

Series 5

1. Moisture content is the most important criterion for quality in the manufacture of a pigment paste. 15 batches with two samples each were randomly taken. Each sample was then analyzed twice. The file `paint.txt` contains the values for BATCH (1 to 15), SAMPLE (1 or 2) and response variable MOISTURE.

```
paint <-
read.table(file="http://stat.ethz.ch/Teaching/Datasets/paint.txt",header=TRUE)
```

- a) Plot the data.
- b) Write down the analysis of variance model. Why are all factors random?
- c) Construct the skeleton analysis of variance table with degrees of freedom and expected values.
- d) Calculate estimates for σ_{BATCH}^2 and σ_{SAMPLE}^2 . Compute the F test statistics, first by hand and then use R.

R Hints: The sum of squares and mean squares can be determined with `aov`:

```
> mod1=aov(MOISTURE ~ BATCH + SAMPLE %in% BATCH, data=paint)
# or: aov(MOISTURE ~ BATCH/SAMPLE, data=paint)
> summary(mod1)
```

The analysis of variance table contains the correct test for the factor `SAMPLE`. To get the correct test for `BATCH`, you have to specify the appropriate error term for the factor `BATCH`:

```
> mod2=aov(MOISTURE ~ BATCH + Error(SAMPLE %in% BATCH), data=paint)
```

2. A plant breeder conducted an experiment to compare the yields of 3 new varieties and a standard variety of peanuts. The varieties were assigned to the plots where it was known that land had a slight sloping from east to west and differences in available nitrogen from north to south. Thus, to account for this variability it was decided to block in the eastwest direction and in the north-south direction. The experiment was run as a Latin square. If we denote the four treatments by T1, T2, T3, T4, the field layout of the experiment was

	Blocks B East-West				
	1	2	3	4	
Blocks	1	T3	T1	T2	T4
C	2	T1	T2	T4	T3
North-South	3	T2	T4	T3	T1
	4	T4	T3	T1	T2

The data are in file `Peanut.txt`.

```
peanut <-
read.table(file="http://stat.ethz.ch/Teaching/Datasets/Peanut.txt",header=TRUE)
```

- a) Plot the data.
- b) Carry out the analysis of variance and report your findings.
- c) Treatment 1 is the control treatment. Even though the anova F is not significant, do any of the other treatments have a significantly higher yield, if a one-sided test is used?

The breeder decided to replicate the experiment. He used three blocks at different locations in the field, but in each location the north-south and east-west gradients were identified. The data are in the file `Peanut2.txt`.

- d) Which factors are nested, which are crossed?
- e) Give an analysis of variance table and test whether any pairwise differences are significant.

3. Tennis elbow is a painful condition that affects many tennis players at some time. A clinical trial was conducted comparing the effectiveness of Motrin, a widely used anti-inflammatory agent, to placebo. Participants received both drugs, but the treatment order was randomized. Group A participants received an initial 3-week course of Motrin, while the other participants in group B received an initial 3-week course of placebo. There followed a 2-week washout period during which they received no study medication. After the washout period, a second period of active administration began, with group A participants receiving 3 weeks of placebo and group B participants receiving 3 weeks Motrin. At the end of each active drug period as well as at the end of the washout period, participants rated their degree of pain compared with baseline (before the beginning of the first active drug period). Degree of pain versus baseline was measured on a 1–6 scale. The comparison was made in four different ways: (1) during maximum activity, (2) 12 hours following maximum activity, (3) during the average day and (4) by overall impression of drug efficacy. The data are in file `TENNIS.dat`, `TENNIS.txt` contains a description of all variables.

```
> tennis <- read.table(file="http://stat.ethz.ch/Teaching/Datasets/TENNIS.dat")
> names(tennis)=c("id", "age", "sex", "order", "max1", "twelve1", "ave1",
  "overall1", "max2", "twelve2", "ave2", "overall2", "max3", "twelve3", "ave3", "overall3")
> ## replace invalid values with NA:
> for (i in 3:16)
  tennis[,i][tennis[,i]==9 | tennis[,i]==0]=NA##identify the invalid values
>
```

Here, `order` stands for the order in which the treatments were received (there are only 2 possibilities for this). The numbers after the variable names *1 *2 *3 stand for the time point at which the measurement was done.

To do an anova fit you can reshape your data as follows.

```
> tennis.max=reshape(tennis[,c(1:5,9,13)],varying=c("max1","max2","max3"),idvar="id",
  timevar="period",v.names="max",direction="long")
> tennis.max$Treatment[tennis.max$order==1 & tennis.max$period==1]="Motrin"
> tennis.max$Treatment[tennis.max$period==2]="Washout"
> tennis.max$Treatment[tennis.max$order==1 & tennis.max$period==3]="Placebo"
> tennis.max$Treatment[tennis.max$order==2 & tennis.max$period==3]="Motrin"
> tennis.max$Treatment[tennis.max$order==2 & tennis.max$period==1]="Placebo"
> tennis.max$Treatment=factor(tennis.max$Treatment)
> tennis.max$id=factor(tennis$id)
> tennis.max$sex=factor(tennis$sex)
> tennis$sex[tennis$sex==1] <- 'male'
> tennis$sex[tennis$sex==2] <- 'female'
> tennis.max$period=factor(tennis.max$period)
```

Is the treatment with Motrin efficient?

To find this out: First do a pairwise `t.test` and `wilcox.test` of the four different outcomes (max, ave, etc.). Then do an anova fit for only the maximum activity comparison taking into account `id`, `period` and `treatment`.

Is the washout period long enough or could there still be a carry-over effect?

Preliminary discussion: 01.12.2014.

Deadline: 08.12.2014.