Linear regression

STATS – Modelação Estatística

PhD Programme in Health Data Science

Cristina Costa Santos & Andreia Teixeira

Comparison of Nested Models

Nested Models: Model M_1 is nested in Model M_2 if the parameters in Model M_1 are a subset of the parameters in Model M_2.

Examples:

M_1: $\text{salary} = \beta_0 + \beta_1 * \text{Experience} + \beta_2 * \text{Management}$
M_2: $\text{salary} = \beta_0 + \beta_1 * \text{Experience} + \beta_2 * \text{Age}$
M_3: $\text{salary} = \beta_0 + \beta_1 * \text{Experience} + \beta_2 * \text{Management} + \beta_3 * \text{Age}$

M_1 is nested in M_3
M_2 is nested in M_3
M_1 is not nested in M_2

Comparison of Nested Models

Requirements: Model M_1 nested in Model M_2

H_0: M_1 and M_2 have the same goodness-of-fit
H_1: M_2 has a better goodness-of-fit than M_1

Reject $H_0 \Rightarrow$ choose M_2
Do not reject $H_0 \Rightarrow$ choose M_1

Instruction in R:

anova(m1,m2)

where m1 and m2 are the models whose goodness of fit is being compared.

Comparison of Non-Nested Models

M_1 and M_2 non-nested models

There are no hypothesis tests for comparison of the model's goodness-of-fit

Instead, **information criteria** can be used:

AIC: Akaike Information Criterion

$$AIC = -2LL + 2p$$

BIC: Bayesian Information Criterion

$$BIC = -2LL + p*\log(n)$$

where LL is the value of the model's log-likelihood function, p is the number of regression parameters and n is the nmber of observations.

The lower the infomation criterion, the better.

Comparison of Non-Nested Models

Instruction in R:

extractAIC(object,…,k)

where

- object is the model
- k is the penalization to be used: k=2 gives AIC; k = log(n) gives BIC

Comparison of non-nested linear models can also use R^2_a.