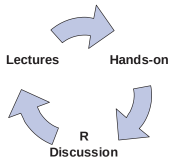
Course 02441: Applied Statistics and Statistical Software

Lasse Engbo Christiansen, IMM

Today's program: outline

- Simple and Multiple Linear Regression
- Hands-on exercises



Methods covered in the course

- Descriptive statistics — Day I
- Comparing treatment means (t-test and non-parametric tests) — Day I
- Multiple regression analysis — Day II
- Analysis of variance — Day III
- Analysis of proportions and counts — Day IV
- The general linear model — Day V

Simple linear regression

- We consider a simple linear regression model on the form

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Intercept Slope

Dependent variable

Independent variable

Residual

Regression models

- Regression models are often applied in order to:
  - Make inference for the relation between X and Y
  - Make predictions of future values of Y

Model assumptions in linear regression

- Linearity: The plot of response means against the explanatory variables should be a straight line
- The model residuals should be independent, having constant variance and be normally distributed

Model assumptions should be checked by appropriate plots before making inference

Inference in linear regression

- Inference in linear regression can be made by testing hypotheses about the model parameters, e.g.

$$H_0: \quad \beta = 0$$
$$H_1: \quad \beta \neq 0$$

Example: Hubble's law

- In 1929, Edwin Hubble investigated the relationship between distance of a galaxy from the earth and the velocity with which it appears to be receding.
- This is thought to be the result of the "Big Bang". The data collected include distances (megaparsecs) to 24 galaxies and their recession velocities (km/sec).
- Hubble's law is as follows:
  Recession Velocity = Ho*Distance
  where Ho is Hubble's constant thought to be about 75 km/sec/Mpc.

Hands-on exercise 1: Simple regression using R

- After Exercise 1 you should be able to
  - Estimate the parameters in a regression model
  - Test parameter significance (model reduction)
  - Perform model diagnostics
  - Use the model for prediction

Transformation of variables

- Both, the dependent and independent variables may be transformed in order to make a better fit and fulfill the assumptions in linear regression
- Often used transformations are logarithm, square root, inverse, square, and cube.

Multiple linear regression

- We now consider models with several independent variables

$$Y_i = \alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_n X_{n,i} + \varepsilon_i$$

Multiple linear regression

- The inference in multiple linear regression must be made with care, i.e. with condition on the variables included or left out of the model
- A problem in multiple linear regression is multi co-linearity, meaning that two (or more) of the independent variables (X's) are mutually correlated