AdvStDaAn, Worksheet, Week 1

Micheal Lappert

05.04.2022

Contents

Exercise 1

```r
path <- file.path('Datasets', 'Softdrink.dat')
df <- read.table(path, header=TRUE)

summary(df)
```

Dataset loading and sanity check:

```
##       Time           volume         distance        location        
##  Min.   : 8.00   Min.   : 2.00   Min.   : 10.8   Length:25         
##  1st Qu.:13.75   1st Qu.: 4.00   1st Qu.: 45.0   Class :character  
##  Median :18.11   Median : 7.00   Median : 99.0   Mode  :character  
##  Mean   :22.38   Mean   : 8.76   Mean   :122.8                     
##  3rd Qu.:21.50   3rd Qu.:10.00   3rd Qu.:181.5                     
##  Max.   :79.24   Max.   :30.00   Max.   :438.0                     
```

```r
head(df)
```

```
##    Time volume distance  location
## 1 16.68      7      168 San Diego
## 2 11.50      3       66 San Diego
## 3 12.03      3      102 San Diego
## 4 14.88      4       24 San Diego
## 5 13.75      6       45 San Diego
## 6 18.11      7       99 San Diego
```

```r
tail(df)
```

```
##     Time volume distance    location
## 20 35.10     17    231.0      Austin
## 21 17.90     10     42.0      Austin
## 22 52.32     26    243.0      Austin
## 23 18.75      9    135.0      Austin
## 24 19.83      8    190.5 Minneapolis
## 25 10.75      4     45.0 Minneapolis
```

Data looks just fine.

Exercise 1.a)

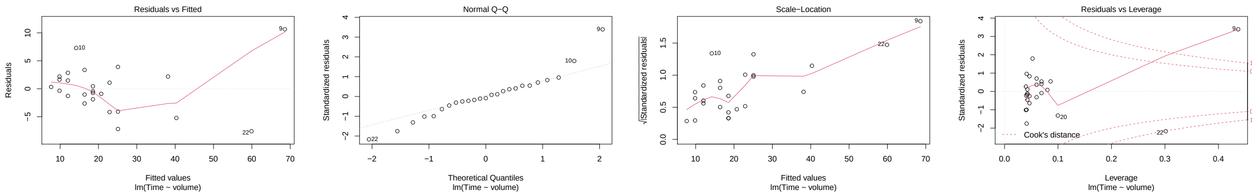
```r
mod1.1 <- lm(Time ~ volume, data = df)
summary(mod1.1)
```

```
##
## Call:
## lm(formula = Time ~ volume, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.5811 -1.8739 -0.3493  2.1807 10.6342
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.321      1.371   2.422   0.0237 *
## volume         2.176      0.124  17.546 8.22e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.181 on 23 degrees of freedom
## Multiple R-squared:  0.9305, Adjusted R-squared:  0.9275
## F-statistic: 307.8 on 1 and 23 DF,  p-value: 8.22e-15
```

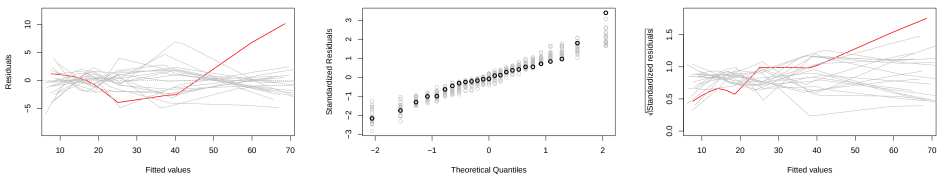
The model looks fine: - Volume is significant on the 5% niveau and the R-squared has a score of 0.93.

We have to do a residual and sensitivity analysis with stochastic simulation to investigate the correctness of the model.

```r
plot(mod1.1)
```




```r
plot.lmSim(mod1.1, SEED = 1)
```



Interpretation:

1. Tukey-Anscombe plot: Shows outlier with index i=9 which affects the smoother. In the simulation it is visible that the original curve is extreme.
   => The expectation of the residuals cannot be constant.
2. Q-Q plot: In the lower as well as in the higher part of the plot some points differ from the straigt line. Most of them are within the stochastic fluctuation except i=9.
   => The assumption of normal distributed residuals seems violated.
3. Scale-location plot: Shows a clear upwards trend. In the simulation it is visible that the original curve is extreme.
   => The variance of the residuals is not constant.
4. Residuals vs. Leverage: Observations i = 9 & 22 have Cooke's Distance > 1 and are therefore too influentious. Both observations have addationally too much leverage.
   => Residuals are not normally distributed.

CONCLUSION: The fit is not satisfactory. Trying transformations of response and explanatory variable. Since the noncanstant variance seems to be the most severe problem, log-transformations might help.

Exercise 1.b)

```r
df$lVolume <- log(df$volume)
df$lTime <- log(df$Time)

head(df)
```

Tukey's first-aid transformations:

```
##    Time volume distance  location  lVolume    lTime
## 1 16.68      7      168 San Diego 1.945910 2.814210
```

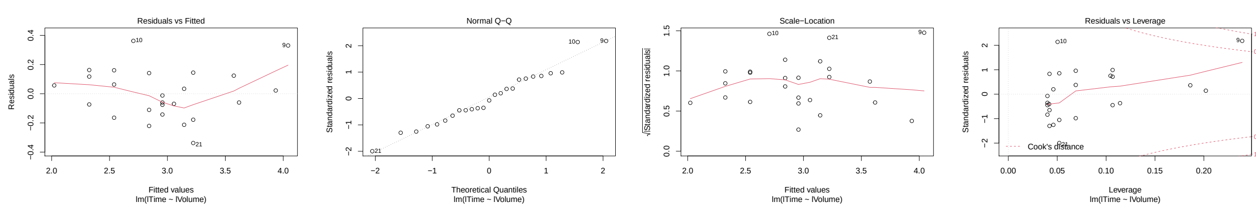
```
## 2 11.50      3       66 San Diego 1.098612 2.442347
## 3 12.03      3      102 San Diego 1.098612 2.487404
## 4 14.88      4       24 San Diego 1.386294 2.700018
## 5 13.75      6       45 San Diego 1.791759 2.621039
## 6 18.11      7       99 San Diego 1.945910 2.896464
```

Model with transformed variables:

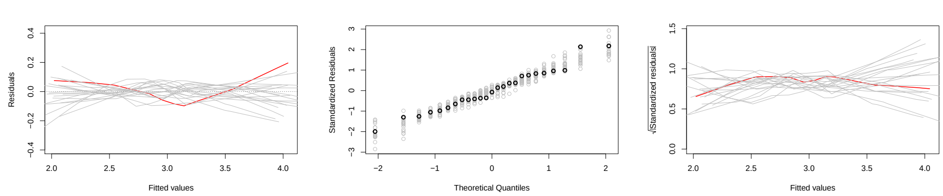
```
mod1.2 <- lm(lTime ~ lVolume, data = df)
summary(mod1.2)
```

```
##
## Call:
## lm(formula = lTime ~ lVolume, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33794 -0.11068 -0.01232  0.12385  0.36222
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.50560    0.10897   13.82 1.26e-12 ***
## lVolume      0.74575    0.05317   14.03 9.25e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1738 on 23 degrees of freedom
## Multiple R-squared:  0.8953, Adjusted R-squared:  0.8908
## F-statistic: 196.7 on 1 and 23 DF,  p-value: 9.252e-13
```

```
plot(mod1.2)
```




```
plot.lmSim(mod1.2, SEED = 1)
```



Interpretation:

1. Tukey-Anscombe plot: The smoother shows a somewhat strange banana form with the low in the middle which is outside the stochastic fluctuation.
   => The assumption of constant expectaion is therefore violated.
2. Q-Q plot: The data scatters nicely around the straight line and seems to be within the stochastic fluctuation.
   => The assumption of Gaussian distributed errors seems not violated.
3. Scale-location plot: The smoother shows a slightly decreasing trend but seems to be ok and lies within the stochastic fluctuation.
   => There is no evidence against the assumption of constant variance of the residuals.
4. Residuals vs. Leverage: All observations have Cook's Distance <1 and therewith no too influential points.
   => No too influential (dangerous) observations

CONCLUSION: The model does still not fit adequately the data, although it is much better than the one before.

An alternative transformation for volume could be the square-root transformation. So let's try out:

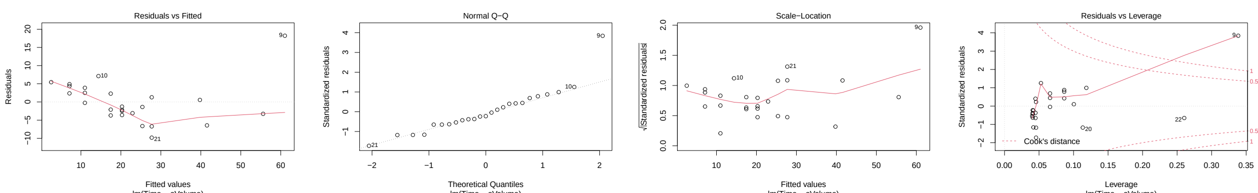
```
df$sVolume <- sqrt(df$volume)

mod1.3 <- lm(Time ~ sVolume, data = df)
summary(mod1.3)
```

```
##
## Call:
## lm(formula = Time ~ sVolume, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.817 -3.266 -1.284  2.509 18.212
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -17.788      3.510  -5.067 3.95e-05 ***
## sVolume       14.390      1.186  12.133 1.77e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.83 on 23 degrees of freedom
## Multiple R-squared:  0.8649, Adjusted R-squared:  0.859
## F-statistic: 147.2 on 1 and 23 DF,  p-value: 1.775e-11
```

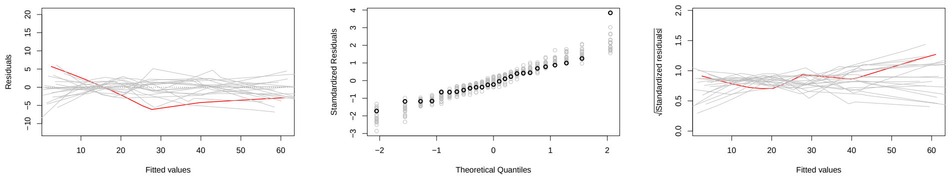
The R^2 is with 0.8649 higher than before.

Residual and Sensitivity Analysis:

```
plot(mod1.3)
```



```r
plot.lmSim(mod1.3, SEED = 1)
```



Interpretation:

1. Tukey-Anscombe plot: The smoother shows still a somewhat strange banana form with the low in the middle but like that it is inside the stochastic fluctuation.
=> The assumption of constant expactation is not violated. 2. Q-Q plot: The data scatters nicely around the straight line (except i=9) and seems to be within the stochastic fluctuation.
=> The assumption of Gaussian distributed errors seems not violated. 3. Scale-location plot: The smoother looks ok and lies within the stochastic fluctuation.
=> There is no evidence against the assumption of constant variance of the residuals. 4. Residuals vs. Leverage: All observations have Cook's Distance <1 and therewith no too influential points.
=> No too influential (dangerous) observations

CONCLUSION: The model does still fit adequately the data.

Exercise 1.c)

The fitte model in 1.b) is:

$$Time_i = exp(\beta_0) + exp(\beta_1) * sqrt(volume_i) + exp(E_i)$$

with

$$\mu = 0$$

$$\sigma = \sigma$$

Exercise 1.d)

Extending the model adequately with the second explanatory variable 'distance':

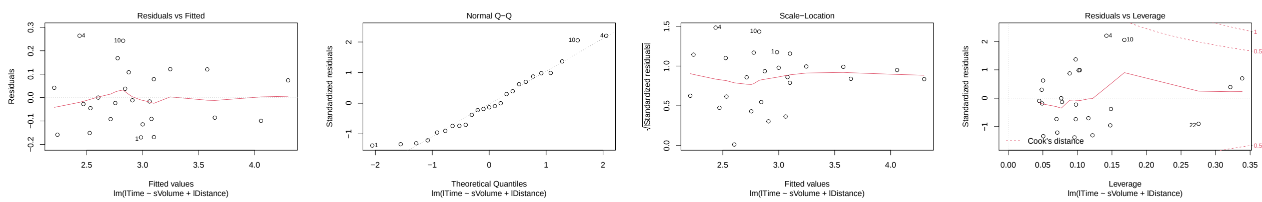
```r
df$lDistance <- log(df$distance)
mod1.4 <- lm(lTime ~ sVolume + lDistance, data = df)
summary(mod1.4)
```

```
##
## Call:
## lm(formula = lTime ~ sVolume + lDistance, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.17012 -0.09203 -0.01658  0.07866  0.26425
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.14704    0.14300   8.021 5.65e-08 ***
## sVolume      0.41553    0.03649  11.389 1.08e-10 ***
```

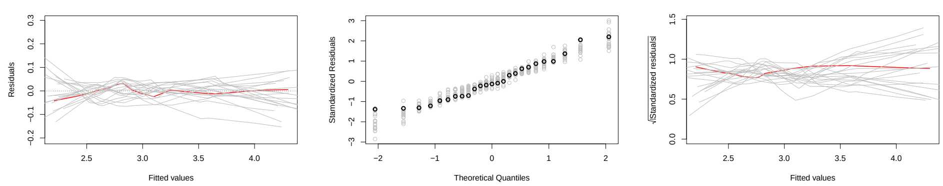
```
## lDistance    0.14401    0.04234   3.401  0.00256 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1296 on 22 degrees of freedom
## Multiple R-squared:  0.9443, Adjusted R-squared:  0.9393
## F-statistic: 186.6 on 2 and 22 DF,  p-value: 1.587e-14
```

The R^2 increases to 0.9443 which is a really good fit. Lets check the model:

```
plot(mod1.4)
```




```
plot.lmSim(mod1.4, SEED = 1)
```



Interpretation:

1. Tukey-Anscombe plot: The smoother shows a really nice almost straight line which is within the stochastic fluctuation.
   => The assumption of constant expectation is not violated.
2. Q-Q plot: The data scatters nicely around the straight line and seems to be within the stochastic fluctuation.
   => The assumption of Gaussian distributed errors seems not violated.
3. Scale-location plot: The smoother shows a almost straight line and is within the stochastic fluctuation.
   => There is no evidence against the assumption of constant variance of the residuals.
4. Residuals vs. Leverage: All observations have Cook's Distance <1 and therewith no too influential points. Some are leverage point with leverage $> 2 * 3/25 = 0.24$ (25 examples in dataset)
   => No too influential (dangerous) observations

CONCLUSION: The model does fit adequately the data.

Exercise 2

```r
path <- file.path('Datasets', 'Windmill.dat')
df <- read.table(path, header = TRUE)

summary(df)
```

Loading and Checking the data

```
##     velocity        DC.output    
##  Min.   : 5.482   Min.   :0.123  
##  1st Qu.: 8.838   1st Qu.:1.144  
##  Median :13.424   Median :1.800  
##  Mean   :13.720   Mean   :1.610  
##  3rd Qu.:18.235   3rd Qu.:2.166  
##  Max.   :22.822   Max.   :2.386
```

```r
dim(df)
```

```
## [1] 25  2
```

```r
head(df)
```

```
##    velocity DC.output
## 1 11.187073     1.582
## 2 13.424487     1.822
## 3  7.607209     1.057
## 4  6.041019     0.500
## 5 22.374145     2.236
## 6 21.702921     2.386
```

```r
tail(df)
```

```
##     velocity DC.output
## 20 12.193909     1.501
## 21 20.360472     2.303
## 22 22.821628     2.310
## 23  9.173400     1.194
## 24  8.837787     1.144
## 25  5.481666     0.123
```

Exercise 2.a)

Start with fitting an ordinary regression model:

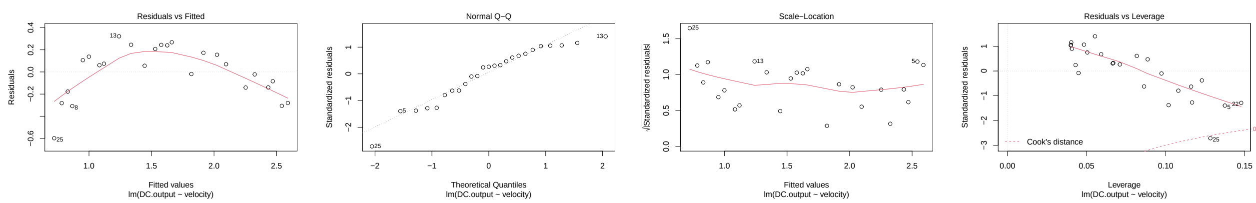
```r
mod2.1 <- lm(DC.output ~ velocity, data = df)
summary(mod2.1)
```

```
## 
## Call:
## lm(formula = DC.output ~ velocity, data = df)
## 
```

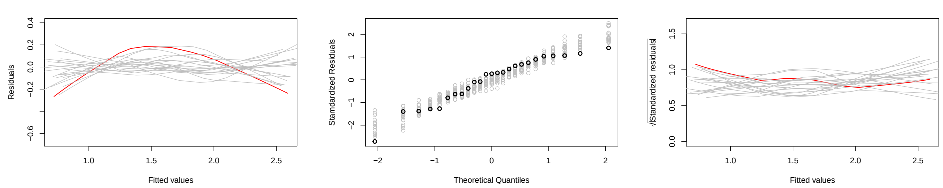
```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59869 -0.14099  0.06059  0.17262  0.32184
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.130875   0.125989   1.039     0.31
## velocity    0.107780   0.008514  12.659 7.55e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2361 on 23 degrees of freedom
## Multiple R-squared:  0.8745, Adjusted R-squared:  0.869
## F-statistic: 160.3 on 1 and 23 DF,  p-value: 7.546e-12
```

The model seems to fit not too bad. Lets check this:

```
plot(mod2.1)
```




```
plot.lmSim(mod2.1, SEED = 1)
```



Interpretation:

1. Tukey-Anscombe plot: The smoother shows a banana form which is outside the stochastic fluctuation. => The assumption of constant expectation is violated.
2. Q-Q plot: The data does not scatter nicely around the straight line and but seems to be within the stochastic fluctuation.
   => The assumption of Gaussian distributed errors seems not violated.
3. Scale-location plot: The smoother shows a almost straight line and is within the stochastic fluctuation. => There is no evidence against the assumption of constant variance of the residuals.
4. Residuals vs. Leverage: All observations have Cook's Distance <1 and therewith no too influential points. There are also no leverage points with leverage $> 2 * 2/25 = 0.16$ (25 examples in dataset, 2 variables in model)
   => No too influential (dangerous) observations

CONCLUSION: The model does not fit adequately the data. Maybe some transfromations would help to remedy the inadequacy. -> Exercise 2.b)

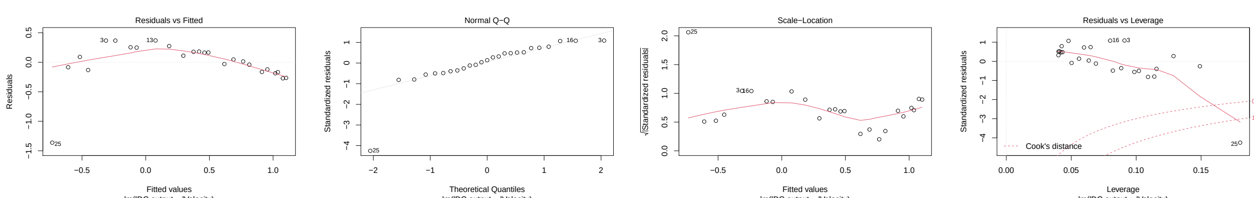
Exercise 2.b)

```
df$lVelocity <- log(df$velocity)
df$lDC.output <- log(df$DC.output)

mod2.2 <- lm(lDC.output ~ lVelocity, data = df)
summary(mod2.2)
```

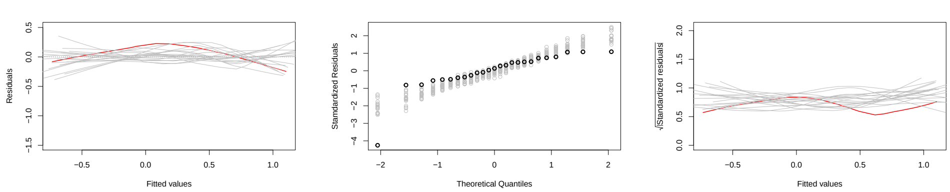
```
##
## Call:
## lm(formula = lDC.output ~ lVelocity, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36184 -0.13163  0.04707  0.18075  0.36880
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.9238     0.4112  -7.110 3.05e-07 ***
## lVelocity     1.2872     0.1603   8.031 4.01e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3537 on 23 degrees of freedom
## Multiple R-squared:  0.7371, Adjusted R-squared:  0.7257
## F-statistic:  64.5 on 1 and 23 DF,  p-value: 4.014e-08
```

Variable seems to be significant and $R^{2} is with 0.7371 not too bad. But we could surely do better with some adjustments. Lets check the model:

```
plot(mod2.2)
```




```
plot.lmSim(mod2.2, SEED = 1)
```



Interpretation:
1. Tukey-Anscombe plot: The smoother still shows a banana form which is outside the stochastic fluctuation.

=> The assumption of constant expactation is violated. 2. Q-Q plot: The data does better scatter around the straight line (except outlier i=25) but is outside the stochastic fluctuation.
=> The assumption of Gaussian distributed errors is violated. 3. Scale-location plot: The smoother shows a wavy line and is within the stochastic fluctuation.
=> There is no evidence against the assumption of constant variance of the residuals. 4. Residuals vs. Leverage: Observation i=25 has Cook's Distance >1 and therewith is too influential. But there are no leverage points with leverage $> 2 * 2/25 = 0.16$ (25 examples in dataset, 2 variables in model)
=> i=25 is too influential observations

CONCLUSION: The model does not fit adequately the data. Maybe some transfromations would help to remedy the inadequacy. -> Exercise 2.c)

Exercise 2.c)

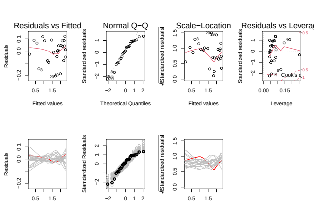
Theroy suggests the following transformation.

```
df$tVelocity <- 1/df$velocity

mod2.3 <- lm(DC.output ~ tVelocity, data = df)
summary(mod2.3)
```

```
##
## Call:
## lm(formula = DC.output ~ tVelocity, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20547 -0.04940  0.01100  0.08352  0.12204
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.9789     0.0449   66.34   <2e-16 ***
## tVelocity   -15.5155     0.4619  -33.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09417 on 23 degrees of freedom
## Multiple R-squared:   0.98,  Adjusted R-squared:  0.9792
## F-statistic:  1128 on 1 and 23 DF,  p-value: < 2.2e-16
```

This model seems to fit the data way better. The transformation from theroy seems pretty adequat. Lets check the model:

```
par(mfrow = c(2,4))
plot(mod2.3)
plot.lmSim(mod2.3, SEED = 1)
```



Interpretation:

1. Tukey-Anscombe plot: The smoother shows now no banana form anymore and is within the stochastic fluctuation.
   => The assumption of constant expectation of the errors is not violated.
2. Q-Q plot: The data does scatter around the straight line and is within the stochastic fluctuation.
   => The assumption of Gaussian distributed errors is not violated.
3. Scale-location plot: The smoother shows a little wavy line but is within the stochastic fluctuation.
   => There is no evidence against the assumption of constant variance of the residuals.
4. Residuals vs. Leverage: No Observation has Cook's Distance >1. And there are no leverage points with leverage $> 2 * 2/25 = 0.16$ (25 examples in dataset, 2 variables in model)
   => i=25 is too influential observations

CONCLUSION: The model does fit adequately the data. The transformation of the velocity variable did indeed remedy the model inadequaties.

Exercise 2.d)

Plotting the model for the interpretation of the parameters β_0 and β_1.

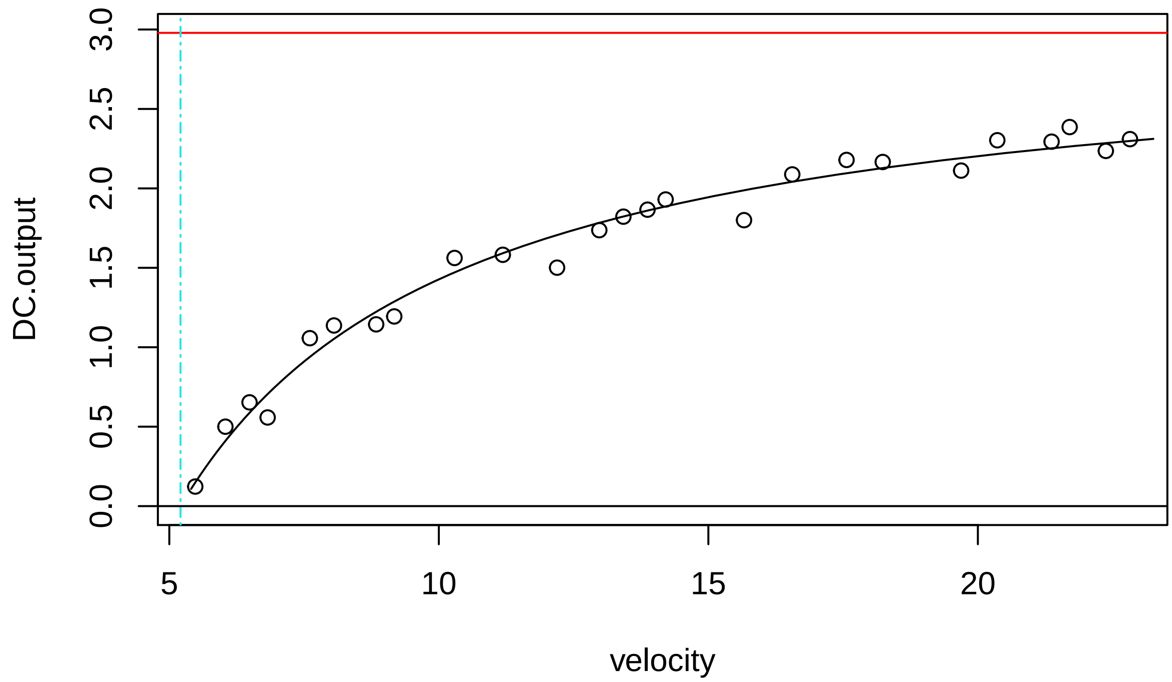
```
par(mfrow=c(1,1))
plot(DC.output ~ velocity, data = df,
     ylim = range(df$DC.output, coef(mod2.3)[1], 0))
# -> The range() takes the larger value of DC.output and the intercept
abline(h = coef(mod2.3)[1], col = "red")

# predicted DC.output
range(df$tVelocity)
```

```
## [1] 0.04381808 0.18242629
```

```
df2 <- data.frame(tVelocity = seq(0.043, 0.185, length = 50))
lines(1/df2$tVelocity, predict(mod2.3, newdata = df2))

# How much wind is needed at least?
(h.minW <- -coef(mod2.3)[2]/coef(mod2.3)[1])  ## = 5.208521 m/s
```

```
## tVelocity
##  5.208521
```

```
abline(v=h.minW, col=5, lty=6)
abline(h=0)
```



Exercise 3

```r
path <- file.path('Datasets', 'NPScosts.dat')
df <- read.table(path, header = TRUE)

summary(df)
```

Loading and checking the data

```
##       cost            date             t1              t2
##  Min.   :207.5   Min.   :67.17   Min.   : 7.00   Min.   :44.00
##  1st Qu.:310.3   1st Qu.:67.90   1st Qu.:11.75   1st Qu.:56.50
##  Median :448.1   Median :68.42   Median :13.00   Median :62.50
##  Mean   :461.6   Mean   :68.58   Mean   :13.75   Mean   :62.38
##  3rd Qu.:612.0   3rd Qu.:68.92   3rd Qu.:15.25   3rd Qu.:70.25
##  Max.   :881.2   Max.   :71.08   Max.   :22.00   Max.   :85.00
##       cap              pr               ne             ct
##  Min.   : 457.0   Min.   :0.0000   Min.   :0.00   Min.   :0.0000
##  1st Qu.: 745.0   1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.0000
##  Median : 822.0   Median :0.0000   Median :0.00   Median :0.0000
##  Mean   : 825.4   Mean   :0.3125   Mean   :0.25   Mean   :0.4062
##  3rd Qu.: 947.2   3rd Qu.:1.0000   3rd Qu.:0.25   3rd Qu.:1.0000
```

```
##  Max.   :1130.0   Max.   :1.0000   Max.   :1.00   Max.   :1.0000
##        bw              cum.n              pt
##  Min.   :0.0000   Min.   : 1.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 3.000   1st Qu.:0.0000
##  Median :0.0000   Median : 7.500   Median :0.0000
##  Mean   :0.1875   Mean   : 8.531   Mean   :0.1875
##  3rd Qu.:0.0000   3rd Qu.:12.500   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :21.000   Max.   :1.0000
```

```
dim(df)
```

```
## [1] 32 11
```

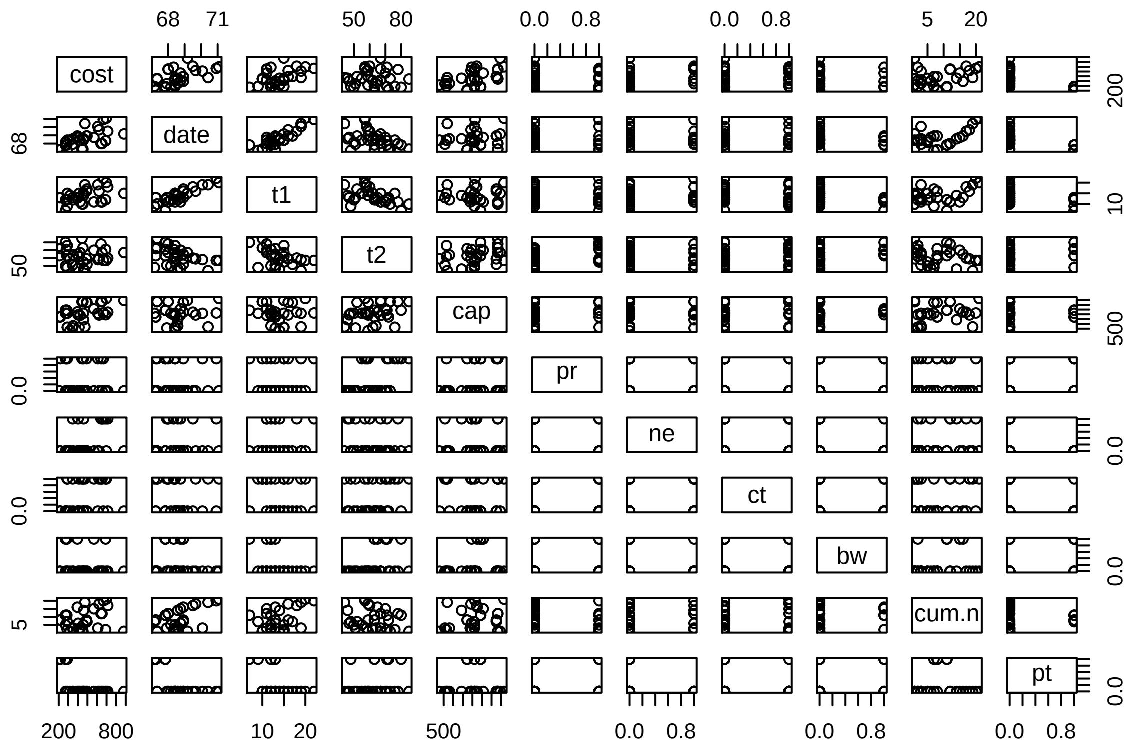
```
head(df)
```

```
##     cost  date t1 t2  cap pr ne ct bw cum.n pt
## 1 460.05 68.58 14 46  687  0  1  0  0    14  0
## 2 452.99 67.33 10 73 1065  0  0  1  0     1  0
## 3 443.22 67.33 10 85 1065  1  0  1  0     1  0
## 4 652.32 68.00 11 67 1065  0  1  1  0    12  0
## 5 642.23 68.00 11 78 1065  1  1  1  0    12  0
## 6 345.39 67.92 13 51  514  0  1  1  0     3  0
```

```
tail(df)
```

```
##      cost  date t1 t2 cap pr ne ct bw cum.n pt
## 27 207.51 67.25 13 63 745  0  0  0  0     8  1
## 28 288.48 67.17  9 48 821  0  0  1  0     7  1
## 29 284.88 67.83 12 63 886  0  0  0  1    11  1
## 30 280.36 67.83 12 71 886  1  0  0  1    11  1
## 31 217.38 67.25 13 72 745  1  0  0  0     8  1
## 32 270.71 67.83  7 80 886  1  0  0  1    11  1
```

```
plot(df)
```



Data looks fine.

Exercise 3.a)

Tukey's first-aid transformations

```
df$lCost <- log(df$cost)
# df$sDate <- sqrt(df$date) -> Time goes in linearly
# df$sT1 <- sqrt(df$t1)
# df$sT2 <- sqrt(df$t2)
df$lCap <- log(df$cap)
df$sCum.n <- sqrt(df$cum.n)

head(df)
```

```
##     cost  date t1 t2  cap pr ne ct bw cum.n pt    lCost     lCap   sCum.n
## 1 460.05 68.58 14 46  687  0  1  0  0    14  0 6.131335 6.532334 3.741657
## 2 452.99 67.33 10 73 1065  0  0  1  0     1  0 6.115870 6.970730 1.000000
## 3 443.22 67.33 10 85 1065  1  0  1  0     1  0 6.094066 6.970730 1.000000
## 4 652.32 68.00 11 67 1065  0  1  1  0    12  0 6.480535 6.970730 3.464102
## 5 642.23 68.00 11 78 1065  1  1  1  0    12  0 6.464946 6.970730 3.464102
## 6 345.39 67.92 13 51  514  0  1  1  0     3  0 5.844674 6.242223 1.732051
```

The transformations look good

Exercise 3.b)

Fitting a regression model including all explanator variables. Does partial turnkey guarantee affect the costs of the plants significantly on the 5% level?

```
mod3.1 <- lm(lCost ~ date + t1 + t2 + lCap + pr + ne + ct + bw + sCum.n + pt,
             data = df)
summary(mod3.1)
```

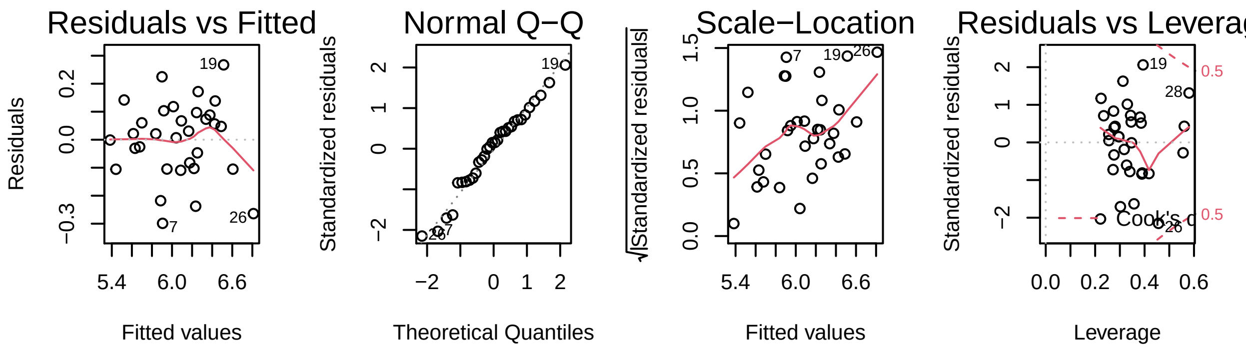
```
##
## Call:
## lm(formula = lCost ~ date + t1 + t2 + lCap + pr + ne + ct + bw +
##     sCum.n + pt, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29896 -0.10332  0.02118  0.09019  0.26731
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.875045   5.404826  -2.567  0.01796 *
## date          0.219318   0.082426   2.661  0.01463 *
## t1            0.006067   0.021990   0.276  0.78531
## t2            0.005273   0.004564   1.155  0.26092
## lCap          0.692542   0.137131   5.050 5.32e-05 ***
## pr           -0.105307   0.082004  -1.284  0.21307
## ne            0.254326   0.078075   3.257  0.00377 **
## ct            0.122969   0.068386   1.798  0.08654 .
## bw            0.029418   0.104469   0.282  0.78101
## sCum.n       -0.069016   0.040985  -1.684  0.10700
## pt           -0.229133   0.128059  -1.789  0.08800 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1666 on 21 degrees of freedom
## Multiple R-squared:  0.8684, Adjusted R-squared:  0.8057
## F-statistic: 13.85 on 10 and 21 DF,  p-value: 3.983e-07
```

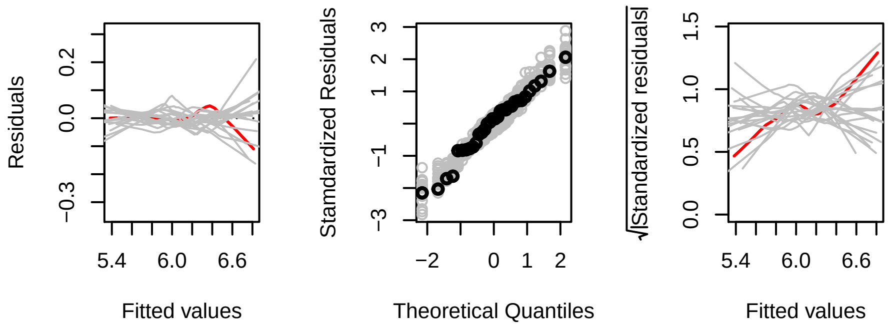
The partial turnkey guarantee does not affect the costs of the plants significantly on the 5% level in this model.

Exercise 3.c)

Assessing the quality of the fitted regression model by a residual and sensitivity analysis:

```
par(mfrow = c(2, 4))
plot(mod3.1)
plot.lmSim(mod3.1, SEED = 1)
```





Interpretation:

1. Tukey-Anscombe plot: The smoother shows a strong decreasing trend in the last third which is, however, within the stochastic fluctuation.
   => The assumption of constant expectation of the errors is not violated.
2. Q-Q plot: The observations lie almost perfect on the straight line and are within the stochastic fluctuation.
   => The assumption of Gaussian distributed errors is not violated.
3. Scale-location plot: The smoother shows a wavy increasing line which is within the stochastic fluctuation.
   => The assumption of constant variance of the errors is not violated.
4. Residuals vs. Leverage: No Observation has Cook's Distance >1. And there are no leverage points with leverage $> 2 * 11/32 = 0.6875$ (32 examples in dataset, 11 variables in model)
   => No dangerous or too influential residuals identified.

CONCLUSION: The model does fit adequately the data.