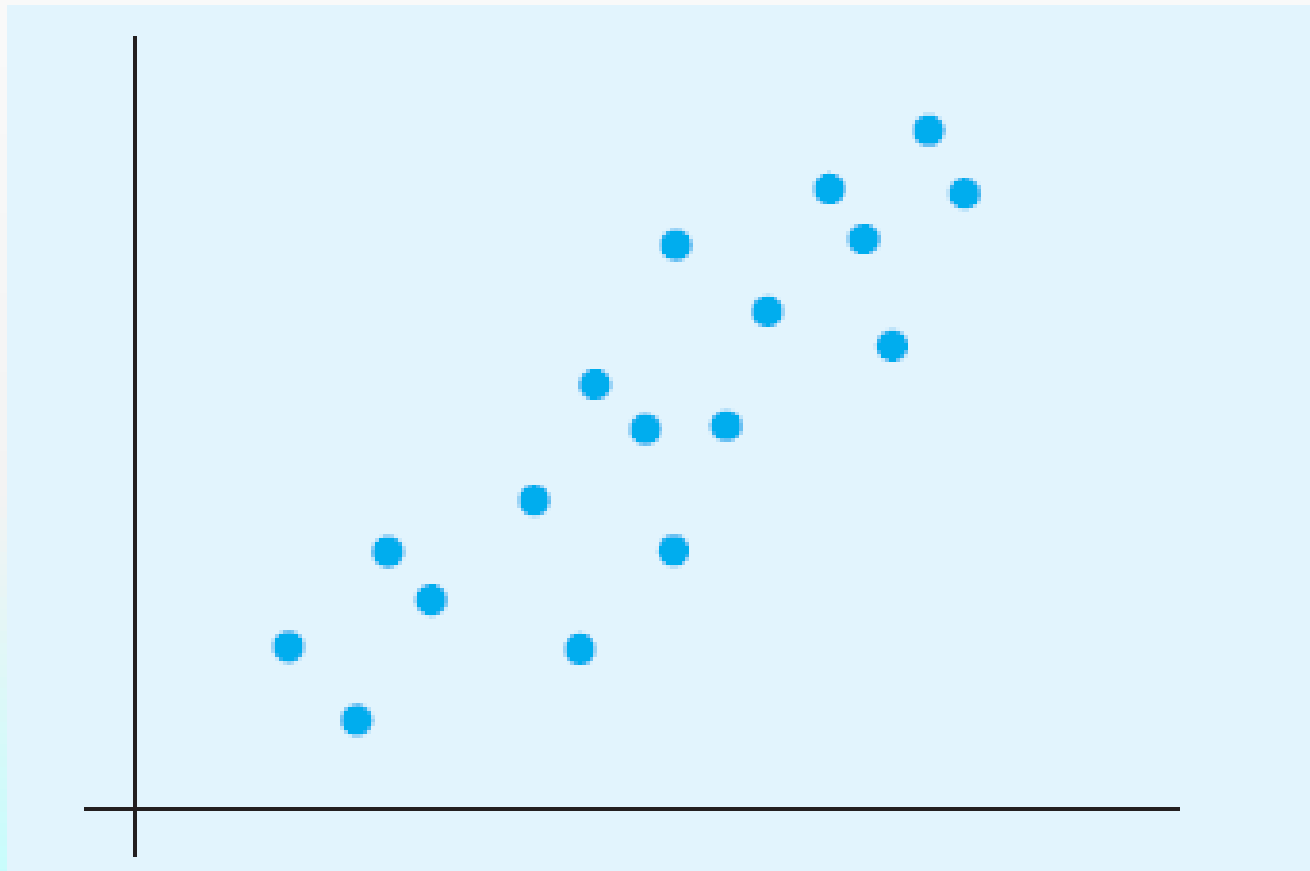


CORRELATION

Dr Parag Chavda
Asst Prof
Community Medicine

|| We are familiar with this diagram....



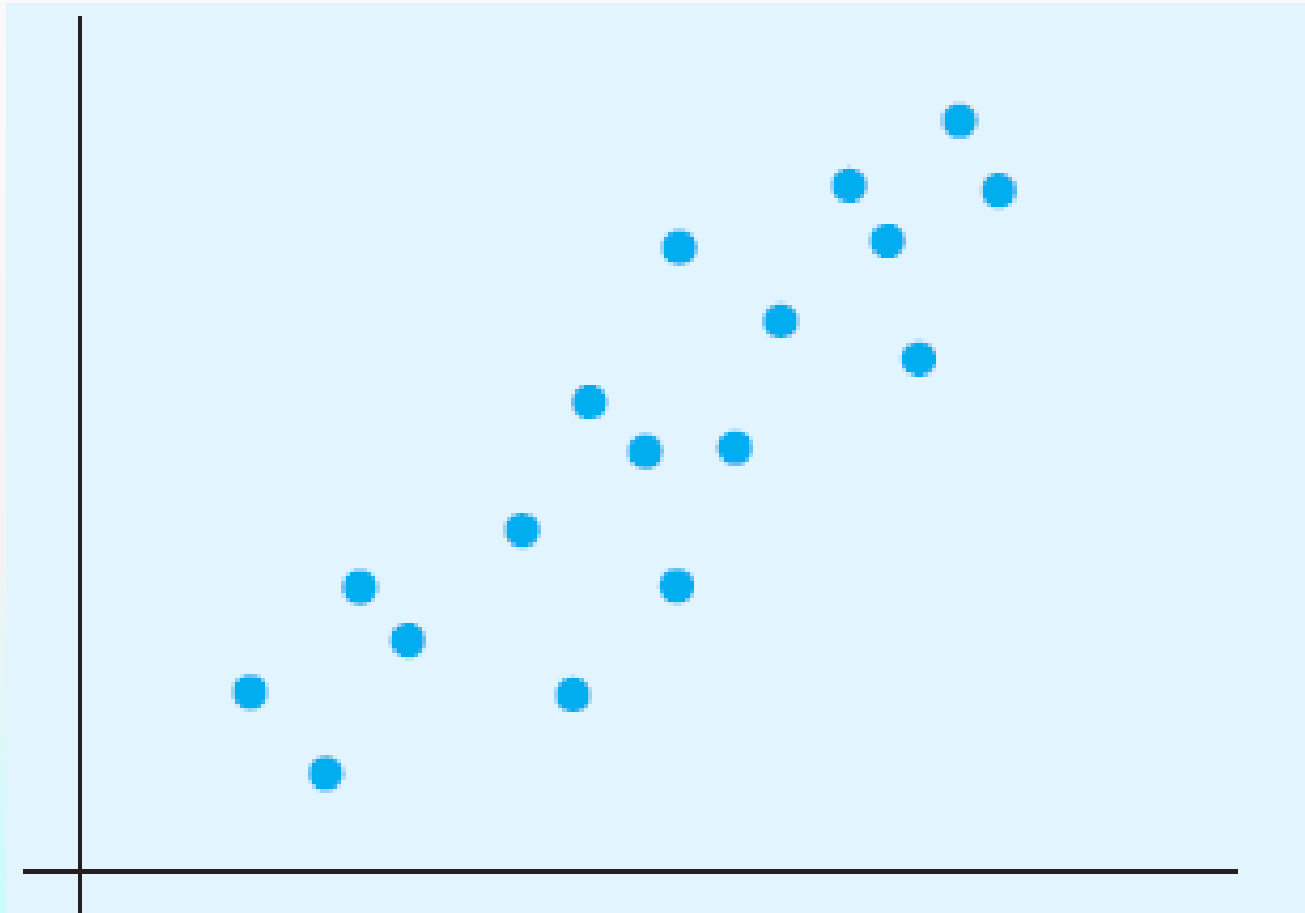
Correlation and Regression

- 'Correlation' indicates the relationship between two quantitative variables in which,
 - with changes in the values of one variable, the values in the other variable also changes.
- Correlation can either be positive or negative.

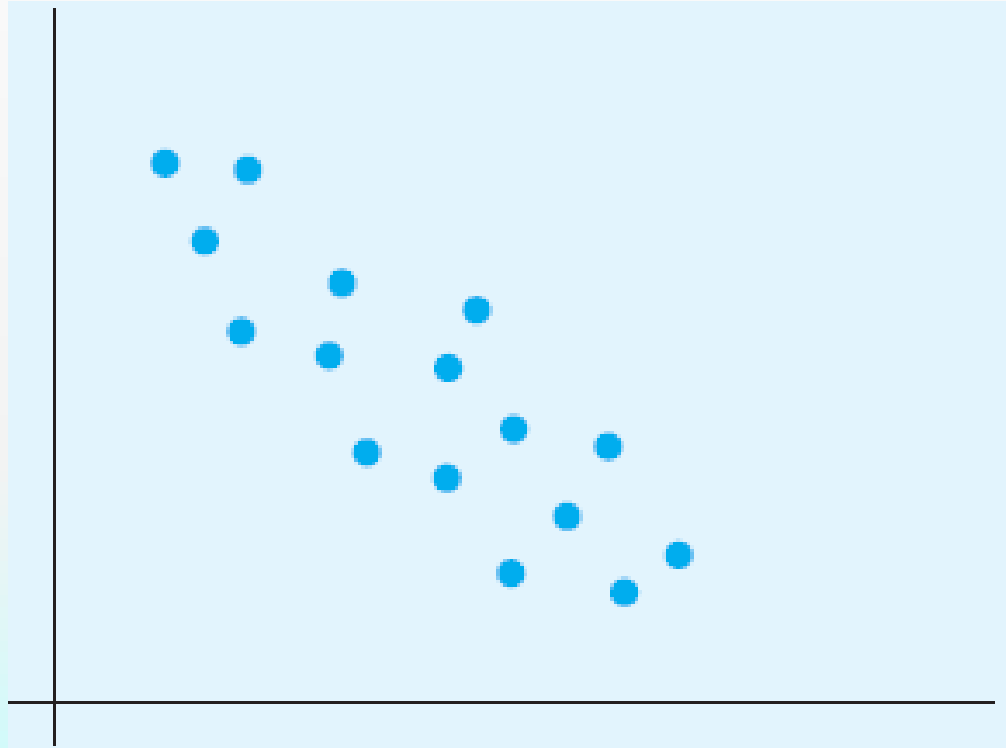
Correlation and Regression

- When the objective is to determine the **strength of relationship** between two such variables,
 - we use **correlation** coefficient (r).
- If the objective is to describe the existing relationship with a view of **prediction**,
 - we use **regression** analysis.

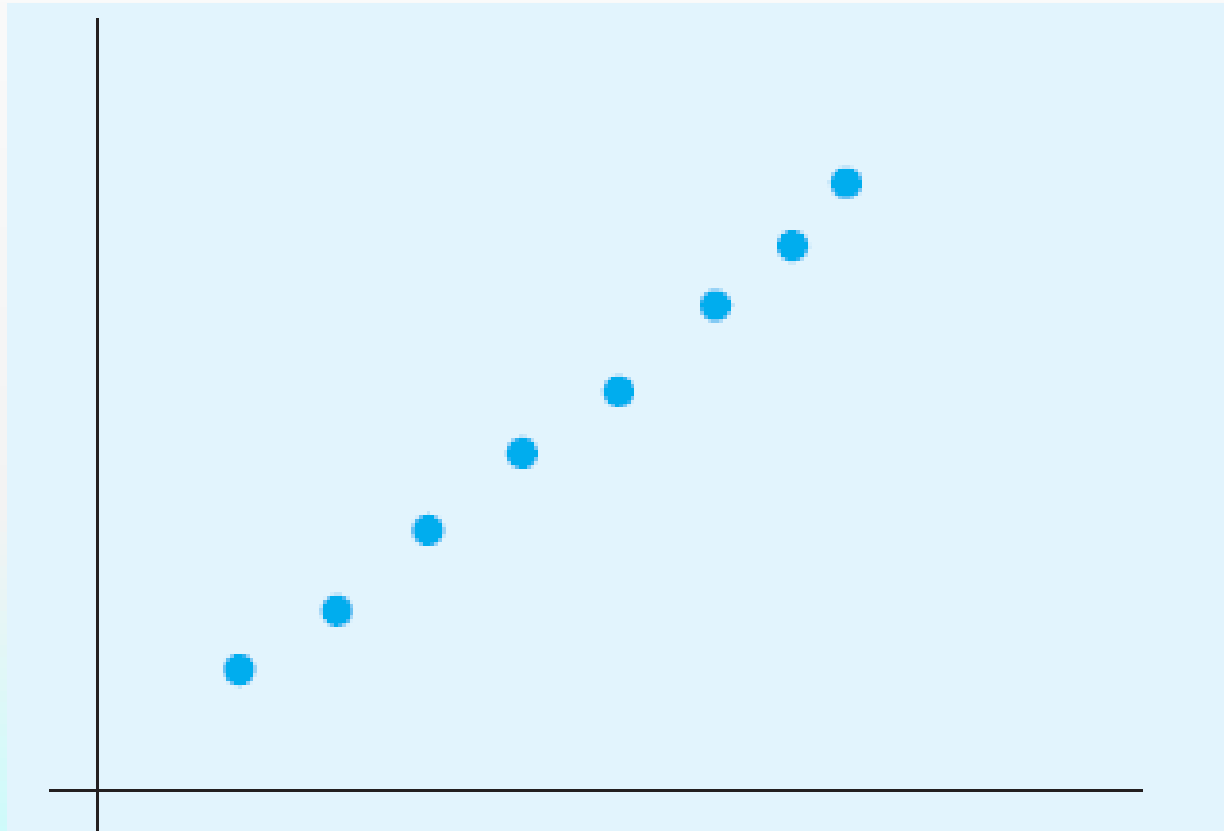
Positive correlation



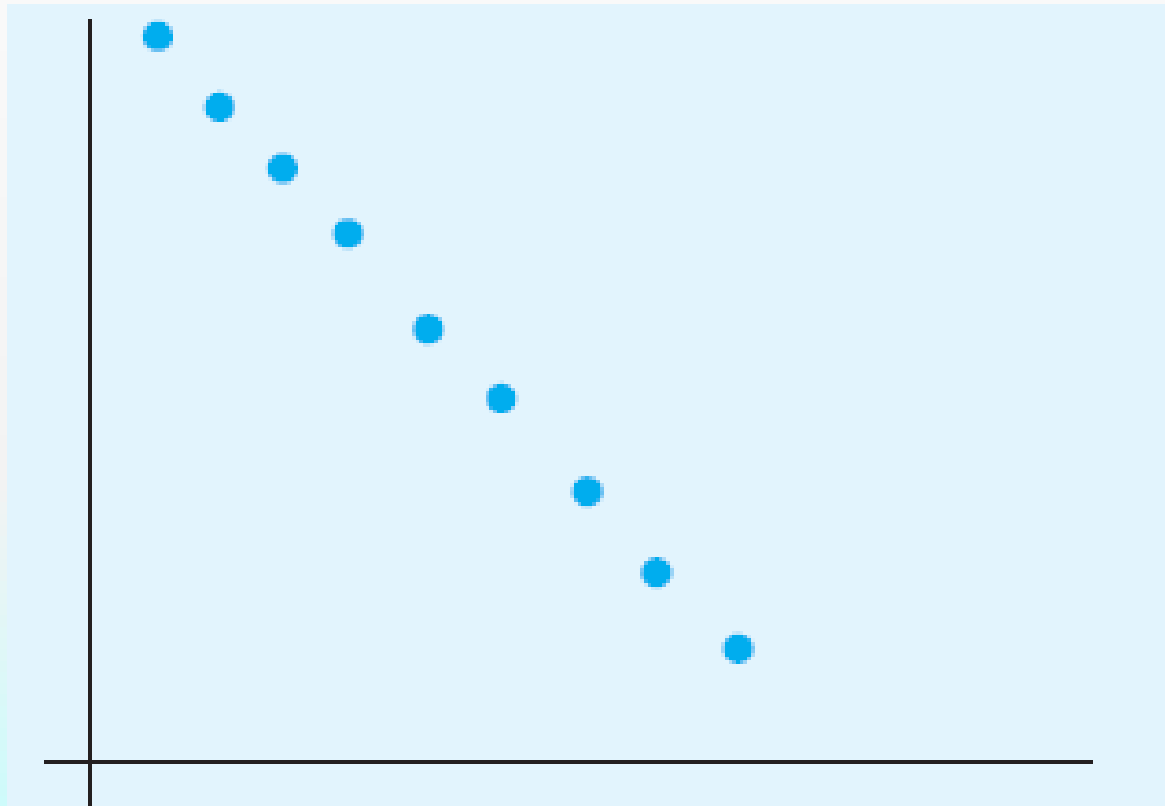
Negative correlation



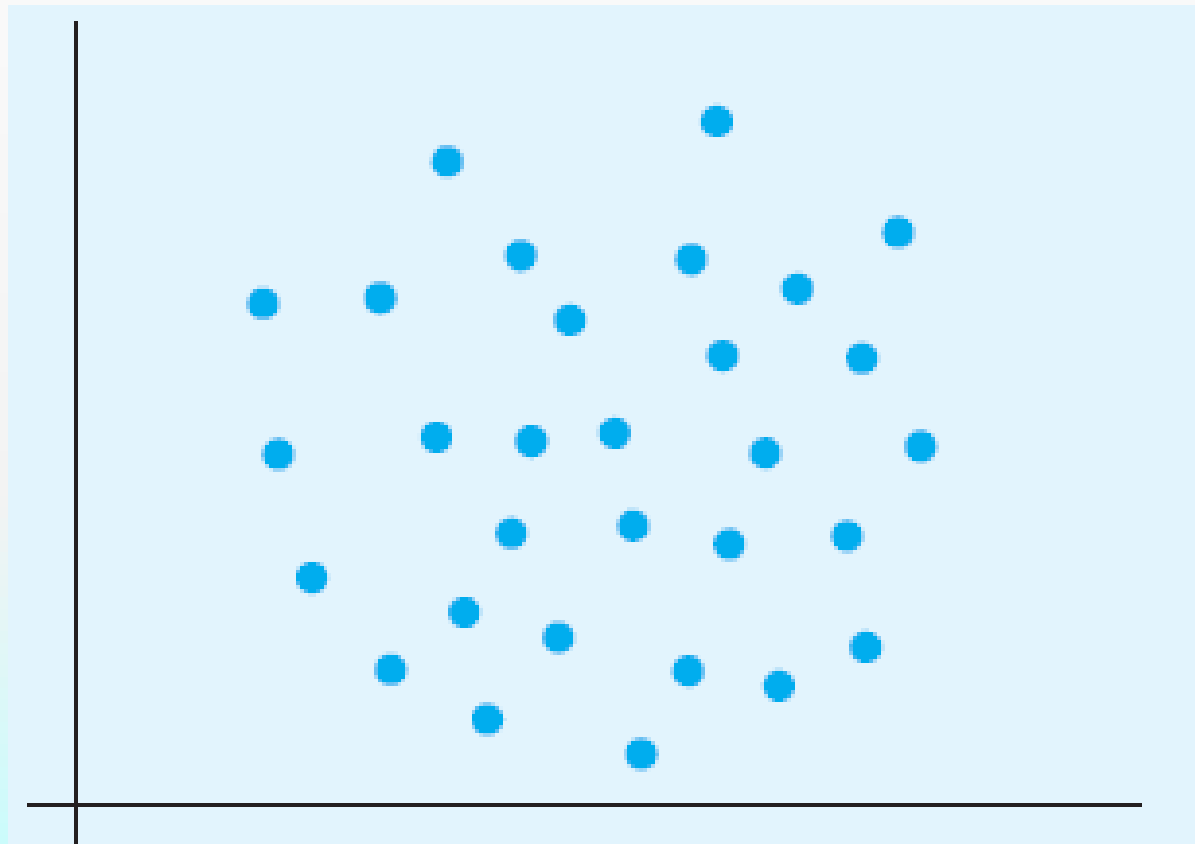
Perfect positive correlation



Perfect negative correlation



No correlation



Correlation Coefficient (r)

- Correlation coefficient measures the degree of correlation between the two variables.
- Range between -1 to +1
- $r = 0$ indicates no association
- $r = +1$ perfect positive association
- $r = -1$ perfect negative association


Types of Correlation Coefficient

- Karl Pearson's Product Moment Correlation Coefficient
 - used when dealing with continuous variables and joint distribution is following normal distribution
- Spearman correlation coefficient
 - Used when variables are not normally distributed

Karl Pearson's Correlation Coefficient^{nt}

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

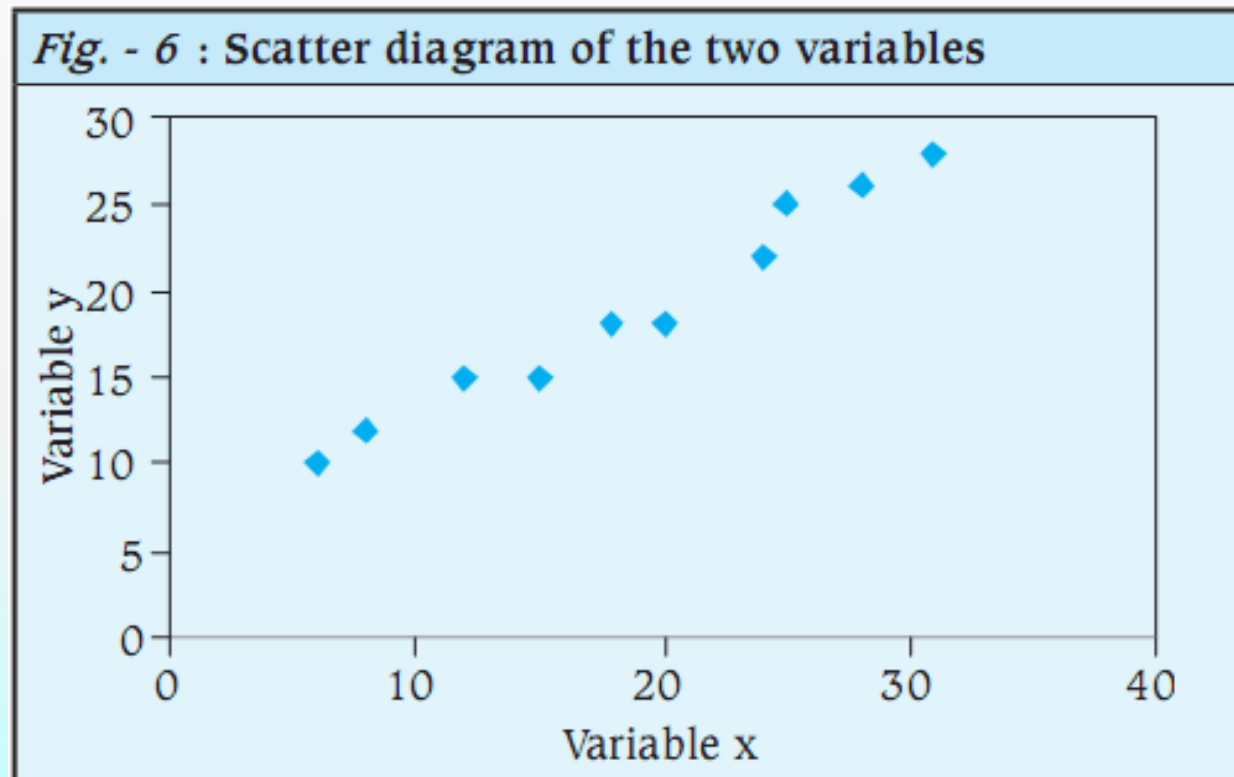
- Numerator - sum of cross products of deviations of both variables
- Denominator - square root of the product of the sum of squares of deviations for both variables.


$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)SD_{(x)}SD_{(y)}}$$

Find out the correlation coefficient of the following two variables

x	y
6	10
8	12
12	15
15	15
18	18
25	25
24	22
28	26
31	28
20	18

- $r = 0.98$




Interpretation of “r”

Value of r	Relationship
If $0 < r < \pm 0.25$ Between 0 and ± 0.25	Little or no linear relationship.
If $\pm 0.25 \leq r < \pm 0.50$ Between ± 0.25 and ± 0.5	Fair degree of linear relationship
If $\pm 0.50 \leq r < \pm 0.75$ Between ± 0.5 and ± 0.75	Moderate to good linear relationship.
If $r \geq \pm 0.75$ More than ± 0.75	Very good linear relationship.



Spearman Rank Correlation Coefficient


- Used when variables are not normally distributed but are ranked in order
 - Spearman correlation coefficient also ranges from -1 to +1
 - Interpreted in the same way as the 'Pearson correlation coefficient'.
- 

Coefficient of determination (r^2)

- It is square of correlation coefficient.
- Represents amount of variation in the dependent variable explained by the independent variable.
- Example:
 - “r” for age and blood pressure = 0.8
 - then coefficient of determination $r^2 = 0.64$
 - Interpretation - 64% of variability in blood pressure is explained by age whereas the remaining 36% is not by age
 - This 36% variability may be because of other factors such as weight, diet and exercise

Hypothesis Testing about Population Correlation Coefficient

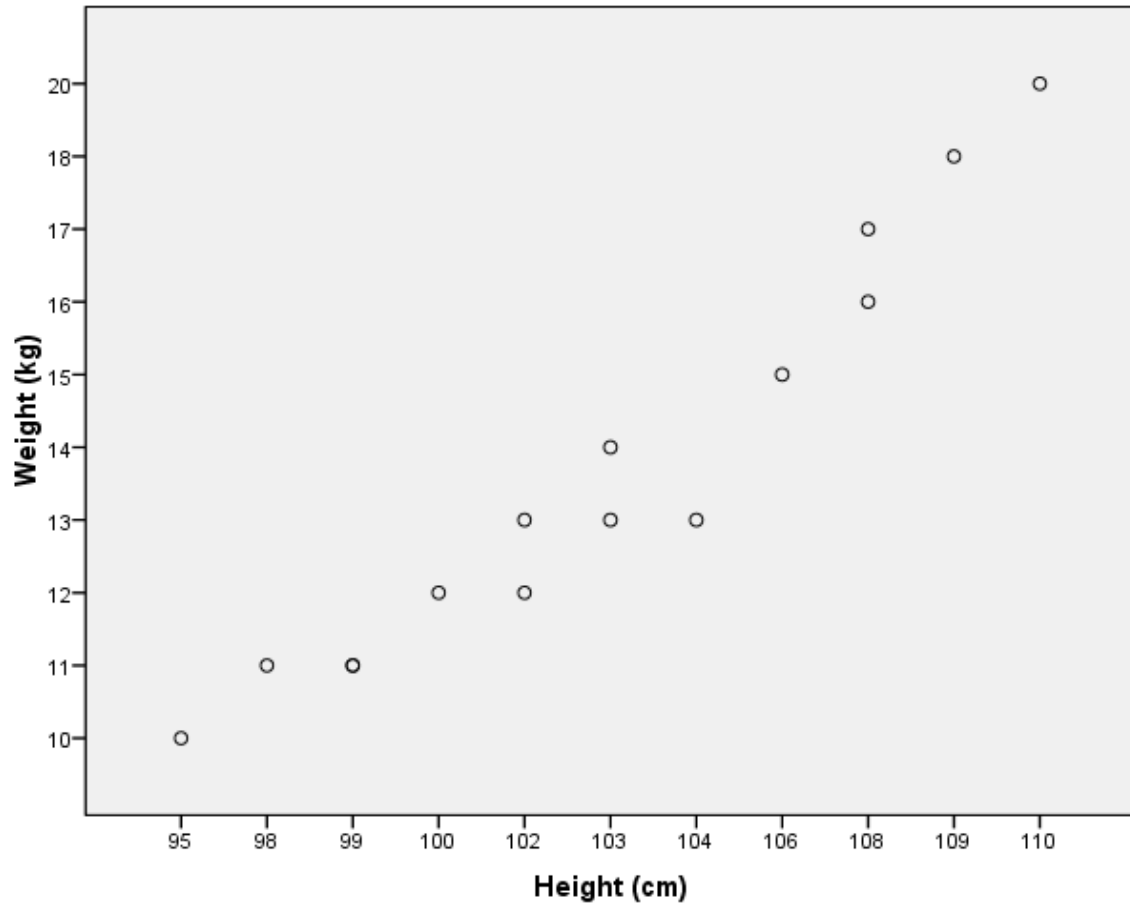
- To know that sample correlation coefficient is not due to chance
- The general steps in testing of hypothesis dealing with population correlation coefficient denoted by ρ are same as discussed in hypothesis testing dealing with mean, proportion or difference between means or proportions.


$$t = \frac{r \sqrt{(n-2)}}{\sqrt{(1-r^2)}}$$

- This “t” is interpreted same as student’s t test with (n-2) degrees of freedom

Exercise

- A researcher collected weights (kgs) and heights (cms) of the 15 children of second standard as 10, 20, 11, 12, 12, 13, 11, 14, 13, 13, 15, 11, 16, 17, 18. If the heights of the same students are 95, 110, 98, 100, 102, 102, 99, 103, 104, 103, 106, 99, 108, 108, 109.
- Plot the observations on a scatter plot. Calculate the correlation coefficient.





Further Reading

- Textbook of Public Health & Community Medicine, RajVir Bhalwar (Edi), Published by Department of Community Medicine, Armed Forces Medical College, Pune