Master BeNeFri in Computer Science

Course: Statistical Learning Methods
Spring 2016

Exercise #3: R programming

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

1. Program the function `secondMax(x)` (where x is a vector) returning the second largest value contained in the vector x. If x is not a vector, return an error message. Explicitly test your implementation with different cases. Test your function with the Mean20.txt dataset.

2. Program a function `mySummary(x)` (where x is a vector) returning the vector composed by the mean, the median, the standard deviation, the minimum, and the maximum value (in that order). Explicitly test your implementation with different cases.

3. You need to generate 20 samples of values. Each sample must contain 20 values and must be drawn from a Gaussian with mean = 2, stdev = 1.5. For each sample, apply the `t.test()` to test if the population mean = 1.5. As you know the real population mean, does the test always return the correct answer? Finally, regroup all the 20 samples into one (big) sample and apply the `t.test()`. What is your conclusion?