Statistical tests and linear models

Sebastian Duchene
University of Sydney

- Parametric tests
- Non-parametric test
- Linear models (formulas)
  - Analysis of variance
  - Linear regression (OLS)

Syntax for main functions in the stats package

- Student's T test
  - t.test(x, y)
  - Returns class "htest"
  - Use function str() to find *T* and the *P-value.*
  - Similar indexing as used in vectors

Syntax for main functions in the stats package

- Shapiro-Wilk test
  - shapiro.test(x)
  - Returns class “htest”
  - Use function str() to find the statistic and *P-value.*
  - Similar indexing as used in vectors

Formulas in R

- Use the syntax: y ~ x
  - In words: variable y is explained by x
  - In linear models both variables can be continuous
  - For ANOVA, variable x should be a factor and y should be continuous

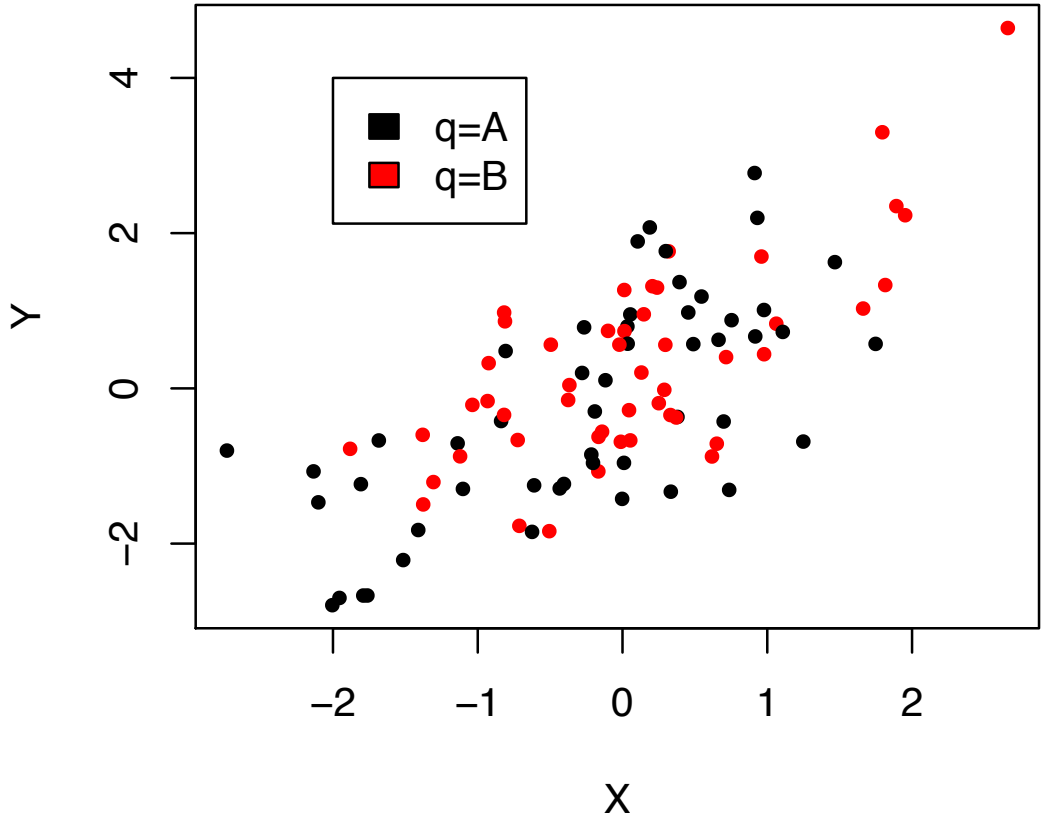
Syntax for linear models

- lm(formula)
  - Returns a model
  - Returns class “lm”

- anova(model):
  - Performs an ANOVA on a model object
  - Returns objects of class “anova” and “data.frame”

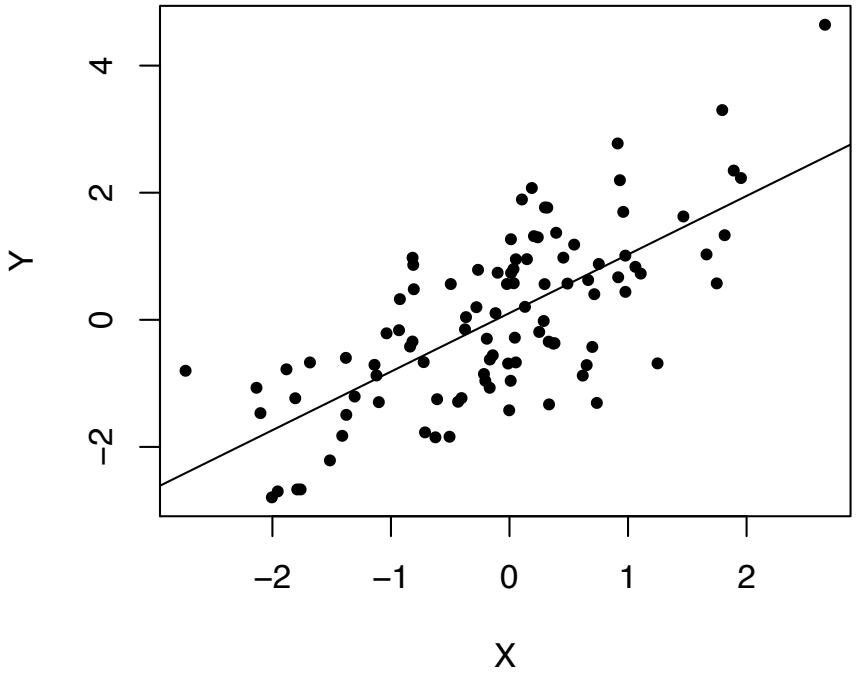
- summary(model)
  - Returns “summary objects” with *P* values and relevant statistics

Example steps for a linear model



- *x* is continuous
- *y* is continuous
- *q* is a categorical variable (factor)

- Linear regression for y ~ x



> model.1 <- lm(y ~ x)

Defines $y = a\,x + b + e$

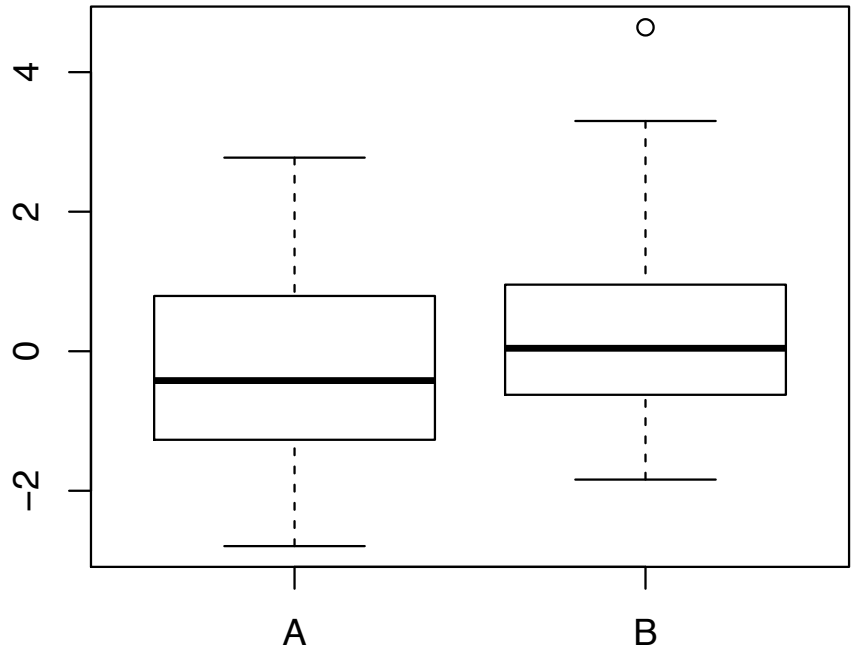
> summary(model.1)

returns regression coefficients *a* and *b*, and the *P*-values associated with the *t* statistic.

> anova(model.1)

returns sum of squares and *P*-values associated with the *F* statistic.

- For ANOVA (y ~ q)



> model.2 <- lm(y ~ q)

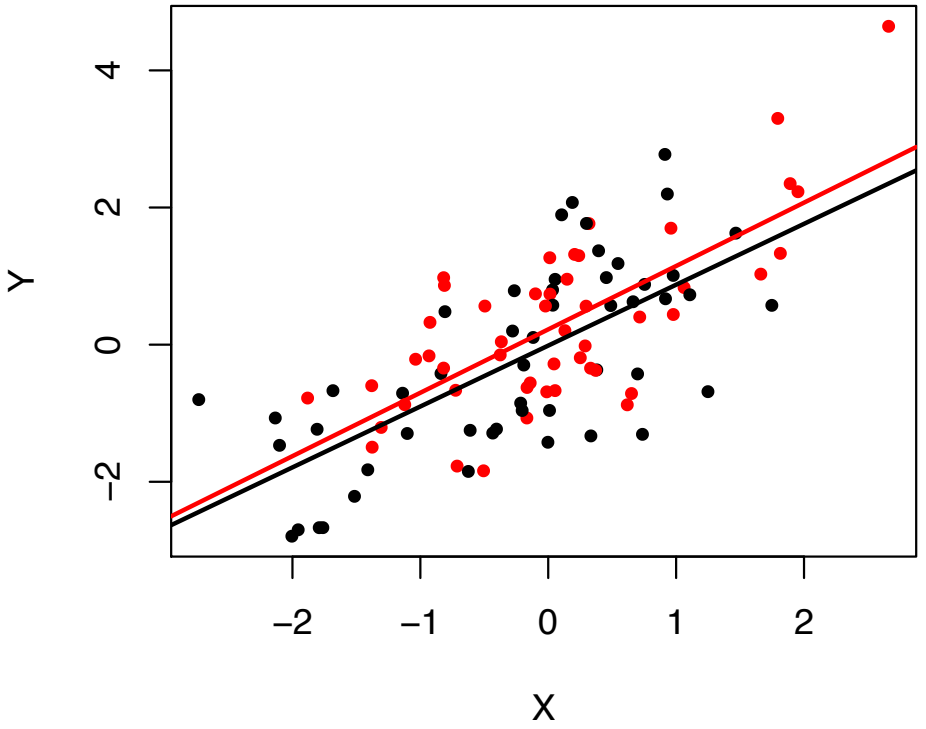
> summary(model.2)

returns the *P*-value associated with the *t* statistic.

> anova(model.2)

returns the *P*-value associated with the *F* statistic.

- For multiple regression



> model.3 <- lm(y ~ x*q)

*# z*z* is short for *x+q+x:q*

> summary(model.3)

returns the *P*-values associated with the *t* statistic.

> anova(model.3)

returns the *P*-values associated with the *F* statistic.

Other classic tests available

- Chi-squared
  - chisq.test(x, y)

- Equal variance tests
  - var.test(x, y, paired = F)

- Non-parametric statistics
  - kruskal.test(x, y)
  - wilcox.test(x, y, paired = F)

  - shapiro.test(x)
  - Returns class “htest”
  - Use function str() to find the statistic and *P-value.*
  - Similar indexing as used in vectors

Tutorial 7: Statistical tests and linear models

- Basic tests and linear models in the stat_modelling.csv data set
- Plotting statistics and regression lines

We learned...

- Many classical statistical tests are implemented in standard R
  - Other tests are available in other packages
- Statistical models typically require formula objects
- The output from statistical analyses can be included in plots

Tutorial 8: Tools for data analysis

- Data analysis of cyst_fibr.csv data set