

Introduction to linear models

Modern statistics are easier than this

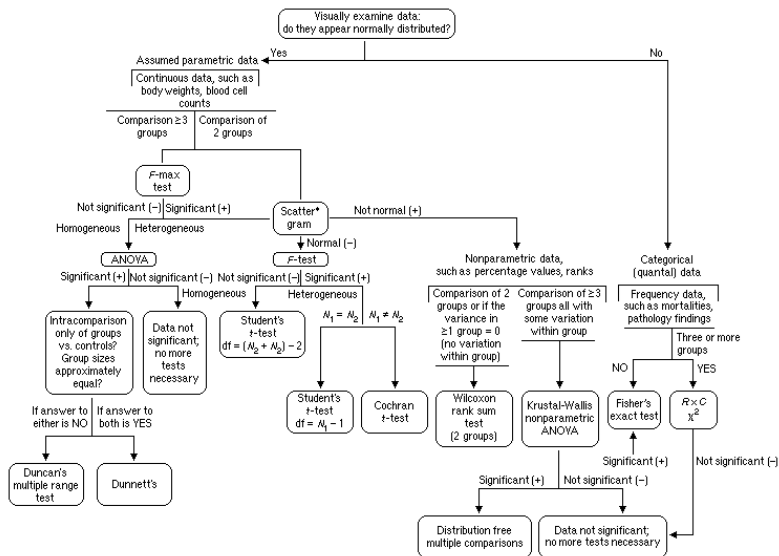
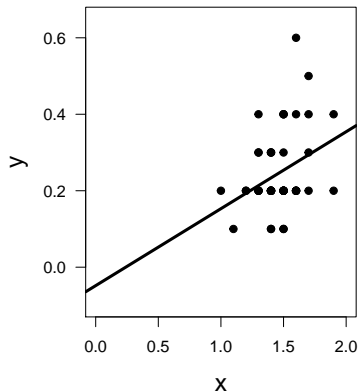


Figure 1:

Our overarching regression framework

$$y_i = a + bx_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$



Data

y = response variable

x = predictor

Parameters

a = intercept

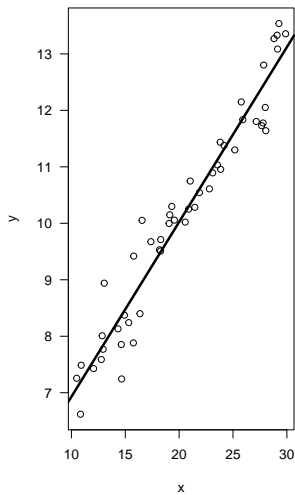
b = slope

σ = residual variation

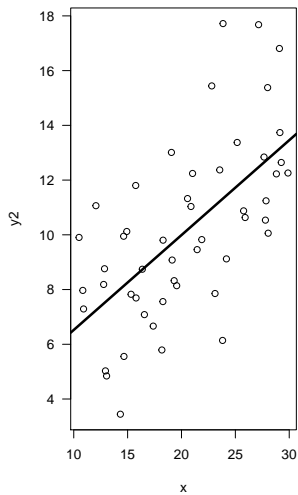
ε = residuals

Residual variation (error)

small



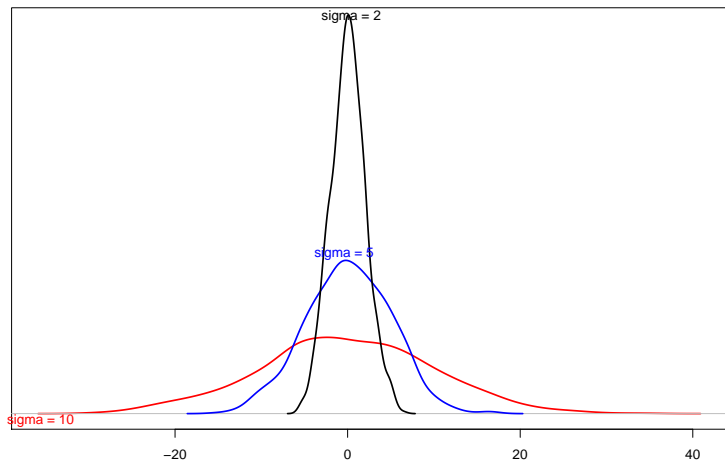
large



Residual variation

$$\varepsilon_i \sim N(0, \sigma^2)$$

Distribution of residuals



In a Normal distribution

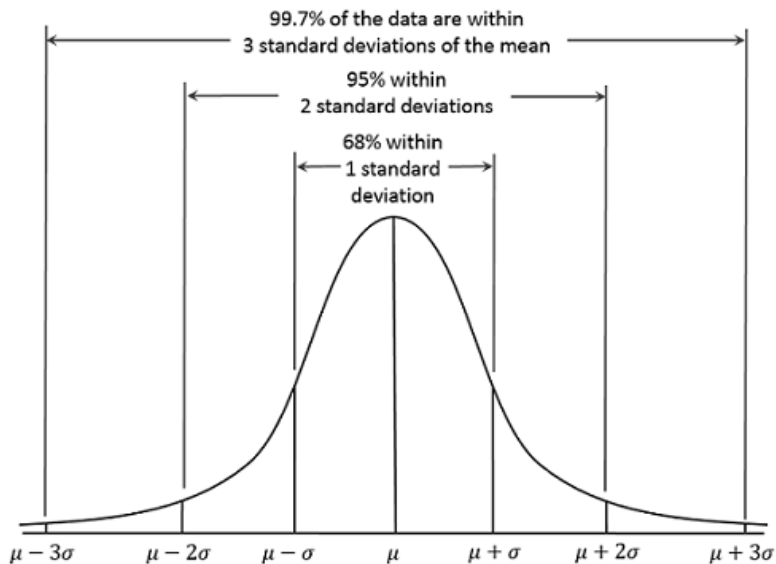


Figure 2: