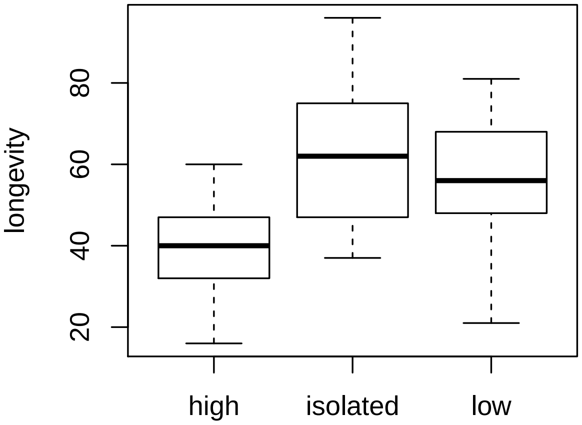
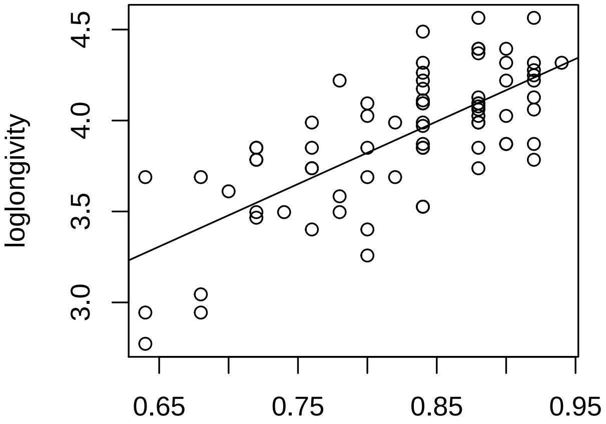
A3_EX1

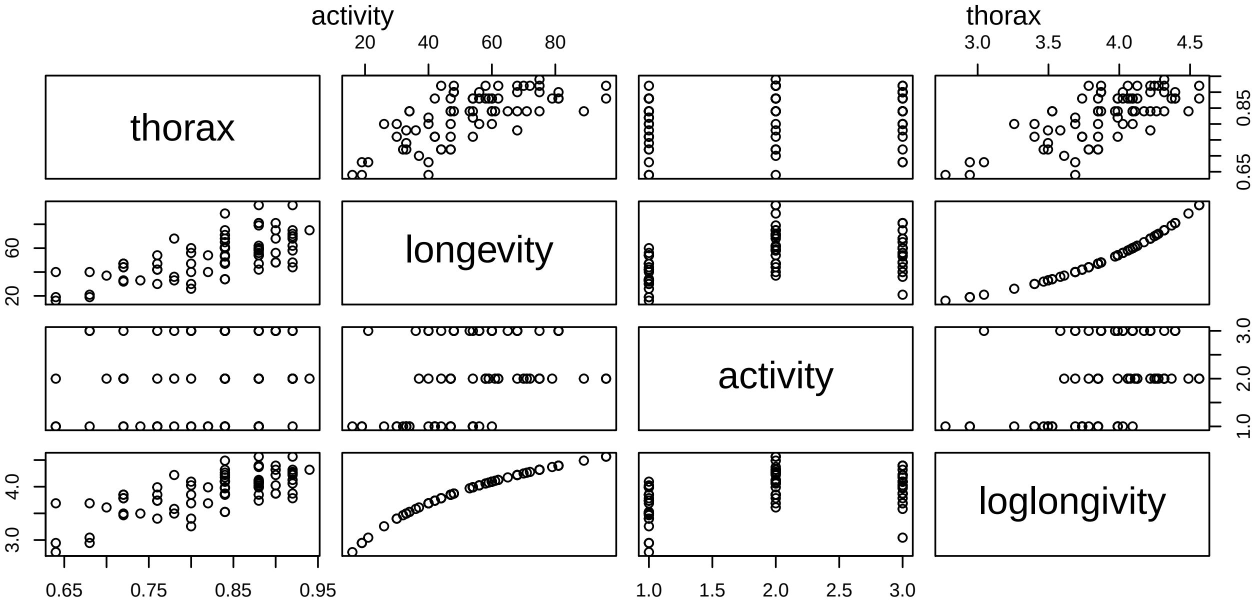
yizhen

3/13/2020

#Exercise 1 ##a) First we add a column 'loglongivity' into dataframe and use this outcome as response variable as follow, and then make some plot of the data. From the boxplot below we could see that the longivity for ftuitflies in group 'isolated' has the longest longevity, followed by 'low' and 'high'. The average longevity in group 'isolated' is the longest one. And group 'low' has the widest range of longevity in these three groups. For the scater plot, the points are distributed beside the line but quite of wide distributed, so a weak linear correlation can be seen between thorax and loglongivity. From the third plot we could see that both longevity and loglongivity have weak linear correlation with thorax. And for fruitflies in the scecond group 'isolated' can be seen gain longer longevity than the other two groups.







In order to investigate whether sexual activity influences longevity we performed one-way anova test. Null hypothesis here is sexual activity doesn’t influence the longevity. According to the p-value below, it is smaller than the significance level 0.05. Therefore we reject H_0 here which means the sexual activity will influence the longevity. According the summary below we could see that for group ‘high’ the estimated longevity is 3.60, for group ‘isolated’ is $3.60 + 0.52 = 4.12$ and for group ‘low’ is $3.60 + 0.39 = 3.99$. And 95% confidence intervals for ‘high’ is [3.48 3.72], for ‘isolated’ is [3.82, 4.41], for ‘low’ is [3.70, 4.29]

```
## Analysis of Variance Table
##
## Response: loglongivity
##           Df Sum Sq Mean Sq F value    Pr(>F)
## activity   2 3.6665  1.8333  19.421 1.798e-07 ***
## Residuals 72 6.7966  0.0944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = loglongivity ~ activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95531 -0.13338  0.02552  0.20891  0.49222
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.60212    0.06145  58.621  < 2e-16 ***
## activityisolated  0.51722    0.08690   5.952 8.82e-08 ***
## activitylow       0.39771    0.08690   4.577 1.93e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3072 on 72 degrees of freedom
## Multiple R-squared:  0.3504, Adjusted R-squared:  0.3324
## F-statistic: 19.42 on 2 and 72 DF,  p-value: 1.798e-07
```

```
##                      2.5 %    97.5 %
## (Intercept)      3.4796296 3.7246190
## activityisolated 0.3439909 0.6904582
## activitylow      0.2244780 0.5709453
```

#b) Here we apply two-way anova considering two factors: activity and thorax. H_0 here are 1) activity has no influence to longevity and 2) thorax has no influence to logevity. 3)there has no interaction between activity and thorax. From the result below, p-values for the first two null hypotheses are all smaller than 0.05, therefore we rejected the first two H_0 which means activity and thorax will influence the longevity. And p-value for the third H_0 is $0.4574 > 0.05$. Therefore, we do not reject the third H_0, which means there is no interaction betweem them. So we changed our model and fit the additive model.

```
## Analysis of Variance Table
##
## Response: loglongivity
##                 Df Sum Sq Mean Sq F value    Pr(>F)
## thorax          13 5.9900 0.46077 12.9892 3.375e-11 ***
## activity         2 2.3734 1.18670 33.4532 1.253e-09 ***
## thorax:activity 14 0.5033 0.03595  1.0135    0.4574
## Residuals       45 1.5963 0.03547
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

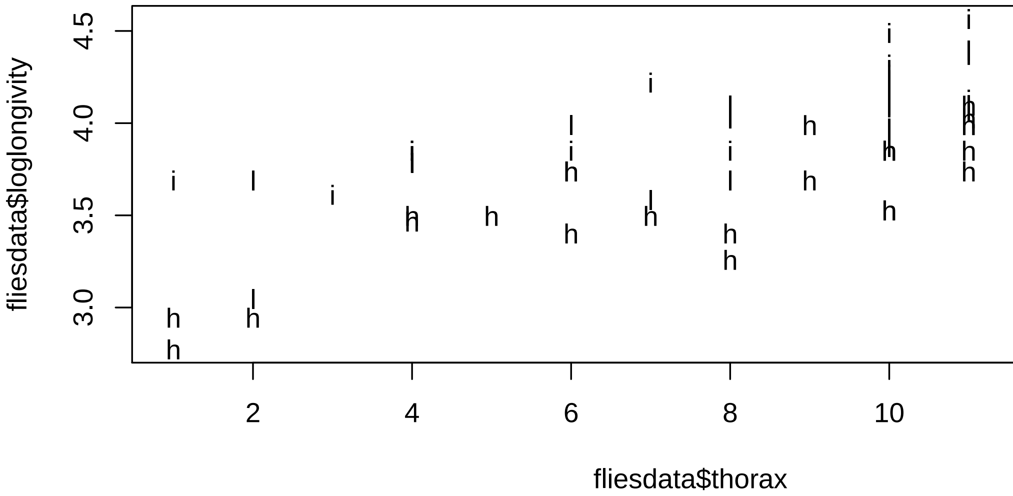
From the result below we could see that the p-values for both activity and thorax are smaller than significant level 0.05. Therefore H_0 here are rejected which means acitivity and thorax will effect the longevity. We calculated the mean of thorax equal to 0.82 and from summary we could see the estimated throax is 2.98. Therefore, estimated longevities for three groups are: ‘high’=(0.82*2.98)+1.22=3.66. ‘isolated’=(0.82*2.98)+1.22+0.41=4.07. ‘low’=(0.82*2.98)+1.22+0.29=3.95. According to the result, we conclude that the higher activity is, the shorter longevity they have, the result is similar in a).

```
## Analysis of Variance Table
##
## Response: loglongivity
##           Df Sum Sq Mean Sq F value    Pr(>F)
## thorax     1 5.4106  5.4106 131.789 < 2.2e-16 ***
## activity   2 2.1376  1.0688  26.033 3.309e-09 ***
## Residuals 71 2.9149  0.0411
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = loglongivity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45369 -0.16746  0.02622  0.15306  0.33443
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.076233   0.067582  45.519  < 2e-16 ***
## thorax           0.067422   0.006934   9.724 1.10e-14 ***
## activityisolated 0.412046   0.058321   7.065 8.92e-10 ***
## activitylow      0.287140   0.058427   4.915 5.52e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2026 on 71 degrees of freedom
## Multiple R-squared:  0.7214, Adjusted R-squared:  0.7096
## F-statistic: 61.29 on 3 and 71 DF,  p-value: < 2.2e-16
```

```
## [1] 0.8245333
```

#c) From the graph below we could see that longevity increase with the throax. Group ‘isolated’ has the longest



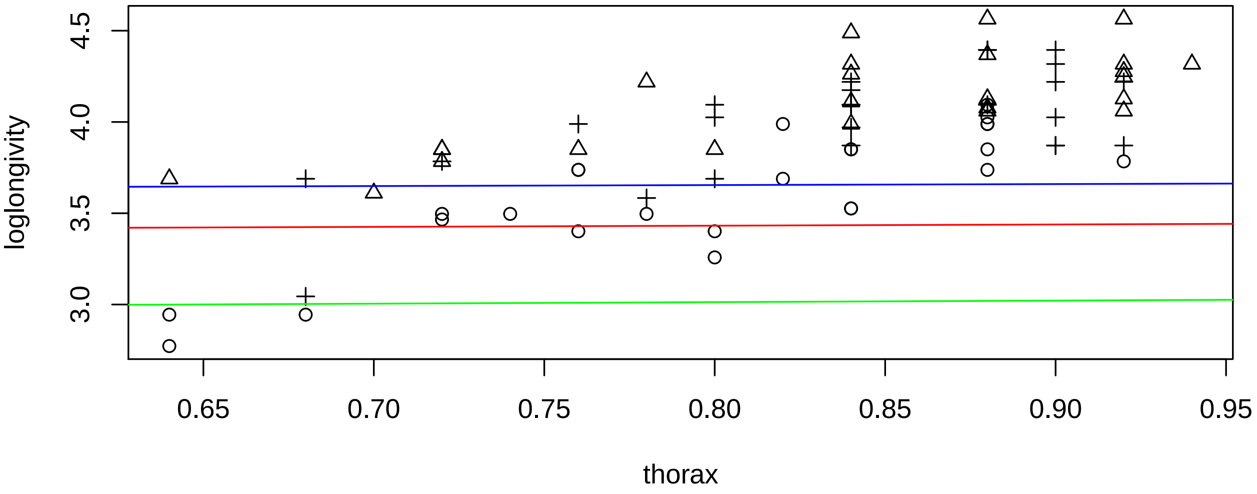
longevity, followed by ‘low’ and ‘high’.

Because thorax will influence the longevity, its dependence on activity is not so clear. Here we apply ANCOVA. Using ‘drop1’ to get the p-value. According to p-values below it confirms our analysis before that both activity and thorax will influence the longevity.

```
## Single term deletions
##
## Model:
## loglongivity ~ thorax + activity
##          Df Sum of Sq    RSS     AIC F value    Pr(>F)
## <none>                2.9149 -235.57
## thorax    1    3.8817 6.7966 -174.08  94.549 1.096e-14 ***
## activity  2    2.1376 5.0525 -198.32  26.033 3.309e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = loglongivity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45369 -0.16746  0.02622  0.15306  0.33443
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.309295   0.065781  50.307  < 2e-16 ***
## thorax       0.067422   0.006934   9.724 1.10e-14 ***
## activity1   -0.233062   0.033904  -6.874 2.00e-09 ***
## activity2    0.178984   0.033264   5.381 9.06e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2026 on 71 degrees of freedom
## Multiple R-squared:  0.7214, Adjusted R-squared:  0.7096
## F-statistic: 61.29 on 3 and 71 DF,  p-value: < 2.2e-16
```

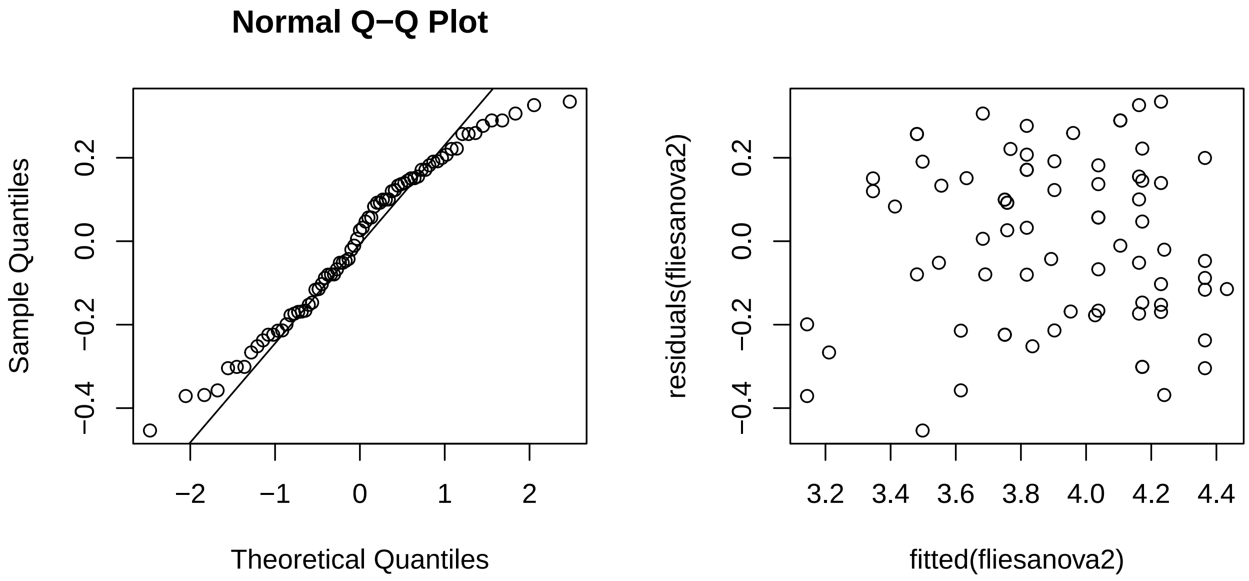
From the plot and summary below we could see that p-values for ‘isolated:thorax’ and ‘low:thorax’ are bigger than significance level 0.05, therefore we do not reject H_0 here which is there is no difference on thorax’s dependence under three activities. So the dependence is similar under all three conditions of sexual activity.



```
##
## Call:
## lm(formula = loglongivity ~ activity * thorax, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46624 -0.15549 -0.00804  0.15749  0.35592
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.312010   0.065954  50.217  < 2e-16 ***
## activity1        -0.366239   0.088099  -4.157 9.11e-05 ***
## activity2         0.298952   0.091817   3.256  0.00175 **
## thorax            0.068066   0.006929   9.823 9.69e-15 ***
## activity1:thorax  0.016082   0.009760   1.648  0.10397
## activity2:thorax -0.013751   0.009405  -1.462  0.14827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2006 on 69 degrees of freedom
## Multiple R-squared:  0.7345, Adjusted R-squared:  0.7153
## F-statistic: 38.18 on 5 and 69 DF,  p-value: < 2.2e-16
```

#d) We prefer to take thorax length into account, due to our analysis above, we know that thorax will influence the longevity of fruitflies. So it is not wise to ignore such a factor when doing analysis. But the first analysis is not wrong. At the begining, we don't know thorax's effect towards longevity and we only take one factor(activity) into account. Therefore, we apply one-way anova. They all get us right results. As the first one only focus on activities' influence to longevity and second one focus on both activity and thorax.

#e) In QQ plot we conclude that normality is ok. For the residuals versus fitted plot there is no clear pattern therefore we conclude that there is no sign of heteroscedasticity.

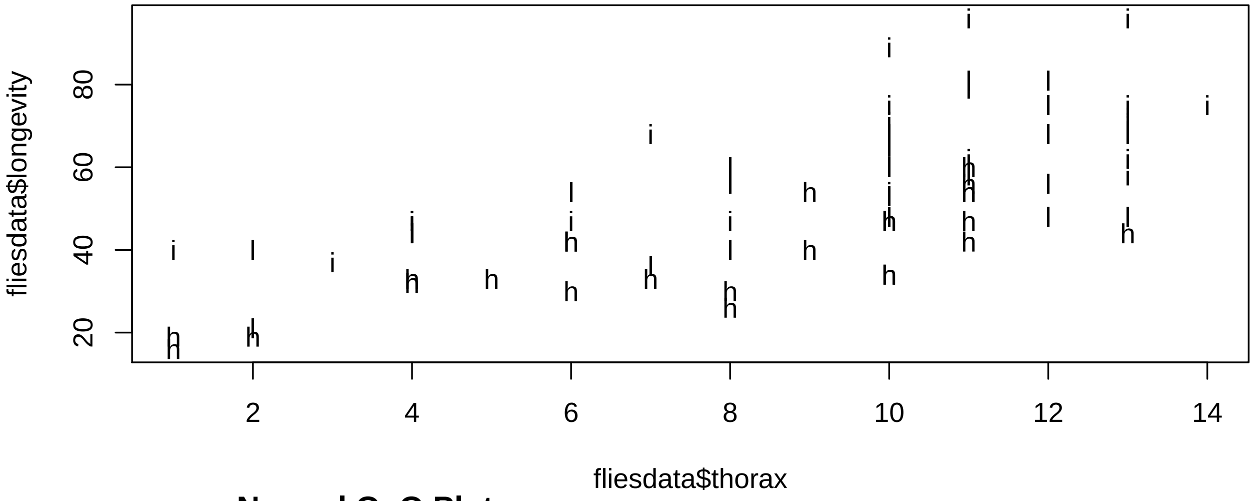


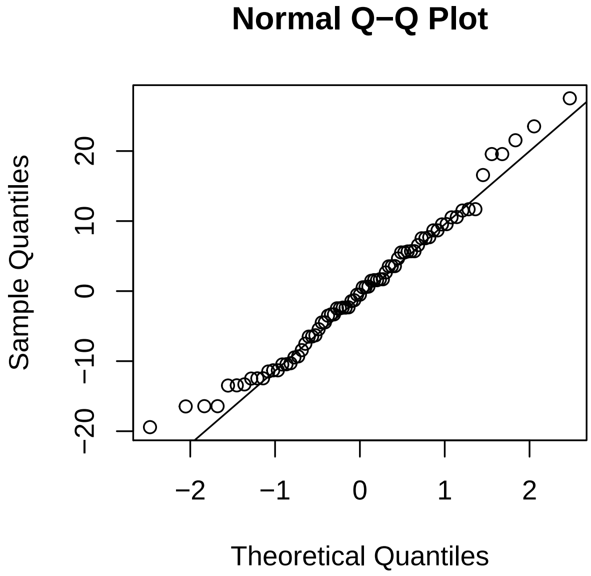
#f) We do the same ancova analysis but use longevity as response variable. From the result below we could see p-values for thorax and activity are smaller than significance level 0.05 therefore we get same conclusion as before that thorax and activity will effect fruitflies' longevity. Also we could see from the first plot that longevity increase with thorax. Then from the qq plot we could see the normality is also good. And from residuals versus fitted plot, we noticed some pattern and residuals seem to be bigger with bigger fitted values. So the inference here is, heteroscedasticity exists. In conclusion, it is wise to use the logarithm as response as we don't see heteroscedasticity in that model.

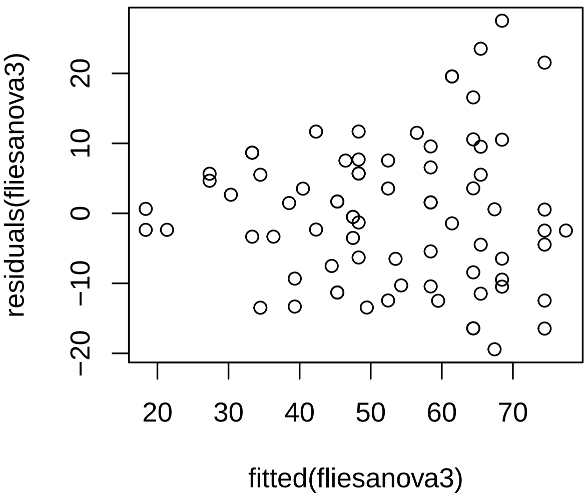
```
## Single term deletions
##
## Model:
## longevity ~ thorax + activity
##          Df Sum of Sq     RSS    AIC F value    Pr(>F)
## <none>                 7701.8 355.38
## thorax    1    7657.9 15359.8 405.15  70.596 3.003e-12 ***
## activity  2    5026.1 12727.9 389.05  23.167 1.800e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = longevity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.4210  -7.9701  -0.5086   6.1300  27.5288
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.4607     3.3813   7.826 3.52e-11 ***
## thorax        2.9947     0.3564   8.402 3.00e-12 ***
## activity1   -11.0990     1.7428  -6.369 1.65e-08 ***
## activity2     9.0693     1.7099   5.304 1.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 10.42 on 71 degrees of freedom
## Multiple R-squared:  0.6736, Adjusted R-squared:  0.6598
## F-statistic: 48.85 on 3 and 71 DF,  p-value: < 2.2e-16
```


fliesdata$longevity
fliesdata$thorax


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles


residuals(fliesanova3)
fitted(fliesanova3)