In real life situations, especially in social sciences and in business, we often to know whether two or more variables are related, and if so, how they are related.
In this case, two questions may arise,

1. Is there any relationship between two or more variables?
2. If so, what is the relationship? Positive or negative or how strong their relationship?

To get answer of these questions, we need to know about the term "Correlation".
Now, what is correlation???

Correlation means "Relationship" or "Association" or "Connection".

Relationship what? Relationship between two or more variables Which gives us strength or degree and direction of association.

At a glance, "Relationship between two or more variables which gives us strength or degree and direction of association".

From this definition, it is clear that, there are two objectives of correlation. What are they?
The primary objective of correlation analysis is to measure,

1. Degree or strength of relationships
2. Direction of relationship

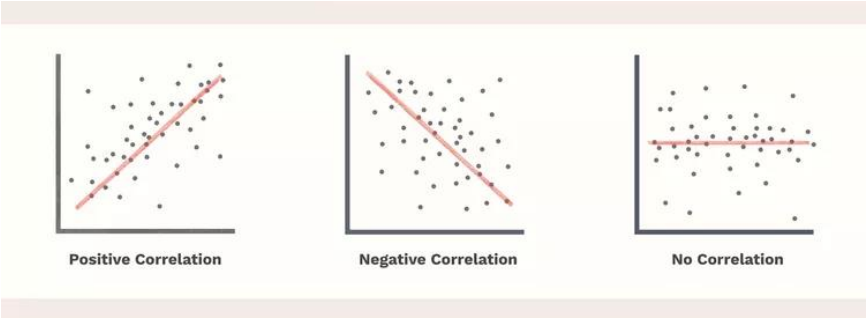
Types of Correlation:

We can classify correlation on the basis of three categories
We can classify correlation based on three categories:

***a) <u>On the basis of direction:</u>*** Two types-

i. **Positive correlation:** A positive correlation signifies a relationship between two variables where they move together, meaning they change in the same direction. This occurs when one variable decreases as the other decreases, or one variable increases as the other increases.

ii. **Negative correlation:** A negative correlation indicates a relationship between two variables where they move in opposite directions. In other words, when one variable increases, the other variable decreases, and vice versa.




***b) <u>On the basis of variables:</u>*** Three types-

i. **<u>Simple correlation:</u>** Simple correlation measures the strength and direction of the linear relationship between two variables. It quantifies how changes in one variable are associated with changes in another variable. For example: relationship between demand and supply.

ii. **<u>Multiple correlation:</u>** Multiple correlation measures the strength and direction of the linear relationship among more than two variables. For example, consider the relationship among rainfall, rice production, and rice prices.

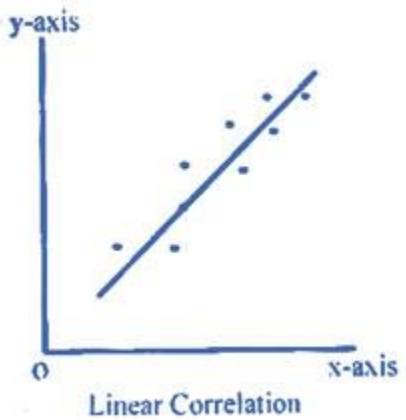
iii. **<u>Partial correlation:</u>** Partial correlation measures the strength and direction of the linear relationship among more than three variables where only two influencing variables are studied and

rest influencing variables are constant. For example: Consider the relationship among demand, supply, and income, assuming income to be constant.

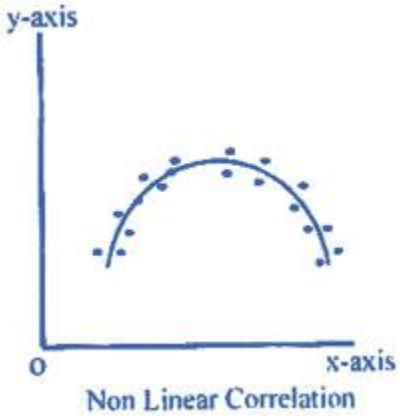
***c) <u>On the basis of linearity:</u>*** Two types-

i. **<u>Linear correlation:</u>** Linear correlation refers to the statistical relationship between two variables that can be represented by a straight line. When changes in one variable consistently correspond with changes in another variable, and these changes can be plotted as a straight line on a graph, it signifies a linear correlation. In other words, if one variable increase and the other also increases at a constant rate, they are linearly correlated.

ii. **<u>Nonlinear correlation:</u>** Nonlinear correlation, on the other hand, occurs when the relationship between two variables is not accurately represented by a straight line. In this case, changes in one variable do not result in proportional changes in the other variable. Instead, the relationship may follow a curve or another non-straight pattern on a graph, indicating a nonlinear correlation. Nonlinear correlations can take various forms, such as quadratic, exponential, or logarithmic relationships, among others. These relationships cannot be adequately described using a single straight line.



Linear Correlation



Non Linear Correlation

Now the question is “Sir, how can we measure the correlation?” or “How can we understand is the variables are related or not”

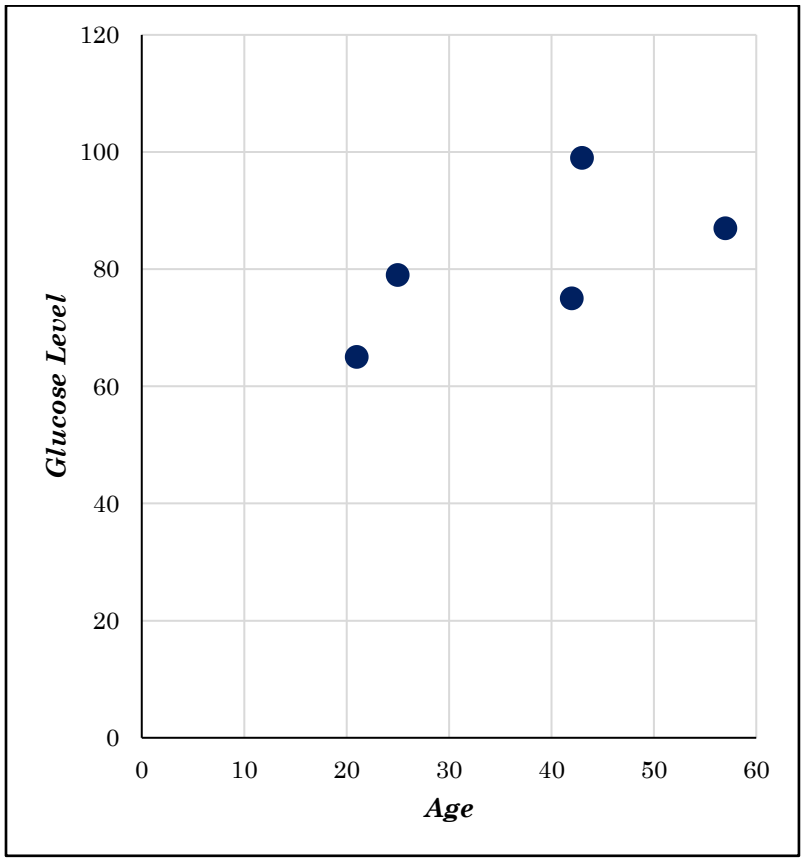
We can measure the correlation in four ways

1. Scatter diagram, which is a graphical method.
2. Karl Pearson’s correlation coefficient
3. Spearman rank correlation coefficient
4. Method of least square

But we just focus on Scatter diagram and Karl Pearson’s correlation coefficient

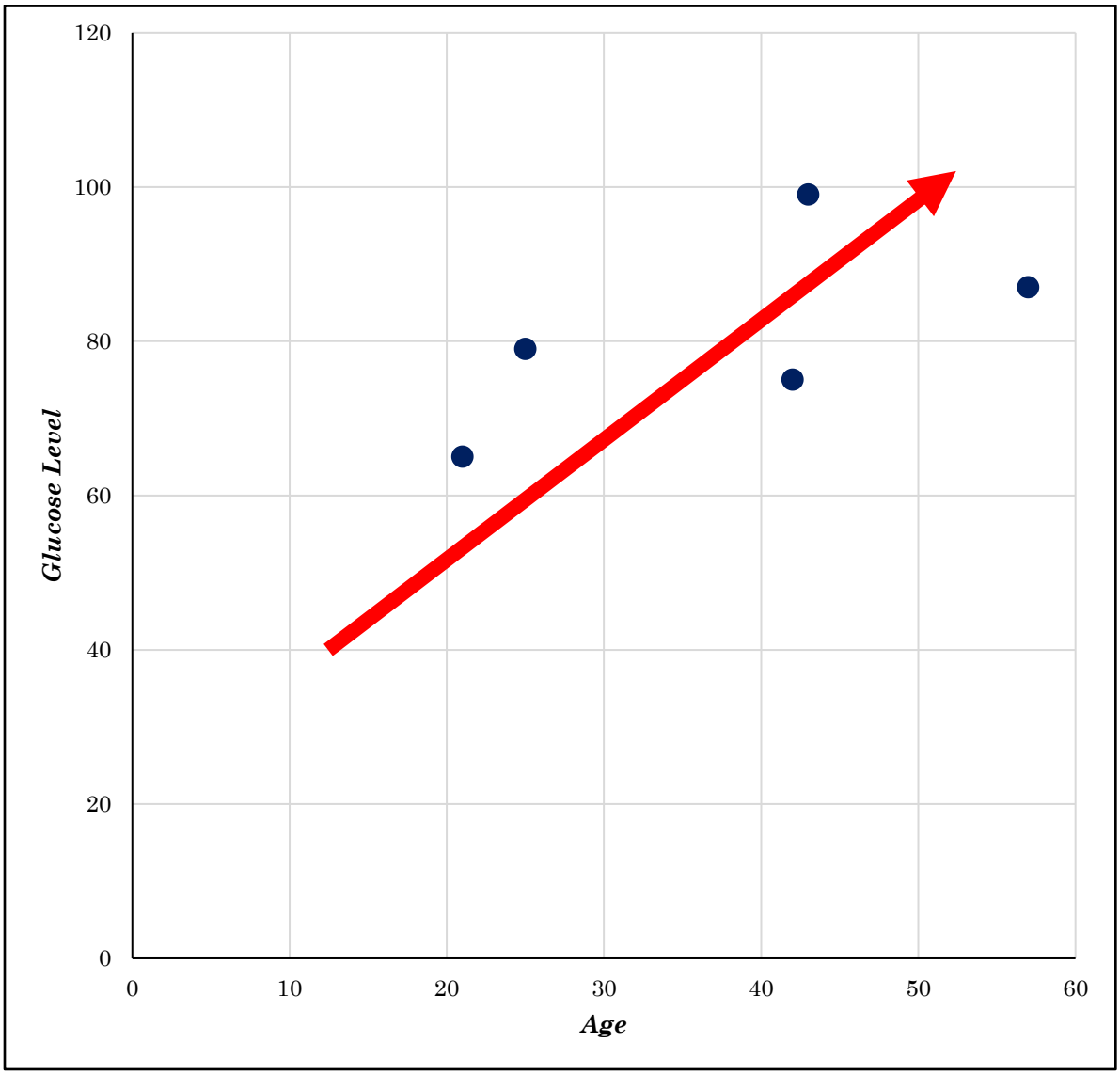
Scatter Diagram:

| Age | 43 | 21 | 25 | 42 | 57 |
|---|---|---|---|---|---|
| Glucose Level | 99 | 65 | 79 | 75 | 87 |



With the increase of age, the Glucose Level also increased. Thus, there is a positive correlation between “Age” and “Glucose Level”.

Limitations: Accurate degree and strength of correlation cannot be obtained by scatter diagram.




In this case, Karl Pearson’s Correlation Coefficient is more suitable than graphical method.

Karl Pearson's Correlation:

1. Both variables are measured in interval or ratio scales
2. Relationship between variables is linear
3. Denoted by $\boldsymbol{r}$

Formula:

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

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}\,\sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

Range of $r$ is $[-1\ to + 1]$

If, $r = +1$, Perfect positive correlation

If, $r = -1$, Perfect negative correlation

If, $r = 0$, No correlation

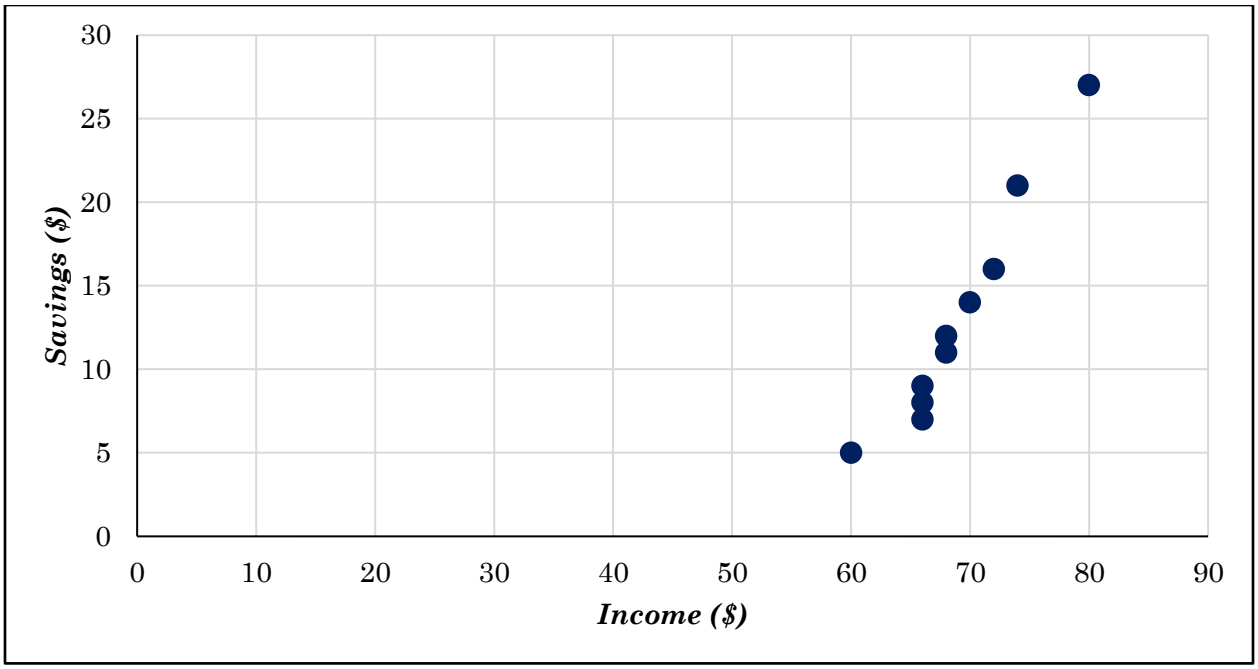
| **Correlation** | **Negative** | **Positive** |
|---|---|---|
| Weak | -0.29 to -0.10 | 0.10 to 0.29 |
| Medium | -0.49 to -0.30 | 0.30 to 0.49 |
| Moderate | -0.50 to -0.79 | 0.50 to 0.79 |
| Strong | -1.00 to -0.80 | 0.80 to 1.00 |

| x (income) | y (Savings) | $x_i^2$ | $y_i^2$ | $x_i \times y_i$ |
|---|---|---|---|---|
| 60 | 5 | 3600 | 25 | 300 |
| 66 | 7 | 4356 | 49 | 462 |
| 66 | 8 | 4356 | 64 | 528 |
| 66 | 9 | 4356 | 81 | 594 |
| 68 | 11 | 4624 | 121 | 748 |
| 68 | 12 | 4624 | 144 | 816 |
| 70 | 14 | 4900 | 196 | 980 |
| 72 | 16 | 5184 | 256 | 1152 |
| 74 | 21 | 5476 | 441 | 1554 |
| 80 | 27 | 6400 | 729 | 2160 |
| $\sum x_i = 690$ | $\sum y_i = 130$ | $\sum x_i^2 = 47876$ | $\sum y_i^2 = 2106$ | $\sum x_i y_i = 9294$ |

Now, the correlation coefficient

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} = \frac{9294 - \frac{690 \times 130}{10}}{\sqrt{47876 - \frac{(690)^2}{10}} \sqrt{2106 - \frac{(130)^2}{10}}} = 0.97$$

The values of $r = 0.97$, suggests a strong positive correlation between income and savings of garments workers. That is, as income increases, there is a strong tendency for saving increase.




Properties:

1. Correlation coefficient has no unit.
2. The sign of correlation coefficient gives the direction of the association.
3. The correlation coefficient is between $-1$ and $+1$.
4. Correlation coefficient is a symmetric measure, $i.e.$, $r_{xy} = r_{yx}$
5. Correlation is sensitive to extreme observations.