Master BeNeFri in Computer Science

Course: Statistical Learning Methods with R
Spring 2022

Exercise #3: *t*-test and R programming

Download from the ILIAS website the dataset "Mean20" (filename: Mean20.txt) This dataset is composed by a single variable (`time`), that records the time delay in minutes between two calls in an info-center.

1. Compute the mean, the median, the standard deviation, the minimum and maximum values of the variable `time`. Do you need to preprocess this list of values?

2. We suppose that the mean delay between two calls is 7.05 minutes. Can you test this hypothesis using the available data? What is your conclusion? Do you see a difference when considering the original values and the preprocessed values?

3. For Mary, the delay cannot be smaller than 7.05 minutes. Thus the only credible alternative hypothesis must take account of this (well-known) fact. How can you test Mary's hypothesis?

4. Define a function `secondMax(x)`, where `x` is a vector, returning the second largest value contained in `x`. If `x` is not a vector, return an error message. Test your implementation in different cases using the Mean20 dataset.

5. Define a function `mySummary(x)`, where `x` is a vector composed by the mean, the median, the standard deviation, the minimum and the maximum values (in this order). Test your implementation in different cases using the Mean20 dataset.