1 Intro Example

We start with a very simple example data set.

1. load the data from the `data1.txt` file using `read.table()`; remember how to use the `header` argument; if you do not remember call the help by typing `?read.table`
2. recode the treatment variable into a factor using the `factor()` function; (Remember: to access the column you have to use the name of the table *and* the name of the column `tablename$colname`)
3. use the appropriate functions to
   (a) get the number of observations/columns
   (b) how many rows per treatment
   (c) what is the mean of `Y` by treatment (`tapply()` might be helpful, but there are many ways)
   (d) plot the respective boxplots using `ggplot()` and `geom_boxplot()`
   (e) we wanna compare these means. What would be the null hypothesis?
4. build a model using the `lm()` function, take `Y` as response variable and `Trt` as explanatory variable
5. extract the coefficients and their confidence intervals
6. is there a difference in means between the two groups?
7. is this difference in means between the two groups statistically significant?
8. What test could you use to confirm the results? Compare! Are the null hypotheses the same?

2 Simple linear regression, prediction: Heart and body weights

In the R package `MASS` there is a dataset called cats. Run the following commands:

```
library(MASS)
data(cats)
```

1. Have a look at the dataset. The variables Bwt and `Hwt` give the weight of the body (kg) and the heart (g), respectively.
2. There are both male and female cats. Make a dataset with the data from males only. 2. Make a scatterplot of the data for the male cats (`Bwt` on x-axis, `Hwt` on y-axis). Does it look reasonable to use a linear regression model for the data?
3. Fit a linear regresison model for the male cats, that allows for prediction of the heart weight given the body weight. Add the fitted regression line to the scatterplot from the previous question.

4. Find the coefficients of the fitted line. How large is the expected difference in heart weight for two cats with a difference of 1 kg in bodyweight? Find a confidence interval for this difference? How large is the expected difference in heart weight for two cats with a difference of 100 g in bodyweight?

5. Use model validation plot to examine if the model is appropriate for the data. (use `plot()` on your model object)

6. Use the estimates to find the expected heart weight for a male cat that weighs 3 kg. Then try the commands (where you replace the name regModel with whatever name you gave the the model fit in question 2).

```
> newObs <- data.frame(Bwt=3)
> newObs
> predict(regModel, newObs)
> predict(regModel, newObs, interval="predict"
```

3 Binary Response

Now we consider the binomial response defined by N (number of trials) and F (number of successes) in the same data set.

1. Set up the appropriate model using `glm()`. Use again `Trt` as explanatory variable. You find a example of a structural similar model at the help page of the `esoph` data set.

2. What are the coefficients? What is the problem?

3. In this context could you explain what's the model scale and what's the data scale?

4. What is the probability of success in both groups?

5. now use Y instead of `Trt` as explanatory variable

6. is there an effect?

7. use `ggplot()` to visualize the model