**Correlation**

A correlation between two variables indicates that there exists a relationship between them.

Ex:

* A persons weight and height
* Number of years of school and future income

We can illustrate a two variable distribution on the Cartesian plane by plotting data points (x and y coordinates). This is called a **Scatter Plot**.

Ex: Age and Diameter of Trees

Diameter (cm)

Age of Tree

A correlation can be positive or negative

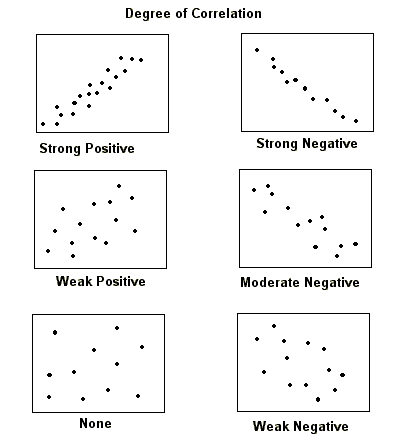
* Positive : when both x and y increase

Ex: tree age increases, so does its diameter

* Negative : when x increases and y decreases

Ex: your golf score decreases as years of experience increases

Correlations are also characterized by their strength



The strength is determined by how closely the scatter plot forms a line.

**Correlation Coefficient**

To measure the strength of a correlation we need to determine a correlation coefficient (r)

**Step 1** – Draw a rectangle around the points **(ignoring the outliers)**

**An OUTLIER is a point that indicates an abnormal piece of data or something out of the ordinary.**

**It is located far from main cloud of points in the scatter plot.**

Diameter (cm)

Age of Tree

**Step 2** – Measure the long side (L) and the short side (l)

l = 3 cm

L = 10 cm

Diameter (cm)

Age of Tree

**Step 3** – Apply correlation formula



r ≈ ( 1 – ) )



r ≈ (1 – )



r ≈ 0.7 🡪 this indicates a moderate positive correlation

**Correlation Strength**

|  |  |  |
| --- | --- | --- |
| Correlation Coefficient | | Meaning |
| Positive | Negative |
| Near 0 | Near 0 | Zero |
| Near 0.5 | Near -0.5 | Weak |
| Near 0.75 | Near -0.75 | Moderate |
| Near 0.87 | Near -0.87 | Strong |
| Near 1 | Near -1 | Perfect |

-1

1

0.87

0.75

0.5

0

-0.5

-0.75

-0.87

Negative Correlation

Positive Correlation