Solution Exercise 4

1. Read in the data:

```
feed <- read.table(file="../feed.txt",header=TRUE)
feed$Feeding <- as.factor(feed$Feeding)
```

a) Test for differences in treatment without taking into account the initial hormone concentration. Estimate the treatment means.

```
modF <- aov(Final~Feeding,data=feed)
summary(modF)

            Df  Sum Sq Mean Sq F value Pr(>F)
Feeding      2  1082.8  541.38  0.6287 0.5404
Residuals   29 24970.9  861.07

TukeyHSD(modF,"Feeding", conf.level=0.95)

Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Final ~ Feeding, data = feed)

$Feeding
         diff       lwr      upr     p adj
2-1 11.555556 -20.40036 43.51147 0.6489302
3-1 14.010101 -18.56238 46.58259 0.5446072
3-2  2.454545 -27.79581 32.70490 0.9781225
```

b) Carry out a one-way analysis of variance for the differences $D_i = Y_i - x_i$ of hormone measurements where Y_i is the response after the treatment and x_i the baseline measurement.

```
modF2 <- aov((Final-Initial)~Feeding,data=feed)
summary(modF2)

            Df Sum Sq Mean Sq F value Pr(>F)
Feeding      2  101.5   50.74  0.1537 0.8582
Residuals   29 9574.4  330.15
```

c) Include the baseline measurement in the model as a covariate and do an analysis of covariance for the responses Y_i. Estimate the adjusted treatment means.

```
modF3 <- aov(Final~Feeding+Initial,data=feed)
summary(modF3)

            Df  Sum Sq Mean Sq F value    Pr(>F)
Feeding      2  1082.8   541.4  10.447 0.0004079 ***
Initial      1 23520.0 23520.0 453.878 < 2.2e-16 ***
Residuals   28  1451.0    51.8
```

d) Compare and comment on the different results.

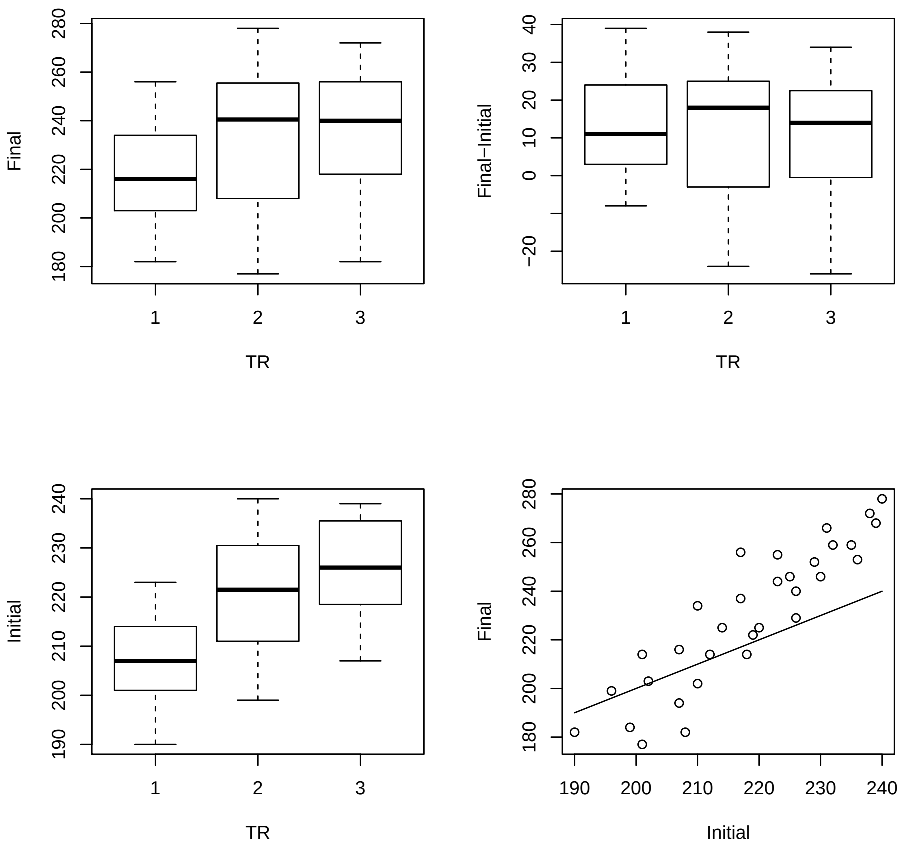
If we test for differences in treatment like in the model of task a) we see that there is no significant difference between food composition. From the one-way analysis of variance in b) we see that the factor feeding is not significant. It is even worse than in a). The output of task c) shows that the factor feeding is quite significant when we take the initial hormone concentration as a covariate into account.

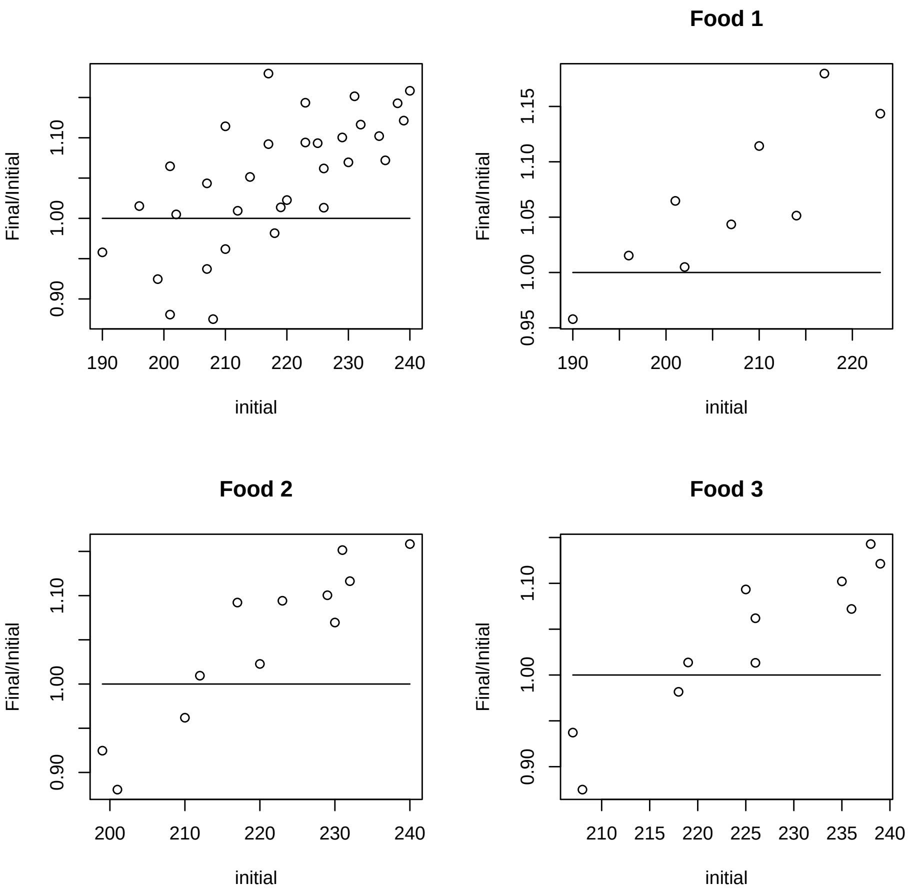
This is quite clear if we look at the data:

In the first and the second plot we can see that feeding does not provide a significant difference. But it is obvious that the animals will all grow up between the initial situation and the final situation for natural reasons. Therefore it makes sense to assume that weight at the end will be a multiplier of the initial weight plus the effect of the feeding. This can be checked in plots 3 and 5-8. In all cases the plot differs clearly from the line (this means the assumption that the only difference in weights is due to the feeding (and the variance) is not really plausible).

Last we look at the initial division of the animals (plot 4). We can see that animals are not really divided randomly (in the first group example we have smaller animals). With a randomized division of the animals we probably would have obtained better results even for the ANOVA-table.

When designing such a study aim at dividing probands at random into the different groups. In any case you have to avoid that all probands with a common feature are in the same group. This could lead to not noticing relevant effects or worse merging effects that do not exist!





2. We take the first replicate of the dataset softdrinkANOVA.txt, i.e.

	score	sugar	soda	water	temp
1	159	-1	-1	-1	-1
3	168	1	-1	-1	-1
5	158	-1	-1	1	-1
7	166	1	-1	1	-1
9	175	-1	1	-1	-1
11	179	1	1	-1	-1
13	173	-1	1	1	-1
15	179	1	1	1	-1
17	164	-1	-1	-1	1
19	187	1	-1	-1	1
21	163	-1	-1	1	1
23	185	1	-1	1	1
25	168	-1	1	-1	1
27	197	1	1	-1	1
29	170	-1	1	1	1
31	194	1	1	1	1

We have $16 = 2^4$ observations.
We wants to divide the observations in $\frac{16}{8} = 2$ different blocks such that we have a new factor (`BLOCK`) with 2 levels.
Construction of the experiment:
call:

```
A=sugar-effect
B=soda-effect
```

```
C=water-effect
D=temp-effect
E=BLOCK-effect
```

The values of A, B, C and D are 1 or -1 (or equivalently $+$ or $-$).
We just have to find the values of the column E to construct our experiment. We know that $E = A \cdot B \cdot C \cdot D$ (because $ABCD$ confounded) hence, the column `BLOCK` will be determined by multiplying the column of A, B, C and D. We let 1 correspond to the $first$ *block* and 2 correspond to the *second block*. We obtain:

```
> softBL
   score sugar soda water temp BLOCK
1    159    -1   -1    -1   -1     2
3    168     1   -1    -1   -1     1
5    158    -1   -1     1   -1     1
7    166     1   -1     1   -1     2
9    175    -1    1    -1   -1     1
11   179     1    1    -1   -1     2
13   173    -1    1     1   -1     2
15   179     1    1     1   -1     1
17   164    -1   -1    -1    1     1
19   187     1   -1    -1    1     2
21   163    -1   -1     1    1     2
23   185     1   -1     1    1     1
25   168    -1    1    -1    1     2
27   197     1    1    -1    1     1
29   170    -1    1     1    1     1
31   194     1    1     1    1     2
```

Note that there is no reason to divide an already performed experiment in different blocks, but if we have to redo the experiment and can, for example, just test 8 combinations per day, the above division in blocks is useful.

Now we make an analysis of variance of the data with the block factor:

```
sB.fit <- aov(score~sugar+soda+water+temp+BLOCK,data=softBL)
> summary(sB.fit)
            Df Sum Sq Mean Sq F value    Pr(>F)
sugar        1 976.56  976.56 26.1375 0.0004557 ***
soda         1 451.56  451.56 12.0860 0.0059561 **
water        1   5.06    5.06  0.1355 0.7204717
temp         1 315.06  315.06  8.4326 0.0157284 *
BLOCK        1   3.06    3.06  0.0820 0.7804926
Residuals   10 373.63   37.36
```

We conclude that sugar and soda are relevant at a 1% level (and temperature is relevant at a 5% level).
If, additionally, we want to compute the 2-way effects we just have to type

```
aov(score~(sugar+soda+water+temp+BLOCK)^2,data=softBL)
```

If we want to compute all the n-way effects we just have to type

```
sB.2k <- aov(score~sugar*soda*water*temp*BLOCK,data=softBL)
```

In this case 3 and 4-way effects are confounded, it follows that we obtain the same result as in the previous two function calls.
Furthermore, if we want to do an analysis of variance we can not look at all the 1 and 2-way effects because otherwise we lose all the degrees of freedom for the residuals!

Remark
With `sB.2k$coef` we can see that the 3 and 4-way effects are confounded (Effects are market with `NA`).

3. We have the following:

- $8 = 2^3 = 2^{k-l}$ runs,
- 5 two-level factors, thus: $k = 5$,
- consequently we need $l = 5 - 3 = 2$ "confounding relations".

Solution:
<u>STEP 1:</u>
Write down the complete 2^3 table.

A	B	C
$-$	$-$	$-$
$+$	$-$	$-$
$-$	$+$	$-$
$+$	$+$	$-$
$-$	$-$	$+$
$+$	$-$	$+$
$-$	$+$	$+$
$+$	$+$	$+$

<u>STEP 2:</u>
Define the "confounding relations". (If not specified otherwise confounding relations can be chosen quite freely).
We try to maximize the resolution without prior information on the dataset and choose: $D = -A \cdot B$ and $E = -A \cdot C$ (The $-$ is not necessary, but doing so our first run will be (1)).
We obtain:

A	B	C	D	E
$-$	$-$	$-$	$-$	$-$
$+$	$-$	$-$	$+$	$+$
$-$	$+$	$-$	$+$	$-$
$+$	$+$	$-$	$-$	$+$
$-$	$-$	$+$	$-$	$+$
$+$	$-$	$+$	$+$	$-$
$-$	$+$	$+$	$+$	$+$
$+$	$+$	$+$	$-$	$-$

<u>STEP 3:</u>
Now read every row of the matrix marking the factors with $+$ for high level:

A	B	C	D	E	Treatm.
$-$	$-$	$-$	$-$	$-$	(1)
$+$	$-$	$-$	$+$	$+$	ade
$-$	$+$	$-$	$+$	-	bd
$+$	$+$	$-$	$-$	$+$	abe
$-$	$-$	$+$	$-$	$+$	ce
$+$	$-$	$+$	$+$	$-$	acd
$-$	$+$	$+$	$+$	$+$	bcde
$+$	$+$	$+$	$-$	$-$	abc

Which effects are confounded with each other?
Because $l = 2$ every effect is confounded with $2^l = 2^2 = 4$ effects.
We know: $D = -A \cdot B$, so the effects of D and AB are not distinguishable (we write $D \cong AB$).

From $D \cong AB$ and $E \cong AC$ we get:

- $I \cong ABD \cong ACE \cong BCDE$

By multiplication we find:

- $A \cong BD \cong CE \cong ABCDE$
- $B \cong AD \cong ABCE \cong CDE$
- $C \cong AE \cong BDE \cong ABCD$
- $D \cong AB \cong BCE \cong ACDE$
- $E \cong AC \cong BCD \cong ABDE$
- $BC \cong ED \cong ABE \cong ACD$
- $BE \cong CD \cong ABC \cong AED$

Remark
The resolution of the experiment can be calculated as follows: take two effects which are confounded and count the number of letters you have. The minimal result you can obtain is the resolution. In our case:

- B & $AD \rightarrow 3$ (letters)
- D & $AB \rightarrow 3$ (letters)
- BCD & $ACDE \rightarrow 7$ (letters)
- ...

The resolution is 3 (not very high).

Can we improve the resolution by changing the relationships $D \cong AB$ and $E \cong AC$?[1]
Let us think about it:
We can make 8 observations (7 degrees of freedom). If we want a resolution of 4 there has to be no confounding between the main effects (with 1 letter) and the 2-way effects (with 2 letters). Naturally we can not have that $A \cong D$ or something similar because otherwise the resolution would be 2. Also we can not have that $AB \cong AC$ because then $B \cong C$. Consequently 3 different 2-way effects can not be confounded all together without having the undesirable consequence that two main effects are confounded.
Summarising: <u>If we want a resolution of 4:</u>
We have 5 main effects and at least 10/2=5 2-way effect which can NOT be confounded! This makes 10 in total. There are just 7 degrees of freedom (we can look at most at 7 different effects), therefore it is impossible to find a structure with resolution 4!

4. a) We have: $8 = 2^3 = 2^{n-k}$ observations. Furthermore we have $n = 4$ different factors and $k = 4 - 3 = 1$.

[1] The answer to this question is not required to solve the exercise and it is not trivial.

b) Let us call the effects of "Side-to-side", "Yarn type", "Pick density" and "Air pressure" A, B, C and D respectively. Then we have

A	B	C	D	Treatm.	Strength
$-$	$-$	$-$	$-$	(1)	24.50
$+$	$-$	$-$	$+$	ad	22.05
$-$	$+$	$-$	$+$	bd	24.52
$+$	$+$	$-$	$-$	ab	25.00
$-$	$-$	$+$	$+$	cd	25.68
$+$	$-$	$+$	$-$	ac	24.51
$-$	$+$	$+$	$-$	bc	24.68
$+$	$+$	$+$	$+$	abcd	24.23

with the alias $D = ABC$.
To find out which terms are aliased together it is enough to multiply the terms by $I = ABCD$. So

$$\begin{aligned} D &= ABC \\ A &= BCD \\ B &= ACD \\ C &= ABD \\ AB &= CD \\ AC &= BD \\ AD &= BC \end{aligned}$$

c) Estimates:

$$\begin{aligned}
\hat{A} &= \frac{1}{4}(-24.5 + 22.05 - 24.52 + 25 - 25.68 + 24.51 - 24.68 + 24.23) = -0.8975 \\
\hat{B} &= \frac{1}{4}(-24.5 - 22.05 + 24.52 + 25 - 25.68 - 24.51 + 24.68 + 24.23) = 0.4225 \\
\hat{C} &= \frac{1}{4}(-24.5 - 22.05 - 24.52 - 25 + 25.68 + 24.51 + 24.68 + 24.23) = 0.7575 \\
\hat{D} &= \frac{1}{4}(-24.5 + 22.05 + 24.52 - 25 + 25.68 - 24.51 - 24.68 + 24.23) = -0.5525 \\
\hat{AB} &= \hat{CD} = \frac{1}{4}(+24.5 - 22.05 - 24.52 + 25 + 25.68 - 24.51 - 24.68 + 24.23) = 0.9125 \\
\hat{AC} &= \hat{BD} = \frac{1}{4}(+24.5 - 22.05 + 24.52 - 25 - 25.68 + 24.51 - 24.68 + 24.23) = 0.0875 \\
\hat{AD} &= \hat{BC} = \frac{1}{4}(+24.5 + 22.05 - 24.52 - 25 - 25.68 - 24.51 + 24.68 + 24.23) = -1.0625
\end{aligned}$$

d) A factor is significant if its absolute value is larger than 0.35. In this case we have that A, B, C, D, AB and AD are significant.

e) With the assumption and the calculations of point c) we do not care about the two (or more)-way-effects.
The effect of A is $-0.8975 < 0$ which means that by changing the level of the factor A from low to high we lose strength. Consequently we choose:

Effect	Estimate	$\gtreqless 0$	Best
A	-0.8975	< 0	low
B	0.4225	>0	high
C	0.7575	> 0	high
D	-0.5525	< 0	low