Applied Statistical Methods - Solution 2

Peter von Rohr

2023-03-06

Problem 1: Measurement Unit

The measurement unit has an influence on the results of a regression model. This is demonstrated by changing the unit for `Breast Circumference` (BC) from centimeters to meters.

Tasks

- Read the original dataset with BC and BW for 10 animals. The dataset is availble as csv-file under: https://charlotte-ngs.github.io/asmss2023/data/asm_bw_bc_reg.csv
- Convert all values of BC from centimeters to meters
- Run the regression model using `lm()`
- Compare the results with the results from the original model

Solution

- Use the function `readr::read_csv()` to read the data

```
s_data_bw_bc_url <- "https://charlotte-ngs.github.io/asmss2023/data/asm_bw_bc_reg.csv"
tbl_bw_bc <- readr::read_csv(file = s_data_bw_bc_path)
```

- Divide all values in column `Breast Circumference` by 100

```
tbl_bw_bc$`Breast Circumference` <- tbl_bw_bc$`Breast Circumference` / 100
```

- Run the regression using `lm()`

```
lm_bw_bc_m <- lm(formula = `Body Weight` ~ `Breast Circumference`, data = tbl_bw_bc)
summary(lm_bw_bc_m)
```

```
##
## Call:
## lm(formula = `Body Weight` ~ `Breast Circumference`, data = tbl_bw_bc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3941  -6.5525  -0.0673   9.3707  13.2594
##
## Coefficients:
```

```
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1065.1      255.5  -4.169 0.003126 **
## `Breast Circumference`    867.3      142.0   6.108 0.000287 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.08 on 8 degrees of freedom
## Multiple R-squared:  0.8234, Adjusted R-squared:  0.8014
## F-statistic: 37.31 on 1 and 8 DF,  p-value: 0.000287
```

- Comparison of results: intercept is the same, because it has the same unit as the response variable `Body Weight`. The regression slope is multiplied by 100 compared to the result of the original dataset where `Breast Circumference` is in cm.
- Although according to the International System of Units (https://en.wikipedia.org/wiki/International_System_of_Units) meter is the official unit for length, it does not make sense to use meter as unit for breast circumference. This is because the range of measurement for breast circumference is very narrow when the values are expressed in meters. Furthermore the increment of 1 meter is very unrealistic, compared to the range of available measurement value for breast circumference.

Problem 2: Significance Level

Do the same type of comparison of regression modelling results when changing the measurement unit for the variable HEI in the complete dataset given in

`https://charlotte-ngs.github.io/asmss2023/data/asm_bw_mult_reg.csv`.

Tasks

- Run the same regression model as in Problem 1 of Exercise 1
- Convert the measurement unit for the variable HEI from centimeter to meter
- Compare the results of the two regression models with a special focus on the significance level

Solution

- Start by reading the data

```
s_sol02_p02_path <- "https://charlotte-ngs.github.io/asmss2023/data/asm_bw_mult_reg.csv"
tbl_sol02_p02 <- readr::read_csv(file = s_sol02_p02_path)
```

- Regression model

```
lm_sol02_p02 <- lm(formula = `Body Weight` ~ `Breast Circumference` + BCS + HEI,
                   data = tbl_sol02_p02)
summary(lm_sol02_p02)
```

```
##
## Call:
## lm(formula = `Body Weight` ~ `Breast Circumference` + BCS + HEI,
##     data = tbl_sol02_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
```

- Convert HEI to meters

```
tbl_sol02_p02_HEI_in_m <- tbl_sol02_p02
tbl_sol02_p02_HEI_in_m$HEI <- tbl_sol02_p02_HEI_in_m$HEI / 100
```

- Run the regression model with the new data

```
lm_sol02_p02_HEI_in_m <- lm(formula = `Body Weight` ~ `Breast Circumference` + BCS + HEI,
                            data = tbl_sol02_p02_HEI_in_m)
summary(lm_sol02_p02_HEI_in_m)
```

```
##
## Call:
## lm(formula = `Body Weight` ~ `Breast Circumference` + BCS + HEI,
##     data = tbl_sol02_p02_HEI_in_m)
##
## 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)           -1313.079    209.331  -6.273 0.000763 ***
## `Breast Circumference`     9.649      1.096   8.805 0.000119 ***
## BCS                        8.633      2.894   2.983 0.024533 *
## HEI                       22.676     17.361   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
```

- Comparison of results: As already seen in Problem 1, the regression coefficients change by a factor of 100. But because also the standard error are 100 times bigger in the second regression model, the significance result does not change.