Biostatistics: Exercise 08

Beate Sick, Lisa Herzog

3.11.2020

Exercise 01: Linear Regression I

We consider the `agefat` dataset from library `HSAUR2`.

a) Investigate the relationship between age (`age`) and body fat percentage (`fat`) as well as between gender (`gender`) and body fat percentage graphically.

b) Fit a linear regression model using the `lm()` function. Use age and gender as covariates. Interpret the estimates for the intercept, age and gender. (**R-Hint:** to fit the model use `mod` $\leftarrow$ `lm()`. To consider the results use `summary(mod)`)

c) Check the model assumptions using a Tukey-Ascombe and a normal QQ-plot (**R-Hint**: To get the fitted values and the residuals for the Tukey-Ascombe plot, you can use `fitted(mod)` and `resid(mod)`. For the QQ-plot use the function `qqPlot()` from library `car`).

d) Write down the equation of the model ($Y = \dots$). Then predict the mean body fat percentage for a 40 year old woman. You can calculate it by hand or by using the R-function `predict()`.

e) Given the model is correct - how much will the body fat change on average if a person gets 2 years older.

Exercise 2: Linear regression II]

The data set of Forbes lists the boiling point of water (in ^{o}F) and the atmospheric pressure (in inches of mercury) at different places in the alps. We want to investigate the association between the temperature (`Temp`) and the pressure (`Press`). You can read the data into R using:

a) Investigate the relationship between pressure and temperature graphically. Is it reasonable to fit a linear regression model?

b) Perform a linear regression. Investigate the influence of the temperature (covariate) on the pressure (outcome).

c) Generate a Tukey-Anscombe plot and a normal Q-Q plot of the residuals. Are there any hints that the model assumptions are violated?

d) To get a better fit, we transform the outcome variable `Press`. Calculate the log of the variable and fit a linear regression model with the transformed outcome variable. Calculate a Tukey-Ascombe and a QQ-plot. What about the model assumptions?

e) Identify and remove the outlier. Calculate a Tukey-Ascombe and a QQ-plot. What about the model assumptions now?