

Übungen für regressionanalyse

Name

Daten

Task 1. Preparation

- Fill the headings in the R file (author name and date)
- Install the packages if necessary: `formatR`, `psych`, `RcmdrMisc` (hint: `install.packages("formatR")`)
- Save the file encoding (UTF-8)

Task 2. Import the dataset

- download and put the *fruitfly.csv* into your working directory
- set your working directory (hint: on the right site click on ... find the directory, OK, after click on More - set Working Directory)
- import fruitfly dataset (hint: on the right site click on *Import Dataset* find the directory, OK...)
- factorize the *groups* variable (hint: `fruitflydollargroups <- as.factor(fruitflydollargroups)` - of course use the dollar symbol instead of writing dollar))
- put this last 3 command into an R code chunk (hint: on the top click on *Insert R*)

Task 3. Descriptive statistics: Graphs, summary

Create a new code chunk and:

- create a **boxplot** of *lifespan* in the different *groups* (hint: `boxplot` command - see help if needed)
- create a **scatterplot** of *lifespan* and *thorax_l* (hint: `plot` command - see help if needed)
- create a **scatterplot** of *lifespan* and *sleep* (hint: `plot` command - see help if needed)
- (extra: this last two could be more 'fancy' with using `pairs.panels` function)
- summarize the variables using `summary`
- (extra: summarize the variable *lifespan* in the different *groups* using `numSummary`)

Task 4. Regression models

Create a new code chunk(s) and (recommendation: use different code chunks for the different models):

- make a regression model without interactions using *lm* command where the outcome variable is *lifespan* and predictor variables are the *groups*, *thorax_l* and *sleep* (hint: model name <- lm(outcome ~ predictor1 + predictor2 + predictor3, data = the dataset) and after summary(model name))
- modify the model with using only the significant predictors
- (extra: check interaction using *_** instead of *+*)
- create the model without intercept (hint: use *-1* after the predictors)
- try to interpret the model (how much change in the lifetime do you expect in average in the different groups)
- diagnose your final linear model (hint: use *_plot*(model name))
- In a new code chunk calculate the confidence interval of the parameters (hint: *confint*(model name))
- (extra: calculate the 95% prediction interval and confidence interval at each group if the thorax length is the mean of *thorax_l* (hint: *predict* function))