Exercise 1 (Linear Regression I)

We consider the `agefat` dataset from library `HSAUR2` and want to investigate the variable body fat (`fat`). The data can be loaded with:

```
# install.packages("HSAUR2")
library(HSAUR2)

## Loading required package:  tools

data("agefat")
dat = agefat
head(dat)

##   age  fat gender
## 1  24 15.5   male
## 2  37 20.9   male
## 3  41 18.6   male
## 4  60 28.0   male
## 5  31 34.7 female
## 6  39 30.2 female
```

  (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 ← 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 ($Y = ...$). 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()`.

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:

```
url = "https://polybox.ethz.ch/index.php/s/uZJZavllfYbxldy/download"
dat = read.table(url, header = TRUE)
```

(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?