

A3_EX1

yizhen

3/13/2020

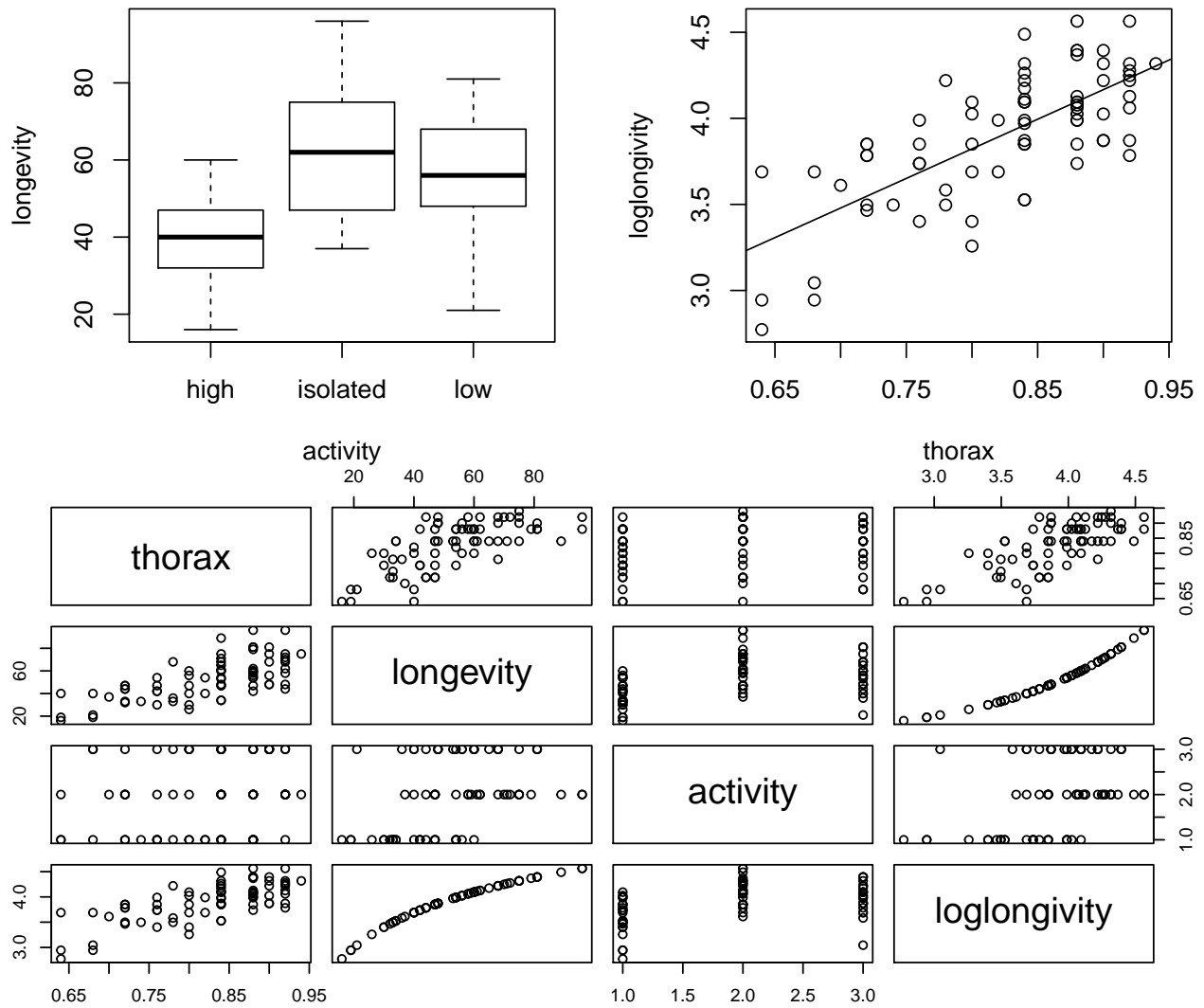
Exercise 1

a)

First we add a column 'loglongevity' 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 longevity for fruitflies 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 scatter plot, the points are distributed beside the line but quite of wide distributed, so a weak linear correlation can be seen between thorax and loglongevity. From the third plot we could see that both longevity and loglongevity have weak linear correlation with thorax. And for fruitflies in the second group 'isolated' can be seen gain longer longevity than the other two groups.

##	thorax	longevity	activity	loglongevity
## 1	0.64	40	isolated	3.688879
## 2	0.70	37	isolated	3.610918
## 3	0.72	44	isolated	3.784190
## 4	0.72	47	isolated	3.850148
## 5	0.72	47	isolated	3.850148
## 6	0.76	47	isolated	3.850148
## 7	0.78	68	isolated	4.219508
## 8	0.80	47	isolated	3.850148
## 9	0.84	54	isolated	3.988984
## 10	0.84	61	isolated	4.110874
## 11	0.84	71	isolated	4.262680
## 12	0.84	75	isolated	4.317488
## 13	0.84	89	isolated	4.488636
## 14	0.88	58	isolated	4.060443
## 15	0.88	59	isolated	4.077537
## 16	0.88	62	isolated	4.127134
## 17	0.88	79	isolated	4.369448
## 18	0.88	96	isolated	4.564348
## 19	0.92	58	isolated	4.060443
## 20	0.92	62	isolated	4.127134
## 21	0.92	70	isolated	4.248495
## 22	0.92	72	isolated	4.276666
## 23	0.92	75	isolated	4.317488
## 24	0.92	96	isolated	4.564348
## 25	0.94	75	isolated	4.317488
## 26	0.68	21	low	3.044522
## 27	0.68	40	low	3.688879
## 28	0.72	44	low	3.784190
## 29	0.76	54	low	3.988984
## 30	0.78	36	low	3.583519
## 31	0.80	40	low	3.688879
## 32	0.80	56	low	4.025352

## 33	0.80	60	low	4.094345
## 34	0.84	48	low	3.871201
## 35	0.84	53	low	3.970292
## 36	0.84	60	low	4.094345
## 37	0.84	60	low	4.094345
## 38	0.84	65	low	4.174387
## 39	0.84	68	low	4.219508
## 40	0.88	60	low	4.094345
## 41	0.88	81	low	4.394449
## 42	0.88	81	low	4.394449
## 43	0.90	48	low	3.871201
## 44	0.90	48	low	3.871201
## 45	0.90	56	low	4.025352
## 46	0.90	68	low	4.219508
## 47	0.90	75	low	4.317488
## 48	0.90	81	low	4.394449
## 49	0.92	48	low	3.871201
## 50	0.92	68	low	4.219508
## 51	0.64	16	high	2.772589
## 52	0.64	19	high	2.944439
## 53	0.68	19	high	2.944439
## 54	0.72	32	high	3.465736
## 55	0.72	33	high	3.496508
## 56	0.74	33	high	3.496508
## 57	0.76	30	high	3.401197
## 58	0.76	42	high	3.737670
## 59	0.76	42	high	3.737670
## 60	0.78	33	high	3.496508
## 61	0.80	26	high	3.258097
## 62	0.80	30	high	3.401197
## 63	0.82	40	high	3.688879
## 64	0.82	54	high	3.988984
## 65	0.84	34	high	3.526361
## 66	0.84	34	high	3.526361
## 67	0.84	47	high	3.850148
## 68	0.84	47	high	3.850148
## 69	0.88	42	high	3.737670
## 70	0.88	47	high	3.850148
## 71	0.88	54	high	3.988984
## 72	0.88	54	high	3.988984
## 73	0.88	56	high	4.025352
## 74	0.88	60	high	4.094345
## 75	0.92	44	high	3.784190



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 3.60212, for group 'isolated' is $3.60212 + 0.51722 = 4.11934$ and for group 'low' is $3.60212 + 0.39771 = 3.99983$. And 95% confidence intervals for 'high' is [3.4796296 3.7246190], for 'isolated' is [3.8236205, 4.4150772], for 'low' is [3.7041076, 4.2955643]

```
## Analysis of Variance Table
##
## Response: loglongevity
##          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 = 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)

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 longevity. 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 between them. So we changed our model and fit the additive model.

```
## Analysis of Variance Table
##
## Response: loglongevity
##              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 activity and thorax will effect the longevity. We calculated the mean of thorax equal to 0.82 and the estimated thorax is 2.98. Therefore, estimated longevity for three groups are: 'high' = $(0.82 \times 2.98) + 1.45 = 5.25$. 'isolated' = $(0.82 \times 2.98) + 1.45 - 0.23 = 5.02$. 'low' = $(0.82 \times 2.98) + 1.45 + 0.18 = 5.43$ According to the result, compared with the result in a) we could see that thorax will increase the longevity for fruitflies.

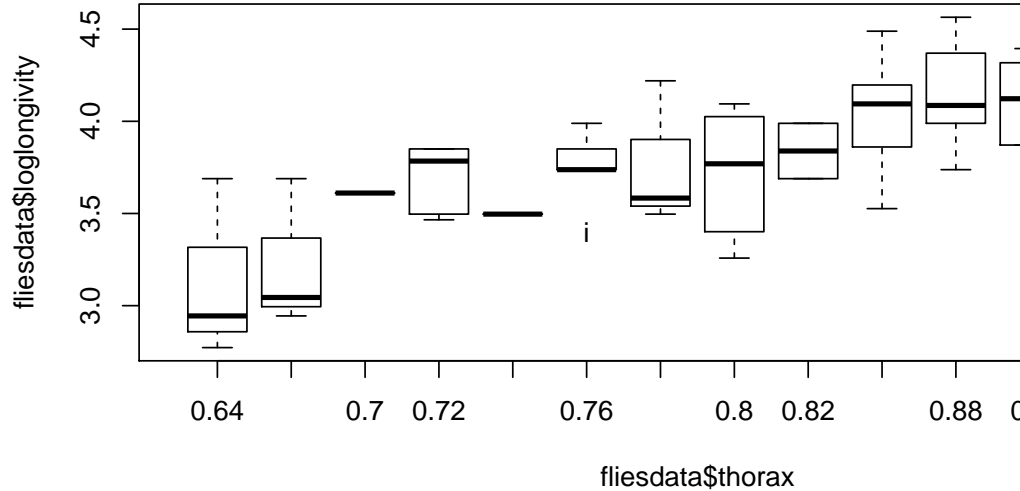
```
## Analysis of Variance Table
##
## Response: loglongevity
##              Df Sum Sq Mean Sq F value    Pr(>F)
## thorax         13  5.9900  0.46077  12.948 7.964e-13 ***
## activity        2  2.3734  1.18670  33.347 2.044e-10 ***
## Residuals     59  2.0996  0.03559
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## lm(formula = loglongevity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30309 -0.14766  0.00601  0.12938  0.34126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.981203   0.110702  26.930 < 2e-16 ***
## thorax0.68      0.001407   0.158256   0.009 0.992938
## thorax0.7       0.167417   0.221404   0.756 0.452562
## thorax0.72      0.431967   0.134520   3.211 0.002141 **
## thorax0.74      0.515305   0.218728   2.356 0.021819 *
## thorax0.76      0.596470   0.138212   4.316 6.16e-05 ***
## thorax0.78      0.509540   0.155406   3.279 0.001751 **
## thorax0.8       0.478914   0.136233   3.515 0.000851 ***
## thorax0.82      0.857729   0.173345   4.948 6.58e-06 ***
## thorax0.84      0.741746   0.121856   6.087 9.27e-08 ***
## thorax0.88      0.901735   0.120832   7.463 4.45e-10 ***
## thorax0.9       0.770324   0.144420   5.334 1.60e-06 ***
## thorax0.92      0.792761   0.129066   6.142 7.50e-08 ***
## thorax0.94      0.873987   0.221404   3.947 0.000213 ***
## activityisolated 0.462298   0.059445   7.777 1.31e-10 ***
## activitylow     0.365006   0.061959   5.891 1.96e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1886 on 59 degrees of freedom
## Multiple R-squared:  0.7993, Adjusted R-squared:  0.7483
## F-statistic: 15.67 on 15 and 59 DF, p-value: 2.69e-15
## [1] 0.8245333
```

c)

From the graph below we could see that longevity increase with the thorax. 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:
## loglongevity ~ thorax + activity
##      Df Sum of Sq  RSS   AIC F value    Pr(>F)
## <none>                 2.0996 -236.18
## thorax   13    4.6970 6.7966 -174.08  10.153 9.453e-11 ***
## activity   2    2.3734 4.4730 -183.46  33.347 2.044e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

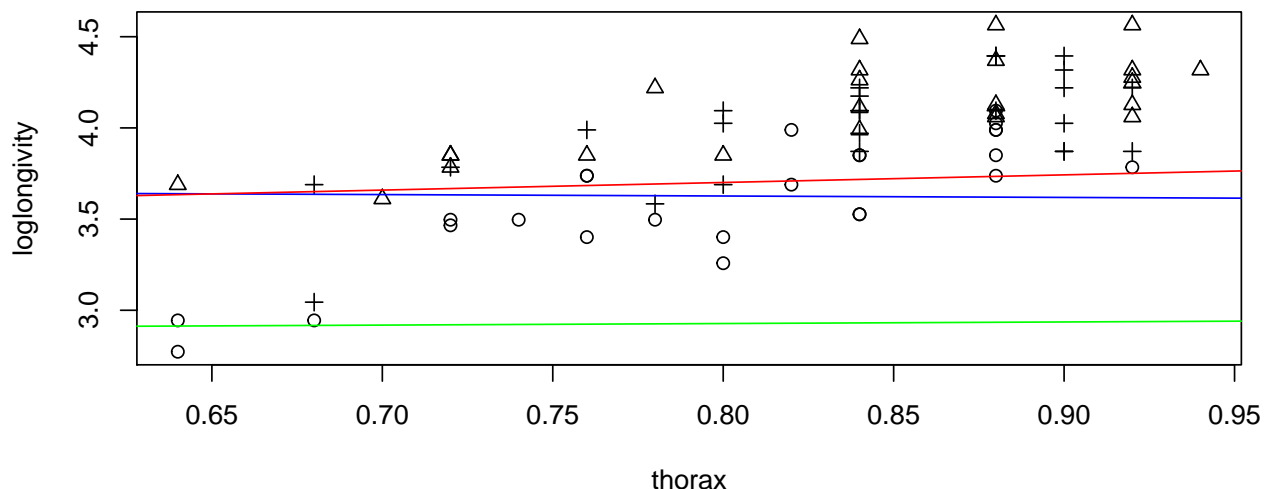
##
## Call:
## lm(formula = loglongevity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30309 -0.14766  0.00601  0.12938  0.34126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.256971   0.110855  29.380 < 2e-16 ***
## thorax0.68   0.001407   0.158256   0.009 0.992938
## thorax0.7    0.167417   0.221404   0.756 0.452562
## thorax0.72   0.431967   0.134520   3.211 0.002141 **
## thorax0.74   0.515305   0.218728   2.356 0.021819 *
## thorax0.76   0.596470   0.138212   4.316 6.16e-05 ***
## thorax0.78   0.509540   0.155406   3.279 0.001751 **
## thorax0.8    0.478914   0.136233   3.515 0.000851 ***
## thorax0.82   0.857729   0.173345   4.948 6.58e-06 ***
## thorax0.84   0.741746   0.121856   6.087 9.27e-08 ***
```

```
## thorax0.88  0.901735  0.120832  7.463 4.45e-10 ***
## thorax0.9   0.770324  0.144420  5.334 1.60e-06 ***
## thorax0.92  0.792761  0.129066  6.142 7.50e-08 ***
## thorax0.94  0.873987  0.221404  3.947 0.000213 ***
## activity1 -0.275768  0.034762 -7.933 7.11e-11 ***
## activity2  0.186530  0.034908  5.343 1.54e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1886 on 59 degrees of freedom
## Multiple R-squared:  0.7993, Adjusted R-squared:  0.7483
## F-statistic: 15.67 on 15 and 59 DF,  p-value: 2.69e-15
```

From the 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.

```
plot(loglongevity~thorax, pch=unclass(activity))
abline(lm(loglongevity~thorax, data=fliesdata[fliesdata$activity == 'isolated',]), col='blue'); abline(
```

```
## Warning in abline(lm(loglongevity ~ thorax, data = fliesdata[fliesdata
## $activity == : only using the first two of 10 regression coefficients
## Warning in abline(lm(loglongevity ~ thorax, data = fliesdata[fliesdata
## $activity == : only using the first two of 11 regression coefficients
## Warning in abline(lm(loglongevity ~ thorax, data = fliesdata[fliesdata
## $activity == : only using the first two of 9 regression coefficients
```



```
aovfliesinter = lm(loglongevity~activity*thorax, data=fliesdata)
summary(aovfliesinter)
```

```
##
## Call:
## lm(formula = loglongevity ~ activity * thorax, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3222 -0.1065  0.0000  0.1000  0.3246
##
## Coefficients: (12 not defined because of singularities)
```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.454886   0.209794  16.468 < 2e-16 ***
## activity1        -0.596372   0.248496  -2.400  0.02060 *
## activity2         0.233994   0.092412   2.532  0.01490 *
## thorax0.68       -0.182319   0.260750  -0.699  0.48802
## thorax0.7        -0.077962   0.266359  -0.293  0.77110
## thorax0.72        0.242939   0.226361   1.073  0.28889
## thorax0.74        0.637994   0.230673   2.766  0.00821 **
## thorax0.76        0.366662   0.230673   1.590  0.11894
## thorax0.78        0.311626   0.236300   1.319  0.19392
## thorax0.8         0.250443   0.226361   1.106  0.27444
## thorax0.82        0.980418   0.188344   5.205 4.63e-06 ***
## thorax0.84        0.542670   0.215509   2.518  0.01543 *
## thorax0.88        0.705706   0.216269   3.263  0.00211 **
## thorax0.9         0.299269   0.492343   0.608  0.54634
## thorax0.92        0.576883   0.203435   2.836  0.00683 **
## thorax0.94        0.628609   0.266359   2.360  0.02267 *
## activity1:thorax0.68 0.268244   0.292793   0.916  0.36447
## activity2:thorax0.68      NA         NA         NA      NA
## activity1:thorax0.7      NA         NA         NA      NA
## activity2:thorax0.7      NA         NA         NA      NA
## activity1:thorax0.72 0.379669   0.273657   1.387  0.17216
## activity2:thorax0.72 -0.103656   0.140383  -0.738  0.46412
## activity1:thorax0.74      NA         NA         NA      NA
## activity2:thorax0.74      NA         NA         NA      NA
## activity1:thorax0.76 0.400336   0.273657   1.463  0.15044
## activity2:thorax0.76 -0.205394   0.171934  -1.195  0.23850
## activity1:thorax0.78 0.326368   0.292231   1.117  0.27000
## activity2:thorax0.78 0.219003   0.179413   1.221  0.22857
## activity1:thorax0.8 0.220690   0.273657   0.806  0.42423
## activity2:thorax0.8 -0.089175   0.166104  -0.537  0.59401
## activity1:thorax0.82      NA         NA         NA      NA
## activity2:thorax0.82      NA         NA         NA      NA
## activity1:thorax0.84 0.287071   0.259108   1.108  0.27378
## activity2:thorax0.84 0.002183   0.115479   0.019  0.98500
## activity1:thorax0.88 0.383360   0.257837   1.487  0.14403
## activity2:thorax0.88 -0.154804   0.116892  -1.324  0.19208
## activity1:thorax0.9      NA         NA         NA      NA
## activity2:thorax0.9      NA         NA         NA      NA
## activity1:thorax0.92 0.348793   0.307564   1.134  0.26278
## activity2:thorax0.92      NA         NA         NA      NA
## activity1:thorax0.94      NA         NA         NA      NA
## activity2:thorax0.94      NA         NA         NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1883 on 45 degrees of freedom
## Multiple R-squared:  0.8474, Adjusted R-squared:  0.7491
## F-statistic: 8.619 on 29 and 45 DF, p-value: 1.52e-10

```


d)

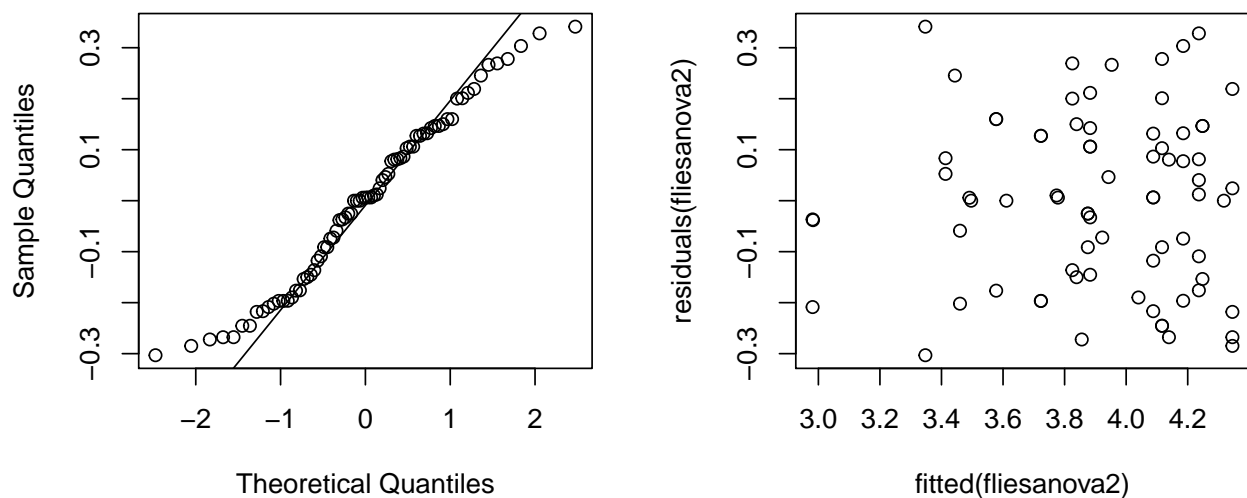
We prefer to take thorax length into account, due to our analysis above, we know that thorax will influence the longevity of fruitflies. So it is not wise to ignore such a factor when doing analysis. But the first analysis is not wrong. At the beginning, we don't know thorax's effect towards longevity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

```
par(mfrow=c(1,2))
qqnorm(residuals(fliesanova2));qqline(residuals(fliesanova2))
plot(fitted(fliesanova2), residuals(fliesanova2))
```

Normal Q-Q Plot



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.

```
fliesanova3 = lm(longevity~thorax+activity, data=fliesdata)
drop1(fliesanova3, test="F")
```

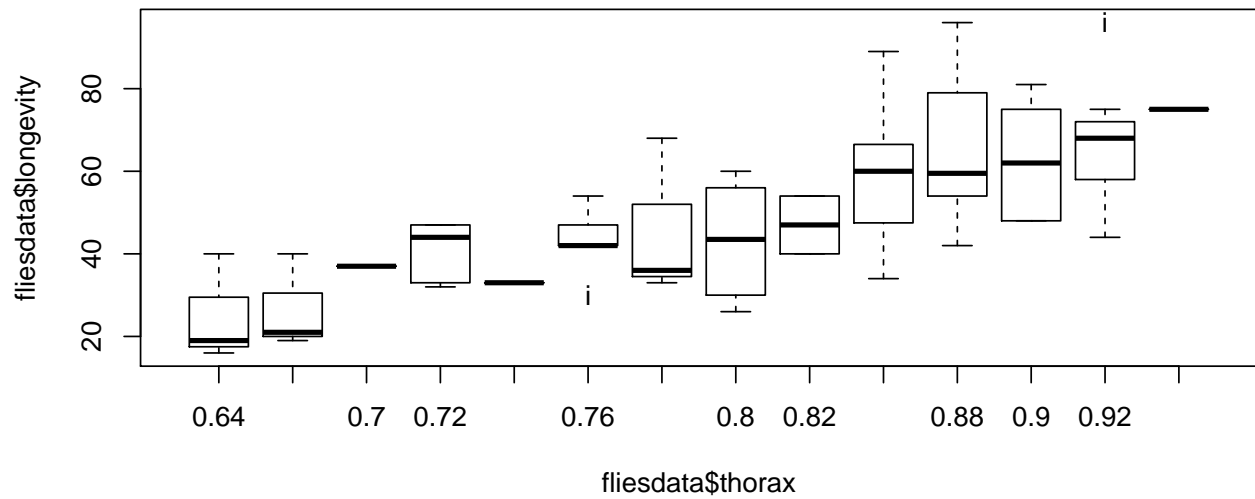
```
## Single term deletions
##
## Model:
```

```
## longevity ~ thorax + activity
##           Df Sum of Sq      RSS      AIC F value    Pr(>F)
## <none>                6560.8 367.35
## thorax    13    8799.0 15359.8 405.15   6.0867 5.053e-07 ***
## activity   2    5376.9 11937.7 408.25  24.1765 2.143e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

contrasts(fliesdata$activity)=contr.sum
fliesanova3 = lm(longevity~thorax+activity, data=fliesdata)
summary(fliesanova3)

##
## Call:
## lm(formula = longevity ~ thorax + activity, data = fliesdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.4254  -7.0082   0.0572   5.2929  26.2525
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   30.419     6.197   4.909 7.58e-06 ***
## thorax0.68    -1.700     8.846  -0.192 0.848253
## thorax0.7     -2.943    12.376  -0.238 0.812879
## thorax0.72     9.721     7.520   1.293 0.201115
## thorax0.74    15.471    12.227   1.265 0.210712
## thorax0.76    17.737     7.726   2.296 0.025260 *
## thorax0.78    15.248     8.687   1.755 0.084418 .
## thorax0.8     13.774     7.615   1.809 0.075598 .
## thorax0.82    29.471     9.690   3.041 0.003510 **
## thorax0.84    26.230     6.812   3.851 0.000292 ***
## thorax0.88    34.483     6.754   5.105 3.71e-06 ***
## thorax0.9     28.881     8.073   3.577 0.000701 ***
## thorax0.92    29.805     7.215   4.131 0.000115 ***
## thorax0.94    35.057    12.376   2.833 0.006308 **
## activity1    -12.891     1.943  -6.634 1.12e-08 ***
## activity2     9.524     1.951   4.881 8.40e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.55 on 59 degrees of freedom
## Multiple R-squared:  0.722, Adjusted R-squared:  0.6513
## F-statistic: 10.21 on 15 and 59 DF, p-value: 2.148e-11

plot(fliesdata$longevity~fliesdata$thorax, pch=as.character(fliesdata$activity))
```



```
par(mfrow=c(1,2))
qqnorm(residuals(fliesanova3));qqline(residuals(fliesanova3))
plot(fitted(fliesanova3), residuals(fliesanova3))
```

Normal Q-Q Plot

