

## #Two-way ANOVA

*#In this tutorial we discuss fitting two-way analysis of variance (ANOVA) in R. As we fit these models using regression methods much of the syntax from the multiple regression tutorial will still be valid. For ANOVA, all the explanatory variables are categorical.*

### #A. Two-way Analysis of Variance

*#Two-way ANOVA is used to compare the means of populations that are classified in two different ways, or the mean responses in an experiment with two factors. One can fit a two-way ANOVA using the aov() function we had introduced in the one-way ANOVA section, However, since the aov() utilizes the Type I sum of squares, we wish to give ourselves more flexibility by being able to use also the other sum of squares types, such as the Type III sum of square implemented in the lm() function. Now we will fit two-way ANOVA models in R using the lm() function. For example, the command:*

*lm(Response ~ FactorA + FactorB) fits a two-way ANOVA model without interactions. In contrast, the command*

*lm(Response ~ FactorA + FactorB + FactorA\*FactorB) includes an interaction term. Here both FactorA and FactorB are categorical variables, while Response is quantitative.*

*#Ex. A study was performed to test the efficiency of a new drug developed to increase high-density lipoprotein (HDL) cholesterol levels in patients. 18 volunteers were split into 3 groups (Placebo, 5 mg, 10 mg) and the difference in HDL levels before and after the treatment was measured. The 18 volunteers were also categorized into two age groups (less than 40 or not). We are interested in determining the answers to the following questions:*

*#Does the amount of drug have an effect on the HDL level?*

*#Does the age of the patient have an effect on the HDL level?*

*#Is there an interaction between age and amount of drug?*

*#We begin by reading in the data and making some exploratory plots of the data.*

*#Data:*

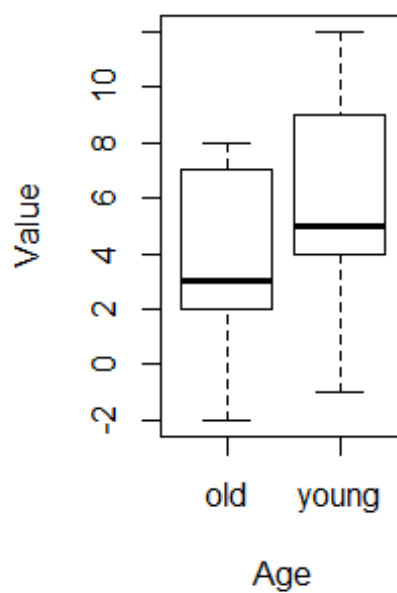
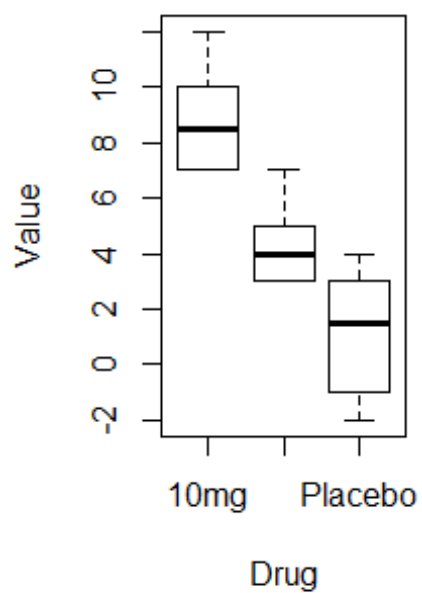
```
Drug = c(rep("Placebo",6), rep("5mg",6),rep("10mg",6))
Age = c(rep(c(rep("young",3), rep("old",3)),3))
Value = c(4, 3, -1, 2, 1, -2, 4, 5, 7, 3, 3, 4, 9, 12, 10, 7, 8, 7)
data<-data.frame(Drug,Age,Value);data
```

```
##      Drug  Age Value
## 1  Placebo young    4
## 2  Placebo young    3
## 3  Placebo young   -1
## 4  Placebo  old    2
## 5  Placebo  old    1
## 6  Placebo  old   -2
## 7     5mg young    4
## 8     5mg young    5
## 9     5mg young    7
## 10    5mg  old    3
## 11    5mg  old    3
## 12    5mg  old    4
## 13   10mg young    9
## 14   10mg young   12
## 15   10mg young   10
## 16   10mg  old    7
## 17   10mg  old    8
## 18   10mg  old    7
```

*#To make side-by-side boxplots:*

```
par(mfrow=c(1,2))
```

```
plot(Value ~ Drug + Age, data=data)
```



*#Judging by the boxplots there appears to be a difference in HDL level for the different drug levels.*

*#However, the difference is less pronounced between the two age groups.*

*#An interaction plot displays the levels of one factor on the x-axis and the mean response for each treatment on the y-axis.*

*#In addition, it shows a separate line connecting the means corresponding to each level of the second factor. When no interaction is present the lines should be roughly parallel.*

*#These types of plots can be used to determine whether an interaction term should be included in our ANOVA model.*

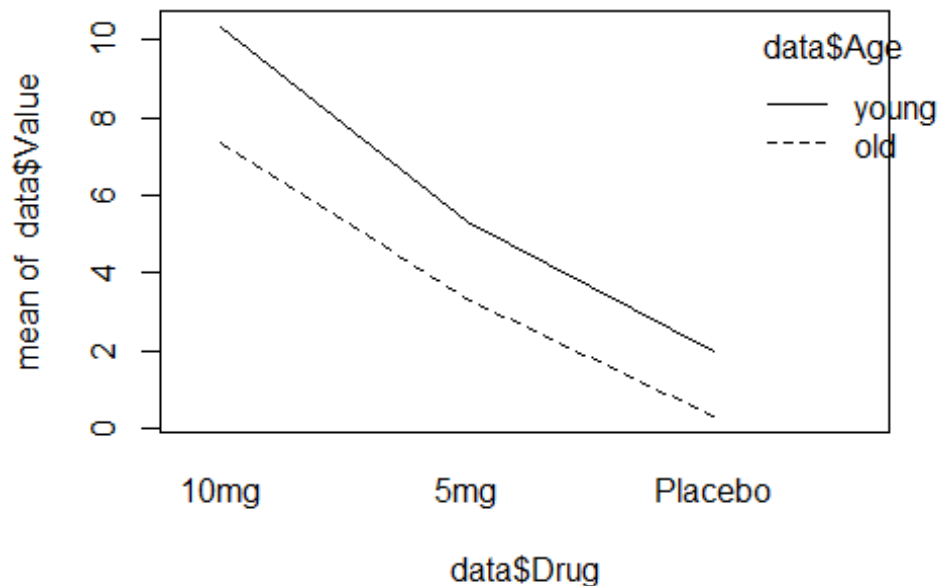
*#Interaction plots can be made in R using the command:*

```
interaction.plot(factorA, factorB, Response)
```

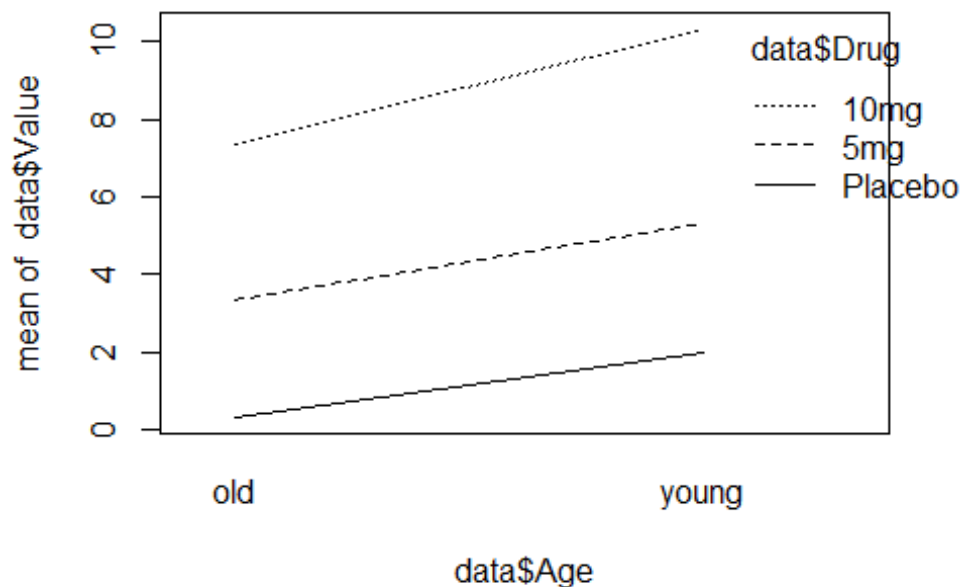
*#If we switch the order of factor A and B we alter which variable is plotted on the x-axis.*

*#Ex. Create an interaction plot for the HDL data.*

```
interaction.plot(data$Drug, data$Age, data$Value)
```



```
interaction.plot(data$Age, data$Drug, data$Value)
```



*#The interaction plots look roughly parallel,  
#but to confirm we fit a two-way ANOVA with an interaction term:*

```
results = lm(Value ~ Drug + Age + Drug*Age, data=data)
anova(results)

## Analysis of Variance Table
##
## Response: Value
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Drug       2 178.111   89.056   32.06 1.535e-05 ***
## Age        1  22.222    22.222    8.00  0.01522  *
## Drug:Age    2   1.444     0.722    0.26  0.77529
## Residuals 12  33.333     2.778
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*#Studying the output of the ANOVA table we see that  
#there is no evidence of a significant interaction effect ( $F=0.26$ ,  $p=0.7753$ ).*

*#We therefore cannot conclude that there is an interaction between a  
ge and the amount of drug taken.*

*#The test for the main effect of treatment ( $F=32.06$ ,  $p<0.0001$ ) shows  
a significant drug effect on the HDL level.*

*#Finally, the test for the main effect of age ( $F=8.00$ ,  $p=0.01522$ ) shows a not strong age effect on the HDL level.*

*#After fitting an ANOVA model it is important to always check the relevant model assumptions.*

*#This includes making QQ-plots and residual plots.*

**Note: You will find the same results using the following slightly different notations:**

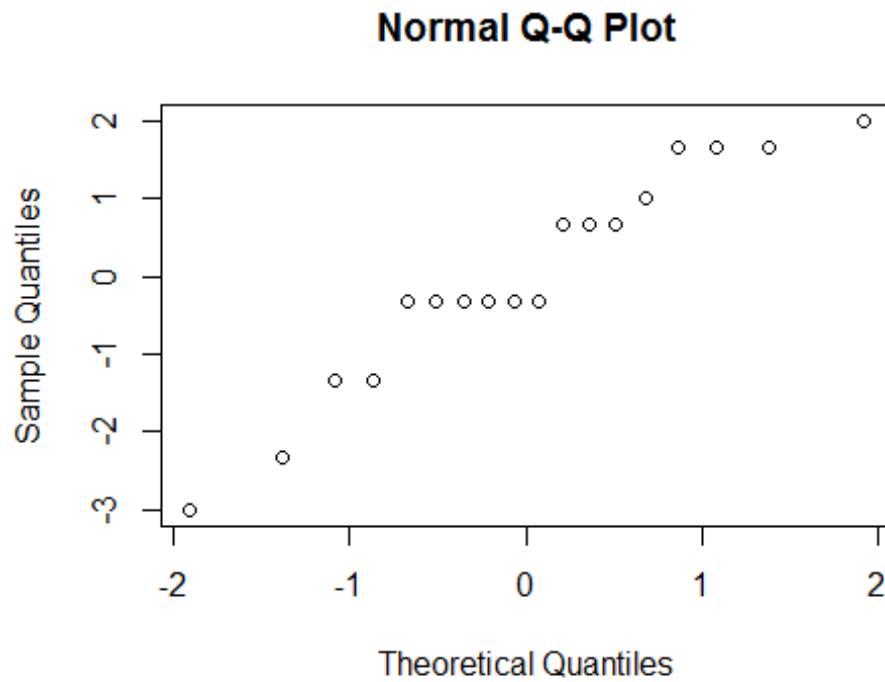
```
results1 = lm(Value ~ Drug + Age + Drug:Age, data=data)
anova(results1)
```

```
results2 = lm(Value ~ Drug*Age, data=data)
anova(results2)
```

Since we have balanced design, the factors are orthogonal, and types I, II and III all give the same results. Therefore, the `aov()` function will give the same results as the `lm()` function in this case.

```
Results3 = aov(Value ~ Drug + Age + Drug*Age, data=data)
results4 = aov(Value ~ Drug + Age + Drug:Age, data=data)
results5 = aov(Value ~ Drug*Age, data=data)
```

```
qqnorm(results$res)
```



```
plot(results$fitted,results$res,xlab="Fitted",ylab="Residuals")
```

