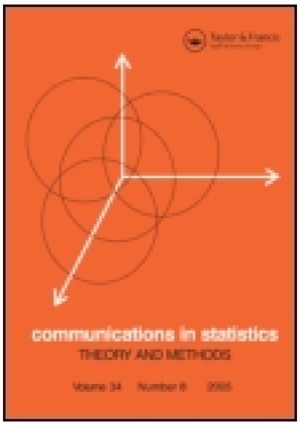

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

New approximations to the exact distribution of the kruskal-wallis test statistic

Ronald L Iman [a] & James M. Davenport [a]
[a] Sandia Laboratories, Albuquerque, New Mexico
[b] Texas Tech University, Lubbock, Texas
Published online: 27 Jun 2007.

COMMUN. STATIST.-THEOR. METH., A5(14), 1335-1348 (1976)

NEW APPROXIMATIONS TO THE EXACT DISTRIBUTION OF THE KRUSKAL-WALLIS TEST STATISTIC

Ronald L. Iman

Sandia Laboratories
Albuquerque, New Mexico

and

James M. Davenport

Texas Tech University
Lubbock, Texas

Key Words & Phrases: F (beta) approximations to the Kruskal-Wallis test; Satterthwaite approximate F test; Monte Carlo simulation; nonparametric tests based on ranks; rank tests.

ABSTRACT

Several approximations to the exact distribution of the Kruskal-Wallis test statistic presently exist. There approximations can roughly be grouped into two classes: (i) computationally difficult with good accuracy, and (ii) easy to compute but not as accurate as the first class. The purpose of this paper is to introduce two new approximations (one in the latter class and one which is computationally more involved), and to compare these with other popular approximations. These comparisons use exact probabilities where available and Monte Carlo simulation otherwise.

1. INTRODUCTION

Let $X_{i1},X_{i2},\ldots,X_{in_i}$ be a random sample of size n_i from population π_i, $i=1,2,\ldots,K$. Let N denote the total number of observations. The ranks $1,2,\ldots,N$ are assigned to the N observations. Let R_i represent the sum of the ranks assigned to the $i^{\underline{th}}$ sample. The Kruskal-Wallis test statistic is then given by

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{K} \frac{R_i^2}{n_i} - 3(N+1). \qquad (1.1)$$

Wallace (1959) gave three simplified beta approximations (B_2-I, B_2-II, B_2-III) for the exact distribution of H and compared his approximations with the beta approximation (B_1) given by Kruskal and Wallis (1952). The B_1 and B_2-I approximations involve computationally difficult procedures for adjusting the degrees of freedom. The B_2-II approximation simplifies the computational difficulty of B_2-I for the case of equal sample sizes. The B_1, B_2-I and B_2-II approximations all do a satisfactory job of approximating the exact distribution of H; however, all three involve fractional degrees of freedom and, hence, may require interpolation within the Beta or F tables. The third approximation, B_2-III, given by Wallace (1959) is applied by computing the ordinary analysis of variance on ranks and has as a test statistic the following F ratio

$$F = \frac{(N-K)(H)}{(K-1)(N-1-H)} \qquad (1.2)$$

with corresponding degrees of freedom K-1 and N-K. This approximation along with the χ^2 approximation with K-1 degrees of freedom given by Kruskal and Wallis (1952) have the advantage of being easy to compute. However, the χ^2 approximation is generally regarded as conservative for small n, a fact reaffirmed in recent papers by Gabriel and Lachenbruch (1969) and Chow, Dickinson and Champagne (1974). As will be demonstrated, the F statistic tends to result in a liberal test (anticonservative) for small n. Hence, an attempt has been made to find test statistics and approximating distributions more accurate than those already noted. Three such procedures are presented and considered in two sets of comparisons that include the χ^2 and F approximations.

2. QUADE APPROXIMATION

An approximation given in Alexander and Quade (1968), herein referred to as the Quade approximation, uses the F statistic with a reduction in the denominator degrees of freedom from N-K to N-K-1. This approximation follows from using the Wallace (1959) B_2-I adjustment factor for denominator degrees of freedom and then rounding the resulting fractional degrees of freedom to the nearest integer. This procedure is denoted by F*.

3. IMAN APPROXIMATION

This approximation is based on techniques given by Iman (1974, 1976) for the Wilcoxon signed rank test and the Wilcoxon-Mann-Whitney rank sum test. A test statistic based on a linear combination of the χ^2 and F approximations is designed to take advantage of the offsetting nature of these conservative and liberal approximations. The statistic has the form

$$J = \frac{(K-1)F+H}{2} = \frac{H}{2}\left(1 + \frac{N-K}{(N-1-H)}\right), \qquad (3.1)$$

where H and F are given by 1.1 and 1.2 respectively. Approximate critical values are given by

$$J_\alpha \approx \frac{(K-1)F_\alpha(K-1,N-K) + \chi^2_\alpha\,(K-1)}{2}, \qquad (3.2)$$

where $F_\alpha(K-1,N-K)$ and $\chi^2_\alpha\,(K-1)$ are right tail critical values from the F and χ^2 tables respectively, with the indicated degrees of freedom and tail probability equal to α.

4. SATTERTHWAITE APPROXIMATION

In certain ANOVA situations, it is desirable to let the data estimate the degrees of freedom of the associated F tests. One such technique makes use of the Satterthwaite (1941, 1946) estimator for the degrees of freedom of an approximate F. This technique is used in this situation in the following manner. Let v_i be the usual sample variance of the ranks of the i^{th} sample.

That is,

$$v_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_i)^2 , \qquad (4.1)$$

where $\bar{R}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}$. Then the linear combination of variance estimates

$$L = \sum_{i=1}^{K} (n_i - 1) v_i = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (R_{ij} - \bar{R}_i)^2 \qquad (4.2)$$

(which is the denominator sum of squares of the F statistic given by (1.2)), has estimated degrees of freedom given by

$$\hat{f} = \frac{\left(\sum_{i=1}^{K} (n_i - 1) v_i\right)^2}{\sum_{i=1}^{K} \frac{[(n_i - 1) v_i]^2}{(n_i - 1)}} . \qquad (4.3)$$

Then F given in (1.2) can be approximated as an F statistic with $K-1$ degrees of freedom in the numerator and $\hat{f}$ degrees of freedom in the denominator. Note that $\hat{f}$ will not usually be integer valued, and, hence, interpolation may be required. However, in most situations a decision can be reached without interpolation. This procedure is denoted by F_s.

Note that the degrees of freedom adjustments given by Wallace (1959) and the one given in Alexander and Quade (1968), are methods that reduce the degrees of freedom of the denominator of the approximating F-distribution for the statistic F given by (1.2). This is necessary, since as was indicated previously, the F approximation associated with (1.2) is a liberal test (anticonservative) for small n. A reduction in size of the approximate test will be accomplished by an appropriate reduction in the degrees of freedom of the denominator associated with the test. This need for reduction in the degrees of freedom is satisfied by (4.3), since an application of the Cauchy-Schwarz inequality shows that $\hat{f} \leq \sum_{i=1}^{K} (n_i - 1) = N-K$. That is, $\hat{f}$ has as a maximum value, the degrees of freedom, N-K, associated with the F approximation discussed in Section 1.

5. COMPARISON OF THE TEST STATISTICS

The four approximations χ^2, F, F* and J are compared in Table I for the balanced cases when K = 3, n = 2,3,...,8; K = 4, n = 2,3,4; K = 5, n = 2,3 and K = 6, n = 2. The exact probabilities used in the comparisons were obtained from the table given by Iman, Quade and Alexander (1975) except for the case for K = 6 which were generated by Iman but not published. Comparisons are made on the basis of the amount that the exact probability associated with each test statistic deviates from the desired probability of Type I error. This comparison could be misleading from the standpoint that the desired alpha levels are not always obtainable with discrete distributions. However, the table does provide some measure of the "surprise" under the null hypothesis. As an example of the construction of Table I, consider the case of the χ^2 approximation. The exact probability referred to in Table I for a given value of α and K is the $P(H \geq \chi^2_\alpha(K-1))$. The value given in Table I under the χ^2 column is then $P(H \geq \chi^2_\alpha(K-1))-\alpha$. The Satterthwaite procedure is not included in this comparison since this approximate test is conditioned on the data.

Three separate computer simulation studies were performed to further investigate these approximations and make comparisons. For each combination of K and N considered, one set of permutations was generated and was used for all alpha levels considered. The first study was performed using an IBM 370/155 to estimate the true alpha levels of the four statistics χ^2, F, F*, and J for the purposes of comparison. Values of K = 3,4,...,8 were considered for the balanced cases where n = 4,5,...,25. For each combination of K and n, 20,000 random permutations of the ranks 1,2,...,Kn were generated using a technique given by Durstenfeld (1964) and the IBM random number generator RAND. For each permutation of ranks the statistics H, F and J were

TABLE I

Summary of the Four Approximations for Small Samples

K	n	α	Exact Probability - Desired α: χ^2	F	F*	T
3	2	.10	-.1000	-.0333	-.0333	-.0333
		.05	-.0500	.0167	-.0500	.0167
	3	.10	.0000	.0321	.0000	.0000
		.05	-.0393	.0357	.0214	.0214
		.025	-.0250	.0036	.0000	.0000
		.01	-.0100	.0007	-.0064	.0007
		.005	-.0050	-.0014	-.0014	-.0014
	4	.10	-.0034	.0042	.0042	.0042
		.05	-.0097	.0075	.0075	.0075
		.025	-.0105	.0077	.0050	.0050
		.01	-.0095	.0059	.0059	.0046
		.005	-.0050	.0026	-.0002	-.0002
		.001	-.0010	.0014	-.0005	-.0005
	5	.10	-.0079	.0015	-.0005	-.0005
		.05	-.0060	.0054	.0009	.0009
		.025	-.0098	.0068	.0048	.0009
		.01	-.0067	.0037	.0023	.0011
		.005	-.0047	.0027	.0018	.0007
		.001	-.0010	.0008	.0002	-.0001
	6	.10	-.0013	.0032	.0010	.0010
		.05	-.0079	.0045	.0020	-.0010
		.025	-.0069	.0039	.0025	.0009
		.01	-.0057	.0035	.0027	.0002
		.005	-.0037	.0021	.0015	.0003
		.001	-.0010	.0008	.0006	.0001
	7	.10	-.0020	.0007	.0007	-.0007
		.05	-.0053	.0031	.0028	-.0009
		.025	-.0065	.0035	.0026	-.0005
		.01	-.0046	.0027	.0019	-.0001
		.005	-.0031	.0019	.0013	.0001
		.001	-.0010	.0007	.0005	.0000
	8	.10	-.0024	.0020	.0002	.0002
		.05	-.0048	.0027	.0014	-.0007
		.025	-.0053	.0028	.0022	-.0004
		.01	-.0039	.0021	.0018	-.0002
		.005	-.0032	.0014	.0012	-.0002
		.001	-.0009	.0006	.0004	-.0000

(continued)

Table I, continued

			Exact Probability - Desired α			
K	n	α	χ^2	F	F*	J
4	2	.10	-.0905	.0238	-.0238	.0143
		.05	-.0500	.0167	-.0119	.0167
		.025	-.0250	-.0155	-.0155	-.0155
		.01	-.0100	-.0005	-.0100	-.0005
		.005	-.0050	.0045	-.0050	.0045
	3	.10	-.0150	.0169	.0027	.0072
		.05	-.0295	.0099	-.0074	.0025
		.025	-.0230	.0049	-.0033	.0007
		.01	-.0100	.0037	-.0062	.0008
		.005	-.0050	.0028	-.0040	-.0006
		.001	-.0010	.0004	-.0010	-.0004
	4	.10	-.0114	.0075	.0040	.0035
		.05	-.0164	.0079	.0047	.0007
		.025	-.0159	.0055	.0032	-.0002
		.01	-.0090	.0030	.0011	-.0009
		.005	-.0050	.0017	.0005	-.0005
		.001	-.0010	.0006	.0003	.0000
5	2	.10	-.0746	.0280	-.0090	.0100
		.05	-.0500	.0135	-.0119	-.0013
		.025	-.0250	.0067	-.0028	.0004
		.01	-.0100	-.0005	-.0047	-.0005
		.005	-.0050	.0003	-.0039	.0003
		.001	-.0010	.0001	-.0010	.0001
	3	.10	-.0280	.0101	.0023	-.0008
		.05	-.0307	.0062	.0005	-.0036
		.025	-.0225	.0042	.0011	-.0021
		.01	-.0100	.0033	.0011	-.0003
		.005	-.0050	.0020	.0003	-.0006
		.001	-.0010	.0004	-.0001	.0002
6	2	.10	-.0729	.0022	-.0092	-.0041
		.05	-.0500	.0085	-.0058	-.0003
		.025	-.0250	.0041	-.0047	-.0022
		.01	-.0100	.0014	-.0034	-.0013
		.005	-.0050	.0005	-.0018	-.0004
		.001	-.0010	.0001	-.0009	.0001

computed. Counts were then maintained for the number of times that (i) $H \geq \chi^2_\alpha(K-1)$, (ii) $F \geq F_\alpha(K-1,N-K)$, (iii) $F \geq F_\alpha(K-1,N-K-1)$ and (iv) $J \geq J_\alpha$. Division of these counts by 20,000 provides an estimate of the true alpha level of each statistic. Values of α considered were .10, .05, .025, .01, .005 and .001.

The second study was performed on an IBM 370/145 to make comparisons of the four statistics χ^2, F, F_s and J. Values of $K = 3,4,\ldots,8$ were considered for the balanced cases where $n = 2(1)10$, $12(2)22$, 25. For each combination of K and n, 4,000 random permutations of the ranks $1,2,\ldots,Kn$ were generated again using the Durstenfeld (1964) algorithm and the random number generator given by Marsaglia and Bray (1968). As before counts were maintained for the number of times rejection would have occurred. Values of α considered were .10, .05, and .025.

Plots of the two above simulation studies are available upon request.

The third study was performed on a CDC/6600 to again make comparisons of the four statistics χ^2, F, F_s, and J. Values of $K = 3$ and 8 were considered for the balanced cases where $n = 2(1)10$, $12(2)22$, 25, with $\alpha = .05$ and $.025$. The Durstenfeld algorithm was again used, and for each combination of K and n, 50,000 random permutations of the ranks $1,2,\ldots,Kn$ were generated. The random number generator used was written by Bell and Holdridge (1967) at Sandia Laboratories. These results are presented in four plots given as Figures 1 through 4.

6. SUMMARY

The tests based on F^*, F_s and J when used in conjunction with the approximate critical values given herein improve upon the accuracy of the F and χ^2 procedures for the balanced cases considered. The F^* procedure has the advantage of requiring no additional calculations beyond the F statistic and no table interpolation is required. For small values of n F^* makes a noticeable improvement over F, but for $n > 10$ the F^* procedure is almost identical to the ordinary F statistic as is demonstrated in the plots that are available upon request. Hence, the results for the F^* procedure are not included in Figures 1 through 4. The J

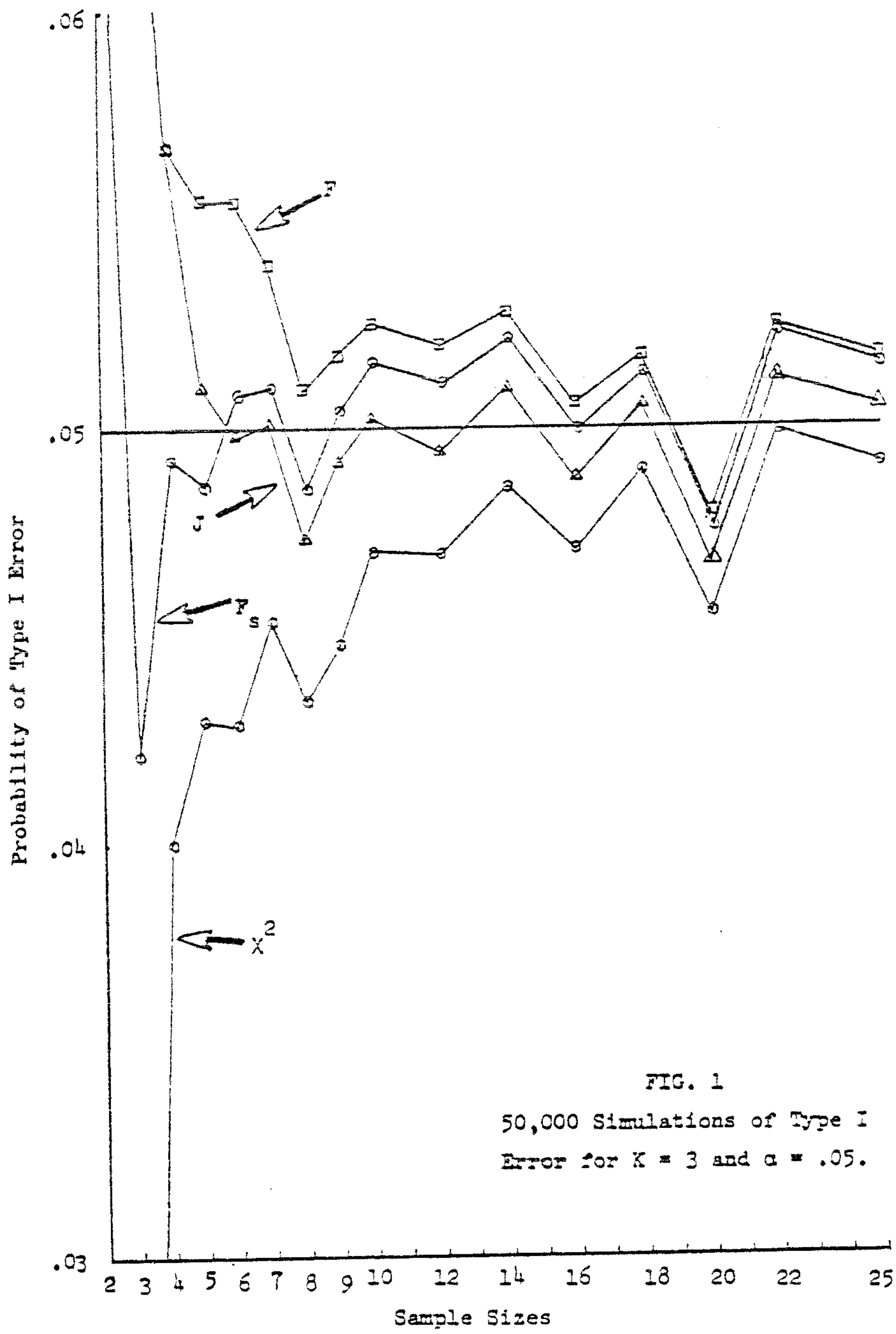

FIG. 1
50,000 Simulations of Type I
Error for K = 3 and α = .05.

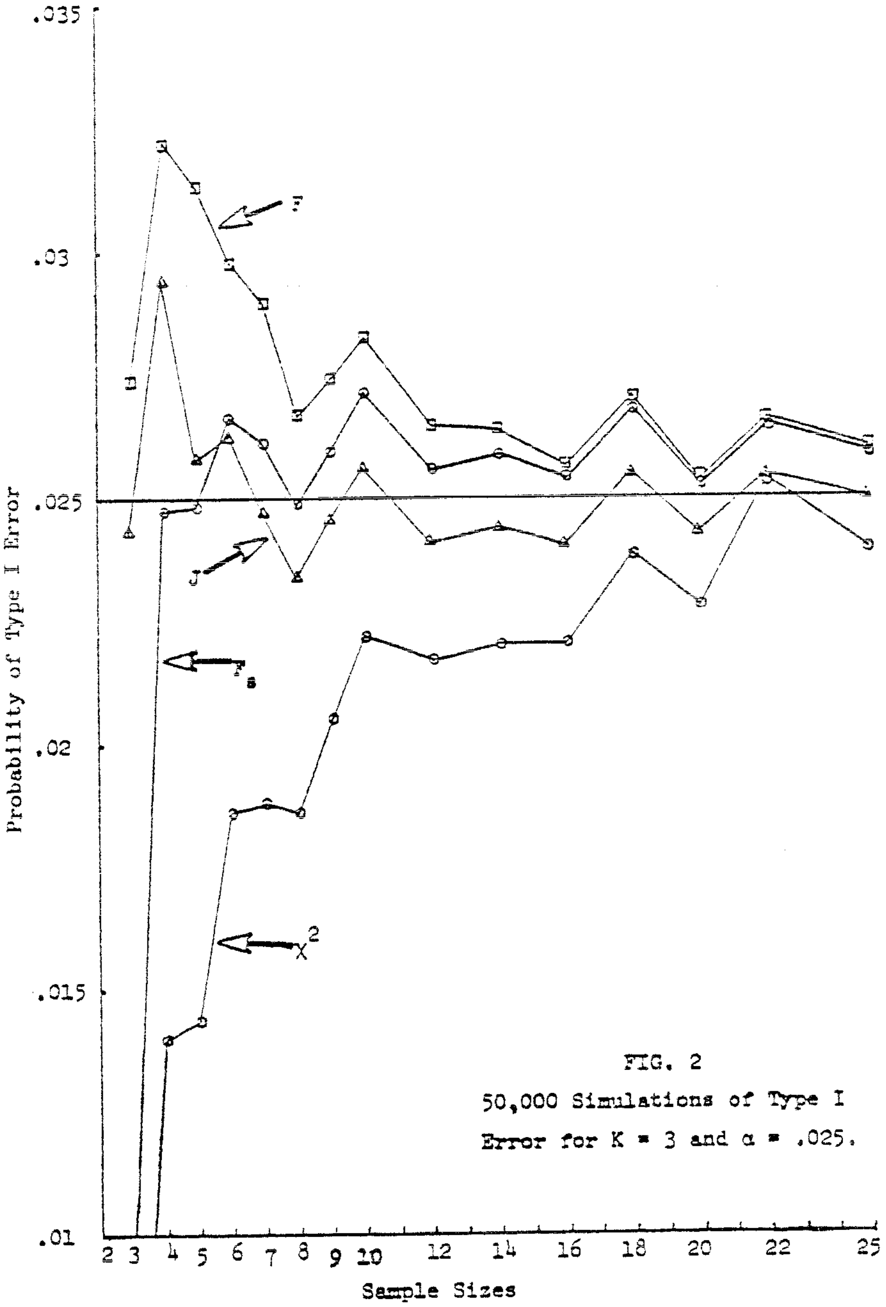

FIG. 2
50,000 Simulations of Type I Error for K = 3 and α = .025.

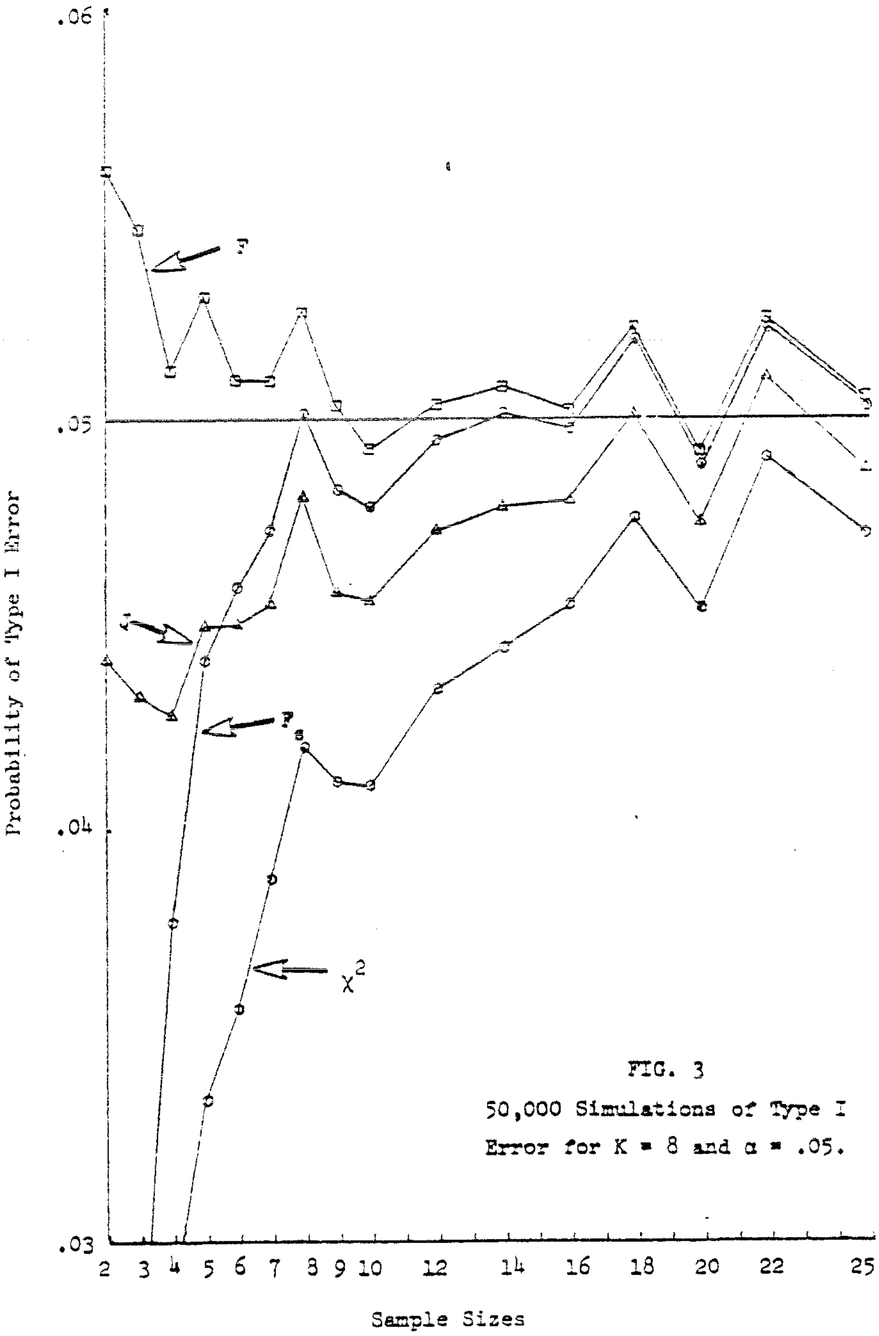

FIG. 3
50,000 Simulations of Type I Error for K = 8 and α = .05.

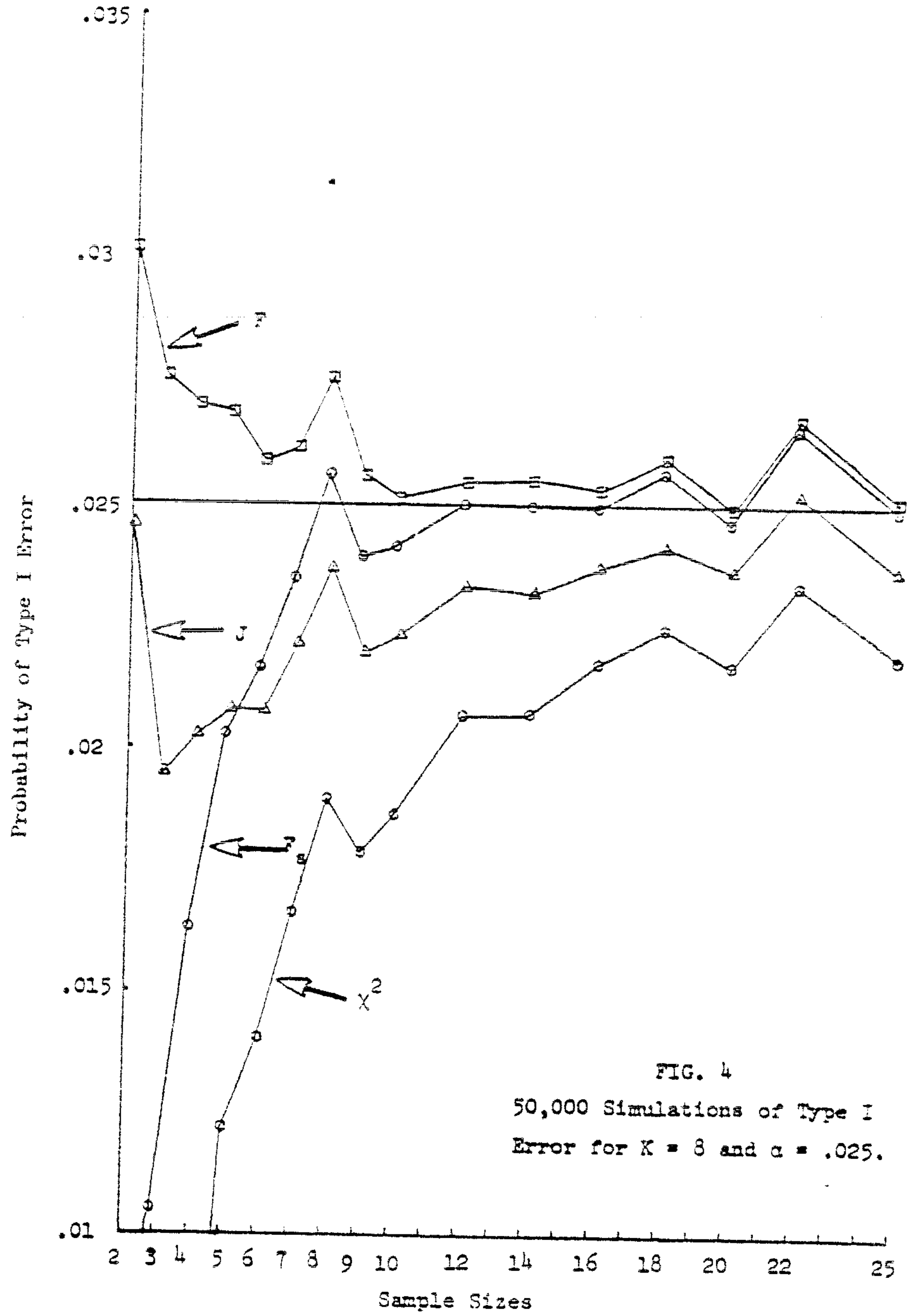

FIG. 4
50,000 Simulations of Type I
Error for K = 8 and α = .025.

approximation requires two table entries to obtain approximate critical values but does seem to exhibit very good agreement with expected type I error rates set by the experimenter. The F_s statistic exhibits a tendency to be conservative for the very small sample sizes, but beyond n = 5 the F_s statistic and the J statistic are very nearly identical. In fact, the size of the F_s test is larger for n greater than 5. For n = 2,3,4, and 5 the size of the J test is larger.

For the unbalanced case the procedures using F* and J showed improvement over the χ^2 and F approximations for the small values of n_i for which the exact probabilities are available in the table of Iman, Quade, and Alexander (1975). However, no computer simulation results were attempted for the unbalanced case due to the large number of permutations of sample sizes.

BIBLIOGRAPHY

Alexander, D. A. and Quade, Dana (1968). On the Kruskal-Wallis three sample H-statistic. Inst. of Statist. Mimeo Series No. 602, Univ. of North Carolina, Chapel Hill.

Bell, S. and Holdridge, D. B. (1967). The random number generators for the Sandia 3600 and their statistical properties. SC-RR-67-65, Sandia Laboratories, Albuquerque, N. M.

Chow, B., Dickinson, P., and Champagne, Q. (1974). An approximation of the critical values of the Kruskal-Wallis H-test using a Monte Carlo sampling technique. J. Quality Control 6, 95-97.

Durstenfeld, R. (1964). Random Permutations. Commun. Assoc. Computing Mach. Algorithm 235.

Gabriel, K. R. and Lachenbruch, P. A. (1969). Non-parametric ANOVA in small samples: A Monte Carlo study of the adequacy of the asymptotic approximation. Biometrics 25, 593-96.

Iman, R. L. (1974). Use of a t-statistic as an approximation to the exact distribution of the Wilcoxon signed ranks test statistic. Commun. Statist. 3, 795-806.

Iman, R. L. (1976). An approximation to the exact distribution of the Wilcoxon-Mann-Whitney rank sum test. Commun. Statist. 5(7).

Iman, R. L., Quade, D., and Alexander, D. (1975). Exact probability levels for the Kruskal-Wallis test. Selected Tables in Mathematical Statistics, Vol. 3. (Harter, H. L. & Owen, D. B., Eds.). Providence, R. I.: Amer. Math. Society

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. J. Amer. Statist. Assoc. 47, 583-621.

Marsaglia, G. and Bray, T. A. (1968). One-line random number generators and their use in combinations. Commun. Assoc. Computing Mach. 11, 757-59.

Satterthwaite, F. E. (1941). Synthesis of Variance. Psychometrika 6, 309-16.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. Biometrics 2, 110-14.

Wallace, D. L. (1959). Simplified Beta-approximations to the Kruskal-Wallis H test. J. Amer. Statist. Assoc. 54, 225-30.

Refereed by:

Ray A. Waller, Los Alamos Scientific Laboratory, N.M.

Gerald R. Hobbs, Jr., State University College at Potsdam, N.Y.