

STATISTICS FOR EES — EXERCISE SHEET 2

1. An experiment was carried out in which 19 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 https://syncandshare.lrz.de/getlink/fiBMyyPQhbix6hFsChtcFX/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.
- adapting and applying it the blue area experiment data of your cohort

(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.

2. Repeat exercise 1 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.

(d) Explore this: If the errors of a certain visualization method is below average in this dataset according to a certain way of measuring errors, does this also hold for any other reasonable way of measuring errors?

3. Two dice are tossed. Given that the sum of the pips is 8, what is the probability that at least one of the dice shows 3 pips?

4. Imagine a test for an infectious disease with a sensitivity of 99.9%, i.e. it detects the disease for 99.9% of the patients that have the disease, and a specificity of 90%, i.e. it indicates the disease for 10% of the uninfected patients. If 2% of the population are infected and the tests indicates the disease for a person that was randomly selected from the population, what is then the probability that the person is indeed infected?

5. You have two dice that look exactly the same, but one is fair and the other one is loaded and gives a six in $1/3$ of the cases (and other results with a probability of $2/15$ each). You forgot which of the two dice is which, so you roll each dice five times. The first dice shows a six only once and with the second dice you obtain two sixes. Given this result, calculate the probability that the first dice is fair.

6. A casino offers the following game: One dice is rolled, and if X is the number of pips, then the player wins $f(X) = 3 - (x - 3)^2$ Euros, where negative values of $f(X)$ mean that the player loses money. Let $G = f(X)$ be the gain of a player in one game. Let T be the total gain of a player who played the game 100 times (where a negative gain is actually a loss).

(a) Calculate $\mathbb{E}G$ in two ways:

(a) once directly with the definition of the expected value $\sum_y y \cdot \Pr(G = y)$

(b) once as $\sum_x f(x) \cdot \Pr(X = x)$

(b) Calculate the expected value and the standard deviation of T .

(c) Which well-known distribution is can be used to approximate the distribution of T ?