Exercise 4 solutions - ST2304

Christoffer Høyvik Hilde

4/24/2018

Problem 1 - Interactions

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.4.3
```

```
library(ggplot2)
library(plyr)
library(cowplot)
```

```
## Warning: package 'cowplot' was built under R version 3.4.3
```

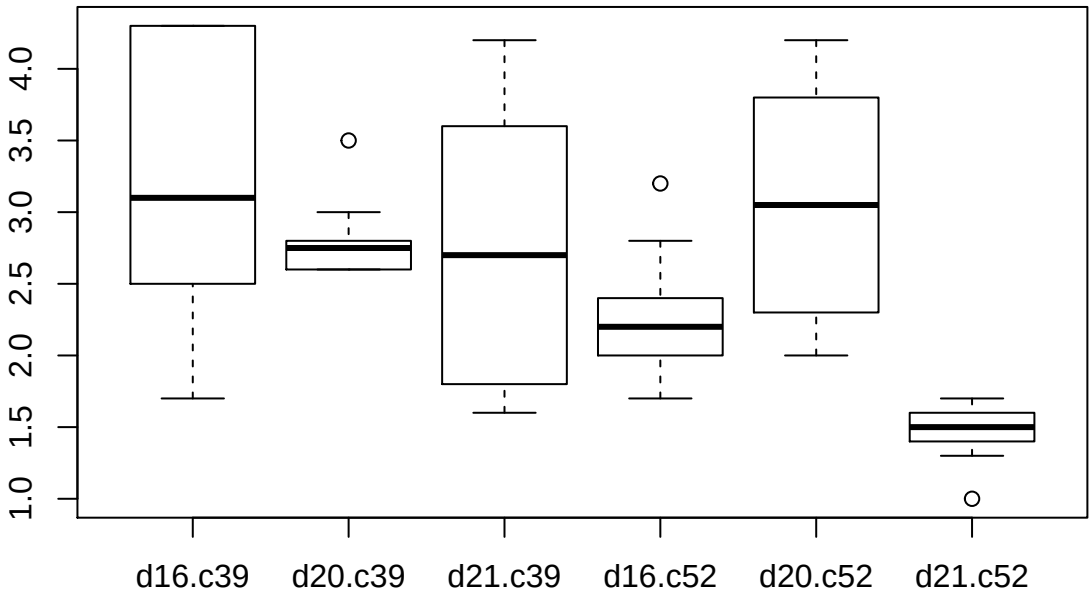
```
library(xtable)
```

1. Are there effects of cultivar ('Cult') and planting date ('Date') on yield ('HeadWt')? Does the effect of cultivar change depending on the planting date?

```
## Add the data to the working environment (data found in the library MASS)
data("cabbages")
```

Visualize the data with an interaction, in this case we want to know whether the effect of cultivar on Head weight depend on the date

```
boxplot(HeadWt~Date*Cult, data=cabbages)
```



It looks like Cultivar 52 on day 21 has lower head weight than the rest, we fit a linear model with Head weight as response and Date, cultivar and their interaction as explanatory variables to investigate whether this is the case:

```
mod1<- lm(HeadWt~Date*Cult, data=cabbages)
```

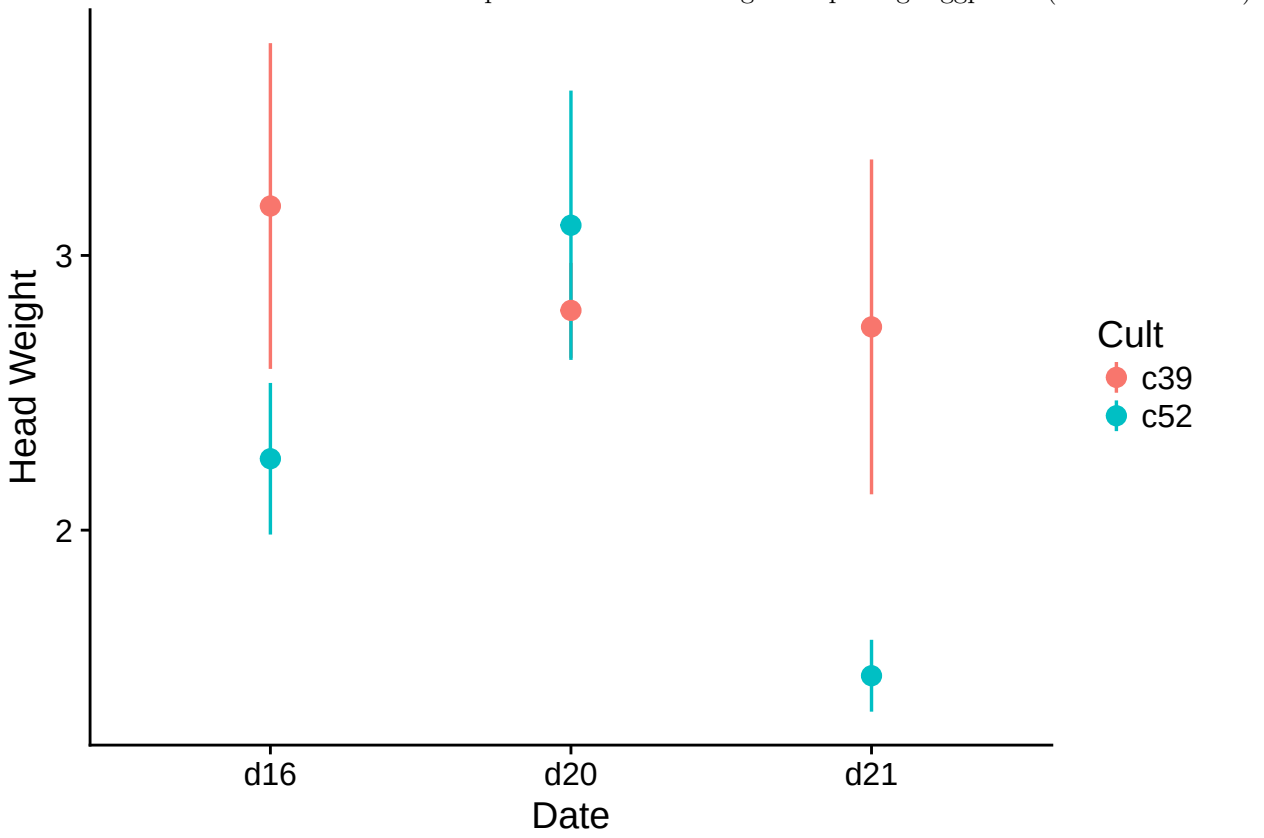
An interaction is specified by putting * between two explanatory variables, this is equivalent to writing Date + Cult + Date:Cult

```
anova(mod1)
```

```
## Analysis of Variance Table
##
## Response: HeadWt
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Date       2  7.7063  3.8532  8.1744 0.0007920 ***
## Cult       1  5.8907  5.8907 12.4969 0.0008451 ***
## Date:Cult  2  6.8863  3.4432  7.3046 0.0015571 **
## Residuals 54 25.4540  0.4714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova says that both Date, Cult and the interaction is statistically significant.This means that the effect of Cultivar on head weight depends on the date of planting

Below is a plot of the estimated means with confidence interval for each combination of date and cultivar. We see that cultivar 52 has a lower head weight on day 16 and 21 than cultivar 39. For those interested the plot was made using the package ggplot2 (not curriculum).



2. Are there effects of cultivar and planting date on vitamin C concentration('VitC')? Does the effect of cultivar change depending on the planting date? For each of these, summarise how well the model explains the data and if there are any problems with the model fit.

To check whether there is an effect of cultivar and date on VitC we fit a model with VitC as response and date + cultivar as explanatory variables:

```
vitmod<-lm(VitC~Date+Cult, data=cabbages)
anova(vitmod)
```

```
## Analysis of Variance Table
##
## Response: VitC
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Date       2  909.3  454.65  9.6609 0.0002486 ***
## Cult       1 2496.2 2496.15 53.0411 1.179e-09 ***
## Residuals 56 2635.4   47.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Both date and cultivar seem to have an effect, but is the effect of cultivar dependent on date? We fit another model with the interaction between these two:

```
vitmod2<-lm(VitC~Date*Cult, data=cabbages)
anova(vitmod2)
```

```
## Analysis of Variance Table
##
## Response: VitC
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Date       2  909.3  454.65  9.8555 0.0002245 ***
## Cult       1 2496.2 2496.15 54.1095 1.089e-09 ***
## Date:Cult  2  144.3   72.15  1.5640 0.2186275
## Residuals 54 2491.1   46.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

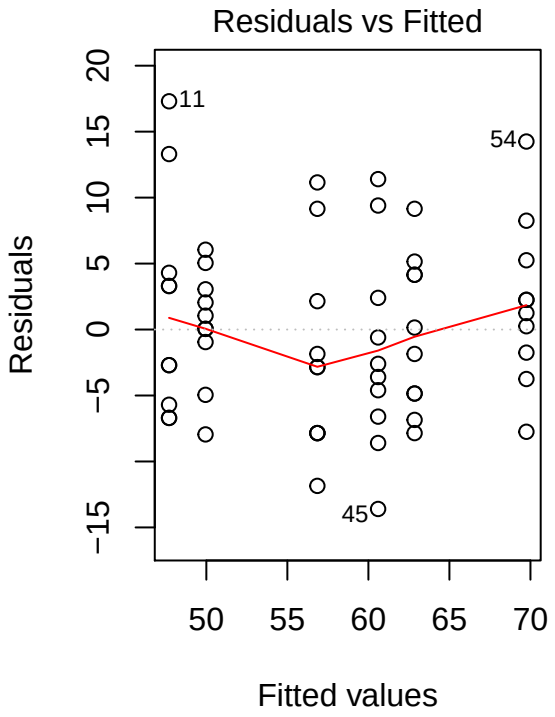
There doesnt seem to be a significant interaction effect between date and cult and looking at the R^2 from the two models the difference is small

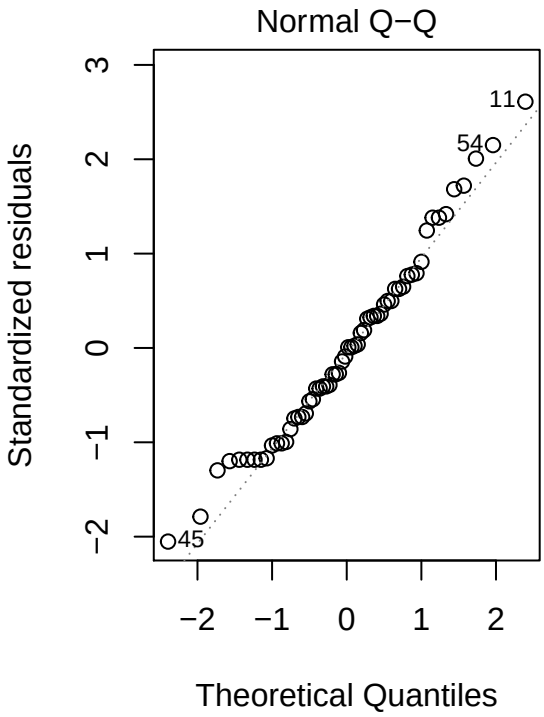
```
tt<-data.frame(Model = c("With Interaction", "Without Interaction"), R2 = c(summary(vitmod)$r.squared,su
table<-xtable(tt)
print(table, comment=F)
```

	Model	R2
1	With Interaction	0.56
2	Without Interaction	0.59

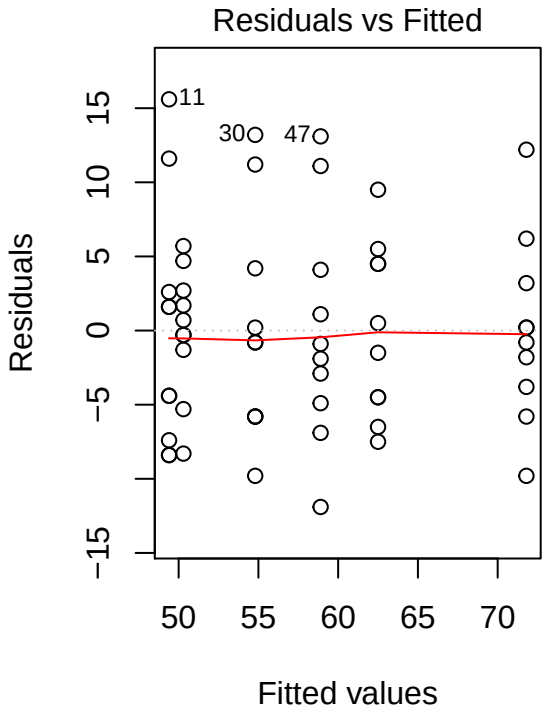
To look at the model fit we plot the diagnostic plots for each model:

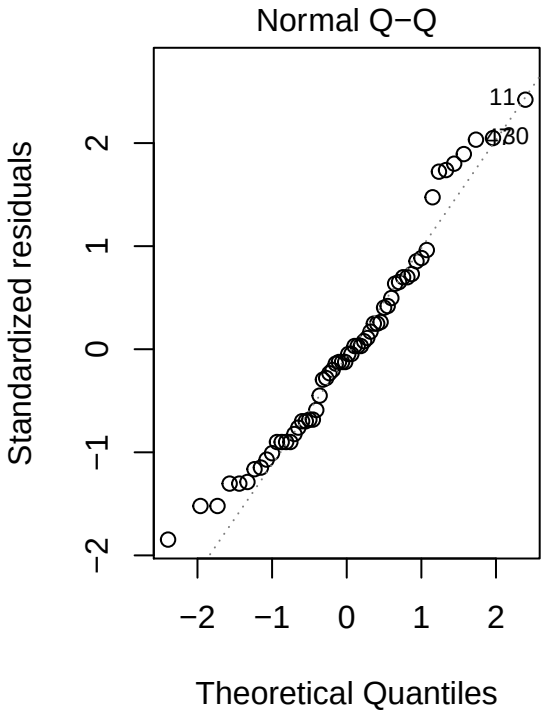
```
par(mfrow=c(1,2))
plot(vitmod, which=c(1:2))
```






```
plot(vitmod2, which=c(1:2))
```





They both look ok, actually the residuals vs fitted plot looks a bit better for the interaction model, but its negligible. The Q-Q plot is better for the model without interaction. In conlcusion: there is no reason to include the interaction in the model.

Problem 2 - CO_2 uptake in plants

1. Fit a model with the source population("Type") and treatment ("Treat") as effects. Do they have an effect, and do their effects interact? If so, how? How well does this model explain the data?

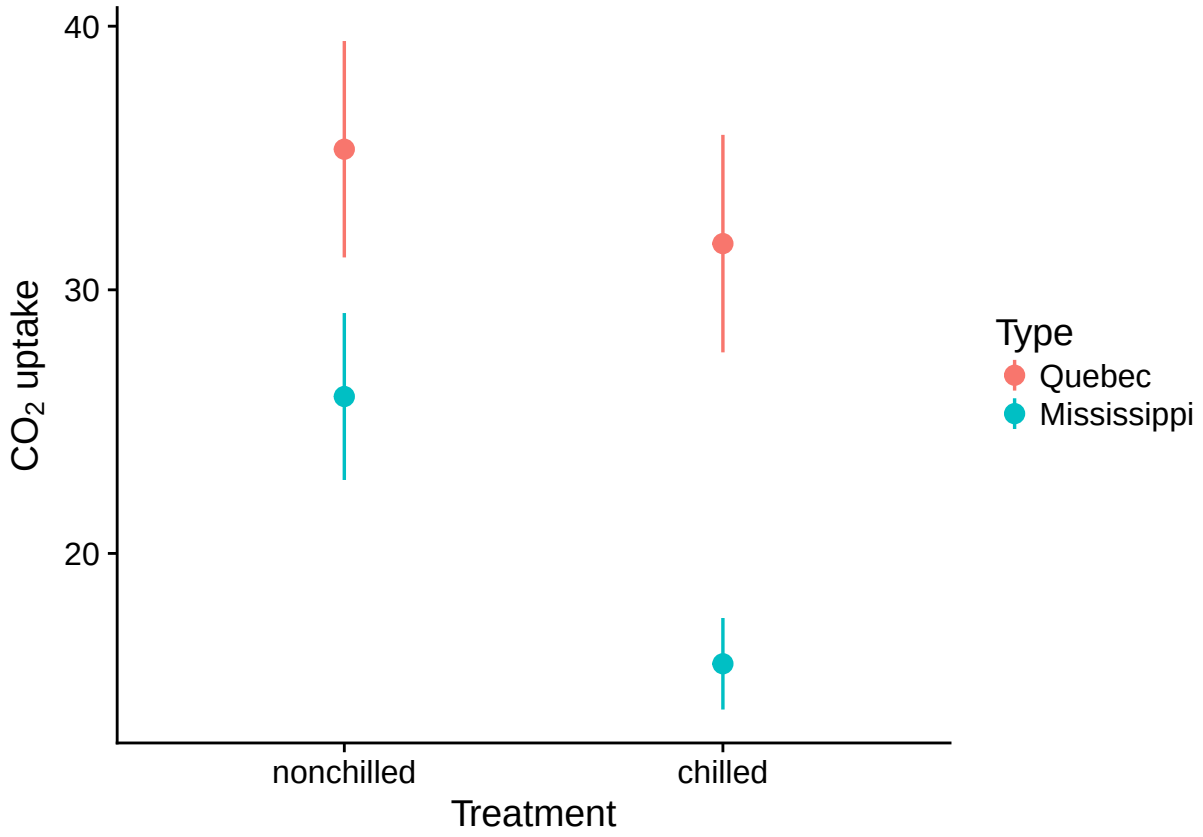
```
data(CO2)
head(CO2)
```

```
##   Plant   Type  Treatment conc uptake
## 1   Qn1 Quebec nonchilled   95   16.0
## 2   Qn1 Quebec nonchilled  175   30.4
## 3   Qn1 Quebec nonchilled  250   34.8
## 4   Qn1 Quebec nonchilled  350   37.2
## 5   Qn1 Quebec nonchilled  500   35.3
## 6   Qn1 Quebec nonchilled  675   39.2
```

```
upmod<-lm(uptake~Treatment*Type, data=CO2)
summary(upmod)
```

```
##
## Call:
## lm(formula = uptake ~ Treatment * Type, data = CO2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.452  -3.624   2.167   5.773  10.648
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       35.333      1.747  20.225  < 2e-16 ***
## Treatmentchilled                  -3.581      2.471  -1.449 0.151141
## TypeMississippi                   -9.381      2.471  -3.797 0.000284 ***
## Treatmentchilled:TypeMississippi  -6.557      3.494  -1.877 0.064213 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.006 on 80 degrees of freedom
## Multiple R-squared:  0.4718, Adjusted R-squared:  0.452
## F-statistic: 23.82 on 3 and 80 DF,  p-value: 4.106e-11
```

We make a model with Treatment, Type and their interaction to explain the CO_2 uptake. From the summary of the model we can see that plants from Mississippi has a lower CO_2 uptake than plants from Quebec and that this difference is even stronger when they are chilled (the interaction effect). A plot makes it easier to



see:

The model explains 47% of the variance, which is pretty good, but since the plants were grown in different CO_2 concentrations it would make sense to include this as an explanatory variable to see if it has an effect on the CO_2 uptake.

2

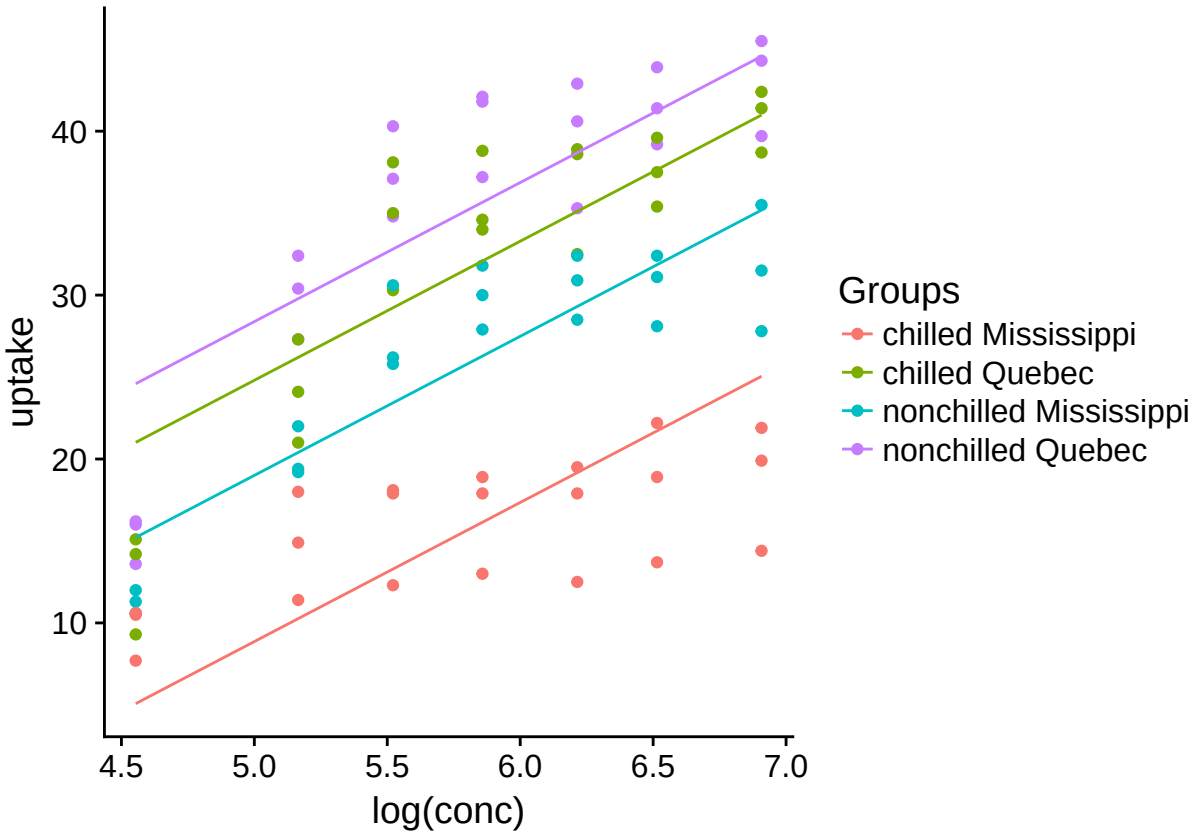
We add the log of CO_2 as a main effect in addition to Type and Treatment:

```
upmodlog<-lm(uptake~Treatment*Type+log(conc), data=CO2)
summary(upmodlog)
```

```
##
## Call:
## lm(formula = uptake ~ Treatment * Type + log(conc), data = CO2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.7166  -2.8960   0.5837   2.7621   8.8745
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -14.0369     4.0768  -3.443 0.000923 ***
## Treatmentchilled                 -3.5810     1.4403  -2.486 0.015017 *
## TypeMississippi                  -9.3810     1.4403  -6.513 6.26e-09 ***
## log(conc)                         8.4839     0.6783  12.507  < 2e-16 ***
## Treatmentchilled:TypeMississippi -6.5571     2.0368  -3.219 0.001866 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.667 on 79 degrees of freedom
## Multiple R-squared:  0.8227, Adjusted R-squared:  0.8138
## F-statistic: 91.67 on 4 and 79 DF,  p-value: < 2.2e-16
```

The R^2 increased from 47% to 82%, a huge increase. We see that the CO_2 uptake increase with the CO_2 concentration in the room the plants are in, makes sense. Apart from that, the difference between the treatments and types are the same as in the previous model, but now the standard errors are much smaller and the P-values are all significant. A crude plot to show the effects:



As we see from the plot, there are 4 lines, one for each combination between type and treatment. The 4 lines have the same slope (i.e. the effect of CO_2 concentration is the same) but different intercepts.

3. Optional: If you are feeling, brave, look to see if there is an effect of the interactions between log(concentration) and the other variables. Can you interpret what is going on?

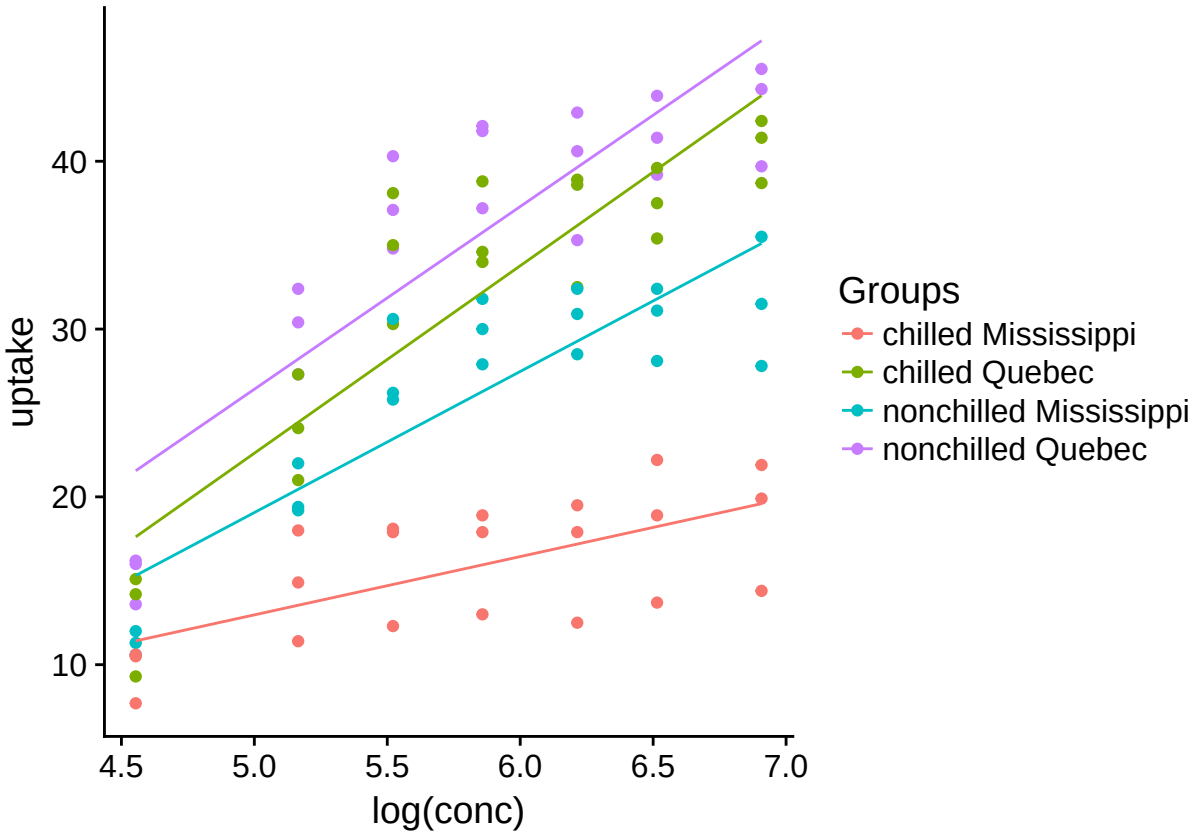
We add log CO_2 in an interaction between Type and Treatment:

```
crazymod<-lm(uptake~Treatment*Type*log(conc), data=CO2)
summary(crazymod)
```

```
##
## Call:
## lm(formula = uptake ~ Treatment * Type * log(conc), data = CO2)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -8.3185 -3.3408  0.3514  2.7150  9.6743
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                              -28.0189     6.9723  -4.019
## Treatmentchilled                          -5.2260     9.8603  -0.530
## TypeMississippi                            5.0530     9.8603   0.512
## log(conc)                                 10.8866     1.1883   9.162
## Treatmentchilled:TypeMississippi          23.7930    13.9446   1.706
## Treatmentchilled:log(conc)                 0.2827     1.6805   0.168
## TypeMississippi:log(conc)                 -2.4804     1.6805  -1.476
## Treatmentchilled:TypeMississippi:log(conc)  -5.2154     2.3766  -2.195
##                                          Pr(>|t|)
## (Intercept)                              0.000137 ***
## Treatmentchilled                         0.597656
## TypeMississippi                          0.609817
## log(conc)                                6.55e-14 ***
## Treatmentchilled:TypeMississippi         0.092046 .
## Treatmentchilled:log(conc)               0.866860
## TypeMississippi:log(conc)                0.144084
## Treatmentchilled:TypeMississippi:log(conc) 0.031254 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.088 on 76 degrees of freedom
## Multiple R-squared:  0.8692, Adjusted R-squared:  0.8571
## F-statistic: 72.13 on 7 and 76 DF,  p-value: < 2.2e-16
```

In the previous model, we had the same intercept for all the groups, but by introducing an interaction between CO_2 concentration and the groups we make it so that they have different slopes as well:



As we see from the plot, it seems that not only do the plants from Mississippi have a lower CO_2 uptake in general, but when they are chilled it seems that the CO_2 uptake doesnt increase with the CO_2 concentration in the room as much as for the three other groups (Chilled/nonchilled Quebec and nonchilled Mississippi).