

Seminar 9

1. An operator measures the length of one given operation in the production process in seconds. He measured the data on four machines. The lengths are available in the `R-script`.
 - (a) Create a boxplot of the time lengths for different machines. Do You think that the expected values of times are different?
 - (b) Try to decide the previous question using the pairwise t-tests for different machines (do not forget to check the assumptions). What is wrong with this approach?
 - (c) Test the hypothesis that the expected time lengths for all four machines are the same against the alternative, that it differs for at least one pair of machines using the analysis of variance (ANOVA) approach. Use the diagnostic graphs to check presumptions of the model.
 - (d) Use the post-hock Tukey test to compare the groups pairwise (if needed). Which pair of machines does probably differ in the expected values?
2. Use the similar data sample (available in the `R-script`) as before and solve the same subtasks (you can skip the subtask (b)). What has changed?
3. Use the data `Computers.csv` from seminar 2 to explore the relationship between categorical variables *RAM* and *screen*.
 - (a) Test the independence of *RAM* and *screen*. Create (and visualize) a contingency table and use the build-in R function to perform the test. What seems to be a problem?
 - (b) Create a new categorical variable for RAM with categories 2 MB, 4 MB, 8 MB, and 16 or more MB. Test the independence of *RAM* and *screen* using this new variable.
 - (c) Compute the Cramer's V coefficient (both manually and using the built-in R function).
 - (d) Perform correspondence analysis using R.