**STAT 575 Final Project Report**

**Introduction**

The purpose of this project is to compare the effectiveness of different tests of multivariate normality through comparing their empirical power.

The power π of a statistical test is the probability that such a test will detect an effect when there is such an effect to be detected, or a little more formally, π = 1-P(Type II Error). Thus, the higher the power of a test, the more effective it is when detecting for true effects.

Formally, the hypothesis test will run as follows, with N being the family of multivariate normal distributions: . The four tests that will be compared are Mardia’s multivariate skewness and kurtosis tests, the Shapiro-Wilk test, and the Energy test.

One way to measure normality is the skewness of the distribution, which is a measure of the asymmetry of a distribution around its mean. Mardia’s multivariate skewness test will be using the following statistic:

where is the MLE of covariance. Its asymptotic distribution is as follows: where d is the number of dimensions. The higher is, the more skewed the distribution is. In other words, if , then the null hypothesis is rejected. (Rizzo, 2019)

Another way to analyze a distribution’s normality is by analyzing its kurtosis, which is a measure of its shape or “tailedness”. Low kurtosis values—platykurtosis—exhibit skinnier tails, less outliers and more values centered around the mean and high values exhibit thicker tails. High kurtosis values—leptokurtosis—exhibit more outliers, thicker tails, and a wider main body.

The statistic for Mardia’s multivariate kurtosis test is as follows:

The statistic is asymptotically distributed . Higher levels of the statistic indicate a higher spread than usual for the distribution and are more significant. The null hypothesis is rejected if . (Rizzo, 2019)

The Shapiro-Wilk test is based on “the regression of the sample order statistics on their expected values under normality.” The critical values are approximated from a normal transformation of the test statistic. (Rizzo, 2019)

Finally, the Energy test is based on “an energy distance between the sampled distribution and normal distribution”. The bigger these values are, the more significant they are. The energy statistic for testing normality is given below. (Rizzo, 2019)

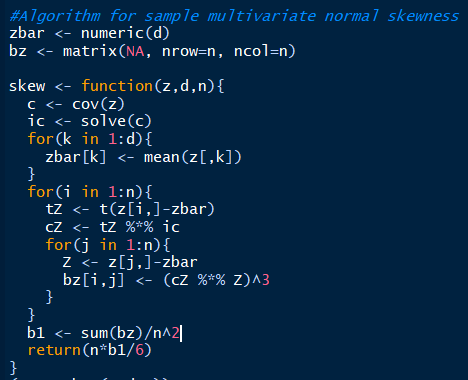
All of these tests will be compared against a bivariate normal contaminated mixture:

When ε=0 or ε=1, the distribution is normal. Otherwise, the distribution is non-normal. These are the alternative distributions against which the empirical tests will be compared. In order to randomly generate a multivariate normal distribution, the mvtnorm package in R was installed in order to use the rmvnorm() function.

**Analysis:**

This section will describe the steps needed to implement this simulation, especially the algorithms needed to make this simulation happen.

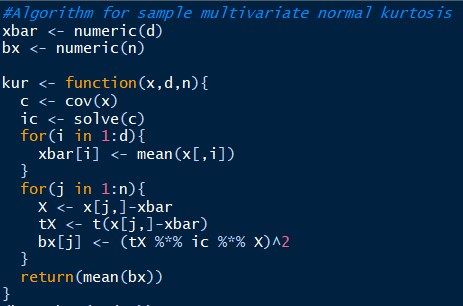
The first algorithm, which gives the multivariate skewness statistic, is given below:



The *zbar* will hold the *1xd* matrix containing the mean for each column, while the *bz* is a *nxn* matrix which will hold each calculation where each is multiplied with each (see skewness equation earlier).

The *skew* algorithm is a function that accepts three parameters: z – a vector, d – number of dimensions, and n – sample size. It starts off by taking the covariance of the vector passed into it, then it takes the inverse of the covariance for the formula. Then it calculates each row mean corresponding to each to be used in the formula inside the first for loop. The next for loop calculates each for each *ith* observation.Before this outer for loop iterates, there is an inner for loop that multiplies the *ith cZ* by each for each *jth* observation and then cubes the whole product. This keeps iterating until all the n rows of the data are multiplied with each other. Each value is summed up, divided by , and then multiplied by n and divided by six for the comparison to the critical value.

The next step was to create a multivariate kurtosis statistic algorithm:

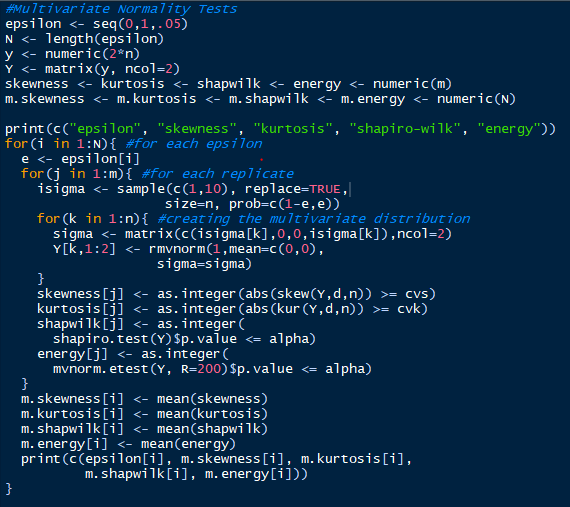


Just like with the skewness algorithm, a *1xd* matrix was created to store the mean of each and an n-length vector *bx* was created to store each calculation for the summation in the kurtosis equation.

The *kur* function passes in the same three parameters as the skew algorithm: x-data, d-number of dimensions, and n-sample size. It first calculates the covariance of the dataset matrix *x* and its corresponding inverse matrix. It then calculates by taking the mean of each column in the first for loop. The second for loop calculates for each *ith* observation in the dataset and stores each iteration in the *bx* vector. The mean of this vector is then returned as the final kurtosis statistic.

The Shapiro-Wilk Test was run using the shapiro.test() function. No special packages needed to be installed. The Energy Test required the use of the mvnorm.etest() function from the *energy* package.

Finally, it was time to run the simulation with all the tests in place. The simulation algorithm is given below.

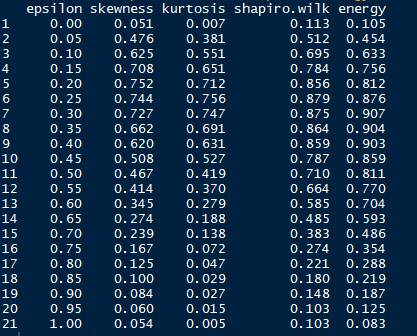


The algorithm will cycle through epsilon values from 0 to 1 in increments of .05. A matrix *Y* was created to store the generated two-variate normal distribution data. Vectors were created to store each test result for each replicate of the algorithm. Vectors were also created to store the mean of each of the test result vectors for each value of epsilon. This was done so that this data could be represented in a data frame after the simulation was run.

The algorithm is a triple nested for loop. The outermost loop represents each value of epsilon being tested on the contaminated mixture. The next loop replicates each test m number of times for each epsilon. For each replication, a sample of size n is created that randomly chooses values of 1 or 10 (possible sigma values) based on probabilities using the current epsilon value. For example, if ε=.2, then about 80% of the values in the sample will be 1 and about 20% will be 10. The innermost loop then takes each value in the sample, creates the sigma matrix with the individual sample values as the sigma values, and generates a multivariate normal data point using the created sigma matrix. This is repeated for all sample points. This creates a bivariate normal distribution dataset of sample size n. The tests are then run for each dataset and stored in their corresponding test storage vector. This is repeated m times. The mean of each test vector for each epsilon is then stored in their corresponding vector. All of this is repeated for each epsilon value.

**Results**

Using n=30 , m=1000, d=2 dimensions, and alpha=0.1, a run of the final algorithm produced the following results:

**A close up of a map

Description automatically generated**

For ε < 0.15, the powers of each test are roughly the same. It is around this epsilon value that the Shapiro-Wilk and Energy tests begin to separate themselves from Mardia’s skewness and kurtosis tests and are generally more powerful than the other two against this contaminated mixture for all other values of epsilon. The peak power for the S-W and Energy tests comes in at around 0.9 between epsilon values of 0.25 and 0.4, while the peak powers for the skewness and kurtosis tests come in at around 0.75. For ε > 0.3 or so, the energy test becomes the most powerful test over the S-W test. Also, around ε = 0.5, the skewness test becomes more powerful than the kurtosis test and stays in third place for most powerful test for the rest of the epsilon values.

**Conclusions and Limitations of Study**

After conducting the simulation, it was found that the energy test was the most powerful test in terms of detecting non-normality effects in a dataset when there was actually non-normality effects to be detected. This is closely followed by the Shapiro-Wilk test, with Mardia’s skewness and kurtosis tests figuring to be significantly less powerful and less effective at detecting true non-normality effects against the mixture than the other two tests.

One of the most notable limitations of this study was the relative inefficiency of the algorithms. For n=30 samples and m=1000 replications, it would take at least 5-10 minutes for the algorithm to completely run its course. One of the potential reasons for this was the way in which the random normal mixture was generated in that the random mixture was generated one point at a time. When this runs 1000 times for the 21 epsilon values tested, this can potentially take a while. This algorithm can thus be possibly rendered more efficient given more time or through the use of random multivariate distribution generation packages.

Also, only a handful of tests were carried out. Future study of this topic would analyze more normality tests in the pursuit of a potentially more powerful test than even the energy test. These tests would also be tested against other distributions to find out if the power results from this contaminated normal mixture distribution are consistent and carry over to other distributions.

**STAT 575 Final Project Report Abstract**

For this project, I set out to learn which of some of the most well-known multivariate normality tests (Mardia’s skewness and kurtosis tests, the Shapiro-Wilk test, and the Energy test) are most powerful at detecting true non-normality through the analysis of power. Each of these were tested against a contaminated bi-variate normal mixture. Algorithms were created for the Mardia skewness and kurtosis test statistics, while packaged R functions were used for the other two tests. A simulation was run with n=30 samples in each bivariate normal mixture and m=1000 replicates at different weights of each part of the mixture that tested the power of each test at each weight. After a run of the simulation, it was found that the energy test was the most powerful at detecting true effects of non-normality, followed by the Shapiro-Wilk test, Mardia’s skewness test, and Mardia’s kurtosis test.

# Bibliography

Rizzo, M. L. (2019). *Statistical Computing with R, Second Edition.* CRC Press.