Journal of Mathematical Extension
Vol. 7, No. 2, (2013), 83-91

# Inference on the Ratio of Variances of Two Independent Populations

**M. R. Mahmoudi***
Shiraz University

**M. Mahmoudi**
Shiraz University

**Abstract.** The asymptotic distribution for the ratio of sample variances in two independent populations is established. The presented method can be used to derive the asymptotic confidence interval and hypothesis testing for the ratio of population variances. The performance of the new interval is comparable with similar confidence intervals in the large sample cases. Then the simulation study is provided to compare our confidence interval with F-statistic method. The proposed confidence set has a good coverage probability witha shorter length.

**AMS Subject Classification:** 62F03; 62F05; 62F12; 62H20
**Keywords and Phrases:** Cramer's theorem, ratio of variances, Slutsky's theorem

## 1. Introduction

It is of interest to make inference about the ratio of variances of two independent populations. The classical tests for comparing variances of k variables were designed for independent variables. These are Bartlett's test, Cochran's test and Hartley's test, all of which were found to be sensitive to departures from normality (Winer, 1971). As an alternative, Box (1953) suggested a test which is fairly robust with respect to departures from normality. This test, also discussed by Scheffe (1959), is

Received October 2012; Accepted: February 2013
*Corresponding author

based on an analysis of variance of the logarithms of the variances. When the variance of a group of items before a treatment is to be compared with the variance of the same group after treatment, the independence is quite apparently not met. As an example, Cochran considered a characteristic was measured both before and after aging. It was of interest to determine whether the variability is changed by aging. When the variables are known to be correlated, Pitman (1939) and Morgan (1939) considered the comparison of variances for bivariate normal variables. The comparison of variances for k $>$ 2 jointly normal correlated variables was discussed by Han (1968, 1969). Choi and Wette (1972) suggested a method for testing the equality of variances of several correlated normal variables when the covariances are unknown. Levy (1976) suggested a test for $k > 2$ jointly normal correlated variables without any restrictive assumptions on the correlation structure. Cohen (1986) considered the comparison of variances for correlated variables. Cacoullos (2001), using the F-representation of t, showed the Pitman-Morgan t-test for homoscedasticity under a bivariate normal setup is equivalent to an F-test on n-2 and n-2 degrees of freedom. This yields an F-test of independence under normality. Brownie and Boos (2004) reviewed the difference between asymptotic properties of normal-theory tests for variances and for means and described several specific methods for the two and k - samples problems. Ojbasic and Tomovich (2007) suggested confidence intervals for the population variance and the difference invariances of two populations based on the ordinary t-statistics combined with the bootstrap method. Guajardo and Lubiano (2012), on the basis of Levene's classical procedure, developed a test for the equality of variances of k fuzzy-valued random elements.Bhandary and Dai (2013) considered the problem of homogeneity of variance in Randomized Complete Block Design (RCBD) and developed a new test for the equality of variances in RCBD.

In this work, the asymptotic distribution for the ratio of sample variances is presented. It will be applied to construct confidence interval and perform test statistics. This method is the most efficient way in comparison with other method, specially when sample size is large.

## 2. Large Samples Inference

Let $X$ and $Y$ be two random variables with means $\mu_X$ and $\mu_Y$, variances $\sigma_X^2$ and $\sigma_Y^2$, Third moments $\mu_{3X}$ and $\mu_{3Y}$, and finite fourth moments $\mu_{4X}$ and $\mu_{4Y}$, respectively. Also assume that $(X, Y)$ has the finite central product moments $\mu_{ij} = E\left[(X-\mu_X)^i (Y-\mu_Y)^j\right]; (i,j) \in \{(1,1)\,,\ (1,3)\,,\ (3,1)\,,\ (2,2)\}$. We are interested to inference about the parameter $\sigma = \frac{\sigma_X^2}{\sigma_Y^2}$. Since $S_X^2 = \overline{X^2} - \overline{X}^2$ and $S_Y^2 = \overline{Y^2} - \overline{Y}^2$ are consistent estimators for $\sigma_X^2$ and $\sigma_Y^2$, $S = \frac{S_X^2}{S_Y^2}$ seems to be a reasonable estimator for the parameter $\sigma$. There is no loss in assuming $m = n$. Also, since the variance doesn't depend on locations, we may as well assume $\mu_X = \mu_Y = 0$ (otherwise work with $X - \mu_X$ and $Y - \mu_Y$).
In the following theorem that is the main theorem of this article, we will give the asymptotic distribution of $S$.

**Theorem 2.1.** *Under the above assumptions,*

$$\sqrt{n}\,(S-\sigma) \xrightarrow{\mathcal{L}} N\left(0, \gamma^2\right) \quad \text{as } n \to \infty,$$

*where*

$$\gamma^2 = \frac{\mu_{4Y}\sigma^2 - 2\mu_{22}\sigma + \mu_{4X}}{\sigma_Y^4}.$$

**Proof.** Define $\mathbf{M}_n = (\overline{X},\ \overline{Y},\ \overline{X^2},\ \overline{Y^2})^T$.Then by central limit theorem,

$$\sqrt{n}\,(\mathbf{M}_n - \mu) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Sigma),$$

where $\mu = (0,\ 0,\ \sigma_X^2,\ \sigma_Y^2)^T$ and

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \mu_{11} & \sigma_{3X} & \sigma_{12} \\ & \mu_Y^2 & \mu_{21} & \mu_{3Y} \\ & & \sigma_{4X} - \mu_X^4 & \sigma_{22} - \sigma_X^2\sigma_Y^2 \\ & & & \sigma_{4Y} - \mu_Y^4 \end{bmatrix}$$

Now, we apply Cramer's theorem(see Ferguson(1996)) for function

$$g\,(\mathbf{M}_n) = (\overline{X^2} - \overline{X}^2)/(\overline{Y^2} - \overline{Y}^2).$$

We define $g : \mathbb{R}^4 \to \mathbb{R}$ as $g\left(x_1, x_2, x_3, x_4\right) = \frac{x_3 - {x_1}^2}{x_4 - {x_2}^2}$ .
Then the gradient function with respect to $g$ is

$$\nabla g\left(x_1, x_2, x_3, x_4\right) = \left( \frac{-2x_1}{x_4 - {x_2}^2}, \frac{2x_2\left(x_3 - {x_1}^2\right)}{\left(x_4 - {x_2}^2\right)^2}, \frac{1}{x_4 - {x_2}^2}, \frac{{x_1}^2 - x_3}{\left(x_4 - {x_2}^2\right)^2} \right).$$

Also $\nabla g\left(0, 0, \sigma_X^2, \sigma_Y^2\right) \acute{O} \left(\nabla g\left(0, 0, \sigma_X^2, \sigma_Y^2\right)\right)^T = \gamma^2$.
Since $\nabla g$ is continuous in neighborhood of $\left(0, 0, \sigma_X^2, \sigma_Y^2\right)$ , therefore, by Cramer's rule we have

$$\sqrt{n}\left(g\left(\overline{X},\ \overline{Y},\ \overline{X^2},\ \overline{Y^2}\right) - g\left(0, 0, \sigma_X^2, \sigma_Y^2\right)\right)$$

$$= \sqrt{n}\left(S - \sigma\right) \xrightarrow{\mathcal{L}} N\left(0, \gamma^2\right) \quad as\ n \to \infty. \quad \square$$

**Corollary 2.2.** *If $X$ and $Y$ are independent, then*

$$\gamma^2 = \left(\frac{\mu_{4Y}}{\sigma_Y^4} - 2\right)\sigma^2 + \frac{\mu_{4X}}{\sigma_Y^4}.$$

*Furthermore, by normality assumption, we have*

$$\gamma^2 = 4\sigma^2.$$

**Proof.** Note that for independent populations $\mu_{22} = \sigma_X^2 \sigma_Y^2$ and for normal populations, $\mu_{4X} = 3\sigma_X^4$ and $\mu_{4Y} = 3\sigma_Y^4$. $\square$

By the theorem we have just proved

$$T_n = \sqrt{n}\left(\frac{S - \sigma}{\gamma}\right) \xrightarrow{\mathcal{L}} N\left(0, 1\right) \quad as\ n \to \infty \tag{1}$$

This result can be used to construct asymptotic confidence interval and hypothesis testing.

## 2.1 Asymptotic Confidence Interval

Since the parameter $\gamma$ in $T_n$ depends on the unknown parameter $\sigma$, it can not be used as a pivotal quantity for the parameter $\sigma$.

**Theorem 2.3.** *If $X$ and $Y$ are independent, then*

$$T_n^* = \sqrt{n}\left(\frac{S-\sigma}{\widehat{\gamma}_n}\right) \xrightarrow{\mathcal{L}} N(0,1) \quad as\ n \to \infty, \tag{2}$$

*where*

$$\widehat{\gamma}_n = \left(\frac{m_{4Y}}{S_Y^4} - 2\right) S^2 + \frac{m_{4X}}{S_Y^4} \quad ,$$

*and $m_{4X}$ and $m_{4Y}$ are the fourth sample moments of $X$ and $Y$, respectively.*

**Proof.** By the weak law of large numbers,

$$S_X^2 \xrightarrow{P} \sigma_X^2 \quad , \quad S_Y^2 \xrightarrow{P} \sigma_Y^2 \quad , \qquad m_{4X} \xrightarrow{P} \mu_{4X} \quad , \qquad m_{4Y} \xrightarrow{P} \mu_{4Y} \quad , \ as\ n \to \infty.$$

From this and the Slutsky'stheorem we have,

$$\begin{pmatrix} S_X^2 \\ S_Y^2 \\ m_{4X} \\ m_{4Y} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \mu_{4X} \\ \mu_{4Y} \end{pmatrix} \quad as\ n \to \infty.$$

By Slutsky'stheorem $\widehat{\gamma}_n \xrightarrow{P} \gamma$, $as\ n \to \infty$. The proof is completed by using Theorem 2.1. □

**Corollary 2.4.** *By normality assumption, we have*

$$\widehat{\gamma}_n = 2S.$$

*Now, $T_n^*$ can be used as a pivotal quantity to construct asymptotic confidence interval for $\sigma$,*

$$\left(S - \frac{\widehat{\gamma}_n}{\sqrt{n}} Z_{\alpha/2}, S + \frac{\widehat{\gamma}_n}{\sqrt{n}} Z_{\alpha/2}\right). \tag{3}$$

**Hypothesis Testing**

Hypothesis testing about $\sigma$ is important in practice. For instance, the assumption $\sigma = 1$ is equivalent to the assumption $\sigma_X^2 = \sigma_Y^2$. In general, to test $H_0 :\ \sigma = \sigma_0$, the test statistic can be

$$T_0 = \sqrt{n}\left(\frac{S-\sigma_0}{\sigma^*}\right), \tag{4}$$

where

$$\sigma^* = \sqrt{\frac{m_{4Y}\sigma_0^2 - 2\sigma_0 m_{22} - m_{4X}}{S_Y^4}}.$$

By similar methodology applied in Theorem 2.3, it can be shown that under null hypothesis, $T_0$ has asymptotic standard normal distribution. Note that, in the case $n \neq m$, it is sufficient to replace n by $n^* = \min(m, n)$ in the above results.

**Remark 2.5.** *By normality assumption,*

$$T_n^* = \sqrt{n}(\frac{S-\sigma}{2S}) \xrightarrow{\mathcal{L}} N(0,1), \quad \text{as } n \to \infty. \tag{5}$$

*This result can be used to construct asymptotic confidence interval and hypothesis testing for the parameter $\sigma$, in two independent normal populations, i.e.,*

$$S \pm \frac{2S}{\sqrt{n}} Z_{\alpha/2} \tag{6}$$

*Also, to test $H_0 : \sigma = \sigma_0$, in two independent normal populations, we can use the test statistic*

$$T_0 = \frac{S - \sigma_0}{2\sigma_0}, \tag{7}$$

*which has asymptotic standard normal distribution.*
*In the case $\sigma_0 = 1$, which is equivalent to $\sigma_X^2 = \sigma_Y^2$, we can also use Fisher statistic as follows:*

$$F = \frac{S_X^2}{s_Y^2},$$

*which has the exact distribution F with degrees of freedom m-1 and n-1, i.e., F(m-1,n-1).*

## 3. Simulation

In this section, we provide the simulation study to compare our confidence interval *(CI-1)* with confidence interval based on $F$-statistic *(CI-2)*, i.e.,

$$\left(\frac{S_X^2}{s_Y^2} F_{1-\alpha/2}(n-1,\ m-1),\ \frac{S_X^2}{s_Y^2} F_{\alpha/2}(n-1,\ m-1)\right),$$

**Table 1:** The empirical probability coverage and the length of the intervals for normal populations

| | | $\sigma_X^2 = 8$ ,$\sigma_Y^2 = 32$ | | $\sigma_X^2 = 8$ ,$\sigma_Y^2 = 10$ | | $\sigma_X^2 = 10$ ,$\sigma_Y^2 = 8$ | | $\sigma_X^2 = 32$ ,$\sigma_Y^2 = 8$ | |
|---|---|---|---|---|---|---|---|---|---|
| **n** | **CI** | **coverage** | **length** | **coverage** | **length** | **coverage** | **length** | **coverage** | **length** |
| **50** | **CI-1** | 0.9536 | 0.2469 | 0.9536 | 0.7912 | 0.9511 | 1.2356 | 0.9487 | 3.9521 |
| | **CI-2** | 0.9512 | 0.2475 | 0.9509 | 0.7920 | 0.9503 | 1.2375 | 0.9509 | 3.9600 |
| **100** | **CI-1** | 0.9498 | 0.1639 | 0.9498 | 0.5309 | 0.9589 | 0.8256 | 0.9496 | 2.6613 |
| | **CI-2** | 0.9499 | 0.1675 | 0.9501 | 0.5360 | 0.9511 | 0.8375 | 0.9502 | 2.6800 |
| **200** | **CI-1** | 0.9499 | 0.1098 | 0.9503 | 0.3615 | 0.9498 | 0.5698 | 0.9501 | 1.7957 |
| | **CI-2** | 0.9507 | 0.1175 | 0.9501 | 0.3760 | 0.9504 | 0.5875 | 0.9500 | 1.8800 |
| **500** | **CI-1** | 0.9499 | 0.0523 | 0.9499 | 0.2034 | 0.9501 | 0.3256 | 0.9499 | 1.0985 |
| | **CI-2** | 0.9500 | 0.0750 | 0.9501 | 0.2400 | 0.9501 | 0.3750 | 0.9499 | 1.2000 |

**Table 2:** The empirical powers of the tests

| | | $\sigma_X^2 = 8$ | $\sigma_X^2 = 8$ | $\sigma_X^2 = 10$ | $\sigma_X^2 = 32$ | $\sigma_X^2 = 8$ |
|---|---|---|---|---|---|---|
| **n** | **Critical Region** | $\sigma_Y^2 = 32$ | $\sigma_Y^2 = 10$ | $\sigma_Y^2 = 8$ | $\sigma_Y^2 = 8$ | $\sigma_Y^2 = 32$ |
| **50** | **T-1** | 0.9930 | 0.9284 | 0.9610 | 0.9503 | 0.9518 |
| | **T-2** | 0.9102 | 0.8927 | 0.8880 | 0.8871 | 0.8990 |
| **75** | **T-1** | 0.9994 | 0.9858 | 0.9433 | 0.9902 | 0.9711 |
| | **T-2** | 0.8986 | 0.9168 | 0.8825 | 0.9156 | 0.9072 |
| **100** | **T-1** | 1 | 0.9969 | 0.9728 | 0.9980 | 0.98094 |
| | **T-2** | 0.9599 | 0.9247 | 0.9523 | 0.9248 | 0.9272 |
| **150** | **T-1** | 1 | 0.9999 | 0.9965 | 0.9999 | 0.9952 |
| | **T-2** | 1 | 0.9298 | 0.9556 | 0.9698 | 0.9335 |

in view of the empirical coverage and average length.

We simulate *50000* times of the above confidence intervals for normal populations with *m=n=50, 100, 200* and *500* for different values of $\sigma_X^2$ and $\sigma_Y^2$. The empirical coverage and mean lengths are summarized in the *Table 1.*

As can be seen, in terms of the empirical probability coverage, two methods have the same empirical probability coverage. In terms of the length of the interval, our method is better. Also, the critical region which is constructed by inverting our confidence interval *(T-1)* has more power than the critical regions corresponding to the other confidence interval *(T-2)*. This subject can be seen by a simulation study for test $H_0$: $\sigma = 1$ , the empirical powers of the tests are presented in *Table 2.*

The power of the presented method is more than the other method and it can be stated that the test statistic (5) has a reasonable power in comparison with competing method.

**Acknowledgment**
The authors are grateful for the comments and suggestions by the referees and Editor-in-Chief. Their comments and suggestions have greatly improved the paper.

## References

[1] M. Bhandary and H. Dai, An alternative test for the equality of variances for several populations in randomised complete block design. *Statistical Methodology*, 11 (2013), 22-35.

[2] G. E. P. Box, Nonnormality and tests on variances. *Biometrika*, 40 (1953), 318-335.

[3] C. Brownie and D. D. Boos, Comparing variances and other measures of dispersion. *Statistical Science*, 19 (4) (2004), 571-578.

[4] T. Cacoullos, The F-test of homoscedasticity for correlated normal variables, *Statistics and Probability Letters*, 54 (2001), 1-3.

[5] S. C. Choi and R. Wette, A test for the homogeneity of variances among variables. *Biometrics, Queries and Notes*, June 1972, 589-591.

[6] W. G. Cochran, Testig two correlated variances. *Technometrics*, 7 (3) (1965), 447-449.

[7] A. Cohen, Comparing variances of correlated variables, *Psychometrika*, 51 (3) (1986), 379-391.

[8] T. S. Ferguson, *A Course in Large Sample Theory*, Chapman and Hall, (1996).

[9] A. B. R. Guajardo and M. A. Lubiano, $K$-sample tests for equality of variances of random fuzzy sets. *Computational Statistics and Data Analysis*, 56 (4) (2012), 956-966.

[10] C. P. Han, Testing the homogeneity of a set of correlated variances, *Biometrika*, 55 (1968), 317-326.

[11] C. P. Han, Testing the homogeneity of variances in a two-way classification, *Biometrics*, 25 (1969), 153-158.

[12] K. J. Levy, A procedure for testing the equality of p-correlated variances, *British Journal of Mathematical*, (1976), 89-93.

[13] W. A. Morgan, A test for the significance of the difference between the two variances in a sample from a normal bivariate population, *Biometrika*, 31 (1939), 13-19.

[14] V. Ojbasic and A. Tomovich, Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples, *Computational Statistics and Data Analysis,* 51 (2007), 5562-5578.

[15] E. J. C. Pitman, A note on normal correlation. *Biometrika*, 31 (1939), 9-12.

[16] H. Scheffe, *The Analysis of Variance*, New York, Wiley, (1959).

[17] B. J. Winer, *Statistical Principles in Experimental Design*, New York, McGraw-Hill, (1971).

**Mohammad Reza Mahmoudi**
Department of Statistics
College of Sciences
Ph.D Student of Statistics
Shiraz University
Shiraz, Iran
E-mail: rezastat@yahoo.com

**Marzieh Mahmoudi**
Department of Statistics
College of Sciences
M.Sc of Statistics
Shiraz University
Shiraz, Iran
E-mail: marzieh_mahmoudi@yahoo.com