Exercise 1

The file `catheter.rda` can be downloaded from the website and can be read with `load()`. The variable `height` describes the height of a patient in cm, the variable `weight` describes his weight in kg. The target variable `catlength` is the optimal length of a catheter that is used for an examination of the heart. The goal is to estimate this quantity from the available data set.

(a) Do a simple linear regression for both `catlength` $\sim$ `height` and `catlength` $\sim$ `weight`. Are the predictors significant?

(b) Fit a multiple linear regression `catlength` $\sim$ `height + weight`. Is there an influence of the predictors on the target overall? Is it significant?

(c) Test the null hypotheses $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$. Compare the results with those from the two simple linear regressions. Comment and explain the differences if there are any.

(d) For a child that is 120cm tall and has a weight of 25kg, compute the 95% prediction interval with the multiple regression model as well as with the simple regression models. In practice, a prediction error of $\pm$ 2cm would be acceptable. Do the data and the models allow for a prediction of `catlength` that is sufficiently precise? Does it make sense to use both predictors?

Exercise 2

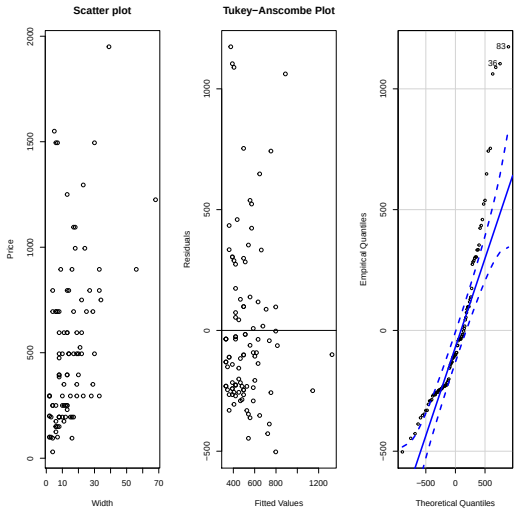
The left figure shows the price of 100 books (y; in Pence) as a function of their width (x; in mm). The data were taken for the estimation of a potential damage loss of a household insurance. The following linear regression model has been fitted to the data:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) iid$$

Here, you see a part of the R output:

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  300.485     57.468   5.229      ???
## Dicke         15.071      3.171   4.752      ???
##
## Residual standard error: ??? on 98 degrees of freedom
## Multiple R-squared: 0.1873,Adjusted R-squared: 0.179
```

```
## Loading required package:  carData
```




```
## [1] 83 36
```

(a) There is a significant correlation between width and price of books (β is significantly different from 0).

  (a) true

  (b) false

(b) Which of the following intervals is an exact 95% confidence interval for β under the the assumption of normally distributed errors?

  (a) $15.071 \pm 1.984 \cdot 3.171$

  (b) $15.071 \pm 1.984 \cdot 4.752$

  (c) $15.071 \pm \frac{1.984}{\sqrt{100}} \cdot 3.171$

  (d) $15.071 \pm \frac{1.984}{\sqrt{100}} \cdot 4.752$

  (e) None of the indicated intervals

(c) What's the estimate for $\hat{\sigma}$ approximately ("?" in the output)?

  (a) $0 \leq \hat{\sigma} < 10$

  (b) $10 \leq \hat{\sigma} < 100$

  (c) $100 \leq \hat{\sigma} < 1000$

  (d) $1000 \leq \hat{\sigma}$

(d) How much does a book of a width of 30 mm approximately cost (in Pence), based on the regression fit?

(a) 500

(b) 750

(c) 1000

(d) 1250

(e) 1500

(e) Do the model assumptions hold for the fitted data set?

(a) The connection between width and price is non-linear.

(b) The errors are not normally distributed.

(c) No deviations from the model assumptions are visible.

Exercise 3

The following data give the income, number of cows and area for a number of American farms.

Income (Dollar)	960	830	1260	610	590	900	820	880	860	760
Number of cows (cows)	18	0	14	6	1	9	6	12	7	2
Size of farm (acres)	60	220	180	80	120	100	170	110	160	230
Income (Dollar)	1020	1080	960	700	800	1130	760	740	980	800
Number of cows (cows)	17	15	7	0	12	16	2	6	12	15
Size of farm (acres)	70	120	240	160	90	110	220	110	160	80

To these data, the linear regression model

$$\text{Dollar}_i = \beta_0 + \beta_1 \text{cows}_i + \beta_2 \text{acres}_i + E_i$$

with iid $E_i \sim \text{N}(0, \sigma^2)$ was fitted.
This is part of the output from R:

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  285.457     81.379   3.508   0.0027 **
## cows          32.569      3.728     ??? 1.08e-07 ***
## acres          2.138      0.394   5.434 4.47e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 76.45 on ??? degrees of freedom
## Multiple R-squared: 0.8179,Adjusted R-squared: 0.7965
## F-statistic: 38.17 on ??? and ??? DF,  p-value: 5.165e-07
```

(a) The size of a farm has a statistically significant influence on its income.
   (a) True
   (b) False

(b) The number of cows on a farm has a statistically significant influence on its income.
   (a) True
   (b) False

(c) What is the outcome of the test of the null hypothesis $H_0 : \beta_2 = 0$ against the alternative $H_A : \beta_2 \neq 0$?
   (a) Keep H_0
   (b) Reject H_0

(d) How many degrees of freedom are there in this model fit?
   (a) ∞
   (b) 20
   (c) 18
   (d) 17
   (e) 3

(e) Which of the following is an exact 95% confidence interval for β_1?
   (a) $32.569 \pm 2.11 \cdot 3.7276$
   (b) $32.569 \pm 1.96 \cdot 3.7276$
   (c) $32.569 \pm \frac{2.11}{\sqrt{17}} \cdot 5.45$
   (d) None of these

(f) How high an income would you predict for a 100-acre farm without cows?
   (a) 285
   (b) 213
   (c) 499
   (d) 325

(g) In a simple linear regression model using the area of a farm as the only explanatory variable, would it (the area) have a significant influence on the income?

(a) Definitely

(b) Definitely not

(c) It isn't clear