Evaluation of Hotelling T^2 and Multivariate Analysis of Variance Tests

G. M. Oyeyemi[1], P.O. Adebayo[2] and A. I.Folorunsho[3].
Department of Statistics, University of Ilorin, Ilorin, Nigeria
Department of Maths and Statistics Osun State Polytechnics Iree
[1]gmoyeyemi@gmail.com
[2]bayooni3@gmail.com
[3]folorunsoidowu@gmail.com

Abstract: *Multivariate Analysis of Variance (MANOVA) which comprises of Wilks' lambda, Pillai's trace, Lawley-Hotelling trace and Roy's largest Root are compared with Hotelling T square when null hypothesis is true. Data were simulated to compared the five (5) test statistic under the two different distribution (Multivariate Gamma and multivariate Normal), sample size (10, 30, 60, 90, 400, 500, 800, 1000), number of variables p = 2 and equal and unequal sample size and variance co – variance matrix. The comparisons were done at two level of significant (α = 0.01 and 0.05) using power of the test and type I error rate. The result showed that Roy's largest Root test statistic is better than all other test statistic considered when sample size are equal but Hotelling T square performed better for unequal sample size.*

Keywords: *Multivariate Analysis of variance, Hotelling T square, Type I error rate, power of the test, normality.*

Introduction

The purpose of a t-test is to assess the likelihood that the means for two groups are sampled from the same sampling distribution of means. The purpose of an ANOVA is to test whether the means of more than two groups are taken from the same sampling distribution. The multivariate equivalent of the t test is Hotelling T square. Hotelling T^2 tests whether the two vectors of means for the two groups are sampled from the same sampling distribution. MANOVA is the multivariate analogue to hotellting'T^2. The purpose of MANOVA is to test whether the vectors of means of more than two groups are sampled from the same sampling distribution.

MANOVA and Hotelling T^2 are rest on three basic assumptions regarding the population: multivariate normality, equality of group population covariance matrices and independence of errors. When these assumptions are violated, MANOVA does not perform well with respect to Type I error and power. Accurate use and interpretation of these multivariate test statistics is dependent upon the assumptions of independent errors, multivariate normality, and homogeneity of group covariance matrices. When these assumptions are met, the tests perform similarly well with respect to controlling Type I error rates and maintaining appropriate statistical power, particularly in studies with relatively large sample sizes (e.g., [1], [6], [10], [16], [19]).

Investigations of Type I error rates and power have suggested that these multivariate tests may not perform well when there are violations in assumptions of multivariate normality and equality of covariance matrices (e.g., [5], [6], [14], [12], [15]). Perhaps most notable is the performance of Hotelling's T^2 in studies of unequal sample sizes when the assumptions of multivariate normality and particularly equality of covariance matrices has not been met. But[9] and [8] have investigated the robustness of Hotelling's T^2 statistic when violating either independent or multivariate normality for large sample sizes. No extensive work has been conducted in this area for smallsample sizes. In such cases, the Hotelling's T^2 demonstrated diminished power as the degree of skewness of the response variables increased [3]. Furthermore, when the groups' covariance matrices were not homogeneous, the Type I error rate of the Hotelling's T^2 was inflated when the groups were not of equal size and the smaller group had the larger variances [5], [6]. These results for Hotelling's T^2 are similar to those reported in studies of the performance of Pillai's Trace, Wilk's Lambda, Hotelling-Lawley's Trace and Roy's Greatest Root when there are violations in the assumption of equality of covariance matrices [4], [14], [18]. When the smaller group had the larger variance the Type I error rates were inflated, whereas when the larger group had the larger elemental covariance elements, there was a reduction in power. Non-normality characterized by relatively severe skewness also resulted in a reduction of power [3] ,[4]. Furthermore, when the assumptions were violated, Pillai's Trace was relatively more robust in terms of Type I error rate control compared to Wilk's Lambda and Hotelling-Lawley's Trace but exhibited somewhat lower power compared to these other tests. Not one of the common MANOVA statistics can be clearly identified as the single best test for use in all situations [12],[15].

The purpose of this study is to identify the conditions under which each of the five is more robust when the assumption of normality and equality of variance co-variance matrix hold or violated using power and type I error rate. All these will be considered when null hypothesis (H_o) is true and when the number groups (g) and random variables (p) are two.

METHODOLOGY

This work targeted at comparison of Hotelling's T^2 and four multivariate analysis of variance (MANOVA) test statistic which are Wilks' Lambda, Pillae's trace, Roy's largest root and Lawley's trace using R statistics.

1. wilks' lambda

$$\Lambda = \frac{|E|}{|H+E|}$$

$$a = N - g - \frac{p-g+2}{2}$$

$$b = \begin{cases} \sqrt{\dfrac{p^2(g-1)-4}{p^2+(g-1)^2-5}}; & if\ p^2+(g-1)-5 > 0 \\ 1 & if\ p^2+(g-1)-5 \le 0 \end{cases}$$

and

$$c = \frac{p(g-1)-2}{2}$$

. then

$$F = \left(\frac{1-\Lambda^{\frac{1}{b}}}{\Lambda^{\frac{1}{b}}}\right)\left(\frac{ab-c}{p(g-1)}\right) \sim F_{p(g-1),ab-c,\alpha}$$

2. Hotelling – lawley Trace

$$T_o^2 = trace(\boldsymbol{HE^{-1}}) = \sum_{i=1}^{s} \lambda_i$$

Lets

$$s = \min(p, g-1)$$

$$t = \frac{|p-g-1|-1}{2}$$

and

$$U = \frac{N-g-p-1}{2}$$

then $F = \frac{2(su+1)}{s^2(2t+s+1)} T_o^2 \sim F_{s(2t+s+1),2(su+1),\alpha}$

3. Pillai Trace

$$V = trace(\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{E})^{-1}) = \sum_{i=1}^{s}\left(\frac{\lambda_i}{1+\lambda_i}\right)$$

here

$$F = \left(\frac{2u+s+1}{2t+s+1}\right)\left(\frac{V}{s-V}\right) \sim F_{s(2t+s+1),s(2u+s+1),\alpha}$$

Where $s = \min(p, g-1)$

4. Roy's largest Root

$$\theta = (\boldsymbol{HE^{-1}}) = \frac{\lambda_1}{1+\lambda_1}$$

$$F = \left(\frac{2U+2}{2t+2}\right)\phi_{max} \sim F_{(2t+2),(2u+2),\alpha}$$

$$\phi_i = \frac{\theta_i}{1-\theta_i} = \frac{1-\lambda_i}{\lambda_i}$$

$$\lambda_i = 1 - \theta_i$$

5. Hotelling T Square

$$T^2 = (\bar{x}_1 - \bar{x}_2)' \left[S_p\left(\frac{1}{n_1} + \frac{1}{2}\right)\right]^{-1} (\bar{x}_1 - \bar{x}_2)$$

$$S_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$F = \frac{n_1 + n_2 - p - 1}{p\,(n_1 + n_2 - 2)} T^2 \sim F_{p, n_1+n_2-p-1, \alpha}$$

Methods

A simulation using R package was conducted in order to estimate the power of the test and Type I error rate for each of the previously discussed multivariate analysis of variance (MANOVA) test statistic (Wilks' Lambda, Pillae's trace, Roy's largest root and Lawley's trace) and Hotelling T square. In each of the four different scenarios, that is ,when: null hypothesis is true, dataset are normal or not ,equality of variance co-variance matrix hold or not. Two factors were varied in the simulation: Sample size (n) and significant levels (α).

Data Generation

In each of the 1000 replications and for each of the factor combination, n_1 x p data matrix X_1 and n_2 x p data matrix X_2 were generated using an R package for Multivariate Normal and Gamma. The programme also performs the [2] test for equality of covariance matrices by using the statistic:

$$M = c \sum_{i=1}^{k} (n_i - 1) \log\left|S_i^{-1} S_p\right| ,$$

Where

$$S_p = \frac{\sum_{i=1}^{k} (n_i - 1) S_i}{n - k}$$

and

$$c = 1 - \frac{2p^2 + 3p - 1}{6(k-1)(p+1)} \left[\sum_{j=1}^{k} \frac{1}{n_j - 1} - \frac{1}{n-k}\right]$$

$$\chi_B^2 = (1 - C)M$$

And $S_i$and S_p are the $i - th$ unbiased covariance estimator and the pooled covariance matrix respectively.

Box's M also has an asymptotic chi-square distribution with $\frac{1}{2}(p+1)(k-1)$ degree of freedom.

Box's approximation seems to be good if each $n_i$exceeds 20 and if g and p do not exceed 5 [13]

H_o is rejected at the significance level α if $\chi_B^2 > \chi_{\alpha(v)}^2$.

Results

Table 1: Multivariate Normal

Power												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Equal variance co-variance matrix	Equal sample	10,10	0.012	0.012	0.012	0.012	0.012	0.143	0.138	0.143	0.130	0.143
		60,60	0.060	0.059	0.060	0.058	0.060	0.103	0.103	0.103	0.103	0.103
		400,400	0.348	0.348	0.348	0.342	0.348	0.055	0.055	0.055	0.055	0.055
		1000,1000	0.110	0.110	0.110	0.110	0.110	0.114	0.114	0.114	0.114	0.114
	Unequal	10,30	0.185	0.179	0.185	0.142	0.128	0.158	0.154	0.158	0.149	0.129
		60,90	0.035	0.035	0.035	0.035	0.034	0.062	0.062	0.062	0.062	0.062

		400,600	0.068	0.068	0.068	0.068	0.067	0.601	0.601	0.601	0.595	0.596
		800,1000	0.152	0.152	0.152	0.152	0.150	0.547	0.547	0.547	0.544	0.541
Type I Error Rate												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Equal variance co-variance matrix	Equal sample	10,10	0.014	0.008	0.014	0.000	0.014	0.058	0.048	0.058	0.009	0.058
		60,60	0.020	0.020	0.020	0.013	0.020	0.096	0.095	0.096	0.085	0.096
		400,400	0.189	0.189	0.189	0.187	0.189	0.389	0.389	0.389	0.385	0.389
		1000,1000	0.582	0.582	0.582	0.580	0.582	0.815	0.815	0.815	0.814	0.815
	Unequal sample	10,30	0.028	0.025	0.028	0.005	0.007	0.101	0.097	0.101	0.074	0.055
		60,90	0.026	0.026	0.026	0.021	0.021	0.110	0.110	0.110	0.101	0.102
		400,600	0.230	0.230	0.230	0.225	0.225	0.423	0.423	0.423	0.422	0.419
		800,1000	0.518	0.517	0.518	0.512	0.509	0.750	0.750	0.750	0.749	0.745
Power												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Unequal variance co-variance matrix	Equal sample	10,10	0.035	0.033	0.035	0.031	0.035	0.115	0.111	0.115	0.108	0.115
		60,60	0.046	0.046	0.046	0.046	0.046	0.149	0.148	0.149	0.147	0.149
		400,400	0.350	0.349	0.350	0.343	0.350	0.086	0.086	0.086	0.086	0.086
		1000,1000	0.112	0.112	0.112	0.111	0.112	0.073	0.073	0.073	0.073	0.073
	Unequal sample	10,30	0.102	0.099	0.102	0.087	0.072	0.593	0.581	0.593	0.461	0.468
		60,90	0.151	0.149	0.151	0.142	0.143	0.077	0.076	0.077	0.076	0.075
		400,600	0.075	0.075	0.075	0.074	0.074	0.535	0.535	0.535	0.530	0.530
		800,1000	0.056	0.056	0.056	0.055	0.055	0.057	0.057	0.057	0.057	0.057
Type I Error Rate												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Unequal	Equal	10,10	0.020	0.018	0.020	0.000	0.020	0.061	0.055	0.061	0.018	0.061
		60,60	0.021	0.019	0.021	0.016	0.021	0.080	0.077	0.080	0.068	0.080
		400,400	0.07	0.072	0.07	0.06	0.07	0.17	0.178	0.17	0.17	0.17

			2		2	9	2	9		9	4	9
		1000,1000	0.179	0.179	0.179	0.176	0.179	0.430	0.430	0.430	0.430	0.430
	Unequal sample	10,30	0.086	0.082	0.086	0.045	0.047	0.209	0.204	0.209	0.163	0.139
		60,90	0.031	0.030	0.031	0.026	0.028	0.082	0.081	0.082	0.074	0.076
		400,600	0.111	0.111	0.111	0.110	0.110	0.245	0.243	0.245	0.241	0.239
		800,1000	0.196	0.196	0.196	0.194	0.192	0.407	0.407	0.407	0.406	0.402

Table 2: Multivariate Gamma

Power												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Equal variance co-variance matrix	Equal sample	10,10	0.051	0.048	0.051	0.043	0.051	0.052	0.051	0.052	0.052	0.052
		60,60	0.024	0.024	0.024	0.024	0.024	0.238	0.236	0.238	0.231	0.238
		400,400	0.029	0.029	0.029	0.029	0.029	0.086	0.086	0.086	0.086	0.086
		1000,1000	0.073	0.073	0.073	0.073	0.073	0.110	0.110	0.110	0.110	0.110
	Unequal sample	10,30	0.429	0.416	0.429	0.272	0.301	0.124	0.122	0.124	0.120	0.104
		60,90	0.046	0.046	0.046	0.045	0.044	0.106	0.106	0.106	0.106	0.104
		400,600	0.203	0.203	0.203	0.200	0.200	0.266	0.266	0.266	0.265	0.263
		800,1000	0.011	0.011	0.011	0.011	0.011	0.069	0.069	0.069	0.069	0.069
Type I Error Rate												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Equal variance co-variance matrix	Equal sample	10,10	0.008	0.006	0.008	0.000	0.008	0.051	0.047	0.051	0.006	0.051
		60,60	0.007	0.007	0.007	0.005	0.007	0.064	0.063	0.064	0.056	0.064
		400,400	0.019	0.019	0.019	0.019	0.019	0.088	0.087	0.088	0.087	0.088
		1000,1000	0.055	0.055	0.055	0.053	0.055	0.151	0.151	0.151	0.149	0.151
	Unequal sample	10,30	0.035	0.034	0.035	0.012	0.015	0.120	0.112	0.120	0.076	0.063
		60,90	0.025	0.025	0.025	0.018	0.022	0.069	0.068	0.069	0.063	0.063
		400,600	0.030	0.030	0.030	0.029	0.029	0.098	0.098	0.098	0.096	0.093
		800,1000	0.053	0.053	0.053	0.053	0.052	0.128	0.128	0.128	0.127	0.124

Power												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Unequal variance co-variance matrix	Equal sample	10,10	0.028	0.027	0.028	0.026	0.028	0.093	0.090	0.093	0.089	0.093
		60,60	0.031	0.031	0.031	0.031	0.031	0.061	0.061	0.061	0.061	0.061
		400,400	0.054	0.054	0.054	0.054	0.054	0.244	0.244	0.244	0.243	0.244
		1000,1000	0.033	0.033	0.033	0.033	0.033	0.089	0.089	0.089	0.089	0.089
	Unequal sample	10,30	0.015	0.015	0.015	0.015	0.014	0.068	0.067	0.068	0.067	0.063
		60,90	0.016	0.016	0.016	0.016	0.016	0.150	0.150	0.150	0.149	0.146
		400,600	0.103	0.103	0.103	0.102	0.101	0.394	0.394	0.394	0.391	0.390
		800,1000	0.016	0.016	0.016	0.016	0.015	0.207	0.207	0.207	0.207	0.205
Type I Error Rate												
			α = 0.01					α = 0.05				
			Wilk	lawley	pillai	Roy	hotel	Wilk	lawley	pillai	Roy	hotel
Unequal variance co-variance matrix	Equal sample	10,10	0.015	0.012	0.015	0.000	0.015	0.062	0.052	0.062	0.010	0.062
		60,60	0.013	0.013	0.013	0.010	0.013	0.063	0.061	0.063	0.057	0.063
		400,400	0.023	0.022	0.023	0.021	0.023	0.113	0.113	0.113	0.111	0.113
		1000,1000	0.050	0.050	0.050	0.048	0.050	0.150	0.150	0.150	0.149	0.150
	Unequal sample	10,30	0.073	0.071	0.073	0.033	0.038	0.224	0.218	0.224	0.174	0.138
		60,90	0.029	0.028	0.029	0.022	0.025	0.076	0.075	0.076	0.073	0.073
		400,600	0.036	0.036	0.036	0.035	0.035	0.102	0.101	0.102	0.100	0.099
		800,1000	0.051	0.051	0.051	0.049	0.047	0.129	0.129	0.129	0.128	0.128

When sample size are equal and data are multivariate normal, Roy has Largest power of the test and type I error rate are better than others, but when sample size are not equal Hotelling T^2 is better than others in terms of power of the test and type I error rate when null hypothesis is true. But when data are multivariate normal and and variance co – variance matrix are not equal, Hotelling T^2 , Wilk lambda and Pillia performance are better in term of power of the test and type I error rate are roughly equivalent while Lawley – Hotelling trace is better than the three mentioned but Roy's largestrootis the best. When sample sizes are not equal and very large, Hotelling T^2 is better than all the other test statistic (Table 1).

When data are multivariate Gamma, the performance of the test statistic are in this order Wilks' lambda ≥ Pillai' trace ≥ Hoteling T square ≥ Lawley – Hotelling trace ≥ Roy's largest root, but when the sample size are not equal the order change to Wilks' lambda ≥ Pillai' trace ≥ Lawley – Hotelling trace ≥ Roy's largest root ≥Hoteling T square. Meaning thatHotelling T^2 is better. (Table 2).

When the three assumption hold with small sample size the power and type I error rate are smaller compare to when equality of variance covariance matrix is violated , but when the sample size are very large it control the power and type I error rate. When assumption of normality isviolated, the power is very high and is reducing as sample size is increasing when sample size are not equal. That is to say when sample size are not equal and very small, the power will be very high, but when the sample size are equal it does not have much effect on power and type I error rate. And when the three assumptions failed the type I error rate will be higher than when one of the assumption failed.
From all indication, Roy's largest root is more robust for two random variables with equal sample size, while Wilk' lambda, Hotelling T^2 and pillai's trace performed the same way in terms of power all through. Hoteling T^2 shows its superiority when sample sizes are not equal and exceptionally large.

Reference

Blair, R. C., Higgins, J. J., Karniski, W., & Kromrey, J. D. (1994). A study of multivariate permutation tests which may replace Hotelling'’sT^2 test in prescribed circumstances. *Multivariate Behavioral Research*, 29, 141-163.

Box, G.E.P.(1949). A general distribution theory for a class of likelihood criteria. *Biometrika, 36, 317-346.*

Everitt, B. S. (1979). A Monte Carlo investigation of the robustness of Hotelling's one and two sample T^2 tests. *Journal of the American Statistical Association,* 74, 48-51.

Finch, H. (2005). Comparison of the performance of the nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology*, 1, 27-38.

Hakstian, A. R., Roed, J. C., & Lind, J. C. (1979). Two-sample T^2 procedure and the assumption for homogeneous covariance matrices. *Psychological Bulletin*, 86, 1255-1263.

Hotelling, H. (1931). The generalized t-test and measure of multivariate dispersion. *In J.*

Hopkins, J. W., & Clay, P. P. F. (1963). Some empirical distributions of bivariate T^2 and homoscedasticity criterion M under unequal variance and leptokurtosis. *Journal of the American Statistical Association*, 58, 1048-1053.

Ito, K. (1969). A Comparison of Powers of Two Multivariate Analysis of Variance Tests. *Biometrika.* 444-456.

Ito, K., & Schull, W.J. (1964)."On the Robustness of the T_0^2 Test in Multivariate Analysis of Variance-Covariance Matrices are Not Equal," *Biometrika,* 51, 71- 82.

Johnson, R.A. & Wichern, D.W. (2002). Applied multivariate statistical analysis. *Upper Saddle River*, NJ: Prentice Hall.

Lawley, D.N. (1938). A generalization of fisher's z-test. *Biometrika, 30, 180 – 187.*

Lee, Y.-S. (1971). asymptotic formulae for the distribution of a multivariate test statistic: Power comparisons of certain multivariate tests. Biometrika, 58, 647- 651.

Mardia, K.V,Kent, J.T., and Bibby, J.M.(1979). *Multivariate Analysis*.London Academic Press.

Olson, C.L. (1974). Comparative robustness of six test in multivariate analysis of variance. Journal of the American Statistical Association, 69, 894-908.

Pillai, K. C. S., & Jaya chandran, K. (1967) Power comparisons of tests of two multivariate hypotheses based on four criteria. *Biometrika*, 54, 195-210.

Ramsey, P. H. (1982). Empirical power of procedures for comparing 2 groups on p-variables. *Journal of Educational Statistics*, 7, 139-156.

Roy, S. N. (1945). The individual sampling distribution of the maximum, the minimum, and any intermediate of the p-statistics on the null hypothesis. *Sankhya*, 7, 133-158.

Sheehan-Holt, J. K. (1998). MANOVA simultaneous test procedures: The power and robustness of restricted multivariate contrasts. *Educational and Psychological Measurement*, 58, 861-881

Stevens, J. (2001). Applied Multivariate Statistics for the Behavioural Sciences. Mahwah, NJ: *Lawrence Erlbaum Associates, Publishers*.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika,* 24, 471-494.