Multiple Regression Assignment

Dr. F.J. Rodenburg

9/19/2021

Exercise 1 (example model)

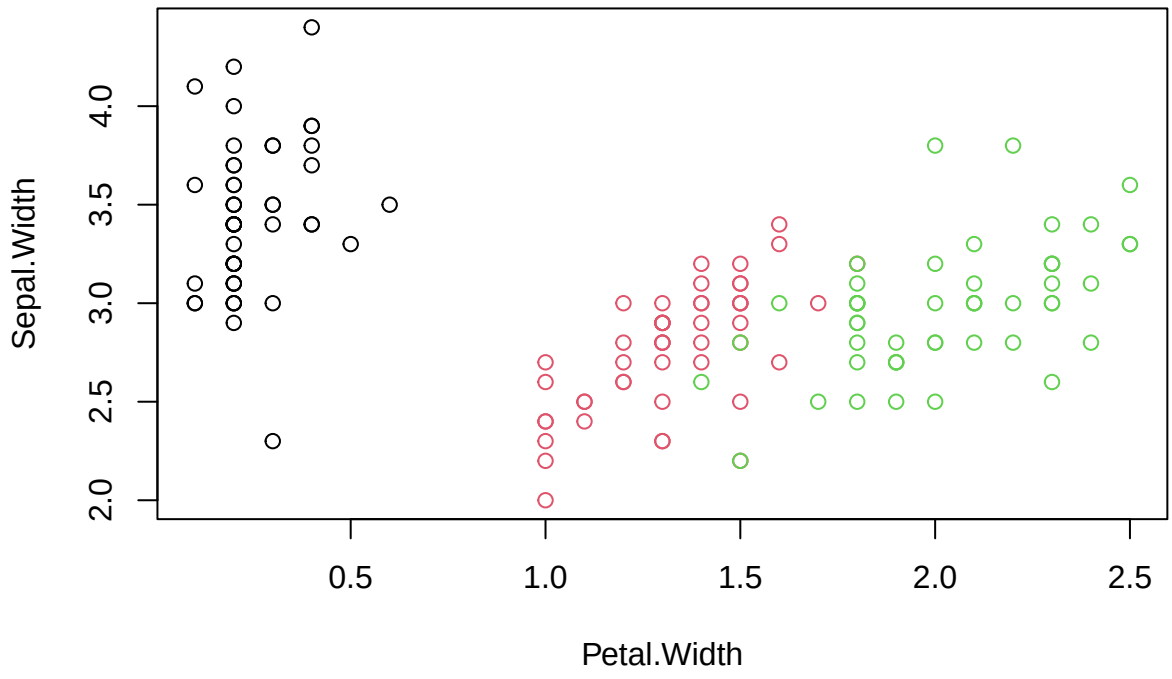
The code below fits two ANCOVA models to the `iris` data set:

- One with interaction
- One without interaction

```
ANCOVA_with    <- lm(Sepal.Width ~ Petal.Width * Species, data = iris)
ANCOVA_without <- lm(Sepal.Width ~ Petal.Width + Species, data = iris)
```

Here is a simple plot of the data:

```
plot(Sepal.Width ~ Petal.Width, data = iris, col = Species)
```



Question:

- Based on this plot, would you expect an interaction or not? Explain.

Answer:

- ...

Question:

- Can you change the code to produce a nicer looking plot?

Answer:

Diagnostics

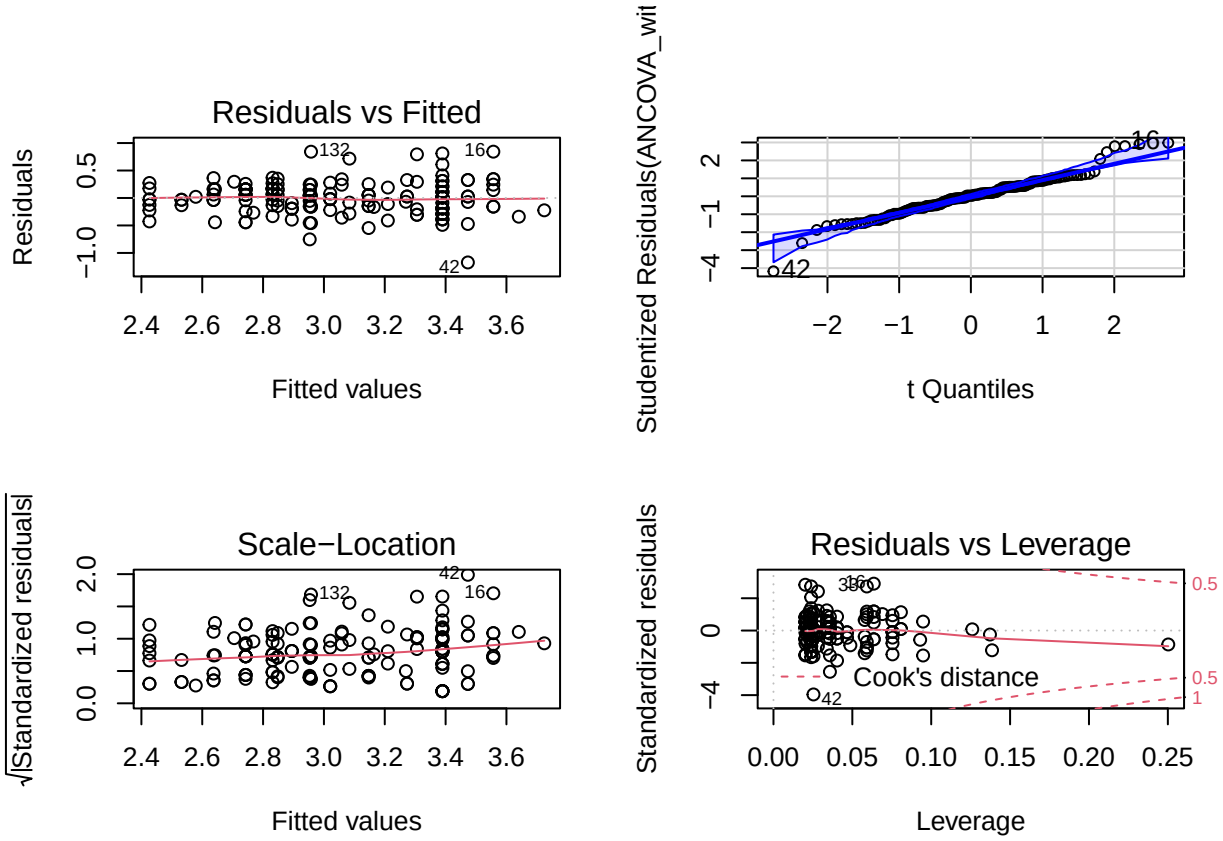
The code below performs visual diagnostics for the model with interaction. If you have trouble installing the package `car`, change `FALSE` to `TRUE` in the first line. *(You can install `car` with `install.packages("car")`.)*

```
problemsinstallingcar <- FALSE

par(mfrow = c(2, 2))
plot(ANCOVA_with, which = 1)
if(problemsinstallingcar){
  plot(ANCOVA_with, which = 2)
} else{
  car::qqPlot(ANCOVA_with)
}
```

```
## [1] 16 42
```

```
plot(ANCOVA_with, which = 3)
plot(ANCOVA_with, which = 5)
```




```
par(mfrow = c(1, 1))
```

In case of an error about **figure margins**, try increasing the size of the lower left pane in RStudio.

Question:

- How do the diagnostic plots look? Is there any cause for concern?

Answer:

- ...

Question:

- Perform the same diagnostics for the model without interaction. Adjust the code above as necessary.

Answer:

Question:

- How do the diagnostic plots look? Do you prefer either model based on the diagnostic plots?

Answer:

- ...

Model Selection

To help choose one of these models, let's perform a goodness-of-fit test:

```
anova(ANCOVA_without, ANCOVA_with)
```

```
## Analysis of Variance Table
##
## Model 1: Sepal.Width ~ Petal.Width + Species
## Model 2: Sepal.Width ~ Petal.Width * Species
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    146 13.207
## 2    144 12.980  2   0.22688 1.2585 0.2872
```

Question:

- Using $\alpha = 0.05$, what do you conclude about this test?

Answer:

- ...

Below are the R-squared and adjusted R-squared of both models:

```
summary(ANCOVA_without)$r.squared
```

```
## [1] 0.5334512
```

```
summary(ANCOVA_with)$r.squared
```

```
## [1] 0.5414662
```

```
summary(ANCOVA_without)$adj.r.squared
```

```
## [1] 0.5238646
```

```
summary(ANCOVA_with)$adj.r.squared
```

```
## [1] 0.5255449
```

Question:

- How much of the total variance in `Sepal.Width` is explained by each model?

Answer:

- ...

Question:

- Which model explains more variance with respect to the model complexity? What number do you use for this?

Answer:

- ...

The code below compares the Akaike and Bayesian information criterion of both models:

```
AIC(ANCOVA_without, ANCOVA_with)
```

```
##                df      AIC
## ANCOVA_without  5 71.19338
## ANCOVA_with     7 72.59410
```

```
BIC(ANCOVA_without, ANCOVA_with)
```

```
##                df      BIC
## ANCOVA_without  5 86.24655
## ANCOVA_with     7 93.66855
```

Question:

- Which model is better according to AIC and BIC? Do AIC and BIC agree?
- What does "better" mean if we choose a model based on these measures?

Answer:

- ...

Summary

Below is the coefficients tab of the model with and without interaction:

```
summary(ANCOVA_without)$coefficients
```

```
##                     Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)        3.2358702 0.05194469  62.294535 4.614309e-107
## Petal.Width        0.7810155 0.12121268   6.443348  1.594001e-09
## Speciesversicolor -1.5014967 0.14406798 -10.422140  2.331471e-19
## Speciesvirginica  -1.8442076 0.22398660  -8.233562  9.349089e-14
```

```
summary(ANCOVA_with)$coefficients
```

```
##                                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)                    3.2220507  0.1087479 29.6286372 3.806608e-63
## Petal.Width                    0.8371922  0.4069786  2.0570914 4.148189e-02
## Speciesversicolor             -1.8491878  0.3103819 -5.9577827 1.874607e-08
## Speciesvirginica              -1.5272777  0.3372340 -4.5288367 1.234386e-05
## Petal.Width:Speciesversicolor  0.2164556  0.4611625  0.4693695 6.395155e-01
## Petal.Width:Speciesvirginica  -0.2057870  0.4359104 -0.4720855 6.375804e-01
```

Question:

- In the previous question, you chose a "best" model. Look at the coefficients tab of this model and comment on the estimates and their uncertainty.

Answer:

- ...

Question:

- If you want to compare *versicolor* to *virginica*, how could you do that? Perform this comparison below:

Answer:

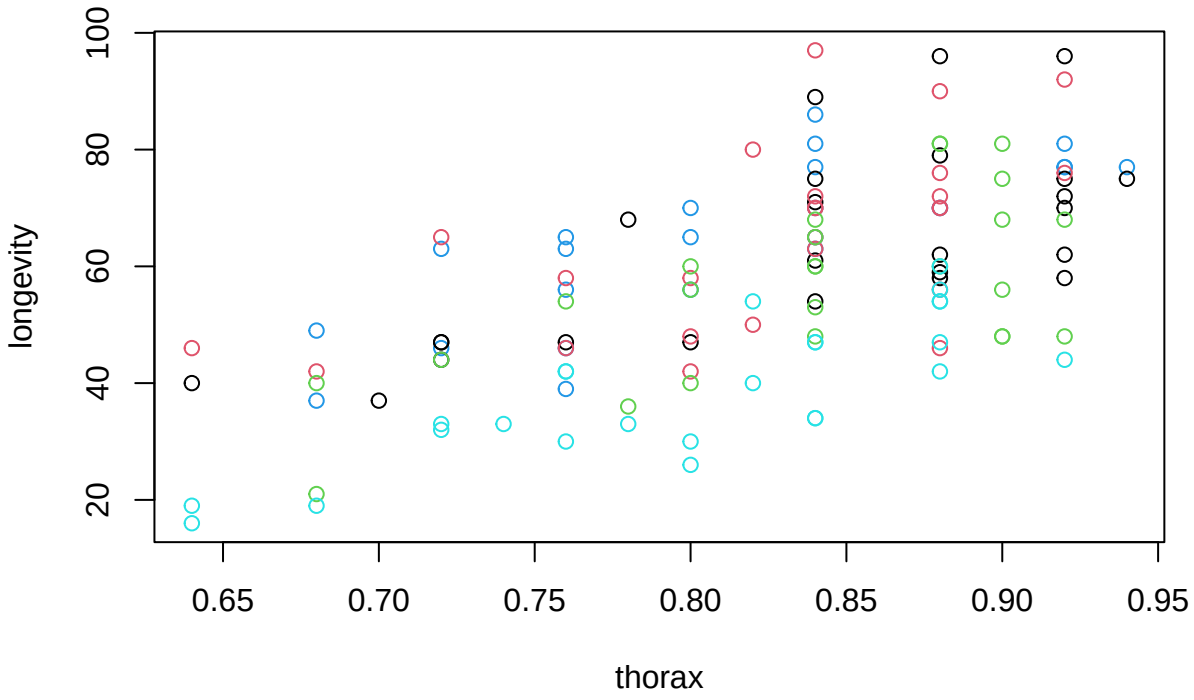
Exercise 2 (fit a model yourself)

Install the package `faraway` and run the code below:

```
library("faraway") # install if missing!
```

```
## Warning: package 'faraway' was built under R version 4.1.1
```

```
plot(longevity ~ thorax, data = fruitfly, col = activity)
```



In case of installation problems, I have included the data in the ZIP folder of the assignment. To read it, go to **Session** -> **Set Working Directory** -> **To Source File Location** and replace the code above by: `fruitfly <- read.csv("fruitfly.csv")`. Note that this will only work if you unzipped the contents of the folder.

Fit an ANCOVA with `longevity` as the outcome variable:

```
# Fit a model of the form:
# ANCOVA <- lm(... ~ ... + ..., data = ...)
# Or:
# ANCOVA <- lm(... ~ ... * ..., data = ...)
```

Then inspect the diagnostic plots (you can copy the code from the previous exercise).

Question:

- How do the diagnostic plots look? Is there any cause for concern?

Answer:

- ...

The code below performs Box-Cox transformation suggestion for the response variable:

```
require("MASS") # This library contains the function boxcox()
boxcox(ANCOVA)
```

Question:

- What does the plot suggest? If you don't remember how to interpret a Box-Cox plot, see ***2.4.2** of Elements of Biostatistics

Answer:

- ...

Question:

Independent of the Box-Cox plot, you can still suggest a transformation of `longevity` based on theory. Can you think of a logical choice for a transformation? What would you suggest? (Also explained in ***2.4.2***)

Answer:

- ...

Question:

- Use the transformation you chose and save the model into an object with a new name.
- Conduct diagnostics on this new model. Do the plots look better than before transformation?

Answer:

- ...

Question:

- Run a `summary` of the model you chose and write a conclusion of a few sentences, using all of the output you understand from the summary.

Answer:

Question (hard):

- Create a plot of your model and the actual observations. As an extra challenge, you can try to plot a confidence, or prediction interval around the model. If you do, explain which one is shown, and how to interpret it.