Simple linear Regression analysis

Terminology and
Basic philosophy

Population and Sample

- Data (=sample) is created by sampling from the population.
- Population (P) is the set of all possible values for the variables existing in the ”universe”. For simple regression this means all the possible pairs (x,Y).
- The sample (S) is obtained by the data generation process (DGP). For classical linear regression model this means that x is kept constant and Y is obtained as $Y=\alpha + \beta x + u$, where u is assumed to follow $N(0, \sigma^2)$. Error term u is random, and therefore also Y is random.

Estimation

- Estimation relates to the calculation of estimates $\hat{\alpha}$ and $\hat{\beta}$ of the parameters α and β in the population regression equation

  $$Y=\alpha + \beta x + u.$$

- This is done based on the sample and ordinarily using the Ordinary Least Squares (OLS) method. See book slides for the details.
- The formula or method to calculate an estimate is called estimator (e.g. OLS estimator)
- Note, that there may exist many alternative estimators for the same parameter.
- The result is the sample regression equation

  $$Y=\hat{\alpha} +\hat{\beta} x + \hat{u}$$

- where $\hat{u}$ is called residual and refers to the estimate of the error term u.