Regression

Master Data Science | Prof. Dr. Marlene Müller
Summer 2018 | marlene.mueller@beuth-hochschule.de

Exercises 6

Exercise 1

Consider normal random variables $Y_1, \ldots, Y_n$ which are iid from the $N(\mu, \sigma^2)$ distribution. We assume that σ^2 is known.

Use the maximum likelihood principle to find an estimate for the expectation parameter μ.

Exercise 2

Generate artificial data (say a sample of size $n = 100$) from the $N(1, 4)$ distribution.

(a) Use the log-likelihood function from the previous Exercise. Plot it as a curve using the sample data and the known value $\sigma^2 = 4$.

(b) Optimize the log-likelihood numerically using R (check for example `?optimize`). Add the estimated value $\widehat{\mu}$ to your plot to check if it is really at the maximum. (You may also add the true value $\mu = 1$.)

Exercise 3

Load the dataset `Affairs` from the R package `AER`. (Check the data documentation.)

To estimate a logit model, we need a dependent variable Y with only to values 1 and 0. A useful approach is to generate Y from the variable `affairs` (say $Y = 1$ if the number of of affairs is positive and $Y = 0$ otherwise).

Do the following analyses:

(a) Explore graphically the effect of single explanatory variables on Y (you may use for example: `spineplot`, `barplot`, or `mosaicplot`).

(b) Fit at least three different logit models (some of them should be nested) and do interpret the estimated coefficients.

(c) Compare your estimated models. Instead of the F test for linear models we do now use χ^2 tests. The syntax is similar: `anova(glm1,glm2,test="Chisq")`

Alternatively, you may also compare AIC values or apply `stepAIC`.