GLOBAL JOURNAL OF MATHEMATICAL SCIENNCES VOL. 6, NO. 2, 2007: 71 -73
 ISSN 1596-6208

COMPUTING POWER AND SAMPLE SIZE FOR HOTELLING T^2 TEST

G. M. OYEYEMI

(Received 18 April 2007; Revision Accepted 15 June 2007)

ABSTRACT

Multiple comparisons often arises in statistical analysis and several methods exist for adjusting statistical significance levels taking into consideration the sample size and the power of the test. In this paper, Hotelling's T^2 method for handling multiplicity in the p-variate two sample cases is discussed. A generalized method is derived for computing the exact power for the test. Apart from sample sizes, the magnitude and the direction of correlation between the variables also contributed to the power size of the test.

KEYWORDS: Bivariate Normal, Correlation, Hotelling's T^2, Power,

1 INTRODUCTION

Situation arises when instead of observing a one response random variable, p-vector (p≥2) random variables are observed And suppose that two independent random samples of sizes n_1 and n_2 are observed, then X_p is said to be multivariate normal $X_p \sim N_p(\mu, \Sigma)$. Therefore, given p-variate observations from two multivariate normally distributed populations with variance-covariance matrix Σ, Hotelling's T^2 statistics may be used to test the equality of the vector of means associated with the multivariate normal distribution (Hotelling, 1949, Norman, 1987).

Simultaneous confidence interval is constructed to determine which components of mean vector differ between the two populations if the null hypothesis is rejected at α-level of significance (Timm, 1985). The rejection of null hypothesis depends on many factors among which are; α (level of significance of the test), the sample sizes n_1 and n_2, the vector of mean differences and the variance-covariance matrix, Σ. It follows that the power or rather the probability of rejecting the null hypothesis when it is false, depends on the factors listed above. This paper will examine the effect of variance-covariance matrix and sample sizes required to attain a desired power for a test.

2 METHODOLOGY

Let X_p be a random p-vector (p≥2) of random variables from multivariate normal, that is $X_p \sim N_p(\mu, \Sigma)$. Assuming we have two populations, and in testing

H$_0$: $\mu_1 = \mu_2$ versus H$_1$: $\mu_1 \neq \mu_2$

The test statistic:

$$T^2 = \frac{N_1 N_2}{N_1 + N_2}(x_i - \bar{x})' S^{-1}(x_i - \bar{x}) \sim T_{P, N_1+N_2-2, \alpha}$$

where, $\bar{x}$ and S are the unbiased estimates of vector of means and the variance-covariance matrix Σ based on the random samples of sizes N_1 and N_2.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_{ij}}{n_j} \qquad i = 1, 2, \ldots, n_j \text{ and } j = 1, 2$$

$$S = \frac{1}{N-1}(x_i - \bar{x})(x_i - \bar{x})'$$

For Hotelling's T^2 to have an F-distribution, the expression for T^2 is weighted by the factor $\frac{N_1 + N_2 - P - 1}{P(N_1 + N_2 - 2)}$ so that

$$T^2 = \frac{N1_1 N_2(N_1 + N_2 - P - 1)}{N_1 + N_2(N_1 + N_2 - 2)P} D^2 \sim F_{P, N_1+N_2-P-1, \alpha}$$

where $D^2 = (x_i - \bar{x})' S^{-1}(x_i - \bar{x})$ is known as Mahalanobis Distance

To obtain the power, let d denote the vector of mean differences between the two populations and S the sample variance-covariance matrix based on the random samples of sizes n_1 and n_2 from the two p-variates normally distributed populations A direct extension of the univariate pooled t-test to multivariate space (Kelsy, et al, 1996, Kramer, 1972) gives the exact power for Hotelling's T^2 test as;

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

where F_{p, n_1+n_2-p-1} is the 100(1-α) percentile of the F-distribution with p and ($n_1 + n_2 - p - 1$) degrees of freedom (Johnson and Wichern 1982). When p = 1, we see that the equation gives the exact power for the univariate pooled t-test

3 EXAMPLE

Since both sample sizes and variance-covariance matrix will to certain affect the power of the test, different values of variance-covariance matrices with varying magnitude of correlation coefficients as well as directions of relationship for bivariate normally distributed random variables for two populations are used. The levels of correlation and directions are given below;

(a) ρ = ± 0.10
(b) ρ = ± 0.20
(c) ρ = ± 0.40
(d) ρ = ± 0.60
(e) ρ = ± 0.80

G. M. Oyeyemi, Department of Statistics, University of Ilorin, Ilorin, Nigeria

Assuming we wish to compare the two populations with respect to their vector of means, the vector of mean difference d, and the estimate of variance-covariance are given as;

$$d = \left(\bar{x}_{11} - \bar{x}_{12}, \bar{x}_{21} - \bar{x}_{22}\right)'$$

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

The SAS Code that was used in this paper is given as follows;

```
Proc iml worksize = 500;
        Start main;
                Alpha=0.05;
                P=2;
                d={ x̄11 - x̄12 , x̄21 - x̄22 };
                s={ s11 s12 , s21 s22 };
                do n1=5 to 30 by 5;
                        do n2=5 to 30 by 5;
                                power=probt(sqrt((n1*n2/(n1+n2))*d*inv(s)*d')
                                        -sqrt(((p*(n1+n2-2))/(n1+n2-1-p))*
                                        finv(1-alpha,p,n1+n2-1-p)),n1+n2-2);
                                print n1 n2 power;
                        end;
                end;
        run;
```

The sample correlation coefficient between the two variables, $r = \frac{s_{12}}{s_1 s_2}$, such that $s_{12} = r s_1 s_2$. Therefore, s_{12} can be determined for various values of r. A SAS Code or S-Plus programme can be used to obtain the power for various values of $n_1 = n_2 = n$ and r.

4. RESULTS AND CONLUSION

It can be seen from the results shown in Table1 that as the sample size increases, the power of the test also increases. For a positive correlation coefficient, the power increases with increase in values of the correlation coefficients between the variables as the sample size increases (Figure 1), but interestingly, the reverse is the case for a negative values of correlation coefficients, with small values of negative correlation coefficients, the test attains higher power than large values of negative correlation coefficients as sample size increases.

Therefore, by taking advantage of the variance-covariance structure among variables, Hotelling's T^2 test presents a simple and easy way to control for multiplicity when comparing the equality of multiple means in two samples.

Table1. Exact Power values for varying levels of correlation coefficients and sample sizes

N1	N2	0.10	-0.10	0.20	-0.20	0.40	-0.40	0.60	-0.60	0.80	-0.80
5	5	0.1365	0.1029	0.1621	0.0917	0.2502	0.0755	0.4511	0.0649	0.8644	0.0586
10	10	0.5900	0.4799	0.6569	0.4359	0.8094	0.3654	0.9484	0.3142	0.9989	0.2822
15	15	0.8425	0.7483	0.8879	0.7032	0.9612	0.6214	0.9997	0.5540	0.9998	0.5083
20	20	0.9456	0.8894	0.9672	0.8575	0.9932	0.7918	0.9998	0.7303	0.9999	0.6848
25	25	0.9825	0.9546	0.9911	0.9358	0.9989	0.8919	0.9999	0.8453	1.0000	0.8077
30	30	0.9947	0.9823	0.9977	0.9725	0.9998	0.9464	1.0000	0.9148	1.0000	0.8871

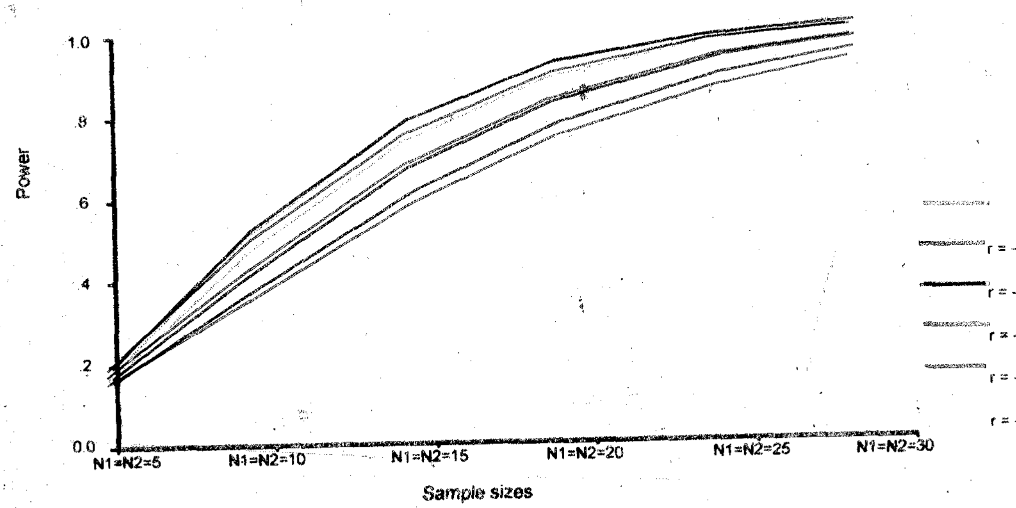

Figure1. Graph of exact power for varying negative correlation coefficients and sample sizes

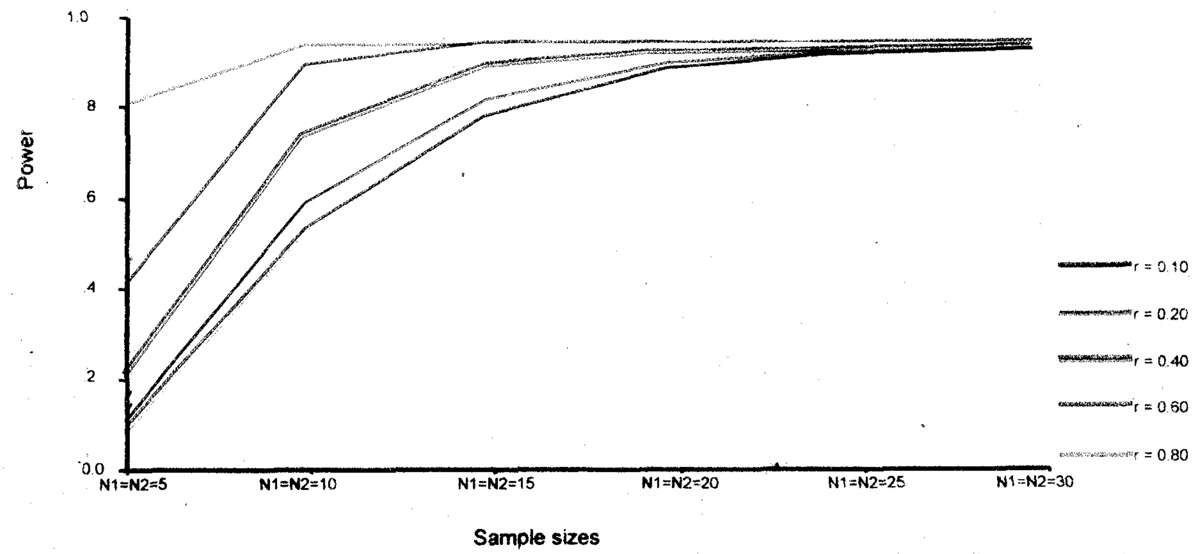

Figure2. Graph of exact power for varying positive correlation coefficients and sample sizes

REFERENCE

Hotelling, H., 1947. Multivariate Quality Control. Techniques of Statistical Analysis. McGraw-Hill.

Johnson, R and Wichern, D., 1982. Applied Multivariate Statistical Analysis. Prentice-Hill, Inc.: Englewood Cliffs, NJ.

Kelsey, J., Whieeemore, A., Evans, A. and Thompson, W., 1996. Methods in Observational Epidemiology. Oxford: New York.

Kramer, C., 1972. A First Course in Methods of Multivariate Analysis. Virginia Polytechnic Institute and State University: Blacksburg, VI.

Norman, C., 1987. Analyzing Multivariate Data. San Diego, CA: Harcourt Brace Jovanovich.

Timm, N. H., 1985. Applied Multivariate Analysis. Prentice-Hill, Inc.: Englewood Cliffs, NJ.
SAS Institute Inc. U.S.A.