Exercises for STK4130.

Exercise 3 A simple linear regression model $Y_i = \alpha + \beta x_i + \varepsilon_i$ where α and β are regression coefficients, x_i are given (non-random) numbers and the ε_i are iid with expectation zero and variance σ^2 (but not necessarily normal). As known the least squares estimator of β is then given by $\hat{\beta} = \sum_{i=1}^n (x_i - \bar{x})Y_i / SXX$ where $SXX = \sum_{i=1}^n (x_i - \bar{x})^2$.

In the lecture we showed that we can write $\hat{\beta} - \beta = \sum_{i=1}^n c_i \varepsilon_i$ where $c_i = (x_i - \bar{x})/SXX$ and argued that

$$\sqrt{SXX}(\hat{\beta} - \beta) \to_d \mathrm{N}(0, \sigma^2)$$

provided that

$$\max_{1 \le i \le n} c_i^2 \to 0 \text{ when } n \to \infty$$

a) Find a sufficient condition for consistency of $\hat{\beta}$ (i.e. $\hat{\beta} \to_p \beta$).

b) Construct a sequence of covariates x_i so that the condition does not hold.

c) Try to make a similar construction for the estimate of the intercept α, i.e. $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$.

d) Investigate whether the approach can be extended to multiple regression

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

with the same assumptions for ε_i.