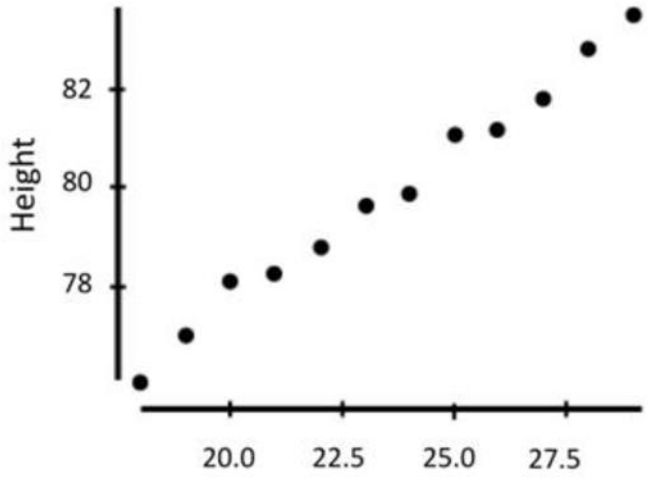
Correlation and Regression

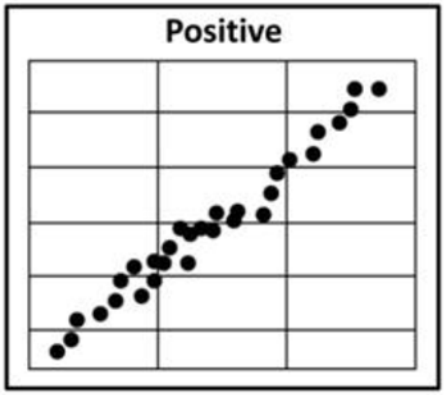
- Correlation analysis is used for investigating the relationship between two quantitative variables
- Goals of correlation analysis :
  - Analyze if two measurement variables have a relation. This means change in one influences change in the other measure
  - Quantify the strength of the relationship between the variables

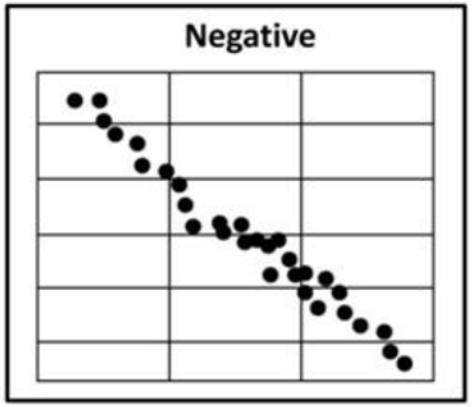
Correlation Examples

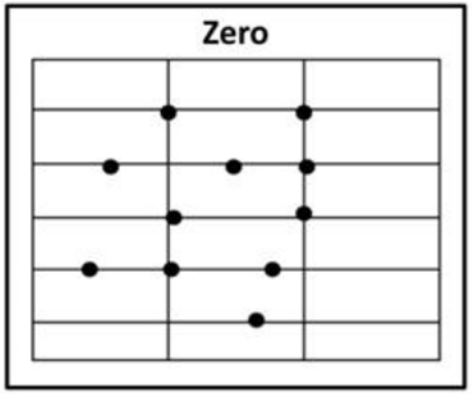
A positive correlation between height of a child and age: As the child grows his or her height increases almost linearly.



Correlation Coefficient (Cont'd)







Correlation and Regression Examples

Analysis of Student Grades in Mathematics and English

- Use Correlation to determine if the students who are good at Mathematics tend to be equally good at English
- Use Regression to determine whether the marks in English can be predicted for given marks in Mathematics

Correlation: Example

Correlation does not imply Causation.

- Myth: India team loses a match if Sachin Tendulkar hits a century
- Correlation: Sachin hits a century and India wins a match
- Does it imply causation? Does Sachin hitting a century causes India to lose the match?

Correlation Coefficient

- Correlation Coefficient (also called Pearson Correlation Coefficient) is a measure of strength and direction of a linear relation between two variables
- Correlation Coefficient r or R is defined as covariance of variables divided by product of Standard Deviations of the variables

$$r = \frac{\sum_{i=1}^{n}\left((x_i - \bar{x})(y_i - \bar{y})\right)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Correlation Coefficient (Cont'd)

The Correlation Coefficient ranges from -1 to 1.

- +1 indicates perfect collinearity, which means, if one value increases, the other also increases in the same proportion
- -1 indicates perfect negative collinearity, which means, if one value decreases, the other increases in the same proportion
- Zero indicates no relationship between the variables