MATH 2101 30.04.17

Frequency Distribution

Frequency : The number of items/distributions

Frequency distribution :

Class limit :-

Class interval :-

Class midpoint :-

Exclusive Method :-

Inclusive Method :-

Marks	No. of students	Marks	
		0 − 9	−0·5 − 9·5
0 − 10	8	10 − 19	9·5 − 19·5
10 − 20	10	20 − 29	19·5 − 29·5
20 − 30	20	30 − 39	29·5 − 39·5
30 − 40	7	40 − 49	39·5 − 49·5
40 − 50	5		
(Exclusive)		(Inclusive Method)	(Corrected class interval)

Class interval : upper limit − lower limit

Correction factor : $\frac{10-0}{2} = 0{\cdot}5 = \frac{\text{upper limit} - \text{lower limit}}{2}$

or, class mid point

Ex: The following data specifies the life of 40 similar car batteries recorded to the nearest tenth of a year:

2.2, 4.1, 3.5, 4.5, 3.2, 3.7, 3.0, 2.6, 3.4, 1.6, 3.1, 3.3, 3.8, 3.1, 4.7, 3.7, 2.5, 4.3, 3.4, 3.6, 2.9, 3.3, 3.9, 3.1, 3.3, 3.1, 3.7, 4.4, 3.2, 4.1, 1.9, 3.4, 4.7, 3.8, 3.2, 2.6, 3.9, 3.0, 4.2, 3.5

Construct a complete frequency distribution / Relative frequency distribution :-

Five step Procedure :-

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

$$2^k \geq n \quad \text{——} \ (1)$$

$n = 40$

$K = 1$
2
3

$K = 6, \ 2^6 = 64 > 40$

$\therefore K = 6$ is the number of group.

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

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

$$= \frac{4.7 - 1.6}{6}$$

$$= 0.51$$

$$= 0.5$$

H = highest value = 4.7

L = Lowest value = 1.6

K = no. of group = 6

Step 3:-

Set the individual class limit.

1·5 – 2·0, 2·0 – 2·5, 2·5 – 3·0, 3·0 – 3·5, 3·5 – 4·0, 4·0 – 4·5, 4·5 – 5·0.

Step 4:- Use tally marks (/) for each number.

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

Table: Frequency distribution of life time of 40 car batteries

Lifetime (Years) class interval	Tally Marks (/)	No. of cars	Cumulative frequency	Relative frequency	%
1·5 – 2·0	//	2	2	2/40 = ·05	5%
2·0 – 2·5	/	1	3	1/40 = ·025	
2·5 – 3·0	////	4	7	4/40 = ·100	
3·0 – 3·5	~~////~~ ~~////~~ ~~////~~	15	22		
3·5 – 4·0	~~////~~ ~~////~~	10	32		
4·0 – 4·5	~~////~~	5	37		
4·5 – 5·0	///	3	40		

$N = \Sigma f_i = 40$

Graphs :-

(i) Histogram
(ii) Frequency Polygon
(iii) Bar diagram
(iv) Cumulative Frequency Polygon (Ogive)
(v) Pie-diagram

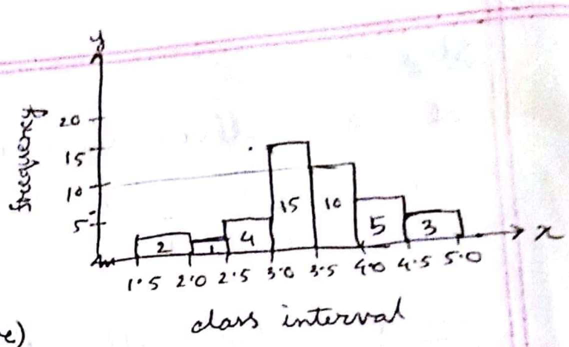


Fig: Histogram of lifetime of car batteries.

07.05.17

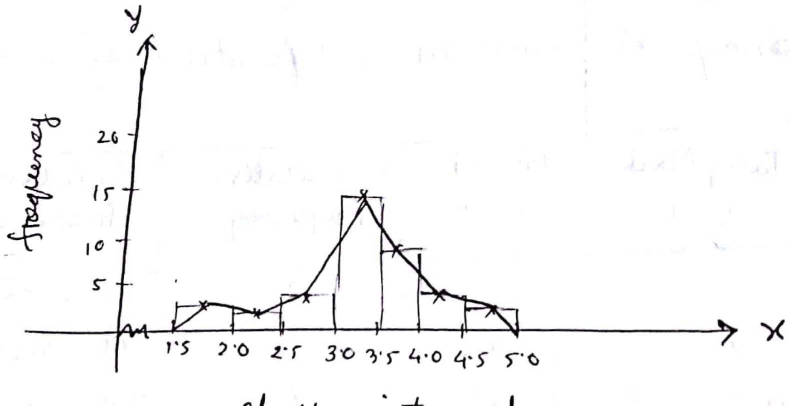
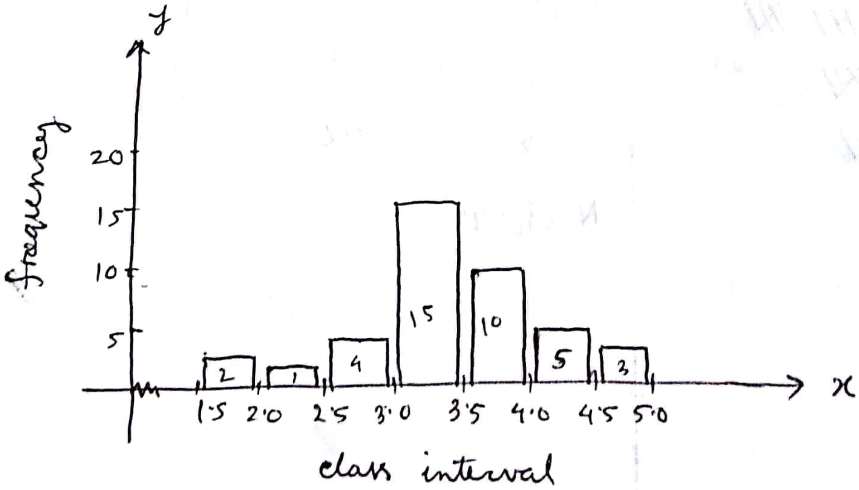


Fig: Frequency polygon



just inside the class limit

Fig: Bar diagram of lifetime of car batteries.

Cumulative frequency Polygon (ogive):-

Liftime	No. of Cars	Cumulative frequency
1.5–2.0	2	2
2.0–2.5	1	3
2.5–3.0	4	7
3.0–3.5	15	22
3.5–4.0	10	32
4.0–4.5	5	37
4.5–5.0	3	40

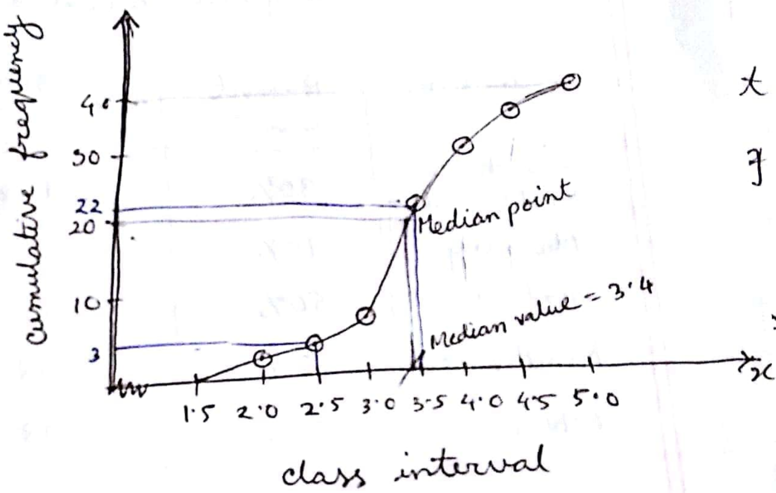


Fig: Cumulative frequency Polygon of lifetime of 40 car batteries.

How many car batteries lasts between 2.5 year and 3.5 year? 22 − 3 = 19 car batteries.

Pie diagram :-

Draw a Pie diagram based on the following information:-

Sector	Percent invested	Angle = $\frac{\text{group value}}{\text{Total}} \times 360$
Research and development	30%	108
planning	10%	36
production	50%	180
Maintenance	5%	18
Others	5%	18

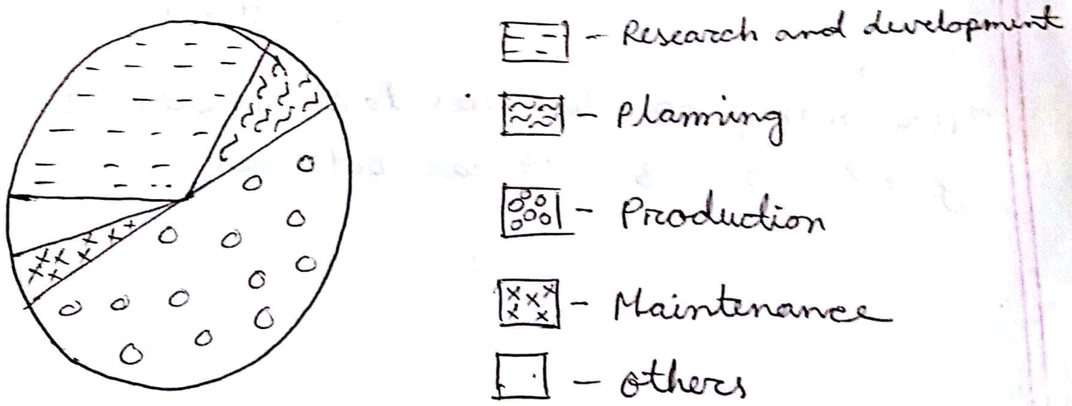


Fig : Pie diagram of investment in different sector.

Measures of central ~~tending~~ tendency :-

Five Measures :-

(i) Arithmetic Mean, Mean (A.M) or $\overline{X}$

(ii) Geometric Mean (G.M)

(iii) Harmonic Mean (H.M.)

(iv) Median (Me)

(v) Mode (Mo)

Ungrouped data

Grouped data

Arithmetic Mean :-

Ungrouped data :-

$$A.M = \overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Example :- The following data represents the lifetime of batteries:

5·0, 3·5, 5·1, 4·5, 4·8, 4·7, 4·3, 3·6, 4·0, 3·8.

Calculate the average lifetime of the batteries.

$$A.M. = \overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$= \frac{5{\cdot}0 + 3{\cdot}5 + \cdots + 3{\cdot}8}{10}$$

$$= 4{\cdot}33 \text{ year}$$

So, the batteries last on average 4·33 year.

~~5.0, 3.5, 5.1, 4.5,~~

Ex: (ii) Also show that the sum of ~~deriva~~ deviation of each value from their mean is zero.

Properties of A.M :-

(i) Every set of interval or ratio level data has a mean.

(ii) All the values are included in computing the mean.

(iii) The mean is unique.

(iv) The sum of deviation of each value from their mean is zero. Symbolically $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$.

14.05.17

x	$\bar{x}$	$(x_i - \bar{x})$
5.0		0.67
3.5		−0.83
5.1		0.77
4.5		0.17
4.8	4.33	0.47
4.7		0.37
4.3		−0.03
3.6		−0.73
4.0		−0.33
3.8		−0.53

$\Sigma (x_i - \bar{x}) = 0$

So, the sum of deviation of each value from their mean is zero

Grouped data

Direct method :-

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

where, f_i = frequency
x_i = class mid-value
$N = \sum f_i$

Shortcut method :-

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

where, $d_i = \frac{x_i - A}{c}$
c = class interval

Ex: From the following frequency distribution calculate the average value.

for short cut method

Lifetime years	No of cars, f_i	Class mid-value x_i	$f_i x_i$	$d_i = \frac{x_i - A}{c}$	$f_i d_i$
1.5 - 2.0	2	1.75	3.5	-3	-6
2.0 - 2.5	1	2.25	2.25	-2	-2
2.5 - 3.0	4	2.75	11	-1	-4
3.0 - 3.5	15	3.25	48.75	0	0
3.5 - 4.0	10	3.75	37.5	1	10
4.0 - 4.5	5	4.25	21.25	2	10
4.5 - 5.0	3	4.75	14.25	3	9
	$N = 40$		$\sum f_i x_i = 138.5$		$\sum f_i d_i = 17$

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

(i) Odd no. of observation's negative

(ii) Product Negative

(iii) 0 in the data

H.M :-

(i) 0 in the data

Harmonic Mean :- (সময়ের সাথে সম্পর্কযুক্ত)

rate, speed, per unit of time

Ungrouped data :-

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

Grouped data :-

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

Ex :- In a certain factory a unit of work completed by A in 4 minutes, by B in 5 minutes, by C in 6 min, by D in 10 min, and by F in 12 min. What is the average number of unit of work completed per minute? (* H.M হবে কারণ info time এর সাথে সম্পর্কযুক্ত)

Calculation table :

	x_i	$1/x_i$
A	4	0·15
B	5	0·20
C	6	0·167
D	10	0·100
E	12	0·083

$$\sum_{i=1}^{n} \frac{1}{x_i} = 0·8$$

Ungrouped data :

We know, $H.M = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

$$= \frac{5}{0·8}$$

$$= 6·25 \text{ minute}$$

of

t

1,

one

4, 6}

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,

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

t

H.W Ex: Check the relationship 10, 20, 30, 40, 50, 60.

Reference :-

(i) Probability and statistics for Engineers and scientists — Ronald E. Walpole, Sharon L. Myers, Keying Ye

(ii) Business statistics — S. P. Gupta, M. P. Gupta

(iii) Statistical Techniques in Business and Economics — Douglas. A. Lind

21.05.17

Median: The middle most value in a set of data arranged in ascending order.

Ungrouped data:

Odd No. of observations:

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

Even No. of observations:

Median = Average of $\left(\frac{n}{2}\right)$ and $\left(\frac{n}{2}+1\right)$th observations in a series arranged in ascending order.

Ex: Following data represent the lifetime of some electronic devices (in years) :-

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

Calculate the median and mode of the above data.

Solution :- At first arrange the data in ascending order.

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

Since, $n = 12$ is even,

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

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

$= \quad$ " $\quad$ " 6th and 7th "

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

$= 8.5$ year

So, the middle most ~~have~~ value have lifetime of 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 = ~~Linear~~ Lower limit of median group

N = Total no of observations

p.c.f = Preceding group cumulative frequency.

f_m = frequency of median group.

c = class interval.

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

Lifetime years	No. of cars	Cumulative frequency
1·5 – 2·0	2	2
2·0 – 2·5	1	3
2·5 – 3·0	4	7
3·0 – 3·5	15	22 Median
3·5 – 4·0	10	32
4·0 – 4·5	5	37
4·5 – 5·0	3	40
	N = 40	

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

$$= 3\cdot0 + \frac{\frac{40}{2} - 7}{15} \times 0\cdot5$$

$$= 3\cdot43 \text{ year}$$

So, the middle most car have lifetime of 3·43 year.

Mode :-

Ungrouped data :- The value in the series x_i which has maximum frequency.

Ex: Lifetime of electronic chips :

x_i	tally	frequency
5	/	1
7	///	3
8	//	2
9	//	2
10	//	2
12	/	1
13	/	1

Since, observation 7 has maximum frequency. So, modal lifetime is 7 years.

Relation between mean, median and mode :-

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

* if two or more observations have same maximum frequency, _

Grouped data :-

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

where, L = Lower limit of modal group.

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

Δ_2 = difference between the frequency of modal group and post modal group.

C = class interval

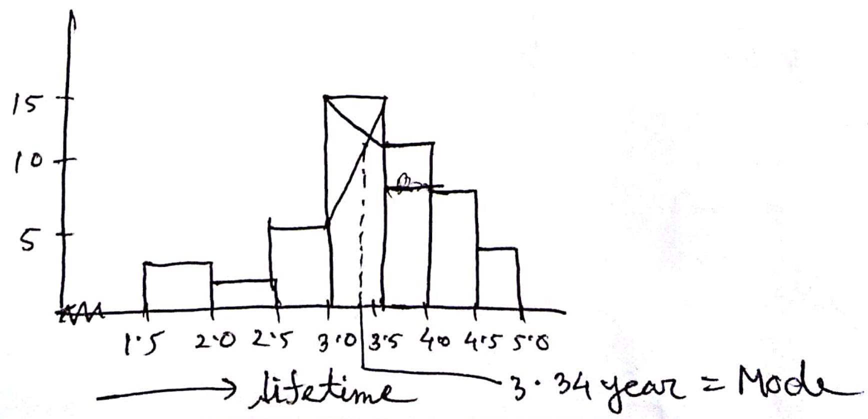
Here, $\Delta_2 = 5$. $\boxed{3.0 - 3.5 \mid 15}$ → Modal group.

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

$$= 3.0 + \frac{11}{11+5} \times 0.5$$

$$= 3.34 \text{ year}$$

∴ Most of the cars have lifetime of 3.34 year.



Weighted Average:

	No. of worker (w_i)	daily Salary (X_i)
Skilled	150	500
Semi-skilled	100	300
Unskilled	50	150

Average salary of the worker

$$\overline{X_w} = \frac{\Sigma w_i X_i}{\Sigma w_i} = \frac{150 \times 500 + 100 \times 300 + 50 \times 150}{150 + 100 + 50} = 375$$

23.05.17

Measures of Dispersion

Scatter, deviation, spread, variability.

Measures of Dispersion

A. Absolute Measure

(i) Range
(ii) Standard deviation
(iii) Mean deviation
(iv) Quartile deviation

B. Relative Measure

(i) Co-efficient of range
(ii) Co-efficient of standard deviation
(iii) Co-efficient of mean deviation.
(iv) Co-efficient of Quartile deviation.

Range :-

Ungrouped data:

Range = Highest value - Lowest value

$= X_H - X_L$

$$\text{Co-efficient of range} = \frac{X_H - X_L}{X_H + X_L} \times 100$$

Grouped data:

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

$= X_U - X_L$

$$\text{Co-efficient of range} = \frac{X_U - X_L}{X_U + X_L} \times 100$$

Ex: Find range and co-efficient of range.

(1) The following measurement were recorded for drying time of certain brand of latex point.

7·4, 4·0, 2·5, 4·8, 2·9, 3·6, 2·8, 3·3, 4·8, 5·2

(11)

Battery Life	Frequency
1·5 - 2·0	2
2·0 - 2·5	1
2·5 - 3·0	4
3·0 - 3·5	15
3·5 - 4·0	10
4·0 - 4·5	5
4·5 - 5·0	3

if we measure the variation from range the variation;

(1) Ungrouped data :- Range $= X_H - X_L = 5{\cdot}2 - 2{\cdot}5$

$=$

~~Grouped~~ co. efficient $= \dfrac{X_H - X_L}{X_H + X_L} \times 100 = \dfrac{5{\cdot}2 - 2{\cdot}5}{5{\cdot}2 + 2{\cdot}5} \times 100$

$=$

(2) Grouped data :- Range $= X_U - X_L = 5{\cdot}0 - 1{\cdot}5 = 3{\cdot}5$

Co-efficient $= \dfrac{X_U - X_L}{X_U + X_L} \times 100 = \dfrac{5{\cdot}0 - 1{\cdot}5}{5{\cdot}0 + 1{\cdot}5} \times 100$

$=$

Standard deviation :-

The arithematic mean of the squared deviations of observations for their arithematic mean is known as variance and the positive square root of variance is known as standard deviation.

	Mean	Standard deviation	denominator
Population	μ	σ	N
Sample	$\bar{X}$	S	$n-1$ ungrouped $N-1$ grouped

Variability	Stability / consistency
Lower	higher
Higher	Lower

Ungrouped Sample Data :-

Sample mean, $\bar{X} = \frac{\sum X_i}{n}$

Sample variance, $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$

Sample standard deviation, $S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$

$= \sqrt{\frac{\sum X_i^2 - n\bar{X}^2}{n-1}}$

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

$$= \sum_{i=1}^{n} (x_i^2 - 2x_i \cdot \bar{x} + \bar{x}^2)$$

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

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

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

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

<u>Grouped Sample data:-</u>

Sample mean, $\bar{x} = \dfrac{\sum_{i=1}^{n} f_i x_i}{N}$

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

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

<u>Ungrouped Population data:-</u>

Population mean, $\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$

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

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

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

Grouped Population data :

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

Population variance, $\sigma^2 = \frac{\sum f_i (x_i - \mu)^2}{N}$

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

Co-efficient of variation :

The ratio of the standard deviation to its arithmetic mean expressed as a ~~perda~~ percentages.

$C.V = \frac{s}{\bar{x}} \times 100$ (Sample)

$C.V = \frac{\sigma}{\mu} \times 100$

When to use C.V :-

(i) When the are in different unit. Such as income in \$, no. of days employees are absent.

(ii) When data are in same unit but the mean are far apart. Such as income of top execute, income of unskilled worker.

28.05.17

Ex: The government awarded grants to the agricultural department of a university to test the yield capabilities of two new varieties of wheat. Each variety was planted on plots equal area at each university and the yields, in kg per plot recorded as follows :-

Compare the variability in the data set using an appropriate measure.

calculator → mode → 3 → 1

University	Variety 1	Variety 2
1	38	45
2	23	25
3	35	31
4	41	38
5	44	50
6	29	33
7	37	36
8	31	40
9	38	43

Variety 1

x_i	x_i^2
38	1444
23	529
35	1225
41	1681
44	1936
29	841
37	1369
31	961
38	1444

$\Sigma x_i = 316$ $\Sigma x_i^2 = 11430$

Variety 2

x_i	x_i^2
45	2025
25	625
31	961
38	1444
50	2500
33	1086
36	1296
40	1600
43	1849

Sample mean, $\overline{X} = \frac{\Sigma X_i}{n} = \frac{316}{9} = 35 \cdot 11$

sample standard deviation, $s = \sqrt{\frac{\Sigma x_i^2 - n\overline{X}^2}{n-1}}$

$= \sqrt{\frac{11430 - 9 \times 35 \cdot 11^2}{9-1}}$

$= 6 \cdot 48$

co-efficient of variation :

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

$= \frac{6 \cdot 47}{35 \cdot 11} \times 100$

$= 18 \cdot 42 \%$

Co-efficient of ~~de~~ variation :

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

$= \frac{7 \cdot 66}{37 \cdot 89} \times 100$

$= 20 \cdot 16\%$

Scanned by CamScanner

~~Variety 2~~

Since C.V. of variety 1 is 18.42 % and C.V of variety 2 is 20.16%. So, there is more variation in the deivation distribution of variety 2 relative to the distribution of variety 1.

Variety 2 is more variable.

Variety 1 is more stable.

Ex:- The following two frequency distribution reports the electricity cost for a sample of 50 two-bed room and three bed-room apartment.

Electricity Cost $	Two bed room frequency	Three bedroom frequency
120-140	3	2
140-160	7	5
160-180	20	10
180-200	8	28
200-220	10	3
220-240	2	2

$N = 50$

11.06.17

Compare the variability in the two distribution using an appropriate measure.

Two bedroom frequency

Electricity cost	Two bedroom frequency (f_i)	x_i	$f_i x_i$	$f_i x_i^2$
120-140	3	130	390	850700
140-160	7	150	1050	157500
160-180	20	170	3400	578000
180-200	8	190	1520	288800
280-220	10	210	2100	441000
220-240	2	230	460	105800

$N = \Sigma f_i = 50$ $\Sigma f_i x_i = 8920$ $\Sigma f_i x_i^2 = 1621800$

$\therefore$ Sample mean, $\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{8920}{50} = 178.4$

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

$= \frac{1621800 - 50 \times (178.4)^2}{50-1}$

$= 24.94$

Co-efficient of variation, $c.V = \frac{s}{\bar{x}} \times 100 = \frac{24.94}{178.4} \times 100\%$

$= 13.97\%$

Scanned by CamScanner

Calculation for three bedroom

Electricity cost	f_i	x_i	$f_i x_i$	$f_i x_i^2$
120 – 140	2	130	260	33800
140 – 160	5	150	750	112500
160 – 180	10	170	1700	289000
180 – 200	28	190	5320	1010800
200 – 220	3	210	630	132300
220 – 240	2	230	460	105800
	$N = \sum f_i = 50$		$\sum f_i x_i = 9120$	$\sum f_i x_i^2 = 16842\cancel{7}00$

Sample mean, $\bar{X} = \dfrac{\sum f_i x_i}{N} = \dfrac{9120}{50} = 182.4$

Sample standard deviation, $s = \sqrt{\dfrac{\sum f_i x_i^2 - N\bar{X}^2}{N-1}}$

$= \sqrt{\dfrac{1684200 - 50 \times (182.4)^2}{50-1}}$

$= 20.56$

co-efficient of variation, $c.V = \dfrac{s}{\bar{X}} \times 100$

$= \dfrac{20.56}{182.4} \times 100$

$= 11.27\%$

Since, C.V of electricity cost of two bedroom apartment is 13.97% and C.V of electricity cost of three bedroom is 15.27% . So, there is more variation in the distribution of two bedroom relative to the distribution of three bedroom apartment. So, distribution of two bedroom apartment is more variable. distribution of three bedroom apartment is more stable.

Ex:

Distribution of test score	Distribution of year of service
Mean test score = 200	Mean year of service 20 years
standard deviation = 40	standard deviation 2 years

Compare the variability in the two distribution using an appropriate measure.

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

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

$$= 20\%$$

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

$$= \frac{2}{20\cancel{0}} \times 100$$

$$= 100\%$$

Probability

The value between zero and one describing the relative possibility that an event will occur.

Approaches to Probability

- Objective
  - Classical probability
    - based ~~of~~ on equally likely outcomes
  - Emperical probability
    - based on past experience
- Subjective
  - based on available information

Classical Probability : Every outcome of the experiment has an equal chance of being selected. (eg : coin, dice)

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

Coin tossing:

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

$P(\text{tail}) = \frac{1}{2}$

Dice Throwing:

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

$\vdots$

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

Experiment :- A process that leads to the occurrence of one and only one of several possible observation.

Random experiment : The experiment which has at ~~lea~~ least two possible outcome. Ex :- ~~cotos~~ coin tossing, dice throwing.

Event :- A collection of one or more possible outcome of an experiment. Ex :- Throwing a dice, even {1, 4, 6} roge and odd { 1, 3, 5 } are event.

Mutually exclusive event : The events which never occur together. Ex :- In a single coin toss head and tail never comes together. $P(H \cap T) = 0$.

Collectively exclusive :- At least one of the outcome must occur in an experiment.

Independent and dependent event :- If two event A and B occur indepedently. They are called independent of each other.

Two events such that A happen if B happen or B happen if A happen. Then A is called dependent on

B and B is called dependent on A.

P(impossible even) = 0

P(Sun rises in the west) = 0

P(Universal truth) = 1

P(The sun rises in the east) = 0

$P(H) = \frac{1}{2}$ $P(T) = \frac{1}{2}$

$P(1) = \ldots$ $P(6) = 1/6$

Law of independence :-

If two events A and B are independent, then

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

Conditional Probability :-

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

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

Law of Addition :-

$P(\text{either A or B or both}) = P(A \cup B)$

$= P(A) + P(B) - P(A \cap B)$ [if A and B are not mutually exclusive]

$= P(A) + P(B)$ [If A and B are mutually exclusive]

Law of multiplication :-

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

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

Properties of probability :-

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

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

Complementary Law :-

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

Empirical Probability :-

$$P(A) = \frac{\text{Number times event occured in the past}}{\text{Total No. of trials}}$$

Subjective probability :- Based on available information

Ex:- A box contains 5 white balls and 8 red balls all of which are of equal size. A ball is drawn from the box at random. What is the probability that the ball is white?

Sol^n :- Let, A = The ball is white

Number of white balls = 5

Total balls = 13

We know, $P(A) = \frac{\text{Number of white balls}}{\text{Total balls}}$

$= \frac{5}{13}$

$= 0.385$ [percentage করলে 38.5%]

So, the ~~number~~ probability that the ball is white is 0.385.

2. If a card is drawn from an ordinary deck of 52 playing cards – find the probability that –

(i) The card is a red $26/52 = 1/2$

(ii) The card is a diamond $13/52 = 1/4$

(iii) The card is an ace $4/52 = 1/13$

(iv) The card is either an ace or a diamond.

Red | Hearts 2-10, K, Q, J, A
Red | Diamond
Black | Spade
Black | clubs

(iv) Let, ~~the~~

A = The card is an ace

B = The card is a diamond

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

There is only one card which is an ace and also a diamond.

$$P(A \cap B) = \frac{1}{52}$$

$\therefore$ P (either an ace or a diamond)

$$= P(A \cup B)$$

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

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

$$= 16/52$$

$$= 4/13$$

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

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

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

Sol^n:- Tossing a coin 3 times, the possible outcomes are –

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

A = at least one head appears

There are 7 outcomes with at least one head.

$$P(\text{at least one head}) = \frac{\text{No. of favourable outcomes}}{\text{Total possible outcomes}}$$

$$= 7/8$$

Ex: If two dice are thrown, write the sample space of all possible outcomes and find the probability of –

i) a double six

ii) a sum of 9 or more.

Throwing two dice the possible outcomes are –

$$S = \{ (1,1), (1,2), \ldots, (1,6) \\ \vdots \\ (6,1), \ldots, (6,6) \}$$

Let, A = a double six

$P(A) = 1/36$

(ii) Let, B = a sum of 9 or more

~~P(B)~~ = $\{(3,6), (6,3), (5,4), (4,5), \ldots\ldots\}$

$\therefore P(B) = 10/36$

16.07.17

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

Sol^n: Tossing two coins the possible outcomes are —

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

Let, A = Two head results = $\{HH\}$

B = There is at least one head = $\{HH, HT, TH\}$

$A \cap B = \{HH\}$

$P(A) = 1/4$

$P(B) = 3/4$

$P(A \cap B) = 1/4$

$\therefore$ P (two head result / there is at least one head)

$= P(A/B)$

$$= \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = 1/3 \quad \text{(Ans.)}$$

9. One bag contains 4 white balls and 2 black balls; another contains 3 white balls and 5 black balls. Let one ball is drawn from each bag, find sol^n of the probability that -

(i) both are white

(ii) both are black

(iii) one is white and one is black.

→

(i) Let,

W_1 = white ball from 1st bag

W_2 = " " " 2nd bag

B_1 = black " " 1st "

B_2 = " " " 2nd "

(i) P(both are white)

$= P(W_1 \cap W_2)$

$= P(W_1) \cdot P(W_2)$ (Law of independence)

$= P \frac{4}{6} \cdot \frac{3}{8}$

$= 1/4$

(ii) P(both are black)

$= P(B_1 \cap B_2)$

$= P(B_1) \cdot P(B_2)$

$= \frac{2}{6} \times \frac{5}{8}$

$= \frac{5}{24}$

(iii) P(one is white and one is black)   07.17

$= P$ [either (one white ball from 1st bag and one black ball 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_1B_2 and B_1W_2 are mutually exclusive]

$= P(W_1) \cdot P(B_2) + P(B_1) \cdot P(W_2)$

$= \frac{4}{6} \times \frac{5}{8} + \frac{2}{6} \times \frac{3}{8}$

$=$

From Book = 10 — 17
S.P Gupta — M.P Gupta
Probability chapter

13. In a certain town, male and female each form 50 percent of the population. It is known that 20 percent of the males and 5 percent of females are unemployed. A research student studying the employment situations and select an ~~unemp~~ unemployed person at random. What is the the probability that the person ~~so~~ selected is (i) male (ii) female.

	Employed	Unemployed	
Male	.400	.100	.50
Female	.475	.025	.50
	.875	.125	

(i) $P(\text{Male}/\text{Unemployed}) = \frac{P(\text{Male and unemployed})}{P(\text{unemployed})}$

$= \frac{.100}{.125} =$

(ii) $P(\text{Female}/\text{Unemployed}) = \frac{P(\text{Female and unemployed})}{P(\text{unemployed})}$

$= \frac{\cancel{.005}}{\cancel{.875}} \; \frac{.025}{.125}$

$=$

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

(i) without replacement

(ii) with replacement

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

$\therefore P(A) = \frac{\text{Number of favorable outcome}}{\text{Total of all possible outcomes}}$

$= \frac{3}{10}$

Let, B = the 2nd roll being found defective

$\therefore P(B/A) = \frac{2}{9}$ [After 1st selection there remain 2 defective rolls out of 9 rolls)

without replacement

$\therefore P(\text{two defective}) = P(A \cap B)$

$= A(A). P(B/A)$

$= \frac{3}{10} \times \frac{2}{9}$

with ~~out~~ replacement

$P(\text{two defective}) = P(A \cap B)$

$= P(A) . P(B/A)$

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

$P(\text{three defective})$

$= P(A). P(B/A). P(C/AB)$

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

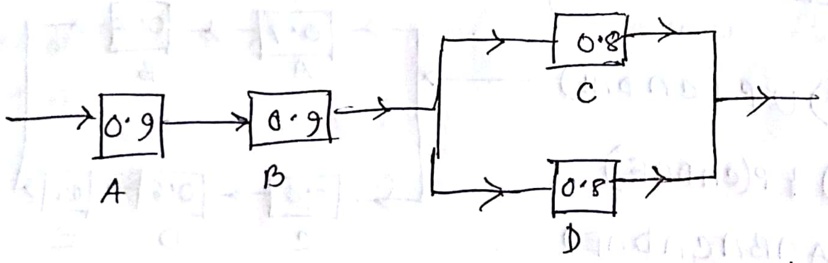
$= \frac{3}{10} \times \frac{3}{10} \times \frac{3}{10}$ (with replacement)

Scanned by CamScanner

23.07.17

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

Assume that all four components work independently.



An ~~entire~~ electrical System.

(i) The probability that the entire system works can be calculated as following :- P(entire system work) $= P[A \cap B \cap (C \cup D)]$

$$= P(A).P(B).P(C \cup D)$$

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

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

$$= P(A.).P(B)\left[1 - P(C').P(D')\right]$$

$$= 0{\cdot}9 \times 0{\cdot}9\left[1 - (1 - 0{\cdot}8)(1 - 0{\cdot}8)\right]$$

(ii) P(System C does not work / the entire system work)

$$= \frac{P(\text{System C does not work and the entire system work})}{P(\text{entire system work})}$$

$$= \frac{P(A \cap B \cap C' \cap D)}{0.7776} = \frac{\cancel{\phi}\, P(A).P(B) \times P(C') \times P(D)}{0.7776}$$

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

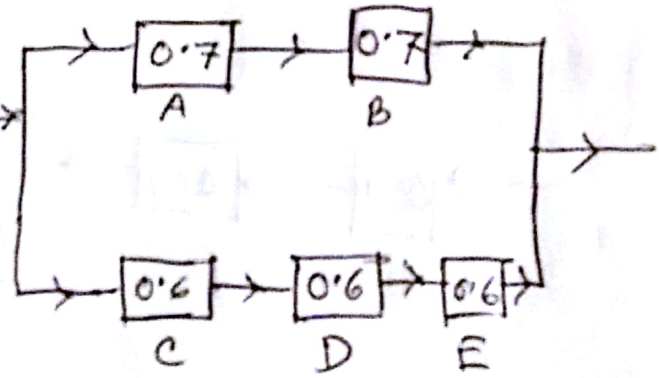
$$= 0.1667$$

Ex: From the following figure find the probability that the entire system work.

$P(A \cup B) \cup P(C \cap D \cap E)$

$P(A \cap B) + P(C \cap D \cap E)$

$- P(A \cap B \cap C \cap D \cap E)$



07.17

Correlation and Regression

Correlation :- Correlation is a statistical technique, which gives us the degree of association or interrelationship that exists between two or more variable.

ex :- (Income, expenditure), (Price, demand), (demand, supply), (Area, resistance), (current, voltage) etc.

Positive and Negative Correlation :-

When ~~two~~ two variables vary together in the same direction then the correlation between two variable is positive.

i.e. X↑ Y↑ or X↓ Y↓

Current ↑ Voltage ↑ current ↓ voltage ↓

Demand ↑ Supply ↑ Demand ↓ Supply ↓

when the changes are in opposite direction then the correlation is negative.

i.e. X↑ Y↓ or X↓ Y↑

Price ↑ Demand ↓ Price ↓ Demand ↑

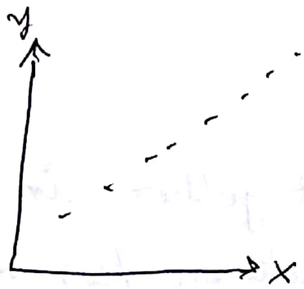
Area ↑ Resistance ↓ Area ↓ Resistance ↑

Karl Pearson's co-efficient of correlation :-

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}}$$

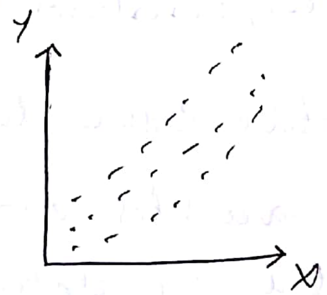
x = value of 1st variable
y = value of 2nd "
$\bar{x}$ = mean of x-values
$\bar{y}$ = mean of y-values

Scatter diagram :

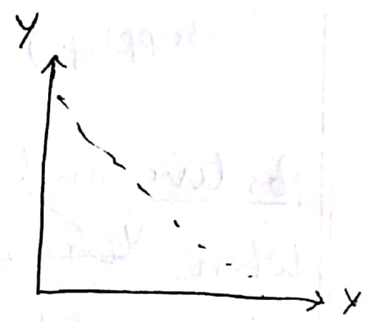


Perfect Positive correlation

$r = +1$

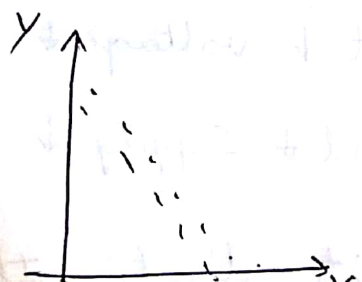


Positive Correlation

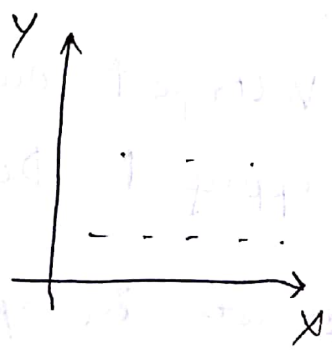


Perfect Negative Correlation

$r = -1$



Negative Correlation



No correlation

$r = 0$

$r = +1$ Perfect and positive
$r = -1$ " " Negative
$r = 0$ & No

$r = 0.7 - 1$ Strong and positive correlation

$0.5 - 0.7$ Moderate " " "

$0 - 0.5$ Weak and " "

$r = (-)$ " " " negative

Ex: Draw a scatter diagram. Find the relationship between demand and supply.

x_i Demand	y_i Supply	x_i^2	y_i^2	$x_i y_i$
52	82			
55	90			
59	95			
62	105			
63	120			
66	132			
70	140			
71	135			
$\Sigma x_i = 498$	$\Sigma y_i = 900$	$\Sigma x_i^2 = 31320$	$\Sigma y_i^2 = 104734$	$\Sigma x_i y_i = 57048$

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

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

$\therefore$ Correlation

$$= \frac{57048 - \frac{498 \times 900}{8}}{\sqrt{(31320 - 8 \times 62.25^2)(104734 - 8 \times 112.5^2)}}$$

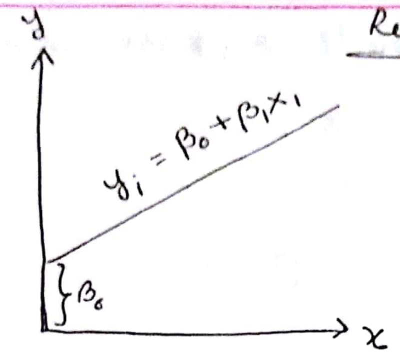
$$= \frac{1023}{\sqrt{319.5 \times 3484}}$$

$$= 0.9696$$

Since, $r = 0.9696$ suggests a strong and positive correlation between demand and supply.

30.07.17

Regression Analysis :-



By regression analysis we can determine the average change in one variable in terms of other.

Simple linear regression Model :-

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

where, y_i = dependent 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 :-

$$\cancel{\hat{\beta}_1 y x} \quad \hat{\beta}_{1yx} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - n\bar{x}^2}$$

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

Scanned by CamScanner

Then the estimated or fitted regression line or equation is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{e}_i = y_i - \hat{y}_i$$

Same way x depend on y **

Properties of regression co-efficients :-

① r is the g.m. of two regression co-efficients

$$r = \sqrt{\hat{\beta}_{1xy} \hat{\beta}_{1yx}}$$

(ii) $\hat{\beta}_{1xy} > 1$, $\hat{\beta}_{1yx} < 1$

(iii) $\hat{\beta}_{1xy} +$, $\hat{\beta}_{1yx} +$, $\hat{\beta}_{1xy} -$, $\beta_{1yx} -$

(iv) ??

Co-efficient of determination :-

explanatory power of regression equation

$$R^2 = r^2 = \hat{\beta}_{1xy} \cdot \hat{\beta}_{1yx}$$

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

Ex: Fit a regression line when

① Demand depend on supply.

② Supply depend on demand.

③ Also determine the explanatory power of the regression equation.

— From demand and supply data (prev. lecture)

$\bar{X} = 62.25$

$\bar{y} = 112.5$

Demand (x_i) depend on Supply (y_i) :-

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

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

$$= \frac{1023}{3484} = 0.2936$$

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

$$= 62.25 - 0.2936 \times 112.5 = 29.22$$

Then the estimated on filtered regression line on equation is,

$$\hat{x}_i = \hat{\beta}_0 + \hat{\beta}_1 y_i$$

$$\hat{x}_i = 29.22 + 0.2936\, y_i$$ $\therefore y \uparrow, x \uparrow$

If the supply is increased by 1 unit there will be an increase in demand (x_i) by 0·2936 unit.

* Estimate the demand for supply of 82.

when $y_i = 82$

$$\hat{x_i} = 29{\cdot}22 + {\cdot}2936 \times 82$$

$$= 53{\cdot}29$$

$\therefore$ $\hat{e_i} = x_i - \hat{x_i} = 52 - 53{\cdot}29 = -1{\cdot}29$

Supply depend on demand :-

$$\hat{\beta}_{1yx} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - n\bar{y}^2}$$

$$57048 - \frac{498 \times 900}{8}$$

Then the estimate line is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

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

If the demand is increased by 1 unit, there will be an increase in supply by 3.20 unit.

* estimate supply for the demand of 52.

$X_i = 52$, $\hat{y}_i = -86.7 + 3.20 \times 52 =$

<u>Co-efficient of determination</u>:

~~Ind~~ $R^2 = \beta_{1xy} \cdot \beta_{1yx}$

$= 0.2936 \times 3.20$

$= 0.9345$

$\therefore$ 93.95% variation in the dependent variable can be explained by the independent variable. So, the regression line best fit the data.

$r = \sqrt{.9395}$

$= 0.9696$

Ex:- The following is a set of coded experimental data on two variable :-

Normal Stress	Shear Ressistance
26.8	26.5
25.4	27.3
28.9	24.2
23.6	27.1
27.7	23.6
23.9	25.9
24.7	26.3
28.1	22.5

(i) Find the relationship between normal stress and shear resistance.

(ii) Estimate normal stress for shear ressistance 26.5.

(iii) Estimate shear ressistance for stress of 26.8.

(iv) Determine the explanatory power.