Exercise 26

A medicine designed to treat clinical depression is suspected of reducing the reaction time as a side effect. Therefore, a clinical trial is conducted with 10 randomly chosen patients, which were administered the medicine in different dosages. The reaction time of the patients was measured by the following experiment: The patient was supposed to press a button, as soon as he received a certain signal. The time that passed from transmitting the signal until the patient pressed the button was used as a measure for the reaction time. For the 10 patients the following dosages X in mg and respective reaction times Y in seconds where recorded:

i	1	2	3	4	5	6	7	8	9	10
x_i	1	5	3	8	2	2	10	8	7	4
y_i	0.5	2.9	0.6	3.0	1.5	1.1	3.9	2.5	3.1	1.2

(a) Visualize the data with a scatter plot. What conclusion about the relationship of X and Y can be drawn from the plot? Compute the correlation of X and Y and relate the result to the scatter plot.

(b) Fit a simple linear model with dose as predictor variable and reaction time as response. Compute the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ and add the corresponding regression line to the plot in (a).

(c) Give an interpretation of the regression coefficient $\hat{\beta}$ and give some comments on the quality of the fit the above linear model provides. Is the relationship of X and Y significant (at level $\alpha = 0.05$, with explanation)?

(d) A patient receives a dose of 5.5 mg of the medicine. What is the reaction time predicted with your fitted model from (b)? Obtain reaction time predictions as well for a dose of 6, 0, and 2 mg, respectively.

(e) Use appropriate (residual) plots to check the model assumptions.

Exercise 27

Consider the dataset `trees`. The aim of the data analysis is to explain/predict the variable `Volume`.

(a) Draw scatter plots for `Volume` vs. `Girth` and `Volume` vs. `Height`. Which variable seems to be the better predictor? Discuss and explain.

(b) Fit two simple linear regression models using either `Girth` or `Height` as the independent variable. Give an interpretation of the regression coefficients and of the value of R^2. What are your conclusions?

(c) Add the regression lines to your plots from (a).

(d) Draw observed vs. fitted values for your models from (b).

(e) Fit a (multiple) regression model with both `Girth` and `Height` as predictors and give an interpretation of the corresponding regression coefficients.

(f) Is the model from (e) better than the simple models from (b)? Compare e.g. the quality of the fit of the different models, and perform a formal model comparison. Write down the corresponding hypotheses you are testing here.

(g) Use appropriate plots to check the model assumptions.

(h) Take the logarithm of `Volume`, `Girth` and `Height`, and fit a (multiple) linear regression model using the transformed values. What happens?

Exercise 28

A member of the UK Building Research Station recorded the weekly gas consumption `Gas` and average external temperature `Temp` at his own house in south-east England for two heating seasons. One series of measurements was taken in the 26 weeks before, and the other series in the 30 weeks after a new insulation `Insul` was installed. The effect of outside temperature X_1 and of the insulation X_2 (a categorical variable with 2 levels `Before` and `After`) on the gas consumption Y was analyzed by a linear regression model. The following (incomplete) output shows the results of the linear regression.

```
Call:
lm(formula = Gas ~ Temp + Insul, data = whiteside)

Residuals:
     Min       1Q   Median       3Q      Max
-0.74236 -0.22291  0.04338  0.24377  0.74314

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.55133    0.11809   55.48   <2e-16 ***
Temp           (?)      0.01776  -18.95   <2e-16 ***
InsulAfter  -1.56520    0.09705   (??)    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

Residual standard error: 0.3574 on 53 degrees of freedom
Multiple R-squared:  0.9097,    Adjusted R-squared:  0.9063
F-statistic: 267.1 on 2 and 53 DF,  p-value: < 2.2e-16
```

(a) Compute the missing quantities in the output.

(b) Write down the corresponding linear model in terms of a mathematical formula (if you did not manage to compute the respective quantities in (a) make up reasonable values in case you need them for the task here) and explain it.

(c) Interpret the value of the estimated coefficient for `InsulAfter` and explain the meaning of the predictor `Insul` in the model.