4-9-2005

Violation of the normality assumption: How serious is it?

Sio F. Kong
Indiana University - Purdue University Fort Wayne

Follow this and additional works at: http://opus.ipfw.edu/stu_symp2005

Part of the Mathematics Commons

Recommended Citation
Sio F. Kong (2005). Violation of the normality assumption: How serious is it?.
http://opus.ipfw.edu/stu_symp2005/12
Violation of the normality assumption: How serious is it?
Sio F. Kong
C. K. Chauhan
Mathematics
Indiana University-Purdue University Fort Wayne

The t test is a well known tool for testing the hypothesis for the unknown mean of a population. Such tests are conducted in the areas of Biology, Psychology, Education, Business, and so forth. In order for the t test to be valid, the following assumptions are made: a) the observations are collected randomly, and b) the population from which the data are collected is normally distributed. This research paper investigates the consequences of the violation of the second assumption.

Many experimenters use the t test without ever checking the normality assumption. Checking the normality assumption is also difficult when the sample is small. However, it is worth noting that the effect of departure from the normality is more serious for smaller samples. How do we measure the effect of the violation from the normality when conducting a t test? Well, in a hypothesis test, two types of errors, known as type 1 and type 2 errors are involved. While an experimenter selects the probability of type 1 error in advance (usually 5%), the probability of type 2 error is calculated for different values of the alternatives of the test. A sound test is considered good if the probabilities of errors are small.

In this project we are going to conduct simulation study in which,
a) We will generate 1000 samples of a size n (n = 6, 8, 10, 12, 15, 20) from several different populations such as normal, uniform, exponential, and heavy tailed.
b) For each distribution, a t test will be conducted, and the probabilities of type 1 and type 2 errors will be calculated.

The above simulation study will help us estimate the probabilities of two types of errors when the data are drawn from non normal distributions. The results thus obtained will be compared with the probabilities of errors when the data were drawn from a normal population. The effects of the sample size and the standard deviation on the probabilities of errors for each of the above distributions will also be investigated.