EDDA - Assignment 3 - Group 77

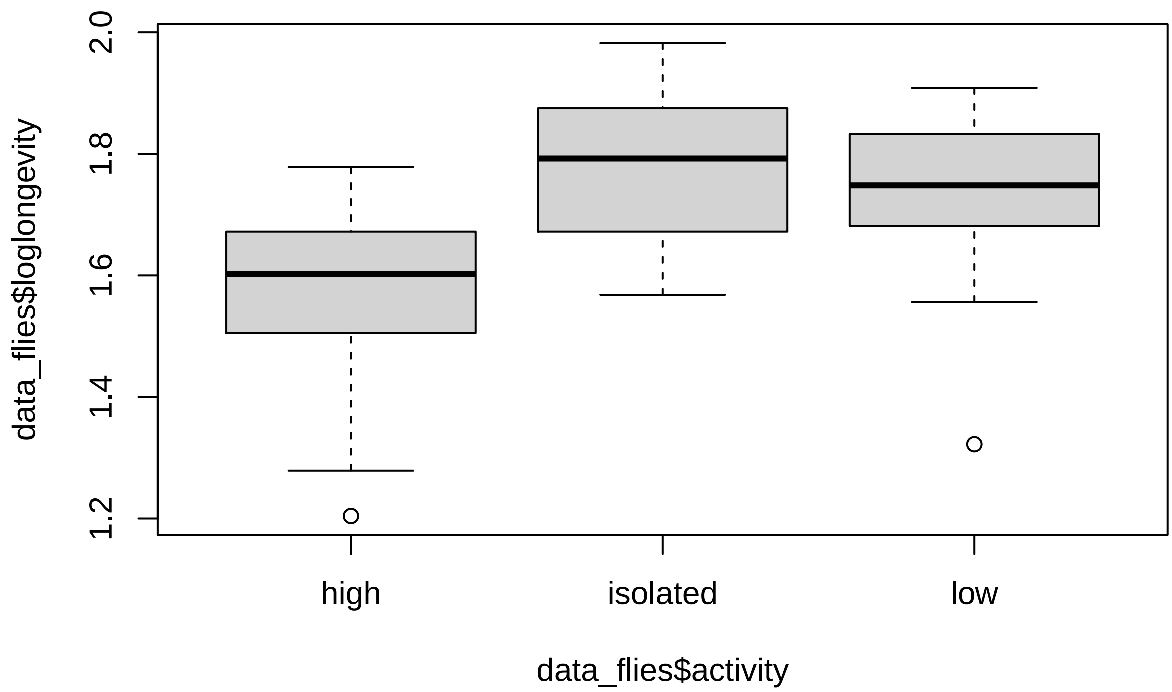
Dante de Lang, Ignas Krikštaponis and Kamiel Gülpen

Exercise 1

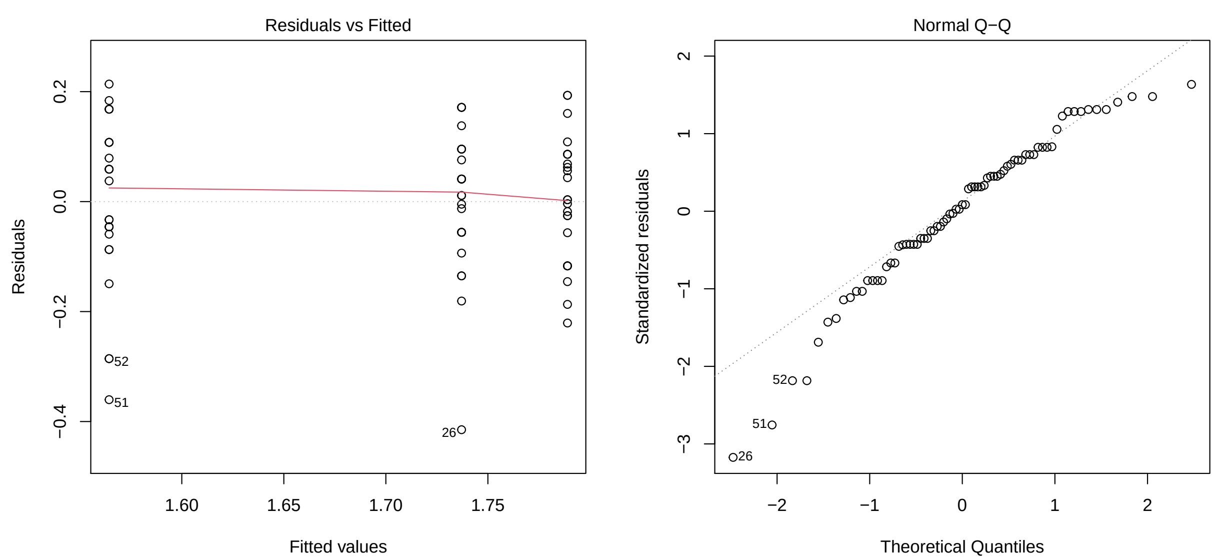
To investigate the effect of sexual activity on longevity of fruit flies, 75 male fruit flies were divided randomly in three groups of 25. The fruit flies in the first group were kept solitary, those in the second were kept together with one virgin female fruit fly per day, and those in the third group were kept together with eight virgin female fruit flies a day. In the data-file fruitflies.txt the three groups are labelled isolated, low and high. The number of days until death (longevity) was measured for all flies. Later, it was decided to measure also the length of their thorax. Add a column loglongevity to the data-frame, containing the logarithm of the number of days until death. Use this as the response variable in the following.

a) Make an informative plot of the data. Investigate whether sexual activity influences longevity by performing a statistical test, without taking the thorax length into account. What are the estimated longevities for the three conditions? Comment.

```
data_flies <- read.table(file="data/fruitflies.txt", header=TRUE)
data_flies$activity <- as.factor(data_flies$activity)
# add loglongevity
data_flies <- data_flies %>% mutate(loglongevity = log10(longevity))
# make informative plot - boxplot
plot(data_flies$loglongevity~data_flies$activity)
```




```
# perform test to see if sexual activity has an effect on longevity
model <- lm(loglongevity~activity, data = data_flies) # prepare model
par(mfrow=c(1,2)); plot(model, 1); plot(model, 2) # investigate normality
```



```r
# perform one-way ANOVA
anova(model); summary(model)$coefficients
```

```
## Analysis of Variance Table
##
## Response: loglongevity
##           Df Sum Sq Mean Sq F value  Pr(>F)
## activity   2  0.692   0.346    19.4 1.8e-07 ***
## Residuals 72  1.282   0.018
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

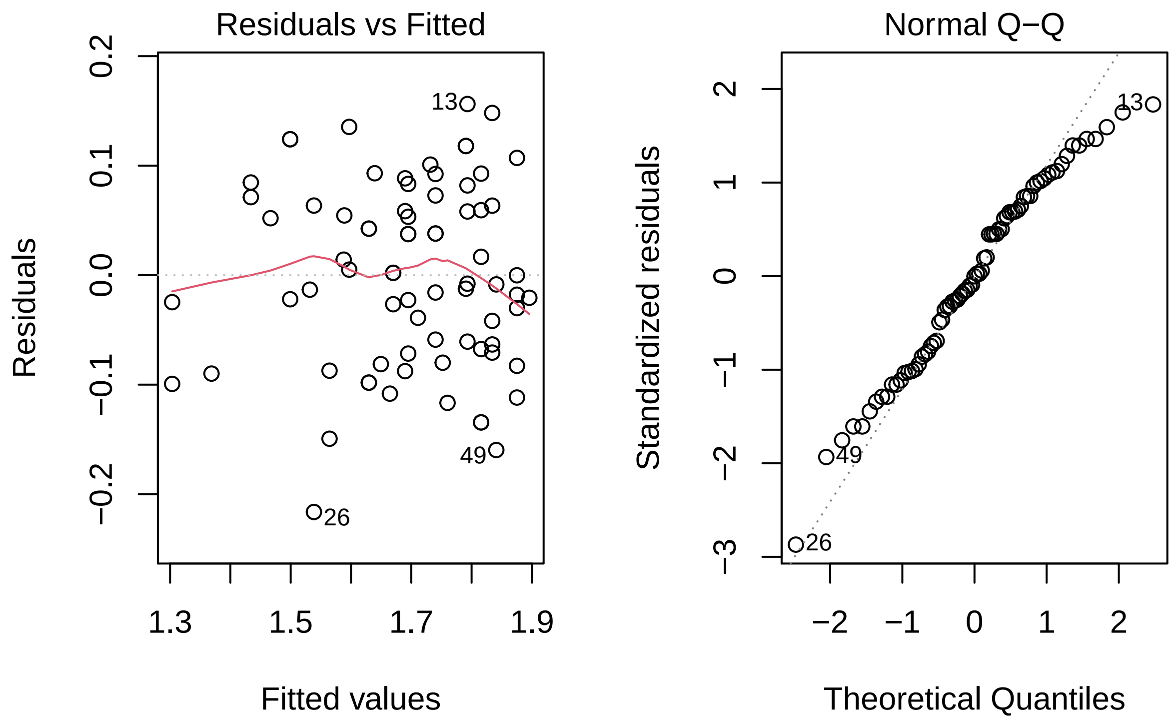
```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.564     0.0267   58.62 1.65e-62
## activityisolated    0.225     0.0377    5.95 8.82e-08
## activitylow         0.173     0.0377    4.58 1.93e-05
```

One-way ANOVA was performed to investigate whether sexual activity has an effect on loglongevity. From the results we can see that the p-values < 0.05 meaning that there sexual activity significantly influences loglongevity. From the summary table we can see that all estimates are significantly different from 0: for high sexual activity the estimate is 1.5644, for isolated it is $1.5644 + 0.2246 = 1.7890$ and low it is $1.5644 + 0.1727 = 1.7371$.

Test diagnostics: no relationship can be observed in the residuals vs fitted plot. QQ-plot seems to follow a straigt line, however there are some outliers at the extremes.

b) Investigate whether sexual activity influences longevity by performing a statistical test, now including thorax length as an explanatory variable into the analysis. Does sexual activity increase or decrease longevity? What are the estimated longevities for the three groups, for a fly with average thorax length?

```r
# perform ANCOVA with interaction analysis
model_interaction <- lm(loglongevity~thorax*activity, data = data_flies) # prepare model
par(mfrow=c(1,2)); plot(model_interaction , 1); plot(model_interaction , 2) # investigate normality
```

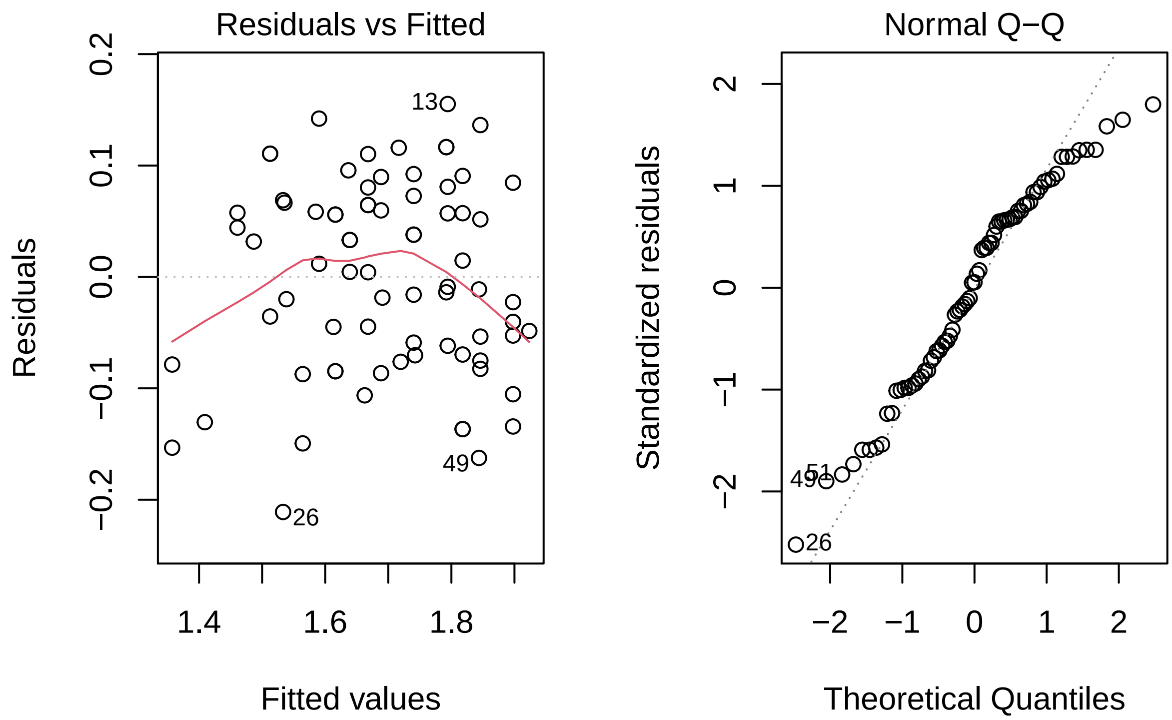



```
anova(model_interaction)
```

```
## Analysis of Variance Table
##
## Response: loglongevity
##                  Df  Sum Sq Mean Sq F value  Pr(>F)
## thorax            1  1.025   1.025   135.62 < 2e-16 ***
## activity          2  0.399   0.199    26.38 3.1e-09 ***
## thorax:activity   2  0.029   0.015     1.93    0.15
## Residuals        69  0.521   0.008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

First we investigate if there is any significant interaction between thorax and activity by performing ANCOVA with interaction. (diagnostics show that data follows the required assumptions). From the results we can see that the interaction factor is insignificant and can be ignored, therefore we can now use the additive ANCOVA model.

```
# perform additive ANCOVA analysis
model <- lm(loglongevity~thorax+activity, data = data_flies) # prepare model
par(mfrow=c(1,2)); plot(model , 1); plot(model , 2) # investigate normality
```




```
anova(model); table <- summary(model)$coefficients; table
```

```
## Analysis of Variance Table
##
## Response: loglongevity
##           Df Sum Sq Mean Sq F value Pr(>F)
## thorax     1  1.025   1.025   132.2 <2e-16 ***
## activity   2  0.399   0.199    25.7  4e-09 ***
## Residuals 71  0.550   0.008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.529     0.1080    4.90 5.79e-06
## thorax               1.294     0.1332    9.71 1.14e-14
## activityisolated     0.178     0.0254    7.02 1.07e-09
## activitylow          0.124     0.0254    4.88 6.18e-06
```

```
# extract model's parameter
intercept <- table[,1][1]; beta <- table[,1][2]
alpha_high <- 0; alpha_low <- table[,1][4]
alpha_isolated <- table[,1][3]
# calculate mean thorax
mean_thorax <- mean(data_flies$thorax)
# calculate estimates
```

```
estimate_high <- 10**(intercept + alpha_high + beta * mean_thorax)
estimate_low <- 10**(intercept + alpha_low + beta * mean_thorax)
estimate_isolated <- 10**(intercept + alpha_isolated + beta * mean_thorax)
estimates <- c(estimate_isolated, estimate_low, estimate_high)
activity_levels <- unique(as.character(data_flies$activity))
knitr::kable(data.frame(Activity = activity_levels,
                        `Longevity estimate` = estimates),
             caption = "Longevity estimates for average thorax fruit fly")
```

Table 1: Longevity estimates for average thorax fruit fly

Activity	Longevity.estimate
isolated	59.5
low	52.5
high	39.5

Model diagnostics: from qq-plot we can see that the data follows a straight line well, however there are some outliers at the extremes. There does not seem to be any obvious relationship in the Residuals vs Fitted plot.

From the ANCOVA analysis results above we can see that sexual activity has a significant effect (p-values $<$ 0.05) on the loglongevity. From the estimates in the summary table we can see that sexual activity decreases longevity of the fruit flies as the estimates from isolated and low are positive and isolated has the highest estimate. Longevity estimates for average thorax fruit fly were calculated by calculating average thorax length (X) and extracting intercept (μ), β and α parameters from the model summary table - the values there plugged into the following formula:

$Y \approx \mu + \alpha + \beta X$