CS306-Data analysis and visualization
Assignment 1

Problem description:

Parametric testing based on t-test requires three assumptions:

1. Assumption of normality
2. Homogeneity of variance
3. Data independence

These are required so that the sampling distribution of t follows the theoretical t-distribution with the corresponding degree of freedom. The goal is to verify the role of first two assumptions. To that end, obtain the sampling distribution of t in four cases, and analyze the role of the said assumptions: (a) normal samples with similar variances, (b) non-normal samples with similar variances, (c) normal samples with very different variances, (d) non-normal samples with very different variances.

Your analysis should provide answers for the following:

1. is sample normality or population normality required for t-test?
2. is homogeneity of variance necessary?
3. does your answer to previous question depend on whether the sample sizes are equal or not?
4. what are the implications if one performs t-test and one or both assumptions are violated. (hint: observe the experimental and theoretical t-distribution and see how the deviations between the two affects the decision about null hypothesis)