Applied Statistical Methods - Solution 1

Peter von Rohr

2023-02-27

Problem 1: Regression Analysis

The following dataset on body weight and on futher observations on a number of animals is given.

Animal	Breast Circumference	Body Weight	BCS	HEI
1	176	471	5.0	161
2	177	463	4.2	121
3	178	481	4.9	157
4	179	470	3.0	165
5	179	496	6.8	136
6	180	491	4.9	123
7	181	518	4.4	163
8	182	511	4.4	149
9	183	510	3.5	143
10	184	541	4.7	130

The data can be read from https://charlotte-ngs.github.io/asmss2023/data/asm_bw_mult_reg.csv. The additional columns contain data on body condition score (BCS) and height (HEI).

Tasks

- Build a regression model of body weight on the other observations using the dataset given above.
- Set up the matrix $\mathbf{X}$ and the vectors $\mathbf{y}$, $\mathbf{b}$ and $\mathbf{e}$.
- Compute estimate for the regression coefficients in the model defined above.

Solution

First, the data has to be read

```
s_sol01_p01_path <- "https://charlotte-ngs.github.io/asmss2023/data/asm_bw_mult_reg.csv"
tbl_sol01_p01 <- readr::read_csv(file = s_sol01_p01_path)
```

- The regression model can be stated as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & 176 & 5 & 161 \\ 1 & 177 & 4.2 & 121 \\ 1 & 178 & 4.9 & 157 \\ 1 & 179 & 3 & 165 \\ 1 & 179 & 6.8 & 136 \\ 1 & 180 & 4.9 & 123 \\ 1 & 181 & 4.4 & 163 \\ 1 & 182 & 4.4 & 149 \\ 1 & 183 & 3.5 & 143 \\ 1 & 184 & 4.7 & 130 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 471 \\ 463 \\ 481 \\ 470 \\ 496 \\ 491 \\ 518 \\ 511 \\ 510 \\ 541 \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_0 \\ b_{BC} \\ b_{BCS} \\ b_{HEI} \end{bmatrix}$$

The residuals are assumed to have expectation $E(\mathbf{e}) = \mathbf{0}$ and a variance-covariance matrix $var(\mathbf{e}) = I_N * \sigma^2$.

- The solution for the estimates $\hat{b}$ can be computed as

$$\hat{\mathbf{b}} = (X^T X)^{-1} X^T y$$

```
mat_xtx <- crossprod(mat_X)
mat_xty <- crossprod(mat_X, vec_y)
mat_sol_bhat <- solve(mat_xtx, mat_xty)
```

The solution vector is

$$\hat{b} = \begin{bmatrix} -1313.079 \\ 9.649 \\ 8.633 \\ 0.227 \end{bmatrix}$$

Problem 2

Use the same dataset as in Problem 1 and verify your results using the function `lm()` in R.

Solution

- Read the data

```
s_sol01_p02_path <- "https://charlotte-ngs.github.io/asmss2023/data/asm_bw_mult_reg.csv"
tbl_sol01_p02 <- readr::read_csv(file = s_sol01_p02_path)
```

- Specify the call to `lm()`

```
lm_bw_mult <- lm(formula = `Body Weight` ~ `Breast Circumference` + `BCS` + `HEI`,
                 data = tbl_sol01_p02)
summary(lm_bw_mult)
```

```
##
## Call:
## lm(formula = `Body Weight` ~ `Breast Circumference` + BCS + HEI,
##     data = tbl_sol01_p02)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.686 -5.001 -2.190  5.715  9.613
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -1313.0788   209.3310  -6.273 0.000763 ***
## `Breast Circumference`     9.6493     1.0958   8.805 0.000119 ***
## BCS                        8.6332     2.8939   2.983 0.024533 *
## HEI                        0.2268     0.1736   1.306 0.239335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.088 on 6 degrees of freedom
## Multiple R-squared:  0.9294, Adjusted R-squared:  0.8942
## F-statistic: 26.35 on 3 and 6 DF,  p-value: 0.0007476
```

- Solutions are obtained by

```
coefficients(lm_bw_mult)
```

```
##            (Intercept) `Breast Circumference`                    BCS                    HEI
##          -1313.0788097              9.6492685              8.6331873              0.2267639
```