PERCENTAGE POINTS FOR THE FISHER-COCHRAN TEST
FOR EQUALITY OF VARIANCES

BY

HERBERT SOLOMON and MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 408
AUGUST 19, 1988

PREPARED UNDER CONTRACT
N00014-86-K-0156 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PERCENTAGE POINTS FOR THE FISHER-COCHRAN TEST
FOR EQUALITY OF VARIANCES

BY

HERBERT SOLOMON and MICHAEL A. STEPHENS

TECHNICAL REPORT NO. 408
AUGUST 19, 1988

Prepared Under Contract
N00014-86-K-0156 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PERCENTAGE POINTS FOR THE FISHER-COCHRAN TEST
FOR EQUALITY OF VARIANCES

by
Herbert Solomon

and
Michael A. Stephens

1. Introduction.

A well known test for equality of normal population variances, based on sample variances, was introduced by Cochran (1941). Suppose \( \sigma_i^2, i = 1, \ldots, n \) represents the population variances of \( n \) normal populations, and let independent sample variances, each based on \( k \) degrees of freedom, be \( s_i^2, i = 1, \ldots, n \). Suppose the \( s^2 \) are ranked, so that the ordered variables are \( s_{(1)}^2, s_{(2)}^2, \ldots, s_{(n)}^2 \). To test \( H_0 \): the \( \sigma_i^2 \) are all equal (suppose the common value is \( \sigma^2 \)), Cochran (1941) introduced the test statistic

\[
Z = \frac{s_{(n)}^2}{\sum_{i=1}^{n} s_i^2},
\]

which compares the largest sample variance with the sum of the sample variances. Clearly the intent is to discover if one variance is an outlier (too large), and, in general, \( H_0 \) will be rejected for large values of \( Z \). However, excessively small values of \( Z \) could be useful also: they will suggest that in some way the sample variances are more homogeneous than should be expected from a set of normal variances all from equi-variable populations, an event which might occur if samples with widely disperse variances have been omitted from a set of data. Such an event would be important, for example, in monitoring quality or process control.

Cochran gave some distributional results for \( Z \), and these were used by Eisenhart and Solomon (1947) to construct tables for the upper 5% and upper 1% points, 95th and 99th percentiles to supplement the very brief table presented by Cochran in his article. Another table with 5% and 1% points, for \( n \leq 20 \) and for slightly different values of \( k \), was given.
by Yamauti (1972). The Eisenhart and Solomon tables have been frequently reproduced and we now augment those tables in what follows by producing the additional percentiles listed in Table 1 for $n = 2(1)10, 12(2)20, 25(5)50, 60, 120$ and $k = 1(1)20, 30$. When $k = 2$, the test statistic serves as the basis of a significance test for any particular term in the harmonic analysis of a series as was demonstrated by Fisher (1929) who also provided a brief table of 95% and 99% points. In another paper, Fisher (1940) provided another brief table that also gave 95% values for the largest and second largest fractions. Tables for the Fisher problem, namely $k = 2$ were supplemented extensively by Nowroozi (1967) but to our knowledge more extensive tables for other values of $k$ have not been produced. It might be useful to have such points, and we give them in the tables below.

An obvious application would be to process control or quality control. Historically, this was mostly examined using, say, the mean of a sample taken daily, but it has become increasingly the practice to examine the variance also, for stability of the process. Thus, for example, the seven variances in a week might be examined to see if these were homogeneous, using the test given below. One might not always wish to be limited to comparing the largest variance with the total, and, for example, the two largest of the week or the four largest of a monthly set of variances, might be tested as too large. The theory given below can be adapted to provide such a test, and work is in progress to provide tables for this more general situation. The test procedure is given in Section 2, followed by theoretical results in Section 3.


The test of $H_0$ thus proceeds as follows.

1. Suppose for the $i$th population the sample variance $s_i^2$ is given, (the unbiased estimator of $\sigma_i^2$), based on $k$ degrees of freedom.

2. Calculate $Z$ from (1).

3. Reject $H_0$ given in Section 1 at significance level $\alpha$ if $Z > Z_\alpha$, where $Z_\alpha$ is given in Table 1, for appropriate values $n$, $k$, and $\alpha$.

3'. On occasion, $H_0$ might be rejected at level $1 - \alpha$ if $Z$ is smaller than $Z_\alpha$.

Table 1 has been constructed, for values $Z_\alpha > 0.5$, from an exact formula for upper tail probabilities, given by Cochran (1941) and used by Eisenhart and Solomon (1947); for critical values $Z_\alpha < 0.5$ Pearson curve approximations have been used. The techniques
used, and comments on the accuracy of the tables, are given in Section 3.

3. Theory of the Tests

3.1. Calculation of critical points $Z_\alpha > 0.5$. Define $r_j = s_j^2 / \sum s_i^2$. Cochran (1941) showed that the probability $P(Z > g)$ is given by

$$P(g) = nP_1(g) - \frac{n(n-1)}{2} P_2(g) + \frac{n(n-1)(n-2)}{3!} P_3(g) \cdots$$  \hspace{1cm} (2)

where $P_1(g)$ is the probability that any one ratio $r_j$ exceeds $g$, $P_2(g)$ is the probability that two of the ratios both exceed $g$, etc., and observed that the upper tail probabilities $P(g)$ will be exactly $nP_1(g)$ when $g$ exceeds 0.5. Eisenhart and Solomon (1947) showed limits for the accuracy of approximating $P(g)$ by $nP_1(g)$ for lower values of $g$. $P_1(g)$ is given by an incomplete Beta function (Cochran, 1941):

$$P_1(g) = \frac{\int_g^1 X^{k/2-1}(1-x)^{(k-1)/2} -1 \, dx}{B(k/2, (k-1)/2)} \quad 0 \leq g \leq 1$$  \hspace{1cm} (3)

where $B(\cdot, \cdot)$ is the Beta function. $P_1(g)$ can also be evaluated from tables of the $F$ distribution. In Table 1, $P_1(Z_\alpha)$ has been used to give critical values $Z_\alpha$, when these are greater than 0.5.

3.2. Pearson curve approximations for $Z_\alpha < 0.5$. For smaller values of $Z_\alpha$ corresponding to higher significance levels, we have approximated the distribution of $Z$ by Pearson curves. For this, the first four moments of $Z^{1/2}$ are used.

Suppose $Z$ is constructed as follows:

(a) Let $y_1, y_2, \ldots, y_n$ be i.i.d. random variables, each with the distribution $\sigma^2 \chi^2_k$, where $\sigma^2$ is any positive value; let $y_{(1)} < y_{(2)} < \cdots < y_{(n)}$ be the order statistics of the set $y_i$.

(b) Let $Y = \Sigma_j y_j$.

(c) Then $Z = y_{(n)} / Y$.

It is clear that the distribution of $Z$ is independent of $\sigma$, the scale parameter of $y_i$; also $Y$ is a completely sufficient statistic for $\sigma^2$. Thus, by the Basu/Hogg/Craig Theorem,
$Z$ and $Y$ are independently distributed. We can henceforth assume that $\sigma = 1$. Then $ZY = y_{(n)}$, and we have

$$E(Z^r) = \frac{E(y_{(n)}^r)}{E(Y^r)},$$

where $E(\cdot)$ denotes expectation. The denominator of (5) is easy to find, since $Y$ is a $\chi^2$-variable with $kn$ degrees of freedom: then

$$E(Y^r) = \frac{2^r \Gamma\{(kn + 2r)/2\}}{\Gamma(kn/2)}$$

(5)

For the distribution of $y_{(n)}$ suppose $G(t)$ is the distribution of $\chi^2_k$; the distribution of $y_{(n)}$ is then $[G(t)]^n = P(y_{(n)} < t)$, and moments are given by

$$E(y_{(n)}^r) = \int_0^\infty t^r n(G(t))^{n-1} g(t) dt$$

(6)

where $g(t)$ is the density of $\chi^2_k$.

Thus the moments of $Z$ or of $Z^{1/2}$ are very easy to calculate, from (4) using (5) and (6). The first four moments of $Z^{1/2}$ have been found and used to fit Pearson curves (see Solomon and Stephens, 1978) to the distribution of $Z^{1/2}$ and hence to obtain significance points $Z_\alpha$ for $Z$.

3.3. The case when $k = 1$.

For $k = 1$, when each sample variance has only one degree of freedom, there is an interesting connection with a distribution in the statistics of directions. Suppose $P$ is a point uniformly distributed on the $n$-sphere with center 0 and radius 1. It is well-known (Marsaglia, 1972) that a method to generate such points $P$ is as follows. Generate $w_1, w_2, \ldots, w_n$ i.i.d. from $N(0,1)$, and calculate $X_i = w_i/Y^{1/2}$ where $Y = \sum_j w_j^2$; then $X_i, i = 1, \ldots, n$ are the components of vector $OP$. Let $S_i = |X_i|$; then $Z = S_{(n)}^2$, that is, $Z$, when $k = 1$, is the square of the largest component of a random unit vector on an $n$-sphere. We can then get some distributional results for $Z$ by finding the distribution of $S_{(n)}$.

$k = 1, n = 2$. For example, when $n = 2$, $OP$ can be fixed by the angle $\theta$ it makes with the $x$-axis, and $\theta$ is uniformly distributed, between 0 and $2\pi$, written $U(0, 2\pi)$. For the distribution of $S_{(n)}$ we can confine attention to the first quadrant, $0 < \theta < \frac{\pi}{2}$; then when $0 < \theta < \frac{\pi}{4}$, $S_{(n)} = \cos \theta$ and when $\frac{\pi}{4} \leq \theta \leq \frac{\pi}{2}$, $S_{(n)} = \sin \theta$. It is easily shown that the
density \( f_S(x) \) of \( S(n) \) is given by

\[
f_S(x) = \frac{4}{\pi \sqrt{1 - x^2}}, \quad \frac{1}{\sqrt{2}} \leq x \leq 1.
\]  

(7)

The moments \( \mu'_r = E(S'_n) \) are \( \int_c^1 x^r f_S(x)dx \), where \( c = 1/\sqrt{2} \); these can be easily calculated to give

\[
\mu'_1 = 0.90032, \quad \mu'_2 = 0.81831, \quad \mu'_3 = 0.75026, \quad \mu'_4 = 0.69331.
\]

Check with formula (2). From (2) and (3), with \( k = 1, n = 2 \), we have for \( g > 0.5 \),

\[
P(g) = \frac{2}{\pi} \int_g^1 x^{-1/2}(1 - x)^{-1/2}dx,
\]

since \( B(\frac{1}{2}, \frac{1}{2}) = \pi \). Let \( x = t^2 \) and we have

\[
P(Z > g) = P(S_{(2)} > \sqrt{g})
\]

\[
= \frac{2}{\pi} \int_0^2 2(1 - t^2)^{-1/2}dt = 2 - \frac{4}{\pi} \sin^{-1} \sqrt{g}, \quad g \geq 0.5.
\]

This is the same result as obtained by integrating (7).

Of course, for \( n = 2 \) there is a simple correspondence with the \( F \)-test for equality of two variances, since \( Z^{-1} = 1 + s^2_{(1)}/s^2_{(2)} \). It quickly follows that \( z_\alpha \), the critical value of \( Z \) at level \( \alpha \), is related to \( F_{k,k}(\alpha/2) \).

For \( n = 3 \), the algebra is more complicated. Let vector \( OP \), where \( P \) is now uniformly distributed on the first orthant of the unit sphere (that is, all coordinates of \( OP \) are positive), have usual spherical coordinates \( \theta, \phi \). Then \( z = \cos \theta \) is \( U(0,1) \), and \( \phi = U(0, \frac{\pi}{2}) \), so that, if we use rectangular axes for \( (\phi, z) \), probability is uniform on the rectangle \( R : 0 < z < 1, 0 < \phi < \frac{\pi}{2} \). We now want \( \Pr(\text{maximum component of } OP < t) = P(t) \). This is found as follows.

\( k = 1, n = 3, t \geq \frac{1}{\sqrt{2}} \). If \( z > t \), and \( t \geq \frac{1}{\sqrt{2}} \), it is clear that \( z \) must be the maximum component, for all \( \phi \). Thus \( P(z > t) = 1 - t \). By symmetry, \( x = \sin \theta \cos \phi \) or \( y = \sin \theta \sin \phi \) (the other two components of \( OP \)) could be the maximum component with equal probability, so that \( P(\text{maximum component} > t) = 3(1 - t) \).

Equations (2) and (3), for \( n = 3, k = 1, g > 0.5 \), give

\[
P(Z > g) = 3 \int_g^1 x^{-1/2}dx / B(\frac{1}{2}, 1) = 3(1 - \sqrt{g});
\]
thus \( P(S_{(3)} > \sqrt{g}) = 3(1 - \sqrt{g}) \), in agreement with the result above.

\[ k = 1, \; n = 3, \; \frac{1}{\sqrt{3}} \leq t \leq \frac{1}{\sqrt{2}}. \]

Then \( P(S_{(3)} < t) = P(\text{all 3 components} < t) \); this is given by the probability over an area bounded by (1) \( z = t \); (2) \( \sqrt{1 - z^2} \cos \phi = t \); (3) \( \sqrt{1 - z^2} \sin \phi = t \), roughly in the middle of rectangle \( R \). Thus

\[
P(S_{(3)} < t) = \frac{4}{\pi} \int_{\phi_1}^{\frac{\pi}{4}} (t - z) d\phi
\]

where \( \cos \phi_1 = t/\sqrt{1 - t^2} \) and where \( \sqrt{(1 - z^2)} \cos \phi = t \); so

\[
P(S_{(3)} < t) = \frac{4}{\pi} \int_{\phi_1}^{\frac{\pi}{4}} \{t - \sqrt{1 - t^2} \sin^2 \phi\} d\phi, \quad \frac{1}{\sqrt{3}} \leq t \leq \frac{1}{\sqrt{2}}.
\]

This last expression must be evaluated numerically.

Similar ideas can be used to give upper tail results for higher values of \( n \), for \( k = 1 \), but again they lead in the end to integrals which must be evaluated numerically.

3.4. Accuracy of Table 1.

Various checks on accuracy have been made for the points in Table 1. In fitting the Pearson curves, the fit was made to \( Z^{1/2} \), which, when \( k = 1 \), is \( S_{(n)} \). The numerator of \( S_{(n)} \) is the largest absolute value of a set of \( n \) standard normals, and the expectation of this variable is known (Biometrika Tables for Statisticians, Vol. 2), since absolute values of standard normals are used in analysis of experiments when main effects and interactions are plotted on half-normal plots (see, for example, Bennett and Franklin, 1954). These known values enabled a check to be made on the accuracy of (6), for \( k = 1 \) and \( r = \frac{1}{2} \). Also, for \( k = 1, \; n = 2 \), the moments of \( S_{(n)} \) could be compared with the exact values given above.

In addition, the exact distributions given above have been used, for \( n = 2 \) and \( n = 3 \), to check the percentage points. A further check was made, for \( k = 2 \), by comparing the values with those given by Nowroozi (1967). Finally, the most extensive check on the Pearson curve points was made by comparing the points \( Z_\alpha \) with those given by the exact formula (2), when \( Z_\alpha > 0.5 \). It was found that the Pearson curve fits were very accurate, with occasional differences from the exact values in the fourth, or sometimes the third, decimal place: however, these differences will make negligible error in the \( \alpha \)-value corresponding to the given point \( Z_\alpha \).
References


Percentage Points For The Fisher-Cochran Test For Equality Of Variances

Herbert Solomon and Michael A. Stephens

Department of Statistics
Stanford University
Stanford, CA 94305

Office of Naval Research
Statistics & Probability Program Code 1111

August 19, 1988

19

tests for equality of variances; Cochran's statistic; tables for tests of variances

PLEASE SEE FOLLOWING PAGE.
20. ABSTRACT

A well known test for equality of normal population variances, based on sample variances, was introduced by Cochran (1941). Suppose \( \sigma_i^2, i=1,\ldots,n \) represents the population variances of \( n \) normal populations, and let independent sample variances, each based on \( k \) degrees of freedom, be \( s_i^2, i=1,\ldots,n \). Suppose the \( s_i^2 \) are ranked, so that the ordered variables are \( s_1^2, s_2^2, \ldots, s_n^2 \). To test \( H_0 : \sigma_i^2 \) are all equal (suppose the common value is \( \sigma^2 \)), Cochran (1941) introduced the test statistic

\[
Z = \frac{s(n)^2}{\sum_{i=1}^{n} s_i^2}
\]

which compares the largest sample variance with the sum of the sample variances. Clearly the intent is to discover if one variance is an outlier (too large), and, in general, \( H_0 \) will be rejected for large values of \( Z \). Tables of various percentiles are given for various values of \( n \) and \( k \).