

Assumptions linear models

Lorenz Gygax

2020-02-13

To check assumptions, we are going back to the brussels sprouts data set that we have analysed using a regression. We first have a look at the automatically generated plots and later see how we can draw these plots by hand based on a second example.

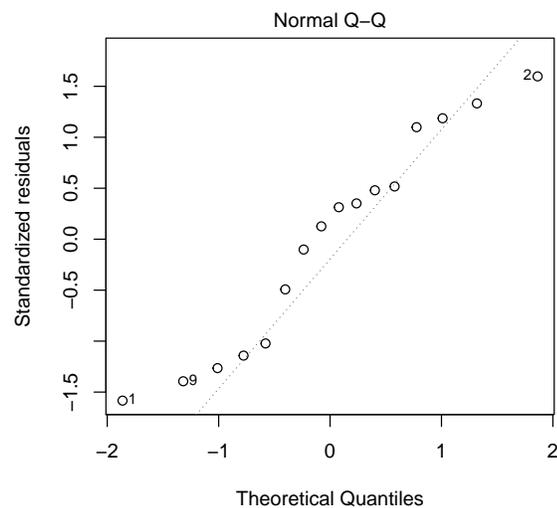
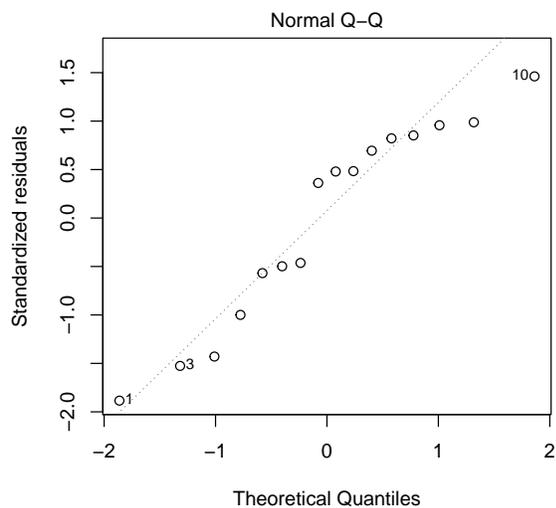
Residuals regression

If no longer available, we re-calculate the linear and the quadratic model from the regression exercise so that we can directly compare the residuals of these two model.

```
rkohl.df <- read.table('Rosenkohl.csv', header= TRUE, sep= ';', dec= '.')
rkohl.df[, 'SW2'] <- rkohl.df[, 'Standweite']^2
rkohl.lin <- lm(Ertrag ~ Standweite, rkohl.df)
rkohl.qua <- lm(Ertrag ~ Standweite + SW2, rkohl.df)
```

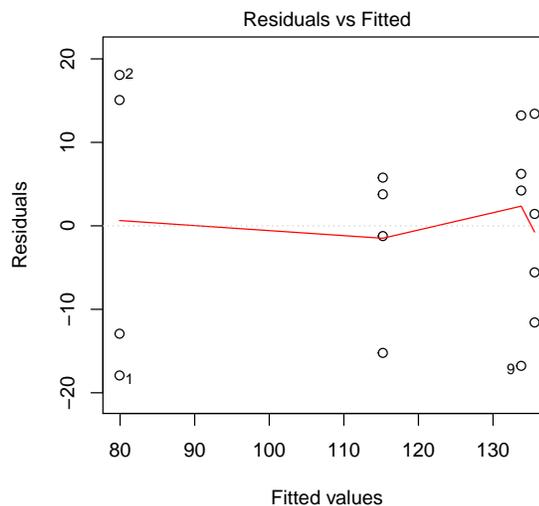
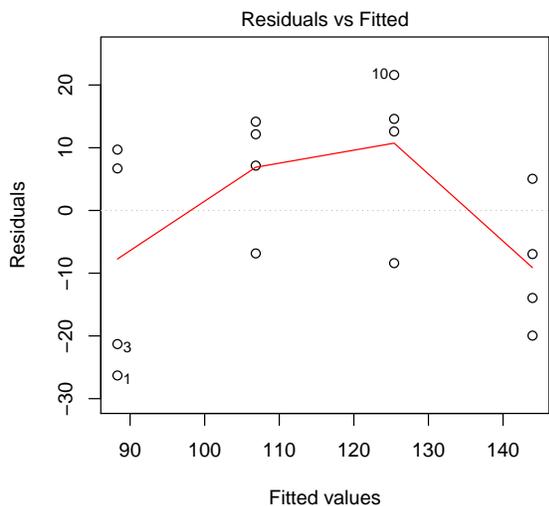
QQ-plot

```
par(mfrow= c(1, 2))
plot(rkohl.lin, which= 2) ## QQ-plot
plot(rkohl.qua, which= 2) ## QQ-plot
```

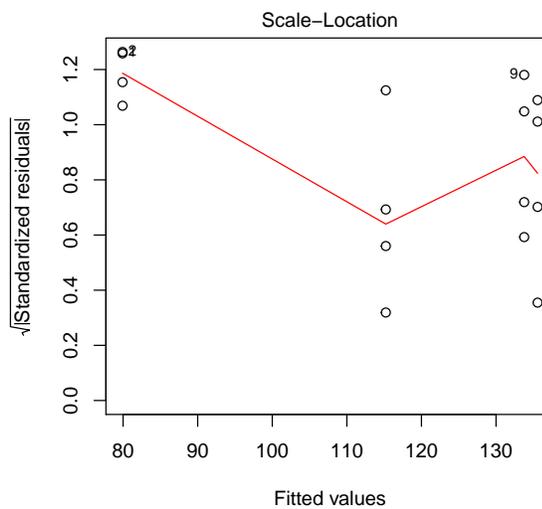
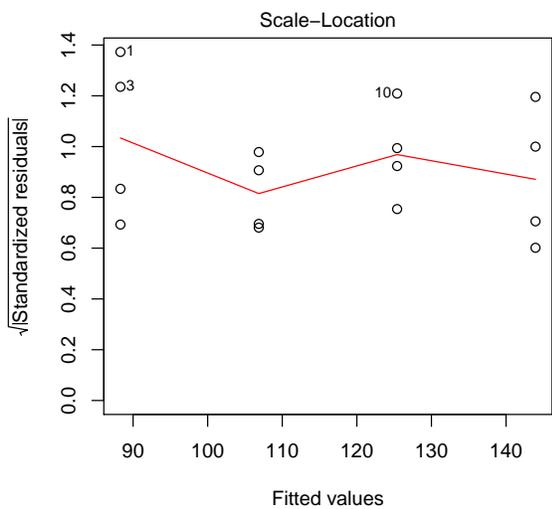


Tukey-Anscombe plot

```
par (mfrow= c (1, 2))
plot (rkohl.lin, which= 1)    ## Tukey-Anscombe Plot
plot (rkohl.qua, which= 1)   ## Tukey-Anscombe Plot
```

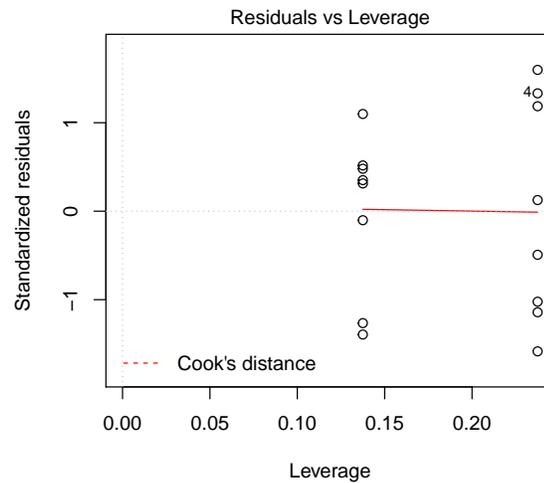
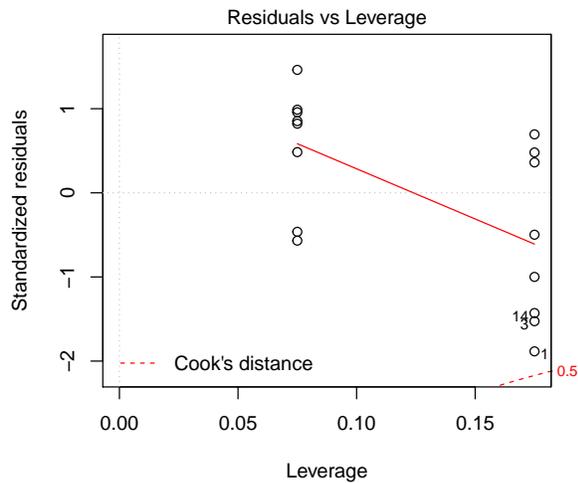


```
plot (rkohl.lin, which= 3)    ## Variant Tukey-Anscombe Plot
plot (rkohl.qua, which= 3)   ## Variant Tukey-Anscombe Plot
```



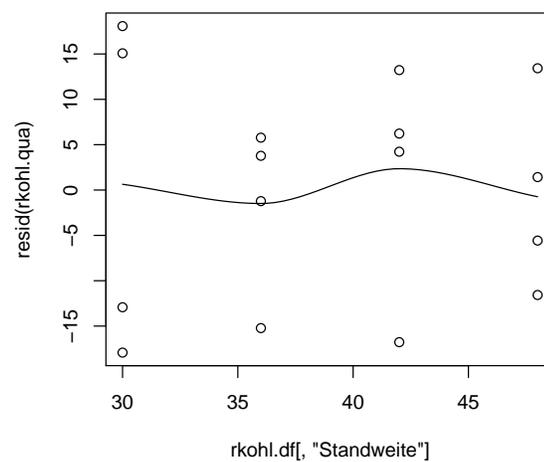
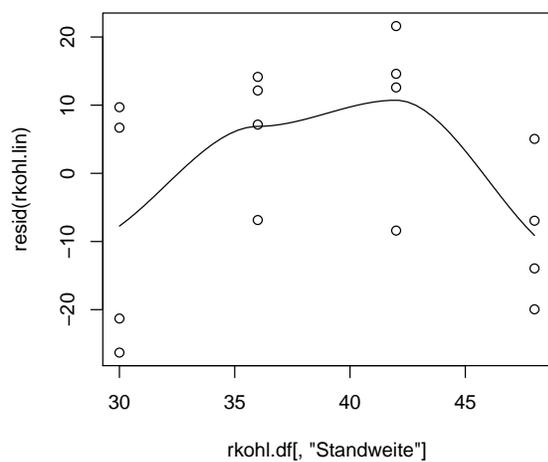
Leverage plots

```
par (mfrow= c (1, 2))
plot (rkohl.lin, which= 5)      ## Leverage Plot
plot (rkohl.qua, which= 5)     ## Leverage Plot
```



Residuals versus explanatory variable

```
par (mfrow= c (1, 2))
scatter.smooth (rkohl.df [, 'Standweite'], resid (rkohl.lin)) ## Resid vs. explanatory
scatter.smooth (rkohl.df [, 'Standweite'], resid (rkohl.qua)) ## Resid vs. explanatory
```



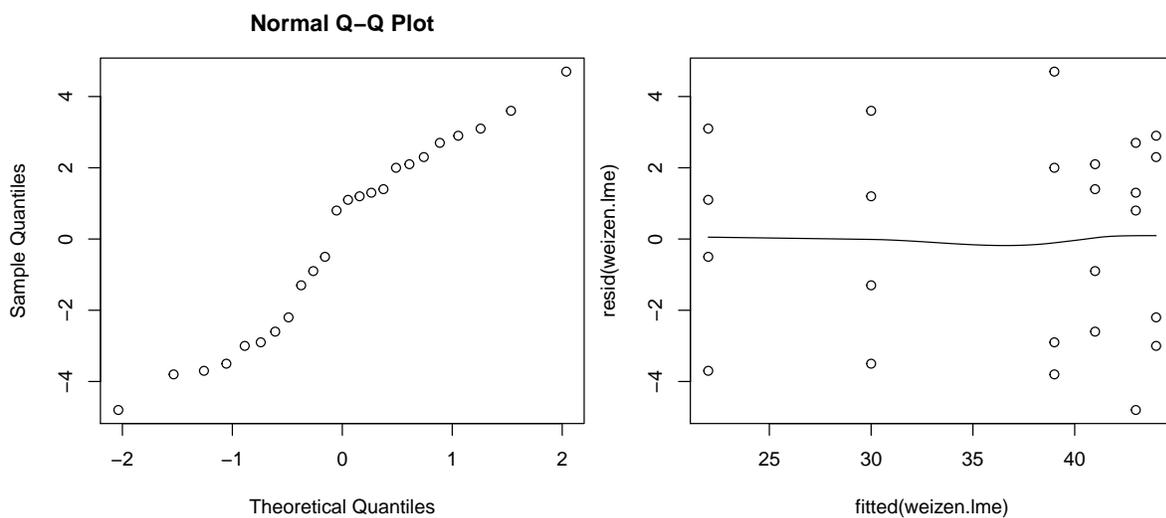
Assess the residuals and compare the residuals of the two models.

Residuals analysis of variance with two explanatory variables

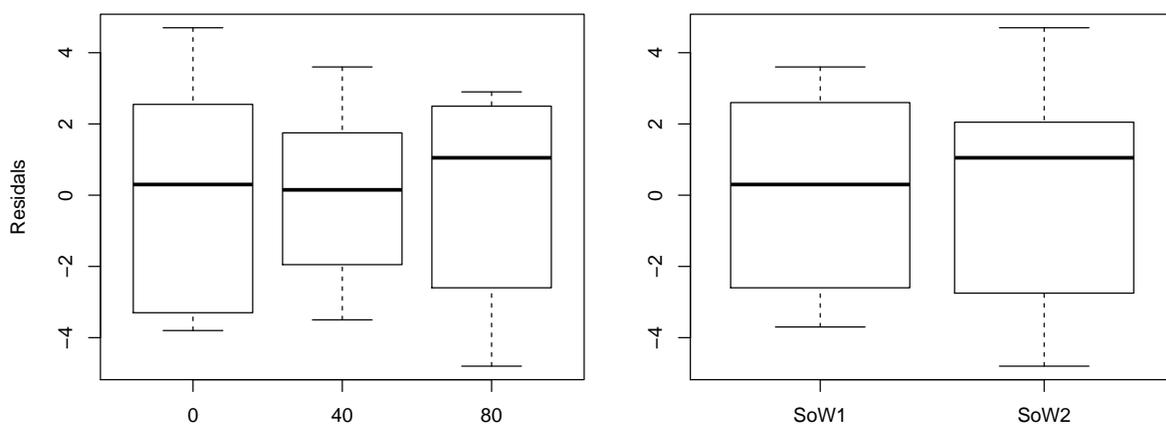
We again look at a data set in which production (“Kornertrag”) of wheat is measured in dependence of the amount of added nitrogen (“Stickstoff”) and the type (“Sorte”). We again assume that all the data are independent.

```
weizen.df <- read.table ('SommerWeizen.csv', header= TRUE, sep= ';', dec= '.')
weizen.df [, 'Stickstoff'] <- factor (weizen.df [, 'Stickstoff'])
weizen.lme <- lm (Kornertrag ~ Stickstoff * Sorte, weizen.df)
```

```
par (mfrow= c (1, 2), mar= c (4, 4, 3, 0.5))
qqnorm (resid (weizen.lme))      ## QQ-plot Residuen
scatter.smooth (fitted (weizen.lme), resid (weizen.lme))
```



```
boxplot (split (resid (weizen.lme), weizen.df [, 'Stickstoff']), ylab= 'Residuals')
boxplot (split (resid (weizen.lme), weizen.df [, 'Sorte']))
```



Assess these residuals. Is any action necessary? What would be the adequate first-aid transformation? Try and re-check the residuals.