Data Analysis with R, fall 2010

Demonstrations 3

1. Compare the tooth length of guinea pigs between 'VC' and 'OJ' groups using t-test. The data is built-in data called *ToothGrowth*. First you have to test can the variances of the groups be assumed equal and then based on the result you can do t-test correctly. Interpret your results.

2. Fit a linear model using *lm* for the *iris*-dataset with the following formulation:

   ```
   Sepal Width = Petal Length + Petal Width
   ```

   Save the residuals of the model and plot them. Draw horizontal lines to y-values -1 and 1. How many of the observed residuals fall outside the lines?

3. Consider the following example on the strength of three different rubber compounds; four specimens of each type were tested for tensile strength (measured in pounds per square inch):

Type	A	B	C
Strength (lb/in^2)	3225, 3320,	3220, 3410,	3545, 3600,
	3165, 3145	3320, 3370	3580, 3485

   Try to explain differences in tensile strength with the type of rubber using an ANOVA model. Have all of the three rubber types same mean tensile strength? If the answer is No, do necessary multiple comparisons to find out which of the rubber types differ from each other. (Tukey's Honestly Significant Differences would be here a good choice for a multiple comparison test).

4. A random sample of 1000 adults was classified according to sex and whether or not they were colour-blind as summarized below:

	Male	Female
Normal	442	514
Colour-blind	38	6

   Test is there a relationship between sex and colour-blind incidence with a chi-square test. Is there a relationship between these two?

5. Dataset for this assignment is R dataset *attitude*. By fitting a linear model determine which of the variables in the dataset have an effect on *rating*-variable. So first fit model and then test the significance of exploratory variables. What are the results? Finally draw a QQ-plot of the residuals of the model. Does the model seem to fit the data?

6. In the previous assignment we assumed that response variable would not be significantly different from a Normal distribution. We didn't base that assumption on anything. Consider now the dataset called *ChickWeight*. Determine is the Normality assumption valid for variable *weight*. Base your answer on two things: A test and a graphical presentation. Based on your result perform either an ANOVA-model or a Kruskal-Wallis-test to determine is there a statistically significant difference between the different diet groups in *weight*. Interpret your results.