Measures of dispersion: சிதறல் அளவைகள்

- சேகரிக்கப்பட்ட விவரங்களின் மதிப்புகள் அதன் மையப்போக்கு அளவையிலிருந்து எந்த அளவிற்கு சிதறியுள்ளது என்பதை தெரிவிப்பது சிதறல் அளவைகள் ஆகும்.
- நான்கு சிதறல் அளவைகள் பயன்படுத்தப்படுகிறது

(i) வீச்சு (range)
(ii) கால்மான விலக்கம் (Quartile deviation)
(iii) சராசரி விலக்கம் (Mean deviation)
(iv) திட்ட விலக்கம் (Standard deviation)

1893-ல் கார்ல் பியர்சன் இதனை முதன்முதலில் அறிமுகப்படுத்தினார்

[illegible])

- The degree to which the individual values of the variate scatter away from the average or the central value, is called a dispersion.
- There are 5 different measures of variance
(i) Range (ii) Quartile deviation
(iii) Mean deviation (iv) Standard deviation
(5) Variance.

Standard deviation: திட்ட விலக்கம் (σ)

SD may be defined as "the square root of the arithmetic mean of the square of deviation from the arithmetic mean".

கொடுக்கப்படுகின்ற மதிப்புகளின் கூட்டுச் சராசரியிலிருந்து பெறப்படுகின்ற விலக்கங்களின் வர்க்கங்களின் கூட்டுச் சராசரியின் வர்க்கமூலம். அவ்வாறான திட்டவிலக்கம் ஆகும். இது வர்க்க மூலச், சராசரி விலக்கம் எனவும் அழைக்கப்படுகிறது.

1 Calculate of S.D. individual observation:

(i) Direct method:

Steps:

1. Find out the actual mean of the Series ($\bar{x}$)
2. Find out the deviation of each value from the mean ($x - \bar{x}$).
3. Square the deviation of each value and take the total of squared deviations $\sum (x - \bar{x})^2$
4. Divide the total by the number of observations $\frac{\sum (x-\bar{x})^2}{N}$
5. Find out the square root of the product.

Formula: $S.D. = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}$ (or) $\sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

eg: The following data are height of 10 Students. Calculate Standard deviation.

60, 60, 61, 62, 63, 63, 64, 64, 70

Height in inches (x)	$\bar{x} = 63$ $(x-\bar{x})$	$(x-\bar{x})^2$
60	60 - 63 = -3	9
60	60 - 63 = -3	9
61	61 - 63 = -2	4
62	62 - 63 = -1	1
63	63 - 63 = 0	0
63	63 - 63 = 0	0
63	63 - 63 = 0	0
64	64 - 63 = 1	1
64	64 - 63 = 1	1
70	70 - 64 = 7	49
$\sum X = 630$		74

* If the number of observation is less than 30 (< 30), instead of N, use n-1

$N = > 30$ observation, $n = < 30$ observation

$$\bar{X} = \frac{\Sigma x}{N} \quad \frac{630}{10} = 63$$

$$S.D = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}} \text{ (because the observation}$$

$$= \sqrt{\frac{74}{9}} \text{ (is less than 30}$$

$$= \sqrt{8.22} = 2.86$$

$$S.D = \underline{2.86}$$

2. **Short-cut method:**

eg: The following data are the height of 10 students. Calculate S.D.

60, 60, 61, 62, 63, 63, 63, 64, 64, 70

Solution:

Step: 1. Assume any one of the values in the data as assumed mean (A)

2. Find out deviation of each value from the assumed mean ($x - A = d$)

3. Square the deviation of each value and take the total squared deviations (d^2)

4. Apply the formula.

$$S.D = \sqrt{\frac{\Sigma d^2}{N} - \left[\frac{\Sigma d}{N}\right]^2}$$

$$= \sqrt{\frac{84}{10} - \left[\frac{10}{10}\right]^2}$$

$$= \sqrt{\frac{84}{10} - (1)^2} = \sqrt{7.4}$$

$$S.D = \underline{2.72}$$

Height in Inches (x)	$A = 62$, $x - A = d$	d^2
60	$60-62=-2$	4
60	$60-62=-2$	4
61	$61-62=-1$	1
62	$62-62=0$	0
63	$63-62=1$	1
63	$63-62=1$	1
63	$63-62=1$	1
64	$64-62=2$	4
64	$64-62=2$	4
70	$70-62=8$	64
	$-5 \; +15$ $\therefore \Sigma d = 10$	$\Sigma d^2 = 84$

திட்ட விலக்கம் கண்டுபிடிக்கும் முறை:

1. கொடுத்த வரிசை விவரங்களின் சராசரி கண்டுபிடிக்கப்படுகிறது.
2. ஒவ்வொரு மதிப்பும் சராசரி இருந்து விலகும் அளவுகள் $(x-\bar{x})$ கண்டுபிடிக்கப்படுகிறது.
3. ஒவ்வொரு வித்தியாசத்தின் $(x-x)$ வர்க்கம் கண்டறிந்து $(x-\bar{x})^2$ அவற்றின் கூட்டுத்தொகை கண்டுபிடிக்கப்படுகிறது.
4. பின்வரும் சூத்திரம் பயன்படுத்தப்படுகிறது.

$$SD = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}}$$

எ.கா: கீழே கொடுக்கப்பட்டுள்ள மீன்களின் எடை அளவுகளின் திட்ட விலக்கம் கண்டுபிடிக்க.

மீன்களின் எடை (கிராம்) = 3, 4, 5, 7, 8, 9, 6

மீன்களின் எடை (g)	$x-\bar{x}$	$(x-\bar{x})^2$
3	-3	9
4	-2	4
5	-1	1
7	1	4
8	2	4
9	3	9
6	0	0
42		

சராசரி $\bar{X} = \dfrac{\Sigma x}{n} = \dfrac{42}{7} = 6$

திட்டவிலக்கம் (SD) $= \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\dfrac{28}{7-1}} = \sqrt{4.67} = \underline{\underline{2.16}}$

மாறுபாடு (Variance):

திட்டவிலக்கத்தின் வர்க்கம் மாறுபாடு என்று அழைக்கப்படும்

சூத்திரம் $\sigma^2 = \dfrac{\Sigma(x-\bar{x})^2}{n-1}$

திட்டப்பிழை (SE)

பல மாதிரிகளின் திட்டவிலக்கங்களை ஒப்பிடுவதற்கு திட்டப்பிழை பயன்படுகிறது. இதை கணக்கிடுவதற்கு

$\dfrac{SD}{\sqrt{n}}$ என்ற சூத்திரம் பயன்படுகிறது

$= \dfrac{2.16}{\sqrt{7}} = \dfrac{2.16}{3.5} = \underline{\underline{0.82}}$

குறுக்கு வழி:

$$SD \text{ (or) } \sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$$

எண்களின் அளவு (x)	x^2
3	9
4	16
5	25
7	49
8	64
9	81
6	36
$\Sigma x = 42$	$\Sigma x^2 = 280$

திட்டவிலக்கம்

$$SD(\sigma) = \sqrt{\frac{280 - \frac{42 \times 42}{7}}{7-1}} = \sqrt{\frac{280-252}{6}}$$

$$= \sqrt{4.677} = \underline{\underline{2.16}}$$

உதாரணம் : 2, கீழ்வரும் மதிப்புகளுக்கு திட்டவிலக்கம் காண்க.

12, 10, 9, 6, 3

தீர்வு : சராசரி $\bar{X} = \frac{12+10+9+6+3}{5} = \frac{40}{5} = 8$

$$\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n}} = \sqrt{\frac{50}{5}} = \sqrt{10} = 3.16$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
12	4	16
10	2	4
9	1	1
6	−2	4
3	−5	25
50	0	50

மாதிரி : (ஊகச்சராசரி முறை)

கீழ்க்காணும் விவரங்களுக்கு திட்டவிலக்கம் காண்க.

12, 18, 20, 15, 8, 10, 6, 9, 7

தீர்வு:

17 ஐ ஊகச்சராசரியாகக் கொள்வோம்

மொத்த உறுப்புகளின் எண்ணிக்கை $n=9$

திட்டவிலக்கம் $\sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$

X	$d = X - A$ $A = 12$	d^2
12	0 = 0	0
18	= 6	36
20	= 8	64
15	= 3	9
8	= −4	16
10	= −2	4
6	= −6	36
9	= −3	9
7	= −5	25
	−3	199

$$\sigma = \sqrt{\frac{199}{n} - \left(\frac{-3}{9}\right)^2} = \sqrt{22.11 - 9/8}$$

$$\sigma = \sqrt{22.11 - 0.11} \quad \sigma = \sqrt{22} = \underline{\underline{4.69}}$$

$\boxed{2}$ **தொடர்ச்சியற்ற தொகுதியில் திட்டவிலக்கம் காணுதல் :**

(அ) நேரடி முறை:

1. ஒவ்வொரு மதிப்பும் (x) சராசரியை $(\bar{x})$ கொண்டு விலக்கங்கள் அவை குறிக்கப்படுகிறது $(x - \bar{x})$

2. சராசரியத்தின் விலக்கங்கள் $(x-\bar{x})^2$ கணக்கிடப்பட்டு பட்டியல் இடப்படுகின்றன.

3. ஒவ்வொரு $(x-\bar{x})^2$ மதிப்பும் அதற்கு நிகரான உள்ள அலைவெண்களால் பெருக்கப்படுகிறது. அவற்றின் கூட்டுத்தொகை கணக்கிடப்படுகிறது.

புள்ளிகளின் எண்ணிக்கை (x)	நபர்கள் (f)	fx	$\bar{x} = 2.88$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
0	2	$0\times2 = 0$	-2.88	8.29	16.58
1	4	$=4$	-1.88	3.53	14.13
2	5	$=10$	-0.88	0.77	3.85
3	6	$=18$	0.12	1.254	5.017
5	1	$=5$	2.12	4.49	4.49
6	2	$=12$	3.12	9.734	19.46
7	1	$=7$	4.12	16.97	16.47
	$\Sigma f = 25$	$\Sigma fx = 72$			$\Sigma f(x-\bar{x})^2 = 80.08$

$$\text{சராசரி} = \frac{\Sigma fx}{\Sigma f} = \frac{72}{25} = 2.88$$

$$SD = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f - 1}} = \frac{80.08}{24} = \sqrt{3.34} = \underline{1.83}$$

$$\text{திட்டப்பிழை } (SE) = \frac{SD}{\sqrt{n}} = \frac{1.83}{\sqrt{25}} = \frac{1.83}{5} = \underline{0.366}$$

ஆ) உத்தேச சராசரி முறை

திட்டவிலக்கம் $\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$

இங்கு $d = (X - A)$

மாதிரி : கொடுக்கப்பட்ட விவரங்களுக்கு திட்ட விலக்கம் காண்

X − 1, 2, 3, 4, 5

f − 3, 7, 10, 3, 2

தீர்வு = 3 ஐ உத்தேச சராசரியாகக் கொள்வோம்

X	f	$(X-A)d$ $A = 3$	fd	fd^2
1	3	−1	−6	12
2	7	−2	−7	7
3	10	0	0	0
4	3	1	3	3
5	2	2	4	8
	25		−6	30

குறிப்பு : fd உடன் d ஐப் பெருக்க fd^2 கிடைக்கும்

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}$$

$$\sigma = \sqrt{\frac{30}{25} - \left(\frac{-6}{25}\right)^2}$$

$$= \sqrt{12 - \left(\frac{36}{625}\right)}$$

$$= \sqrt{12 - 0.05}$$

$$\sigma = \underline{\underline{1.07}}$$

Calculate of S.D Discrete Series:

1. Direct method.

Steps: 1. Calculate the mean

2. Find out deviation of the various values from the mean value $(x-\bar{x})$

3. Square the deviations $(x-\bar{x})^2$

4. Multiply $(x-\bar{x})^2$ with the respective frequencies (f) against various values and add all such values $\Sigma f(x-\bar{x})^2$

5. Divide $\Sigma f(x-\bar{x})^2$ by the number of items N (or) Σf.

6. Apply the formula $S.D = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}}$

eg: Calculate the standard deviation for the following data.

Size of item (X)	6, 7, 8, 9, 10, 11, 12
Frequency	3, 6, 9, 13, 8, 5, 4

solution:

Size of item (x)	Frequency (f)	fx	$\bar{x}=9$ $(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
6	3	18	$6-9=-3$	9	$3\times9=27$
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	$=36$
	$\Sigma f=48$	$\Sigma fx=432$			$\Sigma f(x-\bar{x})^2 = 124$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f (or) N} = \frac{432}{48} = 9$$

$$SD = \sqrt{\frac{\Sigma f (x-\bar{x})^2}{N}} = \sqrt{\frac{124}{48}} = \sqrt{2.58}$$

$$SD = \underline{1.6}$$

Short-cut method:

Steps:

1. Assume any one of the values in the data as assumed mean (A)
2. Find out deviations of each value from the assumed mean $(x - A = d)$
3. Square the deviation d^2
4. Multiply d^2 with the respective frequencies (f) against various values and all such values Σfd^2
5. Apply the formula $SD = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

eg: Find out S.D for the following data:

X - 18, 19, 20, 21, 22, 23, 24, 25, 26, 27

f - 3, 7, 11, 14, 18, 17, 13, 8, 5, 4

x	f	A = 22 x − A = d	fd	d^2	fd^2
18	3	18 − 22 = −4	3×−4 = −12	16	3 × 16 = 48
19	7	= −3	= −21	9	= 63
20	11	= −2	= −22	4	= 44
21	14	= −1	= −14	1	= 14
22	18	= 0	= 0	0	= 0
23	17	= 1	= 17	1	= 17
24	13	= 2	= 26	4	= 52
25	8	= 3	= 24	9	= 72
26	5	= 4	= 20	16	= 80
27	4	= 5	= 20	25	= 100
Σf = 100			Σfd = 38		Σfd^2 = 490

Formula: $SD = \sqrt{\dfrac{\sum fd^2}{N} - \left[\dfrac{\sum fd}{N}\right]^2}$

$= \sqrt{\dfrac{490}{100} - \left[\dfrac{38}{100}\right]^2} = \sqrt{4.9 - 0.144}$

$= \sqrt{4.756}$

$SD = \underline{2.2}$

[3] தொடர் அலைவெண் தொடரை வரையறுக்கவும்:

(அ) நேரடி முறை:

பயிற்சி: நாற்றின் மட்டு நீள அளவுகள் (x) மற்றும் அலைவெண்கள் (f) கொடுக்கப்பட்டுள்ளன. திட்டவிலக்கம் காண்க.

நா.மட்டுகள் நீளம் (செ.மீ)	4-5	5-6	6-7	7-8	8-9	9-10
அலைவெண்	5	7	17	15	7	5

வகுப்பு இடைவெளி (c.m)	மைய மதிப்பு (midx)	அலை வெண் (f)	f(midx)	midx = 6.98 midx $-\bar{x}$	$(midx-\bar{x})^2$	$f(midx-\bar{x})^2$
4 - 5	4.5	5	22.5	6.98-4.5 = -2.48	6.15	30.75
5 - 6	5.5	7	38.5	= -1.48	2.19	15.33
6 - 7	6.5	17	110.5	= -0.48	0.23	3.91
7 - 8	7.5	15	112.5	= 0.52	0.27	4.06
8 - 9	8.5	7	59.5	= 1.52	2.31	16.17
9 - 10	9.5	5	47.5	= 2.52	6.35	31.75
		= 56	= 391		= 17.5	= 101.96

$\bar{x} = \dfrac{\sum f(midx)}{\sum f} = \dfrac{391}{56} = 6.98$

$SD = \sqrt{\dfrac{\sum f\,(midx - \bar{x})^2}{N}} = \sqrt{\dfrac{101.96}{56}} = \sqrt{1.82} = \underline{1.349}$

திட்டவிலக்கம் = $\underline{1.349}$

மாறுபாடு (Variance) $= \frac{\Sigma (mid\,x - \bar{x})^2}{N} = \frac{101.96}{56} = \underline{1.82}$

திட்டப்பிழை (S.E) $= \frac{1.349}{\sqrt{56}} = \frac{1.349}{7.48} = \underline{0.1803}$

(ஆ) இடைவெளி முறை:

மதிப்பு:

மதிப்பு எண்கள்	0-10	10-20	20-30	30-40	40-50
மாணவர்கள்	3	4	7	6	5

CI (X)	f	x - மைய மதிப்பு m	$d = m - A$, $A = 25$	$d_1 = \frac{m-A}{C}$, $C = 10$	fd'	fd_1^2
0-10	3	5	$5-25=-20$	$-\frac{20}{10}=-2$	$3\times-2=-6$	$3(-2)^2=12$
10-20	4	15	$=-10$	$=-1$	$=-4$	$4(1)^2=4$
20-30	7	25	$=0$	$=0$	$=0$	$7(0)^2=0$
30-40	6	35	$=10$	$=1$	$=6$	$6(1)^2=6$
40-50	5	45	$=20$	$=2$	$=10$	$5(2)^2=20$
	25				$=6$	$=42$

$$\sigma = \sqrt{\frac{\Sigma fd_1^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} \times C$$

$$\sigma = \sqrt{\frac{42}{25} - \left(\frac{6}{25}\right)^2} \times 10$$

$$= \sqrt{1.68 - 0.058} \times 10 = \sqrt{1.6224} \times 10 = \sqrt{1.274} \times 10$$

திட்டவிலக்கம் $= \underline{12.74}$

Calculation of S.D - Continuous Series:

1. eg: ovary wt. of 50 fishes and their frequency is given in class interval. Find S.D.

wt. of ovary	2 - 2.9	3 - 3.9	4 - 4.9	5 - 5.9	6 - 6.9
frequency	6	13	11	8	12

Step 1: Find mid point (m) of each class interval.

2. Find mean of the series applying formula $\bar{x} = \frac{\Sigma f x}{\Sigma f}$.
3. Find deviation of each observation $(x - \bar{x})$.
4. Square each deviation.
5. Multiply each squared deviation with their frequency.
6. Sum all the multiplied value of f and x^2.

Class interval	mid point (m)	Frequency (f)	fm	deviation $x - \bar{x} = x$	Deviation square x^2	fx^2
2 - 2.9	2.45	6	14.7	- 2.14	4.579	27.47
3 - 3.9	3.45	13	44.85	- 1.14	1.299	16.88
4 - 4.9	4.45	11	48.95	- 0.14	0.019	0.21
5 - 5.9	5.45	8	43.6	0.86	0.739	5.91
6 - 6.9	6.45	12	77.4	1.86	3.459	41.5
		$\Sigma f = 50$	$\Sigma fm = 229.5$		$\Sigma x^2 = 10.088$	$= 91.97$

$$\bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{229.5}{50} = 4.59$$

$$S = \sqrt{\frac{\Sigma f x^2}{\Sigma f}} = \sqrt{\frac{91.97}{50}} = \sqrt{1.84}$$

$$= \underline{1.35}$$

Variance

- The term variance was first used by R.A Fisher in 1918.
- The square of the Standard deviation is the variance.

$$\text{Variance} = S.D^2 = \sigma^2$$

$$SD^2 = \text{Variance}$$

$$SD = \sqrt{\text{Variance}}$$

Standard Error

- when more than one Sample is drawn from a population, the Standard deviation for each mean will be found to differ slightly.
- In theory, the mean of different Samples drawn from the Same population must be the same or very close to each other, So that a single sample is a reliable true measure of the population. But
- In practice this is not so and means of Samples show variation.

A statistical constant which measures the dispersion of the sample means around the total population mean is called S.E (Standard Error) and it obtained by dividing standard deviation by the sample size

Formula = $S.E = \frac{S.D}{\sqrt{N}}$

where . S.D = Standard deviation
N = Sample Size.

Coefficient of Variance
(Relative Standard Deviation)

- The standard deviation must be converted into a relative measure of dispersion for the purpose of comparison.
- The relative measure is known as the Coefficient of variation (C.V) . This measure is developed by Karl pearson.
- If the coefficient of variation of a group is greater, then it is said that the group is more variable. less stable, less uniform, less consistent or less homogenous and vice versa.

formula : $C.V = \frac{\sigma}{\bar{X}} \times 100$

coefficient of variation = $\frac{\text{Standard deviation}}{\text{mean}} \times 100$

eg 7: The following table gives the wages of employees in two factories: In which factory there is greater variation in the distribution of wages per employee?

	Factory A	Factory B
No. of employees	50	100
Average wage per month per worker	120	85
Variance of the wage per employee per month	9	16

Factory A

$$\sigma^2 = 9 \quad \therefore \sigma = 3$$

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

$$= \frac{3}{120} \times 100 = \underline{\underline{2.5}}$$

Factory B

$$\sigma^2 = 16 \quad \therefore \sigma = 4$$

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

$$= \frac{4}{85} \times 100 = \underline{\underline{4.71}}$$

Since C.V. for factory B is greater, there is greater variation in distribution of wages.

eg 2: Prices of a particular Commodity in five year in two States are given below. from this data, find the State which had more Stable prices.

Price in Tamil Nadu	Price in Kerala
22	20
20	10
23	12
19	18
16	15

$\bar{x} = 20$

$S.D = 2.45$

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

$$= \frac{2.45}{20} \times 100$$

$$= \underline{\underline{12.25}}$$

$\bar{x} = 15$

$S.D = 3.69$

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

$$= \frac{3.69}{15} \times 100$$

$$= \underline{\underline{24.6}}$$

Tamil Nadu had more Stable prices than Kerala because the C.V is lower in Tamil Nadu.

Uses:

- The coefficient of variation is most useful in Comparing the variability of Several different Samples, each with different arithmetic means.
- Higher variability is usually expected when the mean increases and the Coefficient of Variation is a measure that accounts for this variability.
- A more accurate Comparison could be made by Comparing the Coefficient of variation than comparing the standard deviations.

Coefficient of Variation

மாறுபாட்டுக் கெழு

மாறுபாட்டுக் கெழு (அ) மாற்றங்களைக் கணிக்கும் ஒப்பீட்டளவான அளவையாகும். உதாரணமாக

- ஒரு வகுப்பு மாணவர்களின் சராசரி உயரம் 60" என்க. அவர்களின் சராசரி எடை 120 பவுண்டுகள். அவர்களின் உயரத்தின் திட்டவிலக்கம் 3" என்றும் எடையின் திட்டவிலக்கம் 4 பவுண்டுகள் எனவும் கொள்க.

- இங்கு மேலோட்டமாக எடை அதிகமாக மாறி உள்ளது திட்டவிலக்கத்தை ஒப்பிட்டால் தெரிகிறது. ஆனால் உண்மை நிலை என்ன? சராசரியோடு திட்டவிலக்கத்தை ஒப்பிடும்போது தான் உண்மை நிலை புரியும்.

எடை $\frac{4}{120} \times 100 = 3.3\%$

உயரம் $\frac{3}{60} \times 100 = \boxed{5\%}$

- உண்மையில் உயரமே அதிக மாறுதல் உடையது எனப் புரிந்து கொள்கிறோம்.

- ஒரு சராசரிக்கு மாறுதல் அளவையின் மதிப்பு என்ன என்பதைத் தெரிவிப்பது மாறுபாட்டுக் கெழுவாகும்.

- எப்பரவலுக்கு மாறுபாட்டுக் கெழு குறைவாக உள்ளதோ அப்பரவல் மாறுதல் நிலை குறைவானதாகும்.

- மாறுபாட்டுக் கெழு அதிகமாக இருந்தால் மாறுதல் அதிகம் எனக் கொள்ளவேண்டும்.

• நிலைப்புத்தன்மையை அறிந்துகொள்ள இது பயன்படுகிறது.

எ.கா - மாறுபாட்டுக்கெழு அதிகம் (புள்ளி விவரம்) → குறைந்த நிலைப்புத் தன்மையுடையது.

(அதிக மாறுபாடு உடையது)

எ.கா - மாறுபாட்டுக்கெழு குறைவாக இருந்தால் (புள்ளி விவரம்) → அதிக நிலைப்புத் தன்மையுடையது.

(குறைந்த மாறுபாடுடையது)

• எனவே மாறுபாட்டுக்கெழு — நிலைப்புத்தன்மை / ஒப்பீட்டிற்காக கணக்கிடப்படுகிறது.

$$\text{மாறுபாட்டுக்கெழு } [C.V] = \frac{\sigma}{\overline{X}} \times 100$$

கணக்கு 1: A, B என்ற இரண்டு கிரிக்கெட் வீரர்கள் பத்து ஆட்டங்களில் எடுத்த ஓட்டங்கள் கொடுக்கப்பட்டுள்ளது. அதிக நிலைப்புத் தன்மையுடைய ஆட்டக்காரர் யார்?

A	100	60	80	20	30	90	100	5	15	50
B	70	85	80	85	60	65	75	85	60	55

தீர்வு:

'A' எடுத்த ஓட்டங்களை X_A என்றும்

'B' எடுத்த ஓட்டங்களை X_B என்றும் கொள்வோம்.

$$\overline{X}_A = \frac{550}{10} = \boxed{55}$$

$$\overline{X}_B = \frac{700}{10} = \boxed{70}$$

X_A	$(X_A-\bar{X}_A)$	$(X_A-\bar{X}_A)^2$	X_B	$(X_B-\bar{X}_B)$	$(X_B-\bar{X}_B)^2$
100	45	2025	70	0	0
60	5	25	65	-5	25
80	25	625	80	10	100
20	-35	1225	85	15	225
30	-25	625	60	-10	100
90	35	1225	65	-5	25
100	45	2025	75	5	25
5	-50	2500	85	15	225
15	-40	1600	60	-10	100
50	-5	25	55	-15	225
550	0	11900	700	0	1050

திட்ட விலக்கம் $\sigma = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

∴ A-யின் தி. விலக்கம் $= \sqrt{\dfrac{11900}{10}}$

$= \sqrt{11900} = \underline{34.49}$

A யின் மாறுபாட்டுக் கெழு $= \dfrac{34.49}{55} \times 100 = \underline{62.70\%}$

B யின் திட்டவிலக்கம் $= \sqrt{\dfrac{1050}{10}} = \sqrt{105} = \underline{10.24}$

B யின் மாறுபாட்டுக் கெழு $= \dfrac{10.24}{70} \times 100 = \underline{14.63\%}$

B யின் மாறுபாட்டுக் கெழு A ஐ விட குறைவாக இருப்பதால் ஆட்டக்காரர் B அதிக நிலைப்புத் தன்மை உடையவர்.

Correlation (Karl pearson)

ഉൾബന്ധം

- ഒരു ചരത്തിന്റെ മൂല്യത്തിൽ മാറ്റം വരുമ്പോൾ അതിനോട് ബന്ധമുള്ള മറ്റൊരു ചരത്തിന്റെ മൂല്യത്തിലും ഉണ്ടാകുന്ന ആനുപാതിക മാറ്റത്തെ ഉൾബന്ധം (r) കൊണ്ട് സൂചിപ്പിക്കുന്നു എന്ന് പറയുന്നു.

  ഉദാ. മനുഷ്യന്റെ ഭാരം ഉയരത്തെ ആശ്രയിച്ചിരിക്കും.

- Correlation analysis is the study of relationship between two or more variables. The measure of correlation is called correlation co efficient.

  eg: The weight of a man depends on height. when height increases weight also Increases.

- The correlation is measured by a method called pearson's Coefficient of correlation devised by Karl pearson. It explains the degree of relationship between two variables. it was denoted by the symbol 'r', formula for this is

$$\boxed{r = \frac{\Sigma dx\,dy}{\sqrt{\Sigma dx^2 \times \Sigma dy^2}}}$$

$dx = x - \bar{x}$ = deviation from $\bar{x}$

$dy = y - \bar{y}$ = deviation from $\bar{y}$

$\Sigma dx \cdot dy$ = Multiply the deviations of x and y and get the total.

Σdx^2 = The sum of the square of the deviations of x

Σdy^2 = The sum of square of the deviations of y

- கார்ல் பியர்சனின் உடன் தொடர்புக் கெழு முறை இரண்டு மாறிகளுக்கிடையேயுள்ள உடன் தொடர்பினைத் துல்லியமாக அளவிட்டு கூறுவதாகும். இதை 'r' என்ற எழுத்தால் குறிக்கப்படுகிறது.

உடன் தொடர்பின் வகைகள்:

Types of correlation:

(i) நேரிடை, எதிரிடை உடன் தொடர்பு:

Direct, Inverse correlation:
(positive, negative correlation)

- இரண்டு மாறிகளின் மதிப்புகளில் ஏற்படுகின்ற மாற்றம் ஒரே திசையில் நடைபெற்றால். அதாவது ஒரு மாறியின் மதிப்புகள் அதிகரிக்கின்றபோது மற்றமாறியின் மதிப்புகளும் அதிகரிக்க வேண்டும் (அ) குறைந்தால் குறைய வேண்டும்.
- If two variables tend to move together in the same direction. i.e. an increase in the value of one variable is accompanied by an increases in the value of the other variable (or) a decreases in the value of one variable is accompanied by a decrease in the value of the other variable: Then the correlation is called positive correlation.

(ii) நேர்கோட்டு, வளைகோட்டு உடன் தொடர்பு!

Linear and Curve linear correlation.

இரண்டு மாறிகளில் ஏற்படும் மாற்றங்கள் ஒரே விகிதத்தில் இருந்தால் நேர்கோட்டு உடன் தொடர்பு என்றும். இல்லையென்றால் விகிதங்களில் இருந்தால் அவை வளைகோட்டு உடன் தொடர்பு உடையவை என்போம்.

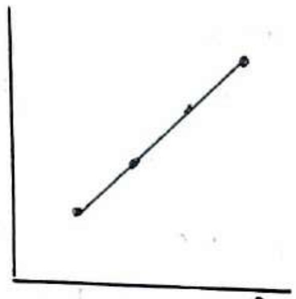
- If the ratio of change between two variable is uniform then there will be linear correlation between them. eg:

x	5	10	15	20	25
y	3	6	9	12	15

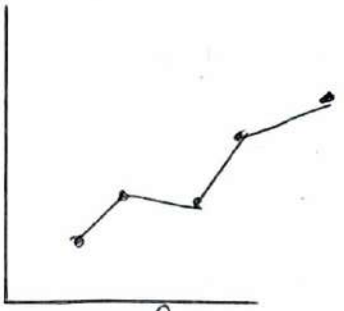
x: $5+5=10$
$10+5=15$
$15+5=20$
$20+5=25$

$y = 3+3=6$
$6+3=9$
$9+3=12$
$12+3=15$

If we plot these on a graph, we get a straight line like this



Linear Correlation



Non-linear Correlation

- In a non-linear correlation, the amount of change in one variable does not bear a constant ratio to the amount of change in the other variables.

X	5	10	15	20	25
Y	3	5	6	9	10

X : $5+5=10$, $10+5=15$, $15+5=20$, $20+5=25$

Y : $3+2=5$, $5+1=6$, $6+3=9$, $9+1=10$

If we plot these on a graph, we get a curve.

iii) எளிமையான, பன்முக பகுதி உடன் தொடர்பு:
Simple, multiple, partial Correlation:

ஏதேனும் இரண்டு மாறிகளுக்கிடையே மட்டும் நிலவும் உடன் தொடர்பினை அவ்வது எளிமையான (அ) சாதாரண உடன் தொடர்பு எனப்படும்.

When we study two variables, the relationship is described as Simple Correlation: eg: Diet, age and weight of the baby.

இரண்டுக்கு மேற்பட்ட மாறிகளுக்கிடையேயுள்ள உடன் தொடர்பு பன்முக உடன் தொடர்புகளாக (அ) பகுதி உடன் தொடர்பாகவோ இருக்கலாம். எ.கா. குழந்தையின் வயது, உணவு முறை, உயரத்தின் அளவு, எடையின் அளவு, பிறவற்றை ஆகிய பலவற்றைச் சார்ந்திருப்பவைகளே பன்முக உடன் தொடர்பு, இதில் எல்லாவற்றையும் இரண்டு மட்டும் தனித்தனியே பகுத்து அவைகள் இணைகளுக்கிடையே உள்ள தொடர்பினைப் பற்றி தனித்தனியே அப்படியே பகுதி உடன் தொடர்பு ஆகும். எ.கா. குழந்தையின் வயதுக்கும் எடையின் அளவுக்கும் உள்ள உடன் தொடர்பை மட்டுமே எடுத்துக்கொள்வது.

உடன் தொடர்பினை அறியும் முறைகள்:
Methods of Studying Correlation:

- கார்ல் பியர்சனின் உடன் தொடர்புக்கெழு.
  Karl Pearson's Coefficient of Correlation.

- இது ஒரு கணித முறையாகும். இவ்விரண்டின் கூட்டு மாறிகளுக்கு - இடையேயான உடன் தொடர்பினைத் துல்லியமாக அளவிடும் கார்ல்பியர்சனின் உடன் தொடர்புக்கெழுவை கணக்கிடுகிறோம். இது 'r' என்ற எழுத்தால் குறிக்கப்படுகிறது.
- கார்ல்பியர்சனின் r-ன் மதிப்பு எப்போதும் 1-க்கும் +1-க்கும் இடையில் இருக்கும்.
- The value of the coefficient of correlation shall always lie between +1 and -1,
- r-ன் மதிப்பு +1 என்றால் முழுமையான நேர்மறை உடன் தொடர்பு ஆகும். When $r = +1$, then there is perfect positive correlation between the two variables.
- r-ன் மதிப்பு -1 எனில் முழுமையான எதிர்மறை உடன் தொடர்பு. When $r = -1$, then there is perfect negative correlation between the two variables.
- r-ன் மதிப்பு 0 என்றால் உடன் தொடர்பு இல்லை. When $r = 0$, then there is no relationship (or) correlation between two variables.
- r-ன் மதிப்பு 0 முதல் +1 க்கும் இடையில் இருப்பின் நேர்மறை உடன் தொடர்பு (positive r)
  r-ன் மதிப்பு -1 க்கும் 0 க்கும் இடையில் இருந்தால் எதிர்மறை (Negative r) உடன் தொடர்பு.
- Theoretically, we get values which lie between +1 and -1; but normally the value lies between +0.8 and -0.5.

उदा. (Illustration): Calculate pearson's correlation Coefficient

X	12	18	16	15	12	10	20	17
Y	6	10	9	8	9	8	12	10

Solution

X	Y	dx ($\bar{x}=15$) $x-\bar{x}$	$(dx)^2$	dy ($\bar{y}=9$) $y-\bar{y}$	$(dy)^2$	dx × dy = dx dy
12	6	12−15 = −3	9	6−9 = −3	9	9
18	10	= 3	9	= 1	1	3
16	9	= 1	1	= 0	0	0
15	8	= 0	0	= −1	1	0
12	9	= −3	9	= 0	0	0
10	8	= −5	25	= −1	1	5
20	12	= 5	25	= 3	9	15
17	10	= 2	4	= 1	1	2
$\Sigma x = 120$	= 72	= 0	= 82	0	= 22	= 34

$$\bar{x} = \frac{\Sigma x}{N} = \frac{120}{8} = 15, \quad \bar{y} = \frac{\Sigma y}{N} = \frac{72}{8} = 9$$

$$r = \frac{\Sigma dx\, dy}{\sqrt{\Sigma dx^2 \times \Sigma dy^2}} = \frac{34}{\sqrt{82 \times 22}} = \frac{34}{9.05 \times 4.69}$$

$$= \frac{34}{42.44} = r = \boxed{+0.80}$$

It is a positive correlation. The Two variables x and y are interdependent.

கார்ல் பியர்சனின் உடன் தொடர்புக்கெழுவின் சிறப்புகளும் குறைகளும் :

Merits and Demerits of Karl pearson Coefficient of correlation :

சிறப்புகள் :

1. இது சுலபமும் துல்லியமான கணக்கிடும் முறையாகும்.
2. ஒரு குறிப்பிட்ட எண்ணின் மூலம் (ஒ-வின் மதிப்பு) உடன் தொடர்பின் போக்கினையும், அளவினையும் அறிந்து கொள்ள முடியும்.
3. பொருளியல், அறிவியல் ஆகிய பல துறைகளிலும் ஆய்வுக்குப் பயன்படுகிறது.
4. உத்தேசப் பிழை (probable error) மற்றும் நிர்ணயக் கெழு (coefficient of determination) ஆகியவற்றைக் காணப் பயன்படுகிறது.

Merits :

1. It is simple to understand.
2. It is easy to calculate.
3. It is very useful in the case of data which are of qualitative.
4. When the actual data are given, this method can also be used.

Demerits :

1. It cannot be used in the case of bivariate destribution.
2. If the no. of items are greater, the calculation becomes tedious and requires a lot of time.

குறைகள்:

1. இதனைக் கணக்கிடுவது மிகவும் கடினமானது.
2. கணக்கிடுவதற்கு அதிக நேரம் தேவைப்படுகிறது
3. மாறிகள் நேர்க்கோட்டு உடன் தொடர்பினைக் கொண்டிருக்கின்றது நடைமுறையில் பல மாறிகள் இத்தகைய தொடர்பினைக் கொண்டி -ருப்பதில்லை.
4. தொகுதியிலுள்ள மிகப்பெரிய (அ) மிகச் சிறிய மதிப்புக்களால் உடன் தொடர்புக் கெழுவின் மதிப்பு பாதிக்கப்படுகிறது.

Regression Analysis (தொடர்புப் போக்கு)

- தெரிந்த ஒரு மாறியின் மதிப்பின் மூலம் அதனோடு தொடர்பு உடைய தெரியாத மற்றொரு மாறியின் மதிப்பை மதிப்பீடு செய்வதற்கு உதவும் புள்ளியியல் முறையே தொடர்பு போக்கு ஆகும்.

- தொடர்பு போக்கு என்பது மாறிகளுக்கிடையே உள்ள தொடர்பினை காணும். அவைகளின் மதிப்பை மதிப்பீடு செய்யவோ அல்லது முன்கணிப்பு செய்யவோ உதவுகின்றது.

- In regression analysis, the value of dependent variable can be calculated from the value of independent variable.

- Thus in regression analysis the value of an unknown variable can be calculated from the value of a known variable.

- Regression is the measures of the average relationship between two or more variables in terms of the original units of the data.

- Literally regression means going back or stepping back.

- One variable is represented as x and the other variable is represented as y.

- The graphic representation of regression is called regression line. There are two regression lines.

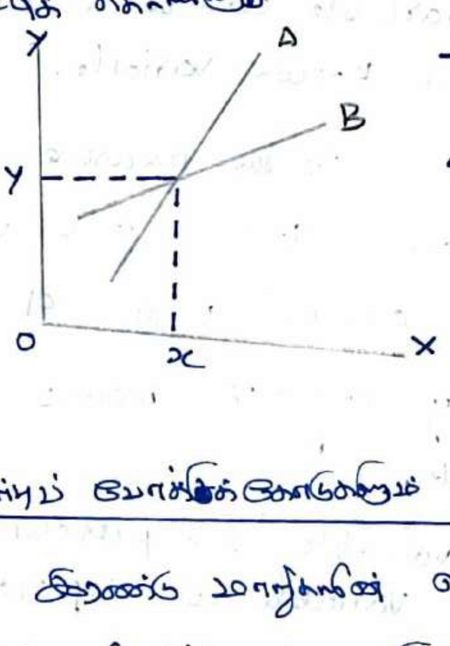
  Regression line of x on y.
  Regression line of y on x.

எளிய நேர்கோட்டுத் தொடர்புப் போக்கு
Simple Linear Regression

- சாரா மாறியினை X என்றும்.
- சார்ந்த மாறியினை Y என்றும் குறிப்போம்
- இரண்டு மாறிகளுக்கிடையே உள்ள தொடர்புப் போக்கினை சமன்பாடுகள் மூலம் குறிக்கலாம். அச்சமன்பாடுகளுக்கு உரிய கோட்டினை வரைபடத்தில் வரையும் போது தொடர்பு போக்கு கோடுகள் கிடைக்கின்றன.

1. X-ன் மதிப்பு Y-ன் மதிப்பை சார்ந்தும் (Regression line of X on Y) Y-ன் மதிப்பு X-ன் மதிப்பை சார்ந்தும் அமைந்தால் (Regression line of Y on X) இரண்டு தொடர்புக் கோடுகள் கிடைக்கும்.

2. இவ்வாறு வரையப்படுகின்ற இரண்டு தொடர்புப் போக்குக் கோடுகளும் ஒன்றையொன்று (X,Y) என்ற புள்ளியில் வெட்டிக் கொள்ளும்



- A என்பது X-ல் Y-ன் தொடர்புப் போக்குக் கோட்டையும்
- B என்பது Y-ல் X-ன் தொடர்புக் கோட்டையும் குறிக்கும்.

தொடர்புப் போக்குக் கோடுகளும் உடன் தொடர்பும்:

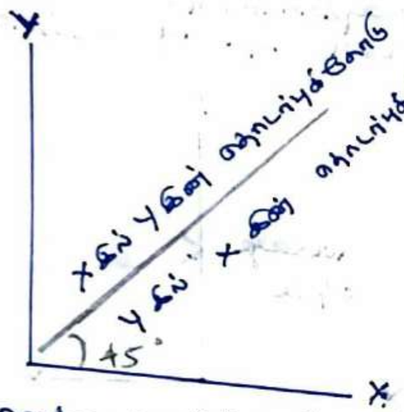
இரண்டு மாறிகளின் தொடர்புப் போக்குக் கோடுகளின் அமைப்புகளிலிருந்து அவைகளுக்கிடையே உள்ள உடன் தொடர்பினைப் பற்றி அறிய முடியும்.

1) X, Y என்ற இரண்டு மாறிகளை எடுத்துக்கொள்வோம். அவைகளின் இரண்டு தொடர்புப் போக்குக் கோடுகளும் ஒன்றின் மீது ஒன்று படிந்தால் அவ்விரண்டு மாறிகளும் முழுமையான உடன் தொடர்புடையன.

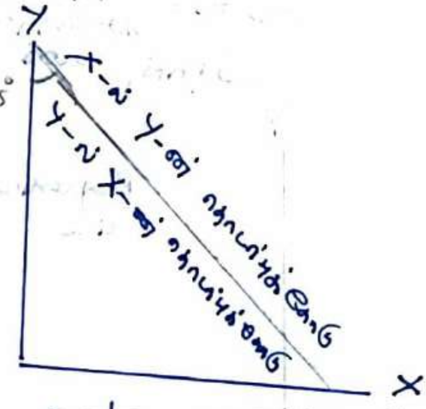
மேலும் கீழை நோக்கி செல்லும். அமைந்து வரைந்தால் அதற்கு முழுமையான நேர்மறை உடன் தொடர்பாகும் $r = +1$.

2. இரண்டு தொடர்புப் போக்குக் கோடுகளும் ஒன்றன் மீது ஒன்று படிந்து மேலிருந்து கீழாக அமைந்து வரைந்தால் இரண்டு மாறிகளுக்கிடையே முழுமையான எதிர்மறை உடன் தொடர்பு நிலவும் $r = -1$.

3. When there is perfect positive correlation or perfect negative correlation, the two regression line will coincide with each other and there will be only one regression line.



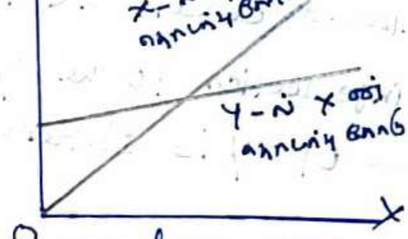
perfect positive r
முழுமையான நேர்மறை r



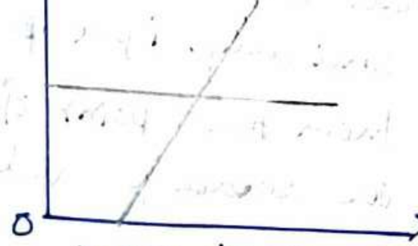
perfect negative r
முழுமையான எதிர்மறை r

4. இரண்டு தொடர்புக் கோடுகளும் நன்றாக நெருங்கி அமைந்திருந்தால் இரண்டு மாறிகளுக்கிடையே அதிக அளவு உடன் தொடர்பு இருக்கும்.

If the two variable regression lines are nearer to each other, then there is a high degree of correlation.



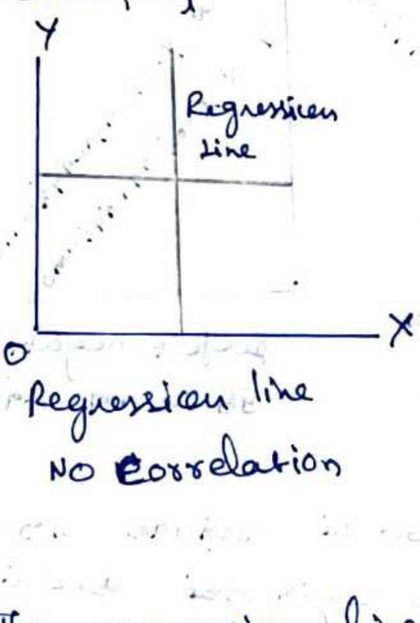
more degree of r
அதிக அளவு உடன் தொடர்பு
$r = 1$



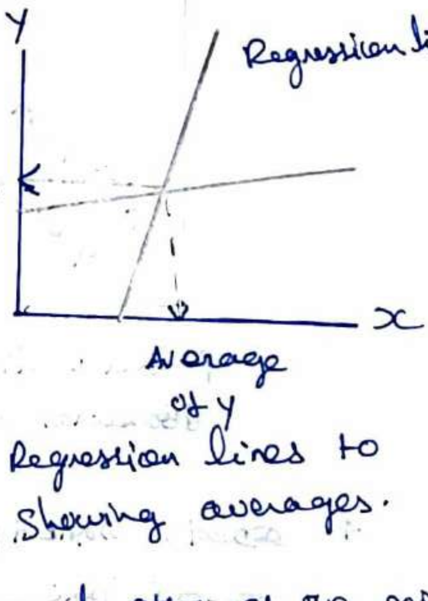
Lesser degree of r
குறைவான அளவு உடன் தொடர்பு
$r = -1$

5. இரண்டு தொடர்புக் கோடுகளும் ஒன்றுக்கொன்று விலகியிருந்தால்
விரிந்த தொலைவிலுள்ள குறைவான அளவு உடன் தொடர்பு
இருக்கும் r-ன் மதிப்பு பூஜ்யத்தை நோக்கி செல்லும்
If the regression line are farther away from each other, then there is lesser degree of correlation.

6. If the two variable are independent, there will be no correlation. So in such cases both the regression lines cut each other at right angles.
இரண்டு தொடர்புக் கோடுகளும் ஒன்றை ஒன்று செங்குத்தாக வெட்டிக்கொண்டால்
இருந்தால், அங்கு உடன் தொடர்பு இருக்காது $r = 0$



Regression line
No Correlation



Regression lines to showing averages.

7. • The regression lines cut each other at the point of average of X and Y.
• If a perpendicular line is drawn from the point of intersection to the X axis the average value of X is obtained.
• Similarly, if a perpendicular line is drawn from the point of intersection to the Y axis, the average value of Y is obtained.

Types of Regression Analysis

தொடர்பு போக்கின் வகைகள்

1. **Simple and multiple: எளிய தொடர்புப் போக்கு:**

• The regression analysis confined to the study of only two variables at a time is termed as simple regression.

• இரண்டு மாறிகளை மட்டும் கொண்டிருந்த தொடர்புப் போக்கு எளிய(து) சாதாரணத் தொடர்புப் போக்கு ஆகும்.

• The regression analysis for studying more than two variables at a time is known as multiple regression.

• தொடர்புப்போக்கின் மூலம் இரண்டிற்கு மேற்பட்ட மாறிகளின் தொடர்பினை ஆய்கின்றபோது அதனைப் பன்மைத் தொடர்புப் போக்கு என்கின்றோம்.

2. **Linear and Non-linear: நேர்கோட்டு மற்றும் வளைக் கோட்டுத் தொடர்புபோக்கு:**

• In linear regression, the regression is a straight line. On the other hand in non-linear regression, the regression is not a straight line.

• ஒரு தொடர்புப்போக்கில் அமைந்துள்ள மாறிகளில் மதிப்புகளின் மாற்ற வீதங்கள் மாறாமல் இருப்பின் அவை நேர்கோட்டுத் தொடர்புபோக்கு ஆகும். அவ்வாறின்றி மாற்ற வீதங்கள் மாறுபட்டு இருப்பின் அவை வளைகோட்டுத் தொடர்புபோக்கு ஆகும்

3. **Total and partial: முழுமையான மற்றும் பகுதி தொடர்பு போக்கு:**

• In total regression, all the important variables are considered. But in case of partial one or more variables are considered.

முழுமையான தொடர்பு போக்கில் அனைத்து உறவுகளும் எடுத்துக்கொள்ளப்படும். அதனால் பகுதியால் ஒன்று (or) இரண்டு தொடர்புடைய உறவுகள் மட்டுமே எடுத்து கொள்ளப்படும்.