

Organisation

Lecturer: Dr. Evelyn Wallor
Email: evelyn.wallor@hnee.de
Exam form: Written exam (**group A**)
Submission deadline: February 09, 2022, midnight (CET)

General announcement

The data and exam announcement is provided on the moodle platform in the corresponding learning room. Related to the present Covid-19 situation the written exam is organized as an open book online examination. Students are divided into groups and work on separate exams. This announcement is relevant for students of group A.

Students work on the exam tasks by using the R software (or any other appropriate statistical software). Results for each exam task have to be transferred into a document and answers have to be formulated precisely and in full sentences! It should be clearly stated to which task answers belong. The final document – including all final solutions – has to be formatted as a PDF file and named as follows:

“Exam_FIT_YourName_WS2022.pdf”

The final PDF file including all final solutions has to be submitted via the moodle platform. Under the tile *Examination* a submission folder for each group is provided for the upload. The deadline for submission is February 09, 2022, at midnight (CET).

Exam tasks

Task 1 (4P)

From systematic observations in a tree nursery it is known that the germination rate of a certain apple tree species is 62 %.

- Which probability function / frequency distribution is appropriate to describe the above statement. Name the properties of the respective frequency distribution!
- What is the probability that from 10 apple tree seeds at least 6 will germinate? Provide the respective parameters and parameter values and write the result in an answer sentence!

Task 2 (6P)

In a certain case study area the variable *soil organic carbon content* is normally distributed with the following parameters: $\mu = 1520 \text{ mg kg}^{-1} \text{ soil}$ and $\sigma = 830 \text{ mg kg}^{-1} \text{ soil}$.

- What is the corresponding *z-value* of the z-standard normal distribution for an observed soil organic carbon content of $1200 \text{ mg kg}^{-1} \text{ soil}$?

- b) What is the probability that soil organic carbon content is between 1200 and 1800 mg kg⁻¹ soil?
- c) Which soil organic nitrogen content value defines the threshold of the upper quantile of 75 %?

Task 3 (8P)

This task is related to the data provided in the text file *DataA1.txt*. It contains measurement values of the diameter at breast height (*BHD*) [cm] for the tree species *Common douglas fir* measured at two different forest districts (BAR = Barnim, UM = Uckermark).

- a) Name the data scales of the features *BHD*, *Species*, and *ForestDistrict*!
- b) Provide a graph that illustrates the distribution of the measured feature *BHD*. Which graph is used and what are the class ranges of the lowest and uppermost *BHD* class?
- c) Apply an appropriate statistical test to proof if *BHD* values are normally distributed. Name the selected test, define the hypothesis, interpret the result, and conclude!
- d) Provide a box-whisker plot of measured *BHD* values for each forest district. Which statistical parameters of location are illustrated in a box-whisker plot? What are the respective *BHD* values for forest district Uckermark?

Task 4 (8P)

This task is again related to the data provided in the text file *DataA1.txt*. It is known, that on average, *Common douglas fir* stands in Northern Europe, and of similar age as the trees observed in Barnim and Uckermark, show a *BHD* in spring of 72 cm.

- a) Which statistical test is appropriate to estimate if the trees in the provided data match with the above mentioned average value? Name it precisely and explain the purpose of the test!
- b) Select one sample of *BHD* values (Barnim or Uckermark) and apply the test mentioned under a). Formulate the hypothesis and conclude from the test result!
- c) Which statistical test is appropriate to estimate if the trees' *BHD* values differ significantly between the two forest districts Barnim and Uckermark? Name it precisely and explain the purpose of the test!
- d) Proceed the test under c). Formulate the hypothesis and conclude from the test result!

Task 5 (3P)

Explain the differences and similarities between the *paired sample t-test* and the *Wilcoxon signed-rank test* with respect to purpose, preconditions, and data properties!

Task 6 (8P)

This task is related to the data provided in the text file *DataA2.txt*. It contains *volume* measurements of a certain tree species after harvest in [m³]. Before harvest, the *health state* of each tree was noted (scale from 5 = healthy to 10 = strongly affected).

- a) Name the data scales of the features *Volume* and *health state*!
- b) Which statistical approach is appropriate to look for differences in the *volume* values between the recorded *health states* of harvested trees? Name the respective approach precisely, explain the purpose, and the required assumptions!
- c) Proceed with the approach defined under b) including necessary statistical tests to proof the assumption. Formulate the hypothesis of the statistical approach mentioned under b). With respect to the result, formulate the conclusion!
- d) Based on the result under c), which additional statistical procedure would you recommend?

Task 7 (9P)

This task is related to the text file *projectdata.txt* provided to students during the group work (cf. moodle learning room, zipped folder on main page). Among others, the data consist of diameter (*DBH* [cm]) and height (*Height* [m]) measurements for several tree species growing on a marteloscope site close to the HNEE forest campus.

- a) Develop a simple regression model that describes the relationship between *Height* ~ *DBH* for the tree species *Copper beech*. Provide the full equation of the resulting regression model including the estimated coefficients!
- b) Explain what the coefficient of determination (R^2) and the residuals tell us about the quality of the developed regression model. How would you assess the quality of the regression model developed under a)?
- c) Visualize the estimated regression model together with the observations in a so-called regression plot!

I wish you much success ☺