MATH411 | Fall 2018 | Homework 3 (Due: Monday in class, 10/31/2018)

Your Name Here

Date Submitted Here

Problem 1

The `Conconi` test measures the endurance performance of a person. It takes place on the 400m-track where one starts running slowly (9 km/h). Every 200 meters the speed is increased by 0.5 km/h. At the end of every 200m section the pulse is measured. The test continues until the speed can no longer be increased. A test was performed in summer 2012 on 19 participants. The data is contained in the file `conconi.rda`.

(a) Visualize the data in a scatter plot with a regression line on it.

(b) Fit a SLR for investigating the relationship between `pulse` (y) and `speed` (x) in R and answer the following questions:

(b.1) Write down the statistical model, in terms of the two variables, that you fitted.

(b.2) To what extent can we explain the scatter in the pulse by the increase in speed?

(b.3) By what amount does the pulse increase on average when the speed is increased by 1 km/h? What other values are also plausible?

(b.4) How large is the resting heart rate (i.e. when there is no movement)? In what interval do you expect this value to be? Does it seem plausible?

(c) Plot the residuals against the predictor as well as the normal plot of the residuals. Decide which of the following four assumptions are fulfilled:

- The regression line captures the relation correctly, i.e. $E(\epsilon_i) = 0$.
- The variance of the error is constant, i.e. $Var(\epsilon_i) = \sigma_\epsilon^2$.
- The errors follow a Normal distribution, i.e. $\mathcal{N}(0, \sigma_\epsilon^2)$.
- The errors are uncorrelated, i.e. $Cov(\epsilon_i, \epsilon_i) = 0 \quad \forall i \neq j$

Problem 2

While fitting and visualizing simple linear regression models as well as conducting the corresponding tests becomes a routine after a while, assessing whether a model fits remains a challenging task. We will practice this with two additional data sets:

(a) The file `gas.rda` contains the gas `consumption` (in kWh) and the differences of `temperature` (in $°C$) inside and outside of 15 houses which are heated with gas. The measurements were collected over a long time span and then averaged. The goal is to explain the gas consumption with the temperature difference. Plot the regression line and perform a residual analysis.

(b) The file `antikeUhren.rda` contains the `age` and the `price` of antique clocks that are auctioned.The goal is to predict the price with the age of the clock. Plot the regression line and perform a residual analysis.

Problem 3

In an experiment marine bacteria were exposed to x-rays during 15 intervals of six minutes. The following table contains the amount of bacteria after each interval

- `Interval`: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
- `Amount`: 255, 211, 197, 166, NA, 106, 104, 60, 56, 38, 36, 32, 21, 19, 15

(a) Show the relation between the number of surviving bacteria and the number of radiation intervals. Does it make sense to fit a OLS regression to the data?

(b) Fit a simple linear regression model and check the model assumptions.

(c) Improve the model by transforming the target variable or/and the predictor. **Hint**: The theory suggests that per radiation interval the proportion of bacteria that is killed remains constant.

(d) Predict the missing value for the fifth interval and compute a 95% prediction interval. In addition, provide the expected number of bacteria at the beginning, i.e. before the first radiation interval. Also compute a 95% confidence interval for this value.