2025-07-18 Analyze data for solving the problems of Exploratory Data Analysis.

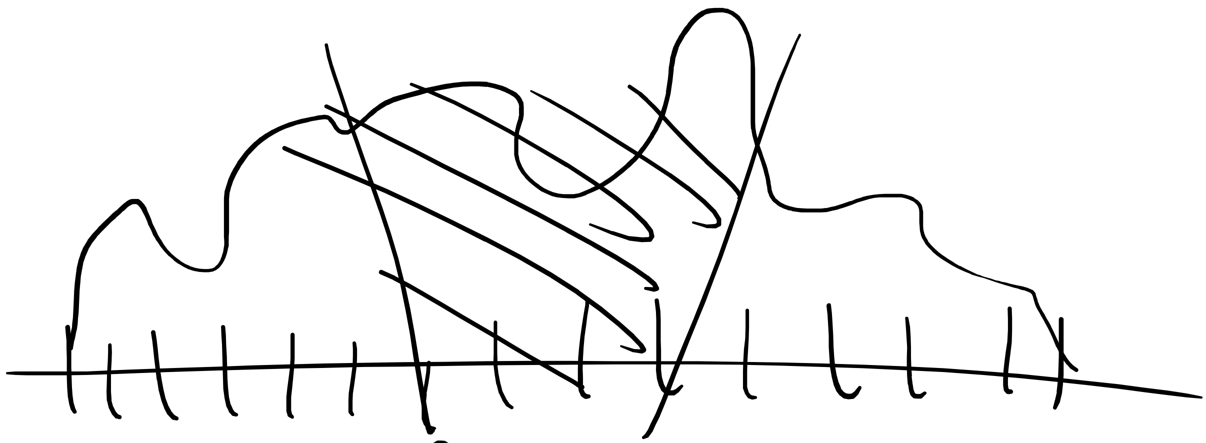
18 July 2025 15:32

Measures of

Central Tendency

Dispersion

=>



Mean, Median & Mode.

=> Measures of Dispersion

-> Standard Deviation, Variance

=> Standard Deviation

Deviation = Error

=> Population : Entire Data N

=> Sample : Desired Sub-space . n

1400000000

1.4B

200M ~ Mah ~ 15M Mum

250M ~ UP

180M ~ RJ

1400

200 MH ~ 15 Mum

250 UP

180 RJ

=> Deviation

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

=> Deviation

$$\rightarrow \sum_{i=1}^{N} (x_i - \bar{x})$$

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) \cdots (x_N - \bar{x})$$

$$\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \sim \text{Deviation (Normalized)}$$

Population S.D

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}}$$

Sample S.D

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$$

=> **Variance :**

Pop Variance

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{n\,1}$$

$\overline{\qquad\quad}$

N

Sample Variance

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

=> Co-variance

$$\boxed{Cov(x,y) = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N-1}}$$

N – Population

n – Sample (N–1)

=> x(2, 4, 6, 8, 10) & y(3, 7, 10, 14, 17)

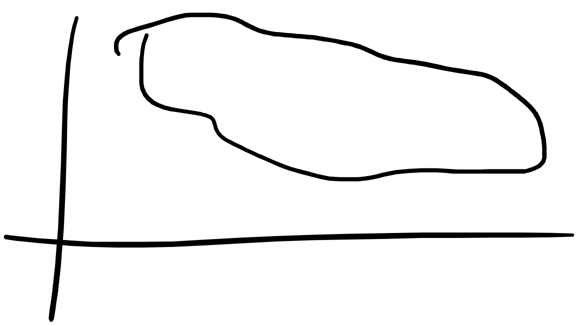
x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
2	3	−4	−7.2	28.8
4	7	−2	−3.2	6.4
6	10	0	−0.2	0
8	14	2	3.8	7.6
10	17	4	6.8	27.2

10	17	4	6.8	27.2
$\bar{x} = 6$	$\bar{y} = 10.2$			$\Sigma = 70$

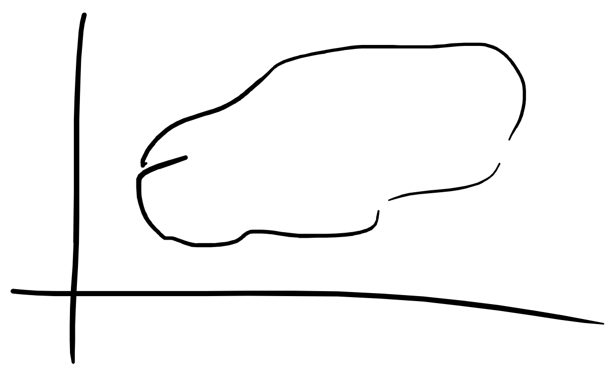
$$Cov_{(x,y)} = \frac{\Sigma (x_i - \bar{x}) \cdot (y_i - \bar{y})}{N-1}$$

$$= \frac{70}{5-1} = \frac{70}{4} = 17.5$$

Cov +ve

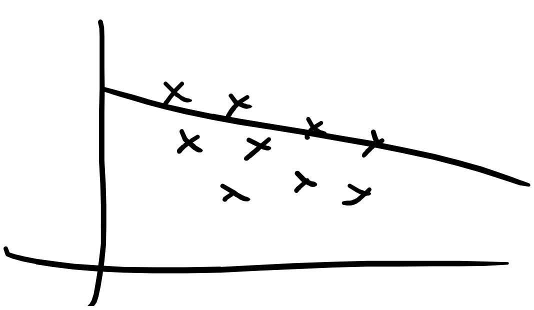


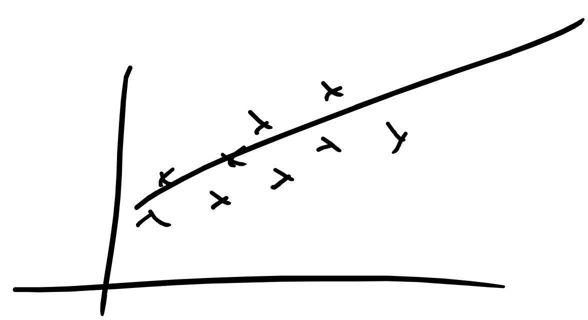
Cov -ve

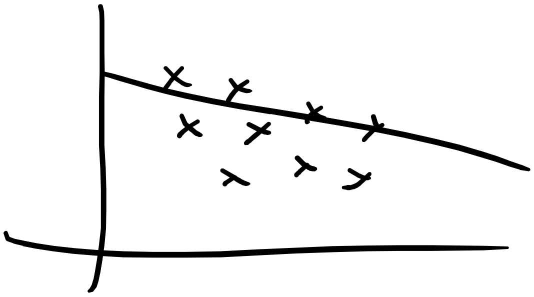


=> Karl Pearson's Correlation Coefficient

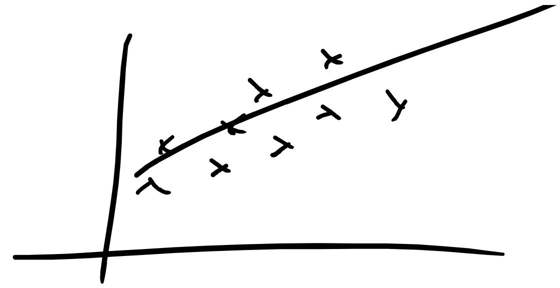
Pop = ρ Sample : r







$0 < \rho < 1$



$-1 > \rho > 0$

Pop ρ

$$\rho = \frac{N \sum xy - \sum x \cdot \sum y}{\sqrt{\left[N \sum x^2 - (\sum x)^2\right] \cdot \left[N \sum y^2 - (\sum y)^2\right]}}$$

$x = 1, 2, 3, 4, 5, 6$

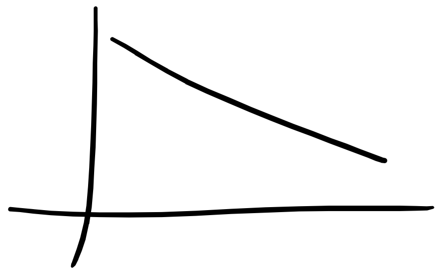
$y = 2, 4, 7, 9, 12, 14$

x	y	xy	x^2	y^2
21	48	211	91	490

$$\rho = \frac{6 \times 211 - 21 \times 48}{\sqrt{\left[6 \times (91) - (21)^2\right] \cdot \left[6 \times (490) - (48)^2\right]}}$$

$\rho = \underline{\underline{0.998}} \approx \underline{\underline{99.8\%}}$

Verdict: Data is



Verdict: Data is
positively correlated