

Statistical Methods for Research II

Shizhe Chen

2020-01-02

Contents

Preface	5
1 Causal Inference	7
1.1 Association and Causality	7
1.2 Potential Outcomes	7
1.3 Experiments v.s. Observational Studies	7
1.4 Learning Objectives	7
2 One-way ANOVA	9
2.1 Simple randomized experiments	9
2.2 One-way ANOVA	9
2.3 Model diagnostics	12
2.4 Learning Objectives	18
3 Two-way ANOVA	21
3.1 Experiments with two (or more) factors	21
3.2 Two-way ANOVA	21
3.3 Learning Objectives	31

Preface

This file contains code and comments in STA 207. Note that this document include only the outline of the in-class lectures. Many materials in this document are adapted from lectures by Professors Prabir Burman, Peng Ding, and Kosuke Imai.

Chapter 1

Causal Inference

1.1 Association and Causality

Discussion

1.2 Potential Outcomes

- Definition of potential outcomes
- Properties of potential outcomes
- Key assumptions in this notation
- Immutable characteristics

1.3 Experiments v.s. Observational Studies

- Advantages of randomized experiments
- Reasons for observational studies

1.4 Learning Objectives

- Students are able to distinguish association and causality.
- Students gain basic familiarity with the potential outcome framework.
- Students are able to recognize the importance of experiment designs.

Chapter 2

One-way ANOVA

2.1 Simple randomized experiments

- Motivation and real world applications of randomized experiments.
- Sampling schemes of a simple randomized experiment.
- Question of interest, null hypotheses, and their causal interpretation.
- Intuition of hypothesis testing.
- Practical concerns in the design of experiments.

2.2 One-way ANOVA

2.2.1 A motivating example: the Spock trial

In 1968 Dr. Benjamin Spock was tried in Boston for conspiring against the government for helping young men to escape the military draft. He was convicted by the Boston federal court, but the judgement was overturned by the Court of Appeals in 1969 for many reasons, one of which was cited as the bias of the presiding judge Francis Ford. Dr. Spock, a pediatrician, was very famous for his books on rearing of children, and thus was widely admired by women. As a matter of fact, the jury in Spock trial has no women. Note that jury panels, though randomly selected, should reflect the demographics. In any particular trial, there may not be any woman on the jury, but it is worthwhile to examine if the jury panels of Judge Ford had fewer women than other judges in Boston in few months before the trial. Data are available for jury panels for 7, but we investigate the data for only 4 judges including Judge Ford.

```
Spock <- read.csv(file="./data/SpockTrial.csv", header=TRUE, sep=",")
Spock$Judge<-as.factor(Spock$Judge);
# Box plot with jittered points (from stackoverflow: https://stackoverflow.com/question)
boxplot(perc.women~Judge,data=Spock)
stripchart(perc.women~Judge, vertical = TRUE, data = Spock,
  method = "jitter", add = TRUE, pch = 20, col = 'blue')
```

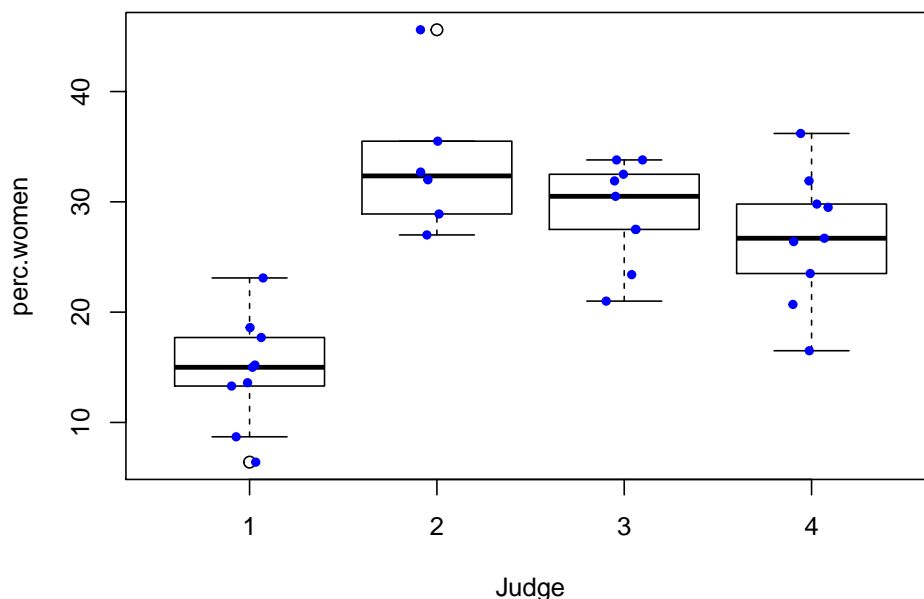


Figure 2.1: Box plot with jittered data points for the Spock trial data.

2.2.2 ANOVA model

- Cell means model
- Estimators of the means
- Decomposition of sum of squares
- Some basic properties

In the Spock trial data, we can use `aov()` to fit a one-way ANOVA model.

```
anova.fit<- aov(perc.women~Judge,data=Spock)
summary(anova.fit)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Judge      3   1591     530   17.6 1.1e-06 ***
## Residuals 29    874      30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can obtain the following.

- $n_1 = 9$, $n_2 = 6$, $n_3 = 9$, $n_4 = 9$, and $n_T = n_1 + n_2 + n_3 + n_4 = 33$.
- $\bar{Y}_{1\cdot} = 14.62$, $\bar{Y}_{2\cdot} = 33.62$, $\bar{Y}_{3\cdot} = 29.1$, and $\bar{Y}_{4\cdot} = 26.8$.
- $SSE = 873.52$, $df = 29$, and $MSE = 30.12$.
- $SSTO = 2464.8$, $df = 32$.
- $SSTR = 1591.28$, $df = 3$, and $MSTR = 530.43$.

2.2.3 Statistical inference

- Null hypothesis
- The F-test
- Testing a linear combination
 - Estimation
 - Hypothesis testing
- (Simultaneous) confidence intervals

To test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against the alternative $H_1 : \text{not all } \mu_i\text{'s are equal}$. We can calculate the F-statistics $F^* = \frac{MSTR}{MSE} =$, $F(0.95; 3, 29) = 2.93$. We can thus reject the null hypothesis at the chosen significance level 0.05.

Consider the quantity $L = \mu_1 - (\mu_2 + \mu_3 + \mu_4)/3$. We can calculate that $\hat{L} = -15.22$, and $s^2(\hat{L}) = 4.65$. Moreover, since $t(1 - 0.01/2; n_T - r) = 2.76$ a 99% confidence interval for L is $(-21.16, -9.27)$.

To test the hypothesis $H_0 : L = 0$ against $H_1 : L \neq 0$. We can calculate the t-statistics $t^* = \hat{L}/s(\hat{L}) = -7.06$. We can calculate the quantile of the t-distribution as before to finish the test.

Next we demonstrate how to construct the simultaneous confidence intervals for all pairwise difference. The code will construct confidence intervals for two pairs, but leave you to finish the rest.

```
# Create vectors for the first two linear combinations
comb.mat<-matrix(0,nrow=2,ncol=4)
comb.mat[1,]=c(1,-1,0,0);comb.mat[2,]=c(1,0,-1,0);

# Obtain the estimates
diff = numeric(dim(comb.mat)[1]);
diff.sd=diff;
mean.tmp=anova.fit$coefficients;mean.tmp[1]=0;
ns=as.numeric(table(Spock$Judge));
for(i in 1:length(diff)){
  diff[i]=sum(comb.mat[i,]*mean.tmp);
  diff.sd[i]=sqrt(sum(comb.mat[i,]^2*ns));
}

alpha=0.05;

# Bonferroni correction:
m=6; # for all pairwise differences, although we only show two here
B.stat=qt(1-alpha/(2*m),anova.fit$df.residual);

# Tukey-Kramer
T.stat=qtukey(1-alpha, nmeans=length(anova.fit$coefficients), df=anova.fit$df.residual),
```

```

# Scheffe
S.stat=sqrt( (length(anova.fit$coefficients)-1)*qf(1-alpha,length(anova.fit$coefficients)

table.stats=matrix(0,1,3);
table.stats[1,]=c(B.stat,T.stat,S.stat);
colnames(table.stats)=c('Bonferroni', 'Tukey', 'Scheffe')
table.stats

##      Bonferroni Tukey Scheffe
## [1,]         2.8   2.7      3

# Then, we can construct the confidence intervals as, e.g.,
CI.bonferroni =matrix(0,nrow=2,ncol=2);
for(i in 1:length(diff)){
  CI.bonferroni[i,]=diff[i]+c(1,-1)*B.stat*diff.sd[i];
}

```

2.2.4 Alternative forms of the ANOVA model

- Factor-effect model
- Regression model

The default of ANOVA in R set weights proportional to the sample size in each cell. You can supply the `weights` to the `aov()` function to force equal weights.

```

# Weights proportional to sample sizes
print(model.tables(anova.fit,"effects"))

```

```

## Tables of effects
##
## Judge
##      1      2      3      4
## -10.72 8.271 3.755 1.455
## rep   9.00 6.000 9.000 9.000

```

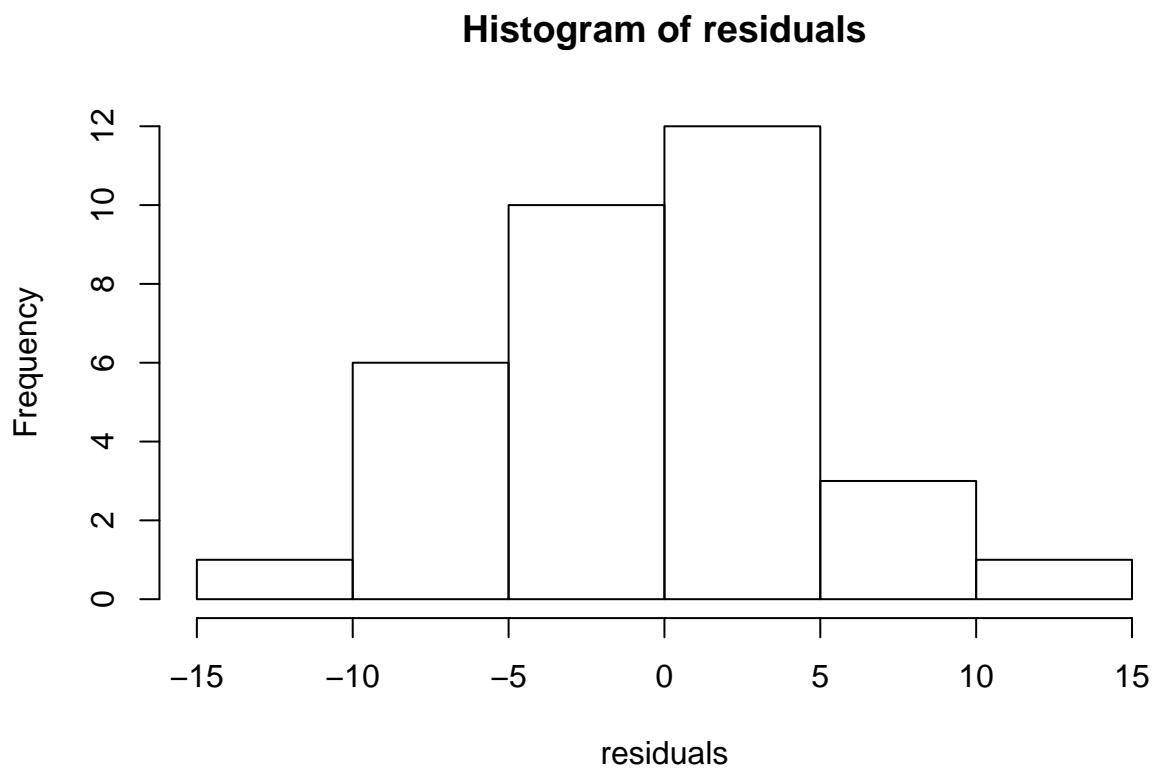
2.3 Model diagnostics

- Possible departures
- Diagnostics with residuals
- Graphical methods
- Formal tests
 - Hartley test
 - Bartlett test
 - Levene test
- Remedial measures

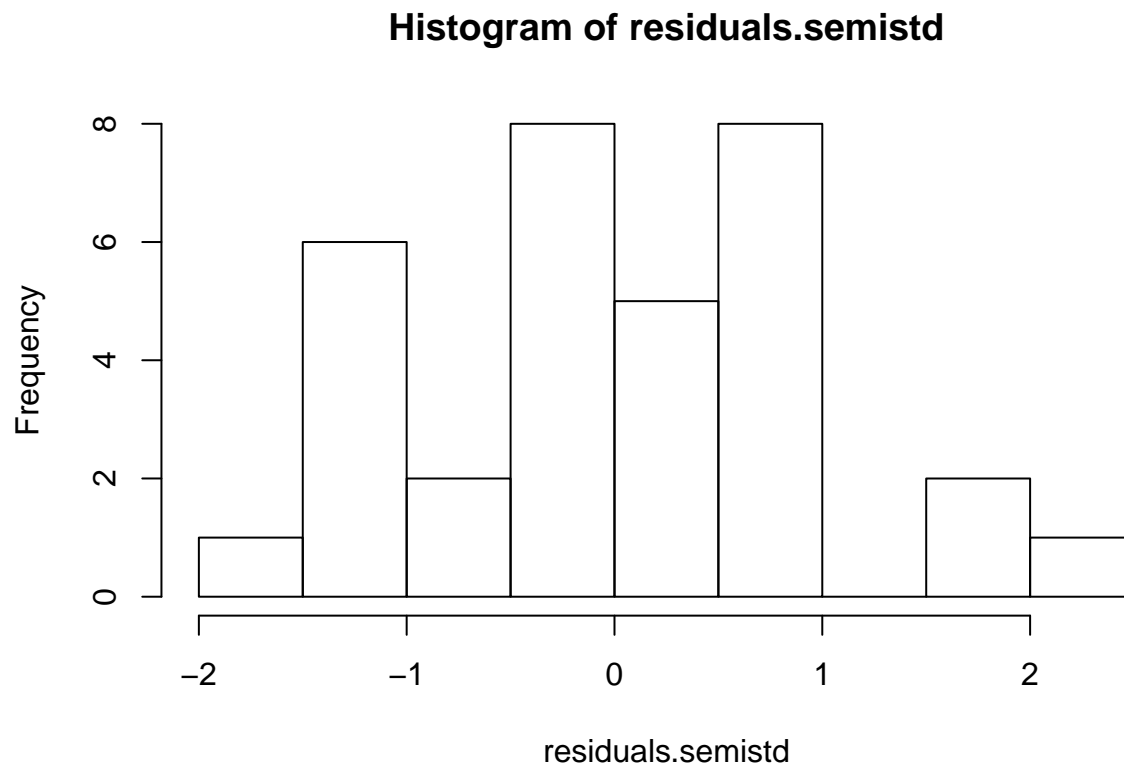
- Weighted least squares
- Nonparametric tests based on ranks: rank test, Kruskal-Wallis test
- Box-Cox transformation

All diagnostics start with the residuals.

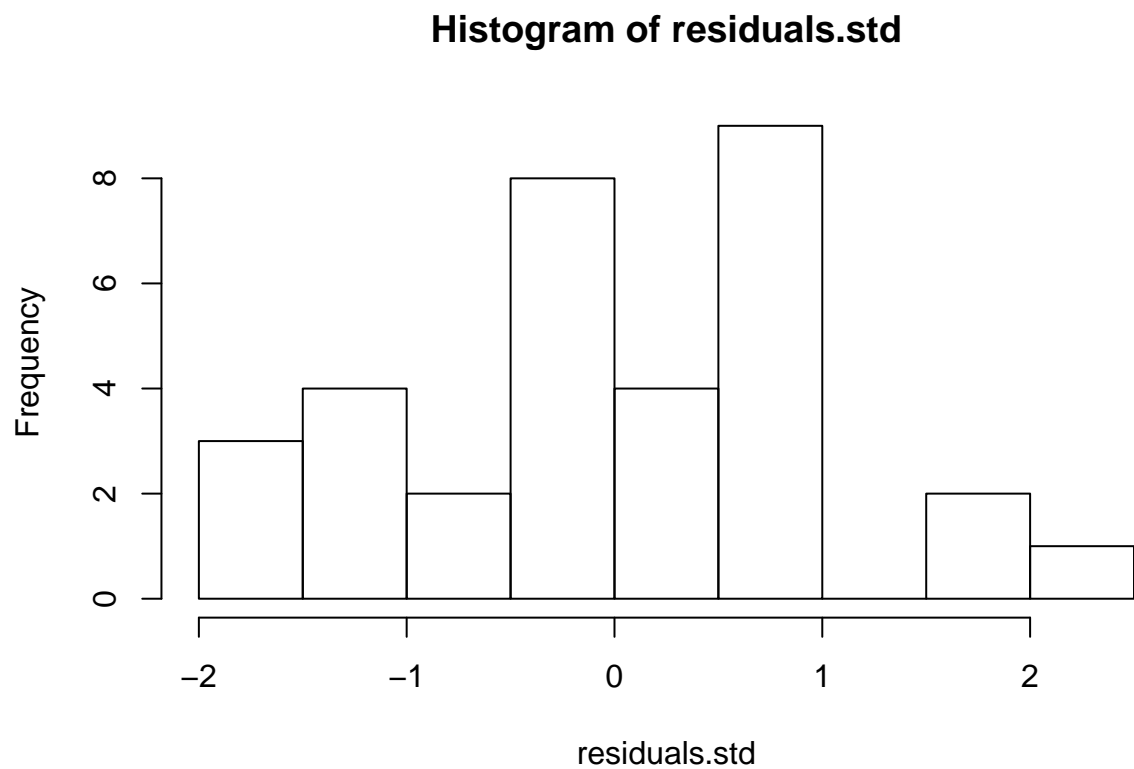
```
# Obtain the residuals from the ANOVA fit  
residuals=anova.fit$residuals;  
hist(residuals)
```



```
# Semistudentized residuals  
residuals.semistd=anova.fit$residuals/sqrt(mse);  
hist(residuals.semistd)
```

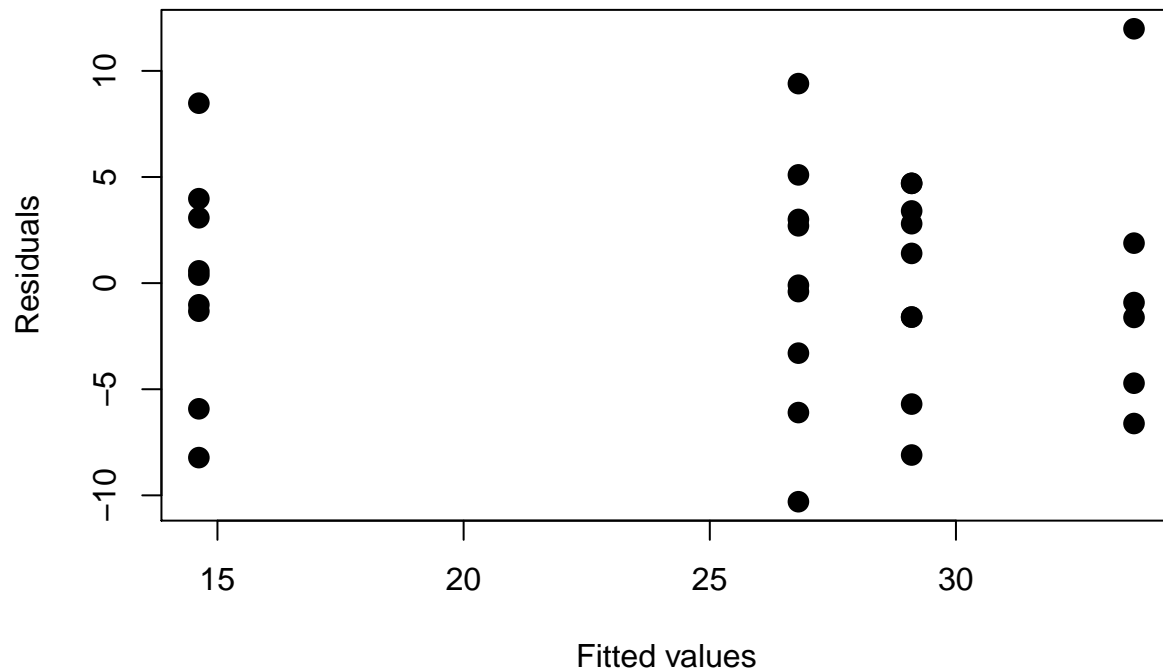


```
# Studentized residuals
weights=1-1/ns[as.numeric(Spock$Judge)];
residuals.std=anova.fit$residuals/sqrt(mse)/sqrt(weights);
hist(residuals.std)
```



Plot the residuals (or the other two versions) against fitted values

```
plot(residuals~anova.fit$fitted.values,type='p',pch=16,cex=1.5,xlab="Fitted values",ylab="Residuals")
```



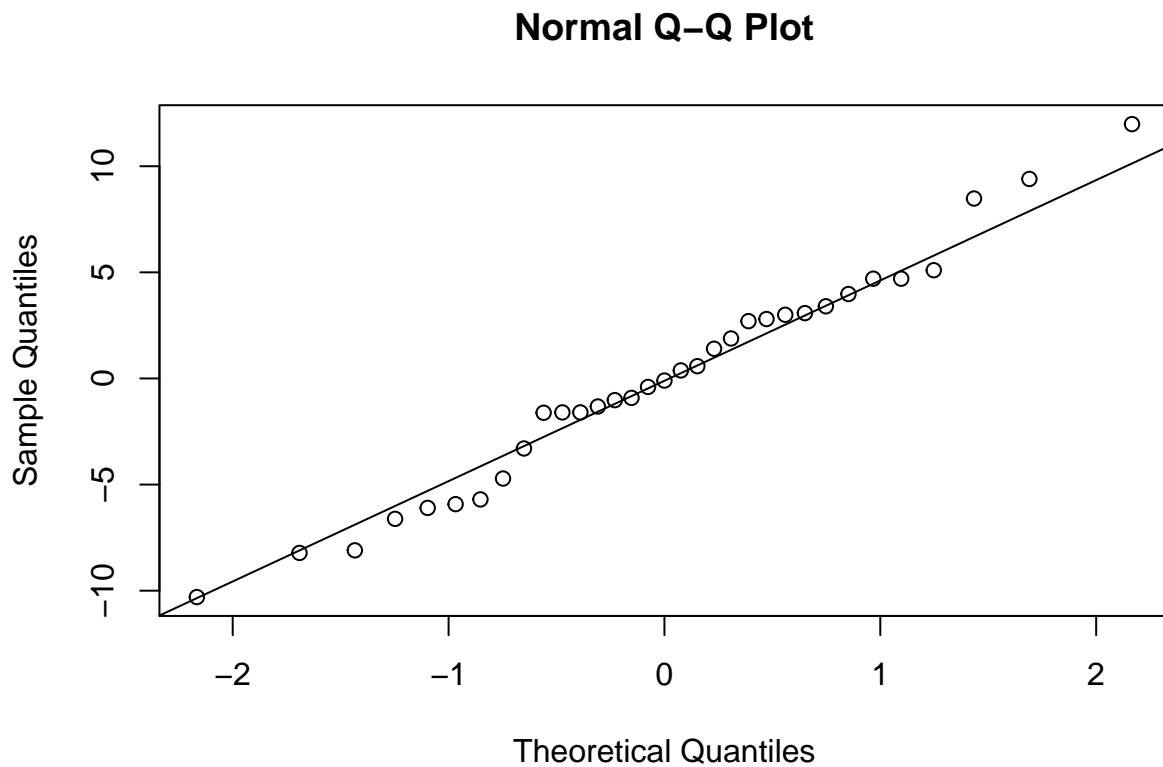
```
# Plot the residual against certain orders
# No clear orders make sense in the Spock trial data
```

```
# Stem-leaf plot (or use histogram, or qq-plot )
stem(residuals)
```

```
##
## The decimal point is at the |
##
## -10 | 3
## -8 | 21
## -6 | 61
## -4 | 977
## -2 | 3
## -0 | 66630941
## 0 | 4649
## 2 | 78014
## 4 | 0771
## 6 |
## 8 | 54
## 10 |
## 12 | 0
```



```
qqnorm(residuals);qqline(residuals)
```



```
# Plot residuals against missing variables  
# Not applicable on Spock trial data
```

We now turn to formal tests of the equality of variances.

```
# Calculate the variances for each group:  
(vars = tapply(Spock$perc.women, Spock$Judge, var))
```

```
##  1  2  3  4  
## 25 43 21 36
```

```
alpha=0.05;
```

```
# Hartley test:  
H.stat=max(vars)/min(vars);  
library(SuppDists) # The distribution is in this package  
# Both df and k only take integers:  
qmaxFratio(1-alpha, df=floor(sum(ns)/length(ns)-1), k=length(ns))
```

```
## [1] 8.4
```

```
qmaxFratio(1-alpha,df=ceiling(sum(ns)/length(ns)-1),k=length(ns))
```

```
## [1] 7.2
```

```
# Bartlett test:
```

```
K.stat= (sum(ns)-length(ns))*log(mse)-sum( (ns-1)*log(vars) );
qchisq(1-alpha,df=length(ns)-1)
```

```
## [1] 7.8
```

```
# Levene test:
```

```
Spock$res.abs=abs(anova.fit$residuals);
summary(aov(res.abs~Judge,data=Spock))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Judge      3    5.6    1.88    0.17  0.91
## Residuals 29  314.7   10.85
```

We leave weighted least squares for exercise. You can either calculate it following the steps discussed in lecture, or use the `weights` option in `lm()` and `aov()`.

We can conduct the nonparametric tests as follows.

```
# The rank test
```

```
Spock$rank.perc=rank(Spock$perc.women)
summary(aov(rank.perc~Judge,data=Spock))
```

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Judge      3   1846     615    15.6 3.1e-06 ***
## Residuals 29   1144      39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Kruskal-Wallis test:
```

```
kruskal.test(perc.women~Judge,data=Spock)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  perc.women by Judge
## Kruskal-Wallis chi-squared = 20, df = 3, p-value = 2e-04
```

For Box-Cox transformation, use the `boxcox` in library `MASS`.

2.4 Learning Objectives

- Students are able to write down a one-way ANOVA model given a new dataset.
- Students understand the basic properties of one-way ANOVA models.

- Students recognize the assumptions associated with each method.
- Students can implement the aforementioned tasks in R.
- Students are comfortable reading R helpfiles related to one-way ANOVA.

Chapter 3

Two-way ANOVA

3.1 Experiments with two (or more) factors

- Randomized experiments with two treatments
- Stratified randomized Experiments (also known as randomized block design)
 - Auditor training data
 - Project STAR
- Reasons for stratification: practical and statistical
- Sampling scheme for a stratified randomized experiment
- Question of interest, null hypotheses, and their causal interpretation.
- Intuition of hypothesis testing.

Description of the auditor training data: There are three training methods for the auditors and the response Y is a proficiency score after the training are completed. Ideally we would like to compare the three methods among those who are as similar as possible in their educational background. How do we achieve this? One way is compare the three different training methods among those whose time since graduation from college are about the same. Suppose then we have ten such groups (of three individuals each). Group 1 consists of those who graduated recently, group 2 people graduated between one and two years ago, and group 10 consists of those who graduated some time in the past (say, ten years or more). Time since graduation is called the block (or a blocking factor) and treatment is the training method.

3.2 Two-way ANOVA

3.2.1 A motivating example: Hey fever relief data set

For the Hay Fever Relief example, 9 compounds for Hay Fever Relief are made by varying levels of the two basic ingredients. Ingredient 1 (factor A) has $a = 3$ levels: low ($i = 1$), medium ($i = 2$) and high ($i = 3$). Similarly, ingredient 2 (factor B) has $b = 3$ levels: low ($j = 1$), medium ($j = 2$) and high ($j = 3$). A total of 36 subjects (suffering from hay fever) are selected and each of the 9 compounds are given to randomly selected $n = 4$ individuals.

3.2.2 A two-way ANOVA model

- Cell mean model
- Decomposition of the means, and their estimators
- Additive models
 - Why additive models?
 - Estimators of the means
- Decomposition of sum of squares, and their properties

3.2.3 Statistical inference

- F-statistics based on sums of squares
- Hypothesis testing
 - Test for interaction effects
 - Test for main effects
 - Alternative test if interaction can be ignored (additive models)
- (Simultaneous) confidence intervals with and without interactions
 - Bonferroni
 - Tukey
 - Scheffe

3.2.4 Model diagnostics

- Similar to those for one-way ANOVA

3.2.5 Strategy for data analysis

Using the Hey Fever data as an example.

```
Hay <- read.csv(file="./data/HayFever.csv", header=TRUE, sep=",")
```

```
# Use a slightly different visualization:
pairs(Hay, pch=16, col='red', cex=1.5)
```

```
# Or draw the interaction plot:
interaction.plot(Hay$Ingredient.1, Hay$Ingredient.2, Hay$Relief)
```

Step 1. Test whether interaction effects are presented.

```
# We can use the regression form here
full_model=lm(Relief~as.factor(Ingredient.1)+as.factor(Ingredient.2)+as.factor(Ingredient.3), data=Hay)
reduced_model=lm(Relief~as.factor(Ingredient.1)+as.factor(Ingredient.2), data=Hay);
anova(reduced_model, full_model)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: Relief ~ as.factor(Ingredient.1) + as.factor(Ingredient.2)
```

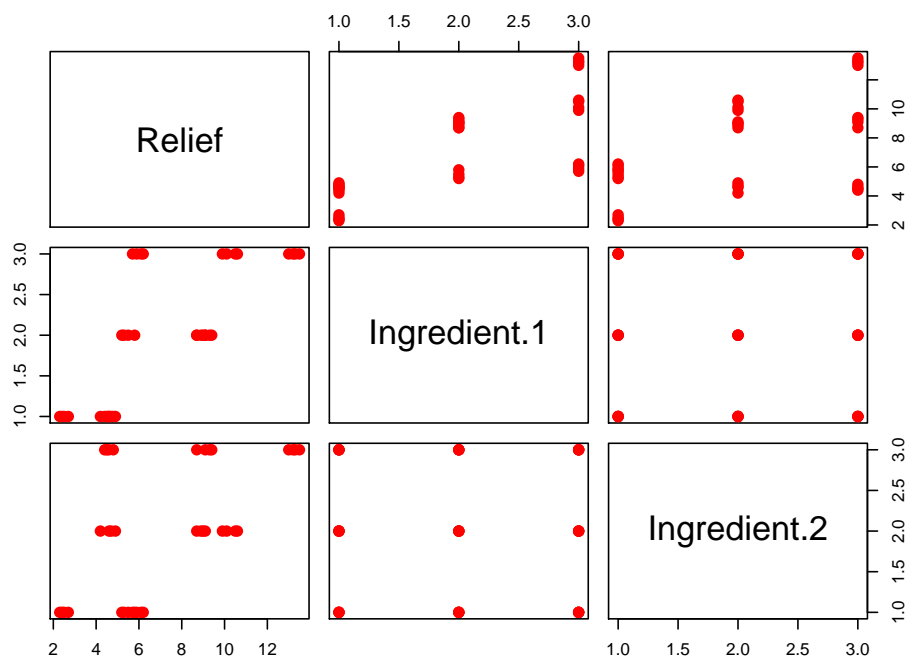


Figure 3.1: Box plot with jittered data points for the Spock trial data.

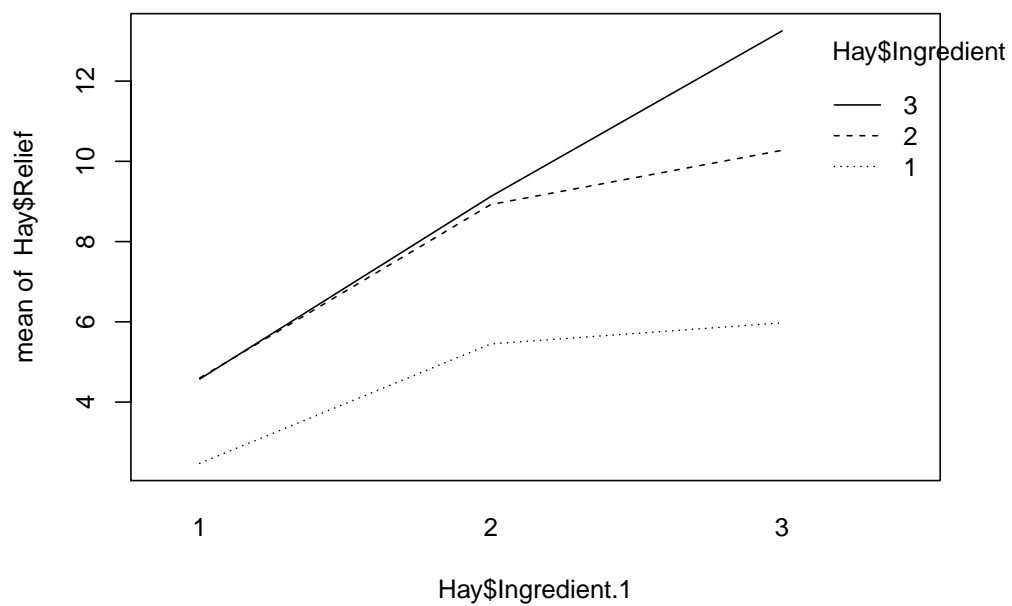


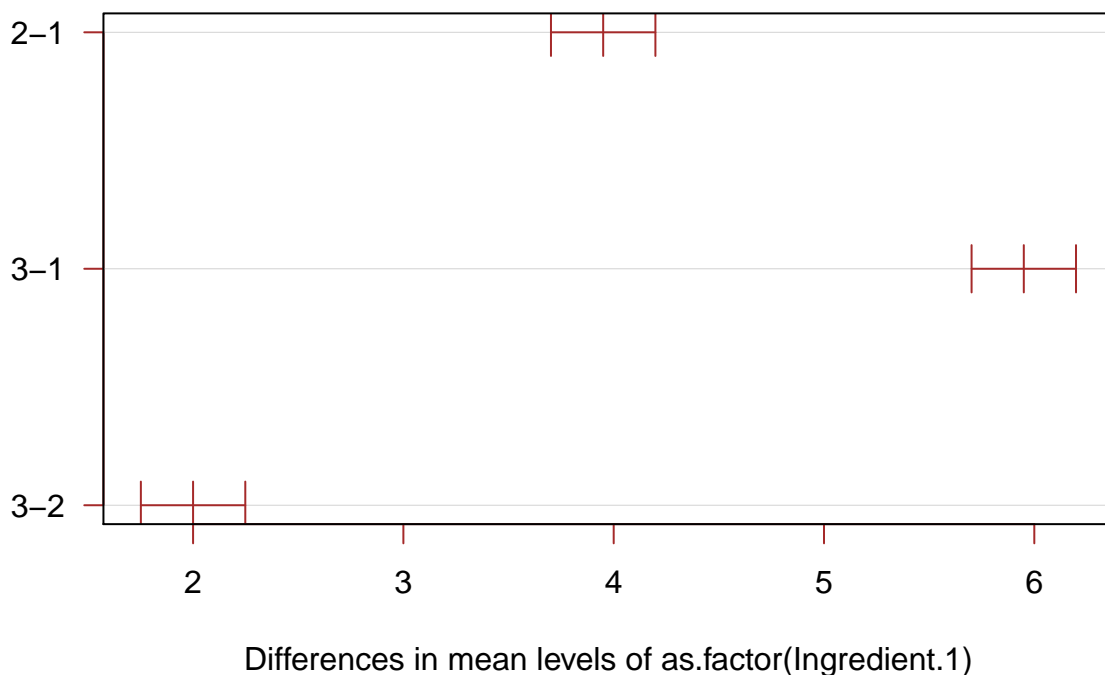
Figure 3.2: Box plot with jittered data points for the Spock trial data.

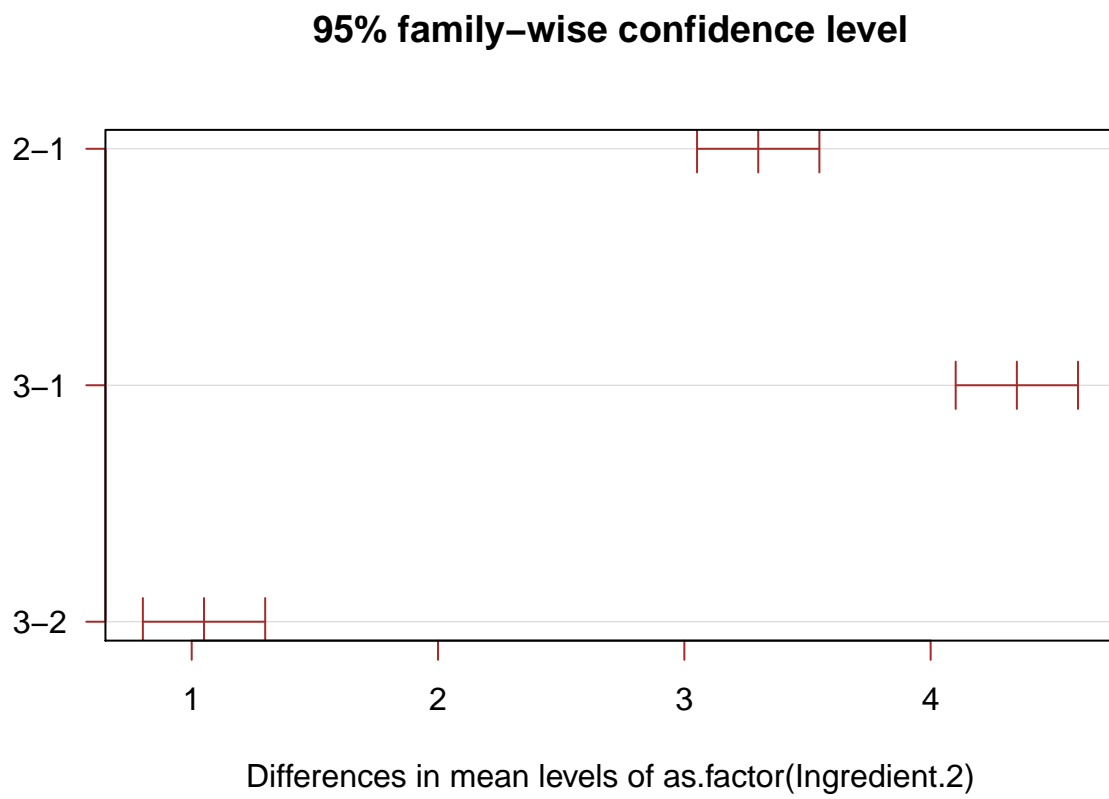
```
## Model 2: Relief ~ as.factor(Ingredient.1) + as.factor(Ingredient.2) +
##           as.factor(Ingredient.1) * as.factor(Ingredient.2)
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      31 31.05
## 2      27  1.63  4      29.4 122 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test result show that interaction effects are very likely to be absent from this data set. This means that we need to treat each combination as a unit, whereas we can compare each type of main effects separately. In the Hay Fever data, we naturally want to find the combination of ingredients that is most effective. We can use the Tukey-Kramer method for this task.

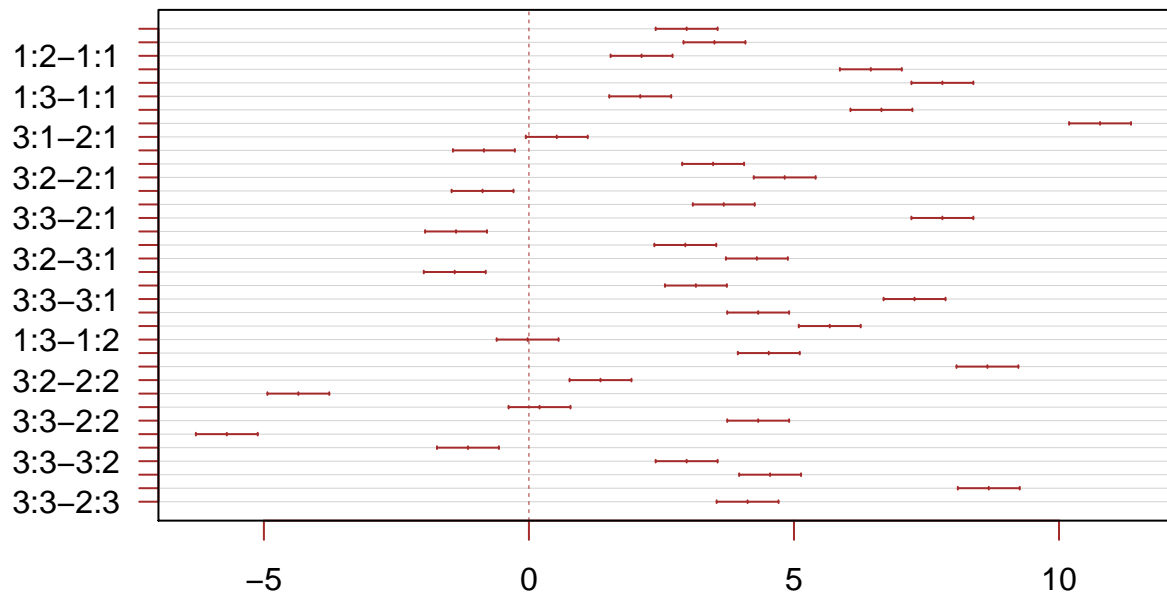
```
library(stats)
alpha=0.05;
anova.fit<-aov(Relief~as.factor(Ingredient.1)+as.factor(Ingredient.2)+as.factor(Ingredient.1):as.factor(Ingredient.2))
T.ci=TukeyHSD(anova.fit,conf.level = 1-alpha)
plot(T.ci, las=1 , col="brown")
```

95% family-wise confidence level





95% family-wise confidence level

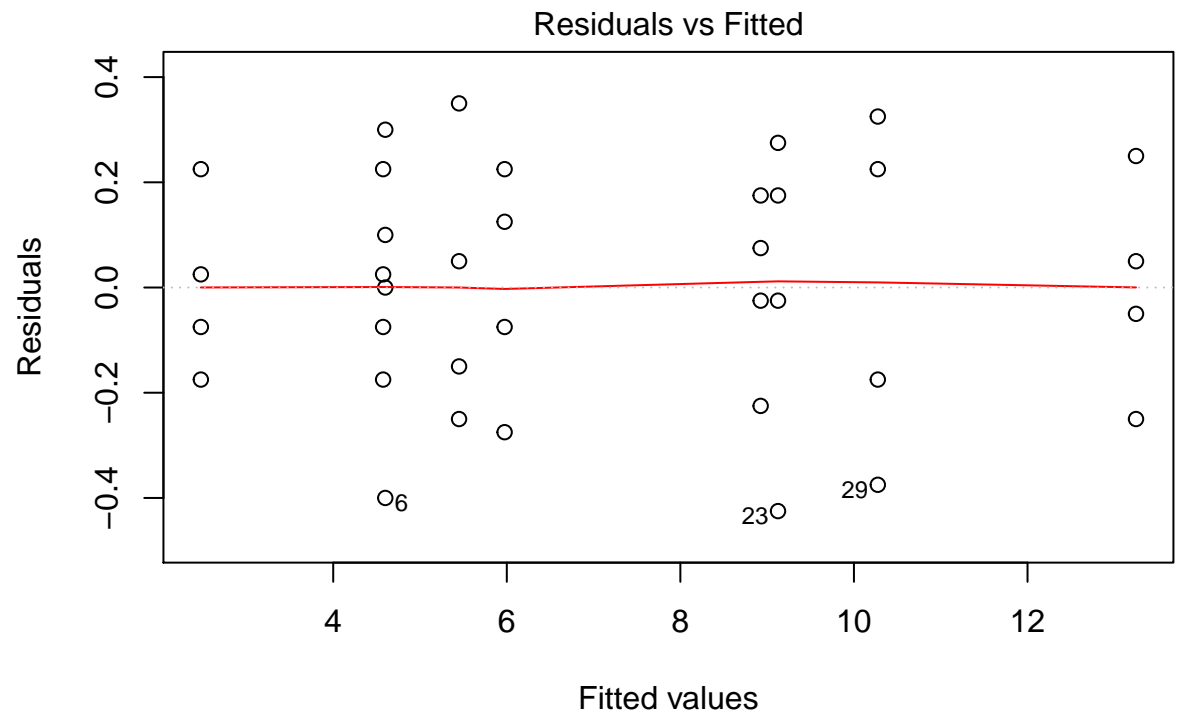


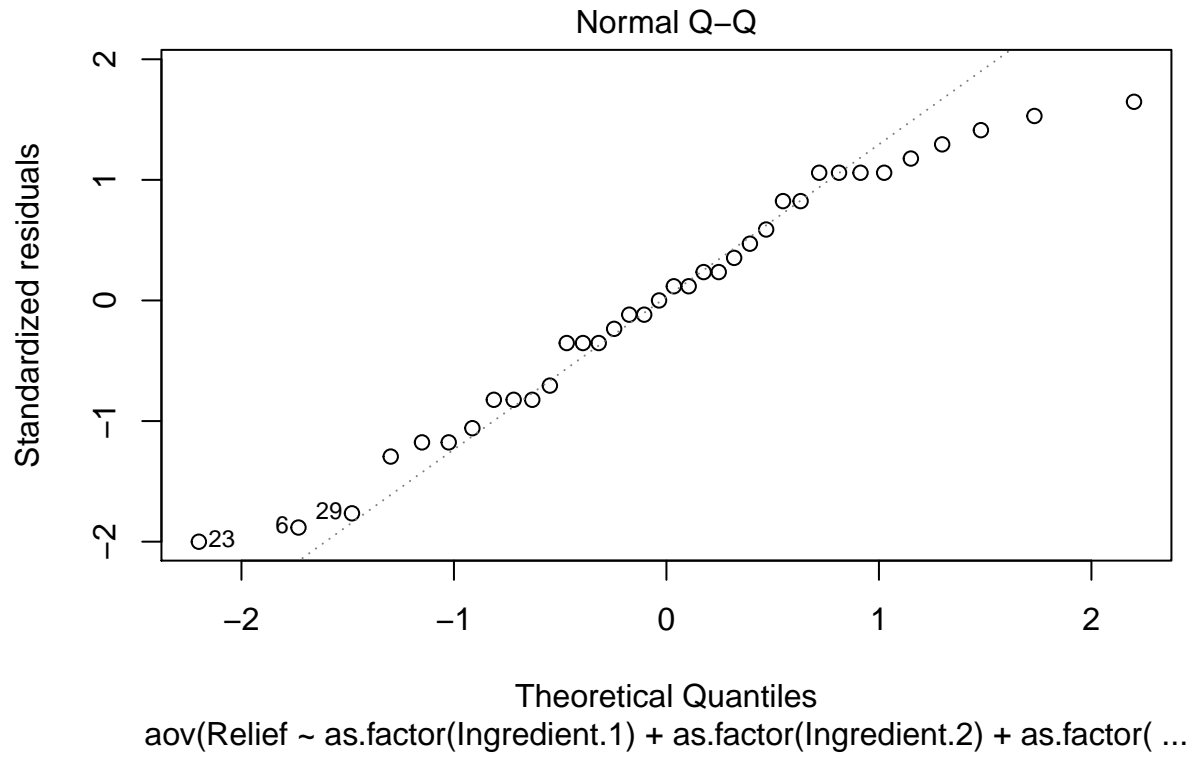
Differences in mean levels of as.factor(Ingredient.1):as.factor(Ingredient.2)

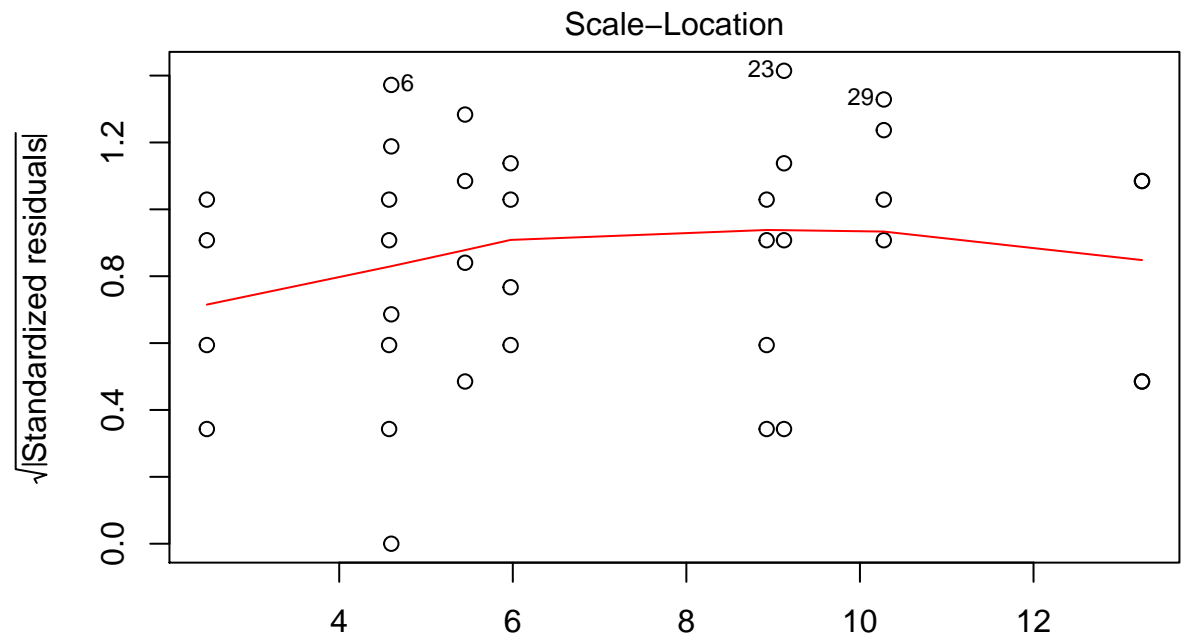
```
# We only need to pay attention to the differences of the two largest means
idx=list();
idx[[1]]=Hay$Ingredient.1;idx[[2]]=Hay$Ingredient.2;
(means.comb=tapply( Hay$Relief, INDEX=idx,mean))
```

```
##      1      2      3
## 1 2.5   4.6   4.6
## 2 5.4   8.9   9.1
## 3 6.0  10.3  13.2
```

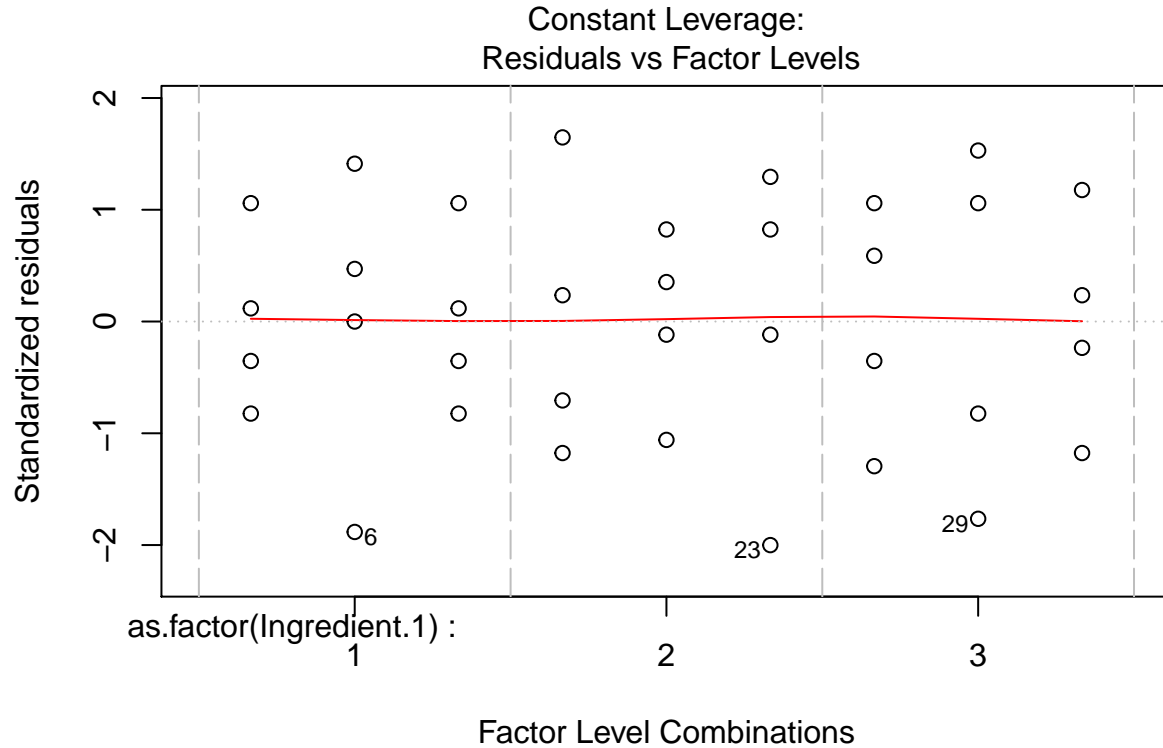
```
# For model diagnostics, we can use the default plotting function of aov()
plot(anova.fit)
```







Fitted values
`aov(Relief ~ as.factor(Ingredient.1) + as.factor(Ingredient.2) + as.factor(...`



3.2.6 Special case: one observation per cell

- Interaction effects are no longer identifiable
- Estimation and testing
- Tukey's test of additivity

3.2.7 Unbalanced two-way ANOVA

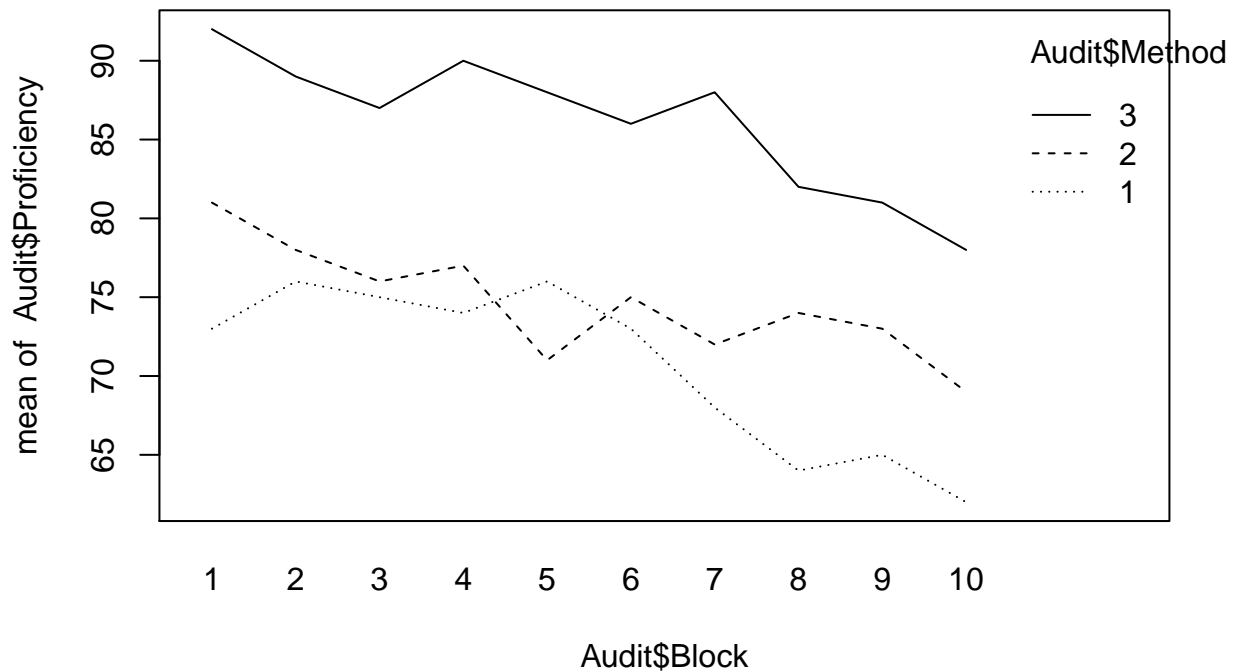
- The model is the same as for the balanced case
- Estimators for the means and variances
- Hypothesis testing
 - Interaction effects
 - Main effects
- Missing data in the one observation per cell case

3.2.8 Analysis of stratified experiments

- Why stratification?

We can check whether stratification is efficient on the auditor training data.

```
Audit <- read.csv(file="./data/AuditorTraining.csv", header=TRUE, sep=",")
# Or draw the interaction plot:
interaction.plot(Audit$Block, Audit$Method, Audit$Proficiency)
```



```
# Fit the model with blocks:
anova.block<-aov(Proficiency~as.factor(Block)+as.factor(Method),data=Audit)
anova.random<-aov(Proficiency~as.factor(Method),data=Audit)

mse.block<-sum(anova.block$residuals^2)/anova.block$df.residual
mse.random<-sum(anova.random$residuals^2)/anova.random$df.residual
(E=mse.random/mse.block)
```

```
## [1] 3.2
```

3.3 Learning Objectives

- Students are able to write down an appropriate two-way ANOVA model given a new dataset.
- Students understand the basic properties of the various types of two-way ANOVA models.

- Students recognize the assumptions associated with each method, and can find appropriate tests to verify the assumptions.
- Students can implement the aforementioned tasks in R.
- Students can seek help in coding using the Internet.