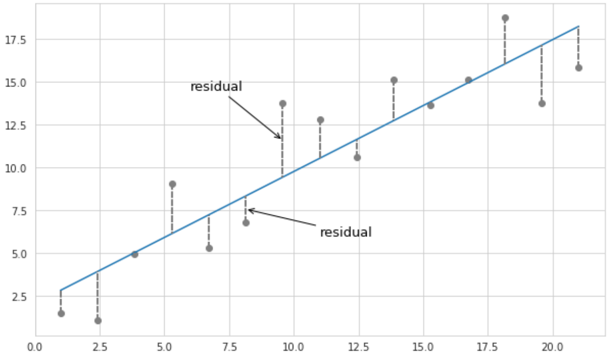
# MATH 205: Statistical methods

Lab 10: Linear regression

# This lab

- Simple fitting of a linear model
- Linear regression
- Verifying model assumptions
- Testing about the slope $\beta_1$

# Simple linear model: lm($y \sim x$)




(Theoretical) Model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad \text{independent}$$

Fitted:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i = \hat{y}_i + e_i$$

# Simple linear model: lm($y \sim x$)

(Theoretical) Model:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad \text{independent}$$

Fitted model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i = \hat{y}_i + e_i$$

- $\hat{y}_i$: fitted values
- $e_i$: residuals
- $\hat{\beta}_0, \hat{\beta}_1$: coefficients

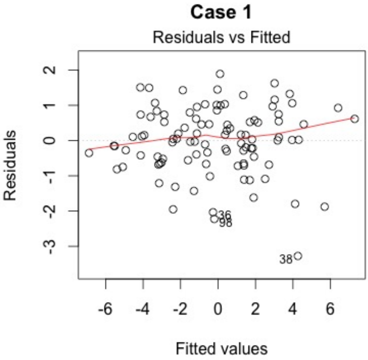
# Verifying model assumptions

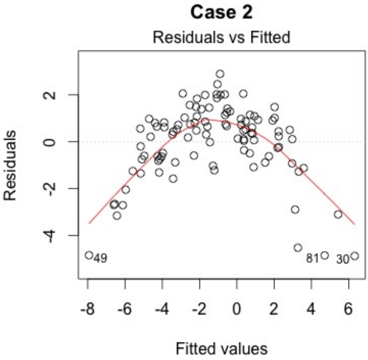
$$plot(lm(y \sim x))$$

Four plots:

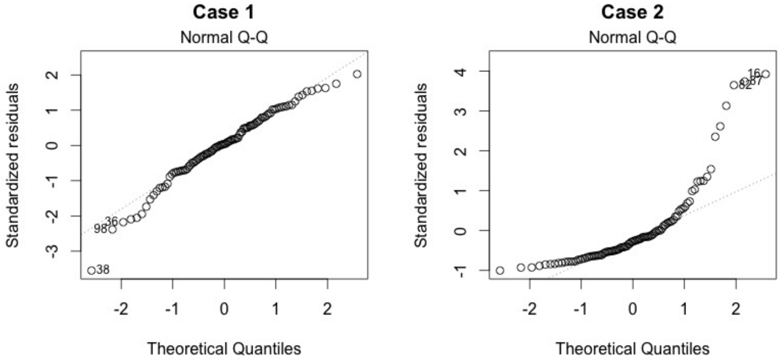
- Residuals vs Fitted (check non-linear patterns in the data)
- Normal Q-Q (check normal assumption)
- Scale-Location (check if the variance are the same across the points)
- Residuals vs Leverage (check for outliers)

# Residuals vs Fitted

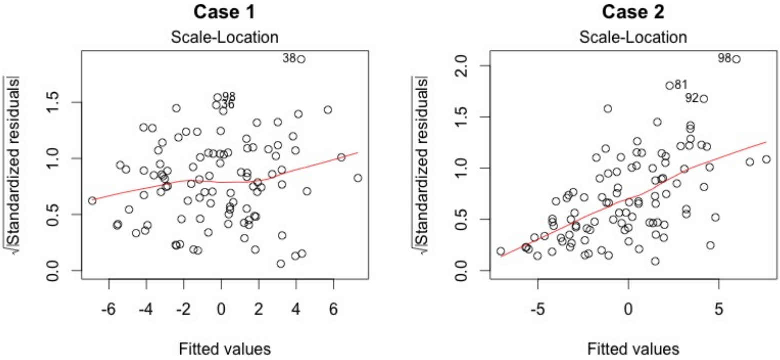


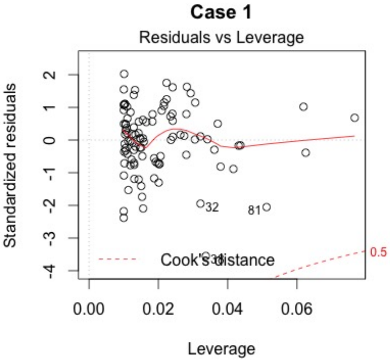
# Normal Q-Q

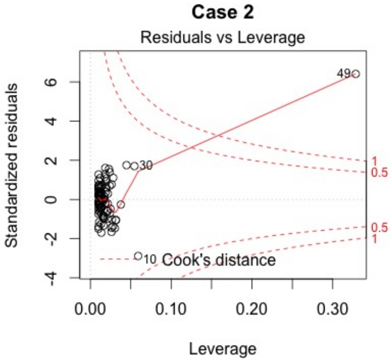

Case 1
Normal Q-Q
Standardized residuals
Theoretical Quantiles
98
36
38
Case 2
Normal Q-Q
Standardized residuals
Theoretical Quantiles
16
87
82

# Scale-Location



# Residuals vs Leverage

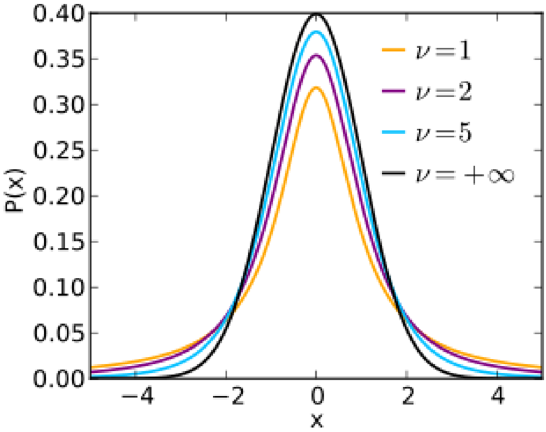


# $t$ distributions with degree of freedom $\nu$

Probability density function

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$



# Probabilistic properties of $\hat{\beta}_1$

## Theorem

*The random variable*

$$\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}}$$

*follows* $t$ *distribution with* $n-2$ *degree of freedom (df), where*

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

*and*

$$S = \frac{1}{n-2} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

Example: We test

$$H_0 : \beta_1 = -1$$
$$H_a : \beta_1 \neq -1$$

by using

$$t = \frac{\hat{\beta}_1 - (-1)}{S/\sqrt{S_{xx}}}$$