Computer Lab Practical n. 2

Today's task

One-way and two-way analysis of variance.

How to start

• Access the course web page on Moodle (/Biology). Download and save the no2.rda and saltBP.dat data files. (To download the latter, you may need to click on the mouse button on the right.)

An example of one-way ANOVA

• Load R's packages called MASS.

```
library(MASS)
```

• Load anorexia data in MASS package

```
data(anorexia)
```

• Inspect the data description.

```
help("anorexia")
```

• We want to investigate the effect of the therapies (Treat) on the difference of weight before (Prewt) and after (Postwt) their onset. To do so, let us create a new variable, say effect as "Postwt – Prewt".

```
anorexia$effect <- anorexia$Postwt - anorexia$Prewt
```

• Create the corresponding boxplots and have a look at the different groups.

```
boxplot(effect~Treat, data = anorexia)
```

Question: what does the figure suggest?

• Obtain the corresponding numerical summaries.

```
numSummary(anorexia$effect, groups = anorexia$Treat)
    or
tapply(anorexia$effect, anorexia$Treat, summary)
```

Try and convince yourself that these summaries, in particular the mean, the median and the quantiles, tell us the same story than the previous graph.

• Try to plot the means

```
plotMeans(anorexia$effect, anorexia$Treat, error.bars="conf.int",
          level=0.95, xlab = "Treat", ylab = "Mean of effect")
```

Comment the picture!

• Let's see what one-way analysis of variance and the Kruskal-Wallis test can tell us.

The former is carried out by

```
fit.aov <- aov(effect ~ Treat, data=anorexia)
summary(fit.aov)
```

The latter is carried out by

```
kruskal.test(effect ~ Treat, data=anorexia)
```

Question: can we conclude that the differences among groups are significant?

• Pairwise comparison can be carried out using, for instance, with the Tukey procedure.

```
TukeyHSD(fit.aov)
```

Pairwise comparison using the t-test or the Wilcoxon test can be carried out using the pairwise.t.test and the pairwise.wilcox.test, respectively.

Questions: How many groups are different? Which groups are different?

• Problem: can we use a similar analysis to try and find out whether the different patients were randomly assigned to the three treatment groups? That is, can we try and find out whether there is some selection bias?

Applying two-way ANOVA

• Load the data genotype from MASS package.

```
data(genotype)
```

Read the summary description of the data.

```
help("genotype")
```

The dataset refers to a litter of rats which right after birth were separated from their natural mother and nourished by an adoptive mother. The Wt variable reports the weight of the litter on day 28. The research question is to investigate whether the genotype of the mother (Mother) or of the litter (Litter) influence weight.

• Apply two-way analysis of variance.

```
weight.aov <- aov(Wt ~ Litter*Mother, data=genotype)
summary(weight.aov)
```

Questions: Is there any effect modification (= interaction)? Are both factors significant? Is pairwise comparison here useful?

A further example of two-way ANOVA

• Load the data provided by the no2.rda file by double clicking on it.
Inspect the data.

```
View(no2)
summary(no2)
```

The data were collected to assess whether exposure to NO_2, a pollutant emitted by automobiles, produces negative effects on the lungs of the Guinea pigs. The following variables are considered:

— Fluorescence: measures the fluorescence of pulmonary tissue; high values highlight a loss of proteins, that is, damages to the lungs;
— NO2: if "Yes", the Guinea pigs were exposed to a concentration of $0,5$ parts per million of NO_2 for the number of days given by Exposition;
— Exposition: number of days of exposure, which can be 10, 12 or 14.

• Let's try and use two-way analysis of variance.

```
no2.aov <- lm(Fluorescence ~ NO2 * Exposition, data=no2)
Anova(no2.aov)
```

Comment the output!

• To understand our findings, have a look at the estimates of the various effects (similarly to what we did in class when anlysing the "poisons and antidotes" dataset).

```
summary(no2.aov)
```

Produce the table of non-zero (= significant) effect estimates. Comment the results.

A very first example of straight line regression

• Load the data of the saltBP.dat file.

```
saltBP <- read.table("~/saltBP.dat",header=TRUE)
```

Where ~ refers to the direction that you stored the data.

• Try and inspect the data.

There are 25 observations on three variables. We will only use blood pressure (BP) and salt consumption (salt). (Both are averages taken on 5 consecutive days.) We will try and investigate whether salt consumption influences blood pressure.

Start off by specifying an equation which allows us to predict blood pressure with salt consumption. Salt consumption is measured in grams and represents

the excess with respect of the generally considered correct dose of 6 grams. (That is, "salt=2" corresponds to a daily consumption of 8 grams.)

- Create the scatter diagram.

```
plot(BP ~ salt, data=saltBP,col="blue")
```

Your comments?

- Calculate the correlation coefficient and verify its significance.

```
cor.test(saltBP$BP, saltBP$salt, alternative="two.sided",
         method="pearson")
```

Your comments?

- Specify the straight line fitted by least squares.

```
fit.lm <- lm(BP~salt, data=saltBP)
summary(fit.lm)
```

Comment the fitted regression model. Are both coefficient significant? What is the blood pressure for a person which consumes exactly 6 grams of salt per day? How much is the blood pressure expected to raise for every extra gram of salt?