Solution to Exercise 3

1. Estimate all effects in the following 3 × 3 designs. Do interactions exist?

a)

		B 1	B 2	B 3	Total
A	1	10	15	20	15
	2	10	15	20	15
	3	10	15	20	15
Total		10	15	20	15

interaction effects		B 1	B 2	B 3	main effects A
A	1	0	0	0	0
	2	0	0	0	0
	3	0	0	0	0
main effects B		-5	0	5	$\hat{\mu} = 15$

b)

		B 1	B 2	B 3	Total
A	1	26	22	21	23
	2	23	19	18	20
	3	17	13	12	14
Total		22	18	17	19

interaction effects		B 1	B 2	B 3	main effects A
A	1	0	0	0	4
	2	0	0	0	1
	3	0	0	0	-5
main effects B		3	-1	-2	$\hat{\mu} = 19$

c)

		B 1	B 2	B 3	Total
A	1	26	23	20	23
	2	18	19	23	20
	3	13	15	14	14
Total		19	19	19	19

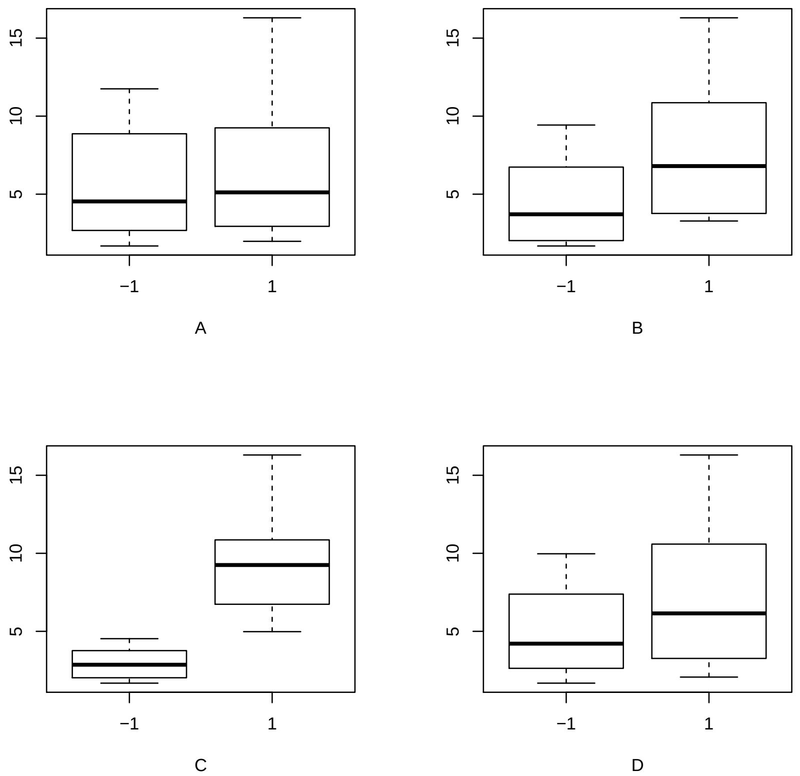
interaction effects		B 1	B 2	B 3	main effects A
A	1	3	0	-3	4
	2	-2	-1	3	1
	3	-1	1	0	-5
main effects B		0	0	0	$\hat{\mu} = 19$

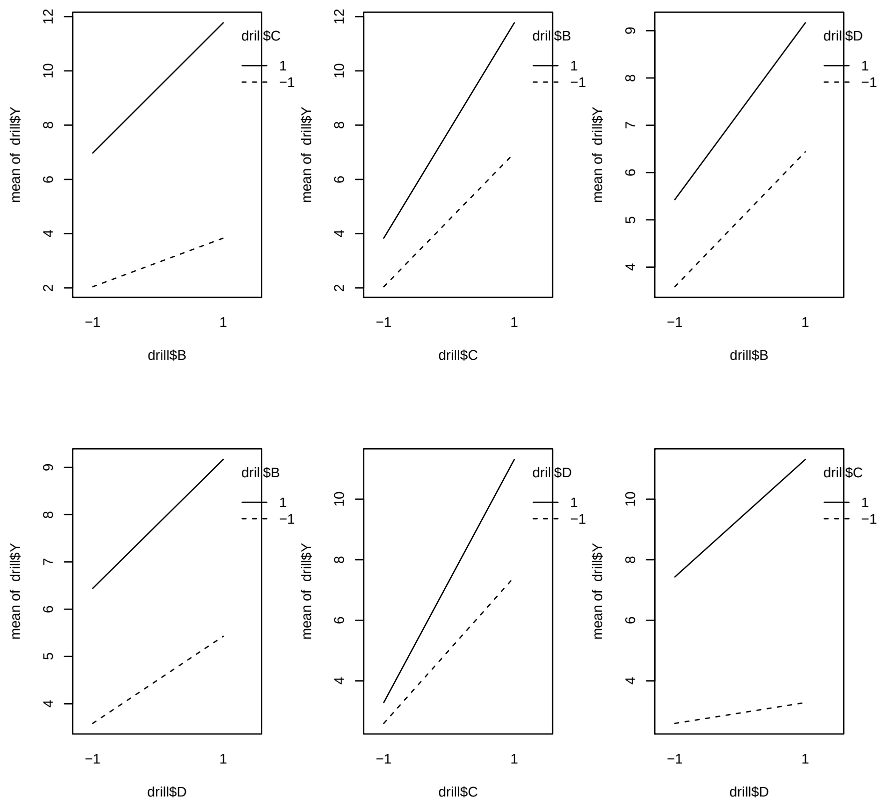
2. Read in the data:

```
drill <- read.table(file="http://stat.ethz.ch/Teaching/Datasets/drill.txt",header=TRUE)

str(drill)

drill$A <- as.factor(drill$A)
drill$B <- as.factor(drill$B)
drill$C <- as.factor(drill$C)
drill$D <- as.factor(drill$D)
```

a) Plot the data with:

```
par(mfrow=c(2,2))
plot(drill$A,drill$Y,xlab="A")
plot(drill$B,drill$Y,xlab="B")
plot(drill$C,drill$Y,xlab="C")
plot(drill$D,drill$Y,xlab="D")
```





From the plots we see that there could be an significant effect for the factors B, C and D but probably not for A. Also the interactions BC and CD look quite promising from the interaction plots.

b) Analysis with all main effects and all interactions:

```
mod1 <- aov(Y~A*B*C*D,data=drill)
summary(mod1)
```

```
            Df  Sum Sq Mean Sq
A            1   3.331   3.331
B            1  43.494  43.494
C            1 165.508 165.508
D            1  20.885  20.885
A:B          1   0.090   0.090
A:C          1   1.416   1.416
B:C          1   9.060   9.060
A:D          1   2.839   2.839
B:D          1   0.783   0.783
C:D          1  10.208  10.208
A:B:C        1   0.112   0.112
A:B:D        1   1.392   1.392
A:C:D        1   2.280   2.280
B:C:D        1   0.130   0.130
A:B:C:D      1   1.156   1.156
```

We get a strange result due to the fact that we do not have any degrees of freedom left for the residuals. We have 15 effects for factors and the overall mean with, that is 16 df and only 16 observations. The model is therefore saturated.

c) Analysis with all main effects and all 2-fold interactions:

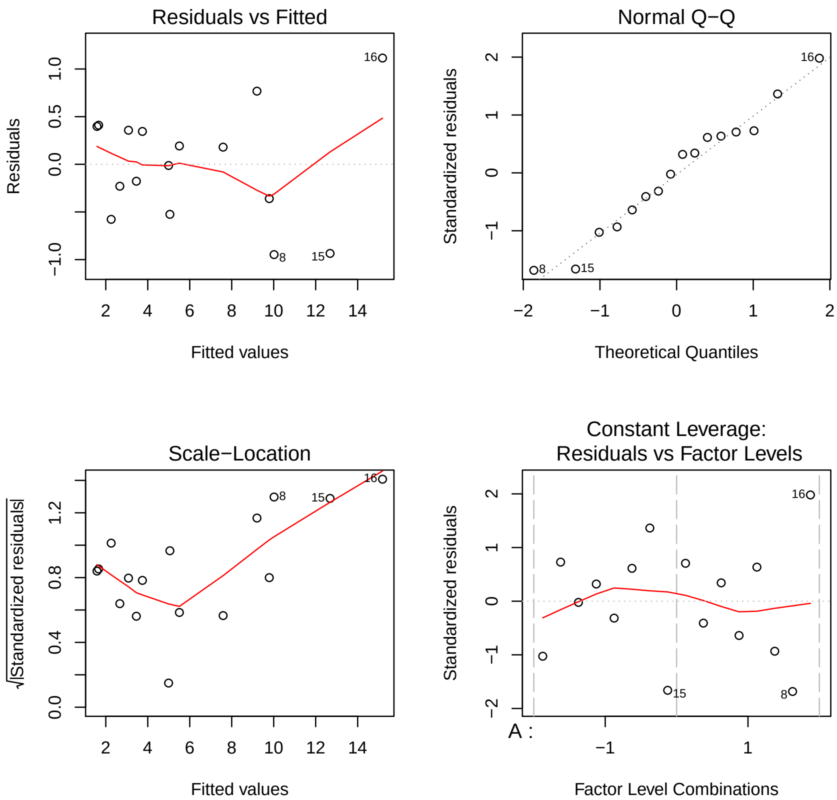
```
mod2 <- aov(Y~A+B+C+D+A:B+A:C+A:D+B:C+B:D+C:D,data=drill)
summary(mod2)
```

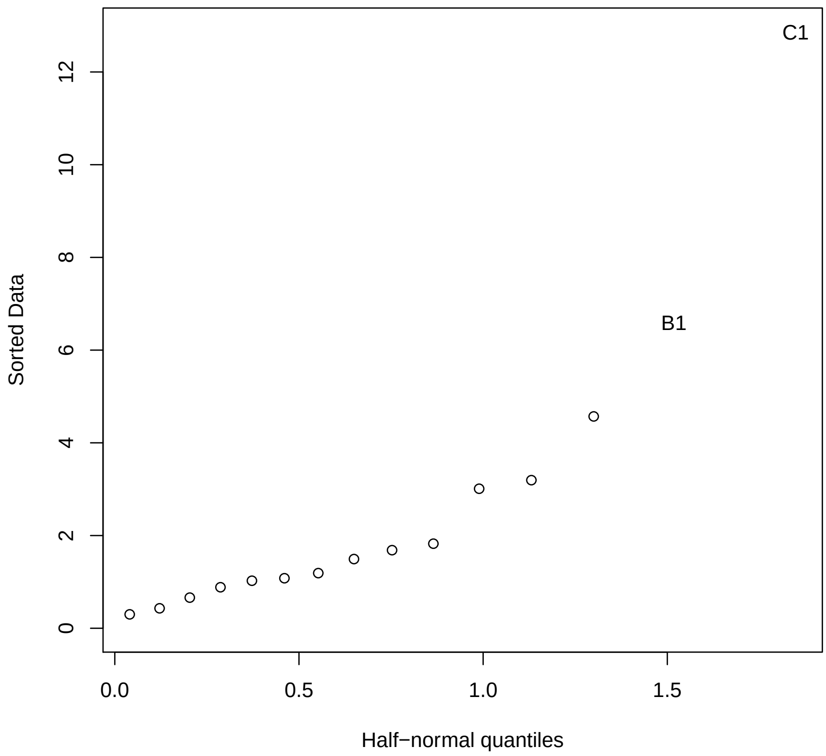
```
            Df  Sum Sq Mean Sq  F value    Pr(>F)
A            1   3.331   3.331   3.2847  0.129677
B            1  43.494  43.494  42.8939  0.001243 **
C            1 165.508 165.508 163.2247 5.227e-05 ***
D            1  20.885  20.885  20.5968  0.006178 **
A:B          1   0.090   0.090   0.0888  0.777744
A:C          1   1.416   1.416   1.3966  0.290438
A:D          1   2.839   2.839   2.8001  0.155118
B:C          1   9.060   9.060   8.9351  0.030477 *
B:D          1   0.783   0.783   0.7724  0.419694
C:D          1  10.208  10.208  10.0672  0.024735 *
Residuals    5   5.070   1.014
```

We see that the main factors B, C and D are significant on a 5% level as well as the 2-fold interactions BC and CD.

d) Check the residuals:

```
par(mfrow=c(2,2))
plot(mod2)

library(faraway)
halfnorm(mod2$effects[-1],labs=names(mod2$effects[-1]))
```

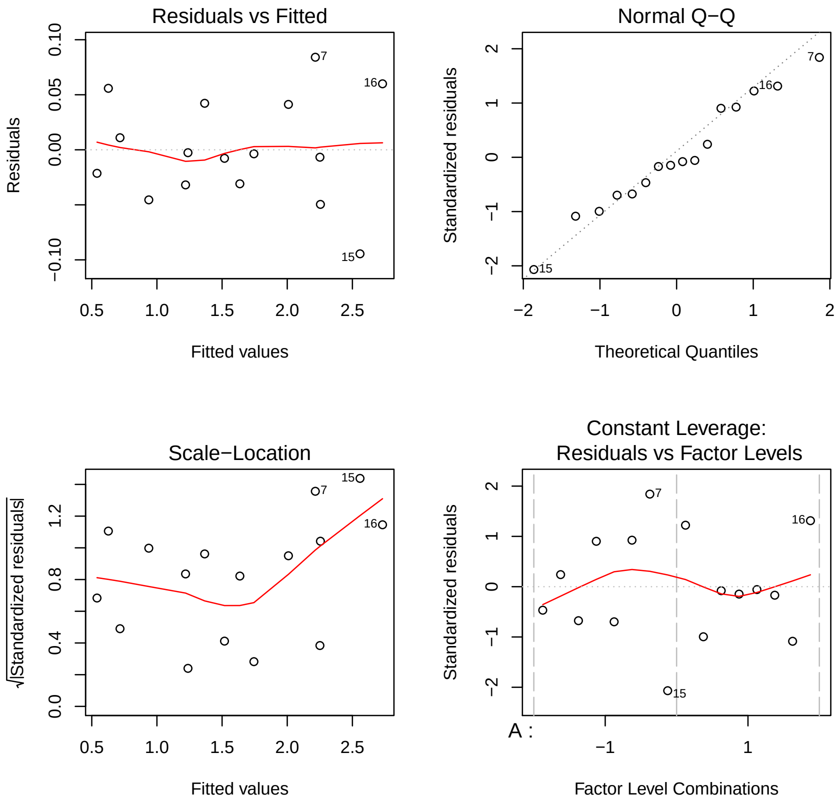




We see that probably the heteroscedasticity assumption is violated. We try a log-transform of the response variable.

```
drill.e <- drill
```

```
drill.e$Y <- log(drill$Y)

mod3 <- aov(Y~A+B+C+D+A:B+A:C+A:D+B:C+B:D+C:D,data=drill.e)
summary(mod3)

par(mfrow=c(2,2))
plot(mod3)
```



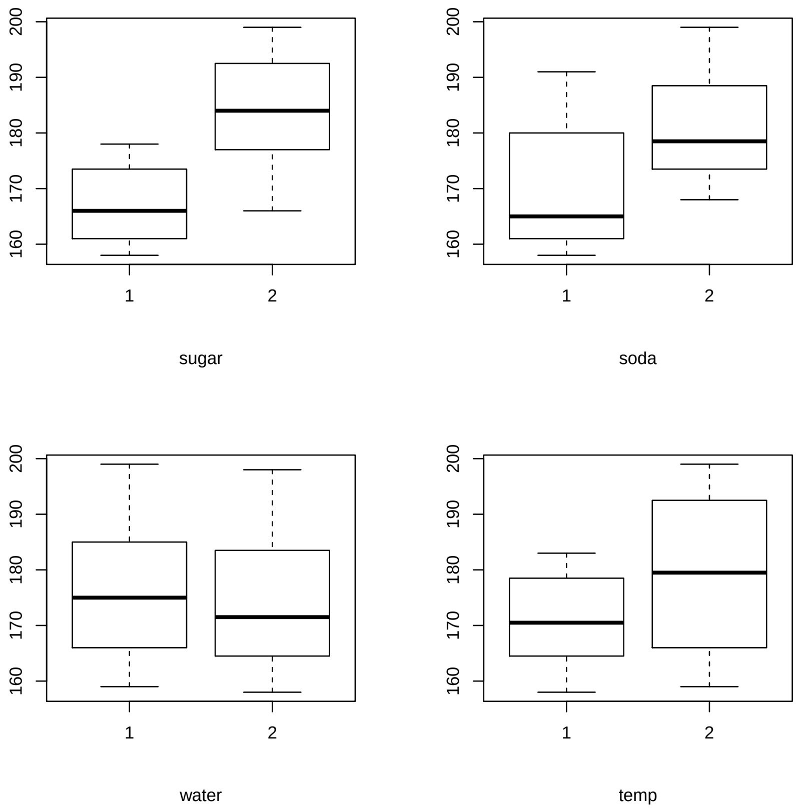
We see that the Tukey-Anscombe plot looks better after the log-transform.

3. Read in the data:

```
soft <- read.table(file="http://stat.ethz.ch/Teaching/Datasets/
softdrinkANOVA.txt",header=TRUE)

soft$sugar <- as.factor(soft$sugar)
soft$soda <- as.factor(soft$soda)
soft$water <- as.factor(soft$water)
soft$temp <- as.factor(soft$temp)
```

a) Plot the data.

```
par(mfrow=c(2,2))
plot(soft$sugar,soft$score,sub="sugar")
plot(soft$soda,soft$score,sub="soda")
plot(soft$water,soft$score,sub="water")
plot(soft$temp,soft$score,sub="temp")
```



From the plots we can say that probably the factors sugar, soda and temp have an significant influence on the flavor of softdrinks.

b) Analyze the data. Which factors are important?

```
modS <- aov(score~sugar*soda*water*temp,data=soft)
summary(modS)
```

```
                     Df  Sum Sq Mean Sq  F value    Pr(>F)
sugar                 1 2312.00 2312.00 241.7778 4.451e-11 ***
soda                  1  946.12  946.12  98.9412 2.958e-08 ***
water                 1   21.12   21.12   2.2092    0.1566
temp                  1  561.13  561.13  58.6797 9.692e-07 ***
sugar:soda            1    3.13    3.13   0.3268    0.5755
sugar:water           1    0.13    0.13   0.0131    0.9104
soda:water            1    0.50    0.50   0.0523    0.8220
sugar:temp            1  666.12  666.12  69.6601 3.187e-07 ***
soda:temp             1   12.50   12.50   1.3072    0.2697
water:temp            1   12.50   12.50   1.3072    0.2697
sugar:soda:water      1    4.50    4.50   0.4706    0.5025
sugar:soda:temp       1    0.00    0.00   0.0000    1.0000
sugar:water:temp      1    2.00    2.00   0.2092    0.6536
soda:water:temp       1    0.13    0.13   0.0131    0.9104
sugar:soda:water:temp 1   21.13   21.13   2.2092    0.1566
Residuals            16  153.00    9.56
```

We see that the factors sugar, soda and temp are very significant. This supports our sugestions from task **a)**. We also see that the interaction of sugar and temp is highly significant.

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

```
library(faraway)
halfnorm(modS$effects[-1],labs=names(modS$effects[-1]))
```