Effects on Pace

Mathias Dåsvand, Peder Eggen Skaug, Ole Kristian Skogly, Isak Ytrøy

May 2022

Introduction

In this experiment we consider how ones pace when running depends on different factors. Here we have looked at: time of day(morning or evening), steepness of track, whether we are running on asphalt or gravel and the effect of warming up prior to running.

This is an interesting question as many runners wish to monitor their running performance over time irrespective of steepness, substrate and other factors. In general, we expect the steepness to play a significant role on the running pace, while there is no clear expectation of the effect of the other 3 variables. The experiment seeks to give estimates of the impact of these variables.

Setup and Data Gathering

The experiment was completed in two workouts, each consisting of 8 1-minute runs. The first session was done at 19:00 12th April 2022, and the second session was done at 13:00 15th April 2022. The runner had some days rest between the sessions so the first session should not affect the second one. Each session was set up as 4 runs followed by a pause, then a warm-up and the last 4 runs. This experiment is a 4-factorial experiment, where the factors we looked at were:

A - the time of day, with morning assigned the value 1 and evening -1 in the factorial design.

B - the steepness, with less steep assigned 1, versus -1 for more steep.

C - the substrate, with asphalt assigned 1, and gravel assigned -1 for the factorial design.

D - warm-up (1) versus no warm-up (-1).

The response variable is defined as the pace, which is the amount of seconds used per kilometer traversed. For example, a response of 260 means that if the pace was continued, it would take 4 minutes and 20 seconds to run a kilometer.

Figure 1: Gravel tracks. The steeper track to the left.

Figure 2: Asphalt tracks. The steeper track to the right.

The data was collected by performing 16 tests, as previously explained, under differing conditions. Responses and conditions for the tests are shown in table 1. We chose not to use blocking in our experiment, due to time constraints. There is no need for replication, since with 4 variables we need 16 trials to test all combinations, which should also suffice for an analysis.

Run	A	B	C	D	Y
1	-1	-1	-1	-1	337
2	-1	1	-1	-1	226
3	-1	1	1	-1	211
4	-1	-1	1	-1	266
5	-1	1	-1	1	313
6	-1	-1	1	1	226
7	-1	1	1	1	284
8	-1	-1	-1	1	430
9	1	-1	-1	-1	377
10	1	1	-1	-1	251
11	1	1	1	-1	238
12	1	-1	1	-1	277
13	1	1	-1	1	253
14	1	-1	1	1	262
15	1	1	1	1	226
16	1	-1	-1	1	285

Table 1: Table showing the parameter level, and the response at each test. A, B, C, D as previously stated.

Analysis

To determine which variables are most significant in a predictive model, a full linear regression is used. Such a model will estimate the influence of not just each parameter individually, but also the parameters which are combinations of multiple parameters. Finally we eliminate the less significant parameters and create a reduced model.

A full linear regression assumes that all of the parameters, both individually and jointly, contribute when calculating the response. This leads to an expression with as many estimators as there are tests. So in this case any response is assumed to be on the form;

$$\hat{Y} = \hat{\beta_0} + \hat{\beta_1}x_1 + ... + \hat{\beta_{15}}x_1x_2x_3x_4, \tag{1}$$

where our test parameters are represented by x_i, and the $\hat{\beta_i}$-estimators relate the combination of parameters to the response. We should note that $\hat{\beta_0}$ is not interesting for our purposes, as it is just the baseline pace. Using the full linear model a few properties can immediately be studied. First is the main effect plot 3, which shows the expected response given the level (1 or -1) of the parameter. Precisely, each value in the plot is the average of all trials when a parameter is at a specific value.

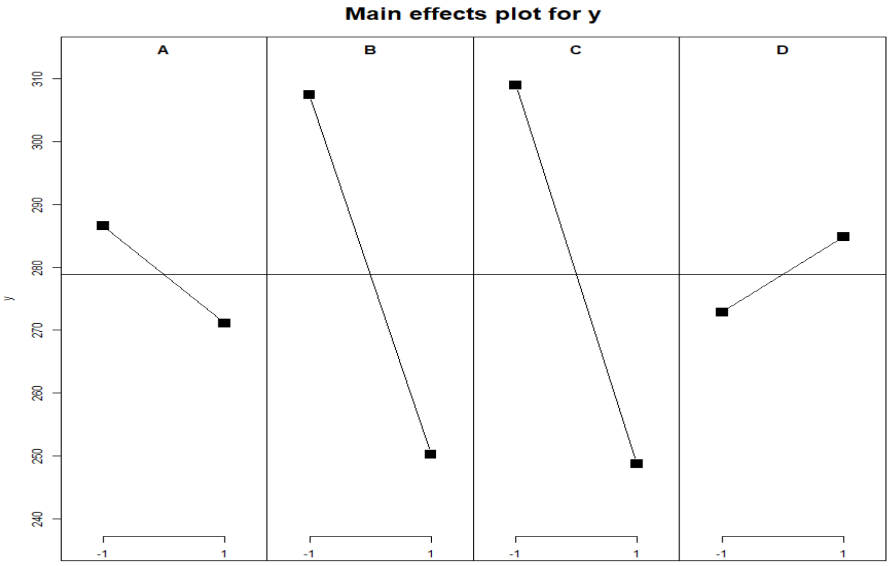

Figure 3: Main effect plot of the full linear model.

Studying figure 3, one can see that all 4 of our parameters seem to have non-trivial effects on the response. This is seen as the difference in response, given the two levels. Any horizontal lines would indicate parameters without significance. One can see that factor B and C is most significant to the response variable. From the plot we can see for both factors B and C that the pace is faster for the value 1 than for the value -1. This tells us that the pace is faster for the cases which are less steep, $B = 1$, and on asphalt, $C = 1$.

The next matter of study is the interaction between the parameters, which has been visualized in plot 4,

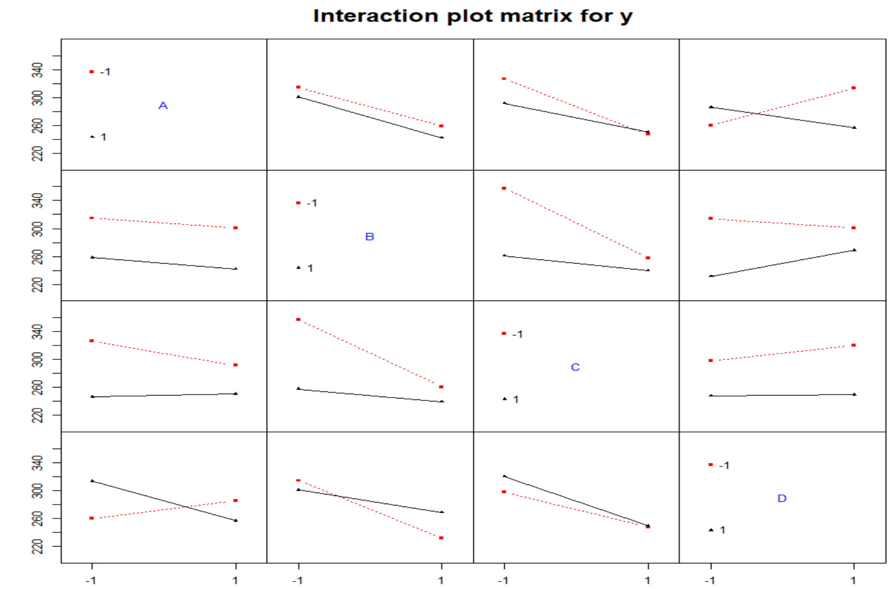

Figure 4: Interaction plot of the full linear model.

where parallel lines indicate little interaction between parameters, as the estimator of their joint effect is negligible. The larger the $\hat{\beta}$ associated with $x_i x_j$, the larger the difference in slopes between the factors will be. In particular B:C and A:D show some potential dependence between the parameters. The initial observations indicate that time of day and incline are independent parameters, while the remaining pairs show varying degrees of dependence. With these observations as a basis, we will now apply Occam's razor, and see if a reduced model can sufficiently explain the response.

To find such a reduced model we perform the following 15 hypothesis tests:

$$H_0 : \hat{\beta}_i = 0 \quad \text{vs.} \quad H_1 : \hat{\beta}_i \neq 0$$

Rather than doing an F-test, which rejects or a accepts all the hypothesis at once, Lenth's method will be employed. This method calculates a "cut-off value", τ^*. The cut-off value is identical for all the tests, and rejects the null hypothesis for the individual estimators if $2\hat{\beta}_i > \tau^*$. Having performed these tests, we rejected several estimators based on the diagram in figure 5.

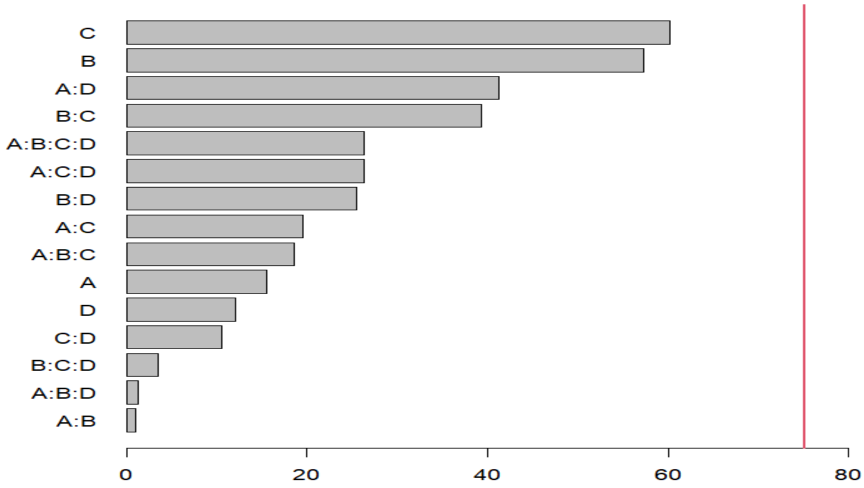

Figure 5: Absolute value of the effects, and the cut-off value in red.

Using a significance level of $\alpha = 0.05$ means that we reject the null hypothesis for any parameter where $2|\hat{\beta}| > 73.80$. Using Lenth's method we find that in fact we do not reject the null hypothesis for any other parameter than the intercept(see figure 5). Regardless, we still continue with the analysis using the parameters C, B, A:D and B:C in order to have something to write about. We acknowledge that this is suboptimal.

These combinations of parameters now construct our reduced model, as this new model is predictive we can look at residuals and quantiles to get an idea of how good it is.

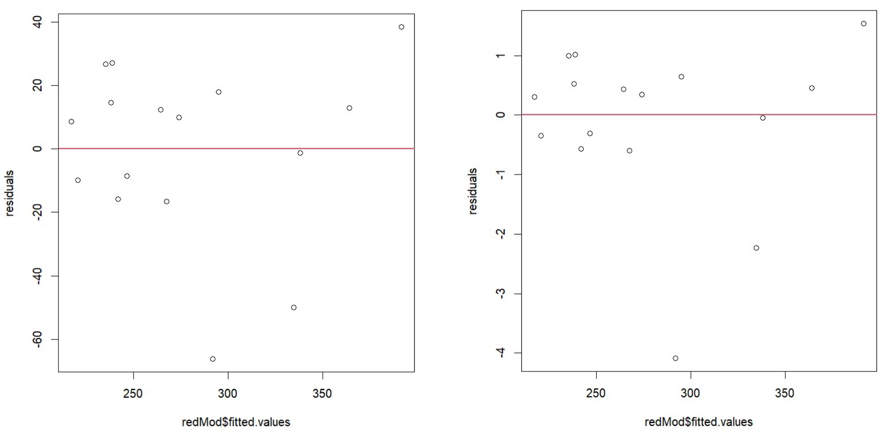

Figure 6: Residuals from the reduced model, both original (left) and studentized (right)

This reduced model has an R^2-value of 0.786, adjusted R^2 of 0.643 and an estimated $\hat{\sigma}^2$ of 35.75 (See appendix). The estimates are given as:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_7 \\ \hat{\beta}_9 \end{bmatrix} = \begin{bmatrix} 278.875 \\ -7.75 \\ -28.625 \\ -30.125 \\ 6 \\ -20.625 \\ 19.625 \end{bmatrix}$$

Looking at figure 6, there are both good and bad news. Many of the residuals appear to behave homoscedastically both before and after studentization, but there are two outliers. Looking at the studentized residuals the largest outlier seems worthy of scrutiny, by naïvely finding the standard deviation of the residuals we get $\sigma_\epsilon = 1.366$. This implies that our outlier deviates from the expected value by $3\sigma_\epsilon$. As residuals are typically modeled as normally distributed noise with expectation 0, this $3\sigma_\epsilon$ deviation from 0 has a roughly 0.13% chance of occurring.

To get another viewpoint, a Quantile-Quantile plot of the studentized residuals would come in handy.

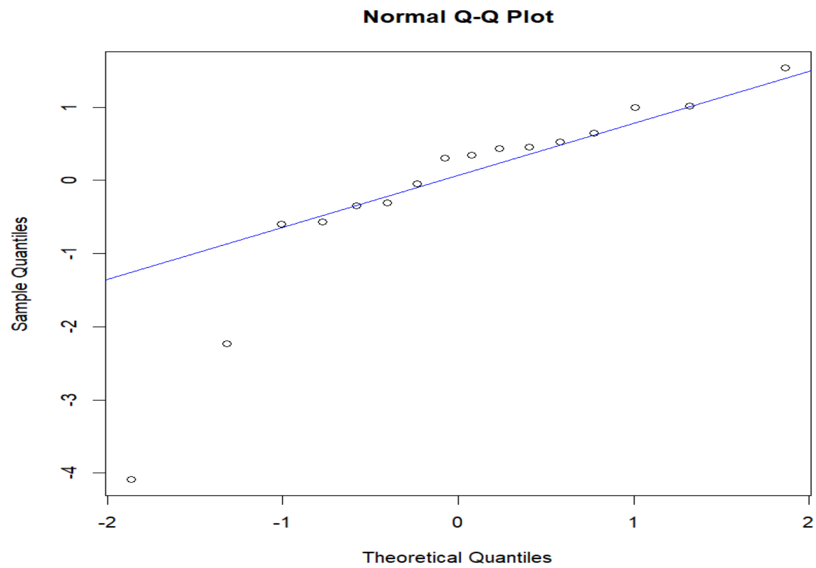

Figure 7: Q-Q plot of the studentized residuals

Looking at figure 7 similar observations as in figure 6 come to mind. Much of the data seems to be normally distributed, but again we see two major outliers. These outliers deviate from the straight line one would expect to see when plotting samples from two normal distributions against each other. These discrepancies and their sources is a matter for later discussion.

Finally, to check if the response data is of an appropriate form, a Box-Cox plot with a corresponding confidence interval can be used to verify if the desired λ-value is within the interval.

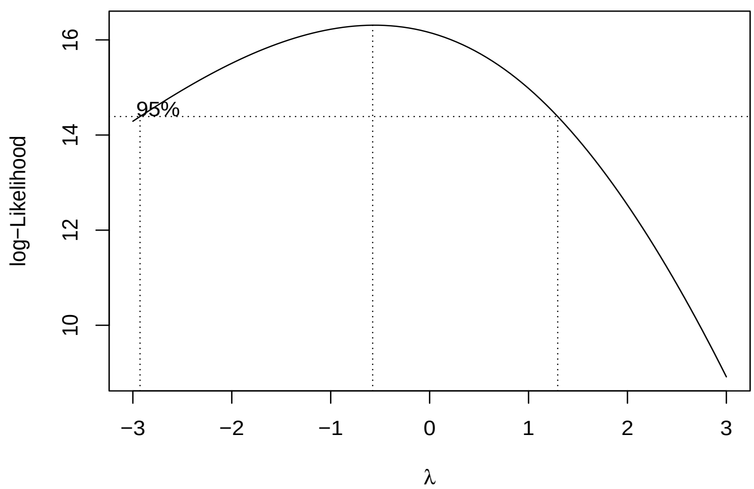

Figure 8: Box-Cox plot.

One can from the Box-Cox plot see that $\lambda = 1$ is within the 95% confidence interval. This tells us that no Box-Cox transformation of our response variables is necessary. A λ value of 1 is equivalent to using the original data. Observe however that several other λ-values, such as $\lambda = -2, -1, 0$, are within our confidence interval, meaning some other transformations would probably work with our data.

Discussion

Having analyzed the data, two subjects of discussion are necessary, the gathering of data, and their influence on the reduced model.

Data Gathering

The trials were all performed in 2 separate sessions which clearly is not ideal in a 2^4 experiment. Originally we intended to have randomization of 16 different trials, however this was cut short due to the original test subject getting injured. It was then decided to gather the data in only two sessions due to time constraints.

Between each trial the new test subject had a break until they felt completely rested. This is of course not enough to ensure that each trial is completely independent of the next as the runner gets more fatigued over time. We argue however that the effect of fatigue was relatively low, as each trial only lasted for a minute.

There are many possible confounders with this design. The inaccuracy of the GPS tracker, effect of training and minor wind are all factors possibly impacting the response variable. The "Polar Ignite"-clock which was used, did return pace as a single data point, but it did also return the travelled distance and used time. This should have let us double check the pace by dividing time with distance, however these 3 data points did rarely correspond. In the worst case we got a pace difference of $10\frac{min}{km}$ from these 3 data points.

While blocking was not used for this experiment, due to time constraints, it may well have been desirable to use it. Some nuisance factors could be weather conditions, form/fatigue or motivation. Blocking could especially prevent the nuisance created by compounding fatigue, as the runner will get progressively more tired after each run. If we were to block our experiment, the nuisance should have less of an effect on the variability. The same applies to changing weather conditions.

The Reduced Model

Based on the full linear model and hypothesis testing with Lenth's method, a reduced model was created. When fitted against the gathered responses the reduced model seemed to perform quite well, with two noticeable exceptions. The largest residual was found to be test 6, which was a steep run on asphalt, after warming up. According to the reduced model such a run should have a moderate pace, but the measured pace was among the fastest. This could simply be a case of our test subject getting a burst of motivation, and pushing themselves. Another possible source of error could, as mentioned above could be the measuring equipment, which proved to have a highly unreliable GPS.

Further, there could be multiple possible interactions between the substrate and the steepness. First of all, the steepness was slight over the whole interval on asphalt, while the steepness went from almost nonexistent to large at the end on the gravel (see figure 1). The gravel track also had some patches of snow at the end. Thus the parameter C could to some degree be another measure of steepness, if the substrate has little impact.

With all of this being said, the reduced model ended up delivering good results over all. Residuals had a homoscedastic trend, and the general data fit gave a coefficient of determination at $R^2 = 0.7855$, which indicates that nearly 80% of the variance in the model was explained by interaction between parameter, rather than unconsidered effects and disturbances. Finally the adjusted coefficient of determination ended up at $\tilde{R}^2 = 0.6426$, which is reassuring as this coefficient is intended to take into account if a model is over-fitted.

Conclusion

From the data results in the experiment no parameters were significant. However, B and C had the most impact on the response variable and were the most significant of all parameters. One can look at the main effects plot for the response variable in figure 3 to see this connection. From figure 6 we decided to use the factors A, B, A:D and B:C in our reduced model. The conclusion of our 2^4 test, is then that steepness and the substrate are the factors which seems to have the most impact on running pace.

Appendix 1: R Code

Plot commands are commented out to not get duplicate plots

```
## Loading required package: DoE.base
```

```
## Loading required package: grid
```

```
## Loading required package: conf.design
```

```
## Registered S3 method overwritten by 'DoE.base':
##   method           from
##   factorize.factor conf.design
```

```
##
## Attaching package: 'DoE.base'
```

```
## The following objects are masked from 'package:stats':
##
##     aov, lm
```

```
## The following object is masked from 'package:graphics':
##
##     plot.design
```

```
## The following object is masked from 'package:base':
##
##     lengths
```

Data from the experiment

```
#Our responses
y = c(337,226,211,266,313,226,284,430,
       377,251,238,277,253,262, 226, 285)

quantile(y,0.25)
```

```
## 25%
## 235
```

```
#The variable combinations
A = c(-1,-1,-1,-1,-1,-1,-1,-1,1,1,1,1,1,1,1,1) #morgen/kveld (1,-1)
B = c(-1,1,1,-1,1,-1,1,-1,-1,1,1,-1,1,-1,1,-1) #steep/steeper (1,-1)
C = c(-1,-1,1,1,-1,1,1,-1,-1,-1,1,1,-1,1,1,-1) #asphalt/gravel (1,-1)
D = c(-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1) #warmup/coldstart (1,-1)
df = data.frame(y,A,B,C,D)
```

Linear regression using all combinations of variables

```
lm4 = lm(y~.^4, data=df)
summary(lm4)
```

```
##
## Call:
## lm.default(formula = y ~ .^4, data = df)
##
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  278.875        NaN     NaN      NaN
## A             -7.750        NaN     NaN      NaN
## B            -28.625        NaN     NaN      NaN
## C            -30.125        NaN     NaN      NaN
## D              6.000        NaN     NaN      NaN
## A:B           -0.500        NaN     NaN      NaN
## A:C            9.750        NaN     NaN      NaN
## A:D          -20.625        NaN     NaN      NaN
## B:C           19.625        NaN     NaN      NaN
## B:D           12.750        NaN     NaN      NaN
## C:D           -5.250        NaN     NaN      NaN
## A:B:C         -9.250        NaN     NaN      NaN
## A:B:D         -0.625        NaN     NaN      NaN
## A:C:D         13.125        NaN     NaN      NaN
## B:C:D          1.750        NaN     NaN      NaN
## A:B:C:D      -13.125        NaN     NaN      NaN
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 15 and 0 DF,  p-value: NA
```

Main effect plot

```
#MEPlot(lm4)
```

#Interaction plot

```
#IAPlot(lm4)
```

Analysis of estimators using Lenth's method

```
effs = 2*lm4$coefficients[-1] # omit intercept
abeffs = abs(effs)
mDabEffs = median(abeffs)
tau = 1.5*mDabEffs
relevant = abeffs[abeffs < 2.5*tau]
betterTau = 1.5*median(relevant)

alpha = 0.05 # significance level
cutoff = qt(1-alpha/2,length(abeffs)/3)*betterTau
cutoff
```

```
## [1] 75.18952
```

```
abs(effs) > cutoff
```

```
##       A       B       C       D     A:B     A:C     A:D     B:C     B:D     C:D
##   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
##   A:B:C   A:B:D   A:C:D   B:C:D A:B:C:D
##   FALSE   FALSE   FALSE   FALSE   FALSE
```

Pareto-type plot for the effects

```
oldmar = par()$mar
par(xpd = FALSE)
par(mar = oldmar+c(0,2,0,0)) # to get a wider left margin in plot
#barplot(sort(abseffects,decreasing = FALSE), las = 1, horiz = TRUE)
#abline(v=cutoff/2, col=2, lwd=2)
```

Reduced model with significant factors

```
redMod=lm(y ~ B + C + A*D + B*C, data = df)
summary(redMod)
```

```
##
## Call:
## lm.default(formula = y ~ B + C + A * D + B * C, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -66.12 -11.38   9.25  15.44  38.38
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  278.875      8.938  31.201 1.75e-10 ***
## B            -28.625      8.938  -3.203  0.01079 *
## C            -30.125      8.938  -3.370  0.00825 **
## A             -7.750      8.938  -0.867  0.40842
## D              6.000      8.938   0.671  0.51889
## A:D          -20.625      8.938  -2.308  0.04642 *
## B:C           19.625      8.938   2.196  0.05573 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.75 on 9 degrees of freedom
## Multiple R-squared:  0.7855, Adjusted R-squared:  0.6426
## F-statistic: 5.494 on 6 and 9 DF,  p-value: 0.01192
```

Residuals for the reduced model

```
#residuals = resid(redMod)
#residuals = rstudent(redMod)
#plot(redMod$fitted.values, residuals)
#abline(h = 0, col = 2, lwd = 2)
```

Checking normal distribution of residuals

```
residuals = rstudent(redMod)
#qqnorm(residuals)
#qqline(residuals, col = "blue")
```

Box cox plot

```
#boxcox(y ~ B + C + A*D + B*C, lambda = seq(-3,3))
```