Biostatistics: Exercise 09

Beate Sick, Lisa Herzog

10.11.2020

Exercise 1

The file `catheter.rda` can be downloaded from the website and can be read with `load()`.

The variables `height (in cm)` and `weight (in kg)` describe the height in centimeter and the weight in kg for a respective patient. The target variable `catlength` is the optimal length of a catheter that is used for an examination of the patient's heart. The goal is to estimate this quantity from the available dataset.

- Do a simple linear regression for both $catlength \sim height$ and $catlength \sim weight$. Is there a significant influence of the predictors on the target?
- Fit a multiple linear regression $catlength \sim height + weight$. Is there an influence of the predictors on the target overall? Is it significant?
- 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 regression models. Comment and explain the differences if there are any.
- For a child with height 120cm and weight 25kg, compute the 95% prediction interval once with the simple regression models and once with the multiple regression model. In practice, a prediction error of $\pm$ 2cm was 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? Why do we use a prediction and not a confidence interval?

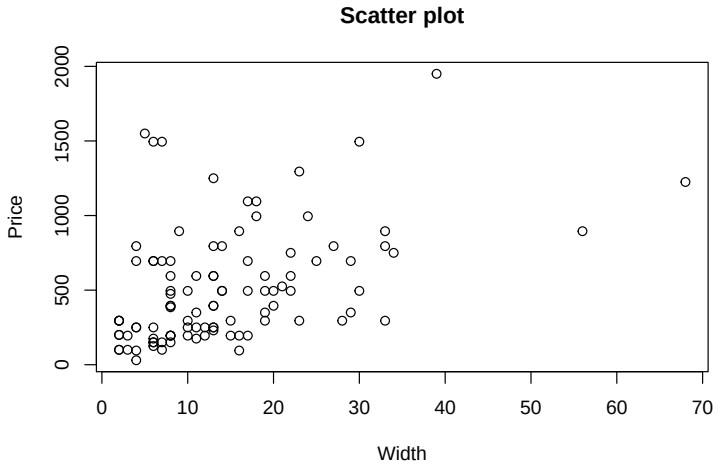
Exercise 2

The figure below 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 was fitted to the data and we assume that the model assumptions are not violated:

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

Consider the R output and the plot to answer the following questions.

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



- There is a significant correlation between width and price of books (β is significantly different from 0).
  - True
  - False
- Which of the following intervals is an exact 95% confidence interval for β under the assumption of normally distributed errors?
  - $15.071 \pm 1.984 \cdot 3.171$
  - $15.071 \pm 1.984 \cdot 4.752$
  - $15.071 \pm \frac{1}{\sqrt{100}}\ 1.984 \cdot 3.171$
  - $15.071 \pm \frac{1}{\sqrt{100}}\ 1.984 \cdot 4.752$
  - None of the indicated intervals
- What does a book of width 30mm on average approximately costs (in pence), based on the regression fit?
  - 500
  - 750
  - 1000
  - 1250
  - 1500

Exercise 3

The following dataset summarizes the income (in dollar), the number of cows and the size of the farm (in acres) for 20 American farms.

```
str(farm)
```

```
## 'data.frame':    20 obs. of  3 variables:
##  $ Dollar: int  960 830 1260 610 590 900 820 880 860 760 ...
##  $ cows  : int  18 0 14 6 1 9 6 12 7 2 ...
##  $ acres : int  60 220 180 80 120 100 170 110 160 230 ...
```

We fit the following linear regression model to the dataset:

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

with $E_i \sim N(0, \sigma^2)$ iid..

Answer the following questions with the information above and this 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
```

- The size of a farm has a statistically significant influence on its income given that the number of cows is kept fixed.
  - True
  - False
- The number of cows on a farm has a statistically significant influence on its income given that the acres are kept fixed.
  - True
  - False
- What is the outcome of the test of the null hypothesis $H_0 : \beta_2 = 0$ against the alternative $H_A : \beta_2 \neq 0$?
  - Keep H_0
  - Reject H_0
- How many degrees of freedom are there in this model fit?
  - ∞
  - 20
  - 18
  - 17
  - 3
- Which of the following is an exact 95% confidence interval for β_1?
  - $32.569 \pm 2.11 \cdot 3.7276$
  - $32.569 \pm 1.96 \cdot 3.7276$
  - $32.569 \pm \frac{1}{\sqrt{17}}\, 2.11 \cdot 5.45$
  - None of the above
- What is the predicted income for a 100 acre farm without cows?
  - 285
  - 213
  - 499
  - 548