Covariance and Correlation

Are the two variables independent?

Prof Hans Georg Schaathun

Høgskolen i Ålesund

12th February 2014

Question

Are taller people also heavier?

- Two stochastic variables:

  height H
  weight W

- We can observe them in pairs (H_i, W_i)
  - each individual i has a height and a weight

Question

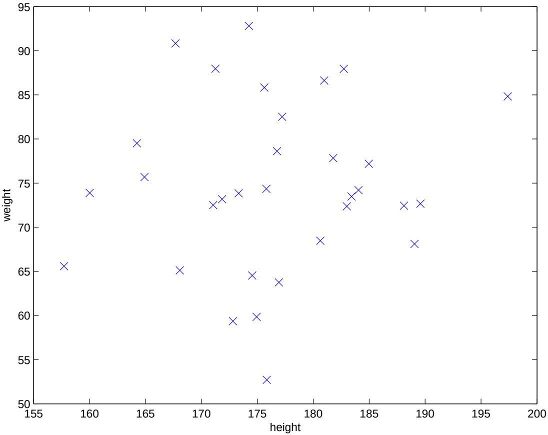
Are taller people also heavier?

- Two stochastic variables:

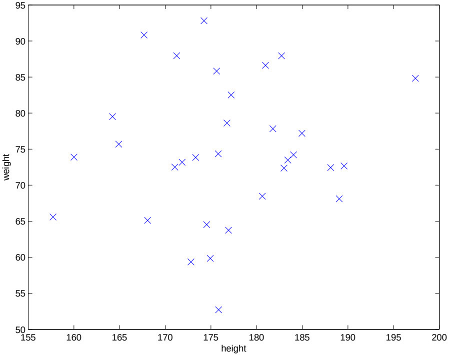
  height H
  weight W

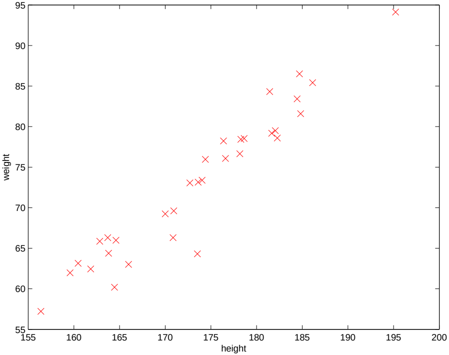
- We can observe them in pairs (H_i, W_i)
  - each individual i has a height and a weight

Scatter plot



What do you expect?





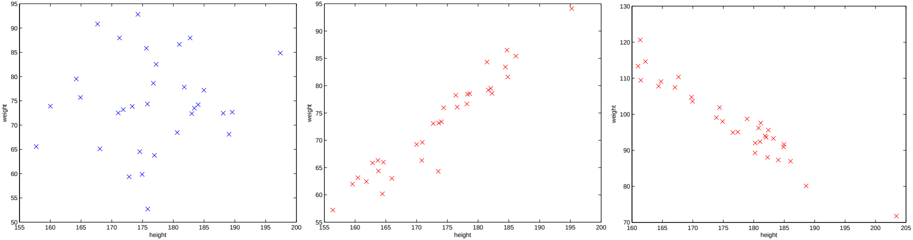
Measuring of Correlation

- Recall the variance $\mathrm{var}(X) = E\big((X - \mu_X)^2\big)$
  - measures the spread of one variable
- Covariance $\mathrm{Cov}(X, Y) = E\big((X - \mu_X)(Y - \mu_Y)\big)$
  - not normalised
  - if σ_X or σ_Y is large, the covariance will have high absolute value
- Normalise $\to$ Correlation Coefficient

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

HØGSKOLEN I ÅLESUND
Aalesund University College

Interpreting the Covariance



HØGSKOLEN I ÅLESUND
Aalesund University College

Sample Correlation Coefficient

1. Population correlation coefficent

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

2. Sample correlation coefficent

$$\hat{\rho} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})/(n-1)}{s_X s_Y}$$

$$\hat{\rho} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]\left[\sum_{i=1}^{n} (Y_i - \bar{Y})^2\right]}}$$

HØGSKOLEN I ÅLESUND
Aalesund University College

Summary

$$\rho = \frac{E\big((X-\mu_X)(Y-\mu_Y)\big)}{\sigma_X\sigma_Y}$$
$$\hat{\rho} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\big[\sum_{i=1}^{n}(X_i-\bar{X})^2\big]\big[\sum_{i=1}^{n}(Y_i-\bar{Y})^2\big]}}$$

- Correlation is dependence between two variables
- The Correlation Coefficient ρ is the principal measure
  - $\rho = 0 \Rightarrow$ independent variables
  - $\rho > 0 \Rightarrow X$ tends to be large when Y is
  - $\rho < 0 \Rightarrow X$ tends to be large when Y is small