Linear Models and Experimental Design

Christine Müller

Literatur

[1] Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Springer, Berlin.
[2] Christensen, R. (2001). *Advanced Linear Modeling: Multivariate, Time Series and Spatial Data - Nonparametric Regression and Response Surface Maximization*. Springer, Berlin.
[3] Christensen, R. (2002). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer, Berlin.
[4] Dalgaard, P. (2002). *Introductory Statistics with R*. Springer, New York.
[5] Fahrmeier, L., Künstler, R., Pigeot, I., und Tutz, G. (1997). *Statistik*. Springer, Berlin.
[6] Faraway, J.J. (2006). *Extending the Linear Model with R*. Chapman & Hall/CRC.
[7] Federer, W.T. (1993). *Statistical Design and Analysis for Intercropping Experiments. Volume I Two Crops*. Springer, Berlin.
[8] Federer, W.T. (1999). *Statistical Design and Analysis for Intercropping Experiments. Volume II Three or More Crops*. Springer, Berlin.
[9] Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J., and Ostrowski, E. (1996). *Handbook of Small Data Sets*. Chapman & Hall, London.
[10] Gomez, K.A., und Gomez, A.A. (1984). *Statistical Procedures for Agricultural Research*. Wiley, New York.
[11] Hartung, J., Elpelt, B., und Klösener, H.P. (1998). *Statistik*. Oldenbourg, München.
[12] Köhler, W., Schachtel, G., Voleske, P. (2002). *Biostatistik*. Springer, Berlin.
[13] Mukerjee, R., und Wu, C.F.J. (2006). *A Modern Theory of Factorial Design*. Springer, Berlin.
[14] Quinn, G.P. und Keough, M.J. (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge University Press.
[15] Rasch, D., Verdooren, L. R. und Gowers, J. (2007). *Planung und Auswertung von Versuchen und Erhebungen*. Oldenbourg, München.
[16] Rencher, A.C. (1995). *Methods of Multivariate Analysis*. Wiley, New York.
[17] Rencher, A.C. (1998). *Multivariate Statistical Inference and Applications*. Wiley, New York.
[18] Ruxton, G.D. und Colegrave, N. (2006). *Experimental Design for the Life Sciences*. Oxford University Press.
[19] Toutenburg, H. (2002). *Lineare Modelle. Theorie und Anwendungen*. Physica-Verlag, Berlin.

Contents

Part I

Methods

1 Preliminaries

1.1 How to download the free software R

In the following it is described how the basis package of R for windows computers (e.g. Windows NT or Windows 98, 2000, XP) can be downloaded. The process is similar for other computers. In particular the R versions for other operating systems can be found under the same internet address. The internet addess is **http://cran.r-project.org/**. For windows computers, you use at first the button `Windows (95 and later)` and then `base`. Then save `R-2.5.1-win32.exe` (or a newer version) in a directory of your computer which shall contain the R program code. For installing, activate `R-2.5.1-win32.exe`. Then R will be installed in subdirectories of the chosen directory. Afterwards you can start R. For working with R, it is of great advantage to link the R with a own working directory. Otherwise all files produced by R are saved in the program directory. It is always good to have program directories and working directories clearly separated. In the directory `\rw2001\doc\manual` (`rw2001` may be substitued by a newer version) you will find the file `R-intro.pdf` which includes a detailed introduction to R in English.

Please note that currently new versions of R appear. Hence differences in the output can be due to different versions.

1.2 Installing and activating the R Package `agricolae`

The R package `agricolae` contains special R functions for agricultural statistics and some agricultural data sets. You will find on the website **http://cran.r-project.org/** on the lefthand site the button `Packages`. This buttom provides a list with over 700 Packages ordered alphabetically. There you find the package `agricolae`. On its site you will find a ZIP file and a PDF file for downloading. The best is to download the ZIP-file in the working directory of your R. When you have started R, then you choose the button `Packages` (`Pakete` in German version) and there the button for installing the package from a local ZIP file (in German: `Installiere Paket(e) aus lokalen ZIP-Dateien`). Activating the agricolae ZIP file installs this package on your computer.

If you have a older version of R, then do not worry about warning messages. They concern missing libraries which are only needed for special routines of the `agricolae` package, we do not need. However, sometimes the R package `combinat` is needed which shall be downloaded like the package `agricolae`. Also other packages which are used here only in very special situations can be downloaded like the package `agricolae`.

To activate the library `agricolae` for your R session, you need for every new session to type

```
> library(agricolae)
```

in the R command window. To test whether the package is available, type for example

```
> ?design.lsd
```

Then in a new window you will find the description of the R routine for generating Latin Squares Designs. The activating of other packages is the same.

Note, that always the help function is activated with `?` and provides the description of the R functions.

1.3 Transferring data into R

If you want to use the data sets from a package, then you must load the data set with the R function `data`. For example the data set `trees` from the agricolae package:

1.3.1 Data Set (TREES)

```
> library(agricolae)
> data(trees)
```

Then the data set `trees` is avalaible. To see how it looks like, type:

```
> trees
```

Then you get the whole data set:

```
   place    species diameter
1      1     LAUREL     18.4
2      2     LAUREL     19.1
3      3     LAUREL       NA
4      4     LAUREL     14.4
5      5     LAUREL     12.9
6      6     LAUREL     14.4
7      1      GUABA     14.3
8      2      GUABA     12.9
9      3      GUABA     15.0
10     4      GUABA     14.6
11     5      GUABA     14.6
12     6      GUABA     12.4
13     1      ROBLE     21.0
14     2      ROBLE     19.7
15     3      ROBLE     13.2
16     4      ROBLE     13.2
17     5      ROBLE     16.8
18     6      ROBLE     14.0
19     1 TERMINALIA     20.9
20     2 TERMINALIA     18.2
21     3 TERMINALIA     19.2
22     4 TERMINALIA     21.7
23     5 TERMINALIA     15.7
24     6 TERMINALIA     18.6
```

It has three variables, one which describes the place, where the tree was measured, one for the tree species and one for the tree height.

To get a data set from a R package is the most convenient way. But usually the data are given in another form. You can transfer many data data formats into R as those from SAS or SPSS. But the simplest format is the ASCII format. This is only explained here.

1.3.2 Data Set (DARWIN)

```
> read.table("DARWIN2.DAT",header=T)
   Pair Cross.fertilized Self.fertilized
1     1             23.5            17.4
2     2             12.0            20.4
3     3             21.0            20.0
4     4             22.0            20.0
5     5             19.1            18.4
6     6             21.5            18.6
7     7             22.1            18.6
8     8             20.4            15.3
9     9             18.3            16.5
10   10             21.6            18.0
11   11             23.3            16.3
12   12             21.0            18.0
13   13             22.1            12.8
14   14             23.0            15.5
15   15             12.0            18.0
```

The argument `header=T` (`T=TRUE`) provides that the first line of the data file is read as header line. Note that Cross-fertilized and Self-fertilized in the data file is converted to `Cross.fertilized` and `Self.fertilized`, since the hyphen is not a allowed character in R.

Explanation of the data set: "These data are from Charles Darwin's study of cross- and self-fertilization. Pairs of seedlings of the same age, one produced by cross-fertilization and the other by self-fertilization, were grown together so that members of each pair were reared under nearly identical conditions. The aim was to demonstrate the greater vigour of the cross-fertilized plants. The data are the final heights of each plant after a fixed period of time. Darwin consulted Galton about the analysis of these data, and they were discussed further in Fisher's *Design of Experiments*."(Hand et al. 1996, P. 2)

1.4 Transforming data sets

Data are often not in the form which is needed for the analysis in R. Hence after reading the data, the data must be transformed in a appropriate form. The form of a data set which can be easily analyzed in R is always a table with several rows and columns as follows:

Data tables

Every **row** belongs to an **experimental unit** (individual, case, field unit).
Every **column** belongs to a **variable** observed, measured, or registered at the experimental units. Variable are usually measurements, treatments, and blocking numbers.

It is sometimes not easy to decide, what the experimental unit is. In the data set 1.3.2 as presented above, the experimental units are the pairs of seedlings. Hence we have three variables: The number of the pairs, the measurement for cross-fertilization, the measurement for self-fertilization. But we will see later that it is sometimes more convenient to regard each seedling as experimental unit. Then we have again three variables: the pair number, the fertilization type, and the final height.

```
> darwin
   Pair Height Fertilization
1     1   23.5         Cross
2     2   12.0         Cross
3     3   21.0         Cross
4     4   22.0         Cross
5     5   19.1         Cross
6     6   21.5         Cross
7     7   22.1         Cross
8     8   20.4         Cross
9     9   18.3         Cross
10   10   21.6         Cross
11   11   23.3         Cross
12   12   21.0         Cross
13   13   22.1         Cross
14   14   23.0         Cross
15   15   12.0         Cross
16    1   17.4          Self
17    2   20.4          Self
18    3   20.0          Self
19    4   20.0          Self
20    5   18.4          Self
21    6   18.6          Self
22    7   18.6          Self
23    8   15.3          Self
24    9   16.5          Self
25   10   18.0          Self
26   11   16.3          Self
27   12   18.0          Self
28   13   12.8          Self
29   14   15.5          Self
30   15   18.0          Self
```

To achieve this form of the data set, type:

```
> darwin0<-read.table("DARWIN2.DAT",header=T)
```

```
> darwinC<-cbind(darwin0[,c(1,2)],"Cross")
> darwinS<-cbind(darwin0[,c(1,3)],"Self")
> names(darwinC)<-c("Pair","Height","Fertilization")
> names(darwinS)<-c("Pair","Height","Fertilization")
> darwin<-rbind(darwinC,darwinS)
> row.names(darwin)<-1:30
```

The first two columns from the data set `darwin` are selected with `darwin[,c(1,2)]`. With `cbind(darwin[,c(1,2)],"Cross")`, a third column which contains everywhere as entry `"Cross"` is added to the two columns so that we then have three columns. With `names(darwinC)<-c("Pair","Height","Fertilization")`, the three columns get the names `"Pair","Height","Fertilization"`. The same is done after selecting the first and third column with `darwin[,c(1,3)]`. Then `darwinC` and `darwinS` are two data tables with three columns and 15 rows. These two data tables are put together with `rbind(darwinC,darwinS)`. If we do not use `row.names(darwin2)<-1:30`, then the row names are strange (you can check this by looking at `darwin2` before using this command).

The variables of a data table are vectors. There are several possibilities in R to generate and combine vectors:

Generation of vectors

Combination with `c`: E.g.: `c(1,3,4,2)` for numbers or `c("self","cross","cross")` for character strings or `c(T,T,F)` for logical values.
Simple sequence: E.g. `1:10` provides the sequence 1,2,3,4,5,6,7,8,9,10.
General sequence with `seq`: E.g. `seq(from=1,to=16,by=3)` provides 1,4,7,10,13,16.
Repetition with `rep`: E.g. `rep(2,10)` provides 2,2,2,2,2,2,2,2,2,2.

Combining vectors and data tables

Combination columnwise with `cbind` and rowwise with `rbind`.
The columns and rows, respectively, must have the same length. However, a vector or data set can be combined with a single value, because then the single value is automatically repeated adequately.
`data.frame` can be used instead of `cbind` if the result should be a data table with different types of columns.

There are also several possibilities to select from a vector or from a data set:

Selection from a vector `x`, eg. $x = (10, 13, 21, 45, 62)$

By components: E.g. `x[2,5,1]` provides 13,62,10
By naming components which should be not used by negative numbers: E.g. `x[c(-1,-4)]` provides 13,21,45.
By names of the components if available: E.g. `x[c("two","five","one")]` provides 13,62,10 if the names coincides with the component numbers.
By logical values: E.g. `x[c(T,T,F,T,F)]` provides 10,13,45.

Selection from a data table or matrix, e.g. `darwin`

Selection of rows by specifying the first components: E.g. `darwin[c(2,5,1),]`
Selection of columns by specifying the second components: E.g. `darwin[,c(3,1)]`
The selection can be done also by negative components, names, and logical values as for vectors.
Additional selection of columns by names for data tables by $: E.g. `darwin$Height` provides the same as `darwin[,"Height"]`.

Producing logical values

`a<b` less than,
`a<=b` less than or equal,
`a==b` equal (2 equality signs!),
`a!=b` not equal,
`a>b` greater than,
`a>=b` greater than or equal.
Logical values are combinded by
& and
`|` or
according to the rules of logics.
`!` is the negation.

For example, if all seedlings from the self-fertilization with height less than 18cm should be selected, then we type:

```
> darwin[darwin$Height<18 & darwin2[,"Fertilization"]=="Self",]
   Pair Height Fertilization
16    1   17.4          Self
23    8   15.3          Self
24    9   16.5          Self
26   11   16.3          Self
28   13   12.8          Self
29   14   15.5          Self
```

1.4.1 Exercise (CHICKEN)
The data file `CHICKENS.DAT` contains the data of a randomized blocks experiment. This experiment "was carried out to investigate a drug added to the feed of chickens in an attempt to promote growth. The comparison is between three treatments: standard feed (control), standard feed plus low dose of drug, standard feed plus high dose of drug. The experimental unit is a group of chicks, reared and fed together in the birdhouse. The experimental units are grouped three to a block, with physically adjacent units going to the same block. The response is the average weight per bird at maturity for the group of birds in each experiment." (Hand et al. 1996, P. 7/8)
The four columns of the data file `CHICKENS.DAT` have the following titles: Block, Control, Low dose, High dose. The average weights of the birds is given in pounds.
Read the data and transform the data to the following form:

```
> chicken
   Block Weight    Feed
1      1   3.93 Control
2      2   3.78 Control
3      3   3.88 Control
4      4   3.93 Control
5      5   3.84 Control
6      6   3.75 Control
7      7   3.98 Control
8      8   3.84 Control
9      1   3.99     Low
10     2   3.96     Low
11     3   3.96     Low
12     4   4.03     Low
13     5   4.10     Low
14     6   4.02     Low
15     7   4.06     Low
16     8   3.92     Low
17     1   3.96    High
18     2   3.94    High
19     3   4.02    High
20     4   4.06    High
21     5   3.94    High
22     6   4.09    High
23     7   4.17    High
24     8   4.12    High
```

1.4.2 Data Set (MUSTARD=Ackersenf)

The data in `MUSTARD.DAT` "come from an experiment to investigate the effect of light on root growth in mustard seedlings. Two groups of seedlings were grown in identical conditions, except that one was kept in the dark while the other had daylight during the day. After germination the stems were cut off some of the seedlings, to allow for the possibility that light affected the vigor of the whole plant through the stem and leaves. Later the root lengths (in mm) of all seedlings were measured. Does light affect root growth; and does this effect depend on whether the stem is cut?" (Hand et al. 1996, P. 74,75).

```
21   27   22   21
39   21   16   39
31   26   20   20
13   12   14   24
52   11   32   20
39    8   28
55        36
50        41
29        17
17        22
```

The first two columns concern the root length grown in the light, the first and third column concern the root length where stems are cut. This data set cannot be readed by `read.table` because `read.table` is expecting equal length of rows. To achieve equal length of rows, we can introduced missing values:

```
21   27   22   21
39   21   16   39
31   26   20   20
13   12   14   24
52   11   32   20
39    8   28    -
55    -   36    -
50    -   41    -
29    -   17    -
17    -   22    -
```

The new data file is called `MUSTARD2.DAT` and can be read as follows where the argument `na.strings="-"` tells R that the missing values are denoted by "-":

```
> mustard0<-read.table("MUSTARD2.DAT",na.strings="-")
> mustard0
   V1 V2 V3 V4
1  21 27 22 21
2  39 21 16 39
3  31 26 20 20
4  13 12 14 24
5  52 11 32 20
6  39  8 28 NA
7  55 NA 36 NA
8  50 NA 41 NA
9  29 NA 17 NA
10 17 NA 22 NA
```

Since the experimental unit is the root and we have 31 roots measures, we have 31 experimental units. We obtain the correct data table with the following commands:

```
> mustard1<-data.frame(mustard0[!is.na(mustard0[,1]),1],"light","cut")
> mustard2<-data.frame(mustard0[!is.na(mustard0[,2]),2],"light","noncut")
> mustard3<-data.frame(mustard0[!is.na(mustard0[,3]),3],"dark","cut")
> mustard4<-data.frame(mustard0[!is.na(mustard0[,4]),4],"dark","noncut")
> names(mustard1)<-c("length","grow.conditions","cutting")
> names(mustard2)<-c("length","grow.conditions","cutting")
> names(mustard3)<-c("length","grow.conditions","cutting")
> names(mustard4)<-c("length","grow.conditions","cutting")
> mustard<-rbind(mustard1,mustard2,mustard3,mustard4)
> row.names(mustard)<-1:length(mustard$length)
> mustard$length<-as.numeric(mustard$length)
```

```
> mustard
   length grow.conditions cutting
1      21           light     cut
2      39           light     cut
3      31           light     cut
4      13           light     cut
5      52           light     cut
6      39           light     cut
7      55           light     cut
8      50           light     cut
9      29           light     cut
10     17           light     cut
11     27           light  noncut
12     21           light  noncut
13     26           light  noncut
14     12           light  noncut
15     11           light  noncut
16      8           light  noncut
17     22            dark     cut
18     16            dark     cut
19     20            dark     cut
20     14            dark     cut
21     32            dark     cut
22     28            dark     cut
23     36            dark     cut
24     41            dark     cut
25     17            dark     cut
26     22            dark     cut
27     21            dark  noncut
28     39            dark  noncut
29     20            dark  noncut
30     24            dark  noncut
31     20            dark  noncut
```

The R function `is.na` provides the logical value `T` (=`TRUE`) for a missing value. With the negation given by `!`, it provides the logical value `F` (=`FALSE`) for the missing values so that they are not used. Without the R function `data.frame` we would get a 31×3 matrix of character strings. The command `data.frame(mustard0[...],...)` is a short hand of `cbind(data.frame(mustard0[...]),...)`. The R function `length` provides length of a vector.

1.5 Data types in R

The structure of a data set can be obtained by the command `str`. For the MUSTARD data set you can see very good with this command the differences between the different objects:

```
> str(cbind(mustard0[,1],"light","cut"))
 chr [1:10, 1:3] "21" "39" "31" "13" "52" "39" "55" "50" "29" "17" "light" ...
```

```
> str(cbind(data.frame(mustard0[,1]),"light","cut"))
'data.frame':   10 obs. of  3 variables:
 $ mustard0...1.: int  21 39 31 13 52 39 55 50 29 17
 $ "light"      : Factor w/ 1 level "light": 1 1 1 1 1 1 1 1 1 1
 $ "cut"        : Factor w/ 1 level "cut": 1 1 1 1 1 1 1 1 1 1
> str(data.frame(mustard0[,1],"light","cut"))
'data.frame':   10 obs. of  3 variables:
 $ mustard0...1.: int  21 39 31 13 52 39 55 50 29 17
 $ X.light.     : Factor w/ 1 level "light": 1 1 1 1 1 1 1 1 1 1
 $ X.cut.       : Factor w/ 1 level "cut": 1 1 1 1 1 1 1 1 1 1
> str(mustard)
'data.frame':   31 obs. of  3 variables:
 $ length         : num  21 39 31 13 52 39 55 50 29 17 ...
 $ grow.conditions: Factor w/ 2 levels "light","dark": 1 1 1 1 1 1 1 1 1 1 ...
 $ cutting        : Factor w/ 2 levels "cut","noncut": 1 1 1 1 1 1 1 1 1 1 ...
```

Look also at the data set TREES:

```
> str(trees)
'data.frame':   24 obs. of  3 variables:
 $ place   : int  1 2 3 4 5 6 1 2 3 4 ...
 $ species : Factor w/ 4 levels "GUABA","LAUREL",..: 2 2 2 2 2 2 1 1 1 1 ...
 $ diameter: num  18.4 19.1 NA 14.4 12.9 14.4 14.3 12.9 15 14.6 ...
```

This data set has three different types of variables: `place` has integer type, `species` is a factor, `diameter` is numeric. For the statistical analysis it is very important to distinguish between the different types.

Data types in R

integer: for integers as counts, sometimes a numeration
numeric: for the results of quantitative measurements
factor: for treatments, blocks

Very important for the experimental design is the type `factor`. The type `factor` is used for nominal values and has a finite number of levels. The levels are obtained by character sequences like `"GUABA"`. As soon as the values of a variable are given by character sequences, they are interpreted as factors. However, R uses internally integers for the levels where by default the integers are ordered in alphabetical order: 1 for `"GUABA"`, 2 for `"LAUREL"`, 3 for `"ROBLE"`, 4 for `"TERMINALIA"`. This can be also seen by converting the factor variable `species` into an integer variable. For that we create a new dummy variable `speciesI`:

```
> speciesI<-as.integer(trees$species)
> speciesI
 [1] 2 2 2 2 2 2 1 1 1 1 1 1 3 3 3 3 3 3 4 4 4 4 4 4
```

The symbol $ means that a special variable of the data set `trees` is used, namely here the variable `species`. The symbol `<-` means that the content of the dummy variable on the right hand side is assigned to the dummy variable at the left hand side. To see the difference between the two dummy variables `speciesI` and `trees$species` type:

```
> attributes(trees$species)
$levels
[1] "GUABA"      "LAUREL"     "ROBLE"      "TERMINALIA"

$class
[1] "factor"

> attributes(speciesI)
NULL
```

By the conversion to an integer variables, all factor attributes are lost. What can we do when the tree species are only given by the numbers 1,2,3,4? We always can convert a integer variable (even a numeric variable but this makes less sense) into a factor variable by the command `as.factor`:

```
> speciesF<-as.factor(speciesI)
> attributes(speciesF)
$levels
[1] "1" "2" "3" "4"

$class
[1] "factor"
```

If we want names for the levels, then type:

```
> levels(speciesF)<-c("GAUBA","LAUREL","ROBLE","TERMINALA")
> attributes(speciesF)
$levels
[1] "GAUBA"     "LAUREL"    "ROBLE"     "TERMINALA"

$class
[1] "factor"

> str(fspecies)
 Factor w/ 4 levels "GAUBA","LAUREL",..: 2 2 2 2 2 2 1 1 1 1 ...
> str(trees$species)
 Factor w/ 4 levels "GUABA","LAUREL",..: 2 2 2 2 2 2 1 1 1 1 ...
```

Hence we have created from the integer variable `speciesI` a factor variable of the same structure and content as the originally variable `trees$species`.

1.5.1 Exercise (SPLIT)
The data file `SPLIT.DAT` contains "a classic data set involving an experiment to investigate the

effect of manure (nitrogen) on the yield of barley (Gerste). Six blocks of three whole plots were used along with three varieties of barley, each whole plot being devoted to one variety only. The whole blocks were each divided into 4 subplots to cater four levels of manure (0,0.01,0.02,and 0.04 tons per acre)."(Hand et al. 1996, P. 253)
The 8 columns of the data file `SPLIT.DAT` have the following names: Block, Variety, Manure, Yield, Block, Variety, Manure, Yield.
Create a data table (data frame) with different rows for different experimental units and columns of correct data type. Use for the variable manure the numeric data type as well as the factor data type. Why make this sense?

1.5.2 Exercise (Pepper)
"An experiment was carried over a two-year period to find the best treatment for growing peppers in glashouses. Three factors were investigated, each at two levels:

Heating: standard (0) or supplementary (1)

Lighting: standard (0) or supplementary (1) Each treatment combination requires a glasshouse

Carbon dioxid: control (0) or added CO_2 (1)

compartment, and 12 compartments, divided into blocks of 6 are available. In the first year of the experiment, all 8 treatment combinations were used. In the second year, the 5 most successful treatments from the first year were retained, and one treatment was replicated in each block.

The response was a measure of the excess of yield over costs."

Heating	0	0	0	0	1	1	1	1
Lighting	0	0	1	1	0	0	1	1
CO2	0	1	0	1	0	1	0	1
Year1 Block 1	11.4	13.2	10.4	-	13.7	-	12.0	12.5
Year 1 Block 2	-	8.4	6.5	6.1	10.8	9.4	-	9.1
Year 2 Block 1	-	13.7	-	-	14.6	16.5	12.8	12.9
						15.4		
Year 2 Block 2	-	10.7	-	-	10.9	10.9	9.0	10.2
							10.1	

(Hand et al. 1996, P. 22,23)
The data are given in the data file `PEPPERS2.DAT`. What are the experimental units and the variables here and of which type are the variables? Prepare the data file so that it can be read with R and transform the data to a data table so that different rows belong to different experimental units and the columns are the variables. The creation of the data table is a more difficult task.

1.6 Saving data sets and self-defined functions

If we have manipulated the data sets it will be good to save the resulting data sets. This can be done with the R function `dump`.

```
>dump(c("chicken","chicken0","darwin","darwin0","mustard","mustard0"),
+ "all_data.asc")
```

Then the data sets are saved in the ASCII file `all_data.asc` in your working directory of R. The data sets can be reloaded in another R session by

```
> source("all_data.asc")
```

If the list of the data sets is too long it is better to write a self-defined function for saving which also includes this function. Functions in R have always the form

`function`(*arguments*) {*function body with R commands*}

They can be created in the R command window or in an additional editor window. For creating the saving function in the command window, type for example

```
dump.data<-function (){dump(c("dump.data","chicken","chicken0","darwin","darwin0",
+ "mustard","mustard0","pepper","pepper0","split","split0"),"all_data.asc")}
```

To create the function in the additional editor window, type

```
> fix(dump.data)
```

Then the additional window appears where you can write:

```
function ()
{
dump(c("dump.data","chicken","chicken0","darwin","darwin0","mustard",
"mustard0","pepper","pepper0","split","split0"),"all_data.asc")
}
```

After saving this window the function `dump.data` is available and can be used by

```
> dump.data()
```

By typing

```
> ls()
```

you see list of all data sets and the function `dump.data`. Data sets and functions are R objects listed by the function `ls`. It has no arguments like `dump.data`. To remove R objects, type for example for deleting the data set `chicken0`

```
> rm(chicken0)
```

You will find the file `all_data.asc` also on the homepage of this lecture. It includes all data sets used in this lecture. If you have problems to create a data set, then you can load it from the file `all_data.asc`. But you should try to create the data sets by yourself because this is a qualification you need in practice.

You can use also self-defined functions for saving several R comands for example some R commands for creating a special data set.

1.7 Descriptive Statistics

The main location estimators are easiliy obtained by the R function `summary`:

```
> str(chicken0)
'data.frame':   8 obs. of  4 variables:
 $ V1: int  1 2 3 4 5 6 7 8
 $ V2: num  3.93 3.78 3.88 3.93 3.84 3.75 3.98 3.84
 $ V3: num  3.99 3.96 3.96 4.03 4.1 4.02 4.06 3.92
 $ V4: num  3.96 3.94 4.02 4.06 3.94 4.09 4.17 4.12
> summary(chicken0)
       V1              V2              V3              V4
 Min.   :1.00   Min.   :3.750   Min.   :3.920   Min.   :3.940
 1st Qu.:2.75   1st Qu.:3.825   1st Qu.:3.960   1st Qu.:3.955
 Median :4.50   Median :3.860   Median :4.005   Median :4.040
 Mean   :4.50   Mean   :3.866   Mean   :4.005   Mean   :4.037
 3rd Qu.:6.25   3rd Qu.:3.930   3rd Qu.:4.037   3rd Qu.:4.098
 Max.   :8.00   Max.   :3.980   Max.   :4.100   Max.   :4.170
```

We see that also for the block numbers the location parameters are calculated, which makes no sense. Using the data table `chicken`, we obtain:

```
> str(chicken)
'data.frame':   24 obs. of  3 variables:
 $ Block : int  1 2 3 4 5 6 7 8 1 2 ...
 $ Weight: num  3.93 3.78 3.88 3.93 3.84 3.75 3.98 3.84 3.99 3.96 ...
 $ Feed  : Factor w/ 3 levels "Control","Low",..: 1 1 1 1 1 1 1 1 2 2 ...
> summary(chicken)
     Block          Weight           Feed
 Min.   :1.00   Min.   :3.750   Control:8
 1st Qu.:2.75   1st Qu.:3.928   Low    :8
 Median :4.50   Median :3.960   High   :8
 Mean   :4.50   Mean   :3.970
 3rd Qu.:6.25   3rd Qu.:4.037
 Max.   :8.00   Max.   :4.170
```

Note that the function `summary` recognize factors and calculate for them only the frequency list. To get the locations estimates of the weight for the different feeding groups from the data set `chicken`, type:

```
> C1<-chicken[chicken$Feed=="Control","Weight"]
> C2<-chicken[chicken$Feed=="Low","Weight"]
> C3<-chicken[chicken$Feed=="High","Weight"]
> summary(cbind(C1,C2,C3))
       C1              C2              C3
 Min.   :3.750   Min.   :3.920   Min.   :3.940
 1st Qu.:3.825   1st Qu.:3.960   1st Qu.:3.955
```

```
 Median :3.860   Median :4.005   Median :4.040
 Mean   :3.866   Mean   :4.005   Mean   :4.037
 3rd Qu.:3.930   3rd Qu.:4.037   3rd Qu.:4.098
 Max.   :3.980   Max.   :4.100   Max.   :4.170
```

These are the same locations estimates as obtained for `chicken0`. Only the derivation is more complicated. However, the data table `chicken` is much more convenient for plotting the box-and-whisker plots.

```
> boxplot(Weight~Feed,data=chicken)
> boxplot(Weight~Block,data=chicken)
```

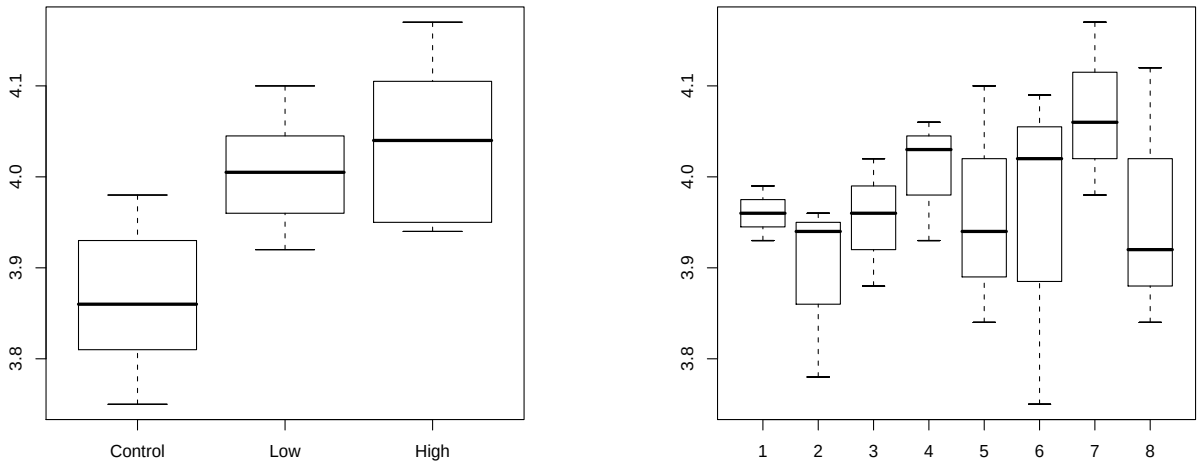

Figure 1.1: Box plots with respect to `Feed` and `Block`

If we interpret the blocks as experimental units so that each block has three weight measurements, we also can plot scatterplots for example the weights of the control group against the weights of the group with low drug:

```
> plot(C1,C2,xlab="Control",ylab="Low")
```

To plot the scatterplots of all possible compinations of two variable we can use also the function `pairs`:

```
> pairs(cbind(C1,C2,C3))
```

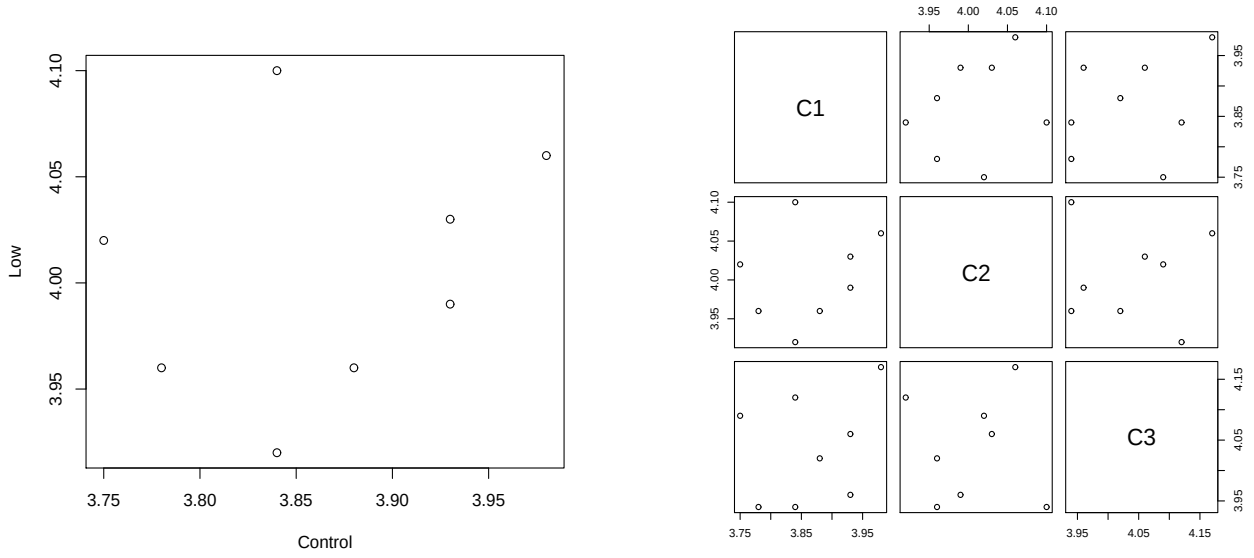

Figure 1.2: Scatter plots for one pair of variables and for all pairs

1.7.1 **Example** (SPLIT)

```
> str(split)
'data.frame':   72 obs. of  4 variables:
 $ Block  : Factor w/ 6 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Variety: Factor w/ 3 levels "1","2","3": 1 1 1 1 2 2 2 2 3 3 ...
 $ Manure : num  0 0.01 0.02 0.04 0 0.01 0.02 0.04 0 0.01 ...
 $ Yield  : num  111 130 157 174 117 114 161 141 105 140 ...
> Y1<-split[split$Variety=="1",]
> Y2<-split[split$Variety=="2",]
> Y3<-split[split$Variety=="3",]
> summary(cbind(Y1$Yield,Y2$Yield,Y3$Yield))
       X1               X2              X3
 Min.   : 53.00   Min.   : 60.0   Min.   : 63.00
 1st Qu.: 74.00   1st Qu.: 85.0   1st Qu.: 96.75
 Median : 94.00   Median :102.5   Median :113.00
 Mean   : 97.63   Mean   :104.5   Mean   :109.79
 3rd Qu.:113.75   3rd Qu.:126.0   3rd Qu.:124.00
 Max.   :174.00   Max.   :161.0   Max.   :156.00
> boxplot(Yield~Variety,data=split)
```

But also the other boxplots are easily obtained:

```
> boxplot(Yield~Block,data=split)
> boxplot(Yield~Block*Variety,data=split)
```

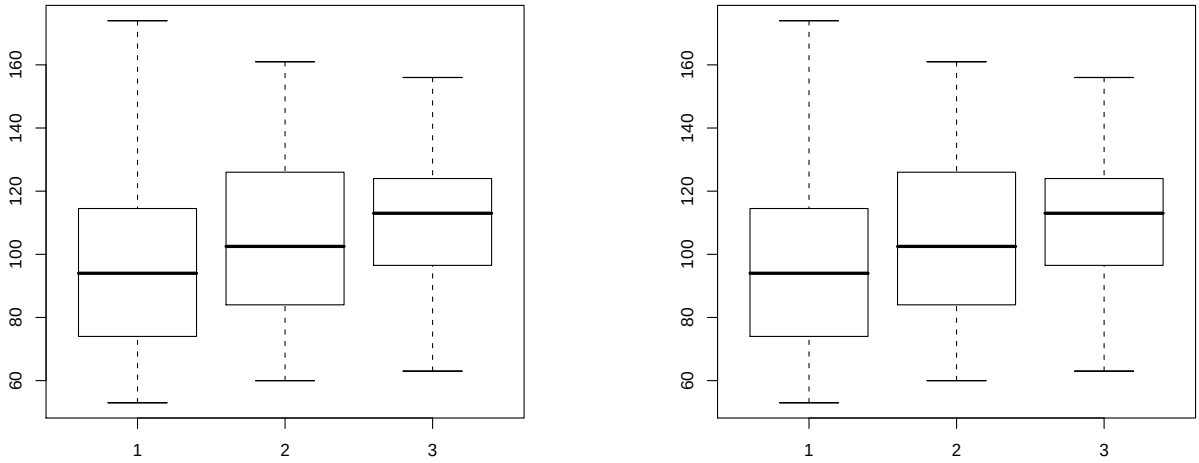

Figure 1.3: Box plots with respect to `Variety` and `Block`

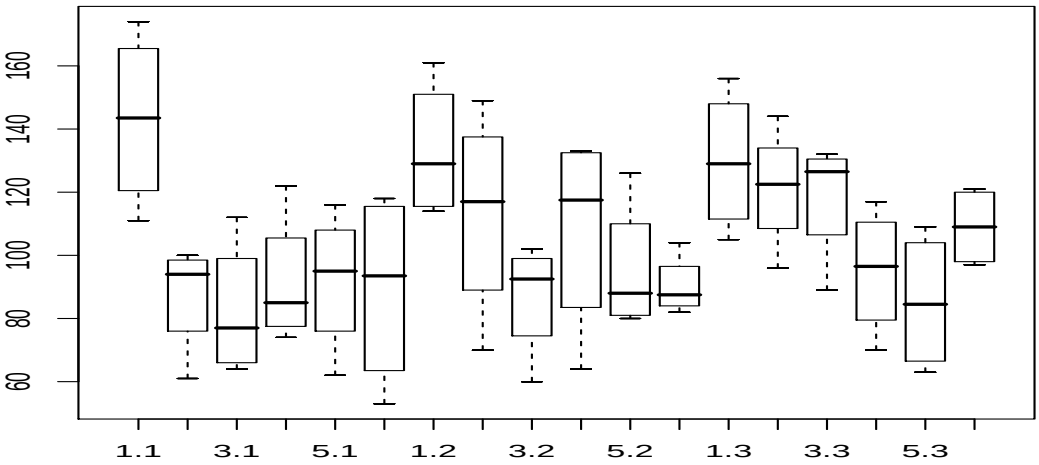

Figure 1.4: Box plots with respect to `Block*Variety`

Since the variables `Yield` and `Manure` are numeric in the data set `split` they can be plotted in a scatter plot:

```
> plot(Yield~Manure,data=split)
```

The plots also can be given separately for the different varieties.

```
> plot(Yield~Manure,data=Y1)
> points(Yield~Manure,data=Y2,pch=2)
> points(Yield~Manure,data=Y3,pch=3)
> legend(0.028,75,c("Variety 1","Variety 2", "Variety 3"),pch=c(1,2,3))
```

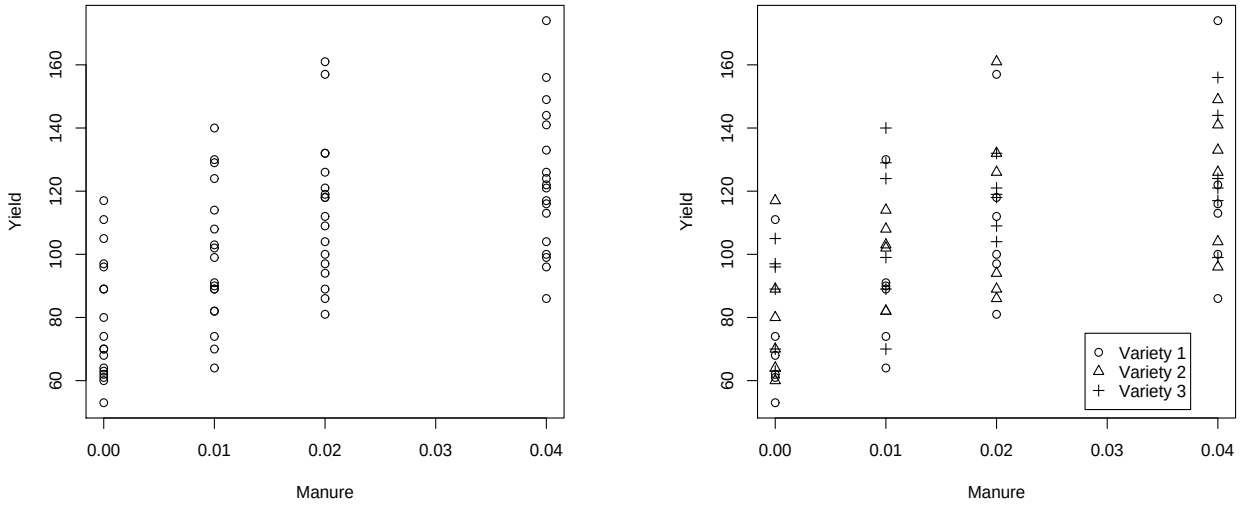

Figure 1.5: Plots for all data and for data separated according to `Variety`

1.7.2 Exercise (Mustard)
Calculate for the data sets `mustard0` and `mustard` from Data set 1.4.2 the summary table of the location estimators. Calculate in particular the location parameters for the different treatment groups by using the data set `mustard`. Plot also the boxplots for all 4 treatment groups. What happens when the two treatment factors are exchanged. Plot also the boxplots for the two treatment groups for each factor separately.

1.7.3 Exercise (Darwin)
Calculate for the data sets `darwin0` and `darwin` from Data set 1.3.2 the summary table of the location estimators. Plot the boxplots for the two treatment groups. Use also the pairs as experimental unit and plot the scatterplot for the height under cross-fertilization and self-fertilization.

2 Test theory and the two-sample problem

In this section, the main concepts for statistical tests and corresponding designs are explained at the example of the two sample problem

If the data set has only two variables where one variable is a treatment or block factor with two levels, then we have a two sample problem. Here we will assume that the second variable is a numeric variable and therefore a measurement. If the treatment or block factor has only two levels we can divide the data set in two samples (two groups), one with the measurements, where the factor variable attains the first level, and the other with the measurements, where the factor variable attains the second level. Such situations we also have, if we are ignoring other variables.

Let $y_{11}, \ldots, y_{1N_1}$ be the measurements of the first sample and $y_{21}, \ldots, y_{2N_2}$ the measurements of the second sample. The sample sizes N_1 and N_2 can be equal or different. The vector of observations/measurements for the first sample is denoted by $y_{1\bullet} = (y_{11}, \ldots, y_{1N_1})^\top$ and the vector for the second sample by $y_{2\bullet} = (y_{21}, \ldots, y_{2N_2})^\top$. Here we will assume that $y_{11}, \ldots, y_{1N_1}$ are realizations of independent identically distributed random variables $Y_{11}, \ldots, Y_{1N_1}$ with normal distribution $\mathcal{N}(\mu_1, \sigma_1^2)$ and that $y_{21}, \ldots, y_{2N_2}$ are realizations of independent identically distributed random variables $Y_{21}, \ldots, Y_{2N_2}$ with normal distribution $\mathcal{N}(\mu_2, \sigma_2^2)$.

2.1 Test theory by means of the two-sample t-test

The two means of the two samples can be estimated by the arithmetic mean

$$\overline{y}_{1\bullet} = \frac{1}{N_1}\sum_{n=1}^{N_1} y_{1n} \quad \text{and} \quad \overline{y}_{2\bullet} = \frac{1}{N_2}\sum_{n=1}^{N_2} y_{2n}$$

and are obtained by R by the function `mean`. Of particular interest is the difference between the means

$$\overline{y}_{1\bullet} - \overline{y}_{2\bullet}.$$

However, a big difference does not automatically mean that the true means μ_1 and μ_2 are different. The difference will be the greater the greater the variability of the data is. Hence for testing

the hypothesis $H_0 : \mu_1 = \mu_2$ versus the alternative $H_1 : \mu_1 \neq \mu_2$

we use the test statistic

$$\widehat{d} = \sqrt{\frac{N_1\, N_2}{N_1 + N_2}}\, \frac{\overline{y}_{1\bullet} - \overline{y}_{2\bullet}}{\widehat{\sigma}_{12}}$$

where

$$\widehat{\sigma}_{12}^2 = \frac{1}{N_1 + N_2 - 2}\left(\sum_{n=1}^{N_1}(y_{1n} - \overline{y}_{1\bullet})^2 + \sum_{n=1}^{N_2}(y_{2n} - \overline{y}_{2\bullet})^2\right).$$

Since $y_{11}, \ldots, y_{1N_1}, y_{21}, \ldots, y_{2N_2}$ are realizations of random variables $Y_{11}, \ldots, Y_{1N_1}, Y_{21}, \ldots, Y_{2N_2}$, also the test statistic $\widehat{d}$ is realization of a random variable $\widehat{D}$.

P-value

If t is a realization of a test statistic T, then the maximum probability under the null hypothesis H_0 that T attains the same or a more extreme value than t is called P-value.

Level α test

The null hypothesis H_0 is rejected if the P-value is not greater than α.

Usual choice of α

$$\alpha = \frac{0.05}{\text{number of tests at the same data set}}.$$

If $\sigma_1^2 = \sigma_2^2$, then the test statistic $\widehat{D}$ has under $H_0 : \mu_1 = \mu_2$ a central t-distribution with $N_1 + N_2 - 2$ degree of freedoms. Then we have the following decision rule

t-test

$$\text{Reject } H_0 : \mu_1 = \mu_2 \quad \text{if} \quad \begin{cases} \text{P-value is not greater than } \alpha, & \text{or} \\ |\widehat{d}| > t_{N_1+N_2-2,1-\frac{\alpha}{2}}, & \text{respectively.} \end{cases}$$

Thereby $t_{N,\alpha}$ denotes the α-quantile of the central t-distribution with N degrees of freedom. Quantiles of distributions are given in R by q*distname*, i.e. the quantiles of the t distribution are given by `qt`. The R function `t.test` calculates the test statistic $|\widehat{d}|$ and the P-value.

The R function `t.test` also calculates the lower and upper bounds of the confidence interval for $\mu_1 - \mu_2$. The $1-\alpha$ interval is given by

$$\begin{aligned} &\widehat{C}_{1-\alpha}(y_{1\bullet}, y_{2\bullet}) \\ &= \left[\overline{y}_{1\bullet} - \overline{y}_{2\bullet} - t_{N_1+N_2-2,1-\frac{\alpha}{2}}\, \widehat{\sigma}_{12} \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}\ ,\ \overline{y}_{1\bullet} - \overline{y}_{2\bullet} + t_{N_1+N_2-2,1-\frac{\alpha}{2}}\, \widehat{\sigma}_{12} \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \right] \end{aligned}$$

It has the property that the probability that the true difference $\mu_1 - \mu_2$ lies in the interval is at least $1-\alpha$. With a $1-\alpha$ confidence interval for $\mu_1 - \mu_2$, we have a third possibility to perform the t-test:

t-test

$$\text{Reject } H_0 : \mu_1 = \mu_2 \quad \text{if} \quad 0 \notin \widehat{C}_{1-\alpha}(y_{1\bullet}, y_{2\bullet}).$$

2.2 Checking the requirements of the t-test

The t-test has two very important requirements: the normal distribution of both samples and the equality of the true variances. These requirements must be checked by tests. Since these tests are only pretests, they are not influencing the level α for the main test(s).

1. Checking the normal distribution: The normal distribution can be checked in R with `shapiro.test`. If the assumption of normal distribution is rejected for at least on sample, then no version of a t-test can be used. Then the Wilcoxon rang sum test `wilcox.test` must be used which makes no assumptions for the distribution.

2. Checking the homogeneity of the variances: The test for $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$ bases on the quotient of the empirical variances for the two samples

$$\widehat{v} = \frac{\widehat{\sigma}^2(y_{1\bullet})}{\widehat{\sigma}^2(y_{2\bullet})}$$

with

$$\widehat{\sigma}^2(y_{1\bullet}) = \frac{1}{N_1 - 1}\sum_{n=1}^{N_1}(y_{1n} - \overline{y}_{1\bullet})^2 \quad \text{and} \quad \widehat{\sigma}^2(y_{2\bullet}) = \frac{1}{N_2 - 1}\sum_{n=1}^{N_2}(y_{2n} - \overline{y}_{2\bullet})^2$$

The test statistic $\widehat{v}$ is a realization of a random variable $\widehat{V}$ which has under the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ a central F distribution with $N_1 - 1$ and $N_2 - 1$ degrees of freedom.

F-test for testing the equality of variances

$$\text{Reject } H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{if} \quad \begin{cases} \text{P-value is not greater than } \alpha, & \text{or} \\ \widehat{v} < F_{N_1-1,N_2-1,\frac{\alpha}{2}} \text{ or } \widehat{v} > F_{N_1-1,N_2-1,1-\frac{\alpha}{2}}, & \text{respectively.} \end{cases}$$

Thereby $F_{N,M,\alpha}$ denotes the α-quantile of the central F distribution with N and M degrees of freedom and is given in R by `qF`. The test statistic and the P-value of the F-test are given in R by `var.test`. If $H_0 : \sigma_1^2 = \sigma_2^2$ is rejected, then the simple t-test cannot be used. However a modified version, the Welch t-test, can be used, which is automatically used in R then. But the Welch t-test also assumes normal distribution for both samples.

2.2.1 Example (Mustard)
It shall be tested whether the length differs under the two growing conditions and whether the length differs under the two cutting treatments. At first we check the assumption of normal distribution for the two growing conditions.

```
> mustardL<-mustard[mustard$grow.conditions=="light","length"]
> mustardD<-mustard[mustard$grow.conditions=="dark","length"]
> shapiro.test(mustardL)$p.value
[1] 0.2552495
> shapiro.test(mustardD)$p.value
[1] 0.0746946

> var.test(mustardL,mustardD)

        F test to compare two variances

data:  mustardL and mustardD
```

```
F = 3.1851, num df = 15, denom df = 14, p-value = 0.03640
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.079940 9.209621
sample estimates:
ratio of variances
          3.185090
```

Hence we can use the Welch t-test.

```
> t.test(mustardL,mustardD)

        Welch Two Sample t-test

data:  mustardL and mustardD
t = 0.7751, df = 23.862, p-value = 0.4459
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.635834 12.410834
sample estimates:
mean of x mean of y
  28.1875   24.8000
```

We can obtain the results of the last two tests also faster:

```
> var.test(length~grow.conditions,data=mustard)$p.value
[1] 0.03640493
> t.test(length~grow.conditions,data=mustard)$p.value
[1] 0.4459191
```

For the two cutting groups we also do not reject the normality assumption. Moreover, we do not reject the assumption of equal variances:

```
> shapiro.test(mustard[mustard$cutting=="noncut","length"])$p.value
[1] 0.5183037
> shapiro.test(mustard[mustard$cutting=="cut","length"])$p.value
[1] 0.1366706

> var.test(length~cutting,data=mustard)$p.value
[1] 0.1835491

> t.test(length~cutting,data=mustard)

        Welch Two Sample t-test

data:  length by cutting
```

```
t = 2.2734, df = 27.788, p-value = 0.03093
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.876303 16.887333
sample estimates:
   mean in group cut mean in group noncut
             29.70000              20.81818
```

Since the equality of the variances is not rejected, it is more accurate to use the nonmodified t-test:

```
> t.test(length~cutting,data=mustard,var.equal=T)

        Two Sample t-test

data:  length by cutting
t = 2.0223, df = 29, p-value = 0.05246
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1007162 17.8643526
sample estimates:
   mean in group cut mean in group noncut
             29.70000              20.81818
```

We can calculate the values also per hand

```
> mustardC<-mustard[mustard$cutting=="cut","length"]
> mustardNC<-mustard[mustard$cutting=="noncut","length"]
> length(mustardC)
[1] 20
> length(mustardNC)
[1] 11
> d<-sqrt(11*20/31)*(mean(mustardC)-mean(mustardNC))/
+ sqrt((19*var(mustardC)+10*var(mustardNC))/29)
> d
[1] 2.022298
> pt(-d,29)+1-pt(d,29)
[1] 0.05245525
```

Not that the p-value is here $P(|\widehat{D}| \geq |\widehat{d}|) = P(D \leq -|d| \text{ or } D \geq |d|) = P(D \leq -|d|)+1-P(D < |d|)$. Hence the p-value can be calculated per hand via the distribution function of the t distribution which is given in R by `pt`. We can also calculate the 0.95 confidence interval per hand:

```
> (mean(mustardC)-mean(mustardNC))-qt(0.975,29)*
+ sqrt((19*var(mustardC)+10*var(mustardNC))/29)/sqrt(11*20/31)
[1] -0.1007162
> (mean(mustardC)-mean(mustardNC))+qt(0.975,29)*
+ sqrt((19*var(mustardC)+10*var(mustardNC))/29)/sqrt(11*20/31)
[1] 17.86435
```

Since the 0.95 confidence interval contains 0, the hypothesis $H_0 : \mu_1 = \mu_0$ is not rejected, if only one test is performed.

But here even two tests are performed at the same data set. Therefore, we have to set $\alpha = 0.05/2 = 0.025$. From

```
> t.test(mustardL,mustardD)$p.value
[1] 0.4459191
> t.test(mustardC,mustardNC,var.equal=T)$p.value
[1] 0.05245525
```

we see that both tests are not rejecting the equality of the means. The Wilcoxon test, which can be always used, provides even worse results

```
> wilcox.test(mustardL,mustardD)$p.value
[1] 0.7216463
Warnmeldung:
cannot compute exact p-value with ties in: wilcox.test.default(mustardL, mustardD)
> wilcox.test(mustardC,mustardNC)$p.value
[1] 0.07209985
Warnmeldung:
cannot compute exact p-value with ties in: wilcox.test.default(mustardC, mustardNC)
```

That both hypothesis of equality of means are not rejected is due to the adjustment of the level α. It would be better two tests both hypothesis with one test. This will be done later.

2.2.2 Exercise (Growing)
"Heights were meausred (to the nearest inch) of maize plants in adjacent rows which differed only in a pollen sterility factor."(Hand et al. 1996, P. 130/131)
The data file `GROWING.DAT` contains the data where the first column concerns the heights for fertile pollen and the second column the heights for sterile pollen. Read the data and create a data set `growing` in which different rows belongs to different plants. Plot the boxplots from this data set. Moreover test whether the two groups with fertile and sterile pollen differs with respect to the means. Use also the necessary pretests. In any case, use also the Wilcoxon test for comparison. Interpret the results.

2.2.3 Exercise (Darwin's fertilization experiment)
Test by means of the data set `darwin` whether the mean heights under cross- and self-fertilitzation are the same. Calculate also with `mean` the difference between the arithmetic means of the two groups and compare the results with the boxplots from Exercise 1.7.3.

2.3 α and β error

Every statistical test for testing a null hypothesis H_0 versus an alternative hypothesis H_1 can make wrong or correct decisions. There two types of wrong decisions called α and β error.

	decision for H_0	decision for H_1
H_0 is true	correct decision	α-error
H_1 is true	β-error	correct decision

Since statistical tests do the decisions based on data which are realizations of random variables, the decisions are done randomly. Hence we have probabilities for correct and wrong decisions. An α-level test is a test where the probability for the α-error is less than α. This means that if we decide according to our data for H_1 we only can have a correct decision or the α-error. Since α is very small (usually less or equal 0.05) the probability for wrong decision is low if we decide for H_1. Therefore we say that the data are speaking significantly for H_1 or versus H_0, respectively, if the decision according to the data is for H_1.

If we decide according to our data for H_0 then we can have a correct decision or a β-error. However, we usually do not know how large the β-error of our test is. Usually the probability for the β-error can be up to $1-\alpha$ where $1-\alpha \geq 0.95$. Hence there is a very high probability for the β-error. Therefore we say that the data are speaking not versus the null hypothesis H_0 if the decision according to the data is for H_0. This can mean in particular that we do not have enough data to reject the null hypothesis H_0.

Interpretation of test results

A decision for H_1 or versus H_0, respectively, based on the data is a significant result.

A decision for H_0 based on the data is a useless result. It could mean in particular that we have too few data to reject H_0.

To see what this means for the t-test, we will simulate data. Random numbers with normal distribution can be easily generated with the command `rnorm`. To generate two normally distributed samples, we can create the self-defined function `twosample` as follows in the additional editor window:

```
function(N1,mu1,sigma1,N2,mu2,sigma2)
{
list(sample1=rnorm(N1,mu1,sigma1),sample2=rnorm(N2,mu2,sigma2))
}
```

To generate for example two samples, one with $N1 = 10$, $\mu_1 = 3$, $\sigma_1 = 2$, and the other with $N2 = 12$, $\mu_2 = 2$, $\sigma_2 = 4$, we have then only to type

```
> twosample(N1=10,mu1=3,sigma1=2,N2=12,mu2=2,sigma=4)
$sample1
 [1] 1.218657 6.299094 5.060729 4.137085 6.324159 6.676449 4.562354 2.024045
 [9] 3.272009 4.857833

$sample2
 [1] -0.3604491  3.4885205  9.6260172  5.9064812 -0.5593599  2.1067412
 [7]  4.5098866  7.1203010  0.4827544  2.0188771 -1.5580092  2.4073573
```

To get the probabilities for α-error and β-error we have only to repeat the generation of the two sample many,many times, say 10.000 times and to count the cases, where we make the error. From

the law of large numbers we know that the relative number of cases approximates the probability if the number of repetitions is high enough. 10.000 repetitions is high enough.

α-error: At first we simulate the α-error. In this case the null hypothesis $H_0 : \mu_1 = \mu_2$ is true. Moreover, we need $\sigma_1 = \sigma_2$ for the t-test. Hence only the sample sizes N_1 and N_2 can be different so that the simulation function `alpha.error` is only a function with the arguments `mu, sigma, N1, N2, M, alpha`, where `M` denotes the number of repetitions which is set by default to 10 000. `alpha` is the level of the test and is set by default to 0.05. If default values are given for the arguments of a function, then these arguments must be not specified by calling the function but can be specified if other values shall be used. The function `alpha.error` is defined as follows:

```
function (mu,sigma,N1,N2,M=10000,alpha=0.05)
{
# Function which simulates the aplha error
error<-0
for(i in 1:M){
  s<-twosample(N1=N1,mu1=mu,sigma1=sigma,N2=N2,mu2=mu,sigma2=sigma)
  s1<-s$sample1
  s2<-s$sample2
  if(t.test(s1,s2,var.equal=T)$p.value<=alpha){
# If decision for H1:
      error<-error+1
  }
}
list(alpha.error=error/M)
}
```

Calling this function we get:

```
> alpha.error(mu=3,sigma=2,N1=10,N2=12)
$alpha.error
[1] 0.0519

> alpha.error(mu=3,sigma=2,N1=10,N2=12)
$alpha.error
[1] 0.0462
```

We see that several calls of the function `alpha.error` provides different α-errors but all α-errors are very close to 0.05. We would get closer results to 0.05 by using larger repetitions numbers `M`. We can interpret the result as follows: If many, many people are using the t-test and the null hypothesis $H_0 : \mu_1 = \mu_2$ is true, then only approximately 5% of the people would reject fasely the null hypothesis, i.e. would falsely decide that $\mu_1 \neq \mu_2$ is true. Hence if you get the decision $\mu_1 \neq \mu_2$, there are two possibilities for you: you may one of the unlucky 5% of people who make a wrong decision or your decision is correct.

β-error: Now we simulate the β-error. In this case the alternative $H_1 : \mu_1 \neq \mu_2$ is true. But still we need $\sigma_1 = \sigma_2$ for the t-test. Hence the simulation function `beta.error` is a function with the arguments `mu1, mu2, sigma, N1, N2, M, alpha` and is defined as follows:

```
function (mu1,mu2,sigma,N1,N2,M=10000,alpha=0.05)
{
# Function which simulates the beta error
error<-0
for(i in 1:M){
  s<-twosample(N1=N1,mu1=mu1,sigma1=sigma,N2=N2,mu2=mu2,sigma2=sigma)
  s1<-s$sample1
  s2<-s$sample2
  if(t.test(s1,s2,var.equal=T)$p.value>alpha){
# If decision for H0:
      error<-error+1
  }
}
list(beta.error=error/M)
}
```

For $\mu_1 = 3$ and $\mu_2 = 4$, which satisfies the alternative, we obtain for example:

```
> fix(beta.error)
> beta.error(mu1=3,mu2=4,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.7965

> beta.error(mu1=3,mu2=4,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.7999
```

This means that if many, many people do the t-test for data coming from the two different normal distribution $\mathcal{N}(3, 4^2)$ and $\mathcal{N}(4, 4^2)$, then approximately 79% of the people make the wrong decision that the two normal distributions are the same.

All self-defined functions used in this lecture can be found on the homepage of this lecture. They are included in the ASCII file `all_funct.asc` which can be loaded with the `source` command like the data ASCII file `all_data.asc` described in Subsection 1.6. But if you are planning to write own R functions in future, then you should implement the functions by yourself to get more experience with this.

The β-error can be also obtained mathematically via the non-central t-distribution. Namely, if $|\mu_1 - \mu_2| = \delta\sigma$, then the β-error is given by:

$$\begin{aligned} & P_{|\mu_1-\mu_2|=\delta\sigma}\,(\text{decision for } H_0) = P_{|\mu_1-\mu_2|=\delta\sigma}\left(|\widehat{D}| \le t_{N_1+N_2-2,1-\alpha/2}\right) \\ &= \; F_{t(N_1+N_2-2,K\,\delta)}(t_{N_1+N_2-2,1-\alpha/2}) - F_{t(N_1+N_2-2,K\,\delta)}(-t_{N_1+N_2-2,1-\alpha/2}) \end{aligned} \tag{1}$$

where $F_{t(N_1+N_2-2,K\,\delta)}$ is the distribution function of the non-central t-distribution with $N_1 + N_2 - 2$ degrees of freedom and non-centrality parameter $K\,\delta$ and $K = \sqrt{\frac{N_1\,N_2}{N_1+N_2}}$. The distribution function of the non-central t-distribution is given in R by `pt` by specifying the non-centrality parameter `ncp`. Hence for $\mu_1 = 3$, $\mu_2 = 4$, $\sigma_1 = \sigma_2 = 2$, $N_1 = 10$, $N_2 = 12$ of the above example we obtain

since $N_1 + N_2 - 2 = 20$, $N_1\,N_2 = 120$, $N_1 + N_2 = 22$, $\delta = |\mu_1 - \mu_2|/\sigma = 0.5$, $K\delta = \sqrt{6}\ 0.5$, and $1 - 0.05/2 = 0.975$:

```
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.5)-pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.5)
[1] 0.800646
>
```

We see that the result is similar to the simulated values. Note that for $\delta = 0$, i.e. the null hypothesis is true, we obtain

```
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0)-pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0)
[1] 0.95
```

i.e. we obtain $1 - \alpha$. Here the noncentrality parameter is equal to 0 so that the noncentral t-ditribution becomes the central t-distribution.

2.3.1 Exercise ($\beta - errors$)
a) Simulate the β-error for $N_1 = 10$, $N_2 = 12$, $\sigma_1 = 2 = \sigma_2$, $\mu_1 = 3$ and the following values for μ_2: 4, 3.5, 3.1, 3.01.
b) Do the same for $N_1 = 10$, $N_2 = 12$, $\sigma_1 = 2 = \sigma_2$, $\mu_1 = 6$ and the following values for μ_2: 7, 6.5, 6.1, 6.01, i.e. all values for μ_1 and μ_2 are added by 3.
c) Do the same for $N_1 = 10$, $N_2 = 12$, $\sigma_1 = 1 = \sigma_2$, $\mu_1 = 6$ and the following values for μ_2: 7, 6.5, 6.1, 6.01, i.e. only $\sigma_1 = \sigma_2$ is changed.
What are your conclusions?
d) Compare all simulated values with the theoretical values given by the noncentral t-distribution.

2.4 Design considerations

Since the α-error of any α-level test is α, the α-error cannot be influenced by the design of the experiment. However, the β-error can be influenced by the design. It should be as small as possible since then the probability is as high as possible to get a significant result, i.e. that H_0 is rejected, if H_1 is indeed true. Thereby note that for the probability of the β-error we have

$$P_{H_1}(\text{decision for } H_0) = 1 - P_{H_1}(\text{decision for } H_1) = 1 - P_{H_1}(\text{significant result})$$

Hence minimization of the β-error, i.e. $P_{H_1}(\text{decision for } H_0)$, means maximizing $P_{H_1}(\text{significant result})$.

Optimal designs

An optimal design is a design which minimizes the β-error of the α-level test.

Optimal allocation of N experiments to two groups: If the total number of experiments shall be $N = N_1 + N_2$, the N_1 and N_2 should be chosen so that the β-error is as small as possible. Formula (1) shows that the β-error depends only on δ, N_1, and N_2. If N is small then the β-error can be easily minimized by playing with δ, N_1, and N_2. To facilitate this task, write the function `beta.error.exact` as follows:

```
function (delta,N1,N2,alpha=0.05)
{
# Calculates the theoretical beta error
K<-sqrt(N1*N2/(N1+N2))
pt(qt(1-alpha/2,20),20,ncp=K*delta)-pt(-qt(1-alpha/2,20),20,ncp=K*delta)
}
```

2.4.1 Exercise
Determine N_1 and N_2, if N should be 30. Does the choice of N_1 and N_2 depends on δ. What is your proposal for the general case?
For mathematicians: Prove the general proposal.

Planning the sample size: Often the total number of all experiments is not given from the beginning and must be chosen by the experimenter. In general we have the general rule:

The larger the sample size N is the smaller the β-error is.

But many experiments produce costs so that the sample size cannot be arbitrary high. Then the aim is to determine the sample size N such that the β-error for a given deviation from the null hypothesis is not greater than a given value β. For the two sample problem we may demand that the β-error is not greater than β if the absolute difference of the means μ_1 and μ_2 is greater a given value κ, i.e. $|\mu_1 - \mu_2| > \kappa$. This means that as soon as $|\mu_1 - \mu_2| > \kappa$ is true, the t-test would reject the null hypothesis, i.e. provide a significant result, with probability of at least $1 - \beta$. Thereby κ denotes a **relevant** difference between the means. This relevant difference is often known or can be specified in practice. Usually β is chosen as α so that the β-error is for $|\mu_1 - \mu_2| > \kappa$ the same as the α-error. Hence the test has the same probability for a wrong decision for $|\mu_1 - \mu_2| > \kappa$ as for $\mu_1 = \mu_2$. However, the probability for a wrong decision is still up to $1 - \alpha$ if $0 < |\mu_1 - \mu_2| < \kappa$ is true.

The aim is now to find the smallest sample size N so that the β-error for $|\mu_1 - \mu_2| > \kappa$ is not greater than β. For simplicity, we will assume that $N_1 = N_2 = \frac{N}{2}$. However, according to 1 the β-error can be only calculated for $\kappa = \delta\ \sigma$ and σ is unknown. Hence relevant alternatives, i.e. relevant differences, must be specified in terms of the unknown standard error. But, as soon as δ is known we can determine the sample size $N = 2\,N_1 = 2\,N_2$ as the minimum number N such that

$$F_{t(N-2,K\,\delta)}(t_{N-2,1-\alpha/2}) - F_{t(N-2,K\,\delta)}(-t_{N-2,1-\alpha/2}) \leq \beta$$

with $K = \sqrt{\frac{N}{2}}$. The calculation of N can be done by try and error by using the self-defined function `beta.error.exact` with `N1=N/2` and `N2=N/2`, i.e. repeat for example for $\delta = 2$

```
> N<-10
> beta.error.exact(2,N/2,N/2)
```

for several values of N.

2.4.2 Exercise
Determine the minimum sample size N so that the β-error is not greater than $\alpha = 0.05$ for $|\mu_1 - \mu_2| > 2\sigma$. Determine also the sample size for $|\mu_1 - \mu_2| > \sigma/2$.

A further design question: Usually the N experiments are done in a specific temporal or spacial order. For examples plants are growing on specific positions of the field, animals are living in specific places of a cot, patients of a hospital arriving in a specific order to the hospital. Then the question is how to assign two different treatments to the experimental units. Since it is never clear if there are special spacial or temporal influences on the measurement, the assignments of the treatments to the experimental units should be done randomly. This allocation can be easily done by the function `design.crd` of the `agricolae` package.

For example to allocate 10 treatments `t1` and 10 treatments `t2` to 20 units, type:

```
> library(agricolae)
> design.crd(c("t1","t2"),c(10,10))
   plots c("t1", "t2")  r
1      1            t2  1
2      2            t2  2
3      3            t1  1
4      4            t2  3
5      5            t1  2
6      6            t1  3
7      7            t2  4
8      8            t1  4
9      9            t1  5
10    10            t2  5
11    11            t1  6
12    12            t1  7
13    13            t2  6
14    14            t2  7
15    15            t1  8
16    16            t2  8
17    17            t2  9
18    18            t1  9
19    19            t2 10
20    20            t1 10
```

Then we get an order how to allocate the two treatments: at first `t2`, then `t2` again, then `t1` and so on.

3 One-way ANOVA

As in Section 2, it is assumed that the data set contains only two variables: one numeric variable concerning measurements and a factor variable concerning treatments or blocks (groups). But here we assume that the factor can has more than two levels, i.e. that we may have more than two treatments or two groups. Let I denote the number of levels (treatments, blocks, groups) and let be

$$\begin{aligned}
&y_{1*} = (y_{11}, \ldots, y_{1N_1})^\top \text{ the vector of observations for level (group) 1,}\\
&y_{2*} = (y_{21}, \ldots, y_{2N_2})^\top \text{ the vector of observations for level (group) 2,}\\
&\vdots\\
&y_{I*} = (y_{I1}, \ldots, y_{IN_2})^\top \text{ the vector of observations for level (group) } I.
\end{aligned}$$

Altogether there are $N = N_1+N_2+\ldots+N_I$ observations. $y_{1*}, \ldots, y_{I*}$ are realizations of independent random vectors $Y_{1*}, \ldots, Y_{I*}$ where $Y_{i*} = (Y_{i1}, \ldots, Y_{iN_i})^\top$ for $i = 1, \ldots, I$.

3.1 The ANOVA test

$y_{i1}, \ldots, y_{iN_i}$ are realizations of independent identically distributed random variables $Y_{i1}, \ldots, Y_{iN_i}$ with normal distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, I$. This can be expressed also as

$$Y_{in} = \mu_i + Z_{in} = \mu + \alpha_i + Z_{in} \quad \text{with} \quad Z_{in} \sim \mathcal{N}(0, \sigma_i^2),$$

where Z_{in} is the measurement error, μ the average mean, and α_i the effect of level (group) i. The aim is to test the null hypothesis that there is no treatment/group effect, i.e. to test

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I \quad \text{versus} \quad H_1 : \text{ there exist } i, j \text{ with } \mu_i \neq \mu_j$$

or, equivalently,

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0 \quad \text{versus} \quad H_1 : \text{ there exist } i \text{ with } \alpha_i \neq 0.$$

The test statistic is based on estimates for the components μ, μ_i, α_i, Z_{in} and σ_i:

$$\begin{aligned}
\mu: &\quad \widehat{\mu} = \overline{y}_{\bullet\bullet} := \frac{1}{N}\sum_{i=1}^{I}\sum_{n=1}^{N_i} y_{in},\\
\mu_i: &\quad \widehat{\mu}_i = \overline{y}_{i\bullet} := \frac{1}{N_i}\sum_{n=1}^{N_i} y_{in},\\
\alpha_i = \mu_i - \mu: &\quad \widehat{\alpha}_i = \overline{y}_{i\bullet} - \overline{y}_{\bullet\bullet},\\
z_{in} = y_{in} - \mu_i: &\quad \widehat{z}_{in} = y_{in} - \overline{y}_{i\bullet},\\
\sigma^2: &\quad \widehat{\sigma}^2 = \frac{1}{N-I}\sum_{i=1}^{I}\sum_{n=1}^{N_i} \widehat{z}_{in}^2 = \frac{1}{N-I}\sum_{i=1}^{I}\sum_{n=1}^{N_i} (y_{in} - \overline{y}_{i\bullet})^2.
\end{aligned}$$

The ANOVA test (AN=analysis,O=of,VA=variance, in German: Varianzanalyse) is based on the decomposition of the general variance

$$\widehat{\sigma}^2_{SSG} := \frac{1}{N-1}\sum_{i=1}^{I}\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{\bullet\bullet})^2$$

as follows

$$\begin{aligned}
\Sigma_{SSG} &:= \sum_{i=1}^{I}\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{\bullet\bullet})^2 = \sum_{i=1}^{I}\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{i\bullet}+\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})^2 \\
&= \sum_{i=1}^{I}\sum_{n=1}^{N_i}\left[(y_{in}-\overline{y}_{i\bullet})^2+2(y_{in}-\overline{y}_{i\bullet})(\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})+(\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})^2\right] \\
&= \sum_{i=1}^{I}\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{i\bullet})^2+2\sum_{i=1}^{I}\underbrace{\left[\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{i\bullet})\right]}_{=0}(\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})+\sum_{i=1}^{I}N_i(\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})^2 \\
&= \sum_{i=1}^{I}\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{i\bullet})^2+\sum_{i=1}^{I}N_i(\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})^2 = \Sigma_{SSE}+\Sigma_{SST}
\end{aligned}$$

with

$$\begin{aligned}
\Sigma_{SSE} &= \sum_{i=1}^{I}\sum_{n=1}^{N_i}(y_{in}-\overline{y}_{i\bullet})^2, \\
\Sigma_{SST} &= \sum_{i=1}^{I}N_i(\overline{y}_{i\bullet}-\overline{y}_{\bullet\bullet})^2.
\end{aligned}$$

Σ_{SSG} is called Grand Sum of Squares, Σ_{SSE} Sum of Squares for errors, and Σ_{SST} Sum of Squares for Treatments. They have the following distributions:

$$\frac{1}{\sigma^2}\Sigma_{SSG}\sim\chi^2_{N-1},\quad \frac{1}{\sigma^2}\Sigma_{SSE}\sim\chi^2_{N-I},\quad \frac{1}{\sigma^2}\Sigma_{SST}\sim\chi^2_{I-1}.$$

Therefore the corresponding variances are $\widehat{\sigma}^2_{SSG}=\frac{1}{N-1}\Sigma_{SSG}$, $\widehat{\sigma}^2_{SSE}=\frac{1}{N-I}\Sigma_{SSE}$, and $\widehat{\sigma}^2_{SST}=\frac{1}{I-1}\Sigma_{SST}$. $\widehat{\sigma}^2_{SSE}$ is called variance within treatments and $\widehat{\sigma}^2_{SST}$ is called variance between treatments. A high variance between treatments compared with the variance within treatments is speaking versus the null hypothesis that the treatment effects are the same. Hence the test statistic is

$$\widehat{V}=\frac{\widehat{\sigma}^2_{SST}}{\widehat{\sigma}^2_{SSE}}.$$

If $\sigma_1^2=\sigma_2^2=\ldots\sigma_I^2$, then $\widehat{V}$ has a F-distribution with $I-1$ and $N-I$ degrees of freedom.

ANOVA test for one-way layout

Reject $H_0:\mu_1=\mu_2=\ldots=\mu_I$ if $\widehat{V}=\dfrac{\widehat{\sigma}^2_{SST}}{\widehat{\sigma}^2_{SSE}}>F_{I-1,N-I,1-\alpha}$.

Thereby $F_{N,M,\alpha}$ denotes the α-quantile of the central F-distribution with N and M degrees of freedom.

Usually, the values for the analysis of variance are summarized in the so-called ANOVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Differences between factor levels	$I-1$	$\Sigma_{SST} = \sum_{i=1}^{I} N_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2$	$\widehat{\sigma}^2_{SST} = \frac{1}{I-1}\Sigma_{SST}$
Measurement error	$N-I$	$\Sigma_{SSE} = \sum_{i=1}^{I}\sum_{n=1}^{N_i}(Y_{in} - \overline{Y}_{i\bullet})^2$	$\widehat{\sigma}^2_{SSE} = \frac{1}{N-I}\Sigma_{SSE}$
Total	$N-1$	$\Sigma_{SSG} = \sum_{i=1}^{I}\sum_{n=1}^{N_i}(Y_{in} - \overline{Y}_{\bullet\bullet})^2$	$\widehat{\sigma}^2_{SSG} = \frac{1}{N-1}\Sigma_{SSG}$

The first two rows of the ANOVA table together with the p-value and the value of the test statistic are given in R by the commands `anova(lm(....))` (`lm` from linear model).

Special case $I = 2$: In the case of $I = 2$, also the t-test can be used. However, it provides the same p-value as the ANOVA test since the ANOVA test statistic $\widehat{V}$ the squared t-test statistic $\widehat{d}$. i.e. $\widehat{d}^2 = \widehat{V}$.

3.1.1 Example (Growing)
Regard the data set `growing` from Exercise 2.2.2. Then the ANOVA test provides:

```
> anova(lm(Height~Pollen,data=growing))
Analysis of Variance Table

Response: Height
          Df Sum Sq Mean Sq F value  Pr(>F)
Pollen     1 192.67  192.67  7.1138 0.01408 *
Residuals 22 595.83   27.08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Recall that the t-test provides:

```
> t.test(Height~Pollen,data=growing,var.equal=T)

        Two Sample t-test

data:  Height by Pollen
t = 2.6672, df = 22, p-value = 0.01408
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  1.260534 10.072800
```

```
sample estimates:
mean in group Fertile mean in group Sterile
             96.58333              90.91667
```

Hnece the p-values coincide. And indeed the value of the squared t-test statistic

```
> (t.test(Height~Pollen,data=growing,var.equal=T)$statistic)^2
       t
7.113846
```

is the value of the test statistic of the ANOVA test which is called `F value` in the ANOVA table of R.

3.2 Checking the requirements of the one-way ANOVA test

As the t-test, the one-way ANOVA test has two requirements:

1. The measurements $y_{i1}, \ldots, y_{iN_i}$ must have a normal distributions for each level (treatment, group) $i = 1, \ldots, I$.
2. The variances of the normal distributions must be equal, i.e. $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_I^2$.

1. Checking the normal distribution: This done as for the t-test with the **Shapiro-Wilk test** given in R by `shapiro.test`.

2. Checking the homogeneity of the variances: If the normal distribution can be assumed for each level (group), then the equality (homogeneity) of the variances can be tested with the **Bartlett-Test**. It bases on the estimates of the variances σ_i^2 given by

$$\widehat{\sigma}_i^2 = \frac{1}{N_i - 1} \sum_{n=1}^{N_i} (Y_{in} - \overline{Y}_{i\bullet})^2$$

and the pooled variance estimate

$$\widehat{\sigma}_{SSE}^2 = \frac{1}{N - I} \sum_{i=1}^{I} \sum_{n=1}^{N_i} (Y_{in} - \overline{Y}_{\bullet})^2 = \frac{1}{N - I} \sum_{i=1}^{I} (N_i - 1) \widehat{\sigma}_i^2.$$

The test statistic has the form

$$\widehat{B} = \frac{1}{c} \left((N - I) \ln \widehat{\sigma}_{SSE}^2 - \sum_{i=1}^{I} (N_i - 1) \ln \widehat{\sigma}_i^2 \right),$$

where

$$c = \frac{1}{3(I-1)} \left(\sum_{i=1}^{I} \frac{1}{N_i - 1} - \frac{1}{N - I} \right) + 1.$$

Bartlett test for testing the homogeneity of the variances

Reject $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_I^2$ if $\widehat{B} > \chi^2_{I-1,1-\alpha}$.

Thereby, $\chi^2_{N,\alpha}$ denotes the α-quantile of the central χ^2-distribution with N degrees of freedom.

The Bartlett test is given in R with the function `bartlett.test`. Note that the Bartlett test statistic is not defined if one of the variance estimates $\widehat{\sigma}_i^2$ is equal to 0 since ln(0) is not defined. If `bartlett.test` is applied to a data set where one $\widehat{\sigma}_i^2$ is equal to 0, then it can happen that the R session breaks down.

Alternative, if the requirements of the ANOVA test are not satisfied: If the hypothesis of normal distribution for each level (group) or the hypothesis of equality of the variances is rejected, then the ANOVA test cannot be used. An alternative is the distribution-free **Kruskal-Wallis test** which generalizes the Wilcoxon rang sum test and is in R available under `kruskal.test`.

3.2.1 Example (Chicken)
Regard the data set `chicken` from Exercise 1.4.1. The aim is to test whether the drug added to the feed has an influence on the weight of chicken. At first we have to check whether the weight measurements for the three drug groups are normally distributed:

```
> shapiro.test(chicken[chicken$Feed=="Control","Weight"])
        Shapiro-Wilk normality test
data:  chicken[chicken$Feed == "Control", "Weight"]
W = 0.9607, p-value = 0.8164
> shapiro.test(chicken[chicken$Feed=="Low","Weight"])
        Shapiro-Wilk normality test
data:  chicken[chicken$Feed == "Low", "Weight"]
W = 0.9782, p-value = 0.9535
> shapiro.test(chicken[chicken$Feed=="High","Weight"])
        Shapiro-Wilk normality test
data:  chicken[chicken$Feed == "High", "Weight"]
W = 0.9234, p-value = 0.4582
```

Nothing speaks versus the normal distribution. Therefore we can test whether the variances are equal:

```
> bartlett.test(Weight~Feed,data=chicken)

        Bartlett test for homogeneity of variances

data:  Weight by Feed
Bartlett's K-squared = 0.9958, df = 2, p-value = 0.6078
```

Hence also the homogeneity of the variances is not rejected, so that the ANOVA test can be applied:

```
> anova(lm(Weight~Feed,data=chicken))
```

```
Analysis of Variance Table

Response: Weight
          Df   Sum Sq  Mean Sq F value    Pr(>F)
Feed       2 0.132358 0.066179  11.492 0.0004254 ***
Residuals 21 0.120938 0.005759
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is less than 0.05, we can conclude that the drug added to the feed has a significant influence on the weight of chicken. The same result we get also with the distribution-free Kruskal-Wallis test. However, this test provides a larger p-value.

```
> kruskal.test(Weight~Feed,data=chicken)

        Kruskal-Wallis rank sum test

data:  Weight by Feed
Kruskal-Wallis chi-squared = 12.0522, df = 2, p-value = 0.002415
```

To produce the complete ANOVA table as above we calculate the general variance and the grand sum of squares:

```
> var(chicken$Weight)
[1] 0.01101286
> 23*var(chicken$Weight)
[1] 0.2532958
```

Then we obtain the following ANOVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Differences between factor levels	2	$\Sigma_{SST} = 0.132358$	$\widehat{\sigma}^2_{SST} = 0.066179$
Measurement error	21	$\Sigma_{SSE} = 0.120938$	$\widehat{\sigma}^2_{SSE} = 0.005759$
Total	23	$\Sigma_{SSG} = 0.253296$	$\widehat{\sigma}^2_{SSG} = 0.01101286$

We see that indeed $0.132358 + 0.120938 = 0.253296$ is satisfied. Moreover, the column `Mean Sq` in the ANOVA table of R containes the variance estimates $\widehat{\sigma}^2_{SST}$ and $\widehat{\sigma}^2_{SSE}$.

3.2.2 Example (Germinating seeds)
The data in the file `GERMIN.DAT` "came from an experiment to study the effect of different amounts of water on the germination of seeds. For each amount of water, four identical boxes were sown with 100 seeds each, and the number of seeds having germinated after two weeks was recordered. The experiment was repeated with boxes covered to slow evaporation. There were six levels of watering, coded 1 to 6, with higher codes corresponding to more water."(Hand et al. 1996, P.1)

Here the number of seeds germinating per box:

Uncovered boxes						Covered boxes					
Amount of water						Amount of water					
1	2	3	4	5	6	1	2	3	4	5	6
22	41	66	82	79	0	45	65	81	55	31	0
25	46	72	73	68	0	41	80	73	51	36	0
27	59	51	73	74	0	42	79	74	40	45	0
23	38	78	84	70	0	43	77	76	62	*	0

Here we analyse only the results for the uncovered boxes. The treatments are here the 6 levels of watering. We want to test whether there is a effect of watering in the uncovered boxes. Hence the null hypothesis is that there is no effect of watering on the seed numbers.

To apply the function `lm`, the data set must have the form where different units belong to different rows. To transform the data from the data file `GERMIN.DAT` into this form, the following function `Germin.funct` was written:

```
function ()
{
germin0<-read.table("GERMIN.DAT",na.string="*")
germin1<-data.frame(germin0[1:4,],"uncovered")
germin2<-data.frame(germin0[5:8,],"covered")
germin11<-cbind(germin1[,c(1,7)],"1")
germin12<-cbind(germin1[,c(2,7)],"2")
germin13<-cbind(germin1[,c(3,7)],"3")
germin14<-cbind(germin1[,c(4,7)],"4")
germin15<-cbind(germin1[,c(5,7)],"5")
germin16<-cbind(germin1[,c(6,7)],"6")
germin21<-cbind(germin2[,c(1,7)],"1")
germin22<-cbind(germin2[,c(2,7)],"2")
germin23<-cbind(germin2[,c(3,7)],"3")
germin24<-cbind(germin2[,c(4,7)],"4")
germin25<-cbind(germin2[-4,c(5,7)],"5")
germin26<-cbind(germin2[,c(6,7)],"6")
names(germin11)<-c("seed.numbers","box","watering")
names(germin11)<-c("seed.numbers","box","watering")
names(germin12)<-c("seed.numbers","box","watering")
names(germin13)<-c("seed.numbers","box","watering")
names(germin14)<-c("seed.numbers","box","watering")
```

```
names(germin15)<-c("seed.numbers","box","watering")
names(germin16)<-c("seed.numbers","box","watering")
names(germin21)<-c("seed.numbers","box","watering")
names(germin22)<-c("seed.numbers","box","watering")
names(germin23)<-c("seed.numbers","box","watering")
names(germin24)<-c("seed.numbers","box","watering")
names(germin25)<-c("seed.numbers","box","watering")
names(germin26)<-c("seed.numbers","box","watering")
germin<-rbind(germin11,germin12,germin13,germin14,germin15,germin16,
germin21,germin22,germin23,germin24,germin25,germin26)
row.names(germin)<-1:47
germin
}
```

Since each group has only 4 measurement it is very unlikely that the normal distribution is rejected in the groups 1,2,3,4,5. The group 6 should be dropped from the data set since there the variance differs clearly from the variances in the other groups. Since this variance is 0, it even would produce a break down of R if the Bartlett test is used.

```
> germin<-germin.funct()
> germin.unc<-germin[germin$box=="uncovered",]
> bartlett.test(seed.numbers~watering,data=germin.unc[germin.unc$watering!="6",])

        Bartlett test for homogeneity of variances

data:  seed.numbers by watering
Bartlett's K-squared = 6.7289, df = 4, p-value = 0.1509
```

Hence the requirements of the ANOVA test are not rejected so that the ANOVA table for the uncovered boxes can be produced:

```
 > anova(lm(seed.numbers~watering,data=germin.unc[germin.unc$watering!="6",]))
Analysis of Variance Table

Response: seed.numbers
          Df Sum Sq Mean Sq F value    Pr(>F)
watering   4 7904.7  1976.2  34.946 1.943e-07 ***
Residuals 15  848.2    56.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With

```
> var(germin.unc[germin.unc$watering!="6","seed.numbers"])
[1] 460.6816
> 19*var(germin.unc[germin.unc$watering!="6","seed.numbers"])
[1] 8752.95
```

we obtain the complete ANOVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Differences between factor levels	4	$\Sigma_{SST} = 7904.7$	$\widehat{\sigma}^2_{SST} = 1976.2$
Measurement error	15	$\Sigma_{SSE} = 848.2$	$\widehat{\sigma}^2_{SSE} = 56.5$
Total	19	$\Sigma_{SSG} = 8752.9$	$\widehat{\sigma}^2_{SSG} = 460.6816$

Since we obtain the very small p-value of 1.943e-07, the null hypothesis is rejected. Hence there is a significant effect of watering in the uncovered boxes.

3.2.3 Exercise (Germinating seeds: Covered boxes)
Test whether there is also a watering effect for the seed numbers in the covered boxes in Example 3.2.2. Check the requirements of the ANOVA test and produce the complete ANOVA table.

3.2.4 Exercise (Trees)
Test with the ANOVA test as well as with the Kruskal-Wallis test whether the stem diameters of trees differ significantly between species in the data set `trees` in the library `agricolae`. Check the requirements of the ANOVA test.

3.3 Multiple comparisons

If the ANOVA test rejects the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$, then one would like to know which μ_i are different. The naive procedure is to perform all pairwise t-tests for $H_0 : \mu_i = \mu_j$ with $i \neq j$. But then $I(I-1)/2$ tests are performed at the same data set so that the Bonferroni adjustment for the levels of the t-test should be used. This means that $\alpha = \frac{2\,0.05}{I\,(I-1)}$ is used for the level of the t-tests. Let here

$$\widehat{d}(y_{i*}, y_{j*}) = \sqrt{\frac{N_i\,N_j}{N_i + N_j}}\;\frac{\overline{y}_{i\bullet} - \overline{y}_{j\bullet}}{\widehat{\sigma}_{ij}}$$

denote the t-test statistic for testing $H_0 : \mu_i = \mu_j$ based on y_{i*} and y_{j*}.

Multiple comparisons with the Bonferroni adjustment

If there are I levels (groups), then for $1 \leq i < j \leq I$:

$$\text{Reject } H_0 : \mu_i = \mu_j \quad \text{if} \quad \begin{cases} \text{P-value of the t-test is not greater than } \frac{2\,\alpha}{I\,(I-1)}, & \text{or} \\ |\widehat{d}(y_{i*}, y_{j*})| > t_{N_1+N_2-2,1-\frac{\alpha}{I\,(I-1)}}, & \text{respectively.} \end{cases}$$

However, the comparisons with the Bonferroni adjustment are very conservative, i.e. the β-error is high. A better method is Tukey's Honest Significant Difference (HSD) method. It uses the fact

that the maximum relative range

$$\widehat{R} = \frac{\max_{n=1,\ldots,N} Y_n - \min_{n=1,\ldots,N} Y_n}{\widehat{\sigma}}$$

has a studentized range distribution $q_{N,v}$ if $Y_1, \ldots, Y_N$ are independent and identically distributed with normal distribution $\mathcal{N}(\mu, \sigma)$ and $\widehat{\sigma}^2$ is independently distributed as χ-squared with v degrees of freedom.

Multiple comparisons with the Tukey's Honest Significant Difference method

If there are I levels (groups), then for $1 \leq i < j \leq I$:

$$\text{Reject } H_0 : \mu_i = \mu_j \quad \text{if} \quad \sqrt{2}\sqrt{\frac{N_i\, N_j}{N_i + N_j}}\,\frac{|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}|}{\widehat{\sigma}_{SSE}} > q_{I,N-I,1-\alpha}$$

Thereby $q_{N,v,\alpha}$ denotes the α-quantile of the studentized range distribution $q_{N,v}$. These quantiles are given in R by `qtukey`. R offers also a function for making the comparisons. This function is called `TukeyHSD` and bases on the results of the function `aov`, which produces a reduced ANOVA table. Besides $\overline{y}_{i\bullet} - \overline{y}_{j\bullet}$ and the p-values for the pairwise comparisons, `TukeyHSD` provides also the lower and upper bounds of the simultaneous confidence intervals for the mean differences $\mu_i - \mu_j$, i.e. it calculates

$$\overline{y}_{i\bullet} - \overline{y}_{j\bullet} \pm q_{I,N-I,1-\alpha}\frac{\widehat{\sigma}_{SSE}}{\sqrt{2}}\sqrt{\frac{1}{N_i} + \frac{1}{N_j}}$$

for all $1 \leq i < j \leq I$.

When the sample sizes N_i are very unequal, Tukey's HSD can be also very conservative. Therefore, the R package `agricolae` provides several other multiple comparison method as the **Waller-Duncan** method given in `waller.test`. However, to run them without errors, a newer version of the `agricolae` package must be used, namely `agricolae_1.0-6.zip` from `http://tarwi.lamolina.edu.pe/~fmendiburu/`.

3.3.1 Example (Germinating seeds: Uncovered boxes)
To see which watering levels provides really different seed numbers we use Tukey's Honest Significant Difference method. Before using this, we compare at first the commands `anova(lm...)` and `aov(...)`:

```
> germin.unc<-germin[germin$box=="uncovered",]
> anova(lm(seed.numbers~watering,data=germin.unc[germin.unc$watering!="6",]))
Analysis of Variance Table

Response: seed.numbers
          Df Sum Sq Mean Sq F value    Pr(>F)
watering   4 7904.7  1976.2  34.946 1.943e-07 ***
Residuals 15  848.2    56.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> aov(seed.numbers~watering,data=germin.unc[germin.unc$watering!="6",])
```

```
Call:
   aov(formula = seed.numbers ~ watering, data = germin.unc[germin.unc$watering !=
    "6", ])

Terms:
                watering Residuals
Sum of Squares   7904.70    848.25
Deg. of Freedom        4        15

Residual standard error: 7.519973
Estimated effects may be unbalanced
```

We see that indeed `aov` provides a reduced ANOVA table.

```
> germin.unc<-germin[germin$box=="uncovered",]
> TukeyHSD(aov(seed.numbers~watering,data=germin.unc[germin.unc$watering!="6",]))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = seed.numbers ~ watering,
   data = germin.unc[germin.unc$watering != "6", ])

$watering
     diff        lwr      upr     p adj
2-1 21.75   5.330196 38.16980 0.0073044
3-1 42.50  26.080196 58.91980 0.0000075
4-1 53.75  37.330196 70.16980 0.0000004
5-1 48.50  32.080196 64.91980 0.0000014
3-2 20.75   4.330196 37.16980 0.0105229
4-2 32.00  15.580196 48.41980 0.0001967
5-2 26.75  10.330196 43.16980 0.0012013
4-3 11.25  -5.169804 27.66980 0.2637507
5-3  6.00 -10.419804 22.41980 0.7894955
5-4 -5.25 -21.669804 11.16980 0.8569113
```

We can conclude that almost all pairs of watering levels produce different seed numbers. The only exceptions are the levels 4-3, 5-3, and 5-4. This result corresponds also to the boxplots:

```
> boxplot(seed.numbers~watering,data=germin.unc[germin.unc$watering!="6",])
```

Now we compare the results of Tukey's Honest Significant Difference method with the Bonferroni adjustment method:

```
> gerS.unc<-germin.unc[,"seed.numbers"]
> gerW.unc<-germin.unc[,"watering"]
> t.test(gerS.unc[gerW.unc=="1"],gerS.unc[gerW.unc=="2"])$p.value
[1] 0.01557484
```

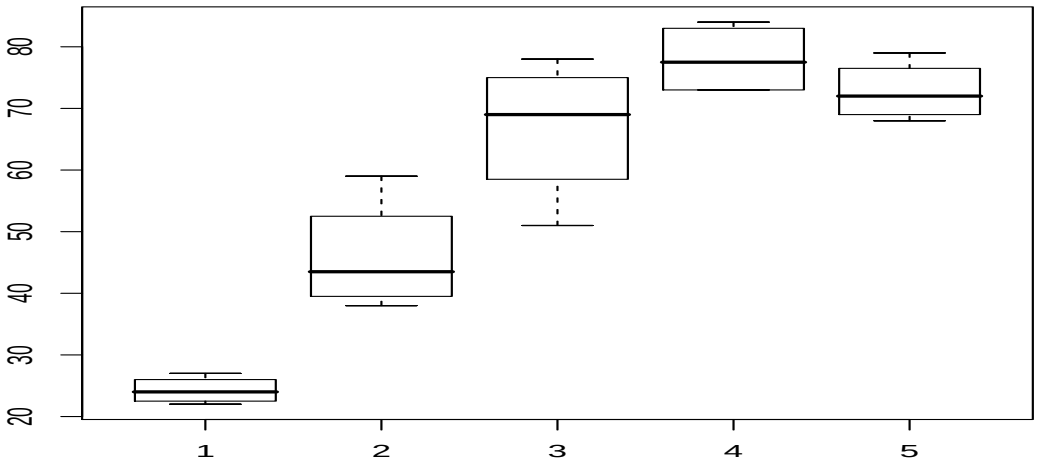

Figure 3.1: Box plots for the uncovered boxes

```
> t.test(gerS.unc[gerW.unc=="1"],gerS.unc[gerW.unc=="3"])$p.value
[1] 0.004334964
> t.test(gerS.unc[gerW.unc=="1"],gerS.unc[gerW.unc=="4"])$p.value
[1] 8.701039e-05
> t.test(gerS.unc[gerW.unc=="1"],gerS.unc[gerW.unc=="5"])$p.value
[1] 3.762982e-05
> t.test(gerS.unc[gerW.unc=="2"],gerS.unc[gerW.unc=="3"])$p.value
[1] 0.03292381
> t.test(gerS.unc[gerW.unc=="2"],gerS.unc[gerW.unc=="4"])$p.value
[1] 0.002008712
> t.test(gerS.unc[gerW.unc=="2"],gerS.unc[gerW.unc=="5"])$p.value
[1] 0.004932404
> t.test(gerS.unc[gerW.unc=="3"],gerS.unc[gerW.unc=="4"])$p.value
[1] 0.1508645
> t.test(gerS.unc[gerW.unc=="3"],gerS.unc[gerW.unc=="5"])$p.value
[1] 0.393299
> t.test(gerS.unc[gerW.unc=="4"],gerS.unc[gerW.unc=="5"])$p.value
[1] 0.2172643
```

Since the adjusted significance level is $0.05/10 = 0.005$, we obtain only significant differences for the watering combinations 1-3, 1-4, 1-5, 2-4, 2-5. Hence the differences for 1-2, 2-3, which appeared significant with Tukey's Honest Significant Difference method, are not any more significant. This shows that the Bonferroni method is more conservative.

We can use also `HSD.test` and `waller.test` of the newest `agricolae` package:

```
> library(agricolae)
> attach(germin.unc[germin.unc$watering!="6",])
> model<-aov(seed.numbers~watering)
> df<-df.residual(model)
> MSerror<-deviance(model)/df
> comparison<-HSD.test(seed.numbers,watering,df,MSerror,group=TRUE,main="title")
```

```
> comparison
  trt means   M N  std.err
1   4 78.00   a 4 2.915476
2   5 72.75   a 4 2.428134
3   3 66.75   a 4 5.793315
4   2 46.00   b 4 4.636809
5   1 24.25   c 4 1.108678
```

The lengthy printout of `HSD.test` is dropped here. Since `HSD.test` uses Tukey's Honest Significant Difference, it provides the same results as `TukeyHSD` but in a different form: different letters in the column `M` indicate significantly different means. Hence 3-4, 3-5, and 4-5 do not concern different means. The Waller-Duncan method provides even more significantly different means:

```
> comparison<-waller.test(seed.numbers,watering,df,MSerror,
+ Fc,group=TRUE,main="title")
> comparison
  trt means    M N  std.err
1   4 78.00    a 4 2.915476
2   5 72.75   ab 4 2.428134
3   3 66.75    b 4 5.793315
4   2 46.00    c 4 4.636809
5   1 24.25    d 4 1.108678
```

Here additionally 3-4 shows significantly different means.

3.3.2 Exercise (Germinating seeds: Covered boxes)
Find out for the covered boxes which watering levels provides significantly different seed numbers. Use Tukey's Honest Significant Difference method and the Bonferroni adjustment method and compare the results. Use also `HSD.test` and `waller.test` from the newest version of the `agricolae` package. Visualize the results with boxplots.

3.3.3 Exercise (Trees)
Find out with the data set `trees` from the `agricolae` package which species have significantly different stem diameters. Visualize the result with boxplots.

3.4 Designing the one-way ANOVA

Like for the t-test, the smallest β-error is achieved if the samples sizes N_1, $N_2, \ldots, N_I$ are equal or as equal as possible. Moreover, the allocations of the levels (treatments) to the experimental units should be done randomly. Such designs are called **completely randomized designs for one factor** and can be created, for example, with the function `design.crd` of the `agricolae` package.

3.4.1 Exercise (One-way ANOVA design)
Determine a good design for 3 treatments `TR1`, `TR2`, `TR3`, `TR4` applied at 12 experimental units.

4 Two-way ANOVA

Here it is assumed that the data set contains three variables: one numeric variable concerning measurements and two factor variables A and B concerning treatments or blocks (groups). The factor variable A has A levels and the factor variable B has B levels. Let N_{ab} denote the sample size for factor level combination (a, b). Because of $A\,B$ level combinations, we have the following table for the sample sizes:

		B			
		1	2	...	B
A	1	N_{11}	N_{12}	...	N_{1B}
	2	N_{21}	N_{22}	...	N_{2B}
	⋮	⋮	⋮		⋮
	A	N_{A1}	N_{A2}	...	N_{AB}

Table 1: Numbers of repetitions for each level combination

In **balanced designs** the sample sizes are all equal, i.e. $N_{ab} = M$ for $a = 1, \ldots, A$ and $b = 1, \ldots, B$. But **unbalanced designs** with different sample sizes are also considered. In particular, same of the sample sizes N_{ab} can be zero.

In the general case, we have the following measurements:

$$
\begin{aligned}
&y_{11*} = (y_{111}, \ldots, y_{11N_{11}})^\top \text{ the vector of observations for level combination } (1,1),\\
&y_{12*} = (y_{121}, \ldots, y_{12N_{12}})^\top \text{ the vector of observations for level combination } (1,2),\\
&\vdots\\
&y_{1B*} = (y_{1B1}, \ldots, y_{1BN_{1B}})^\top \text{ the vector of observations for level combination } (1,B),\\
&y_{21*} = (y_{211}, \ldots, y_{21N_{21}})^\top \text{ the vector of observations for level combination } (2,1),\\
&\vdots\\
&y_{2B*} = (y_{2B1}, \ldots, y_{2BN_{2B}})^\top \text{ the vector of observations for level combination } (2,B),\\
&\vdots\\
&y_{A1*} = (y_{A11}, \ldots, y_{A1N_{A1}})^\top \text{ the vector of observations for level combination } (A,1),\\
&\vdots\\
&y_{AB*} = (y_{AB1}, \ldots, y_{ABN_{AB}})^\top \text{ the vector of observations for level combination } (A,B).
\end{aligned}
$$

Altogether there are $N = N_{11} + N_{12} + \ldots + N_{1B} + N_{21} + \ldots + N_{2B} + \ldots + N_{A1} + \ldots + N_{AB}$ observations. $y_{11*}, \ldots, y_{AB*}$ are realizations of independent random vectors $Y_{11*}, \ldots, Y_{AB*}$ where $Y_{ab*} = (Y_{ab1}, \ldots, Y_{abN_{ab}})^\top$ for $a = 1, \ldots, A$ and $b = 1, \ldots, B$.

The ANOVA test assumes that y_{abn} is a realization of a random variable Y_{abn} with normal distribution $\mathcal{N}(\mu_{ab}, \sigma^2)$ and that all measurement variables Y_{abn} are stochastically independent for $n = 1, \ldots, N_{ab}$, $a = 1, \ldots, A$, $b = 1, \ldots, B$. Note that all variables Y_{abn} have the same variance as

it should be for the one-way ANOVA.

4.1 Model with interactions

If $N_{ab} \leq 2$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$, then the following model for the distribution of the measurements variables can be used:

$$Y_{abn} = \mu_{ab} + Z_{in} = \mu + \alpha_a + \beta_b + \gamma_{ab} + Z_{abn} \quad \text{with} \quad Z_{abn} \sim \mathcal{N}(0, \sigma^2),$$

where

1. Z_{abn} is the measurement error,
2. μ the **average mean**,
3. α_a the **main effect** of level a of factor A,
4. β_b the **main effect** of level b of factor B,
5. γ_{ab} the **interaction** between the levels a and b of the factors A and B.

If the parameters μ, $\alpha_1, \ldots, \alpha_A$, $\beta_1, \ldots, \beta_B$, $\gamma_{11}, \ldots, \gamma_{1B}, \ldots, \gamma_{AB}$ can attain arbitrary values, then the model is overparametrized since there are $1 + A + B + AB$ different parameters while there are only AB level combinations. Hence side conditions for the parameters are needed, which are:

$$\sum_{a=1}^{A} \alpha_a = 0, \quad \sum_{b=1}^{B} \beta_b = 0, \quad \sum_{a=1}^{A} \gamma_{ab} = 0 \text{ for all } b = 1, \ldots, B, \quad \sum_{b=1}^{B} \gamma_{ab} = 0 \text{ for all } a = 1, \ldots, A. \quad (2)$$

Let be $N_{a\bullet} = \sum_{b=1}^{B} N_{ab}$ and $N_{\bullet b} = \sum_{a=1}^{A} N_{ab}$. The parameters have the following estimates:

$$\begin{aligned}
\mu: \quad & \widehat{\mu} = \overline{y}_{\bullet\bullet\bullet} := \frac{1}{N}\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} y_{abn}, \\
\mu_{ab}: \quad & \widehat{\mu}_{ab} = \overline{y}_{ab\bullet} := \frac{1}{N_{ab}}\sum_{n=1}^{N_{ab}} y_{abn}, \\
\alpha_a: \quad & \widehat{\alpha}_a = \frac{1}{N_{a\bullet}}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} y_{abn} - \overline{y}_{\bullet\bullet\bullet} = \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet}, \\
\beta_b: \quad & \widehat{\beta}_b = \frac{1}{N_{\bullet b}}\sum_{a=1}^{A}\sum_{n=1}^{N_{ab}} y_{abn} - \overline{y}_{\bullet\bullet\bullet} = \overline{y}_{\bullet b\bullet} - \overline{y}_{\bullet\bullet\bullet}, \\
\gamma_{ab} = \mu_{ab} - \mu - \alpha_a - \beta_b: \quad & \widehat{\gamma}_{ab} = \widehat{\mu}_{ab} - \widehat{\mu} - \widehat{\alpha}_a - \widehat{\beta}_b = \overline{y}_{ab\bullet} - \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet b\bullet} + \overline{y}_{\bullet\bullet\bullet}, \\
z_{abn} = y_{abn} - \mu_{ab}: \quad & \widehat{z}_{abn} = y_{abn} - \overline{y}_{ab\bullet}, \\
\sigma^2: \quad & \widehat{\sigma}^2_{SSE} = \frac{1}{N - A\,B}\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} \widehat{z}^2_{abn} \\
& = \frac{1}{N - A\,B}\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} (y_{abn} - \overline{y}_{ab\bullet})^2.
\end{aligned} \tag{3}$$

Note that the estimates $\widehat{\alpha}_a$, $\widehat{\beta}_b$, and $\widehat{\gamma}_{ab}$ satisfy the side conditions 2.

4.2 Model without interactions

If some of the N_{ab} are zero or equal to 1, then the model with interactions cannot be used. Then a model without interaction shall be used:

$$Y_{abn} = \mu_{ab} + Z_{in} = \mu + \alpha_a + \beta_b + Z_{abn} \quad \text{with} \quad Z_{abn} \sim \mathcal{N}(0, \sigma^2),$$

where μ, α_a and β_b have the same interpretations as before and also the same estimates. Only the estimate for the variance changes:

$$\begin{aligned}
z_{abn} = y_{abn} - \mu - \alpha_a - \beta_b: \quad & \widehat{z}_{abn} = y_{abn} - \widehat{\mu} - \widehat{\alpha}_a - \widehat{\beta}_b, \\
\sigma^2: \quad & \widehat{\sigma}^2_{SSE|A+B} = \frac{1}{N - A\,B}\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} \widehat{z}^2_{abn} \\
& = \frac{1}{N - A - B + 1}\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} (y_{abn} - \widehat{\mu} - \widehat{\alpha}_a - \widehat{\beta}_b)^2.
\end{aligned} \tag{4}$$

4.3 ANOVA tests

Hypotheses
In the model with interactions, there are three hypotheses, which should be tested:

$$\begin{aligned} H_0^I &: \gamma_{ab} = 0 \text{ for all } (a,b) \quad \text{versus} \quad H_1^I : \text{ there exists } (a,b) \text{ with } \gamma_{ab} \neq 0, \\ H_0^A &: \alpha_a = 0 \text{ for all } a \quad \text{versus} \quad H_1^A : \text{ there exists } a \text{ with } \alpha_a \neq 0, \\ H_0^B &: \beta_b = 0 \text{ for all } b \quad \text{versus} \quad H_1^B : \text{ there exists } b \text{ with } \beta_b \neq 0. \end{aligned}$$

The tests are derived by regarding modified hypotheses which can be tested sequentially:

$$\begin{aligned} \widetilde{H}_0^{A+B} &: \mu_{ab} = \mu + \alpha_a + \beta_b \quad \text{versus} \quad \widetilde{H}_1^{A+B} : \mu_{ab} = \mu + \alpha_a + \beta_b + \gamma_{ab}, \\ \widetilde{H}_0^{A|A+B} &: \mu_{ab} = \mu + \beta_b \quad \text{versus} \quad \widetilde{H}_1^{A|A+B} : \mu_{ab} = \mu + \alpha_a + \beta_b, \\ \widetilde{H}_0^{B} &: \mu_{ab} = \mu \quad \text{versus} \quad \widetilde{H}_1^{B} : \mu_{ab} = \mu + \beta_b. \end{aligned}$$

The hypotheses $\widetilde{H}_0^{A+B}$, $\widetilde{H}_0^{A|A+B}$, and $\widetilde{H}_0^{B}$ can be tested separately by

$$\begin{aligned} \text{Reject } \widetilde{H}_0^{A+B} \quad &\text{if} \quad \frac{\Sigma_{SSI}/(A-1)(B-1)}{\Sigma_{SSE}/(N-A\,B)} > F_{(A-1)(B-1),N-A\,B,1-\alpha}. \\ \text{Reject } \widetilde{H}_0^{A|A+B} \quad &\text{if} \quad \frac{\Sigma_{SSA|A+B}/(A-1)}{\Sigma_{SSE|A+B}/(N-A-B+1)} > F_{A-1,N-A-B+1,1-\alpha}. \\ \text{Reject } \widetilde{H}_0^{B} \quad &\text{if} \quad \frac{\Sigma_{SSB}/(B-1)}{\Sigma_{SSE|B}/(N-B)} > F_{B-1,N-B,1-\alpha}. \end{aligned} \tag{5}$$

Thereby $F_{N,M,\alpha}$ denotes again the α-quantile of the central F-distribution with N and M degrees of freedom. The constructions of the tests bases on the fact that the used sum of squares are stochastically independent with χ^2-distribution.

Note, that the last test is the ANOVA test for the one-way layout. Thus, we have

$$\begin{aligned} \Sigma_{SSB} &= \sum_{b=1}^{B} N_{\bullet b}\,(\overline{y}_{\bullet b\bullet} - \overline{y}_{\bullet\bullet\bullet})^2 \qquad \text{is the sum of squares for factor B,} \\ \Sigma_{SSE|B} &= \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}} (\overline{y}_{abn} - \overline{y}_{\bullet b\bullet})^2 \qquad \text{is the sum of squares for errors in the one-way layout.} \end{aligned}$$

The other sum of squares are defined as follows:

$$\Sigma_{SSA|A+B} \quad \text{is the sum of squares for factor A in model A+B,}$$

$$\Sigma_{SSE|A+B} = \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}}(y_{abn} - \widehat{\mu} - \widehat{\alpha}_a - \widehat{\beta}_b)^2 \quad \text{is the sum of squares for errors in model A+B,}$$

$$\Sigma_{SSI} = \sum_{a=1}^{A}\sum_{b=1}^{B} N_{ab}\,(\overline{y}_{ab\bullet} - \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet b\bullet} + \overline{y}_{\bullet\bullet\bullet})^2 \quad \text{is the sum of squares for interactions,}$$

$$\Sigma_{SSE} = \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}}(y_{abn} - \overline{y}_{ab\bullet})^2 \quad \text{is the sum of squares for errors in the full model.}$$

The grand sum of squares

$$\Sigma_{SSG} := \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}}(y_{abn} - \overline{y}_{\bullet\bullet\bullet})^2.$$

can be decomposed as follows

$$\begin{aligned}
\Sigma_{SSG} &= \Sigma_{SSB} + \Sigma_{SSE|B},\\
\Sigma_{SSG} &= \Sigma_{SSB} + \Sigma_{SSA|A+B} + \Sigma_{SSE|A+B},\\
\Sigma_{SSG} &= \Sigma_{SSB} + \Sigma_{SSA|A+B} + \Sigma_{SSI} + \Sigma_{SSE},
\end{aligned}$$

where

Σ_{SSB}, $\Sigma_{SSE|B}$ are stochastically independent,
Σ_{SSB}, $\Sigma_{SSA|A+B}$, $\Sigma_{SSE|A+B}$ are stochastically independent,
Σ_{SSB}, $\Sigma_{SSA|A+B}$, Σ_{SSI}, Σ_{SSE} are stochastically independent.

Note that there is no simple expression for $\Sigma_{SSA|A+B}$. Moreover, the tests depend on the order of the factors. We obtain other tests if the factors A and B are exchanged. This means in practice that we have to decide which factor should take the role of factor A and which factor the role of factor B.

Balanced designs
However, for balanced designs with $N_{ab} = M$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$, we do not have this problem. There it holds

$$\begin{aligned}
\Sigma_{SSB} &= M\,A\,\sum_{b=1}^{B}(\overline{y}_{\bullet b\bullet} - \overline{y}_{\bullet\bullet\bullet})^2,\\
\Sigma_{SSA|A+B} = \Sigma_{SSA} &= M\,B\,\sum_{a=1}^{A}(\overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2,\\
\Sigma_{SSI} &= M\,\sum_{a=1}^{A}\sum_{b=1}^{B}(\overline{y}_{ab\bullet} - \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet b\bullet} + \overline{y}_{\bullet\bullet\bullet})^2,
\end{aligned}$$

so that the decomposition of the grand sum of squares becomes

$$\Sigma_{SSG} = \Sigma_{SSA} + \Sigma_{SSB} + \Sigma_{SSE|A+B} \quad (6)$$
$$\Sigma_{SSG} = \Sigma_{SSA} + \Sigma_{SSB} + \Sigma_{SSI} + \Sigma_{SSE}. \quad (7)$$

Since the sum of squares for factors A and B are of analogous form, the factors A and B can be exchanged.

4.3.1 Exercise (For mathematicians)
Prove that the decompositions (6) and (7) really hold for balanced designs.

ANOVA test for the two-way layout with interactions and general design

Using the tests in (5), the order of factors A and B would be also important for balanced designs since the sum of squares for errors $\Sigma_{SSE|A+B}$ and $\Sigma_{SSE|B}$ are different. Therefore, always Σ_{SSE} is used in the denominator of the ANOVA test statistic. Moreover, only then it is not necessary to adjust the level α of the tests. I.e. we can use for all three tests $\alpha = 0.05$ if these are the only tests at the data set except for some pretests. This is due to the decomposition of the grand sum of squares. The following variance estimators are used in the test statistics besides $\widehat{\sigma}^2_{SSE}$ given in (3):

$$\widehat{\sigma}^2_{SSB} = \frac{1}{B-1}\Sigma_{SSB}, \quad \widehat{\sigma}^2_{SSA|A+B} = \frac{1}{A-1}\Sigma_{SSA|A+B}, \quad \widehat{\sigma}^2_{SSI} = \frac{1}{(A-1)(B-1)}\Sigma_{SSI}.$$

This leads to the following tests:

ANOVA tests for the two-way layout with interactions

Reject $H_0^I : \gamma_{ab} = 0$ for all (a,b) if $\widehat{V}_I = \dfrac{\widehat{\sigma}^2_{SSI}}{\widehat{\sigma}^2_{SSE}} > F_{(A-1)(B-1),N-AB,1-\alpha}$.

Reject $H_0^A : \alpha_a = 0$ for all a if $\widehat{V}_A = \dfrac{\widehat{\sigma}^2_{SSA|A+B}}{\widehat{\sigma}^2_{SSE}} > F_{A-1,N-AB,1-\alpha}$.

Reject $H_0^B : \beta_b = 0$ for all b if $\widehat{V}_B = \dfrac{\widehat{\sigma}^2_{SSB}}{\widehat{\sigma}^2_{SSE}} > F_{B-1,N-AB,1-\alpha}$.

Again, the values for the analysis of variance are summarized in the so-called ANOVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Factor B	$B-1$	Σ_{SSB}	$\widehat{\sigma}^2_{SSB} = \frac{1}{B-1}\Sigma_{SSB}$
Factor A	$A-1$	$\Sigma_{SSA\|A+B}$	$\widehat{\sigma}^2_{SSA\|A+B} = \frac{1}{A-1}\Sigma_{SSA\|A+B}$
Interaction	$(A-1)(B-1)$	Σ_{SSI}	$\widehat{\sigma}^2_{SSI} = \frac{1}{(A-1)(B-1)}\Sigma_{SSI}$
Measurement error	$N-A\,B$	Σ_{SSE}	$\widehat{\sigma}^2_{SSE} = \frac{1}{N-A\,B}\Sigma_{SSE}$
Total	$N-1$	Σ_{SSG}	$\widehat{\sigma}^2_{SSG} = \frac{1}{N-1}\Sigma_{SSG}$

ANOVA test for the two-way layout without interactions and with general design

Here $\Sigma_{SSE|A+B}$ is used in the denominator of the test statistics since then we have again the decomposition of the grand sum of squares.

ANOVA tests for the two-way layout without interactions

Reject $H_0^A : \alpha_a = 0$ for all a if $\widehat{V}_A = \dfrac{\widehat{\sigma}^2_{SSA|A+B}}{\widehat{\sigma}^2_{SSE|A+B}} > F_{A-1,N-A-B+1,1-\alpha}$.

Reject $H_0^B : \beta_b = 0$ for all b if $\widehat{V}_B = \dfrac{\widehat{\sigma}^2_{SSB}}{\widehat{\sigma}^2_{SSE|A+B}} > F_{B-1,N-A-B+1,1-\alpha}$.

Again, the values for the analysis of variance are summarized in the so-called ANOVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Factor B	$B-1$	Σ_{SSB}	$\widehat{\sigma}^2_{SSB} = \frac{1}{B-1}\Sigma_{SSB}$
Factor A	$A-1$	$\Sigma_{SSA\|A+B}$	$\widehat{\sigma}^2_{SSA\|A+B} = \frac{1}{A-1}\Sigma_{SSA\|A+B}$
Measurement error	$N-A-B+1$	$\Sigma_{SSE\|A+B}$	$\widehat{\sigma}^2_{SSE\|A+B} = \frac{1}{N-A-B+1}\Sigma_{SSE\|A+B}$
Total	$N-1$	Σ_{SSG}	$\widehat{\sigma}^2_{SSG} = \frac{1}{N-1}\Sigma_{SSG}$

Checking the requirements of the two-way ANOVA tests
As for the one-way ANOVA test the normal distribution for each factor combination and the homogeneity of the variances must be checked. However this can be done only if there are many measurements for each factor combination. In such situations, the same methods as for the one-way layout can be used by regarding each level combination separately. But in designs with few or even zero observations for some factor combinations this makes no sense. The normal distribution, however, can be always tested by testing the normal distribution of the residuals with `shapiro.test(lm(...)$residuals`. This method should be also used if there are many levels in the one-way layout.

4.3.2 Example (Germinating seeds)
We have seen in Example 3.2.2 and Exercise 3.2.3 that there is a significant watering effect for the uncovered boxes as well as for the covered boxes. To test whether there is also a box effect, we could use the t-test. But then we have to take into account that we are doing three tests at the same data set `germin`. Therefore, it is better to use the ANOVA test. With this test we also can test whether there is an interaction between the type of the box and the watering level. But at first we test whether we can assume the normal distribution and the homogeneity of the variances:

```
> shapiro.test(lm(seed.numbers~watering*box,data=germin[germin$watering!="6",]
+ )$residuals)$p.value
[1] 0.9119584
```

Hence there is no evidence that the normal distribution is not satisfied.

```
> anova(lm(seed.numbers~watering*box,data=germin[germin$watering!="6",]))
Analysis of Variance Table

Response: seed.numbers
             Df Sum Sq Mean Sq F value    Pr(>F)
watering      4 6675.8  1668.9 34.7094 1.142e-10 ***
box           1    0.2     0.2  0.0035     0.953
watering:box  4 6068.9  1517.2 31.5540 3.492e-10 ***
Residuals    29 1394.4    48.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can conclude that there is a significant watering effect and a significant interaction. The type of the box has no significant effect. With

```
> var(germin[germin$watering!="6","seed.numbers"])
[1] 372.085
> 38*var(germin[germin$watering!="6","seed.numbers"])
[1] 14139.23
```

we can produce the complete ANOVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Factor B	4	$\Sigma_{SSB} = 6675.8$	$\widehat{\sigma}^2_{SSB} = 1668.9$
Factor A	1	$\Sigma_{SSA\|A+B} = 0.2$	$\widehat{\sigma}^2_{SSA\|A+B} = 0.2$
Interaction	4	$\Sigma_{SSI} = 6068.9$	$\widehat{\sigma}^2_{SSI} = 1517.2$
Measurement error	29	$\Sigma_{SSE} = 1394.4$	$\widehat{\sigma}^2_{SSE} = 48.1$
Total	38	$\Sigma_{SSG} = 14139.23$	$\widehat{\sigma}^2_{SSG} = 372.085$

If the order of the factors `watering` and `box` are exchanged, then almost the same table is obtained. This is due to the fact that we have almost a balanced design with $N_{ab} = 4$. Only one observation is missing in the covered boxes for watering level 5:

```
> anova(lm(seed.numbers~box*watering,data=germin[germin$watering!="6",]))
Analysis of Variance Table

Response: seed.numbers
             Df Sum Sq Mean Sq F value    Pr(>F)
box           1    0.2     0.2  0.0037    0.9522
watering      4 6675.8  1668.9 34.7093 1.142e-10 ***
box:watering  4 6068.9  1517.2 31.5540 3.492e-10 ***
Residuals    29 1394.4    48.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Although there is a significant interaction and $N_{ab} \leq 3$ always, we demonstrate here the use of `anova(lm(...))` also for the one-way layout without interactions:

```
> shapiro.test(lm(seed.numbers~box+watering,data=germin[germin$watering!="6",]
+ )$residuals)$p.value
[1] 0.0704208
```

The p-value provided by the Shapiro-Wilks test is now much smaller than in the model with interactions. This is due to the strong interactions. But still it is larger than 0.05.

```
> anova(lm(seed.numbers~box+watering,data=germin[germin$watering!="6",]))
```

```
Analysis of Variance Table

Response: seed.numbers
          Df Sum Sq Mean Sq F value    Pr(>F)
box        1    0.2     0.2  0.0008 0.9779438
watering   4 6675.8  1668.9  7.3795 0.0002320 ***
Residuals 33 7463.3   226.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that the test statistics, the F values, have changed. But we see also that the sum of squares for `box:watering` and `Residuals` satisfy $6068.9 + 1394.4 = 7463.3$.

4.3.3 Exercise (Mustard)
In Example 2.2.1, the two t-tests for testing for an influence of cutting and of the growing conditions provided no significant results.
a) Now test with the ANOVA test whether there are effects of the cutting and the growing conditions and whether there are interactions between the two treatment factors. Test also the requirements of the ANOVA test and produce a complete ANOVA table. What do you conclude?
b) Change also the order of the treatment factors and compare the results.
c) Use the test for interactions as pretest. If the hypothesis of no interactions is not rejected, use also the ANOVA test for the two-way layout without interactions. Test then also its requirements. Compare the results with the ANOVA test for the two-way layout with interactions. What happens with the test statistics and P values? Explain the result.

4.4 Designs for two treatment factors

Table 4 regarded as matrix is called **incidence matrix** and has the form

$$\begin{pmatrix} N_{11} & N_{12} & \dots & N_{1B} \\ N_{21} & N_{22} & \dots & N_{2B} \\ \vdots & \vdots & & \vdots \\ N_{A1} & N_{A2} & \dots & N_{AB} \end{pmatrix} \tag{8}$$

If there are two treatments, then also the interaction of the treatments is of interest. This means that each N_{ab} should be at least 2. We have seen that balanced designs have the advantage that the order of the treatment has no influence on the analysis. Hence a good design is a balanced design with $N_{ab} = M$ for all $a = 1, \dots, A$, $b = 1, \dots, B$ and $M \geq 2$. We know that the higher M is the smaller the β-error of the test is.

The allocation of the level combinations should be done randomly to the $N = M\,A\,B$ experimental units. Such designs are called **randomized designs for two factors** and can be created with the function `design.ab` of the library `agricolae`.

4.4.1 Example
If factor A has the levels `A1, A2, A3` and factor B has the levels `B1, B2, B3, B4` and $M = 2$, then

we obtain for example the following allocation:

```
> library(agricolae)
> design.ab(c("A1","A2","A3"),c("B1","B2","B3","B4"),2)
   plots block c("A1", "A2", "A3") c("B1", "B2", "B3", "B4")
1      1     1                  A1                        B4
2      2     1                  A3                        B1
3      3     1                  A2                        B1
4      4     1                  A3                        B3
5      5     1                  A3                        B2
6      6     1                  A2                        B3
7      7     1                  A1                        B1
8      8     1                  A1                        B2
9      9     1                  A2                        B2
10    10     1                  A2                        B4
11    11     1                  A3                        B4
12    12     1                  A1                        B3
13    13     2                  A2                        B2
14    14     2                  A3                        B3
15    15     2                  A3                        B2
16    16     2                  A1                        B2
17    17     2                  A2                        B4
18    18     2                  A2                        B1
19    19     2                  A2                        B3
20    20     2                  A3                        B1
21    21     2                  A1                        B4
22    22     2                  A1                        B3
23    23     2                  A3                        B4
24    24     2                  A1                        B1
```

We see that the second repetitions are given in a second block. This makes sense since if the experiment must be stopped before all measurements are done, then at least all level combinations were used at least one time.

4.5 Designs for one treatment and one block factor

Usually experiments cannot be done under the same conditions. There are temporal and spacial conditions. The random allocation of the treatments to the experimental units in the one-way layout described in Subsection 3.4 aims to reduce unknown temporal and spacial influences. But sometimes these temporal and spacial influences are known and therefore cannot be neglected. This is in particular the case if the experiments are done in different years with different weather or at different experimental stations with different climatic conditions and different ground. Then these different conditions must be regarded as block factor. In our notation, the second factor B will denote this block factor. Since there are usually many block factors, the numbers N_{ab} of sample sizes for the level combinations of the treatment and the block factor are small. Very often the incidence matrix given in (8) consists only of zeros and ones. Since the numbers N_{ab} are such small, only the ANOVA test for the one-way layout without interaction is used. This makes sense, since

one wants to know whether there is a block effect and interactions are of less interest. The order of the treatment factor and the block factor should be such that the treatment is the factor A and the block factor is the factor B. Then at first in the whole model it is tested whether there is a treatment effect, and then in a model without treatment effects it is tested whether the blocks have significant influence.

The treatment factor is the factor A and the block factor is the factor B in the ANOVA test of Subsection 4.3.

For example a incidence matrix can have the form:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

This is a block design with $A = 4$ treatments and $B = 6$ blocks. It is a incomplete block design since some N_{ab} are equal to zero.

A block design is called **complete block design** if $N_{ab} \geq 0$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$.

A block design is called **incomplete block design** if there exists level combinations (a, b) with $N_{ab} = 0$.

A block design is called **balanced block design** if the number of treatments $N_{\bullet b}$ is equal in each block b ($N_{\bullet 1} = N_{\bullet 2} = \ldots = N_{\bullet B}$), each treatment level a appears in the same number of blocks ($N_{1\bullet} = N_{2\bullet} = \ldots = N_{A\bullet}$) and each pair of treatments $a_1, a_2 \in \{1, \ldots, A\}$ appears in the same number of blocks.

Obviously, if $N_{ab} = M$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$, then this is a balanced complete block design. But there are also balanced incomplete block designs.

A block incomplete design which is balanced is called **balanced incomplete block design (BIBD)**.

Balanced complete block designs

Although with the block factor a temporal or spacial influence is taking into account, there may be also unknown temporal or spacial influence. Therefore the treatments should be allocated in a block randomly. Such designs are called **randomized complete block designs (RCBD)** and can be constructed with `design.rcbd` of the `agricolae` package.

4.5.1 Example

If there are 3 levels `T1, T2, T3` for the treatment and 4 levels for the block factor, we can use for example the following randomized complete block design:

```
> library(agricolae)
> design.rcbd(c("T1","T2","T3"),4)
   plots block c("T1", "T2", "T3")
1      1     1                  T1
```

```
2      2     1                  T2
3      3     1                  T3
4      4     2                  T2
5      5     2                  T3
6      6     2                  T1
7      7     3                  T3
8      8     3                  T2
9      9     3                  T1
10    10     4                  T2
11    11     4                  T1
12    12     4                  T3
```

In randomized complete block designs, we have $N_{ab} = 1$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$. Besides the ANOVA test for two-way layout without interactions, also the **Friedman rank sum test** can be used. This is a distribution-free test and should be used if the normal distribution is rejected.

4.5.2 Example (A soil experiment)
The data in the file `SOIL.DAT` "are part of a larger experiment to determine the effectiveness of blast furnace slags (German: Hochofenschlacke) as agricultural liming material (German: Scheidungsmaterial) on three types of soil, sandy loam (German: Lehm) (I), sandy clay loam (clay in German: Ton) (II) and loamy sand (III). The treatments were all applied at 4000 lbs per acre, and what was measured was the corn yield in bushels per acre." (Hand et al. 1996, P. 220)
The tree types of soil are regarded as blocks. There were 7 levels of the treatment: none slag `None`, coarse slag `Coarse`, medium slag `Medium`, agricultural slag `slag`, agricultural limestone `lime`, agricultural slag + minor elements `slag.plus`, agricultural limestone + minor elements `lime.plus`. At first we read the data:

```
> soil0<-read.table("SOIL.DAT")
> soil1<-c(soil0[,1],soil0[,2],soil0[,3])
> soil2<-data.frame(rep(c("none","coarse","medium","slag","lime","slag.plus",
+ "lime.plus"),3),soil1)
> soil<-data.frame(c(rep("I",7),rep("II",7),rep("III",7)),soil2)
> names(soil)<-c("soil","slag","yield")
> soil
   soil      slag yield
1     I      none  11.1
2     I    coarse  15.3
3     I    medium  22.7
4     I      slag  23.8
5     I      lime  25.6
6     I slag.plus  31.2
7     I lime.plus  25.8
8    II      none  32.6
9    II    coarse  40.8
10   II    medium  52.1
11   II      slag  52.8
12   II      lime  63.1
```

```
13   II slag.plus  59.5
14   II lime.plus  55.3
15  III      none  63.3
16  III    coarse  65.0
17  III    medium  58.8
18  III      slag  61.4
19  III      lime  41.1
20  III slag.plus  78.1
21  III lime.plus  60.2
```

This is not a randomized complete block design. But maybe the randomization was lost by putting the data in file `SOIL.DAT`.

Then we test for normality:

```
> shapiro.test(lm(yield~soil+slag,data=soil)$residuals)$p.value
[1] 0.2199803
```

Hence the normal distribution is not rejected so that the ANOVA test can be used. Since `soil` is the block factor it should come at first:

```
> anova(lm(yield~soil+slag,data=soil))
Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value   Pr(>F)
soil       2 5696.3  2848.2 36.0743 8.41e-06 ***
slag       6  731.1   121.8  1.5432   0.2457
Residuals 12  947.4    79.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence the types of slags have no significant influence, however the soil has. But this was probably known before the experiment. Now we can also test the hypothesis of the influence of slag types with the Friedman rank sum test:

```
> friedman.test(yield~slag|soil,data=soil)

        Friedman rank sum test

data:  yield and slag and soil
Friedman chi-squared = 8.1429, df = 6, p-value = 0.2278
```

It provides almost the same p-value although usually distribution-free tests provides larger p-values and have larger β-error. Note the different writing of the formula: `yield~soil+slag` in the ANOVA test and `yield~slag|soil` in the Friedman rank sum test.

Balanced incomplete block designs

Balanced incomplete block designs are needed when the block size is too small so that not all treatments can be applied in the block. The block size can be even 2 in the extreme case. This is for example the case when the experimental units are the eyes persons and $k > 2$ eye drops should be studied. Then each person provides a block of block size is 2 and the number of treatments levels is higher than the block size.

The question is, for which block sizes and for which numbers of treatment levels a balanced incomplete block design exists. Here some necessary conditions for the existence are given. For balanced designs, let be

$$\begin{array}{ll} R & = N_{1\bullet} = \ldots = N_{A\bullet} \text{ the block size,} \\ K & = N_{\bullet 1} = \ldots = N_{\bullet B} \text{ the total number of repetitions of the treatment levels,} \\ \lambda & \text{the number of blocks in which a pair of different treatments } a_1,\ a_2 \in \{1, \ldots, A\} \text{ appears.} \end{array}$$

Obviously, it holds:

$$A\,R = B\,K \tag{9}$$

Moreover, there are

$$\begin{pmatrix} A \\ 2 \end{pmatrix} = \frac{A\,(A-1)}{2} \qquad \text{different pairs of treatment levels,}$$
$$\begin{pmatrix} K \\ 2 \end{pmatrix} = \frac{K\,(K-1)}{2} \qquad \text{different pairs of treatments in each block,}$$

so that

$$\lambda\,A\,(A-1) = B\,K\,(K-1).$$

Substituting $B\,K$ by $A\,R$ and dividing by A, we obtain

$$\lambda\,(A-1) = R\,(K-1). \tag{10}$$

Conditions (9) and (10) are only necessary conditions for a balanced incomplete block designs but not sufficient conditions. For example $A = 16$, $R = 3$, $B = 8$, $K = 6$, and $\lambda = 1$ satisfies with 16 3=8 6 and 1 15=3 5 the conditions (9) and (10) but there exists no balanced incomplete block design.

In the function `design.bib` of the `agricolae` package one can only specify the number of treatments and the block size.

4.5.3 Example (Balanced incomplete block design)
The balanced incomplete block design for $A = 3$ and $B = 2$ has the form:

```
> design.bib(c("T1","T2","T3"),k=2)

Parameters BIB
==============
Lambda     : 1
treatmeans : 3
Block size : 2
Blocks     : 3
Replication: 2

Efficiency factor 0.75

<<< Book >>>
  plots block c("T1", "T2", "T3")
1     1     1                  T1
2     2     1                  T2
3     3     2                  T1
4     4     2                  T3
5     5     3                  T2
6     6     3                  T3
```

This means that we have $R = 2$, $B = 3$ and $\lambda = 1$. A BIB design with $A = 16$ and $K = 6$ is so large that we give here only the automatic printout:

```
> ddd<-design.bib(as.factor(1:16),k=6)

Parameters BIB
==============
Lambda     : 1001
treatmeans : 16
Block size : 6
Blocks     : 8008
Replication: 3003

Efficiency factor 0.8888889

<<< Book >>>
```

4.5.4 Exercise (Designs for two factors)
a) Create a randomized design for two treatments A and B where treatment A has 6 levels, treatment B has 4 levels, and $M = 2$.
b) Create a randomized complete block design for a treatment with 6 levels and a block factor with 4 levels.
c) Is it possible so create a randomized incomplete block design for 6 treatments levels, 4 block levels and block size 3? d) Create a randomized incomplete block design for 6 treatments levels and block size 3.

4.5.5 Exercise (Pepper)
Consider the data set `pepper` in Exercise 1.5.2. Treat the three treatment factors Heating, Lighting, Carbon dioxid, where each has 2 levels, as one treatment factor with $8 = 2^3$ levels. The condition of the experiment was that only 6 compartments per block are available. Is the design of data set `pepper` randomized incomplete block design if the block factor year is neglected? Find a randomized incomplete block design for the experimental conditions with minimum number N of experimental units. Neglect again the block factor year. How many blocks are needed? And if only two blocks per year can be realized, how many years are needed for a randomized incomplete block design?

5 Multi-way layouts and further models

If a data set has more than two factor variables, then we have a multi-way layout. The factor variables can consist of several block variables $\mathrm{B}^1,\ldots,\mathrm{B}^I$ and several treatment variables $\mathrm{A}^1,\ldots,\mathrm{A}^J$. Still we assume that there is only one numeric variable.

Since we have more than two factors, we do not have only interactions between two factors but also higher order interactions between several factors.

Interactions of higher order
The effect, which appears when several factors are simultaneous at certain levels, is called higher order interactions.

Example: four-way layout
Consider for example four factors A, B, C, and D. The factor A has levels $a_1,\ldots,a_A$, factor B has levels $b_1,\ldots,b_B$, factor C has levels $c_1,\ldots,c_C$, and the factor D has levels $d_1,\ldots,d_D$. The full model with all interactions is than

$$\begin{aligned} Y_{abcd} = {} & \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} \\ & + \alpha\beta\gamma_{abc} + \alpha\beta\delta_{abd} + \beta\gamma\delta_{bcd} + \alpha\beta\gamma\delta_{abcd} + Z_{abcd} \end{aligned}$$

with $Z_{abcd} \sim \mathcal{N}(0,\sigma^2)$ for all $a \in \{a_1,\ldots,a_A\}$, $b \in \{b_1,\ldots,b_B\}$, $c \in \{c_1,\ldots,c_C\}$, $d \in \{d_1,\ldots,d_D\}$. For simplicity we also write: $a \in \{1,\ldots,A\}$, $b \in \{1,\ldots,B\}$, $c \in \{1,\ldots,C\}$, $d \in \{1,\ldots,D\}$. Thereby,

μ	is the average mean,
α_a, β_b, γ_c, δ_d	are the main effects of the factors a, B, C, D,
$\alpha\beta_{ab}$	is the second order interaction of factor A at level a and factor B at level b,
$\alpha\gamma_{ac}$, $\beta\gamma_{bc}$, $\alpha\delta_{ad}$, $\beta\delta_{bd}$	are the other second order interactions,
$\alpha\beta\gamma_{abc}$	is the third order interaction of factor A at level a, factor B at level b, and factor C at level c,
$\alpha\beta\delta_{abd} + \beta\gamma\delta_{bcd}$	are the other third order interactions,
$\alpha\beta\gamma\delta_{abcd}$	is the forth order interaction of factor A at level a, factor B at level b, factor C at level c, and factor D at level d.

The main effects and the second order interactions should satisfy the same side conditions as for the two-way layout. Moreover, the higher order interactions should satisfy for example

$$\begin{aligned} 0 &= \sum_{a=1}^{A} \alpha\beta\gamma_{abc} = \sum_{b=1}^{B} \alpha\beta\gamma_{abc} = \sum_{c=1}^{C} \alpha\beta\gamma_{abc}, \\ 0 &= \sum_{a=1}^{A} \alpha\beta\gamma\delta_{abcd} = \sum_{b=1}^{B} \alpha\beta\gamma\delta_{abcd} = \sum_{c=1}^{C} \alpha\beta\gamma\delta_{abcd} = \sum_{d=1}^{D} \alpha\beta\gamma\delta_{abcd}, \end{aligned}$$

for all a, b, c, d. This means that we have

$$\begin{aligned} & 1 + (A-1) + (B-1) + (C-1) + (D-1) + (A-1)(B-1) + (A-1)(C-1) \\ & \quad + (B-1)(C-1) + (A-1)(D-1) + (B-1)(D-1) + (C-1)(D-1) + \\ & \quad + (A-1)(B-1)(C-1) + (A-1)(B-1)(D-1) + (A-1)(C-1)(D-1) \\ & \quad + (B-1)(C-1)(D-1) + (A-1)(B-1)(C-1)(D-1) \\ = \; & A \cdot B \cdot C \cdot D \end{aligned}$$

parameters.

Estimability in the full model of the four-way layout
All of the $A\,B\,C\,D$ parameters of the full model are estimable if the total sample size N is at least $A \cdot B \cdot C \cdot D$, i.e. $N \geq A \cdot B \cdot C \cdot D$, and each level combimation a, b, c, d is observed at least once, i.e. $N_{abcd} \geq 1$.
Testability in the full model of the four-way layout
All hypotheses about main effects and interactions are testable if the total sample size N is greater than $A \cdot B \cdot C \cdot D$, i.e. $N > A \cdot B \cdot C \cdot D$, and each level combimation a, b, c, d is observed at least once, i.e. $N_{abcd} \geq 1$.

This means that for testing, we need at least one observation more than for estimation. This is due to the fact that for testing we additionally need an estimate for the variance σ^2, which is not allowed to be zero.

5.1 The ANOVA test for the multi-way layout

The ANOVA test tests the interactions and main effects again in a sequential order. For example, for the four-way layout as follows:

$$\begin{aligned} 1)\; H_0^{ABCD}: \quad & \mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} \\ & + \alpha\beta\gamma_{abc} + \alpha\beta\delta_{abd} + \beta\gamma\delta_{bcd} \end{aligned}$$

versus

$$\begin{aligned} H_1^{ABCD}: \quad & \mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} \\ & + \alpha\beta\gamma_{abc} + \alpha\beta\delta_{abd} + \beta\gamma\delta_{bcd} + \alpha\beta\gamma\delta_{abcd} \end{aligned}$$

$$\begin{aligned} 2)\; H_0^{BCD}: \quad & \mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} \\ & + \alpha\beta\gamma_{abc} + \alpha\beta\delta_{abd} \end{aligned}$$

versus

$$\begin{aligned} H_1^{BCD}: \quad & \mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} \\ & + \alpha\beta\gamma_{abc} + \alpha\beta\delta_{abd} + \beta\gamma\delta_{bcd} \end{aligned}$$

3) H_0^{ABD} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} + \alpha\beta\gamma_{abc}$

versus

H_1^{ABD} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} + \alpha\beta\gamma_{abc} + \alpha\beta\delta_{abd}$

4) H_0^{ABC} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd}$

versus

H_1^{ABC} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd} + \alpha\beta\gamma_{abc}$

5) H_0^{BD} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad}$

versus

H_1^{BD} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad} + \beta\delta_{bd}$

6) H_0^{AD} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc}$

versus

H_1^{AD} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc} + \alpha\delta_{ad}$

7) H_0^{BC} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \beta\gamma_{bc}$

versus

H_1^{BC} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac} + \beta\gamma_{bc}$

8) H_0^{AC} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab}$

versus

H_1^{AC} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab} + \alpha\gamma_{ac}$

9) H_0^{AB} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d$

versus

H_1^{AB} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d + \alpha\beta_{ab}$

10) H_0^{D} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c$

versus

H_1^{D} : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c + \delta_d$

11) H_0^C : $\mu_{abcd} = \mu + \alpha_a + \beta_b$

versus

H_1^C : $\mu_{abcd} = \mu + \alpha_a + \beta_b + \gamma_c$

12) H_0^B : $\mu_{abcd} = \mu + \alpha_a$

versus

H_1^B : $\mu_{abcd} = \mu + \alpha_a + \beta_b$

13) H_0^A : $\mu_{abcd} = \mu$

versus

H_1^A : $\mu_{abcd} = \mu + \alpha_a$

All these hypotheses can be only tested if the sample size is greater than $A \cdot B \cdot C \cdot D$. If this is not the case some interaction must be dropped.

5.1.1 Example (Pepper)
Now we regard all variables of the Example 1.5.2 separately. This means that there are two block variables (`Year` and `Block`) and three treatment variables (`Heating, Lighting CO2`).

```
> aov(Excess~Year*Block*Heating*Lighting*CO2,data=pepper)
Call:
   aov(formula = Excess ~ Year * Block * Heating * Lighting * CO2,
    data = pepper)

Terms:
                    Year    Block  Heating Lighting      CO2 Year:Block
Sum of Squares  24.40167 92.04167  7.62881 20.36507  1.19428    0.17964
Deg. of Freedom        1        1        1        1        1          1
                Year:Heating Block:Heating Year:Lighting Block:Lighting
Sum of Squares       0.44831       0.00373       0.02709        0.49988
Deg. of Freedom            1             1             1              1
                Heating:Lighting Year:CO2 Block:CO2 Heating:CO2 Lighting:CO2
Sum of Squares           0.08403  0.26779   0.84438     0.17992      0.15277
Deg. of Freedom                1        1         1           1            1
                Year:Block:Heating Year:Block:Lighting Block:Heating:Lighting
Sum of Squares             1.64638             0.34362                0.47883
Deg. of Freedom                  1                   1                      1
                Year:Block:CO2 Block:Heating:CO2 Block:Lighting:CO2 Residuals
Sum of Squares         0.72043           0.00600            0.51571   1.21000
Deg. of Freedom              1                 1                  1         2

Residual standard error: 0.7778175
10 out of 32 effects not estimable
Estimated effects may be unbalanced
```

Hence not all of the parameters of the full model are estimable. Hence we must reduce the model, for example, to:

```
> aov(Excess~Year+Block+Heating*Lighting*CO2,data=pepper)
Call:
   aov(formula = Excess ~ Year + Block + Heating * Lighting * CO2,
    data = pepper)

Terms:
                    Year    Block  Heating Lighting      CO2 Heating:Lighting
Sum of Squares  24.40167 92.04167  7.62881 20.36507  1.19428          0.20271
Deg. of Freedom        1        1        1        1        1                1
                Heating:CO2 Lighting:CO2 Heating:Lighting:CO2 Residuals
Sum of Squares      0.03074      0.13298              1.17600   6.06607
Deg. of Freedom           1            1                    1        14

Residual standard error: 0.6582484
Estimated effects may be unbalanced
```

In this model, all parameters are estimable. Before we use the ANOVA test, we test for normal distribution:

```
> shapiro.test(aov(Excess~Year+Block+Heating*Lighting*CO2,data=pepper)$residuals)$p.value
[1] 0.7816656
> anova(lm(Excess~Year+Block+Heating*Lighting*CO2,data=pepper))
Analysis of Variance Table

Response: Excess
                     Df Sum Sq Mean Sq  F value    Pr(>F)
Year                  1 24.402  24.402  56.3170 2.859e-06 ***
Block                 1 92.042  92.042 212.4246 7.441e-10 ***
Heating               1  7.629   7.629  17.6067 0.0008974 ***
Lighting              1 20.365  20.365  47.0009 7.861e-06 ***
CO2                   1  1.194   1.194   2.7563 0.1190947
Heating:Lighting      1  0.203   0.203   0.4678 0.5051472
Heating:CO2           1  0.031   0.031   0.0709 0.7938500
Lighting:CO2          1  0.133   0.133   0.3069 0.5883288
Heating:Lighting:CO2  1  1.176   1.176   2.7141 0.1217170
Residuals            14  6.066   0.433
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this model the following tests were done in the following order. Thereby the following abbreviations are used: Y =Year, B =Block, H =Heating, L =Lighting, C =CO2.

1) H_0^{H*L*C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L + H*C + L*C$

versus

H_1^{H*L*C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L + H*C + L*C + H*L*C$

H_0^{H*L*C} is not rejected.

2) H_0^{L*C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L + H*C$

versus

H_1^{L*C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L + H*C + L*C$

H_0^{L*C} is not rejected.

3) H_0^{H*C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L$

versus

H_1^{H*C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L + H*C$

H_0^{H*C} is not rejected.

4) H_0^{H*L} : $\mu_{ybhlc} = \mu + Y + B + H + L + C$

versus

H_1^{H*L} : $\mu_{ybhlc} = \mu + Y + B + H + L + C + H*L$

H_0^{H*L} is not rejected.

5) H_0^{C} : $\mu_{ybhlc} = \mu + Y + B + H + L$

versus

H_1^{C} : $\mu_{ybhlc} = \mu + Y + B + H + L + C$

H_0^{C} is not rejected.

6) H_0^{L} : $\mu_{ybhlc} = \mu + Y + B + H$

versus

H_1^{L} : $\mu_{ybhlc} = \mu + Y + B + H + L$

There is a significant lighting effect.

7) H_0^{H} : $\mu_{ybhlc} = \mu + Y + B$

versus

H_1^{H} : $\mu_{ybhlc} = \mu + Y + B + H$

There is a significant heating effect.

8) H_0^{B} : $\mu_{ybhlc} = \mu + Y$

versus

H_1^{B} : $\mu_{ybhlc} = \mu + Y + B$

There is a significant block effect.

$$9)\ H_0^Y: \quad \mu_{ybhlc} = \mu$$
$$\text{versus}$$
$$H_1^Y: \quad \mu_{ybhlc} = \mu + Y$$

There is a significant year effect.

5.1.2 Exercise (Pepper)
Study for the data set `pepper` also the following models: `Excess~Year+Block*Heating*Lighting*CO2`, `Excess~Year*Block+Heating*Lighting*CO2`, `Excess~Year+Block+Heating+Lighting+CO2`. In which model are all parameters estimable and in which model are all hypothesis testable? Indicate the parameters which are not estimable. Do the ANOVA for the testable models and check the normal distribution. Which model should be used in practice?

5.1.3 Exercise (Huasahuasi)
The package `agricolae` contains the data set `huasahuasi`:

```
> library(agricolae)
> data(huasahuasi)
> huasahuasi
   Block    Treat     Clon Comercial  y1da  y2da y3ra yield   AUDPC
1      I     40mm 386209.1     18.80 10.80  8.00 6.10 24.90  442.40
2      I     40mm 387164.4     28.25 22.75  5.50 4.14 32.39    2.10
3      I     40mm Cruza148     13.30  1.60 11.70 4.15 17.45   30.80
4      I     40mm    Musuq      8.60  3.50  5.10 2.40 11.00 1424.85
5      I     40mm   Yungay     20.82 10.92  9.90 5.20 26.02  404.95
6      I    7dias 386209.1     23.00 10.50 12.50 3.60 26.60  895.65
7      I    7dias 387164.4     28.98 21.98  7.00 7.60 36.58    7.70
8      I    7dias Cruza148     11.95  2.80  9.15 4.90 16.85   13.65
9      I    7dias    Musuq      7.15  2.55  4.60 3.50 10.65 1147.30
10     I    7dias   Yungay     26.80 18.00  8.80 5.90 32.70  359.10
11     I SinAplic 386209.1     13.60 10.75  2.85 2.50 16.10 2071.30
12     I SinAplic 387164.4     31.80 29.30  2.50 6.50 38.30   20.30
13     I SinAplic Cruza148     16.80 10.50  6.30 4.20 21.00  156.45
14     I SinAplic    Musuq      0.80  0.00  0.80 1.65  2.45 2590.70
15     I SinAplic   Yungay     15.05  8.45  6.60 5.60 20.65 1790.60
16    II     40mm 386209.1     15.80 12.00  3.80 3.10 18.90 1254.05
17    II     40mm 387164.4     50.70 46.60  4.10 2.40 53.10    6.65
18    II     40mm Cruza148     12.40  6.40  6.00 5.70 18.10   39.90
19    II     40mm    Musuq      0.85  0.00  0.85 1.30  2.15 3317.30
20    II     40mm   Yungay     24.80 14.10 10.70 1.95 26.75 1125.25
21    II    7dias 386209.1     20.55 15.65  4.90 6.30 26.85  476.70
22    II    7dias 387164.4     37.30 32.10  5.20 2.00 39.30    1.05
23    II    7dias Cruza148     12.30  5.00  7.30 8.00 20.30   21.00
24    II    7dias    Musuq      9.60  6.00  3.60 2.30 11.90  428.05
```

```
25    II    7dias   Yungay     20.20 14.30  5.90 2.70 22.90  514.50
26    II SinAplic 386209.1      9.50  5.40  4.10 2.70 12.20 1785.00
27    II SinAplic 387164.4     35.50 30.30  5.20 1.60 37.10    0.35
28    II SinAplic Cruza148     13.40  6.30  7.10 5.80 19.20   74.20
29    II SinAplic    Musuq      0.00  0.00  0.00 0.40  0.40 3168.90
30    II SinAplic   Yungay     11.60  5.50  6.10 2.15 13.75 2072.35
31   III     40mm 386209.1     10.95  5.10  5.85 3.20 14.15  872.20
32   III     40mm 387164.4     33.10 29.30  3.80 4.90 38.00    2.45
33   III     40mm Cruza148     18.70  8.10 10.60 5.80 24.50   29.05
34   III     40mm    Musuq      1.50  0.50  1.00 1.10  2.60 3069.50
35   III     40mm   Yungay     24.00 15.70  8.30 1.40 25.40  754.95
36   III    7dias 386209.1     18.75 10.40  8.35 3.05 21.80  517.30
37   III    7dias 387164.4     37.80 32.10  5.70 3.10 40.90    3.50
38   III    7dias Cruza148     17.90  9.30  8.60 3.55 21.45   15.05
39   III    7dias    Musuq      3.40  0.70  2.70 3.70  7.10 1352.40
40   III    7dias   Yungay     35.90 25.90 10.00 3.20 39.10  318.50
41   III SinAplic 386209.1     11.35  4.85  6.50 4.40 15.75 1520.05
42   III SinAplic 387164.4     34.70 29.20  5.50 4.65 39.35    2.45
43   III SinAplic Cruza148     14.70  3.80 10.90 4.70 19.40   33.60
44   III SinAplic    Musuq      0.25  0.00  0.25 0.70  0.95 2903.60
45   III SinAplic   Yungay     25.80 13.70 12.10 1.10 26.90 1055.95
```

Regard only the measurement `yield` but all factors `Block`, `Treat`, `Clon`. Find the largest reasonable model in which all parameters are estimable and the largest reasonable model in which all hypothesis are testable. Write done the null hypotheses, alternatives and the conclusions as in Example 5.1.1.

5.2 Latin square and graeco-latin square designs

If there are two block factors and one treatment factor and all three factors have the same number of levels, than a latin square design is the best design.

Latin square design
A latin square design is a design which allocates to each level combination of two block factors exactly one treatment level such that for each block level all treatment levels are used. Thereby the numbers of treatment and block levels are the same.

Latin square designs are produced by `design.lsd` of the `agricolae` package. For example,

```
> library(agricolae)
> design.lsd(c("A","B","C","D"))
   plots row col c("A", "B", "C", "D")
1      1   1   1                     A
2      2   1   2                     D
3      3   1   3                     C
4      4   1   4                     B
5      5   2   1                     C
6      6   2   2                     B
```

```
7       7   2   3                     A
8       8   2   4                     D
9       9   3   1                     D
10     10   3   2                     C
11     11   3   3                     B
12     12   3   4                     A
13     13   4   1                     B
14     14   4   2                     A
15     15   4   3                     D
16     16   4   4                     C
```

provides the following latin square

	1	2	3	4
1	A	D	C	B
2	C	B	A	D
3	D	C	B	A
4	B	A	D	C

Here the rows are the levels of the first block factor and the columns are the levels of the second block factor. The capital letters A,B,C,D denote the four levels of the treatment.

If there are two block factors and two treatment factors and all four factors have the same number of levels, than a graeco-latin square design is the best design.

Graeco-latin square design
A graeco-latin square design is a design which allocates to each level combination of two block factors exactly one combination of levels of two treatment factors such that for each block level all levels of the first and the second treatment factor are used. Thereby the numbers of treatment and block levels are the same.

Latin square designs are produced by `design.graeco` of the `agricolae` package. For example,

```
> library(agricolae)
> design.graeco(c("A","B","C","D"),c("a","b","c","d"))
   plots row col c("A", "B", "C", "D") c("a", "b", "c", "d")
1      1   1   1                     A                     b
2      2   1   2                     D                     a
3      3   1   3                     C                     d
4      4   1   4                     B                     c
5      5   2   1                     D                     c
6      6   2   2                     A                     d
7      7   2   3                     B                     a
8      8   2   4                     C                     b
9      9   3   1                     C                     a
10    10   3   2                     B                     b
11    11   3   3                     A                     c
12    12   3   4                     D                     d
```

```
13    13   4   1                B                d
14    14   4   2                C                c
15    15   4   3                D                b
16    16   4   4                A                a
```

provides the following graeco-latin square

	1	2	3	4
1	Ab	Da	Cd	Bc
2	Dc	Ad	Ba	Cb
3	Ca	Bb	Ac	Dd
4	Bd	Cc	Db	Aa

Here the small letters stands for the graeco letters.

5.2.1 Exercise (Graeco-latin square design)
Produce a graeco-latin square design for the case that the numbers of block and treatment levels are 5. Write the design also as graeco-latin square.

The level combinations of a latin square or graeco-latin square design can have repetitions or not. If they have repetitions, the number of repetitions should be equal for all combinations of the design. Many latin and graeco latin square designs have no repetitions. Then not all interactions can be tested. Even in the case of 3 levels for each factor, only an additive model, i.e. a model without interaction, can be used.

5.2.2 Exercise (A vandalized experiment)
The data file `VANDAL.DAT` contains the data of the following experiment. "'Six varieties of turnip were grown in 36 plots arranged in a latin square design. The response variable is the fresh weight (roots plus tops) of turnips (German: Rübe) in pounds per plot (15ft $\times$ 15ft). Three plots in one corner of the experiment had been attacked by vandals and therefore did not yield any usuable data. Do the varieties of turnip differ in mean weight per plot...? The data below are laid out in the pattern of the experiment. The letters denote the varieties, A to F."' (Hand et al. 1996, P. 61)

```
E, 29.0 F, 14.5 D, 20.5 A, 22.5 B, 16.0  C,  6.5
B, 17.5 A, 29.5 E, 12.0 C,  9.0 D, 33.0  F, 12.5
F, 17.0 B, 30.0 C, 13.0 D, 29.0 A, 27.0  E, 12.0
A, 31.5 D, 31.5 F, 24.0 E, 19.5 C, 10.5  B, 21.0
D, 25.0 C, 13.0 B, 31.0 F, 26.0 E, 19.5  A,   NA
C, 12.2 E, 13.0 A, 34.0 B, 20.0 F,   NA  D,   NA
```

Analyze the experiment. Does the order of the factors influence the analysis? What happens, if the missing values are substituted by 21.5 for F, 20.8 for A, 13.5 for D?

5.3 Factorial designs

Latin square designs and graeco-latin square designs are special fractional factorial designs.

Factorial and fractional factorial design
Assume that there are k factors, each with n levels.
A **complete factorial design** is a design where each of the n^k level combinations is realized $M \geq 1$ times.
A **fractional factorial design** is a design where only some of the n^k level combinations are realized $M \geq 1$ times.

As in latin square and graeco-latin square designs, not all interactions can be estimated in fractional factorial designs. However, some linear combinations of the interaction parameters can be estimated. If interaction parameters are only estimable within such linear combinations, then they are called **confounded** (in German: vermengt).

However, in complete factorial, all interactions are estimable. The function `fact.nk` of the `agricolae` package provides complete factorial designs, where the allocation to the experimental units within blocks is done randomly.

5.3.1 Example (Complete factorial design)
If there are three factors, each with two levels, which should be allocated to the experimental units of four blocks, type

```
> fact.nk(2,3,4)
   plots blocks A B C
1      1      1 0 1 1
2      2      1 1 0 1
3      3      1 1 1 1
4      4      1 1 1 0
5      5      1 0 0 1
6      6      1 0 0 0
7      7      1 1 0 0
8      8      1 0 1 0
9      9      2 1 1 1
10    10      2 1 0 0
11    11      2 0 1 0
12    12      2 1 0 1
13    13      2 0 0 0
14    14      2 1 1 0
15    15      2 0 1 1
16    16      2 0 0 1
17    17      3 0 1 1
18    18      3 1 0 1
19    19      3 1 0 0
20    20      3 0 0 0
21    21      3 1 1 0
22    22      3 1 1 1
23    23      3 0 1 0
```

```
24    24      3 0 0 1
25    25      4 0 1 1
26    26      4 1 0 1
27    27      4 0 1 0
28    28      4 0 0 1
29    29      4 1 0 0
30    30      4 1 1 0
31    31      4 0 0 0
32    32      4 1 1 1
```

In each block we have $2^3 = 8$ level combinations of the three factors.

5.3.2 Example (Pepper)
The Example 1.5.2 provides a design where the treatments H =`Heating`, L =`Lighting`, C =`CO2` follow a 2^3 factorial design. However, the allocation of the 2^3 level combinations to the blocks is not complete since each block can have only 6 experimental units.

It is clear as soon as we have more and more factors, then the number of experimental units within each block explodes. For example, for 8 factors, each with 2 levels, we need $2^8 = 256$ units for each block. This is usually not possible, the reason why fractional factorial designs are needed. The R package `conf.design` provides fractional factorial designs. In particular with this package designs can be constructed where specified treatment contrasts are confounded with blocks.

5.4 Hierachical models and split plot designs

5.4.1 Example
The data set `plots` of the package `agricolae` is based on a split plot design.

```
> library(agricolae)
> data(plots)
> plots
   block plot  A  B yield
1      1   p1 a1 b1     4
2      1   p1 a1 b2     1
3      1   p1 a1 b3     9
4      1   p2 a2 b1     6
5      1   p2 a2 b2    10
6      1   p2 a2 b3     2
7      2   p3 a1 b1     5
8      2   p3 a1 b2     3
9      2   p3 a1 b3    10
10     2   p4 a2 b1     4
11     2   p4 a2 b2    14
12     2   p4 a2 b3     1
13     3   p5 a1 b1     2
```

```
14     3   p5 a1 b2     2
15     3   p5 a1 b3    15
16     3   p6 a2 b1     3
17     3   p6 a2 b2    12
18     3   p6 a2 b3     1
```

There are 3 blocks and each block is divided in two subblocks called plots. Hence there are 6 plots but each plot belongs only to one block. The blocks and plots are nested. Without the plots, this would be a balanced complete block design. Models, where factors are nested, are called hierarchical models or nested models.

The nested structure of the design can be expressed in the formula for the ANOVA by `block/plot`:

```
> anova(lm(yield~block/plot+A*B,data=plots))
Analysis of Variance Table

Response: yield
           Df  Sum Sq Mean Sq F value    Pr(>F)
block       1   0.750   0.750  0.1742 0.6873870
A           1   0.222   0.222  0.0516 0.8259782
B           2  29.778  14.889  3.4581 0.0827438 .
block:plot  3   5.472   1.824  0.4237 0.7413290
A:B         2 300.444 150.222 34.8903 0.0001119 ***
Residuals   8  34.444   4.306
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(lm(yield~block/plot+A*B,data=plots))

Call:
lm(formula = yield ~ block/plot + A * B, data = plots)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3333 -0.7778  0.1111  0.9167  3.0000

Coefficients: (2 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.8333     2.1318   0.860  0.41483
block          0.8333     0.8471   0.984  0.35406
Aa2           -0.5556     2.1872  -0.254  0.80590
Bb2           -1.6667     1.6942  -0.984  0.35406
Bb3            7.6667     1.6942   4.525  0.00194 **
block:plotp2   2.3333     2.3960   0.974  0.35865
block:plotp3   0.2500     0.7336   0.341  0.74205
block:plotp4   0.9167     0.9471   0.968  0.36145
block:plotp5       NA         NA      NA       NA
block:plotp6       NA         NA      NA       NA
Aa2:Bb2        9.3333     2.3960   3.895  0.00457 **
```

```
Aa2:Bb3      -10.6667     2.3960  -4.452  0.00213 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.075 on 8 degrees of freedom
Multiple R-Squared: 0.9072,     Adjusted R-squared: 0.8028
F-statistic: 8.688 on 9 and 8 DF,  p-value: 0.002851
```

That some block/plots parameters cannot be estimated does not depend on the interactions of A and B. We obtain the same problem without them:

```
> summary(lm(yield~block/plot+A+B,data=plots))

Call:
lm(formula = yield ~ block/plot + A + B, data = plots)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8889 -4.4722  0.1111  3.2778  8.1111

Coefficients: (2 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.0556     5.6239   0.366    0.722
block          0.8333     2.3625   0.353    0.732
Aa2           -1.0000     4.7250  -0.212    0.837
Bb2            3.0000     3.3411   0.898    0.390
Bb3            2.3333     3.3411   0.698    0.501
block:plotp2   2.3333     6.6822   0.349    0.734
block:plotp3   0.2500     2.0460   0.122    0.905
block:plotp4   0.9167     2.6414   0.347    0.736
block:plotp5       NA         NA      NA       NA
block:plotp6       NA         NA      NA       NA

Residual standard error: 5.787 on 10 degrees of freedom
Multiple R-Squared: 0.0976,     Adjusted R-squared: -0.5341
F-statistic: 0.1545 on 7 and 10 DF,  p-value: 0.9893
```

5.5 Models with random effects

If factors are block factors, then it could be that one is not interested in the effect of a specific level since the levels are for example regions or years and one is not interested in the effect of a specific randomly chosen region or specific year which will never come again. Then one is only interested whether the region or year at all has a effect. Since the regions or years are chosen randomly, their effect is also random. Usually, we do not have only block factors but also treatment factors so that we have factors with random effects and factors with fixed effects. Such models are called **mixed models** (German: Gemischte Modelle). The analysis of variance (ANOVA) is the same as for models with fixed effects if the designs are balanced. The main difference between

models with random effects and models with only fixed effects is that observations are not anymore stochastically independent. This provides different β-errors. Moreover, different parameters are estimated. Instead of the effects of the fixed factor levels, **variance components** of the random factors are estimated.

6 Regression

In the previous sections, it was assumed that there is one numeric measurement variable and one or several factor variables. The factor variables can be also considered as **explanatory variables** for the measurement variable. Here, we will assume that there are further numeric variables besides the measurement variable and that these other numeric variable can be used as explanatory variable for the measurement variable. These explanatory numeric variables are considered as **independent variables** while the measurement variable is a **dependent variable** which depends on the explanatory variables. The aim of regression is to specify this dependence.

6.1 Linear regression

The simplest dependence is that of linear regression. In this case we have only one explanatory variable:

$$x = (x_1, \ldots, x_N)^\top.$$

The dependent measurement variable is denoted by

$$y = (y_1, \ldots, y_N)^\top$$

and is a realization of a random vector $Y = (Y_1, \ldots, Y_N)^\top$ of stochastically independent variables $Y_1, \ldots, Y_N$. Often $x = (x_1, \ldots, x_N)^\top$ is given by the experimenter for example if this variable concerns some concentrations or doses of a drug or fertilizer. But it also can consist of measurements. Than it is also a realization of a random variable. But we will only consider the conditional distribution of $Y = (Y_1, \ldots, Y_N)^\top$ given $x = (x_1, \ldots, x_N)^\top$ so that $x = (x_1, \ldots, x_N)^\top$ is regarded always as nonrandom.

In a linear regression model, the dependence of the measurement Y_n on the explanatory variable x_n is given by

$$Y_n = \beta_0 + \beta_1 x_n + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2),$$

for all $n = 1, \ldots, N$.

Estimates for the unknown parameters β_0 and β_1 are obtained by the method of least squares, i.e. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are those values so that the sum of squares

$$\sum_{n=1}^{N} (y_n - \beta_0 - \beta_1 x_n)^2$$

is as small as possible. It can be proved that this minimization problem is solved by

$$\begin{aligned} \widehat{\beta}_0 &= \overline{y} - \widehat{\beta}_1 \overline{x}, \\ \widehat{\beta}_1 &= \frac{s_{xy}}{s_x^2}, \end{aligned}$$

where

$$\overline{y} = \frac{1}{N}\sum_{n=1}^{N} y_n, \quad \overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

are the arithmetic means of y and x, respectively,

$$s_{xy} = \frac{1}{N-1}\sum_{n=1}^{N}(y_n - \overline{y})(x_n - \overline{x})$$

is the empirical covariance between y and x, and

$$s_x^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \overline{x})^2$$

is the empirical variance of x. The estimate for the unknown variance σ^2 is based on the sum of squares for errors which is the sum of squares with the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$, i.e.

$$\widehat{\sigma}^2 = \widehat{\sigma}^2_{SSE} = \frac{1}{N-2}\sum_{n=1}^{N}(y_n - \widehat{\beta}_0 - \widehat{\beta}_1\, x_n)^2.$$

These estimators are unbaised estimators, i.e. their expectations are the values which they are estimating, i.e.

$$E_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_0) = \beta_0, \quad E_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_1) = \beta_1, \quad E_{\beta_0,\beta_1,\sigma^2}(\widehat{\sigma}^2) = \sigma^2,$$

for all $\beta_0,\ \beta_1,\ \sigma^2$. The variances of the estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are

$$\sigma^2_{\beta_0} = \mathrm{var}_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_0) = \frac{\overline{x^2}}{(N-1)s_x^2}\,\sigma^2 = \frac{\sum_{n=1}^{N} x_n^2}{N\ \sum_{n=1}^{N}(x_n - \overline{x})^2}\,\sigma^2$$

and

$$\sigma^2_{\beta_1} = \mathrm{var}_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_1) = \frac{1}{(N-1)s_x^2}\,\sigma^2 = \frac{1}{\sum_{n=1}^{N}(x_n - \overline{x})^2}\,\sigma^2.$$

These variances are estimated by

$$\widehat{\sigma}^2_{\beta_0} = \frac{\sum_{n=1}^{N} x_n^2}{N\ \sum_{n=1}^{N}(x_n - \overline{x})^2}\,\widehat{\sigma}^2 \quad \text{and} \quad \widehat{\sigma}^2_{\beta_1} = \frac{1}{\sum_{n=1}^{N}(x_n - \overline{x})^2}\,\widehat{\sigma}^2.$$

6.1.1 Exercise (Designs for linear regression)
Assume that the design region is [0,1] and that $N = 12$. Which constellations of $x_1, \ldots, x_{12} \in [0,1]$ provides the smallest variance $\sigma^2_{\beta_0}$ and which provides the smallest variance $\sigma^2_{\beta_1}$? Try different constellations like

$$
\begin{aligned}
x &= (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.5, 0.6, 0.7, 0.8, 0.9, 1)^\top, \\
x &= (0, 0, 0.2, 0.2, 0.4, 0.4, 0.6, 0.6, 0.8, 0.8, 1, 1)^\top, \\
x &= (0, 0, 0, 0.3, 0.3, 0.3, 0.7, 0.7, 0.7, 1, 1, 1)^\top, \\
x &= (0, 0, 0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1, 1, 1)^\top, \\
x &= (0, 0, 0, 0, 0.5, 0.5, 0.5, 0.5, 1, 1, 1, 1)^\top, \\
x &= (0, 0, 0, 0, 0, 0.5, 0.5, 1, 1, 1, 1, 1)^\top, \\
x &= (0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1)^\top.
\end{aligned}
$$

What are your conclusions?
Hint: for example, the variance $\sigma^2_{\beta_1}$ can be calculated for the first design as follows:

```
> x<-c(0,0.1,0.2,0.3,0.4,0.5,0.5,0.6,0.7,0.8,0.9,1)
> mean(x^2)/(9*var(x))
```

Testing whether the intercept β_0 is b_0
For testing

$$H_0^0 : \beta_0 = b_0 \quad \text{versus} \quad H_0^0 : \beta_0 \neq b_0,$$

the test statistic is

$$\widehat{d}_0 = \frac{\widehat{\beta}_0 - b_0}{\widehat{\sigma}_{\beta_0}}.$$

It has a t-distribution with $N - 2$ degrees of freedom.

t-test for the intercept

Reject $H_0^0 : \beta_0 = b_0$ if $|\widehat{d}_0| > t_{N-2,1-\frac{\alpha}{2}}$.

Squaring the test statistic leads to a test statistic with F-distribution with 1 and N_2 degrees of freedom.

F-test for the intercept

Reject $H_0^0 : \beta_0 = b_0$ if $\widehat{d}_0^2 > F_{1,N-2,1-\alpha}$.

Testing whether the slope β_1 is b_1
For testing

$$H_0^1 : \beta_1 = b_1 \quad \text{versus} \quad H_0^1 : \beta_1 \neq b_1,$$

the test statistic is

$$\widehat{d}_1 = \frac{\widehat{\beta}_1 - b_1}{\widehat{\sigma}_{\beta_1}}.$$

It has again a t-distribution with $N-2$ degrees of freedom.

t-test for the slope

Reject $H_0^1 : \beta_1 = b_1$ if $|\widehat{d}_1| > t_{N-2,1-\frac{\alpha}{2}}$.

Squaring the test statistic leads also to a test statistic with F-distribution with 1 and N_2 degrees of freedom.

F-test for the slope

Reject $H_0^1 : \beta_1 = b_1$ if $\widehat{d}_1^2 > F_{1,N-2,1-\alpha}$.

The estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are calculated in R with `lsfit` or `summary(lm(...))`. The t-tests for $H_0^0 : \beta_0 = 0$ and $H_0^1 : \beta_1 = 0$ are obtained with `summary(lm(...))` and the F-test for $H_0^1 : \beta_1 = 0$ with `anova(lm(...))`. The requirement of normal distributed errors and residuals can be tested again with `shapiro.test(aov(...))`.

6.1.2 Example (Protein content in ground wheat)
The data file `GROUND.DAT` contains "the results of an experiment to calibrate a near infrared reflectance instrument for the measurement of protein content of ground wheat (German: gemahlener Weizen) samples. The second column shows the protein content, measured by the standard Kjeldahl method. The final six columns show measurements of the reflectance of near infrared radiation of the wheat samples at six wavelengths in range 1680-2310. In the source paper the aim was to find a linear combination of the last six columns which could be used to predict protein content." (Hand et al. 1996, P. 411)

```
> ground<-read.table("GROUND.DAT")
> names(ground)<-c("sample","protein", "L1","L2","L3","L4","L5","L6")
> ground
   sample protein  L1  L2  L3  L4  L5  L6
1       1    9.23 468 123 246 374 386 -11
2       2    8.01 458 112 236 368 383 -15
3       3   10.95 457 118 240 359 353 -16
4       4   11.67 450 115 236 352 340 -15
5       5   10.41 464 119 243 366 371 -16
6       6    9.51 499 147 273 404 433   5
7       7    8.67 463 119 242 370 377 -12
8       8    7.75 462 115 238 370 353 -13
9       9    8.05 488 134 258 393 377  -5
10     10   11.39 483 141 264 384 398  -2
11     11    9.95 463 120 243 367 378 -13
12     12    8.25 456 111 233 365 365 -15
13     13   10.57 512 161 288 415 443  12
14     14   10.23 518 167 293 421 450  19
```

```
15     15   11.87 552 197 324 448 467  32
16     16    8.09 497 146 271 407 451  11
17     17   12.55 592 229 360 484 524  51
18     18    8.38 501 150 274 406 407  11
19     19    9.64 483 137 260 385 374  -3
20     20   11.35 491 147 269 389 391   1
21     21    9.70 463 121 242 366 353 -13
22     22   10.75 507 159 285 410 445  13
23     23   10.75 474 132 255 376 383  -7
24     24   11.47 496 152 276 396 404   6
```

We regard here only the columns `protein` and `L1` where `L1` is the explanatory variable. The estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are obtained by

```
> lsfit(ground$L1,ground$protein)$coef
 Intercept          X
0.24069916 0.01995496
```

or by

```
> summary(lm(protein~L1,data=ground))

Call:
lm(formula = protein ~ L1, data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0683 -0.8799  0.1663  0.9453  2.4496

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.240699   3.938416   0.061   0.9518
L1          0.019955   0.008063   2.475   0.0215 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.282 on 22 degrees of freedom
Multiple R-Squared: 0.2178,     Adjusted R-squared: 0.1822
F-statistic: 6.125 on 1 and 22 DF,  p-value: 0.02151
```

We can conclude that the slope of the regression line differs significantly from zero, i.e. the variable `L1` has a significant influence on the variable `protein`. However, there is no evidence that the regression line has an intercept different from zero. `summary(lm(...))` provides also the estimated variances $\widehat{\sigma}_{\beta_0}$ and $\widehat{\sigma}_{\beta_1}$ in the column `Std.Error` while `anova(lm(...))` produces only the following ANOVA table:

```
> anova(lm(protein~L1,data=ground))
```

```
Analysis of Variance Table

Response: protein
          Df Sum Sq Mean Sq F value  Pr(>F)
L1         1 10.069  10.069   6.125 0.02151 *
Residuals 22 36.165   1.644
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that we get exactly the same p-value for testing $H_0^1 : \beta_1 = \beta_{L1} = 0$. The squared t-value is also the F-value since $2.475^2 = 6.125625$. We should also test whether the requirement of normally distributed errors is satisfied:

```
> shapiro.test(aov(protein~L1,data=ground)$residuals)$p.value
[1] 0.613258
```

The function `lsfit` is useful to draw the estimated line in the scatter plot:

```
> plot(ground$L1,ground$protein,xlab="L1",ylab="Protein")
> abline(lsfit(ground$L1,ground$protein)$coef)
```

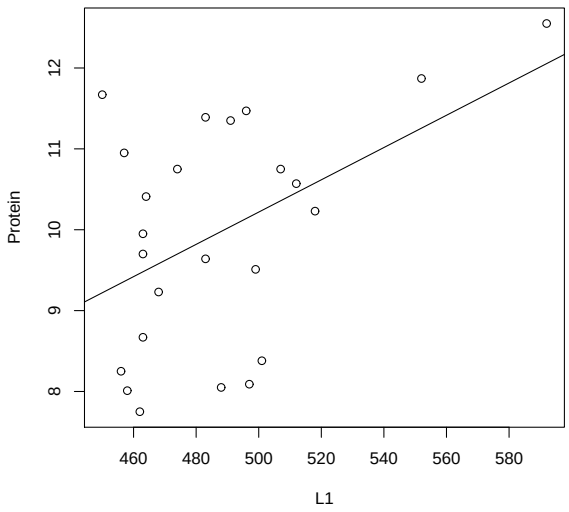

Figure 6.1: Scatter plot with regression line

6.1.3 Exercise (Protein content in ground wheat)
Tests whether the other variables `L2, L3, L4, L5, L6` of the data set `ground` have significant influence on the variable `protein` and test whether the regression lines have an intercept different from zero. Take into account that six tests with the test of Example 6.1.2 are done at the same data set. Check also the normal distribution. Plot the scatter plot with the regression line for the variables `L2` and `protein`.

6.2 Polynomial regression

The dependence of the variable y on the variable x cannot described always by a linear function. Sometimes a better description of the dependence is given by a quadratic or cubic function or even by polynomial function of higher order.

If we assume a polynomial function of order r, then we use the model

$$Y_n = \beta_0 + \beta_1 x_n + \beta_1 x_n^2 + \ldots + \beta_R x_n^R + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2), \tag{11}$$

for all $n = 1, \ldots, N$. The unknown parameters $\beta_0, \beta_1 \ldots, \beta_R$ are estimated again by the method of least squares. This means that their estimates $\widehat{\beta}_0, \widehat{\beta}_1 \ldots, \widehat{\beta}_R$ are those values $\beta_0, \beta_1 \ldots, \beta_R$ for which the sum of squares

$$\sum_{n=1}^{N} (y_n - \beta_0 - \beta_1 x_n - \beta_1 x_n^2 - \ldots - \beta_R x_n^R)^2$$

is as small as possible. There are not any more simple forms for the estimates $\widehat{\beta}_0, \widehat{\beta}_1 \ldots, \widehat{\beta}_R$ as for linear regression. However, if the **design matrix** X is defined as

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^R \\ 1 & x_2 & x_2^2 & \ldots & x_2^R \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_N & x_N^2 & \ldots & x_N^R \end{pmatrix}$$

then the vector $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1 \ldots, \widehat{\beta}_R)^\top$ of estimates is given by

$$\widehat{\beta} = (X^\top X)^{-1} X^\top y.$$

The unknown error variance σ^2 is estimated as before by

$$\widehat{\sigma}^2 = \widehat{\sigma}^2_{SSE} = \frac{1}{N - R - 1} \sum_{n=1}^{N} (y_n - \widehat{\beta}_0 - \widehat{\beta}_1 x_n - \widehat{\beta}_1 x_n^2 - \ldots - \widehat{\beta}_R x_n^R)^2.$$

The covariance matrix of the estimator $\widehat{\beta}$ is

$$\mathrm{Cov}(\widehat{\beta}) = (X^\top X)^{-1} \sigma^2$$

and is estimated by

$$\widehat{\mathrm{Cov}}(\widehat{\beta}) = (X^\top X)^{-1}\widehat{\sigma}^2.$$

The variances for the single estimators $\widehat{\beta}_0$, $\widehat{\beta}_1 \ldots, \widehat{\beta}_R$ and their estimated values are given by the diagonal elements of the covariance matrices $\mathrm{Cov}(\widehat{\beta})$ and $\widehat{\mathrm{Cov}}(\widehat{\beta})$, respectively. In particular the sum of the variances is the trace of the covariance matrix $\mathrm{Cov}(\widehat{\beta})$.

6.2.1 Example (Quadratic regression)
In a quadratic regression model the degree of the polynomial is $R = 2$. Regard the following explanatory variable or design, respectively,:

$$x = (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)^\top$$

The matrix $(X^\top X)^{-1}$ can be calculated in R with the function `ginv` of the contributed library `MASS`:

```
> x<-c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9)
> X<-cbind(rep(1,10),x,x^2)
> library(MASS)
> ginv(t(X)%*%X)
           [,1]       [,2]       [,3]
[1,]  0.6181818  -2.590909   2.272727
[2,] -2.5909091  16.553030 -17.045455
[3,]  2.2727273 -17.045455  18.939394
```

Hence $0.6181818\,\sigma^2$ is the variance of the estimate $\widehat{\beta}_0$, $16.553030\,\sigma^2$ is the variance of the estimate $\widehat{\beta}_1$, and $18.939394\,\sigma^2$ is the variance of the estimate $\widehat{\beta}_2$, for the design given by x. The sum of the variances can be obtained by

```
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 36.11061
```

so that $36.11061\,\sigma^2$ is the sum of the variances.

6.2.2 Exercise (Designs for quadratic regression)
Assume as in Example 6.1.1 that the design region is [0,1] and that $N = 12$. Which constellations of $x_1, \ldots, x_{12} \in [0, 1]$ provides the smallest sum of variances? Try for example again the first six designs of Example 6.1.1 for x. Note that the 7th design cannot be used for quadratic regression since a quadratic function cannot be determined with measurements only at two different points. What is your proposal concerning a good design for quadratic regression?

Often it is unknown which degree of the polynomial shall be used. There are two possibilities to find the degree:

1. Start with a high degree and test whether this degree has a significant influence. If it has no significant influence, reduce the degree and test whether this has a significant influence.

And so on. As soon as you have find a degree with significant influence you should use a polynomial model with this degree.

2. Start with a model with low degree where the largest degree has significant influence. Add a degree and test whether it has a significant influence. Stop when the added degree has no significant influence. Use then the polynomial model without this degree.

6.2.3 Example (Protein content in ground wheat)
For the variables `protein` and `L1` of the data set `ground` of Example 6.1.2, we start with a polynomial model with degree 5:

```
> summary(lm(protein~L1+I(L1^2)+I(L1^3)+I(L1^4)+I(L1^5),data=ground))

Call:
lm(formula = protein ~ L1 + I(L1^2) + I(L1^3) + I(L1^4) + I(L1^5),
    data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9889 -0.8412  0.1932  0.9740  1.5638

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.625e+05  2.311e+05   1.136    0.271
L1          -2.568e+03  2.266e+03  -1.133    0.272
I(L1^2)      1.003e+01  8.871e+00   1.131    0.273
I(L1^3)     -1.954e-02  1.732e-02  -1.128    0.274
I(L1^4)      1.900e-05  1.688e-05   1.126    0.275
I(L1^5)     -7.374e-09  6.563e-09  -1.124    0.276

Residual standard error: 1.324 on 18 degrees of freedom
Multiple R-Squared: 0.3177,     Adjusted R-squared: 0.1282
F-statistic: 1.677 on 5 and 18 DF,  p-value: 0.1912
```

In the column `Estimate` are the estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$ $\ldots$, $\widehat{\beta}_5$ defined as above. They can be also obtained with `lsfit`:

```
> lsfit(cbind(ground$L1,ground$L1^2,ground$L1^3,ground$L1^4,ground$L1^5),
+ ground$protein)$coef
    Intercept            X1            X2            X3            X4
 2.625338e+05 -2.568146e+03  1.002919e+01 -1.954335e-02  1.900193e-05
           X5
-7.374407e-09
```

Since the term `L1^5` has no significant influence, we reduce the model to a polynomial of degree 4 and then further and further:

```
> summary(lm(protein~L1+I(L1^2)+I(L1^3)+I(L1^4),data=ground))
```

```
Call:
lm(formula = protein ~ L1 + I(L1^2) + I(L1^3) + I(L1^4), data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8092 -1.0254  0.1343  0.9793  1.8553

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.206e+03  1.178e+04   0.272    0.788
L1          -2.379e+01  9.173e+01  -0.259    0.798
I(L1^2)      6.610e-02  2.671e-01   0.247    0.807
I(L1^3)     -8.132e-05  3.447e-04  -0.236    0.816
I(L1^4)      3.742e-08  1.664e-07   0.225    0.824

Residual standard error: 1.333 on 19 degrees of freedom
Multiple R-Squared: 0.2699,     Adjusted R-squared: 0.1162
F-statistic: 1.756 on 4 and 19 DF,  p-value: 0.1795

> summary(lm(protein~L1+I(L1^2)+I(L1^3),data=ground))

Call:
lm(formula = protein ~ L1 + I(L1^2) + I(L1^3), data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-1.84300 -1.04798  0.06961  0.94494  1.81930

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.623e+02  7.808e+02   0.720    0.480
L1          -3.190e+00  4.571e+00  -0.698    0.493
I(L1^2)      6.076e-03  8.884e-03   0.684    0.502
I(L1^3)     -3.811e-06  5.728e-06  -0.665    0.513

Residual standard error: 1.301 on 20 degrees of freedom
Multiple R-Squared: 0.2679,     Adjusted R-squared: 0.1581
F-statistic:  2.44 on 3 and 20 DF,  p-value: 0.09425

> summary(lm(protein~L1+I(L1^2)+I(L1^3),data=ground))

Call:
lm(formula = protein ~ L1 + I(L1^2) + I(L1^3), data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-1.84300 -1.04798  0.06961  0.94494  1.81930
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.623e+02  7.808e+02   0.720    0.480
L1          -3.190e+00  4.571e+00  -0.698    0.493
I(L1^2)      6.076e-03  8.884e-03   0.684    0.502
I(L1^3)     -3.811e-06  5.728e-06  -0.665    0.513

Residual standard error: 1.301 on 20 degrees of freedom
Multiple R-Squared: 0.2679,     Adjusted R-squared: 0.1581
F-statistic:  2.44 on 3 and 20 DF,  p-value: 0.09425

> summary(lm(protein~L1+I(L1^2),data=ground))

Call:
lm(formula = protein ~ L1 + I(L1^2), data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-1.85132 -1.03552  0.01018  0.87267  2.08337

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 43.7098394 44.7032358   0.978    0.339
L1          -0.1508808  0.1751872  -0.861    0.399
I(L1^2)      0.0001668  0.0001708   0.976    0.340

Residual standard error: 1.284 on 21 degrees of freedom
Multiple R-Squared: 0.2517,     Adjusted R-squared: 0.1805
F-statistic: 3.532 on 2 and 21 DF,  p-value: 0.0476

> summary(lm(protein~L1,data=ground))

Call:
lm(formula = protein ~ L1, data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0683 -0.8799  0.1663  0.9453  2.4496

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.240699   3.938416   0.061   0.9518
L1          0.019955   0.008063   2.475   0.0215 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.282 on 22 degrees of freedom
```

```
Multiple R-Squared: 0.2178,     Adjusted R-squared: 0.1822
F-statistic: 6.125 on 1 and 22 DF,  p-value: 0.02151
```

Hence the linear regression model is the correct model. We also see that the t-values and the p-values changes every time when the model is refitted. In particular the linear term is not significant in the polynomial models with degree larger than 1, but is significant in the linear regression model.

For large degrees, there can be problems with numerical stability. Orthogonal polynomials get around this problem by using

$$
\begin{aligned}
z_1 &= a_1 + b_1\, x,\\
z_2 &= a_2 + b_2\, x + c_2\, x^2,\\
z_3 &= a_3 + b_3\, ^x + c_3\, x^2 + d_4\, x^3,
\end{aligned}
$$

etc. where the coefficinets $a, b, c, , \ldots$ are chosen such that $z_r^\top z_s = 0$ when $r \neq s$. $z_1, z_2, z_3, \ldots, z_R$ are called orthogonal polynomials. If $z_r = (z_{1r}, \ldots, z_{Nr})^\top$ for $r = 1, \ldots, R$, then we have a new parametrization of the model (11):

$$Y_n = \alpha_0 + \alpha_1\, z_{n1} + \alpha_2\, z_{n2} + \ldots + \alpha_R\, z_{nR} + Z_n. \tag{12}$$

The design matrix of this parametrization is

$$
Z = \begin{pmatrix}
1 & z_{11} & z_{12} & \ldots & z_{1R}\\
1 & z_{21} & z_{22} & \ldots & z_{2R}\\
\vdots & \vdots & \vdots & & \vdots\\
1 & z_{N1} & z_{N2} & \ldots & z_{NR}
\end{pmatrix}
$$

and satisfies that $Z^\top Z$ is a diagonal matrix. This means that the estimates $\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_R$ for $\alpha_1, \alpha_2, \ldots, \alpha_R$ do not depend on the degree R of the model since $\widehat{\alpha} = (\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_R)^\top$ is again given by

$$\widehat{\alpha} = (Z^\top Z)^{-1} Z^\top y.$$

The `poly()` function constructs orthogonal polynomials.

6.2.4 Example (Protein content in ground wheat: Continuation of Example 6.2.3)

```
> summary(lm(protein~poly(L1,5),data=ground))
```

```
Call:
lm(formula = protein ~ poly(L1, 5), data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9889 -0.8412  0.1932  0.9740  1.5638

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.9663     0.2702  36.882   <2e-16 ***
poly(L1, 5)1   3.1731     1.3238   2.397   0.0276 *
poly(L1, 5)2   1.2530     1.3238   0.946   0.3564
poly(L1, 5)3  -0.8655     1.3238  -0.654   0.5215
poly(L1, 5)4   0.2997     1.3238   0.226   0.8234
poly(L1, 5)5  -1.4875     1.3238  -1.124   0.2759
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.324 on 18 degrees of freedom
Multiple R-Squared: 0.3177,     Adjusted R-squared: 0.1282
F-statistic: 1.677 on 5 and 18 DF,  p-value: 0.1912
```

Here we see at once that only the linear term has a significant influence so that the linear model is the correct model. Now the estimates in column `Estimate` are the estimates $\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_5$ for the orthogonal polynomials. Note that they differ from the estimates $\widehat{\beta}_0$, $\widehat{\beta}_1 \ldots, \widehat{\beta}_5$. However , the estimates $\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3, \widehat{\alpha}_4$ do not change if a model of degree 4 is used:

```
> summary(lm(protein~poly(L1,4),data=ground))

Call:
lm(formula = protein ~ poly(L1, 4), data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8092 -1.0254  0.1343  0.9793  1.8553

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.9663     0.2721  36.630   <2e-16 ***
poly(L1, 4)1   3.1731     1.3329   2.381   0.0279 *
poly(L1, 4)2   1.2530     1.3329   0.940   0.3590
poly(L1, 4)3  -0.8655     1.3329  -0.649   0.5239
poly(L1, 4)4   0.2997     1.3329   0.225   0.8245
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.333 on 19 degrees of freedom
Multiple R-Squared: 0.2699,     Adjusted R-squared: 0.1162
```

```
F-statistic: 1.756 on 4 and 19 DF,  p-value: 0.1795
```

Only the p-values changes slightly since the variance estimates $\widehat{\sigma}^2_{SSE}$ changes slightly because the estimate for `poly(L1, 5)5`, i.e. for α_5, is with -1.4875 not zero but rather small. The same p-values are obtained also with the ANOVA approach:

```
> anova(lm(protein~L1+I(L1^2)+I(L1^3)+I(L1^4),data=ground))
Analysis of Variance Table

Response: protein
          Df Sum Sq Mean Sq F value  Pr(>F)
L1         1 10.069  10.069  5.6673 0.02791 *
I(L1^2)    1  1.570   1.570  0.8836 0.35901
I(L1^3)    1  0.749   0.749  0.4216 0.52391
I(L1^4)    1  0.090   0.090  0.0506 0.82448
Residuals 19 33.756   1.777
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Once one has found an appropriate model, one would like to plot the estimated polynomial function in the scatter plot. This is done by plotting estimated functions values

$$\widehat{y}_m = \widehat{f}(\chi_m) = \widehat{\beta}_0 + \widehat{\beta}_1 \chi_m + \widehat{\beta}_2 \chi_m^2 + \ldots + \widehat{\beta}_R \chi_m^R$$

at several points $\chi_1, \ldots, \chi_M$ of the range of $x_1, \ldots, x_N$. For these points, a design matrix can be created as

$$\mathcal{X} = \begin{pmatrix} 1 & \chi_1 & \chi_1^2 & \cdots & \chi_1^R \\ 1 & \chi_2 & \chi_2^2 & \cdots & \chi_2^R \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \chi_M & \chi_M^2 & \cdots & \chi_M^R \end{pmatrix}$$

Then the vector $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_M)^\top$ of the M estimated functions values at the points $\chi_1, \ldots, \chi_M$ is given by

$$\widehat{y} = \mathcal{X}\widehat{\beta}.$$

6.2.5 Example (Protein content in ground wheat: Continuation of Example 6.2.4)
Although the quadratic term is not significant, we plot the estimated quadratic function and compare this function with the estimated linear function:

```
> plot(ground$L1,ground$protein,xlab="L1",ylab="Protein")
> abline(lsfit(ground$L1,ground$protein)$coef)
> x<-seq(440,600,by=2)
> X<-cbind(rep(1,length(x)),x,x^2)
> beta<-lsfit(cbind(ground$L1,ground$L1^2),ground$protein)$coef
> y<-X%*%beta
> lines(x,y)
```

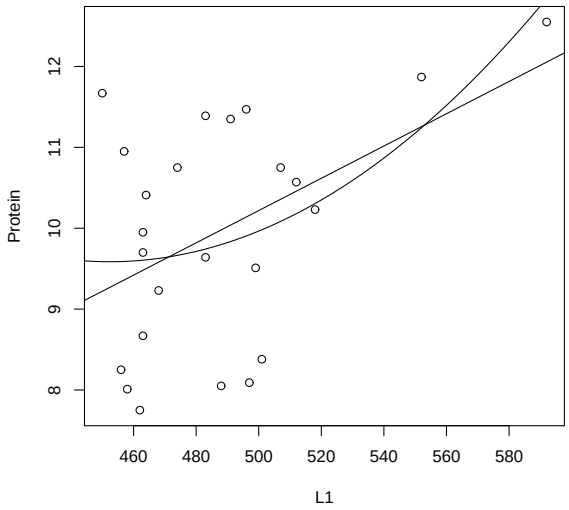

Figure 6.2: Scatter plot with linear and quadratic regression line

6.2.6 Exercise (Split)
Regard the data set `split` from Exercise 1.5.1 and consider the variable `Yield` as dependent variable and the variable `Manure` as explanatory variable, i.e. neglect the variables `Block` and `Variety`. Find an appropriate polynomial model for the dependence of `Yield` on `Manure`. Find the model with and without the function `poly`. What are the estimates for the parameters of the model? Plot the estimated function in the scatter plot and compare the result with the linear regression line.

6.3 Multiple regression

The Subsections 6.1 and 6.2 allowed only one explanatory variable. But often there several explanatory variables as in the Example 6.1.2. Let $x_{\bullet 1} = (x_{11}, \ldots, x_{N1})^2$, $x_{\bullet 2} = (x_{12}, \ldots, x_{N2})^2$, ..., $x_{\bullet R} = (x_{1R}, \ldots, x_{NR})^2$ denote R explanatory variables.

Multiple Regression without interactions
The model without interactions, also called additive model, is given by

$$Y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_R x_{nR} + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2),$$

for all $n = 1, \ldots, N$. Estimates $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_R$ are given again by those values $\beta_0, \beta_1, \beta_2, \ldots, \beta_R$ which minimizes the sum of squares

$$\sum_{n=1}^{N} (y_n - \beta_1 x_{n1} - \beta_2 x_{n2} - \ldots - \beta_R x_{nR})^2$$

and the estimate for the error variance σ^2 is

$$\widehat{\sigma}^2 = \widehat{\sigma}^2_{SSE} = \frac{1}{N-R-1} \sum_{n=1}^{N} (y_n - \widehat{\beta}_1 x_{n1} - \widehat{\beta}_2 x_{n2} - \ldots - \widehat{\beta}_R x_{nR})^2.$$

The vector $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_R)^\top$ of parameter estimates can be calculated as before by

$$\widehat{\beta} = (X^\top X)^{-1} X^\top y$$

where the design matrix X is here

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1R} \\ 1 & x_{21} & x_{22} & \ldots & x_{2R} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N1} & x_{N2} & \ldots & x_{NR} \end{pmatrix}.$$

As soon as the columns of the design matrix are orthogonal to each other which can be satisfied by special designs then $X^\top X$ is a diagonal matrix and the parameter estimates are $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_R$ independent of the number R of explanatory variables. I.e. for a smaller model with a smaller R we would get the same estimates. Moreover, the order of the tests

$$\begin{aligned} &H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0, \\ &H_0 : \beta_2 = 0 \quad \text{versus} \quad H_1 : \beta_2 \neq 0, \\ &\vdots \\ &H_0 : \beta_R = 0 \quad \text{versus} \quad H_1 : \beta_R \neq 0, \end{aligned}$$

does not depend on the order of the explanatory variables.

However, in the general case all this is not true. Hence we should take into account in which model an estimator $\widehat{\beta}_r$ is obtained and what are the null and alternative hypotheses which are tested. The

function `summary(lm(...))` provides the t-tests for the following hypotheses:

$$
\begin{aligned}
0)\quad & H_0^0 : \mu_n = \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR} \\
& \qquad \text{versus} \\
& H_1^0 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR}, \\
1)\quad & H_0^1 : \mu_n = \beta_0 + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR} \\
& \qquad \text{versus} \\
& H_1^1 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR}, \\
2)\quad & H_0^2 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \ldots + \beta_R\, x_{nR} \\
& \qquad \text{versus} \\
& H_1^2 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR}, \\
& \vdots \\
r)\quad & H_0^r : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_{R-1}\, x_{n(R-1)} \\
& \qquad \text{versus} \\
& H_1^r : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR}.
\end{aligned}
$$

Additionally it provides in the last line of the output the p-value for testing

$$
H_0 : \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_R \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{versus} \quad H_1 : \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_R \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
$$

The function `anova(lm(...))` provides the F-tests for the following sequential hypotheses:

$$
\begin{aligned}
1)\quad & H_0^1 : \mu_n = \beta_0 \\
& \qquad \text{versus} \\
& H_1^1 : \mu_n = \beta_0 + \beta_1, \\
2)\quad & H_0^2 : \mu_n = \beta_0 + \beta_1\, x_{n1} \\
& \qquad \text{versus} \\
& H_1^2 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2}, \\
3)\quad & H_0^3 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} \\
& \qquad \text{versus} \\
& H_1^3 : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \beta_3\, x_{n3}, \\
& \vdots \\
r)\quad & H_0^r : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_{R-1}\, x_{n(R-1)} \\
& \qquad \text{versus} \\
& H_1^r : \mu_n = \beta_0 + \beta_1\, x_{n1} + \beta_2\, x_{n2} + \ldots + \beta_R\, x_{nR}.
\end{aligned}
$$

We see that the last hypotheses coincide so that the p-values should be the same. However, the other tests are different. Moreover, the ANOVA tests depend very much on the order of the explanatory variables.

6.3.1 Example (Protein content in ground wheat: Continuation of Example 6.1.2)
Here we regard additionally the explanatory variables `L2` and `L3` besides `L1`.

```
> summary(lm(protein~L1+L2+L3,data=ground))

Call:
lm(formula = protein ~ L1 + L2 + L3, data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-1.09579 -0.29380  0.02471  0.29865  1.24967

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 87.14739   12.53757   6.951 9.52e-07 ***
L1          -0.32804    0.04195  -7.820 1.66e-07 ***
L2           0.17296    0.12302   1.406    0.175
L3           0.22074    0.13928   1.585    0.129
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5607 on 20 degrees of freedom
Multiple R-Squared: 0.864,      Adjusted R-squared: 0.8436
F-statistic: 42.36 on 3 and 20 DF,  p-value: 7.484e-09
```

We see from the last line that $\beta = (\beta_0, \beta_{L1}, \beta_{L2}, \beta_{L3})^\top$ differs significantly from zero in the model with explanatory variables `L1`, `L2`, `L3`. But the analysis for the single variables shows that only the variable `L1` has significant influence.

```
> anova(lm(protein~L1+L2+L3,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq F value    Pr(>F)
L1         1 10.0688 10.0688 32.0322 1.539e-05 ***
L2         1 29.0890 29.0890 92.5423 6.046e-09 ***
L3         1  0.7896  0.7896  2.5118    0.1287
Residuals 20  6.2866  0.3143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here we can conclude that the variable `L2` has a significant influence in the model with the explanatory variables `L1` and `L2` and variable `L1` has a significant influence in the model with the only explanatory variables `L1`. To see whether variable `L1` has also a significant influence in the model with the explanatory variables `L1` and `L2`, wee need again `summary(lm(...))`:

```
> summary(lm(protein~L1+L2,data=ground))

Call:
lm(formula = protein ~ L1 + L2, data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-1.11384 -0.27161  0.01101  0.29607  1.49950

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 98.37271   10.71122   9.184 8.42e-09 ***
L1          -0.28482    0.03300  -8.630 2.39e-08 ***
L2           0.35876    0.03861   9.291 6.91e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5805 on 21 degrees of freedom
Multiple R-Squared: 0.8469,     Adjusted R-squared: 0.8324
F-statistic:  58.1 on 2 and 21 DF,  p-value: 2.759e-09
```

Indeed both variables have significant influence in the model with the two variables `L1` and `L2`.

What happens with the ANOVA tests when we change the order of the variables:

```
> anova(lm(protein~L2+L1+L3,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq F value    Pr(>F)
L2         1 14.0641 14.0641 44.7427 1.645e-06 ***
L1         1 25.0937 25.0937 79.8318 2.029e-08 ***
L3         1  0.7896  0.7896  2.5118    0.1287
Residuals 20  6.2866  0.3143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence this result is similar to that before. But we get a completely different result when we type:

```
> anova(lm(protein~L2+L3+L1,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq F value    Pr(>F)
L2         1 14.0641 14.0641  44.743 1.645e-06 ***
L3         1  6.6634  6.6634  21.199 0.0001717 ***
L1         1 19.2199 19.2199  61.145 1.655e-07 ***
Residuals 20  6.2866  0.3143
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This means that L1 has significant influence in the model with the variables L1, L2, L3, variable L3 has significant influence in the model with variables L2, L3, and variable L2 has significant influence in the model with only variable L2.

```
> summary(lm(protein~L2+L3,data=ground))

Call:
lm(formula = protein ~ L2 + L3, data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-2.18978 -0.56816 -0.06679  0.77975  1.78713

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  61.7833    23.8062   2.595   0.0169 *
L2            0.5487     0.2226   2.465   0.0224 *
L3           -0.4873     0.2080  -2.342   0.0291 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.102 on 21 degrees of freedom
Multiple R-Squared: 0.4483,     Adjusted R-squared: 0.3958
F-statistic: 8.533 on 2 and 21 DF,  p-value: 0.001940
```

It seems that it was not a good idea to drop variable L1 from the model. We have done 6 tests which means that we should use $\alpha = 0.05/6 = 0.008333333$ so that L2 and L3 has no significant influence. Since the other p-values are so small, we are here glad that we still have some significant results. But doing so many test is very dangerous.

The strategy for testing in the additive multiple regression model
1. Choose an order of the variables so that the first variables have the highest chances to have an influence in your opinion.
2. Then do only one ANOVA test.
3. Drop all the last variables from the model which show no significant influence.
4. Test in the reduced model with `summary(lm(...))` whether β differs significantly from zero.
5. If β differs significantly from zero, then test with `summary(lm(...))` for significant influence of the single variables.

6.3.2 Exercise (Protein content in ground wheat: Continuation of Example 6.1.2)
Regard now all variables L1, L2, L3, L4, L5, L6 as explanatory variables and assume that you expect

decreasing influence of the variables with increasing number, i.e. you expect that variable `L1` has the most influence. Do the analysis within an additive model.

Multiple Regression with first order interactions
The model with first order interactions is given by

$$Y_n = \beta_0 + \sum_{r=1}^{R} \beta_r \, x_{nr} + \sum_{r=1}^{R} \sum_{s=r+1}^{R} \beta_{rs} \, x_{nr} \, x_{ns} + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2),$$

for all $n = 1, \ldots, N$. This model has $1 + R + R(R-1)/2$ model parameters $\beta_0, \beta_1, \ldots, \beta_r, \beta_{12}, \ldots, \beta_{(R-1)R}$. Estimates $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_R, \widehat{\beta}_{12}, \ldots, \widehat{\beta}_{1R}, \ldots, \widehat{\beta}_{(R-1)R}$ are given again by those values $\beta_0, \beta_1, \ldots, \beta_r, \beta_{12}, \ldots, \beta_{(R-1)R}$ which minimizes the sum of squares

$$\sum_{n=1}^{N} \left(y_n - \sum_{r=1}^{R} \beta_r \, x_{nr} - \sum_{r=1}^{R} \sum_{s=r+1}^{R} \beta_{rs} \, x_{nr} \, x_{ns} \right)^2$$

and the estimate for the error variance σ^2 is

$$\widehat{\sigma}^2 = \widehat{\sigma}^2_{SSE} = \frac{1}{N - R - R(R-1)/2 - 1} \sum_{n=1}^{N} \left(y_n - \sum_{r=1}^{R} \widehat{\beta}_r \, x_{nr} - \sum_{r=1}^{R} \sum_{s=r+1}^{R} \widehat{\beta}_{rs} \, x_{nr} \, x_{ns} \right)^2.$$

Multiple Regression with first order interactions and quadratic terms
In the model we can use additionally quadratic terms besides the first order interactions so that the model is given by

$$Y_n = \beta_0 + \sum_{r=1}^{R} \beta_r \, x_{nr} + \sum_{r=1}^{R} \sum_{s=r}^{R} \beta_{rs} \, x_{nr} \, x_{ns} + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2),$$

for all $n = 1, \ldots, N$. Here we have $1 + R + R^2 - R(R-1)/2$ model parameters. These are estimated as before with the method of least squares.

Multiple Regression with higher order interactions
Besides first order interactions also higher order interactions can be included in the model. If interactions up to the order S are included, then the model is given by

$$\begin{aligned} Y_n = \beta_0 &+ \sum_{r=1}^{R} \beta_r \, x_{nr} + \sum_{r=1}^{R} \sum_{s=r+1}^{R} \beta_{rs} \, x_{nr} \, x_{ns} + \ldots \\ &+ \sum_{r_1=1}^{R} \sum_{r_2=r+1}^{R} \cdots \sum_{r_S=r+S-1}^{R} \beta_{r_1 r_2 \ldots r_S} \, x_{nr_1} \, x_{nr_2} \, \ldots, x_{nr_S} + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2), \end{aligned}$$

for all $n = 1, \ldots, N$. This model has

$$\binom{R}{0} + \binom{R}{1} + \binom{R}{2} + \binom{R}{3} + \ldots + \binom{R}{S} = 1 + R + \frac{R(R-1)}{2} + \frac{R!}{3!(R-3)!} + \ldots + \frac{R!}{S!(R-S)!}$$

model parameters.

General multiple regression
Also models with arbitrary combinations of specific interactions and higher order terms can be used. But note that only a model can be used where the number of model parameters is less than the sample size N.

6.3.3 Example (Protein content in ground wheat: Continuation of Example 6.3.1)
Let us regard as explanatory variables the variables `L1`, `L2`, `L3`, `L4`. If we want to use a model with all higher order interactions we can use

```
> anova(lm(protein~L1*L2*L3*L4,data=ground))
Analysis of Variance Table

Response: protein
            Df  Sum Sq Mean Sq  F value    Pr(>F)
L1           1 10.0688 10.0688 134.1987 2.803e-06 ***
L2           1 29.0890 29.0890 387.7053 4.605e-08 ***
L3           1  0.7896  0.7896  10.5234   0.01181 *
L4           1  5.2074  5.2074  69.4053 3.258e-05 ***
L1:L2        1  0.0292  0.0292   0.3894   0.55000
L1:L3        1  0.0139  0.0139   0.1849   0.67853
L2:L3        1  0.0001  0.0001   0.0012   0.97344
L1:L4        1  0.0884  0.0884   1.1788   0.30923
L2:L4        1  0.1017  0.1017   1.3549   0.27796
L3:L4        1  0.0579  0.0579   0.7718   0.40527
L1:L2:L3     1  0.0575  0.0575   0.7664   0.40684
L1:L2:L4     1  0.0001  0.0001   0.0016   0.96867
L1:L3:L4     1  0.0645  0.0645   0.8597   0.38095
L2:L3:L4     1  0.0628  0.0628   0.8373   0.38692
L1:L2:L3:L4  1  0.0029  0.0029   0.0385   0.84925
Residuals    8  0.6002  0.0750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If a model with first order interactions and quadratic terms should be used, then type:

```
> anova(lm(protein~L1+L2+L3+L4+I(L1^2)+I(L2^2)+I(L3^2)+I(L4^2)+
+ L1:L2+L1:L3+L1:L4+L2:L3+L2:L4+L3:L4,data=ground))
Analysis of Variance Table

Response: protein
          Df    Sum Sq   Mean Sq   F value    Pr(>F)
```

```
L1          1   10.0688   10.0688  122.3152 1.538e-06 ***
L2          1   29.0890   29.0890  353.3735 1.567e-08 ***
L3          1    0.7896    0.7896    9.5915   0.01278 *
L4          1    5.2074    5.2074   63.2594 2.319e-05 ***
I(L1^2)     1    0.0265    0.0265    0.3219   0.58433
I(L2^2)     1    0.0150    0.0150    0.1824   0.67933
I(L3^2)     1 2.211e-06 2.211e-06 2.686e-05   0.99598
I(L4^2)     1    0.0376    0.0376    0.4569   0.51606
L1:L2       1    0.1679    0.1679    2.0400   0.18698
L1:L3       1    0.0872    0.0872    1.0588   0.33035
L1:L4       1    0.0011    0.0011    0.0134   0.91053
L2:L3       1    0.0005    0.0005    0.0066   0.93687
L2:L4       1    0.0023    0.0023    0.0277   0.87159
L3:L4       1    0.0003    0.0003    0.0031   0.95713
Residuals   9    0.7409    0.0823
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The analysis of both extended models shows that the additive model is already an appropriate model.

6.3.4 Exercise (Protein content in ground wheat: Continuation of Example 6.3.3)
Regard now all explanatory variables `L1`, `L2`, `L3`, `L4`, `L5`, `L6` and consider the models

`protein`∼`L1*L2*L3*L4*L5*L6`,
`protein`∼`L1*L2*L3*L4*L5+L6`,
`protein`∼`L1*L2*L3*L4+L5*L6`,
`protein`∼`L1*L2*L3+L4*L5*L6`,
`protein`∼`L1*L2*L3*L4+L5+L6`,
`protein`∼`L1*L2*L3+L4+L5+L6`.

Determine for each model the number of model parameters. Which is the largest model which can be used? Explain why some models cannot be used. Determine for the models which can be used appropriate submodels.

6.4 Analysis of covariance

If there are numeric explanatory variables and explanatory variables which are factors then the Analysis of Variance is also called Analysis of Covariance (ANCOVA). The numeric explanatory variables are also called **covariates**. For the analysis in R, there is nothing new except the requirement that block factors should be given at first and that they are usually given as additive variables. Treatment variables should be given at last and it makes sense to use interactions between the treatment and the numeric variables. It can be that the dependence of the measurement on the numeric explanatory variable is different for different treatment levels.

If a linear regression line describes the dependence of the measurement on the numeric explanatory variable, then a significant effect of a factor variable means that the intercept of the regression lines is different for the different levels of the factor. A significant interaction between factor and numeric explanatory variable means that the slopes of the regressions lines are different for the different

levels of the factor.

Example: Model for one covariate with linear influence and one factor
The model for one covariate with linear influence, one factor A with A levels, and with interactions between A and the covariate has the form

$$Y_n = \mu + \alpha_a + \beta\, x_n + \gamma_a\, x_n + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2), \tag{13}$$

for all $n = 1, \ldots, N$. Thereby, μ is the average mean, $\alpha_1, \ldots, \alpha_A$ are the main effects of the factor A, β_1 is the main slope of the regression line, and $\gamma_1, \ldots, \gamma_A$ are the interactions between covariate and factor A. Again we have

$$\sum_{a=1}^{A} \alpha_a = 0 \quad \sum_{a=1}^{A} \gamma_a = 0,$$

which means

$$-\alpha_1 = \sum_{a=2}^{A} \alpha_a, \quad -\gamma_1 = \sum_{a=2}^{A} \gamma_a.$$

Setting $\beta_0 = \mu - \alpha_1$ and $\beta_1 = \beta - \gamma_1$, the model (13) is equivalent with

$$Y_n = \beta_0 + \alpha_a + \beta_1\, x_n + \gamma_a\, x_n + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2),$$

for all $n = 1, \ldots, N$. Here, β_0 and β_1 are the intercept and the slope of the regression line for factor level $a = 1$, and $\beta_0 + \alpha_a$ and $\beta_1 + \gamma_a$ are the intercept and slope of the factor levels $a = 2, \ldots, A$.

6.4.1 Example (Ground cover under apple trees)
The data in data file `APPLE.DAT`, "which were first published by Professor Pearce in 1953, come from an experiment to study the best way of forming ground cover in an apple plantation. Treatment O represents what was the usual treatment, keeping the land clear during the growing season but letting the weeds grow up towards the end. Treatments A,B,C,D and E represent the growing of various permanent crops (German: Feldfrüchte) under the trees. There were four randomized blocks. The response Y was the total crop weight in pounds over a four-year period after the treatments were begun. The trees were old and their crop seizes would be likely to vary considerable from one tree to the next. However, records were available of cropping before the experiment began. These were used to provide a covariate X, the total volume of crop bushels over a four-year period before the new treatments began." (Hand et al. 1996, P. 66)

```
> apple0<-read.table("APPLE.DAT")
> apple1<-c(apple0[,1],apple0[,3],apple0[,5],apple0[,7])
> apple2<-c(apple0[,2],apple0[,4],apple0[,6],apple0[,8])
> apple<-data.frame(c(rep("1",6),rep("2",6),rep("3",6),rep("4",6)),
+ rep(c("A","B","C","D","E","O"),4),apple1,apple2)
> names(apple)<-c("Block","Treat","X","Y")
> apple
```

```
   Block Treat    X   Y
1      1     A  8.2 287
2      1     B  8.2 271
3      1     C  6.8 234
4      1     D  5.7 189
5      1     E  6.1 210
6      1     O  7.6 222
7      2     A  9.4 290
8      2     B  6.0 209
9      2     C  7.0 210
10     2     D  5.5 205
11     2     E  7.0 276
12     2     O 10.1 301
13     3     A  7.7 254
14     3     B  9.1 243
15     3     C  9.7 286
16     3     D 10.2 312
17     3     E  8.7 279
18     3     O  9.0 238
19     4     A  8.5 307
20     4     B 10.1 348
21     4     C  9.9 371
22     4     D 10.3 375
23     4     E  8.1 344
24     4     O 10.5 357
```

At first we do an analysis without the covariate:

```
> anova(lm(Y~Block+Treat,data=apple))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
Block      3  47853   15951  10.211 0.0006492 ***
Treat      5    750     150   0.096 0.9914606
Residuals 15  23432    1562
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that there is no treatment effect. Now we add the covariate X in the analysis:

```
> anova(lm(Y~Block+X*Treat,data=apple))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
Block      3  47853   15951  80.819 7.913e-07 ***
```

```
X          1  15944   15944  80.781 8.633e-06 ***
Treat      5   4353     871   4.411   0.02622 *
X:Treat    5   2109     422   2.137   0.15207
Residuals  9   1776     197
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now there is a significant treatment effect. We can also look at the estimated effects:

```
> summary(lm(Y~Block+X*Treat,data=apple))

Call:
lm(formula = Y ~ Block + X * Treat, data = apple)

Residuals:
    Min      1Q  Median      3Q     Max
-17.902  -6.654   0.775   6.135  16.891

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  317.699    111.672   2.845  0.01925 *
Block2         3.147      9.386   0.335  0.74509
Block3       -39.750     11.434  -3.476  0.00698 **
Block4        32.134     11.715   2.743  0.02273 *
X             -3.797     13.127  -0.289  0.77895
TreatB      -267.757    129.926  -2.061  0.06938 .
TreatC      -360.563    130.653  -2.760  0.02212 *
TreatD      -304.252    121.095  -2.513  0.03317 *
TreatE      -350.886    137.274  -2.556  0.03088 *
TreatO      -344.749    126.273  -2.730  0.02322 *
X:TreatB      30.015     15.367   1.953  0.08254 .
X:TreatC      42.028     15.456   2.719  0.02364 *
X:TreatD      36.342     14.385   2.526  0.03243 *
X:TreatE      45.476     16.809   2.706  0.02417 *
X:TreatO      36.879     14.576   2.530  0.03222 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.05 on 9 degrees of freedom
Multiple R-Squared: 0.9753,     Adjusted R-squared: 0.937
F-statistic: 25.43 on 14 and 9 DF,  p-value: 1.623e-05
```

The estimates can be obtained by:

```
> coefficients(lm(Y~Block+X*Treat,data=apple))
(Intercept)      Block2      Block3      Block4           X      TreatB
 317.699450    3.147023  -39.750123   32.133547   -3.796694 -267.757215
```

```
     TreatC      TreatD      TreatE      TreatO    X:TreatB    X:TreatC
-360.563086 -304.252285 -350.885716 -344.748648   30.015275   42.027954
   X:TreatD    X:TreatE    X:TreatO
  36.341832   45.476112   36.879122
```

To plot the regression lines for the different treatments, we drop the factor `block` from the model although it has significant influence:

```
> plot(apple$X,apple$Y,type="n",xlab="X",ylab="Y")
> text(apple$X,apple$Y,as.character(apple$Treat))
> co<-coefficients(lm(Y~X*Treat,data=apple))
> co
(Intercept)           X      TreatB      TreatC      TreatD      TreatE
 126.545752   18.692810  -95.045206 -162.518533 -106.254956 -104.695471
     TreatO    X:TreatB    X:TreatC    X:TreatD    X:TreatE    X:TreatO
-252.128756    9.600537   18.579379   12.847783   15.474376   24.864502
> abline(co[1],co[2])
> abline(co[1]+co[3],co[2]+co[8],lty=2)
> abline(co[1]+co[4],co[2]+co[9],lty=3)
> abline(co[1]+co[5],co[2]+co[10],lty=4)
> abline(co[1]+co[6],co[2]+co[11],lty=5)
> abline(co[1]+co[7],co[2]+co[12],lty=6)
> legend(6,370,c("A","B","C","D","E","O"),lty=1:6)
```

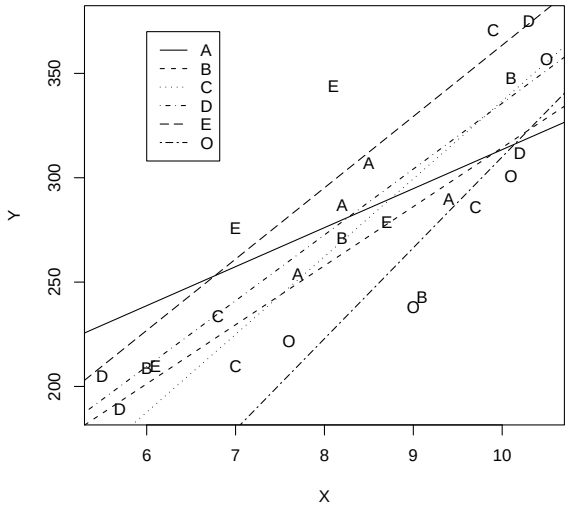

Figure 6.3: Scatter plot with regression lines for the 6 treatments

We see that the 6 regression lines have different slopes and different intercepts. The different slopes are due to the fact that the interactions between the treatment and the covariate X are included in the model. If the interactions are not included, then the regression lines are parallel:

```
> plot(apple$X,apple$Y,type="n",xlab="X",ylab="Y")
> text(apple$X,apple$Y,as.character(apple$Treat))
> co<-coefficients(lm(Y~X+Treat,data=apple))
> co
(Intercept)           X      TreatB      TreatC      TreatD      TreatE
   6.064324   32.950968  -13.454903   -5.954903    3.049258   24.877193
     TreatO
 -33.008322
> abline(co[1],co[2])
> abline(co[1]+co[3],co[2],lty=2)
> abline(co[1]+co[4],co[2],lty=3)
> abline(co[1]+co[5],co[2],lty=4)
> abline(co[1]+co[6],co[2],lty=5)
> abline(co[1]+co[7],co[2],lty=6)
> legend(6,370,c("A","B","C","D","E","O"),lty=1:6)
```

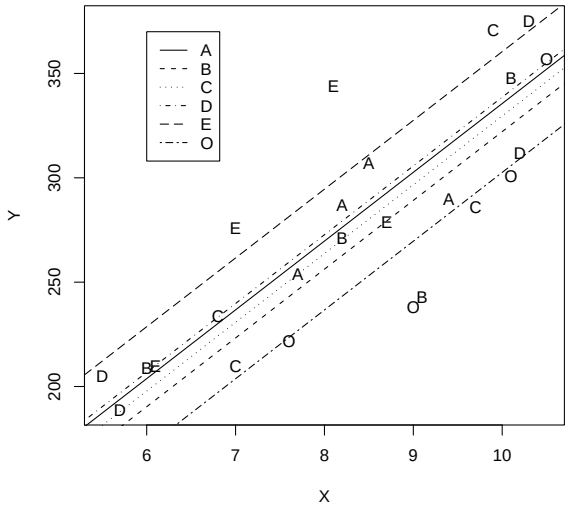

Figure 6.4: Scatter plot with regression lines for the 6 treatments

6.4.2 Exercise (Split: Continuation of Exercise 6.2.6)
Regard in the data set `split` the variables `Yield`, `Manure`, `Variety`. Investigate whether the variety has an influence on the linear dependence of the yield on the manure. Plot the different regression lines for the three varieties also when the varieties have no significant influence and there is no significant interaction between variety and manure.

6.5 Designing regression experiments

Orthogonal polynomials have the advantage that they provide a design matrix Z so that the columns of the matrix Z are orthogonal to each other, so that $Z^\top Z$ is diagonal matrix. This orthogonality property of the design matrix is also the aim of designing multiple regression and the analysis of covariance.

Designing multiple regression
Multiple regression can be designed as soon as the input of explanatory variables are given by the experimenter. This is for example the case when the explanatory variable consists of some drug doses or some concentrations of some fertilizer or insecticide. To derive the orthogonality property it is useful to subtract the mean of the explanatory variables from the variable, i.e. regard $\widetilde{x}_{\bullet r} = (x_{1r} - \overline{x}_{\bullet r}, \ldots, x_{Nr} - \overline{x}_{\bullet r})^2$ instead of $x_{\bullet r} = (x_{1r}, \ldots, x_{Nr})^2$ for $r = 1, \ldots, R$. The model

$$Y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_R x_{nR} + Z_n,$$

becomes then the model

$$\begin{aligned} Y_n &= \beta_0 + \beta_1 \overline{x}_{\bullet 1} + \ldots + \beta_R \overline{x}_{\bullet R} + \beta_1 (x_{n1} - \overline{x}_{\bullet 1}) + \ldots + \beta_R (x_{nR} - \overline{x}_{\bullet R}) + Z_n, \\ &= \widetilde{\beta}_0 + \beta_1 \widetilde{x}_{n1} + \ldots + \beta_R \widetilde{x}_{nR} + Z_n, \end{aligned}$$

so that only the intercept has changed. Then each vector $\widetilde{x}_{\bullet r}$ is orthogonal to the vector $(1, 1, \ldots, 1, 1)^\top$ consisting only of ones. As soon as the inputs of the vectors $x_{\bullet r}$ are chosen such that $\widetilde{x}_{\bullet 1}, \ldots, \widetilde{x}_{\bullet R}$ are mutually orthogonal, i.e. $\widetilde{x}_{\bullet r}^\top \widetilde{x}_{\bullet s}$ for all $r \neq s$, then the design matrix

$$\widetilde{X} = \begin{pmatrix} 1 & \widetilde{x}_{11} & \widetilde{x}_{12} & \ldots & \widetilde{x}_{1R} \\ 1 & \widetilde{x}_{21} & \widetilde{x}_{22} & \ldots & \widetilde{x}_{2R} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \widetilde{x}_{N1} & \widetilde{x}_{N2} & \ldots & \widetilde{x}_{NR} \end{pmatrix}$$

has orthogonal columns and $\widetilde{X}^\top \widetilde{X}$ is a diagonal matrix. This means that each of the parameters $\beta_1, \ldots, \beta_R$ can be estimated independently from the other parameters and that the ANOVA tests does not depend on the order of the variables. Designs with this orthogonality property can be obtained for example by `fact.nk` for factorial designs.

6.5.1 Example
If there are two explanatory variables and each should be realized at 3 points (levels), then use:

```
> x<-fact.nk(3,2,1)
> x
  plots blocks A B
1     1      1 0 0
2     2      1 1 0
3     3      1 2 0
4     4      1 1 1
5     5      1 0 2
6     6      1 2 2
7     7      1 0 1
8     8      1 1 2
9     9      1 2 1
```

The last two columns provide the design points. If the design region is not $[0, 2]$, then the design points must be shifted and scaled appropriately. These design points are given again in a random

order to reduce the influence of unknown factors. To see that the design points indeed provide the orthogonal property, type:

```
> t(x$A-mean(x$A))%*%(x$B-mean(x$B))
     [,1]
[1,]    0
```

Sorting the design point with respect to the first variable, the orthogonality property can be seen already from the design if one has some experience with this:

```
> x[order(x$A),c(3,4)]
  A B
1 0 0
5 0 2
7 0 1
2 1 0
4 1 1
8 1 2
3 2 0
6 2 2
9 2 1
```

Regression designs based on complete factorial designs need like the complete block designs a high sample size N. If there are many explanatory variables than complete factorial designs with even 2 levels for each variable are not possible. Then fractional factorial designs are needed. In these designs not all unknown parameters are estimable since some are confounded with other parameters. But the R package `conf.design` allows the construction of fractional factorial designs where the confounded parameters can be specified.

6.5.2 Exercise (Split: Continuation of Exercise 6.4.2)
Compare the models `Yield~Manure*Variety` and `Yield~Variety*Manure` with respect to the estimators and the p-values of the ANOVA tests. Do the same for the models

`Yield~Block+Manure*Variety`, `Yield~Block+Variety*Manure`
`Yield~Block*Manure*Variety`, `Yield~Variety*Block*Manure`

Regard in the last comparison only the p-values of the ANOVA tests. Does the order of the variables have an influence? Explain the result.

Designing the analysis of covariance
The model for the analysis of covariance is the most general model since in includes numeric explanatory variables as well as factor variables. The question is whether also here a design matrix with orthogonal columns can be created. For that we need to know how to get the columns of the design matrix if the variable is a factor. Note that if the factor has A levels, then there are A main effects $\alpha_1, \ldots, \alpha_A$. Since they should satisfy $\sum_{a=1}^{A} \alpha_a = 0$ there are indeed only $A-1$ free parameters. There are two main possibilities for coding the $A-1$ parameters in the design matrix: treatment coding and Helmert coding. For example for 4 levels the **treatment coding** is

		Dummy coding		
		1	2	3
	1	0	0	0
levels	2	1	0	0
	3	0	1	0
	4	0	0	1

The **Helmert coding** is given by

		Dummy coding		
		1	2	3
	1	-1	-1	-1
levels	2	1	-1	-1
	3	0	2	-1
	4	0	0	3

We see that in both codings the columns are mutually orthogonal and that the columns are orthogonal to the vector $(1,1,1,1)^\top$. Hence a design where the levels 1,2,3,4 are repeated with the same number, M say, provides a design matrix where the corresponding columns are mutually orthogonal and which are orthogonal to the first column consisting of ones. This means that balances designs satisfy the orthogonality property.

A- and D-optimal designs
Besides the orthogonality criterion, there are also other design criteria. These are based on the covariance matrix of the estimators which is proportional to

$$(X^\top X)^{-1},$$

where X is again the design matrix. The sum of the diagonal elements of the matrix $(X^\top X)^{-1}$ is the sum of the variances of the single estimators. Thus minimizing the sum of the diagonal elements of the matrix $(X^\top X)^{-1}$ means that an average of the variances is minimized.

A-optimal designs
If a design minimizes the sum of the diagonal elements of the matrix $(X^\top X)^{-1}$ within all possible designs, then the design is called A-optimal (A from average).

The power of the ANOVA tests however depends on the determinant of $(X^\top X)^{-1}$.

D-optimal designs
If a design minimizes the determinant of the matrix $(X^\top X)^{-1}$ within all possible designs, then the design is called D-optimal (D from determinant).

Exercise 6.2.2 shows that the A-optimal designs for polynomial regression can be rather strange. In general it is not easy to find A- and D-optimal designs. However, A- and D-optimal designs can be found with the R package `AlgDesign`.

7 Multivariate Analysis

7.1 Multivariate analysis of variance (MANOVA)

Up to now, only one measurement variable was considered. But often several measurement variables exists. It is no good strategy to analyze them separately since these are tests at the same data set so that the level of the tests must be adjusted by the number of the tests, i.e.

$$\alpha = \frac{0.05}{\text{number of tests}}$$

must be used as level for the tests. This is avoided by using the multivariate analysis of variance (MANOVA). In R, the analysis is done with `summary(manova(...))`. Its usage is the same as for `anova(lm(...)))` with the exception that several measurement variables should be given.

7.1.1 Example (Diet supplements)
"Fifteen guinea pigs were given a growth inhibiting substance and body weight measurements (in grams) were recorded at the ends of weeks 1,3,4,5,6, and 7. At the beginning of week 5 vitamin E therapy was started, the guinea pigs being divided into three groups of five to receive zero, low, or high doses of vitamin E." (Hand et al. 1996, P. 325)

```
> diet0<-read.table("DIET.DAT")
> diet<-data.frame(diet[,1],c(rep("1",5),rep("2",5),rep("3",5)),diet0[,2:7])
> names(diet)<-c("No","Group","Week1","Week3","Week4","Week5","Week6","Week7")
> diet
   No Group Week1 Week3 Week4 Week5 Week6 Week7
1   1     1   455   460   510   504   436   466
2   2     1   467   565   610   596   542   587
3   3     1   445   530   580   597   582   619
4   4     1   485   542   594   583   611   612
5   5     1   480   500   550   528   562   576
6   6     2   514   560   565   524   552   597
7   7     2   440   480   536   484   567   569
8   8     2   495   570   569   585   576   677
9   9     2   520   590   610   637   671   702
10 10     2   503   555   591   605   649   675
11 11     3   496   560   622   622   632   670
12 12     3   498   540   589   557   568   609
13 13     3   478   510   568   555   576   605
14 14     3   545   565   580   601   633   649
15 15     3   472   498   540   524   532   583
```

No apply the multivariate analysis of variance:

```
> summary(manova(cbind(Week1,Week3,Week4,Week5,Week6,Week7)~Group, data=diet))
          Df Pillai approx F num Df den Df  Pr(>F)
```

```
Group      2 1.4033   3.1358     12     16 0.01760 *
Residuals 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that there is a significant group effect. If we would do 6 ANOVA tests for each week separately, we even would get never a significant group effect. These 6 ANOVA tests can called easily with `summary(aov(...))`

```
> summary(aov(cbind(Week1,Week3,Week4,Week5,Week6,Week7)~Group, data=diet))
 Response Week1 :
            Df Sum Sq Mean Sq F value Pr(>F)
Group        2 2969.2  1484.6  2.1006 0.1651
Residuals   12 8481.2   706.8

 Response Week3 :
            Df  Sum Sq Mean Sq F value Pr(>F)
Group        2  2497.6  1248.8  0.8728 0.4427
Residuals   12 17170.4  1430.9

 Response Week4 :
            Df  Sum Sq Mean Sq F value Pr(>F)
Group        2   302.5   151.3  0.1397  0.871
Residuals   12 12992.4  1082.7

 Response Week5 :
            Df  Sum Sq Mean Sq F value Pr(>F)
Group        2   260.4   130.2  0.0541 0.9476
Residuals   12 28906.0  2408.8

 Response Week6 :
            Df Sum Sq Mean Sq F value Pr(>F)
Group        2   8551    4275  1.3905 0.2863
Residuals   12  36898    3075

 Response Week7 :
            Df Sum Sq Mean Sq F value Pr(>F)
Group        2  13730    6865  2.4563 0.1276
Residuals   12  33539    2795
```

To check that `summary(aov(...))` provides the univariate ANOVA test, type:

```
> anova(lm(Week7~Group, data=diet))
Analysis of Variance Table

Response: Week7
          Df Sum Sq Mean Sq F value Pr(>F)
```

```
Group      2  13730    6865  2.4563 0.1276
Residuals 12  33539    2795
```

Even if we regard the weeks before the therapy started and the weeks after start separately, there is no group effect:

```
> summary(manova(cbind(Week5,Week6,Week7)~Group, data=diet))
          Df  Pillai approx F num Df den Df Pr(>F)
Group      2 0.70386  1.99117      6     22 0.1105
Residuals 12
> summary(manova(cbind(Week1,Week3,Week4)~Group, data=diet))
          Df  Pillai approx F num Df den Df Pr(>F)
Group      2 0.71841  2.05538      6     22 0.1007
Residuals 12
```

This means that effects can become only significant if all measurements are analyzed simultaneously. This is the great advantage of MANOVA.

R uses the method of Pillai by default. But there are also other methods as those of Wilks, Hotelling-Lawley, Roy. The method of Roy, also called Roy's union intersection test, has a very simple idea: Let

$$y^1 = \begin{pmatrix} y_1^1 \\ y_2^1 \\ \vdots \\ y_N^1 \end{pmatrix}, \quad y^2 = \begin{pmatrix} y_1^2 \\ y_2^2 \\ \vdots \\ y_N^2 \end{pmatrix}, \quad \ldots, \quad y^p = \begin{pmatrix} y_1^p \\ y_2^p \\ \vdots \\ y_N^p \end{pmatrix}$$

the p measurements of the p measurements variables. These p measurements can be combined to a univariate variable by using a linear combination given by a vector $a = (a_1, a_2, \ldots, a_p)^\top \in \mathbb{R}^p$:

$$y^a = a_1 \begin{pmatrix} y_1^1 \\ y_2^1 \\ \vdots \\ y_N^1 \end{pmatrix} + a_2 \begin{pmatrix} y_1^2 \\ y_2^2 \\ \vdots \\ y_N^2 \end{pmatrix} + \ldots + a_p \begin{pmatrix} y_1^p \\ y_2^p \\ \vdots \\ y_N^p \end{pmatrix}.$$

If $\widehat{t}(y)$ is the test statistic of a univariate variable y, then it also can be used for the univariate variable y^a. Roy's idea was to regard

$$\sup_{a \in \mathbb{R}^p} \widehat{t}(y^a)$$

as test statistic for the multivariate case. The rejection set of this test is a union since

$$\left\{ \sup_{a \in \mathbb{R}^p} \widehat{t}(y^a) \geq c \right\} = \bigcup_{a \in \mathbb{R}^p} \left\{ \widehat{t}(y^a) \geq c \right\}.$$

If $\widehat{t}(y^a)$ is the test statistic for the univariate test for $H_0^a : \theta \in \Theta^a$ versus $H_1^a : \theta \notin \Theta^a$, then $\sup_{a\in\mathbb{R}^p} \widehat{t}(y^a)$ is the test statistic for testing $H_0 : \theta \in \bigcap_{a\in\mathbb{R}^p} \Theta^a$ versus $H_1 : \theta \notin \bigcap_{a\in\mathbb{R}^p} \Theta^a$. This is the reason why this method is called Roy's union intersection test.

7.1.2 Example (Diet supplements: Continuation of Example 7.1.1)
The test results for the other testing methods are obtained by:

```
> summary(manova(cbind(Week1,Week3,Week4,Week5,Week6,Week7)~Group, data=diet))
          Df Pillai approx F num Df den Df  Pr(>F)
Group      2 1.4033   3.1358     12     16 0.01760 *
Residuals 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(manova(cbind(Week1,Week3,Week4,Week5,Week6,Week7)~Group, data=diet),
+ test="Wilks")
          Df   Wilks approx F num Df den Df  Pr(>F)
Group      2 0.08793  2.76773     12     14 0.03630 *
Residuals 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(manova(cbind(Week1,Week3,Week4,Week5,Week6,Week7)~Group, data=diet),
+ test="Hotelling-Lawley")
          Df Hotelling-Lawley approx F num Df den Df Pr(>F)
Group      2           4.7859   2.3930     12     12 0.0724 .
Residuals 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(manova(cbind(Week1,Week3,Week4,Week5,Week6,Week7)~Group, data=diet),
+ test="Roy")
          Df    Roy approx F num Df den Df  Pr(>F)
Group      2 2.7666   3.6888      6      8 0.04637 *
Residuals 12
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that the p-values of the different methods are different and that the method of Pillai provides the smallest p-value. But this is not the case in general.

7.2 Crossover designs

In Example 7.1.1, the treatment, the dose of vitamin E, was not changed over the weeks. But sometimes it makes sense to change the treatment so that each experimental units is treated with several levels of the treatment, for example over the time. This is in particular useful for fields where several crops can grow in different seasons or years and the yield is measures for each season or year. Then it makes not sense to use for each subfield the same crops. It is much better to change the crop in each subfield from season to season or year to year, respectively. Such designs are called crossover designs or carryover designs. They can be constructed with the R package `cross.des`

which needs the R packages `AlgDesign` and `gtools`. Hence all three packages must be installed. It needs also the library `MASS` but this is a contributed library which must be not installed additionally.

7.2.1 Example (Crossover Design)
To create a balanced crossover design for 4 treatments in 3 time periods, type

```
> library(crossdes)
Lade nötiges Paket: gtools
Lade nötiges Paket: MASS
> all.combin(4,3)
      [,1] [,2] [,3]
 [1,]    1    2    3
 [2,]    1    2    4
 [3,]    1    3    2
 [4,]    1    3    4
 [5,]    1    4    2
 [6,]    1    4    3
 [7,]    2    1    3
 [8,]    2    1    4
 [9,]    2    3    1
[10,]    2    3    4
[11,]    2    4    1
[12,]    2    4    3
[13,]    3    1    2
[14,]    3    1    4
[15,]    3    2    1
[16,]    3    2    4
[17,]    3    4    1
[18,]    3    4    2
[19,]    4    1    2
[20,]    4    1    3
[21,]    4    2    1
[22,]    4    2    3
[23,]    4    3    1
[24,]    4    3    2
```

Since there are $4 \cdot 3 \cdot 2 = 24$ combinations of the 4 treatments, at least 24 experimental units are needed.

Part II

Mathematical Foundations

8 Preliminaries

8.1 Special methods from linear algebra

8.1.1 Definition (Idempotent matrix)
A $n \times n$ matrix A is called idempotent if and only if $A\,A = A$.

8.1.2 Lemma
If A is a symmetric and idempotent matrix of rank r, then A has r eigenvalues equal to 1 and $n-r$ eigenvalues equal to 0.

Proof. With the spectral decomposition of A. See e.g. Rencher 1998, P. 414. □

8.1.3 Lemma
Let $tr(A)$ denote the trace of a the matrix $A \in \mathbb{R}^{N\times N}$, i.e. the sum of the diagonal elements of A. Then we have:

a) $tr(AB) = tr(BA)$ for all matrices $A \in \mathbb{R}^{N\times M}$, $B \in \mathbb{R}^{M\times N}$.

b) If A is symmetric, then $tr(A)$ is the sum of the eigenvalues of the matrix A.

Proof.
a) Let $A = (A_{nm})_{n=1,\ldots,N,m=1,\ldots,M}$ and $B = (B_{mn})_{m=1,\ldots,M,n=1,\ldots,N}$. Then the n'the diagonal element of $AB \in \mathbb{R}^{N\times N}$ is $\sum_{m=1}^{M} A_{nm}B_{mn}$ and the m'th diagonal element of $BA \in \mathbb{R}^{M\times M}$ is $\sum_{n=1}^{N} B_{mn}A_{nm}$ so that

$$\mathrm{tr}(AB) = \sum_{n=1}^{N}\sum_{m=1}^{M} A_{nm}B_{mn} = \sum_{m=1}^{M}\sum_{n=1}^{N} B_{mn}A_{nm} = \mathrm{tr}(BA).$$

b) A has the spectral decomposition $PDP^\top$ where $P^\top P$ is the identity matrix I and D is a diagonal matrix consisting of the eigenvalues. According to a) it holds

$$\mathrm{tr}(A) = \mathrm{tr}(PDP^\top) = \mathrm{tr}(DP^\top P) = \mathrm{tr}(DI) = \mathrm{tr}(D)$$

so that $\mathrm{tr}(A)$ is the sum of its eigenvalues. □

8.1.4 Definition (g-inverse)
$A^- \in \mathbb{R}^{m\times n}$ is called g-inverse (generalized inverse) of $A \in \mathbb{R}^{n\times m}$ if and only if $A\,A^-\,A = A$.

If A is a regular matrix, then $A^- = A^{-1}$ and A^{-1} is the only g-inverse. Hence the g-inverse is really a generalization of the inverse for regular matrices. But note that, if A is not a regular matrix, then the g-inverse of A is not unique, i.e. there are several g-inverses. For the g-inverse of $A^\top A$ the following lemma holds.

8.1.5 Lemma
Let $(A^\top A)^-$ be a g-inverse of $A^\top A$. Then it holds:
a) $((A^\top A)^-)^\top$ is g-inverse of $A^\top A$.
b) $A^\top A(A^\top A)^- A^\top = A^\top$ and $A(A^\top A)^- A^\top A = A$.
c) $A(A^\top A)^- A^\top$ is idempotent, i.e. $A(A^\top A)^- A^\top \; A(A^\top A)^- A^\top = A(A^\top A)^- A^\top$.
d) $A(A^\top A)^- A^\top$ is independent of the choice of the g-inverse.
e) $A(A^\top A)^- A^\top$ is a symmetric matrix.

Proof.
a) $A^\top A((A^\top A)^-)^\top A^\top A = (A^\top A(A^\top A)^- A^\top A)^\top = (A^\top A)^\top = A^\top A$.
b) In general, it holds: $BD^\top D = CD^\top D$ implies $BD^\top = CD^\top$. For $BD^\top D = CD^\top D$ implies

$$\begin{aligned} 0 &= (BD^\top D - CD^\top D)(B-C)^\top = (BD^\top - CD^\top)D(B-C)^\top \\ &= (BD^\top - CD^\top)\left((B-C)D^\top\right)^\top = (BD^\top - CD^\top)\,(BD^\top - CD^\top)^\top \end{aligned}$$

Multiplying the last expression from both sides with an arbitrary vector of appropriate dimension yields $0 = x^\top (BD^\top - CD^\top)\,(BD^\top - CD^\top)^\top x$. This means $0 = (BD^\top - CD^\top)^\top x$ for all x and therefore $0 = BD^\top - CD^\top$.
Because of the definition of the g-Inverse, it holds $A^\top A(A^\top A)^- A^\top A = A^\top A$. Setting $B = A^\top A(A^\top A)^-$, $C = I$ the identity matrix, and $D = A$ provides the first part of the assertion b). The second part follows from the first part by transposing the matrices and using a).
c) follows from b).
d) Let $(A^\top A)^\sim$ be another g-inverse of $A^\top A$. Assertion b) implies $A(A^\top A)^\sim A^\top A = A = A(A^\top A)^- A^\top A$. Setting $B = A(A^\top A)^\sim$, $C = A(A^\top A)^-$ and $D = A$, then the assertion shown in b) provides $A(A^\top A)^\sim A^\top = A(A^\top A)^- A^\top$. This means that $A(A^\top A)^- A^\top$ does not depend on the choice of the g-inverse.
e) The assertion a) implies $(A(A^\top A)^- A^\top)^\top = A((A^\top A)^-)^\top A^\top = A(A^\top A)^- A^\top$. □

8.1.6 Lemma (g-inverse of partitioned matrix)
If the symmetric matrix M is a partitioned matrix given by

$$M = \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix}$$

with $A \in \mathbb{R}^{K\times K}$, $B \in \mathbb{R}^{L\times K}$, and $C \in \mathbb{R}^{K\times K}$, where A is non-singular. Then the g-inverse of M is given by

$$M^- = \begin{pmatrix} A^{-1} + A^{-1}B^\top E^- B\, A^{-1} & -A^{-1}B^\top E^- \\ -E^- B\, A^{-1} & E^- \end{pmatrix}$$

with $E = C - B\, A^{-1}B^\top$.

Proof.

$$M\,M^- = \begin{pmatrix} A\,A^{-1} + A\,A^{-1}B^\top E^- B\,A^{-1} - B^\top E^- B\,A^{-1} & -A\,A^{-1}B^\top E^- + B^\top E^- \\ B\,A^{-1} + B\,A^{-1}\,B^\top E^- B\,A^{-1} - C\,E^- B\,A^{-1} & -B\,A^{-1}B^\top E^- + C\,E^- \end{pmatrix}$$

and

$$\begin{aligned} M\,M^-M = & \begin{pmatrix} A\,A^{-1}A + A\,A^{-1}B^\top E^- B\,A^{-1}A - B^\top E^- B\,A^{-1}A - A\,A^{-1}B^\top E^- B + B^\top E^- B \\ B\,A^{-1}A + B\,A^{-1}\,B^\top E^- B\,A^{-1}A - C\,E^- B\,A^{-1}A - B\,A^{-1}B^\top E^- B + C\,E^- B \end{pmatrix} \\ & \qquad \begin{pmatrix} A\,A^{-1}B^\top + A\,A^{-1}B^\top E^- B\,A^{-1}B^\top - B^\top E^- B\,A^{-1}B^\top - A\,A^{-1}B^\top E^- C + B^\top E^- C \\ B\,A^{-1}B^\top + B\,A^{-1}\,B^\top E^- B\,A^{-1}B^\top - C\,E^- B\,A^{-1}B^\top - B\,A^{-1}B^\top E^- C + C\,E^- C \end{pmatrix} \\ = & \begin{pmatrix} A + B^\top E^- B - B^\top E^- B - B^\top E^- B + B^\top E^- B \\ B + B\,A^{-1}\,B^\top E^- B - C\,E^- B - B\,A^{-1}B^\top E^- B + C\,E^- B \end{pmatrix} \\ & \qquad \begin{pmatrix} B^\top + B^\top E^- B\,A^{-1}B^\top - B^\top E^- B\,A^{-1}B^\top - B^\top E^- C + B^\top E^- C \\ B\,A^{-1}B^\top + B\,A^{-1}\,B^\top E^- B\,A^{-1}B^\top - C\,E^- B\,A^{-1}B^\top - B\,A^{-1}B^\top E^- C + C\,E^- C \end{pmatrix} \\ = & \begin{pmatrix} A & B^\top \\ B & C \end{pmatrix} = M \end{aligned}$$

since

$$\begin{aligned} & B\,A^{-1}B^\top + (B\,A^{-1}\,B^\top - C)E^- B\,A^{-1}B^\top - (B\,A^{-1}B^\top - C)E^- C \\ & \quad = B\,A^{-1}B^\top + (B\,A^{-1}\,B^\top - C)E^-(B\,A^{-1}B^\top - C) \\ & \quad = B\,A^{-1}B^\top + B\,A^{-1}\,B^\top - C = C. \end{aligned}$$

□

8.1.7 Definition (Column space)
Let be $X \in \mathbb{R}^{N\times R}$. Then

$$C(X) := \{X\beta;\ \beta \in \mathbb{R}^R\}$$

is called the column space of X.

8.1.8 Definition (Perpendicular projection matrix)
Let be U a subspace of $\mathbb{R}^N$. $P \in \mathbb{R}^{N\times N}$ is called perpendicular projection matrix onto U if and only if

$$\begin{aligned} & Pu = u \text{ for all } u \in U, \\ & Pv = 0 \text{ for all } v \in U^\perp = \{w \in \mathbb{R}^N;\ w^\perp u \text{ for all } u \in U\}. \end{aligned}$$

8.1.9 Lemma
a) The perpendicular projection matrix $P \in \mathbb{R}^{N\times N}$ is idempotent.
b) Every idempotent and symmetric matrix $A \in \mathbb{R}^{N\times N}$ is a perpendicular projection matrix onto $C(A)$.

Proof.
a) It holds

$$PPu = Pu = u \text{ for all } u \in U,$$
$$PPv = P0 = 0 \text{ for all } v \in U^{\perp}.$$

Let be $w \in \mathbb{R}^N$ arbitrary. Then there exists $u \in U$ and $v \in U^{\perp}$ with $w = u + v$. Then we obtain

$$PPw = PP(u+v) = Pu = u = P(u+v) = Pw$$

and thus $PP = P$.
b) For $u = A\beta$ it holds $Au = AA\beta = A\beta = u$. If

$$v \in C(A)^{\perp} = \{w \in \mathbb{R}^N;\ w^{\top}A\beta = 0 \text{ for all } \beta \in \mathbb{R}^N\},$$

then $v^{\top}A\beta = 0$ for all $\beta \in \mathbb{R}^N$. The symmetry of A implies $\beta^{\top}A\,v = \beta^{\top}A^{\top}v = 0$ for all $\beta \in \mathbb{R}^N$ and thus $Av = 0$. □

8.1.10 Lemma
Let be $X \in \mathbb{R}^{N\times R}$.
a) $X(X^{\top}X)^{-}X^{\top}$ is the perpendicular projection matrix onto $C(X)$.
b) $I_{N\times N} - X(X^{\top}X)^{-}X^{\top}$ is the perpendicular projection matrix onto $C(X)^{\perp}$.
($I_{N\times N}$ denotes the $N \times N$ identity matrix).

Proof.
a) Lemma 8.1.5 c) and e) and Lemma 8.1.9 imply that $X(X^{\top}X)^{-}X^{\top}$ is the perpendicular projection matrix onto $C(X(X^{\top}X)^{-}X^{\top})$. It remains to show $C(X) = C(X(X^{\top}X)^{-}X^{\top})$. It is clear that $C(X(X^{\top}X)^{-}X^{\top}) \subset C(X)$. For the opposite inclusion note that for any $u \in C(X)$ there exists $\beta \in \mathbb{R}^p$ with $u = X\beta$. Then Lemma 8.1.5 b) implies

$$X(X^{\top}X)^{-}X^{\top}u = X(X^{\top}X)^{-}X^{\top}X\beta = X\beta = u$$

and thus $C(X) \subset C(X(X^{\top}X)^{-}X^{\top})$.
b) Part a) implies $X(X^{\top}X)^{-}X^{\top}u = 0$ for all $u \in C(X)^{\perp} = C(X(X^{\top}X)^{-}X^{\top})^{\perp}$ and thus

$$(I_{N\times N} - X(X^{\top}X)^{-}X^{\top})u = u \text{ for all } u \in C(X)^{\perp}.$$

If $v \in (C(X)^{\perp})^{\perp} = C(X)$, then Part a) implies

$$(I_{N\times N} - X(X^{\top}X)^{-}X^{\top})v = v - v = 0.$$

□

8.1.11 Lemma
For any $X \in \mathbb{R}^{N\times R}$, it holds

a) $rk(X(X^{\top}X)^{-}X^{\top}) = rk(X)$.

b) $tr(X(X^{\top}X)^{-}X^{\top}) = rk(X)$.

c) $tr(I_{N\times N} - X(X^\top X)^- X^\top) = N - rk(X)$.

d) $rk(I_{N\times N} - X(X^\top X)^- X^\top) = N - rk(X)$.

Thereby $rk(A)$ *denotes the rank of the matrix* A.

Proof.
a) It is clear that $\mathrm{rk}(X(X^\top X)^- X^\top) \leq \mathrm{rk}(X)$ holds. Because of Lemma 8.1.5 b) also the converse inequality holds:

$$\mathrm{rk}(X) = \mathrm{rk}(X(X^\top X)^- X^\top X) \leq \mathrm{rk}(X(X^\top X)^- X^\top).$$

b) $X(X^\top X)^- X^\top$ is a perpendicular projection matrix according to Lemma 8.1.10 a). Hence it is idempotent according to Lemma 8.1.9 a). This means according to Lemma 8.1.2 that $X(X^\top X)^- X^\top \in \mathbb{R}^{N\times N}$ has r eigenvalues equal to 1 and $N-r$ eigenvalues equal to 0, where $r = \mathrm{rk}(X(X^\top X)^- X^\top)$. According to a) we have $r = \mathrm{rk}(X)$ and according to Lemma 8.1.3 b) $\mathrm{tr}(X(X^\top X)^- X^\top) = r$ such that $\mathrm{tr}(X(X^\top X)^- X^\top) = \mathrm{rk}(X)$.

c) The linearity of the trace provides

$$\mathrm{tr}(I_{N\times N} - X(X^\top X)^- X^\top) = \mathrm{tr}(I_{N\times N}) - \mathrm{tr}(X(X^\top X)^- X^\top) = N - \mathrm{rk}(X).$$

d) Since $I_{N\times N} - X(X^\top X)^- X^\top$ is also a perpendicular matrix according to Lemma 8.1.10 b), its rank coincide with its trace as in b) so that the assertion follows from c). □

8.1.12 Definition
Let be $A = (A_{nm})_{n=1,\ldots,N,m=1,\ldots,M} \in \mathbb{R}^{N\times M}$ *and* $B \in \mathbb{R}^{I\times J}$. *The Kronecker product* $A \otimes B$ *is defined as*

$$A \otimes B = \begin{pmatrix} A_{11}B & A_{12}B & \ldots & A_{1M}B \\ A_{21}B & A_{22}B & \ldots & A_{2M}B \\ \vdots & \vdots & & \vdots \\ A_{N1}B & A_{N2}B & \ldots & A_{NM}B \end{pmatrix} \in \mathbb{R}^{N\,I\times M\,J}.$$

8.1.13 Definition
Let be $A = (A_{nm})_{n=1,\ldots,N,m=1,\ldots,M} \in \mathbb{R}^{N\times M}$. *The vec operator* $vec : \mathbb{R}^{N\times M} \longrightarrow \mathbb{R}^{N\cdot M}$ *is defined as*

$$vec(A) = (A_{11}, A_{21}, \ldots, A_{N1}, A_{12}, \ldots, A_{N2}, \ldots, A_{1M}, \ldots, A_{NM})^\top.$$

Note that if A is given columnwise by $A = (A_{\bullet 1}|A_{\bullet 2}|\ldots|A_{\bullet M})$, then

$$vec(A) = \begin{pmatrix} A_{\bullet 1} \\ A_{\bullet 2} \\ \vdots \\ A_{\bullet M} \end{pmatrix}.$$

8.1.14 Lemma
If $A \in \mathbb{R}^{N\times M}$, $B \in \mathbb{R}^{I\times J}$, $C \in \mathbb{R}^{M\times L}$, $D \in \mathbb{R}^{J\times K}$, *then*

a) $(A \otimes B)\,(C \otimes D) = A\,C \otimes B\,D \in \mathbb{R}^{N\,I\times L\,K}$,

b) $(A \otimes B)^\top = A^\top \otimes B^\top$.

Proof. Exercise. □

8.2 Random vectors and random matrices

8.2.1 Definition
a) $Y = (Y_1, \ldots, Y_p)^\top$ *is a random vector if* $Y_1, \ldots, Y_p$ *are random variables.*
b)

$$Z = \begin{pmatrix} Z_{11} & Z_{12} & \ldots & Z_{1p} \\ Z_{21} & Z_{22} & \ldots & Z_{2p} \\ \vdots & \vdots & & \vdots \\ Z_{N1} & Z_{N2} & \ldots & Z_{Np} \end{pmatrix}$$

is a random matrix if $Z_{11}, Z_{12}, \ldots, Z_{1p}, Z_{21}, \ldots, Z_{2p}, \ldots, Z_{N1}, \ldots, Z_{Np}$ *are random variables.*

8.2.2 Definition (Expectation of random vectors and random matrices)
a) If $Y = (Y_1, \ldots, Y_p)^\top$ *is a random vector, then*

$$\mu = \mu_Y = E(Y) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \in \mathbb{R}^p$$

is the expectation of Y.
b) If $Z = (Z_{ni})_{n=1,\ldots,N,i=1,\ldots,p}$ *is a random matrix, then*

$$E(Z) = \begin{pmatrix} E(Z_{11}) & E(Z_{12}) & \ldots & E(Z_{1p}) \\ E(Z_{21}) & E(Z_{22}) & \ldots & E(Z_{2p}) \\ \vdots & \vdots & & \vdots \\ E(Z_{N1}) & E(Z_{N2}) & \ldots & E(Z_{Np}) \end{pmatrix} \in \mathbb{R}^{N\times p}$$

is the expectation of Z.

8.2.3 Definition (Covariance matrix of a random vector)
If $Y = (Y_1, \ldots, Y_p)^\top$ is a random vector, then

$$\Sigma = Cov(Y) = \begin{pmatrix} cov(Y_1,Y_1) & cov(Y_1,Y_2) & \ldots & cov(Y_1,Y_p) \\ cov(Y_2,Y_1) & cov(Y_2,Y_2) & \ldots & cov(Y_2,Y_p) \\ \vdots & \vdots & & \vdots \\ cov(Y_p,Y_1) & cov(Y_p,Y_2) & \ldots & cov(Y_p,Y_p) \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{pmatrix} \in \mathbb{R}^{p\times p}$$

with $\sigma_{ij} = cov(Y_i, Y_j) = E((Y_i - \mu_i)(Y_j - \mu_j))$ for $i, j = 1, \ldots, p$ is the covariance matrix of Y.

Recall that $\sigma_i^2 = \sigma_{ii} = \mathrm{cov}(Y_i, Y_i) = \mathrm{E}((Y_i - \mu_i)^2)$ is the variance of Y_i, i.e. $\mathrm{cov}(Y_i, Y_i) = \mathrm{var}(Y_i)$.

8.2.4 Definition (Covariance matrix of two random vectors)
If $X = (X_1, \ldots, X_q)^\top$ and $Y = (Y_1, \ldots, Y_p)^\top$ are random vectors, then

$$\Sigma_{XY} = Cov(X,Y) = \begin{pmatrix} cov(X_1,Y_1) & cov(X_1,Y_2) & \ldots & cov(X_1,Y_p) \\ cov(X_2,Y_1) & cov(X_2,Y_2) & \ldots & cov(X_2,Y_p) \\ \vdots & \vdots & & \vdots \\ cov(X_q,Y_1) & cov(X_q,Y_2) & \ldots & cov(X_q,Y_p) \end{pmatrix} \in \mathbb{R}^{q\times p}$$

with $cov(X_i, Y_j) = E((X_i - E(X_i))(Y_j - E(Y_j)))$ for $i = 1, \ldots, q, j = 1, \ldots, p$ is the covariance matrix of X and Y.

8.2.5 Lemma
We have
a) $Cov(Y, Y) = Cov(Y)$,
b) $Cov(X, Y) = E\left((X - \mu_X)(Y - \mu_Y)^\top\right)$.

Proof.
a) is clear.

b)

$$\mathrm{E}\left((X-\mu_X)(Y-\mu_Y)^\top\right) = \mathrm{E}\left(\begin{pmatrix} (X_1-\mathrm{E}(X_1)) \\ \vdots \\ (X_q-\mathrm{E}(X_q)) \end{pmatrix}((Y_1-\mathrm{E}(Y_1)),\ldots,(Y_p-\mathrm{E}(Y_p)))\right)$$

$$= \mathrm{E}\left(\begin{pmatrix} (X_1-E(X_1))(Y_1-E(Y_1)) & \ldots & (X_1-E(Y_1))(Y_p-E(Y_p)) \\ \vdots & & \vdots \\ (X_q-E(X_q))(Y_1-E(Y_1)) & \ldots & (X_q-E(X_q))(Y_p-E(Y_p)) \end{pmatrix}\right)$$

$$= \begin{pmatrix} \mathrm{E}((X_1-E(X_1))(Y_1-E(Y_1))) & \ldots & \mathrm{E}((X_1-E(Y_1))(Y_p-E(Y_p))) \\ \vdots & & \vdots \\ \mathrm{E}((X_q-E(X_q))(Y_1-E(Y_1))) & \ldots & \mathrm{E}((X_q-E(X_q))(Y_p-E(Y_p))) \end{pmatrix}$$

$$= \mathrm{Cov}(X,Y).$$ □

8.2.6 Lemma
Let $X=(X_1,\ldots,X_q)^\top$ and $Y=(Y_1,\ldots,Y_p)^\top$ be random vectors and Z be a $q\times p$ random matrix. Then:
a) $E(X+Y)=E(X)+E(Y)$ if $q=p$.
b) $E(AY+b)=A\,E(Y)+b$ if $A\in\mathbb{R}^{q\times p}$, $b\in\mathbb{R}^q$.
c) $E(A\,Z\,B)=A\,E(Z)\,B$ if $A\in\mathbb{R}^{m\times q}$, $B\in\mathbb{R}^{p\times n}$.
d) $Cov(Y)=Cov(Y)^\top$.
e) $Cov(Y)$ is positive semidefinite.
f) $\Sigma_{XY}=Cov(X,Y)=E(X\,Y^\top)-E(X)\,E(Y)^\top$.
g) $Cov(AX+a,BY+b)=A\,Cov(X,Y)\,B^\top$ if $A\in\mathbb{R}^{m\times q}$, $B\in\mathbb{R}^{n\times p}$, $a\in\mathbb{R}^m$, $b\in\mathbb{R}^n$.
h) $Cov(AY+b)=A\,Cov(Y)\,A^\top$ if $A\in\mathbb{R}^{q\times p}$, $b\in\mathbb{R}^q$.
i) $E(X^\top A\,Y)=tr(A\,\Sigma_{YX})+\mu_X^\top A\,\mu_Y$ if $A\in\mathbb{R}^{q\times p}$.
j) $Cov(X+Y)=Cov(X)+Cov(Y)$ if $q=p$ and X and Y are stochastically independent.

Proof.
a) to c) follow from the linearity of the expectation.
d) Since

$$\mathrm{cov}(X_i,Y_j)=\mathrm{E}\left((X_i-E(X_i))(Y_j-E(Y_j))\right)=\mathrm{E}\left((Y_j-E(Y_j)(X_i-E(X_i)))\right)=\mathrm{cov}(Y_j,X_i)$$

for $i=1,\ldots,q, j=1,\ldots,p$, the assertion follows.

f) Lemma 8.2.5 b) provides with the linearity of the expectation

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}\left((X-\mu_X)(Y-\mu_Y)^\top\right) = \mathrm{E}\left(X\,Y^\top - X\,\mu_Y^\top - \mu_X\,Y^\top + \mu_X\,\mu_Y^\top\right)\\ &= \mathrm{E}(X\,Y^\top) - \mathrm{E}(X)\,\mu_Y^\top - \mu_X\,E(Y^\top) + \mu_X\,\mu_Y^\top = \mathrm{E}(X\,Y^\top) - E(X)\,E(Y)^\top.\end{aligned}$$

g) It follows from the above assertions:

$$\begin{aligned}\mathrm{Cov}(AX+a, BY+b) &\overset{f)}{=} \mathrm{E}\left((AX+a)\,(BY+b)^\top\right) - \mathrm{E}(AX+a)\,\mathrm{E}(BY+b)^\top\\ &\overset{b)}{=} \mathrm{E}\left(A\,X\,Y^\top B^\top + a\,Y^\top B^\top + A\,X\,b^\top + a\,b^\top\right) - (A\,\mathrm{E}(X)+a)\,(B\,\mathrm{E}(Y)+b)^\top\\ &\overset{a),b)}{=} \mathrm{E}(A\,X\,Y^\top\,B^\top) - A\,E(X)\,E(Y)^\top\,B^\top \overset{c)}{=} A\,\mathrm{E}(X\,Y^\top)\,B^\top - A\,E(X)\,E(Y)^\top\,B^\top\\ &= A\left(\mathrm{E}(X\,Y^\top) - E(X)\,E(Y)^\top\right)B^\top \overset{f)}{=} A\,\mathrm{Cov}(X,Y)\,B^\top.\end{aligned}$$

h) follows from g).
e) Assertion h) implies for all $a \in \mathbb{R}^p$

$$a^\top\,\mathrm{Cov}(Y)\,a = \mathrm{Cov}(a^\top Y) = \mathrm{var}(a^\top Y) \geq 0.$$

i) Lemma 8.1.3 and the linearity of the expectation provide

$$\begin{aligned}\mathrm{E}(X^\top A\,Y) &= \mathrm{E}(\mathrm{tr}(X^\top A\,Y)) = \mathrm{E}(\mathrm{tr}(A\,Y\,X^\top)) = \mathrm{tr}(\mathrm{E}(A\,Y\,X^\top))\\ &\overset{c)}{=} \mathrm{tr}(A\,\mathrm{E}(Y\,X^\top)) \overset{f)}{=} \mathrm{tr}(A\,(\Sigma_{YX} + E(Y)\,E(X)^\top)) = \mathrm{tr}(A\,\Sigma_{YX}) + \mathrm{tr}(A\,E(Y)\,E(X)^\top)\\ &= \mathrm{tr}(A\,\Sigma_{YX}) + \mathrm{tr}(E(X)^\top\,A\,E(Y)) = \mathrm{tr}(A\,\Sigma_{YX}) + \mu_X^\top\,A\,\mu_Y.\end{aligned}$$

j) Regard the (i,j) component of $\mathrm{Cov}(X+Y)$:

$$\begin{aligned}\mathrm{Cov}(X+Y)_{i,j} &= \mathrm{cov}(X_i+Y_i, X_j+Y_j) = \mathrm{E}((X_i+Y_i)(X_j+Y_j)) - \mathrm{E}(X_i+Y_i)\,\mathrm{E}(X_j+Y_j)\\ &= \mathrm{E}(X_iX_j + Y_iX_j + X_iY_j + Y_iY_j) - (\mathrm{E}(X_i)+\mathrm{E}(Y_i))\,(\mathrm{E}(X_j)+\mathrm{E}(Y_j))\\ &= \mathrm{E}(X_iX_j) + \mathrm{E}(Y_i)\mathrm{E}(X_j) + \mathrm{E}(X_i)\mathrm{E}(Y_j) + \mathrm{E}(Y_iY_j)\\ &\quad - \mathrm{E}(X_i)\mathrm{E}(X_j) - \mathrm{E}(Y_i)\mathrm{E}(X_j) - \mathrm{E}(X_i)\mathrm{E}(Y_j) - \mathrm{E}(Y_i)\mathrm{E}(Y_j)\\ &= = \mathrm{cov}(X_i,X_j) + \mathrm{cov}(Y_i,Y_j) = (\mathrm{Cov}(X) + \mathrm{Cov}(Y))_{i,j}.\end{aligned}$$

□

8.3 The normal distribution and related distributions

8.3.1 Definition

The random vector Y has a p dimensional normal distribution with parameters $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$, i.e. $Y \sim \mathcal{N}_p(\mu,\Sigma)$, if and only if Y has the density

$$f(y) = \frac{1}{(2\pi)^{p/2}\sqrt{\det\Sigma}}\,e^{-\frac{1}{2}(y-\mu)^\top\Sigma^{-1}(y-\mu)}.$$

8.3.2 Lemma
a) If Y has a p dimensional normal distribution with parameters $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$, then

$$E(Y) = \mu, \quad Cov(Y) = \Sigma.$$

b) If Y has a p dimensional normal distribution with parameters $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p\times p}$, $A \in \mathbb{R}^{p\times p}$, and $b \in \mathbb{R}^p$, then

$$AY + b \sim \mathcal{N}_p(A\mu + b\ ,\ A\Sigma A^\top) \tag{14}$$

Proof. See books which give introductions in probability theory. □

8.3.3 Lemma
If $Y = (Y_1, \ldots, Y_p)^\top \sim \mathcal{N}_p(\mu, \Sigma)$ and the components $Y_1, \ldots, Y_p$ are pairwise stochastically independent, i.e. Y_i and Y_j are stochastically independent for $i \neq j$, then $Y_1, \ldots, Y_p$ are stochastically independent.

Proof. Since Σ is the covariance matrix of Y, all components of Σ are the covariances $\mathrm{cov}(Y_i, Y_j)$. The pairwise independence of $Y_1, \ldots, Y_p$ implies $\mathrm{cov}(Y_i, Y_j) = 0$ for $i \neq j$. Hence Σ is a diagonal matrix $\mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. This means with $\mu = (\mu_1, \ldots, \mu_p)^\top$ that the density of Y has the form

$$\begin{aligned} f_Y(y) &= \frac{1}{(2\pi)^{p/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)\right) \\ &= \frac{1}{(2\pi)^{p/2}\sqrt{\prod_{i=1}^p \sigma_i^2}} \exp\left(-\frac{1}{2}\sum_{i=1}^p \frac{(y_i-\mu_i)^2}{\sigma_i^2}\right) \\ &= \prod_{i=1}^r \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(y_i-\mu_i)^2\right) = \prod_{i=1}^r f_{\mathcal{N}(\mu_i, \sigma_i^2)}(y_i). \end{aligned}$$

Hence the common density is a product of densities of the single densities of Y_i, so that $Y_1, \ldots, Y_p$ are stochastically independent. □

8.3.4 Theorem
If $Y \sim \mathcal{N}_p(\mu, \Sigma)$, $A \in \mathbb{R}^{q\times p}$, $B \in \mathbb{R}^{p-q\times p}$, $\binom{A}{B}$ is not singular with $A\Sigma B^\top = 0$, $a \in \mathbb{R}^q$, $b \in \mathbb{R}^{p-q}$, then:

$$\begin{aligned} AY + a &\sim \mathcal{N}_q(A\mu + a\ ,\ A\Sigma A^\top) \\ BY + b &\sim \mathcal{N}_{p-q}(B\mu + b\ ,\ B\Sigma B^\top) \end{aligned}$$

and $AY + a$ and $BY + b$ are stochastically independent.

Proof. Let be $X = AY + a$ and $Z = BY + b$. According to (14) we have

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} Y + \begin{pmatrix} a \\ b \end{pmatrix} \sim \mathcal{N}_p \left(\begin{pmatrix} A \\ B \end{pmatrix} \mu + \begin{pmatrix} a \\ b \end{pmatrix} , \ \begin{pmatrix} A \\ B \end{pmatrix} \Sigma \begin{pmatrix} A \\ B \end{pmatrix}^\top \right).$$

Then

$$\begin{pmatrix} A \\ B \end{pmatrix} \Sigma \begin{pmatrix} A \\ B \end{pmatrix}^\top = \begin{pmatrix} A\Sigma A^\top & A\Sigma B^\top \\ B\Sigma A^\top & B\Sigma B^\top \end{pmatrix} = \begin{pmatrix} A\Sigma A^\top & 0 \\ 0 & B\Sigma B^\top \end{pmatrix}$$

implies

$$\det \left(\begin{pmatrix} A \\ B \end{pmatrix} \Sigma \begin{pmatrix} A \\ B \end{pmatrix}^\top \right) = \det A\Sigma A^\top \ \det B\Sigma B^\top$$

and thus

$$\begin{aligned}
& f_{\binom{X}{Y}}(x,z) \\
&= \frac{1}{(2\pi)^{p/2} \, (\det A\Sigma A^\top)^{1/2} \, (\det B\Sigma B^\top)^{1/2}} \\
&\quad \cdot \exp \left(-\frac{1}{2} \left(\begin{pmatrix} x \\ z \end{pmatrix} - \left(\begin{pmatrix} A \\ B \end{pmatrix} \mu + \begin{pmatrix} a \\ b \end{pmatrix} \right) \right)^\top \begin{pmatrix} A\Sigma A^\top & 0 \\ 0 & B\Sigma B^\top \end{pmatrix}^{-1} \right. \\
&\qquad \left. \cdot \left(\begin{pmatrix} x \\ z \end{pmatrix} - \left(\begin{pmatrix} A \\ B \end{pmatrix} \mu + \begin{pmatrix} a \\ b \end{pmatrix} \right) \right) \right) \\
&= \frac{1}{(2\pi)^{q/2} \, (\det A\Sigma A^\top)^{1/2}} \exp \left(-\frac{1}{2} \left(x - (A\mu + a) \right)^\top (A\Sigma A^\top)^{-1} \left(x - (A\mu + a) \right) \right) \\
&\quad \cdot \frac{1}{(2\pi)^{(p-q)/2} \, (\det B\Sigma B^\top)^{1/2}} \exp \left(-\frac{1}{2} \left(z - (B\mu + b) \right)^\top (B\Sigma B^\top)^{-1} \left(z - (B\mu + b) \right) \right) \\
&= f_{\mathcal{N}_q(A\mu + a, A\Sigma A^\top)}(x) \ \cdot \ f_{\mathcal{N}_{p-q}(B\mu + b, B\Sigma B^\top)}(z).
\end{aligned}$$

Since the density $f_{\binom{X}{Y}}(x,z)$ is a product of the densities of X and Z, $X = AY + a$ and $Z = BY + b$ are stochastically independent. The product form of the densities provides also the distribution of $X = AY + a$ and $Z = BY + b$. □

8.3.5 Corollary
If $Y \sim \mathcal{N}_p(\mu, \Sigma)$, $A \in \mathbb{R}^{q \times p}$, $B \in \mathbb{R}^{r \times p}$, $rkA = q$, $rkB = r$ and $A\Sigma B^\top = 0$, $a \in \mathbb{R}^q$, $b \in \mathbb{R}^r$, then:

$$\begin{aligned}
AY + a &\sim \mathcal{N}_q(A\mu + a \, , \ A\Sigma A^\top) \\
BY + b &\sim \mathcal{N}_r(B\mu + b \, , \ B\Sigma B^\top)
\end{aligned}$$

and $AY + a$ and $BY + b$ are stochastically independent.

Proof. Since Σ is positive definite, there exists $V \in \mathbb{R}^{p\times p}$ with $\Sigma = VV^\top$ (linear algebra). Because A and B are of full rank (rk =rank), there exists also $C \in \mathbb{R}^{p-q-r\times p}$ with $AV\,V^\top\binom{B}{C} = 0$ and

$$\mathrm{rk}\begin{pmatrix} AV \\ BV \\ CV \end{pmatrix} = \mathrm{rk}\left(\begin{pmatrix} A \\ B \\ C \end{pmatrix} V\right) = p.$$

Then it holds

$$\mathrm{rk}\begin{pmatrix} A \\ B \\ C \end{pmatrix} = p \text{ and } A\Sigma\begin{pmatrix} B \\ C \end{pmatrix}^\top = 0.$$

According to Theorem 8.3.4, we obtain for every $c \in \mathbb{R}^{p-q-r}$ that $AY + a$ and $\binom{B}{C}Y + \binom{b}{c}$ are stochastically independent and $AY + a \sim \mathcal{N}_q(A\mu + a, A\Sigma A^\top)$. Then $AY + a$ and $BY + b$ are stochastically independent as well, since $BY + b$ can be obtained by projection onto the first r components of $\binom{B}{C}Y + \binom{b}{c}$. By analogous extension of A to $\binom{A}{C'}$, we also obtain $BY + b \sim \mathcal{N}_q(B\mu + b, B\Sigma B^\top)$. □

8.3.6 Theorem (Theorem of Craig and Sakamoto)
If $Y \sim \mathcal{N}_p(\mu, \Sigma)$ and $A, B \in \mathbb{R}^{p\times p}$ are positive semidefinite with $A\Sigma B^\top = 0$, then:

a) AY and BY are stochastically independent.

b) $Y^\top AY$ and BY are stochastically independent.

c) $Y^\top AY$ and $Y^\top BY$ are stochastically independent.

Proof. Because A and B are positive semidefinite, there exists $L \in \mathbb{R}^{p\times q}$, $M \in \mathbb{R}^{p\times r}$ with $LL^\top = A$, $MM^\top = B$, $\mathrm{rk}L = q = \mathrm{rk}L^\top L$, $\mathrm{rk}M = r = \mathrm{rk}M^\top M$ (linear algebra). Then $0 = A\Sigma B^\top$ implies

$$0 = (L^\top L)^{-1}\,L^\top L\,L^\top\,\Sigma\,M\,M^\top M\,(M^\top M)^{-1} = L^\top \Sigma M.$$

Corollary 8.3.5 provides that $L^\top Y$ and $M^\top Y$ are stochastically independent. Since functions of independent random variables are also independent, i.e. $h_1(X)$ and $h_2(Z)$ are stochastically independent if X and Z are stochastically independent, we obtain the independence of
a) $LL^\top Y = AY$ and $MM^\top Y = BY$,
b) $Y^\top LL^\top Y = Y^\top AY$ and $MM^\top Y = BY$,
c) $Y^\top LL^\top Y = Y^\top AY$ and $Y^\top MM^\top Y = Y^\top BY$. □

8.3.7 Definition (χ^2-distribution)
a) X has a χ^2-distribution with N degrees of freedom and non-centrality parameter $\mu^\top\mu$, abbreviated by $X \sim \chi^2(N, \mu^\top\mu)$, if and only if there is a random vector $Y = (Y_1, \ldots, Y_N)^\top$ with $Y \sim \mathcal{N}_N(\mu, I_{N\times N})$ such that

$$X = Y^\top Y = \sum_{n=1}^{N} Y_n^2.$$

b) X has a central χ^2-distribution with N degrees of freedom, abbreviated by $X \sim \chi^2(N,0)$, if and only if there is a random vector $Y = (Y_1, \ldots, Y_N)^\top$ with $Y \sim \mathcal{N}_N(0, I_{N\times N}$ (or if there are stochastically independent random variables $Y_1, \ldots, Y_N$ with $Y_n \sim \mathcal{N}_1(0,1)$ for $n = 1, \ldots, N$) such that

$$X = Y^\top Y = \sum_{n=1}^{N} Y_n^2.$$

c) X has a $\sigma^2\chi^2(N)$-distribution, if and only if if there are stochastically independent random variables $Y_1, \ldots, Y_N$ with $Y_n \sim \mathcal{N}_1(0, \sigma^2)$ for $n = 1, \ldots, N$ such that

$$X = \sum_{n=1}^{N} Y_n^2.$$

8.3.8 Definition (t-distribution)
T has a t-distribution with N degrees of freedom and non-centrality parameter δ, abbreviated by $T \sim t(N,\delta)$, if and only if there are stochastically independent random variables X and Y with $X \sim \chi^2(N,0)$ and $Y \sim \mathcal{N}_1(\delta,1)$ such that

$$T = \frac{Y}{\sqrt{\frac{1}{N}X}}.$$

If $\delta = 0$, then T has a central t-distribution with N degrees of freedom.

8.3.9 Definition (F-distribution)
V has a F-distribution with M and N degrees of freedom and non-centrality parameter δ, abbreviated by $V \sim F(M,N,\delta)$, if and only if there are stochastically independent random variables X and Y with $X \sim \chi^2(N,0)$ and $Y \sim \chi^2(M,\delta)$ such that

$$V = \frac{\frac{1}{M}Y}{\frac{1}{N}X}.$$

If $\delta = 0$, then V has a central F-distribution with M and N degrees of freedom.

8.3.10 Lemma
If X and Y are stochastically independent with $X \sim \chi^2(N,0)$ and $Y \sim \mathcal{N}_1(\delta,1)$ then

$$\frac{Y}{\sqrt{\frac{1}{N}X}} \sim t(N,\delta) \quad \text{and} \quad \frac{Y^2}{\frac{1}{N}X} \sim F(1,N,\delta^2).$$

Proof. The first part is the definition of the t-distribution. The second part follows from $Y^2 \sim \chi^2(1,\delta^2)$ and the definition of the F-distribution. □

8.3.11 Theorem
Let be $Y = (Y_1, \ldots, Y_N)^\top$ a random vector satisfying $Y \sim \mathcal{N}_N(\mu, I_{N\times I})$ with $\mu \in \mathbb{R}^N$ and $A \in \mathbb{R}^{N\times N}$ a symmetric and idempotent matrix of rank $rk(A)$. Then

$$Y^\top A Y \sim \chi^2(rk(A), \mu^\top A \mu).$$

Proof. Since A is symmetric and idempotent with rank $r = \mathrm{rk}(A)$, A has the spectral decomposition (see Lemma 8.1.2)

$$A = \sum_{i=1}^{r} v_i \, v_i^\top = V^\top V,$$

where $v_1, \ldots, v_r \in \mathbb{R}^N$ are mutually orthogonal and normed and $V = (v_1, \ldots, v_r)^\top$. This implies

$$Y^\top A Y = Y^\top \sum_{i=1}^{r} v_i \, v_i^\top Y = \sum_{i=1}^{r} Y^\top v_i (Y^\top v_i)^\top = Z^\top Z$$

with $Z = (Y^\top v_1, \ldots, Y^\top v_r)^\top = VY$. Then we have that V is of full rank and

$$V V^\top = \begin{pmatrix} v_1^\top \\ \vdots \\ v_r^\top \end{pmatrix} (v_1, \ldots, v_r) = \begin{pmatrix} v_1^\top v_1 & v_1^\top v_2 & \ldots & v_1^\top v_r \\ v_2^\top v_1 & v_2^\top v_2 & \ldots & v_2^\top v_r \\ \vdots & \vdots & & \vdots \\ v_r^\top v_1 & v_r^\top v_2 & \ldots & v_r^\top v_r \end{pmatrix} = I_{r \times r}$$

since $v_i^\top v_j = 0$ if $i \neq j$ and $v_i^\top v_i = 1$. Hence Lemma 8.3.4 provides

$$Z = VY \sim \mathcal{N}_r(V\mu, V \, I_{N \times N} V^\top) = \mathcal{N}_r(V\mu, V \, V^\top) = \mathcal{N}_r(V\mu, I_{r \times r})$$

so that with Definition 8.3.7 a)

$$Y^\top A \, Y = Z^\top Z \sim \chi^2(r, \mu^\top V^\top V \, \mu) = \chi^2(\mathrm{rk}(A), \mu^\top A \, \mu). \qquad \square$$

8.3.12 Theorem
If X and V are stochastically independent with $Y \sim \chi^2(r, 0)$, $V \sim \chi^2(q, 0)$, then

$$Y + V \sim \chi^2(r + q, 0).$$

Proof. According to Definition 8.3.7, there are stochastically independent random variables $Y_1, \ldots, Y_r, W_1, \ldots, W_q$ with $Y_i \sim \mathcal{N}(0, 1)$ for $i = 1, \ldots, r$, $W_j \sim \mathcal{N}(0, 1)$ for $j = 1, \ldots, q$ and $X = \sum_{i=1}^{r} Y_i^2$, $V = \sum_{j=1}^{q} W_j^2$. Then it holds $X + V = \sum_{i=1}^{r} Y_i^2 + \sum_{j=1}^{q} W_j^2 \sim \chi^2(r + q, 0)$. $\square$

8.4 Foundations of statistical tests

In the general statistical setup, it is assumed that the data $y_1, \ldots, y_N \in \mathcal{Y}$ are realizations of random variables $Y_1, \ldots, Y_N : \Omega \longrightarrow \mathcal{Y}$. The vector $y = (y_1, \ldots, y_N)^\top \in \mathcal{Y}^N$ is called observation vector or **sample** and $\mathcal{Y}^N$ is called **sample space**. In the parametric setup, it is assumed that

the distribution of the random vector $Y = (Y_1, \ldots, Y_N)^\top : \Omega \longrightarrow \mathcal{Y}^N$ is known up to a unknown parameter θ, hence

$$P^Y \in \left\{P_\theta^Y;\ \theta \in \Theta\right\}.$$

Statistical tests are used if there are hypotheses about the unknown parameter:

null hypotheses $H_0 :\ \theta \in \Theta_0$ versus alternative $H_1 :\ \theta \in \Theta_1 = \Theta \setminus \Theta_0$.

8.4.1 Definition (Decision rule between H_0 and H_1)
φ is called a decision rule between $H_0 :\ \theta \in \Theta_0$ and $H_1 :\ \theta \in \Theta_1$ if and only if

$$\varphi : \mathcal{Y}^N \longrightarrow \{0, 1\}.$$

If $\varphi(y) = 0$, then a decision for H_0 is made based on the sample y. If $\varphi(y) = 1$, then a decision for H_1 is made based on the sample y.

8.4.2 Remark (α and β error probability)
A decision rule between $H_0 :\ \theta \in \Theta_0$ and $H_1 :\ \theta \in \Theta_1$ has two error probabilities:
the probability for the α-error

$$P_\theta(\varphi(Y) = 1) \ \text{ with } \theta \in \Theta_0,$$

and the probability for the β-error

$$P_\theta(\varphi(Y) = 0) \ \text{ with } \theta \in \Theta_1.$$

8.4.3 Definition (α-level test)
Let be $\alpha \in (0, 1)$. A decision rule φ between $H_0 :\ \theta \in \Theta_0$ and $H_1 :\ \theta \in \Theta_1$ is called α-level test if and only if

$$P_\theta(\varphi(Y) = 1) \leq \alpha \ \text{ for all } \theta \in \Theta_0,$$

i.e. the probability for the α-error is always not larger than α.

8.4.4 Remark
An α-level test φ has usually the form

$$\varphi(y) = 1\!\!1_{\{\widehat{T}(y) > c\}}(y),$$

where $\widehat{T}(y)$ is called **test statistic** and c is called **critical value**. $1\!\!1$ denotes here the indicator function. The main task for the development of tests is to determine the test statistic and the critical value.

8.5 Tests for one and two samples

Tests for one or two samples of normally distributed random variables use the quantiles of the t-distribution, the χ^2-distribution, and the F-distribution as critical values. Therefore let be

$$t_{M,\alpha} = F_{t_M}^{-1}(\alpha)$$

the α-quantile of the central t-distribution with M degrees of freedom,

$$\chi^2_{M,\alpha} = F^{-1}_{\chi^2_M}(\alpha)$$

the α-quantile of the central χ^2-distribution with M degrees of freedom, and

$$F_{N,M,\alpha} = F^{-1}_{F_{N,M}}(\alpha)$$

the α-quantil of the central F-distribution with N and M degrees of freedom.

Tests for one sample

If the random variables $Y_1, \ldots, Y_N$ are stochastically independent and identically distributed with $Y_n \sim \mathcal{N}(\mu, \sigma^2)$, then there are two main test problems:

$$\begin{array}{lll} a) & H_0 : \mu = \mu_0 & \text{versus} \quad H_1 : \mu \neq \mu_0, \\ b) & H_0 : \mu \leq \mu_0 & \text{versus} \quad H_1 : \mu > \mu_0, \\ c) & H_0 : \mu \geq \mu_0 & \text{versus} \quad H_1 : \mu < \mu_0, \end{array}$$

where μ_0 is a given value, and

$$\begin{array}{lll} a) & H_0 : \sigma^2 = \sigma_0^2 & \text{versus} \quad H_1 : \sigma^2 \neq \sigma_0^2, \\ b) & H_0 : \sigma^2 \leq \sigma_0^2 & \text{versus} \quad H_1 : \sigma^2 > \sigma_0^2, \\ c) & H_0 : \sigma^2 \geq \sigma_0^2 & \text{versus} \quad H_1 : \sigma^2 < \sigma_0^2, \end{array}$$

where σ_0^2 is a given value. The tests base on estimates for μ and σ^2. The estimate for μ is the arithmetic mean

$$\overline{y} := \frac{1}{N}\sum_{n=1}^{N} y_n = \frac{1}{N} 1_N^\top y,$$

where $1_N \in \mathbb{R}^N$ denotes the N dimensional vector consisting only of ones, and the estimate for σ^2 is the empirical variance

$$\widehat{\sigma}^2 := \widehat{\sigma}^2(y) := \frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \overline{Y})^2.$$

Thereby note (Exercise!) that

$$\widehat{\sigma}^2 = \frac{1}{N-1} Y^\top \left(I_{N\times N} - 1_N (1_N^\top 1_N)^{-1} 1_N^\top \right) Y = \frac{1}{N-1} Y^\top \left(I_{N\times N} - \frac{1}{N} 1_N 1_N^\top \right) Y,$$

where $I_{N\times N}$ is the $N \times N$ identity matrix.

8.5.1 Theorem
If $Y_1, \ldots, Y_N$ are stochastically independent with $Y_n \sim \mathcal{N}(\mu, \sigma^2)$ für $n = 1, \ldots, N$, then

a) $\overline{Y} \sim \mathcal{N}(\mu, \frac{1}{N}\sigma^2)$,

b) $\frac{1}{\sigma^2}\sum_{n=1}^{N}(Y_n - \overline{Y})^2 \sim \chi^2(N-1, 0)$,

c) $\overline{Y}$ *and* $\sum_{n=1}^{N}(Y_n - \overline{Y})^2$ *are stochastically independent.*

Proof.
a) Note that the stochastic independence of $Y_1, \ldots, Y_N$ implies $Y = (Y_1, \ldots, Y_N)^\top \sim \mathcal{N}_N(\mu\, 1_N, \sigma^2 I_{N\times N})$ so that the assertion follows from (14) with $A = \frac{1}{N}\, 1_N^\top$ and $b = 0$ since $\overline{y} = \frac{1}{N} 1_N^\top y$.
b) The above exercise provides

$$\sum_{n=1}^{N}(Y_n - \overline{Y})^2 = Y^\top A\, Y$$

with $A = \left(I_{N\times N} - \frac{1}{N} 1_N 1_N^\top\right)$. A is a symmetric and idempotent matrix because

$$\begin{aligned} A \cdot A &= \left(I_{N\times N} - \frac{1}{N} 1_N 1_N^\top\right)\left(I_{N\times N} - \frac{1}{N} 1_N 1_N^\top\right) \\ &= I_{N\times N} - \frac{1}{N} 1_N 1_N^\top - \frac{1}{N} 1_N 1_N^\top + \frac{1}{N^2} 1_N 1_N^\top 1_N 1_N^\top = A. \end{aligned}$$

Moreover, we have

$$A\, 1_N = \left(I_{N\times N} - \frac{1}{N} 1_N 1_N^\top\right) 1_N = 1_N - \frac{1}{N} 1_N 1_N^\top\, 1_N = 0.$$

If v is orthogonal to 1_N, i.e. $1_N^\top v = 0$, then

$$A\, v = I_{N\times N}\, v - \frac{1}{N} 1_N 1_N^\top\, v = v.$$

This means that A has rank $N-1$. With Theorem 8.3.11 we obtain

$$\begin{aligned} \frac{1}{\sigma^2}\sum_{n=1}^{N}(Y_n - \overline{Y}.)^2 &= \frac{1}{\sigma^2} Y^\top A\, Y = \frac{1}{\sigma^2} Y^\top A^\top A\, Y = \frac{1}{\sigma^2}(Y - \mu\, 1_N)^\top A^\top A\,(Y - \mu\, 1_N) \\ &= \begin{pmatrix} \frac{Y_1-\mu}{\sigma} \\ \vdots \\ \frac{Y_N-\mu}{\sigma} \end{pmatrix}^\top A \begin{pmatrix} \frac{Y_1-\mu}{\sigma} \\ \vdots \\ \frac{Y_N-\mu}{\sigma} \end{pmatrix} \sim \chi^2(N-1, 0), \end{aligned}$$

since $\frac{Y_n-\mu}{\sigma} \sim \mathcal{N}(0,1)$ and $Y_1, \ldots, Y_N$ are stochastically independent.
c) Because of

$$Y \sim \mathcal{N}(\mu\, 1_N\, ,\, \sigma^2\, I_{N\times N}).$$

and

$$A\,\sigma^2\,I_{N\times N}\,1_N\,1_N^\top = A\,1_N\,1_N^\top\,\sigma^2 = 0,$$

Theorem 8.3.6 implies the stochastic independence of $Y^\top A\,Y$ and $1_N\,1_N^\top Y = N\,1_N\,\overline{Y}$. Then also $\sum_{n=1}^N (Y_n - \overline{Y})^2$ and $\overline{Y}$ are stochastically independent. □

8.5.2 Theorem (t-test for one sample)
Let $Y_1, \ldots, Y_N \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. with unknown $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$, $\mu_0 \in \mathbb{R}$ be known, and

$$\widehat{d}(y) := \sqrt{N}\,\frac{\overline{y} - \mu_0}{\widehat{\sigma}(y)}.$$

Then:

a) $\varphi(y) = \mathbb{1}_{\{|\widehat{d}(y)| > t_{N-1,1-\alpha/2}\}}(y)$ is α-level test for $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

b) $\varphi(y) = \mathbb{1}_{\{\widehat{d}(y) > t_{N-1,1-\alpha}\}}(y)$ is α-level test for $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$.

c) $\varphi(y) = \mathbb{1}_{\{\widehat{d}(y) < t_{N-1,\alpha}\}}(y)$ is α-level test for $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0$.

Proof. We show at first that $\widehat{d}(y)$ has a central t-distribution with $N-1$ degrees of freedom if $Y_n \sim \mathcal{N}(\mu_0, \sigma^2)$ for all $n = 1, \ldots, N$. According to Theorem 8.5.1, $\frac{N-1}{\sigma^2}\widehat{\sigma}^2(Y)$ has a central χ^2-distribution with $N-1$ degrees of freedom and is stochastically independent from the arithmetic mean $\overline{Y}$. Then $\frac{N-1}{\sigma^2}\widehat{\sigma}^2(Y)$ is also independent of $\sqrt{N}\frac{\overline{Y}-\mu_0}{\sigma}$. Because of $Y \sim \mathcal{N}(\mu_0 1_N, \sigma^2 I_{N\times N})$ and $\sqrt{N}\frac{\overline{Y}-\mu_0}{\sigma} = \frac{1}{\sqrt{N}\sigma}1_N^\top Y - \frac{\sqrt{N}}{\sigma}\mu_0$, Theorem 8.3.4 or Corollary 8.3.5, respectively, provide

$$\sqrt{N}\frac{\overline{Y}-\mu_0}{\sigma} \sim \mathcal{N}\left(\frac{1}{\sqrt{N}\sigma}1_N^\top\,\mu_0 1_N - \frac{\sqrt{N}}{\sigma}\mu_0, \frac{1}{\sqrt{N}\sigma}1_N^\top\,\sigma^2 I_{N\times N}\,(\frac{1}{\sqrt{N}\sigma}1_N^\top)^\top\right) = \mathcal{N}(0,1).$$

We can see also the $\mathcal{N}(0,1)$-distribution by calculating the expectation and variance of $\sqrt{N}\frac{\overline{Y}-\mu_0}{\sigma}$ knowing that linear combinations of normal distributed random variables are always normal distributed. Since $\sqrt{N}\,\frac{\overline{Y}-\mu_0}{\sigma}$ and $\frac{N-1}{\sigma^2}\widehat{\sigma}^2(Y)$ are independent, the central t-distribution of $\widehat{d}(Y)$ follows from Definition 8.3.8. This implies at once a).

b) Since μ_0 was arbitrary in the above considerations, we have for all $\mu \in \mathbb{R}$ that $\sqrt{N}\,\frac{\overline{Y}-\mu}{\widehat{\sigma}(Y)}$ has a $t(N-1,0)$-distribution if $Y_n \sim \mathcal{N}(\mu, \sigma^2)$ for all $n = 1, \ldots, N$. In particular we obtain for arbitrary $\mu \leq \mu_0$

$$\begin{aligned}
P_\mu\,(\varphi(Y) = 1) &= P_\mu\left(\sqrt{N}\,\frac{\overline{Y}-\mu_0}{\widehat{\sigma}(Y)} > t_{N-1,1-\alpha}\right) \\
&= P_\mu\left(\sqrt{N}\,\frac{\overline{Y}-\mu}{\widehat{\sigma}(Y)} > t_{N-1,1-\alpha} + \sqrt{N}\,\frac{\mu_0-\mu}{\widehat{\sigma}(Y)}\right) \\
&\leq P_\mu\left(\sqrt{N}\,\frac{\overline{Y}-\mu}{\widehat{\sigma}(Y)} > t_{N-1,1-\alpha}\right) = 1 - F_{t(N-1,0)}(t_{N-1,1-\alpha}) = 1-(1-\alpha) = \alpha.
\end{aligned}$$

c) This assertion follows analogously to the proof in b). □

8.5.3 Example (Gepaarte Zwei-Stichproben-Probleme)
Zum Beispiel werde der Blutdruck von N Personen vor und nach einer Therapie gemessen. Die n'te Beobachtung Y_n ist dann die Differenz der Blutdruckwerte V_n vor und des Blutdruckwertes W_n nach der Therapie, d.h. $Y_n = V_n - W_n$. Gilt $V_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $W_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$, so kann man auch annehmen, dass der Zufallsvektor $(V_n, W_n)^\top$ eine zweidimensionale Normalverteilung besitzt. Nach Satz 8.3.4 bzw. Folgerung 8.3.5 besitzt dann $Y_n = V_n - W_n$ auch eine Normalverteilung mit Erwartungswert

$$\mathrm{E}(Y_n) = \mathrm{E}(V_n) - \mathrm{E}(W_n) = \mu_1 - \mu_2 =: \mu$$

und Varianz

$$\mathrm{var}(Y_n) = \mathrm{var}(V_n) + \mathrm{var}(W_n) - 2\mathrm{cov}(V_n, W_n) = \sigma_1^2 + \sigma_2^2 - 2\mathrm{cov}(V_n, W_n) =: \sigma^2,$$

d.h. wir haben $Y_n \sim \mathcal{N}(\mu, \sigma^2)$. Hat die Therapie eine Blutdruck senkende Wirkung, so gilt $\mu > 0$. Um das zu belegen, muss man dann $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$ testen. Führt ein Test zum Niveau α zur Ablehnung der Nullhypothese, sprechen die Daten zum Signifikanzniveau α dafür, dass die Therapie eine Wirkung hat. Natürlich kann es dann immer noch sein, dass die Therapie nicht wirkt. Aber dann würde die Nullhypothese höchstens mit einer Wahrscheinlichkeit α abgelehnt werden, und bei kleinem α wäre das sehr unwahrscheinlich.

$V_1, \ldots, V_N$ und $W_1, \ldots, W_N$ bilden zwei Stichproben, die aber gepaart sind, da V_n und W_n von der gleichen Person stammen und damit nicht unabhängig sind. Aus diesem Grund werden solche Probleme **gepaarte Zwei-Stichproben-Probleme** genannt.

8.5.4 Theorem (Variance test for one sample)
Let $Y_1, \ldots, Y_N \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. with unknown $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$, $\sigma_0^2 \in \mathbb{R}^+$ be known, and

$$\widehat{v}(y) = \frac{N-1}{\sigma_0^2}\, \widehat{\sigma}^2(y).$$

Then:

a) $\varphi(y) = \mathbb{1}_{\{T(y) > \chi^2_{N-1,1-\alpha/2} \text{ or } T(y) < \chi^2_{N-1,\alpha/2}\}}(y)$ is α-level test for $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$.

b) $\varphi(y) = \mathbb{1}_{\{T(y) > \chi^2_{N-1,1-\alpha}\}}(y)$ is α-level test for $H_0 : \sigma^2 \leq \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$.

c) $\varphi(y) = \mathbb{1}_{\{T(y) < \chi^2_{N-1,\alpha}\}}(y)$ is α-level test for $H_0 : \sigma^2 \geq \sigma_0^2$ versus $H_1 : \sigma^2 < \sigma_0^2$.

Proof. According to Theorem 8.5.1, $\widehat{v}(Y)$ has a central χ^2-distribution with $N-1$ degrees of freedom, if $Y_n \sim \mathcal{N}(\mu, \sigma_0^2)$ for all $n = 1, \ldots, N$. This implies at once the assertion a).

b) Since for every $\sigma^2 \in \mathbb{R}^+$, $\frac{N-1}{\sigma^2}\widehat{\sigma}^2(Y)$ has a central χ^2-distribution with $N-1$ degrees of freedom,

if $Y_n \sim \mathcal{N}(\mu, \sigma^2)$ for all $n = 1, \ldots, N$, we have for arbitrary $\sigma^2 \leq \sigma_0^2$

$$\begin{aligned} P_{\sigma^2}(\varphi(Y) = 1) &= P_{\sigma^2}\left(\frac{N-1}{\sigma_0^2}\,\widehat{\sigma}^2(Y) > \chi^2_{N-1,1-\alpha}\right) \\ &= P_{\sigma^2}\left(\frac{N-1}{\sigma^2}\,\widehat{\sigma}^2(Y) > \frac{\sigma_0^2}{\sigma^2}\chi^2_{N-1,1-\alpha}\right) \\ &\leq P_{\sigma^2}\left(\frac{N-1}{\sigma^2}\,\widehat{\sigma}^2(Y) > \chi^2_{N-1,1-\alpha}\right) = 1 - F_{\chi^2(N-1,0)}(\chi^2_{N-1,1-\alpha}) = 1-(1-\alpha) = \alpha. \end{aligned}$$

The assertions c) follows like that in b). □

The probabilities of the β-error can be calculated for the t-tests given in Theorem 11.1.3. We will give here only the β-error for the two-sided alternative. The other β-errors can be obtained completely similarly.

8.5.5 Theorem (β-error of the two-sided one-sample t-test)
Let $Y_1, \ldots, Y_N \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. and $\varphi(y) = \mathbb{1}_{\{|\widehat{d}(y)| > t_{N-1,1-\alpha/2}\}}(y)$ be the α-level test of Theorem 11.1.3 for $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Then the probabilities of the β-error are given by

$$P_\mu(\varphi(Y) = 0) = F_{t(N-1,\delta(\mu))}(t_{N-1,1-\alpha/2}) - F_{t(N-1,\delta(\mu))}(-t_{N-1,1-\alpha/2}),$$

where $\delta(\mu) = \sqrt{N}\frac{\mu-\mu_0}{\sigma}$.

Proof. If μ is the true value, then Theorem 8.3.4 or Corollary 8.3.5, respectively, provide (compare with the proof of Theorem 11.1.3 a))

$$\begin{aligned} \sqrt{N}\frac{\overline{Y}-\mu_0}{\sigma} &\sim \mathcal{N}\left(\frac{1}{\sqrt{N}\sigma}1_N^\top\,\mu\,1_N - \frac{\sqrt{N}}{\sigma}\mu_0\,,\ \frac{1}{\sqrt{N}\sigma}1_N^\top\,\sigma^2 I_{N\times N}\,(\frac{1}{\sqrt{N}\sigma}1_N^\top)^\top\right) \\ &= \mathcal{N}\left(\sqrt{N}\frac{\mu-\mu_0}{\sigma}\,,\ 1\right) = \mathcal{N}(\delta(\mu), 1). \end{aligned}$$

This means with Theorem 8.5.1 that

$$\widehat{d}(y) = \sqrt{N}\,\frac{\overline{y}-\mu_0}{\widehat{\sigma}(y)}.$$

has a $t(N-1, \delta(\mu))$-distribution if μ is the true parameter. Hence

$$\begin{aligned} P_\mu(\varphi(Y) = 0) &= P_\mu\left(|\widehat{d}(y)| > t_{N-1,1-\alpha/2}\right) \\ &= P_\mu\left(\widehat{d}(y) < -t_{N-1,1-\alpha/2}\right) + P_\mu\left(\widehat{d}(y) > t_{N-1,1-\alpha/2}\right) \\ &= F_{t(N-1,\delta(\mu))}(t_{N-1,1-\alpha/2}) - F_{t(N-1,\delta(\mu))}(-t_{N-1,1-\alpha/2}). \end{aligned}$$

□

8.5.6 Example (Einhaltung von Produktionsvorgaben)
Bei der industriellen Produktion ist es nicht nur wichtig, dass die Produkte, die eine Maschine erzeugt, eine bestimmte Größe μ_0 einhalten sondern dass die Größe der Produkte auch nicht zu sehr um den geforderten Mittelwert schwankt. So sollten Schrauben, die einen bestimmten Durchmesser haben sollen, in ihrem Durchmesser nicht zu sehr von dem vorgegebenen Durchmesser abweichen. Bei der Überprüfung der Maschine wird man also zuerst

$$H_0 : \mu = \mu_0 \text{ gegen } H_1 : \mu \neq \mu_0$$

testen und dann

$$H_0 : \sigma^2 \leq \sigma_0^2 \text{ gegen } H_1 : \sigma^2 > \sigma_0^2$$

überprüfen, wobei σ_0^2 der vorgegebene Genauigkeitswert ist. Wird eine der beiden Hypothesen abgelehnt, darf die Maschine nicht weiterbenutzt werden.

Auch bei chemischen Analyse-Geräten und Methoden gilt diese Anforderung. Sie müssen im Mittel das Richtige liefern und die Ergebnisse dürfen nicht zu sehr um den Mittelwert schwanken.

Tests for two samples

Let $y_{11}, \ldots, y_{1N_1}$ be the measurements of the first sample and $y_{21}, \ldots, y_{2N_2}$ the measurements of the second sample. The sample sizes N_1 and N_2 can be equal or different. The vector of observations/measurements for the first sample is denoted by $y_{1\bullet} = (y_{11}, \ldots, y_{1N_1})^\top$ and the vector for the second sample by $y_{2\bullet} = (y_{21}, \ldots, y_{2N_2})^\top$. Here we will assume that $y_{11}, \ldots, y_{1N_1}$ are realizations of independent identically distributed random variables $Y_{11}, \ldots, Y_{1N_1}$ with normal distribution $\mathcal{N}(\mu_1, \sigma_1^2)$ and that $y_{21}, \ldots, y_{2N_2}$ are realizations of independent identically distributed random variables $Y_{21}, \ldots, Y_{2N_2}$ with normal distribution $\mathcal{N}(\mu_2, \sigma_2^2)$. The two samples are stochastically independent, i.e. $Y_{1\bullet} = (Y_{11}, \ldots, Y_{1N_1})^\top$ and $Y_{2\bullet} = (Y_{21}, \ldots, Y_{2N_2})^\top$ are stochastically independent. Here we have again two main test problems:

$$\begin{aligned}
&a) \quad H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2,\\
&b) \quad H_0 : \mu_1 \leq \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2,\\
&c) \quad H_0 : \mu_1 \geq \mu_2 \text{ versus } H_1 : \mu_1 < \mu_2,
\end{aligned}$$

and

$$\begin{aligned}
&a) \quad H_0 : \sigma_1^2 = \sigma_2^2 \text{ versus } H_1 : \sigma_1^2 \neq \sigma_2^2,\\
&b) \quad H_0 : \sigma_1^2 \leq \sigma_2^2 \text{ versus } H_1 : \sigma_1^2 > \sigma_2^2,\\
&c) \quad H_0 : \sigma_1^2 \geq \sigma_2^2 \text{ versus } H_1 : \sigma_1^2 < \sigma_2^2.
\end{aligned}$$

Let be

$$\overline{y}_{1\bullet} = \frac{1}{N_1} \sum_{n=1}^{N_1} y_{1n} \quad \text{and} \quad \overline{y}_{2\bullet} = \frac{1}{N_2} \sum_{n=1}^{N_2} y_{2n}$$

the estimates for μ_1 and μ_2, respectively,

$$\widehat{\sigma}^2(y_{1\bullet}) = \frac{1}{N_1 - 1}\sum_{n=1}^{N_1}(y_{1n} - \overline{y}_{1\bullet})^2 \text{ and } \widehat{\sigma}^2(y_{2\bullet}) = \frac{1}{N_2 - 1}\sum_{n=1}^{N_2}(y_{2n} - \overline{y}_{2\bullet})^2$$

the estimates for σ_1^2 and σ_2^2, respectively, and

$$\widehat{\sigma}_{12}^2 = \frac{1}{N_1 + N_2 - 2}\left(\sum_{n=1}^{N_1}(y_{1n} - \overline{y}_{1\bullet})^2 + \sum_{n=1}^{N_2}(y_{2n} - \overline{y}_{2\bullet})^2\right).$$

the pooled variance estimate.

8.5.7 Theorem (t-test for two samples)
Let be $Y_{11}, \ldots, Y_{1N_1}, Y_{21}, \ldots, Y_{2N_2}$ *stochastically independent with* $Y_{2n} \sim \mathcal{N}(\mu_1, \sigma^2)$ *for* $n = 1, \ldots, N_1$ *and* $Y_{2m} \sim \mathcal{N}(\mu_2, \sigma^2)$ *for* $m = 1, \ldots, N_2$ *and*

$$\widehat{d}(y) = \sqrt{\frac{N_1\, N_2}{N_1 + N_2}}\,\frac{\overline{y}_{1\bullet} - \overline{y}_{2\bullet}}{\widehat{\sigma}_{12}}.$$

Then:

a) $\varphi(y) = 1_{\{|\widehat{d}(y)| > t_{N_1+N_2-2,1-\alpha/2}\}}(y)$ *is* α*-level test for* $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 \neq \mu_2$.

b) $\varphi(y) = 1_{\{\widehat{d}(y) > t_{N_1+N_2-2,1-\alpha}\}}(y)$ *is* α*-level test for* $H_0 : \mu_1 \leq \mu_2$ *versus* $H_1 : \mu_1 > \mu_2$.

c) $\varphi(y) = 1_{\{\widehat{d}(y) < t_{N_1+N_2-2,\alpha}\}}(y)$ *is* α*-level test for* $H_0 : \mu_1 \geq \mu_2$ *versus* $H_1 : \mu_1 < \mu_2$.

Proof. Acoording to Theorem 8.5.1, $\frac{N_1+N_2-2}{\sigma^2}\widehat{\sigma}_{12}^2(Y)$ has a central χ^2-distribution with $N_1 + N_2 - 2$ degrees of freedom. Moreover $\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$ has a normal distribution. The parameters of this normal distribution can be determined by calculating the expectation and the variance:

$$\mathrm{E}\left(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}\right) = \mu_1 - \mu_2$$

and

$$\begin{aligned}
&\mathrm{var}\left(\sqrt{\frac{N_1\, N_2}{N_1 + N_2}}\left(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}\right)\right) \\
&= \frac{N_1\, N_2}{N_1 + N_2}\left(\mathrm{var}(\overline{Y}_{1\bullet}) + \mathrm{var}(\overline{Y}_{2\bullet})\right) \\
&= \frac{N_1\, N_2}{N_1 + N_2}\,\sigma^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right) \\
&= \sigma^2.
\end{aligned}$$

Hence

$$\sqrt{\frac{N_1\, N_2}{N_1 + N_2}}\,\frac{\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}}{\sigma^2}$$

has a $\mathcal{N}(0,1)$-distribution under $H_0 : \mu_1 = \mu_2$. According to Theorem 8.5.1 c) and the stochastic independence of Y^1 and Y^2, we have that $\overline{Y}_{1\bullet}$, $\overline{Y}_{2\bullet}$, $\sum_{n=1}^{N_1}(Y_{1n} - \overline{Y}_{1\bullet})^2$, $\sum_{n=1}^{N_2}(Y_{2n} - \overline{Y}_{2\bullet})^2$ are stochastically independent. This implies that also $\sqrt{\frac{N_1 N_2}{N_1+N_2}}\left(\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}\right)$ and $\widehat{\sigma}^2_{12}(y)$ are stochastically independent such that Definition 8.3.8 provides that $\widehat{d}(Y)$ has a central t-distribution with $N_1 + N_2 - 2$ degrees of freedom under $H_0 : \mu_1 = \mu_2$. This implies at once the assertion a).

The proof of b) is an exercise since it is similar to the proof for the t-test for one sample. The only difference is that $\mu_1 - \mu_2$ must be subtracted from $\overline{y}_{1\bullet} - \overline{y}_{2\bullet}$ so that the numerator has a $\mathcal{N}(0,1)$-distribution. The also the proof of part b) of Theorem 8.5.8. The assertion c) follows similarly like the assertion b). □

8.5.8 Theorem (Variance test for two samples)
Let be $Y_{11}, \ldots, Y_{1N_1}, Y_{21}, \ldots, Y_{2N_2}$ *stochastically independent with* $Y_{2n} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ *for* $n = 1, \ldots, N_1$ *and* $Y_{2m} \sim \mathcal{N}(\mu_2, \sigma_2^2)$ *for* $m = 1, \ldots, N_2$ *and*

$$\widehat{v}(y) = \frac{\widehat{\sigma}^2(y_{1\bullet})}{\widehat{\sigma}^2(y_{2\bullet})}.$$

Then:

a) $\varphi(y) = 1_{\{\widehat{v}(y) > F_{N_1-1,N_2-1,1-\alpha/2} \text{ or } \widehat{v}(y) < F_{N_1-1,N_2-1,\alpha/2}\}}(y)$ *is a* α*-level test for* $H_0 : \sigma_1^2 = \sigma_2^2$ *versus* $H_1 : \sigma_1^2 \neq \sigma_2^2$.

b) $\varphi(y) = 1_{\{\widehat{v}(y) > F_{N_1-1,N_2-1,1-\alpha}\}}(y)$ *is a* α*-level test for* $H_0 : \sigma_1^2 \leq \sigma_2^2$ *versus* $H_1 : \sigma_1^2 > \sigma_2^2$.

c) $\varphi(y) = 1_{\{\widehat{v}(y) < F_{N_1-1,N_2-1,\alpha}\}}(y)$ *is a* α*-level test for* $H_0 : \sigma_1^2 \geq \sigma_2^2$ *versus* $H_1 : \sigma_1^2 < \sigma_2^2$.

Proof. According to Theorem 8.5.1, $\frac{N_1-1}{\sigma_1^2}\widehat{\sigma}^2(y_{1\bullet})$ has a χ^2-distribution with $N_1 - 1$ degrees of freedom and $\frac{N_2-1}{\sigma_2^2}\widehat{\sigma}^2(y_{2\bullet})$ has χ^2-distribution with $N_2 - 1$ degrees of freedom.

a) Under the null hypotheses, we have $\sigma_1^2 = \sigma_2^2$ and that $\widehat{\sigma}^2(y_{1\bullet})$ and $\widehat{\sigma}^2(y_{2\bullet})$ are stochastically independnently, since $Y_{1\bullet}$ und $Y_{2\bullet}$ are stochastically independent. This means that $\widehat{v}(Y)$ has a F-distribution with $N_1 - 1$ and $N_2 - 1$ degrees of freedom according to Definition 8.3.9. This implies assertion a).

b) If $\sigma_1^2 \neq \sigma_2^2$, then

$$\frac{\frac{1}{\sigma_1^2}\widehat{\sigma}^2(y_{1\bullet})}{\frac{1}{\sigma_2^2}\widehat{\sigma}^2(y_{2\bullet})}$$

has a F-distribution with N_1-1 and N_2-1 degrees of freedom. Hence for $\sigma_1^2 \le \sigma_2^2$, we have

$$\begin{aligned} P_{\sigma_1^2,\sigma_2^2}(\varphi(Y)=1) &= P_{\sigma_1^2,\sigma_2^2}\left(\frac{\widehat{\sigma}^2(y_{1\bullet})}{\widehat{\sigma}^2(y_{2\bullet})} > F_{N_1-1,N_2-1,1-\alpha}\right) \\ &= P_{\sigma_1^2,\sigma_2^2}\left(\frac{\frac{1}{\sigma_1^2}\widehat{\sigma}^2(y_{1\bullet})}{\frac{1}{\sigma_2^2}\widehat{\sigma}^2(y_{2\bullet})} > \frac{\sigma_2^2}{\sigma_1^2}F_{N_1-1,N_2-1,1-\alpha}\right) \\ &\le P_{\sigma_1^2,\sigma_2^2}\left(\frac{\frac{1}{\sigma_1^2}\widehat{\sigma}^2(y_{1\bullet})}{\frac{1}{\sigma_2^2}\widehat{\sigma}^2(y_{2\bullet})} > F_{N_1-1,N_2-1,1-\alpha}\right) = 1 - F_{F_{N_1-1,N_2-1}}(F_{N_1-1,N_2-1,1-\alpha}) = \alpha. \end{aligned}$$

The proof of assertion c) is similar to that of b). □

8.5.9 Theorem (β-error of the two-sided two-sample t-test)
Let be $Y_{11},\dots,Y_{1N_1},Y_{21},\dots,Y_{2N_2}$ *stochastically independent with* $Y_{2n} \sim \mathcal{N}(\mu_1,\sigma^2)$ *for* $n=1,\dots,N_1$ *and* $Y_{2m} \sim \mathcal{N}(\mu_2,\sigma^2)$ *for* $m=1,\dots,N_2$ *and* $\varphi(y) = \mathbb{1}_{\{|\widehat{d}(y)|>t_{N_1+N_2-2,1-\alpha/2}\}}(y)$ *be the* α*-level test of Theorem 8.5.7 for* $H_0: \mu_1=\mu_2$ *versus* $H_1: \mu_1 \ne \mu_2$. *Then the probabilities of the* β*-error are given by*

$$P_{\mu_1,\mu_2}(\varphi(Y)=0) = F_{t(N_1+N_2-2,\delta(\mu_1,\mu_2))}(t_{N_1+N_2-2,1-\alpha/2}) - F_{-t(N_1+N-2-2,\delta(\mu_1,\mu_2))}(t_{N_1+N_2-2,1-\alpha/2})),$$

where $\delta(\mu_1,\mu_2) = \sqrt{\frac{N_1N_2}{N_1+N_2}}\frac{\mu_1-\mu_2}{\sigma}$.

Proof. If $\mu_1-\mu_2$ is the true difference, then

$$\mathrm{E}\left(\sqrt{\frac{N_1\,N_2}{N_1+N_2}}\;\frac{\overline{Y}_{1\bullet}-\overline{Y}_{2\bullet}}{\sigma}\right) = \sqrt{\frac{N_1\,N_2}{N_1+N_2}}\frac{\mu_1-\mu_2}{\sigma} = \delta(\mu_1,\mu_2)$$

and (see the proof of Theorem 8.5.7)

$$\mathrm{var}\left(\sqrt{\frac{N_1\,N_2}{N_1+N_2}}\;\frac{\overline{Y}_{1\bullet}-\overline{Y}_{2\bullet}}{\sigma}\right) = 1.$$

Hence

$$\sqrt{\frac{N_1\,N_2}{N_1+N_2}}\;\frac{\overline{Y}_{1\bullet}-\overline{Y}_{2\bullet}}{\sigma}$$

has a $\mathcal{N}(\delta(\mu_1,\mu_2),1)$-distribution. This means with Theorem 8.5.1 that

$$\widehat{d}(y) = \sqrt{\frac{N_1\,N_2}{N_1+N_2}}\;\frac{\overline{y}_{1\bullet}-\overline{y}_{2\bullet}}{\widehat{\sigma}_{12}}.$$

has a $t(N_1 + N_2 - 2, \delta(\mu_1, \mu_2))$-distribution if $\mu_1,\ \mu_2$ are the true parameters. Hence

$$\begin{aligned} P_{\mu_1,\mu_2}(\varphi(Y) = 0) &= P_\mu\left(|\widehat{d}(y)| > t_{N_1+N_2-2,1-\alpha/2}\right) \\ &= P_\mu\left(\widehat{d}(y) < -t_{N_1+N_2-2,1-\alpha/2}\right) + P_\mu\left(\widehat{d}(y) > t_{N_1+N_2-2,1-\alpha/2}\right) \\ &= F_{t(N_1+N_2-2,\delta(\mu_1,\mu_2))}(t_{N_1+N_2-2,1-\alpha/2}) - F_{t(N_1+N_2-2,\delta(\mu_1,\mu_2))}(-t_{N_1+N_2-2,1-\alpha/2}). \end{aligned}$$

□

9 The general linear model

In the general linear model it is assumed that the data $y_1, \ldots, y_N$ are realizations of stochastically independent random variables $Y_1, \ldots, Y_N$ which satisfy

$$\mathrm{E}(Y_n) = x(t_n)^\top \beta \tag{15}$$

or, respectively,

$$Y_n = x(t_n)^\top \beta + Z_n \quad \text{with } \mathrm{E}(Z_n) = 0 \tag{16}$$

for $n = 1, \ldots, N$. Thereby $\beta \in \mathbb{R}^R$ is an unknown parameter vector, $t_1, \ldots, t_N \in \mathcal{T}$ are known **experimental conditions**, also called **design points**, in the **design region** $\mathcal{T}$ and $x : \mathcal{T} \longrightarrow \mathbb{R}^R$ a known **regression function**. $Z_1, \ldots, Z_N$ are error variables which usually satisfy $\mathrm{var}(Z_n) = \sigma^2$ for all $n = 1, \ldots, N$.

Setting $Y = (Y_1, \ldots, Y_N)^\top$, $Z = (Z_1, \ldots, Z_N)^\top$, $X = (x(t_1), \ldots, x(t_N))^\top$, the model (15) or (16), respectively, can be written as

$$Y = X\beta + Z \quad \text{with } \mathrm{E}(Z) = 0_N \quad \text{and} \quad \mathrm{Cov}(Z) = \sigma^2 I_{N\times N},$$

where $0_N \in \mathbb{R}^N$ is the N dimensional vector consisting only of zeros and $I_{N\times N}$ is the $N \times N$ identity matrix. $X = (x(t_1), \ldots, x(t_N))^\top$ is also called **design matrix**. Sometimes we also write X_d to express the dependence on the **design** $d = (t_1, \ldots, t_N)$.

9.0.10 Example (One sample problem)
The one sample problem given in 8.5 is a special linear model where $\beta = \mu \in \mathbb{R}$ and $x(t) = 1$ for all t. Hence

$$Y_n = x(t_n)^\top \beta + Z_n = \mu + Z_n.$$

If $Z_n \sim \mathcal{N}(0, \sigma^2)$ for $n = 1, \ldots, N$, then $\mathrm{E}(Z_n) = 0$ and $\mathrm{var}(Z_n) = \sigma^2$ so that $Y_n \sim \mathcal{N}(\mu, \sigma^2)$ for $n = 1, \ldots, N$.

9.0.11 Example (Two sample problem)
The two sample problem given in 8.5 is also a special linear model where $\beta = (\mu_1, \mu_2)^\top \in \mathbb{R}^2$ and $x(t) = (\mathbb{1}_{\{1\}}(t), \mathbb{1}_{\{2\}}(t))^\top$ for $t \in \{1, 2\}$. Thereby $t_n \in \{1, 2\}$ denotes from which sample the n'the observation is. If the n'the observation is from the first sample, then $t_n = 1$ and

$$Y_n = x(t_n)^\top \beta + Z_n = (\mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n)) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + Z_n = (1, 0) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + Z_n = \mu_1 + Z_n.$$

$Z_n \sim \mathcal{N}(0, \sigma^2)$ implies then $Y_n \sim \mathcal{N}(\mu_1, \sigma^2)$ for $n = 1, \ldots, N$. If the n'the observation is from the second sample, then $t_n = 2$ and

$$Y_n = x(t_n)^\top \beta + Z_n = (\mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n)) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + Z_n = (0, 1) \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + Z_n = \mu_2 + Z_n.$$

Here, $Z_n \sim \mathcal{N}(0, \sigma^2)$ implies $Y_n \sim \mathcal{N}(\mu_2, \sigma^2)$ for $n = 1, \ldots, N$.

9.0.12 Example (Linear regression)
In the linear regression model, a linear relation between quantitative experimental conditions t_n and the observations/measurements y_n is assumed, i.e.

$$Y_n = \beta_0 + \beta_1 t_n + Z_n = (1, t_n) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + Z_n = x(t_n)^\top \beta + Z_n$$

with $x(t) = (1, t)^\top$ and $\beta = (\beta_0, \beta_1)^\top$.

9.1 Identifiability

In many examples, it happens that the design matrix $X = (x(t_1), \ldots, x(t_N))^\top \in \mathbb{R}^{N \times R}$ is of full rank R. However, there are also many examples where the rank of X is less than R, i.e. $\mathrm{rk}(X) < R$.

9.1.1 Example (One-way layout)
In the one-way layout, we assume that a qualitative factor A can attain A levels. These A levels provides A samples so that the two sample problem is a special case of the one-way layout with $A = 2$. For the one-way layout several parameterizations are possible.

Non-singular parameterization:

$$Y_n = \mu_a + Z_n, \quad \text{if } t_n = a, \quad \text{for } n = 1, \ldots, N,$$

i.e.

$$x(t_n) = (\mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^A,$$
$$\beta = (\mu_1, \mu_2, \ldots, \mu_A)^\top \in \mathbb{R}^A.$$

As soon as each level is observed at least once, then $X = (x(t_1), \ldots, x(t_N))^\top \in \mathbb{R}^{N \times A}$ is of full rank A. Sorting the observations/measurements with respect to the levels and assuming that each level is observed M times (i.e. we have balance design), then the design matrix can be written with the Kronecker product as

$$X = I_{A\times A} \otimes 1_M = \begin{pmatrix} 1_M & 0_M & 0_M & \ldots & 0_M & 0_M \\ 0_M & 1_M & 0_M & \ldots & 0_M & 0_M \\ 0_M & 0_M & 1_M & \ldots & 0_M & 0_M \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0_M & 0_M & 0_M & \ldots & 1_M & 0_M \\ 0_M & 0_M & 0_M & \ldots & 0_M & 1_M \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N \times A}.$$

Control parameterization:
Assume without loss of generality that the first level is the control level, for example a placebo in clinical studies or the standard crop in agricultural studies. Then we can set

$$\begin{aligned} Y_n &= \mu + Z_n \quad \text{if } t_n = 1, \\ Y_n &= \mu + \alpha_a + Z_n \quad \text{if } t_n = a, \quad \text{for } a = 2, \ldots, A. \end{aligned}$$

Then we have

$$\begin{aligned} x(t_n) &= (1, \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^A, \\ \beta &= (\mu, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^A. \end{aligned}$$

As soon as each level is observed at least once, then $X = (x(t_1), \ldots, x(t_N))^\top \in \mathbb{R}^{N \times A}$ is of full rank A. In balanced designs, the design matrix has now the form

$$X = \begin{pmatrix} 1_M & 0_{M\times(A-1)} \\ 1_{M(A-1)} & I_{(A-1)\times(A-1)} \otimes 1_M \end{pmatrix}$$

$$= \begin{pmatrix} 1_M & 0_M & 0_M & \ldots & 0_M & 0_M \\ 1_M & 1_M & 0_M & \ldots & 0_M & 0_M \\ 1_M & 0_M & 1_M & \ldots & 0_M & 0_M \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_M & 0_M & 0_M & \ldots & 1_M & 0_M \\ 1_M & 0_M & 0_M & \ldots & 0_M & 1_M \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N \times A}.$$

Singular parameterization:
This parameterization is preferred in applications although the corresponding design matrix is not of full rank:

$$\begin{aligned} Y_n &= \mu + \alpha_a + Z_n \quad \text{if } t_n = a, \quad \text{for } a = 1, \ldots, A, \\ x(t_n) &= (1, \mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^{A+1}, \\ \beta &= (\mu, \alpha_1, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^{A+1}. \end{aligned}$$

Here the design matrix $X \in \mathbb{R}^{N\times(A+1)}$ is never of full rank since it has $A+1$ columns. In balanced

design it has the form:

$$X = \begin{pmatrix} 1_{M\,A} & I_{A\times A} \otimes 1_M \end{pmatrix}$$

$$= \begin{pmatrix} 1_M & 1_M & 0_M & 0_M & \dots & 0_M & 0_M \\ 1_M & 0_M & 1_M & 0_M & \dots & 0_M & 0_M \\ 1_M & 0_M & 0_M & 1_M & \dots & 0_M & 0_M \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_M & 0_M & 0_M & 0_M & \dots & 1_M & 0_M \\ 1_M & 0_M & 0_M & 0_M & \dots & 0_M & 1_M \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{N\times(A+1)}.$$

X is not of full rank because there are two many parameters. Namely there are $A+1$ parameters where there are only A different experimental conditions. This means that not all of these $A+1$ parameter can be estimated by data, i.e. they are **not identifiable**. To avoid this problem one can use the side condition that

$$\sum_{a=1}^{A} \alpha_a = 0.$$

However, this requirement is not convenient mathematically. It is more convenient to use no side condition for the parameters. This is possible since the interest lies not in estimating (identifying) all parameters. The interest is here in estimating and testing the difference of level effects, i.e. we want to know whether the A levels provide different effects. This means that we are only interested in specific **aspects** of the unknown parameter vector β. For example, we may only interested in $\lambda(\beta) = \alpha_1 - \alpha_2$, the difference of the effects of the first and the second level. Statistical methods as estimators and tests should provide for such aspects the same results independently of the parameterization which is used.

9.1.2 Definition (Linear Aspect)
If $L \in \mathbb{R}^{S\times R}$, then $\lambda(\beta) = L\beta$ is called linear aspect of $\beta \in \mathbb{R}^R$.

9.1.3 Definition (Linear identifiability)
A linear aspect $\lambda(\beta) = L\beta$ is called linear identifiable at X ($d = (t_1, \dots, t_N)$, respectively) if and only if for all $\beta \in \mathbb{R}^R$ it holds

$$X\beta = 0 \implies L\beta = 0$$
$$(X_d\beta = 0 \implies L\beta = 0, \quad \textit{respectively}).$$

9.1.4 Theorem
The linear aspect $\lambda(\beta) = L\beta$ is linear identifiable at X if and only if there exists $K \in \mathbb{R}^{S\times N}$ such that $L = KX$.

Proof.
$\Leftarrow$: Clear.
$\Rightarrow$: Let be $b \in \mathbb{R}^R$ arbitrary and set

$$\beta = b - (X^\top X)^- X^\top X\, b.$$

Then Lemma 8.1.5 b) provides

$$X\beta = X(b - (X^\top X)^- X^\top X\, b) = Xb - X(X^\top X)^- X^\top X\, b = Xb - Xb = 0.$$

The linear identifiability of $\lambda(\beta) = L\beta$ implies

$$0 = L\beta = Lb - L(X^\top X)^- X^\top X\, b$$

and thus

$$Lb = L(X^\top X)^- X^\top X\, b$$

for all $b \in \mathbb{R}^R$. This means

$$L = L(X^\top X)^- X^\top X = KX$$

with $K = L(X^\top X)^- X^\top$. □

9.2 Estimators

9.2.1 Theorem
If $Z \sim \mathcal{N}_N(0_N, \sigma^2 I_{N\times N})$, $Y = X\beta + Z$, where $\sigma^2 \in \mathbb{R}^+$ and $\beta \in \mathbb{R}^R$ are unknown, then $(X\widehat{\beta}, \widehat{\sigma}^2)$ with

$$\widehat{\beta} = (X^\top X)^- X^\top y,$$
$$\widehat{\sigma}^2 = \frac{1}{N} y^\top \left(I_{N\times N} - X(X^\top X)^- X^\top\right) y = \frac{1}{N}(y - X\widehat{\beta})(y - X\widehat{\beta})^\top$$

is the unique maximum likelihood estimator for $(X\beta, \sigma^2)$.

Proof. Since $Y \sim \mathcal{N}_N(X\beta, \sigma^2 I_{N\times N})$, the density of Y is given by

$$f_{\beta,\sigma^2}(y) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma^2}(y-X\beta)^\top (y-X\beta)}$$

so that

$$L(\beta, \sigma^2, y) := \log f_{\beta,\sigma^2}(y) = -\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta) - \frac{N}{2}\log(2\pi\sigma^2)$$
$$= -\frac{1}{2\sigma^2}\left(y^\top y - 2y^\top X\beta + \beta^\top X^\top X\beta\right) - \frac{N}{2}\log(2\pi\sigma^2).$$

With $\frac{\partial}{\partial\beta}\beta^\top A\beta = 2\,A\,\beta$ for any $A \in \mathbb{R}^{R\times R}$, we obtain

$$\frac{\partial}{\partial\beta}L(\beta,\sigma^2,y) = -\frac{1}{\sigma^2}\left(X^\top X\,\beta - X^\top y\right) = 0$$

$$\iff \quad X^\top X\,\beta = X^\top y. \tag{17}$$

Any solution $\widetilde{\beta}$ of (17) satisfies according to Lemma 8.1.5 b)

$$X\,\widetilde{\beta} = X\,(X^\top X)^- X^\top X\,\widetilde{\beta} = X\,(X^\top X)^- X^\top y = X\,\widehat{\beta}.$$

Since $X\,(X^\top X)^- X^\top$ is independent of the choice of the g-inverse according to Lemma 8.1.5 d), we see that $X\,\widehat{\beta}$ is unique. Hence every $\widetilde{\beta}$ which maximizes $f_{\beta,\sigma^2}(y)$ satisfies $X\,\widetilde{\beta} = X\,\widehat{\beta}$.

Moreover, for each $\widetilde{\beta}$ which satisfies (17), we have

$$\frac{\partial}{\partial\sigma^2}L(\beta,\sigma^2,y)\Big|_{\beta=\widetilde{\beta}} = \frac{1}{2\sigma^4}(y - X\widetilde{\beta})^\top(y - X\widetilde{\beta}) - \frac{N}{2}\,\frac{1}{\sigma^2}$$

$$= \quad \frac{1}{2\sigma^4}(y - X\widehat{\beta})^\top(y - X\widehat{\beta}) - \frac{N}{2}\,\frac{1}{\sigma^2} \;=\; 0$$

$$\iff$$

$$\sigma^2 = \frac{1}{N}\left(y - X\widehat{\beta}\right)^\top\left(y - X\widehat{\beta}\right) = \frac{1}{N}\left(y^\top y - 2\widehat{\beta}^\top X^\top y + \widehat{\beta}^\top X^\top X\,\widehat{\beta}\right)$$

$$= \quad \frac{1}{N}\left(y^\top y - 2y^\top X(X^\top X)^- X^\top y + y^\top \underbrace{X(X^\top X)^- X^\top X}_{=X\ \text{(Lem. 8.1.5)}}(X^\top X)^- X^\top y\right)$$

$$= \quad \frac{1}{N}\left(y^\top y - y^\top X(X^\top X)^- X^\top y\right)$$

$$= \quad \frac{1}{N}y^\top\left(I_{N\times N} - X(X^\top X)^- X^\top\right)y.$$

Then for each $\widetilde{\beta}$ which satisfies (17), it holds

$$\frac{\partial^2}{\partial^2(\beta,\sigma^2)}L(\beta,\sigma^2,y)\Big|_{(\beta,\sigma^2)=(\widetilde{\beta},\widehat{\sigma}^2)} = \begin{pmatrix} -\frac{1}{\widehat{\sigma}^2}\,X^\top X & 0_{R\times 1} \\ 0_{1\times R} & -\frac{1}{N}\frac{1}{\widehat{\sigma}^4} \end{pmatrix}$$

which is a negative semidefinit matrix. Hence $L(\beta,\sigma^2,y)$ has a local maximum at each $(\widetilde{\beta},\widehat{\sigma}^2)$. With $(\widetilde{\beta},\widehat{\sigma}^2)$ also $(X\,\widetilde{\beta},\widehat{\sigma}^2) = (X\,\widehat{\beta},\widehat{\sigma}^2)$ is a maximum likelihood estimator and this estimator is unique. □

9.2.2 Remark
The estimator $X\,\widehat{\beta}$ of Theorem 9.2.1 satisfies

$$X\,\widehat{\beta} = P\,y,$$

where P is the perpendicular projection matrix onto $C(X)$, i.e. $X\,\widehat{\beta}$ is perpendicular projection of y onto $C(X) = \{X\beta;\ \beta \in \mathbb{R}^R\}$.

9.2.3 Theorem
Let be $\widehat{\beta} = (X^\top X)^- X^\top y$. *Any estimate* $\widetilde{\beta}$ *with* $X\widetilde{\beta} = X\widehat{\beta}$ *satisfies*

$$\widetilde{\beta} \in \arg\min_{\beta \in \mathbb{R}^R} (y - X\beta)^\top (y - X\beta),$$

i.e. $\widetilde{\beta}$ *is a least squares estimator.*

Proof. Because of $X(X^\top X)^- X^\top X = X$ (Lemma 8.1.5 b)), it holds for all $\beta \in \mathbb{R}^R$

$$\begin{aligned}
(y - X\beta)^\top (y - X\beta) &= (y - X\widehat{\beta} + X\widehat{\beta} - X\beta)^\top (y - X\widehat{\beta} + X\widehat{\beta} - X\beta) \\
&= (y - X\widehat{\beta})^\top (y - X\widehat{\beta}) + (y - X\widehat{\beta})^\top (X\widehat{\beta} - X\beta) \\
&\quad + (X\widehat{\beta} - X\beta)^\top (y - X\widehat{\beta}) + (X\widehat{\beta} - X\beta)^\top (X\widehat{\beta} - X\beta) \\
&= (y - X\widehat{\beta})^\top (y - X\widehat{\beta}) + (y - X(X^\top X)^- X^\top y)^\top (X\widehat{\beta} - X\beta) \\
&\quad + (X\widehat{\beta} - X\beta)^\top (y - X(X^\top X)^- X^\top y) + (X\widehat{\beta} - X\beta)^\top (X\widehat{\beta} - X\beta) \\
&= (y - X\widehat{\beta})^\top (y - X\widehat{\beta}) + (y^\top - y^\top X(X^\top X)^- X^\top) X (\widehat{\beta} - \beta) \\
&\quad + (\widehat{\beta} - \beta)^\top X^\top (y - X(X^\top X)^- X^\top y) + (X\widehat{\beta} - X\beta)^\top (X\widehat{\beta} - X\beta) \\
&= (y - X\widehat{\beta})^\top (y - X\widehat{\beta}) + y^\top (X - \underbrace{X(X^\top X)^- X^\top X}_{=X}) (\widehat{\beta} - \beta) \\
&\quad + (\widehat{\beta} - \beta)^\top (X^\top - \underbrace{X^\top X(X^\top X)^- X^\top}_{X^\top}) y + (X\widehat{\beta} - X\beta)^\top (X\widehat{\beta} - X\beta) \\
&= (y - X\widehat{\beta})^\top (y - X\widehat{\beta}) + (\widehat{\beta} - \beta)^\top X^\top X (\widehat{\beta} - \beta) \\
&\geq (y - X\widehat{\beta})^\top (y - X\widehat{\beta}).
\end{aligned}$$

Thereby, we have equality if and only if $X\beta = X\widehat{\beta} = X(X^\top X)^- X^\top y$. □

9.2.4 Theorem
If $Y = X\beta + Z$, $E(Z) = 0_N$, $Cov(Z) = \sigma^2 I_{N\times N}$, $\widehat{\beta}(y) = (X^\top X)^- X^\top y$, *and* $\lambda(\beta) = L\beta$ *is identifiable at* X, *then:*

a) $L\widehat{\beta}(y)$ *is unbiased estimator for* $L\beta$.

b) $\widehat{\sigma}^2(y) = \frac{1}{N - rk(X)}\, y^\top \left(I_{N\times N} - X(X^\top X)^- X^\top\right) y$ *is unbiased estimator for* σ^2.

c) $L\widehat{\beta}(y)$ *and* $\widehat{\sigma}^2(y)$ *do not dependent on the choice of the g-inverse.*

(unbiased estimator = erwartungstreue Schätzfunktion in German)

Proof.
a) The linear identifiability implies the existence of $K \in \mathbb{R}^{S\times N}$ with $L = KX$. Hence according to the linearity of the expectation and Lemma 8.1.5 b)

$$\begin{aligned}
\mathrm{E}(L\widehat{\beta}(Y)) &= \mathrm{E}(KX(X^\top X)^- X^\top Y) = KX(X^\top X)^- X^\top \mathrm{E}(Y) \\
&= K \underbrace{X(X^\top X)^- X^\top X}_{=X} \beta = KX\beta = L\beta
\end{aligned}$$

for all $\beta \in \mathbb{R}^r$.
b) $Y = X\beta + Z$, $\mathrm{E}(Z) = 0_N$, $\mathrm{Cov}(Z) = \sigma^2 I_{N\times N}$ imply $\mathrm{E}(Y) = X\beta + \mathrm{E}(Z) = X\beta$ and $\mathrm{Cov}(Y) = \mathrm{Cov}(Z) = \sigma^2 I_{N\times N}$ according to Lemma 8.2.6 b) and h). Set $P = I_{N\times N} - X(X^\top X)^- X^\top$. Then we have $\mathrm{tr}(P) = N - \mathrm{rk}(X)$ according to Lemma 8.1.11 c) and $PX = (I_{N\times N} - X(X^\top X)^- X^\top)X = X - X(X^\top X)^- X^\top X = 0_N$ according to 8.1.5 b). Lemma 8.2.6 i) provides then

$$\begin{aligned}\mathrm{E}(\widehat{\sigma}^2(Y)) &= \frac{1}{N-\mathrm{rk}(X)}\,\mathrm{E}(Y^\top P\,Y) = \frac{1}{N-\mathrm{rk}(X)}\,\Big(\mathrm{tr}(P\,\mathrm{Cov}(Y)) + \mathrm{E}(Y)^\top P\,\mathrm{E}(Y)\Big)\\ &= \frac{1}{N-\mathrm{rk}(X)}\,\Big(\mathrm{tr}(P\,\mathrm{Cov}(Y)) + \mathrm{E}(Y)^\top P\,\mathrm{E}(Y)\Big)\\ &= \frac{1}{N-\mathrm{rk}(X)}\,\Big(\mathrm{tr}(P\,\sigma^2 I_{N\times N}) + \beta^\top X^\top P\,X\,\beta\Big) \overset{PX=0}{=} \frac{\sigma^2}{N-\mathrm{rk}(X)}\,\mathrm{tr}(P) = \sigma^2.\end{aligned}$$

c) The uniqueness of $L\widehat{\beta}(y) = KX\,(X^\top X)^- X^\top y$ (see a)) and $\widehat{\sigma}^2(y)$ follows from the fact that $X\,(X^\top X)^- X^\top$ does not depend on the g-inverse according to Lemma 8.1.5 d). □

9.2.5 Lemma
If $Y = X\beta + Z$, $E(Z) = 0_N$, $Cov(Z) = \sigma^2 I_{N\times N}$, $\widehat{\beta}(y) = (X^\top X)^- X^\top y$, and $\lambda(\beta) = L\beta$ is identifiable at X, then

$$Cov(L\widehat{\beta}(Y)) = L\,(X^\top X)^- L^\top\,\sigma^2.$$

Proof. Since $\lambda(\beta) = L\beta$ is identifiable at X, there exists $K \in \mathbb{R}^{S\times N}$ with $L = KX$. Lemma 8.2.6 h) and Lemma 8.1.5 b) imply

$$\begin{aligned}\mathrm{Cov}(L\widehat{\beta}(Y)) &= \mathrm{Cov}(KX\,(X^\top X)^- X^\top Y) = KX\,(X^\top X)^- X^\top \mathrm{Cov}(Y) X\,(X^\top X)^- X^\top\,K^\top\\ &= KX\,(X^\top X)^- X^\top \mathrm{Cov}(Z) X\,(X^\top X)^- X^\top\,K^\top = KX\,(X^\top X)^- X^\top \sigma^2 I_{N\times N} X\,(X^\top X)^- X^\top\,K^\top\\ &= K\underbrace{X\,(X^\top X)^- X^\top X}_{=X}\,(X^\top X)^- X^\top\,K^\top\,\sigma^2 = KX\,(X^\top X)^- X^\top\,K^\top\,\sigma^2 = L\,(X^\top X)^- L^\top\,\sigma^2.\end{aligned}$$ □

The question is whether $L\widehat{\beta}(y) = L(X^\top X)^- X^\top y$ is the best estimator for $L\beta$. Estimators can be compared by their covariance matrices. Since there is no natural ordering of matrices we define for matrices $A, B \in \mathbb{R}^{N\times N}$:

9.2.6 Definition

$$A \le B :\Longleftrightarrow\ c^\top A\,c \le c^\top B\,c\ \ \textit{for all } c \in \mathbb{R}^N \Longleftrightarrow B - A\ \ \textit{is positive semidefinite.}$$

This is no complete ordering of matrices. It could be that matrices are not comparable with respect to this ordering.

9.2.7 Theorem (Gauss-Markov theorem)
If $Y = X\beta + Z$, $E(Z) = 0_N$, $Cov(Z) = \sigma^2 I_{N\times N}$, $\widehat{\beta}(y) = (X^\top X)^- X^\top y$, and $\lambda(\beta) = L\beta$ is identifiable

at X, then for all linear unbiased estimators Ay for $\lambda(\beta) = L\beta$ we have

$$Cov(L\widehat{\beta}(Y)) \leq Cov(AY).$$

Proof. Set $B = L(X^\top X)^- X^\top$ so that $L\widehat{\beta}(y) = By$. Since $L\beta$ is identifiable, there exists $K \in \mathbb{R}^{S\times N}$ with $L = KX$. Since Ay is unbiased estimator for $L\beta$, we have

$$KX\beta = L\beta = \mathrm{E}(AY) = A\,\mathrm{E}(Y) = AX\beta$$

for all $\beta \in \mathbb{R}^R$. This implies $KX = AX$ and

$$\begin{aligned}(A-B)B^\top &= (A - KX\,(X^\top X)^- X^\top)\,X\,(X^\top X)^- X^\top K^\top \\ &= AX\,(X^\top X)^- X^\top K^\top - K\underbrace{X\,(X^\top X)^- X^\top X}_{=X}(X^\top X)^- X^\top K^\top \\ &= AX\,(X^\top X)^- X^\top K^\top - KX\,(X^\top X)^- X^\top K^\top \\ &= KX(X^\top X)^- X^\top K^\top - KX\,(X^\top X)^- X^\top K^\top = 0_S.\end{aligned}$$

We obtain with Lemma 8.2.6 h)

$$\begin{aligned}&\mathrm{Cov}(AY) = A\,\mathrm{Cov}(Y)\,A^\top = A\,\mathrm{Cov}(Z)\,A^\top = A\,\sigma^2 I_{N\times N}\,A^\top = \sigma^2 AA^\top \\ &= \sigma^2(A-B+B)(A-B+B)^\top = \sigma^2\left((A-B)(A-B)^\top + \underbrace{(A-B)B^\top}_{=0} + \underbrace{B(A-B)^\top}_{=0} + BB^\top\right) \\ &= \sigma^2\left((A-B)(A-B)^\top + BB^\top\right) \\ &\geq \sigma^2\,BB^\top \overset{(\text{as for } AY)}{=} \mathrm{Cov}(BY) = \mathrm{Cov}(L\widehat{\beta}(Y)),\end{aligned}$$

since $c^\top(A-B)(A-B)^\top c \geq 0$ for all $c \in \mathbb{R}^N$. □

9.2.8 Example (One-way layout: Continuation of Example 9.1.1)
Assume that level a_1 (shortly level 1) of factor A is the control level (the placebo, the standard crop etc.) and that the effects of the $A-1$ other levels of the factor should be estimated as additional effect to the effect of level 1. These additional effects can be positive or negative. Assume that the observations/measurements are ordered according to the factor levels so that

$$y = (y_1, \ldots, y_N)^\top = (y_{11}, \ldots, y_{1N_1}, \ldots, y_{A1}, \ldots, y_{AN_A})^\top = (y_{1*}^\top, y_{2*}^\top, \ldots, y_{A*}^\top)^\top,$$

where $N = N_1 + N_2 + \ldots + N_A$ and

$$y_{a*} = (y_{a1}, \ldots, y_{aN_a})^\top$$

for all $a = 1, \ldots, A$. Set also

$$\begin{aligned}y_{a\bullet} &= 1_{N_a}^\top y_{a*} = \sum_{n=1}^{N_a} y_{an}, \\ \overline{y}_{a\bullet} &= \frac{1}{N_a}\,y_{a\bullet},\end{aligned}$$

for $a = 1, \ldots, N$, and

$$y_{\bullet\bullet} = \sum_{a=1}^{A} \sum_{n=1}^{N_a} y_{an} = \sum_{n=1}^{N} y_n,$$
$$\overline{y} = \overline{y}_{\bullet\bullet} = \frac{1}{N}\, y_{\bullet\bullet}.$$

Non-singular parameterization:
If

$$x(t_n) = (\mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^A,$$
$$\beta = (\mu_1, \mu_2, \ldots, \mu_A)^\top \in \mathbb{R}^A,$$

then the interesting aspect $\lambda(\beta)$ is

$$\lambda(\beta) = \begin{pmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = L\beta \in \mathbb{R}^{A-1}$$

with

$$L = \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times A}.$$

The design matrix X has here in the general case the form

$$X = \begin{pmatrix} 1_{N_1} & 0_{N_1} & 0_{N_1} & \ldots & 0_{N_1} & 0_{N_1} \\ 0_{N_2} & 1_{N_2} & 0_{N_2} & \ldots & 0_{N_2} & 0_{N_2} \\ 0_{N_3} & 0_{N_3} & 1_{N_3} & \ldots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \ldots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 0_{N_A} & 0_{N_A} & 0_{N_A} & \ldots & 0_{N_A} & 1_{N_A} \end{pmatrix}$$

so that

$$(X^\top X)^- = \begin{pmatrix} N_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & N_2 & 0 & \ldots & 0 & 0 \\ 0 & 0 & N_3 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & N_{A-1} & 0 \\ 0 & 0 & 0 & \ldots & 0 & N_A \end{pmatrix}^{-1}$$
$$= \quad \operatorname{diag}(N_1, N_2, \ldots, N_A)^{-1} = \operatorname{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_A}\right),$$

where $\mathrm{diag}(b_1, b_2, \ldots, b_N) \in \mathbb{R}^{N\times N}$ denotes a diagonal matrix with diagonal elements $b_1, b_2, \ldots, b_N$. Since

$$X^\top y = \begin{pmatrix} y_{1\bullet} \\ y_{2\bullet} \\ \vdots \\ y_{A\bullet} \end{pmatrix},$$

we obtain

$$\widehat{\beta} = (X^\top X)^- X^\top y = (X^\top X)^{-1} X^\top y = \begin{pmatrix} \frac{1}{N_1} y_{1\bullet} \\ \frac{1}{N_1} y_{2\bullet} \\ \vdots \\ \frac{1}{N_1} y_{A\bullet} \end{pmatrix} = \begin{pmatrix} \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} \\ \vdots \\ \overline{y}_{A\bullet} \end{pmatrix},$$

which is the unique estimator for β. Then

$$L(X^\top X)^{-1} X^\top y = \begin{pmatrix} \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} \in \mathbb{R}^{A-1}$$

is the unique estimator for $\lambda(\beta) = L\beta$. That $\lambda(\beta) = L\beta$ is identifiable at X follows with Theorem 9.1.4 from the fact that $X^\top X$ is non-singular since

$$L = L(X^\top X)^{-1} X^\top X = KX.$$

Control parameterization:
Here we have (see Example 9.1.1)

$$x(t_n) = (1, \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^A,$$
$$\beta = (\mu, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^A,$$

so that

$$\lambda(\beta) = (\alpha_2, \ldots, \alpha_A)^\top = L\beta$$

with

$$L = \left(0_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times A}$$

is the interesting aspect. The design matrix X has here in the general case the form

$$X = \begin{pmatrix} 1_{N_1} & 0_{N_1} & 0_{N_1} & \dots & 0_{N_1} & 0_{N_1} \\ 1_{N_2} & 1_{N_2} & 0_{N_2} & \dots & 0_{N_2} & 0_{N_2} \\ 1_{N_3} & 0_{N_3} & 1_{N_3} & \dots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \dots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 1_{N_A} & 0_{N_A} & 0_{N_A} & \dots & 0_{N_A} & 1_{N_A} \end{pmatrix}$$

so that

$$\begin{aligned} (X^\top X)^- &= \begin{pmatrix} N & N_2 & N_3 & \dots & N_{A-1} & N_A \\ N_2 & N_2 & 0 & \dots & 0 & 0 \\ N_3 & 0 & N_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ N_{A-1} & 0 & 0 & \dots & N_{A-1} & 0 \\ N_A & 0 & 0 & \dots & 0 & N_A \end{pmatrix}^{-1} \\ &= \begin{pmatrix} N & b^\top \\ b & \operatorname{diag}(N_2, N_3, \dots, N_A) \end{pmatrix}^{-1}, \end{aligned}$$

where $b = (N_2, N_3, \dots, N_A)^\top$. The inverse of $X^\top X$ is given by Lemma 8.1.6. For applying this lemma, set $B = b$, $A = \operatorname{diag}(N_2, N_3, \dots, N_A)$, and $C = N$. Then

$$\begin{aligned} E &= N - b^\top \operatorname{diag}(N_2, N_3, \dots, N_A)^{-1} b = N - b^\top \operatorname{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \dots, \frac{1}{N_A}\right) b \\ &= N - \sum_{a=2}^{A} N_a = N_1, \\ E^- B^\top A^{-1} &= \frac{1}{N_1} b^\top \operatorname{diag}(N_2, N_3, \dots, N_A)^{-1} \\ &= \frac{1}{N_1} b^\top \operatorname{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \dots, \frac{1}{N_A}\right) = \frac{1}{N_1} 1_{A-1}^\top \\ A^{-1} + A^{-1} B\, E^- B^\top A^{-1} &= \operatorname{diag}(N_2, N_3, \dots, N_A)^{-1} + \operatorname{diag}(N_2, N_3, \dots, N_A)^{-1} b \frac{1}{N_1} b^\top \operatorname{diag}(N_2, N_3, \dots, N_A)^{-1} \\ &= \operatorname{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \dots, \frac{1}{N_A}\right) + \frac{1}{N_1} 1_{A-1} 1_{A-1}^\top \\ &= \operatorname{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \dots, \frac{1}{N_A}\right) + \frac{1}{N_1} 1_{(A-1)\times(A-1)} \end{aligned}$$

such that

$$(X^\top X)^{-1} = \begin{pmatrix} \frac{1}{N_1} & -\frac{1}{N_1} 1_{A-1}^\top \\ -\frac{1}{N_1} 1_{A-1} & \operatorname{diag}\left(\frac{1}{N_2}, \frac{1}{N_3}, \dots, \frac{1}{N_A}\right) + \frac{1}{N_1} 1_{(A-1)\times(A-1)} \end{pmatrix}.$$

With

$$X^\top y = \begin{pmatrix} y_{\bullet\bullet} \\ y_{2\bullet} \\ \vdots \\ y_{A\bullet} \end{pmatrix},$$

we obtain

$$\begin{aligned}
\widehat{\beta} = (X^\top X)^{-1} X^\top y &= \begin{pmatrix} \frac{y_{\bullet\bullet}}{N_1} - \frac{1}{N_1} 1_{A-1}^\top (y_{2\bullet}, \ldots, y_{A\bullet})^\top \\ -\frac{y_{\bullet\bullet}}{N_1} 1_{A-1} + \operatorname{diag}(\frac{y_{2\bullet}}{N_2}, \frac{y_{3\bullet}}{N_3}, \ldots, \frac{y_{A\bullet}}{N_A}) + \frac{1}{N_1} 1_{A-1}\, 1_{A-1}^\top (y_{2\bullet}, \ldots, y_{A\bullet})^\top \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{N_1} \left(y_{\bullet\bullet} - \sum_{a=2}^A y_{a\bullet} \right) \\ \operatorname{diag}(\overline{y}_{2\bullet}, \overline{y}_{3\bullet}, \ldots, \overline{y}_{A\bullet}) - 1_{A-1} \frac{1}{N_1} \left(y_{\bullet\bullet} - \sum_{a=2}^A y_{a\bullet} \right) \end{pmatrix} \\
&= \begin{pmatrix} \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{4\bullet} - \overline{y}_{1\bullet} \end{pmatrix}
\end{aligned}$$

which is the unique estimator for β. Then

$$L(X^\top X)^{-1} X^\top y = \left(0_{A-1} \mid I_{(A-1)\times(A-1)}\right) \begin{pmatrix} \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} = \begin{pmatrix} \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} \in \mathbb{R}^{A-1}$$

is the same unique estimator for $\lambda(\beta) = L\beta$ as we obtained for the non-singular parametrization. That $\lambda(\beta) = L\beta$ is identifiable at X follows as for the non-singular parametrization from the fact that $X^\top X$ is non-singular.

Singular parameterization:
Here we have (see Example 9.1.1)

$$\begin{aligned}
&x(t_n) = (1, \mathbb{1}_{\{1\}}(t_n), \mathbb{1}_{\{2\}}(t_n), \ldots, \mathbb{1}_{\{A\}}(t_n))^\top \in \mathbb{R}^{A+1}, \\
&\beta = (\mu, \alpha_1, \alpha_2, \ldots, \alpha_A)^\top \in \mathbb{R}^{A+1}.
\end{aligned}$$

so that

$$\lambda(\beta) = (\alpha_2 - \alpha_1, \ldots, \alpha_A - \alpha_1)^\top = L\beta$$

with

$$L = \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times(A+1)}$$

is the interesting aspect. The design matrix X has here in the general case the form

$$X = \begin{pmatrix} 1_{N_1} & 1_{N_1} & 0_{N_1} & 0_{N_1} & \dots & 0_{N_1} & 0_{N_1} \\ 1_{N_2} & 0_{N_2} & 1_{N_2} & 0_{N_2} & \dots & 0_{N_2} & 0_{N_2} \\ 1_{N_3} & 0_{N_3} & 0_{N_3} & 1_{N_3} & \dots & 0_{N_3} & 0_{N_3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 1_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & 0_{N_{A-1}} & \dots & 1_{N_{A-1}} & 0_{N_{A-1}} \\ 1_{N_A} & 0_{N_A} & 0_{N_A} & 0_{N_A} & \dots & 0_{N_A} & 1_{N_A} \end{pmatrix}$$

so that

$$X^\top X = \begin{pmatrix} N & N_1 & N_2 & N_3 & \dots & N_{A-1} & N_A \\ N_1 & N_1 & 0 & 0 & \dots & 0 & 0 \\ N_2 & 0 & N_2 & 0 & \dots & 0 & 0 \\ N_3 & 0 & 0 & N_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ N_{A-1} & 0 & 0 & 0 & \dots & N_{A-1} & 0 \\ N_A & 0 & 0 & 0 & \dots & 0 & N_A \end{pmatrix}$$

$$= \begin{pmatrix} N & b^\top \\ b & \mathrm{diag}(N_1, N_2, \dots, N_A) \end{pmatrix} \in \mathbb{R}^{(A+1)\times(A+1)},$$

where $b = (N_1, N_2, \dots, N_A)^\top$. Here $X^\top X$ is singular so that only a g-inverse can be calculated. The g-inverse of $X^\top X$ is given by Lemma 8.1.6. For applying this lemma, set $B = b$, $A = \mathrm{diag}(N_1, N_2, \dots, N_A)$, and $C = N$. Then

$$E = N - b^\top \mathrm{diag}(N_1, N_2, \dots, N_A)^{-1} b = N - b^\top \mathrm{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \dots, \frac{1}{N_A}\right) b$$

$$= N - \sum_{a=1}^{A} N_a = N - N = 0,$$

Then E^- can be any value $c \in \mathbb{R}$ since $0c0 = 0$. Hence set $E^- = c \in \mathbb{R}$. Then

$$\begin{aligned}
&E^- B^\top A^{-1} = c\, b^\top \mathrm{diag}(N_1, N_2, \ldots, N_A)^{-1} \\
&= c\, b^\top \mathrm{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_A}\right) = c\, 1_A^\top \\
&A^{-1} + A^{-1} B\, E^- B^\top A^{-1} \\
&= \mathrm{diag}(N_1, N_2, \ldots, N_A)^{-1} + \mathrm{diag}(N_1, N_2, \ldots, N_A)^{-1} b\, c\, b^\top \mathrm{diag}(N_1, N_2, \ldots, N_A)^{-1} \\
&= \mathrm{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_A}\right) + c\, 1_A 1_A^\top \\
&= \mathrm{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_A}\right) + c\, 1_{A\times A}
\end{aligned}$$

such that

$$(X^\top X)^- = \begin{pmatrix} c & -c\, 1_A^\top \\ -c\, 1_A & \mathrm{diag}\left(\frac{1}{N_1}, \frac{1}{N_2}, \ldots, \frac{1}{N_A}\right) + c\, 1_{A\times A} \end{pmatrix}.$$

With

$$X^\top y = \begin{pmatrix} y_{\bullet\bullet} \\ y_{1\bullet} \\ \vdots \\ y_{A\bullet} \end{pmatrix},$$

we obtain that

$$\begin{aligned}
\widehat{\beta} = (X^\top X)^- X^\top y &= \begin{pmatrix} c\, y_{\bullet\bullet} - c\, 1_A^\top\, (y_{1\bullet}, \ldots, y_{A\bullet})^\top \\ -c\, y_{\bullet\bullet}\, 1_A + \mathrm{diag}(\frac{y_{1\bullet}}{N_1}, \frac{y_{2\bullet}}{N_2}, \ldots, \frac{y_{A\bullet}}{N_A}) + c\, 1_A\, 1_A^\top\, (y_{1\bullet}, \ldots, y_{A\bullet})^\top \end{pmatrix} \\
&= \begin{pmatrix} c\, \left(y_{\bullet\bullet} - \sum_{a=1}^A y_{a\bullet}\right) \\ \mathrm{diag}(\overline{y}_{1\bullet}, \overline{y}_{2\bullet}, \ldots, \overline{y}_{A\bullet}) - 1_A c\, \left(y_{\bullet\bullet} - \sum_{a=1}^A y_{a\bullet}\right) \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ \overline{y}_{1\bullet} \\ \overline{y}_{2\bullet} \\ \vdots \\ \overline{y}_{A\bullet} \end{pmatrix}
\end{aligned}$$

is a least squares estimator for β for all $c \in \mathbb{R}$. Hence $\widehat{\beta}$ is unique although $(X^\top X)^-$ is not unique. However, $\widehat{\beta}$ is not the only least squares estimator since $\widehat{\beta} + \mu\gamma$ with $\gamma = (1, -1, -1, \ldots, -1)^\top$

and $\mu \in \mathbb{R}$ is also a least squares estimator according to Theorem 9.2.3 since $X\gamma = 0$ so that $X\widehat{\beta} = X(\widehat{\beta} + \mu\gamma)$. The property $X\gamma = 0$ for $\gamma \neq 0_{A+1}$ means according to Definition 9.1.3 that β is not identifiable at X. However, $\lambda(\beta) = L\beta$ is identifiable since

$$L = \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) = K\,X$$

for

$$K = \begin{pmatrix} -\frac{1}{N_1}1_{N_1}^\top & \frac{1}{N_2}1_{N_2}^\top & 0 & 0 & \dots & 0 \\ -\frac{1}{N_1}1_{N_1}^\top & 0 & \frac{1}{N_3}1_{N_3}^\top & 0 & \dots & 0 \\ -\frac{1}{N_1}1_{N_1}^\top & 0 & 0 & \frac{1}{N_4}1_{N_4}^\top & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ -\frac{1}{N_1}1_{N_1}^\top & 0 & 0 & 0 & \dots & \frac{1}{N_A}1_{N_A}^\top \end{pmatrix} \in \mathbb{R}^{(A-1)\times N}.$$

In particular for any $\widetilde{\beta} = \widehat{\beta} + \mu\gamma$ with $\mu \in \mathbb{R}$ we have

$$\begin{aligned} L\widetilde{\beta} &= L\left((X^\top X)^{-1}X^\top y + \mu\gamma\right) \\ &= \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \begin{pmatrix} \mu \\ \overline{y}_{1\bullet} - \mu \\ \overline{y}_{2\bullet} - \mu \\ \vdots \\ \overline{y}_{A\bullet} - \mu \end{pmatrix} = \begin{pmatrix} \overline{y}_{2\bullet} - \overline{y}_{1\bullet} \\ \overline{y}_{3\bullet} - \overline{y}_{1\bullet} \\ \vdots \\ \overline{y}_{A\bullet} - \overline{y}_{1\bullet} \end{pmatrix} \in \mathbb{R}^{A-1}. \end{aligned}$$

This is the same unique estimator for $\lambda(\beta) = L\beta$ as we obtained for the other parametrizations.

Since all estimators for the additional effects of the $A-1$ factor levels which are not the control are unique and do not depend on the parametrization we can calculate the covariance matrix of the estimator with the non-singular parametrization:

$$\begin{aligned} \mathrm{Cov}(L\widehat{\beta}) &= \sigma^2\, L(X^\top X)^{-1}L^\top \\ &= \sigma^2 \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \mathrm{diag}\left(\frac{1}{N_1}, \dots, \frac{1}{N_A}\right) \begin{pmatrix} -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \\ &= \sigma^2 \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \begin{pmatrix} -\frac{1}{N_1}\,1_{A-1}^\top \\ \mathrm{diag}\left(\frac{1}{N_2}, \dots, \frac{1}{N_A}\right) \end{pmatrix} \\ &= \sigma^2 \left(\frac{1}{N_1}\,1_{(A-1)\times(A-1)} + \mathrm{diag}\left(\frac{1}{N_2}, \dots, \frac{1}{N_A}\right)\right). \end{aligned}$$

For the special case of $A = 2$ we obtain

$$\mathrm{Cov}(L\widehat{\beta}) = \mathrm{var}(\overline{Y}_{2\bullet} - \overline{Y}_{1\bullet}) = \sigma^2 \left(\frac{1}{N_1} + \frac{1}{N_2}\right).$$

Hence minimizing the variance of the estimator $\overline{y}_{2\bullet} - \overline{y}_{1\bullet}$ leads to the same optimization problem as in Exercise 2.4.1.

9.3 Tests

9.3.1 Example (One-way layout: Continuation of Example 9.2.8)
The aim is to test whether the different levels of the factor A have different effects for the oberservations. If they do not have different effects then we say that the factor A has no influence. Hence we have to decide between

H_0 : Factor A has no influence

versus

H_1 : Factor A has an influence.

The null hypothesis H_0 can be expressed in different forms for the different parametrizations:

Non-singular parametrization:

$$
\begin{aligned}
& H_0 : \mu_1 = \mu_2 = \ldots = \mu_A \\
\Longleftrightarrow \quad & \\
& H_0 : L\beta = \begin{pmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = 0 \text{ with } L = \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times A}. \\
\Longleftrightarrow \quad & \\
& H_0 : Y_n = \mu + Z_n \text{ for all } n = 1, \ldots, N \\
\Longleftrightarrow \quad & \\
& H_0 : \mathrm{E}(Y) = \mu\, 1_N \\
\Longleftrightarrow \quad & \\
& H_0 : \mathrm{E}(Y) \in C(1_N).
\end{aligned}
$$

Control parametrization:

$$
\begin{aligned}
& H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_A = 0 \\
\iff & \\
& H_0 : L\beta = \begin{pmatrix} \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_A \end{pmatrix} = 0 \text{ with } L = \left(0_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times A}. \\
\iff & \\
& H_0 : Y_n = \mu + Z_n \text{ for all } n = 1, \ldots, N \\
\iff & \\
& H_0 : \mathrm{E}(Y) = \mu\, 1_N \\
\iff & \\
& H_0 : \mathrm{E}(Y) \in C(1_N).
\end{aligned}
$$

Singular parametrization:

$$
\begin{aligned}
& H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_A \\
\iff & \\
& H_0 : L\beta = \begin{pmatrix} \alpha_2 - \alpha_1 \\ \alpha_3 - \alpha_1 \\ \vdots \\ \alpha_A - \alpha_1 \end{pmatrix} = 0 \text{ with } L = \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \in \mathbb{R}^{(A-1)\times(A+1)}. \\
\iff & \\
& H_0 : Y_n = \mu + Z_n \text{ for all } n = 1, \ldots, N \\
\iff & \\
& H_0 : \mathrm{E}(Y) = \mu\, 1_N \\
\iff & \\
& H_0 : \mathrm{E}(Y) \in C(1_N).
\end{aligned}
$$

We see that the formulation of the null hypothesis via $\mathrm{E}(Y) \in C(1_N)$ does not depend on the parametrization. Moreover, we have $C(1_N) \subset C(X)$ for each parametrization.

The general testing problem in linear models
In the model $Y = X\beta + Z$ with $Z \sim \mathcal{N}_N(0_N, \sigma^2 I_{N\times N})$, the following testing problem is considered:

$$
\begin{aligned}
& H_0 : \mathrm{E}(Y) \in C(X_0) \quad \text{versus} \quad H_1 : \mathrm{E}(Y) \notin C(X_0), \\
\iff & \\
& H_0 : \mathrm{E}(Y) = X_0\gamma \text{ for some } \gamma \in \mathbb{R}^Q \quad \text{versus} \quad H_1 : \mathrm{E}(Y) \neq X_0\gamma \text{ for all } \gamma \in \mathbb{R}^Q,
\end{aligned}
$$

where $C(X_0) \subset C(X)$ and $X_0 \in \mathbb{R}^{N\times Q}$.

9.3.2 Theorem
Every likelihood ratio test for $H_0: E(Y) \in C(X_0)$ *versus* $H_1: E(Y) \notin C(X_0)$ *depends on* y *only via*

$$\frac{y^\top (P - P_0)\, y}{y^\top (I_{N\times N} - P)\, y},$$

where P *and* P_0 *are perpendicular projection matrices onto* X *and* X_0*, respectively.*

Proof. Note that $P = X(X^\top X)^- X^\top$ and $P_0 = X_0(X_0^\top X_0)^- X_0^\top$ according to Lemma 8.1.10. In the full model $Y = X\beta + Z$, we have according to Theorem 9.2.1 and its proof

$$\sup_{\beta\in\mathbb{R}^R, \sigma^2\in\mathbb{R}^+} f_{\beta,\sigma^2}(y) = \frac{1}{(2\pi\widehat{\sigma}^2)^{N/2}} e^{\frac{1}{2\widehat{\sigma}^2}(y-X\widehat{\beta})^\top (y-X\widehat{\beta})} = \frac{1}{(2\pi\widehat{\sigma}^2)^{N/2}} e^{\frac{N}{2}}$$

since $\widehat{\sigma}^2 = \frac{1}{N}(y - X\widehat{\beta})^\top (y - X\widehat{\beta}) = \frac{1}{N} y^\top (I_{N\times N} - P) y$. Analogously we obtain for $Y = \gamma\, X_0 + Z$

$$\sup_{\gamma\in\mathbb{R}^Q, \sigma^2\in\mathbb{R}^+} f^0_{\beta,\sigma^2}(y) = \frac{1}{(2\pi\widehat{\sigma}_0^2)^{N/2}} e^{\frac{N}{2}}$$

for $\widehat{\sigma}_0^2 = \frac{1}{N}(y - X_0\widehat{\gamma})^\top (y - X_0\widehat{\gamma}) = \frac{1}{N} y^\top (I_{N\times N} - P_0) y = \frac{1}{N} y^\top (I_{N\times N} - P + P - P_0) y$. This implies

$$\begin{aligned} &\frac{\sup_{\beta\in\mathbb{R}^R, \sigma^2\in\mathbb{R}^+} f_{\beta,\sigma^2}(y)}{\sup_{\gamma\in\mathbb{R}^Q, \sigma^2\in\mathbb{R}^+} f^0_{\beta,\sigma^2}(y)} = \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{\frac{N}{2}} \\ &= \left(\frac{y^\top (I_{N\times N} - P) y + y^\top (P - P_0) y}{y^\top (I_{N\times N} - P) y}\right)^{\frac{N}{2}} = \left(1 + \frac{y^\top (P - P_0) y}{y^\top (I_{N\times N} - P) y}\right)^{\frac{N}{2}}. \end{aligned}$$

□

9.3.3 Theorem
Let be $I = I_{N\times N}$.

a) $\frac{1}{\sigma^2}\, Y^\top (P - P_0) Y \sim \chi^2\left(rk(P - P_0),\ \frac{\beta^\top X^\top (P - P_0) X\beta}{\sigma^2}\right),$

b) $\frac{1}{\sigma^2}\, Y^\top (I - P) Y \sim \chi^2\left(rk(I - P), 0\right),$

c) $\frac{\frac{1}{rk(P-P_0)}\, Y^\top (P - P_0) Y}{\frac{1}{rk(I-P)}\, Y^\top (I - P) Y} \sim F\left(rk(P - P_0),\ rk(I - P),\ \frac{\beta^\top X^\top (P - P_0) X\beta}{\sigma^2}\right),$

d) *Under* H_0: $\frac{1}{\sigma^2}\, Y^\top (P - P_0) Y \sim \chi^2\left(rk(P - P_0),\ 0\right),$

e) *Under* H_0: $\frac{\frac{1}{rk(P-P_0)}\, Y^\top (P - P_0) Y}{\frac{1}{rk(I-P)}\, Y^\top (I - P) Y} \sim F\left(rk(P - P_0),\ rk(I - P),\ 0\right).$

Proof.
a) $C(X_0) \subset C(X)$ implies $P\,P_0 = P_0$ and with the symmetry $P_0\,P = P_0$ as well so that

$$(P - P_0)(P - P_0) = P\,P - P\,P_0 - P_0 P + P_0 P_0 = P - P_0 - P_0 + P_0 = P - P_0.$$

Hence also $P - P_0$ is symmetric and idempotent. Hence the assertion follows from $\frac{1}{\sigma}Y \sim \mathcal{N}_N(\frac{1}{\sigma}X\beta, I_{N\times N})$ and Theorem 8.3.11.
b) The assertion follows similarly as in a) since $I - P$ is also symmetric and idempotent and $(I - P)X = 0$.
c) $Y^\top(P - P_0)Y$ and $Y^\top(I - P)Y$ are stochastically independent according to Theorem 8.3.6 since

$$(I - P)I(P - P_0) = (I - P)(P - P_0) = P - P_0 - P\,P + P\,P_0 = P - P_0 - P + P_0 = 0.$$

Then the assertion follows from a) and b) and the definition of the F-distribution.
d) Under H_0 we have

$$\beta^\top X^\top (P - P_0)X\beta = \gamma^\top X_0^\top (P - P_0)X_0\gamma = \gamma^\top X_0^\top (X_0 - X_0)\gamma = 0.$$

e) follows from d) and c). □

9.3.4 Remark
$\frac{1}{\mathrm{rk}(I-P)}Y^\top(I-P)Y$ is the unbiased estimator for σ^2 of Theorem 9.2.4. Thereby we have $\mathrm{rk}(I-P) = N - \mathrm{rk}(X)$ according to Lemma 8.1.11.

9.3.5 Corollary
Let be $q(1-\alpha)$ the $1-\alpha$-quantile of the central F-distribution with $rk(P - P_0)$ and $rk(I - P)$ degrees of freedom, then

$$\mathbb{1}_{\{\widehat{V}(y) > q(1-\alpha)\}}(y) \quad \textit{with} \quad \widehat{V}(y) = \frac{\frac{1}{rk(P-P_0)}\, y^\top(P - P_0)y}{\frac{1}{rk(I-P)}\, y^\top(I - P)y}$$

is α-level test for $H_0:\ E(Y) \in C(X_0)$ versus $H_1:\ E(Y) \notin C(X_0)$. Its β-error is given by

$$F_{F\left(rk(P-P_0),\, rk(I-P),\, \frac{\beta^\top X^\top (P-P_0)X\beta}{\sigma^2}\right)}(q(1-\alpha))\,.$$

Sometimes it is easier to formulate hypotheses in form of

$$H_0:\ L\beta = l \quad \text{versus} \quad H_1:\ L\beta \neq l$$

with $L \in \mathbb{R}^{S\times R}$, $l = Lb \in \mathbb{R}^S$ for some $b \in \mathbb{R}^R$. Thereby, $\lambda(\beta) = L\beta$ shall be identifiable at X. To

express this null hypothesis in the above form note that we have for the model under H_0:

$$
\begin{aligned}
& Y = X\beta + Z, \;\; L\,\beta = L\,b \\
\overset{L=KX}{\Longleftrightarrow} \quad & \\
& \mathrm{E}(Y) = X\beta, \;\; KX(\beta - b) = 0 \\
\overset{PX=X}{\Longleftrightarrow} \quad & \\
& \mathrm{E}(Y - Xb) = X(\beta - b), \;\; KPX(\beta - b) = 0 \\
\Longleftrightarrow \quad & \\
& \mathrm{E}(Y - Xb) \in C(X), \;\; \mathrm{E}(Y - Xb) \perp C(P^\top K^\top) \\
\Longleftrightarrow \quad & \\
& \mathrm{E}(Y - Xb) \in C(X) \cap C(P\,K^\top)^\perp.
\end{aligned}
$$

9.3.6 Lemma
$X_0 = P - P_K$ with $P_K = P\,K^\top(K\,P\,K^\top)^- K\,P$ satisfies

$$C(P - P_K) = C(X_0) = C(X) \cap C(P\,K^\top)^\perp$$

and $P - P_K$ is perpendicular projection matrix onto $C(P - P_K)$.

Proof. Let be $x \in C(X) \cap C(P\,K^\top)^\perp$. Then $x = X\beta$ for some $\beta \in \mathbb{R}^R$ and $0 = KP^\top x = KPx$ so that $P_K x = 0$ and

$$(P - P_K)X\beta = PX\beta - P_K x = X\beta - 0 = x,$$

which means $x \in C(P - P_K)$.
Conversely, let be $x \in C(P - P_K)$. Then $x = (P - P_K)v$ for some $v \in \mathbb{R}^N$ and with this

$$
\begin{aligned}
x = (P - P_K)v &= (X(X^\top X)^- X^\top - X(X^\top X)^- X^\top K^\top (K\,P\,K^\top)^- K\,P)v \\
&= X\left((X^\top X)^- X^\top - (X^\top X)^- X^\top K^\top (K\,P\,K^\top)^- K\,P\right)v,
\end{aligned}
$$

which implies $x \in C(X)$. Moreover,

$$
\begin{aligned}
K\,P^\top x &= K\,P^\top (P - P_K)v \\
&= K\,P^\top P\,v - K\,P^\top P\,K^\top (K\,P\,K^\top)^- K\,Pv \\
&= K\,P^\top v - \underbrace{K\,P^\top P\,K^\top (K\,P^\top P\,K^\top)^- K\,P^\top}_{= KP^\top \text{(Lem. 8.1.5 b))}} v \\
&= K\,P^\top v - K\,P^\top v = 0,
\end{aligned}
$$

so that $x \in C(P\,K^\top)^\perp$.
Furthermore, $P_K = P\,K^\top(K\,P^\top P\,K^\top)^- K\,P^\top$ is perpendicular projection matrix onto $P\,K^\top$ according to Lemma 8.1.10, so that $P_K P_K = P_K$. Since $P\,P_K = P_K$ and $P_K P = P_K$ because of

symmetry, we have

$$(P - P_K)(P - P_K) = P\,P - P_K P - P\,P_K + P_K P_K = P - P_K - P_K + P_K = P_K$$

so that $P - P_K$ is perpendicular projection matrix onto $C(P - P_K)$. □

9.3.7 Lemma
Let be $P_0 = P - P_K$ *with* $P_K = P\,K^\top (K\,P\,K^\top)^- K\,P$, $L = KX$, $l = L\,b$, *and* $\widehat{\beta} = (X^\top X)^- X^\top y$. *Then*

$$\begin{aligned}
&a) \quad (y - X\,b)^\top (P - P_0)(y - X\,b) = (L\widehat{\beta} - l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta} - l),\\
&b) \quad (\beta - b)^\top X^\top (P - P_0) X (\beta - b) = (L\beta - l)^\top [L(X^\top X)^- L^\top]^- (L\beta - l),\\
&c) \quad rk(P - P_0) = rk(P_K) = rk(L).
\end{aligned}$$

Proof.
a)

$$\begin{aligned}
&(y - X\,b)^\top (P - P_0)(y - X\,b) = (y - X\,b)^\top P_K (y - X\,b)\\
&= (y - X\,b)^\top P\,K^\top (K\,P\,K^\top)^- K\,P\,(y - X\,b)\\
&= (K\,P\,y - K\,P\,X\,b)^\top (K\,P\,K^\top)^- (K\,P\,y - K\,P\,X\,b)\\
&= (K\,X(X^\top X)^- X^\top\,y - K\,X\,b)^\top (K\,X(X^\top X)^- X^\top\,K^\top)^- (K\,X(X^\top X)^- X^\top\,y - K\,X\,b)\\
&= (L\widehat{\beta} - l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta} - l).
\end{aligned}$$

The assertion b) follows analogously.
c) Moreover,

$$\mathrm{rk}(P_K) \le \mathrm{rk}(K\,P) = \mathrm{rk}(K\,X(X^\top X)^- X^\top) \le \mathrm{rk}(KX) = \mathrm{rk}(L)$$

and

$$\begin{aligned}
&\mathrm{rk}(L) = \mathrm{rk}(KX) = \mathrm{rk}(KX(X^\top X)^- X^\top X) \le \mathrm{rk}(KX(X^\top X)^- X^\top)\\
&\overset{\mathrm{rk}(A)=\mathrm{rk}(A\,A^\top)}{=} \mathrm{rk}(KX(X^\top X)^- X^\top X\,(X^\top X)^- X^\top K^\top) = \mathrm{rk}(KX(X^\top X)^- X^\top K^\top)\\
&= \mathrm{rk}(K\,P\,K^\top) = \mathrm{rk}(K\,P\,K^\top (K\,P\,K^\top)^- K\,P\,K^\top)\\
&\le \mathrm{rk}(P\,K^\top (K\,P\,K^\top)^- K\,P) = \mathrm{rk}(P_K).
\end{aligned}$$ □

9.3.8 Theorem

$$\begin{aligned}
&a) \quad \frac{1}{\sigma^2}\,(L\widehat{\beta} - l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta} - l) \sim \chi^2\left(rk(L), \frac{(L\beta - l)^\top [L(X^\top X)^- L^\top]^- (L\beta - l)}{\sigma^2}\right),\\
&b) \quad \textit{Under } H_0 : L\beta = l : \frac{1}{\sigma^2}\,(L\widehat{\beta} - l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta} - l) \sim \chi^2\,(rk(L), 0)\,.
\end{aligned}$$

Proof. Since $\frac{1}{\sigma}(Y - X\,b) \sim \mathcal{N}_N(\frac{1}{\sigma}X(\beta - b), I_{N\times N})$, the assertion a) follows from Theorem 9.3.3 and Lemma 9.3.7. The assertion b) is a immediate consequence of a). □

9.3.9 Corollary
Let be $q(1-\alpha)$ the $1-\alpha$-quantile of the central F-distribution with $rk(L)$ and $rk(I-P)$ degrees of freedom, $L = KX$, and $l = L\,b$, then

$$\mathbb{1}_{\{\widehat{V}(y)>q(1-\alpha)\}}(y) \quad \textit{with} \quad \widehat{V}(y) = \frac{\frac{1}{rk(L)}\,(L\widehat{\beta} - l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta} - l)}{\frac{1}{rk(I-P)}\,y^\top (I-P)y}$$

is α-level test for $H_0:\ L\beta = l$ versus $H_1:\ L\beta \neq l$. Its β-error is given by

$$F_{F\left(rk(L),\, rk(I-P),\, \frac{(L\beta-l)^\top [L(X^\top X)^- L^\top]^- (L\beta-l)}{\sigma^2}\right)}\,(q(1-\alpha))\,.$$

9.3.10 Remark (Designing experiments)
The aim of a good design is to minimize the β-error of the test. This means here that the non-centrality parameter

$$(L\widehat{\beta} - l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta} - l)$$

should be as large as possible for all β with $L\beta \neq l$. This is achieved if

$$L(X^\top X)^- L^\top$$

is as small as possible. Since $L(X^\top X)^- L^\top$ is the covariance matrix of the estimator $L(X^\top X)^- X^\top y$ (see Lemma 9.2.5), we realize that testing and estimation leads to the optimization problem for designing experiments.

10 ANOVA models

10.1 Further lemmas from linear algebra

10.1.1 Lemma
For $A \in \mathbb{N}$ and $P_A := \left(I_{A\times A} - \frac{1}{A} 1_A 1_A^\top\right)$ it holds:

$$a) \quad \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix}^- = \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix},$$

$$b) \quad P_A^- = P_A.$$

Proof.
a)

$$\begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix}$$
$$= \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} 0 & 0_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} = \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix}.$$

b) Since P_A is perpendicular projection matrix onto $C(1_A)^\perp$ (see Lemma 8.1.10 b)) we have $P_A P_A^- P_A = P_A P_A P_A = P_A$. □

10.1.2 Lemma
If $A \in \mathbb{R}^{A\times A}$ is regular and symmetric, $b \in \mathbb{R}^A$ with $1 + b^\top A^{-1} b \neq 0$, then

$$(A + b\,b^\top)^{-1} = A^{-1} - \frac{A^{-1} b\, b^\top A^{-1}}{1 + b^\top A^{-1} b}.$$

Proof.

$$\begin{aligned} & (A + b\,b^\top) \left(A^{-1} - \frac{A^{-1} b\, b^\top A^{-1}}{1 + b^\top A^{-1} b}\right) \\ &= I_{A\times A} + b\,b^\top A^{-1} - \frac{b\,b^\top A^{-1}}{1 + b^\top A^{-1} b} - \frac{b\,(b^\top A^{-1} b)\, b^\top A^{-1}}{1 + b^\top A^{-1} b} \\ &= I_{A\times A} + b\,b^\top A^{-1} - \frac{b\,b^\top A^{-1}(1 + b^\top A^{-1} b)}{1 + b^\top A^{-1} b} \\ &= I_{A\times A}. \end{aligned}$$

□

10.1.3 Lemma
For $A, B \in \mathbb{N}$ it holds

$$\begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}^- = \begin{pmatrix} 0 & 0_A^\top & 0_B^\top \\ 0_A & \frac{1}{B}\,I_{A\times A} & 0_{A\times B} \\ 0_B & 0_{B\times A} & \frac{1}{A}\,P_B \end{pmatrix},$$

where $P_B = (I_{B\times B} - \frac{1}{B}\,1_B\,1_B^\top)$.

Proof. Since P_B is perpendicular projection matrix onto $C(1_B)^\perp$ (see Lemma 8.1.10 b)), we have $P_B 1_B = 0_B$ and $1_B^\top P_B = 0_B^\top$. This implies

$$\begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix} \begin{pmatrix} 0 & 0_A^\top & 0_B^\top \\ 0_A & \frac{1}{B}\,I_{A\times A} & 0_{A\times B} \\ 0_B & 0_{B\times A} & \frac{1}{A}\,P_B \end{pmatrix} \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1_A^\top & 0_B^\top \\ 0_A & I_{A\times A} & 0_{A\times B} \\ 0_B & \frac{1}{B} 1_B 1_A^\top & P_B \end{pmatrix} \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}$$

$$= \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & \underbrace{\frac{A}{B}\,1_B 1_B^\top + A\,P_B}_{=A\,I_{B\times B}} \end{pmatrix}.$$

□

10.2 Balanced one-way ANOVA

In the balanced one-way layout, there is only one factor A with A levels and each level is observed M times so that $N = M\,A$ is the total sample size. We will use here the singular parametrization:

$$Y_{an} = \mu + \alpha_a + Z_{an}, \quad \text{for } n = 1, \ldots, M.$$

Then (see Example 9.1.1)

$$X = (1_{AM} \mid I_{A\times A} \otimes 1_M) = (1_A \mid I_{A\times A}) \otimes 1_M \in \mathbb{R}^{AM\times(A+1)}.$$

The null hypothesis

$$\begin{aligned} & H_0: \text{ Factor A has no influence} \\ \Longleftrightarrow & \\ & H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_A \\ \Longleftrightarrow & \\ & H_0: L\beta = 0 \text{ with } L = \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \end{aligned}$$

shall be tested via the test statistic

$$\frac{\frac{1}{\mathrm{rk}(L)}\,(L\widehat{\beta}-l)^\top [L(X^\top X)^- L^\top]^- (L\widehat{\beta}-l)}{\frac{1}{N-\mathrm{rk}(X)}\,y^\top (I-P)y}.$$

We have

$$X^\top X = \begin{pmatrix} 1_A^\top \\ I_{A\times A} \end{pmatrix} \otimes 1_M^\top \;\cdot\; (1_A \mid I_{A\times A}) \otimes 1_M = \begin{pmatrix} A & 1_A^\top \\ 1_A & I_{A\times A} \end{pmatrix} \otimes M$$

so that according to Lemma 10.1.1

$$(X^\top X)^- = \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix} \otimes \frac{1}{M} = \frac{1}{M} \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix}.$$

This implies

$$\begin{aligned} L\,(X^\top X)^- L^\top &= \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \frac{1}{M} \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} 0_{A-1}^\top \\ -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \\ &= \frac{1}{M} \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \begin{pmatrix} 0_{A-1}^\top \\ -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} = \frac{1}{M} \left(1_{A-1}1_{A-1}^\top + I_{(A-1)\times(A-1)}\right). \end{aligned}$$

Lemma 10.1.2 provides

$$\begin{aligned} \left(L\,(X^\top X)^- L^\top\right)^{-1} &= M \left(I_{(A-1)\times(A-1)} - \frac{1_{A-1}1_{A-1}^\top}{1 + 1_{A-1}^\top 1_{A-1}}\right) \\ &= M \left(I_{(A-1)\times(A-1)} - \frac{1_{A-1}1_{A-1}^\top}{1+A-1}\right) = M \left(I_{(A-1)\times(A-1)} - \frac{1}{A}\, 1_{A-1}1_{A-1}^\top\right). \end{aligned}$$

Moreover, we have

$$\begin{aligned} L\,(X^\top X)^- X^\top &= \frac{1}{M} \left(0_{A-1} \mid -1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \begin{pmatrix} 1_A^\top \\ I_{A\times A} \end{pmatrix} \otimes 1_M^\top \\ &= \frac{1}{M} \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \otimes 1_M^\top \end{aligned}$$

so that

$$\begin{aligned}
& X\,(X^\top X)^- L^\top \left(L\,(X^\top X)^- L^\top\right)^{-1} L\,(X^\top X)^- X^\top \\
&= \begin{pmatrix} -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \left(I_{(A-1)\times(A-1)} - \frac{1}{A}\,1_{A-1}1_{A-1}^\top\right) \left(-1_{A-1} \mid I_{(A-1)\times(A-1)}\right) \otimes \frac{1}{M}\,1_M\,1_M^\top \\
&= \begin{pmatrix} -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \left(-1_{A-1} + \frac{A-1}{A}1_{A-1} \mid I_{(A-1)\times(A-1)} - \frac{1}{A}\,1_{A-1}1_{A-1}^\top\right) \otimes \frac{1}{M}\,1_M\,1_M^\top \\
&= \begin{pmatrix} -1_{A-1}^\top \\ I_{(A-1)\times(A-1)} \end{pmatrix} \left(-\frac{1}{A}1_{A-1} \mid I_{(A-1)\times(A-1)} - \frac{1}{A}\,1_{A-1}1_{A-1}^\top\right) \otimes \frac{1}{M}\,1_M\,1_M^\top \\
&= \begin{pmatrix} \frac{A-1}{A} & -1_{A-1}^\top + \frac{A-1}{A}1_{A-1}^\top \\ -\frac{1}{A}1_{A-1} & I_{(A-1)\times(A-1)} - \frac{1}{A}\,1_{A-1}1_{A-1}^\top \end{pmatrix} \otimes \frac{1}{M}\,1_M\,1_M^\top \\
&= \left(I_{A\times A} - \frac{1}{A}\,1_A 1_A^\top\right) \otimes \frac{1}{M}\,1_M\,1_M^\top \\
&= P_A \otimes \frac{1}{M}\,1_M\,1_M^\top.
\end{aligned}$$

Hence, we obtain for $y = (y_{11}, \ldots, y_{1M}, \ldots, y_{A1}, \ldots, y_{AM})^\top = (y_1, \ldots, y_N)^\top$

$$\begin{aligned}
& \widehat{\beta}^\top L^\top \left(L\,(X^\top X)^- L^\top\right)^{-1} L\,\widehat{\beta} = y^\top \left(P_A \otimes \frac{1}{M}\,1_M\,1_M^\top\right) y \\
&= M\,y^\top \left(P_A \otimes \frac{1}{M}\,1_M\right) \left(P_A \otimes \frac{1}{M}\,1_M^\top\right) y = M \sum_{a=1}^{A} (\overline{y}_{a\bullet} - \overline{y}_{\bullet\bullet})^2
\end{aligned}$$

with $\overline{y}_{a\bullet} = \frac{1}{M}\sum_{m=1}^{M} y_{am}$ and $\overline{y}_{\bullet\bullet} = \frac{1}{AM}\sum_{a=1}^{A}\sum_{m=1}^{M} y_{am}$. This provides the numerator of the test statistic. For the denominator of the test statistic, we calculate

$$\begin{aligned}
P = X\,(X^\top X)^- X^\top &= ((1_A \mid I_{A\times A}) \otimes 1_M) \left(\begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix} \otimes \frac{1}{M}\right) \left(\begin{pmatrix} 1_A^\top \\ I_{A\times A} \end{pmatrix} \otimes 1_M^\top\right) \\
&= (1_A \mid I_{A\times A}) \begin{pmatrix} 0 & 0_A^\top \\ 0_A & I_{A\times A} \end{pmatrix} \begin{pmatrix} 1_A^\top \\ I_{A\times A} \end{pmatrix} \otimes \frac{1}{M}\,1_M\,1_M^\top = I_{A\times A} \otimes \frac{1}{M}\,1_M\,1_M^\top,
\end{aligned}$$

so that

$$\begin{aligned}
I - P &= I_{A\times A} \otimes I_{M\times M} - I_{A\times A} \otimes \frac{1}{M}\,1_M\,1_M^\top \\
&= I_{A\times A} \otimes \left(I_{M\times M} - \frac{1}{M}\,1_M\,1_M^\top\right) = I_{A\times A} \otimes P_M
\end{aligned}$$

and

$$y^\top (I - P)\,y = \sum_{a=1}^{A}\sum_{m=1}^{M} (y_{am} - \overline{y}_{a\bullet})^2.$$

Since $\mathrm{rk}(I_{A\times A} \otimes P_M) = A(M-1)$ and $\mathrm{rk}(P_A \otimes \frac{1}{M} 1_M 1_M^\top) = A-1$ we obtain the ANOVA table:

Factor	Matrix	Sum of squares	Rank
A	$P_A \otimes \frac{1}{M} 1_M 1_M^\top$	$M \sum_{a=1}^{A} (\overline{y}_{a\bullet} - \overline{y}_{\bullet\bullet})^2$	$A-1$
Error	$I_{A\times A} \otimes P_M$	$\sum_{a=1}^{A} \sum_{m=1}^{M} (y_{am} - \overline{y}_{a\bullet})^2$	$A(M-1)$

Since the Grand Sum of Squares can be expressed as

$$\sum_{a=1}^{A} \sum_{m=1}^{M} (y_{am} - \overline{y}_{\bullet\bullet})^2 = y^\top \left(I_{A\times A} \otimes I_{M\times M} - \left(\frac{1}{A} 1_A 1_A^\top \right) \otimes \left(\frac{1}{M} 1_M 1_M^\top \right) \right) y$$

and

$$\begin{aligned} & I_{A\times A} \otimes P_M + P_A \otimes \left(\frac{1}{M} 1_M 1_M^\top \right) \\ &= I_{A\times A} \otimes I_{M\times M} - I_{A\times A} \otimes \left(\frac{1}{M} 1_M 1_M^\top \right) + I_{A\times A} \otimes \left(\frac{1}{M} 1_M 1_M^\top \right) - \left(\frac{1}{A} 1_A 1_A^\top \right) \otimes \left(\frac{1}{M} 1_M 1_M^\top \right) \\ &= I_{A\times A} \otimes I_{M\times M} - \left(\frac{1}{A} 1_A 1_A^\top \right) \otimes \left(\frac{1}{M} 1_M 1_M^\top \right), \end{aligned}$$

we see the decomposition of the Grand Sum of Squares also via the matrices.

The same result is obtained if the other parametrizations are used. In particular, the derivation via the non-singular parametrization is easy (Exercise!)

10.3 Balanced two-way ANOVA

In the two-way layout, we have two factors, factor A with A levels and factor B with B factors. The observations are here:

$y_{11*} = (y_{111}, \ldots, y_{11N_{11}})^\top$ the vector of observations for level combination $(1,1)$,
$y_{12*} = (y_{121}, \ldots, y_{12N_{12}})^\top$ the vector of observations for level combination $(1,2)$,
$\vdots$
$y_{1B*} = (y_{1B1}, \ldots, y_{1BN_{1B}})^\top$ the vector of observations for level combination $(1,B)$,
$y_{21*} = (y_{211}, \ldots, y_{21N_{21}})^\top$ the vector of observations for level combination $(2,1)$,
$\vdots$
$y_{2B*} = (y_{2B1}, \ldots, y_{2BN_{2B}})^\top$ the vector of observations for level combination $(2,B)$,
$\vdots$
$y_{A1*} = (y_{A11}, \ldots, y_{A1N_{A1}})^\top$ the vector of observations for level combination $(A,1)$,
$\vdots$
$y_{AB*} = (y_{AB1}, \ldots, y_{ABN_{AB}})^\top$ the vector of observations for level combination (A,B).

If the design is balanced, then each level combination a and b is observed M times so that $N_{ab} = M$ for all $a = 1, \ldots, A$, $b = 1, \ldots, B$ and the total sample size is $N = M\,A\,B$. Here we will only consider balanced designs.

Non-singular parametrization

$$\mathrm{E}(Y_{abm}) = \mu_{ab} \text{ for all } a = 1, \ldots, A,\ b = 1, \ldots, B,\ m = 1, \ldots, M$$

and

$$\begin{aligned}
&\beta = (\mu_{11}, \ldots, \mu_{1B}, \mu_{21}, \ldots, \mu_{2B}, \ldots, \mu_{A1}, \ldots, \mu_{AB})^\top \in \mathbb{R}^{A \cdot B},\\
&X = I_{A\times A} \otimes I_{B\times B} \otimes 1_M.
\end{aligned}$$

Singular parametrization

$$\mathrm{E}(Y_{abm}) = \mu + \alpha_a + \beta_b + \gamma_{ab} \text{ for all } a = 1, \ldots, A,\ b = 1, \ldots, B,\ m = 1, \ldots, M$$

and

$$\begin{aligned}
&\beta = (\mu, \alpha_1, \ldots, \alpha_A, \beta_1, \ldots, \beta_B, \gamma_{11}, \ldots, \gamma_{1B}, \ldots, \gamma_{A1}, \ldots, \gamma_{AB})^\top \in \mathbb{R}^{1+A+B+A\cdot B},\\
&X = (1_A \otimes 1_B \otimes 1_M \mid I_{A\times A} \otimes 1_B \otimes 1_M \mid 1_A \otimes I_{B\times B} \otimes 1_M \mid I_{A\times A} \otimes I_{B\times B} \otimes 1_M)\,.
\end{aligned}$$

The following hypotheses are considered:

$$\begin{aligned}
&H_0^I: \text{ There is no interaction between factor A and factor B}\\
\Longleftrightarrow\\
&H_0^I: \ \gamma_{11} = \ldots = \gamma_{1B} = \ldots = \gamma_{A1} = \ldots = \gamma_{AB} = 0,\\
&H_0^A: \text{ Factor A has no effect}\\
\Longleftrightarrow\\
&H_0^A: \ \alpha_1 = \ldots = \alpha_A = 0,\\
&H_0^B: \text{ Factor B has no effect}\\
\Longleftrightarrow\\
&H_0^B: \ \beta_1 = \ldots = \beta_B = 0.
\end{aligned}$$

The advantage of the singular parametrization is that the hypotheses can be easily formulated.

We start with the hypothesis H_0^I. We use the design matrix of the non-singular parametrization, i.e. $X = I_{A\times A} \otimes I_{B\times B} \otimes 1_M$ for the full model and express the reduced model of H_0^I via the design matrix

$$\begin{aligned}
X_I &= (1_A \otimes 1_B \otimes 1_M \mid I_{A\times A} \otimes 1_B \otimes 1_M \mid 1_A \otimes I_{B\times B} \otimes 1_M)\\
&= \widetilde{X}_I \otimes 1_M
\end{aligned}$$

with

$$\widetilde{X}_I := (1_A \otimes 1_B \mid I_{A\times A} \otimes 1_B \mid 1_A \otimes I_{B\times B})\,.$$

Then we test in the model $\mathrm{E}(Y) = X\beta$

$$H_0^I : \mathrm{E}(Y) \in C(X_I) \quad \text{versus} \quad H_1^I : \mathrm{E}(Y) \notin C(X_I).$$

To calculate the test statistic, note that

$$\begin{aligned}\widetilde{X}_I^\top \widetilde{X}_I &= \begin{pmatrix} 1_A^\top \otimes 1_B^\top \\ I_{A\times A} \otimes 1_B^\top \\ 1_A^\top \otimes I_{B\times B} \end{pmatrix} (1_A \otimes 1_B \mid I_{A\times A} \otimes 1_B \mid 1_A \otimes I_{B\times B}) \\ &= \begin{pmatrix} A\otimes B & 1_A^\top \otimes B & A \otimes 1_B^\top \\ 1_A \otimes B & I_{A\times A}\otimes B & 1_A \otimes 1_B^\top \\ A\otimes 1_B & 1_A^\top \otimes 1_B & A \otimes I_{B\times B} \end{pmatrix} = \begin{pmatrix} A\,B & B\,1_A^\top & A\,1_B^\top \\ B\,1_A & B\,I_{A\times A} & 1_A 1_B^\top \\ A\,1_B & 1_B 1_A^\top & A\,I_{B\times B} \end{pmatrix}\end{aligned}$$

Lemma 10.1.3 provides

$$\left(\widetilde{X}_I^\top \widetilde{X}_I\right)^- = \begin{pmatrix} 0 & 0_A^\top & 0_B^\top \\ 0_A & \frac{1}{B} I_{A\times A} & 0_{A\times B} \\ 0_B & 0_{B\times A} & \frac{1}{A} P_B \end{pmatrix} = \begin{pmatrix} 0\otimes 0 & 0_A^\top \otimes 0 & 0 \otimes 0_B^\top \\ 0_A \otimes 0 & I_{A\times A}\otimes \frac{1}{B} & 0_A \otimes 1_B^\top \\ 0\otimes 0_B & 0_A^\top \otimes 0_B & \frac{1}{A} \otimes P_B \end{pmatrix}$$

so that

$$\begin{aligned}\widetilde{P}_I &:= \widetilde{X}_I \left(\widetilde{X}_I^\top \widetilde{X}_I\right)^- \widetilde{X}_I^\top \\ &= \widetilde{X}_I \begin{pmatrix} 0\otimes 0 & 0_A^\top \otimes 0 & 0 \otimes 0_B^\top \\ 0_A \otimes 0 & I_{A\times A}\otimes \frac{1}{B} & 0_A \otimes 1_B^\top \\ 0\otimes 0_B & 0_A^\top \otimes 0_B & \frac{1}{A} \otimes P_B \end{pmatrix} \begin{pmatrix} 1_A^\top \otimes 1_B^\top \\ I_{A\times A} \otimes 1_B^\top \\ 1_A^\top \otimes I_{B\times B} \end{pmatrix} \\ &= (1_A \otimes 1_B \mid I_{A\times A} \otimes 1_B \mid 1_A \otimes I_{B\times B}) \begin{pmatrix} 0_A^\top \otimes 0_B^\top \\ I_{A\times A} \otimes \frac{1}{B} 1_B^\top \\ \frac{1}{A} 1_A^\top \otimes P_B \end{pmatrix} \\ &= I_{A\times A} \otimes \frac{1}{B} 1_B\, 1_B^\top + \frac{1}{A} 1_A\, 1_A^\top \otimes P_B.\end{aligned}$$

For $\widetilde{X} := I_{A\times A} \otimes I_{B\times B}$, it holds

$$\widetilde{P} := \widetilde{X}(\widetilde{X}^\top \widetilde{X})^- \widetilde{X}^\top = I_{A\times A} \otimes I_{B\times B}$$

so that

$$\begin{aligned}\widetilde{P} - \widetilde{P}_I &= I_{A\times A} \otimes I_{B\times B} - I_{A\times A} \otimes \frac{1}{B} 1_B 1_B^\top - \frac{1}{A} 1_A 1_A^\top \otimes P_B \\ &= I_{A\times A} \otimes P_B - \frac{1}{A} 1_A 1_A^\top \otimes P_B = P_A \otimes P_B.\end{aligned}$$

Since any $X_0 = \widetilde{X}_0 \otimes 1_M$ satisfies

$$\begin{aligned}X_0(X_0^\top X_0)^- X_0^\top &= \widetilde{X}_0 \otimes 1_M (\widetilde{X}_0^\top \widetilde{X}_0 \otimes M)^- \widetilde{X}_0^\top \otimes 1_M^\top \\ &= \widetilde{X}_0 \otimes 1_M \left((\widetilde{X}_0^\top \widetilde{X}_0)^- \otimes \frac{1}{M} \right) \widetilde{X}_0^\top \otimes 1_M^\top = \widetilde{X}_0 (\widetilde{X}_0^\top \widetilde{X}_0)^- \widetilde{X}_0 \otimes \frac{1}{M} 1_M 1_M^\top,\end{aligned}$$

we obtain for $P = X(X^\top X)^- X^\top$ and $P_I = X_I (X_I^\top X_I)^- X_I^\top$

$$P - P_I = \left(\widetilde{P} - \widetilde{P}_I \right) \otimes \frac{1}{M} 1_M 1_M^\top = P_A \otimes P_B \otimes \frac{1}{M} 1_M 1_M^\top.$$

Moreover

$$\begin{aligned}I_{A\,B\,M\times A\,B\,M} - P &= I_{A\times A} \otimes I_{B\times B} \otimes I_{M\times M} - I_{A\times A} \otimes I_{B\times B} \otimes \frac{1}{M} 1_M 1_M^\top \\ &= I_{A\times A} \otimes I_{B\times B} \otimes P_M\end{aligned}$$

and

$$\begin{aligned}&\mathrm{rk}(I_{A\times A} \otimes I_{B\times B} \otimes P_M) = A\,B\,(M-1), \\ &\mathrm{rk}(P_A \otimes P_B \otimes \frac{1}{M} 1_M 1_M^\top) = (A-1)(B-1)\end{aligned}$$

so that

$$\frac{\frac{1}{(A-1)(B-1)} y^\top \left(P_A \otimes P_B \otimes \frac{1}{M} 1_M 1_M^\top \right) y}{\frac{1}{A\,B\,(M-1)} y^\top (I_{A\times A} \otimes I_{B\times B} \otimes P_M)\, y}$$

is the test statistic for testing $H_0^I : \mathrm{E}(Y) \in C(X_I)$ versus $H_1^I : \mathrm{E}(Y) \notin C(X_I)$. It has a central F-distribution with $(A-1)(B-1)$ and $A\,B\,(M-1)$ degrees of freedom under H_0^I.

In next step, we derive the numerator of the test statistic for testing

$$H_0^B : \text{ Factor B has no effect } \iff H_0^B : E(Y) \in C(X_B),$$

where

$$X_B = (1_A \otimes 1_B \otimes 1_M \mid I_{A\times A} \otimes 1_B \otimes 1_M) = (1_A \mid I_{A\times A}) \otimes 1_{B\,M}.$$

The full model is the model without interactions, i.e. we test H_0^B in the model $\mathrm{E}(Y) = X_I\beta_I$. From the results for the one-way layout with the singular parametrization, we obtain

$$X_B(X_B^\top X_B)^- X_B^\top = I_{A\times A} \otimes \frac{1}{B\,M}\, 1_{B\,M}\, 1_{B\,M}^\top = I_{A\times A} \otimes \frac{1}{B}\, 1_B\, 1_B^\top \otimes \frac{1}{M}\, 1_M\, 1_M^\top$$

so that

$$\begin{aligned} P_I - P_B &= X_I(X_I^\top X_I)^- X_I^\top - X_B(X_B^\top X_B)^- X_B^\top \\ &= \left(I_{A\times A} \otimes \frac{1}{B}\, 1_B\, 1_B^\top + \frac{1}{A}\, 1_A\, 1_A^\top \otimes P_B - I_{A\times A} \otimes \frac{1}{B}\, 1_B\, 1_B^\top \right) \otimes \frac{1}{M}\, 1_M\, 1_M^\top \\ &= \frac{1}{A}\, 1_A\, 1_A^\top \otimes P_B \otimes \frac{1}{M}\, 1_M\, 1_M^\top. \end{aligned}$$

For testing

$$H_0^A: \text{ Factor A has no effect } \iff H_0^A:\ E(Y) \in C(X_A)$$

in the model $\mathrm{E}(Y) = X_I\beta_I$, where

$$X_A = (1_A \otimes 1_B \otimes 1_M \mid 1_A \otimes I_{B\times B} \otimes 1_M)\,,$$

we obtain analogously

$$\begin{aligned} P_I - P_A &= X_I(X_I^\top X_I)^- X_I^\top - X_A(X_A^\top X_A)^- X_A^\top \\ &= P_A \otimes \frac{1}{B}\, 1_B\, 1_B^\top \otimes \frac{1}{M}\, 1_M\, 1_M^\top. \end{aligned}$$

However, we obtain this projection matrix also for testing

$$H_0^A: \text{ Factor A has no effect } \iff H_0^A:\ E(Y) \in C(1_N)$$

in the model $\mathrm{E}(Y) = X_B\beta_B$ since this is the test problem of the one-way layout. Similarly,

$$\frac{1}{A}\, 1_A\, 1_A^\top \otimes P_B \otimes \frac{1}{M}\, 1_M\, 1_M^\top$$

is the projection matrix for testing

$$H_0^B: \text{ Factor B has no effect } \iff H_0^B:\ E(Y) \in C(1_N)$$

in the model $\mathrm{E}(Y) = X_A\beta_A$. Hence, it does not matter in which order we regard the submodels. Setting $J_L = \frac{1}{L}\, 1_L\, 1_L^\top$ for any $L \in \mathbb{N}$, the ANOVA table has now the form

Factor	Matrix	Sum of squares	Rank
B	$J_A \otimes P_B \otimes J_M$	$M\,A\,\sum_{b=1}^{B}(\overline{y}_{\bullet b\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$B-1$
A	$P_A \otimes J_B \otimes J_M$	$M\,B\,\sum_{a=1}^{A}(\overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$A-1$
AB	$P_A \otimes P_B \otimes J_M$	$M\sum_{a=1}^{A}\sum_{b=1}^{B}(\overline{y}_{ab\bullet} - \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet b\bullet} + \overline{y}_{\bullet\bullet\bullet})^2$	$(A-1)(B-1)$
Error	$I_{A\times A} \otimes I_{B\times B} \otimes P_M$	$\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{m=1}^{M}(y_{abm} - \overline{y}_{ab\bullet})^2$	$AB(M-1)$

Again, we obtain for the Grand Sum of Squares with $P_L = I_{L\times L} - J_L$

$$\begin{aligned}
&\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{m=1}^{M}(y_{abm} - \overline{y}_{\bullet\bullet\bullet})^2 = y^\top\,(I_{A\times A}\otimes I_{B\times B}\otimes I_{M\times M} - J_A\otimes J_B\otimes J_M)\,y\\
&= y^\top\,(J_A\otimes I_{B\times B}\otimes J_M - J_A\otimes J_B\otimes J_M\\
&\quad + I_{A\times A}\otimes J_B\otimes J_M - J_A\otimes J_B\otimes J_M\\
&\quad + I_{A\times A}\otimes I_{B\times B}\otimes J_M - I_{A\times A}\otimes J_B\otimes J_M - J_A\otimes I_{B\times B}\otimes J_M + J_A\otimes J_B\otimes J_M\\
&\quad + I_{A\times A}\otimes I_{B\times B}\otimes I_{M\times M} - I_{A\times A}\otimes I_{B\times B}\otimes J_M)\,y\\
&= y^\top\,(J_A\otimes P_B\otimes J_M + P_A\otimes J_B\otimes J_M + P_A\otimes P_B\otimes J_M + I_{A\times A}\otimes I_{B\times B}\otimes P_M)\,y.
\end{aligned}$$

10.4 Balanced hierarchical models with two factors

Here we assume that the factor B is nested with the factor A. This means that the levels of factor B appear only for specific levels of factor A. If for example the levels of A are some species and B are subspecies of these species, then the subspecies belong only to one species. We can test then whether the species have an effect and whether the subspecies have an effects. Hence we have a hierarchy of the factors. Here we will regard only balanced hierarchical models which means for the examples of species that the number of regarded subspecies is always the same. Then we have the following model

$$Y_{abm} = \mu + \alpha_a + \beta_{ab} + Z_{abm} \quad \text{for } a = 1,\ldots,A,\ b = 1,\ldots,B,\ m = 1,\ldots,M.$$

This is a singular model so that side conditions are necessary:

$$\sum_{a=1}^{A}\alpha_a = 0,\ \sum_{b=1}^{B}\beta_{ab} = 0 \quad \text{for all } a = 1,\ldots,A.$$

The model can be also written as

$$Y = X\beta + Z,$$

with

$$\begin{aligned}
X &= (1_A\otimes 1_B\otimes 1_M \mid I_{A\times A}\otimes 1_B\otimes 1_M \mid I_{A\times A}\otimes I_{B\times B}\otimes 1_M) \in \mathbb{R}^{A\,B\,M\times 1+A+A\,B}\\
\beta &= (\mu,\alpha_1,\ldots,\alpha_A,\beta_{11},\ldots,\beta_{1B},\ldots,\beta_{A1},\ldots,\beta_{AB}) \in \mathbb{R}^{1+A+A\,B}.
\end{aligned}$$

The following hypothesis can be tested:

$$\begin{aligned} H_0^{B(A)}: \quad & \beta_{ab} = 0 \quad \text{for all } a = 1, \ldots, A,\ b = 1, \ldots, B, \\ \iff \quad & E(Y) \in C(X_{B(A)}) \quad \text{with } X_{B(A)} = (I_{A\times A} \otimes 1_B \otimes 1_M)\,, \end{aligned}$$

$$\begin{aligned} H_0^{A}: \quad & \alpha_a = 0 \quad \text{for all } a = 1, \ldots, A, \\ \iff \quad & E(Y) \in C(X_A) \quad \text{with } X_A = 1_A \otimes 1_B \otimes 1_M. \end{aligned}$$

Since $C(X) = C(\overline{X})$ with $\overline{X} = (I_{A\times A} \otimes I_{B\times B} \otimes 1_M)$, the perpendicular projection matrices can be obtained from those for the one-way ANOVA:

$$\begin{aligned} P &= \overline{X}(\overline{X}^\top \overline{X})^- \overline{X}^\top = I_{A\times A} \otimes I_{B\times B} \otimes \frac{1}{M} 1_M\, 1_M^\top, \\ P_{B(A)} &= X_{B(A)}(X_{B(A)}^\top X_{B(A)})^- X_{B(A)}^\top = I_{A\times A} \otimes \frac{1}{B} 1_B\, 1_B^\top \otimes \frac{1}{M} 1_M\, 1_M^\top, \\ P_A &= X_A(X_A^\top X_A)^- X_A^\top = \frac{1}{A} 1_A\, 1_A^\top \otimes \frac{1}{B} 1_B\, 1_B^\top \otimes \frac{1}{M} 1_M\, 1_M^\top, \end{aligned}$$

so that with the same notations P_A, P_B, P_M as before

$$\begin{aligned} I - P &= I_{A\times A} \otimes I_{B\times B} \otimes P_M, \\ P - P_{B(A)} &= I_{A\times A} \otimes P_B \otimes \frac{1}{M} 1_M\, 1_M^\top, \\ P_{B(A)} - P_A &= P_A \otimes \frac{1}{B} 1_B\, 1_B^\top \otimes \frac{1}{M} 1_M\, 1_M^\top. \end{aligned}$$

This provides the following ANOVA table (J_A, J_B, and J_M defined as before)

Factor	Matrix	Sum of squares	Rank
A	$P_A \otimes J_B \otimes J_M$	$M\,B \sum_{a=1}^{A} (\overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$	$A-1$
B(A)	$I_{A\times A} \otimes P_B \otimes J_M$	$M \sum_{a=1}^{A} \sum_{b=1}^{B} (\overline{y}_{ab\bullet} - \overline{y}_{a\bullet\bullet})^2$	$A\,(B-1)$
Error	$I_{A\times A} \otimes I_{B\times B} \otimes P_M$	$\sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{m=1}^{M} (y_{abm} - \overline{y}_{ab\bullet})^2$	$AB(M-1)$

Note that again the matrices sum up to $I_{A\times A} \otimes I_{B\times B} \otimes I_{M\times M} - J_A \otimes J_B \otimes J_M$.

10.5 General ANOVA models

In unbalanced models the order of the factors is important. Therefore the following hypotheses are considered for example in the two-way layout

$$\widetilde{H}_0^{A+B}: \mu_{ab} = \mu + \alpha_a + \beta_b \quad \text{versus} \quad \widetilde{H}_1^{A+B}: \mu_{ab} = \mu + \alpha_a + \beta_b + \gamma_{ab},$$

$$\widetilde{H}_0^{A|A+B}: \mu_{ab} = \mu + \beta_b \quad \text{versus} \quad \widetilde{H}_1^{A|A+B}: \mu_{ab} = \mu + \alpha_a + \beta_b,$$

$$\widetilde{H}_0^{B}: \mu_{ab} = \mu \quad \text{versus} \quad \widetilde{H}_1^{B}: \mu_{ab} = \mu + \beta_b.$$

Expressing the hypotheses via submodels, we obtain

$$\widetilde{H}_0^{A+B} : \mathrm{E}(Y) \in C(X_{A+B}) \quad \text{versus} \quad \widetilde{H}_1^{A+B} : \mathrm{E}(Y) \in C(X) \setminus C(X_{A+B}),$$
$$\widetilde{H}_0^{A|A+B} : \mathrm{E}(Y) \in C(X_{A|A+B}) \quad \text{versus} \quad \widetilde{H}_1^{A|A+B} : \mathrm{E}(Y) \in C(X_{A+B}) \setminus C(X_{A|A+B}),$$
$$\widetilde{H}_0^{B} : \mathrm{E}(Y) \in C(1_N) \quad \text{versus} \quad \widetilde{H}_1^{B} : \mathrm{E}(Y) \in C(X_{A|A+B}) \setminus C(1_N),$$

where X is the design matrix for the model with interactions. Let be P, P_{A+B}, $P_{A|A+B}$, P_0 the perpendicular projection matrices onto $C(X)$, $C(X_{A+B})$, $C(X_{A|A+B})$, and $C(1_N)$ respectively. Then the ANOVA uses the following test statistics

$$\widehat{V}_I = \frac{\widehat{\sigma}^2_{SSI}}{\widehat{\sigma}^2_{SSE}} = \frac{\frac{1}{\mathrm{rk}(P-P_{A+B})}\, y^\top (P - P_{A+B}) y}{\frac{1}{\mathrm{rk}(I-P)}\, y^\top (I-P) y} \qquad \text{for testing } \widetilde{H}_0^{A+B},$$

$$\widehat{V}_A = \frac{\widehat{\sigma}^2_{SSA|A+B}}{\widehat{\sigma}^2_{SSE}} = \frac{\frac{1}{\mathrm{rk}(P_{A+B}-P_{A|A+B})}\, y^\top (P_{A+B} - P_{A|A+B}) y}{\frac{1}{\mathrm{rk}(I-P)}\, y^\top (I-P) y} \qquad \text{for testing } \widetilde{H}_0^{A|A+B},$$

$$\widehat{V}_B = \frac{\widehat{\sigma}^2_{SSB}}{\widehat{\sigma}^2_{SSE}} = \frac{\frac{1}{\mathrm{rk}(P_{A|A+B}-P_0)}\, y^\top (P_{A|A+B} - P_0) y}{\frac{1}{\mathrm{rk}(I-P)}\, y^\top (I-P) y} \qquad \text{for testing } \widetilde{H}_0^{B}.$$

Although the denominator of the last two test statistics is not given by the full model for these test problems, the test statistics have F-distribution, since

$$(P_{A+B}) - P_{A|A+B})(I-P) = P_{A+B} - P_{A|A+B} - P_{A+B}\,P + P_{A|A+B}\,P = 0_{N\times N},$$
$$(P_{A|A+B} - P_0)(I-P) = P_{A|A+B} - P_0 - P_{A|A+B}\,P + P_0\,P = 0_{N\times N}.$$

Since $P_0 = J_N = \frac{1}{N}\,1_N\,1_N^\top$ and $I = I_{N\times N}$, we still have the decomposition of the Grand Sum of Squares

$$\sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{m=1}^{N_{ab}} (y_{abm} - \overline{y}_{\bullet\bullet\bullet})^2 = y^\top\, (I_{N\times N} - J_N)\, y$$
$$= \; y^\top \left((I_{N\times N} - P) + (P - P_{A+B}) + (P_{A+B} - P_{A|A+B}) + (P_{A|A+B} - P_0) \right) y.$$

For more than two factors, the Grand Sum of Squares can be decomposite analogously and the test statistics are given for the corresponding sequence of submodels.

Note that in the balanced two-way model we have

$$P - P_{A+B} = P - P_I, \;\; P_{A+B} - P_{A|A+B} = P_I - P_A, \;\; P_{A|A+B} - P_0 = P_I - P_B.$$

10.6 Models with random effects

It is often assumed that block factors have not fixed effects but random effects. Since there are usually also treatment factors which should have fixed effects, we have models with random and fixed effects. These models are called **mixed models** (German: Gemischte Modelle) and are given by

$$\begin{aligned} Y &= X\beta + V\mathcal{C} + Z \\ &= X\beta + (V_1, V_2, \ldots, V_Q) \begin{pmatrix} \mathcal{C}_1 \\ \mathcal{C}_2 \\ \vdots \\ \mathcal{C}_Q \end{pmatrix} + Z \\ &= X\beta + \sum_{q=1}^{Q} V_q \, \mathcal{C}_q + Z, \end{aligned}$$

where $X \in \mathbb{R}^{N\times R}$ is the known design matrix for the fixed effects, $\beta = (\beta_1, \ldots, \beta_R)^\top \in \mathbb{R}^R$ is the unknown vector of fixed effects, $V_1 \in \mathbb{R}^{N\times R_1}$, $V_2 \in \mathbb{R}^{N\times R_2}$, ..., $V_Q \in \mathbb{R}^{N\times R_Q}$ are the known design matrices for the random effects vectors $\mathcal{C}_1$, $\mathcal{C}_2$, ..., $\mathcal{C}_Q$. Each random effects vector corresponds to a factor so that we have Q factors with random effects. R_q is the number of observed levels of the q'th factor with random effects. The simplest approach is to assume for the random vectors $\mathcal{C}_1$, $\mathcal{C}_2$, ..., $\mathcal{C}_Q$, Z

$$\begin{aligned} &\mathcal{C}_q \sim \mathcal{N}_{R_q}\left(0_{R_q}, \sigma_q^2 \, I_{R_q\times R_q}\right) \quad \text{for} \quad q = 1, \ldots, Q, \\ &Z \sim \mathcal{N}_N(0_N, \sigma^2 \, I_{N\times N}), \end{aligned}$$

where $\sigma_1^2, \ldots, \sigma_Q^2, \sigma^2 \in \mathbb{R}^+$ are unknown fixed parameters, and that

$\mathcal{C}_1$, $\mathcal{C}_2$, ..., $\mathcal{C}_Q$, Z are stochastically independent .

Then

$$\theta = (\beta_1, \ldots, \beta_R, \sigma_1^2, \ldots, \sigma_Q^2, \sigma^2)^\top \in \mathbb{R}^{R+Q+1}$$

is the unknown parameter vector. Hence we have $R + Q + 1$ unknown parameters instead of $R + \sum_{q=1}^{Q} R_q + 1$ unknown parameters when we would assume fixed effects for the same design matrices. This means that using random effects reduces the number of unknown parameters drastically. The parameters $\sigma_1^2, \ldots, \sigma_Q^2$ of the random effects are called **variance components**. If a variance component satisfies $\sigma_q^2 = 0$ then the q'th factor with random effects has no influence on the measurement variable Y, since $\mathcal{C}_q \sim \mathcal{N}_{R_q}\left(0_{R_q}, \sigma_q^2 \, I_{R_q\times R_q}\right) = \mathcal{N}_{R_q}\left(0_{R_q}, 0_{R_q\times R_q}\right)$ means that $\mathcal{C}_q = 0_{R_q}$ almost surely.

The estimation and testing of the variance components is more complicated than in a linear model with only fixed effects. There are several approaches. Here only some of them:

1. The ANOVA (Analysis Of Variance) approach which is treated here.
2. The MINQUE (Minimum Norm Quadratic Estimation) method (C.R. Rao, 1972, Journal of the American Statistical Association).
3. The REML (Restricted Maximum Likelihood) method (R.R. Corbeil, and S.R. Searle, 1976, Technometrics).
4. The minimum bias invariant estimation method (J. Hartung 1981, Annals of Statistics.).

The ANOVA approach
Under the above mentioned assumptions we have with Lemma 8.2.6

$$\begin{aligned}
&\mathrm{E}(Y) = X\beta, \\
&\mathrm{Cov}(Y) = \sum_{q=1}^{Q} \sigma_q^2 \, V_q V_q^\top + \sigma^2 \, I_{N\times N} =: \sum_{l=1}^{L} \rho_l \, T_l,
\end{aligned}$$

where $T_1, \ldots, T_L$ are pairwise orthogonal matrices, i.e. $T_l \, T_k = 0_{N\times N}$ if $l \neq k$, which are symmetric and idempotent, i.e. they are perpendicular projection matrices satisfying $T_l \, T_l = T_l$ for all $l = 1, \ldots, L$. The coefficients $\rho_1, \ldots, \rho_L \in \mathbb{R}$ are linear combinations of the variance components.

10.6.1 Theorem
If ρ_l and ρ_k are positive, then

$$a) \quad \frac{1}{\rho_l} Y^\top T_l Y \sim \chi^2 \left(tr(T_l), \; \frac{\beta^\top X^\top T_l X \beta}{\rho_l} \right),$$

$$b) \quad \textit{if } T_k X\beta = 0 \textit{ then, } \quad \frac{\rho_k \frac{1}{tr(T_l)} Y^\top T_l Y}{\rho_l \frac{1}{tr(T_k)} Y^\top T_k Y} \sim F \left(tr(T_l), \; tr(T_k), \; \frac{\beta^\top X^\top T_l X \beta}{\rho_l} \right),$$

$$c) \quad \textit{if } T_k X\beta = 0 \textit{ and } \rho_l = \rho_k \textit{ then, } \quad \frac{\frac{1}{tr(T_l)} Y^\top T_l Y}{\frac{1}{tr(T_k)} Y^\top T_k Y} \sim F \left(tr(T_l), \; tr(T_k), \; \frac{\beta^\top X^\top T_l X \beta}{\rho_l} \right),$$

$$d) \quad E(Y^\top T_l Y) = \rho_l \, tr(T_l) + \beta^\top X^\top T_l X \beta.$$

Proof.
a) Note that $Y \sim \mathcal{N}_N(X\beta, W)$ with $W = \sum_{l=1}^{L} \rho_l \, T_l$.

$$W^{1/2} := \sum_{l=1}^{L} \sqrt{\rho_l} \, T_l,$$

satisfies

$$\begin{aligned} W^{1/2}W^{1/2} &= \left(\sum_{l=1}^{L}\sqrt{\rho_l}\,T_l\right)\left(\sum_{l=1}^{L}\sqrt{\rho_l}\,T_l\right) \\ &= \sum_{l=1}^{L}\sum_{k=1}^{L}\sqrt{\rho_l}\sqrt{\rho_k}\,T_l\,T_k \\ &\overset{T_l\,T_k=0,\ \text{if } l\neq k}{=} \sum_{l=1}^{L}\rho_l\,T_l\,T_l = \sum_{l=1}^{L}\rho_l\,T_l = W. \end{aligned}$$

The symmetry of the matrices T_l implies the symmetry of $V^{1/2}$. Since $W = \sum_{q=1}^{Q}\sigma_q^2\,V_qV_q^\top + \sigma^2\,I_{N\times N}$ is positive definite and thus regular, there exists W^{-1}. Moreover,

$$\begin{aligned} W\,\left(\sum_{l=1}^{L}\frac{1}{\rho_l}\,T_l\right)W &= W\,\left(\sum_{l=1}^{L}\sum_{k=1}^{L}\frac{1}{\rho_l}\,\rho_k\,T_l\,T_k\right) \\ &\overset{T_l\,T_k=0,\ \text{if } l\neq k}{=} W\,\sum_{l=1}^{L}T_l = \sum_{l=1}^{L}\sum_{k=1}^{L}\rho_k\,T_l\,T_k \overset{T_l\,T_k=0,\ \text{if } l\neq k}{=} \sum_{l=1}^{L}\rho_l\,T_l = W, \end{aligned}$$

which implies $W^{-1} = W^- = \sum_{l=1}^{L}\frac{1}{\rho_l}\,T_l$. Similarly, we obtain

$$W^{-1/2} = \sum_{l=1}^{L}\frac{1}{\sqrt{\rho_l}}\,T_l.$$

Then it holds

$$W^{-1/2}\,Y \sim \mathcal{N}_N(W^{-1/2}\,X\beta,\ W^{-1/2}W^{1/2}W^{1/2}W^{-1/2}) = \mathcal{N}_N(W^{-1/2}\,X\beta, I_{N\times N}).$$

Moreover, we have

$$\begin{aligned} \left(\frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}\right)\left(\frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}\right) &= \left(\frac{1}{\rho_l}\right)^2\,W^{1/2}\,T_l\,W\,T_l\,W^{1/2} \\ &= \left(\frac{1}{\rho_l}\right)^2\,W^{1/2}\,T_l\,(\sum_{k=1}^{L}\rho_k\,T_k)\,T_l\,W^{1/2} = \left(\frac{1}{\rho_l}\right)^2\,W^{1/2}\,T_l\,\rho_l\,T_l, T_l\,W^{1/2} \\ &\overset{T_lT_l=T_l}{=} \frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}, \end{aligned}$$

so that $\frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}$ is idempotent. Theorem 8.3.11 implies

$$\begin{aligned} \frac{1}{\rho_l}Y^\top T_l Y &= \left(W^{-1/2}Y\right)^\top\left(\frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}\right)\left(W^{-1/2}Y\right) \\ &\sim \chi^2\left(\mathrm{rk}\left(\frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}\right),\ \beta^\top X^\top\,W^{-1/2}\,\left(\frac{1}{\rho_l}\,W^{1/2}\,T_l\,W^{1/2}\right)\,W^{-1/2}\,X\beta\right) \\ &= \chi^2\left(\mathrm{rk}(T_l),\ \frac{\beta^\top X^\top T_l X\beta}{\rho_l}\right). \end{aligned}$$

b) and c) follow from a), the definition of the F-distribution and the fact that $Y^\top T_l Y$ and $Y^\top T_k Y$ are stochastically independent according to Theorem 8.3.6 because of

$$T_l\, W\, T_k = T_l \left(\sum_{j=1}^{L} \rho_j\, T_j \right) T_k = T_l\, \rho_k\, T_k = 0_{N\times N}$$

for $l \neq k$.
d) Lemma 8.2.6 provides

$$\begin{aligned} &\mathrm{E}(Y^\top T_l Y) = \mathrm{tr}(T_l\, W) + \beta^\top X^\top T_l X\beta \\ &= \quad \mathrm{tr}(T_l \sum_{k=1}^{L} \rho_k\, T_k) + \beta^\top X^\top T_l X\beta = \rho_l\, \mathrm{tr}(T_l) + \beta^\top X^\top T_l X\beta. \end{aligned}$$ □

The quadratic forms $Y^\top T_l Y$ are always nonnegative since the matrices T_l are positive semidefinite. Hence $\frac{1}{\mathrm{tr}(T_l)}\, Y^\top T_l Y$ can be used as estimator for ρ_l. This estimator is even unbiased for ρ_l according to Theorem MixedModelTheo d) if $T_l\, X = 0$. Since $\rho_1, \ldots, \rho_L$ are linear combinations of the variance components $\sigma_1^2, \ldots, \sigma_Q^2$, we also obtain estimators for the variance components by solving the linear equation system. However, the it could happen that some variance component estimators are negative which makes no sense since the variance components as variances should be nonnegative. The following example of the one-way layout with random effects demonstrate this problem.

10.6.2 Example (One-way layout with random effects)
The factor A has A levels which are chosen randomly so that their effects are random as well. We assume again a balanced design so that we have

$$\begin{aligned} &Y_{am} = \mu + \widetilde{\mathcal{A}}_a + Z_{am} \quad \text{for } a = 1, \ldots, A,\ m = 1, \ldots, M, \\ &X = 1_N, \\ &V = V_1 = I_{A\times A} \otimes 1_M, \\ &\mathcal{A} = \mathcal{A}_1 = (\widetilde{\mathcal{A}}_1, \ldots, \widetilde{\mathcal{A}}_A)^\top \sim \mathcal{N}(0_A, \sigma_A^2\, I_{A\times A}), \\ &Z = (Z_{11}, \ldots, Z_{1M}, \ldots, Z_{A1}, \ldots, Z_{AM})^\top \sim \mathcal{N}(0_N, \sigma^2\, I_{N\times N}) \quad \text{with } N = A\,M, \\ &W = \mathrm{Cov}(Y) = \sigma_A^2\, I_{A\times A} \otimes 1_M\, 1_M^\top + \sigma^2\, I_{A\times A} \otimes I_{M\times M}. \end{aligned}$$

The spectral decomposition of W is

$$W = (M\sigma_A^2 + \sigma^2)\, \frac{1}{A}\, 1_A\, 1_A^\top \otimes \frac{1}{M}\, 1_M\, 1_M^\top + (M\sigma_A^2 + \sigma^2)\, P_A \otimes \frac{1}{M}\, 1_M\, 1_M^\top + \sigma^2\, I_{A\times A} \otimes P_M$$

with $P_A = I_{A\times A} - \frac{1}{A}\, 1_A\, 1_A^\top$, $P_M = I_{M\times M} - \frac{1}{M}\, 1_M\, 1_M^\top$ again, and

$$\rho_1 = \rho_2 = M\sigma_A^2 + \sigma^2 \quad \text{and} \quad \rho_3 = \sigma^2.$$

Estimators for these coefficients are

$$\widehat{\rho}_2 = \frac{1}{A-1}\, y^\top \left(P_A \otimes \frac{1}{M}\, 1_M\, 1_M^\top \right) y = \frac{M}{A-1} \sum_{a=1}^{A} (\overline{y}_{a\bullet} - \overline{y}_{\bullet\bullet})^2,$$

$$\widehat{\rho}_3 = \widehat{\sigma}^2 = \frac{1}{A(M-1)}\, y^\top\, (I_{A\times A} \otimes P_M)\, y = \frac{1}{A(M-1)} \sum_{a=1}^{A} \sum_{m=1}^{M} (y_{am} - \overline{y}_{a\bullet})^2.$$

Then

$$\widehat{\sigma}^2 = \widehat{\rho}_3,$$
$$\widehat{\sigma}_A^2 = \frac{1}{M}\left(\widehat{\rho}_2 - \widehat{\rho}_3\right)$$

are estimators for σ^2 and σ_A^2, respectivly. The estimate $\widehat{\sigma}^2$ is always nonnegative, while this is not the case for $\widehat{\sigma}_A^2$ which can be seen with a simple example: Let be $y_{11} = 1$, $y_{12} = 5$, $y_{21} = 4$, $y_{22} = 2$. Then $3 = \overline{y}_{\bullet\bullet} = \overline{y}_{1\bullet} = \overline{y}_{2\bullet}$ so that $\widehat{\rho}_2 = 0$ and $\widehat{\sigma}^2 = \widehat{\rho}_3 = 5$, which implies $\widehat{\sigma}_A^2 = -2.5$.

Although the estimators are often not reasonable, we can derive reasonable tests for the following hypotheses:

$$H_0^1 : \mu = 0,$$

$$H_0^A : \text{ factor A has no effect} \iff \sigma_A^2 = 0 \iff M\,\sigma_A^2 + \sigma^2 = \sigma^2 \iff \rho_2 = \rho_3.$$

According to Theorem 10.6.1, we can use the following test statistics:
for testing $H_0^1 : \mu = 0$

$$F_{12} := \frac{\frac{1}{\rho_1}\,\frac{1}{\operatorname{tr}(\frac{1}{A}1_A 1_A^\top \otimes \frac{1}{M}1_M 1_M^\top)}\, y^\top \left(\frac{1}{A}1_A 1_A^\top \otimes \frac{1}{M}1_M 1_M^\top\right) y}{\frac{1}{\rho_2}\,\frac{1}{\operatorname{tr}(P_A \otimes \frac{1}{M}1_M 1_M^\top)}\, y^\top \left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) y} = \frac{y^\top \left(\frac{1}{N}1_N 1_N^\top\right) y}{\frac{1}{\operatorname{tr}(A-1)}\, y^\top \left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) y}$$

$$\sim F\left(1, A-1, \frac{\mu^2\, 1_N^\top \left(\frac{1}{A}1_A 1_A^\top \otimes \frac{1}{M}1_M 1_M^\top\right) 1_N}{\rho_1}\right) = F\left(1, A-1, \frac{\mu^2\, N}{\rho_1}\right),$$

where $F_{12} \sim F(1, A-1, 0)$ under H_0^1,

for testing $H_0^A : \sigma_A^2 = 0 \iff \rho_2 = \rho_3$

$$F_{23} := \frac{\frac{1}{\rho_2}\,\frac{1}{\operatorname{tr}(P_A \otimes \frac{1}{M}1_M 1_M^\top)}\, y^\top \left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) y}{\frac{1}{\rho_3}\,\frac{1}{\operatorname{tr}(I_{A\times A} \otimes P_M)}\, y^\top \left(I_{A\times A} \otimes P_M\right) y} = \frac{\frac{1}{\rho_2}\,\frac{1}{\operatorname{tr}(A-1)}\, y^\top \left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) y}{\frac{1}{\rho_3}\,\frac{1}{\operatorname{tr}(A(M-1))}\, y^\top \left(I_{A\times A} \otimes P_M\right) y} = \frac{\rho_3}{\rho_2} D_{23}$$

with $D_{23} \sim F\left(A-1, A(M-1), \frac{\mu^2\, 1_N^\top \left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) 1_N}{\rho_1}\right) = F\left(A-1, A(M-1), 0\right),$

since $\left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) 1_N = \left(P_A \otimes \frac{1}{M}1_M 1_M^\top\right) 1_A \otimes 1_M = P_A \cdot 1_A \;\otimes\; 1_M = 0,$

where $F_{23} \sim F(A-1, A(M-1), 0)$ under H_0^A.

This provides the following ANOVA table

Factor	Matrix	Sum of squares	Rank
μ	$\frac{1}{A}1_A 1_A^\top \otimes \frac{1}{M}1_M 1_M^\top$	$\overline{y}_{\bullet\bullet}^2$	1
A	$P_A \otimes \frac{1}{M}1_M 1_M^\top$	$M\ \sum_{a=1}^A (\overline{y}_{a\bullet} - \overline{y}_{\bullet\bullet})^2$	$A-1$
Error	$I_{A\times A} \otimes P_M$	$\sum_{a=1}^A \sum_{m=1}^M (y_{am} - \overline{y}_{a\bullet})^2$	$A(M-1)$

This is the same ANOVA table as for the one-way layout with fixed effects.

10.6.3 Example (Two-way layout with random effects)
For

$$Y_{abm} = \mu + \widetilde{\mathcal{A}}_a + \widetilde{\mathcal{B}}_b + \widetilde{\mathcal{AB}}_{ab} + Z_{abm} \quad \text{for} \quad a = 1, \ldots, A,\ b = 1, \ldots, B,\ m = 1, \ldots, M,$$

we obtain the same ANOVA table as for the two-way layout with fixed effects.

We obtain the same ANOVA tables also for mixed models and hierarchical models if the design is balanced.

11 Regression

11.1 Linear regression

The linear regression model is given by

$$Y_n = \beta_0 + \beta_1 t_n + Z_n, \quad \text{with } Z_n \sim \mathcal{N}(0, \sigma^2),$$

for all $n = 1, \ldots, N$, so that

$$x(t) = (1, t)^\top \in \mathbb{R}^2 \quad \text{and} \quad \beta = (\beta_0, \beta_1)^\top \in \mathbb{R}^2.$$

We assume here always that there exists $n,\ m \in \{1, \ldots, N\}$ with $t_n \neq t_m$. Then estimators for β_0, β_1, and σ^2 are given by

$$\widehat{\beta}_0 = \widehat{\beta}_0(y) = \overline{y} - \widehat{\beta}_1 \, \overline{t}, \quad \widehat{\beta}_1 = \widehat{\beta}_1(y) = \frac{s_{ty}}{s_t^2}, \tag{18}$$

$$\widehat{\sigma}^2 = \widehat{\sigma}^2(y) = \widehat{\sigma}^2_{SSE} = \frac{1}{N-2} \sum_{n=1}^{N} (y_n - \widehat{\beta}_0 - \widehat{\beta}_1 \, t_n)^2, \tag{19}$$

where

$$s_{ty} = \frac{1}{N-1} \sum_{n=1}^{N} (y_n - \overline{y})(t_n - \overline{t}), \quad s_t^2 = \frac{1}{N-1} \sum_{n=1}^{N} (t_n - \overline{t})^2.$$

11.1.1 Theorem (Least squares estimator for linear regression)
The estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)^\top$ with $\widehat{\beta}_0$ and $\widehat{\beta}_1$ given by (18) is the least squares estimator for β, i.e. satisfies

$$\widehat{\beta} \in \arg\min_{\beta \in \mathbb{R}^2} \sum_{n=1}^{N} (y_n - \beta_0 - \beta_1 \, t_n)^2.$$

Proof. Here we have

$$X = \begin{pmatrix} 1 & t_1 \\ 2 & t_2 \\ \vdots & \vdots \\ N & t_N \end{pmatrix}$$

and

$$\widetilde{\beta} \in \arg\min_{\beta \in \mathbb{R}^2} \sum_{n=1}^{N} (y_n - \beta_0 - \beta_1 \, t_n)^2 = \arg\min_{\beta \in \mathbb{R}^R} (y - X\beta)^\top (y - X\beta).$$

According to Theorem 9.2.3, $\widetilde{\beta}$ is given by $\widehat{\beta} = (X^\top X)^- X^\top y$. Because there exists $n,\ m \in \{1,\dots,N\}$ with $t_n \neq t_m$, we have $(X^\top X)^- = (X^\top X)^{-1}$. Hence we have to calculate $(X^\top X)^{-1}$ and $X^\top y$:

$$
\begin{aligned}
(X^\top X)^{-1} &= \begin{pmatrix} N & \sum_{n=1}^N t_n \\ \sum_{n=1}^N t_n & \sum_{n=1}^N t_n^2 \end{pmatrix}^{-1} \\
&= \frac{1}{N\ \sum_{n=1}^N t_n^2 - (\sum_{n=1}^N t_n)^2} \begin{pmatrix} \sum_{n=1}^N t_n^2 & -\sum_{n=1}^N t_n \\ -\sum_{n=1}^N t_n & N \end{pmatrix} \\
&= \frac{1}{N\ \sum_{n=1}^N t_n^2 - N^2\,\overline{t}^2} \begin{pmatrix} \sum_{n=1}^N t_n^2 & -\sum_{n=1}^N t_n \\ -\sum_{n=1}^N t_n & N \end{pmatrix} \\
&= \frac{1}{N\,(N-1)s_t^2} \begin{pmatrix} \sum_{n=1}^N t_n^2 & -\sum_{n=1}^N t_n \\ -\sum_{n=1}^N t_n & N \end{pmatrix},
\end{aligned}
$$

$$
X^\top y = \begin{pmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N t_n\, y_n \end{pmatrix}.
$$

Hence

$$
\begin{aligned}
\widehat{\beta} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} = (X^\top X)^- X^\top y &= \frac{1}{N\,(N-1)s_t^2} \begin{pmatrix} \sum_{n=1}^N t_n^2 & -\sum_{n=1}^N t_n \\ -\sum_{n=1}^N t_n & N \end{pmatrix} \begin{pmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N t_n\, y_n \end{pmatrix} \\
&= \frac{1}{N\,(N-1)s_t^2} \begin{pmatrix} (\sum_{n=1}^N t_n^2)\ (\sum_{n=1}^N y_n) - (\sum_{n=1}^N t_n)\ (\sum_{n=1}^N t_n\, y_n) \\ N\ \sum_{n=1}^N t_n\, y_n - (\sum_{n=1}^N t_n)\ (\sum_{n=1}^N y_n) \end{pmatrix} \\
&= \frac{1}{N\,(N-1)s_t^2} \begin{pmatrix} (\sum_{n=1}^N t_n^2)\ N\ \overline{y} - N^2\,\overline{t}^2\,\overline{y} + N^2\,\overline{t}^2\,\overline{y} - N\,\overline{t}\ (\sum_{n=1}^N t_n\, y_n) \\ N\ \sum_{n=1}^N t_n\, y_n - N^2\,\overline{y}\,\overline{t} \end{pmatrix} \\
&= \frac{1}{N\,(N-1)s_t^2} \begin{pmatrix} N\,(N-1)\ \overline{y}\ s_t^2 - N\,(N-1)\ \overline{t}\, s_{ty} \\ N\ (N-1)s_{ty} \end{pmatrix} \\
&= \begin{pmatrix} \overline{y} - \overline{t}\,\frac{s_{ty}}{s_t^2} \\ \frac{s_{ty}}{s_t^2} \end{pmatrix}.
\end{aligned}
$$

□

11.1.2 Theorem
The estimators given by (18) and (19) satisfy

$$E_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_0(Y)) = \beta_0, \quad E_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_1(Y)) = \beta_1, \quad E_{\beta_0,\beta_1,\sigma^2}(\widehat{\sigma}^2(Y)) = \sigma^2,$$

and

$$\sigma^2_{\beta_0} = var_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_0(Y)) = \frac{\overline{t^2}}{(N-1)s_t^2}\,\sigma^2 = \frac{\sum_{n=1}^N t_n^2}{N\,\sum_{n=1}^N (t_n-\overline{x})^2}\,\sigma^2,$$
$$\sigma^2_{\beta_1} = var_{\beta_0,\beta_1,\sigma^2}(\widehat{\beta}_1(Y)) = \frac{1}{(N-1)s_t^2}\,\sigma^2 = \frac{1}{\sum_{n=1}^N (t_n-\overline{t})^2}\,\sigma^2.$$

for all $\beta_0,\ \beta_1,\ \sigma^2$.

Proof. The unbiasedness of $\widehat{\beta}_0(Y)$, $\widehat{\beta}_1(Y)$, and $\widehat{\sigma}^2(Y)$ follows at once from Theorem 9.2.4. Note that $\mathrm{rk}(X) = 2$. Lemma 9.2.5 provides that

$$\mathrm{Cov}(L\widehat{\beta}(Y)) = L\,(X^\top X)^- L^\top\,\sigma^2.$$

Since

$$(X^\top X)^{-1} = \frac{1}{N\,(N-1)s_t^2}\begin{pmatrix} \sum_{n=1}^N t_n^2 & -\sum_{n=1}^N t_n \\ -\sum_{n=1}^N t_n & N \end{pmatrix}$$

(see the proof of Theorem 11.1.1), the formulas for the variances follow with $L = (1,0)$ and $L = (0,1)$. □

We define the following test statistics

$$\widehat{d}_0(y) = \frac{\widehat{\beta}_0(y) - b_0}{\widehat{\sigma}_{\beta_0}(y)} \quad \text{with} \quad \widehat{\sigma}^2_{\beta_0}(y) = \frac{\sum_{n=1}^N t_n^2}{N\,\sum_{n=1}^N (t_n-\overline{t})^2}\,\widehat{\sigma}^2(y)$$

and

$$\widehat{d}_1(y) = \frac{\widehat{\beta}_1(y) - b_1}{\widehat{\sigma}_{\beta_1}(y)} \quad \text{with} \quad \widehat{\sigma}^2_{\beta_1}(y) = \frac{1}{\sum_{n=1}^N (t_n-\overline{t})^2}\,\widehat{\sigma}^2(y).$$

11.1.3 Theorem (t-tests for for linear regression)
Let $Y \sim \mathcal{N}(X\beta, \sigma^2 I_{N\times N})$. *Then:*

a) $\varphi(y) = \mathbb{1}_{\{|\widehat{d}_0(y)| > t_{N-2,1-\alpha/2}\}}(y)$ *is* α*-level test for* $H_0 : \beta_0 = b_0$ *versus* $H_1 : \beta_0 \neq b_0$.

b) $\varphi(y) = \mathbb{1}_{\{|\widehat{d}_1(y)| > t_{N-2,1-\alpha/2}\}}(y)$ *is* α*-level test for* $H_0 : \beta_1 = b_1$ *versus* $H_1 : \beta_1 \neq b_1$.

Proof. According to Corollary 9.3.9,

$$\mathbb{1}_{\{\widehat{V}(y)>q(1-\alpha)\}}(y) \quad \text{with} \quad \widehat{V}(y) = \frac{\frac{1}{\mathrm{rk}(L)}\,(L\widehat{\beta}-l)^\top[L(X^\top X)^- L^\top]^-(L\widehat{\beta}-l)}{\frac{1}{\mathrm{rk}(I-P)}\,y^\top(I-P)y}$$

is an α-level test for $H_0:\ L\beta = l$ versus $H_1:\ L\beta \neq l$, where $q(1-\alpha)$ is the $1-\alpha$-quantile of the central F-distribution with $\mathrm{rk}(L)$ and $\mathrm{rk}(I-P) = N - \mathrm{rk}(X)$ degrees of freedom, $L = KX$, and $l = Lb$.
a) Set $L = (1,0)$ and $l = b_0$. Then $\mathrm{rk}(L) = 1$, $N - \mathrm{rk}(X) = N-2$, $\frac{1}{\mathrm{rk}(I-P)}\,y^\top(I-P)y = \widehat{\sigma}^2(y)$, and

$$\frac{\frac{1}{\mathrm{rk}(L)}\,(L\widehat{\beta}-l)^\top[L(X^\top X)^- L^\top]^-(L\widehat{\beta}-l)}{\frac{1}{\mathrm{rk}(I-P)}\,y^\top(I-P)y} = \frac{(\widehat{\beta}_0(y)-b_0)\left[\frac{\sum_{n=1}^N t_n^2}{N\sum_{n=1}^N (t_n-\bar{t})^2}\right]^{-1}(\widehat{\beta}_0(y)-b_0)}{\widehat{\sigma}^2(y)} = (\widehat{d}_0(y))^2.$$

Since $(\widehat{d}_0(y))^2$ has a F-distribution with 1 and $N-2$ degrees of freedom under $H_0:\beta_0 = b_0$, we obtain that $\widehat{d}_0(y)$ has a t-distribution with $N-2$ degrees of freedom under $H_0:\beta_0 = b_0$ which provides the assertion.
b) Setting $L = (0,1)$ and $l = b_1$, we have

$$\frac{\frac{1}{\mathrm{rk}(L)}\,(L\widehat{\beta}-l)^\top[L(X^\top X)^- L^\top]^-(L\widehat{\beta}-l)}{\frac{1}{\mathrm{rk}(I-P)}\,y^\top(I-P)y} = \frac{(\widehat{\beta}_1(y)-b_1)\left[\frac{1}{\sum_{n=1}^N (t_n-\bar{t})^2}\right]^{-1}(\widehat{\beta}_1(y)-b_1)}{\widehat{\sigma}^2(y)} = (\widehat{d}_1(y))^2.$$

Hence the assertion follows as in a). □

11.2 Regression with random regressors

If the explanatory variables (regressors) $t_1,\ldots,t_N$ are random, i.e. they are realizations of random variables $T_1,\ldots,T_N$, then we can model the conditional expectation with a linear model, i.e.

$$\mathrm{E}(Y_n|T_n = t_n) = x(t_n)^\top\beta,$$

where again $\beta \in \mathbb{R}^R$ is the unknown parameter vector and $x:\mathcal{T}\to\mathbb{R}^R$ is the known regression function. A justification for this approach is given by the following theorem.

11.2.1 Theorem

If Y_n is a p-dimensional random vector and T_n is a q-dimensional random vector with

$$(Y_n^\top, T_n^\top)^\top \sim \mathcal{N}_{p+q}\left(\begin{pmatrix}\mu_Y\\ \mu_T\end{pmatrix},\ \begin{pmatrix}\Sigma_{YY} & \Sigma_{YT}\\ \Sigma_{TY} & \Sigma_{TT}\end{pmatrix}\right),$$

where $E(Y_n) = \mu_Y$, $E(T_n) = \mu_T$, $Cov(Y_n) = \Sigma_{YY}$, $Cov(T_n) = \Sigma_{TT}$, $Cov(Y_n,T_n) = \Sigma_{YT}$, then the conditional distribution of Y_n given $T_n = t_n$ is a multivariate normal distribution with

$$E(Y_n|T_n = t_n) = \mu_Y + \Sigma_{YT}\Sigma_{TT}^{-1}(t_n - \mu_T).$$

Proof. See for example Rencher (1998), P. 47, Theorem 2.2E. □

In particular, if Y_n is univariate, then

$$\begin{aligned} \mathrm{E}(Y_n|T_n = t_n) &= \mu_Y + \Sigma_{YT}\Sigma_{TT}^{-1}(t_n - \mu_T) \\ &= \mu_Y - \Sigma_{YT}\,\Sigma_{TT}^{-1}\,\mu_T + \Sigma_{YT}\,\Sigma_{TT}^{-1}\,t_n = \beta_0 + (\beta_1, \ldots, \beta_q)\,t_n = x(t_n)^\top \beta \end{aligned}$$

with $\beta = (\beta_0, \beta_1, \ldots, \beta_q)^\top$, $\beta_0 = \mu_Y - \Sigma_{YT}\,\Sigma_{TT}^{-1}\,\mu_T$, $(\beta_1, \ldots, \beta_q) = \Sigma_{YT}\,\Sigma_{TT}^{-1} \in \mathbb{R}^q$, $x(t_n) = (1, \tau_{11}, \ldots, \tau_{qN})^\top \in \mathbb{R}^{q+1}$ for $t_n = (\tau_{11}, \ldots, \tau_{qN})^\top$, i.e. we have a model of multiple regression.

If $(Y_1, T_1), \ldots, (Y_N, T_N)$ are independent and identically distributed, we have two possibilities to estimate β:

1. Estimate μ_Y, μ_T, Σ_{YY}, Σ_{YT}, and Σ_{TT} with methods of multivariate analysis.
2. Estimate β with the methods of linear models by using the conditional distribution of $Y = (Y_1, \ldots, Y_N)^\top$ given $T = (T_1, \ldots, T_N)^\top = t = (t_1, \ldots, t_N)^\top$. However, assertions obtained with this approach concern only the conditional distribution. Under this condition also the tests given by Corollary 9.3.5 and 9.3.9 can be used.

The second approach can be used also without assuming a normal distribution or a multiple regression model since the following approach is always possible:

$$f_{(Y_n,T_n)}(y_n, t_n) = f_{(Y_n|T_n=t_n)}(y_n) \cdot f_{T_n}(t_n)$$

where $f_{(Y_n,T_n)}$ is common density of (Y_n, T_n) and $f_{(Y_n|T_n=t_n)}$ is the density of the condition distribution of Y_n given $T_n = t_n$ with

$$\mathrm{E}(Y_n|T_n = t_n) = \int y_n\, f_{(Y_n|T_n=t_n)}(y_n)\, d(y_n) = x(t_n)^\top \beta.$$

This makes in particular sense if the random process for choosing the experimental conditions is independent from the observation process. This is the case for "randomized designs" where the experimental conditions are chosen according a random process. An example is the allocation of medicaments according to the patient number, e.g. patients of a specific disease with even number get medicament A and patients of the same disease with odd number the medicament B.

However, for many other problems, the approach $f_{(Y_n,T_n)}(y_n, t_n) = f_{(Y_n|T_n=t_n)}(y_n) \cdot f_{T_n}(t_n)$ with $\mathrm{E}(Y_n|T_n = t_n) = x(t_n)^\top \beta$ makes no sense. If for example Y_n is the height of a person and T_n its weight, then the role of Y_n and T_n is exchangeable. Then the approach of errors-in-variables can be used.

Errors-in-variables

Here we assume

$$\begin{pmatrix} Y_n \\ T_n \end{pmatrix} = \begin{pmatrix} V_n \\ W_n \end{pmatrix} + Z,$$

where Z is a q-dimensional random vector with $\mathrm{E}(Z) = 0_q$, V_n a one-dimensional random variable, W_n a q-dimensional random vector so that

$$a^\top \begin{pmatrix} V_n \\ W_n \end{pmatrix} = a_0 V_n + a_q^\top W_n = b \quad \text{almost surely} , \tag{20}$$

i.e.

$$P\left(a^\top \begin{pmatrix} V_n \\ W_n \end{pmatrix} = b\right) = 1,$$

where $a = (a_0, a_q^\top)^\top \in \mathbb{R}^{q+1}$ and $b \in \mathbb{R}$ are unknown. It is further assumed that $\begin{pmatrix} V_n \\ W_n \end{pmatrix}$ and Z are stochastically independent. The assumption (20) means that $\begin{pmatrix} V_n \\ W_n \end{pmatrix}$ lies in the hyperplane

$$H_{a,b} = \{x \in \mathbb{R}^{q+1},\ a^\top x = b\} = \{x_0 + x;\ a^\top x = 0\} \quad \text{with} \quad a^\top x_0 = b.$$

If $a_0 \neq 0$, then

$$V_n = \frac{b}{a_0} - \frac{1}{a_0} a_q^\top W_n = x(W_n)^\top \beta$$

with $\beta = \left(\frac{b}{a_0}, -\frac{1}{a_0} a_q^\top\right)^\top$, $x(t_n) = (1, \tau_{11}, \ldots, \tau_{qN})^\top$. But $a_0 = 0$ is also possible. The condition $a^\top \begin{pmatrix} V_n \\ W_n \end{pmatrix} = b$ implies in particular for $q = 1$ the exchangeability of Y_n and T_n.

The aim is then to find $(a^\top, b)^\top \in \mathbb{R}^{q+2}$ such the perpendicular distance between the points $(y_1, t_1^\top)^\top, \ldots, (y_N, t_N^\top)^\top$ and the hyperplane $H_{a,b}$ is as small as possible. Thereby $\|a\| = 1$ is required since otherwise $(a^\top, b)^\top \in \mathbb{R}^{q+2}$ is not identifiable. The perpendicular projection $P((y_n, t_n^\top)^\top)$ of $(y_n, t_n^\top)^\top$ onto $H_{a,b}$ is given by

$$P((y_n, t_n^\top)^\top) = P_a\ (y_n, t_n^\top)^\top + (x_0 - P_a\ x_0)$$

where

$$P_a = (I - a(a^\top a)^- a^\top) \overset{\|a\|=1}{=} (I - a^\top a)$$

is the perpendicular projection matrix onto $C(a)^\perp = \{x \in \mathbb{R}^{q+1};\ a^\top x = 0\}$. Hence

$$\begin{aligned}
&\|(y_n, t_n^\top)^\top - P((y_n, t_n^\top)^\top)\| \\
&= \|(y_n, t_n^\top)^\top - (I - a^\top a)(y_n, t_n^\top)^\top - x_0 + (I - a^\top a)\, x_0\| \\
&= \left\| a\, a^\top \begin{pmatrix} y_n \\ t_n \end{pmatrix} - a\, a^\top x_0 \right\| = \|a\| \left| a^\top \begin{pmatrix} y_n \\ t_n \end{pmatrix} - a^\top x_0 \right| \\
&\overset{\|a\|=1}{=} \left| a^\top \begin{pmatrix} y_n \\ t_n \end{pmatrix} - b \right|
\end{aligned}$$

is the smallest distance between $(y_n, t_n^\top)^\top$ and $H_{a,b}$. Now $(a^\top, b)^\top$ is determined such that the sum of the squared distances is minimized.

11.2.2 Definition (Least squares estimator for orthogonal regression)
$\widehat{\beta} = (\widehat{a}^\top, \widehat{b})^\top$ *is least squares estimator for orthogonal regression if and only if*

$$\widehat{\beta} = (\widehat{a}^\top, \widehat{b})^\top \in \arg\min \left\{ \sum_{n=1}^{N} \left(a^\top \begin{pmatrix} y_n \\ t_n \end{pmatrix} - b \right)^2 ;\ (a^\top, b)^\top \in \mathbb{R}^{q+2},\ \|a\| = 1 \right\}.$$

The estimator $\widehat{\beta} = (\widehat{a}^\top, \widehat{b})^\top$ must be determined numerically. Therefore, its distribution is unknown and this is the case also if Z has normal distribution. Hence tests about a and b can be only obtained via asymptotic distributions.

12 Experimental design

The first aim of an experimental design is that all interesting aspects of the assumed model are identifiable. If there are many factors of quantitative and qualitative type, then this is no simple task. The theory of fractional factorial designs leads to designs where given aspects of the model are identifiable (see the book of Mukerjee and Wu, 2006). For more complex models also algebraic methods are helpful. In particular it can decided with the theory of Gröbner bases which models and which aspects of models are identifiable if a design is already given (see the book of Pistone, Riccomagno, and Wynn 2001). These concepts however are beyond this lecture.

Another practical aspect of designing experiments are balanced designs. Balanced complete designs provide a design matrix where the columns are orthogonal so that the parameter can be estimated independently of the other parameters. Because of this property, the analysis of variance is not dependent of the order of the factors. Moreover models with random effects can be treated with the analysis of variance like models with fixed effects.

Additionally, a good design should provide precise estimates and a small β error for testing. How this can be achieved, it treated in this section.

12.1 Generalized designs

12.1.1 Definition

Let be $\mathcal{T}$ the experimental region and $x : \mathcal{T} \longrightarrow \mathbb{R}^R$ the known regression function.

$$d = (t_1, \ldots, d_N) \in \mathcal{T}^N$$

is called **concrete design** *and*

$$X = X_d = \begin{pmatrix} x(t_1)^\top \\ x(t_2)^\top \\ \vdots \\ x(t_N)^\top \end{pmatrix} \in \mathbb{R}^{N \times R}$$

is the corresponding design matrix.

The aim of a good design d is to maximize the power of the test for testing $H_0 : L\beta = l$ or/and to minimize the covariance matrix of the Gauss-Markov estimator $L(X_d^\top X_d)^- X_d^\top y$ for $\lambda(\beta) = L\beta$. Fortunately the problem for testing as well as the problem for estimation leads to the same optimization problem, namely to minimize

$$L(X_d^\top X_d)^- L^\top,$$

see Remark 9.3.10. The only problem is, that $L \in \mathbb{R}^{S \times R}$ implies $L(X_d^\top X_d)^- L^\top \in \mathbb{R}^{S \times S}$ so that we have to minimize a matrix as soon as $S > 1$. On the set of $S \times S$-matrices we have only a partial

ordering by

$$A \le B \iff c^\top (B - A)\, c \ge 0 \text{ for all } c \in \mathbb{R}^S.$$

Partial ordering means that there are matrices which cannot be compared. If we have a special set of matrices $\{A;\ A \in \mathcal{A}\}$, it could be that there is no matrix $A_0 \in \mathcal{A}$ with

$$A_0 \le A \quad \text{for all} \quad A \in \mathcal{A}.$$

This happens in particular, if we compare different designs.

To reduce the dimension S, we always will assume here, that L is of full rank and that $\lambda(\beta) = L\beta$ is identifiable at d.

12.1.2 Lemma
If $\lambda(\beta) = L\beta$ is identifiable at d and $L \in \mathbb{R}^{S\times R}$ is of full rank, i.e. $rk(L) = S$, then $L(X_d^\top X_d)^- L^\top$ is invertible.

Proof. We always have $\mathrm{rk}(L(X_d^\top X_d)^- L^\top) \le \mathrm{rk}(L) = S$. Conversely, the identifiability implies with Theorem 9.1.4

$$\begin{aligned} S = \mathrm{rk}(L) = \mathrm{rk}(K\, X_d) &\overset{(\text{ Lemma 8.1.5 b)})}{=} \mathrm{rk}(K\, X_d (X_d^\top X_d)^- X_d^\top X_d) \\ &\le \mathrm{rk}(K\, X_d (X_d^\top X_d)^- X_d^\top) \overset{\mathrm{rg}(AA^\top)=\mathrm{rg}(A)}{=} \mathrm{rk}(K\, X_d (X_d^\top X_d)^- X_d^\top X_d\, (X_d^\top X_d)^- X_d^\top\, K^\top) \\ &= \mathrm{rk}(L(X_d^\top X_d)^- L^\top). \end{aligned}$$

Hence $L(X_d^\top X_d)^- L^\top$ is of rank S and thus invertible. □

The design problem
Find a design

$$d \in \Delta \subset \Delta_{N,\lambda} := \left\{ d_N \in \mathcal{T}^N;\ \lambda(\beta) \text{ is identifiable at } d \right\}$$

such that $(L(X_d^\top X_d)^- L^\top)^{-1}$ is maximal.

12.1.3 Definition (Information matrix)
a) $I_\lambda(d) := (L(X_d^\top X_d)^- L^\top)^{-1}$ is called information matrix for $\lambda(\beta) = L\beta$ at d.
b) $I_\beta(d) := X_d^\top X_d$ is called information matrix for β at d.

12.1.4 Remark
If $Y \sim \mathcal{N}(X_d\beta,\ \sigma^2 I_{N\times N})$ and σ^2 is known, then

$$\sigma^{-2}\, I_\beta(d) = \left(\mathrm{E}_\beta \left(\frac{\partial \ln f_\beta(y)}{\partial \beta_s} \cdot \frac{\partial \ln f_\beta(y)}{\partial \beta_r} \right) \right)_{r,s=1,\dots,R},$$

i.e. $\sigma^{-2}\, I_\beta(d)$ is the Fisher information matrix.

The maximization of the information matrix $I_\lambda(d)$ within the concrete designs is a complicated task. Therefore the designs are generalized:

$$d = (t_1, \ldots, t_N) \longrightarrow \delta_N = \frac{1}{N} \sum_{n=1}^{N} e_{t_n} \longrightarrow \delta \text{ probability measure } on(\mathcal{T}, \mathcal{D}),$$

where $\mathcal{D}$ is a σ-algebra on $\mathcal{T}$ and e_t denotes the Dirac measure, the one-point measure, at t, i.e. $e_t(A) = 1\!\!1_A(t)$ for all $A \in \mathcal{D}$.

12.1.5 Definition (Generalized design)
A probability measure δ on $(\mathcal{T}, \mathcal{D})$ is called generalized design.

12.1.6 Definition (Information matrix for generalized designs)
a) $I_\beta(\delta) := \int x(t)\, x(t)^\top \delta(dt)$ is called information matrix for β at δ.

b) $I_\lambda(\delta) := \left(L\, I_\beta(\delta)^- L^\top\right)^{-1}$ is called information matrix for $\lambda(\beta) = L\beta$ at δ.

To define the identifiability for generalized designs, note:

12.1.7 Lemma
$\lambda(\beta) = L\beta$ is identifiable at the concrete design d if and only if there exists $K \in \mathbb{R}^{S \times R}$ such that $L = K\, I_\beta(d)$.

Proof. According to Theorem 9.1.4, $\lambda(\beta) = L\beta$ is identifiable at d if and only if $L = K_0\, X_d$ for some $K_0 \in \mathbb{R}^{S \times N}$. Hence, if $L = K\, I_\beta(d) = K\, X_d^\top X_d = K_0\, X_d$, then $\lambda(\beta) = L\beta$ is identifiable at d. Conversely, if $\lambda(\beta) = L\beta$ is identifiable at d, then there exists $K_0 \in \mathbb{R}^{S \times N}$ with

$$L = K_0\, X_d \overset{(\text{Lemma 8.1.5 b}))}{=} K_0\, X_d (X_d^\top X_d)^- X_d^\top X_d = K\, I_\beta(d). \qquad \square$$

12.1.8 Definition (Identifiability at generalized designs)
$\lambda(\beta) = L\beta$ is called identifiable at the generalized design δ if and only if there exists $K \in \mathbb{R}^{S \times R}$ such that $L = K\, I_\beta(\delta)$.

12.1.9 Lemma
If $rk(L) = S$ and $L = K\, I_\beta(\delta)$, then $rk(L\, I_\beta(\delta)^- L^\top) = S$ and $L\, I_\beta(\delta)^- L^\top$ is independent of the choice of the g-inverse.

Proof. At first note, that the identifiability implies

$$L\, I_\beta(\delta)^- L^\top = K\, I_\beta(\delta) I_\beta(\delta)^- I_\beta(\delta)\, K^\top = K\, I_\beta(\delta)\, K^\top,$$

so that $L\, I_\beta(\delta)^- L^\top$ does not depend on the choice of the g-inverse. Since

$$a^\top I_\beta(\delta) a = \int a^\top x(t)\, x(t)^\top a\, \delta(dt) = \int (a^\top x(t))^2\, \delta(dt) \geq 0,$$

$I_\beta(\delta)$ is positive semidefinite and symmetric so that there exists $A \in \mathbb{R}^{Q\times R}$ such that $I_\beta(\delta) = A^\top A$. Hence the assertion follows as in the proof of Lemma 12.1.2. □

The design problem for generalized designs
Find a generalized design

$$\delta \in \Delta \subset \Delta_\lambda := \{\delta;\ \lambda(\beta) \text{ is identifiable at } \delta\}$$

such that $(L\, I_\beta(\delta) - L^\top)^{-1}$ is maximal.

If an optimal generalized design δ is found and if $x : \mathcal{T} \to \mathbb{R}^R$ is continous, $(\mathcal{T}, d_m)$ is a compact metric space with metric d_m and corresponding Borel-σ-algebra $\mathcal{D}$, then there exists a discrete probability measure (discrete design) $\overline{\delta}$ with

$$I_\beta(\delta) = I_\beta(\overline{\delta})$$

and finite support $\{\tau_1, \ldots, \tau_I\}$ with $I \leq \frac{R(R+1)}{2}$. This is a consequence of the Theorem of Caratheodory (see e.g. the book of Silvey 1980, P. 72) and the fact that the set of all probability measures with finite support is dense within all probability measures under the weak topology on the space of all probability measures on $(\mathcal{T}, \mathcal{D})$ (see e.g. the book of Billingsley 1968, P. 237).

Often it is also possible to find a concrete design d for a discrete design $\overline{\delta}$ such that

$$L\, I_\beta(\overline{\delta})^- L^\top = N\, L\, I_\beta(d)^- L^\top.$$

If this is not possible, then $\overline{\delta}$ must be approximated by an appropriate concrete design d.

The main reason for regarding the generalized designs is that the set of generalized designs is convex.

12.1.10 Lemma
If $rk(L) = S$, then

$$\Delta_\lambda := \{\delta;\ \lambda(\beta) \textit{ is identifiable at } \delta\}$$

is convex. In particular we have

$$\alpha\, \delta_1 + (1-\alpha)\, \delta_2 \in \Delta_\lambda$$

for all $\alpha \in (0,1)$, if $\delta_1 \in \Delta_\lambda$ and $I_\beta(\delta_2)$ is finite.

Proof. If $\delta_2 \in \Delta_\lambda$, then $I_\beta(\delta_2)$ is in particular finite. Since $I_\beta(\delta_2)$ is finite, we have

$$I_\beta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2) = \alpha\, I_\beta(\delta_1) + (1-\alpha)\, I_\beta(\delta_2) \geq \alpha\, I_\beta(\delta_1)$$

since $I_\beta(\delta_2)$ is positive semidefinite. In general, it holds for symmetric matrices A, B:

$$A \geq B \geq 0 \implies C(B) \subset C(A). \tag{21}$$

For, if $x \in C(A)^\perp$, then $0 = x^\top A$ and thus $0 = x^\top A\, x \geq x^\top B\, x$, so that $0 = x^\top B$. Hence $x \in C(A)^\perp$ implies $x \in C(B)^\perp$. Since $C(A)^\perp \subset C(B)^\perp$ implies $C(B) \subset C(A)$, the assertion (21) is proved. This means that we have $C(I_\beta(\delta_1)) \subset C(I_\beta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2))$ so that with $L = K\, I_\beta(\delta_1)$ also a $\overline{K}$ exists with $L = \overline{K}\, I_\beta(\alpha\,\delta_1 + (1-\alpha)\,\delta_2)$. Hence, $\alpha\,\delta_1 + (1-\alpha)\,\delta_2 \in \Delta_\lambda$. □

12.2 Optimality criteria for designs

12.2.1 Definition

Let be $\Delta \subset \Delta_\lambda$. The generalized design δ_* is called

a) U_λ-optimal in Δ $:\Longleftrightarrow$ $I_\lambda(\delta_*)^{-1} \leq I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$,

b) D_λ-optimal in Δ $:\Longleftrightarrow$ $\det I_\lambda(\delta_*)^{-1} \leq \det I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$,

c) A_λ-optimal in Δ $:\Longleftrightarrow$ $tr\, I_\lambda(\delta_*)^{-1} \leq tr\, I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$,

d) E_λ-optimal in Δ $:\Longleftrightarrow$ $\lambda_{max} I_\lambda(\delta_*)^{-1} \leq \lambda_{max} I_\lambda(\delta)^{-1}$ *for all* $\delta \in \Delta$.

Thereby det *denotes the determinat, tr the trace, and λ_{max} the maximum eigenvalue of a matrix.*

12.2.2 Lemma

Let A and B be symmetric $S \times S$ matrices and A positive definite, i.e. $A > 0$. Then there exists a regular matrix $U \in \mathbb{R}^{S\times S}$ with

$$A = U^\top U \quad \textit{and} \quad B = U^\top\, \textit{diag}(\mu_1, \ldots, \mu_S)\, U,$$

where $\mu_s \overset{>}{(-)} 0$ for $s = 1, \ldots, S$ if $B \overset{>}{(-)} 0$.

Proof. According to the spectral decomposition, there exists an orthogonal matrix P with

$$A = P^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_S)\, P,$$

where $\lambda_s > 0$ for $s = 1, \ldots, S$ because of $A > 0$. Set

$$D^{1/2} := \mathrm{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_S}\right), \quad D^{-1/2} = (D^{1/2})^{-1},$$

and

$$C := D^{-1/2} P\, B\, P^\top D^{-1/2}.$$

C is symmetric and $C \overset{>}{(-)} 0$ if $B \overset{>}{(-)} 0$. There exists also an orthogonal matrix Q with

$$Q\,C\,Q^\top = \mathrm{diag}(\mu_1, \ldots, \mu_S),$$

where $\mu_s \overset{>}{(-)} 0$ for $s = 1, \ldots, S$ if $B \overset{>}{(-)} 0$. Set $U := Q\,D^{1/2}P$. Then we have

$$U^\top U = P^\top D^{1/2} Q^\top Q\,D^{1/2} P = P^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_S) P = A$$

and

$$\begin{aligned} &U^\top \mathrm{diag}(\mu_1, \ldots, \mu_S)\,U = U^\top\,Q\,C\,Q^\top\,U = P^\top D^{1/2} Q^\top\,Q\,C\,Q^\top\,Q\,D^{1/2}P \\ &= \quad P^\top D^{1/2}\,C\,D^{1/2}P = P^\top D^{1/2}\,D^{-1/2}P\,B\,P^\top D^{-1/2}\,D^{1/2}P = P^\top P\,B\,P^\top P = B. \end{aligned}$$

□

12.2.3 Lemma
Let A and B be symmetric $S \times S$ matrices with $A \geq B > 0$. Then it holds

a) $A^{-1} \leq B^{-1}$,

b) $\det A \geq \det B$,

c) $tr\,A \geq tr\,B$.

Proof. According to Lemma 12.2.2, there exists a regular matrix $U \in \mathbb{R}^{S \times S}$ with

$$A = U^\top U \quad \text{and} \quad B = U^\top \mathrm{diag}(\mu_1, \ldots, \mu_S)\,U.$$

$A \geq B$ implies

$$I_{S\times S} = (U^\top)^{-1} A\,U^{-1} \geq (U^\top)^{-1} B\,U^{-1} = \mathrm{diag}(\mu_1, \ldots, \mu_S),$$

i.e. $1 \geq \mu_s > 0$ for $s = 1, \ldots, S$.
$I_{S\times S} \leq \mathrm{diag}(\mu_1^{-1}, \ldots, \mu_S^{-1})$ implies

$$A^{-1} = U^{-1}\,I_{S\times S}\,(U^\top)^{-1} \leq U^{-1}\,\mathrm{diag}(\mu_1^{-1}, \ldots, \mu_S^{-1})\,(U^\top)^{-1} = B^{-1}.$$

Moreover, we have

$$\det A = (\det U)^2\,\det I_{S\times S} \geq (\det U)^2\,\det \mathrm{diag}(\mu_1, \ldots, \mu_S) = \det B,$$

and

$$\mathrm{tr}\,A = \sum_{s=1}^{S} e_s^\top A\,e_s \geq \sum_{s=1}^{S} e_s^\top B\,e_s = \mathrm{tr}\,B,$$

where $e_1, \ldots, e_S$ are the unit vectors. □

12.2.4 Remark
If

$$L = \begin{pmatrix} L_1 \\ \dots \\ L_S \end{pmatrix},$$

then

$$\operatorname{tr} I_\lambda(\delta)^{-1} = \operatorname{tr} L\, I_\beta(\delta)^- L^\top = \sum_{s=1}^{S} L_s\, I_\beta(\delta)^- L_s^\top.$$

Hence the minimization of $\operatorname{tr} I_\lambda(\delta)^{-1}$ is equivalent with the minimization of $\frac{1}{S}\sum_{s=1}^{S} L_s\, I_\beta(\delta)^- L_s^\top$, the "average" of $L_s\, I_\beta(\delta)^- L_s^\top$. The is the reason that a design which minimizes $\operatorname{tr} I_\lambda(\delta)^{-1}$ is called A_λ-optimal.

For the estimation of $L\beta$ it is reasonable that all components $L_s\beta$ of $L\beta$ are estimated with high precision, i.e. with small variance. Since $\operatorname{tr} I_\lambda(\delta)^{-1}$ is the sum of the variances of the estimators for $L_s\beta$, the A_λ-optimal designs are in particular appropriate for estimation.

However, the A_λ-optimal designs are not appropriate for testing since they are not invariant with respect to regular transformations of $\lambda(\beta) = L\beta$, i.e. to transformations of $\widetilde{\lambda}(\beta) = \widetilde{L}\beta$ with $\widetilde{L} = H\,L$ where H is a regular matrix. They are only invariant with respect to orthogonal transformations where H is an orthogonal matrix. For testing, the optimal design should not depend on the special form how the hypothesis is formulated. Since $H_0 : L\beta = l$ is equivalent to $H_0 : \widetilde{L}\beta = \widetilde{l}$ with $\widetilde{L} = H\,L$ and $\widetilde{l} = H\,l$ if H is regular, the optimal design should be invariant with respect to transformations with regular designs. This is satisfies for D_λ-optimal designs.

12.2.5 Theorem
Let be $\lambda(\beta) = L\,\beta$ with $rk(L) = S$ and $\widetilde{\lambda}(\beta) = \widetilde{L}\beta$ with $\widetilde{L} = H\,L$ for $H \in \mathbb{R}^{S\times S}$.

a) If H is a regular matrix, then δ_* is D_λ-optimal in Δ if and only if δ_* is $D_{\widetilde{\lambda}}$-optimal in Δ.

b) If H is an orthogonal matrix, then δ_* is A_λ-optimal in Δ if and only if δ_* is $A_{\widetilde{\lambda}}$-optimal in Δ.

Proof.

$$\begin{aligned} a)\quad & \det I_{\widetilde{\lambda}}(\delta)^{-1} = \det \widetilde{L}\, I_\beta(\delta)^- \widetilde{L}^\top = \det H\,L\, I_\beta(\delta)^- L^\top H^\top \\ & = (detH)^2\, detL\, I_\beta(\delta)^- L^\top = (detH)^2\, detI_\lambda(\delta)^{-1}. \\ b)\quad & \operatorname{tr} I_{\widetilde{\lambda}}(\delta)^{-1} = \operatorname{tr} \widetilde{L}\, I_\beta(\delta)^- \widetilde{L}^\top = \operatorname{tr} H\,L\, I_\beta(\delta)^- L^\top H^\top \\ & \overset{\text{Lemma 8.1.3}}{=} \operatorname{tr} L\, I_\beta(\delta)^- L^\top H^\top H \overset{H\text{orthogonal}}{=} \operatorname{tr} L\, I_\beta(\delta)^- L^\top = \operatorname{tr} I_\lambda(\delta)^{-1}. \end{aligned}$$

□

Besides the invariance with respect to regular transformations of $\lambda(\beta) = L\,\beta$, the D_λ-optimal designs have the advantage that they minimize the volume of the confidence ellipsoid which is derived from

the F-test. Because of the relation between tests and confidence regions, this confidence ellipsoid is given according to Corollary 9.3.9 by

$$\widehat{B}_d(y) = \left\{ l \in \mathbb{R}^S;\ \frac{(L\widehat{\beta} - l)^\top I_\lambda(d)\,(L\widehat{\beta} - l)\,/\,\mathrm{rk}(L)}{\widehat{\sigma}^2(y)} \le q(1-\alpha) \right\},$$

where $q(1-\alpha)$ is the $1-\alpha$-quantile of the central F-distribution with $\mathrm{rk}(L)$ and $N - \mathrm{rk}(X_d)$ degrees of freedom. The volume of this confidence ellipsoid depends only via $\det I_\lambda(d)^{-1}$ on the design, since in general the volume $V^S(E)$ of an ellipsoid

$$E = \left\{ x \in \mathbb{R}^S;\ (x-\mu)^\top \Sigma^{-1}\,(x-\mu) \le q \right\}$$

is

$$V^S(E) = (q\pi)^{S/2} \left(\Gamma\left(\frac{S}{2} + 1 \right) \right)^{-1} (\det\,\Sigma)^{1/2},$$

where Γ is the Γ-function. (see e.g. the book of Pazman 1986, P. 79).

Moreover, D_λ-optimal designs minimizes the volume of ellipsoids where the power of the F-test given in Corollary 9.3.9 is bounded by given values. Namely, on the ellipsoid

$$E_d(q) := \left\{ L\beta \in \mathbb{R}^s;\ (L\beta - l)^\top\, I_\lambda(d)\,(L\beta - l) \le \sigma^2\, k \right\},$$

the power function (German: Gütefunktion) is given by

$$\gamma_d(\beta) \le 1 - F_{F(\mathrm{rk}(L),\,\mathrm{rk}(I-P),\,k)}(q(1-\alpha))$$

(see Corollary 9.3.9).

12.3 Characterizations of optimal designs

The characterizations of optimal designs based on the fact that the optimality criteria leads to convex functionals on the set of generalized designs. We consider here the following functionals:

$$\Phi_{A,\lambda}: \quad \Delta_\lambda \ni \delta \longrightarrow \Phi_{A,\lambda}(\delta) := \mathrm{tr}\, I_\lambda(\delta)^{-1} = \mathrm{tr}\; L\, I_\beta(\delta)^- L^\top \in \mathbb{R},$$
$$\Phi_{D,\lambda}: \quad \Delta_\lambda \ni \delta \longrightarrow \Phi_{D,\lambda}(\delta) := \ln \det I_\lambda(\delta)^{-1} = \ln \det\; L\, I_\beta(\delta)^- L^\top \in \mathbb{R}.$$

Note that minimizing $\Phi_{D,\lambda}(\delta)$ leads to the D_λ-optimal designs since the logarithm is a monotone increasing function. However the logarithm is necessary to provide the convexity of the functional. To prove the convexity of the functionals, we need the following lemmas.

12.3.1 Lemma
If $M_1,\ M_2 \in \mathbb{R}^{R\times R}$ are symmetric and positive semidefinite and $L \in \mathbb{R}^{S\times R}$ with $L = K_1 M_1$ and $L = K_2 M_2$, then

$$L\ (\alpha\, M_1 + (1-\alpha)\, M_2)^-\ L^\top\ \le\ \alpha\, L\, M_1^-\, L^\top + (1-\alpha)\, L\, M_2^-\, L^\top.$$

Proof. At first let be $M_1,\ M_2$ positive definite. Then also $M := \alpha\, M_1 + (1-\alpha)\, M_2$ is symmetric and positive definite. Hence we have for all $x, y \in \mathbb{R}^R$

$$0 \le \left(x^\top - y^\top M^{-1}\right)\ M\ \left(x - M^{-1}y\right) = y^\top M^{-1} y - (2\, x^\top y - x^\top M\, x),$$

where equality holds if and only if $x - M^{-1}y = 0$, i.e. $x = M^{-1}y$. This means

$$y^\top M^{-1} y = \max\left\{2\, x^\top y - x^\top M\, x;\ x \in \mathbb{R}^R\right\},$$

which implies for all $l \in \mathbb{R}^S$

$$\begin{aligned}
l^\top L\, M^{-1}\, L^\top l &= \max\left\{2\, x^\top L^\top l - x^\top M\, x;\ x \in \mathbb{R}^R\right\} \\
&= \max\left\{\alpha\,(2\, x^\top L^\top l - x^\top M_1\, x) + (1-\alpha)\,(2\, x^\top L^\top l - x^\top M\, x);\ x \in \mathbb{R}^R\right\} \\
&\le \alpha\ \max\left\{2\, x^\top L^\top l - x^\top M_1\, x;\ x \in \mathbb{R}^R\right\} \\
&\quad + (1-\alpha)\ \max\left\{2\, x^\top L^\top l - x^\top M_2\, x;\ x \in \mathbb{R}^R\right\} \\
&\overset{x=M_1^{-1}L^\top l \text{ bzw. } x=M_2^{-1}L^\top l}{=} \alpha\ \left(2\, l^\top L\, M_1^{-1} L^\top l - l^\top L\, M_1^{-1} L^\top l\right) \\
&\quad + (1-\alpha)\ \left(2\, l^\top L\, M_2^{-1} L^\top l - l^\top L\, M_2^{-1} L^\top l\right) \\
&= \alpha\, l^\top L\, M_1^{-1} L^\top l + (1-\alpha)\, l^\top L\, M_2^{-1} L^\top l.
\end{aligned}$$

For the case that M_1 or M_2 is not positive definite, see Kiefer (Journal of the Royal Statistical Society, B 21, P. 272ff) or Gaffke/Krafft (Modern Applied Mathematics - Optimization and Operation Research, Korte (eds.), North Holland 1981). □

12.3.2 Lemma
If $A,\ B \in \mathbb{R}^{R\times R}$ are symmetric and positive definite and $\alpha \in (0,1)$, then

$$\det(\alpha\, A + (1-\alpha)\, B) \ge (\det A)^\alpha\ (\det B)^{1-\alpha}.$$

Proof. If A and B are diagonal matrices, then the concavity of the logarithm provides

$$\begin{aligned}
\ln(\det(\alpha\, A + (1-\alpha)\, B)) &= \ln \prod_{r=1}^{R} (\alpha\, A_{rr} + (1-\alpha)\, B_{rr}) \\
&= \sum_{r=1}^{R} \ln\, (\alpha\, A_{rr} + (1-\alpha)\, B_{rr}) \;\geq\; \sum_{r=1}^{R} (\alpha\, \ln\, A_{rr} + (1-\alpha)\, \ln\, B_{rr}) \\
&= \alpha\, \ln \prod_{r=1}^{R} A_{rr} + (1-\alpha)\, \ln \prod_{r=1}^{R} B_{rr} \;=\; \ln \left(\prod_{r=1}^{R} A_{rr} \right)^{\alpha} + \ln \left(\prod_{r=1}^{R} B_{rr} \right)^{1-\alpha}.
\end{aligned}$$

To prove the assertion for the general case, we use the fact that according to Lemma 12.2.2 there exists a regular matrix U and diagonal matrix D such that

$$A = U^\top U \quad \text{and} \quad B = U^\top D\, U.$$

Then we obtain with the above result

$$\begin{aligned}
&\det(\alpha\, A + (1-\alpha)\, B) = \det(U^\top (\alpha\, I_{R\times R} + (1-\alpha)\, D)\, U) \\
&= (\det U)^2\; \det(\alpha\, I_{R\times R} + (1-\alpha)\, D) \;\geq\; (\det U)^2\, (\det I_{R\times R})^{\alpha}\, (\det D)^{1-\alpha} \\
&= \left((\det U)^2\; \det I_{R\times R}\right)^{\alpha} \left((\det U)^2\; \det D\right)^{1-\alpha} \;=\; (\det U^\top U)^{\alpha}\, (\det U^\top D\, U)^{1-\alpha} \\
&= (\det A)^{\alpha}\, (\det B)^{1-\alpha}.
\end{aligned}$$

□

12.3.3 Theorem

a) $\Phi_{A,\lambda}$ *is convex on* Δ_λ.

b) $\Phi_{D,\lambda}$ *is convex on* Δ_λ.

Proof.
a) Lemma 12.3.1 provides for all $\delta_1, \delta_2 \in \Delta_\lambda$

$$L\, I_\beta(\alpha\delta_1 + (1-\alpha)\delta_2)^- L^\top \;\leq\; \alpha\, L\, I_\beta(\delta_1)^- L^\top + (1-\alpha)\; L\, I_\beta(\delta_2)^- L^\top$$

so that with Lemma 12.2.3 we obtain

$$\begin{aligned}
\Phi_{A,\lambda}(\alpha\delta_1 + (1-\alpha)\delta_2) \;&=\; \operatorname{tr} L\, I_\beta(\alpha\delta_1 + (1-\alpha)\delta_2)^- L^\top \\
&\leq\; \operatorname{tr} \left(\alpha\, L\, I_\beta(\delta_1)^- L^\top + (1-\alpha)\; L\, I_\beta(\delta_2)^- L^\top\right) \\
&=\; \alpha \operatorname{tr} L\, I_\beta(\delta_1)^- L^\top + (1-\alpha) \operatorname{tr} L\, I_\beta(\delta_2)^- L^\top \;=\; \alpha\, \Phi_{A,\lambda}(\delta_1) + (1-\alpha)\, \Phi_{A,\lambda}(\delta_2).
\end{aligned}$$

b) We show here the assertion only for $\lambda(\beta) = \beta$. Lemma 12.3.2 and the definition of $I_\beta(\delta)$ provide for all $\delta_1, \delta_2 \in \Delta_\beta$

$$\begin{aligned}
&\Phi_{D,\beta}(\alpha\delta_1 + (1-\alpha)\delta_2) \;=\; \ln \det\left(I_\beta(\alpha\delta_1 + (1-\alpha)\delta_2)^{-1}\right) \\
&= \ln\left(\det\; I_\beta(\alpha\delta_1 + (1-\alpha)\delta_2)\right)^{-1} \;=\; -\ln\left(\det\; I_\beta(\alpha\delta_1 + (1-\alpha)\delta_2)\right) \\
&= -\ln\left(\det\; \alpha\, I_\beta(\delta_1) + (1-\alpha)\, I_\beta(\delta_2)\right) \;\le\; -\ln\left((\det\; I_\beta(\delta_1))^{\alpha}\;\; (\det\; I_\beta(\delta_2))^{1-\alpha}\right) \\
&= -\alpha \ln\left(\det\; I_\beta(\delta_1)\right) \;-\; (1-\alpha)\; \ln\left(\det\; I_\beta(\delta_2)\right) \\
&= \alpha \ln\left(\det\; I_\beta(\delta_1)^{-1}\right) \;+\; (1-\alpha)\; \ln\left(\det\; I_\beta(\delta_2)^{-1}\right) \\
&= \alpha\; \Phi_{D,\beta}(\delta_1) + (1-\alpha)\; \Phi_{D,\beta}(\delta_2).
\end{aligned}$$

The proof for general $\lambda(\beta) = L\beta$ is much more complicated and can be found in the books of Pázman (1986) and Pukelsheim (1993). □

Since $I_\beta(\delta)$ is a linear function in δ and the inverse, the trace, and the determinant are differentiable functions, the functionals $\Phi_{A,\lambda}$ and $\Phi_{D,\lambda}$ are Fréchet differentiable with respect to matric on Δ_λ which provides the weak topology. The directional derivatives, the Gâteaux derivatives, have rather simple forms. To derive these forms, we need the following lemma.

12.3.4 Lemma

a) If $A : \mathbb{R} \ni t \to A(t) \in \mathbb{R}^{N\times M}$ and $B : \mathbb{R} \ni t \to A(t) \in \mathbb{R}^{M\times K}$ are differentiable in t_0, then

$$\frac{\partial}{\partial t}\; A(t)\; B(t)\Big|_{t=t_0} = \left(\frac{\partial}{\partial t}\; A(t)\Big|_{t=t_0}\right)\; B(t_0) \;+\; A(t_0)\; \left(\frac{\partial}{\partial t}\; B(t)\Big|_{t=t_0}\right).$$

b) If $A : \mathbb{R} \ni t \to A(t) \in \mathbb{R}^{N\times N}$ is differentiable in t_0 and $A(t_0)$ is regular, then

$$\frac{\partial}{\partial t}\; A(t)^{-1}\Big|_{t=t_0} = -A(t_0)^{-1}\; \left(\frac{\partial}{\partial t}\; A(t)\Big|_{t=t_0}\right)\; A(t_0)^{-1}$$

and

$$\frac{\partial}{\partial t}\; \ln\det A(t)\Big|_{t=t_0} = tr\; \left(A(t_0)^{-1}\; \left(\frac{\partial}{\partial t}\; A(t)\Big|_{t=t_0}\right)\right).$$

Proof.
a) The assertion follows from the product rule.

b) Set $B(t) = A(t)\,A(t)^{-1} = I_{N\times N}$. The assertion a) implies

$$\begin{aligned}
\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0} &= \frac{\partial}{\partial t}\,A(t)^{-1}B(t)\Big|_{t=t_0} \\
&= \frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0} + A(t_0)^{-1}\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)A(t_0)^{-1} + A(t_0)^{-1}\,A(t_0)\left(\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0}\right) \\
&= 2\,\frac{\partial}{\partial t}\,A(t)^{-1}\Big|_{t=t_0} + A(t_0)^{-1}\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)A(t_0)^{-1}.
\end{aligned}$$

To prove the second assertion in b), let Π the set of all permutations of $\{1,\ldots,N\}$. Then we have

$$\begin{aligned}
\det A(t) &= \sum_{\pi\in\Pi}\operatorname{sgn}(\pi)\;A_{1\pi(1)}(t)\cdot\ldots\cdot A_{N\pi(N)}(t) \\
&= \sum_{n=1}^{N}A_{kn}(t)\sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t) \\
&= \sum_{n=1}^{N}A_{kn}(t)\;\alpha_{kn}(t)
\end{aligned}$$

for all $k = 1,\ldots,N$, where

$$\alpha_{kn}(t) = \sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t)$$

is the cofactor of $A(t)$ with respect to (k,n). It follows

$$\begin{aligned}
\frac{\partial}{\partial t}\,\ln\det A(t)\Big|_{t=t_0} &= \frac{1}{\det A(t_0)}\;\frac{\partial}{\partial t}\,\det A(t)\Big|_{t=t_0} \\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{\pi\in\Pi}\operatorname{sgn}(\pi)\left(\frac{\partial}{\partial t}\,A_{k\pi(k)}(t)\Big|_{t=t_0}\right)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t) \\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{n=1}^{N}\sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\left(\frac{\partial}{\partial t}\;A_{kn}(t)\Big|_{t=t_0}\right)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t) \\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{n=1}^{N}\left(\frac{\partial}{\partial t}\,A_{kn}(t)\Big|_{t=t_0}\right)\sum_{\pi\in\Pi,\;\pi(k)=n}\operatorname{sgn}(\pi)\prod_{m=1,\;m\neq k}^{N}A_{m\pi(m)}(t) \\
&= \frac{1}{\det A(t_0)}\sum_{k=1}^{N}\sum_{n=1}^{N}\left(\frac{\partial}{\partial t}\,A_{kn}(t)\Big|_{t=t_0}\right)\alpha_{kn} \\
&= \operatorname{tr}\left(A(t_0)^{-1}\left(\frac{\partial}{\partial t}\,A(t)\Big|_{t=t_0}\right)\right)
\end{aligned}$$

since $A(t_0)^{-1} = \frac{1}{\det A(t_0)}(\alpha_{kn})_{k,n=1,\ldots,N}$. □

To define Gâteaux differentiability, let Δ be a subset of all probability measure on $(\mathcal{T}, \mathcal{D})$ and define for $\delta_* \in \Delta$

$$\Delta(\delta_*) := \{\delta \in \Delta; \text{ there exists } k > 0 \text{ with } (1-\alpha)\,\delta_* + \alpha\,\delta \in \Delta \text{ for all } \alpha \le k\}\,.$$

12.3.5 Definition (Directional derivative and Gâteaux differentiability)
a) The directional derivative of the functional $\Phi : \Delta \longrightarrow \mathbb{R}$ at δ_* in direction of δ is defined as

$$\Phi'(\delta_*, \delta) := \lim_{\alpha \downarrow 0} \frac{\Phi((1-\alpha)\,\delta_* + \alpha\,\delta) - \Phi(\delta_*)}{\alpha} = \lim_{\alpha \downarrow 0} \frac{\Phi(\delta_* + \alpha\,(\delta - \delta^*)) - \Phi(\delta_*)}{\alpha}.$$

b) The functional $\Phi : \Delta \longrightarrow \mathbb{R}$ is Gâteaux differentiable at δ_* if and only if $\Phi'(\delta_*, \delta)$ exists for all $\delta \in \Delta(\delta_*)$ and

$$\Phi'(\delta_*, \delta) = \int \Phi'(\delta_*, e_t)\ \delta(dt)$$

for all $\delta \in \Delta(\delta_*)$, where e_t is the Dirac measure on t, i.e. $e_t(A) = \mathbb{1}_A(t)$ for all $A \in \mathcal{D}$.

12.3.6 Theorem
If $rk(L) = s$, then we have for all δ with $(1-\alpha)\,\delta_* + \alpha\,\delta \in \Delta_\lambda$ for sufficient small α the following directional derivatives

$$a) \quad \Phi'_{A,\lambda}(\delta_*, \delta) = tr\, L\, I_\beta(\delta_*)^- L^\top - \int |L\, I_\beta(\delta_*)^- x(t)|^2\ \delta(dt),$$

$$b) \quad \Phi'_{D,\lambda}(\delta_*, \delta) = S - \int x(t)^\top I_\beta(\delta_*)^- L^\top\ (L\, I_\beta(\delta_*)^- L^\top)^{-1}\ L\, I_\beta(\delta_*)^- x(t)\ \delta(dt).$$

Proof. Let be $\delta(\alpha) = (1-\alpha)\,\delta_* + \alpha\,\delta$. Since $I_\beta(\delta)$ is linear in δ we have

$$\begin{aligned}
\frac{\partial}{\partial\alpha} I_\beta(\delta(\alpha))\Big|_{\alpha=0} &= \lim_{\alpha\downarrow 0} \frac{I_\beta((1-\alpha)\,\delta_* + \alpha\,\delta) - I_\beta(\delta_*)}{\alpha} \\
&= \lim_{\alpha\downarrow 0} \frac{\int x(t)\,x(t)^\top\ ((1-\alpha)\,\delta_* + \alpha\,\delta)(dt) - \int x(t)\,x(t)^\top\ \delta_*(dt)}{\alpha} \\
&= \lim_{\alpha\downarrow 0} \frac{\alpha\,(\int x(t)\,x(t)^\top\ \delta(dt) - \int x(t)\,x(t)^\top\ \delta_*(dt))}{\alpha} \\
&= I_\beta(\delta) - I_\beta(\delta_*). \qquad (22)
\end{aligned}$$

At first we assume that $I_\beta(\delta_*)$ is regular.

a) Lemma 12.3.4 provides

$$\begin{aligned}
\Phi'_{A,\lambda}(\delta_*,\delta) &= \frac{\partial}{\partial\alpha}\mathrm{tr}\, L\, I_\beta(\delta(\alpha))^{-1}L^\top\Big|_{\alpha=0} \\
&= \mathrm{tr}\, L\left(\frac{\partial}{\partial\alpha}\, I_\beta(\delta(\alpha))^{-1}\Big|_{\alpha=0}\right) L^\top \\
&\overset{\text{Lemma 12.3.4}}{=} -\mathrm{tr}\, L\, I_\beta(\delta_*)^{-1}\left(\frac{\partial}{\partial\alpha}\, I_\beta(\delta(\alpha))\Big|_{\alpha=0}\right) I_\beta(\delta_*)^{-1}\, L^\top \\
&\overset{(22)}{=} -\mathrm{tr}\, L\, I_\beta(\delta_*)^{-1}\, (I_\beta(\delta) - I_\beta(\delta_*))\, I_\beta(\delta_*)^{-1}\, L^\top \\
&= \mathrm{tr}\, L\, I_\beta(\delta_*)^{-1}L^\top \;-\; \mathrm{tr}\, L\, I_\beta(\delta_*)^{-1}\, I_\beta(\delta)\, I_\beta(\delta_*)^{-1}\, L^\top \\
&= \mathrm{tr}\, L\, I_\beta(\delta_*)^{-1}L^\top \;-\; \mathrm{tr}\int L\, I_\beta(\delta_*)^{-1}\, x(t)\, x(t)^\top\, I_\beta(\delta_*)^{-1}\, L^\top\, \delta(dt) \\
&= \mathrm{tr}\, L\, I_\beta(\delta_*)^{-1}L^\top - \int |L\, I_\beta(\delta_*)^{-1}x(t)|^2\, \delta(dt).
\end{aligned}$$

b) Lemma 12.3.4 and Lemma 8.1.3 provide

$$\begin{aligned}
\Phi'_{D,\lambda}(\delta_*,\delta) &= \frac{\partial}{\partial\alpha}\ln\det\, L\, I_\beta(\delta(\alpha))^{-1}L^\top\Big|_{\alpha=0} \\
&\overset{\text{Lemma 12.3.4}}{=} \mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, \frac{\partial}{\partial\alpha}\, L\, I_\beta(\delta(\alpha))^{-1}L^\top\Big|_{\alpha=0}\right) \\
&= \mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\left(\frac{\partial}{\partial\alpha}\, I_\beta(\delta(\alpha))^{-1}\Big|_{\alpha=0}\right) L^\top\right) \\
&\overset{\text{Lemma 12.3.4}}{=} -\mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\beta(\delta_*)^{-1}\left(\frac{\partial}{\partial\alpha}\, I_\beta(\delta(\alpha))\Big|_{\alpha=0}\right) I_\beta(\delta_*)^{-1}\, L^\top\right) \\
&\overset{(22)}{=} -\mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\beta(\delta_*)^{-1}\, (I_\beta(\delta) - I_\beta(\delta_*))\, I_\beta(\delta_*)^{-1}\, L^\top\right) \\
&= \mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\beta(\delta_*)^{-1}L^\top\right) \\
&\quad - \mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\beta(\delta_*)^{-1}\, I_\beta(\delta)\, I_\beta(\delta_*)^{-1}\, L^\top\right) \\
&= \mathrm{tr}\, I_{S\times S} - \int \mathrm{tr}\left((L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\beta(\delta_*)^{-1}\, x(t)\, x(t)^\top\, I_\beta(\delta_*)^{-1}\, L^\top\right)\, \delta(dt) \\
&\overset{\text{Lemma 8.1.3}}{=} S - \int x(t)^\top I_\beta(\delta_*)^{-1}\, L^\top (L\, I_\beta(\delta_*)^{-1}L^\top)^{-1}\, L\, I_\beta(\delta_*)^{-1}\, x(t)\, \delta(dt).
\end{aligned}$$

The proof for singular $I_\beta(\delta_*)$ follows from the above properties with the fact that there exists always a regression function $\widetilde{x}:\mathcal{T}\to\mathbb{R}^Q$ and $\widetilde{L}\in\mathbb{R}^{S\times Q}$ so that $\widetilde{I}_\beta(\delta_*)=\int\widetilde{x}(t)\,\widetilde{x}(t)^\top\,\delta_*(dt)$ is regular and $L\,I_\beta(\delta_*)^-L^\top=\widetilde{L}\,\widetilde{I}_\beta(\delta_*)^{-1}\widetilde{L}^\top$ and $L\,I_\beta(\delta_*)^-x(t)=\widetilde{L}\,\widetilde{I}_\beta(\delta_*)^{-1}\widetilde{x}(t)$ for all $t\in\mathrm{supp}((1-\alpha)\,\delta_*+\alpha\,\delta)$, where $\mathrm{supp}(\delta)$ denotes the support of δ, i.e. the smallest set $A\in\mathcal{D}$ with $\delta(A)=1$. □

12.3.7 Corollary
If $rk(L)=s$, then $\Phi'_{A,\lambda}$ and $\Phi'_{D,\lambda}$ are Gâteaux differentiable at δ_*.

Proof. Theorem 12.3.6 implies

$$\Phi'_{A,\lambda}(\delta_*, \delta) = \int \Phi'_{A,\lambda}(\delta_*, e_t)\, \delta(dt)$$

and

$$\Phi'_{D,\lambda}(\delta_*, \delta) = \int \Phi'_{D,\lambda}(\delta_*, e_t)\, \delta(dt),$$

which are the additional conditions for Gâteaux differentiability. □

12.3.8 Theorem (Theorem of Whittle)
Let be Δ a convex subset of all probability measure on $(\mathcal{T}, \mathcal{D})$, $\delta_* \in \Delta$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and $\Phi : \Delta \to \mathbb{R}$ convex and Gâteaux differentiable at δ_*. Then the following assertions are equivalent:

a) $\Phi(\delta_*) = \min\limits_{\delta \in \Delta} \Phi(\delta)$,

b) $\Phi'(\delta_*, \delta) \geq 0 \quad$ *for all* $\quad \delta \in \Delta$,

c) $\Phi'(\delta_*, e_t) \geq 0 \quad$ *for all* $\quad t \in \mathcal{T}$.

Each of the assertion a), b), and c) implies

d) $\Phi'(\delta_*, e_t) = 0 \quad$ *for all* $\quad t \in supp(\delta_*)$.

Thereby, $supp(\delta)$ denotes the support of δ, i.e. the smallest set $A \in \mathcal{D}$ with $\delta(A) = 1$.

Proof.
a) $\Longrightarrow$ *b)* : If $\Phi(\delta_*) = \min_{\delta \in \Delta} \Phi(\delta)$, then

$$\Phi(\delta_*) \leq \Phi((1-\alpha)\, \delta_* + \alpha\, \delta)$$

for all $\delta \in \Delta(\delta_*)$ for sufficient small α. This implies

$$\Phi'(\delta_*, \delta) = \lim_{\alpha \downarrow 0} \frac{\Phi((1-\alpha)\, \delta_* + \alpha\, \delta) - \Phi(\delta_*)}{\alpha} \geq 0$$

for all $\delta \in \Delta(\delta_*)$.
b) $\Longrightarrow$ *c)* : This follows at once with $\delta = e_t$ and $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$.
c) $\Longrightarrow$ *a)* : Let be $\delta \in \Delta$ arbitrary. Since Δ and Φ are convex, it holds $(1-\alpha)\, \delta_* + \alpha\, \delta \in \Delta$ for all $\alpha \in [0, 1]$ and

$$\Phi(\delta) - \Phi(\delta_*) = \frac{(1-\alpha)\, \Phi(\delta_*) + \alpha\, \Phi(\delta) - \Phi(\delta_*)}{\alpha} \geq \frac{\Phi((1-\alpha)\, \delta_* + \alpha\, \delta) - \Phi(\delta_*)}{\alpha}$$

for all $\alpha \in (0,1)$. The Gâteaux differentiability implies then

$$\Phi(\delta) - \Phi(\delta_*) \geq \lim_{\alpha \downarrow 0} \frac{\Phi((1-\alpha)\,\delta_* + \alpha\,\delta) - \Phi(\delta_*)}{\alpha} = \Phi'(\delta_*, \delta) = \int \Phi'(\delta_*, e_t)\; \delta(dt) \overset{c)}{\geq} 0.$$

$c) \Longrightarrow d)$: $\Phi'(\delta_*, e_t) \geq 0$ for all $t \in \mathcal{T}$ and

$$0 = \Phi'(\delta_*, \delta_*) = \int \Phi'(\delta_*, e_t)\; \delta_*(dt)$$

implies $\Phi'(\delta_*, e_t) = 0$ for all $t \in \text{supp}(\delta_*)$. □

12.3.9 Theorem (Equivalence theorem for A-optimality)
Let be $\Delta \subset \Delta_\lambda$ convex, $\delta_* \in \Delta$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and $rk(L) = S$. Then the following assertions are equivalent:

a) δ_* *is A_λ optimal in Δ,*

b) $|L\, I_\beta(\delta_*)^- x(t)|^2 \leq tr\, L\, I_\beta(\delta_*)^- L^\top \quad$ *for all* $\quad t \in \mathcal{T}$.

If δ_* is A_λ-optimal in Δ, then

c) $|L\, I_\beta(\delta_*)^- x(t)|^2 = tr\, L\, I_\beta(\delta_*)^- L^\top \quad$ *for all* $\quad t \in supp(\delta_*)$.

Proof. The assertion follows at once from Theorem 12.3.6 a) and Theorem 12.3.8. □

12.3.10 Theorem (Equivalence theorem for D-optimality)
Let be $\Delta \subset \Delta_\lambda$ convex, $\delta_* \in \Delta$, $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$, and $rk(L) = S$. Then the following assertions are equivalent:

a) δ_* *is D_λ optimal in Δ,*

b) $x(t)^\top I_\beta(\delta_*)^- L^\top\, (L\, I_\beta(\delta_*)^- L^\top)^{-1}\, L\, I_\beta(\delta_*)^- x(t) \leq S \quad$ *for all* $\quad t \in \mathcal{T}$.

If δ_* is D_λ-optimal in Δ, then

c) $x(t)^\top I_\beta(\delta_*)^- L^\top\, (L\, I_\beta(\delta_*)^- L^\top)^{-1}\, L\, I_\beta(\delta_*)^- x(t) = S \quad$ *for all* $\quad t \in supp(\delta_*)$.

Proof. The assertion follows at once from Theorem 12.3.6 b) and Theorem 12.3.8. □

12.3.11 Remark
The condition $e_t \in \Delta(\delta_*)$ for all $t \in \mathcal{T}$ is satisfied for example for $\Delta = \Delta_\lambda$ (see Lemma 12.1.10).

12.4 Optimal designs for linear regression

In the linear regression model we have

$$x(t) = (1, t)^\top \quad \text{with} \quad t \in \mathcal{T} \subset \mathbb{R}$$

and

$$\beta = (\beta_0, \beta_1)^\top,$$

where β_0 is the intercept and β_1 the slope of the regression line.

12.4.1 Lemma
If $\mathcal{T} = [-a, a]$ for $0 < a \in \mathbb{R}$, then $\delta_* = \frac{1}{2}(e_{-a} + e_a)$ is A_β-optimal and D_β-optimal in Δ_β.

Proof. We have $x(t) = (1, t)^\top$ so that

$$\begin{aligned} I_\beta(\delta_*) &= \int x(t)\, x(t)^\top \, \delta_*(dt) = \frac{1}{2}\left(x(-a)\, x(-a)^\top + x(a)\, x(a)^\top\right) \\ &= \frac{1}{2}\left(\begin{pmatrix} 1 \\ -a \end{pmatrix} (1 \ -a) + \begin{pmatrix} 1 \\ a \end{pmatrix} (1 \ a)\right) = \begin{pmatrix} 1 & 0 \\ 0 & a^2 \end{pmatrix}. \end{aligned}$$

Since $L = I_{2\times 2}$, we obtain for all $t \in [-a, a]$

$$\begin{aligned} |L\, I_\beta(\delta_*)^{-1} x(t)|^2 &= \left| \begin{pmatrix} 1 & 0 \\ 0 & a^{-2} \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix} \right|^2 \\ &= 1 + a^{-4} t^2 \le 1 + a^{-2} = \operatorname{tr} I_\beta(\delta_*)^{-1} \end{aligned}$$

and

$$\begin{aligned} &x(t)^\top I_\beta(\delta_*)^{-1} L^\top \left(L\, I_\beta(\delta_*)^- L^\top\right)^{-1} L\, I_\beta(\delta_*)^- x(t) = x(t)^\top I_\beta(\delta_*)^{-1} x(t) \\ &= (1 \ t) \begin{pmatrix} 1 & 0 \\ 0 & a^{-2} \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix} = 1 + a^{-2} t^2 \le 2 \end{aligned}$$

so that δ_* is A_β-optimal and D_β-optimal in Δ_β according to Theorem 12.3.9 and Theorem 12.3.10, respectively. □

12.4.2 Lemma
If $\mathcal{T} = [0, a]$ for $0 < a \in \mathbb{R}$, then $\delta_D = \frac{1}{2}(e_0 + e_a)$ is D_β-optimal in Δ_β but not A_β-optimal. The A_β-optimal design in Δ_β is $\delta_A = \frac{1}{\sqrt{1+a^2}+1}\left(\sqrt{1+a^2}\, e_0 + e_a\right)$.

Proof. Because of

$$\begin{aligned} I_\beta(\delta_D) &= \int x(t)\,x(t)^\top\,\delta_D(dt) = \frac{1}{2}\left(x(0)\,x(0)^\top + x(a)\,x(a)^\top\right) \\ &= \frac{1}{2}\left(\begin{pmatrix}1\\0\end{pmatrix}(1\;0) + \begin{pmatrix}1\\a\end{pmatrix}(1\;a)\right) = \frac{1}{2}\begin{pmatrix}2 & a\\ a & a^2\end{pmatrix}, \\ I_\beta(\delta_D)^{-1} &= \frac{2}{a^2}\begin{pmatrix}a^2 & -a\\ -a & 2\end{pmatrix}, \end{aligned}$$

we have for all $t \in [0,a]$

$$\begin{aligned} & x(t)^\top I_\beta(\delta_D)^{-1}L^\top\,(L\,I_\beta(\delta_D)^- L^\top)^{-1}\,L\,I_\beta(\delta_D)^- x(t) = x(t)^\top I_\beta(\delta_D)^{-1}x(t) \\ &= (1\;t)\,\frac{2}{a^2}\begin{pmatrix}a^2 & -a\\ -a & 2\end{pmatrix}\begin{pmatrix}1\\t\end{pmatrix} \\ &= \frac{2}{a^2}\,(1\;t)\begin{pmatrix}a^2 - a\,t\\ 2\,t-a\end{pmatrix} = \frac{2}{a^2}\,(a^2 - a\,t + 2t^2 - a\,t) = \frac{2}{a^2}\,(a^2 + 2\,t(t-a)) \le 2, \end{aligned}$$

so that δ_D is D_β-optimal according to Theorem 12.3.10. Moreover, we obtain

$$\begin{aligned} & |L\,I_\beta(\delta_D)^{-1}x(t)|^2 = |I_\beta(\delta_D)^{-1}x(t)|^2 \\ &= \left|\frac{2}{a^2}\begin{pmatrix}a^2 & -a\\ -a & 2\end{pmatrix}\begin{pmatrix}1\\t\end{pmatrix}\right|^2 = \left|\frac{2}{a^2}\begin{pmatrix}a^2 - ta\\ 2t-a\end{pmatrix}\right|^2 \\ &= \frac{4}{a^4}\,\left((a^2-ta)^2 + (2t-a)^2\right) \overset{t=0}{=} \frac{4}{a^4}\,\left(a^4 + a^2\right) = \frac{1}{a^2}\,\left(4a^2+4\right) \\ &\ge \frac{1}{a^2}\,\left(2a^2+4\right) = \frac{2}{a^2}(a^2+2) = \operatorname{tr} I_\beta(\delta_D)^{-1}, \end{aligned}$$

so that condition b) of Theorem 12.3.9 is violated which means that δ_D is not A_β-optimal. To show that δ_A is A_β-optimal, set $\xi = \frac{1}{\sqrt{1+a^2}+1}$. Then we have

$$\begin{aligned} I_\beta(\delta_A) &= \int x(t)\,x(t)^\top\,\delta_A(dt) = \left((1-\xi)\,x(0)\,x(0)^\top + \xi\,x(a)\,x(a)^\top\right) \\ &= \left((1-\xi)\begin{pmatrix}1\\0\end{pmatrix}(1\;0) + \xi\begin{pmatrix}1\\a\end{pmatrix}(1\;a)\right) = \begin{pmatrix}1 & \xi\,a\\ \xi\,a & \xi\,a^2\end{pmatrix}, \\ I_\beta(\delta_D)^{-1} &= \frac{1}{\xi\,a^2(1-\xi)}\begin{pmatrix}\xi\,a^2 & -\xi\,a\\ -\xi\,a & 1\end{pmatrix}, \end{aligned}$$

so that

$$\operatorname{tr} I_\beta(\delta_A)^{-1} = \frac{\xi\,a^2+1}{\xi\,a^2(1-\xi)}.$$

Now the condition b) of Theorem 12.3.9,

$$|L\, I_\beta(\delta_A)^{-1} x(t)|^2 = |I_\beta(\delta_A)^{-1} x(t)|^2 \le \text{tr}\; I_\beta(\delta_A)^{-1}$$

for all $t \in [0, a]$, is equivalent with

$$\begin{aligned}
&\left| \frac{1}{\xi\, a^2(1-\xi)} \begin{pmatrix} \xi\, a^2 & -\xi\, a \\ -\xi\, a & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \end{pmatrix} \right|^2 = \left| \frac{1}{\xi\, a^2(1-\xi)} \begin{pmatrix} \xi\, a^2 - \xi\, a\, t \\ t - \xi\, a \end{pmatrix} \right|^2 \\
&\qquad = \frac{1}{\xi^2\, a^4 (1-\xi)^2} \left((\xi\, a^2 - \xi\, a\, t)^2 + (t - \xi\, a)^2 \right) = \frac{\xi^2\, a^2 (a-t)^2 + (t - \xi\, a)^2}{\xi^2\, a^4 (1-\xi)^2} \le \frac{\xi\, a^2 + 1}{\xi\, a^2 (1-\xi)} \\
&\iff \quad \xi^2\, a^2 (a-t)^2 + (t - \xi\, a)^2 \le \left(\xi\, a^2 + 1 \right) \left(\xi\, a^2 (1-\xi) \right) \\
&\iff \quad \xi^2\, a^2 (a^2 - 2\, a\, t + t^2) + (t^2 - 2\, t\, \xi\, a + \xi^2\, a^2) \le \left(\xi\, a^2 + 1 \right) \left(a^2 (\xi - \xi^2) \right) \\
&\iff \quad t^2 (\xi^2\, a^2 + 1) - 2t\, \xi\, a\, (\xi\, a^2 + 1) + \xi^2\, a^4 + \xi^2\, a^2 \le \xi^2\, a^4 + \xi\, a^2 - \xi^3\, a^4 - \xi^2\, a^2 \\
&\iff \quad t^2 (\xi^2\, a^2 + 1) - 2t\, \xi\, a\, (\xi\, a^2 + 1) + 2\, \xi^2\, a^2 - \xi\, a^2 + \xi^3\, a^4 \le 0. \qquad (23)
\end{aligned}$$

Using the special form of

$$\xi = \frac{1}{\sqrt{1+a^2}\; + 1} = \frac{\sqrt{1+a^2}\; - 1}{(\sqrt{1+a^2}\; + 1)(\sqrt{1+a^2}\; - 1)} = \frac{\sqrt{1+a^2}\; - 1}{a^2}$$

we obtain

$$\begin{aligned}
2\, \xi^2\, a^2 - \xi\, a^2 + \xi^3\, a^4 &= \frac{2\, (\sqrt{1+a^2} - 1)^2}{a^2} - \left(\sqrt{1+a^2} - 1 \right) + \frac{(\sqrt{1+a^2} - 1)^2\, (\sqrt{1+a^2} - 1)}{a^2} \\
&= \frac{2(1 + a^2 - 2\sqrt{1+a^2} + 1) - a^2 \sqrt{1+a^2} + a^2 + (1 + a^2 - 2\sqrt{1+a^2} + 1)(\sqrt{1+a^2} - 1)}{a^2} \\
&= \frac{1}{a^2} \Big(4 + 2\, a^2 - 4\sqrt{1+a^2} - a^2 \sqrt{1+a^2} + a^2 \\
&\qquad + \sqrt{1+a^2} + a^2 \sqrt{1+a^2} - 2 - 2\, a^2 + \sqrt{1+a^2} - 1 - a^2 + 2\sqrt{1+a^2} - 1 \Big) \\
&= 0.
\end{aligned}$$

Hence the inequality (23) is equivalent with

$$t^2 (\xi^2\, a^2 + 1) - 2t\, \xi\, a\, (\xi\, a^2 + 1) \le 0.$$

We see at once that equality holds if $t = 0$. The second root of the quadratic function

$$f(t) = t^2 (\xi^2\, a^2 + 1) - 2t\, \xi\, a\, (\xi\, a^2 + 1)$$

is given by

$$t = \frac{2\xi\, a\,(\xi\, a^2+1)}{\xi^2\, a^2+1} = \frac{2\, a\, \frac{\sqrt{1+a^2}-1}{a^2}\sqrt{1+a^2}}{\frac{(\sqrt{1+a^2}-1)^2}{a^2}+1} = \frac{2\, a\,(\sqrt{1+a^2}\ -1)\,\sqrt{1+a^2}}{(\sqrt{1+a^2}\ -1)^2+a^2}$$

$$= \frac{a\ 2\ \left(1+a^2-\sqrt{1+a^2}\right)}{1+a^2\ -2\sqrt{1+a^2}+1+a^2} = \frac{a\ 2\ \left(1+a^2-\sqrt{1+a^2}\right)}{2+2\,a^2\ -2\sqrt{1+a^2}} = a.$$

This means that the quadratic function f is zero for $t=0$ and $t=a$ and smaller than zero for all $t \in [0,a]$. Hence inequality (23) holds for all $t \in [0,a]$, so that the criterion b) of Theorem 12.3.9 for A_β-optimality of δ_A is satisfied. □

12.4.3 Remark
The somehow surprising result of Lemma 12.4.2 for the A_β-optimal design can be explained as follows: The A-optimality criterion is in particular appropriate for estimation since it is the average of the variance of the estimators for the single components. Since the intercept can be estimated only with observations at 0, while the slope can be estimated only with observations at 0 and a, it is advantageous to have more observations at 0.
Nevertheless the A_β-optimal design has the disadvantage that it depends on the scale of the experimental region $\mathcal{T}$. If we use another unit of measurement, we obtain a different optimal design.

12.4.4 Lemma
If $\mathcal{T} = [0,a]$ for $0 < a \in \mathbb{R}$ and the interesting aspect is the slope of the regression line, i.e. $\lambda(\beta) = \beta_1$, then $\delta_* = \frac{1}{2}(e_0 + e_a)$ is A_λ-optimal and D_λ-optimal in Δ_λ.

Proof. In the proof of Lemma 12.4.2 it was shown

$$I_\beta(\delta_*)^{-1} = \frac{2}{a^2}\begin{pmatrix} a^2 & -a \\ -a & 2 \end{pmatrix},$$

so that with $L = (0,1)$ we obtain

$$\left|L\, I_\beta(\delta_*)^{-1} x(t)\right|^2 = \left|\frac{2}{a^2}\,(-a\ 2)\,\begin{pmatrix} 1 \\ t \end{pmatrix}\right|^2 = \frac{4}{a^4}\,(2\,t-a)^2$$

$$= \frac{4}{a^4}\,(4\,t^2 - 4t\,a + a^2) \le \frac{4}{a^2} = L\, I_\beta(\delta_*)^{-1} L^\top$$

for all $t \in [0,a]$ since the $4\,t^2 - 4t\,a \le 0$ for $t \in [0,1]$. Hence Theorem 12.3.9 provides the A_λ-optimality of δ_*. Since $L\, I_\beta(\delta_*)^{-1} L^\top$ is one-dimensional, the A_λ-optimal design coincides with the D_λ-optimal design. □

12.5 Optimal designs for the one-way layout

In the one-way layout, we have only one factor with A levels so that

$$x(t) = (\mathbb{1}_1(t), \ldots, \mathbb{1}_A(t))^\top \in \mathbb{R}^A \quad \text{for} \quad t \in \mathcal{T} = \{1, \ldots, A\},$$
$$\beta = (\mu_1, \ldots, \mu_A)^\top \in \mathbb{R}^A.$$

If the first level is a control level (the standard crop, the placebo), then the interesting aspect of β is

$$\lambda(\beta) = \begin{pmatrix} \mu_2 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = L\,\beta \quad \text{with} \quad L = (-1_{A-1} \mid I_{A-1\times A-1}) \in \mathbb{R}^{A-1\times A}. \tag{24}$$

12.5.1 Lemma
If $\mathcal{T} = \{1, \ldots, A\}$ and the interesting aspect is given by (24), then $\delta_D = \frac{1}{A}\sum_{a=1}^A e_a$ is D_λ-optimal in Δ_λ and $\delta_A = \frac{1}{\sqrt{A-1}+A-1}\left(\sqrt{A-1}\,e_1 + \sum_{a=2}^A e_a\right)$ is A_λ-optimal in Δ_λ.

Proof. For proving the D_λ-optimality of δ_D, note

$$I_\beta(\delta_D) = \frac{1}{A} I_{A\times A},$$

$$L\,I_\beta(\delta_D)^{-1}L^\top = A\,L\,L^\top = A\,(-1_{A-1} \mid I_{A-1\times A-1}) \begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix},$$
$$= A\,(1_{A-1\times A-1} + I_{A-1\times A-1}),$$

$$\left(L\,I_\beta(\delta_D)^{-1}L^\top\right)^{-1} \overset{\text{Lemma 10.1.2}}{=} \frac{1}{A}\left(I_{A-1\times A-1} - \frac{1}{A}1_{A-1\times A-1}\right)$$

$$I_\beta(\delta_D)^{-1}L^\top \left(L\,I_\beta(\delta_D)^{-1}L^\top\right)^{-1} L\,I_\beta(\delta_D)^{-1}$$
$$= \frac{A^2}{A}\begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix}\left(I_{A-1\times A-1} - \frac{1}{A}1_{A-1\times A-1}\right)(-1_{A-1} \mid I_{A-1\times A-1})$$
$$= A\begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix}\left(-1_{A-1} + \frac{A-1}{A}\,1_{A-1} \mid I_{A-1\times A-1} - \frac{1}{A}1_{A-1\times A-1}\right)$$
$$= A\begin{pmatrix} A-1-\frac{(A-1)^2}{A} & -1_{A-1}^\top + \frac{A-1}{A}\,1_{A-1}^\top \\ -1_{A-1} + \frac{A-1}{A}\,1_{A-1} & I_{A-1\times A-1} - \frac{1}{A}1_{A-1\times A-1} \end{pmatrix}$$
$$= \begin{pmatrix} A-1 & -1_{A-1}^\top \\ -1_{A-1} & A\,I_{A-1\times A-1} - 1_{A-1\times A-1} \end{pmatrix}.$$

Hence for all $t \in \{1, \ldots, A\}$, we have

$$x(t)^\top \, I_\beta(\delta_D)^{-1} L^\top \left(L \, I_\beta(\delta_D)^{-1} L^\top \right)^{-1} L \, I_\beta(\delta_D)^{-1} \, x(t) = A - 1,$$

so that δ_D is D_λ-optimal according to Theorem 12.3.10. The proof of the A_λ-optimality of δ_A bases on the following calculations:

$$\begin{aligned}
I_\beta(\delta_A) &= \frac{1}{\sqrt{A-1} + A - 1} \begin{pmatrix} \sqrt{A-1} & 0_{A-1}^\top \\ 0_{A-1} & I_{A-1\times A-1} \end{pmatrix}, \\
I_\beta(\delta_A)^{-1} &= \left(\sqrt{A-1} + A - 1 \right) \begin{pmatrix} \frac{1}{\sqrt{A-1}} & 0_{A-1}^\top \\ 0_{A-1} & I_{A-1\times A-1} \end{pmatrix}, \\
L \, I_\beta(\delta_A)^{-1} &= \left(\sqrt{A-1} + A - 1 \right) (-1_{A-1} \mid I_{A-1\times A-1}) \begin{pmatrix} \frac{1}{\sqrt{A-1}} & 0_{A-1}^\top \\ 0_{A-1} & I_{A-1\times A-1} \end{pmatrix} \\
&= \left(\sqrt{A-1} + A - 1 \right) \left(-\frac{1}{\sqrt{A-1}} \, 1_{A-1} \mid I_{A-1\times A-1} \right), \\
L \, I_\beta(\delta_A)^{-1} L^\top &= \left(\sqrt{A-1} + A - 1 \right) \left(-\frac{1}{\sqrt{A-1}} \, 1_{A-1} \mid I_{A-1\times A-1} \right) \begin{pmatrix} -1_{A-1}^\top \\ I_{A-1\times A-1} \end{pmatrix} \\
&= \left(\sqrt{A-1} + A - 1 \right) \left(\frac{1}{\sqrt{A-1}} 1_{A-1\times A-1} + I_{A-1\times A-1} \right), \\
\operatorname{tr} L \, I_\beta(\delta_A)^{-1} L^\top &= \left(\sqrt{A-1} + A - 1 \right) \left(\frac{1}{\sqrt{A-1}} (A-1) + (A-1) \right) = \left(\sqrt{A-1} + A - 1 \right)^2, \\
\left| L \, I_\beta(\delta_A)^{-1} x(1) \right|^2 &= \left| \left(\sqrt{A-1} + A - 1 \right) \left(-\frac{1}{\sqrt{A-1}} \, 1_{A-1} \mid I_{A-1\times A-1} \right) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right|^2 \\
&= \left(\sqrt{A-1} + A - 1 \right)^2 \frac{1}{A-1} \, (A-1), \\
\left| L \, I_\beta(\delta_A)^{-1} x(t) \right|^2 &= \left(\sqrt{A-1} + A - 1 \right)^2 \quad \text{for} \quad t \in \{2, \ldots, A\}.
\end{aligned}$$

Hence for all $t \in \{1, \ldots, A\}$ it holds

$$\left| L \, I_\beta(\delta_A)^{-1} x(t) \right|^2 = \operatorname{tr} L \, I_\beta(\delta_A)^{-1} L^\top$$

so that Theorem 12.3.9 provides the A_λ-optimality of δ_A. □

12.5.2 Remark

For testing the hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_A$ the design should not depend on equivalent formulations of the hypothesis. Since

$$H_0 : \begin{pmatrix} \mu_2 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix} = 0_{A-1} \quad \text{and} \quad H_0 : \begin{pmatrix} \mu_1 - \mu_A \\ \vdots \\ \mu_{A-1} - \mu_A \end{pmatrix} = 0_{A-1}$$

are two equivalent formulation of $H_0 : \mu_1 = \mu_2 = \ldots = \mu_A$, the optimal design should not depend which level is chosen as control. This is satisfied by the D-optimal design, so that this shows again that D-optimal designs are in particular appropriate for testing. Since the D-optimal design is the balanced design, we see that balanced designs have also optimality properties with respect to the power of the test.

However, if we want to estimate the additional effects of the new treatments compared with the standard treatment given by level 1, then the interest lies really in estimating

$$L\beta = \begin{pmatrix} \mu_2 - \mu_1 \\ \vdots \\ \mu_A - \mu_1 \end{pmatrix}$$

and this should be done as precisely as possible. Hence the sum of the variances of the single component estimates should be as small as possible. Since μ_1 is involved in each component, the precision of each component estimate will be high if the precision of the estimate of μ_1 is high. Hence a high precision of the estimate of μ_1 influences the precision of all component estimates. This is achieved by more observations at level 1 than at the other levels. Since A-optimal designs provide this property, we see again that A-optimal designs are in particular appropriate for estimation.

Further literature

1. Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs.* Clarendon Press, Oxford.

2. Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York.

3. Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

4. Pistone, G., Riccomagno, E., and Wynn, H.P. (2001). *Algebraic Statistics: Computational Commutative Algebra in Statistics.* Chapman & Hall/CRC.

5. Pukelsheim, F. (1993). *Optimal Design of Experiments.* John Wiley, New York.

6. Silvey, S.D. (1980). *Optimal Design.* Chapman and Hall, London.

Part III

Solutions

13 Solutions of methodical exercises

13.1 Solution of Exercise 1.4.1

```
> chickenC<-chicken0[,c(1,2)]
> chickenC<-cbind(chicken0[,c(1,2)],"Control")
> chickenL<-cbind(chicken0[,c(1,3)],"Low")
> chickenH<-cbind(chicken0[,c(1,4)],"High")
> names(chickenC)<-c("Block","Weight","Feed")
> names(chickenL)<-c("Block","Weight","Feed")
> names(chickenH)<-c("Block","Weight","Feed")
> chicken<-rbind(chickenC,chickenL,chickenH)
> row.names(chicken)<-1:24
```

13.2 Solution of Exercise 1.5.1

```
> split0<-read.table("SPLIT.DAT")
> str(split0)
'data.frame':   36 obs. of  8 variables:
 $ V1: int  1 1 1 1 1 1 1 1 1 1 ...
 $ V2: int  1 1 1 1 2 2 2 2 3 3 ...
 $ V3: int  0 1 2 4 0 1 2 4 0 1 ...
 $ V4: int  111 130 157 174 117 114 161 141 105 140 ...
 $ V5: int  4 4 4 4 4 4 4 4 4 4 ...
 $ V6: int  1 1 1 1 2 2 2 2 3 3 ...
 $ V7: int  0 1 2 4 0 1 2 4 0 1 ...
 $ V8: int  74 89 81 122 64 103 132 133 70 89 ...
> split01<-split0[1:4]
> split02<-split0[5:8]
> names(split01)<-c("Block","Variety","Manure","Yield")
> names(split02)<-c("Block","Variety","Manure","Yield")
> split<-rbind(split01,split02)
> str(split)
'data.frame':   72 obs. of  4 variables:
 $ Block  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Variety: int  1 1 1 1 2 2 2 2 3 3 ...
 $ Manure : int  0 1 2 4 0 1 2 4 0 1 ...
 $ Yield  : int  111 130 157 174 117 114 161 141 105 140 ...
> row.names(split)<-1:72
> split$Block<-as.factor(split$Block)
> split$Variety<-as.factor(split$Variety)
```

```
> split$Manure<-as.numeric(split$Manure)/100
> split$Yield<-as.numeric(split$Yield)
> str(split)
'data.frame':   72 obs. of  4 variables:
 $ Block  : Factor w/ 6 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Variety: Factor w/ 3 levels "1","2","3": 1 1 1 1 2 2 2 2 3 3 ...
 $ Manure : num  0 0.01 0.02 0.04 0 0.01 0.02 0.04 0 0.01 ...
 $ Yield  : num  111 130 157 174 117 114 161 141 105 140 ...
> split.b<-split
> split.b$Manure<-as.factor(split.b$Manure)
> str(split.b)
'data.frame':   72 obs. of  4 variables:
 $ Block  : Factor w/ 6 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Variety: Factor w/ 3 levels "1","2","3": 1 1 1 1 2 2 2 2 3 3 ...
 $ Manure : Factor w/ 4 levels "0","0.01","0.02",..: 1 2 3 4 1 2 3 4 1 2 ...
 $ Yield  : num  111 130 157 174 117 114 161 141 105 140 ...
```

13.3 Solution of Exercise 1.5.2

At first a data file `PEPPERS3.DAT` of the following form is created:

```
   0     0     0     0     1     1     1     1
   0     0     1     1     0     0     1     1
   0     1     0     1     0     1     0     1
11.4  13.2  10.4     -  13.7     -  12.0  12.5
   -   8.4   6.5   6.1  10.8   9.4     -   9.1
   -  13.7     -     -  14.6  16.5  12.8  12.9
   -     -     -     -     -  15.4     -     -
   -  10.7     -     -  10.9  10.9   9.0  10.2
   -     -     -     -     -     -  10.1     -
```

```
> pepper0<-read.table("PEPPERS3.DAT",na.strings="-")
> pepper1<-data.frame(pepper0[-c(1,2,3),],c("Y1","Y1","Y2","Y2","Y2","Y2"),
+ c("B1","B2","B1","B1","B2","B2"))
> str(pepper1)
'data.frame':   6 obs. of  10 variables:
 $ V1                                    : num  11.4 NA NA NA NA NA
 $ V2                                    : num  13.2 8.4 13.7 NA 10.7 NA
 $ V3                                    : num  10.4 6.5 NA NA NA NA
 $ V4                                    : num  NA 6.1 NA NA NA NA
 $ V5                                    : num  13.7 10.8 14.6 NA 10.9 NA
 $ V6                                    : num  NA 9.4 16.5 15.4 10.9 NA
 $ V7                                    : num  12 NA 12.8 NA 9 10.1
 $ V8                                    : num  12.5 9.1 12.9 NA 10.2 NA
 $ c..Y1....Y1....Y2....Y2....Y2....Y2..: Factor w/ 2 levels "Y1","Y2": 1 1 2 2 2 2
 $ c..B1....B2....B1....B1....B2....B2..: Factor w/ 2 levels "B1","B2": 1 2 1 1 2 2
```

```
> P1<-data.frame(pepper1[,c(1,9,10)],"0","0","0")
> P2<-data.frame(pepper1[,c(2,9,10)],"0","0","1")
> P3<-data.frame(pepper1[,c(3,9,10)],"0","1","0")
> P4<-data.frame(pepper1[,c(4,9,10)],"0","1","1")
> P5<-data.frame(pepper1[,c(5,9,10)],"1","0","0")
> P6<-data.frame(pepper1[,c(6,9,10)],"1","0","1")
> P7<-data.frame(pepper1[,c(7,9,10)],"1","1","0")
> P8<-data.frame(pepper1[,c(8,9,10)],"1","1","1")
> names(P1)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P2)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P3)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P4)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P5)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P6)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P7)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> names(P8)<-c("Excess","Year","Block","Heating","Lighting","CO2")
> pepper<-rbind(P1,P2,P3,P4,P5,P6,P7,P8)
> pepper<-pepper[!is.na(pepper[,1]),]
> row.names(pepper)<-1:length(pepper[,1])
> str(pepper)
'data.frame':   24 obs. of  6 variables:
 $ Excess  : num  11.4 13.2 8.4 13.7 10.7 10.4 6.5 6.1 13.7 10.8 ...
 $ Year    : Factor w/ 2 levels "Y1","Y2": 1 1 1 2 2 1 1 1 1 1 ...
 $ Block   : Factor w/ 2 levels "B1","B2": 1 1 2 1 2 1 2 2 1 2 ...
 $ Heating : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 2 2 ...
 $ Lighting: Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 2 1 1 ...
 $ CO2     : Factor w/ 2 levels "0","1": 1 2 2 2 2 1 1 2 1 1 ...
> pepper
   Excess Year Block Heating Lighting CO2
1    11.4   Y1    B1       0        0   0
2    13.2   Y1    B1       0        0   1
3     8.4   Y1    B2       0        0   1
4    13.7   Y2    B1       0        0   1
5    10.7   Y2    B2       0        0   1
6    10.4   Y1    B1       0        1   0
7     6.5   Y1    B2       0        1   0
8     6.1   Y1    B2       0        1   1
9    13.7   Y1    B1       1        0   0
10   10.8   Y1    B2       1        0   0
11   14.6   Y2    B1       1        0   0
12   10.9   Y2    B2       1        0   0
13    9.4   Y1    B2       1        0   1
14   16.5   Y2    B1       1        0   1
15   15.4   Y2    B1       1        0   1
16   10.9   Y2    B2       1        0   1
17   12.0   Y1    B1       1        1   0
18   12.8   Y2    B1       1        1   0
19    9.0   Y2    B2       1        1   0
```

```
20   10.1   Y2    B2       1        1   0
21   12.5   Y1    B1       1        1   1
22    9.1   Y1    B2       1        1   1
23   12.9   Y2    B1       1        1   1
24   10.2   Y2    B2       1        1   1
```

13.4 Solution of Exercise 1.7.2

```
> str(mustard0)
'data.frame':   10 obs. of  4 variables:
 $ V1: int  21 39 31 13 52 39 55 50 29 17
 $ V2: int  27 21 26 12 11 8 NA NA NA NA
 $ V3: int  22 16 20 14 32 28 36 41 17 22
 $ V4: int  21 39 20 24 20 NA NA NA NA NA
> summary(mustard0)
       V1              V2              V3              V4
 Min.   :13.00   Min.   : 8.00   Min.   :14.00   Min.   :20.0
 1st Qu.:23.00   1st Qu.:11.25   1st Qu.:17.75   1st Qu.:20.0
 Median :35.00   Median :16.50   Median :22.00   Median :21.0
 Mean   :34.60   Mean   :17.50   Mean   :24.80   Mean   :24.8
 3rd Qu.:47.25   3rd Qu.:24.75   3rd Qu.:31.00   3rd Qu.:24.0
 Max.   :55.00   Max.   :27.00   Max.   :41.00   Max.   :39.0
                 NA's   : 4.00                   NA's   : 5.0
> str(mustard)
'data.frame':   31 obs. of  3 variables:
 $ length         : num  21 39 31 13 52 39 55 50 29 17 ...
 $ grow.conditions: Factor w/ 2 levels "light","dark": 1 1 1 1 1 1 1 1 1 1 ...
 $ cutting        : Factor w/ 2 levels "cut","noncut": 1 1 1 1 1 1 1 1 1 1 ...
> summary(mustard)
     length      grow.conditions   cutting
 Min.   : 8.00   light:16        cut   :20
 1st Qu.:18.50   dark :15        noncut:11
 Median :22.00
 Mean   :26.55
 3rd Qu.:34.00
 Max.   :55.00
```

Starting from the data table `mustard`, we obtain the same results for the 4 treatment groups as using `mustard0` using the following steps:

```
> mustard11<-mustard[mustard$grow.conditions=="light" &
+ mustard$cutting=="cut","length"]
> mustard12<-mustard[mustard$grow.conditions=="light"&mustard $
+ cutting=="noncut","length"]
> mustard21<-mustard[mustard$grow.conditions=="dark"&mustard $
+ cutting=="cut","length"]
> mustard22<-mustard[mustard$grow.conditions=="dark" &
```

```
+ mustard$cutting=="noncut","length"]
> cbind(summary(mustard11),summary(mustard12),
+ summary(mustard21),summary(mustard22))
         [,1]  [,2]  [,3] [,4]
Min.    13.00  8.00 14.00 20.0
1st Qu. 23.00 11.25 17.75 20.0
Median  35.00 16.50 22.00 21.0
Mean    34.60 17.50 24.80 24.8
3rd Qu. 47.25 24.75 31.00 24.0
Max.    55.00 27.00 41.00 39.0
```

The box-and-whisker plots are easily calculated.

```
> boxplot(length~grow.conditions*cutting,data=mustard)
> boxplot(length~cutting*grow.conditions,data=mustard)
```

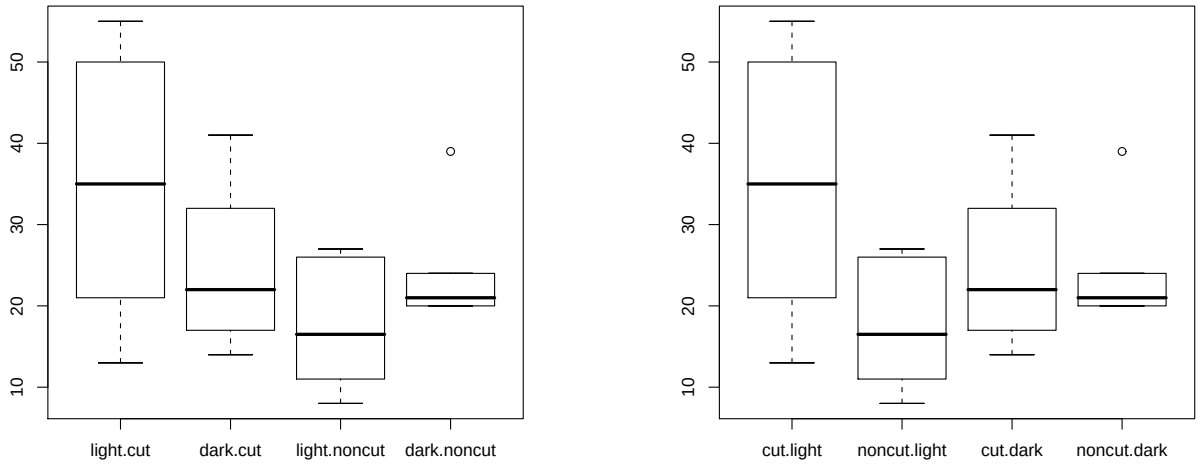

Figure 13.1: Box plots with `grow.conditions*cutting` and `cutting*grow.conditions`

```
> boxplot(length~grow.conditions,data=mustard)
> boxplot(length~cutting,data=mustard)
```

13.5 Solution of Exercise 1.7.3

```
> str(darwin)
'data.frame':   30 obs. of  3 variables:
 $ Pair         : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Height       : num  23.5 12 21 22 19.1 21.5 22.1 20.4 18.3 21.6 ...
 $ Fertilization: Factor w/ 2 levels "Cross","Self": 1 1 1 1 1 1 1 1 1 1 ...
> summary(darwin)
      Pair           Height      Fertilization
 Min.   : 1.00   Min.   :12.00   Cross:15
```

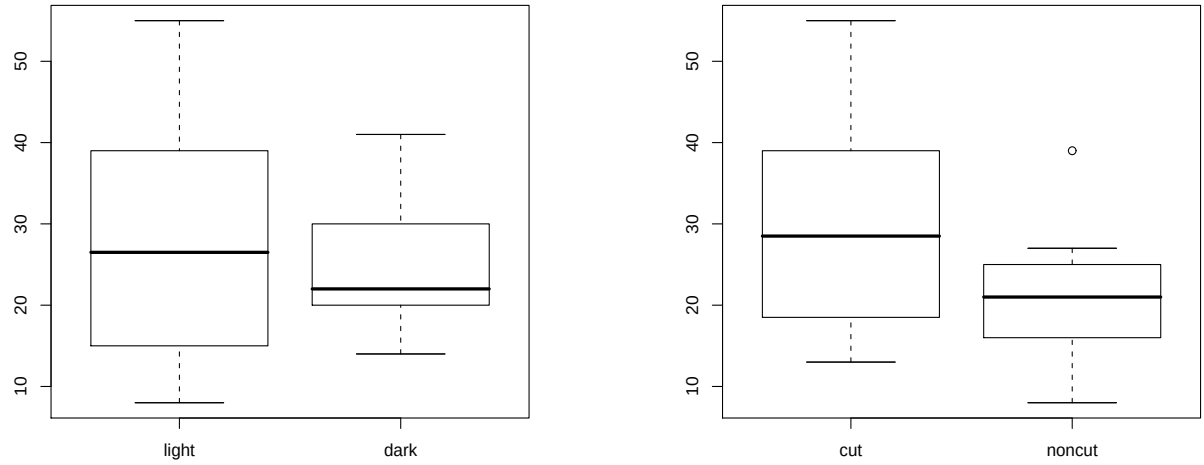

Figure 13.2: Box plots for `grow.conditions` and `cutting`

```
 1st Qu.: 4.25   1st Qu.:17.55   Self :15
 Median : 8.00   Median :18.85
 Mean   : 8.00   Mean   :18.89
 3rd Qu.:11.75   3rd Qu.:21.38
 Max.   :15.00   Max.   :23.50
```

The heights for the two different kinds of fertilization can be also be obtained as follows:

```
> summary(darwin[darwin$Fertilization=="Cross","Height"])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.00   19.75   21.50   20.19   22.10   23.50
> summary(darwin[darwin$Fertilization=="Self","Height"])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.80   16.40   18.00   17.59   18.60   20.40
```

This we would also obtain with the originally form of the data set:

```
> str(darwin0)
'data.frame':   15 obs. of  3 variables:
 $ Pair            : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Cross.fertilized: num  23.5 12 21 22 19.1 21.5 22.1 20.4 18.3 21.6 ...
 $ Self.fertilized : num  17.4 20.4 20 20 18.4 18.6 18.6 15.3 16.5 18 ...
> summary(darwin0)
      Pair      Cross.fertilized Self.fertilized
 Min.   : 1.0   Min.   :12.00    Min.   :12.80
 1st Qu.: 4.5   1st Qu.:19.75    1st Qu.:16.40
 Median : 8.0   Median :21.50    Median :18.00
 Mean   : 8.0   Mean   :20.19    Mean   :17.59
```

```
 3rd Qu.:11.5   3rd Qu.:22.10    3rd Qu.:18.60
 Max.   :15.0   Max.   :23.50    Max.   :20.40
```

```
> boxplot(Height~Fertilization,darwin)
> plot(darwin[darwin$Fertilization=="Cross","Height"],
+ darwin[darwin$Fertilization=="Self","Height"],
+ xlab="Height for Cross",ylab="Height for Self")
```

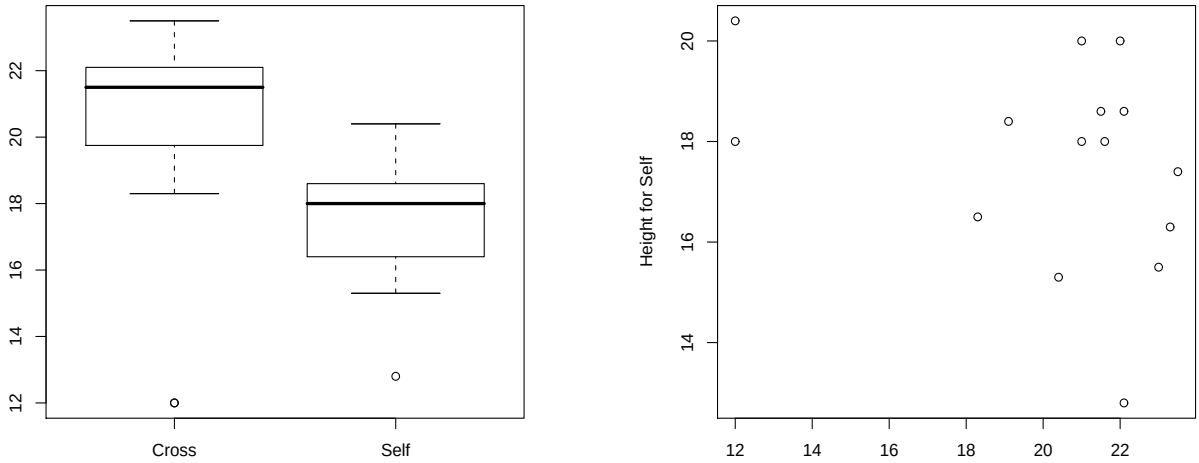

Figure 13.3: Box plot and scatter plot

13.6 Solution of Exercise 2.2.2

```
> growing0<-read.table("GROWING.DAT")
> Fert<-growing0[,1]
> Ster<-growing0[,2]
> Fert0<-data.frame(Fert,"Fertile")
> Ster0<-data.frame(Ster,"Sterile")
> names(Fert0)<-c("Height","Pollen")
> names(Ster0)<-c("Height","Pollen")
> growing<-rbind(Fert0,Ster0)
> str(growing)
'data.frame':   24 obs. of  2 variables:
 $ Height: int  92 107 98 97 95 94 92 96 98 104 ...
 $ Pollen: Factor w/ 2 levels "Fertile","Sterile": 1 1 1 1 1 1 1 1 1 1 ...
> boxplot(Height~Pollen,growing)
```

The analysis can be done in a short version:

```
> shapiro.test(Fert)$p.value
[1] 0.4524324
```

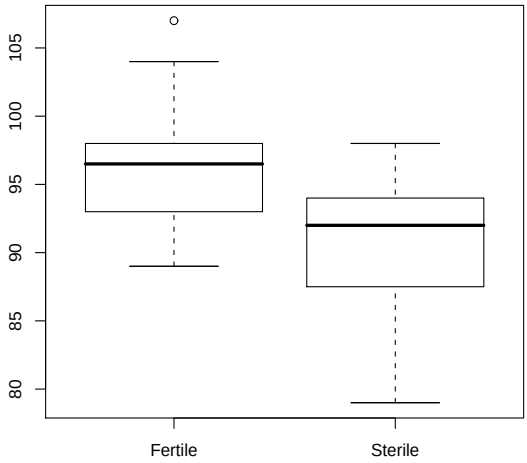

Figure 13.4: Box plots for the growing data

```
> shapiro.test(Ster)$p.value
[1] 0.51354
> var.test(Height~Pollen,growing)$p.value
[1] 0.8190027
> t.test(Height~Pollen,growing,var.equal=T)$p.value
[1] 0.01407609
> wilcox.test(Height~Pollen,growing,var.equal=T)$p.value
[1] 0.02567722
Warnmeldung:
cannot compute exact p-value with ties in: wilcox.test.default(x = c(92, 107, 98,
 97, 95, 94, 92, 96, 98,
```

or in an extended version:

```
> shapiro.test(Fert)
        Shapiro-Wilk normality test
data:  Fert
W = 0.9364, p-value = 0.4524
> shapiro.test(Ster)
        Shapiro-Wilk normality test
data:  Ster
W = 0.9412, p-value = 0.5135
> var.test(Fert,Ster)
        F test to compare two variances
data:  Fert and Ster
F = 0.8683, num df = 11, denom df = 11, p-value = 0.819
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2499653 3.0162273
```

```
sample estimates:
ratio of variances
         0.8683042
> t.test(Fert,Ster,var.equal=T)
        Two Sample t-test
data:  Fert and Ster
t = 2.6672, df = 22, p-value = 0.01408
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  1.260534 10.072800
sample estimates:
mean of x mean of y
 96.58333  90.91667
> wilcox.test(Fert,Ster)
        Wilcoxon rank sum test with continuity correction
data:  Fert and Ster
W = 111, p-value = 0.02568
alternative hypothesis: true mu is not equal to 0
Warnmeldung:
cannot compute exact p-value with ties in: wilcox.test.default(Fert, Ster)
```

The Wilcoxon tests provides a worse p-value, but still would also reject the hypothesis of equal means.

13.7 Solution of Exercise 2.2.3

At first we check the normality assumtion:

```
> shapiro.test(darwin[darwin$Fertilization=="Cross","Height"])$p.value
[1] 0.0009706594
> shapiro.test(darwin[darwin$Fertilization=="Self","Height"])$p.value
[1] 0.3838259
```

Hence the t-test cannot be used.

```
> wilcox.test(Height~Fertilization,data=darwin)$p.value
[1] 0.002608089
Warnmeldung:
cannot compute exact p-value with ties in: wilcox.test.default(x = c(23.5, 12, 21,
 22, 19.1, 21.5, 22.1,
```

We can conclude that the heights under cross- and self-fertilitzation differ significantly. The difference between the means is given by

```
> darwinC<-darwin[darwin$Fertilization=="Cross","Height"]
> darwinS<-darwin[darwin$Fertilization=="Self","Height"]
```

```
> mean(darwinC)-mean(darwinS)
[1] 2.606667
```

The boxplots in Figure 13.3 show the difference between the two groups. The extreme outlier in the cross-fertilized plants explains that the normal distribution is rejected there.

13.8 Solution of Exercise 2.3.1

```
> beta.error(mu1=3,mu2=4,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.8028
> beta.error(mu1=3,mu2=3.5,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.916
> beta.error(mu1=3,mu2=3.1,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.947
> beta.error(mu1=3,mu2=3.01,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.9484
> beta.error(mu1=6,mu2=7,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.8002
> beta.error(mu1=6,mu2=6.5,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.9192
> beta.error(mu1=6,mu2=6.1,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.9469
> beta.error(mu1=6,mu2=6.01,sigma=2,N1=10,N2=12)
$beta.error
[1] 0.9505
> beta.error(mu1=6,mu2=7,sigma=1,N1=10,N2=12)
$beta.error
[1] 0.3974
> beta.error(mu1=6,mu2=6.5,sigma=1,N1=10,N2=12)
$beta.error
[1] 0.8037
> beta.error(mu1=6,mu2=6.1,sigma=1,N1=10,N2=12)
$beta.error
[1] 0.9448
> beta.error(mu1=6,mu2=6.01,sigma=1,N1=10,N2=12)
$beta.error
[1] 0.9499
```

The β-error depends only on the absolute difference $|\mu_1 - \mu_2|$ and on the variance σ^2. The smaller the absolute difference $|\mu_1 - \mu_2|$ is the greater the β-error is and the β-error approaches 0.95=1-0.05.

If the variance is smaller then the β-error is also smaller.

```
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*1/2)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*1/2)
[1] 0.800646
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.5/2)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.5/2)
[1] 0.9138582
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.1/2)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.1/2)
[1] 0.9485803
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.01/2)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.01/2)
[1] 0.9499858
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*1/1)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*1/1)
[1] 0.3964526
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.5/1)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.5/1)
[1] 0.800646
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.1/1)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.1/1)
[1] 0.9443069
> pt(qt(0.975,20),20,ncp=sqrt(120/22)*0.01/1)
+ -pt(-qt(0.975,20),20,ncp=sqrt(120/22)*0.01/1)
[1] 0.9499433
```

The simulated values are very similar to the theoretical values although there are some small differences due to randomness.

13.9 Solution of Exercise 2.4.1

```
> beta.error.exact(1,15,15)
[1] 0.2593674
> beta.error.exact(1,16,14)
[1] 0.2612445
> beta.error.exact(1,17,13)
[1] 0.2669429
> beta.error.exact(1,18,12)
[1] 0.2766668
> beta.error.exact(1,19,11)
[1] 0.2907635
> beta.error.exact(1,20,10)
[1] 0.3097332
> beta.error.exact(1,25,5)
[1] 0.5067606
> beta.error.exact(1,29,1)
```

```
[1] 0.8450542
> beta.error.exact(3,15,15)
[1] 2.888766e-09
> beta.error.exact(3,16,14)
[1] 3.201754e-09
> beta.error.exact(3,17,13)
[1] 4.35778e-09
> beta.error.exact(3,20,10)
[1] 3.71341e-08
> beta.error.exact(3,25,5)
[1] 5.898017e-05
> beta.error.exact(3,29,1)
[1] 0.19891
```

The parameter δ has no influence. The best choice of N_1 and N_2 is $N_1 = 15$ and $N_2 = 15$. One can even prove that the minimum β-error is always attained by $N_1 = \frac{N}{2} = N_2$.

13.10 Solution of Exercise 2.4.2

```
> N<-10
> beta.error.exact(2,N/2,N/2)
[1] 0.1474736
> N<-100
> beta.error.exact(2,N/2,N/2)
[1] 3.164278e-14
> N<-50
> beta.error.exact(2,N/2,N/2)
[1] 1.045841e-06
> N<-20
> beta.error.exact(2,N/2,N/2)
[1] 0.01104945
> N<-18
> beta.error.exact(2,N/2,N/2)
[1] 0.01916283
> N<-16
> beta.error.exact(2,N/2,N/2)
[1] 0.03277304
> N<-14
> beta.error.exact(2,N/2,N/2)
[1] 0.05516538
> N<-15
> beta.error.exact(2,N/2,N/2)
[1] 0.04261038
> beta.error.exact(2,7,8)
[1] 0.04335619
> beta.error.exact(2,7,7)
[1] 0.05516538
```

Hence the minimum sample size is $N = 15 = 7+8$ for $|\mu_1 - \mu_2| > 2\sigma$. We see that the t-distribution is also defined for degrees of freedom which are not integers. With the same procedure we obtain

```
> N<-230
> beta.error.exact(1/2,N/2,N/2)
[1] 0.05009967
> N<-231
> beta.error.exact(1/2,N/2,N/2)
[1] 0.0492988
> beta.error.exact(1/2,115,116)
[1] 0.04930224
```

Hence the minimum sample size is $N = 231 = 115 + 116$ for $|\mu_1 - \mu_2| > \sigma/2$.

13.11 Solution of Exercise 3.2.3

As for the uncovered boxes, we can expect that the normal distribution is not rejected based on four measurements so that we have only to test the homogeneity of the variances:

```
> germin.c<-germin[germin$box=="covered",]
> bartlett.test(seed.numbers~watering,data=germin.c[germin.c$watering!="6",])
        Bartlett test for homogeneity of variances
data:  seed.numbers by watering
Bartlett's K-squared = 6.6995, df = 4, p-value = 0.1526
```

Since the homogeneity of the variances is not rejected, we can apply the ANOVA test:

```
> anova(lm(seed.numbers~watering,data=germin.c[germin.c$watering!="6",]))
Analysis of Variance Table

Response: seed.numbers
          Df Sum Sq Mean Sq F value    Pr(>F)
watering   4 4839.9  1210.0  31.016 8.037e-07 ***
Residuals 14  546.2    39.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With

```
> var(germin.c[germin.c$watering!="6","seed.numbers"])
[1] 299.2281
> 18*var(germin.c[germin.c$watering!="6","seed.numbers"])
[1] 5386.105
```

we obtain the complete ANOAVA table:

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
Differences between factor levels	4	$\Sigma_{SST} = 4839.9$	$\widehat{\sigma}^2_{SST} = 1210.0$
Measurement error	14	$\Sigma_{SSE} = 546.2$	$\widehat{\sigma}^2_{SSE} = 39.0$
Total	18	$\Sigma_{SSG} = 5386.1$	$\widehat{\sigma}^2_{SSG} = 299.2281$

Since the p-value is 8.037e-07, there is again a significant watering effect.

13.12 Solution of Exercise 3.2.4

```
> library(agricolae)
> data(trees)
> shapiro.test(trees[trees$species=="GUABA","diameter"])$p.value
[1] 0.1548390
> shapiro.test(trees[trees$species=="LAUREL","diameter"])$p.value
[1] 0.2396723
> shapiro.test(trees[trees$species=="ROBLE","diameter"])$p.value
[1] 0.1887661
> shapiro.test(trees[trees$species=="TERMINALIA","diameter"])$p.value
[1] 0.8486412
> bartlett.test(diameter~species,data=trees)$p.value
[1] 0.1361938
```

Hence the assumptions of the ANOVA test are not rejected.

```
> anova(lm(diameter~species,data=trees))
Analysis of Variance Table

Response: diameter
          Df  Sum Sq Mean Sq F value  Pr(>F)
species    3  79.097  26.366  4.2975 0.01787 *
Residuals 19 116.569   6.135
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can conclude that the four species have significantly different stem diameters.

```
> kruskal.test(diameter~species,data=trees)
```

```
        Kruskal-Wallis rank sum test

data:  diameter by species
Kruskal-Wallis chi-squared = 7.8414, df = 3, p-value = 0.04941
```

Again the distribution-free test provides a larger p-value although it is still less 0.05.

13.13 Solution of Exercise 3.3.2

```
> germin.c<-germin[germin$box=="covered",]
> TukeyHSD(aov(seed.numbers~watering,data=germin.c[germin.c$watering!="6",]))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = seed.numbers ~ watering, data = germin.c[germin.c$watering != "6", ])

$watering
          diff        lwr         upr     p adj
2-1  32.500000  18.738266  46.2617340 0.0000301
3-1  33.250000  19.488266  47.0117340 0.0000233
4-1   9.250000  -4.511734  23.0117340 0.2753769
5-1  -5.416667 -20.281038   9.4477049 0.7857254
3-2   0.750000 -13.011734  14.5117340 0.9997884
4-2 -23.250000 -37.011734  -9.4882660 0.0009555
5-2 -37.916667 -52.781038 -23.0522951 0.0000126
4-3 -24.000000 -37.761734 -10.2382660 0.0007077
5-3 -38.666667 -53.531038 -23.8022951 0.0000100
5-4 -14.666667 -29.531038   0.1977049 0.0538955
```

There are only significant differences for the watering levels 1-2, 1-3, 2-4, 2-5, 3-4, 3-5. The same result is obtained via `HSD.test` from the newest version of the `agricolae` package:

```
> library(agricolae)
> attach(germin.c[germin.c$watering!="6",])
> model<-aov(seed.numbers~watering)
> df<-df.residual(model)
> MSerror<-deviance(model)/df
> comparison<-HSD.test(seed.numbers,watering,df,MSerror,group=TRUE,main="title")
> comparison
  trt    means   M    N   std.err
1   3 76.00000   a 3.75 1.7795130
2   2 75.25000   a 3.75 3.4731110
3   4 52.00000   b 3.75 4.6007246
4   1 42.75000  bc 3.75 0.8539126
5   5 37.33333   c 3.75 4.0960686
```

The Waller-Duncan method provides:

```
> Fc<-anova(model)[1,4]
> comparison<-waller.test(seed.numbers,watering,df,MSerror,
+ Fc,group=TRUE,main="title")
> comparison
  trt    means   M    N   std.err
1   3 76.00000   a 3.75 1.7795130
2   2 75.25000   a 3.75 3.4731110
3   4 52.00000   b 3.75 4.6007246
4   1 42.75000   c 3.75 0.8539126
5   5 37.33333   c 3.75 4.0960686
```

Here additionally the watering levels 1-4 show significant differences.

```
> boxplot(seed.numbers~watering,data=germin.c[germin.c$watering!="6",])
```

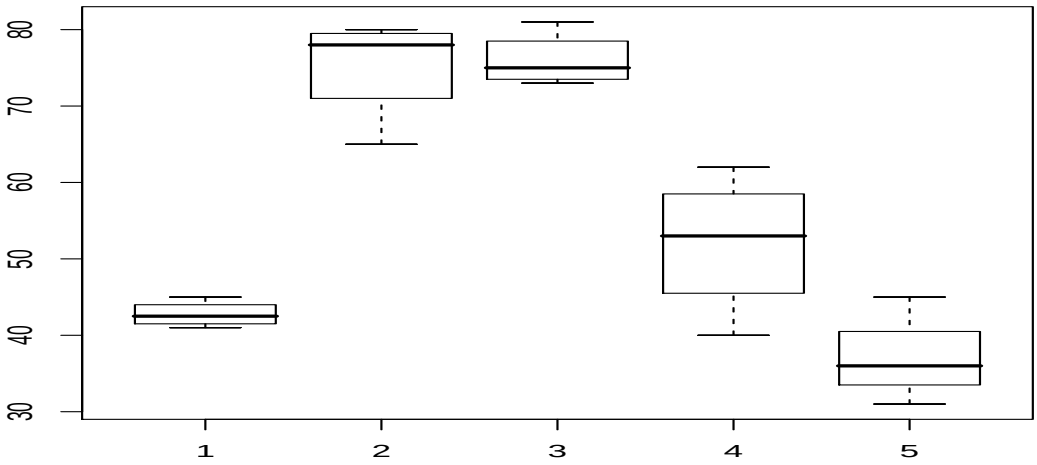

Figure 13.5: Box plots for the covered boxes

```
> gerS.c<-germin.c[,"seed.numbers"]
> gerW.c<-germin.c[,"watering"]
> t.test(gerS.c[gerW.c=="1"],gerS.c[gerW.c=="2"])$p.value
[1] 0.001763613
> t.test(gerS.c[gerW.c=="1"],gerS.c[gerW.c=="3"])$p.value
[1] 4.241045e-05
> t.test(gerS.c[gerW.c=="1"],gerS.c[gerW.c=="4"])$p.value
[1] 0.1365746
> t.test(gerS.c[gerW.c=="1"],gerS.c[gerW.c=="5"])$p.value
[1] 0.3158338
> t.test(gerS.c[gerW.c=="2"],gerS.c[gerW.c=="3"])$p.value
[1] 0.8560103
> t.test(gerS.c[gerW.c=="2"],gerS.c[gerW.c=="4"])$p.value
```

```
[1] 0.007963777
> t.test(gerS.c[gerW.c=="2"],gerS.c[gerW.c=="5"])$p.value
[1] 0.001479392
> t.test(gerS.c[gerW.c=="3"],gerS.c[gerW.c=="4"])$p.value
[1] 0.008917755
> t.test(gerS.c[gerW.c=="3"],gerS.c[gerW.c=="5"])$p.value
[1] 0.004431519
> t.test(gerS.c[gerW.c=="4"],gerS.c[gerW.c=="5"])$p.value
[1] 0.06346417
```

With the significant level $0.05/10 = 0.005$, we obtain only significant differences for the watering levels 1-2, 1-3, 2-5, 3-5. Again we obtain less significant differences than with Tukey's Honest Significant Difference method and the Waller-Duncan method.

13.14 Solution of Exercise 3.3.3

```
> library(agricolae)
> data(trees)
> TukeyHSD(aov(diameter~species,data=trees))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diameter ~ species, data = trees)

$species
                       diff       lwr      upr     p adj
LAUREL-GUABA      1.8733333 -2.344033 6.090700 0.6047337
ROBLE-GUABA       2.3500000 -1.671101 6.371101 0.3795158
TERMINALIA-GUABA  5.0833333  1.062232 9.104435 0.0104236
ROBLE-LAUREL      0.4766667 -3.740700 4.694033 0.9885329
TERMINALIA-LAUREL 3.2100000 -1.007367 7.427367 0.1764053
TERMINALIA-ROBLE  2.7333333 -1.287768 6.754435 0.2565468

> boxplot(diameter~species,data=trees)
```

Only the species Terminalia and Guaba have significantly different stem diameters.

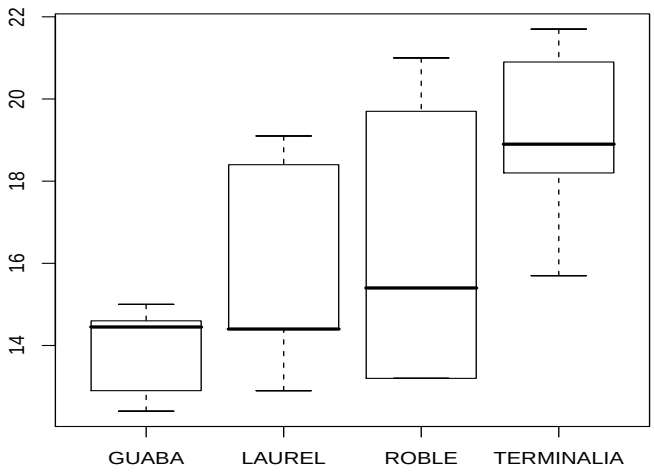

Figure 13.6: Box plots for the stem diameters

13.15 Solution of Exercise 3.4.1

Since we have 12 units and 4 treatments, the replication number of each treatment is 3. Then we obtain for example the following design:

```
> library(agricolae)
> design.crd(c("TR1","TR2","TR3","TR4"),3)
   plots c("TR1", "TR2", "TR3", "TR4") r
1      1                           TR2 1
2      2                           TR3 1
3      3                           TR1 1
4      4                           TR3 2
5      5                           TR4 1
6      6                           TR1 2
7      7                           TR4 2
8      8                           TR4 3
9      9                           TR1 3
10    10                           TR2 2
11    11                           TR3 3
12    12                           TR2 3
```

13.16 Solution of Exercise 4.3.3

a)

```
> shapiro.test(lm(length~cutting*grow.conditions,data=mustard)$residuals)$p.value
[1] 0.5210626
```

Hence the ANOVA test can be used:

```
> anova(lm(length~cutting*grow.conditions,data=mustard))
Analysis of Variance Table

Response: length
                       Df Sum Sq Mean Sq F value  Pr(>F)
cutting                 1  559.8   559.8  4.5198 0.04280 *
grow.conditions         1  109.5   109.5  0.8842 0.35539
cutting:grow.conditions  1  516.0   516.0  4.1660 0.05113 .
Residuals              27 3344.3   123.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence only the cutting has a significant influence on the length. The hypothesis of no interactions between cutting and growing conditions is not rejected.

The ANOVA table is obtained as follows:

```
> length(mustard[,"length"])
[1] 31
> var(mustard[,"length"])
[1] 150.9892
> 30*var(mustard[,"length"])
[1] 4529.677
```

Cause of variability	Degrees of freedom	Sum of squares	Variance estimates
cutting	1	$\Sigma_{SSA} = 559.8$	$\widehat{\sigma}^2_{SSA} = 559.8$
grow.conditions	1	$\Sigma_{SSB\|A+B} = 109.5$	$\widehat{\sigma}^2_{SSB\|A+B} = 109.5$
Interaction	1	$\Sigma_{SSI} = 516.0$	$\widehat{\sigma}^2_{SSI} = 516.0$
Measurement error	27	$\Sigma_{SSE} = 3344.3$	$\widehat{\sigma}^2_{SSE} = 123.9$
Total	30	$\Sigma_{SSG} = 4529.677$	$\widehat{\sigma}^2_{SSG} = 150.9892$

b)

```
> anova(lm(length~grow.conditions*cutting,data=mustard))
Analysis of Variance Table
```

```
Response: length
                        Df Sum Sq Mean Sq F value  Pr(>F)
grow.conditions          1   88.8    88.8  0.7172 0.40449
cutting                  1  580.5   580.5  4.6868 0.03940 *
grow.conditions:cutting  1  516.0   516.0  4.1660 0.05113 .
Residuals               27 3344.3   123.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here the order of the treatments is important since the design is very unbalanced with samples sizes N_{ab} of 10, 6, 10, and 5.

c) Since the hypothesis of no interactions is not rejected, the two-way layout without interactions can be used. At first the assumption of normal distribution is tested:

```
> shapiro.test(lm(length~grow.conditions+cutting,data=mustard)$residuals)$p.value
[1] 0.2221986
```

The p-value is now smaller than in the model with interactions but still greater than 0.05. Hence the ANOVA test can be used.

```
> anova(lm(length~grow.conditions+cutting,data=mustard))
Analysis of Variance Table

Response: length
                Df Sum Sq Mean Sq F value  Pr(>F)
grow.conditions  1   88.8    88.8  0.6444 0.42889
cutting          1  580.5   580.5  4.2107 0.04963 *
Residuals       28 3860.3   137.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test statistics, F values, are smaller and the p-values are larger. This is due to the fact that $(516.0 + 3344.3)/28 = 3860.3/28 = 137.9$ instead of $3344.3/27 = 123.9$ is used in the denominator of the test statistic.

13.17 Solution of Exercise 4.5.4

a)

```
> design.ab(1:6,1:4,2)
   plots block 1:6 1:4
1      1     1   2   3
2      2     1   4   1
3      3     1   3   4
```

```
4       4       1   1   2
5       5       1   2   4
6       6       1   5   4
7       7       1   6   4
8       8       1   2   2
9       9       1   2   1
10     10       1   6   3
11     11       1   4   2
12     12       1   1   4
13     13       1   3   1
14     14       1   6   1
15     15       1   3   2
16     16       1   4   4
17     17       1   5   2
18     18       1   1   3
19     19       1   4   3
20     20       1   3   3
21     21       1   5   3
22     22       1   1   1
23     23       1   6   2
24     24       1   5   1
25     25       2   1   1
26     26       2   3   2
27     27       2   4   2
28     28       2   1   2
29     29       2   3   1
30     30       2   6   4
31     31       2   2   2
32     32       2   2   4
33     33       2   3   4
34     34       2   1   3
35     35       2   4   3
36     36       2   4   4
37     37       2   2   3
38     38       2   1   4
39     39       2   5   4
40     40       2   3   3
41     41       2   6   1
42     42       2   5   2
43     43       2   6   3
44     44       2   5   3
45     45       2   6   2
46     46       2   2   1
47     47       2   4   1
48     48       2   5   1
```

b)

```
> design.rcbd(1:6,4)
   plots block 1:6
1      1     1   3
2      2     1   5
3      3     1   4
4      4     1   2
5      5     1   1
6      6     1   6
7      7     2   3
8      8     2   1
9      9     2   2
10    10     2   4
11    11     2   5
12    12     2   6
13    13     3   4
14    14     3   3
15    15     3   2
16    16     3   1
17    17     3   6
18    18     3   5
19    19     4   6
20    20     4   2
21    21     4   3
22    22     4   4
23    23     4   5
24    24     4   1
```

c) There exists no balanced incomplete block design since condition (9) means $6 \cdot r = 4 \cdot 3$ which implies $r = 2$. Then condition (10) means $\lambda \cdot 5 = 2 \cdot 2$. But there exists no λ in the integers which satisfies this equation.
d)

```
> design.bib(1:6,3)

Parameters BIB
==============
Lambda     : 4
treatmeans : 6
Block size : 3
Blocks     : 20
Replication: 10

Efficiency factor 0.8

<<< Book >>>
   plots block 1:6
1      1     1   1
2      2     1   5
```

3	3	1	6
4	4	2	1
5	5	2	3
6	6	2	5
7	7	3	4
8	8	3	2
9	9	3	5
10	10	4	1
11	11	4	2
12	12	4	4
13	13	5	2
14	14	5	3
15	15	5	5
16	16	6	5
17	17	6	4
18	18	6	3
19	19	7	6
20	20	7	1
21	21	7	4
22	22	8	4
23	23	8	3
24	24	8	2
25	25	9	6
26	26	9	3
27	27	9	4
28	28	10	4
29	29	10	1
30	30	10	3
31	31	11	6
32	32	11	2
33	33	11	5
34	34	12	6
35	35	12	2
36	36	12	4
37	37	13	5
38	38	13	6
39	39	13	3
40	40	14	6
41	41	14	4
42	42	14	5
43	43	15	1
44	44	15	2
45	45	15	5
46	46	16	3
47	47	16	6
48	48	16	1
49	49	17	1
50	50	17	2

```
51    51    17   6
52    52    18   2
53    53    18   3
54    54    18   1
55    55    19   4
56    56    19   5
57    57    19   1
58    58    20   3
59    59    20   6
60    60    20   2
```

13.18 Solution of Exercise 4.5.5

```
> pepper.design<-design.bib(1:8,6)

Parameters BIB
==============
Lambda     : 15
treatmeans : 8
Block size : 6
Blocks     : 28
Replication: 21

Efficiency factor 0.952381

<<< Book >>>
```

$168 = 8 * 21 = 6 * 28$ blocks are needed, which means that the experiment would have a duration of 84 years.

```
> anova(lm(yield~Block*Clon*Treat, data=huasahuasi))
Analysis of Variance Table

Response: yield
                 Df Sum Sq Mean Sq F value Pr(>F)
Block             2    7.5     3.8
Clon              4 5435.8  1359.0
Treat             2  280.6   140.3
Block:Clon        8  299.7    37.5
Block:Treat       4   65.8    16.4
Clon:Treat        8  194.1    24.3
Block:Clon:Treat 16  280.8    17.6
Residuals         0    0.0
```

13.19 Solution of Exercise 5.1.2

```
> aov(Excess~Year+Block*Heating*Lighting*CO2,data=pepper)
Call:
   aov(formula = Excess ~ Year + Block * Heating * Lighting * CO2,
    data = pepper)

Terms:
                    Year    Block  Heating Lighting      CO2 Block:Heating
Sum of Squares  24.40167 92.04167  7.62881 20.36507  1.19428       0.03541
Deg. of Freedom        1        1        1        1        1             1
                Block:Lighting Heating:Lighting Block:CO2 Heating:CO2
Sum of Squares         0.26877          0.22369   0.74872     0.02326
Deg. of Freedom              1                1         1           1
                Lighting:CO2 Block:Heating:Lighting Block:Heating:CO2
Sum of Squares       0.14940                1.65698           0.26978
Deg. of Freedom            1                      1                 1
                Block:Lighting:CO2 Residuals
Sum of Squares             1.32250   2.91000
Deg. of Freedom                  1         9

Residual standard error: 0.5686241
2 out of 17 effects not estimable
Estimated effects may be unbalanced
```

In this model not all parameters are estimable. The interactions `Heating*Lighting*CO2` and `Block*Heating*Lighting*CO2` are not estimable.

```
> aov(Excess~Year*Block+Heating*Lighting*CO2,data=pepper)
Call:
   aov(formula = Excess ~ Year * Block + Heating * Lighting * CO2,
    data = pepper)

Terms:
                    Year    Block  Heating Lighting      CO2 Year:Block
Sum of Squares  24.40167 92.04167  7.62881 20.36507  1.19428    0.17964
Deg. of Freedom        1        1        1        1        1          1
                Heating:Lighting Heating:CO2 Lighting:CO2 Heating:Lighting:CO2
Sum of Squares           0.11168     0.04689      0.13298               1.17600
Deg. of Freedom                1           1            1                     1
                Residuals
Sum of Squares    5.96131
Deg. of Freedom        13

Residual standard error: 0.6771723
Estimated effects may be unbalanced
> shapiro.test(aov(Excess~Year*Block+Heating*Lighting*CO2,data=pepper)$residuals
```

```
+ )$p.value
[1] 0.6033543
> anova(lm(Excess~Year*Block+Heating*Lighting*CO2,data=pepper))
Analysis of Variance Table

Response: Excess
                    Df Sum Sq Mean Sq  F value    Pr(>F)
Year                 1 24.402  24.402  53.2134 6.050e-06 ***
Block                1 92.042  92.042 200.7179 2.787e-09 ***
Heating              1  7.629   7.629  16.6364  0.001304 **
Lighting             1 20.365  20.365  44.4107 1.554e-05 ***
CO2                  1  1.194   1.194   2.6044  0.130567
Year:Block           1  0.180   0.180   0.3918  0.542213
Heating:Lighting     1  0.112   0.112   0.2436  0.629886
Heating:CO2          1  0.047   0.047   0.1022  0.754228
Lighting:CO2         1  0.133   0.133   0.2900  0.599329
Heating:Lighting:CO2 1  1.176   1.176   2.5645  0.133294
Residuals           13  5.961   0.459
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> aov(Excess~Year+Block+Heating+Lighting+CO2,data=pepper)
Call:
   aov(formula = Excess ~ Year + Block + Heating + Lighting + CO2,
    data = pepper)

Terms:
                    Year    Block  Heating Lighting      CO2 Residuals
Sum of Squares  24.40167 92.04167  7.62881 20.36507  1.19428   7.60850
Deg. of Freedom        1        1        1        1        1        18

Residual standard error: 0.6501496
Estimated effects may be unbalanced
> shapiro.test(aov(Excess~Year+Block+Heating+Lighting+CO2,data=pepper)$residuals
+ )$p.value
[1] 0.5144351
> anova(lm(Excess~Year+Block+Heating+Lighting+CO2,data=pepper))
Analysis of Variance Table

Response: Excess
          Df Sum Sq Mean Sq  F value    Pr(>F)
Year       1 24.402  24.402  57.7289 5.067e-07 ***
Block      1 92.042  92.042 217.7499 1.695e-11 ***
Heating    1  7.629   7.629  18.0480 0.0004835 ***
Lighting   1 20.365  20.365  48.1792 1.739e-06 ***
CO2        1  1.194   1.194   2.8254 0.1100532
Residuals 18  7.609   0.423
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In practice, the model `Excess~Year*Block+Heating*Lighting*CO2` should be used since it allows the most interactions, even one more interaction, namely the interaction between `Year` and `Block`, than the model `Excess~Year+Block+Heating*Lighting*CO2` used in Example 5.1.1. Since no interaction has a significant effect, also the additive model `Excess~Year+Block+Heating+Lighting+CO2` might be used. But that the interactions are not significant, cannot be known before analyzing the data. Hence the additive model should be not used.

13.20 Solution of Exercise 5.1.3

```
> aov(yield~Block*Clon*Treat,data=huasahuasi)
Call:
   aov(formula = yield ~ Block * Clon * Treat, data = huasahuasi)

Terms:
                   Block     Clon    Treat Block:Clon Block:Treat Clon:Treat
Sum of Squares     7.509 5435.843  280.645    299.699      65.754    194.067
Deg. of Freedom        2        4        2          8           4          8
                Block:Clon:Treat
Sum of Squares           280.840
Deg. of Freedom               16

Estimated effects may be unbalanced
```

Hence in the largest possible model, which includes all interactions, all parameters are estimable.

```
> shapiro.test(aov(yield~Block*Clon*Treat,data=huasahuasi)$residuals)$p.value
Fehler in shapiro.test(aov(yield ~ Block * Clon * Treat, data = huasahuasi)$residuals) :
        all 'x' values are identical
> anova(lm(yield~Block*Clon*Treat,data=huasahuasi))
Analysis of Variance Table

Response: yield
                 Df Sum Sq Mean Sq F value Pr(>F)
Block             2    7.5     3.8
Clon              4 5435.8  1359.0
Treat             2  280.6   140.3
Block:Clon        8  299.7    37.5
Block:Treat       4   65.8    16.4
Clon:Treat        8  194.1    24.3
Block:Clon:Treat 16  280.8    17.6
Residuals         0    0.0
```

Since there are $3 \cdot 5 \cdot 3 = 45$ parameter and only 45 observations, all hypotheses of the largest model are not testable. This indicates also the residual sum of squares (the sum of squares for errors

σ^2_{SSE}) which is zero with zero degree of freedom. Therefore the Shapiro-Wilk test provides an error message.

```
> aov(yield~Block+Clon*Treat,data=huasahuasi)
Call:
   aov(formula = yield ~ Block + Clon * Treat, data = huasahuasi)

Terms:
                   Block     Clon    Treat Clon:Treat Residuals
Sum of Squares     7.509 5435.843  280.645    194.067   646.293
Deg. of Freedom        2        4        2          8        28

Residual standard error: 4.804362
Estimated effects may be unbalanced
> aov(yield~Block*Clon+Treat,data=huasahuasi)
Call:
   aov(formula = yield ~ Block * Clon + Treat, data = huasahuasi)

Terms:
                   Block     Clon    Treat Block:Clon Residuals
Sum of Squares     7.509 5435.843  280.645    299.699   540.662
Deg. of Freedom        2        4        2          8        28

Residual standard error: 4.394239
Estimated effects may be unbalanced
```

Both models have $2 + 4 + 2 + 4 * 2 = 16$ parameters. The model with interactions could be more reasonable because an interaction between treatment and clone may be of interest. However, the normal distribution is rejected in this model:

```
> shapiro.test(aov(yield~Block+Clon*Treat,data=huasahuasi)$residuals)$p.value
[1] 0.04501483
> shapiro.test(aov(yield~Block*Clon+Treat,data=huasahuasi)$residuals)$p.value
[1] 0.2980998
```

Hence the model with interactions between blocks and clones should be used.

```
> anova(lm(yield~Block*Clon+Treat,data=huasahuasi))
Analysis of Variance Table

Response: yield
           Df Sum Sq Mean Sq F value    Pr(>F)
Block       2    7.5     3.8  0.1944   0.82439
Clon        4 5435.8  1359.0 70.3784 3.377e-14 ***
Treat       2  280.6   140.3  7.2671   0.00287 **
Block:Clon  8  299.7    37.5  1.9401   0.09316 .
Residuals  28  540.7    19.3
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Set B =Block, C =Clon, and T =Treat.

$$1)\ H_0^{B*C}: \quad \mu_{bct} = \mu + B + C + T$$
$$\text{versus}$$
$$H_1^{B*C}: \quad \mu_{bct} = \mu + B + C + T + B*C$$

H_0^{B*C} is not rejected.

$$2)\ H_0^{T}: \quad \mu_{bct} = \mu + B + C$$
$$\text{versus}$$
$$H_1^{T}: \quad \mu_{bct} = \mu + B + C + T$$

There is a significant treatment effect.

$$3)\ H_0^{C}: \quad \mu_{bct} = \mu + B$$
$$\text{versus}$$
$$H_1^{C}: \quad \mu_{bct} = \mu + B + C$$

There is a significant clone effect.

$$4)\ H_0^{B}: \quad \mu_{bct} = \mu$$
$$\text{versus}$$
$$H_1^{B}: \quad \mu_{bct} = \mu + B$$

There is a no significant block effect.

13.21 Solution of Exercise 5.2.1

```
> design.graeco(c("A","B","C","D","E"),c("a","b","c","d","e"))
   plots row col c("A", "B", "C", "D", "E") c("a", "b", "c", "d", "e")
1      1   1   1                          B                          c
2      2   1   2                          E                          d
3      3   1   3                          C                          b
4      4   1   4                          A                          e
5      5   1   5                          D                          a
6      6   2   1                          E                          b
7      7   2   2                          C                          e
8      8   2   3                          A                          a
9      9   2   4                          D                          c
10    10   2   5                          B                          d
```

```
11    11   3   1                    C                    a
12    12   3   2                    A                    c
13    13   3   3                    D                    d
14    14   3   4                    B                    b
15    15   3   5                    E                    e
16    16   4   1                    A                    d
17    17   4   2                    D                    b
18    18   4   3                    B                    e
19    19   4   4                    E                    a
20    20   4   5                    C                    c
21    21   5   1                    D                    e
22    22   5   2                    B                    a
23    23   5   3                    E                    c
24    24   5   4                    C                    d
25    25   5   5                    A                    b
```

	1	2	3	4	5
1	Bc	Ed	Cb	Ae	Da
2	Eb	Ce	Aa	Dc	Bd
3	Ca	Ac	Dd	Bb	Ee
4	Ad	Db	Be	Ea	Cc
5	De	Ba	Ec	Cd	Ab

13.22 Solution of Exercise 5.2.2

```
> vandal0<-read.table("VANDAL.DAT",na.string="Ä")
> vandal1<-as.factor(c(vandal0[ ,1],vandal0[ ,3],vandal0[ ,5],
+ vandal0[ ,7],vandal0[ ,9],vandal0[ ,11]))
> levels(vandal1)<-c("A","B","C","D","E","F")
> vandal2<-c(vandal0[ ,2],vandal0[ ,4],vandal0[ ,5],vandal0[ ,8],
+ vandal0[ ,10],vandal0[ ,12])
> vandal3<-rep(as.factor(1:6),6)
> vandal4<-as.factor(c(rep(1,6),rep(2,6),rep(3,6),rep(4,6),rep(5,6),rep(6,6)))
> vandal<-data.frame(vandal3,vandal4,vandal1,vandal2)
> names(vandal)<-c("row","col","treat","weight")
> vandal[1:3,]
  row col treat weight
1   1   1     E   29.0
2   2   1     B   17.5
3   3   1     F   17.0
> shapiro.test(aov(weight~row+col+treat,data=vandal)$residuals)$p.value
[1] 0.1449412
> anova(lm(weight~row+col+treat,data=vandal))
Analysis of Variance Table

Response: weight
```

```
          Df  Sum Sq Mean Sq F value    Pr(>F)
row        5  232.15   46.43  1.6471  0.201324
col        5 1651.86  330.37 11.7199 5.066e-05 ***
treat      5  672.78  134.56  4.7734  0.006592 **
Residuals 17  479.21   28.19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(weight~col+treat+row,data=vandal))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
col        5 1636.95  327.39 11.6142 5.363e-05 ***
treat      5  741.19  148.24  5.2588  0.004239 **
row        5  178.64   35.73  1.2674  0.322716
Residuals 17  479.21   28.19
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because of the missing values the design is not balanced and therefore the order of the factors matters. However, adding the missing value, provides a balanced design:

```
> vandal.complete<-vandal
> vandal.complete[30,"weight"]<-21.5
> vandal.complete[35,"weight"]<-20.8
> vandal.complete[36,"weight"]<-13.5
> anova(lm(weight~row+col+treat,data=vandal.complete))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
row        5  185.93   37.19  1.4000  0.266699
col        5 1631.66  326.33 12.2860 1.552e-05 ***
treat      5  731.56  146.31  5.5085  0.002393 **
Residuals 20  531.23   26.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(weight~col+row+treat,data=vandal.complete))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
col        5 1631.66  326.33 12.2860 1.552e-05 ***
row        5  185.93   37.19  1.4000  0.266699
treat      5  731.56  146.31  5.5085  0.002393 **
Residuals 20  531.23   26.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(lm(weight~row+treat+col,data=vandal.complete))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
row        5  185.93   37.19  1.4000  0.266699
treat      5  731.56  146.31  5.5085  0.002393 **
col        5 1631.66  326.33 12.2860 1.552e-05 ***
Residuals 20  531.23   26.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(weight~col+treat+row,data=vandal.complete))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
col        5 1631.66  326.33 12.2860 1.552e-05 ***
treat      5  731.56  146.31  5.5085  0.002393 **
row        5  185.93   37.19  1.4000  0.266699
Residuals 20  531.23   26.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(weight~treat+row+col,data=vandal.complete))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
treat      5  731.56  146.31  5.5085  0.002393 **
row        5  185.93   37.19  1.4000  0.266699
col        5 1631.66  326.33 12.2860 1.552e-05 ***
Residuals 20  531.23   26.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(weight~treat+col+row,data=vandal.complete))
Analysis of Variance Table

Response: weight
          Df  Sum Sq Mean Sq F value    Pr(>F)
treat      5  731.56  146.31  5.5085  0.002393 **
col        5 1631.66  326.33 12.2860 1.552e-05 ***
row        5  185.93   37.19  1.4000  0.266699
Residuals 20  531.23   26.56
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the complete latin square design, the order of the factors does not matter.

13.23 Solution of Exercise 6.1.1

At first we regard $\sigma^2_{\beta_0}$. Since the variance σ^2 does not matter, we calculate only $\frac{1}{(N-1)s_x^2}$:

```
> x<-c(0,0.1,0.2,0.3,0.4,0.5,0.5,0.6,0.7,0.8,0.9,1)
> 1/(9*var(x))
[1] 1.111111
> x<-c(0,0,0.2,0.2,0.4,0.4,0.6,0.6,0.8,0.8,1,1)
> 1/(9*var(x))
[1] 0.8730159
> x<-c(0,0,0,0.3,0.3,0.3,0.7,0.7,0.7,1,1,1)
> 1/(9*var(x))
[1] 0.7024266
> x<-c(0,0,0,0.5,0.5,0.5,0.5,0.5,0.5,1,1,1)
> 1/(9*var(x))
[1] 0.8148148
> x<-c(0,0,0,0,0.5,0.5,0.5,0.5,1,1,1,1)
> 1/(9*var(x))
[1] 0.6111111
> x<-c(0,0,0,0,0,0.5,0.5,1,1,1,1,1)
> 1/(9*var(x))
[1] 0.4888889
> x<-c(0,0,0,0,0,0,1,1,1,1,1,1)
> 1/(9*var(x))
[1] 0.4074074
```

The proposal is that the design with smallest variance $\sigma^2_{\beta_0}$ is the design which puts half of the observations at 0 and the other half of the observations at 1. For $\sigma^2_{\beta_1}$ we calculate $\frac{\frac{1}{N}\sum_{n=1}^{N} x_n^2}{(N-1)s_x^2}$:

```
> x<-c(0,0.1,0.2,0.3,0.4,0.5,0.5,0.6,0.7,0.8,0.9,1)
> mean(x^2)/(9*var(x))
[1] 0.3796296
> x<-c(0,0,0.2,0.2,0.4,0.4,0.6,0.6,0.8,0.8,1,1)
> mean(x^2)/(9*var(x))
[1] 0.3201058
> x<-c(0,0,0,0.3,0.3,0.3,0.7,0.7,0.7,1,1,1)
> mean(x^2)/(9*var(x))
[1] 0.2774585
> x<-c(0,0,0,0.5,0.5,0.5,0.5,0.5,0.5,1,1,1)
> mean(x^2)/(9*var(x))
[1] 0.3055556
> x<-c(0,0,0,0,0.5,0.5,0.5,0.5,1,1,1,1)
> mean(x^2)/(9*var(x))
[1] 0.2546296
> x<-c(0,0,0,0,0,0.5,0.5,1,1,1,1,1)
> mean(x^2)/(9*var(x))
[1] 0.2240741
```

```
> x<-c(0,0,0,0,0,0,1,1,1,1,1,1)
> mean(x^2)/(9*var(x))
[1] 0.2037037
```

Again, the proposal is that the design with smallest variance $\sigma^2_{\beta_1}$ is the design which puts half of the observations at 0 and the other half of the observations at 1.

13.24 Solution of Exercise 6.1.3

At first the normal distributions are checked:

```
> shapiro.test(aov(protein~L2,data=ground)$residuals)$p.value
[1] 0.8573912
> shapiro.test(aov(protein~L3,data=ground)$residuals)$p.value
[1] 0.8441344
> shapiro.test(aov(protein~L4,data=ground)$residuals)$p.value
[1] 0.3232782
> shapiro.test(aov(protein~L5,data=ground)$residuals)$p.value
[1] 0.366428
> shapiro.test(aov(protein~L6,data=ground)$residuals)$p.value
[1] 0.6450972
```

Hence the t- and F-tests can be performed.

```
> summary(lm(protein~L2,data=ground))

Call:
lm(formula = protein ~ L2, data = ground)

Residuals:
     Min       1Q   Median       3Q      Max
-2.02800 -0.75285  0.09001  1.02295  2.40733

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.089647   1.274136   4.779    9e-05 ***
L2          0.027591   0.008897   3.101  0.00521 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.209 on 22 degrees of freedom
Multiple R-Squared: 0.3042,     Adjusted R-squared: 0.2726
F-statistic: 9.618 on 1 and 22 DF,  p-value: 0.005209

> summary(lm(protein~L3,data=ground))
```

```
Call:
lm(formula = protein ~ L3, data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0385 -0.7785  0.1001  0.9945  2.4207

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.320697   2.237690   1.484  0.15200
L3          0.025121   0.008406   2.988  0.00677 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.223 on 22 degrees of freedom
Multiple R-Squared: 0.2887,     Adjusted R-squared: 0.2564
F-statistic: 8.931 on 1 and 22 DF,  p-value: 0.006773

> summary(lm(protein~L4,data=ground))

Call:
lm(formula = protein ~ L4, data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1671 -1.0127  0.1710  0.9547  2.3899

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.02711    3.58059   0.845   0.4070
L4           0.01776    0.00914   1.944   0.0648 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.339 on 22 degrees of freedom
Multiple R-Squared: 0.1466,     Adjusted R-squared: 0.1078
F-statistic: 3.778 on 1 and 22 DF,  p-value: 0.06484

> summary(lm(protein~L5,data=ground))

Call:
lm(formula = protein ~ L5, data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4576 -1.1006  0.1763  1.1459  2.3939

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.381606   2.557850   2.104   0.0470 *
L5          0.011454   0.006353   1.803   0.0851 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.353 on 22 degrees of freedom
Multiple R-Squared: 0.1287,     Adjusted R-squared: 0.08913
F-statistic: 3.251 on 1 and 22 DF,  p-value: 0.08512

> summary(lm(protein~L6,data=ground))

Call:
lm(formula = protein ~ L6, data = ground)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2824 -0.9135  0.1954  1.0541  2.2762

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.95841    0.26408  37.710   <2e-16 ***
L6           0.03764    0.01587   2.372   0.0268 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.294 on 22 degrees of freedom
Multiple R-Squared: 0.2037,     Adjusted R-squared: 0.1675
F-statistic: 5.628 on 1 and 22 DF,  p-value: 0.02684
```

Since we have 6 tests, we compare the results of the tests with $\alpha = 0.05/6 = 0.008333333$. The following conclusions are possible then: The variable `L2` has a significant influence on `protein` and the intercept of the regression line differs significantly from zero. Variable `L3` has also a significant influence however there is no evidence that the intercept differs from zero. There is no evidence that `L4` has an influence on `protein` nor that the corresponding regression line has an intercept different from zero. The same holds for the variable `L5`. Variable `L6` shows no significant influence on `protein`, but the regression line has an intercept which differs significantly from zero.

```
> plot(ground$L2,ground$protein,xlab="L2",ylab="Protein")
> abline(lsfit(ground$L2,ground$protein)$coef)
```

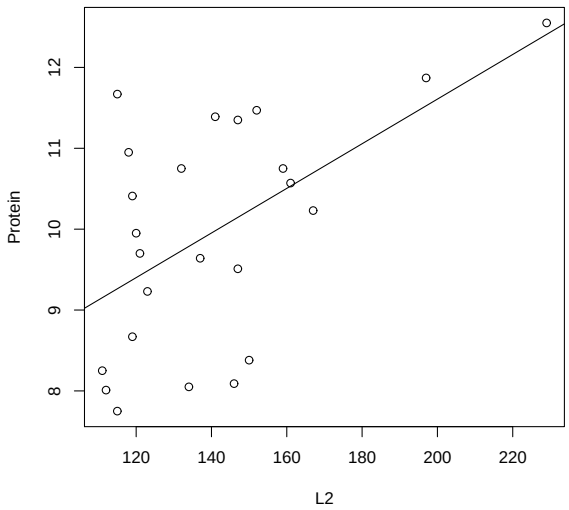

Figure 13.7: Scatter plot with regression line

13.25 Solution of Exercise 6.2.2

```
> x<-c(0,0.1,0.2,0.3,0.4,0.5,0.5,0.6,0.7,0.8,0.9,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 22.5437
> x<-c(0,0,0.2,0.2,0.4,0.4,0.6,0.6,0.8,0.8,1,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 17.86607
> x<-c(0,0,0,0.3,0.3,0.3,0.7,0.7,0.7,1,1,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 16.00222
 x<-c(0,0,0,0.5,0.5,0.5,0.5,0.5,0.5,1,1,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 11.66667
> x<-c(0,0,0,0,0.5,0.5,0.5,0.5,1,1,1,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 12.75
> x<-c(0,0,0,0,0,0.5,0.5,1,1,1,1,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 19.8
> x<-c(0,0,0,0.5,0.5,0.5,0.5,0.5,1,1,1,1)
```

```
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 12.31667
> x<-c(0,0,0,0.5,0.5,0.5,0.5,0.5,0.6,1,1,1)
> X<-cbind(rep(1,12),x,x^2)
> ginv(t(X)%*%X)[1,1]+ginv(t(X)%*%X)[2,2]+ginv(t(X)%*%X)[3,3]
[1] 11.72335
```

The best design is that with $x = (0, 0, 0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1, 1, 1)^\top$, i.e. a design with three observations at 0, six at 0.5, and three at 1.

13.26 Solution of Exercise 6.2.6

```
> summary(lm(Yield~poly(Manure,3),data=split))

Call:
lm(formula = Yield ~ poly(Manure, 3), data = split)

Residuals:
    Min      1Q  Median      3Q     Max
-37.389 -16.889  -2.306  15.486  50.611

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       103.972      2.555  40.691  < 2e-16 ***
poly(Manure, 3)1  133.507     21.681   6.158  4.5e-08 ***
poly(Manure, 3)2  -46.724     21.681  -2.155   0.0347 *
poly(Manure, 3)3   -3.641     21.681  -0.168   0.8671
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.68 on 68 degrees of freedom
Multiple R-Squared: 0.3851,     Adjusted R-squared: 0.358
F-statistic:  14.2 on 3 and 68 DF,  p-value: 2.782e-07
```

Hence a quadratic model is an appropriate model. This can be seen also without the function `poly`:

```
> summary(lm(Yield~Manure+I(Manure^2)+I(Manure^3),data=split))

Call:
lm(formula = Yield ~ Manure + I(Manure^2) + I(Manure^3), data = split)

Residuals:
    Min      1Q  Median      3Q     Max
-37.389 -16.889  -2.306  15.486  50.611
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      79.39       5.11  15.535   <2e-16 ***
Manure         2083.33    1708.76   1.219    0.227
I(Manure^2)   -9583.33  128713.08  -0.074    0.941
I(Manure^3) -375000.00 2233236.75  -0.168    0.867
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.68 on 68 degrees of freedom
Multiple R-Squared: 0.3851,     Adjusted R-squared: 0.358
F-statistic:  14.2 on 3 and 68 DF,  p-value: 2.782e-07

> summary(lm(Yield~Manure+I(Manure^2),data=split))

Call:
lm(formula = Yield ~ Manure + I(Manure^2), data = split)

Residuals:
    Min      1Q  Median      3Q     Max
-37.471 -16.834  -2.101  15.057  50.529

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     79.143      4.862  16.277   <2e-16 ***
Manure        2350.606    617.211   3.808   0.0003 ***
I(Manure^2) -31060.606  14311.241  -2.170   0.0334 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.53 on 69 degrees of freedom
Multiple R-Squared: 0.3849,     Adjusted R-squared: 0.367
F-statistic: 21.58 on 2 and 69 DF,  p-value: 5.244e-08
```

Here we see advance that the estimate for β_0 is 79.143, the estimate for β_1 is 2350.606, and the estimate for β_2 is -31060.606.

```
> plot(split$Manure,split$Yield,xlab="Manure",ylab="Yield")
> abline(lsfit(split$Manure,split$Yield)$coef)
> x<-seq(-0.1,0.5,by=0.002)
> X<-cbind(rep(1,length(x)),x,x^2)
> beta<-lsfit(cbind(split$Manure,split$Manure^2),split$Yield)$coef
> y<-X%*%beta
> lines(x,y)
```

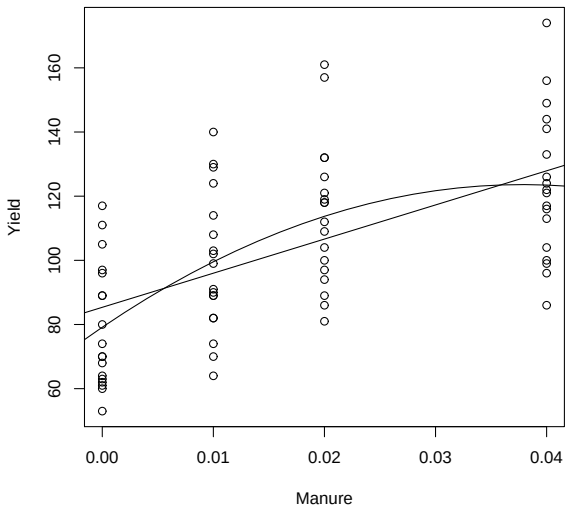

Figure 13.8: Scatter plot with linear and quadratic regression line

13.27 Solution of Exercise 6.3.2

```
> anova(lm(protein~L1+L2+L3+L4+L5+L6,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq  F value    Pr(>F)
L1         1 10.0688 10.0688 207.3930 5.900e-11 ***
L2         1 29.0890 29.0890 599.1666 1.077e-14 ***
L3         1  0.7896  0.7896  16.2630 0.0008637 ***
L4         1  5.2074  5.2074 107.2602 9.241e-09 ***
L5         1  0.2243  0.2243   4.6193 0.0463073 *
L6         1  0.0297  0.0297   0.6108 0.4452214
Residuals 17  0.8253  0.0485
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence a model with the variables L1, L2, L3, L4, L5 is appropriate.

```
> summary(lm(protein~L1+L2+L3+L4+L5,data=ground))

Call:
lm(formula = protein ~ L1 + L2 + L3 + L4 + L5, data = ground)

Residuals:
       Min         1Q     Median         3Q        Max
-0.3734980 -0.1297629  0.0006976  0.1089368  0.3370486
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 18.66952    8.04964   2.319  0.03234 *
L1           0.07756    0.05180   1.497  0.15165
L2          -0.04578    0.06186  -0.740  0.46880
L3           0.24552    0.07537   3.258  0.00437 **
L4          -0.28204    0.03375  -8.357 1.31e-07 ***
L5           0.01286    0.00592   2.173  0.04339 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2179 on 18 degrees of freedom
Multiple R-Squared: 0.9815,     Adjusted R-squared: 0.9764
F-statistic: 191.1 on 5 and 18 DF,  p-value: 6.084e-15
```

Since the p-value for testing $H_0 : \beta = 0$ is 6.084e-15, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top$ differs significantly from the zero vector. Hence we can do the tests for the single variables. It turns out that the intercept differs significantly from zero and that the variables `L3`, `L4`, `L5` have significant influence.

13.28 Solution of Exercise 6.3.4

```
> anova(lm(protein~L1*L2*L3*L4*L5*L6,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq F value Pr(>F)
L1         1 10.0688 10.0688
L2         1 29.0890 29.0890
L3         1  0.7896  0.7896
L4         1  5.2074  5.2074
L5         1  0.2243  0.2243
L6         1  0.0297  0.0297
L1:L2      1  0.0052  0.0052
L1:L3      1  0.0232  0.0232
L2:L3      1  0.0050  0.0050
L1:L4      1  0.2500  0.2500
L2:L4      1  0.1841  0.1841
L3:L4      1  0.0014  0.0014
L1:L5      1  0.0232  0.0232
L2:L5      1  0.1010  0.1010
L3:L5      1  0.0042  0.0042
L4:L5      1  0.1179  0.1179
L1:L6      1  0.0459  0.0459
L2:L6      1  0.0040  0.0040
L3:L6      1  0.0010  0.0010
L4:L6      1  0.0010  0.0010
```

```
L5:L6      1  0.0282  0.0282
L1:L2:L3   1  0.0197  0.0197
L1:L2:L4   1  0.0105  0.0105
Residuals  0  0.0000
> anova(lm(protein~L1*L2*L3*L4*L5+L6,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq F value Pr(>F)
L1         1 10.0688 10.0688
L2         1 29.0890 29.0890
L3         1  0.7896  0.7896
L4         1  5.2074  5.2074
L5         1  0.2243  0.2243
L6         1  0.0297  0.0297
L1:L2      1  0.0052  0.0052
L1:L3      1  0.0232  0.0232
L2:L3      1  0.0050  0.0050
L1:L4      1  0.2500  0.2500
L2:L4      1  0.1841  0.1841
L3:L4      1  0.0014  0.0014
L1:L5      1  0.0232  0.0232
L2:L5      1  0.1010  0.1010
L3:L5      1  0.0042  0.0042
L4:L5      1  0.1179  0.1179
L1:L2:L3   1  0.0730  0.0730
L1:L2:L4   1  0.0034  0.0034
L1:L3:L4   1  0.0001  0.0001
L2:L3:L4   1  0.0001  0.0001
L1:L2:L5   1  0.0174  0.0174
L1:L3:L5   1  0.0086  0.0086
L2:L3:L5   1  0.0078  0.0078
Residuals  0  0.0000
> anova(lm(protein~L1*L2*L3*L4+L5*L6,data=ground))
Analysis of Variance Table

Response: protein
            Df  Sum Sq Mean Sq   F value    Pr(>F)
L1           1 10.0688 10.0688  905.7282 7.598e-07 ***
L2           1 29.0890 29.0890 2616.6841 5.397e-08 ***
L3           1  0.7896  0.7896   71.0238 0.0003858 ***
L4           1  5.2074  5.2074  468.4273 3.907e-06 ***
L5           1  0.2243  0.2243   20.1735 0.0064499 **
L6           1  0.0297  0.0297    2.6677 0.1633348
L1:L2        1  0.0052  0.0052    0.4689 0.5239244
L1:L3        1  0.0232  0.0232    2.0835 0.2084936
L2:L3        1  0.0050  0.0050    0.4471 0.5333354
L1:L4        1  0.2500  0.2500   22.4842 0.0051421 **
```

```
L2:L4        1  0.1841  0.1841   16.5569 0.0096433 **
L3:L4        1  0.0014  0.0014    0.1259 0.7371602
L5:L6        1  0.1666  0.1666   14.9890 0.0117417 *
L1:L2:L3     1  0.1135  0.1135   10.2129 0.0241092 *
L1:L2:L4     1  0.0032  0.0032    0.2904 0.6131020
L1:L3:L4     1  0.0039  0.0039    0.3464 0.5817152
L2:L3:L4     1  0.0080  0.0080    0.7217 0.4343597
L1:L2:L3:L4  1  0.0057  0.0057    0.5155 0.5049145
Residuals    5  0.0556  0.0111
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(protein~L1*L2*L3+L4*L5*L6,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq  F value    Pr(>F)
L1         1 10.0688 10.0688 243.9593 7.935e-08 ***
L2         1 29.0890 29.0890 704.8080 7.379e-10 ***
L3         1  0.7896  0.7896  19.1304  0.001787 **
L4         1  5.2074  5.2074 126.1716 1.349e-06 ***
L5         1  0.2243  0.2243   5.4338  0.044671 *
L6         1  0.0297  0.0297   0.7185  0.418595
L1:L2      1  0.0052  0.0052   0.1263  0.730478
L1:L3      1  0.0232  0.0232   0.5612  0.472903
L2:L3      1  0.0050  0.0050   0.1204  0.736545
L4:L5      1  0.0822  0.0822   1.9908  0.191871
L4:L6      1  0.0830  0.0830   2.0111  0.189834
L5:L6      1  0.1549  0.1549   3.7527  0.084691 .
L1:L2:L3   1  0.0966  0.0966   2.3397  0.160467
L4:L5:L6   1  0.0039  0.0039   0.0950  0.764876
Residuals  9  0.3715  0.0413
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(protein~L1*L2*L3*L4+L5+L6,data=ground))
Analysis of Variance Table

Response: protein
            Df  Sum Sq Mean Sq   F value    Pr(>F)
L1           1 10.0688 10.0688 1086.4467 5.188e-08 ***
L2           1 29.0890 29.0890 3138.7870 2.172e-09 ***
L3           1  0.7896  0.7896   85.1951 9.129e-05 ***
L4           1  5.2074  5.2074  561.8918 3.700e-07 ***
L5           1  0.2243  0.2243   24.1987  0.002659 **
L6           1  0.0297  0.0297    3.1999  0.123852
L1:L2        1  0.0052  0.0052    0.5625  0.481618
L1:L3        1  0.0232  0.0232    2.4992  0.164985
L2:L3        1  0.0050  0.0050    0.5363  0.491566
L1:L4        1  0.2500  0.2500   26.9705  0.002028 **
```

```
L2:L4        1  0.1841  0.1841   19.8605  0.004299 **
L3:L4        1  0.0014  0.0014    0.1511  0.710937
L1:L2:L3     1  0.2798  0.2798   30.1932  0.001522 **
L1:L2:L4     1  0.0030  0.0030    0.3201  0.592038
L1:L3:L4     1  0.0044  0.0044    0.4786  0.514931
L2:L3:L4     1  0.0080  0.0080    0.8607  0.389345
L1:L2:L3:L4  1  0.0058  0.0058    0.6234  0.459845
Residuals    6  0.0556  0.0093
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(protein~L1*L2*L3+L4+L5+L6,data=ground))
Analysis of Variance Table

Response: protein
          Df  Sum Sq Mean Sq  F value    Pr(>F)
L1         1 10.0688 10.0688 294.2838 2.608e-10 ***
L2         1 29.0890 29.0890 850.1973 3.129e-13 ***
L3         1  0.7896  0.7896  23.0766 0.0003445 ***
L4         1  5.2074  5.2074 152.1986 1.498e-08 ***
L5         1  0.2243  0.2243   6.5546 0.0237299 *
L6         1  0.0297  0.0297   0.8668 0.3688215
L1:L2      1  0.0052  0.0052   0.1524 0.7026080
L1:L3      1  0.0232  0.0232   0.6770 0.4254704
L2:L3      1  0.0050  0.0050   0.1453 0.7092506
L1:L2:L3   1  0.3472  0.3472  10.1478 0.0071639 **
Residuals 13  0.4448  0.0342
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The models `protein~L1*L2*L3*L4*L5*L6` and `protein~L1*L2*L3*L4*L5+L6` cannot be used since they have

$$\binom{6}{0}+\binom{6}{1}+\binom{6}{2}+\ldots+\binom{6}{6}=1+6+15+20+15+6+1=64$$

and

$$\binom{5}{0}+\binom{5}{1}+\ldots+\binom{5}{5}+1=1+5+10+10+5+1+1=33,$$

respectively, model parameters, but the sample size is only $N=24$. All other models can be used. The largest model is `protein~L1*L2*L3*L4+L5*L6` which has

$$\binom{4}{0}+\binom{4}{1}+\ldots+\binom{4}{4}+2+1=1+4+6+4+1+2+1=19$$

model parameters. It can be reduced to the model

`protein~L1*L2+L1*L3+L2*L3+L1*L4+L2*L4+L3*L4+L5*L6+L1:L2:L3`.

The model `protein~L1*L2*L3+L4*L5*L6` has

$$\binom{3}{0}+\binom{3}{1}+\binom{3}{2}+\binom{3}{3}+\binom{3}{1}+\binom{3}{2}+\binom{3}{3}=1+3+3+1+3+3+1=15$$

model parameter and can be reduced to the model
`protein~L1+L2+L3+L4+L5`.
The model `protein~L1*L2*L3*L4+L5+L6` has

$$\binom{4}{0}+\binom{4}{1}+\ldots+\binom{4}{4}+2=1+4+6+4+1+2=18$$

model parameters. It can be reduced to the model
`protein~L1*L2+L1*L3+L2*L3+L1*L4+L2*L4+L3*L4+L5+L6+L1:L2:L3`.
The model `protein~L1*L2*L3+L4+L5+L6` has

$$\binom{3}{0}+\binom{3}{1}+\binom{3}{2}+\binom{3}{3}+3=1+3+3+1+3=11$$

model parameters. This model cannot be reduced.

13.29 Solution of Exercise 6.4.2

```
> anova(lm(Yield~Manure*Variety,data=split))
Analysis of Variance Table

Response: Yield
               Df Sum Sq Mean Sq F value    Pr(>F)
Manure          1  17824   17824 36.5035 7.842e-08 ***
Variety         2   1786     893  1.8292    0.1686
Manure:Variety  2    149      74  0.1522    0.8591
Residuals      66  32227     488
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The variety has no significant influence and there is also no significant interaction between variety and manure.

```
> plot(split$Manure,split$Yield,type="n",xlab="Manure",ylab="Yield")
> text(split$Manure,split$Yield,as.character(split$Variety))
> co<-coefficients(lm(Yield~Manure*Variety,data=split))
> abline(co[1],co[2])
> abline(co[1]+co[3],co[2]+co[5],lty=2)
> abline(co[1]+co[4],co[2]+co[6],lty=3)
> legend(0.025,70,c("1","2","3"),lty=c(1,2,3))
```

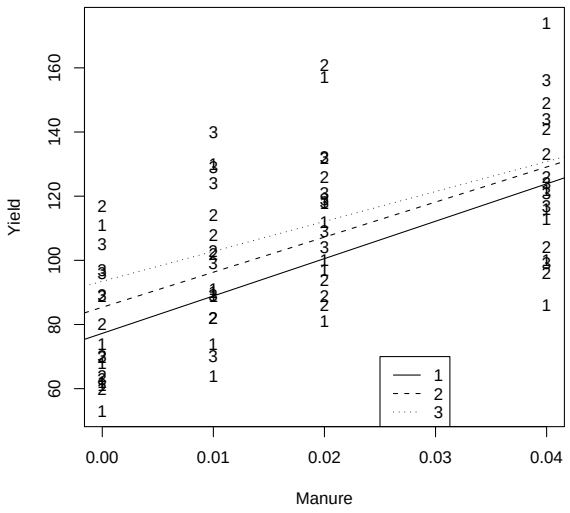

Figure 13.9: Scatter plot with regression lines for the four varieties

Since the three varieties have no significant influence, the intercept of the three lines are very similar. Since also the interactions between variety and manure are not significant also the slopes of the three lines are very similar.

13.30 Solution of Exercise 6.5.2

```
> coefficients(lm(Yield~Manure*Variety,data=split))
    (Intercept)          Manure        Variety2        Variety3 Manure:Variety2
      77.233333     1165.238095        8.133333       16.233333      -71.904762
Manure:Variety3
    -232.380952
> coefficients(lm(Yield~Variety*Manure,data=split))
    (Intercept)        Variety2        Variety3          Manure Variety2:Manure
      77.233333        8.133333       16.233333     1165.238095      -71.904762
Variety3:Manure
    -232.380952
> anova(lm(Yield~Manure*Variety,data=split))
Analysis of Variance Table

Response: Yield
              Df Sum Sq Mean Sq F value    Pr(>F)
Manure         1  17824   17824 36.5035 7.842e-08 ***
Variety        2   1786     893  1.8292    0.1686
Manure:Variety 2    149      74  0.1522    0.8591
Residuals     66  32227     488
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(Yield~Variety*Manure,data=split))
Analysis of Variance Table

Response: Yield
              Df Sum Sq Mean Sq F value    Pr(>F)
Variety        2   1786     893  1.8292    0.1686
Manure         1  17824   17824 36.5035 7.842e-08 ***
Variety:Manure 2    149      74  0.1522    0.8591
Residuals     66  32227     488
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coefficients(lm(Yield~Block+Manure*Variety,data=split))
    (Intercept)          Block2          Block3          Block4          Block5
     108.594444      -28.083333      -39.416667      -37.166667      -44.416667
         Block6          Manure        Variety2        Variety3 Manure:Variety2
     -39.083333     1165.238095        8.133333       16.233333      -71.904762
Manure:Variety3
    -232.380952
> coefficients(lm(Yield~Block+Variety*Manure,data=split))
    (Intercept)          Block2          Block3          Block4          Block5
     108.594444      -28.083333      -39.416667      -37.166667      -44.416667
         Block6        Variety2        Variety3          Manure Variety2:Manure
     -39.083333        8.133333       16.233333     1165.238095      -71.904762
Variety3:Manure
    -232.380952
> anova(lm(Yield~Block+Manure*Variety,data=split))
Analysis of Variance Table

Response: Yield
              Df  Sum Sq Mean Sq F value    Pr(>F)
Block          5 15875.3  3175.1 11.8446 5.032e-08 ***
Manure         1 17824.1 17824.1 66.4935 2.387e-11 ***
Variety        2  1786.4   893.2  3.3320   0.04233 *
Manure:Variety 2   148.6    74.3  0.2772   0.75885
Residuals     61 16351.6   268.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(Yield~Block+Variety*Manure,data=split))
Analysis of Variance Table

Response: Yield
              Df  Sum Sq Mean Sq F value    Pr(>F)
Block          5 15875.3  3175.1 11.8446 5.032e-08 ***
Variety        2  1786.4   893.2  3.3320   0.04233 *
Manure         1 17824.1 17824.1 66.4935 2.387e-11 ***
Variety:Manure 2   148.6    74.3  0.2772   0.75885
```

```
Residuals      61 16351.6   268.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> anova(lm(Yield~Block*Manure*Variety,data=split))
Analysis of Variance Table

Response: Yield
                     Df  Sum Sq Mean Sq F value    Pr(>F)
Block                 5 15875.3  3175.1 14.1487 1.136e-07 ***
Manure                1 17824.1 17824.1 79.4279 1.227e-10 ***
Variety               2  1786.4   893.2  3.9802   0.02744 *
Block:Manure          5   817.5   163.5  0.7286   0.60660
Block:Variety        10  6013.3   601.3  2.6797   0.01462 *
Manure:Variety        2   148.6    74.3  0.3311   0.72027
Block:Manure:Variety 10  1442.1   144.2  0.6426   0.76763
Residuals            36  8078.6   224.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(Yield~Variety*Block*Manure,data=split))
Analysis of Variance Table

Response: Yield
                     Df  Sum Sq Mean Sq F value    Pr(>F)
Variety               2  1786.4   893.2  3.9802   0.02744 *
Block                 5 15875.3  3175.1 14.1487 1.136e-07 ***
Manure                1 17824.1 17824.1 79.4279 1.227e-10 ***
Variety:Block        10  6013.3   601.3  2.6797   0.01462 *
Variety:Manure        2   148.6    74.3  0.3311   0.72027
Block:Manure          5   817.5   163.5  0.7286   0.60660
Variety:Block:Manure 10  1442.1   144.2  0.6426   0.76763
Residuals            36  8078.6   224.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The order of the variables has no influenced since it is balanced design.

14 Solutions of mathematical exercises

14.1 Solution of Exercise 2.4.1

Proposal: $N_1 = \frac{N}{2} = N_2$

Proof: The larger the noncentrality parameter is the smaller the β-error is. Hence we have to maximize $K = \sqrt{\frac{N_1 N_2}{N_1+N_2}}$. The maximization of $\frac{N_1 N_2}{N_1+N_2}$ is equivalent to the minimization of $f(n) = \frac{1}{n} + \frac{1}{N-n}$ with respect to $n = N_1$. Differentiation of f yields

$$f'(n) = \frac{-1}{n^2} + \frac{1}{(N-n)^2} = \frac{-(N-n)^2 + n^2}{n^2(N-n)^2} = 0$$
$$\Leftrightarrow \quad n^2 = (N-n)^2 \Leftrightarrow n = N - n \Leftrightarrow n = \frac{N}{2}$$

Since

$$f''(n) = \frac{2}{n^2} + \frac{2}{(N-n)^3} > 0$$

the minimum is attained at $n = \frac{N}{2}$.

14.2 Solution of Exercise 4.3.1

$$\begin{aligned}\Sigma_{SSG} &:= \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}}(y_{abn} - \overline{y}_{\bullet\bullet\bullet})^2 \\ &= \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{n=1}^{N_{ab}}(y_{abn} - \overline{y}_{ab\bullet} + \overline{y}_{ab\bullet} - \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet b\bullet} + \overline{y}_{\bullet\bullet\bullet} + \overline{y}_{a\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet} + \overline{y}_{\bullet b\bullet} - \overline{y}_{\bullet\bullet\bullet})^2 \\ &= \Sigma_{SSE} + \Sigma_{SSI} + \Sigma_{SSA} + \Sigma_{SSB}.\end{aligned}$$