MATH 531T-A : Exam 4

Louis Bensard

June 24, 2018

Problem 1:

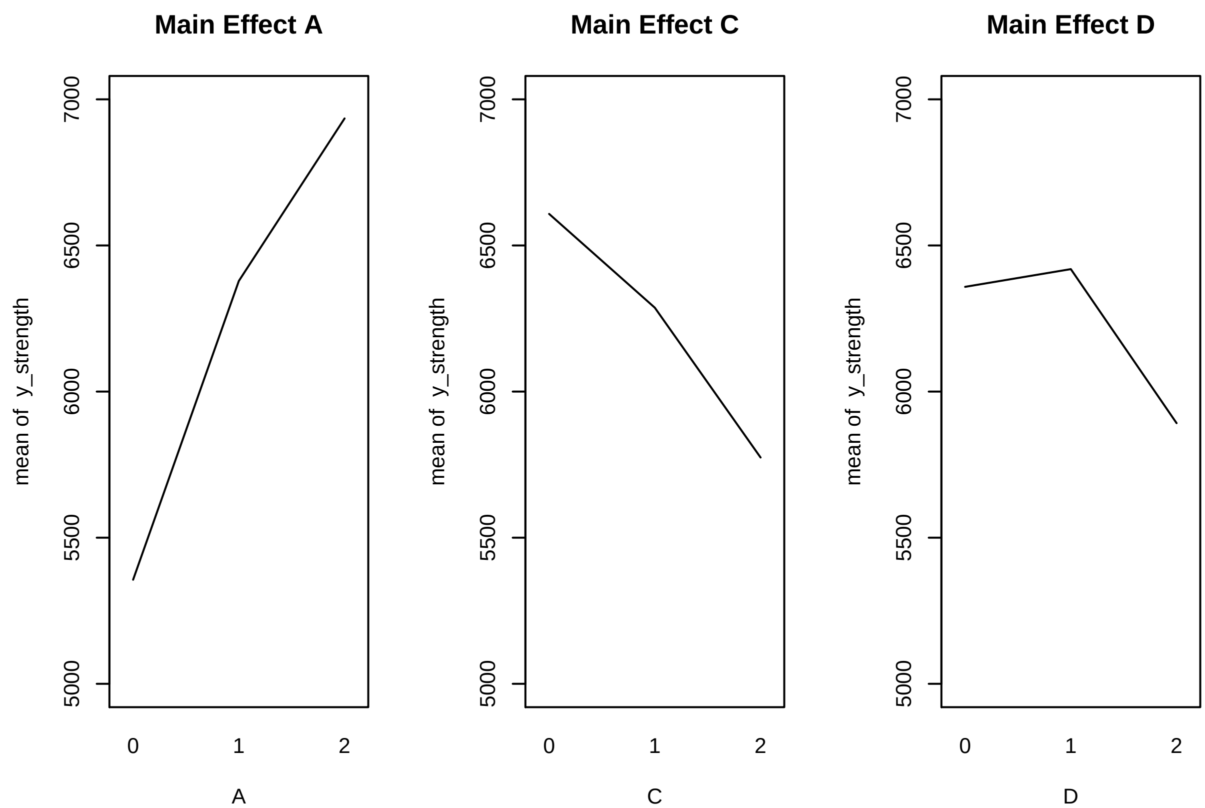
(a)

Strength location

First, let's show (at least for part (a)) how we choose the factor to consider, here is the ANOVA of the strength location model:

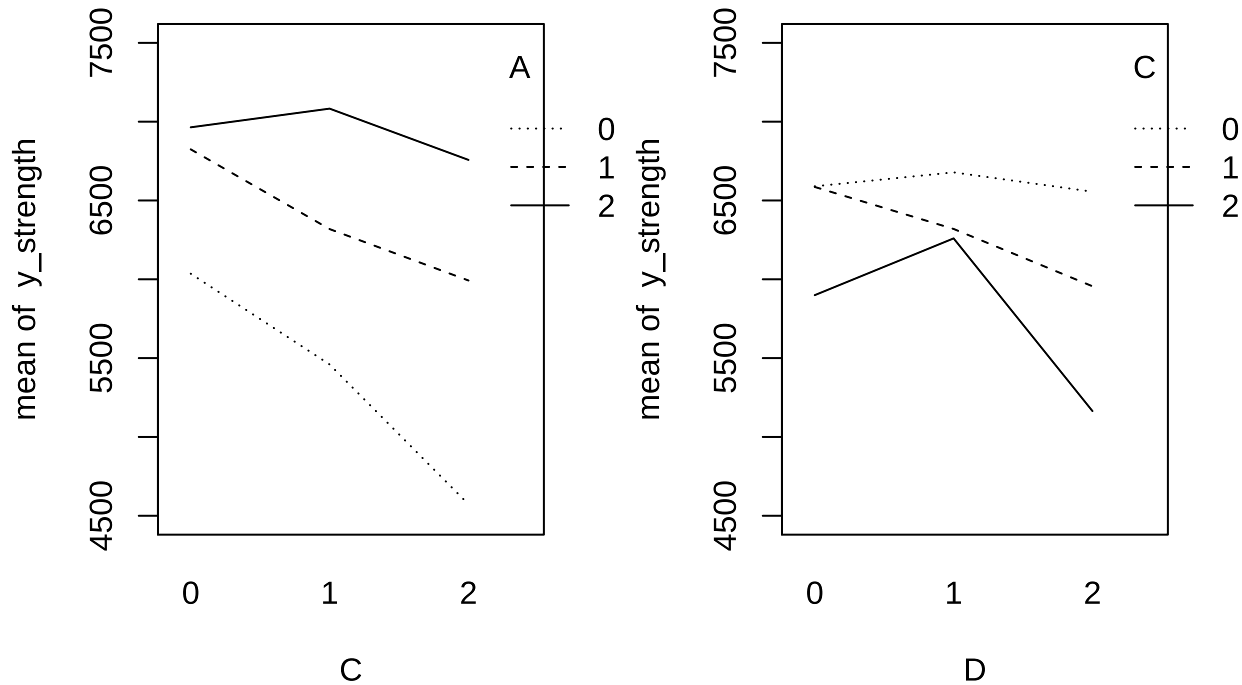
```
## Analysis of Variance Table
##
## Response: y_strength
##           Df   Sum Sq  Mean Sq F value  Pr(>F)
## A          2 34621746 17310873   85.58 < 2e-16 ***
## B          2   938539   469270    2.32  0.1080
## AB         2  2727451  1363725    6.74  0.0024 **
## AB2        2   570795   285397    1.41  0.2528
## C          2  9549481  4774741   23.61 4.3e-08 ***
## AC         2  2985591  1492796    7.38  0.0015 **
## AC2        2   886587   443294    2.19  0.1216
## BC         2   427214   213607    1.06  0.3549
## BC2        2    21134    10567    0.05  0.9491
## D          2  4492927  2246464   11.11 9.1e-05 ***
## AD         2   245439   122720    0.61  0.5488
## BD         2   205537   102768    0.51  0.6045
## CD         2   263016   131508    0.65  0.5260
## Residuals 54 10922599   202270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can clearly see that the significant main effects are A, C, D. Now let's find the optimal seeting of these factors to maximize the strength location. Here are the Main effects plots that we need:



Therefore, to maximize the strength, A needs to be at level 2, C at level 0 and D at level 1.

The table above showed that $AB = CD^2$ and $AC = BD^2$ are significant as well, since the significant main effects are A,C and D and not B, then let's take a look at the interaction plots of CD and AC:



Those plot suggest A at level 2, C at level 1 or level 0 and D at level 1. Since the main effect plot of C recommends C at level 0, C at level 0 will be the optimal setting to maximize Strength location.

So to sum up:

$$Max(Strength\,location) = A(2),\, C(0),\, D(1)$$

Strength dispersion

Since A is the only significant factor, we just have to look at the main effect plot of A for the dispersion of Strength. The plot in the book suggests to set A to level 2 to minimize dispersion.

$$Min(Strength\,dispersion) = A(2)$$

There is no conflict between maximizing location and minimizing dispersion for Strength. The optimal factor level for A in both case is 2.

(b)

Flash location

The same way as part (a), we see that factors A and C are significant, if we take a look at their main effects plots, we can see that to minimize Flash location, we should set A to level 0 and C to level 2.

The table above showed that BC^2 and $AC = BD^2$ are significant as well, since the significant main effects are A and C and not B and D, then we take a look at the interaction plots of BC and AC. Those plots suggest to set A to level 0 and C to level 2.

Thus:

$$Min(Flash\, location) = A(0),\, C(2)$$

Flash dispersion

Since A is the only significant factor, we just have to look at the main effect plot of A for the dispersion of Flash. The plot in the book suggests to set A to level 0 to minimize dispersion.

$$Min(Flash\, dispersion) = A(0)$$

There is no conflict between minimizing location and minimizing dispersion for Flash. The optimal factor level for A in both case is 0.

However, we can clearly see a conflict between maximizing Strength and Minimizing Flash as the optimal level settings for factor A are different for each response.

Problem 2:

(a)

$I = ABCD^2$, $I = AB^2C^2E$, $I = I \cdot I = ADE^2$ and $I = I \cdot I^2 = BCDE$, therefore:

$$DCS = \{I, ADE^2, ABCD^2, AB^2C^2E, BCDE\}$$

Therefore the design as resolution III.

(b)

Here are the 8 effects aliased with A:

$$\begin{aligned} A &= AB^2C^2D \\ &= BCD^2 \\ &= ABCE^2 \\ &= BCE^2 \\ &= AD^2E \\ &= DE^2 \\ &= ABCDE \\ &= AB^2C^2D^2E^2 \end{aligned}$$

(c)

These are all the aliasing relations between main effects and two-factor interaction components, only the 3-fi ADE^2 and A^2D^2E in the DCS can generate such relation

$$A = A \cdot A^2D^2E = DE^2$$
$$D = D \cdot A^2D^2E = AE^2$$
$$E = E \cdot ADE^2 = AD$$

(d)

From the previous question (which list the unclear main effects of the design), we find that only the ME B and C are clear. Then, by listing all the 2-fi's, we found that only the 2-fi BC^2 is clear.

Problem 3:

(a)

Even though the units come from the "same cell culture" for both replicates, we should block the "replicate" factor in case the variation within blocks is absmaller than the variation between blocks.

Therefore, it is a randomized block design experiment.

(b)

The dataset corresponds to the defining relation I = ABCDE. Therefore, this design has resolution 5 (from table in the book).

(c)

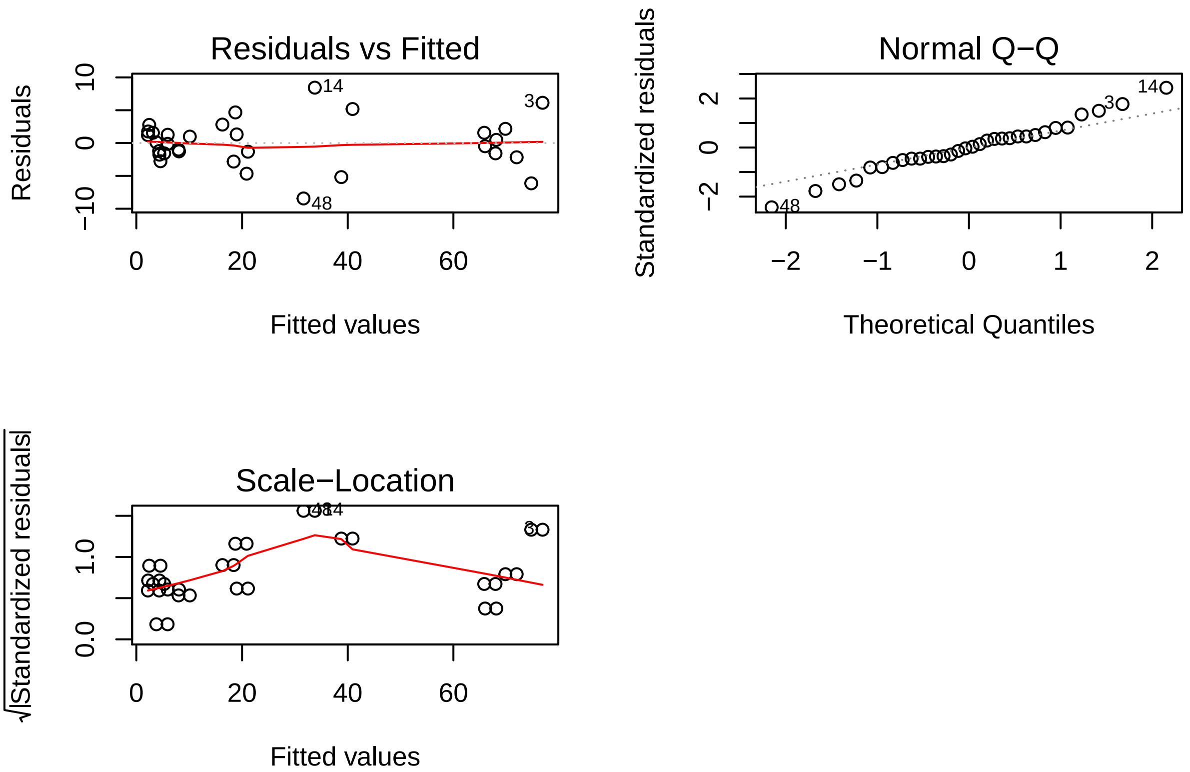
In this part, the 5 factors have only 2 levels with 16 observations, so we can afford considering the 2-fi's in the model. Then, let's first consider that model. (we will consider a main effect model only if that one does not work).

Note that even though the units come from the "same cell culture" for both replicates, we should block the "replicate" factor in case the variation within blocks is absmaller than the variation between blocks.

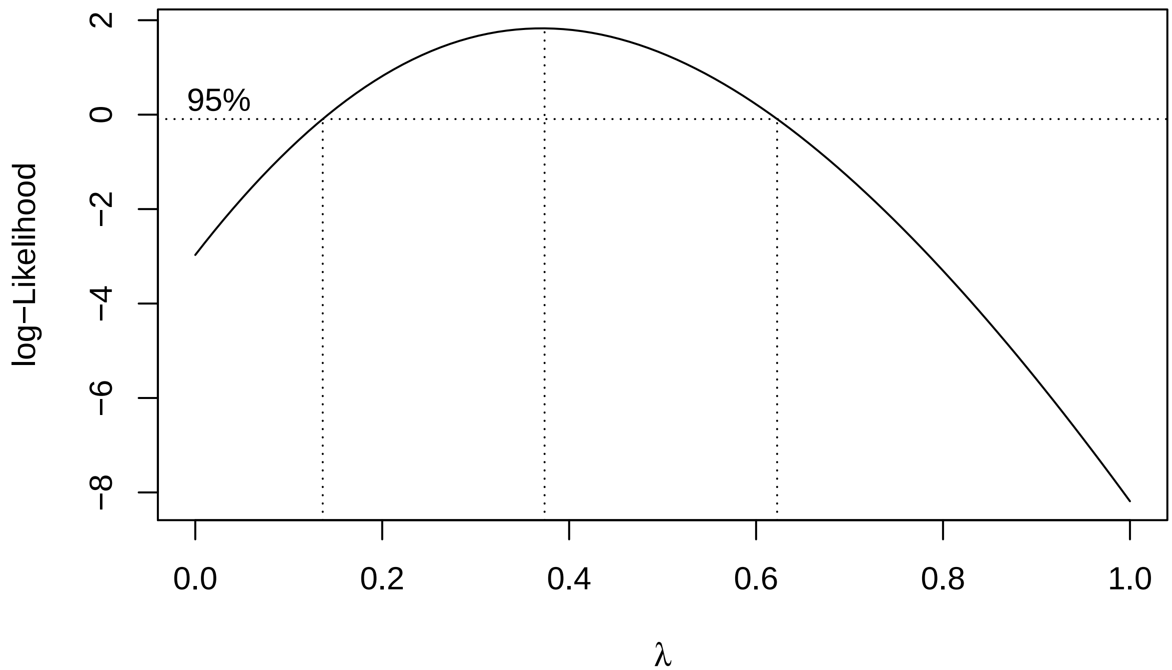
Here is the corresponding model:

```
model2 = lm(y~(A+B+C+D+E)^2 + replicate, data=data_16)
```

Here is the residual plots of the model:

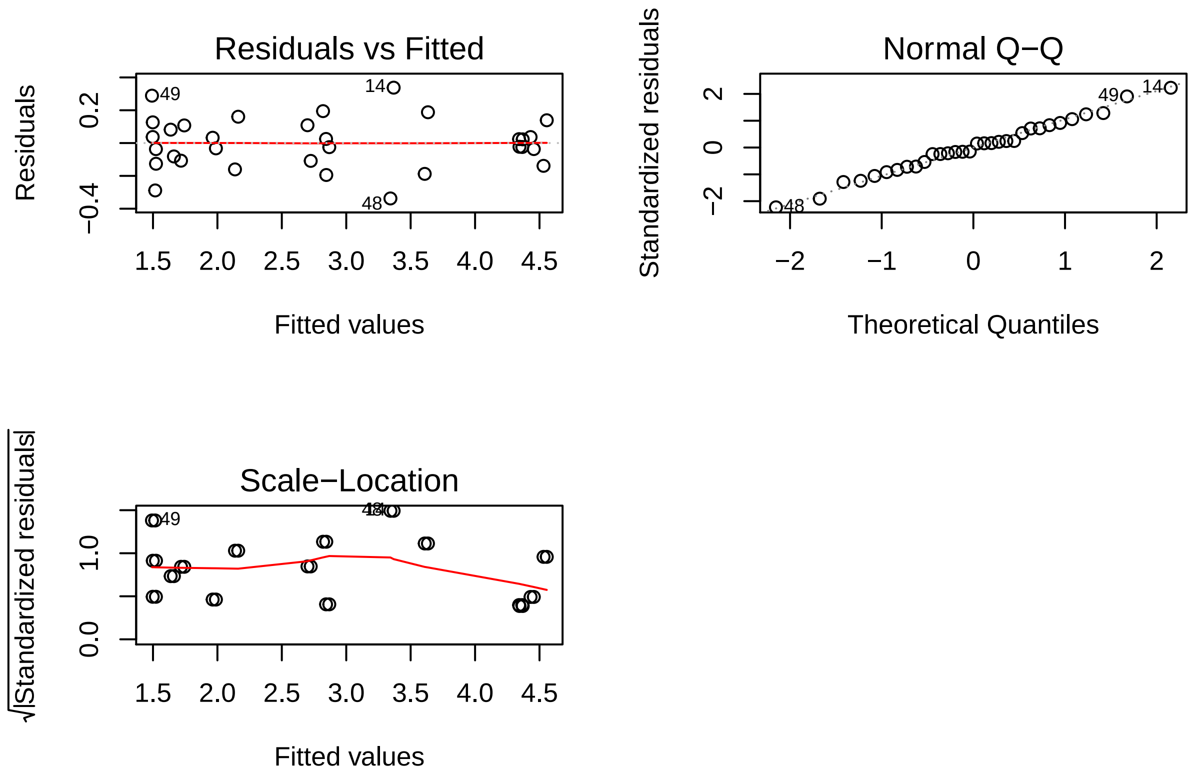


The residuals vs Fitted look really good, but the Normal QQ plot not that much, let's transform the response to see if it can improve the model:



We select $\lambda = 0.35$ to transform the response, here are the resulting residual plots:

```
## hat values (leverages) are all = 0.5312
##  and there are no factor predictors; no plot no. 5
```



We fixed the Normal QQ plot, we can now use this model for analysis of significance:

```
model4 = lm(y^(0.35)~(A+B+C+D+E)^2 + replicate, data=data_16)
```

Here is the summary of this model:

```
##
## Call:
## lm(formula = y^(0.35) ~ (A + B + C + D + E)^2 + replicate, data = data_16)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.338 -0.108  0.000  0.108  0.338
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.8192     0.0391   72.11  < 2e-16 ***
## A            -0.1125     0.0391   -2.88  0.01149 *
## B            -0.1310     0.0391   -3.35  0.00439 **
## C            -0.0683     0.0391   -1.75  0.10090
## D            -0.9613     0.0391  -24.59  1.6e-13 ***
## E            -0.4791     0.0391  -12.26  3.2e-09 ***
## replicate    -0.0126     0.0391   -0.32  0.75246
## A:B           0.0631     0.0391    1.61  0.12743
## A:C           0.0798     0.0391    2.04  0.05927 .
## A:D           0.0759     0.0391    1.94  0.07131 .
## A:E          -0.0456     0.0391   -1.17  0.26153
```

```
## B:C           0.1184     0.0391    3.03  0.00845 **
## B:D          -0.0452     0.0391   -1.16  0.26608
## B:E           0.0630     0.0391    1.61  0.12788
## C:D          -0.0660     0.0391   -1.69  0.11184
## C:E           0.0348     0.0391    0.89  0.38683
## D:E           0.1651     0.0391    4.22  0.00074 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.221 on 15 degrees of freedom
## Multiple R-squared:  0.982,  Adjusted R-squared:  0.963
## F-statistic: 51.5 on 16 and 15 DF,  p-value: 3.31e-10
```

First, note that the replicate block factor is clearly not significant (not surprising). We found that factors A,B,D and E are significant, as well as interaction BC and DE.

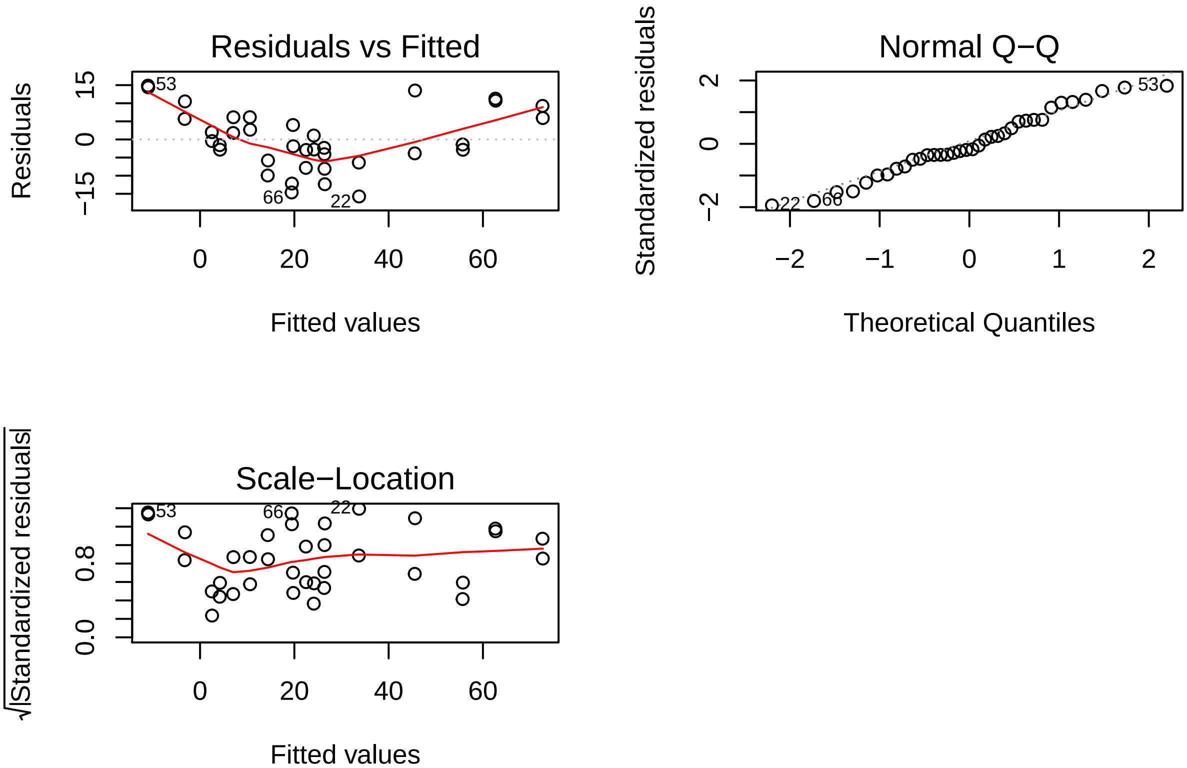
(d)

In this part, the 5 factors have 3 levels with only 18 observations, so we cannot afford to consider the 2-fi's in the model but we can afford the quadratic components of the main effects. Here is the model we're starting with:

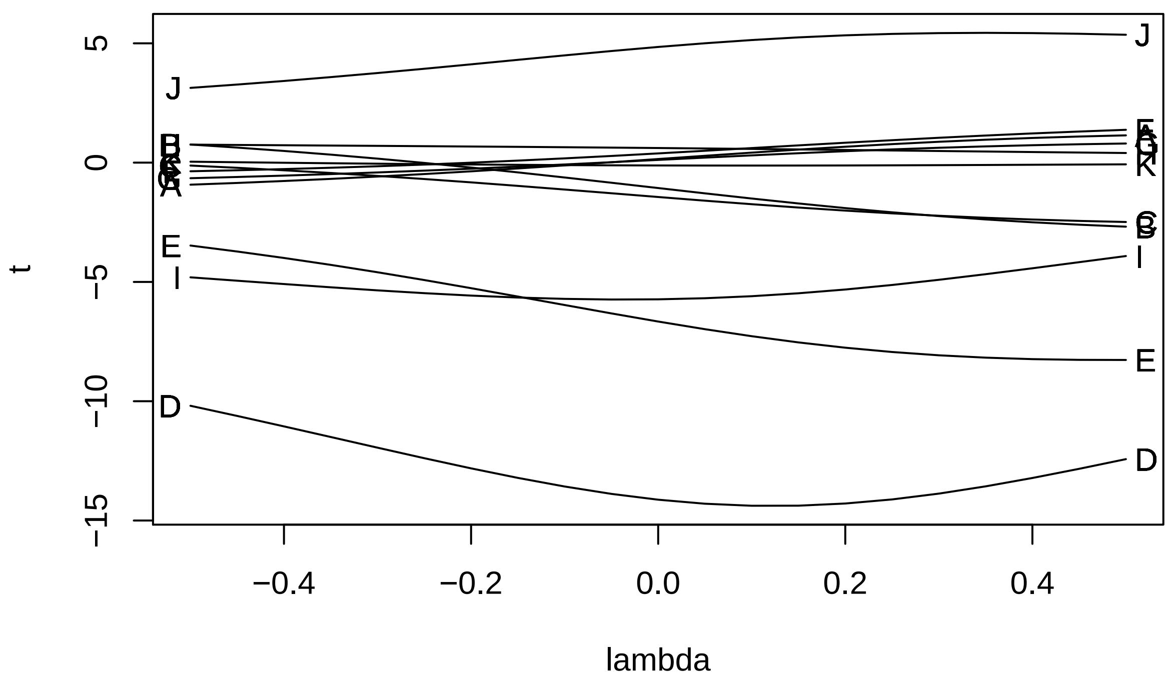
```r
model1 = lm(y~A+B+C+D+E + I(A^2) + I(B^2) + I(C^2) + I(D^2) + I(E^2) +replicate, data=data_18)
```

Here are its corresponding residual plots:

```
## hat values (leverages) are all = 0.3333
##  and there are no factor predictors; no plot no. 5
```



The residual vs Fitted is clearly bad. Let's use lambdaPlot on the model to find a good transformation of the response:



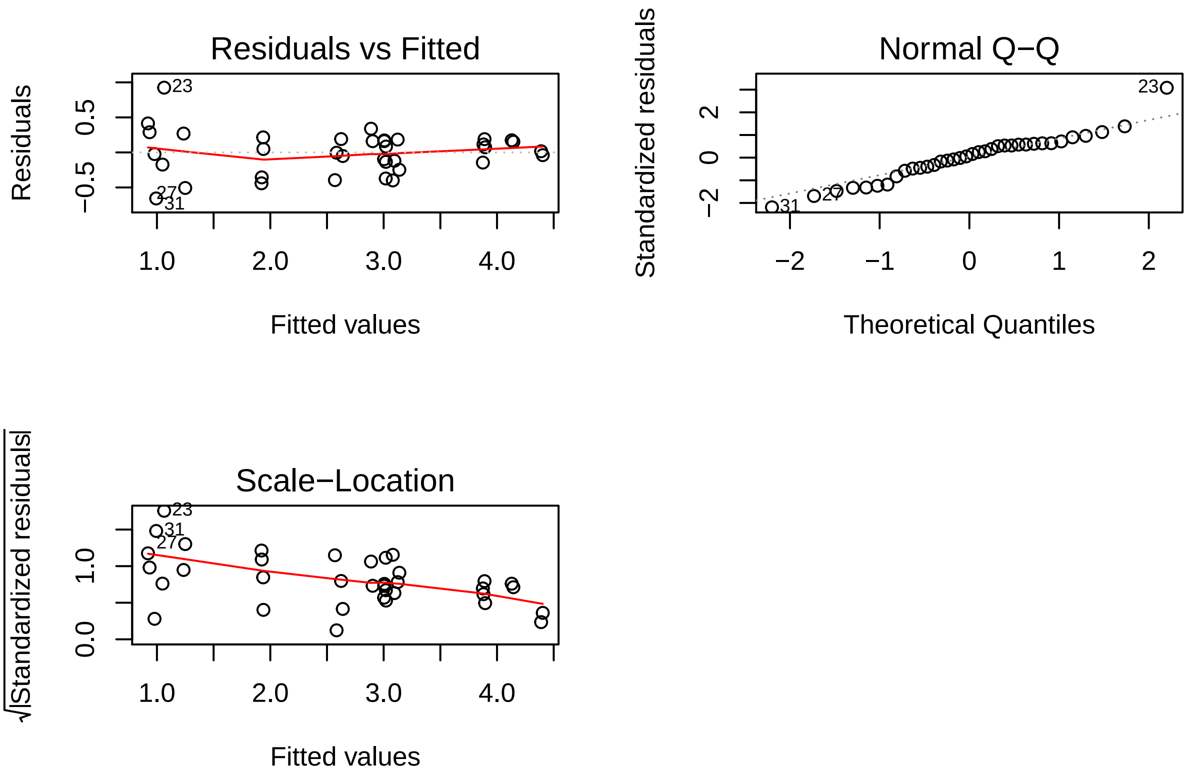

```
##           term label
## 1  (Intercept)
## 2            A     A
## 3            B     B
## 4            C     C
## 5            D     D
## 6            E     E
## 7     `I(A^2)`     F
## 8     `I(B^2)`     G
## 9     `I(C^2)`     H
## 10    `I(D^2)`     I
## 11    `I(E^2)`     J
## 12   replicate     K
```

λ is very close to 0, we will then use a log transformation on the response, here is the new model and its residual plots:

```
model3 = lm(log(y)~A+B+C+D+E+I(A^2) + I(B^2) + I(C^2) + I(D^2) + I(E^2)+replicate, data=data_18)
```

```
par(mfrow = c(2,2))
plot(model3)
```

```
## hat values (leverages) are all = 0.3333
##  and there are no factor predictors; no plot no. 5
```



The residual plots are fixed, we can then use this model for our analysis:

```
##
## Call:
## lm(formula = log(y) ~ A + B + C + D + E + I(A^2) + I(B^2) + I(C^2) +
##     I(D^2) + I(E^2) + replicate, data = data_18)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6570 -0.1520  0.0325  0.1756  0.9242
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.63216    0.20291   12.97  2.5e-12 ***
## A            0.01130    0.07493    0.15     0.88
## B           -0.07984    0.07493   -1.07     0.30
## C           -0.10792    0.07493   -1.44     0.16
## D           -1.05855    0.07493  -14.13  4.0e-13 ***
## E           -0.49900    0.07493   -6.66  6.9e-07 ***
## I(A^2)       0.05015    0.12978    0.39     0.70
## I(B^2)       0.01479    0.12978    0.11     0.91
## I(C^2)       0.07946    0.12978    0.61     0.55
## I(D^2)      -0.74346    0.12978   -5.73  6.7e-06 ***
## I(E^2)       0.62913    0.12978    4.85  6.1e-05 ***
## replicate   -0.00727    0.06118   -0.12     0.91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.367 on 24 degrees of freedom
## Multiple R-squared:  0.927,  Adjusted R-squared:  0.893
## F-statistic: 27.6 on 11 and 24 DF,  p-value: 5.87e-11
```

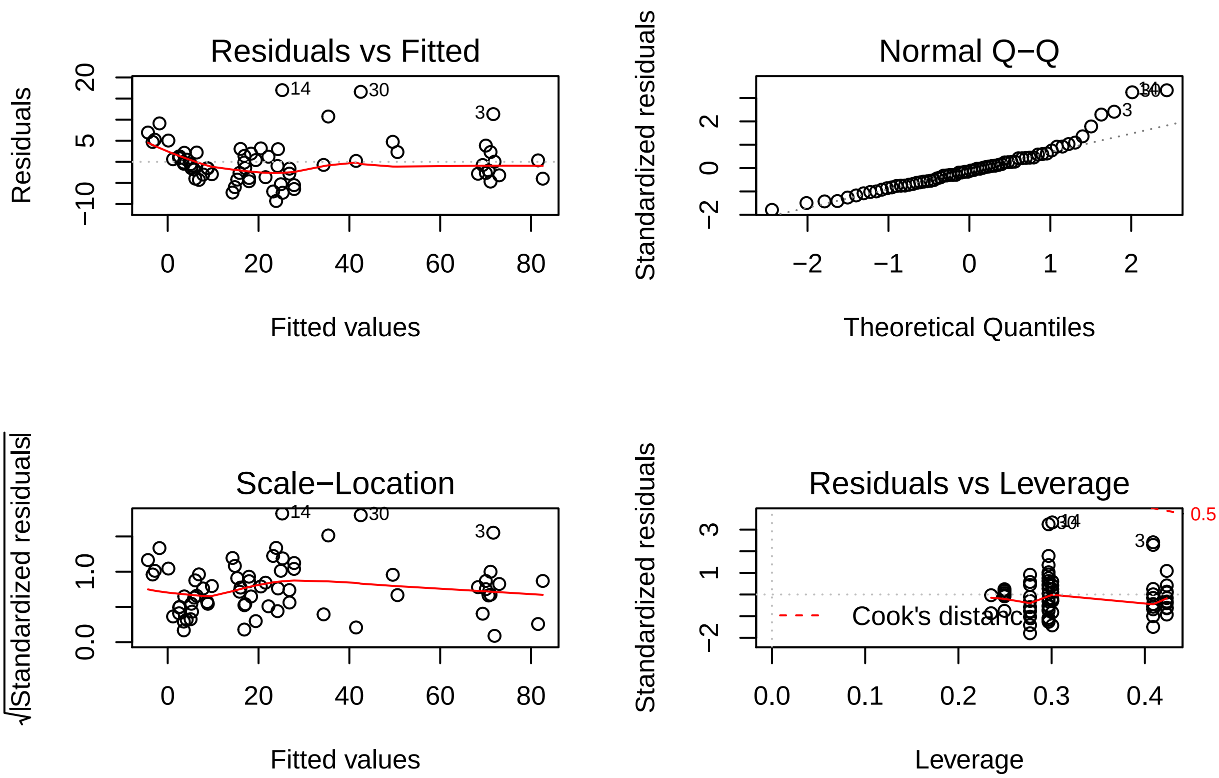
The linear and quadratic components of factor D and E are significant here.

(e)

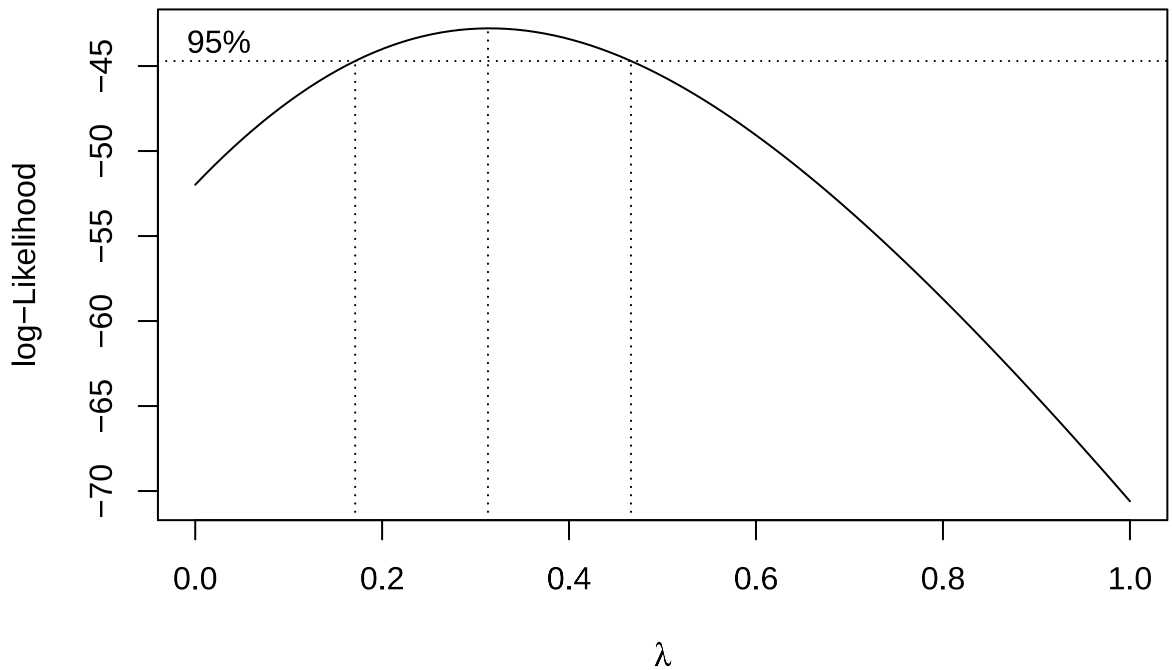
In this part, we have enough observations so we can afford considering the main effects (along with quadratic components) and the 2-fi's:

```
model5 = lm(y~(A + B + C + D + E)^2 + I(A^2) + I(B^2) + I(C^2) + I(D^2) + I(E^2) + replicate,
            data=data_all)
```

Here are its residual plots:



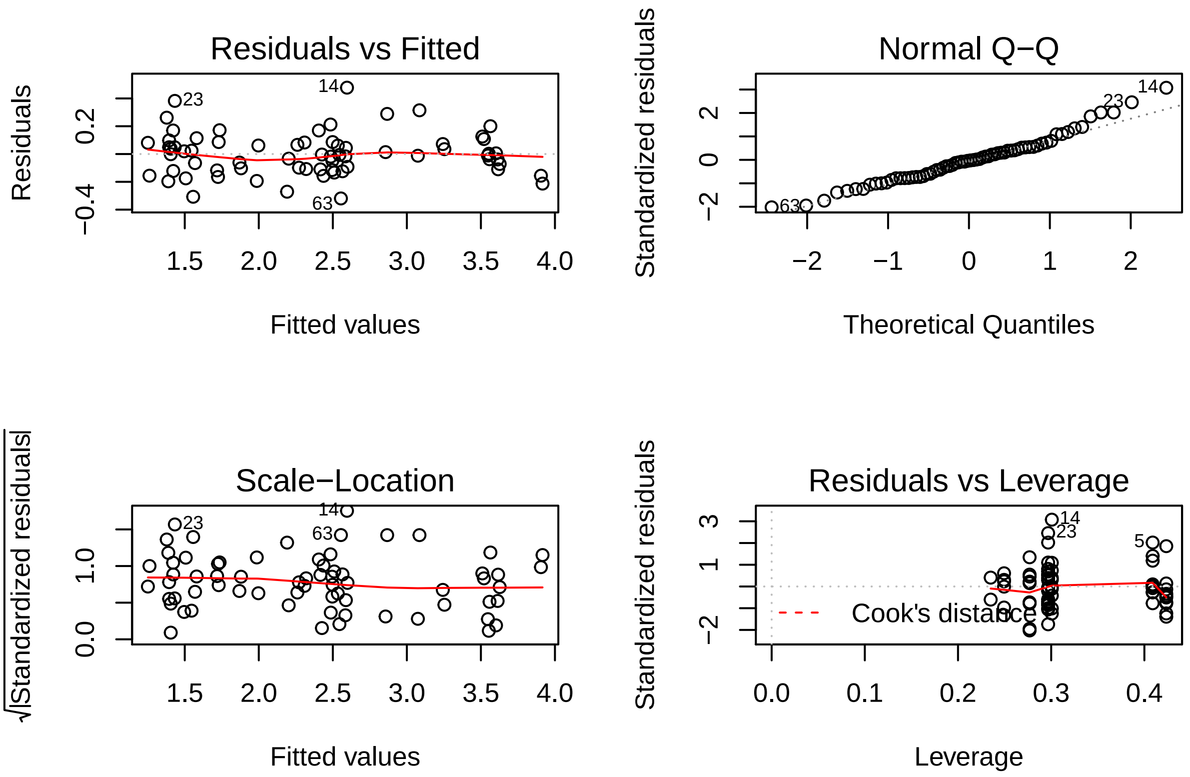
Again, we are using a boxcox plot to transform the response:



Thus, our final model is:

```
model6 = lm(y^0.3~(A + B + C + D + E)^2 + I(A^2) + I(B^2) + I(C^2) + I(D^2) + I(E^2) + replicate,
            data=data_all)
```

With its fixed residuals:



Here is its summary:

```
##
## Call:
## lm(formula = y^0.3 ~ (A + B + C + D + E)^2 + I(A^2) + I(B^2) +
##     I(C^2) + I(D^2) + I(E^2) + replicate, data = data_all)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3196 -0.1099 -0.0052  0.0813  0.4774
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2663     0.0872   26.00  < 2e-16 ***
## A            -0.0368     0.0259   -1.42  0.16268
## B            -0.0741     0.0259   -2.86  0.00639 **
## C            -0.0659     0.0259   -2.54  0.01442 *
## D            -0.6970     0.0259  -26.88  < 2e-16 ***
## E            -0.3816     0.0259  -14.72  < 2e-16 ***
## I(A^2)        0.0709     0.0633    1.12  0.26883
## I(B^2)        0.0328     0.0633    0.52  0.60752
## I(C^2)        0.0225     0.0633    0.36  0.72363
## I(D^2)       -0.4516     0.0633   -7.13  5.8e-09 ***
## I(E^2)        0.4516     0.0633    7.13  5.8e-09 ***
## replicate    -0.0054     0.0225   -0.24  0.81139
## A:B           0.0504     0.0284    1.78  0.08235 .
## A:C           0.0864     0.0284    3.05  0.00381 **
```

```
## A:D           0.0146     0.0284    0.52  0.60897
## A:E          -0.0366     0.0284   -1.29  0.20347
## B:C           0.0442     0.0284    1.56  0.12599
## B:D          -0.0337     0.0284   -1.19  0.24086
## B:E           0.0441     0.0284    1.55  0.12694
## C:D          -0.0420     0.0284   -1.48  0.14556
## C:E           0.0214     0.0284    0.76  0.45407
## D:E           0.1141     0.0284    4.02  0.00021 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.186 on 46 degrees of freedom
## Multiple R-squared:  0.963,  Adjusted R-squared:  0.946
## F-statistic: 56.7 on 21 and 46 DF,  p-value: <2e-16
```

Only linear components of factors B,C are significant while both linear and quadratic components of D and E are found significant here. Also, the interaction DE and AC are found significant as well.

(f)

We found that factors D and E significant in all three analysis (their quadratic components in 2 of those 3), and the interaction DE significant in two of the three analysis (the second one did not consider 2fi's). Therefore, we are confident to say that Ribavirin and Acyclovir are two effective antiviral drugs, and also a mix of those 2 drugs is significant as well for treating the HSV-1.

I would suggest doing a follow up experiement, with only drugs D and E but with more levels (like 4 or 5 level for each).