CHI-SQUARE AND F APPROXIMATIONS

OF HOTELLING'S GENERALIZED T_0^2

CHI-SQUARE AND F APPROXIMATIONS

OF HOTELLING'S GENERALIZED T_0^2

By

Davin Y.N. Yip, B.Sc.

A Project

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the degree

Master of Science

McMaster University

MASTER OF SCIENCE (Statistics)

McMASTER UNIVERSITY Hamilton, Ontario

TITLE: Chi-Square and F Approximations of Generalized T_0^2

AUTHOR: Davin Y.N. Yip, B.Sc. (McMaster University)

SUPERVISOR: Professor M.L. Tiku

ABSTRACT

Chi-square approximation of Hotelling's generalized T_0^2 have been investigated by Professor Tiku. In this project, further results on the approximation are presented and examined. Also, an F approximation for the distribution of Hotelling's generalized T_0^2 is proposed. The results are compared with other approximations and discussed.

PREFACE

In multivariate analysis problems one often confronts the testing of different kinds of linear hypotheses. For example, the testing of the equality of p-dimensional mean vectors of p-variate normal population is a familiar one.

There are a number of test statistics suitable for this purpose. In general, test statistics for the multivariate linear hypothesis based on multinormal models can be expressed as functions of non-zero roots of determinantal equations of the form (see Johnson and Kotz, 1972)

$$\det(S_1 - \lambda\, S_2) = 0,$$

where S_1 and S_2 are independent $p \times p$ Wishart matrices having n_1 and n_2 degrees of freedom, respectively. These characteristic roots are the invariants of S_1 and S_2 under certain transformations on the observations and design matrix (see Schatzoff, 1966). The λ's are the roots of $\det(S_1 S_2^{-1} - \lambda I) = 0$. Some invariant test statistics can be represented as

(a) $\Lambda = \dfrac{\det(S_2)}{\det(S_1+S_2)} = \prod_{i=1}^{p} \left(\dfrac{1}{1+\lambda_i}\right)$ (Wilks)

(b) $T_0^2/n_2 = \mathrm{tr}(S_1 S_2^{-1}) = \sum_{i=1}^{p} \lambda_i$ (Hotelling)

(c) $\max(\lambda_i)$ (Roy)

(d) $\min(\lambda_i)$ (Roy)

(e) $T = \left[\sum_{i=1}^{p} \left(\frac{1+\lambda_i}{\lambda_i}\right)\right]^{-1}$ (Pillai)

(f) $V = \mathrm{tr}[S_1(S_1+S_2)^{-1}] = \sum_{i=1}^{p} \left(\frac{\lambda_i}{1+\lambda_i}\right)$ (Pillai)

The test statistic (a) is Wilks' (1932) Λ criterion. The test (b) was studied by Hotelling (1951) whereas (c) and (d) were discussed by Roy (1953). The tests (e) and (f) were investigated by Pillai (1955). Because of mathematical complexities little is known about the exact distributions of the statistics defined above except for special cases. Among these statistics, Hotelling's generalized T_0^2 and Wilk's Λ criterion have been most extensively studied and used. However, very little is known about the relative power of these test statistics. Schatzoff (1966), however, showed that Hotelling's generalized T_0^2 and Wilk's Λ criterion are relatively more sensitive to a large number of alternative hypotheses and have more or less the same power; see also Lee (1971).

The exact distribution of Hotelling's generalized T_0^2 is not known except for s=2. For other values of s different authors have suggested different methods of approximation. Davis (1968) showed that the distribution of T_0^2 satisfies an ordinary linear differential equation of degree p(number of characteristic roots). Davis (1970) tabulated the exact percentage points of T_0^2 by carrying out an analytic continuation of Constantine's series using the Newton-Raphson method. The task involved is formidable. Other approximations of T_0^2 by Ito (1960), Pillai and Samson (1959) also involve considerable computations. On the other hand, Tiku's (1971) chi-square approximations exhibit two main advantages.

(1) The chi-square approximations are easier to compute than the approximations discussed above.

(2) Tables of both the percentage points and of the probability integrals of chi-square are readily available.

In this project, Tiku's chi-square approximations are discussed in detail. A new F approximation is introduced. The F approximation of the 5% and 1% points of T_0^2 are calculated and compared with Davis' (1970) exact values and values based on Tiku's (1971) chi-square methods. It is found that for approximating the 1% points of T_0^2, Tiku's inverse chi-square approximation is slightly more accurate than the F approximation when n is small, but for large n, the F approximation is generally more accurate than Tiku's chi-square approximation. For approximating the 5% points of T_0^2, the F approximation almost always gives more accurate results than Tiku's chi-square approximations for all s and n.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor M.L. Tiku, my advisor, for his valuable guidance, suggestions, helpful criticisms and discussions throughout the project.

Appreciated financial support was provided by the Department of Applied Mathematics, McMaster University.

Finally, thanks are due to Sally Coutts for her excellent typing.

1. Summary

Chi-square approximations for the distribution of Hotelling's generalized T_0^2 statistic have been obtained by Tiku (1971). The present project extends Tiku's tabulated values and further examines his approximations. An F approximation is also introduced. The accuracy of its 5% and 1% points approximations of T_0^2 are discussed.

2. Introduction

Hotelling's generalized T_0^2 statistic is defined as

$$T_0^2 = n_2 \; tr(S_1 S_2^{-1}), \tag{1}$$

where S_1 and S_2 are two independent $p \times p$ Wishart matrices on n_1 and n_2 degrees of freedom respectively, estimating the same dispersion matrix. Following Pillai and Samson (1959), we denote T_0^2/n_2 by $U^{(s)}$, then

$$U^{(s)} = tr(S_1 S_2^{-1}), \tag{2}$$

which equals the sum of the characteristic roots of $S_1 S_2^{-1}$. The superscript s denotes the number of non-zero roots of $S_1 S_2^{-1}$.

The exact distribution of T_0^2 is not known except for s=2. For other values of s different methods of approximation have been suggested. Direct formulae for approximating $U^{(s)}$ are given by Pillai (1954, 1956) and Ito (1956). Although the first three moments of $U^{(s)}$ are known for any s, the exact fourth moment for general s has not yet been obtained because of heavy algebraic manipulations. Pillai and Samson (1959) worked out the exact fourth moment in explicit forms for s=2, 3 and 4 and approximated the upper 5% and 1% points of $U^{(s)}$ by fitting 4-moment curves of the Pearson system from tables given by Johnson, Nixon, Amos and Pearson (1963). For higher values of s, the fourth moment can be calculated from Pillai and Mijares (1959) but the task is formidable. Davis (1970) gave a large sample approximation for the fourth moment and tabulated the exact percentage points of $U^{(s)}$ by using an equivalent system of first order linear homogeneous differential equations to carry out an analytic continuation of Constantine's series. Pearson 4-moment curves can be fitted for any s using the fourth moment by Davis (1970). However, the Johnson et al tables (1963) do not provide the values when the probability integrals and not the percentage points are required.

Tiku (1971) used chi-square and inverse chi-square distributions evaluated at the exact first three moments to approximate the percentage points of $U^{(s)}$. This method provides the values of the probability integrals and the values of the percentage points of $U^{(s)}$ since the probability integrals and percentage points of chi-square distribution are readily available. In this project, we examine Tiku's chi-square approximations for larger values of s. Approximations based on F distribution having the correct first four moments are obtained. Approximations for the 5% and 1% points of $U^{(s)}$ by an F distribution are compared with the values based on chi-square and inverse chi-square distributions to the exact values for s=2 and to Davis' (1970) values for s=3, 4 and 5. The accuracy of these approximations is discussed.

3. Chi-square approximation for $U^{(s)}$

$$\text{Let } \chi^2 \simeq \{U^{(s)} + a\}/h, \tag{3}$$

where χ^2 is a chi-square variate having v degrees of freedom. Equate both sides of (3) at the first three moments, we have

$$\mu_1'(\chi^2) = \{\mu_1'(U) + a\}/h$$

$$\mu_2(\chi^2) = \mu_2(U)/h^2$$

$$\mu_3(\chi^2) = \mu_3(U)/h^3$$

Since $\beta_1 = \mu_3^2/\mu_2^3$, (4) gives

$$\beta_1(\chi^2) = \beta_1(U) \tag{4}$$

By defining two parameters

$$m = \frac{1}{2}(n_1-s-1)$$

$$n = \frac{1}{2}(n_2-s-1)$$

and noting that

$$\beta_1(\chi^2) = 8/v$$

$$\beta_1(U) = \frac{16(n+2m+s+1)^2(n+s)^2(n-1)(2n+1)}{s(2m+s+1)(2n+2m+s+1)(2n+s)(n-2)^2(n+1)^2} \tag{5}$$

we obtain

$$v = \frac{8}{\beta_1} = \frac{s(2m+s+1)(2n+2m+s+1)(2n+s)(n-2)^2(n+1)^2}{2(n+2m+s+1)^2(n+s)^2(n-1)\ (2n+1)} \tag{6}$$

$$h = \frac{1}{4}\ \frac{\mu_3}{\mu_2} = \frac{(n+2m+s+1)(n+s)}{2n(n-2)(n+1)}$$

$$a = vh-\mu_1'$$

where $\mu_1' = \dfrac{s(2m+s+1)}{2n}$.

To justify that chi-square approximation produces percentage points of $U^{(s)}$ close to the values based on Pearson curves, Tiku (1971) compared β_2 values of the chi-square line $\beta_2 = 3+1.5\beta_1$ with the true β_2 values of

$U^{(s)}$ for s = 2, 3 and 4. In this project we extend this comparison for s = 5, 6 and 7 using large sample approximation for β_2 from Davis (1970).

$$\beta_2(U) = \frac{3(c-2)(c+1)A}{s(2m+s+1)(2n+s)(c-6)(c-4)(c-1)(c+2)(c+3)(c+2m+s+1)}, \quad (7)$$

where

$$c = 2n$$

$$A = (2m+s+1)(c+2m+s+1)\{s(c+s)\ (c^3-5c^2+78c-72) + 4c^2(5c-6)\}$$

$$+ 4c^2\{s(c+s)\ (5c-6) + c(c^2-c+2)\}\ .$$

The true β_1 and β_2 values together with the difference of the true β_2 values and the β_2 values of the chi-square line are given in table 1. The results show that the differences between β_2 values of $U^{(s)}$ and β_2 values of the chi-square line do not increase as s increases (see also Tiku (1971) table 1, which is reproduced in table 3) and in general the differences decrese for larger s. For n > 15 and for any s, the differences are small and thus chi-square approximation should give percentage points of $U^{(s)}$ close to the values based on curves of the Pearson system. Note that in general Pearson curves give percentage points which differ from the true values by not more than 2 units in the third significant digit. A comparison of the chi-square approximation with Pillai's (1954) approximation to that based on Pearson curves is shown in table 6 (values for n=15, 20 and 60 are reproduced from Tiku's (1971) tabulated values for the sake of comparison); see also Pillai and Samson (1959). It is clear that the chi-square approximation gives percentage points of $U^{(s)}$ close to the values based on Pearson curves.

Table 1. β_1 and β_2 of $U^{(s)}$ and the difference $\beta_2-(3+1.5\beta_1)$.

		$s=5$			$s = 6$			$s = 7$		
m	n	β_1	β_2	Diff.	β_1	β_2	Diff.	β_1	β_2	Diff.
0	5	3.65	13.14	4.66	3.31	12.31	4.34	3.09	11.75	4.12
	10	1.15	5.41	0.68	0.98	5.09	0.62	0.86	4.87	0.57
	15	0.75	4.45	0.32	0.61	4.20	0.29	0.52	4.04	0.26
	30	0.46	3.81	0.11	0.36	3.64	0.10	0.29	3.52	0.09
	60	0.36	3.58	0.05	0.27	3.44	0.04	0.21	3.35	0.04
	100	0.32	3.50	0.03	0.23	3.37	0.02	0.18	3.29	0.02
10	5	2.97	11.49	4.03	2.81	11.07	3.86	2.69	10.78	3.74
	10	0.76	4.69	0.54	0.69	4.55	0.51	0.64	4.45	0.48
	15	0.42	3.88	0.24	0.38	3.79	0.22	0.34	3.72	0.21
	30	0.20	3.38	0.08	0.17	3.33	0.07	0.15	3.29	0.07
	60	0.12	3.21	0.03	0.10	3.18	0.03	0.09	3.15	0.02
	100	0.09	3.16	0.02	0.08	3.13	0.02	0.07	3.11	0.01
50	5	2.90	11.32	3.96	2.74	10.91	3.80	2.63	10.64	3.68
	10	0.71	4.59	0.52	0.65	4.46	0.49	0.60	4.37	0.47
	15	0.37	3.79	0.23	0.33	3.71	0.21	0.30	3.66	0.20
	30	0.15	3.30	0.07	0.13	3.26	0.07	0.12	3.23	0.06
	60	0.07	3.13	0.03	0.06	3.12	0.02	0.05	3.10	0.02
	100	0.05	3.08	0.02	0.04	3.07	0.01	0.03	3.06	0.01

Table 2. Difference of true β_2 values and the β_2 values of inverse chi-square distribution.

		$s = 5$	$s = 6$	$s = 7$
m	n	Diff.	Diff.	Diff.
0	5	1.94	2.00	2.01
	10	0.12	0.16	0.18
	15	-0.01	0.02	0.04
	30	-0.08	-0.05	-0.03
	60	-0.10	-0.07	-0.05
	100	-0.10	-0.07	-0.05
10	5	2.03	2.02	2.01
	10	0.20	0.20	0.20
	15	0.07	0.07	0.07
	30	0.00	0.00	0.01
	60	-0.01	-0.01	-0.01
	100	-0.02	-0.01	-0.01
50	5	2.04	2.02	2.01
	10	0.21	0.21	0.21
	15	0.08	0.08	0.08
	30	0.01	0.02	0.02
	60	0.00	0.00	0.00
	100	-0.00	-0.00	-0.00

4. Inverse chi-square approximation for $U^{(s)}$

It is clear from tables 1 and 3 that for small values of n the β_2 values of $U^{(s)}$ lie away from the chi-square line and are close to the Pearson type V line if β_1 and β_2 values are plotted. Thus we expect that for small n inverse chi-square approximation for the percentage points of $U^{(s)}$ is more appropriate.

Let $\frac{1}{x}$ be a chi-square variate having f degrees of freedom, then x is an inverse chi-square variate with

$$\text{mean} = 1/(f-2) \tag{8}$$

$$\text{variance} = 2/\{(f-2)^2(f-4)\}$$

$$\beta_1(x) = \frac{32(f-4)}{(f-6)^2}$$

$$\beta_2(x) = \frac{3(f-4)(f+10)}{(f-6)\ (f-8)} .$$

Now consider

$$x \simeq \{U^{(s)} + c\}/k . \tag{9}$$

Equate both sides at the first three moments and, using (6) and (8), we obtain

$$f = 6 + 2v\{1 + \sqrt{1+2/v}\}$$

$$k = \tfrac{1}{2}h(f-2)(f-6)$$

$$c = \tfrac{1}{2}h(f-6) - \mu_1'$$

where v, h and μ_1' are given in (6). To see how close the β_2 values based on inverse chi-square distribution are to the true β_2 values of $U^{(s)}$, their differences are calculated for s = 5, 6 and 7 for several values of m and n. The results are tabulated in table 2. Again the differences of true β_2 values and the β_2 values of inverse chi-square approximation are small for small n. A comparison of tables 1 and 2 indicates that for s = 5, 6 and 7 inverse chi-square approximation of $U^{(s)}$ should give closer values to the values given by Pearson curves than chi-square approximation when $n \lesssim 30$. Tables from Tiku's paper (1970) are included in tables 3 and 4 for similar comparison. The two comparisons (table 1 with table 2 and table 3 with table 4) suggest the appropriate approximations in the following table.

Approximations of % points of $U^{(s)}$	Inverse Chi-square Approximation	Chi-square Approximation
s = 2, 3, 4	$n \lesssim 15$	$n > 15$
s = 5, 6, 7	$n \lesssim 30$	$n > 30$

Table 3

Difference of the true β_2 values and the β_2 values of the chi-square approximation.

m	n	$s = 2$ Diff.	$s = 3$ Diff.	$s = 4$ Diff.
0	5	7.91	6.05	5.17
	10	1.32	0.96	0.78
	15	0.66	0.47	0.38
	30	0.25	0.18	0.14
	60	0.11	0.08	0.06
	100	0.06	0.04	0.03
10	5	5.68	4.75	4.29
	10	0.86	0.68	0.60
	15	0.41	0.32	0.27
	30	0.15	0.11	0.09
	60	0.07	0.05	0.04
	100	0.04	0.03	0.02
50	5	5.56	4.66	4.22
	10	0.83	0.66	0.57
	15	0.39	0.30	0.26
	30	0.14	0.10	0.08
	60	0.06	0.04	0.03
	100	0.03	0.02	0.02

Table 4

Difference of the true β_2 value and the β_2 values of the inverse chi-square approximation.

m	n	$s = 2$ Diff.	$s = 3$ Diff.	$s = 4$ Diff.
0	5	-0.79	1.30	1.79
	10	-0.84	-0.14	0.05
	15	-0.77	-0.24	-0.08
	30	-0.72	-0.28	-0.14
	60	-0.70	-0.29	-0.16
	100	-0.69	-0.29	-0.16
10	5	1.78	2.00	2.03
	10	0.14	0.18	0.20
	15	0.07	0.05	0.06
	30	-0.04	-0.02	-0.01
	60	-0.05	0.10	-0.02
	100	-0.20	0.09	-0.10
50	5	1.85	2.03	2.04
	10	0.19	0.20	0.21
	15	0.07	0.07	0.08
	30	0.01	0.02	0.01
	60	0.01	0.00	0.00
	100	-0.01	0.01	-0.01

The upper 5% and 1% points of $U^{(s)}$ using chi-square and inverse chi-square approximations are calculated for $s=2$. These values are compared with the exact values in table 5. Pillai's (1959) values of approximation are also included. The results suggest that Tiku's approximations of the percentage points of $U^{(s)}$ are closer to the exact values than Pillai's.

For $s=3$ and 4, chi-square and inverse chi-square approximations of the percentage points of $U^{(s)}$ are compared with those based on the Pearson curves in table 6. It is clear that chi-square approximation gives percentage points of $U^{(s)}$ close to the values based on Pearson curves. For large n, the chi-square approximation generally gives percentage points of $U^{(s)}$ which differ from the values based on Pearson curves only slightly in the fourth significant digit.

For $s=5$ and 6 approximations of percentage points of $U^{(s)}$ based on chi-square and inverse chi-square distributions are compared with the exact values by Davis (1970) in tables 7 and 8. It is clear that for small n, approximations based on chi-square and inverse chi-square distributions give percentage points of $U^{(s)}$ which generally differ from Davis'(1970) values only slightly in the third significant digit. Agreement is in general good.

Table 5. Exact and approximate percentage points of $U^{(2)}$

5%	$m = 0$			$m = 3$			$m = 4$			$m = 5$		
n	Exact	Tiku	Pillai	Exact	Tiku	Pillai	Exact	Tiku	Pillai	Exact	Tiku	Pillai
5	1.43	1.44*	1.39	—	3.5883*	—	—	4.275*	—	—	4.9611*	—
10	0.671	0.673*	0.664	—	1.6246*	—	1.927	1.928*	1.886	—	2.2275*	—
15	0.443	0.437*	0.435	—	1.0406*	—	1.236	1.233*	1.219	—	1.4229*	—
20	0.328	0.331	0.323	0.7690	0.7728	0.7600	0.908	0.912	0.898	1.046	1.050	1.034
30	0.2155	0.2169	0.2124	0.5020	0.5037	0.4983	0.592	0.5939	—	0.681	0.683	—

1%	$m = 0$			$m = 3$			$m = 4$			$m = 5$		
n	Exact	Tiku	Pillai	Exact	Tiku	Pillai	Exact	Tiku	Pillai	Exact	Tiku	Pillai
5	2.17	2.26*	2.05	—	5.2041*	—	—	6.1471*	—	—	7.0822*	—
10	0.979	0.979*	0.931	—	2.1512*	—	2.516	2.520*	2.415	—	2.8856*	—
15	0.624	0.621*	0.602	—	1.3352*	—	1.563	1.561*	1.526	—	1.7847*	—
20	0.455	0.457	0.443	0.9698	0.9686	0.9499	1.130	1.129	1.108	1.288	1.286	1.263
30	0.2960	0.2959	0.2904	0.6235	0.6231	0.6165	0.725	0.725	—	0.825	0.824	—

* values based on inverse Chi-square distribution.

- values not given by the authors.

Table 6 (a). Approximate values of the percentage points of $U^{(3)}$

$s=3$	5%						1%					
	$m=0$			$m=5$			$m=0$			$m=5$		
n	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
15	0.747	-0.006	—	2.120	0.010	0.044	0.9827	0.0003	—	2.560	0.005	0.098
20	0.5517	-0.003	—	1.857	-0.004	0.024	0.7155	0.0008	—	1.852	0.004	0.054
35	0.3097	-0.0001	0.0042	0.8638	-0.0016	0.0067	0.3939	0.0006	0.1072	1.009	0.002	0.0147
40	0.2692	-0.0009	0.0021	0.7520	-0.0014	0.0037	0.3412	-0.0007	0.0036	0.8759	-0.014	0.0087
50	0.2144	-0.0005	0.0011	0.5977	-0.0007	0.0020	0.2714	-0.0001	0.0028	0.6927	0.0008	0.0041
60	0.1787	-0.000	0.001	0.4959	-0.000	0.002	0.4635	-0.0007	0.0111	0.5728	0.000	0.004
80	0.1331	-0.0001	0.0004	0.3688	-0.0001	0.0001	0.1674	0.0000	0.0012	0.4238	-0.0015	-0.0002
100	0.1062	-0.0001	0.0002	0.2949	0.0002	0.0002	0.1333	0.0000	0.0005	0.3384	0.0002	0.0002

(1) Values based on Pearson curves.

(2) Difference: (1)-values based on the chi-square approximation.

(3) Difference: (1)-values based on Pillai's approximation.

\- Pillai did not work out his approximation

Table 6(b). Approximate values of the percentage points of $U^{(4)}$

s=4	5%						1%					
	m=0			m=5			m=0			m=5		
n	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
15	1.122	-0.007	0.038	2.877	-0.000	0.065	1.413	0.003	0.079	3.395	0.009	0.149
20	0.8272	-0.004	0.019	2.113	-0.004	0.028	1.028	-0.002	0.044	2.458	0.006	0.071
35	0.4629	-0.005	0.0057	1.174	-0.001	0.002	0.5646	0.0005	0.0116	1.341	0.001	0.012
40	0.4033	-0.0005	0.0039	1.023	0.000	0.009	0.4911	0.0006	0.0092	1.165	0.002	0.009
50	0.3209	-0.0003	0.0025	0.8130	0.0005	0.0035	0.3893	0.0004	0.0072	0.9218	0.0011	0.0084
60	0.2666	-0.000	0.002	0.6745	0.0000	0.001	0.3224	0.001	0.004	0.7627	0.000	0.003
80	0.1990	-0.0001	0.0006	0.5030	0.0002	-0.0009	0.2399	0.0000	0.0024	0.5670	0.0003	-0.0004
100	0.1588	0.0000	0.0003	0.4010	0.0002	-0.0008	0.1909	-0.0002	0.0014	0.4512	0.0002	-0.0006

(1) Values based on Pearson curves.

(2) Difference: (1) - values based on the chi-square approximation.

(3) Difference: (1) - values based on Pillai's approximation.

Table 7. Exact and approximate percentage points of $U^{(s)}$

$s=5$

	5%					1%			
	$m=0$		$m=4.5$			$m=0$		$m=4.5$	
n	Davis	Tiku	Davis	Tiku	n	Davis	Tiku	Davis	Tiku
5	5.3396	5.3775*	12.1069	12.2081*	5	7.302	7.344*	15.8559	16.0619*
7	3.6234	3.6387*	8.1555	8.1967*	7	4.7181	4.7447*	10.1438	10.1960*
7.5	2.5747	2.5748*	5.7637	5.7746*	7.5	3.2470	3.2452*	6.933	6.937*
12	1.9938	1.9938*	4.4482	4.4534*	12	2.469	2.448*	5.25	5.25*
14.5	1.6258	1.6255*	3.6187	3.6217*	14.5	1.9898	1.9913*	4.2194	4.2207*
17	1.3720	1.3707*	3.0489	3.0450*	17	1.6658	1.6658	3.5252	3.5254*
22	1.0453	1.0441*	2.3174	2.3211*	22	1.2058	1.2563*	2.6508	2.6467*
27	0.8441	0.8430*	1.8685	1.8706*	27	1.0076	1.0083*	2.1233	2.1205*
32	0.7077	0.7089	1.5651	1.5656	32	0.0412	0.0401	1.7706	1.7674
37	0.6093	0.6101	1.3464	1.3465	37	0.72196	0.72115	1.5182	1.5157
47	0.4767	0.4771	1.0522	1.0524	47	0.5625	0.5619	1.1814	1.1802
97	0.2282	0.2283	0.5027	0.5026	97	0.2672	0.2671	0.5599	0.5595

* Inverse chi-square approximation

Table 8. Exact and approximate percentage points of $U^{(s)}$

$s=6$

	5%					1%			
	$m=0.5$		$m=4$			$m=0.5$		$m=4$	
n	Davis	Tiku	Davis	Tiku	n	Davis	Tiku	Davis	Tiku
6.5	5.91524	5.9491*	10.4242	10.4854*	6.5	7.5543	7.6056*	12.9505	13.0309*
9	4.0915	4.0988*	7.1848	7.2011*	9	5.0358	5.0393*	8.5941	8.6003*
11.5	3.12141	3.1257*	5.4694	5.4784*	11.5	3.7648	3.7684*	6.4082	6.4122*
14	2.5213	2.5233*	4.4115	4.4150*	16.5	2.4963	2.4977*	4.2367	4.2374*
16.5	2.1141	2.1148*	3.6951	3.6974*	16.5	2.4963	2.4977*	4.2367	4.2314*
26.5	1.2833	1.2851	2.2377	2.2387	26.5	1.4891	1.4871	2.5206	2.5160
31.5	1.0723	1.0735	1.8687	1.8692	31.5	1.2389	1.2374	2.0956	2.0924
36.5	0.9209	0.9217	1.6040	1.6046	36.5	1.0607	1.0596	1.7931	1.7911
46.5	0.7181	0.7184	1.2499	1.2501	46.5	0.8236	0.8228	1.3914	1.3901
96.5	0.3417	0.3416	0.5739	0.5939	96.5	0.3889	0.3886	0.6561	0.6557
248.5	0.1328	0.1328	0.2307	0.2307	248.5	0.1505	0.1504	0.2537	0.2537

* Inverse chi-square approximation

Although chi-square approximations of the percentage points of $U^{(s)}$ are compared with the exact values for s=2, with the values based on Pearson curves for s=3 and 4 and with Davis' values for s=5. Remarkably enough, the accuracy of the approximations improves with increasing s. Thus we expect that chi-square and inverse chi-square approximations for the percentage points of $U^{(s)}$ are generally useful for all s and also the approximations are easy to compute. It may be noted that it is a formidable task to work out the exact distributions of T_0^2 for $s > 2$.

5. Central F approximation for $U^{(s)}$

Chi-square approximations involve equating the first three moments on both sides of (3). The closeness of the approximations to the values from Pearson curves depends on the closeness of the β_2 values of $U^{(s)}$ to the β_2 values of chi-square and inverse chi-square distributions. The differences of these beta values are checked in tables 1, 2, 3 and 4 to justify the use of these approximations. The comparing of β_2 values involves comparing the fourth moments, since $\beta_2 = \mu_4/\mu_2^2$. This suggests the use of an F approximation evaluated at the first four moments.

$$\text{Let } F \simeq \frac{U^{(s)} + g}{h} \text{ having } v_1 \text{ and } v_2 \text{ degrees of freedom.} \qquad (10)$$

Equate both sides of (10) at the first four moments, we have

$$\mu_1'(F) = \{\mu_1'(U) + g\}/h$$

$$\beta_1(F) = \beta_1(U)$$

$$\beta_2(F) = \beta_2(U) .$$

$\beta_1(U)$ and $\beta_2(U)$ are given in (5) and (7).

$$\beta_1(F) = \frac{8(v_2-4)(2v_1+v_2-2)^2}{v_1(v_2-6)^2(v_1+v_2-2)}$$

A simple linear relation between $\beta_1(F)$ and $\beta_2(F)$ from Pearson and Tiku (1970) is used, which greatly simplifies algebraic manipulations,

$$\beta_2(F) = \frac{3}{v_2-8}\{v_2-4+\tfrac{1}{2}(v_2-6)\ \beta_1(F)\}.$$

After some algebra, we obtain

$$v_2 = 2\left[3+\frac{\beta_2+3}{\beta_2-(3+1.5\beta_1)}\right]$$

$$v_1 = \frac{(v_2-2)}{2}\left\{-1+\sqrt{1-\frac{1}{1-\frac{1}{32}\frac{(v_2-6)^2}{(v_2-4)}\beta_1}}\right\}$$

$$h = \sqrt{\frac{v_1(v_2-2)^2(v_2-4)}{2v_2^2\,(v_1+v_2-2)}\mu_2}$$

$$g = \frac{v_2}{v_2-2}h - \mu_1',$$

where β_1, β_2, μ_1' and μ_2 are obtained previously.

Limitations to the use of the F approximation are that v_1 and h are not defined if the terms inside the square roots are negative. For large n, the values of v_1 and h are generally well defined. The degrees of freedom v_1 and v_2 are generally large for large n and the approximate percentage points of $U^{(s)}$ are thus readily obtainable from F tables.

Let $z = 1/\{1 + v_1 F/v_2\}$,

then z is a beta variate with parameters $a = \frac{v_2}{2}$, $b = \frac{v_1}{2}$. The integral values corresponding to specific values of the statistics are readily obtainable from tables of incomplete beta function by Pearson (1948). However the lower percentage points of z are not available to more than 2 significant figures. These values are used as initial values and an iterative method is used until the values z_α make the significant level of α to 6 significant figures of accuracy.

Now

$$\text{Prob } (z \leq z_\alpha) = \alpha$$

$$\text{i.e. Prob}\left[\frac{1}{1 + \frac{v_1 F}{v_2}} \leq z_\alpha\right] = \alpha$$

$$\text{i.e. Prob}\left[\frac{1}{1 + \frac{v_1}{v_2}\left(\frac{U^{(s)} + g}{h}\right)} \leq z_\alpha\right] = \alpha$$

$$\text{i.e. Prob}\left[U^{(s)} \geq \left(\frac{1}{z_\alpha} - 1\right) \frac{v_2}{v_1} h - g\right] = \alpha$$

Thus $U_\alpha = (\frac{1}{z_\alpha} - 1) \frac{v_2}{v_1} h - g$ correspond to the upper percentage points of $U^{(s)}$.

Since the F approximation is obtained by equating (10) at the first four moments while chi-square approximations are obtained by equating (3) or (9) at the first three moments, we expect the F approximation to be generally more accurate than the chi-square and inverse chi-square approximations.

Table 9 compares the F approximation of the upper 5% and 1% points of $U^{(s)}$ with the chi-square approximations to the exact values for s=2. The results show that for $n < 15$ inverse chi-square distribution gives slightly more accurate percentage points of $U^{(s)}$ than the F distribution. But the F approximation is more accurate for $n \geq 15$ than both chi-square and inverse chi-square approximations. Even for small n, agreement between the F approximation and the exact values is good.

The F approximation of percentage points of $U^{(s)}$ are also compared with chi-square and inverse chi-square approximations to Davis' values in tables 10, 11 and 12 for s=3, 4 and 5 respectively. These comparisons seem to show that the F distribution approximating the upper 5% points of $U^{(s)}$ almost always gives more accurate results than both chi-square and inverse chi-square distributions for all n and s. For approximating the 1% points of $U^{(s)}$, the F approximation seems in general slightly less accurate than inverse chi-square approximation when n is small but generally more accurate than chi-square approximation when n is large.

The overall agreement of the 5% and 1% points approximation by F with the exact percentage points of $U^{(s)}$ is good. Thus it is expected that for the upper 5%, 1% points of $U^{(s)}$, the F approximation is useful for practically all n and all s. Also tables 10, 11 and 12 indicate that agreement of the approximate percentage points of $U^{(s)}$ by chi-square and inverse chi-square distributions with Davis' (1970) exact values are excellent.

Table 9. Exact and approximate percentage points of $U^{(s)}$ by the F approximation.

$s=2$

	5%						1%					
	$m=0$			$m=4$			$m=0$			$m=4$		
n	Exact	Chi-square	F	Exact	Chi-square	F	Exact	Chi-square	F	Exact	Chi-square	F
5	1.43	1.44*	1.45	—	4.275*	X	2.17	2.26*	2.27	—	6.1471*	X
10	0.671	0.673*	0.680	1.927	1.928*	X	0.979	0.979*	0.986	2.516	2.520*	X
15	0.443	0.437*	0.443	1.236	1.233*	1.236	0.624	0.621*	0.624	1.563	1.561*	1.564
20	0.328	0.331	0.328	0.908	0.912	0.908	0.455	0.457	0.456	1.130	1.129	1.131
30	0.2155	0.2169	0.2158	0.592	0.594	0.592	0.2960	0.2959	0.2959	0.725	0.725	0.726

* Inverse chi-square approximation.

X F approximation not obtainable.

\- exact values not given by Davis.

Table 10. Exact and approximate percentage points of $U^{(s)}$

$s = 3$ 5%

n	$m = 0$ Davis	Chi-square	F	$m = 0.5$ Davis	Chi-square	F	$m = 2$ Davis	Chi-square	F
8	1.4730	1.4709*	1.4712	1.7628	X	X	2.60896	X	X
10.5	1.0945	1.0914*	1.0940	1.3679	1.3061*	1.3079	1.9292	1.9274*	X
13	0.8699	0.8667*	0.8698	1.0385	1.0348*	1.0384	1.5285	1.5269*	1.5286
15.5	0.7215	0.7278	0.7215	0.8608	0.8674	0.8608	1.2649	1.2728	1.2651
18	0.6162	0.6205	0.6162	0.7348	0.7393	0.7348	1.0786	1.0838	1.0787
23	0.4768	0.4793	0.4769	0.5682	0.5707	0.5683	0.8328	0.8356	0.8329
28	0.3888	0.3904	0.3888	0.4631	0.4657	0.4632	0.6781	0.6798	0.6781
33	0.3282	0.3292	0.3282	0.3908	0.3919	0.3908	0.5718	0.5729	—

$s = 3$ 1%

n	$m = 0$ Davis	Chi-square	F	$m = 0.5$ Davis	Chi-square	F	$m = 2$ Davis	Chi-square	F
8	2.030	2.034*	2.035	2.3863	X	X	3.4197	X	X
10.5	1.4731	1.4734*	1.4746	1.7268	1.7294*	1.7299	2.4637	2.4620*	X
13	1.1525	1.1531*	1.1540	1.3504	1.3488*	1.3522	1.9213	1.9213*	1.9239
15.5	0.9463	0.9469	0.9472	1.1078	1.1077	1.1089	1.5732	1.5717	1.5747
18	0.8023	0.8022	0.8029	0.9389	0.9381	0.9395	1.3313	1.3295	1.3323
23	0.6149	0.6144	0.6152	0.7189	0.7182	0.7193	1.0176	1.0161	1.0180
28	0.4983	0.4979	0.4985	0.5824	0.5817	0.5825	0.8233	0.8221	0.8235
33	0.4188	0.4184	0.4189	0.4893	0.4888	0.4894	0.6911	0.6902	—

* Inverse chi-square approximation

X F approximation not obtainable

\- values not computed.

Table 11. Exact and approximate percentage points of $U^{(s)}$

$s = 4$ 5%

n	$m = 0$ Davis	Chi-square	F	$m = 0.5$ Davis	Chi-square	F	$m = 1.5$ Davis	Chi-square	F
12.5	1.3638	1.3603*	1.3637	1.5853	1.5844*	—	2.0209	2.0205*	X
15	1.1217	1.1200*	1.1218	1.3033	1.3012*	1.3033	1.6601	1.6599*	—
17.5	0.9524	0.9574	0.9524	1.1062	1.1114	1.1063	1.4083	1.4139	1.4085
22.5	0.7313	0.7339	0.7313	0.8490	0.8517	0.8491	1.0800	1.0829	1.0801
27.5	0.5934	0.5950	0.5934	0.6887	0.6904	0.6888	0.8757	0.8774	0.8757

$s = 4$ 1%

n	$m = 0$ Davis	Chi-square	F	$m = 0.5$ Davis	Chi-square	F	$m = 1.5$ Davis	Chi-square	F
12.5	1.7348	1.7321*	1.7373	1.9902	1.9912	—	2.4902	2.3896*	X
15	1.4114	1.4124*	1.4130	1.6184	1.6176*	1.6201	2.0231	2.0244*	—
17.5	1.1892	1.1879	1.1902	1.3631	1.3612	1.3642	1.7028	1.7001	1.7041
22.5	0.9040	0.9028	0.9044	1.0357	1.0342	1.0362	1.2927	1.2907	1.2933
27.5	0.7289	0.7280	0.7292	0.8349	0.8338	0.8352	1.0415	1.0400	1.0418

* Inverse chi-square approximation

X F approximation not obtainable.

\- values not computed.

Table 12. Exact and approximate percentage points of $U^{(s)}$

$s = 5$ 5%

	$m = 0$			$m = 1.5$		
n	Davis	Chi-square	F	Davis	Chi-square	F
17	1.3720	1.3707*	1.3722	1.7523	1.7518*	—
22	1.0453	1.0441*	1.0454	1.3342	1.3335*	1.3343
27	0.8441	0.8430*	0.8441	1.0769	1.0763*	1.0769

$s=5$ 1%

	$m=0$			$m=1.5$		
n	Davis	Chi-square	F	Davis	Chi-square	F
17	1.6658	1.6658*	1.6673	2.0896	2.0900*	—
22	1.2858	1.2563*	1.2564	1.5745	1.5752*	1.5752
27	1.0076	1.0083*	1.0079	1.2628	1.2636*	1.2631

* Inverse chi-square approximation

- values not computed.

6. An example on application

The $U^{(s)}$ statistics is useful in testing hypotheses in multivariate analysis problems. An example given below uses the $U^{(s)}$ criterion to test the equality of p-dimensional mean vectors from p-variate normal populations.

Measurements of (1) head length, (2) height and (3) weight of 140 school boys of almost the same age from 6 different schools in an Indian city are given by Rao (1952, p. 263). The problem is to test any significant differences in the mean characteristics between schools. Let S_1 and S_2 be the variance-covariance matrices between and within schools for the three characteristics. Then

$$S_1 = \begin{array}{cl} \begin{array}{ccc} 1 & 2 & 3 \end{array} & \\ \begin{bmatrix} 752.0 & 214.2 & 521.3 \\ . & 151.3 & 401.2 \\ . & . & 161.7 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

$$S_2 = \begin{array}{cl} \begin{array}{ccc} 1 & 2 & 3 \end{array} & \\ \begin{bmatrix} 12809.3 & 1003.7 & 2671.2 \\ & 1499.6 & 4123.6 \\ & & 21009.60 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}$$

Here
$$n_1 = 5 \quad , \quad n_2 = 134$$
$$s = 3 \quad , \quad U^{(3)} = tr(s_1 s_2^{-1}) = 0.2016$$
$$m = \tfrac{1}{2}(5-3-1) = 0.5$$
$$n = \tfrac{1}{2}(134-3-1) = 65$$

Using the F approximation for $U^{(3)}$, we have

$$pr(U^{(3)} \geq 0.2016) = 0.042 \; .$$

Thus the differences between schools are significant at the 5% level of significance but not at the 1% level. This is in agreement with the findings of Rao (1952), who examined the data using Wilks' Λ criterion.

REFERENCES

Davis, A.W. (1968). A system of linear differential equations for the distribution of Hotelling's generalized T_0^2. Ann. of Math. Statist. 39, 815-32.

Davis, A.W. (1970). Exact distributions of Hotelling's generalized T_0^2. Biometrika 57, 187-91.

Davis, A.W. (1970). Further applications of a differential equation for Hotelling's generalized T_0^2. Ann. of the Inst. of Stat. Math. 22, 77-87.

Hotelling, H. (1951). A generalized T test and measure of multivariate dispersion. Proc. Second Berkeley Symp. 23-42.

Ito, K (1956). Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic. Ann. Math. Statist. 27, 1091-1105.

Johnson, N.L., Nixon, E., Amos, D.E. and Pearson, E.S. (1963). Table of percentage points of Pearson curves. Biometrika 50, 459-98.

Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics: continuous multivariate distributions. John Wiley and Sons, Inc., New York.

Lee, Y.S. (1971). Asymptotic formulae for the distribution of a multivariate test statistic: power comparisons of certain multivariate tests. Biometrika 58, 647-51.

Pearson, K. (1948). Tables of the incomplete beta-function. ed. K. Pearson, Cambridge: University Press.

Pearson, E.S. and Tiku, M.L. (1970). Some notes on the relationship between the distributions of central and non-central F. Biometrika 57, 175-9.

Pillai, K.C.S. (1955). Some new test criteria in multivariate analysis. Ann. Math. Statist. 26, 117-21.

Pillai, K.C.S. and Mijares, T.A. (1959). On the moments of the trace of a matrix and approximations to its distribution. Ann. Math. Statist. 30, 1136-40.

Pillai, K.C.S. and Samson, P. (1959). On Hotelling's generalization of T_0^2. Biometrika 46, 160-8.

Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. New York: Wiley.

Roy, S.N. (1953). On a heuristic method of test construction and its use in multivariate analysis. Ann. Math. Statist. 24, 220-38.

Schatzoff, M. (1966). Exact distributions of Wilks' likelihood ratio criterion. Biometrika 53, 347-58.

Schatzoff, M. (1966). Sensitivity comparisons among tests of the general linear hypothesis. Jour. of Amer. Stat. Assoc. 415-35.

Tiku, M.L. (1971). A note on the distribution of Hotelling's generalized T_0^2. Biometrika 58, 237-41.

Wilks, S.S. (1932). Certain generalizations in the analysis of variance. Biometrika 24, 471-94.