This assignment was locked 6 Mar at 23:59.

- *Before submitting your report please read [Guidelines for assignment reports to submit](), the page limit for your report is **20**.*
- *Throughout this assignment tests should be performed using a confidence level*
- *=0.05, unless otherwise specified.*
- *Abbreviation CI means* confidence interval*.*
- *Where appropriate, motivate your answers and check the model assumptions by using relevant diagnostic tools.*
- *Beware that the levels of some factors for some data sets may be coded by numbers.*

Exercise 1. Waiting time

A researcher measured (in minutes) how long patients have to wait in the waiting room of a doctor's office: 15.4, 17.9, 19.0, 0.5, 15.9, 2.7, 6.2, 2.5, 4.7, 6.9, 10.8, 24.3, 5.6, 23.0, 10.7. Denote the mean waiting time by

.

a) Check normality of the data. Assuming normality (irrespective of your conclusion about normality of the data), construct a 97%-CI for

. Evaluate the sample size needed to provide that the length of the 97%-CI is at most 2. Compute a bootstrap 97%-CI for

and compare it to the above CI.

b) The doctor claims that the mean waiting time is less than 15 minutes. Under an assumption, verify this claim by a relevant t-test, explain the meaning of the CI in the R-output for this test. Propose and perform a suitable sign tests for this problem. Can we use yet another test based on ranks?

c) Propose a way to compute the powers of the t-test and sign test from b) at

= 14 and

= 13, comment.

d) Let *p* be the probability that a patient has to wait longer than 15.5 minutes. Using asymptotic normality, the researcher computed the right end

= 0.53 of the confidence interval [

] for *p*. Recover the whole confidence interval and its confidence level.

e) The researcher also reported that there were 3 men and 2 women among 5 patients who had to wait more than 15.5 minutes, 4 men and 6 women among the remaining 10 patients.

The researcher claims that the waiting time is different for men and women. Verify this claim by an appropriate test.

Exercise 2. Seeded clouds

To improve rain fall in dry areas, an experiment was carried out with 52 clouds. Scientists investigated whether the addition of silver nitrate leads to more rainfall. They chose 26 out of a sample of 52 clouds and seeded it with silver nitrate. The file clouds.txt Download clouds.txt contains the precipitation values (records the rainfall in feet per acre) of seeded and unseeded clouds.

a) Test whether silver nitrate has an effect by performing three tests: the two samples t-test (argue whether the data are paired or not), the Mann-Whitney test and the Kolmogorov-Smirnov test. Indicate whether these tests are actually applicable for our research question. Comment on your findings.

b) Repeat the procedures from a) first on the square root of the values in *clouds.txt*, then on the square root of the square root of the values in *clouds.txt*. Comment on your findings.

c) Let

be the sample for seeded clouds (column *seeded*). Assuming and using the central limit theorem, find an estimate of and construct a 95%-CI for . By using a bootstrap test with the test statistic , test the hypothesis with the parameter

. Test this also by the Kolmogorov-Smirnov test.

d) Using an appropriate test, verify whether the median precipitation for seeded clouds is less than 300. Next, design and perform a test to check whether the fraction of the seeded clouds with the precipitation less than 30 is at most 25%.

Exercise 3. Concentrations of epinephrine

The concentrations (in nanograms per millimeter) of plasma epinephrine were measured for 10 dogs under *isofluorane*, *halothane*, and *cyclopropane* anesthesia, represented as three columns in data frame dogs.txt Download dogs.txt . We are interested in differences in the concentration for the different drugs.

a) Is it reasonable to assume that the three columns of *dogs.txt* were taken from normal populations?

b) Investigate whether the columns *isofluorane* and *halothane* are correlated. Apply relevant tests to verify whether the distributions of these columns are different. Is a permutation test applicable?

c) Conduct a one-way ANOVA to determine whether the type of drug has an effect on the concentration of plasma epinephrine. Give the estimated concentrations for each of the three anesthesia drugs.

d) Does the Kruskal-Wallis test arrive at the same conclusion about the effect of drug as the test in c)? Explain possible differences between conclusions of the Kruskal-Wallis and ANOVA tests.

Exercise 4. Hemoglobin in trout

Hemoglobin is measured (g/100 ml.) in the blood of brown trout after 35 days of treatment with four rates of sulfamerazine: the daily rates of 0, 5, 10 and 15 g of sulfamerazine per 100 pounds of fish, denoted as rates 1, 2, 3 and 4, respectively. (Beware that the levels of the factor rate are coded by numbers.) Two methods (denoted as A and B) of administering the sulfamerazine were used. The data is collected in data set hemoglobin.txt Download hemoglobin.txt .

a) Present an R-code for the randomization process to distribute 80 fishes over all combinations of levels of factors *rate* and *method*.

b) Perform the two-way ANOVA to test for effects of factors *rate*, *method* and their interaction on the response variable *hemoglobin*. Comment on your findings.

c) Which of the two factors has the greatest influence? Is this a good question? Consider the additive model. Which combination of rate and method yield the highest hemoglobin? Estimate the mean hemoglobin value for *rate 3* by using *method A*. What rate leads to the highest mean hemoglobin?

d) Test the null hypothesis that the hemoglobin is the same for all rates by a one-way ANOVA test, ignoring the variable *method*. Is it right/wrong or useful/not useful to perform this test on this dataset?

Exercise 5. Sour cream

The file cream.txt Download cream.txt contains data on an experiment to produce sour cream. Yogurt was placed in sweet cream, and yogurt bacteria were allowed to develop. Bacteria produce lactic acid, and as a surrogate for the number of yogurt bacteria, the acidity of the cream was measured. Interest was in the effect of the type of yogurt (denoted as *starter*) on *acidity*. The mixtures of yogurt and sweet cream were kept at constant temperature in a yogurt maker, in which five different positions could be used. The experiment was carried out with five batches of sweet cream, which were meant to have the same composition. With each batch each of five types of starter was used, with the yogurt placed in one of the five positions. The combinations of levels of three factors form a three-dimensional latin square. (You may need to install the R-package *lme4*, which is not included in the standard distribution of R.)

a) Analyze the data in a three-way experiment without interactions with *acidity* as response and *starter*, *batch* and *position* as factors. By using *summary* command, can you tell whether there is a significant difference between the effects of *starter 1* and *starter 2* on *acidity*?

b) Recall that the main interest is in the effect of starter on the acidity; factors *positions* and *batches* represent the block variables. Remove insignificant block variable(s) if there are such, and perform an ANOVA for the resulting "fixed effects" model. Which starter(s) lead to significantly different acidity?

c) For the model from b), can we also apply the Friedman test to test whether there is an effect of starter on acidity? Motivate your answer.

d) Repeat b) by performing a mixed effects analysis, modeling the block variable(s) (if there are any) as a random effect by using the function *lmer*. Compare your results to the results found by using the fixed effects model in b).