**Correlation**

* Correlation

*Question: What is correlation? Why is it required?*

Examples of Some Relationships or Phenomenon

* As summer heat rises, hills stations are crowded with more and more visitors
* Icre-cream sales are higher during summer and lower during winter
* Production of tomato increases, retail price falls

**So, what is Correlation analysis?**

Correlation analysis is a means for examining relationship between two variables systematically.

What does correlation analysis deal with?

* Is there any relationship between two variables?
* If the value of one variable changes, does the value of the other also change?
* Do both the variables move in the same direction?
* How strong is the relationship?

**Spurious Correlation**

*Question: What is spurious correlation?*

Where the relationship between variables is merely coincidence, any correlation found is spurious.

* Arrival of birds in a sanctuary and the birth rates in the locality.
* Size of your shoe and money in your pocket

**Confounding Variable**

Where a third variable is affecting the relationship between 2 variables, the third variable is the confounding variable.

*Question: Can you think of an example?*

* Brisk sale of ice-creams and high number of deaths due to drowning

*Question: What is the confounding variable here?*

* High temperature is leading more people to use swimming pools, oceans etc causing higher incidence of drowning
* Temperature is the confounding variable

**What does Correlation Measure?**

* Correlation measures co-variation
* It should never be interpreted as implying cause and effect relation
* The presence of correlation between two variables X and Y simply means that when the value of one variable is found to change in one direction, the value of the other variable is found to change either in the same direction or in the opposite direction.
* When it is found to change in the same direction, the correlation is called positive correlation.
* When it is found to change in the opposite direction, the correlation is called negative correlation.

**Types of Correlation**

**Positive Correlation**

* Correlation is said to be positive when the variables move together in the same direction

*Question: Can you give examples of positive correlation?*

* + Example: score in exams and no. of hours studied
  + when income increases, consumption increases as well

**Negative Correlation**

*Question: Can you give examples of negative correlation?*

* Correlation is said to be negative when the variables move in opposite directions
  + Prices of apples fall, demand increases
  + Absenteeism and marks scored in exam
  + Speed and travel time

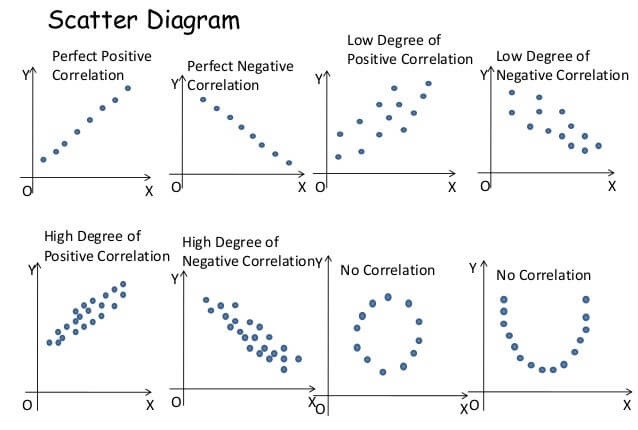
**Techniques for Measuring Correlation**

* Scatter diagram
* Karl Pearson’s coefficient of correlation
* And one Spearman’s rank correlation

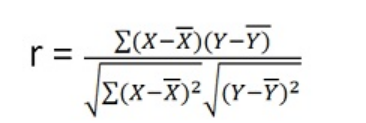
**Scatter Diagram**

*Question: What is a scatter plot?*

* Visual examination of relationship between variables without measurement
* The values of the two variables are plotted as points on a graph
* If all the points lie on a line the correlation is said to be ‘perfect’
* If the scatter points are widely dispersed around, the correlation is low.



**Karl Pearson’s Correlation Coefficient**



**Example**

|  |  |
| --- | --- |
| **No. of years of schooling of farmer** | **Annual Yield per acre** |
| **0** | **4** |
| **2** | **4** |
| **4** | **6** |
| **6** | **10** |
| **8** | **10** |
| **10** | **8** |
| **12** | **7** |

**R = 0.644**

**Properties of Correlation Coefficient**

1. R has no unit. It is a pure number
2. A negative value of r indicates an inverse relation
3. A positive r indicates the two variables move in the same direction.
4. R lies between -1 and 1.
5. If r = 0 the two variables are uncorrelated.
6. R=1 correlation is perfectly positive
7. R=-1 correlation is perfectly negative.
8. A high value of r indicates strong linear relationship
9. A low value of r indicates a weak linear relation.

**Spearman’s Rank Correlation**

1. Suppose we estimate the correlation between the heights and weights of students in a remote village where neither measuring rods nor weighting machines are available.
2. In such a situation, we cannot measure height or weight but we can rank the students according to weight and height.
3. These ranks can then be used to calculate Spearman’s rank correlation coefficient.