the two samples are known to have come from different populations, with variances σ_1^2 and σ_2^2 respectively, we can test by a similar procedure whether the two populations may have the same mean. In this case the S.E. of their difference is given by

$$E^2=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

On the assumption that the two populations have the same mean, the distribution of the difference of the means of the samples has zero for its expected value, and is approximately normal for large samples. The significance is tested in usual manners as above.

*From *Statistics* by Weatherburn.

Examples

Problem 469.—A coin is tossed 1000 times and the head comes out 550 times. Can the deviation from expected value be due to fluctuations of simple sampling ?

Solution :

Chance of getting a head, $p = \frac{1}{2}$

Expected frequency $= np = 1000 \times \frac{1}{2} = 500$

Standard deviation $= \sqrt{npq} = \sqrt{\frac{1}{4} \times 1000}$

$= 15 \cdot 81$

The difference between observed frequency and expected frequency $= 550 - 500 = 50$

Since difference is greater than three times the standard error and so it cannot be accounted for by fluctuations of simple sampling.

Problem 470.—A certain cubical die was thrown 9000 times, and a 5 or a 6 was obtained 3240 times. On the assumption of random throwing, do the data indicate an unbiassed die ?

Solution :

Assuming that die is unbiassed we have chance of throwing a 5 or a 6 is $\frac{1}{3}$

$\therefore \quad p = \frac{1}{3}, \quad q = \frac{2}{3}$

Expected number of successes $= np = 9000 \times \frac{1}{3}$

$= 3000$

The standard error of the number of successes is

$$= \sqrt{npq} = \sqrt{9000 \times \frac{2}{9}} = 44 \cdot 72$$

The difference between observed frequency and expected frequency $= 3240 - 3000 = 240$

Since difference is greater than 3 times standard error, hence the deviation is not due to simple sampling fluctuations but the die is biassed one.

Problem 471.—In some dice throwing experiments Weldon threw dice 49152 times, and of these 25145 yielded a 4, 5, or 6. Is this consonant with the hypothesis that the dice were unbiassed ?

Solution :

The probability of getting 4, 5, or 6 with a die is $\frac{1}{2}$, hence the expected proportion of successes is $p = \frac{1}{2} = \cdot 5$.

The observed proportion of successes is $= \frac{25145}{49152}$

$= \cdot 5115$

The standard error of the proportion is

$$= \sqrt{\frac{pq}{n}} = \sqrt{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{49152}} = \sqrt{\cdot 00000508}$$

$$= \cdot 00225$$

Difference between the observed and expected proportions of successes $= \cdot 5115 - \cdot 5 = \cdot 0115$

3 times standard error of proportions $= \cdot 00675$

Since difference is greater than 3 times standard error hence the deviation is not due to simple sampling fluctuations but the dice were biassed and not unbiassed.

Problem 472.—Certain crosses of the pea, Pisum sativum, gave 5321 yellow and 1804 green seeds. The expectation is 25 per cent of green seeds on a Mendelian hypothesis. Can the divergence from expected value have arisen from fluctuations of simple sampling only ? (*Quoted*)

Solution :

Total number of pea seeds examined $= 5321 + 1804$

$= 7125$

Observed proportion of getting green seeds is

$$= \frac{1804}{7125} = \cdot 25223$$

Expected proportion of getting green seeds is 25 per cent;

therefore $p = \frac{1}{4} = \cdot 25$

Standard error of proportions of getting green seeds

$$=\sqrt{\frac{pq}{n}}$$

$$=\sqrt{\frac{1}{4}\times\frac{3}{4}\times\frac{1}{7152}}$$

$$=\sqrt{\cdot 00002621}$$

$$=\cdot 0051$$

Difference between observed and expected proportions

$$=\cdot 25223-\cdot 25$$

$$=\cdot 00223$$

3 times standard error $=\cdot 0153$

Since difference is less than 3 times standard error hence the deviation is due to simple sampling fluctuations.

Problem 473.—A coin is tossed 10000 times, and head turns up 5195 times. Is it reasonable to think that the coin is unbiassed ? (*P.C.S.*)

Solution :

Probability of getting a head is $=\frac{1}{2}$.

Observed frequency of getting heads $=5195$

Expected frequency of getting heads $=np=5000$

Standard error $=\sqrt{npq}=\sqrt{10000\times\frac{1}{4}}=50$

Difference between observed and expected frequency

$$=5195-5000=195$$

This is greater than 3 times the standard error hence deviation is not due to simple sampling fluctuation but the die is biassed.

Problem 474.—A coin is tossed 400 times and it turns head 216 times. Discuss whether the coin may be an unbiassed one, and explain briefly the theoretical principles you would use for this purpose. (*I.A.S.*)

Solution :

$$p=\frac{1}{2}$$

$\therefore$ Expected frequency of getting heads $=400\times\frac{1}{2}=200$

Observed frequency of getting heads $=240$

Standard error $=\sqrt{npq}=\sqrt{\frac{1}{4}\times 400}=10.$

Difference between observed and expected frequency $=40$

Since difference is greater than 3 times standard error hence dice is biassed one and the deviation is not due to simple sampling fluctuations.

Problem 475.—A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Estimate the proportion of the bad pineapples in the consignment, as well as the standard error of the estimate. Deduce that the percentage of bad pineapples in the consignment almost certainly lies between 8·5 and 17·5. (*I.A.S. 1954*)

Solution :

The proportion of bad pineapples in the sample is

$$p=\frac{65}{500}=\frac{13}{100}=\cdot 13$$

$$\therefore \quad q=1-\frac{13}{100}=\frac{87}{100}$$

Standard error of proportion of bad pineapples

$$=\sqrt{\frac{pq}{n}}=\sqrt{\frac{13}{100}\times\frac{87}{100}\times\frac{1}{500}}$$

$$=\cdot 015$$

Hence the proportion of bad pineapples lies between $\cdot 13\pm 3\times\cdot 015$ (or between ·085 and ·175).

Hence the percentage of bad pineapples lies between 8·5 and 17·5.

Problem 476.—A sample of 900 days is taken from Meteorological records of a certain district, and 100 of them are found to be foggy. What are the probable limits to the percentage of foggy days in the district ? (*P.C.S. 1952*)

Solution :

The proportion of foggy days in the sample is

$$p=\frac{100}{900}=\frac{1}{9}=\cdot 111$$

$$\therefore \qquad q=1-\frac{1}{9}=\frac{8}{9}.$$

Standard error of proportion of foggy days

$$=\sqrt{\frac{pq}{n}}=\sqrt{\frac{1}{9}\times\frac{8}{9}\times\frac{1}{900}}$$

$$=\cdot 0105$$

Hence the proportion of foggy days lies between $(\cdot 111-3\times\cdot 0105)$ and $(\cdot 111+3\times\cdot 0105)$ that is between ·08 and ·1425.

Hence the percentage of foggy days lies between 8 and 14·25.

Problem 477.—In a locality containing 18,000 families, a sample of 840 families was selected at random. Of these 840 families, 206 families were found to have a monthly income of Rs. 50 or less. It is desired to estimate how many out of the 18,000 families have a monthly income of Rs. 50 or less. Within what limits would you place your estimate? *(U.P.C.S. 1948)*

Solution :

Proportion of families having a monthly income of Rs. 50 or less is $p=\frac{206}{840}=\frac{103}{420}=\cdot 245$

$$q=1-\frac{103}{420}=\frac{317}{420}$$

∴ Standard error of proportion of families having a monthly income Rs. 50 or less $=\sqrt{\frac{pq}{n}}$

$$=\sqrt{\frac{103}{420}\times\frac{317}{420}\times\frac{1}{840}}$$

$$=\cdot 015$$

Hence proportion of families earning Rs. 50 or less p.m. lies between $(\cdot 245-3\times\cdot 015)$ and $(\cdot 245+3\times\cdot 015)$ that is between $\cdot 2$ and $\cdot 29$.

Hence number of families, out of 18,000 families, earning Rs. 50 or less p.m. lies between $\cdot 2\times 18000$ and $\cdot 29\times 18000$ that is between 360 and 5220 families.

Problem 478.—Out of a simple sample of 1000 individuals from the inhabitants of a country we find that 36 per cent of them have blue eyes and the remainder have eyes of some other colour. What can we infer about the proportion of blue eyes individuals of the whole population.

Solution :

Proportion of blue eyes individuals is 36 per cent, therefore $p=\cdot 36$ and $q=\cdot 64$

Standard error of proportion of blue eyes individuals

$$=\sqrt{\frac{pq}{n}}=\sqrt{\frac{\cdot 36\times\cdot 64}{1000}}$$

$$=\frac{\cdot 6\times\cdot 8}{10\times 3\cdot 162}=\frac{\cdot 48}{31\cdot 62}$$

$$=\cdot 0152$$

Hence proportion of blue eyes individuals lies between $(\cdot 36-3\times\cdot 0152)$ and $(\cdot 36+3\times\cdot 0152)$ that is between $(\cdot 3144)$ and $(\cdot 4056)$.

Hence the percentage of blue eyes individual lies between 31·44 and 40·56.

Problem 479.—Of 10,000 babies born in U.P. 5200 are male childs. Taking this to be a random sample of the births in U.P., show that it throws considerable doubt on the hypothesis that the sexes are born in equal proportions.

Solution :

Probability that child is male is $p = \frac{1}{2}$

$\therefore$ Expected number of male childs is $= np = 5000$

Observed number of male childs is $= 5200$

$$\text{Standard Error} = \sqrt{npq} = \sqrt{10{,}000 \times \frac{1}{2} \times \frac{1}{2}} = 50$$

Difference between observed and expected number $= 200$

Since this is greater than 3 times the standard error hence the difference is not due to sampling fluctuation but our hypothesis that sexes are born in equal proportions is doubtful.

Problem 480.—In a simple sample of 600 men from a certain large city, 400 are found to be smokers. In one of 900 from another city 450 are found smokers. Do the data indicates that cities are significantly different with respect to prevalence of smoking among men ?

Solution :

$$p_1 = \frac{400}{600} = \frac{2}{3} \quad \text{and} \quad p_2 = \frac{450}{900} = \frac{1}{2}$$

$$\text{Then} \quad p_1 - p_2 = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} = \cdot 166$$

On the assumption that the cities are like with respect to the prevalence of smoking among men we get

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

$$= \frac{\frac{2}{3} \times 600 + \frac{1}{2} \times 900}{600 + 900}$$

$$= \frac{17}{30}$$

$$\therefore \quad q_0 = 1 - \frac{17}{30} = \frac{13}{30}$$

Then $E^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

$$= \frac{17}{30} \times \frac{13}{30} \times \left(\frac{1}{600} + \frac{1}{900}\right)$$

$$= \cdot 000682$$

$\therefore$ E $= \cdot 026$

Since $p_1 - p_2 > 3E$, our hypothesis is wrong *i.e.*, the assumption that two cities are similar is wrong.

Problem 481.—In a random sample of 500 men from a particular district of U.P., 300 are found to be smokers. In one of 1000 men from another district, 550 are smokers. Do the data indicate that the two districts are significantly different with respect to the prevalence of smoking among men ?

(P.C.S. 1953)

Solution :

$$p_1 = \frac{300}{500} = \frac{3}{5} = \cdot 6$$

$$p_2 = \frac{550}{1000} = \frac{11}{20} = \cdot 55$$

Then $p_1 - p_2 = \cdot 6 - \cdot 55 = \cdot 05$

On the assumption that the districts are like with respect to the prevalence of smoking among men we get

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

$$= \frac{300 + 550}{500 + 1000}$$

$$= \frac{850}{1500}$$

$$= \frac{17}{30}$$

$\therefore$ $q_0 = 1 - \frac{17}{30} = \frac{13}{30}$

Then $E^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

$$= \frac{17}{30} \times \frac{13}{30} \left(\frac{1}{500} + \frac{1}{1000}\right)$$

$$= \cdot 000736$$

$\therefore$ E $= \cdot 027$

Since $p_1 - p_2 < 3E$ hence the data do not indicate that the two districts are significantly different with respect to the prevalence of smoking among men.

Problem 482.—In a random sample of 800 adults from the population of a certain large city, 600 are found to have dark hair. In a random sample of 1000 adults from the inhabitants of another large city, 700 are dark haired. Show that the difference of the proportions of dark haired people is nearly 2·4 times the S.E. of this difference for samples of the above sizes.

Solution :

Proportion of dark haired adults in first is $p_1 = \frac{600}{800} = \frac{3}{4}$

" " " second is $p_2 = \frac{700}{1000} = \frac{7}{10}$

Then difference $p_1 - p_2 = \frac{3}{4} - \frac{7}{10} = \frac{1}{20} = \cdot 05$

On the assumption that the two samples are from same parent universe we have

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$
$$= \frac{600 + 700}{800 + 1000}$$
$$= \frac{13}{18}$$

$\therefore \quad q_0 = \frac{5}{18}$

$$E^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$
$$= \frac{13}{18} \times \frac{5}{18} \left(\frac{1}{800} + \frac{1}{1000} \right)$$
$$= \cdot 000451$$

$\therefore \quad E = \cdot 021$

Hence $\quad p_1 - p_2 = 2 \cdot 4$ E nearly.

Hence the difference of the proportions of dark haired people is nearly 2·4 times the S.E. of this difference.

Problem 483.—The following table gives the proportion of dark colour people in two cities.

City	Number of People	Percentage of dark coloured people
Bombay	450	35
Jhansi	600	45

Can the difference observed in the percentage of dark-coloured people be due to the fluctuations of the sampling ?

Solution :

Proportion of dark-coloured people in first is $p_1 = \frac{35}{100} = \cdot 35$

" " " second is $p_2 = \frac{45}{100} = \cdot 45$

Difference $p_2 - p_1 = \cdot 45 - \cdot 35 = \cdot 1$

On the assumption that the two cities do not differ with respect to the colour of people we have

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

$$= \frac{\cdot 35 \times 450 + \cdot 45 \times 600}{450 + 600}$$

$$= \frac{157 \cdot 5 + 270}{1050}$$

$$= \frac{427 \cdot 5}{1050}$$

$$= \cdot 407$$

$$\therefore \quad q_0 = \cdot 593$$

$$E^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$= \cdot 407 \times \cdot 593 \times \left(\frac{1}{450} + \frac{1}{600} \right)$$

$$= \cdot 000912$$

$$\therefore \quad E = \cdot 03 \text{ nearly}$$

Since difference is greater than 3E slightly, hence to some extent we can say that difference is not due to sampling fluctuations.

Problem 484.—In a random sample of 500 persons from town A, 200 are found to be consumers of cheese. In a sample of 400 from town B, 200 are also found to be consumers of cheese. Discuss the question whether the data reveal a significant difference between A and B so far as the proportion of cheese consumed is concerned.

(*Yule and Candall*, Page 370)

Solution :

A $\quad p_1 = \frac{200}{500} = \frac{2}{5}$

B $\quad p_2 = \frac{200}{400} \quad \frac{1}{2}$

$$\therefore \quad p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

$$= \frac{200+200}{500+400} = \frac{4}{9}$$

$$q_0 = \frac{5}{9}$$

$$E^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$= \frac{4}{9} \times \frac{5}{9} \left(\frac{1}{500} + \frac{1}{400} \right)$$

$$= \cdot 001111$$

$$\therefore \quad E = \cdot 033$$

Now $\quad p_2 - p_1 = \frac{1}{2} - \frac{2}{5} = \frac{1}{10} = \cdot 1.$

Since the difference is greater than 3 times the standard erro hence this is not due to sampling fluctuations but is due to som other reason.

Problem 485.—In a certain association table the followin frequencies were obtained :—

$(AB) = 309 \qquad (\alpha B) = 132$

$(A\beta) = 214 \qquad (\alpha\beta) = 119$

Can the association of the table have arisen as a fluctuation o simple sampling, the true association being zero ?

(*Yule and Candall* Page 370

Solution :

$$(A) = 309 + 214 = 523$$

$$(B) = 309 + 132 = 441$$

$$(\beta) = 214 + 119 = 333$$

$$\therefore \quad N = 441 + 333 = 774$$

Given that the true association is zero, hence

$$\frac{(AB)}{(B)} = \frac{(A\beta)}{(\beta)}$$

The proportion of A's in B's *i.e.* $p_1 = \frac{(AB)}{(B)} = \frac{309}{441} = \cdot 701$

" " A's in β's *i e.* $p_2 = \frac{(A\beta)}{(\beta)} = \frac{214}{333} = \cdot 643$

$\therefore \qquad p_1 - p_2 = \cdot 058$

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{(A)}{N}$$

$$= \frac{523}{774} = \cdot 676$$

$\therefore \qquad q_0 = \cdot 324$

$$E^2 = p_0 q_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

$$= \cdot 676 \times \cdot 324 \left(\frac{1}{441} + \frac{1}{333} \right)$$

$$= \cdot 001156$$

$\therefore \qquad E = \cdot 034.$

Since difference is less than 3E hence the association of table have arisen due to fluctuation of the sampling.

Problem 486.—The subject under investigation is the measure of dependence of Tamil on words of Sanskrit origin. One newspaper article reporting the proceedings of the Constituent Assembly contained 2025 words of which 729 words were declared by a literary critic to be of Sanskrit origin. A second article by the same author describing atomic research contained 1600 words of which 640 words were declared by the same critic to be of Sanskrit origin. Assuming that simple sampling conditions held, estimate the limits for the proportion of Sanskrit terms in the writers vocabulary, and examine whether there is any significant difference in the independence of this writer on words of Sanskrit origin in writing on these two subjects.

(*I.A.S. 1947*)

Solution :

$$p_1 = \frac{729}{2025} = \cdot 36$$

$$p_2 = \frac{640}{1600} = \cdot 4$$

$\therefore \qquad p_2 - p_1 = \cdot 04$

$$p_0 = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

$$= \frac{729 + 640}{2025 + 1600}$$

$$= \cdot 3777$$

$$q_0 = \cdot 6223.$$

The standard error of the proportion of Sanskrit terms in the writers vocabulary is $=\sqrt{\dfrac{p_0q_0}{n}}$

$$=\sqrt{\frac{\cdot3777\times\cdot6223}{3625}}$$

$$=\cdot0081$$

Hence the proportion of Sanskrit terms in writers vocabulary lies between $\cdot3777-3\times\cdot0081$ and $\cdot3777+3\times\cdot0081$ that is between $\cdot3534$ and $\cdot402$ or between 35·34% and 40·2%.

Again $$E^2=p_0q_0\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

$$=\cdot3777\times\cdot6223\times\left(\frac{1}{2025}+\frac{1}{1600}\right)$$

$$=\cdot000286$$

$$\therefore \quad E=\cdot017.$$

Since the difference is less than 3 times standard error (E), hence the difference is due to sampling fluctuations.

Problem 487.—In two large populations there are 35 and 30 per cent of fair haired people. Is the difference likely to be revealed by simple samples of 1500 and 1000 respectively from the two populations ?

(*Weatherburn*, Page 114)

Solution :

Here $p_1=\cdot35, \quad p_2=\cdot3$

So that $p_1-p_2=\cdot05.$

The variance of the difference of the proportions in the sample is

$$E^2=\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}$$

$$=\frac{\cdot35\times\cdot65}{1500}+\frac{\cdot3\times\cdot7}{1000}$$

$$=\cdot000362$$

$$E=\cdot019$$

$$\therefore \quad p_1-p_2=2\cdot6E \text{ nearly.}$$

The probability that the real difference between the populations will be hidden is approximately the probability that, for a random value of normal variate, the deviation from the mean will be on the negative side and greater than 2·6 times the S. D. Since this is less than ½ per cent, it is unlikely that the difference will be hidden.

Problem 488.—In two large populations there are 30 and 25 per cent respectively of blue-eyes peoples. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations ?

Solution :

$$p_1 = \cdot 3$$

$$p_2 = \cdot 25$$

$\therefore$ difference $p_1 - p_2 = \cdot 05$

The variance of the difference of the populations in the sample is

$$E^2 = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

$$= \frac{\cdot 3 \times \cdot 7}{1200} + \frac{\cdot 25 \times \cdot 75}{90}$$

$$= \cdot 000175 + \cdot 000208$$

$$= \cdot 000383$$

$\therefore$ $E = \cdot 0195$

Since the difference, ·05, in the proportions is more than 2·5 times the standard error E of this difference, hence it is unlikely that the real difference will be hidden.

Problem 489.—Calculate the standard error of the mean from the following data collected in one of the many random sample inquiries conducted to find out average earnings of a particular class :

Earnings p.m. in rupees	*Number of Persons*
Upto 10	50
,, 20	150
,, 30	300
,, 40	500
,, 50	700
,, 60	800
,, 70	900
,, 80	1000

(*M. Com. Allad.*, 1951)

Solution :

Calculation of standard deviation of monthly earnings.

Monthly earnings in Rs.	Mid-Value x	Frequency f	Step deviation $u=\frac{x-35}{10}$	fu	fu^2
0—10	5	50	−3	−150	450
10—20	15	100	−2	−200	400
20—30	25	150	−1	−150	150
30—40	35	200	0	0	0
40—50	45	200	1	200	200
50—60	55	100	2	200	400
60—70	65	100	3	300	900
70—80	75	100	4	400	1600
Total		n=1000		600	4100

Now
$$\sigma=\sqrt{\frac{\Sigma fu^2}{n}-\left(\frac{\Sigma fu}{n}\right)^2}\times 10$$
$$=\sqrt{\frac{4100}{1000}-\left(\frac{600}{1000}\right)^2}\times 10$$
$$=\sqrt{4{\cdot}1-{\cdot}36}\times 10$$
$$=\sqrt{3{\cdot}64}\times 10$$
$$=19{\cdot}34$$

Standard error of the mean $=\frac{\sigma}{\sqrt{n}}$

$$=\frac{19{\cdot}34}{\sqrt{1000}}$$
$$={\cdot}61.$$

Problem 490.—A sample of 1000 members is found to have a mean 3·42 cm. Could it be reasonably regarded as a simple sample from a large population whose mean is 3·3 cm. and standard deviation 2·6 cm. ?

Solution :

The standard error of mean $=\frac{\sigma}{\sqrt{n}}$

$$=\frac{2{\cdot}6}{\sqrt{1000}}$$
$$={\cdot}082$$

Deviation of the mean of the sample from that of the population is $=3{\cdot}42-3{\cdot}3={\cdot}12$

This deviation is less than twice the S.E. of the mean, and is therefore not significant.

We conclude that the given sample might be one drawn from the population specified.

Problem 491.—A simple sample of heights of 6400 English men has a mean 67·85 and standard deviation 2·56 inches while a simple sample of heights of 1600 Austrians has a mean of 68·5 and standard deviation as 2·52. Do the data indicate that the Austrians are on the average taller than Englishmen?

Solution :

The standard error of the mean of a sample of heights of 6400 English people is

$$=\frac{\sigma_1}{\sqrt{n_1}}=\frac{2{\cdot}56}{\sqrt{6400}}={\cdot}032 \text{ inches}$$

The standard error of the mean of a sample of heights of 1600 Austrians is

$$=\frac{\sigma_2}{\sqrt{n_2}}=\frac{2{\cdot}52}{\sqrt{1600}}={\cdot}063 \text{ inches}$$

∴ Standard error of the difference of the mean

$$E=\sqrt{({\cdot}032)^2+({\cdot}063)^2}$$
$$={\cdot}07.$$

The observed difference between the means of the sample

$$68{\cdot}15-67{\cdot}85={\cdot}70$$

Since difference is ten times of its S.E., hence the data are inconsistent with the assumption that the means of the two populations are equal and we conclude that Austrians on average are taller than Englishmen.

Problem 492.—It is known that the mean and standard deviation of a variable are respectively 100 and 10 in the universe. It is, however, considered sufficient to draw a sample of sufficient size but such as to ensure that the mean of the sample would be, in all probability, within ·01% of the true value. How much would the cost be (exclusive of overhead charges) if the charges for drawing 100 members of a sample be one rupee?

(*I.A.S. 1947*)

Solution :

On the assumption that simple sampling conditions hold good, the mean of the sample cannot differ from the mean of the universe by more than 3 times standard error of the mean of the sample.

So, in a universe with a mean of 100, if the mean of the sample is to be within ·01 limits of n should be such that

$$3. \frac{\sigma}{\sqrt{n}} = \cdot 01$$

or

$$3 \frac{10}{\sqrt{n}} = \cdot 01$$

$$\therefore \quad n = 9000000$$

Thus 9000000 members have to draw from the universe to get the given condition :

Charges are=90000 rupees.

Problem 493.—What is meant by the standard error and what are its practical uses?

Intelligence tests on two groups of boys and girls, give the following results. Examine if the difference is significant :—

Girls	Mean 84 ;	S.D. 10 ;	No. 121
Boys	Mean 81 ;	S.D. 12 ;	No. 81

(P.C.S. 1943)

Solution :

Assuming that the two samples are independent and come from different universes under simple sampling conditions, we have

$$E^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$= \frac{(10)^2}{121} + \frac{(12)^2}{81}$$

$$= 2 \cdot 6042$$

$$\therefore \quad E = 1 \cdot 61$$

Difference between the means=84−81=3 which is less than 3 times the standard error, hence the difference is due to sampling fluctuations, and the difference is not significant.

Problem 494.—A random sample of 200 villages was taken from Gorakhpur district and the average population per village was found to be 485 with a standard deviation of 50. Another random sample of 200 villages from the same district gave an average population of 510 per village with a standard deviation of 40. Is the difference between the averages of the two samples statistically significant ? Give reasons. *(U.P.C.S. 1949)*

Solution :

Assuming that the two samples are drawn quite independently, we have

$$E^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$= \frac{(50)^2}{200} + \frac{(40)^2}{200}$$

$$= 20 \cdot 5$$

$$\therefore \quad E = 4 \cdot 53$$

The observed difference between the two averages$=510-485=25$ which is greater than 3 times the standard error, hence the difference is not due to fluctuations of sampling. Hence the difference between the averages of the two samples are statistically significant.

Problem 495.—The means of simple samples of 1000 and 2000 are 67·5 and 68·0 inches respectively. Can the samples be regarded as drawn from the same population of S.D. 2·5 ? (*M.Sc. Agra, 1952*)

Solution :

Assuming that the two samples are drawn from the same population we have

$$E^2=\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$
$$=(2{\cdot}5)^2\left(\frac{1}{1000}+\frac{1}{2000}\right)$$
$$=6{\cdot}25\times{\cdot}0015$$
$$={\cdot}009375$$

$\therefore$ $E={\cdot}097$ nearly

Difference between the two means$=68-67{\cdot}5={\cdot}5$

Since difference is greater than 3 times the standard error, hence the difference is not due to sampling fluctuations but the two samples are drawn not from same population.

Problem 496.—Given that for a universe $M_u=66$, $\sigma_u=5{\cdot}5$, what sample size, n, must be used in order that for similar test conditions, the probability that the average value of the sample will be in error by not more than 5% of the average value of the universe shall be ?

(*J. C. Chaturvedi*, Page 413)

Solution :

The error which can be allowed is 5 per cent of 66 or 3·3 and this must correspond to a probability of ·9.

$$\therefore \qquad Q=2\int_0^{\delta}\phi\,(t)\,dt={\cdot}9$$

or

$$=\int_0^{\delta}\phi\,(t)\,dt={\cdot}45$$

From the tables we find $\delta=1{\cdot}645$.

Again since $\delta=\dfrac{M_s-M_u}{}\ \sqrt{n}$

We have $$1{\cdot}645=\frac{3{\cdot}3}{5{\cdot}5}\sqrt{n}$$

$$\therefore \quad n=\left(\frac{5}{3}\times 1{\cdot}645\right)^2$$
$$=8 \text{ approximately.}$$

Problem 497.—A sample of 900 members found to have a mean 5·7, could it be reasonably as a simple sample from a large population whose mean is 4·5 and standard deviation is 2·8.

Solution :

The standard error of the mean

$$=\frac{\sigma}{\sqrt{n}}$$
$$=\frac{2{\cdot}8}{\sqrt{900}}$$
$$={\cdot}093$$

Deviation of the mean of the sample from that of the population $=5{\cdot}7-4{\cdot}5$
$$=1{\cdot}2$$

Since the difference is greater than 3 times the standard error, hence the difference is not due to simple sampling fluctuations.

Problem 498.—To study the correlation between the stature of the father and the stature of the son, a sample of 1200 is taken from the universe of fathers and sons. The sample study gives the correlation between the two to be ·46. Within what limits does it hold true for the universe ?

Solution :

The standard error of the Coefficient of correlation is given by

$$E=\frac{1-r^2}{\sqrt{n}}$$
$$=\frac{1-({\cdot}46)^2}{\sqrt{1200}}$$
$$=\frac{1-{\cdot}2116}{34{\cdot}64}$$
$$=\frac{{\cdot}7884}{34{\cdot}64}$$
$$={\cdot}0227$$

If the sampling was simple, the correlation in the universe cannot differ from the correlation in the sample by more than three times the standard error. Hence the correlation in the universe lies between $\cdot 46-3\times\cdot 0227$ and $\cdot 46+3\times\cdot 0227$ that is between ·3919 and ·5281.

Logarithms

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170						5	9	13	17	21	26	30	34	38
						0212	0253	0294	0334	0374	4	8	12	16	20	24	28	32	36
11	0414	0453	0492	0531	0569						4	8	12	16	20	23	27	31	35
						0607	0645	0682	0719	0755	4	7	11	15	18	22	26	29	33
12	0792	0828	0864	0899	0934						3	7	11	14	18	21	25	28	32
						0969	1004	1038	1072	1106	3	7	10	14	17	20	24	27	31
13	1139	1173	1206	1239	1271						3	6	10	13	16	19	23	26	29
						1303	1335	1367	1399	1430	3	7	10	13	16	19	22	25	29
14	1461	1492	1523	1553	1584						3	6	9	12	15	19	22	25	28
						1614	1644	1673	1703	1732	3	6	9	12	14	17	20	23	26
15	1761	1790	1818	1847	1875						3	6	9	11	14	17	20	23	26
						1903	1931	1959	1987	2014	3	6	8	11	14	17	19	22	25
16	2041	2068	2095	2122	2148						3	6	8	11	14	16	19	22	24
						2175	2201	2227	2253	2279	3	5	8	10	13	16	18	21	23
17	2304	2330	2355	2380	2405						3	5	8	10	13	15	18	20	23
						2430	2455	2480	2504	2529	3	5	8	10	12	15	17	20	22
18	2553	2577	2601	2625	2648						2	5	7	9	12	14	17	19	21
						2672	2695	2718	2742	2765	2	4	7	9	11	14	16	18	21
19	2788	2810	2833	2856	2878						2	4	7	9	11	13	16	18	20
						2900	2923	2945	2967	2989	2	4	6	8	11	13	15	17	19
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8

Logarithms

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

Anti-Logarithms

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
·01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
·02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
·03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
·04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
·05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
·06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
·07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
·08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
·09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
·10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
·11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
·12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
·13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
·14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
·15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
·16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
·17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
·18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
·19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
·20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
·21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
·22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
·23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
·24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
·25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
·26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
·27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
·28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
·29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
·30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
·31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
·32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
·33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
·34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
·35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
·36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
·37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
·38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
·39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
·40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
·41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
·42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
·43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
·44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
·45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
·46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
·47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
·48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
·49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

Anti-Logarithms

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
·50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
·51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
·52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
·53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
·54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
·55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
·56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
·57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
·58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
·59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
·60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
·61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
·62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
·63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
·64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
·65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
·66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
·67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
·68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
·69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
·70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
·71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
·72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
·73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
·74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
·75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
·76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
·77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
·78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
·79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
·80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
·81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
·82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
·83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
·84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
·85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
·86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
·87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
·88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
·89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
·90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
·91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
·92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
·93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
·94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
·95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
·96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
·97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
·98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
·99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

APPENDIX

SQUARES, SQUARE-ROOTS AND RECIPROCALS

No. n	Square n^2	Square Root $\sqrt{n}$	Reciprocal $1/n$	No. n	Square n^2	Square root $\sqrt{n}$	Reciprocal $1/n$
			0·				0·
1	1	1·000		26	676	5·099	0384
2	4	1·414	5000	27	729	5·196	0370
3	9	1·732	3333	28	784	5·291	0357
4	16	2·000	2500	29	841	5·385	0345
5	25	2·236	2000	30	900	5·477	0333
6	36	2·449	1666	31	961	5·567	0322
7	49	2·645	1428	32	1024	5·656	0312
8	64	2·828	1250	33	1089	5·744	0303
9	81	3·000	1111	34	1156	5·830	0294
10	100	3·162	1000	35	1225	5·916	0285
11	121	3·316	0909	36	1296	6·000	0277
12	144	3·464	0833	37	1369	6·082	0270
13	169	3·605	0769	38	1444	6·164	0263
14	196	3·741	0714	39	1521	6·244	0256
15	225	3·872	0666	40	1600	6·324	0250
16	256	4·000	0625	41	1681	6·403	0243
17	289	4·123	0588	42	1764	6·480	0238
18	324	4·242	0555	43	1849	6·557	0232
19	361	4·358	0526	44	1936	6·633	0227
20	400	4·472	0500	45	2025	6·708	0222
21	441	4·582	0476	46	2116	6·782	0217
22	484	4·690	0454	47	2209	6·855	0212
23	529	4·795	0434	48	2304	6·928	0208
24	576	4·898	0416	49	2401	7·000	0204
25	625	5·000	0400	50	2500	7·071	0200

Squares, Square-Roots and Reciprocals (Contd.)

No. n	Square n^2	Square root $\sqrt{n}$	Reciprocal $1/n$	No. n	Square n^2	Square root $\sqrt{n}$	Reciprocal $1/n$
			0·				0·
51	2601	7·141	1960	76	5776	8·717	1315
52	2704	7·211	1923	77	5929	8·747	1298
53	2809	7·280	1886	78	6084	8·831	1282
54	2916	7·348	1851	79	6241	8·888	1265
55	3025	7·416	1818	80	6400	8·944	1250
56	3136	7·483	1785	81	6561	9·000	1234
57	3249	7·549	1754	82	6724	9·055	1219
58	3364	7·615	1724	83	6889	9·110	1204
59	3481	7·681	1694	84	7056	9·165	1190
60	3600	7·745	1666	85	7225	9·219	1176
61	3721	7·810	1639	86	7396	9·273	1162
62	3844	7·874	1612	87	7569	9·327	1149
63	3969	7·937	1587	88	7744	9·380	1136
64	4096	8·000	1562	89	7921	9·433	1123
65	4225	8·062	1538	90	8100	9·486	1111
66	4356	8·124	1515	91	8281	9·539	1098
67	4489	8·185	1492	92	8464	9·591	1086
68	4624	8·246	1470	93	8649	9·643	1075
69	4761	8·306	1449	94	8836	9·695	1063
70	4900	8·366	1428	95	9025	9·746	1052
71	5041	8·426	1408	96	9216	9.797	1041
72	5184	8·485	1388	97	9409	9·848	1030
73	5329	8·544	1369	98	9604	9·899	1020
74	5476	8·602	1351	99	9801	9·949	1010
75	5625	8·660	1333	100	10000	10·00	1000

H_2O
H_2O_2

Casteroil

Glucose

$AgNO_3$

K. Na. tartrate

$SnCl_2$

Cu ... Acid

Formaldehyde

Amonia

125
125
725
250
125
15725