Statistical Learning, Exercise 3 : R programming

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

Question 1

Create a function called **secondMinSquared(x)** (where x is a vector) returning the second smallest value squared 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 data set.

Question 2

Create a function called **vecSummary(x)** (where x is a vector) returning a vector composed of the mean, median, variance, the minimum, and the maximum value (in that order). Explicitly test your implementation with different cases.

Question 3

Create a function called **generateSample(n_samples, mean, stdev)** returning *n_samples* values drawn from a Gaussian with mean *mean* and standard deviation *stdev*. Test your implementation with different cases.

Question 4

Generate 20 samples of values with the function **generateSample** from the preceding question with $mean = 2$ and $stdev = 1.5$. For each sample, apply a 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? Regroup all the 20 sample into one big sample and apply the t-test. What is your conclusion?