UPPSALA UNIVERSITET
Matematiska Institutionen
Statistics for biology. Spring semester 2015.
Jordi-Lluís Figueras (figueras@math.uu.se)

Computer Lab 4.

1. Read the files `l1.csv2` and `circ.csv2`. These files contain paired data.

2. There are several commands that can be used while working with paired data.

   - Plotting data: `plot`.
   - Covariance: `cov`.
   - Correlation coefficient test: `cor.test`.
   - Construction of a linear model: `lm(y~x)`, where `y` is the dependent variable and `x` is the independent one.
   - Summarizing the linear model: `summary`.
   - Plotting the linear regression: just after plotting the data we can use `abline` for plotting the linear regression.

3. For the two data read above, execute all the previous commands.

   Remark: Notice that the command `summary` gives a lot of information: Residuals, the coefficients, F-statistic (correlation coefficient test).

We work for a company that produces candies. We performed two studies about the preferences of our clients with respect several candies we produce: lakrits, marshmallows and lollipops. These two studies are found in the files `candy.csv2` and `candy_2.csv2`. The goal of the first study was to see if all these three candies are equally preferred. The goal of the second was the same as in the first one, but also considering how much candies our study individuals consumed.

4. To study if the means are equal for the first study, we will perform a one way ANOVA. Read the file `candy.csv2` and save it in a variable (`fst`). Have a look at the data and observe how it is organized. Type also `boxplot(fst$score~fst$candy)`, which shows the boxplot of the score preference with respect the candy.

   The one way anova is perform with the command `aov(score~candy, fst)`. Save its output in a variable and then read it with the function `summary`. The number at the right-most column gives us the p-value of the hypothesis test. What do you think? Can we accept that all the means are the same?

5. To perform the second study, we will use a two way ANOVA. Read the file `candy_2.csv2` and save it in a variable (`sdn`). Have a look at the data and observe how it is organized. Also, plot the two boxplots of the score with respect the variable `candy`, and with respect `quantity`.

   The two way anova is perform with the command

`aov(score~candy+quantity+candy*quantity, sdn)`.

Save its output in a variable and then read it with the function `summary`. The numbers at the right-most column gives us the p-value of the hypothesis test (the candy population means are equal, the quantity population means are equal, and the interaction between them). What do you think?