Monte Carlo Simulation Study on the Power of Two-Sample Hotelling's Test

Gulumbe S.U.[1], Danbaba A[1]. and Boyi A.D.[2]
[1]Mathematics Department, Usmanu Danfodiyo University, Sokoto
[2]Mathematics Department, Polytechnic Birnin Kebbi, Kebbi State

Abstract: Hotelling's T^2 statistic is one of the multivariate tests that are based on certain fundamental assumptions, one of which is the multivariate normality. The test is usually applied to investigate the equality of means associated with two normal multivariate populations. The application of the test when the normality assumption is violated may affect the power of the test. In this paper, a Monte Carlo simulation is conducted to compare the exact powers of T^2 based on multivariate normal and four multivariate non-normal populations. The results show insignificant differences in the average power performance between the normal and non-normal populations for both small and large sample sizes. It was also observed that the results from multivariate normal distribution do not show any unique performance from non-normal distributions when samples sizes are large.

Keywords: Hotelling's T^2. Power. Two-sample. Monte Carlo Simulation.

INTRODUCTION

In multivariate analysis, when *p*-variate observations come from two multivariate normally distributed populations with common variance-covariance matrix, Hotelling's T^2-statistic may be used to test the equality of the vectors of means associated with the two samples. Hotelling's T^2 has been one of the most important statistical tools for comparison of means between two multivariate samples, Hotelling (1931). Many practitioners and applied statisticians are increasingly using Hotelling's T^2 in their comparative analyses with samples dimensions, p =2. Works such as Huizenga *et al.* (2007), Wu *et al.*, (2006) and Timm (1975) shows that Hotelling's T^2 is insensitive if departures from independence, multivariate normality and variance-covariance matrices assumptions do not greatly affect the significance level.

The purpose of this paper is to investigate the power of the Hotelling's T^2 when the normality assumption is violated. That is to say, we depart from normality by applying the test to non-normal populations. Zech and Asian (2003) reported that the performances of various statistical tests were usually assessed for finite sample sizes by Monte Carlo simulation. In so doing, Monte Carlo simulation study was conducted using pseudo-random numbers generated from multivariate exponential, uniform, lognormal, poison and normal distributions to compute the exact power of Hotelling's T^2 Since efficiency and effectiveness in the use of any test statistic can be measured in terms of its power, the formula proposed by Kelsey *et al* (1996) and Kramer (1972) for computing exact power of T^2 can serve as a tool to work out what happens when Hotelling's T^2 is used to compare samples from multivariate non-normal populations. The formula used to compute exact power of T^2 according to Kelsey *et al* (1996) and Kramer (1972) is

$$T_{power} = \sqrt{\frac{(n_1 n_2)}{(n_1 + n_2)} ds^{-1} d^1} - \sqrt{\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p,n_1,+n_2-p-1,\alpha}}$$

where n_1, n_2, d, s, p and are size of the first sample, size of the second sample, difference between first and second samples mean vectors, pool variance-covariance matrix of the samples, dimension and level of significance respectively.

Simulation Study

In order to asses the exact power of Hotelling's T^2 simulations were conducted. The powers were calculated and average power were plotted against the sample sizes for each population generated under the following conditions: significant level of the test (0.01, 0.05), sample sizes n_1, n_2 (3, 6, 9, 12, 15, 18, 21, 24) = 3(3)24 and n_1, n_2 (25, 50, 75, 100, 125, 150, 175, 200) = 25(25)200, and dimensions of random variables p (2, 5, 10). A total of 5000 random samples of multivariate normal, exponential, uniform, lognormal and Poisson data sets were generated for each combination of , n_1, n_2, and p. For example, in computing with a sample of size 3, will require at alpha level 0.01 and p for all the five distributions. This is same for alpha = 0.05 for all the five distributions. However, a program is developed in *R* programming system, so that a run of the program gives an average power of Hotelling's T^2 with respect to each of the distributions.

RESULTS

For each replication of sample size in 3(3)24 and 25(25)200 at 5% and 1% significance levels, the average power generated were plotted against sample sizes as in Figures 1 through 6. These figures shows that all the five distributions have almost equal power to reject the null hypothesis when it is false for both small and large sample sizes. It was also observed that the power at = 0.05 is more than that at = 0.01 level of significance.

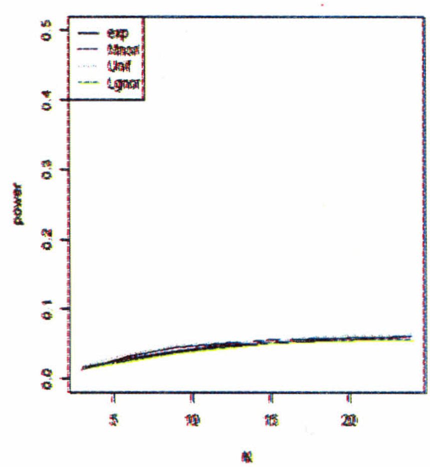

Fig . 1:Average exact power of T^2 at $\alpha = 0.01$ with N =3(3)24 & 2-dimensional variables

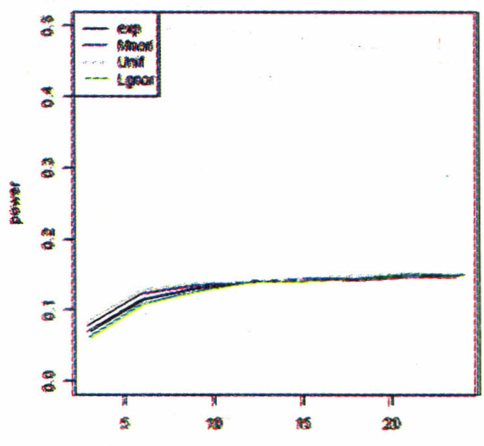

Fig . .2:Average exact power of T^2 at $\alpha = 0.05$ with N =3(3)24 & 2-dimensional variables

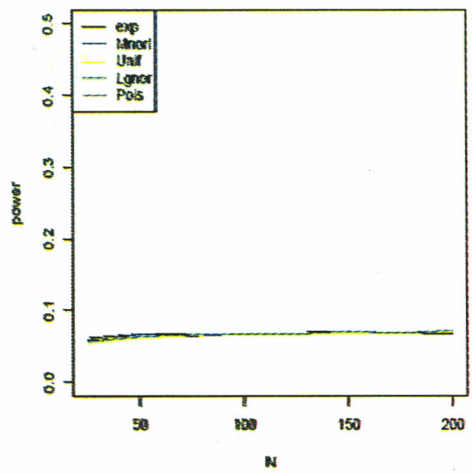

Fig . 3:Average exact power of T^2 at = 0.01
with N =25(25)200 & 2-dimensional variables

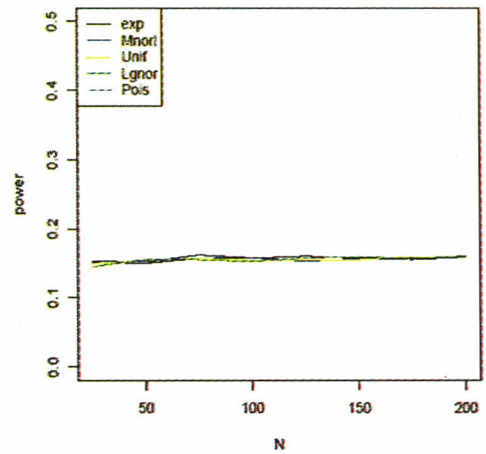

Fig . 4:Average exact power of T^2 at = 0.05
with N =25(25)200 & 2-dimensional variables

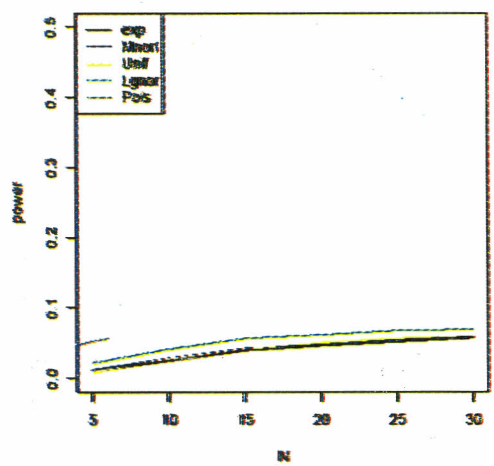

Fig . 5:Average exact power of T^2 at = 0.01 \
with N =25(25)200 & 5-dimensional variables

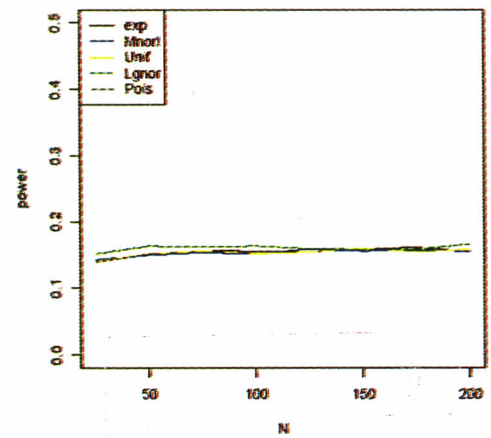

Fig ..6:Average exact power of T^2 at = 0.05\
\ with N =25(25)200 & 5-dimensional variables

The averages of exact power of Hotelling's T^2 on each 5000 multivariate samples generated from five distributions are presented in Table 1. Tnorl, Texp, Tunif, Tlog and Tpoi denote the power generated under multivariate normal, exponential, uniform, lognormal and Poisson data respectively. It was observed from this table that for fairly large samples, average power is the same. In addition average power is higher when significance level at =0.05 is used than =0.01. The results for p = 10 are similar to that of p = 5 and therefore not reported.

Table 1: Average exact power of Hotelling's T^2 on each 5000 Multivariate Samples

	Texp		Tnorl		Tunif		Tlog		Tpoi	
Samples and p	α = 0.01	α = 0.05	α = 0.01	α = 0.05	α = 0.01	α = 0.05	α = 0.01	α = 0.05	α = 0.01	α = 0.05
3(3)24 p =2	0.04	0.13	0.05	0.13	0.05	0.14	0.04	0.13	-	-
25(25)200 p =2	0.07	0.16	0.07	0.16	0.07	0.16	0.07	0.16	0.07	0.15
25(25)200 p = 5	0.07	0.15	0.07	0.15	0.07	0.15	0.08	0.16	0.07	0.15

DISCUSSION

Result in Fig.1 shows an insignificant variation in the average power when , but stabilizes when . It also shows that power increases insignificantly and become almost equal as the sample size increases. However, result in Fig.2 shows a fairly higher performance in average powers of the distributions than in Fig.1. For instance, Fig.1 result shows that a sample of size will have power while in the result of Fig.2, a sample of will have power. This is because the significance level is taken at 0.05 in Fig. 2 result. On the other hand, results in Fig.3, Fig.4, Fig.5 and Fig.6, where sample sizes are used do not show any significant difference in the average power among the distributions. This is because most distributions converge to normal when sample sizes are large as stated by the central limit theorem. Also results generally show higher performance in average power when significant level is 0.05. For instance, when sample of sizes produced powerat 0.01 level of significant as in Fig.3 a sample of size produced poweras in Fig.4. The same observations apply to results in Fig.5 and Fig.6.

CONCLUSION

In this study, Monte Carlo simulation study was conducted to investigate the exact power performance of Hotelling's T^2 test involving two sets of sequences of sample sizes, 3(3)24 and 25(25)200 each repeated 5000 times from multivariate normal, exponential, uniform, lognormal and Poisson populations. It was found that the average exact power performance of the on all the populations are almost the same. In addition, some insignificant differences in the average exact power performance among the populations occur when sample sizes are very small. However, when sample sizes are large, the average exact power performance tends to be exactly the same for all the populations. This conforms to the central limit theorem. Furthermore, it was generally observed that the average power of appears to be higher at 5% level of significance than at 1% level of significance in all the sample sizes used for the simulation.

REFERENCES

Hotelling, H. (1931). The Generalization of Student's Ratio. *Annals of Mathematical Statistics,* 2, 360 378.

Huizenga, H.M., Smeding, H., Grasman, R.P.P.P., Schmand, B. (2007), Multivariate Normative Comparisons. *Neuropsychologia* 45, 2534 2542

Kelsey, J., Whittemore A., Evans, A., and Thompson, W. (1996). *Methods in Observational Epidemology*. Oxford, New York

Kramer, C. Y. (1972). *A course in Methods of Multivariate Analysis*. Virginia Polytechnic institute and State University, Blamlburg VI

Timm, N. H. (1975). *Multivariate Analysis with Applications in Education and Psychology*, Brooks/Cole, Monterey, California.

Wu, S.S., Li, H. and Casella, G. (2006). Tests with Optimal Average Power in Multivariate Analysis, *Statistica Sinica,* 16, 255 266

Zech, G. and Asian, B. (2003). A Multivariate Two-Sample Test Based on the Concept of Minimum Energy. *Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology, SLAC-R-703,* Stanford, California