Lecture 9
sample size

Prior information

- Some prior information is necessary for a sample size calculation to be possible.
  - Clinically important difference, or an expected difference between groups
  - Estimate of variability (continuous response) or control group ‘success’ proportion (binary response)

Sample size for mean

Standard error of mean = $\sigma / \sqrt{n}$

95% confidence interval = $\bar{x} \pm 1.96\,\sigma / \sqrt{n}$

If we know a value for σ (the standard deviation) and the desired confidence interval width, we can obtain *n*, the number of observations required.

Example

95% confidence interval width = $1.96\sigma / \sqrt{n}$

Standard deviation: 48 pcs./min

Desired confidence interval width: ± 20 litres/min

CI width = $20 = 1.96 \times 48 / \sqrt{n}$ and solving for *n* gives

$$n = (1.96 \times 48 / 20)^2 \approx 22$$

i.e. a sample of 22 would enable us to estimate the population mean to within 20 pcs./min (with 95% CI).

Sample size for proportion

- Problem: the standard error of a proportion depends on the proportion itself, the quantity we are trying to estimate!
  - We need an initial estimate

$$95\% \text{ confidence interval} = p \pm 1.96\sqrt{p(1-p)/n}$$

$$90\% \text{ confidence interval} = p \pm 1.645\sqrt{p(1-p)/n}$$

Example

- Customer fail to service about 3%, and we want the 95% confidence interval to be 0.5% on either side

$$\text{CI width} = \ 0.005 = 1.96\sqrt{0.03(1-0.03)/n}$$

$$n = 1.96^2 \times 0.03(1-0.03)/0.005^2 = 4472$$

Correlation Analysis Basic

It used to find relation between 2 variables.

Quantitative variable only.

Such:

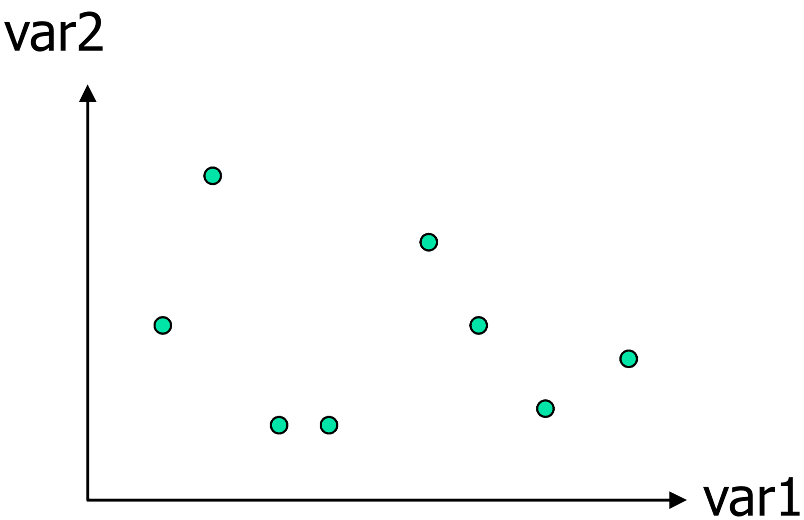
Ingredient VS Property.

Number of operator VS Output.

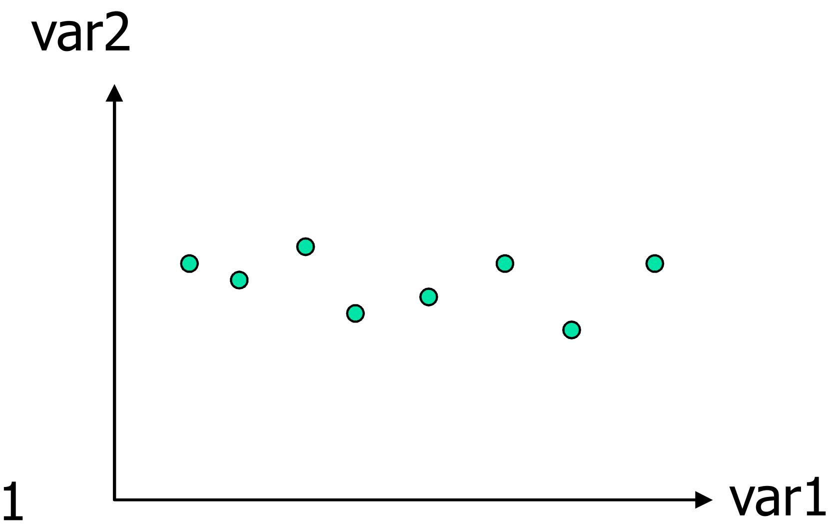
Speed of machine VS %defect.

Etc.

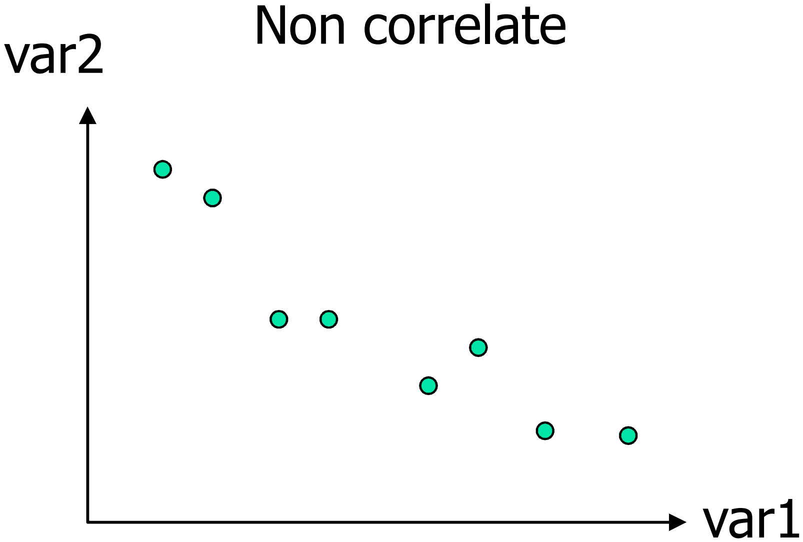
Graphical description of correlation



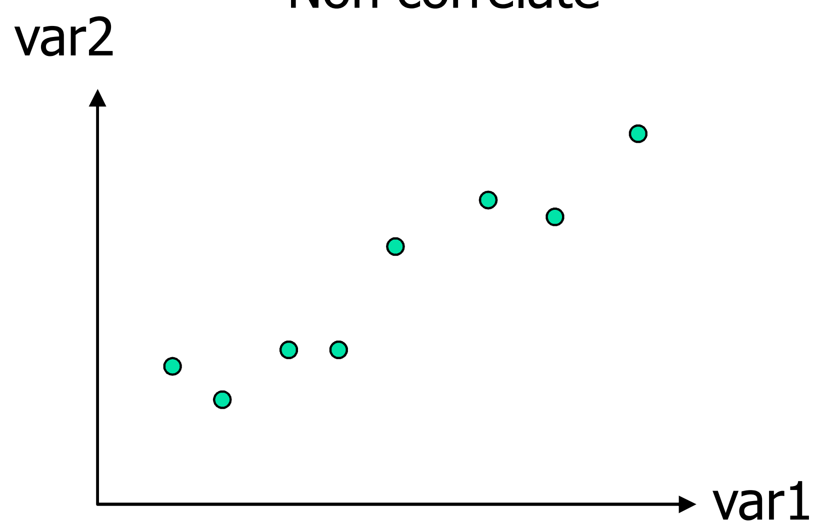
Non correlate



Non correlate



Negative correlate



Positive correlate