EDDA - Assignment 3 - Group 77

Dante de Lang, Ignas Krikštaponis and Kamiel Gülpen

Exercise 1

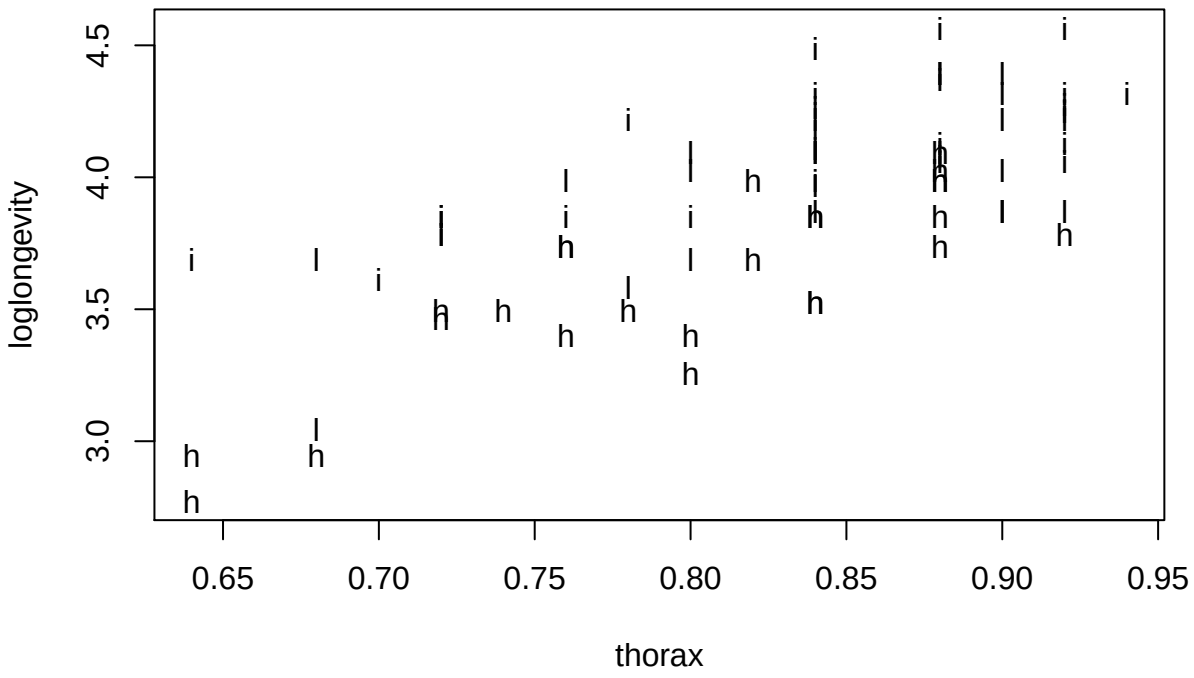
To investigate the effect of sexual activity on longevity of fruit flies, 75 male fruit flies were divided randomly in three groups of 25. The fruit flies in the first group were kept solitary, those in the second were kept together with one virgin female fruit fly per day, and those in the third group 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$loglongevity = log(data_flies$longevity)

plot(loglongevity~thorax,pch=as.character(activity), data=data_flies)
```




```
data_flies$activity=as.factor(data_flies$activity)
aov = lm(loglongevity~activity,data=data_flies)
anova(aov)
```

```
## Analysis of Variance Table
##
## Response: loglongevity
##           Df Sum Sq Mean Sq F value  Pr(>F)
## activity   2   3.67   1.833    19.4 1.8e-07 ***
## Residuals 72   6.80   0.094
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov)$coefficients
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.602     0.0614   58.62 1.65e-62
## activityisolated    0.517     0.0869    5.95 8.82e-08
## activitylow         0.398     0.0869    4.58 1.93e-05
```

We see that a high activity a estimate gives of 3.602, low = 3.602 +0.398 = 4.0 and isolated = 3.602+0.517 = 4.19

(b)investigate whether sexual activity influences longevity by performing a statistical test, now includingthorax 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?

```
aov1 = lm(loglongevity~thorax+activity,data=data_flies)
anova(aov1)
```

```
## Analysis of Variance Table
##
## Response: loglongevity
##           Df Sum Sq Mean Sq F value Pr(>F)
## thorax     1   5.43    5.43   132.2 <2e-16 ***
## activity   2   2.11    1.06    25.7  4e-09 ***
## Residuals 71   2.92    0.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov1)$coefficients
```

```
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.219     0.2486    4.90 5.79e-06
## thorax               2.979     0.3067    9.71 1.14e-14
## activityisolated     0.410     0.0584    7.02 1.07e-09
## activitylow          0.286     0.0585    4.88 6.18e-06
```

They still look significant with high = 1.219 low = 1.219 + 0.286 = 1.505 and isolated = 1.219 + 0.410 = 1.629

The plots look ok **(c)** How does thorax length influence longevity? Investigate graphically and by using an appropriate test whether this dependence is similar under all three conditions of sexual activity

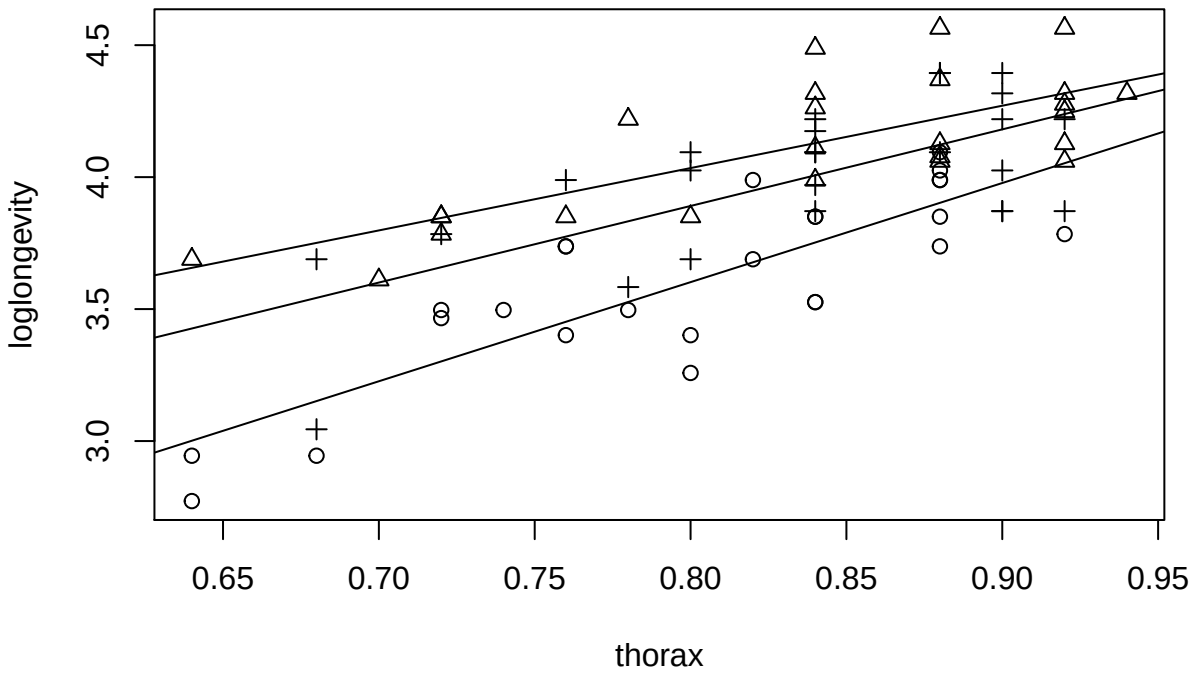
```
data_flies
```

```
##    thorax longevity activity loglongevity
## 1    0.64        40 isolated         3.69
## 2    0.70        37 isolated         3.61
## 3    0.72        44 isolated         3.78
## 4    0.72        47 isolated         3.85
## 5    0.72        47 isolated         3.85
## 6    0.76        47 isolated         3.85
## 7    0.78        68 isolated         4.22
## 8    0.80        47 isolated         3.85
## 9    0.84        54 isolated         3.99
## 10   0.84        61 isolated         4.11
## 11   0.84        71 isolated         4.26
## 12   0.84        75 isolated         4.32
## 13   0.84        89 isolated         4.49
## 14   0.88        58 isolated         4.06
## 15   0.88        59 isolated         4.08
## 16   0.88        62 isolated         4.13
## 17   0.88        79 isolated         4.37
## 18   0.88        96 isolated         4.56
## 19   0.92        58 isolated         4.06
## 20   0.92        62 isolated         4.13
## 21   0.92        70 isolated         4.25
## 22   0.92        72 isolated         4.28
```

```
## 23   0.92        75 isolated         4.32
## 24   0.92        96 isolated         4.56
## 25   0.94        75 isolated         4.32
## 26   0.68        21       low        3.04
## 27   0.68        40       low        3.69
## 28   0.72        44       low        3.78
## 29   0.76        54       low        3.99
## 30   0.78        36       low        3.58
## 31   0.80        40       low        3.69
## 32   0.80        56       low        4.03
## 33   0.80        60       low        4.09
## 34   0.84        48       low        3.87
## 35   0.84        53       low        3.97
## 36   0.84        60       low        4.09
## 37   0.84        60       low        4.09
## 38   0.84        65       low        4.17
## 39   0.84        68       low        4.22
## 40   0.88        60       low        4.09
## 41   0.88        81       low        4.39
## 42   0.88        81       low        4.39
## 43   0.90        48       low        3.87
## 44   0.90        48       low        3.87
## 45   0.90        56       low        4.03
## 46   0.90        68       low        4.22
## 47   0.90        75       low        4.32
## 48   0.90        81       low        4.39
## 49   0.92        48       low        3.87
## 50   0.92        68       low        4.22
## 51   0.64        16      high        2.77
## 52   0.64        19      high        2.94
## 53   0.68        19      high        2.94
## 54   0.72        32      high        3.47
## 55   0.72        33      high        3.50
## 56   0.74        33      high        3.50
## 57   0.76        30      high        3.40
## 58   0.76        42      high        3.74
## 59   0.76        42      high        3.74
## 60   0.78        33      high        3.50
## 61   0.80        26      high        3.26
## 62   0.80        30      high        3.40
## 63   0.82        40      high        3.69
## 64   0.82        54      high        3.99
## 65   0.84        34      high        3.53
## 66   0.84        34      high        3.53
## 67   0.84        47      high        3.85
## 68   0.84        47      high        3.85
## 69   0.88        42      high        3.74
## 70   0.88        47      high        3.85
## 71   0.88        54      high        3.99
## 72   0.88        54      high        3.99
## 73   0.88        56      high        4.03
## 74   0.88        60      high        4.09
## 75   0.92        44      high        3.78
```

```r
activities = c("isolated", "low", "high")

plot(loglongevity~thorax, pch = unclass(activity), data = data_flies);for (i in 1:3) abline(lm(loglonge
```




```r
aov2=lm(loglongevity~activity*thorax,data=data_flies); anova(aov2)
```

```
## Analysis of Variance Table
##
## Response: loglongevity
##                 Df Sum Sq Mean Sq F value  Pr(>F)
## activity         2   3.67    1.83   45.77 2.2e-13 ***
## thorax           1   3.88    3.88   96.83 9.0e-15 ***
## activity:thorax  2   0.15    0.08    1.93    0.15
## Residuals       69   2.76    0.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(aov2)
```

```
##
## Call:
## lm(formula = loglongevity ~ activity * thorax, data = data_flies)
##
## Residuals:
```

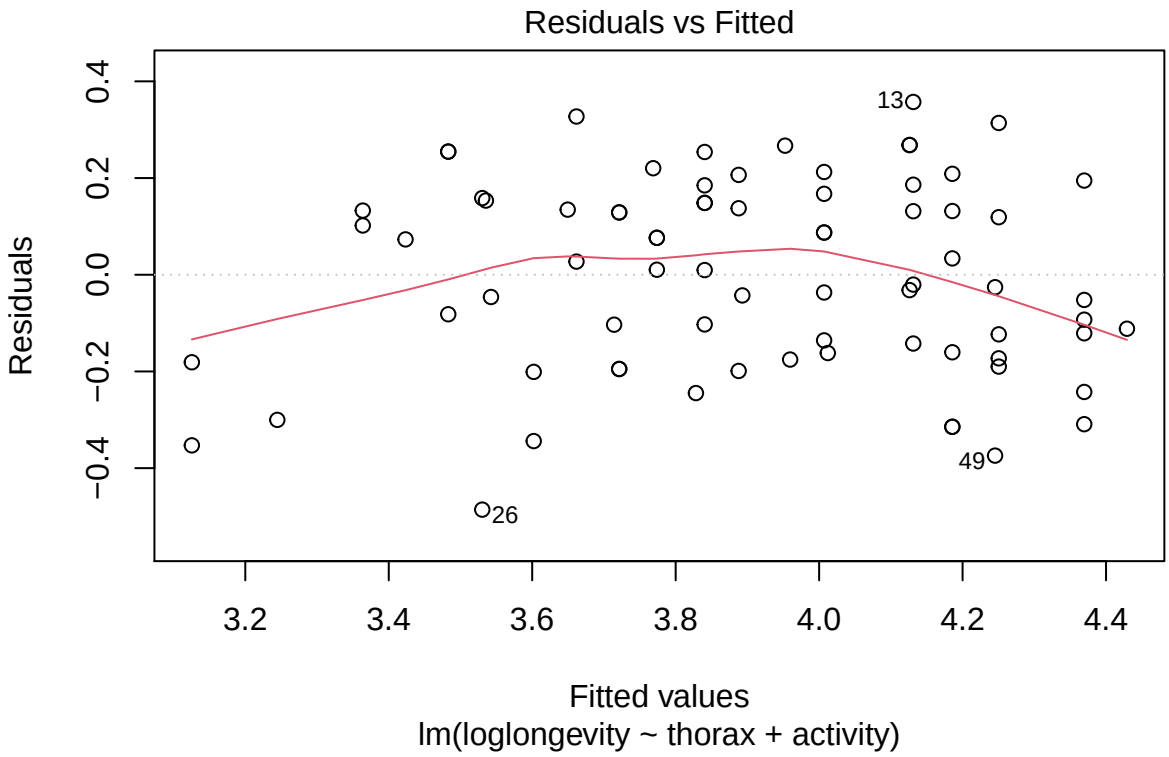
```
##     Min      1Q  Median      3Q     Max
## -0.4980 -0.1592 -0.0003  0.1462  0.3598
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.598      0.419    1.43    0.158
## activityisolated            1.546      0.584    2.65    0.010 *
## activitylow                 0.972      0.642    1.51    0.135
## thorax                      3.755      0.522    7.20  5.8e-10 ***
## activityisolated:thorax    -1.393      0.712   -1.96    0.055 .
## activitylow:thorax         -0.854      0.779   -1.10    0.277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2 on 69 degrees of freedom
## Multiple R-squared:  0.736,  Adjusted R-squared:  0.717
## F-statistic: 38.4 on 5 and 69 DF,  p-value: <2e-16
```

We see that in the graphical analysis that the thorax has a influence on the loglongevity, as we see that for all three types of activities the loglongevity rises when the thorax increases. After looking at the statistical test we see indeed that the thorax has a significant influence on the loglongevity as it has a p-value of 9.0e-15.

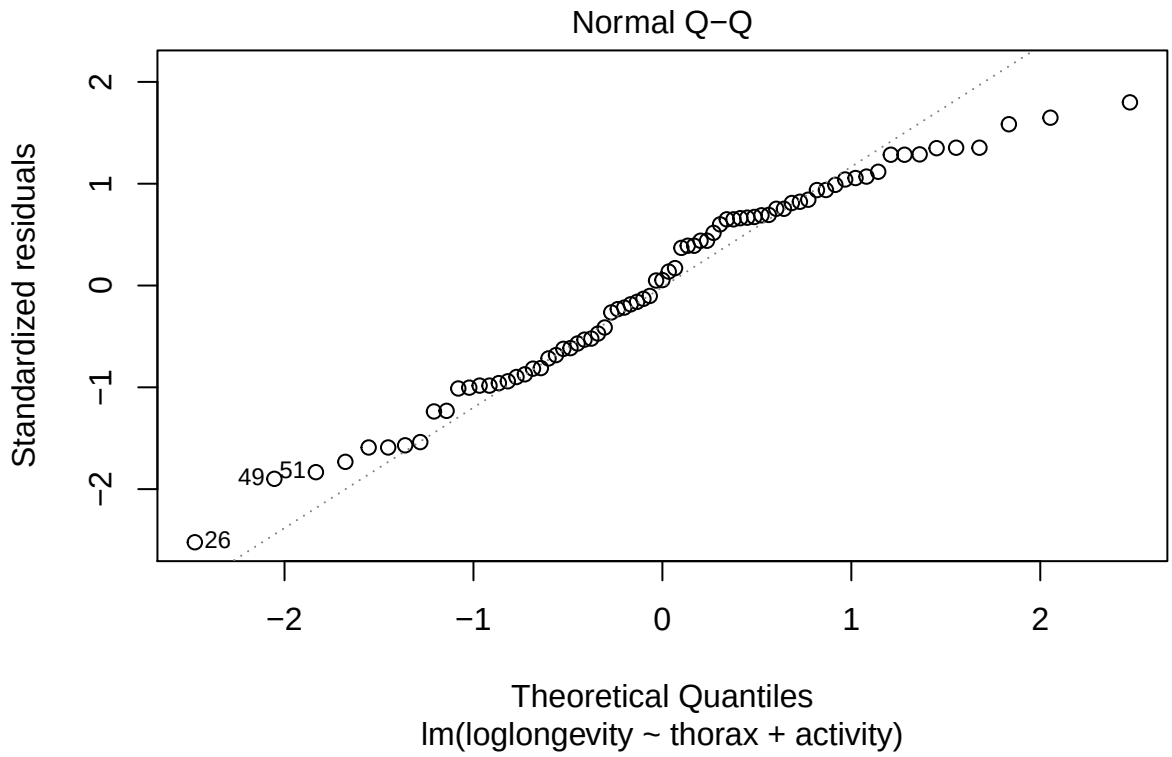
d) Which of the two analyses, without or with thorax length, do you prefer? Is one of the analyses wrong? I prefer the analysis without the thorax as the the thorax and activity do not have a significant interaction. The second model is wrong if we want to look at the influence of activity becasue of this.

e)Verify normality and heteroscedasticity by making a normal QQ-plot of the residuals, and a residualsversus fitted plot, for the analysis that includes thorax length.

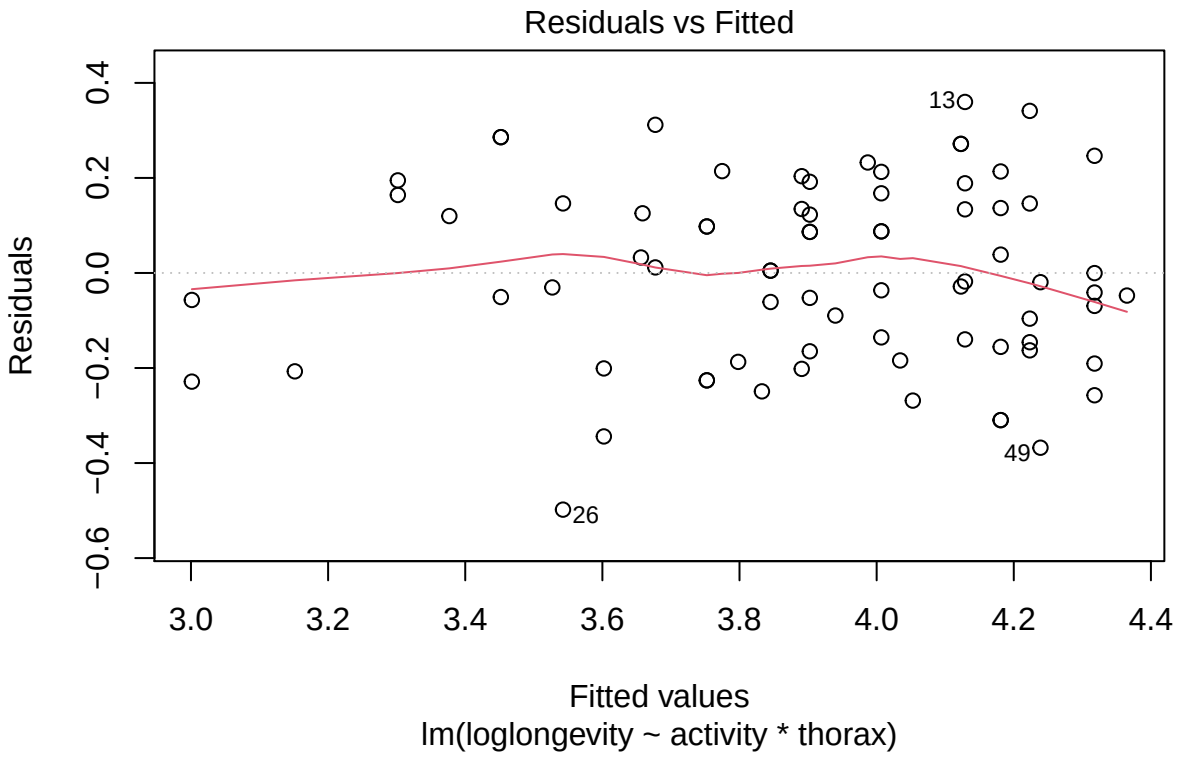
```
plot(aov1, 1)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
13
49
26
3.2
3.4
3.6
3.8
4.0
4.2
4.4
Fitted values
lm(loglongevity ~ thorax + activity)

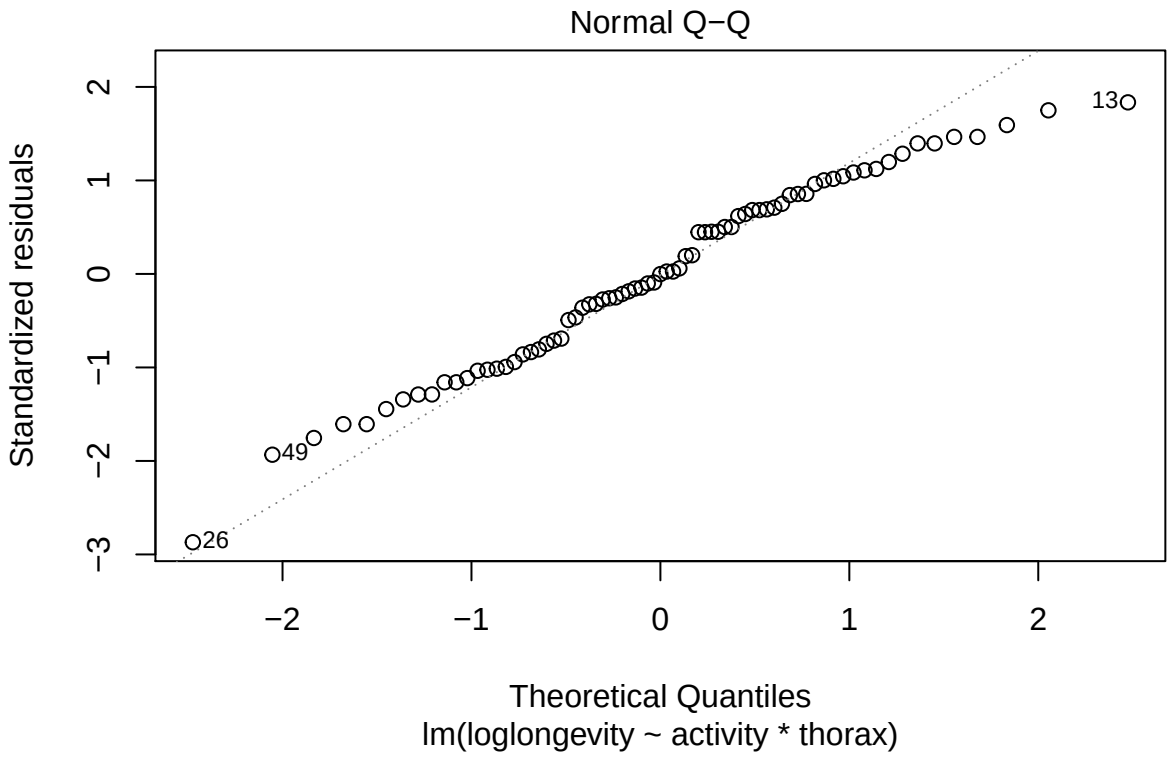
```
plot(aov1, 2)
```


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
49
51
26
−2
−1
0
1
2
Theoretical Quantiles
lm(loglongevity ~ thorax + activity)

```
plot(aov2,1)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
−0.6
13
49
26
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
Fitted values
lm(loglongevity ~ activity * thorax)

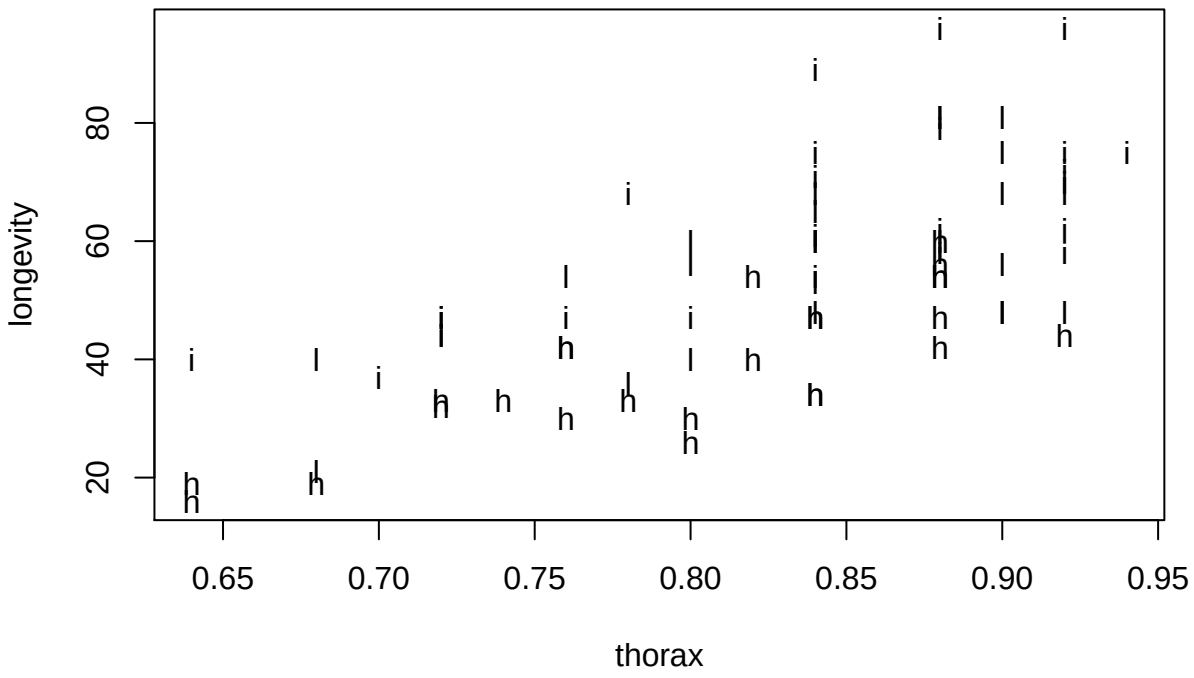
```
plot(aov2,2)
```



The plots look somewhat ok. It does show some outliers.

f)Perform the ancova analysis with the number of days as the response, rather than its logarithm. Verifynormality and homoscedasticity of the residuals of this analysis. Was it wise to use the logarithm as response?

```
plot(longevity~thorax,pch=as.character(activity), data=data_flies)
```



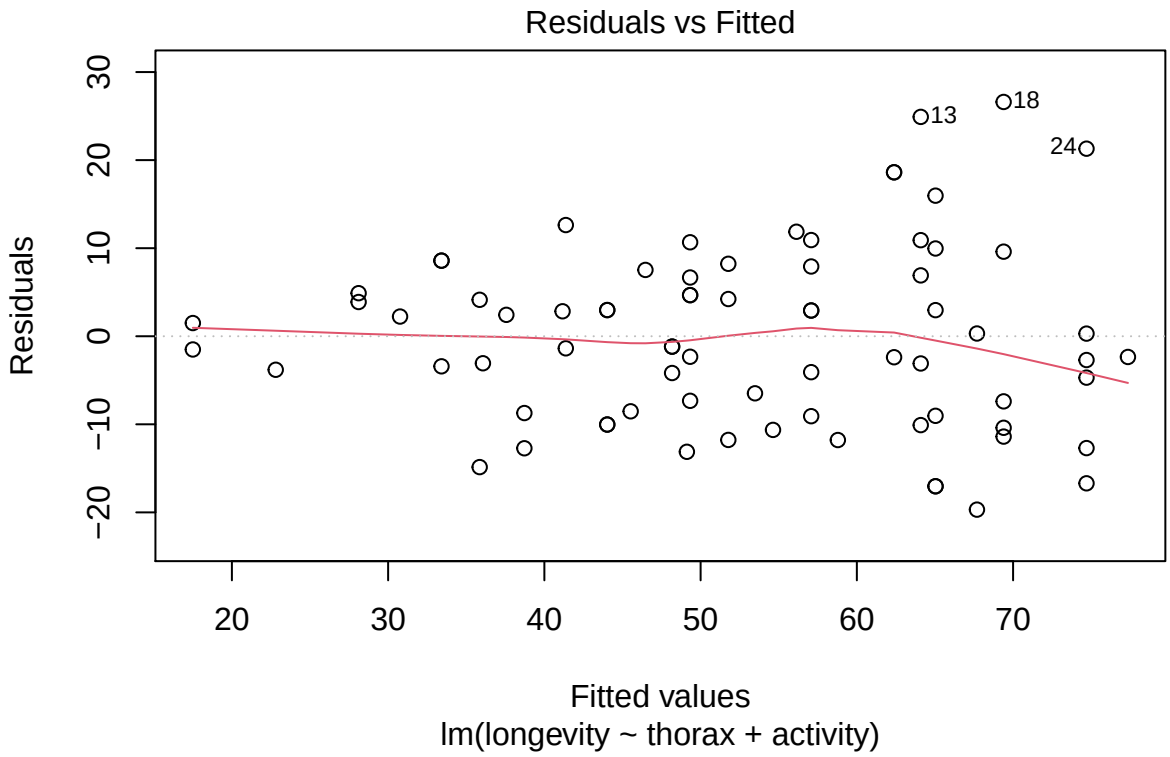

```
aov3 = lm(longevity~thorax+activity,data=data_flies)
drop1(aov3, test="F")
```

```
## Single term deletions
##
## Model:
## longevity ~ thorax + activity
##          Df Sum of Sq   RSS AIC F value  Pr(>F)
## <none>                 7673 355
## thorax    1      7687 15360 405    71.1 2.6e-12 ***
## activity  2      4967 12640 389    23.0 2.0e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

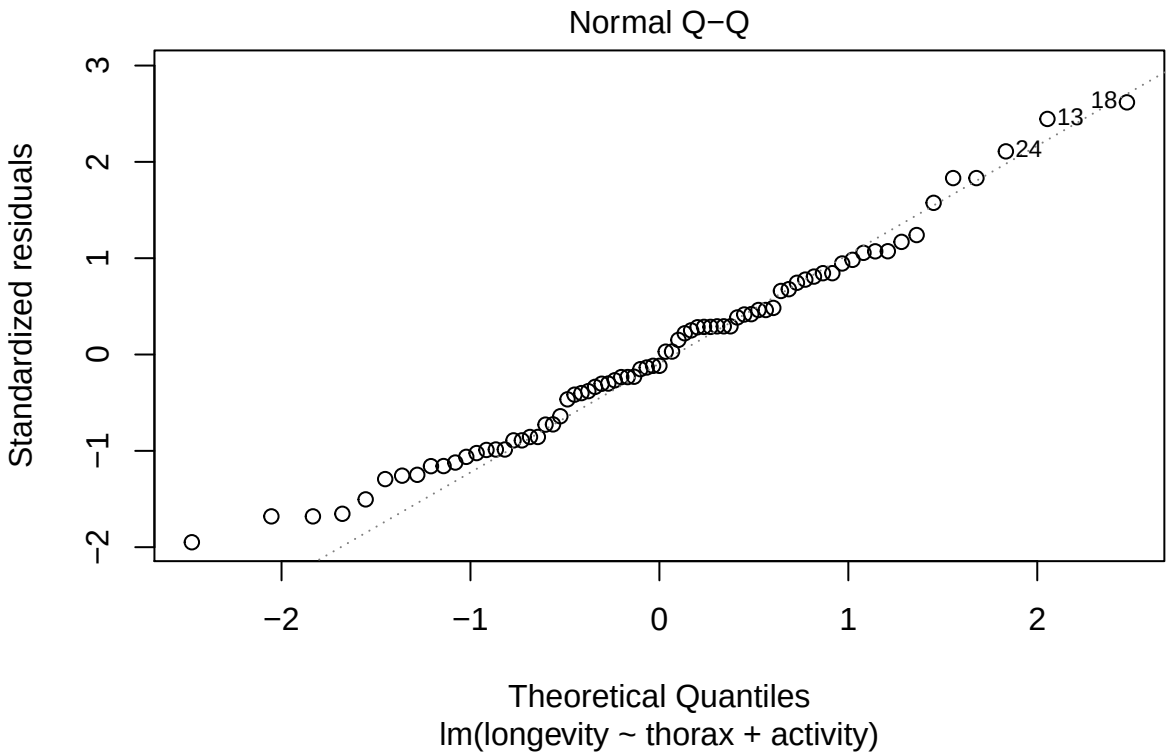
```
summary(aov3)$coefficients
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -67.4      12.75   -5.28 1.33e-06
## thorax              132.6      15.72    8.43 2.62e-12
## activityisolated     20.1       2.99    6.70 4.13e-09
## activitylow          13.1       3.00    4.35 4.43e-05
```

```
plot(aov3,1)
```


Residuals vs Fitted
Residuals
30
20
10
0
−10
−20
20
30
40
50
60
70
13
18
24
Fitted values
lm(longevity ~ thorax + activity)

```
plot(aov3,2)
```



It looks normal also with the normal data but the intercept looks off as it has a estimate has a value of -61.26 which seems to be not correct.

Exercise 2.

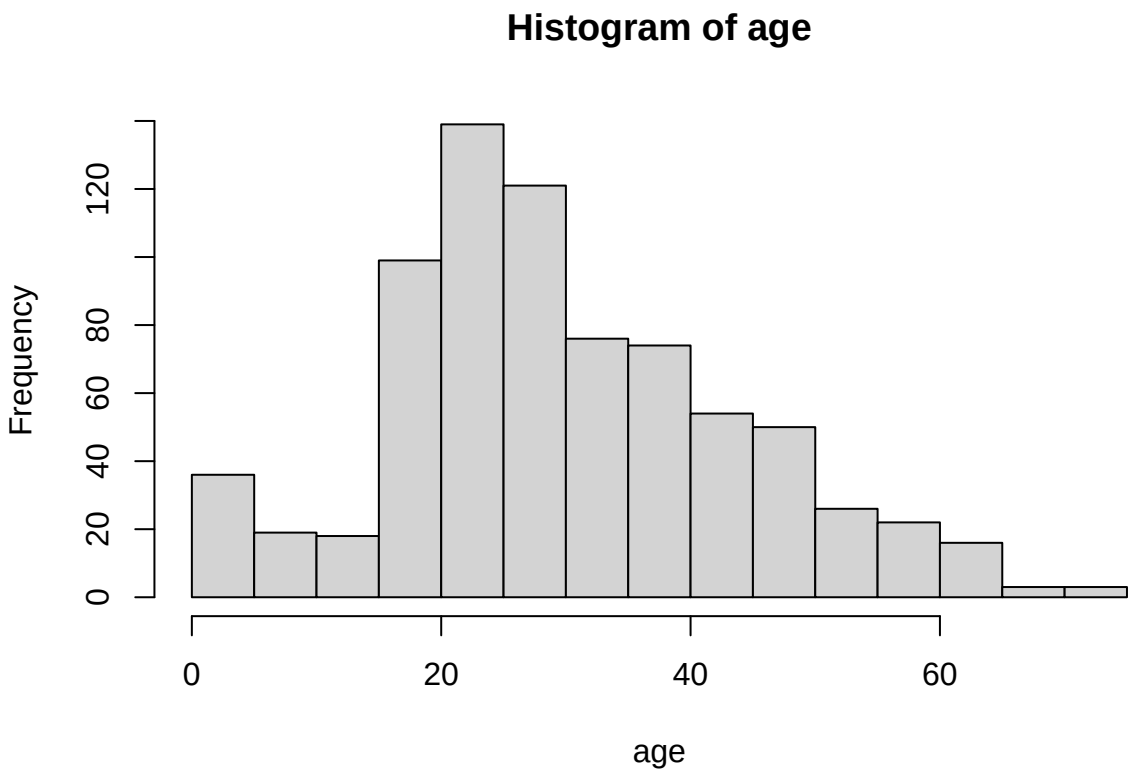
On April 15, 1912, British passenger liner Titanic sank after colliding with an iceberg. There were notenough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew. Thedata filetitanic.txtgives the survival status of passengers on the Titanic, together with their names,age, sex and passenger class. (About half of the ages for the 3rd class passengers are missing, althoughmany of these could be filled in from the original source.) The columns:Name– name of passenger;PClass– passenger class (1st, 2nd or 3rd),Age– age in years,Sex– male or female,Survived– survival status(1=Yes or 0=No)

a) Study the data and give a few (>1) summaries (graphics or tables).

```
data_titanic <- read.table(file="data/titanic.txt", header=TRUE)
data_titanic[2,]
```

```
##                          Name PClass Age    Sex Survived
## 2 Allison, Miss Helen Loraine    1st   2 female        0
```

```
age = data_titanic[,3]
hist(age)
```




```
class_sex =  xtabs(~PClass+ Sex, data=data_titanic)
class_sex
```

```
##       Sex
## PClass female male
##    1st    143  179
##    2nd    107  173
##    3rd    212  499
```

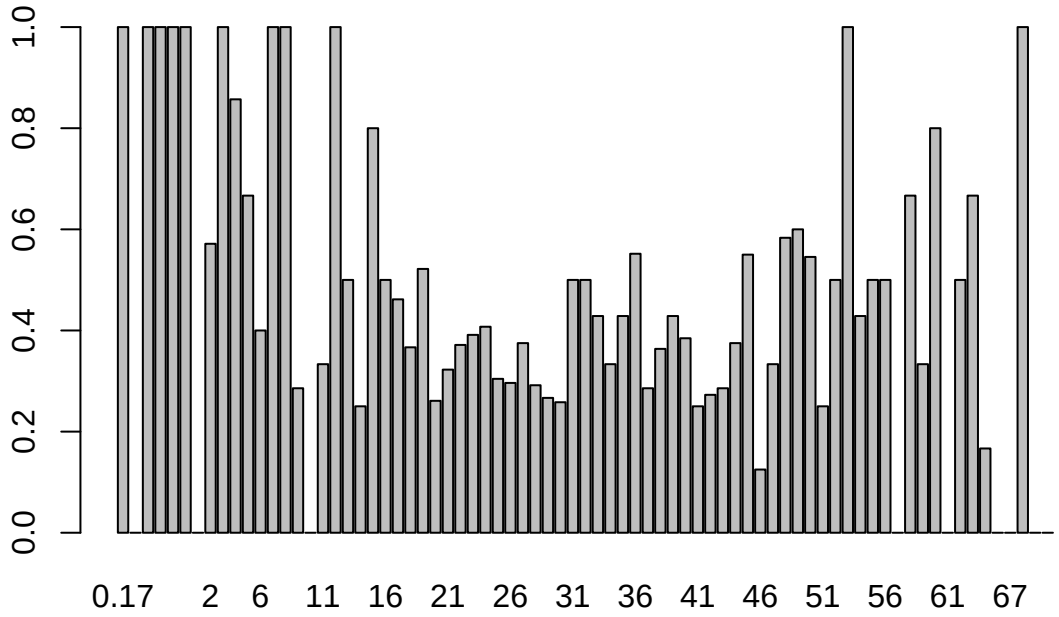
```
tot = xtabs(Survived~PClass+ Sex, data=data_titanic)
tot
```

```
##       Sex
## PClass female male
##    1st    134   59
##    2nd     94   25
##    3rd     80   58
```

```
round(tot/class_sex, 2)
```

```
##       Sex
## PClass female male
##    1st   0.94 0.33
##    2nd   0.88 0.14
##    3rd   0.38 0.12
```

```
totage=xtabs(~age,data=data_titanic)
barplot(xtabs(Survived~age, data=data_titanic)/totage)
```



The histogram shows the ages on board, while the first The table shows the total numbers of individuals for each combination of levels of class and sex. The second table shows the percentage of survivers based on sex and class. The last barplot shows the percentage per age group that survived.

(b) Fit a logistic regression model (no interactions yet) to investigate the association between the survival status and the predictors PClass,Age and Sex. Interpret the results in terms of odds, comment

```
titanicglm=glm(Survived~PClass+Sex+age,data=data_titanic,family=binomial)
titanicglm=glm(Survived~PClass,data=data_titanic,family=binomial)
titanicglm
```

```
##
## Call:  glm(formula = Survived ~ PClass, family = binomial, data = data_titanic)
##
## Coefficients:
## (Intercept)    PClass2nd    PClass3rd
##       0.403       -0.705       -1.827
##
## Degrees of Freedom: 1312 Total (i.e. Null);  1310 Residual
## Null Deviance:       1690
## Residual Deviance: 1520  AIC: 1520
```

```
summary(titanicglm)$coefficients
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.403      0.114    3.54 3.96e-04
## PClass2nd     -0.705      0.166   -4.25 2.15e-05
## PClass3rd     -1.827      0.148  -12.34 5.84e-35
```

```
drop1(titanicglm,test="Chisq")
```

```
## Single term deletions
##
## Model:
## Survived ~ PClass
##        Df Deviance  AIC LRT Pr(>Chi)
## <none>        1515 1521
## PClass  2     1688 1690 173   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
odds_male_example =  exp(3.7597 + 1 * -1.2920 + 0 * -2.5214 + 1 * -2.6314 + 25 * -0.0392)
odds_male_example
```

```
## [1] 0.319
```

```
odds_female_example =  exp(3.7597 + 0 * -1.2920 + 0 * -2.5214 + 0 * -2.6314 + 25 * -0.0392)
odds_female_example
```

```
## [1] 16.1
```

The odds can be defined as the probability of success devided by the probability of failure. The summary show that the odds of surviving is exp(3.7597 + PClass2nd * -1.2920 + PClass3rd * -2.5214 + Sexmale * -2.6314 + age * -0.0392). This means that the odds of survival while being a female in the first class is exp(3.7597 + age * -0.0392). This shows that the odds of survival for a man who is 25 years old in the 2nd class is 0.319. While a woman who is in the first class and has a age of 25 has the odds of survival of 16.1. The drop1 table also shows that all estimators are significant.

c) Investigate for interaction of predictor Age with factors PClass and Sex. From this and b), choose(and justify) a resulting model. For this model, report the estimate for the probability of survival for each combination of levels of the factors PClass and Sex for a person of age 53

```
data_titanic$age = as.numeric(data_titanic$Age)
glm3=glm(Survived~age*Sex,data=data_titanic,family=binomial);drop1(glm3,test="Chisq")
```

```
## Single term deletions
##
## Model:
## Survived ~ age * Sex
##         Df Deviance AIC LRT Pr(>Chi)
## <none>          771 779
## age:Sex  1      796 802  25  5.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(glm3)
```

```
##
## Call:
## glm(formula = Survived ~ age * Sex, family = binomial, data = data_titanic)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.126  -0.735  -0.519   0.770   2.263
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.3011     0.2990    1.01   0.3139
## age           0.0294     0.0101    2.91   0.0036 **
## Sexmale      -0.5999     0.4080   -1.47   0.1415
## age:Sexmale  -0.0657     0.0137   -4.80  1.6e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1025.57  on 755  degrees of freedom
## Residual deviance:  770.56  on 752  degrees of freedom
##   (557 observations deleted due to missingness)
## AIC: 778.6
##
## Number of Fisher Scoring iterations: 4
```

```r
glm3=glm(Survived~age*PClass,data=data_titanic,family=binomial);drop1(glm3,test="Chisq")
```

```
## Single term deletions
##
## Model:
## Survived ~ age * PClass
##           Df Deviance AIC  LRT Pr(>Chi)
## <none>             909 921
## age:PClass  2      910 918 1.17     0.56
```

```r
summary(glm3)
```

```
##
## Call:
## glm(formula = Survived ~ age * PClass, family = binomial, data = data_titanic)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.999  -0.909  -0.669   1.075   2.202
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.92298    0.43625    4.41  1.0e-05 ***
```

```
## age           -0.03584    0.00996   -3.60  0.00032 ***
## PClass2nd     -0.74428    0.57155   -1.30  0.19284
## PClass3rd     -2.29007    0.54057   -4.24  2.3e-05 ***
## age:PClass2nd -0.01321    0.01587   -0.83  0.40519
## age:PClass3rd  0.00464    0.01594    0.29  0.77090
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1025.57  on 755  degrees of freedom
## Residual deviance:  908.75  on 750  degrees of freedom
##   (557 observations deleted due to missingness)
## AIC: 920.8
##
## Number of Fisher Scoring iterations: 4
```

d) Propose a method to predict the survival status and a quality measure for your prediction and describe how you would implement that method (you do not need to implement it).

e)

```
tab1 = table(data_titanic$Survived,data_titanic$Sex)
tab1
```

```
##
##     female male
##   0    154  709
##   1    308  142
```

```
rowSums(tab1)
```

```
##   0   1
## 863 450
```

```
colSums(tab1)
```

```
## female   male
##    462    851
```

```
chisq.test(tab1)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tab1
## X-squared = 330, df = 1, p-value <2e-16
```

```
tab2 = table(data_titanic$Survived,data_titanic$PClass)
tab2
```

```
##
##     1st 2nd 3rd
##   0 129 161 573
##   1 193 119 138
```

```
rowSums(tab2)
```

```
##   0   1
## 863 450
```

```
colSums(tab2)
```

```
## 1st 2nd 3rd
## 322 280 711
```

```
chisq.test(tab2)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab2
## X-squared = 172, df = 2, p-value <2e-16
```

The contingence table shows for both gender as class a significant result.

f) Is the second approach in e) wrong? Name both an advantage and a disadvantage of the two approaches,relative to each other

An advantage of the contingency table in relative to the logistic regression is that the table is easy to read and gives a insight right away. The main advantage of a logistic regression is that non existing data can be used in order to forecast the outcome.

Exercise 3.Military coups jn Africa To study the influence of different political and geographical variables on the number of military coups, thesedata are collected for several Sub Saharan African countries in the fileafrica.txt.

a)

```
data_africa <- read.table(file="data/africa.txt", header=TRUE)

afrglm=glm(miltcoup~oligarchy+pollib+parties+pctvote+popn + size, family=poisson,data=data_africa)

summary(afrglm)$coefficients
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.112687   0.516303  -0.218 0.827228
## oligarchy    0.085962   0.025910   3.318 0.000908
## pollib      -0.689403   0.227857  -3.026 0.002481
```

```
## parties      0.029194   0.010195   2.863 0.004190
## pctvote      0.014159   0.009198   1.539 0.123723
## popn         0.006274   0.005399   1.162 0.245272
## size        -0.000195   0.000242  -0.804 0.421378
```

We see that not all precictive variables are significant. The significant variables are oliarchy, pollib and parties.

b) Use the step down approach (using output of the functionsummary) to reduce the number of explanatoryvariables. Compare the resulting model with your findings in a).

```
afrglm=glm(miltcoup~oligarchy+pollib+parties+pctvote+popn + size, family=poisson,data=data_africa)

summary(afrglm)$coefficients # Shows that size is the least significant
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.112687   0.516303  -0.218 0.827228
## oligarchy    0.085962   0.025910   3.318 0.000908
## pollib      -0.689403   0.227857  -3.026 0.002481
## parties      0.029194   0.010195   2.863 0.004190
## pctvote      0.014159   0.009198   1.539 0.123723
## popn         0.006274   0.005399   1.162 0.245272
## size        -0.000195   0.000242  -0.804 0.421378
```

```
afrglm2=glm(miltcoup~oligarchy+pollib+parties+pctvote+popn, family=poisson,data=data_africa)

summary(afrglm2)$coefficients # Shows that popn is the least significant
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.24447    0.49571  -0.493  0.62190
## oligarchy    0.08317    0.02544   3.270  0.00108
## pollib      -0.65283    0.22123  -2.951  0.00317
## parties      0.02980    0.01029   2.895  0.00379
## pctvote      0.01384    0.00928   1.491  0.13591
## popn         0.00559    0.00538   1.039  0.29883
```

```
afrglm3=glm(miltcoup~oligarchy+pollib+parties+pctvote, family=poisson,data=data_africa)

summary(afrglm3)$coefficients
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.0937    0.46328  -0.202 8.40e-01
## oligarchy     0.0954    0.02242   4.253 2.11e-05
## pollib       -0.6666    0.21756  -3.064 2.18e-03
## parties       0.0256    0.00950   2.697 6.99e-03
## pctvote       0.0121    0.00906   1.340 1.80e-01
```

```
afrglm3=glm(miltcoup~oligarchy+pollib+parties, family=poisson,data=data_africa)
```

We see that after the step down method 3 variables are significantly influential instead of 2. Y = exp(-0.0937 + oliarchy * 0.0954 + pollib * -0.6666 + parties * 0.0256 + pctvote * 0.0121)

```r
m_oligarchy = mean(data_africa$oligarchy)
m_parties = mean(data_africa$parties)
m_pctvote = mean(data_africa$pctvote)

exp(-0.0937 + m_oligarchy*0.0954 + 0*-0.6666 + m_parties*0.0256 + m_pctvote* 0.0121)
```

```
## [1] 3.42
```

```r
exp(-0.0937 + m_oligarchy*0.0954 + 1*-0.6666 + m_parties*0.0256 + m_pctvote* 0.0121)
```

```
## [1] 1.76
```

```r
exp(-0.0937 + m_oligarchy*0.0954 + 2*-0.6666 + m_parties*0.0256 + m_pctvote* 0.0121)
```

```
## [1] 0.902
```

We see that the amount of coupes decrease when the pollitical liberalization increases