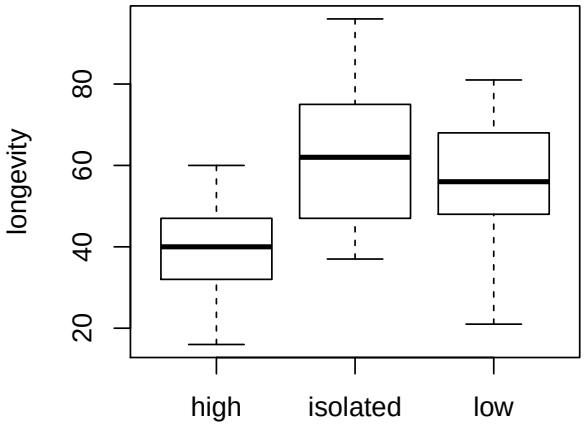
A3_EX1

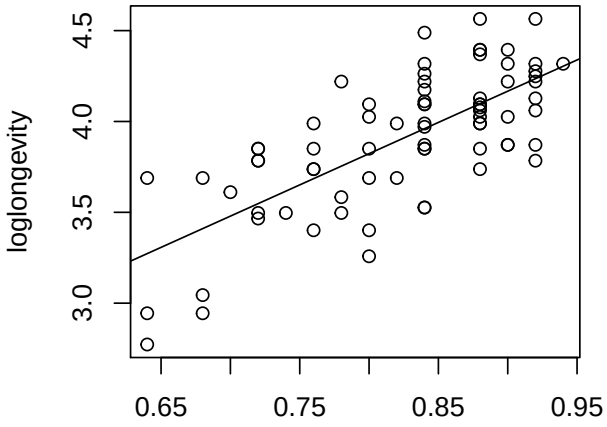
yizhen

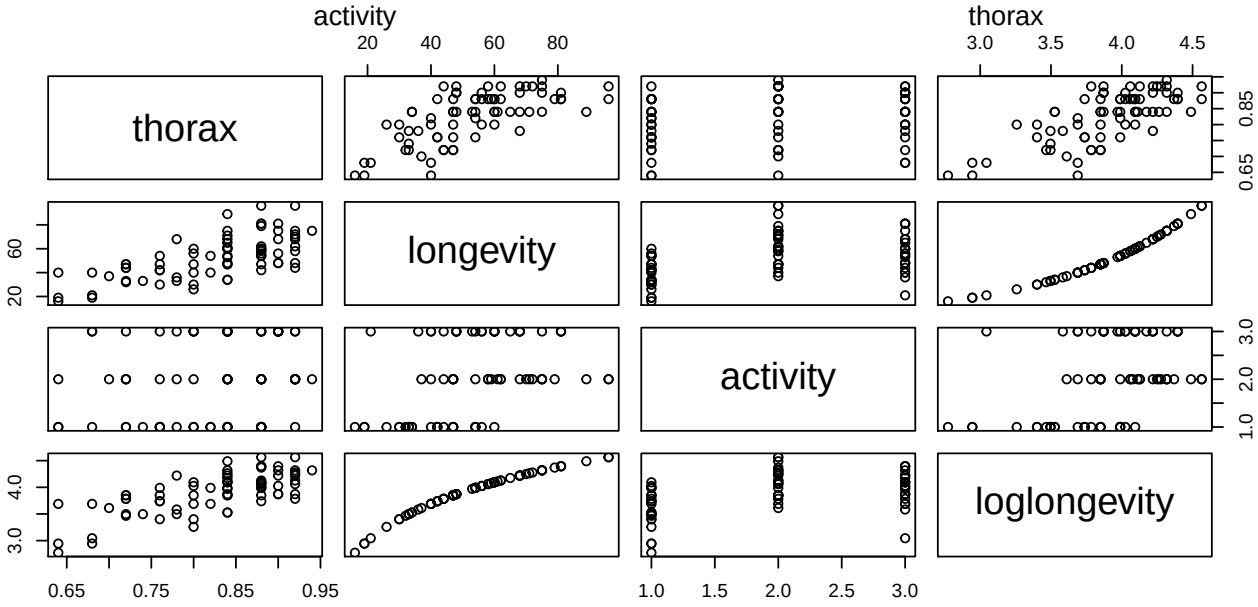
3/13/2020

Exercise 1

a) First, we add a column 'loglongevity' which will be used as a response variable. Next, we plot the longevity data in a separate boxplot for each activity. We observe that the longevity for fruitflies of the activity 'isolated' is the longest, followed by the activity 'low' and the activity 'high' has the lowest longevity. Looking at the scater plot of loglongevity and thorax, we observe a weak linear correlation. The points follow a linear pattern, however, they are relatively widely spread. Third, we could observe a weak linear correlation bewteen longevity and thorax.







In order to investigate whether sexual activity influences longevity, we performed an one-way Anova test. The null hypothesis states that sexual activity does not influence the longevity. The test results in a p-value smaller than the significance level of 0.05. Therefore, we reject H_0 and thus conclude that the sexual activity will influence the longevity. According to the summary, the estimated longevity for group 'high' is 3.60, and for group 'isolated' 3.60 + 0.52 = 4.12 and for group 'low' 3.60 + 0.39 = 3.99. With a 95% confidence interval, the logevity for 'high' is [3.48 3.72], for 'isolated' [3.82, 4.41] and for 'low' [3.70, 4.29]. From this, we confirm that a high sexual acitivty has a negative impact on the longevity.

```
##
## Call:
## lm(formula = loglongevity ~ 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) For this exercise, we apply two-way Anova with the two factors: activity and thorax. With H_0 (1) activity does not influence longevity, (2) thorax does not influence logevity, and (3) there is no interaction between activity and thorax. The output of this test shows that the p-values for the first two null hypotheses are all smaller than 0.05, therefore, we reject the first two null hypotheses. This means that activity and thorax influence the longevity. The p-value for the third null hypothesis is 0.4574. This is larger than 0.05, therefore, we do not reject the third null hypothesis. This means that there is no interaction betweem them. Then our model fit the additive model.

Now from the result we could know 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).

```
##
## Call:
## lm(formula = loglongevity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45369 -0.16746  0.02622  0.15306  0.33443
##
```

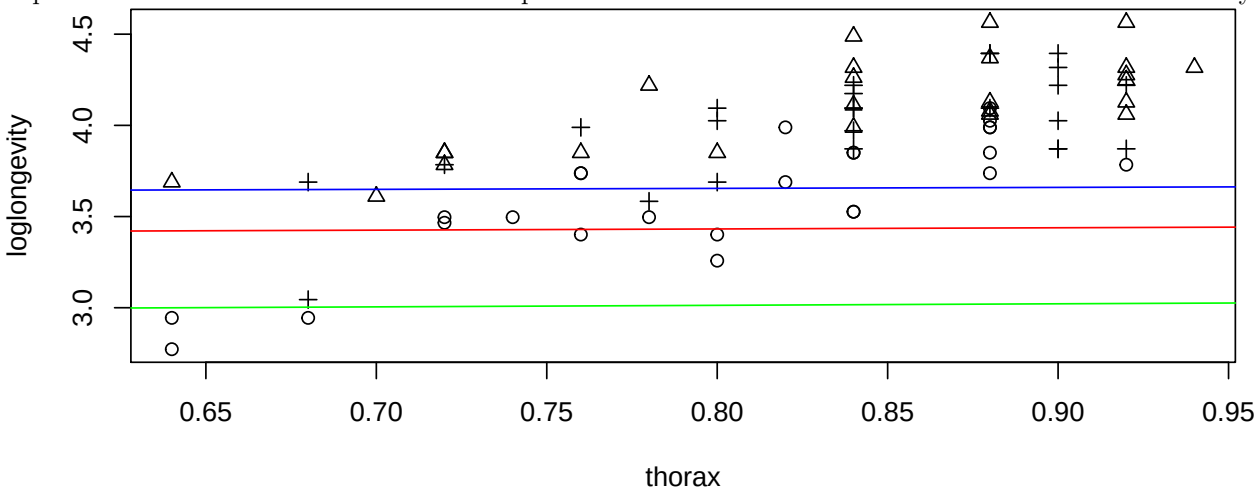
```
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.076233   0.067582  45.519  < 2e-16 ***
## thorax           0.067422   0.006934   9.724 1.10e-14 ***
## activityisolated 0.412046   0.058321   7.065 8.92e-10 ***
## activitylow      0.287140   0.058427   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
```

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 result, p-values are all less than significance level 0.05. It confirms our analysis before that both activity and thorax will influence the longevity.

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.

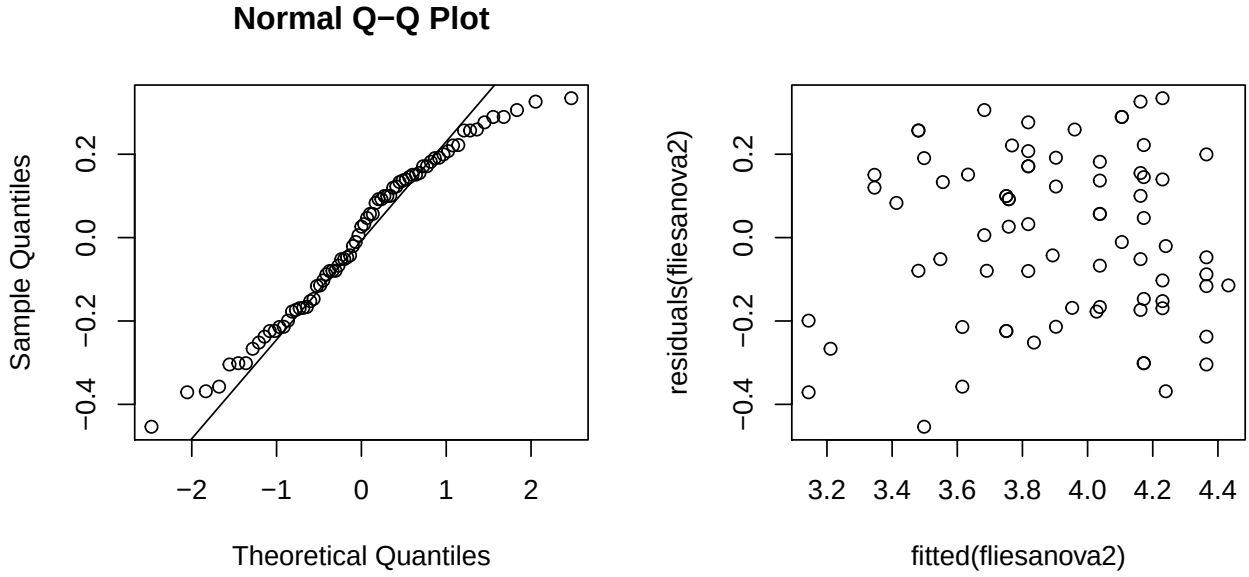



```
##
## Call:
```

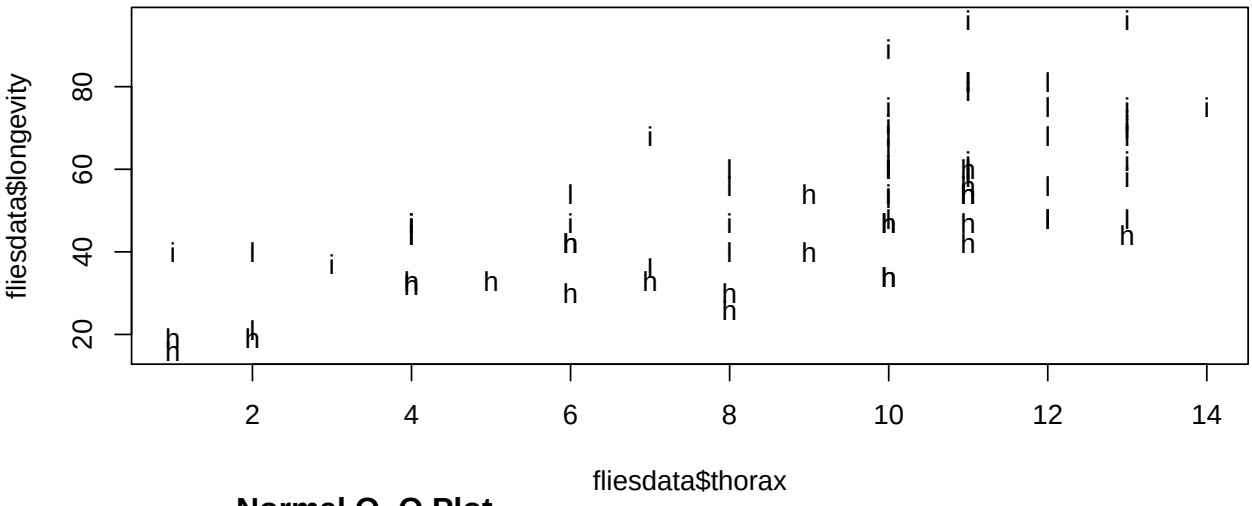
```
## lm(formula = loglongevity ~ activity * thorax, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46624 -0.15549 -0.00804  0.15749  0.35592
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.312010   0.065954  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
```

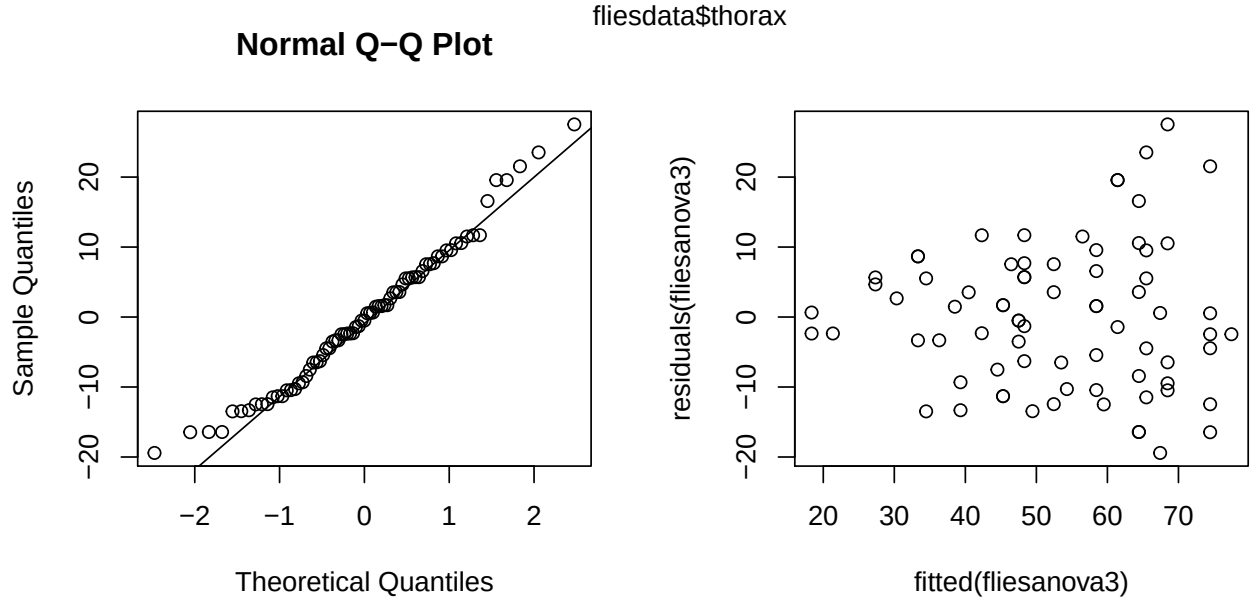
d) We prefer to take thorax length into account, due to our analysis above, we know that thorax will influence the longevity of fruitflies. So it is not wise to ignore such a factor when doing analysis. But the first analysis is not wrong. At the begining, we don't know thorax's effect towards longevity 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 we could know 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.


fliesdata$longevity
20
40
60
80
2
4
6
8
10
12
14
fliesdata$thorax


Normal Q−Q Plot
Sample Quantiles
−20
−10
0
10
20
−2
−1
0
1
2
Theoretical Quantiles
residuals(fliesanova3)
−20
−10
0
10
20
20
30
40
50
60
70
fitted(fliesanova3)