

# Applied Machine Learning and Predictive Modelling 1 - Notes

Nils Rechberger

2026-02-09

## Week 01: Linear Models

### Regression

$$y = \beta_0 + \beta_1 \cdot x_1 + \epsilon$$

$\beta_0$  and  $\beta_1$  are the regression parameters:

- $\beta_0$  is also the intercept
- $\beta_1$  ist the slope

### Coefficients

The coefficients are estimated from data. Estimated regression coefficients are denoted with a hat (e.g.  $\hat{\beta}_1$ ). Fitted values (i.e. what the model predicts) are denoted with a hat as well (i.e.  $\hat{y}$ ). Residuals are the difference between observed and predicted values.

$$\text{res} = y - \hat{y}$$

If the errors are normally distributed, regression coefficients can be tested with t-tests

#### Caution

Dichotomising p-values into "significant"/"non-significant" is very bad practice!

The grade of fit can be quantified with  $R^2$

$$R^2 = \text{corr}(y, \hat{y})^2$$

Note:  $R^2$  is not used to formally compare model.

## **p-values**

The p-value quantifies the probability of observing the value of the test statistic, or a more extreme value, under the null hypothesis. Low p-values are coherent with a rejection of the null hypothesis stating that there is no effect. Large p-values (close to 1) do not imply the we can accept the null hypothesis.

## **Testing**

### **Categorical Values**

Categorical values can be tested via F-tests by using the `drop1()` function or by comparing two models via the `anova()` function. Comparisons among levels of a factor (i.e. "contrasts") can be performed by using the `glht()` function.

### **Continuous or discrete variables**

Continuous (and discrete) variables can be tested via F-tests (with `drop1()`) or by t-tests (with `summary()`). Sometimes the inferential results for continuous variables are best displayed and communicated with confidence intervals.

## **Interactions**

If an interaction term is shown to be significant, then all terms involved in this interaction play a relevant role. An interaction involving two predictors is called "two-fold interaction" (e.g. `age * species`)

Note: An interaction can involve more than two predictor.