7.3 Two-Way ANOVA

In many cases, the response can be explained by more than one factors. For example, if we want to investigate sales of a product, both selling price (factor A) and the type of promotional campaign (factor B) have an effect on the number of products sold. Or the effect of a blood pressure lowering drug might depend on both the different drug types (factor A) of the drug and the gender of the participants (factor B). Therefore, there is a need to extend the statistical methods we discussed for one factor Analysis of Variance to two *or more*.

When the response is continuous and we have two categorical factors as predictors, then we perform a **two-way analysis of variance (ANOVA)**, and consequently when we have more than two factors we have a **multi-way ANOVA**.

One would argue that a multi-factor analysis is not necessary, because we can perform several one-factor-at-a-time analyses. But, many one-factor-at-a-time analyses do not explore the entire space of treatment combinations, do not allow us to estimate interactions and when it comes to an experiment full randomization is not possible.

In this section, we are going to present the two-way ANOVA theory and methods with the following experiment:

The Rats Experiment

The `rats` data set in the `faraway` library consistes of the *survival times* (the response) of rats that are randomly allocated to 3 poisons (I, II, III) and four treatments (A, B, C, D). This is an experiment with **2 factors**, *poison* and *treatment*, each having 3 and 4 levels respectively. The structure of this experiment is what we call *factorial* since each of the 3 types of poison appears together with each of the 4 treatments (*crossed effects*).

We can look at the data below:

```
library(faraway)
library(ggplot2)
?rats
```

```
## Help on topic 'rats' was found in the following packages:
##
##   Package               Library
##   survival              /Library/Frameworks/R.framework/Versions/3.6/Resources
##   faraway               /Library/Frameworks/R.framework/Versions/3.6/Resources
##
##
## Using the first match ...
```

```
head(rats)
```

```
##   time poison treat
## 1 0.31      I     A
## 2 0.82      I     B
## 3 0.43      I     C
## 4 0.45      I     D
## 5 0.45      I     A
## 6 1.10      I     B
```

ANOVA and Experimental Design Jargon

Each combination of the factor level means is called **treatment**. So For example, a *treatment* is the combination of poison type I and treatment B and so on and so forth. In the `rats` experiment, the factors are (fully) crossed, since each combination of their levels is represented, which results in $3 \times 4 = 12$ treatments in total.

Remark: Here we need to make a discrimination between the treatment understood as a factor level combination (which is a statistical term) and the treatment that we give to rats (in this example) to "treat" the effect of the poison (which is used in layman's terms). Unfortunately, this term is used in both ways in studies, textbooks and analysis and therefore we have to be careful with its meaning every time.

When an experiment is replicated n times and when all treatment groups (i.e. factor level combinations) have an equal number of experimental units, the design is called **balanced**. In this section, we **assume that we have a balanced ANOVA**. This assumption will be lifted later.

In terms of notation, we denote by

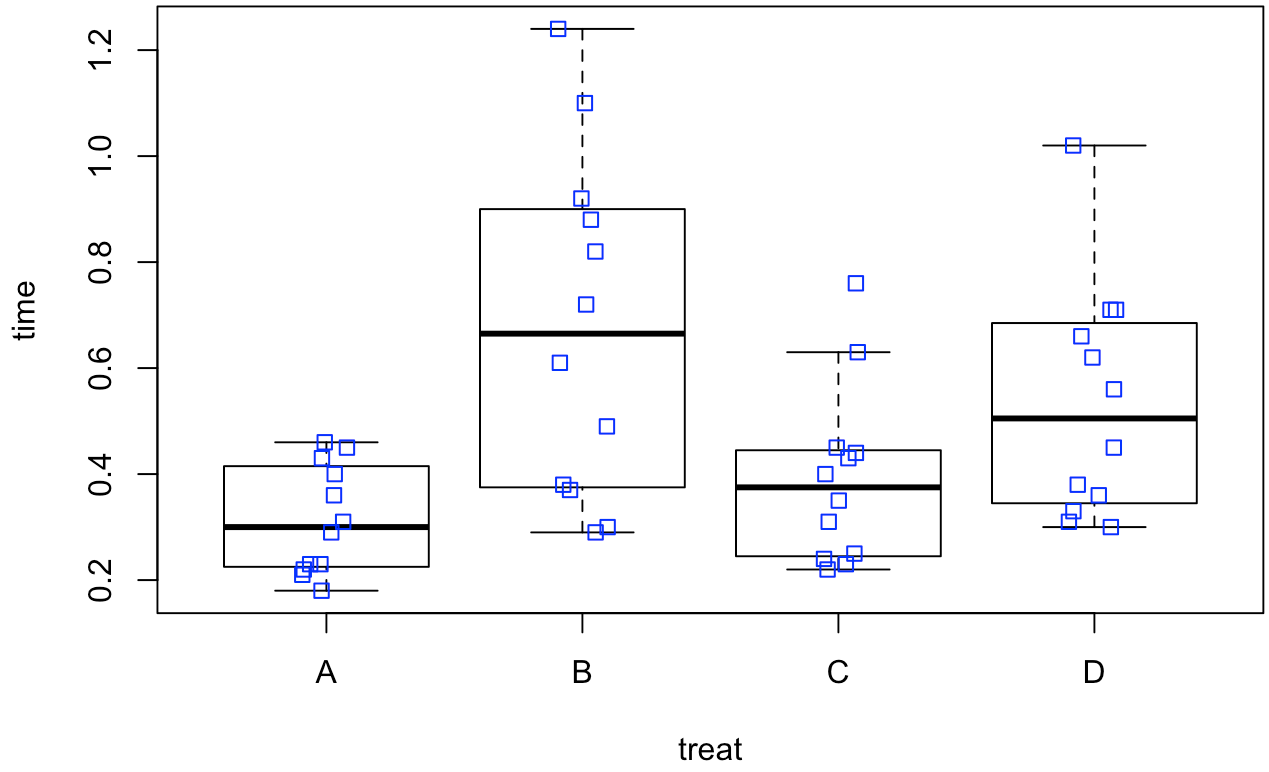
$$y_{ijk} = \text{the response of the } k^{th} \text{experimental unit that was assigned to level } i \text{of factor A and }$$

So, in the `rats` experiment y_{ijk} is the survival time for the kth animal that receives the ith poison (factor A) and the jth treatment (factor B). Obviously, which factor is *"factor A"* and which factor is *"factor B"* is arbitrary and can change.

As we discussed before, the best way to visualize a study with more than one factors is a side-by-side boxplot. When we have more than one factors, we construct a plot for **each** one of the factors.

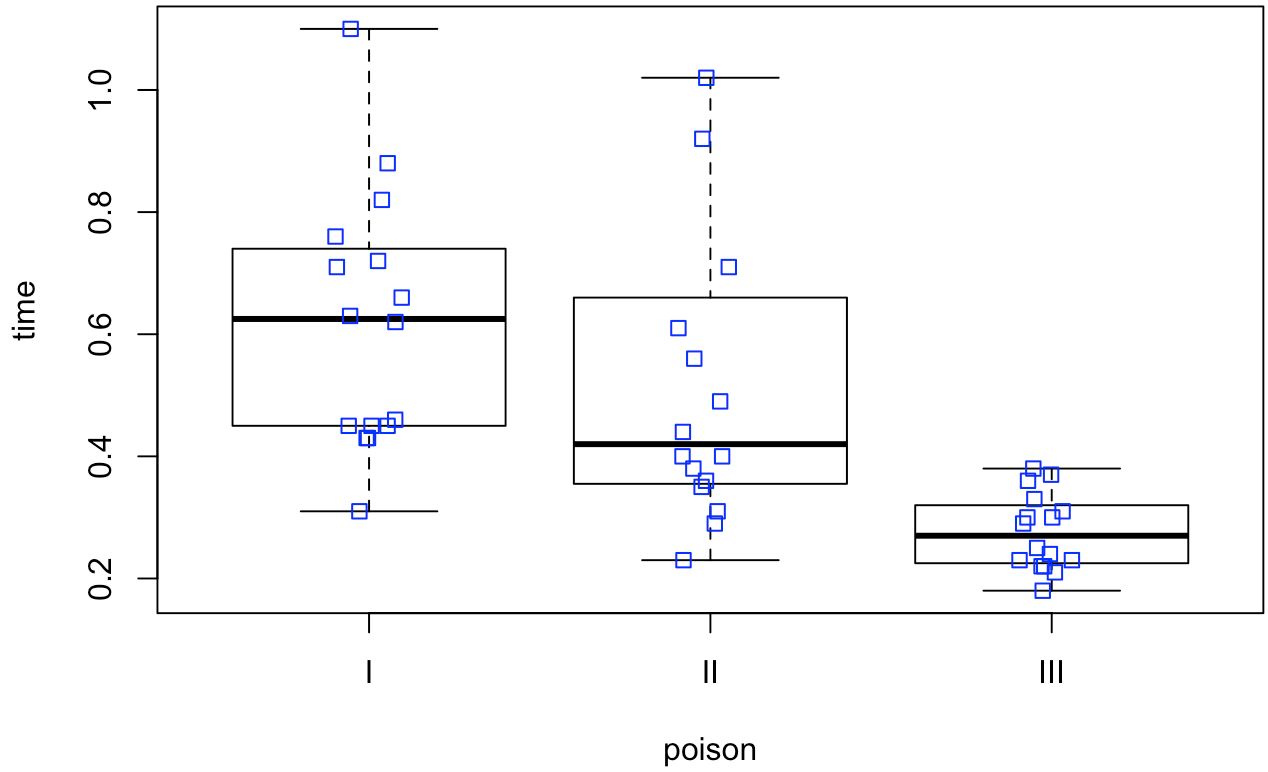
The Rats Experiment

```
boxplot(time~treat, data=rats, outline=FALSE)
stripchart(time~treat, data=rats, method="jitter",
           col="blue", vertical=TRUE, add=TRUE)
```



Based on this plot, we observe that treatments A and D are at the same level and B and D are at the same level as well and slightly higher than A and D, although we cannot tell whether this observed difference is statistically significant.

```
boxplot(time~poison, data=rats, outline=FALSE)
stripchart(time~poison, data=rats, method="jitter",
           col="blue", vertical=TRUE, add=TRUE)
```



Based on this plot, we observe that the survival time for type I poison is higher than type II poison and higher than type III poison. However, the boxes overlap, so we need to further investigate this relationship.

The boxplot is an informative wayto understand the relationship among the different levels of the factors, but it is constructed for one factor at a time (so it does not include interactions) and it does not help us make any statistical conclusions.

In order to formulate hypothesis tests, we need to define a model that will help us model the problem and construct appropriate test statistics. As in the one-way ANOVA there are many ways that we can express our model:

Cell Means Model for Two Factors

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

- μ_{ij} are the mean of the ith level of factor A and the jth level of factor B
- ε_{ijk} are independent $\mathcal{N}(0, \sigma^2)$
- $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, n$, where $n > 1$,
- $n_T = nab$: total sample size

We can decompose the means μ_{ij} as follows:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

where

- μ is the overall mean
- α_i are the factor A, `poison`, effects (fixed)
- β_j are the factor B, `"treatment"`, effects (fixed)
- $(\alpha\beta)_{ij}$ are the interaction effects (fixed)

In this case, we obtain the factor effects formulation of the ANOVA model:

Factor Effects Model for Two Factors

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

- ε_{ijk} are independent $\mathcal{N}(0, \sigma^2)$

This model is equivalent to the cell means model, but it is more useful when we want to understand the significance of each term in a software like `R`. And as before, here again we need to impose constraints to ensure that the estimators for the effects are *unique*.

Before proceeding let us fix the notation that will be used in the remaining chapter:

	Sum	Average
Cell (i,j)	$y_{ij\cdot} = \sum_{k=1}^{n} y_{ijk}$	$\bar{y}_{ij\cdot} = \frac{y_{ij\cdot}}{n}$
Row i	$y_{i\cdot\cdot} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$	$\bar{y}_{i\cdot\cdot} = \frac{y_{i\cdot\cdot}}{bn}$
Column j	$y_{\cdot j\cdot} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}$	$\bar{y}_{\cdot j\cdot} = \frac{y_{\cdot j\cdot}}{an}$
Overall	$y_{\cdots} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}$	$\bar{y}_{\cdots} = \frac{y_{\cdots}}{nab}$

Fitting of ANOVA

Using least squares method, the estimated treatment means are:

$$\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$$

The factor effects estimators depend on the constraints that we impose. For example, under the *sum-constraints* we have

$$\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}, \quad \hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$$

$$\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}$$

The fitted values and residuals compute as usual as

$$\hat{y}_{ijk} = \bar{y}_{ij\cdot}, \quad r_{ij} = y_{ijk} - \hat{y}_{ijk}$$

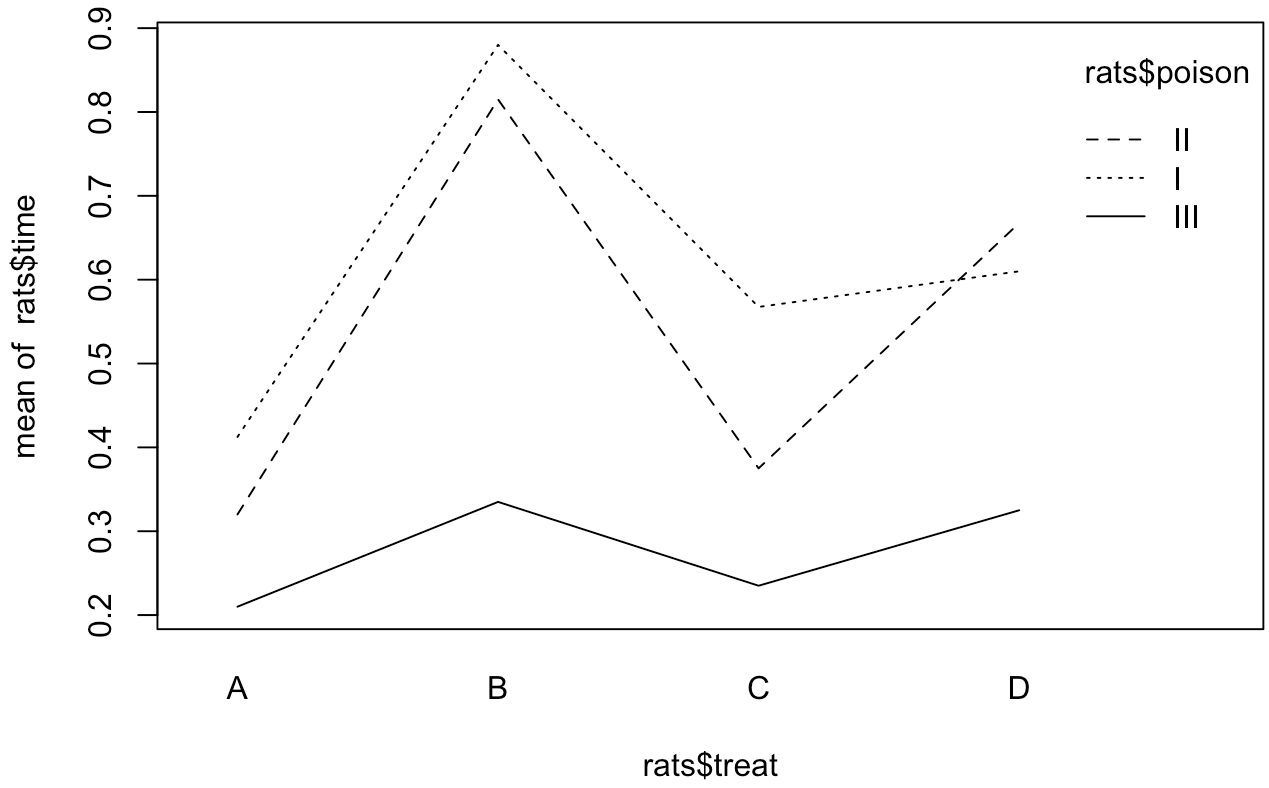
Since the model assumptions are the same, we can still use the residuals to do diagnostic plots and tests. We can also use all the remedial measures that we talked about before.

As far as it concerns the model selection process, then we use the hierarchical rule and we start testing for the significance of the higher order terms *first*. For the interactions, we also have a plot that helps us undestand whether interactions **are present or not**. This is the so-called interaction plot and is constructed as follows:

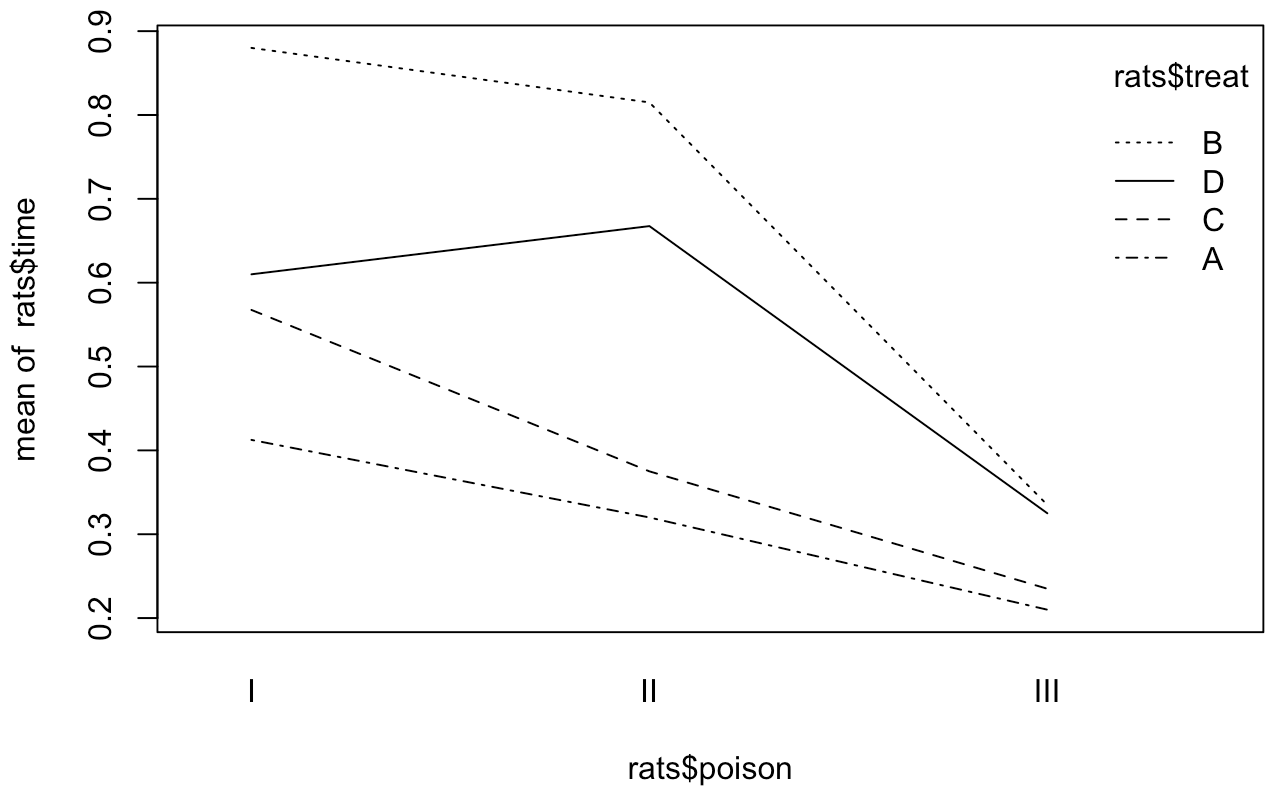
In this example, we found that the interactions are statistically significant.

The Rats Experiment

```
interaction.plot(rats$treat, rats$poison, rats$time)
```




```
interaction.plot(rats$poison, rats$treat, rats$time)
```



There are **intersecting** lines, so interactions are present although they seem to be not very strong.

However, in order to quantify whether the interactions are significant or not, we fir the ANOVA model and we perform an F test for the interaction term. We can see this process of model selection in the following example:

The Rats Experiment

We start with the model with interactions:

```
rats.full = lm(time ~ poison*treat, rats)
summary(rats.full)
```

```
##
## Call:
## lm(formula = time ~ poison * treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32500 -0.04875  0.00500  0.04312  0.42500
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.41250    0.07457   5.532 2.94e-06 ***
## poisonII         -0.09250    0.10546  -0.877   0.3862
## poisonIII        -0.20250    0.10546  -1.920   0.0628 .
## treatB            0.46750    0.10546   4.433 8.37e-05 ***
## treatC            0.15500    0.10546   1.470   0.1503
## treatD            0.19750    0.10546   1.873   0.0692 .
## poisonII:treatB   0.02750    0.14914   0.184   0.8547
## poisonIII:treatB -0.34250    0.14914  -2.297   0.0276 *
## poisonII:treatC  -0.10000    0.14914  -0.671   0.5068
## poisonIII:treatC -0.13000    0.14914  -0.872   0.3892
## poisonII:treatD   0.15000    0.14914   1.006   0.3212
## poisonIII:treatD -0.08250    0.14914  -0.553   0.5836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1491 on 36 degrees of freedom
## Multiple R-squared:  0.7335, Adjusted R-squared:  0.6521
## F-statistic:  9.01 on 11 and 36 DF,  p-value: 1.986e-07
```

```
anova(rats.full)
```

```
## Analysis of Variance Table
##
## Response: time
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## poison        2 1.03301 0.51651 23.2217 3.331e-07 ***
## treat         3 0.92121 0.30707 13.8056 3.777e-06 ***
## poison:treat  6 0.25014 0.04169  1.8743    0.1123
## Residuals    36 0.80073 0.02224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Looking at the p-value, we observe that the interaction term is not statistically significant, so we remove it and then fit the additive model:

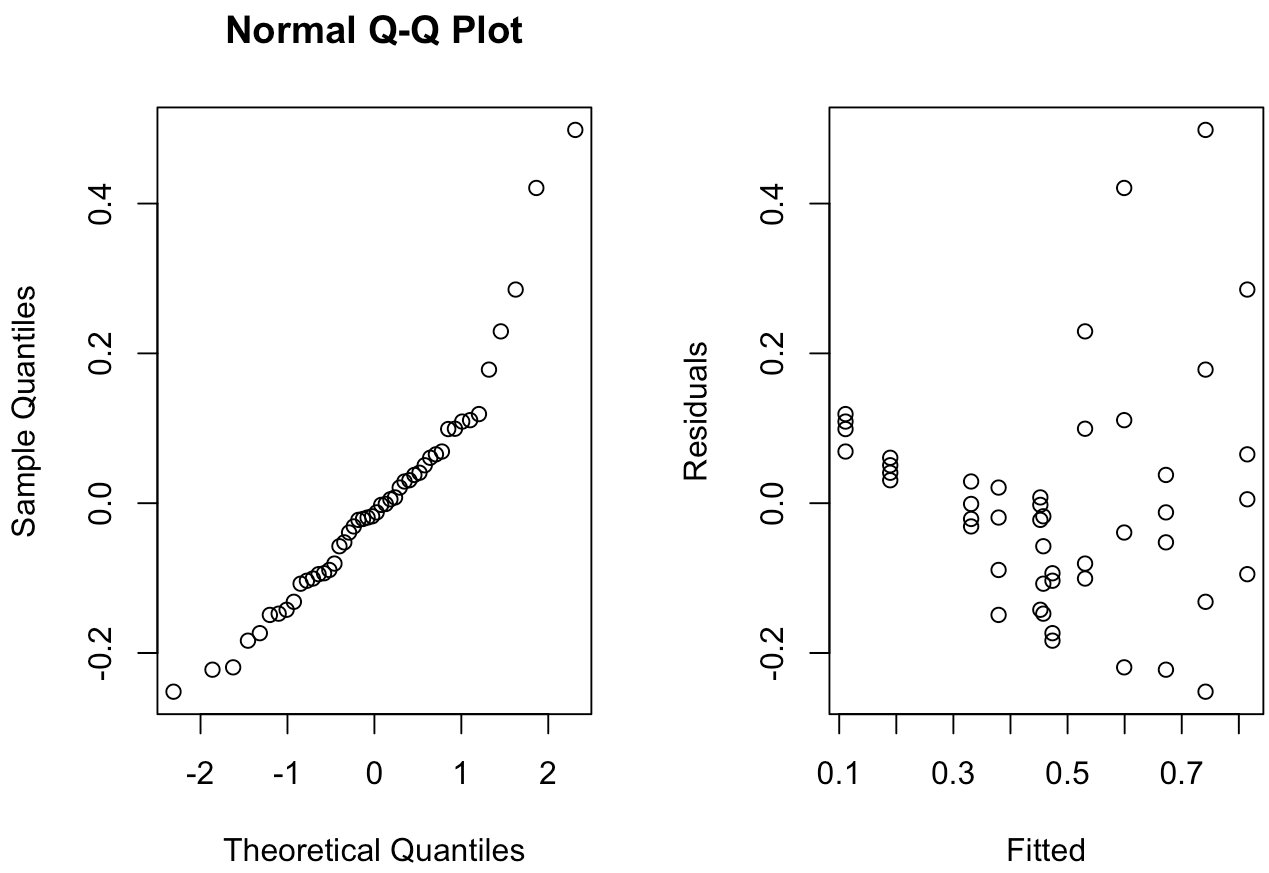
```
rats.additive = lm(time ~ poison+treat, rats)
anova(rats.additive)
```

```
## Analysis of Variance Table
##
## Response: time
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## poison     2 1.03301 0.51651  20.643 5.704e-07 ***
## treat      3 0.92121 0.30707  12.273 6.697e-06 ***
## Residuals 42 1.05086 0.02502
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the additive model, both factors are statistically significant, so this is the final model.

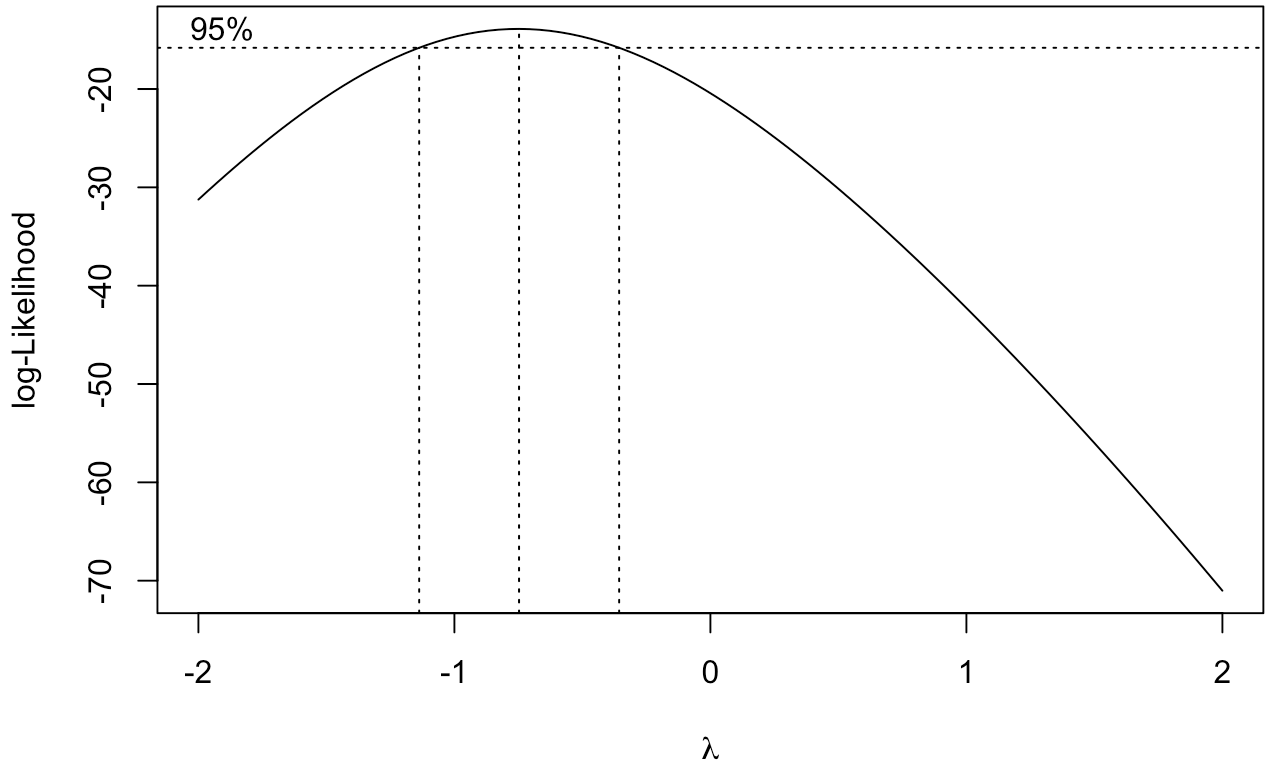
We are now ready to check the model assumptions:

```
par(mfrow=c(1,2))
qqnorm(rats.additive$res)
plot(rats.additive$fitted, rats.additive$res, xlab="Fitted", ylab="Residuals")
```



Notice the *trumpet* pattern for the residuals, so try Box-cox transformation.

```
library(MASS)
boxcox(rats.additive)
```



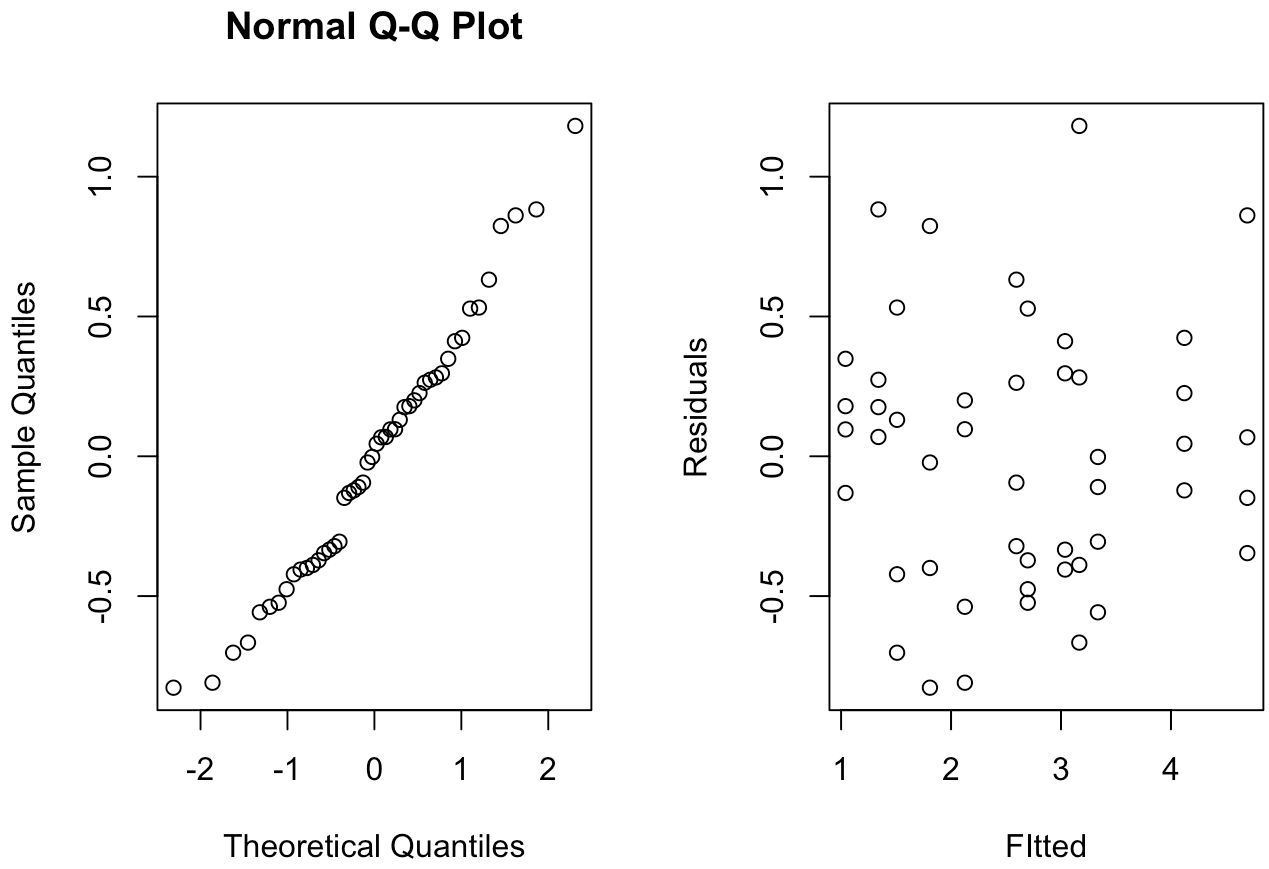
Try the reciprocal transformation ($\lambda = -1$)

```
rats.additive.inv=lm(1/time ~ poison+treat, rats)
anova(rats.additive.inv)
```

```
## Analysis of Variance Table
##
## Response: 1/time
##           Df Sum Sq Mean Sq F value    Pr(>F)
## poison     2 34.877 17.4386  71.708 2.865e-14 ***
## treat      3 20.414  6.8048  27.982 4.192e-10 ***
## Residuals 42 10.214  0.2432
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We check the model assumptions again to see whether the transfomration worked:

```
par(mfrow=c(1,2))
qqnorm(rats.additive.inv$res)
plot(rats.additive.inv$fitted, rats.additive.inv$res, xlab="FItted", ylab="Residu
```



Everything looks much better now. Based on these plots, the normality and constant variance assumptions are satisfied.

Let us now understand how the tests in the ANOVA tables are constructed and how the various degrees of freedom are computed:

Partitioning of Total Sum of Squares

$$\underbrace{y_{ijk} - \bar{y}_{...}}_{\text{Total Deviation}} = \underbrace{\bar{y}_{ij.} - \bar{y}_{...}}_{\substack{\text{Deviation of estimated} \\ \text{treatment mean around} \\ \text{overall mean}}} + \underbrace{y_{ijk} - \bar{y}_{ij.}}_{\substack{\text{Deviation} \\ \text{around estimated} \\ \text{treatment mean}}}$$

$$TSS = FSS + RSS$$

where

$$TSS = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2$$
$$FSS = n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{...})^2$$
$$RSS = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 = \sum_i \sum_j \sum_k e_{ijk}^2$$

However, the FSS is not very informative, since it does not allow us to understand the significance of each term in the model. Therefore, we partition it even further;

$$\underbrace{\bar{y}_{ij.} - \bar{y}_{...}}_{\text{Deviation of estimated treatment mean around overall mean}} = \underbrace{\bar{y}_{i..} - \bar{y}_{...}}_{\text{A main effect}} + \underbrace{\bar{y}_{.j.} - \bar{y}_{...}}_{\text{B main effect}} + \underbrace{\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}}_{\text{A B interaction effect}}$$

$$FSS = SSA + SSB + SSAB \text{ (Orthogonal Decomposition)}$$

where

$$SSA = nb \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$$
$$SSB = na \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$$
$$SSAB = n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

This leads us to the following ANOVA table

Source of Variation	SS	df	MS
Factor A	SSA	$a-1$	$MSA = \frac{SSA}{a-1}$
Factor B	SSB	$b-1$	$MSB = \frac{SSB}{b-1}$
AB Interactions	$SSAB$	$(a-1)(b-1)$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$
Error	RSS	$ab(n-1)$	$MSE = \frac{RSS}{ab(n-1)}$
Total	TSS	$nab-1$	

The degrees of freedom are calculated as usual:

- Factor A has a levels which means $a-1$ degrees of freedom..
- Factor B has b levels which means $b-1$ degrees of freedom..
- Interaction AB has ab levels which means $(a-1)(b-1)$ degrees of freedom.
- The total degrees of freedom are $n_T - 1 = nab - 1$.
- The residual degrees of freedom are $(nab-1) - (a-1) - (b-1) - (a-1)(b-1)$.

F Tests

The test for the significance of the interaction term can be done either using partial F tests or using the ANOVA F tests.

1. **Partial F-tests**
   We fit a main effects model (i.e. no interactions)

   $$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

   and we compare it with the *full* model (i.e. with interactions):

   $$\begin{cases} H_0: \text{ smaller model with } p_0 \text{ coefficients} \\ H_\alpha: \text{ larger model with } p_\alpha \text{ coefficients} \end{cases}$$

   The F-test is formulated as

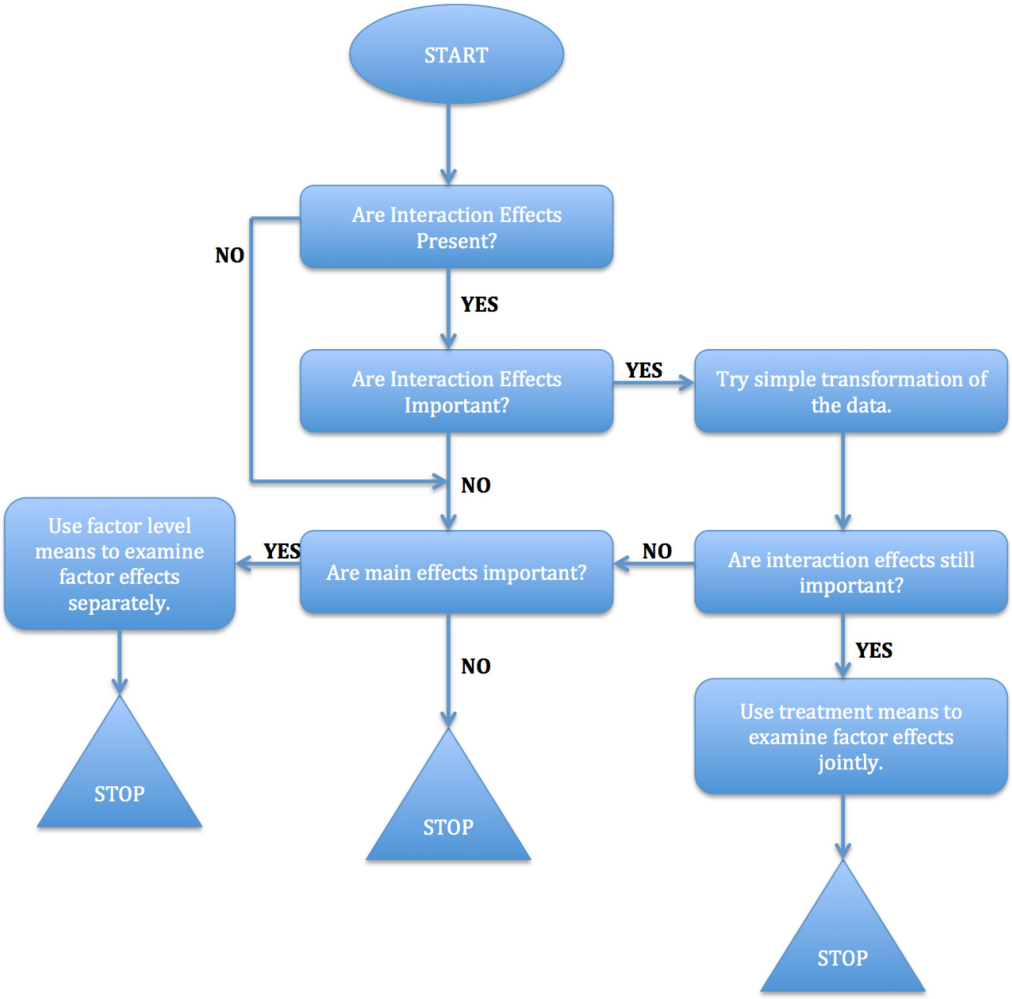
   $$F = \frac{(RSS_0 - RSS_\alpha)/(p_\alpha - p_0)}{MSE_\alpha} \sim F_{p_\alpha - p_0, n - p_\alpha} \text{ under the } H_0$$

2. **ANOVA Table F-tests**
   We can also perform F-tests directly using the ANOVA table, where for the interaction term we have:

   $$F_{AB} = \frac{MSAB}{MSE} \sim F_{(a-1)(b-1), nab-1}$$

   Hierarchy principle: we test for main effects *only if* the interaction term is not statistically significant.

In general, the analysis of 2 or more factor studies is summarized in the following flowchart:



As we can see finding the best model to the data is the first step. How we are going to proceed depends on one thing: whether the interactions are statistically significant or not. If the interactions are **not** statistically significant, then we focus on understanding the effect of a factor on the response for each factor separately while when the interactions **are** statistically significant, we work with the combinations of factor levels and we try to quantify their effect on the response.

7.3.1 Estimation of Factor Level Means

When interactions are *not statistically significant*, we analyze the factor level means[19]:

Estimation of Factor Level Means

- **Factor Level Means**: $\hat{\mu}_{i\cdot} = \bar{y}_{i\cdot\cdot}, \ s^2_{\hat{\mu}_{i\cdot}} = \frac{MSE}{bn}$

- **Differences of Factor Level Means**:

$$\hat{\mu}_{i\cdot} - \hat{\mu}_{i'\cdot} = \bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}, \; s^2_{\hat{D}} = \frac{2MSE}{bn}$$

- **Contrasts of Factor Level Means**:

$$\hat{L} = \sum c_i \hat{\mu}_{i\cdot} = \sum c_i \bar{y}_{i\cdot\cdot}, \; s^2_{\hat{L}_{(i)}} = \frac{MSE}{bn} \sum c_i^2$$

  where $\sum c_i = 0$.

For individual hypothesis test and CIs, the multiplier is $T_{(n-1)ab}(\alpha/2)$.

For family hypothesis tests/intervals, we select the desired family multiplier:

– *Tukey Multiplier*: $\frac{1}{\sqrt{2}} q_{a,(n-1)ab}(1-\alpha)$

– *Bonferroni Multiplier*: $B = T_{(n-1)ab}(1-\alpha/2m)$, where m refers to the number of multiple comparisons.

– *Scheff'e Multiplier*: $S^2 = (b-1)F_{b-1,(n-1)ab}(1-\alpha)$, if the contrasts involve the $\mu_{i\cdot}$ and $S^2 = (a-1)F_{a-1,(n-1)ab}(1-\alpha)$, if the contrasts involve the $\mu_{\cdot j}$.

The Rats Experiment

Just to get an idea of the effects of the levels of the factors for the inverse response, we fit a factor effects model with sum constraints. Note that we need to define the *sum constraints* for each of the factors separately:

```
contrasts(rats$poison) <- contr.sum
contrasts(rats$treat) <- contr.sum
```

Looking at the following summary output:

```
rats.additive.inv2 = lm(1/time ~ poison+treat, rats)
summary(rats.additive.inv2)
```

```
##
## Call:
## lm(formula = 1/time ~ poison + treat, data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82757 -0.37619  0.02116  0.27568  1.18153
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.62238    0.07118  36.842  < 2e-16 ***
## poison1     -0.82169    0.10066  -8.163 3.32e-10 ***
## poison2     -0.35305    0.10066  -3.507  0.00109 **
## treat1       0.89697    0.12328   7.276 5.88e-09 ***
## treat2      -0.76043    0.12328  -6.168 2.28e-07 ***
## treat3       0.32483    0.12328   2.635  0.01174 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4931 on 42 degrees of freedom
## Multiple R-squared:  0.8441, Adjusted R-squared:  0.8255
## F-statistic: 45.47 on 5 and 42 DF,  p-value: 6.974e-16
```

we compute the effects as follows:

Remark: In the notation below, we use y^T to denote that we work with the transformed y, i.e. $y^T = 1/y$.

- Grand Mean

$$\hat{\mu} = 2.62238 = \bar{y}^T_{\cdots}$$

- Poison -> Factor A

From the output we obtain:

$$\hat{\alpha}^T_I = -0.82169$$

$$\hat{\alpha}_{II}^T = -0.35305$$

$$\hat{\alpha}_{III}^T = 0.82169 + 0.35305 = 1.1747$$

The last estimator is obtained using the sum constraint: $\sum_i \alpha_i^T = 0$.

- Treatment -> Factor B

From the output we obtain:

$$\hat{\beta}_A^T = 0.89697$$

$$\hat{\beta}_B^T = -0.76043$$

$$\hat{\beta}_C^T = 0.32483$$

$$\hat{\beta}_D^T = -0.89697 + 0.76043 - 0.32483 = -0.46137$$

The last estimator is obtained using the sum constraint: $\sum_j \beta_j^T = 0$ and we can interpret the estimators in a similar way.

Remark: It is important to note here that the above interpretation quantifies exactly the average effect of a factor level on the response, because *interactions are **not** statistically significant*.

Based on the point estimators, we can see that (for factor A for example), the mean survival time of rats who were given poison I is higher compared to the mean survival time of rats that were given poison II, and that was higher compared to the rats that were given poison III.

However, this is just an observation based on point estimators. To establish this fact with a statistical confidence, we need to perform a family test of all pairwise differences *for each factor separately*. Therefore, we use Tukey's test as follows:

```
TukeyHSD(aov(1/time ~ poison + treat, data=rats), "poison")
```

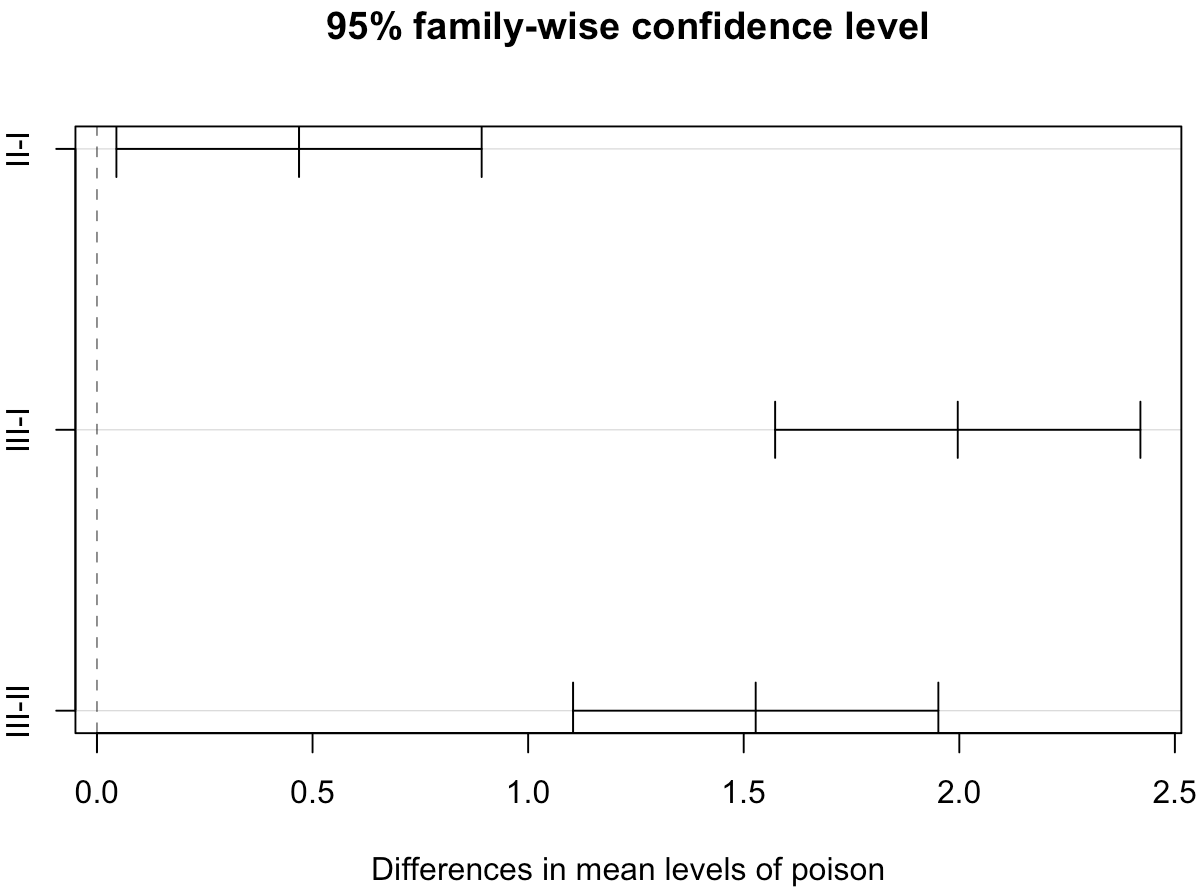
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = 1/time ~ poison + treat, data = rats)
##
## $poison
##             diff        lwr       upr     p adj
## II-I   0.4686413 0.04505584 0.8922267 0.0271587
## III-I  1.9964249 1.57283950 2.4200103 0.0000000
## III-II 1.5277837 1.10419824 1.9513691 0.0000000
```

Based on this output, by looking at the p-values, we can say that:

- All poison levels are statistically different.

We can also plot the intervals:

```
factorA_CI = TukeyHSD(aov(1/time ~ poison + treat, data=rats), "poison")
plot(factorA_CI)
```



Similarly for factor B (the treatment), we have

```
TukeyHSD(aov(1/time ~ poison + treat, data=rats), "treat")
```

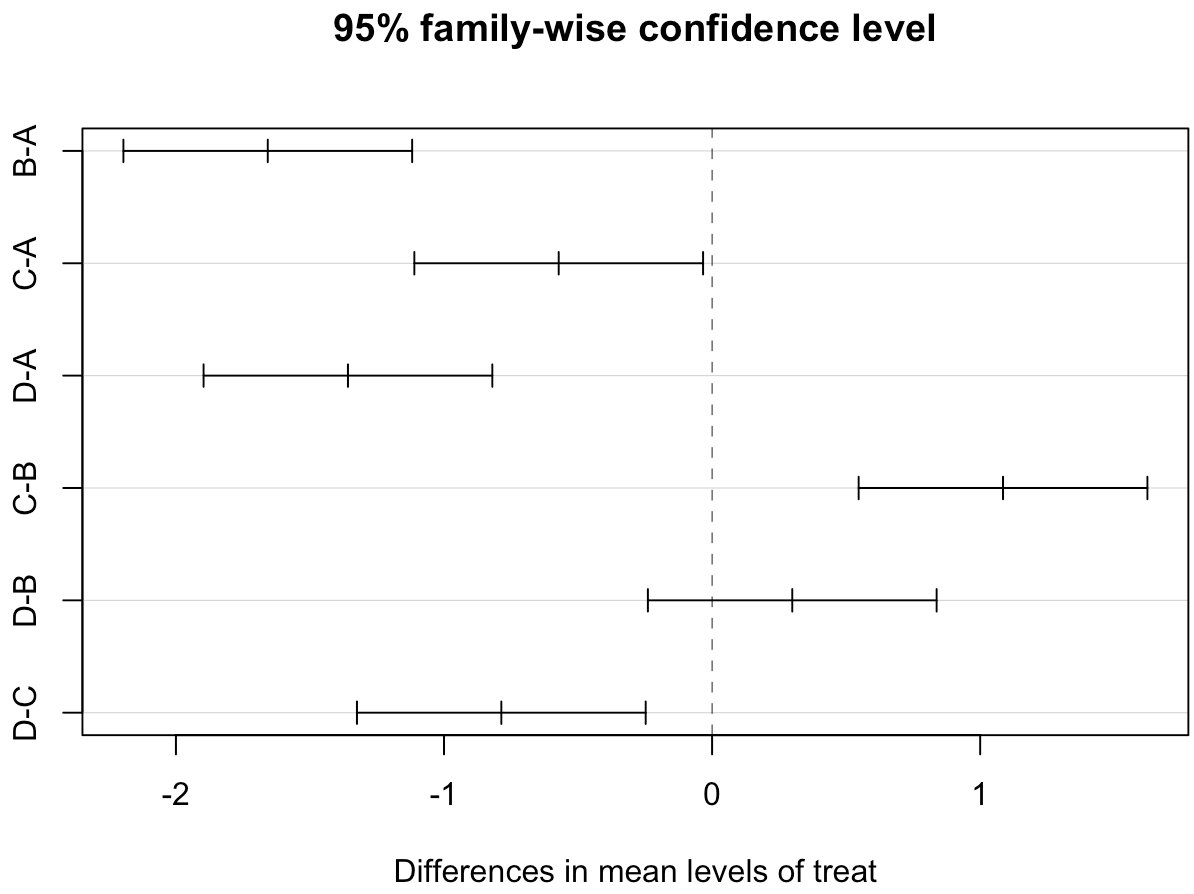
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = 1/time ~ poison + treat, data = rats)
##
## $treat
##           diff        lwr         upr     p adj
## B-A -1.6574024 -2.1959343 -1.11887050 0.0000000
## C-A -0.5721354 -1.1106673 -0.03360355 0.0335202
## D-A -1.3583383 -1.8968702 -0.81980640 0.0000002
## C-B  1.0852669  0.5467351  1.62379883 0.0000172
## D-B  0.2990641 -0.2394678  0.83759598 0.4550931
## D-C -0.7862029 -1.3247347 -0.24767096 0.0018399
```

Based on this output, by looking at the p-values, we can say that:

- Treatments A and B, C and A, D and A, C and B, D and C are statistically different, while D and B are the (statistically) the same.

We can also plot the intervals:

```
factorB_CI = TukeyHSD(aov(1/time ~ poison + treat, data=rats), "treat")
plot(factorB_CI)
```



7.3.2 Estimation of Treatment Means

Estimation of Treatment Means

When interactions are statistically significant, we analyze the treatment means:

- **Treatment Means**: $\hat{\mu}_{ij} = \bar{y}_{ij\cdot}, \ s^2_{\hat{\mu}_{ij}} = \frac{MSE}{n}$
- **Differences of Treatment Means**:

$$\hat{D} = \hat{\mu}_{ij} - \hat{\mu}_{i'j'} = \bar{y}_{ij\cdot} - \bar{y}_{i'j'\cdot}, \ \ i, j \neq i', j' \text{ and } s^2_{\hat{D}} = \frac{2MSE}{n}$$

- **Contrasts of Treatment Means**:

$$\hat{L} = \sum\sum c_{ij}\hat{\mu}_{ij} = \sum\sum c_{ij}\bar{y}_{ij\cdot}, \text{ where } \sum\sum c_{ij} = 0$$

with variance $s^2_{\hat{L}} = \frac{MSE}{n}\sum c^2_{ij}$.

For individual hypothesis test and CIs, the multiplier is $T_{(n-1)ab}(\alpha/2)$.

For family hypothesis tests/intervals, we select the desired family multiplier:

– *Tukey Multiplier*: $\frac{1}{\sqrt{2}}q_{ab,(n-1)ab}(1-\alpha)$

– *Bonferroni Multiplier*: $B = T_{(n-1)ab}(1-\alpha/2m)$, where m refers to the number of multiple comparisons.

– *Scheff'e Multiplier*: $S^2 = (ab-1)F_{ab-1,(n-1)ab}(1-\alpha)$.

Let's work on an example where the interactions are statistically significant:

The Hay Fever Relief Study

A research laboratory was developing a new compound for the relief of severe cases of hay fever. In an experiment with 36 volunteers, the amounts of the two active ingredients (factor A and B) in the compound were varied at three levels each. Randomization was used in assigning 4 volunteers to each of the nine treatments. The data on hours of relief are in the `hayfever.txt` data set.

```
fever= read.table("data/ch7/hayfever.txt")
colnames(fever)[1] = "hours"
colnames(fever)[2] = "ingredientA"
colnames(fever)[3] = "ingredientB"
colnames(fever)[4] = "rep"
```

Since the three levels of the two factors are numerical values, we need to alert `R` that they should be treated as factors:

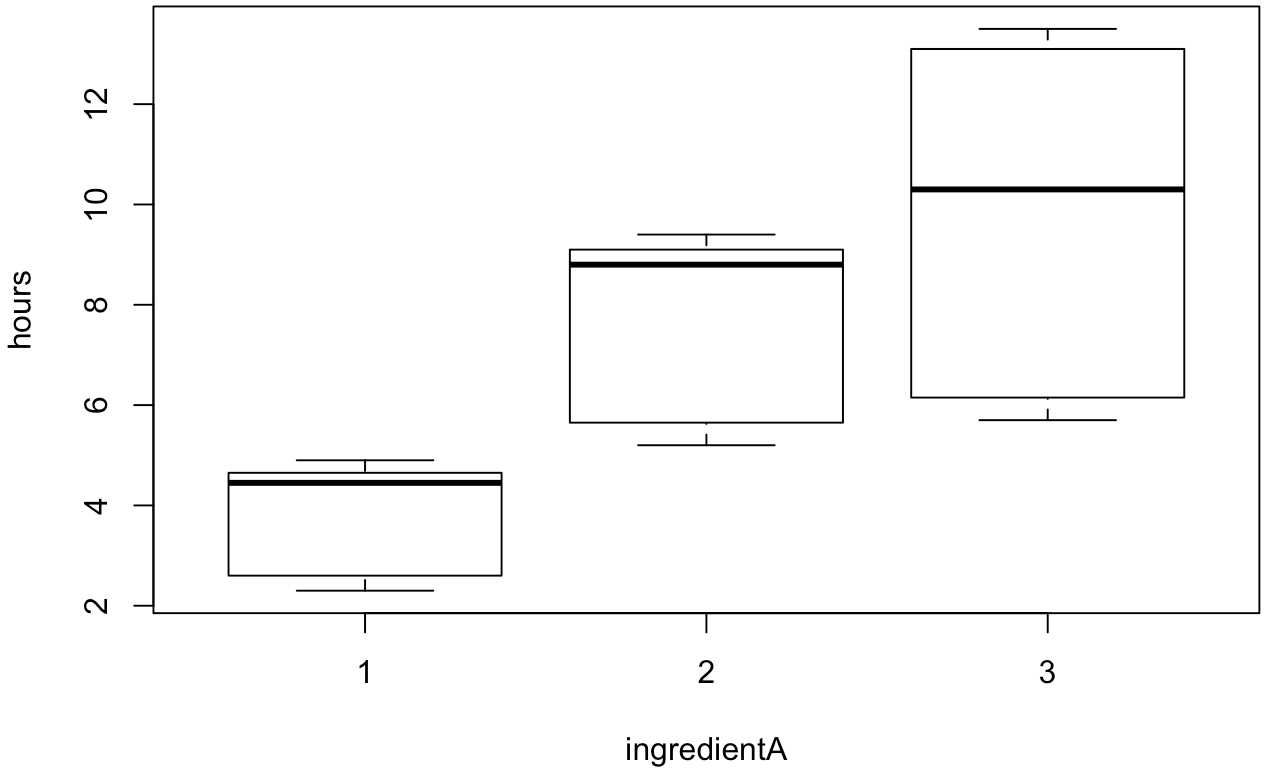
```r
fever$ingredientA= as.factor(fever$ingredientA)
fever$ingredientB = as.factor(fever$ingredientB)
fever$rep = as.factor(fever$rep)

head(fever)
```

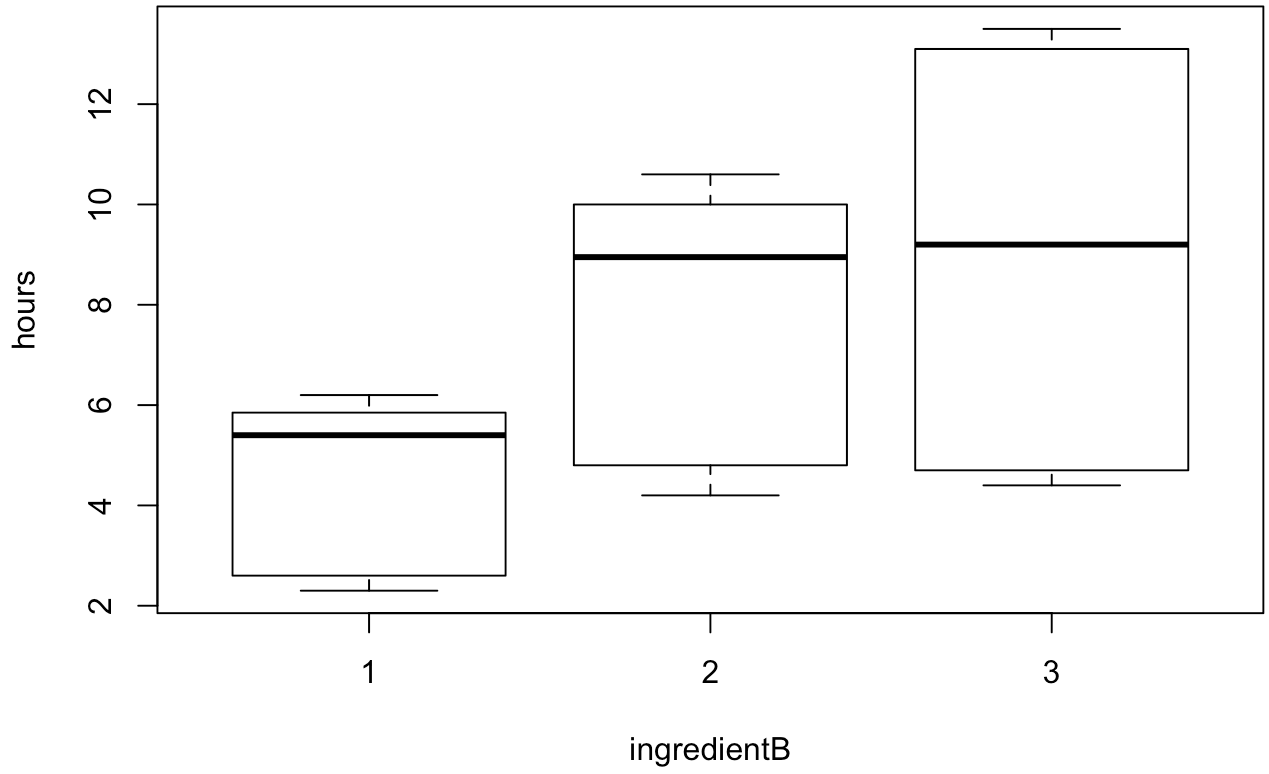
```
##   hours ingredientA ingredientB rep
## 1   2.4           1           1   1
## 2   2.7           1           1   2
## 3   2.3           1           1   3
## 4   2.5           1           1   4
## 5   4.6           1           2   1
## 6   4.2           1           2   2
```

We visualize our data with side-by-side boxplots although for each factor separately:

```r
boxplot(hours~ingredientA, data=fever)
```

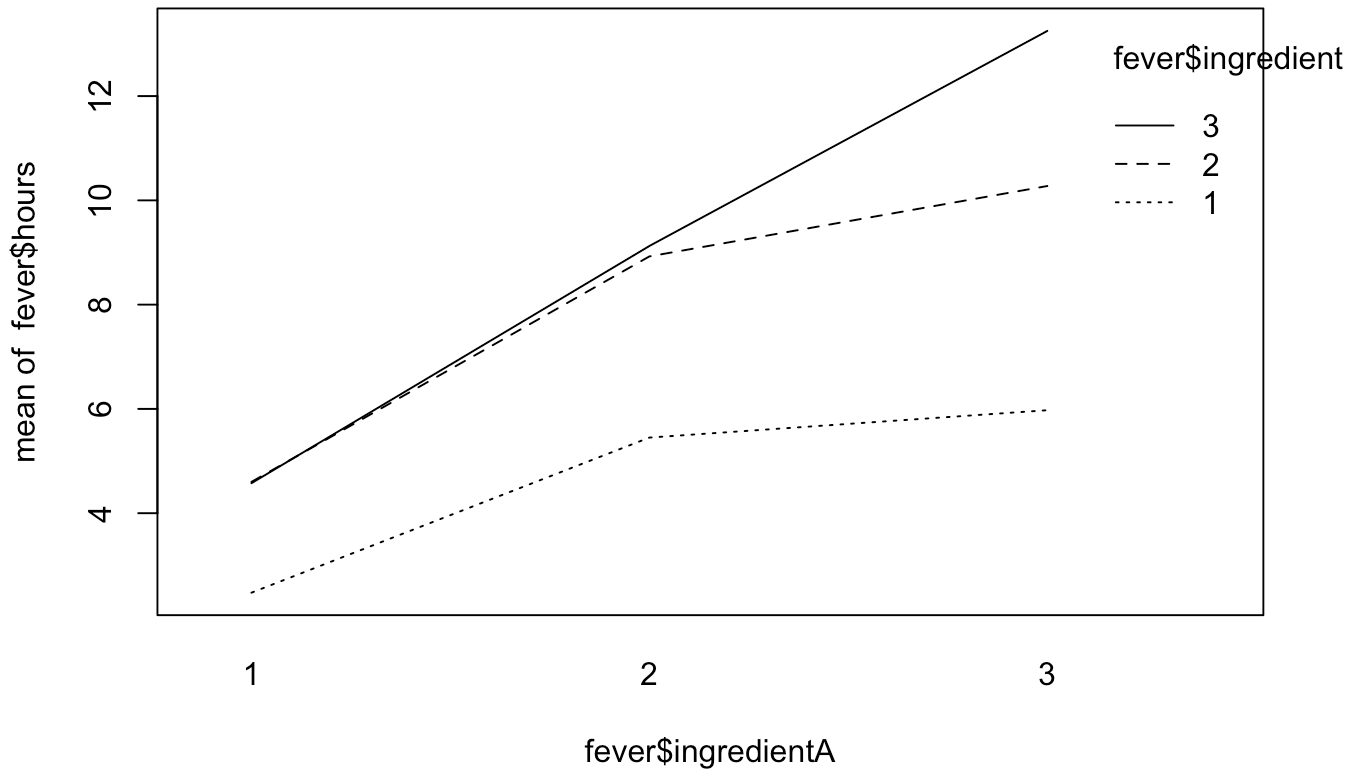



```
boxplot(hours~ingredientB, data=fever)
```

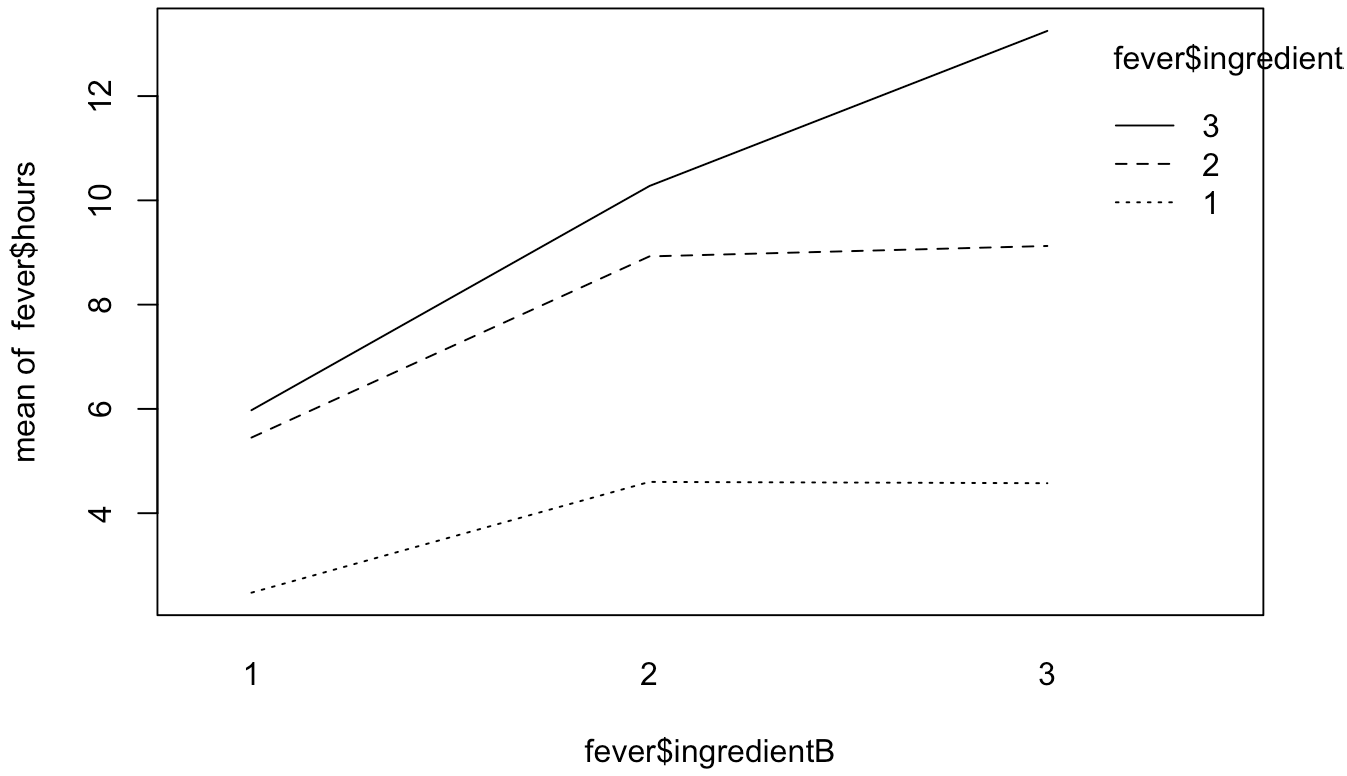


To investigate whether *interaction* are *present*, we construct the interaction plots:

```
interaction.plot(fever$ingredientA, fever$ingredientB, fever$hours)
```




```
interaction.plot(fever$ingredientB, fever$ingredientA, fever$hours)
```



Since lines are intersecting, we conclude that interactions are present.

We start by fitting the full model with both factors and interaction term:

```
my.contrasts <- list(ingredientA = "contr.sum", ingredientB = "contr.sum" )
fever.full = lm(hours ~ ingredientA*ingredientB, fever, contrasts=my.contrasts)
anova(fever.full)
```

```
## Analysis of Variance Table
##
## Response: hours
##                          Df  Sum Sq Mean Sq F value    Pr(>F)
## ingredientA               2 220.020 110.010 1827.86 < 2.2e-16 ***
## ingredientB               2 123.660  61.830 1027.33 < 2.2e-16 ***
## ingredientA:ingredientB   4  29.425   7.356  122.23 < 2.2e-16 ***
## Residuals                27   1.625   0.060
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe that the interaction term is statistically significant, so the full model is the model we are going to choose.

If we want to compute the estimators for effects and interaction terms, we need to fit a factor effects model with appropriate constraints. We choose the following sum constraints here:

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$

$$\sum_i \alpha_i = 0, \ \sum_j \beta_j = 0$$

```
summary(fever.full)
```

```
##
## Call:
## lm(formula = hours ~ ingredientA * ingredientB, data = fever,
##     contrasts = my.contrasts)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4250 -0.1750  0.0125  0.1875  0.3500
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                7.18333    0.04089 175.684  < 2e-16 ***
## ingredientA1              -3.30000    0.05782 -57.070  < 2e-16 ***
## ingredientA2               0.65000    0.05782  11.241 1.09e-11 ***
## ingredientB1              -2.55000    0.05782 -44.099  < 2e-16 ***
## ingredientB2               0.75000    0.05782  12.970 4.10e-13 ***
## ingredientA1:ingredientB1  1.14167    0.08178  13.961 7.22e-14 ***
## ingredientA2:ingredientB1  0.16667    0.08178   2.038 0.051446 .
## ingredientA1:ingredientB2 -0.03333    0.08178  -0.408 0.686767
## ingredientA2:ingredientB2  0.34167    0.08178   4.178 0.000276 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2453 on 27 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9944
## F-statistic: 774.9 on 8 and 27 DF,  p-value: < 2.2e-16
```

Let's call Ingredient 1, factor A and Ingredient 2, factor B. So, the effects are:

$$\hat{\mu} = 7.18333$$

$$\hat{\alpha}_1 = -3.3,\ \hat{\alpha}_2 = 0.65,\ \hat{\alpha}_3 = 3.3 - 0.65 = 2.65$$

$$\hat{\beta}_1 = -2.55,\ \hat{\beta}_2 = 0.75,\ \hat{\beta}_3 = 2.55 - 0.75 = 1.8$$

$$\widehat{(\alpha\beta)}_{11} = 1.14167, \widehat{(\alpha\beta)}_{21} = 0.16667$$

$$\widehat{(\alpha\beta)}_{12} = -0.03333, \widehat{(\alpha\beta)}_{22} = 0.34167$$

The remaining effects will be estimated from the sum constraints for the interaction terms.

When interactions are present, it is not easy to clearly interpret the effects, since the effect of ingredient 1 on the response is not only $\hat{\alpha}_1$, since there is also a contributionn from the interaction term that we cannot clearly quantify.

So, the way that we interpret estimated values here would be more in terms of the means μ_{ij}. For example, the estimated mean hours of relief when level i from ingredient 1 is chosen and level j from ingredient 1 is chosen is equal to

$$\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$$

To further investigate the effects of the two ingredients, we can also perform family pairwise comparisons. Using Tukey's family coefficient, we compute all pairwise family comparisons:

```
TukeyHSD(aov(hours ~ ingredientA * ingredientB, data=fever), "ingredientA:ingredi
```

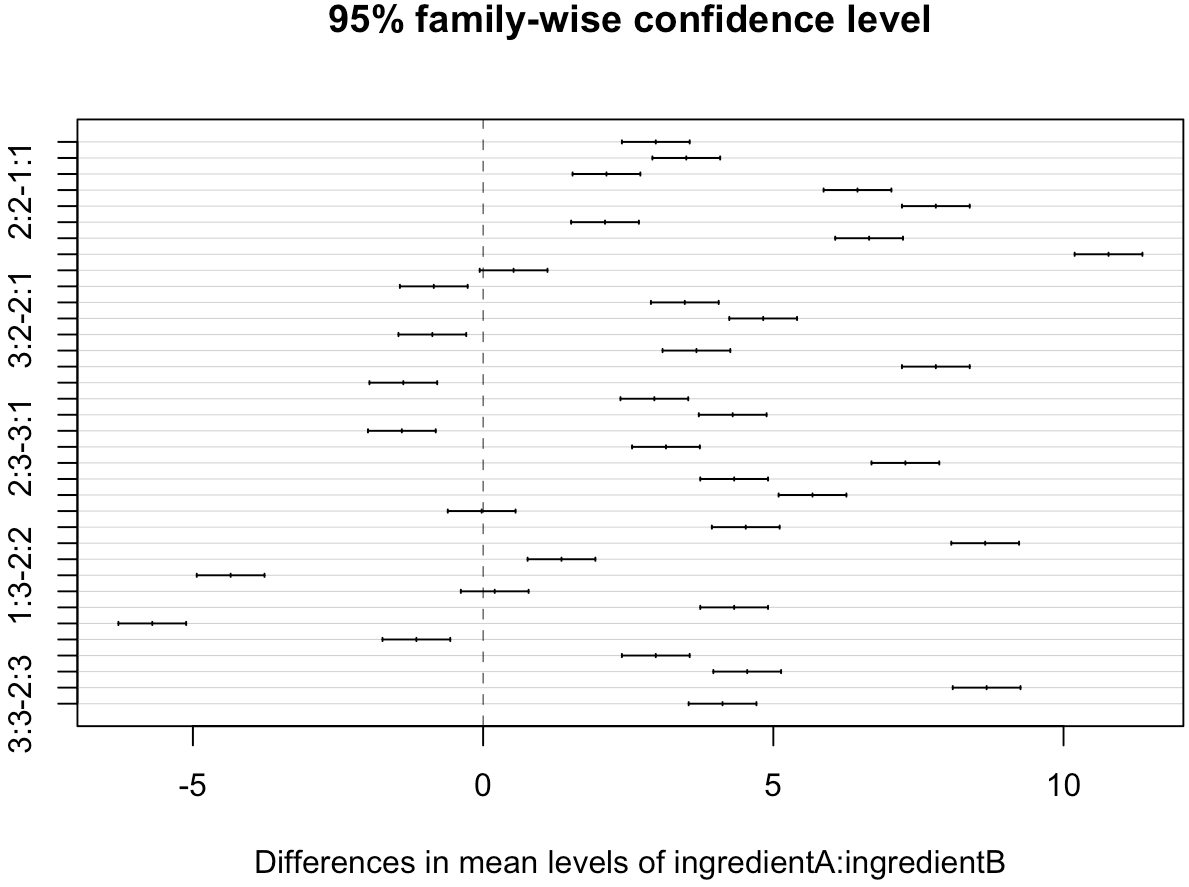
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = hours ~ ingredientA * ingredientB, data = fever)
##
## $`ingredientA:ingredientB`
##           diff        lwr        upr     p adj
## 2:1-1:1  2.975  2.3913187  3.5586813 0.0000000
## 3:1-1:1  3.500  2.9163187  4.0836813 0.0000000
## 1:2-1:1  2.125  1.5413187  2.7086813 0.0000000
## 2:2-1:1  6.450  5.8663187  7.0336813 0.0000000
## 3:2-1:1  7.800  7.2163187  8.3836813 0.0000000
## 1:3-1:1  2.100  1.5163187  2.6836813 0.0000000
## 2:3-1:1  6.650  6.0663187  7.2336813 0.0000000
## 3:3-1:1 10.775 10.1913187 11.3586813 0.0000000
## 3:1-2:1  0.525 -0.0586813  1.1086813 0.1033088
## 1:2-2:1 -0.850 -1.4336813 -0.2663187 0.0011424
## 2:2-2:1  3.475  2.8913187  4.0586813 0.0000000
## 3:2-2:1  4.825  4.2413187  5.4086813 0.0000000
## 1:3-2:1 -0.875 -1.4586813 -0.2913187 0.0007862
## 2:3-2:1  3.675  3.0913187  4.2586813 0.0000000
## 3:3-2:1  7.800  7.2163187  8.3836813 0.0000000
## 1:2-3:1 -1.375 -1.9586813 -0.7913187 0.0000005
## 2:2-3:1  2.950  2.3663187  3.5336813 0.0000000
## 3:2-3:1  4.300  3.7163187  4.8836813 0.0000000
## 1:3-3:1 -1.400 -1.9836813 -0.8163187 0.0000004
## 2:3-3:1  3.150  2.5663187  3.7336813 0.0000000
## 3:3-3:1  7.275  6.6913187  7.8586813 0.0000000
## 2:2-1:2  4.325  3.7413187  4.9086813 0.0000000
## 3:2-1:2  5.675  5.0913187  6.2586813 0.0000000
## 1:3-1:2 -0.025 -0.6086813  0.5586813 1.0000000
## 2:3-1:2  4.525  3.9413187  5.1086813 0.0000000
## 3:3-1:2  8.650  8.0663187  9.2336813 0.0000000
## 3:2-2:2  1.350  0.7663187  1.9336813 0.0000007
## 1:3-2:2 -4.350 -4.9336813 -3.7663187 0.0000000
```

```
## 2:3-2:2  0.200 -0.3836813  0.7836813 0.9596929
## 3:3-2:2  4.325  3.7413187  4.9086813 0.0000000
## 1:3-3:2 -5.700 -6.2836813 -5.1163187 0.0000000
## 2:3-3:2 -1.150 -1.7336813 -0.5663187 0.0000131
## 3:3-3:2  2.975  2.3913187  3.5586813 0.0000000
## 2:3-1:3  4.550  3.9663187  5.1336813 0.0000000
## 3:3-1:3  8.675  8.0913187  9.2586813 0.0000000
## 3:3-2:3  4.125  3.5413187  4.7086813 0.0000000
```

The treatment levels that are not statistically significant are: $\mu_{23} and \mu_{22}$, $\mu_{31} and \mu_{21}$, $\mu_{13} and \mu_{12}$.

If we want to select the optimal levels for the two factors, that lead to the “fastest” fever relief, then we need to look at the plot of all the intervals:

```
Interaction_CI = TukeyHSD(aov(hours ~ ingredientA * ingredientB, data=fever), "ir
plot(Interaction_CI)
```


95% family-wise confidence level
3:3-2:3
1:3-2:2
2:3-3:1
3:2-2:1
2:2-1:1
-5
0
5
10
Differences in mean levels of ingredientA:ingredientB