Dirk Metzler SS 2021

STATISTICS FOR EES AND MEME — EXERCISE SHEET 5

1. A breeder crossed 200 pairs of plants of some crop. The yield of the father plant, the mother plant and the offspring (F1) is given in table yield.csv (in some measuring unit). For the next generation, the breeder crosses plants from the F1 generation, but selects for this only plants of a yield of more than 20.

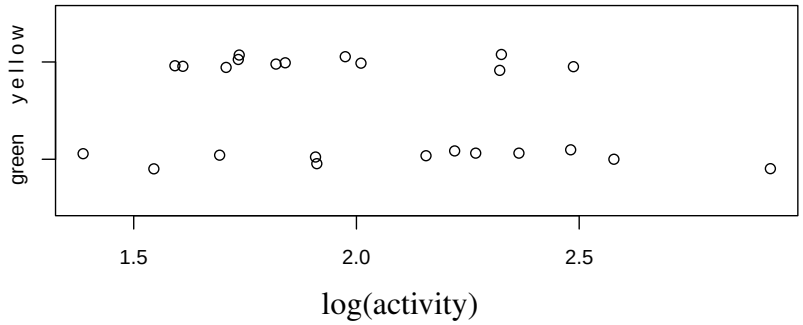
(a) Visualize the distributions of the yield in the parent population, in the offspring population before selection and in the offspring population after selection.

(b) Add lines indicating the mean values to your plot.

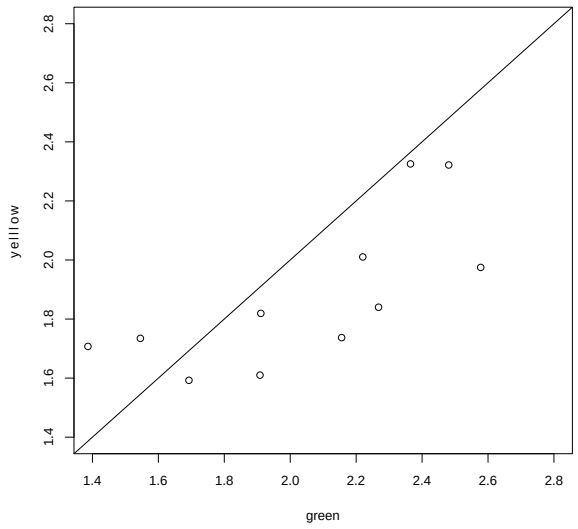
(c) Predict the average yield of the plants of the F2 generation and show it in your plot.

2. Sample $n = 10$ values from a normal distribution with $\mu = 0$ and $\sigma = 5$. Apply the z-test to this sample for the null hypothesis $\mu = 0$, first assuming $\sigma = 5$ and then using the σ that is estimated from the sample. Repeat this 10.000 times. How often is the null hypothesis rejected on the 5% level if the true σ is used and how often if the estimated σ is used? How does this change when the sample size is $n = 100$?

3. **a)** Can the following values be true or are some of them obviously wrong?
$\overline{x}$(yellow) = 1.93, σ(yellow) = 0.30, $\overline{x}$(green) = 2.12 and σ(green) = 0.45



b) Adam claims that the activity with green light is significantly higher. Do you agree?



c) Eve responds: None of the means is significantly different from 2.0. Is she right?

The data come from 12 birds. Each bird gave a activity value for green and one for yellow. The paired samples are shown in the second graph.

d) How do you describe the activities corresponding to the two colors?

e) Compute the t-statistic to test if the means are equal. The standard deviation of the differences is s=0.27.

f) Summarise the result of the test.

4. Student's classical hypnotics dataset: Two hypnotics were tested with 10 test persons. The sleep time was measured in hours relative to the average in a control group. Thus, a negative value indicates that the test person slept less than the average in the control group. You get the data in R in a data frame `sleep` after typing `data(sleep)`. The column `group` shows which hypnotic was used, and row n refers to the same test person as test person $n + 10$. Perform an appropriate t-test to compare the efficacy of the two hypnotics. Do this first without using the R command t.test. Then apply this command to check your results.

5. The t-test requires that the sample consists of independent observations. Explore how robust the t-test is against violations of this requirement. To this end, simulate data `x` by
`x <- rnorm(20) + sample(rnorm(10),20,replace=TRUE).`

(a) Explain what this command does.

(b) What is the expectation value of `x`?

(c) Perform t-tests on x for the null hypothesis that it is sampled from a normal distribution with mean 0. Repeat this with 10000 different simulated data sets `x`. How often is the null hypothesis rejected?