Survey Review Vol. 27, 210, October 1983

A NOTE ON THE CALCULATION OF THE χ^2 AND F STATISTICS

K. S. Milford

Department of Surveying, University of Witwatersrand

Abstract

For many survey applications, the calculation of both the χ^2 and F statistics are required for the analysis of the confidence regions pertaining to individual or correlated statistics. To facilitate the use and computation of these parameters, several selected approximations are given and methods are suggested for the inclusion during the routine least squares adjustment techniques.

Introduction

The use of least squares adjustment techniques has become standard practice for the processing of survey data related to precision control surveys in engineering and allied fields. Associated with the adjustment procedures is the analysis of the results and in this regard, the use of both the χ^2 and F statistics play an important part [**6**]. The use of relatively simple approximations to the rigorous formula describing these distributions, allows the inclusion of an elementary statistical analysis to be programmed as part of a computer adjustment routine.

1. The χ^2 Distribution

For n stochastically independent variables $u_1, u_2, u_3, \ldots, u_n$, each of which are normally distributed, the distribution of the sum of the squares of these variables is denoted by the χ^2 distribution, where

$$\chi^2 = \sum_{i=1}^{v} u_i^2 \qquad \text{for } v \text{ degrees of freedom.}$$

The probability distribution function (p.d.f.) for χ^2 is given by

$$p\{\chi^2\}d(\chi^2) = \frac{1}{2^{v/2}\Gamma\left(\frac{v}{2}\right)} (\chi^2)^{(v/2)-1} e^{-(\chi^2/2)} d(\chi^2)$$

where for $v > 2$, the Gamma-function $\Gamma^{(n/2)}$ is given by

$$\Gamma\left(\frac{v}{2}\right) = \left(\frac{v}{2}-1\right)! = \begin{cases} \left(\frac{v}{2}-1\right)\left(\frac{v}{2}-2\right) & \ldots 3.2.1 \text{ for } n \text{ even} \\ \\ \left(\frac{v}{2}-1\right)\left(\frac{v}{2}-2\right) & \ldots \frac{3}{2}.\frac{1}{2}.\sqrt{\pi} \text{ for } n \text{ odd} \end{cases}$$

and

$$\Gamma(1) = 1 \quad \text{and} \quad \Gamma(\tfrac{1}{2}) = \pi$$

The calculation of the χ^2 statistic at a given probability level (α) and number of degrees of freedom (ν) becomes complicated if the exact formula as given above has to be used. However, a commonly used approximation [3], [5] was found to be adequate for survey purposes as it is simple to calculate and takes up very little space in a computer program. The formula used is:

$$\chi^2_{\nu,\alpha} = \nu\left\{1 - \frac{2}{9^\nu} + U_\alpha \sqrt{\frac{2}{9^\nu}}\right\}^3$$

where U_α is the normal variate and has the value of 1·645 for $\alpha = 0{\cdot}95$ and 1·960 for $\alpha = 0{\cdot}975$.

As can be seen from the accompanying tables relating to the χ^2 distribution, the percentage differences between the values calculated from the approximate formula, and values tabulated in any standard set of statistical tables, showed deviations of less than 2% under the following conditions:

Probability level	*Degrees of freedom*
$\alpha = 0{\cdot}025$	$\nu > 7$
$\alpha = 0{\cdot}05$	$\nu > 5$
$\alpha = 0{\cdot}95$	$\nu > 2$
$\alpha = 0{\cdot}975$	$\nu > 2$

2. The Fischer's (F) Distribution

The F distribution describes the statistical distribution of the ratio of variances, and hence can be derived from the χ^2 distribution since:

$$s_1^2 = \sigma^2 \frac{\chi^2}{\nu_2} \quad \text{and} \quad s_2^2 = \sigma^2 \frac{\chi^2}{\nu_2}$$

The variance ratio F is then given by

$$F = \frac{s_1^2}{s_2^2}\frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2}$$

and the distribution function of F is

$$p\{F\} = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \nu_1^{\nu_1/2}\nu_2^{\nu_2/2} \frac{F^{((\nu_1/2)-1)}}{(\nu_2+F\nu_1)^{(\nu_1+\nu_2/2)}}$$

This function may be transformed into a less complicated expression by using the distribution $z_{\nu_1,\nu_2} = \frac{1}{2}\log F_{\nu_1,\nu_2}$ [7] and using the moment generating function:

$$E\{e^{tz}\} = E\{F^{\frac{1}{2}t}\} = \left(\frac{\nu_2}{\nu_1}\right)^{\frac{1}{2}t} \frac{\Gamma(\frac{1}{2}(\nu_1+t))\Gamma(\frac{1}{2}(\nu_2-t))}{\Gamma(\frac{1}{2}\nu_1)\Gamma(\frac{1}{2}\nu_2)}$$

The moments ($t = 1, 2, 3, \ldots, n$) are all finite.
The need for seeking a relatively simple approximate formula for calculating F becomes obvious when considering the formula explicitly describing the F distribution.

2.1 *Approximations to the F distribution*

Of the various approximations to F, only four methods were selected for this study, as these methods provide for solutions within the range of the degrees of freedom normally associated with survey. Method 1 is an exception to these criteria but was included because of its simplicity.

2.1.1 *Method* 1

For large values of ν_1 and ν_2, the z distribution tends towards the normal distribution with an expected value of $\frac{1}{2}(\nu_1^{-1}-\nu_2^{-1})$ $(=\delta)$ and a variance of $\frac{1}{2}(\nu_1^{-1}+\nu_2^{-1})$ $(=\sigma^2)$ [5]. The approximate formula is accredited to Fischer and is given by:

$$z_{\nu_1,\nu_2,\alpha} = \tfrac{1}{2}(\nu_2^{-1}-\nu_1^{-1}) + U_\alpha\{\tfrac{1}{2}(\nu_1^{-1}+\nu_2^{-1})\}^{\frac{1}{2}}$$

or

$$z_{\nu_1,\nu_2,\alpha} = -\delta + U_\alpha \cdot \sigma$$

2.1.2 *Method* 2

Applying a Cornish–Fischer type expansion to the z distribution (see [5]), a more complex expression for the cumutant z is obtained:

$$\begin{aligned} z_{\nu_1,\nu_2,\alpha} = {} & U_\alpha\sigma + \tfrac{1}{3}\delta(U_\alpha^2+2) \\ & + \sigma\left\{\frac{\sigma^2}{12}(U_\alpha^3+3U_\alpha) + \frac{1}{36}\left(\frac{\delta}{\sigma}\right)^2(U_\alpha^3+11U_\alpha)\right. \\ & + \frac{1}{30}\,\delta\sigma^2(U_\alpha^4+9U_\alpha^2+8) \\ & \left. - \frac{1}{810}\frac{\delta^3}{\sigma^2}(3U_\alpha^4+7U_\alpha^2-10) + \ldots\right\} \end{aligned}$$

This formulae gave values of F $(=e^{2z})$ which consistently differed from the required values. A scale factor of 1·046957 was applied to the F values as calculated from the approximate formula and the expression used in calculating approximate F values was

$$F = 1{\cdot}046957e^{2z}$$

2.1.3 *Method* 3

Cochran [2], [5] has suggested the use of a modification of the basic formula from Method 2 and gives as an expression for z:

$$z_{\nu_1,\nu_2,\alpha} = \delta\left\{1+\frac{1}{3}(U_\alpha^2-1)\right\} + U_\alpha\sigma\left\{1-\frac{1}{6}(U_\alpha^2+3)\sigma^2\right\}^{-\frac{1}{2}}$$

2.1.4 *Method* 4

Scheffe and Tukey [7] give an alternate method of calculating F when only one of ν_1 and ν_2 is large. This method is based on the fact that F_{ν_1,ν_2} will be distributed approximately as $\chi^2_{\nu_1/\nu_1}$ for ν_2 large. The formula is then:

$$F_{\nu_1,\nu_2,\alpha} = \frac{\nu_1}{\chi^2_{\nu_1,\alpha}} + \frac{\nu_1}{\nu_2}\left(\frac{\frac{1}{2}\nu_1^{-1}}{\chi^2_{\nu_1,\alpha}} - \frac{1}{2}\right)^{-1}$$

For $\nu_1 = 2$, $\chi^2_{\nu_1,\cdot 95} = 5{\cdot}99$ and the formula then becomes:

$$F_{2,\nu_2,\cdot 95} = \frac{5{\cdot}99}{(2\nu_2 - 5{\cdot}99)}$$

2.2 *Evaluation of Formulae*

In order to evaluate the formulae, a computer routine has been used in which ν_1 was fixed at 2 and ν_2 varied from 5 to 100. The value of F calculated from each method are compared with standard tabulated values of the F distribution (see table 3) and plotted in graph 1. Graph 2 illustrates the percentage deviation of the calculated values of F from the true values.

From the tabulated and graphical data, the following results were obtained:

(i) Method 1 produces values of F with percentage deviations of less than 2% for $9 < \nu_2 < 16$.

(ii) For $\nu_2 < 8$, methods 1, 3 and 4 showed deviations of greater than 2%, with the error becoming asymptotic as ν_2 approaches 1.

(iii) Method 4 produces values of F which are within 1% of the tabulated values for $\nu_2 > 20$ and within 2% for $15 < \nu_2 < 19$.

(iv) The modified method 2 gave results which differed by less than 0·3% for $\nu_2 > 15$.

3. Conclusion

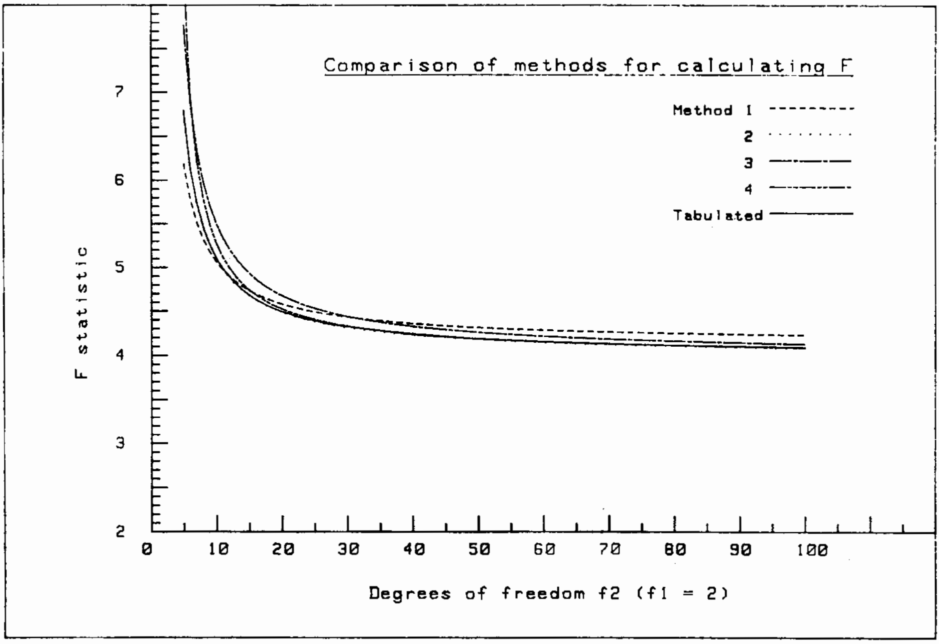
The approximate formula for the χ^2 distribution produces results which are within 2% of the correct values within the range of the degrees of freedom normally associated with precision surveys. In addition, the simplicity of the formula lends itself to ease of programming and can therefore be recommended for general use.

The modification applied to method 2 produced the most accurate results and the programming of this formula takes no more than 2 to 3 lines depending upon the computer and language being used. Method 4 is simple in format and can be recommended for use in the analysis of networks containing more than 15 degrees of freedom.

References

1. Beyer, 2nd Ed. *CRC Handbook of Tables for Probability and Statistics.*
2. Cochran, W. G., 1940. Note on an Approximate Formula for the Significance Levels of *z*. *Annals of Mathematical Statistics*, 11.
3. Hald, A. *Statistical Theory with Engineering Applications.* John Wiley & Sons, 1952.
4. Hemmerk, W. J. *Statistical Computations on a Digital Computer.* Blaisdell Pub. Comp., 1967.
5. Johnson, N. I. and Kotz, S. *Distributions in Statistics—Continuous Univariate Distributions.* Houghton Mifflin Company, 1970.
6. Milford, K. S. The Application of Statistics to the Error (Confidence) Ellipse. *S.A.S.J.* (in print), 1982.
7. Scheffe, H. and Tukey, J. W., 1944. A Formula for Sample Sizes for Population Tolerance Limits. *Annals of Mathematical Statistics*, 15.

GRAPH 1



GRAPH 2

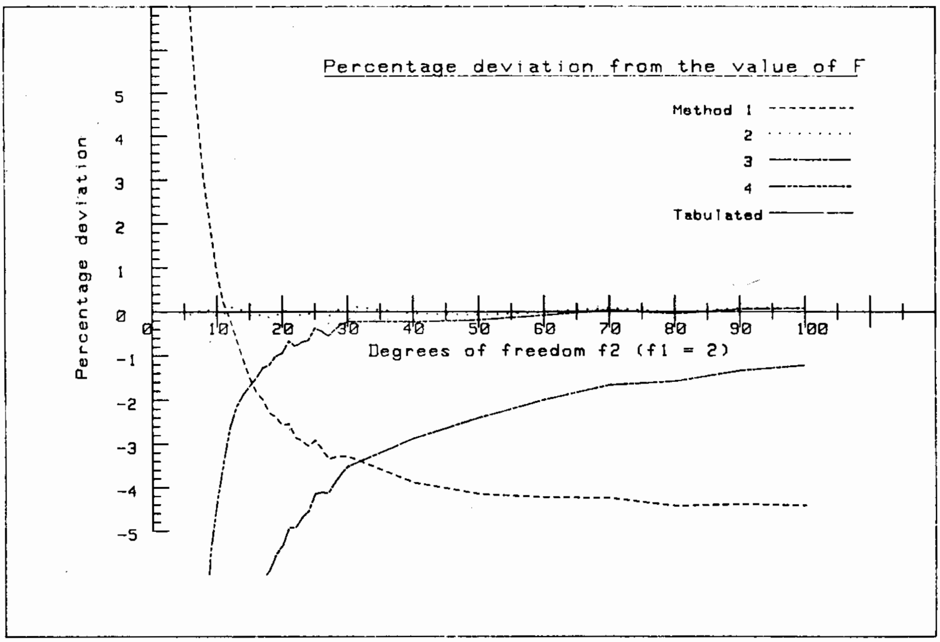


TABLE 1. Approximations of the Chi-squared statistic

Df. \ %	Calc. ·025	Tab. ·025	Percent diff.	Calc. ·975	Tab. ·975	Percent diff.
1	−·00	0·00	412·33	4·93	5·02	1·83
2	0·03	0·05	48·34	7·34	7·38	0·59
3	0·18	0·22	16·03	9·32	9·35	0·27
4	0·45	0·48	7·18	11·13	11·10	−·25
5	0·80	0·83	4·01	12·82	12·80	−·17
6	1·21	1·24	2·75	14·44	14·40	−·29
7	1·66	1·69	1·75	16·01	16·00	−·05
8	2·15	2·18	1·27	17·53	17·50	−·18
9	2·67	2·70	0·94	19·02	19·00	−·11
10	3·22	3·25	0·84	20·48	20·50	0·09
11	3·79	3·82	0·71	21·92	21·90	−·09
12	4·38	4·40	0·41	23·34	23·30	−·16
13	4·99	5·01	0·44	24·74	24·70	−·15
14	5·61	5·63	0·38	26·12	26·10	−·08
15	6·24	6·26	0·27	27·49	27·50	0·04
16	6·89	6·91	0·30	28·85	28·80	−·16
17	7·55	7·56	0·18	30·19	30·20	0·02
18	8·21	8·23	0·20	31·53	31·50	−·09
19	8·89	8·91	0·22	32·85	32·90	0·14
20	9·57	9·59	0·16	34·17	34·20	0·08
21	10·27	10·30	0·32	35·48	35·50	0·05
22	10·97	11·00	0·30	36·78	36·80	0·05
23	11·67	11·70	0·22	38·08	38·10	0·06
24	12·39	12·40	0·11	39·37	39·40	0·08
25	13·11	13·10	−·04	40·65	40·60	−·12
26	13·83	13·80	−·22	41·93	41·90	−·06
27	14·56	14·60	0·27	43·20	43·20	0·01
28	15·29	15·30	0·03	44·46	44·50	0·08
29	16·03	16·00	−·21	45·73	45·70	−·06
30	16·78	16·80	0·13	46·98	47·00	0·04
40	24·42	24·40	−·09	59·35	59·30	−·08
50	32·35	32·40	0·16	71·42	71·40	−·03
60	40·47	40·50	0·07	83·30	83·30	−·00
70	48·75	48·80	0·10	95·03	95·00	−·03
80	57·15	57·20	0·09	106·63	106·60	−·03
90	65·64	65·60	−·06	118·14	118·10	−·03
100	74·22	74·20	−·02	129·56	129·60	0·03

TABLE 2. Approximations of the Chi-squared statistic

df. \ %	Calc. ·05	Tab. ·05	Percent diff.	Calc. ·95	Tab. ·95	Percent diff.
1	0·00	0·00	100·00	3·75	3·84	2·47
2	0·08	0·10	23·31	5·94	5·99	0·88
3	0·33	0·35	6·79	7·78	7·81	0·44
4	0·69	0·71	2·93	9·46	9·49	0·35
5	1·13	1·15	1·91	11·04	11·10	0·50
6	1·62	1·64	1·20	12·57	12·60	0·24
7	2·15	2·17	0·72	14·05	14·10	0·37
8	2·72	2·73	0·32	15·49	15·50	0·06
9	3·31	3·33	0·45	16·90	16·90	−·02
10	3·93	3·94	0·22	18·29	18·30	0·04
11	4·57	4·57	0·08	19·66	19·70	0·19
12	5·22	5·23	0·22	21·01	21·00	−·07
13	5·88	5·89	0·09	22·35	22·40	0·22
14	6·56	6·57	0·09	23·67	23·70	0·11
15	7·26	7·26	0·07	24·99	25·00	0·06
16	7·96	7·96	0·05	26·29	26·30	0·05
17	8·67	8·67	0·04	27·58	27·60	0·08
18	9·39	9·39	0·05	28·86	28·90	0·14
19	10·11	10·10	−·12	30·14	30·10	−·12
20	10·85	10·90	0·49	31·40	31·40	−·01
21	11·59	11·60	0·11	32·66	32·70	0·11
22	12·33	12·30	−·28	33·92	33·90	−·05
23	13·09	13·10	0·10	35·17	35·20	0·10
24	13·84	13·80	−·32	36·41	36·40	−·02
25	14·61	14·60	−·05	37·65	37·70	0·14
26	15·38	15·40	0·16	38·88	38·90	0·05
27	16·15	16·20	0·32	40·11	40·10	−·02
28	16·92	16·90	−·15	41·33	41·30	−·08
29	17·71	17·70	−·03	42·55	42·60	0·11
30	18·49	18·50	0·06	43·77	43·80	0·07
40	26·51	26·50	−·03	55·76	55·80	0·08
50	34·76	34·80	0·11	67·50	67·50	−·00
60	43·19	43·20	0·03	79·08	79·10	0·03
70	51·74	51·50	−·07	90·53	90·50	−·03
80	60·39	60·40	0·02	101·88	101·90	0·02
90	69·12	69·10	−·04	113·14	113·10	−·04
100	77·93	77·90	−·04	124·34	124·30	−·03

TABLE 3. Approximations of the F statistic—95% level

Degrees of freedom		Method 1	Method 2	Method 3	Method 4	Tabulated
2	5	5·19	5·78	6·77	7·47	5·79
2	6	4·79	5·14	5·90	5·98	5·14
2	7	4·52	4·74	5·36	5·23	4·74
2	8	4·32	4·46	4·98	4·79	4·46
2	9	4·18	4·26	4·71	4·49	4·26
2	10	4·06	4·10	4·50	4·28	4·10
2	11	3·97	3·98	4·34	4·12	3·98
2	12	3·90	3·89	4·21	3·99	3·89
2	13	3·83	3·81	4·10	3·89	3·81
2	14	3·78	3·74	4·01	3·81	3·74
2	15	3·74	3·68	3·93	3·74	3·68
2	16	3·70	3·63	3·87	3·68	3·63
2	17	3·66	3·59	3·81	3·64	3·59
2	18	3·63	3·55	3·76	3·59	3·55
2	19	3·60	3·52	3·72	3·56	3·52
2	20	3·58	3·49	3·68	3·52	3·49
2	21	3·56	3·47	3·64	3·49	3·47
2	22	3·54	3·44	3·61	3·47	3·44
2	23	3·52	3·42	3·58	3·44	3·42
2	24	3·50	3·40	3·55	3·42	3·40
2	25	3·49	3·38	3·53	3·40	3·39
2	26	3·47	3·37	3·51	3·38	3·37
2	27	3·46	3·35	3·49	3·37	3·35
2	28	3·45	3·34	3·47	3·35	3·34
2	29	3·44	3·33	3·45	3·34	3·33
2	30	3·43	3·32	3·44	3·33	3·32
2	40	3·36	3·23	3·32	3·24	3·23
2	50	3·31	3·18	3·26	3·19	3·18
2	60	3·28	3·15	3·21	3·15	3·15
2	70	3·26	3·13	3·18	3·13	3·13
2	80	3·25	3·11	3·16	3·11	3·11
2	90	3·24	3·10	3·14	3·10	3·10
2	100	3·23	3·09	3·13	3·09	3·09