

Data Terms

Hypothesis

An educated guess before you perform an experiment or survey

Primary Data

Data gathered by the person who does the data analysis

Secondary Data

Data that has been gathered by some other source that the person analysing must reference in their report

Independent Variable

Graphed on the x-axis
The variable that is controlled, usually time

Dependent Variable

Graphed on the y-axis
The variable that is measured or observed

Linear Relationship

The points on the scatter plot appear to follow a straight line

Non-Linear Relationship

The points on the scatter plot appear to follow a curve or have no pattern at all.

Trends

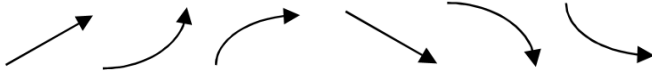
"As $\frac{\quad}{X}$ increases, the $\frac{\quad}{Y}$ increases/decreases"

Increasing

The graph is moving upward toward the right

Decreasing

The graph is moving downward toward the right



Line of Best Fit (Regression Equation)

- Follows the trend
- Is in the middle of the data such that you have the same # of points above and below the line
- Extends from one end of the grid to the other
- Is a straight line – use a ruler!
- May ignore the obvious outlier

Outlier

A point that does not fit with the rest of the data

Correlation

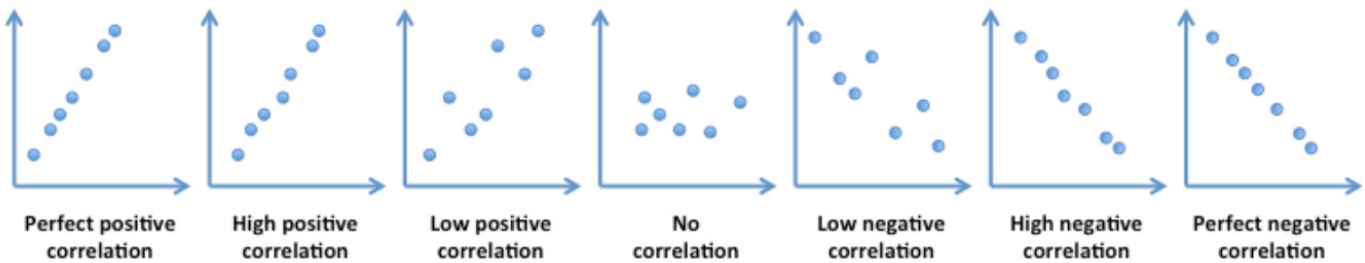
Correlation coefficient, r , specifies how good the line/curve of best fit suits the data

Interpolation

Predicting a value *inside* of your data set

Extrapolation

Predicting a value *outside* of your data set



Correlation coefficient

1 0.8 0.3 0 -0.3 -0.8 -1

Choosing a Scale

- If all numbers are far from zero, create a break in the graph
- Count the number of available squares (after you create a break, if applicable)
- Find the Range = Max – Min
- Divide

$$\frac{\text{Range}}{\text{Squares}} \uparrow \text{round up}$$

$$\frac{\text{Squares}}{\text{Range}} \downarrow \text{round down}$$