

- (4 points) 1. The seed rate of a mechanical seeder (measured as number of seeds distributed per square meter) is known to be normal distributed with variance $\sigma^2 = 10000$. The aim is to estimate the unknown expected mean value m . On 30 plots of size 1 m^2 the number of seeds are counted

231	308	344	200	127	180
172	456	191	296	387	240
142	373	316	226	163	261
173	188	259	228	212	293
197	110	161	260	12	241

- (1 point) (a) Compute the point estimation for m .
- (1 point) (b) Compute the 95% Confidence interval for m .
- (1 point) (c) The farmer states that the mean seed rate is 250 seeds/ m^2 . Formulate the statistical hypotheses to show that the mean seed rate is unequal to 250 seeds/ m^2 .
- (1 point) (d) Illustrate "type I error" and "type II error" using the given problem based on the hypotheses formulated in 1 c).
- (e) Optional (1 bonus point): Produce a box plot and histogram of the seed rate and add a line for the point estimate from 1a) and for the upper and lower limit of the confidence interval from 1b).
- (9 points) 2. Assume you have the weights for 27 spinach leafs. In the literature for spinach a mean leaf weight of 10g is given. It is to be tested if the sample belongs to a population of leaves with a higher weight than cited in literature.
- (2 points) (a) Read the dataset from the file "SpinachLeafWeight.txt", describe the data by statistical parameters, and visualize them in a suitable way. Deal with outliers if appropriate.
- (1 point) (b) Compute the 99% Confidence interval for the leaf weight and interpret the results in regard to the hypotheses.
- (1 point) (c) Formulate the statistical hypotheses to answer the given problem.
- (1 point) (d) Give the right testing procedure and its assumptions.
- (2 points) (e) Test the assumptions. Use a significance level of $\alpha = 0.01$.
- (2 points) (f) Run the test procedure at significance level $\alpha = 0.01$ and interpret the results.
- (4 points) 3. Take another look at the OrchardSprays dataset from Assignment 2. Is there a significant difference in the decrease of sugar solution between treatment C and D?
- (2 points) (a) Check the data for outliers and normal distribution using visualizations and a statistical test. Use level of significance 0.05. Note your conclusions.
- (2 points) (b) Formulate the statistical hypotheses for this problem and compute the appropriate test. Interpret the result.
- (3 points) 4. Formulate the statistical hypotheses for the following problems and decide on the appropriate test procedure. Normal distribution of parameters is always assumed:
- (1 point) (a) The β -carotene content per 100 g carrots before and after cooking has been measured twenty times. Formulate the statistical hypotheses to test if the mean β -carotene content is decreased after cooking.
- (1 point) (b) The manufacturer states that his tins of beans contain on average 450g. Quality control has to test whether the mean content of the tinned beans differs from 450g. The variance of the content is assumed to be 36.

- (1 point) (c) The body length of 50 moths from an unknown species showed a mean of 2.75mm. The body length of 30 horse chestnut leaf miner moths showed a mean of 2.45mm. Is - on significance level 0.01 - the mean body length of the unknown species more than 10% greater than the mean body length of the horse chestnut leaf miner moth?

This exercise has 4 questions, for a total of 20 points and 1 bonus point.