15.05.16

Frequency Distribution

num of value

Frequency :—

distribution

Frequency Distribution :—

upper limit, lower limit

Class limit :—

difference between upper limit & lower limit

Class interval :— upper limit − lower limit

class mid-point :— upper limit + lower limit / 2

Exclusive Method :—

Inclusive Method :—

Frequency distribution of marks

Marks	No of students		corrected class interval
0 − 10	4	0 − 9	−0.5 − 9.5
10 − 20	10	10 − 19	9.5 − 19.5
20 − 30	20	20 − 29	19.5 − 29.5
30 − 40	7	30 − 39	29.5 − 39.5
40 − 50	5	40 − 49	39.5 − 49.5
N =	50		

Exclusive Method

inclusive method

$$\text{Correction factor} = \frac{\text{lower limit of 2nd group} - \text{upper limit of 1st group}}{2}$$

$$= \frac{10-9}{2}$$

$$= 0.5$$

How to construct a complete frequency distribution:-

Ex :- The profits in lac of taka of 50 companies for the year 2014-15 are given below :—

66, 62, 45, 79, 32, 51, 56, 60, 51, 49, 25, 42, 54, 54, 58, 70, 43, 58, 52, 50, 38, 67, 80, 59, 48, 65, 71, 30, 46, 55, 82, 51, 63, 45, 53, 40, 35, 56, 70, 52, 67, 55, 57, 30, 63, 42, 74, 58, 44, 55.

Construct a complete frequency distribution / Relative frequency distribution using 5-step procedure.

group the data / balance the data in a proper method

Step 1 :- Decide on the number of classes or groups.

$$2^k \geq n \quad \text{—— (1)}$$

$K = 1, \quad 2^1 = 2 < 50$

$K = 2, \quad 2^2 = 4 < 50$

$\vdots$

$K = 5, \quad 2^5 = 32 < 50$

$K = 6, \quad 2^6 = 64 > 50$

$K = 6$ is the number of groups.

Step 2 :- Determine the class interval or class-width.

$$i \geq \frac{H - L}{K}$$

$$= \frac{82 - 25}{6}$$

$$= 9.5$$

$$\approx 10$$

if possible multiple of 5 or 10
avoid complex value
eg: 9, 22, 11, 27

Step 3 :— Set the individual class limits :

25 - 35, 35 - 45, 45-55, 55 - 65, 65-75,
75 - 85

Step 4 :— use tally marks (।) for each number.

Step 5 :— Count the number of item in each class.

Table :- Frequency Distribution of Profits

Profit lac taka class interval	Tally Marks (।)	No-of Companies frequency, fi	Cumulative frequency	Relative frequency
25 - 35	||||	4	4	4/50 = 0.8
35 - 45	$\cancel{||||}$ ||	7	11	7/50 =
45-55	$\cancel{||||}$ $\cancel{||||}$ $\cancel{||||}$	15	26	15/50 =
55 - 65	$\cancel{||||}$ $\cancel{||||}$ ||||	14	40	14/50 =
65- 75	$\cancel{||||}$ |||	8	48	8/50 =
75 - 85	||	2	50	2/50 =

$$\text{Relative frequency} = \frac{\text{group frequency}}{\text{total frequency}}$$

Graphical Representation :-

(i) Histogram

(ii) Frequency Polygon

(iii) Bar diagram

(iv) ~~Cuma~~ Cumulative frequency Polygon (ogive)

(v) Pie diagram

* graph next class

22.05.16

Histogram :

series of rectangle

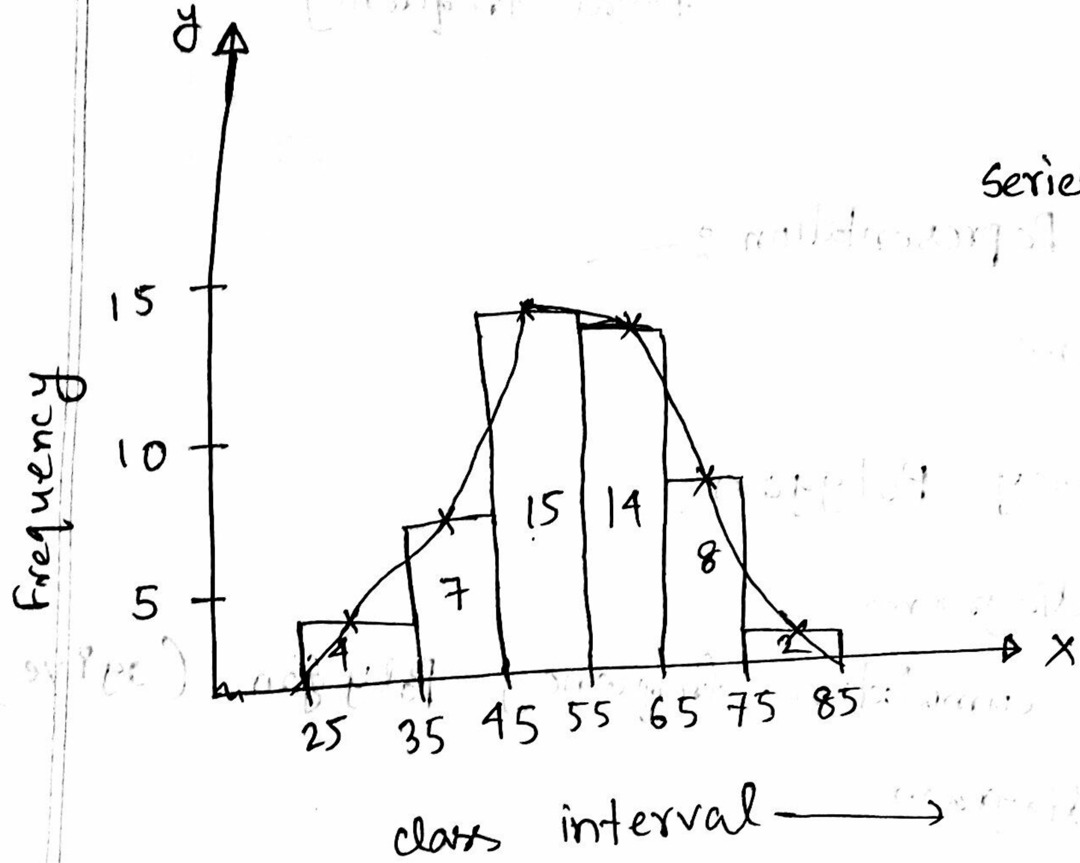


Fig :- Histogram of frequency distribution of profits

Bar diagram:

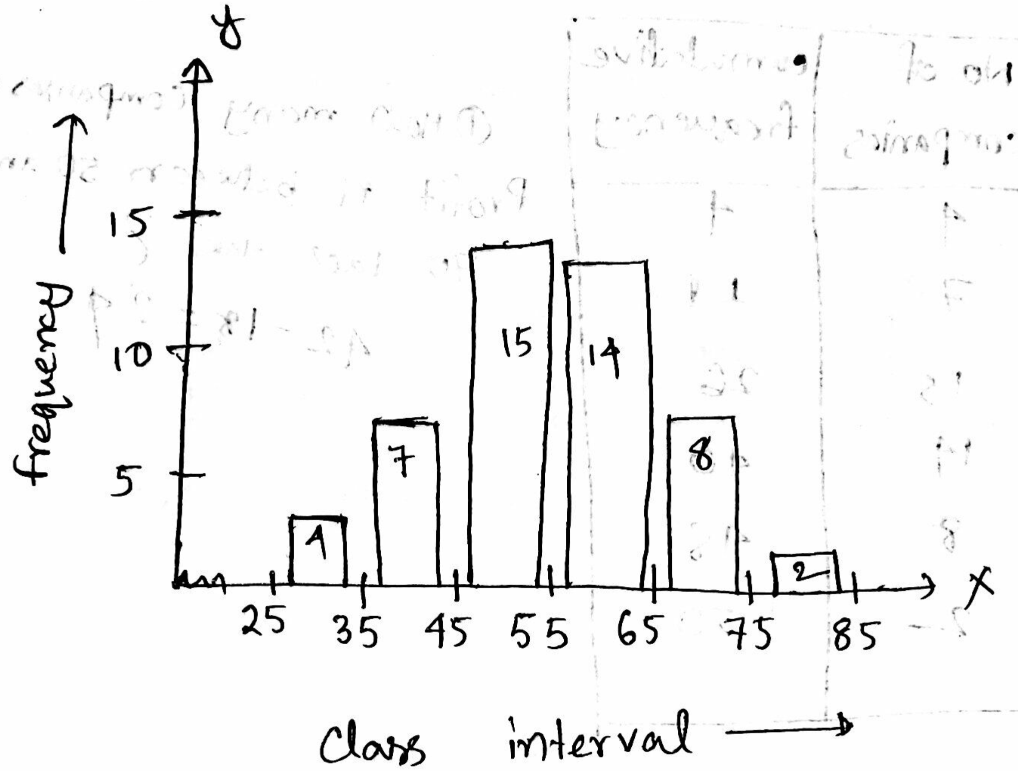


fig: Bar diagram of frequency distribution of profits

Cumulative frequency polygon (ogive)

profit Lac taka	No of companies	cumulative frequency
25-35	4	4
35-45	7	11
45-55	15	26
55-65	14	40
65-75	8	48
75-85	2	50

① How many companies make profit of between 50 and 70 lac taka?

42 - 18 = 24

Median value = $\frac{50}{2}$ = 25th value

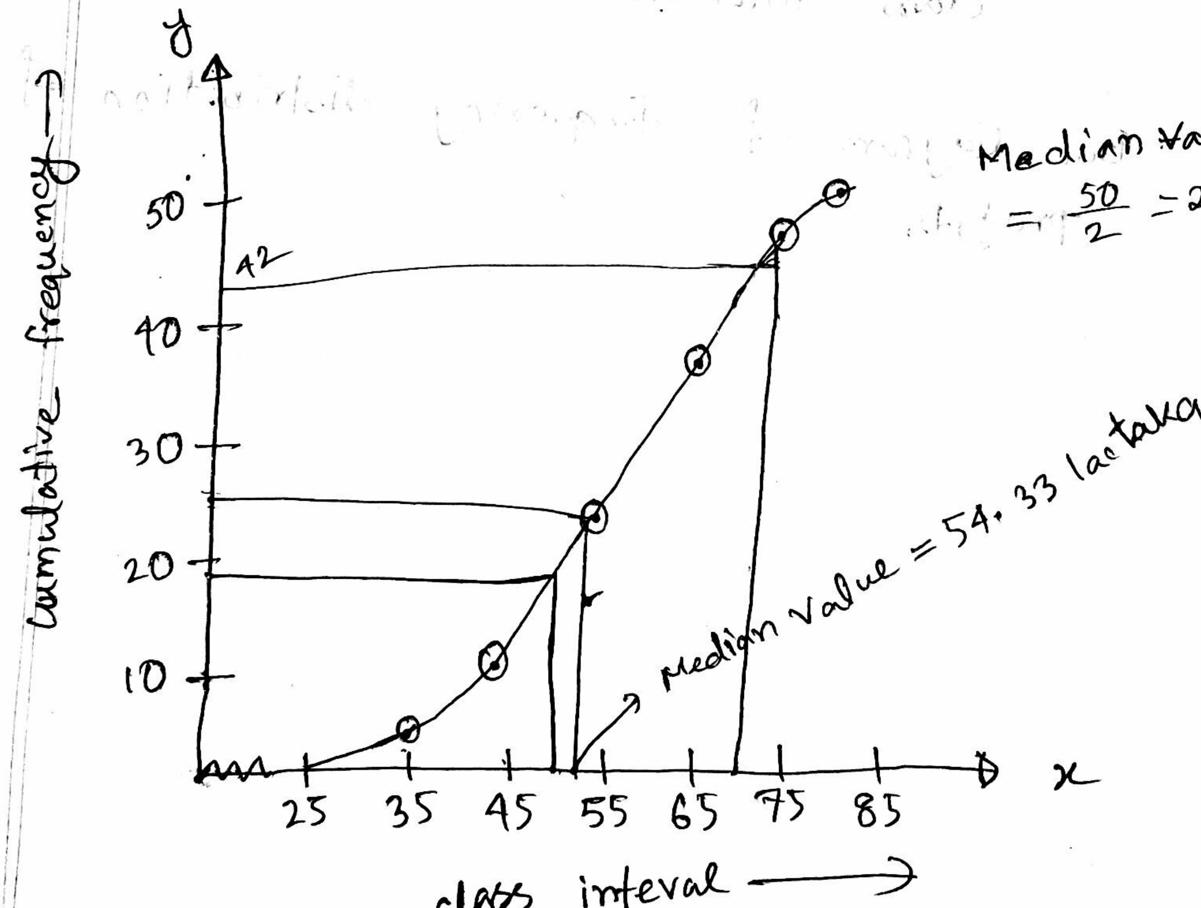


Fig:- cumulative frequency polygon

Pre-diagram :

Draw a Pie-diagram based on the following information :—

$$\text{Angle} = \frac{\text{group value}}{\text{total value}} \times 360°$$

Forest types	Area in square km	Angle
Evergreen	7820.45	173.12
Most-deciduous	1029.4	22.79
Mangrove	7412.41	164.09
total :-	16262.26	360

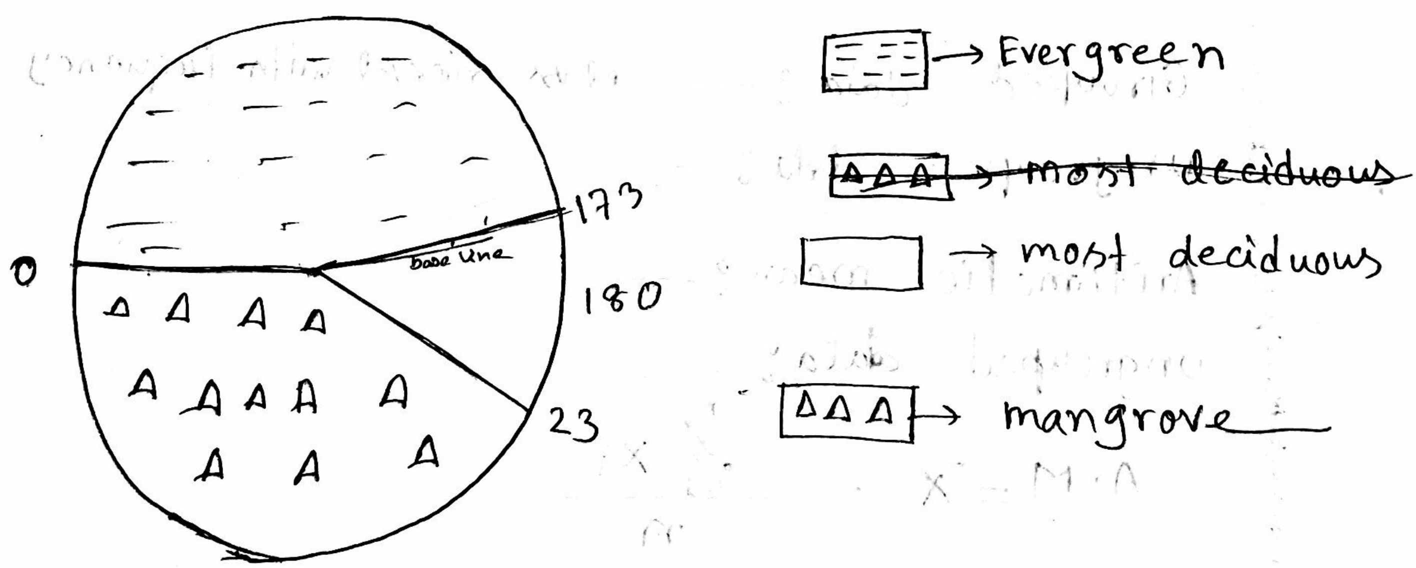


Fig : Pie diagram ~~of~~ of different forest types

Measures of central tendency

Central tendency :—

Five Measure :—

1. Arithmetic Mean (A.M) or Mean.
2. Geometric Mean (G.M)
3. Harmonic Mean (H.M)
4. Median (Me)
5. Mode (Mo)

Grouped data :— class interval with frequency

ungrouped data :—

Arithmetic mean :—

ungrouped data :—

$$A.M = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

29.05.16

Ex: The following data represent the length of life, in seconds, for a sample of 50 fruit flies subject to a new spray in a controlled laboratory experiment:

17, 10, 20, 9, 23, 13, 12, 19, 18, 24

Calculate the mean of the above sample.

Soln: We know,

$$A.M = \bar{X} = \frac{\sum X_i}{n}$$

$$= \frac{17+10+20+\cdots+24}{10}$$

$$= 16.5$$

Properties of A.M :-

1. Every set of interval or ratio level data has a mean.

2. All the values are included in computing the mean.

3. The mean is unique.

4. The sum of derivation of each value from their mean is zero. Symbolically.

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

X_i	$\bar{X}$	$(X_i - \bar{X})$
17		0.5
10		-6.5
20		3.5
9	16.5	-7.5
23		6.5
13		-3.5
12		-4.5
19		2.5
18		1.5
24		7.5
		$\sum(X_i - \bar{X}) = 0$

: Grouped data :

Shortcut Method :-

$$A.M = \bar{x} = A + \frac{\sum f_i d_i}{N} \times C$$

where,

A = middle value of the class midclass.

$$d_i = \frac{x_i - A}{C}$$

C = class interval

directed Method :-

$$A.M = \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

x_i = Class-mid value

f_i = frequency

$N = \sum f_i$ = Total number of observation.

Exa: From the following frequency distribution of battery life calculate the mean life.

class interval year	frequency, f_i	corrected class interval	x_i = class mid-values	$f_i x_i$
1.5 – 1.9	2	1.45 – 1.95	1.7	3.4
2.0 – 2.4	1	1.95 – 2.45	2.2	2.2
2.5 – 2.9	4	2.45 – 2.95	2.7	10.8
3.0 – 3.4	15	2.95 – 3.45	3.2	48
3.5 – 3.9	10	3.45 – 3.95	3.7	37
4.0 – 4.4	~~10~~ 5	3.95 – 4.45	4.2	21
4.5 – 4.9	3	4.45 – 4.95	4.7	14.1
	$\sum f_i = 40$			$\sum f_i x_i = 136$

$\therefore$ A.M $= \bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$

$= \frac{136.5}{40}$

$= 3.41$ year

So, the mean lifetime of the battery is 3.41 year.

Geometric Mean :— It is useful in finally the average change in percentage ratio, growth rate etc.

Ungrouped data :—

~~A~~ ~~A~~.M = nth root of product of nth observation

$= (x_1, x_2 \cdots x_n)^{1/n}$

$\log G.M = \frac{1}{n} \log (x_1, x_2 \cdots x_n)$

$= \frac{1}{n} (\log x_1 + \log x_2 + \cdots + \log x_n)$

$\log G.M = \frac{\sum_{i=1}^{n} \log x_i}{n}$

$G.M = \text{Antilog} \left(\frac{\sum \log x_i}{n} \right)$

Grouped data :-

$\cancel{G.M = (x_1 f_1 \cdot x_2 f_2}$

$$G.M = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)^{1/N}$$

$$\log G.M = \frac{\sum_{i=1}^{n} f_i \log x_i}{N}$$

Ex :- The annual rates of growth of output of a factory in 5 years are 5.0 %, 7.5 %, 2.5 %, 5.0 % and 10.0 % respectively. what is tee average rate of growth of output for tee period?

* Value % এ থাকে, তবে geometric mean appropriate.

Let, the initial output is 100

x_i	$\log x_i$
105.0	2.021
107.5	2.031
102.5	2.011
105.0	2.021
110.0	2.041
	$\sum \log x_i = 10.125$

we know,

$$\log G.M = \frac{\sum \log x_i}{n}$$

$$= \frac{10.125}{5}$$

$$= 2.025$$

$$G.M = \text{Antilog}(2.025)$$

$$= 105.93$$

12/06/15

1st ↔ center tendency

So, the average Rate of growth of output for the period is $105.93 - 100 = 5.93\%$

* when it is not possible to find G.M and H.M:-

G.M

(i) Zero

(ii) odd values negative

(iii) product negative

H.M

zero values

Harmonic Mean :- Per unit of time rate/speed

Ungrouped date :-

$$H.M = \frac{n}{\sum \frac{1}{x_i}}$$

$$H.M = \frac{N}{\sum \frac{f_i}{x_i}}$$

05.06.16

Ex : In a certain factory a unit of work is completed by A in 4 min, by B in 5 min, by C in 6 min, by D in 10 min and by E in 12 min. what is the average number of unit of work completed per minute?

	x_i	$1/x_i$
A	4	.25
B	5	.20
C	6	.167
D	10	.100
E	12	.083

$\sum 1/x_i = 0.8$

We know,

$$H.M = \frac{n}{\sum_{i=1}^{n} 1/x_i}$$

$$= \frac{5}{0.8}$$

$= 6.25$ minute

So, the average number of unit of work completed per minute is $\frac{1}{6.25} = 0.16$ unit

Relation between A.M, G.M and H.M :—

For same data,

$$A.M \geq G.M \geq H.M$$

check :— (i) 10, 20, 30, 40, 60

(ii)

Profit. Lac taka	No of companies, f_i	x_i	$f_i x_i$	$f_i \log x_i$	f_i / x_i
25-35	4	30			
35-45	7	40			
45-55	15	50			
55-65	14	60			
65-75	8	70			
75-85	2	80			
			$\Sigma f_i x_i =$	$\Sigma f_i \log x_i =$	$\Sigma f_i/x_i =$

Median :— Middle most value in a set of data arranged in ascending order.

Ungrouped data :—

Even No. of observations : Media = Average of $\left(\frac{n}{2}\right)$ and $\left(\frac{n}{2}+1\right)^{th}$ observation in a series arranged in ascending order.

For odd No of observations :—

Median = $\left(\frac{n+1}{2}\right)^{th}$ observation in a series arranged in ascending order.

Ex: Following data represents the ages of children currently attending child-care :—

8, 5, 9, 10, 9, 12, 7, 7, 12, 13, 7, 8

Calculate median and mode of the age of the children.

Solⁿ :— At first arrange the data in ascending order

5 , 7, 7 , 7 , 8, 8, 9, 9 , 10, 12 , 12, 13

Ungrouped data :—

Since $n = 12$ is even,

Average of $\left(\frac{n}{2}\right)$ and $\left(\frac{n}{2}+1\right)$th observations

$= \text{" " } \left(\frac{12}{2}\right)$ and $\left(\frac{12}{2}+1\right)$ " "

= " " 6th and 7th " "

$= \frac{8+9}{2}$

= 8.5 years

Grouped data :—

Median group = Size of $\frac{N}{2}$

$$\text{Median} = L + \frac{\frac{N}{2} - p.c.f}{f_m} \times C$$

where, L = Lower Limit of median group

$N = \sum f_i$ = Total No of observations

P.c.f = Cumulative frequency of group just preceding median group

C = class interval

Ex :— Find the median and mode of the following frequency distribution :—

Profit Lac taka	No of companies	Cumulative frequency, f
25 − 35	4	4
35 − 45	7	11
$\boxed{45 - 55}$	15	26
55 − 65	14	40
65 − 75	8	48
75 − 85	2	50
	$\sum f_i = 50$	

→ median group

Grouped data :—

Median group = size of $\frac{N}{2} = \frac{50}{2} = 25^{th}$

$$\text{Median} = L + \frac{N/2 - p.c.f}{f_m} \times C$$

$$= 45 + \frac{50/2 - 11}{15} \times 10$$

$$= 54.33 \text{ lac taka}$$

[illegible]

Mode :—

Ungrouped data :—

Mode = the value in the series x_i which maximum frequency.

Ex — Child — care :— data

x_i	Tally	f_i
5	/	1
7	///	3
8	//	2
9	//	2
10	/	1
12	//	2
13	/	[illegible]

Since observation 7 has maximum frequency.

So, mode age is 7 years

Grouped data :—

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

$\Delta_1 \to 15-7$, $\Delta_2 \to 15-14$

$$= 45 + \frac{8}{8+1} \times 10$$

$= 53.89$ Lac taka

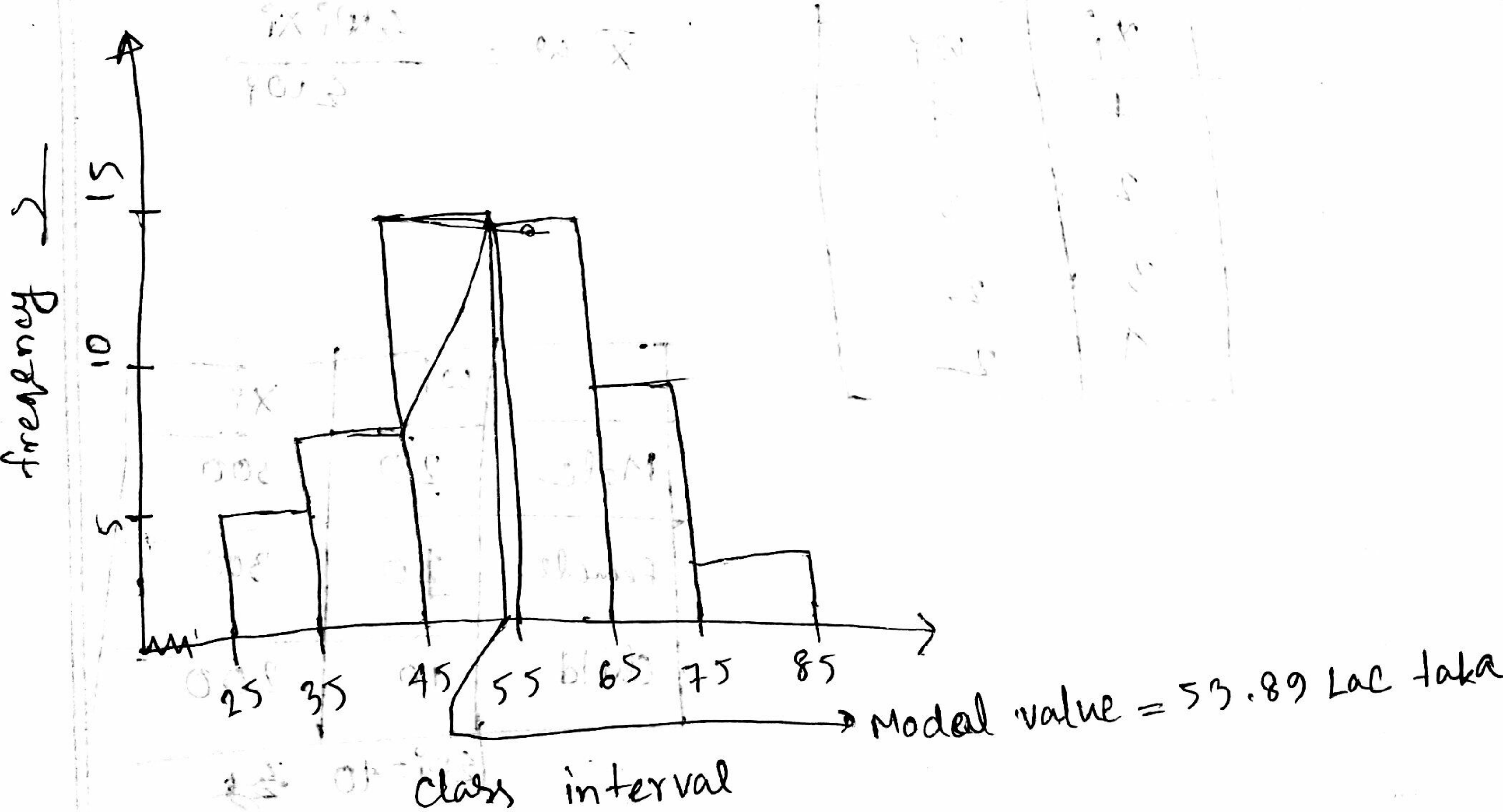
where,

L = Lower Limit of modal group

Δ_1 = difference between the frequency of modal group and pre-modal group

Δ_2 = " " " modal group and post modal group

C = class interval



Relation between Mean (A.M), Median and Mode:

Mode:

$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

Weighted Average:

1111 222 33 44

x_i	w_i
1	4
2	3
3	2
4	2

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

	w_i	x_i
Male	20	300
Female	10	300
Child	10	200

$\sum w_i = 40$

19.06.16

Measures of Dispersion

Absolute Measure : (i) Range
(ii) standard deviation
(iii) Mean deviation
(iv) Quartiled deviation

Relative Measure : (i) Co-efficient of Range
(ii) co-efficient standard deviation
(iii) " Mean deviation
(iv) " ~~Quart~~ Quartile deviation

Range :—

Ungrouped data :—

Range = Highest Value − Lowest value

$$x_H - x_L$$

$$\text{Co-efficient of Range} = \frac{x_H - x_L}{x_H + x_L} \times 100$$

* Standard deviation কম হওয়া মানে, consistancy, stability বেশি হওয়া, variation কম হওয়া।

Grouped data :—

Range = Upper limit of last class interval − Lower limit of 1st class interval

$= X_u - X_L$

Co-efficient of Range :—

$$\frac{X_U - X_L}{X_U + X_L} \times 100$$

Standard deviation :—

The arithmetic mean of the squared deviation of observation from their arithmetic mean is known as variance. The positive square root of variance is known as standard deviation.

	Mean	Standard deviation	denominator
population	μ	σ	N
sample	$\overline{X}$	s	$n-1$, $N-1$

Standard deviation	lower →	Stability, consistency higher	Variability ~~low~~ low
"	higher	" lower	" high

Ungrouped sample data :-

Sample Mean, $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Sample Variance, $S^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

Sample Standard deviation, $S = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}}$

$\sum (x_i - \bar{x})^2$

$= \sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2)$

$= \sum x_i^2 - 2 \sum x_i . \bar{x} + n\bar{x}^2$

$= \sum x_i^2 - 2n\bar{x} . \bar{x} + n\bar{x}^2$

$= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$

$= \sum x_i^2 - n\bar{x}^2$

$= \sqrt{\dfrac{\sum x_i^2 - n\bar{x}^2}{n-1}}$

Grouped sample data :-

Sample mean, $\bar{x} = \dfrac{\Sigma f_i x_i}{N}$

Sample variance, $s^2 = \dfrac{\Sigma f_i (x_i - \bar{x})^2}{N-1}$

Sample standard deviation, $s = \sqrt{\dfrac{\Sigma f_i (x_i - \bar{x})^2}{N-1}}$

$= \sqrt{\dfrac{\Sigma f_i x_i^2 - N\bar{x}^2}{N-1}}$

Ungrouped population data :-

population Mean, $\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$

Population variance, $\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$

Population standard deviation, $\sigma = \sqrt{\dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$

$= \sqrt{\dfrac{\Sigma x_i^2 - N\mu^2}{N}}$

Grouped population data :—

$$\text{Population mean, } \mu = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

$$\text{Population variance, } \sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \mu)^2}{N}$$

$$\text{Population standard deviation, } \sigma = \sqrt{\frac{\sum f_i (x_i - \mu)^2}{N}}$$

$$= \sqrt{\frac{\sum f_i x_i^2 - N\mu^2}{N}}$$

Co-efficient of variation :—

The ratio of standard deviation to its arithmetic Mean expressed as a percentage.

$$C.V = \frac{S}{\overline{X}} \times 100 \quad \text{(for sample)}$$

$$C.V = \frac{\sigma}{\mu} \times 100 \quad \text{(for population)}$$

when do we use C.V :—

1. when data are in different unit such as ~~in com~~ income in dollars No. of days absent.

2. when data are in same unit, but mean are far apart :

such as income of top executive

income of unskilled employee.

17.07.16

Ex:– the following data represent the daily production of coal in two mines :–

Mine 1 :– 4500 4400 4000 4560

Mine 2 : 5150 5250 5500 5555 5000

Compare the variaability in the two data using an appropriate measure. consider (i) sample
(ii) population

Calculation for sample

Mine 1

x_i	$\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
4500	$\frac{\sum x_i}{n}$	135	
4400	$= \frac{17460}{4}$	35	
4000		−365	
4560	$= 4365$	195	

$\sum x_i = 17460$ $\sum (x_i - \bar{x})^2 = 190700$

Sample Mean,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{17460}{4} = 4365$$

Sample standard deviation

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{190700}{4-1}} = 252.124$$

Co-efficient of variation,

$$C.V = \frac{s}{\overline{x}} \times 100$$

$$= \frac{252.124}{4365} \times 100$$

$$= 5.77\%$$

Mine 2

x_i	$\overline{x}$	$(x_i-\overline{x})$	$(x_i-\overline{x})^2$
5150	$\frac{\sum x_i}{n}$	-141	19881
5250	$= \frac{26455}{5}$	-41	1681
5500	$=5291$	209	43681
5555		264	69696
5000		-291	84681

$\sum x_i = 26455$ $\quad\quad$ $\sum (x_i-\overline{x})^2 = 219620$

~~Sam~~ Sample Mean,

$$\overline{x} = \frac{\sum x_i}{n}$$

$$= \frac{26455}{5}$$

$$= 5291$$

Sample standard deviation,

$$s = \sqrt{\frac{\sum(x_i-\overline{x})^2}{n-1}}$$

$$= \sqrt{\frac{219620}{4}}$$

$$= 234.318$$

Co-efficient of variation,

$$C.V = \frac{s}{\overline{x}} \times 100$$

$$= \frac{234.318}{5291} \times 100$$

$$= 4.43\%$$

Since the C.V of sample of mine 1 is 5.77% and C.V of sample of mine 2 is 4.43%

So, there is more variation in the sample of mine 1 relative to distribution of mine 2. So, sample of mine 1 is more variable.

Grouped data :—

Ex:— The following data represent the electricity cost for a sample of 50 two bedroom apartment and 50 three bedroom apartment in a city of New Mexico during the month of May last year:

Electricity cost	Two bed room frequency	Three bed-room
80 – 100	4	3
100–120	5	10
120 – 140	20	25
140 – 160	5	5
160 –180	12	3
180 –200	4	4

(i) sample
(ii) population

Electricity cost of which distribution show more consiste

Considering Population :—

Electricity cost	two bed room frequency, f_i	x_i = mid-value	$f_i x_i$	$f_i x_i^2$
				32400
80–100	4	90	360	
100–120	5	110	550	
120–140	20	130	2600	
140–160	5	150	750	
160–180	12	170	2040	
180–200	4	190	760	
			$\Sigma f_i x_i = 7060$	$\Sigma f_i x_i^2 = 1034600$

Population Mean, $\mu = \frac{\Sigma f_i x_i}{N}$

$= \frac{7060}{50}$

$= 141.2$ \$

Population standard deviation ;

$$\sigma = \sqrt{\frac{\Sigma f_i x_i^2 - N\mu^2}{N}}$$

$$= \sqrt{\frac{1034600 - 50(141.2)^2}{50}} = 27.47$$

* population. complete
* sample
* variability compare

co-efficient of variation, $C.V = \frac{\sigma}{\mu} \times 100$

$= \frac{27.47}{141.2} \times 100$

$= 19.45\%$

Ex :- Distribution of test score

Mean test score 200

Standard deviation 40

Distribution of year of Service

Mean year of service 20 years

Standard deviation 2 year

compare variability using an appropriate measure

C.V of test score

$C.V = \frac{S}{\bar{X}} \times 100$

$= \frac{40}{200} \times 100$

$= 20\%$

C.V of year of service

$C.V = \frac{S}{\bar{X}} \times 100$

$= \frac{2}{20} \times 100$

$= 10\%$

24.07.16

Probability : A value between 0 and 1 describing the relative possibility that an event will occur.

Probability
- objective
  - Classical
  - Empirical
- subjective → based on ~~a~~available information.

Classial probability : Every outcome has an equal chance of being selected.

$$P(A) = \frac{\text{no of favourable outcome}}{\text{Total of all possible outcomes}}$$

$P(\text{head}) = \frac{1}{2}$, $P(\text{tai}) = \frac{1}{2}$

$P(1) = \frac{1}{6} \; \{1, 2, 3, 4, 5, 6\}$

!

$P(6) = \frac{1}{6}$

P (impossible event) = 0

P (The ~~sum~~ sum will rise in the west) = 0

P (universal truth) = 1

P (the sun rises in the east) = 1

Experiment:

Out come:

Random Experiment:

Event

mutually Exclusive Event:

collectively Exhaustive:

Independent and dependent Event:

$\cap$ — and / both , $\cup$ – either or

Conditional Probability :—

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

law of independence :– if two events are independent

$$P(A \cap B) = P(A) \cdot P(B)$$

Law of Addition :—

P (either A or B or both)

$= P(A \cup B) = P(A) + P(B) - P(A \cap B)$

[if A and B are not mutually exclusive]

$= P(A) + P(B)$

Law of multiplication :—

$$P(A \cap B) = P(A) \cdot P(B/A)$$

$$P(A \cap B) = P(B) \cdot P(A/B)$$

Complementary Law :-

$$P(A) + P(\bar{A}) = 1$$

Properties of probability :-

(i) $0 \leq P(A_i) \leq 1$

(ii) $\sum_{i=1}^{s} P(A_i) = 1$

Empirical Probability :

$$= \frac{\text{No of times event occured in the past}}{\text{Total No of observation}}$$

subjective probability :-

based on available information

Ex :— A box contain 5 white balls and 8 red balls all of which are of equal size. A ball is drawn from the ~~box~~ box at random. what is the probability that the ball is white?

Soln : Let, A = The ball is white

No of white balls = 5

Total balls = 13

We know,

$$P(A) = \frac{\text{No of favorable outcomes}}{\text{Total of all possible outcomes}}$$

$$= \frac{5}{13}$$

$$= 0.385$$

So, the probability that the ball is white is 0.385

Ex :- If a card is drawn from an ordinary deck of 52 playing cards. Find the probability that (i) The card is a ~~read~~ red
(ii) The card is a diamon
(iii) The card is an ace
(iv) The card is either an ace or a diamond

Soln:
(iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$

$= \frac{16}{52}$

$= \frac{4}{13}$

Let, A = The card is an ace

B = The card is a diamond

$P(A) = \frac{4}{52}$, $P(B) = \frac{13}{52}$

There is only one card which is an ace and also a diamond. $P(A \cap B) = \frac{1}{52}$.

$2^1 \{H, T\}$

$2^2 \{HH, HT, TH, TT\}$

$2^3 \{HHH, HHT, HTH, THH, TTH, THT, HTT,$
$TTT\}$

$6^1 \{1, 2, 3, 4, 5, 6\}$

$6^2 \{$

31.07.16

3. A fair coin is tossed three times. What is the probability that at least one head appears?

Soln: Tossing a coin three times, the possible outcomes are —

$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

Let, A = at least one head appears

There are 7 outcomes with at least one head.

$$P(A) = \frac{\text{No of favourable outcome}}{\text{Total of all possible outcome}}$$

$$= \frac{7}{8}$$

5. If two dice are thrown, write down the sample space of all possible outcomes and find the probability of —

(i) a double six

(ii) a sum of 9 or more.

(i) Let A = a double six

There is only one outcome with a double six.

$$P(A) = \frac{\text{No of favorable outcomes}}{\text{Total of all possible outcomes}}$$

$$= \frac{1}{36}$$

Throwing two dice the possible outcomes are—

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2						
3						
4						
5						
6	(6,1)					(6,6)

(ii) B = a sum of 9 or more

$= \{(3,6), (6,3), (4,6), (6,4), (4,5), (5,4), (5,5)$

$(6,5), (5,6), (6,6)\}$

$$P(B) = \frac{10}{36}$$

7. Two coins are tossed. What is the conditional probability that two head results given that there is at least one head?

$\Rightarrow$ Tossing two coins, the possible outcomes are —

$$S = \{HH, HT, TH, TT\}.$$

Let, A = two head results.

B = There is atleast one head

$A = \{HH\}$, $P(A) = \frac{1}{4}$

$B = \{HH, TH, TH\}$, $P(B) = \frac{3}{4}$

$A \cap B = \{HH\}$, $P(A \cap B) = \frac{1}{4}$

P(Two head results / There is atleast one head)

$$= P(A/B)$$

$$= \frac{P(A \cap B)}{P(B)}$$

$$= \frac{1/4}{3/4}$$

$$= \frac{1}{3}$$

9. One bag contains 4 white balls and 2 black balls and another bag contains 3 white balls and 5 black balls. If one ball is drawn from each bag, find the probability that

(i) both are white

(ii) both are black

(iii) one is white and one is black.

⇒ Let, W_1 = white ball from 1st bag

W_2 = White ball from 2nd bag

B_1 = Black ball from 1st bag

B_2 = Black ball from 2nd bag.

(i) P(both are white)

$= P(W_1 \cap W_2)$

$= P(W_1) \cdot P(W_2)$

$= \frac{4}{6} \times \frac{3}{8}$

(iii) P(one is white and one is black)

$= P$ [either (one white ball from 1st bag and one black from 2nd bag) or (one black ball from 1st bag and one white ball from 2nd bag)]

$= P[(W_1 \cap B_2) \cup (B_1 \cap W_2)]$

$= P(W_1 \cap B_2) + P(B_1 \cap W_2)$ [since $W_1 B_2$ and $B_1 W_2$ are mutually exclusive]

$= P(W_1) \times P(B_2) + P(B_1) \times P(W_2)$ [since the events are independent]

Example: Let us consider a problem with 10 rolls of a film in a box, 3 of which are found defective. Two rolls are to be selected one after the other. What is the probability of selecting a defective roll followed by another defective roll?

(i) without replacement

(ii) with replacement.

(i) without replacement;

Let, A = the first roll of a film is defective

$$P(A) = \frac{\text{No. of favourable outcome}}{\text{Total of all possible}}$$

$$= \frac{3}{10}$$

Let, B = the 2nd roll being and defective

$P(B/A) = \frac{2}{9}$ [After 1st selection there remains 2 defectives out of 9 rolls]

$$P[\text{two defective}] = P(A \cap B)$$
$$= P(A) \cdot P(B/A)$$
$$= \frac{3}{10} \times \frac{2}{9}$$
$$= \frac{1}{15}$$

(ii) with replacement,

$$P(\text{two defectives}) = P(A \cap B)$$
$$= P(A) \cdot P(B/A)$$
$$= \frac{3}{10} \times \frac{3}{10}$$

* $P(\text{three defectives}) = P(A \cap B \cap C)$

$$= P(A) \cdot P(B/A) \cdot P(C/AB)$$

$$= \frac{3}{10} \times \frac{3}{10} \times \frac{3}{10} \text{ (with replacement)}$$

$$= \frac{3}{10} \times \frac{2}{9} \times \frac{1}{8} \text{ (without replacement)}$$

07.08.16

* Ex :- An electrical system consists of four components as illustrated in the following figure. The system work if A and B work either of the components C or D work. The reliability (Probability of working) of each component is also shown in the figure. Find the probability that a) the entire system work
b) the component C does not work, given that the entire system work.

Assume that four components work independently

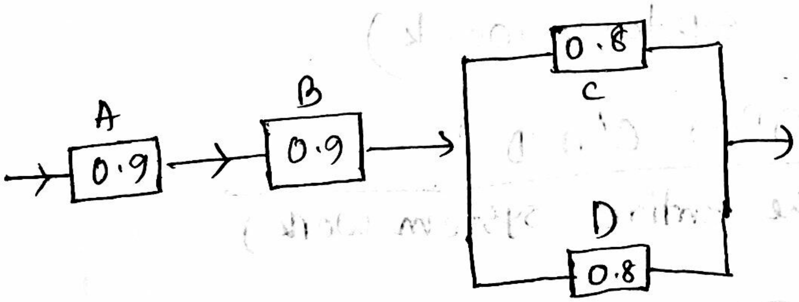


Fig : An electrical system.

ⓐ Clearly the probability that the entire system works can be calculated as the following

$$P(A \cap B \cap (C \cup D)) = P(A) \cdot P(B) \cdot P(C \cup D)$$

$$= P(A) \cdot P(B) [1 - P(C' \cap D')] \quad \left[C \cup D = 1 - (C' \cap D') \right]$$

$$= P(A) \cdot P(B) [1 - P(C') \cdot P(D')]$$

$$= 0.9 \times 0.9 \, [1 - (1 - 0.8)(1 - 0.8)]$$

$$= 0.7776$$

ⓑ P(the system does not work / the entire system work)

$$= \frac{P(A \cap B \cap C' \cap D)}{P(\text{the entire system work})}$$

$$= \frac{P(A) \cdot P(B) \cdot P(C') \cdot P(D)}{0.7776}$$

$$= \frac{0.9 \times 0.9 \times (1 - 0.8) \times 0.8}{0.7776}$$

$$= 0.1667$$

Regression Analysis

A simple Linear Regression Model :-

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \ldots, n$$

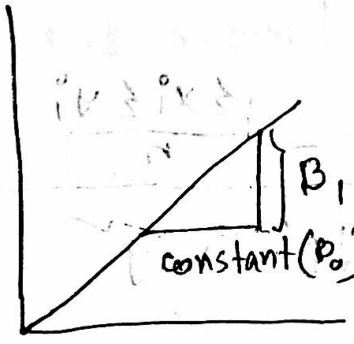
y_i = Dependent or response variable

x_i = Independent or explanatory variable

β_0 = intercept (constant)

β_1 = slope, co-efficient of independent variable

ε_i = error term.



Calculation of regression co-efficients :—

y depend on x :—

$$\hat{\beta}_1 yx = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - n\bar{x}^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Then the estimated or fitted regression line or equation is $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Error, $e_i = y_i - \hat{y}_i$

x depend on y :—

$$\hat{\beta}_1 xy = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum y_i^2 - n\bar{y}^2}$$

$$\hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \bar{y}$$

Then the estimated or fitted regression line or equation is $\hat{x}_i = \hat{\beta}_0 + \hat{\beta}_1 y_i$

Error, $e_i = x_i - \hat{x}_i$

Properties of regression co-efficients :

~~ca~~ co-efficient relation

1. $r = \sqrt{\beta_1 xy \cdot \beta_1 yx}$

2. $\hat{\beta}_1 xy > 1, \quad \hat{\beta}_1 yx < 1$

3. same sign

4. average r এর বেশি হবে

5. +, − করলে effect হবে না, *, / করলে effect হবে.

6.

Co-efficient of determination :

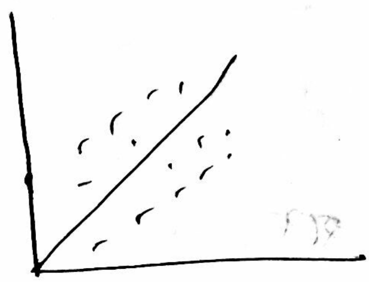
explanatory Power

$$R^2 = r^2 = \beta_1 xy \cdot \beta_1 yx$$

It determines the proportion of total variation in the dependent variable explained by the independent variable.

Standard Error of estimates :—

Variability or scatter of the observed value around the regression line.



Standard Error of y :—

$$SE(y) = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

$$= \sqrt{\frac{\sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i}{n-2}}$$

$$SE(x) = \sqrt{\frac{\sum (x_i - \hat{x}_i)^2}{n-2}}$$

$$= \sqrt{\frac{\sum x_i^2 - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i y_i}{n-2}}$$

14.08.18

Calculate regression co-efficients and fit a regression line when —

① height depends on weight

② weights depends on height.

③ Explanatory power of the regression (co-efficient of determination)

④ standard error

Person	height	weight	x_i^2	y_i^2	$x_i y_i$
1	52	82			
2	55	90			
3	54	95			
4	62	106			
5	63	120			
6	66	132			
7	70	140			
8	71	135			

$\sum x_i = 498$, $\sum y_i = 900$, $\sum x_i^2 = 31320$,

$\sum y_i^2 = 104734$, $\sum x_i y_i = 57048$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{498}{8} = 62.25$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{900}{8} = 112.5$$

X depend on Y :—

$$\hat{\beta}_{1\,xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum y_i^2 - n\bar{y}^2}$$

$$= \frac{57048 - \frac{498 \times 900}{8}}{104734 - 8 \times (112.5)^2}$$

$$= 0.2936$$

$$\hat{\beta}_0 = \bar{x} - \hat{\beta}_1 \bar{y}$$

$$= 62.25 - 0.2936 \times 112.5$$

$$= 29.22$$

then the estimated on fitted regression line is,

$$\hat{x}_i = \hat{\beta}_0 + \hat{\beta}_{1xy}\, y_i$$

$$\hat{x}_1 = 29.22 + 0.2936\, y_i$$

If the value of weight (y_i) is increased in 1 until there will be an increase in height (x_i) by 0.2936

$$y_i = 82, \quad \hat{x}_i = 53.29$$

error, $e_i = x_1 - \hat{x}_1 = 52 - 53.29 = -1.29$

Y depend on X :-

$\hat{\beta}_1 yx = 3.20$

$\hat{\beta}_0 = -86.7$

$\hat{y}_i = -86.7 + 3.20 X_i$

Co-efficient of determinationation (explanatory Power)

$R^2 = r^2 = \hat{\beta}_1 xy \times \hat{\beta}_1 yx$

$= 0.293 \times 3.20$

$= 0.9395$

93.95% variation in the dependent variable can be explained by the indepent variable.

Standard Error :-

Y on X :-

$$SE(y) = \sqrt{\frac{\sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i}{n-2}}$$

$$= \sqrt{\frac{104734 - (-86.7) \times 900 - 3.20 \times 57048}{8-2}}$$

$=$

X on Y :—

$$SF(x) = \sqrt{\frac{\sum x_i^2 - \hat{\beta}_0 \sum x_i - \hat{\beta}_1 \sum x_i y_i}{n-2}}$$

$$= \sqrt{\frac{31320 - 29.22 \times 498 - 0.2936 \times 57048}{8-2}}$$

$=$

* Probability
* degression

$\Big\} \rightarrow$