Solution to Series 3

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

a)

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

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

b)

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

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

c)

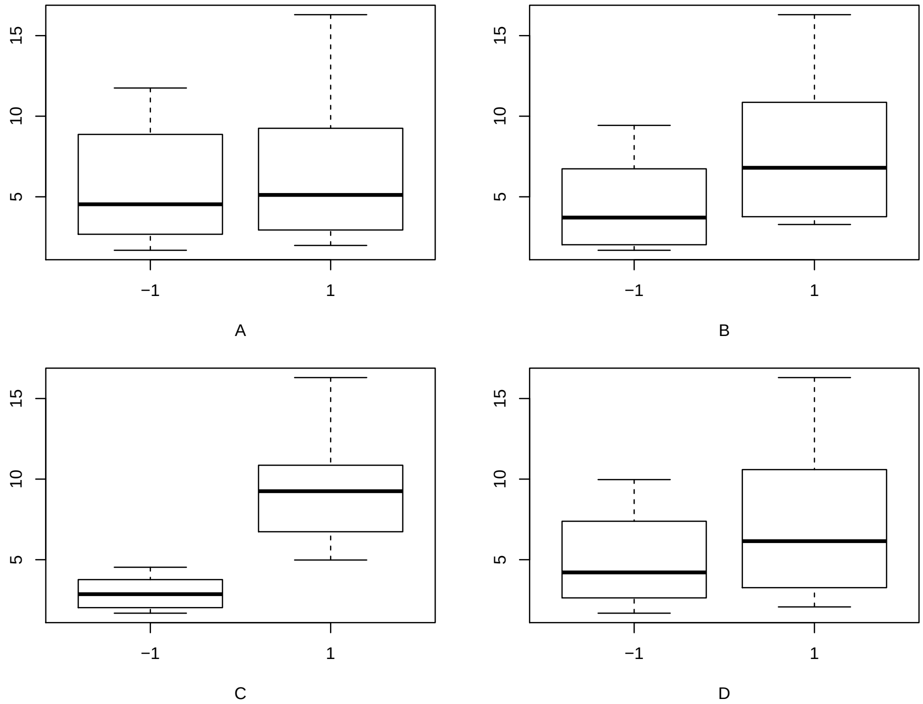
		B			Total
		1	2	3	
	1	26	23	20	23
A	2	18	19	23	20
	3	13	15	14	14
Total		19	19	19	19

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

2. a) Plot the data.

```
> drill <- read.table(file="http://stat.ethz.ch/Teaching/Datasets/drill.txt",header=TRUE)
> drill$A <- as.factor(drill$A)
> drill$B <- as.factor(drill$B)
> drill$C <- as.factor(drill$C)
> drill$D <- as.factor(drill$D)

> 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) Do an 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   3.33    3.33
B            1  43.49   43.49
C            1 165.51  165.51
D            1  20.88   20.88
A:B          1   0.09    0.09
A:C          1   1.42    1.42
B:C          1   9.06    9.06
A:D          1   2.84    2.84
B:D          1   0.78    0.78
C:D          1  10.21   10.21
A:B:C        1   0.11    0.11
A:B:D        1   1.39    1.39
A:C:D        1   2.28    2.28
B:C:D        1   0.13    0.13
A:B:C:D      1   1.16    1.16
```

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) Do an 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)
```

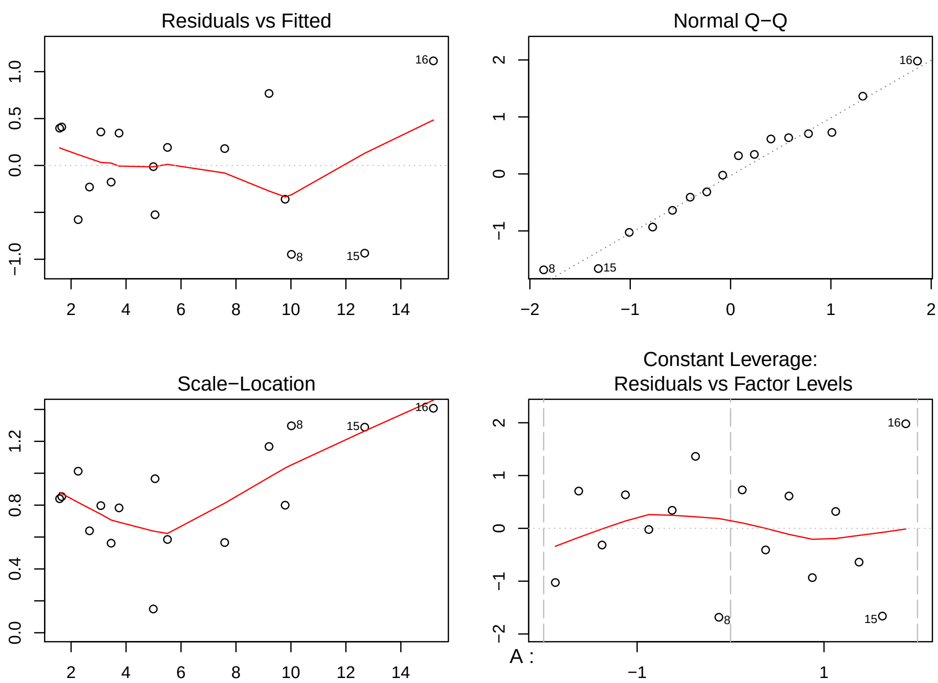
```
            Df Sum Sq Mean Sq F value   Pr(>F)
A            1   3.33    3.33   3.285  0.12968
B            1  43.49   43.49  42.894  0.00124 **
C            1 165.51  165.51 163.225 5.23e-05 ***
D            1  20.88   20.88  20.597  0.00618 **
A:B          1   0.09    0.09   0.089  0.77774
A:C          1   1.42    1.42   1.397  0.29044
```

```
A:D          1   2.84    2.84   2.800  0.15512
B:C          1   9.06    9.06   8.935  0.03048 *
B:D          1   0.78    0.78   0.772  0.41969
C:D          1  10.21   10.21  10.067  0.02473 *
Residuals    5   5.07    1.01
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

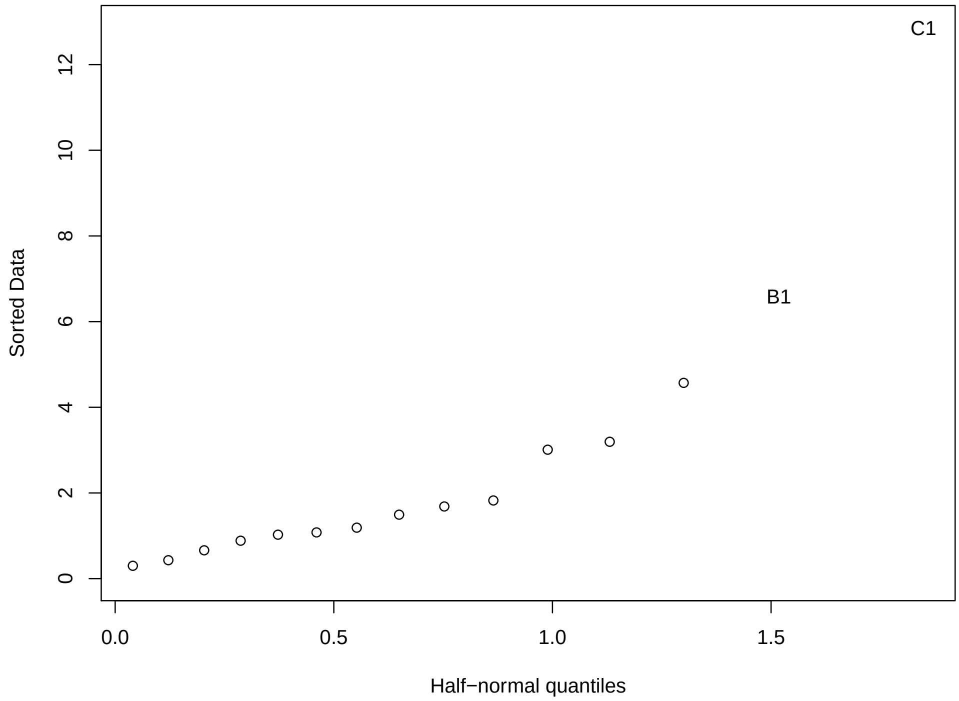
We see that the main factors B, C and D are significant on a 5% level as well as the 2-fold interactions BC and CD.

d) Check the residuals and improve your model if necessary.

```
> par(mfrow=c(2,2),mar=c(3,2,3,2))
> plot(mod2)
```

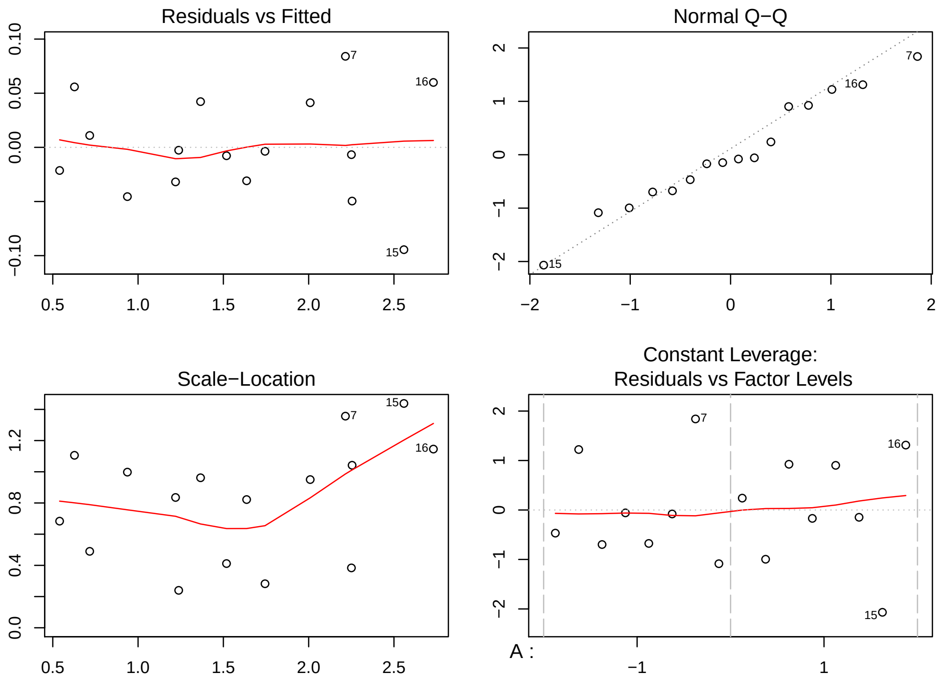



```
> library(faraway)
> halfnorm(mod2$effects[-1],labs=names(mod2$effects[-1]))
```



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)
            Df Sum Sq Mean Sq F value   Pr(>F)
A            1  0.068   0.068  10.119 0.024511 *
B            1  1.346   1.346 201.504 3.12e-05 ***
C            1  5.331   5.331 798.098 1.04e-06 ***
D            1  0.427   0.427  63.854 0.000496 ***
A:B          1  0.005   0.005   0.707 0.438754
A:C          1  0.000   0.000   0.064 0.810121
A:D          1  0.018   0.018   2.680 0.162530
B:C          1  0.010   0.010   1.509 0.273906
B:D          1  0.001   0.001   0.134 0.729647
C:D          1  0.039   0.039   5.768 0.061498 .
Residuals    5  0.033   0.007
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> par(mfrow=c(2,2),mar=c(3,2,3,2))
> plot(mod3)
```

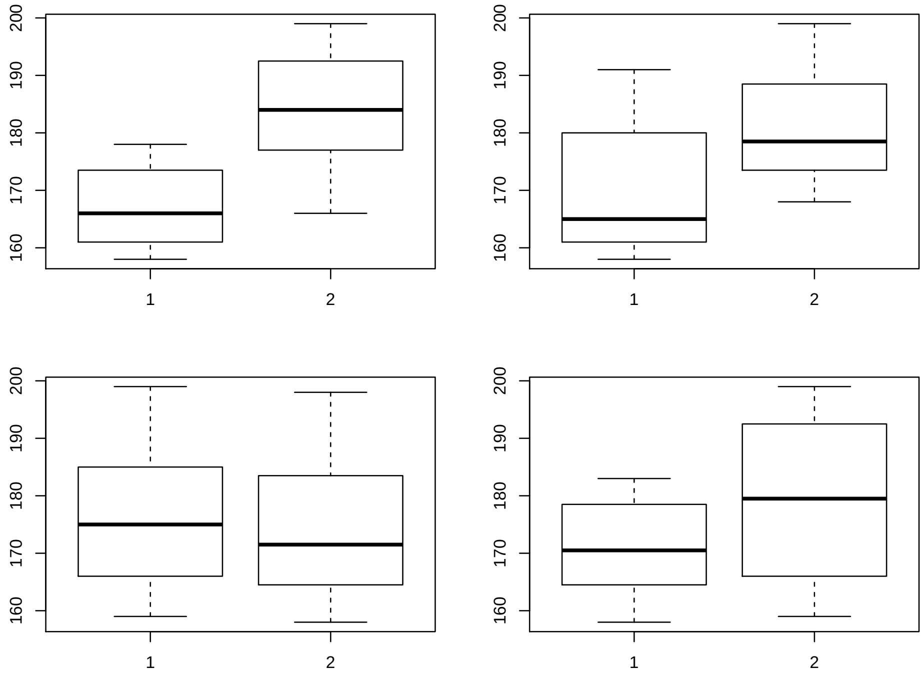


We see that the Tukey-Anscombe plot looks better after the log-transform.

3. **a)** Plot the data.

```
> soft <- read.table(file="http://stat.ethz.ch/Teaching/Datasets/softdrinkANOVA.txt",header=TR
> soft$sugar <- as.factor(soft$sugar)
> soft$soda <- as.factor(soft$soda)
> soft$water <- as.factor(soft$water)
> soft$temp <- as.factor(soft$temp)

> 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?

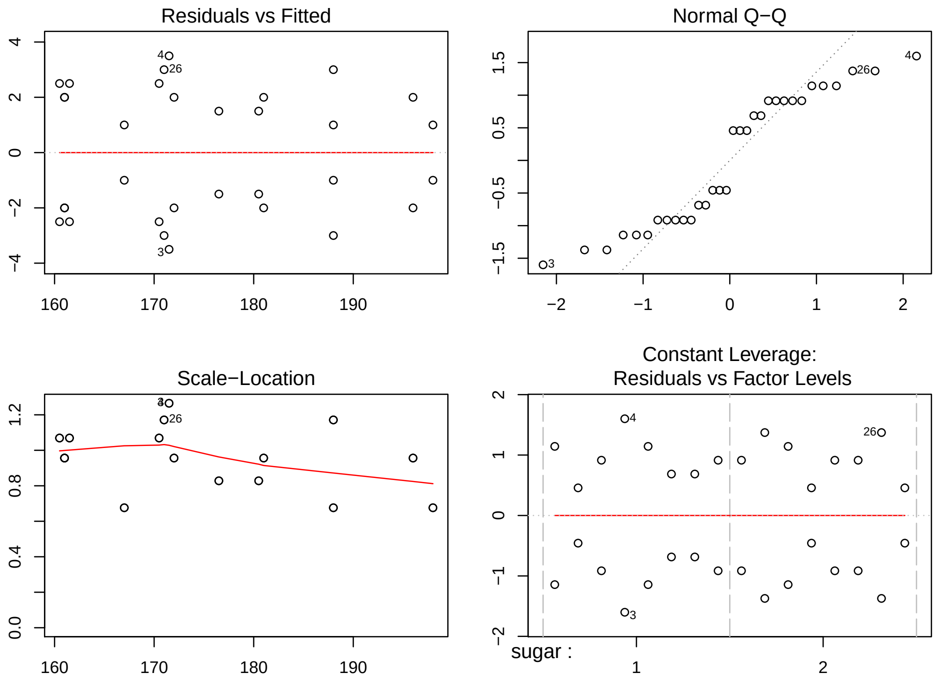
```
> modS <- aov(score~sugar*soda*water*temp,data=soft)
> summary(modS)
                      Df Sum Sq Mean Sq F value   Pr(>F)
sugar                  1 2312.0  2312.0 241.778 4.45e-11
soda                   1  946.1   946.1  98.941 2.96e-08
water                  1   21.1    21.1   2.209    0.157
temp                   1  561.1   561.1  58.680 9.69e-07
sugar:soda             1    3.1     3.1   0.327    0.575
sugar:water            1    0.1     0.1   0.013    0.910
soda:water             1    0.5     0.5   0.052    0.822
sugar:temp             1  666.1   666.1  69.660 3.19e-07
soda:temp              1   12.5    12.5   1.307    0.270
water:temp             1   12.5    12.5   1.307    0.270
sugar:soda:water       1    4.5     4.5   0.471    0.503
sugar:soda:temp        1    0.0     0.0   0.000    1.000
sugar:water:temp       1    2.0     2.0   0.209    0.654
soda:water:temp        1    0.1     0.1   0.013    0.910
sugar:soda:water:temp  1   21.1    21.1   2.209    0.157
Residuals             16  153.0     9.6

sugar                 ***
soda                  ***
water
temp                  ***
sugar:soda
sugar:water
soda:water
sugar:temp            ***
soda:temp
water:temp
sugar:soda:water
sugar:soda:temp
sugar:water:temp
```

```
soda:water:temp
sugar:soda:water:temp
Residuals
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

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),mar=c(3,2,3,2))
> plot(modS)
```




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