Statistics Examen_NeuroBIM

Maxime Houtekamer

16 november 2015

Loading the data file into R

First of all, the data was loaded into R from the textfile. A summary of the data was obtained as an indication whether the data was correctly loaded into R.

```
data<-read.table("lesionsBIM.txt",header=TRUE)
data1<-read.table("lesionsBIM.txt",header=TRUE)
```

Let's add a collumn for the difference between time needed

```
diff<-vector()
for(i in 1:(length(data[,1]))){
  if(as.character(data[i,1]) == "D5 "){
    diff<-c(diff,(data[i,22] - data[i,26]))
      } else if(as.character(data[i,1]) == "D3 "){
    diff<-c(diff,(data[i,22] - data[i,24]))
    }
}
data<-cbind(data,diff)
```

Creating seperate files for the 4 conditions

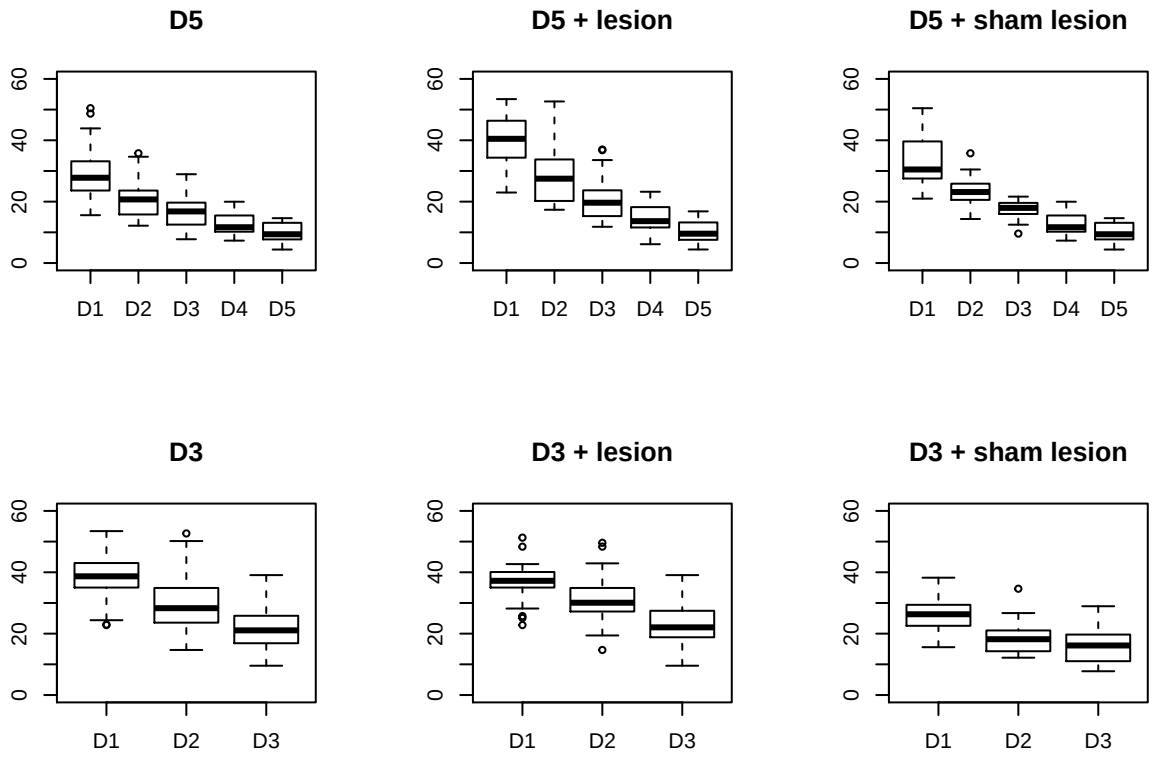
The mice were either trained in 3 sessions (D3) or in 5 sessions (D5). Within each of these two groups, the animals were either lesioned in the dorsal hippocampus (H) or they were given a sham lesion (SH). These groups were originally stored in the datafile, but will now be sorted in order to easily be able to display them seperately.

```
d3<-data[1:48,]
d5<-data[49:90,]
d3sh<-data[49:72,]
d5sh<-data[73:90,]
d3h<-data[25:48,]
d5h<-data[1:24,]
```

Learning time

The mice were given a task, and the time they spent in the dark is a measure of how well they learnt it.

```
par(mfrow=c(2,3))
boxplot(d5[22:26],ylim=c(0, 60))
title(main="D5")
boxplot(d5h[22:26],ylim=c(0, 60))
title(main="D5 + lesion")
boxplot(d5sh[22:26],ylim=c(0, 60))
title(main="D5 + sham lesion")
boxplot(d3[22:24],ylim=c(0, 60))
title(main="D3")
boxplot(d3h[22:24],ylim=c(0, 60))
title(main="D3 + lesion")
boxplot(d3sh[22:24],ylim=c(0, 60))
title(main="D3 + sham lesion")
```




```
mean(d5h[,27])
```

```
## [1] 29.5587
```

```
mean(d5sh[,27])
```

```
## [1] 22.88382
```

```r
mean(d3h[,27])
```

```
## [1] 13.70844
```

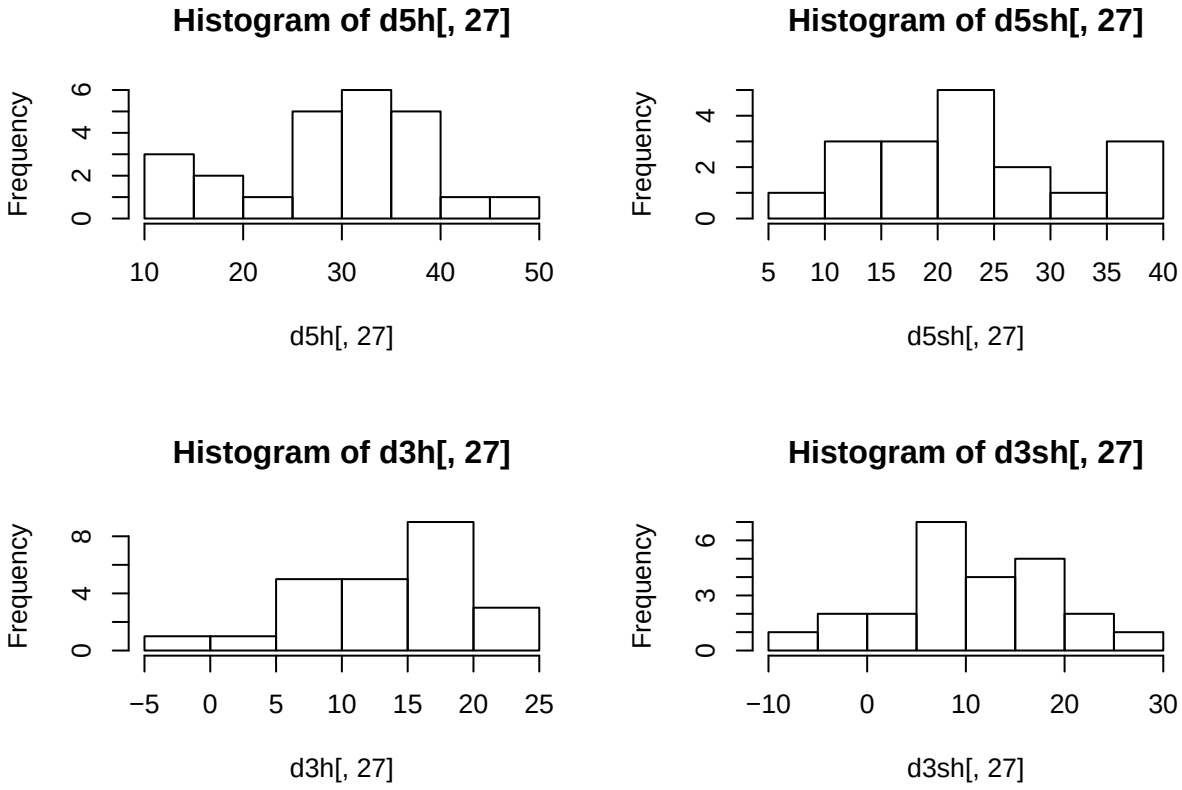
```r
mean(d3sh[,27])
```

```
## [1] 10.73177
```

Is the data normally distributed?

First, Let's look at histograms of each group.

```r
par(mfrow=c(2,2))
hist(d5h[,27])
hist(d5sh[,27])
hist(d3h[,27])
hist(d3sh[,27])
```



NORMAL DISTIBUTION?

We will carry out the shapiro-Wilk test. If p>a (bigger than 0.05 generally), the data is normal.

```
shapiro.test(d5h[,27])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  d5h[, 27]
## W = 0.96506, p-value = 0.5482
```

```
shapiro.test(d5sh[,27])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  d5sh[, 27]
## W = 0.94456, p-value = 0.3461
```

```
shapiro.test(d3h[,27])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  d3h[, 27]
## W = 0.95398, p-value = 0.3297
```

```
shapiro.test(d3sh[,27])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  d3sh[, 27]
## W = 0.98565, p-value = 0.9735
```

All the values are higher than p=0.05, so the data is normally distributed.

Making a new dataframe for ANOVA

We will make a list of the factors (d5h, d5sh, d3h, d3sh), and a list with the "learned" decrease in time needed to explore the matrix.

```
factorlist<-c((rep("d5h",24)),(rep("d5sh",18)),(rep("d3h",24)),(rep("d3sh",24)))
variablelist<-c(d5h[,27],d5sh[,27],d3h[,27],d3sh[,27])
d1<-data.frame(factorlist,variablelist)
colnames(d1)<-c("exp","values")
f1<-d1$values~d1$exp
```

Homogeneity of Variance

```
bartlett.test(f1)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  d1$values by d1$exp
## Bartlett's K-squared = 3.6572, df = 3, p-value = 0.3009
```

ANOVA

Maybe we should instead to a repeated measures anova where we follow the animal over the different learning trials.

```
aov1<-aov(f1)
summary(aov1)
```

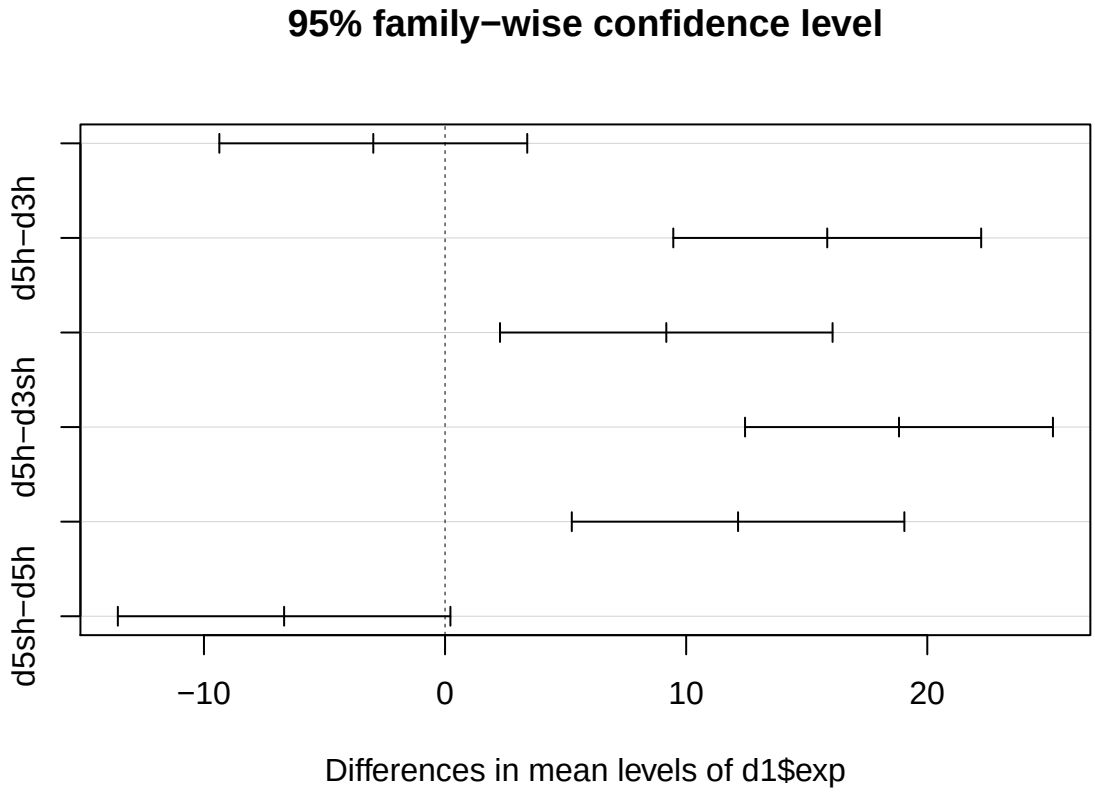
```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## d1$exp       3   5260  1753.3    24.6 1.38e-11 ***
## Residuals   86   6129    71.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova is highly significant at p<0.05. Let's do a post-hoc Tukey test to find where the differences are #Tukey posthoc

```
t1<-TukeyHSD(aov1)
t1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = f1)
##
## $`d1$exp`
##                 diff        lwr        upr     p adj
## d3sh-d3h  -2.976667  -9.361469  3.4081355 0.6150204
## d5h-d3h   15.850260   9.465458 22.2350626 0.0000000
## d5sh-d3h   9.175382   2.279007 16.0717565 0.0042380
## d5h-d3sh  18.826927  12.442125 25.2117293 0.0000000
## d5sh-d3sh 12.152049   5.255674 19.0484232 0.0000790
## d5sh-d5h  -6.674878 -13.571253  0.2214961 0.0614873
```

```
plot(t1)
```



#Repeated measures ANOVA

```
id<-vector()
for(i in 1:42){
  id<-c(id,(rep(i,5)))
  }
for(i in 43:90){
  id<-c(id,(rep(i,3)))
  }


group<-c((rep("d5h",(24*5))),(rep("d5sh",(18*5))),(rep("d3h",(24*3))),(rep("d3sh",(24*3))))



tasktime<-vector()
for(i in 1:24){
     tasktime<-c(tasktime,data1[i,22])
     tasktime<-c(tasktime,data1[i,23])
     tasktime<-c(tasktime,data1[i,24])
     tasktime<-c(tasktime,data1[i,25])
     tasktime<-c(tasktime,data1[i,26])
  }
for(i in 73:90){
    tasktime<-c(tasktime,data1[i,22])
    tasktime<-c(tasktime,data1[i,23])
    tasktime<-c(tasktime,data1[i,24])
```

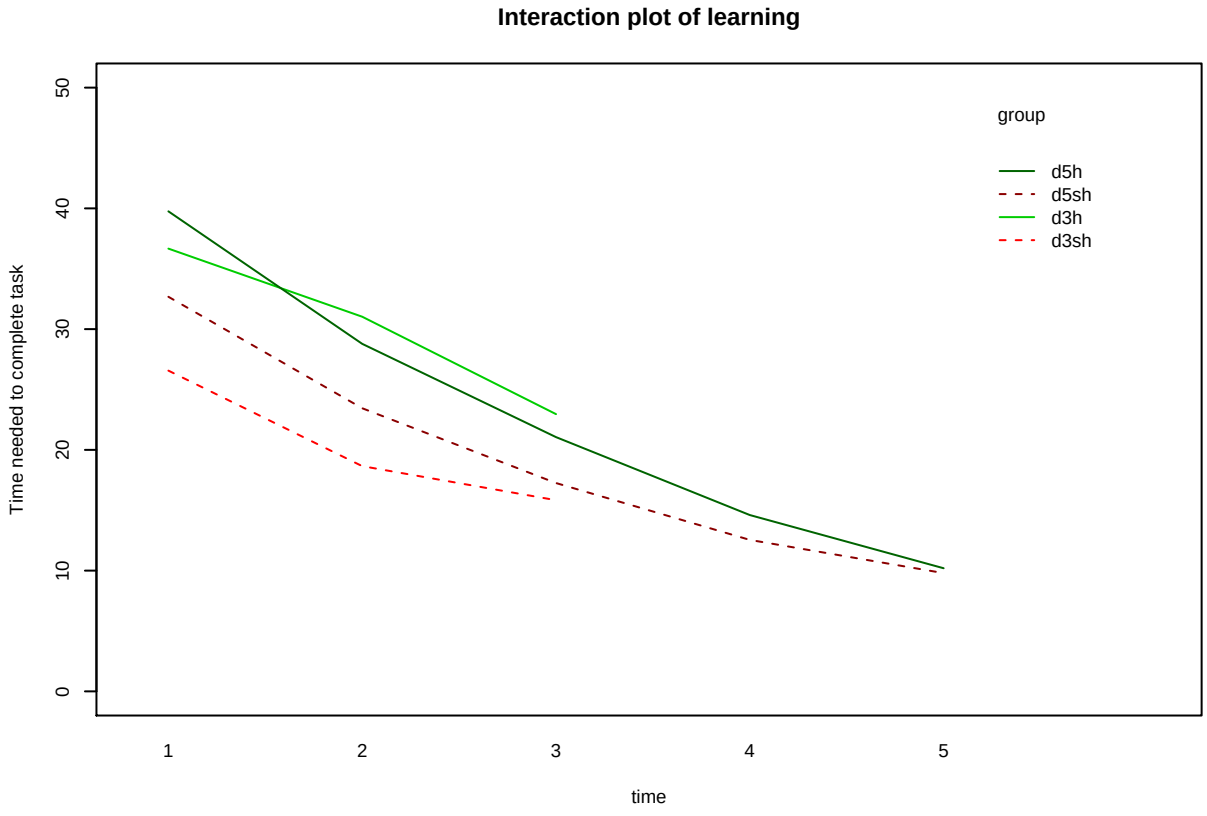
```
    tasktime<-c(tasktime,data1[i,25])
    tasktime<-c(tasktime,data1[i,26])
  }
for(i in 25:72){
      tasktime<-c(tasktime,data1[i,22])
      tasktime<-c(tasktime,data1[i,23])
      tasktime<-c(tasktime,data1[i,24])
  }

time<-c(rep(1:5,(42)),rep(1:3,(48)))

d2<-data.frame(id,group,tasktime,time)

##convert variables to factor
d2<-within(d2, {
  group<-factor(group)
  time<-factor(time)
  id<-factor(id)
})
par(cex = .6)

with(d2,interaction.plot(time, group, tasktime,
                         ylim= c(0,50), lty=c(1,20,1,20),col = c(3,2,"darkgreen","darkred"),
                         ylab= "Time needed to complete task", xlab= "time",trace.label="group",main="I
```



```r
d2.aov<-aov(tasktime ~ group * time + Error(id), data=d2)
summary(d2.aov)
```

```
##
## Error: id
##           Df Sum Sq Mean Sq F value   Pr(>F)
## group      3   5594  1864.7   27.11 1.95e-12 ***
## Residuals 86   5914    68.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##             Df Sum Sq Mean Sq F value  Pr(>F)
## time         4  22320    5580 169.998 < 2e-16 ***
## group:time   8    751      94   2.861 0.00462 **
## Residuals  252   8272      33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The between group tests indicates that the variable group is significant. consequently, int he graph we see that the liens for the two groups are rather far apart. The within subject test indicates that there is a significant time effect, inother words, the groups do change other time, both groups are taking less time to complete the task over time. Morover, the interaction of time and group is significant which means that the groups are changing over time but are changing in different ways, which means that in the graph, the liens will not be parallel.