STATISTICS FOR EES — EXERCISE SHEET 2

---

1. An experiment was performed in which two groups of chicks were fed with two different diets. After three weeks the weights of the chicks were as follows (in g):

Diet A

           175 74 331 167 265 251 192 233 309 150

Diet B

           196 204 281 238 205 200 322 237 264

(a) Compute for each group the mean and the standard error.

(b) Assume that you have to design an experiment, in which the mean weight of chicks fed with diet A has to be determined with a precision of $\pm 5g$. How many chicks do you need?

(c) How many chicks do you need to solve the same problem for diet B?

(data taken from R example data set `ChickWeight`, cf. M. Crowder, D. Hand, *Analysis of Repeated Measures*, Chapman and Hall (1990).)

2. An experiment was carried out in which 23 test persons estimated the size of a blue area (relative to a total area of 100) in five different types of plots. Download the R file http://evol.bio.lmu.de/_statgen/StatEES/16SS/bluearea_first_analysis.R that contains the commands for downloading the data and visualizing certain aspects of the data.

(a) Understand what this R file does and how it does it by

- trying it out, also parts of it,
- using the online help system, and
- changing some commands and options and check how this affects the results.

(b) Add similar R commands to the file to visualize how the estimation errors depend on the type of the plot **and** on the true value. Try to do this in just one plot, and explore several possibilities for this.

(c) What (preliminary) conclusions would you draw from your new plots?

(d) Discuss the experimental design. Should the experiment be repeated in a different way? Discuss the advantages and disadvantages of several possible approaches.

3. Repeat exercise 2 with various alternative error measures, e.g. squared error, relative error, …

(a) Discuss and explore with the data how sensitive these error measures are for outliers. Discuss how sensitive they *should be* for outliers.

(b) Discuss and explore with the data how the different error measures depend on the true value of the blue area.

(c) Search by visualization for further possible effects in the data.