Statistical Inference for the Coefficient of Variation in Normally Distributed Data

Johannes Forkman

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2006:2
ISSN 1651-8543

Statistical inference for the coefficient of variation in normally distributed data

Johannes Forkman[1]

Centre of Biostochastics
Swedish University of Agricultural Sciences
Box 7013, SE-750 07 Uppsala, Sweden

Abstract

Methods for analysing coefficients of variation in normally distributed data are studied. An approximate F-test for equality of two coefficients of variation is introduced. The approximate F-test is compared with eight other tests in a simulation study. The new test performs well, also when the sample sizes are small. A generalized version of the approximate F-test is defined for the case that there are several independent estimates of each coefficient of variation, calculated with different averages. The test is applied to a real immunoassay dataset from diagnostic research. All moments of the proposed test statistic are shown to be approximately equal to the moments of an F-distribution. The distribution of the logarithm of the test statistic equals the distribution of the logarithm of an F- distribution plus some error variables that are in probability of small orders.

Key words: coefficient of variation, normal distribution, McKay's approximation, approximate F-test

[1]E-mail address to the correspondence author: Johannes.Forkman@bt.slu.se

1 Introduction

In statistical theory the second moment of the mean is a convenient measure of dispersion. When the observations $y_1, y_2, ..., y_n$ are normally distributed it is usually estimated by the sample variance s^2, defined by

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - m)^2, \quad m = \frac{1}{n}\sum_{j=1}^{n} y_j. \tag{1}$$

The standard deviation s is a measure of the variability in the original scale. In many problems it is, however, necessary to go one step further and relate the variation to the level of the observations. If the standard deviation is *e.g.* 25 it may possibly be small if the average is 1000, but large if the average is 100. For this reason variability in data is often summarised by the coefficient of variation

$$c = \frac{s}{m}. \tag{2}$$

The coefficient of variation is a ratio between two outcomes of random variables. Theoretically this measure of dispersion is not as convenient as the variance, but from a practical point of view it provides useful information. In many fields of interest, often in biological and medical research, the coefficient of variation is preferred to the variance or standard deviation.

Pearson (1896) defined the coefficient of variation and used it for comparison of various measurements on females with corresponding measurements on males. Schimmerl-Metz et al. (1999) provide a modern example from morphology. They calculate coefficients of variation on measurements of the scapholunate joint intercortical width of wrists.

In laboratory analytical procedures the standard deviation of repeated measurements are often proportional to the concentration being measured. The precision of an analytical method is usually described by coefficients of variation between and within assays. DeSilva et al. (2003) accordingly recommend that precision shall be expressed by coefficients of variation. Comparing the performance of *e.g.* two laboratories or two instruments thus involves the problem of comparing two coefficients of variation.

In clinical trials not only the average effect of a treatment but also the variation in the effect is considered. It is not always appropriate to assume independence between effect size and variance. Often data indicate a constant coefficient of variation. In crossover trials treatments are compared within individuals. An individual is first given one treatment, and then a second

treatment and so on. Sometimes each individual receives each treatment several times. The individuals may respond very differently on the treatments, and the standard deviation in the replicated measurements is often proportional to the response. In this case the coefficient of variation is a natural measure of dispersion. The Food and Drug Administration (2001) establish that coefficients of variation shall be reported in bioequivalence studies.

The reaction time of a task may differ much between a group of patients and a control group. The coefficients of variation may, however, be similar or equal in the two groups (Schafer and Sullivan, 1986).

Despite the large number of applications the properties of the coefficient of variation are seldom discussed in statistical textbooks. As a consequence there is among practitioners often an inadequate knowledge on how to make proper statistical inference concerning the measure. We shall in this article bring some light on the subject in general and particularly discuss statistical tests for the coefficient of variation when the observations are normally distributed. Many approximate tests have been suggested for the hypothesis that two coefficients of variation are equal. They are however not well known, and maybe for this reason coefficients of variation are often reported in scientific work without any use of statistical methods. The present article compares eight proposed methods with a new approximate F-test in a simulation study. This study should help the researcher to choose a test for the comparison of two coefficients of variation.

We shall see that all moments of the approximate F-test statistic are close to the moments of an F-distributed random variable. The distribution of the logarithm of the test statistic approximately equals the distribution of the logarithm of an F-distributed random variable. The test is easily generalized to the case that there are several estimates per coefficient of variation. We shall show by a real data example from diagnostic research how the approximate F-test and its generalized version can be applied. We shall also suggest a method for the problem considered by Tian (2005) of making inference for an *a priori* common coefficient of variation.

2 Inference on a single coefficient of variation

Let $y_j = \mu + e_j$, where e_j are independently distributed $\mathrm{N}(0, \gamma^2\mu^2), j = 1, 2, \ldots, n$, with positive population coefficient of variation γ and positive expected value μ. Let m denote the average as defined in (1), c the sample coefficient of variation as defined in (2).

In the well-known t-test of the hypothesis that the expected value of a normally distributed random variable equals zero, the test statistic

$$t = \frac{m}{s/\sqrt{n}} = \frac{\sqrt{n}}{c}$$

is t distributed with $n-1$ degrees of freedom under assumption that the hypothesis is true. Generally t follows a noncentral t distribution with $n-1$ degrees of freedom and noncentrality parameter $\tau = \sqrt{n}/\gamma$. Owen (1968) discusses this and other applications of the noncentral t distribution. With modern statistical software the percentiles of the noncentral t distribution are accessible and it is easy to test the hypothesis that τ, and thus also γ, equals a specified value. A confidence set for τ can be constructed by inverting the acceptance region of a test of the hypothesis about τ (Shao, 2003). Thus, if $\Pr(t < \sqrt{n}/c \mid \tau = \tau_1) = \alpha/2$ and $\Pr(t > \sqrt{n}/c \mid \tau = \tau_2) = \alpha/2$ then $[\tau_2, \tau_1]$ is a $100(1-\alpha)\%$ confidence interval for τ. An exact finite confidence interval for γ is easily obtained from the confidence interval for τ provided that the latter does not include zero, which it seldom does. The exact finite confidence interval is $[\sqrt{n}/\tau_1, \sqrt{n}/\tau_2]$.

If the percentiles of the noncentral t distribution are not available there are several ways to calculate approximate confidence intervals. McKay (1932) shows that if γ is small, i.e., less than $1/3$, and if $\theta = (n-1)/n$, then

$$(n-1)\frac{c^2/(1+\theta c^2)}{\gamma^2/(1+\gamma^2)} \tag{3}$$

is approximately χ^2 distributed with $n-1$ degrees of freedom. Note that in applications the condition $\gamma < 1/3$ is often reasonable since it makes negative observations unlikely. The condition is fulfilled when the observations are necessarily positive though well described by the normal distribution. This is often the case when measuring, e.g., length, mass, time, blood pressure or concentration. Fieller (1932), Pearson (1932), Iglewicz and Myers (1970) and Umphrey (1983) all confirm the adequacy of McKay's χ^2 approximation. Since (3) is an approximate pivotal quantity (Shao, 2003) it can be used as an approximate test for the hypothesis that $\gamma = \gamma_0$, where γ_0 is a specified constant, or for calculating an approximate confidence interval.

Vangel (1996) proposes a small modification of McKay's approximation useful for calculating approximate confidence intervals that are accurate also for small sample sizes. The confidence interval based on this approximation can be written

$$\left[c\left(\frac{u_1}{n-1} + c^2\left(\frac{u_1+2}{n} - 1\right)\right)^{-\frac{1}{2}},\ c\left(\frac{u_2}{n-1} + c^2\left(\frac{u_2+2}{n} - 1\right)\right)^{-\frac{1}{2}} \right], \quad (4)$$

where u_1 denote the $100(1-\alpha/2)$:th percentile of a χ^2 distribution with $n-1$ degrees of freedom, and where u_2 denote the $100(\alpha/2)$:th percentile of a χ^2 distribution with $n-1$ degrees of freedom. Another approximate method, developed by Wong and Wu (2002), for calculating confidence intervals is based on the modified signed log likelihood ratio statistic defined by Barndorff-Nielsen (1986, 1991). This method is also claimed to give accurate results in case of small sample sizes.

There is a strong tradition among statisticians to use the logarithmic transformation when the standard deviation is proportional to the mean. A Taylor series expansion of $\log y$ about $y = \mu$ gives

$$\log y \approx \log \mu + \frac{1}{\mu}(y - \mu),$$

so that $\mathrm{Var}(\log y) \approx \mathrm{Var}(y)/\mu^2$. Thus the standard deviation in log scale roughly equals the coefficient of variation in the original scale. In terms of changes in μ the logarithmic transformation is variance stabilising when the coefficient of variation in the original scale is constant. After having transformed all data into log values the statistical analyst often proceeds by modelling an expected value under assumption of a normally distributed error term. This additive error is normally distributed in log scale. In the original scale the error is multiplicative with a lognormal distribution. The lognormal distribution is however not symmetric, but positively skewed. The final analysis does for this reason not conform to an initial assumption of a symmetric distribution with approximately normally distributed errors. In *e.g.* blood test systems the measurement errors, detected by measuring the same blood sample repeatedly, are often approximately normally distributed as a result of approximately normally distributed error sources such as variation in pipetted volume.

3 Review of tests for equality of two coefficients of variation

Various test statistics have been proposed for the hypothesis that two coefficients of variation are equal. The most well known are collected in this section. We investigate their performances in Section 6.

Let $y_{ij} = \mu_i + e_{ij}$, where e_{ij} are independently distributed $\mathrm{N}(0, \gamma_i^2\mu_i^2), i = 1, 2$ and $j = 1, 2, \ldots, n_i$, with positive population coefficients of variation γ_i and positive expected values μ_i. Let c_i and m_i denote the sample coefficient of variation and the average in sample i, respectively. We study tests of the null hypothesis $H_0 : \gamma_1 = \gamma_2$ of equal population coefficients of variation.

3.1 Likelihood ratio test

Several authors explore the likelihood ratio test of the hypothesis. Miller and Karson (1977) and Bhoj and Ahsanullah (1993) deal with the special case of equal sample sizes. Lohrding (1975), Bennett (1977) and Doornbos and Dijkstra (1983) consider the general case of unequal sample sizes. According to Gerig and Sen (1980), the maximum likelihood estimates of μ_1, μ_2 and γ are

$$\hat{\mu}_1 = \frac{n_1 m_1 \hat{\mu}_2}{(n_1 + n_2)\hat{\mu}_2 - n_2 m_2}, \quad \hat{\mu}_2 = -\frac{q}{2p} + \sqrt{\frac{q^2}{4p^2} - \frac{r}{p}}$$

and

$$\hat{\gamma} = \sqrt{\frac{1}{\hat{\mu}_2}\left(\frac{n_2 - 1}{n_2} c_2^2 m_2^2 + m_2^2 - m_2\hat{\mu}_2\right)} \tag{5}$$

respectively, where $p = (n_1 + n_2)c_1^2 + n_2$, $q = -(2n_2c_1^2 + 2n_2 - n_1)m_2$ and $r = (n_2^2(c_1^2 + 1) - n_1^2(c_2^2 + 1))m_2^2/(n_1 + n_2)$. The likelihood ratio test statistic can be written

$$R = -2\log\lambda = n_1 \log\frac{n_1(\hat{\gamma}\hat{\mu}_1)^2}{(n_1 - 1)c_1^2 m_1^2} + n_2 \log\frac{n_2(\hat{\gamma}\hat{\mu}_2)^2}{(n_2 - 1)c_2^2 m_2^2}, \tag{6}$$

where λ is the likelihood ratio. Asymptotically R is χ^2 distributed with 1 degree of freedom.

3.2 Bennett's test

Bennett (1976) utilise McKay's approximation (3) and applies a test according to Pitman (1939) of the hypothesis of equal scale parameters of gamma variables. Shafer and Sullivan (1986) note that Bennett by mistake uses a variance with devisor $n - 1$ where McKay (1932) uses a variance with devisor n. For this reason they modify Bennett's test correspondingly. The modified

Bennett's test statistic is

$$\begin{aligned} B &= (n_1+n_2-2)\log\left(\frac{1}{n_1+n_2-2}\left(\frac{n_1\theta_1 c_1^2}{1+\theta_1 c_1^2}+\frac{n_2\theta_2 c_2^2}{1+\theta_2 c_2^2}\right)\right) \\ &\quad -(n_1-1)\log\left(\frac{n_1\theta_1 c_1^2}{(n_1-1)(1+\theta_1 c_1^2)}\right) \\ &\quad -(n_2-1)\log\left(\frac{n_2\theta_2 c_2^2}{(n_2-1)(1+\theta_2 c_2^2)}\right) \end{aligned} \tag{7}$$

where $\theta_i = (n_i-1)/n_i, i = 1, 2$. The value of the test statistic shall be compared with a χ^2 distribution with 1 degree of freedom.

3.3 Miller's test

When there are many observations, the sample coefficient of variation has an approximate normal distribution. Miller (1991) gives a test based on this asymptotic normality. The population coefficient of variation γ is estimated by a weighted average, $\gamma_W = ((n_1-1)c_1 + (n_2-1)c_2)/(n_1+n_2-2)$. This estimate is employed in the calculation of a test statistic

$$M = \frac{c_1 - c_2}{\sqrt{\frac{\gamma_W^2}{2(n_1-1)} + \frac{\gamma_W^4}{n_1-1} + \frac{\gamma_W^2}{2(n_2-1)} + \frac{\gamma_W^4}{n_2-1}}}, \tag{8}$$

which shall be compared with a standard normal distribution. Feltz and Miller (1996, 1997) give more information about this test.

3.4 Wald test

Rao and Vidya (1992) give the Wald statistic for the case of equal sample sizes. Gupta and Ma (1996) modify it to the general case of unequal sample sizes. The test statistic

$$W = \frac{(c_1 - c_2)^2}{\frac{c_1^2}{2n_1} + \frac{c_1^4}{n_1} + \frac{c_2^2}{2n_2} + \frac{c_2^4}{n_2}} \tag{9}$$

is approximately χ^2 distributed with 1 degree of freedom. This test statistic is obviously closely related to Miller's statistic (8). Bhoj and Ahsanullah (1993) give a third statistic on the same theme, but only for the case of equal sample sizes.

3.5 Score test

Gupta and Ma (1996) derive the score test, based on the maximum likelihood estimates (5). Its explicit value is given by

$$S = \left(\frac{1}{2}\,\hat{\gamma}^2 + \hat{\gamma}^4\right)\left(\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2}\right), \tag{10}$$

where

$$a_i = \hat{\mu}_i^{-2}\hat{\gamma}^{-3}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i)^2 - n_i\hat{\gamma}^{-1}, \quad i = 1, 2.$$

The test statistic (10) shall be compared with a χ^2 distribution with 1 degree of freedom.

3.6 Doornbos and Dijkstra's test

Doornbos and Dijkstra (1983) develop a test based on the distribution of the inverse of the sample coefficient of variation. Let $b_i = 1/c_i, b_W = (n_1 b_1 + n_2 b_2)/(n_1 + n_2)$. The total sum of squares $T = n_1(b_1 - b_w)^2 + n_2(b_2 - b_w)^2$ is sensitive to deviations from the null hypothesis. Doornbos and Dijkstra estimate the expectation of T by

$$\begin{aligned}\hat{E}[T] &= \frac{n_2(n_1-1)}{(n_1+n_2)(n_1-3)} + \frac{n_1(n_2-1)}{(n_1+n_2)(n_2-3)} \\ &+ \frac{1}{c_p^2(n_1+n_2)}\left(\frac{n_1 n_2(n_1-1)}{n_1-3} + \frac{n_1 n_2(n_2-1)}{n_2-3} + n_1^2 e_1^2 + n_2^2 e_2^2 - (n_1 e_1 + n_2 e_2)^2\right)\end{aligned}$$

where

$$c_p^2 = \frac{\frac{n_1(n_1-1)}{n_1-3} + \frac{n_2(n_2-1)}{n_2-3}}{n_1 b_1^2 + n_2 b_2^2 - \frac{n_1-1}{n_1-3} + \frac{n_2-1}{n_2-3}}, \quad e_i = \sqrt{\frac{n_i-1}{2}}\,\frac{\Gamma[\frac{n_i-2}{2}]}{\Gamma\left[\frac{n_i-1}{2}\right]}, \quad i = 1, 2.$$

The test statistic

$$D = \frac{T}{\hat{E}(T)} \tag{11}$$

is approximately χ^2 distributed with 1 degree of freedom.

3.7 Log test

A test based on the logarithmic approach discussed in Section 2 can be made in the following way. Take the logarithm of all observations and calculate the standard deviation s_{L1} in sample 1 and the standard deviation s_{L2} in sample 2. Then compare

$$L = \frac{s_{L1}^2}{s_{L2}^2} \tag{12}$$

with an F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

3.8 Naive test

With the "naive" test the sample coefficients of variation are compared by an F-test in the same way as standard deviations are compared, that is,

$$N = \frac{c_1^2}{c_2^2} \tag{13}$$

is compared with an F-distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

4 An approximate F-test for equality of two coefficients of variation

4.1 The approximate F-test

In the previous section we reviewed eight tests for the hypothesis of equal coefficients of variation. Many of them require large sample sizes. When the numbers of observations are not large it is not clear which test should be preferred. For this reason we now introduce an approximate F-test plausible to work well also for small sample sizes. It is natural to look for an F-test, since such tests are used for comparing variances. The ordinary test statistic for comparing two variances is the ratio between the two variances. If we, for the comparison of two coefficients of variation, analogously take the ratio between the two coefficients of variation we get the naive test (13). This test does not take into account the variation in the estimated averages in the denominators of the coefficients of variation. Therefore it is reasonable to suppose that it is better to build the test on McKay's transformation (3). According to McKay (1932),

$$(n_i - 1)\frac{c_i^2/(1 + \theta_i c_i^2)}{\gamma_i^2/(1 + \gamma_i^2)}, \quad i = 1, 2,$$

is approximately χ^2 distributed with $n_i - 1$ degrees of freedom when $\theta_i = (n_i - 1)/n_i$. The only requirement is that the coefficients of variation are smaller than $1/3$, which is fulfilled when negative observations are unlikely. Consequently we can, if $H_0 : \gamma_1 = \gamma_2$ is true, anticipate

$$F = \frac{c_1^2/(1 + \theta_1 c_1^2)}{c_2^2/(1 + \theta_2 c_2^2)} \tag{14}$$

to be approximately F-distributed with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. The statistic F is an increasing function of c_1 and a decreasing function of c_2. Large deviations between c_1 and c_2 result in large deviations of F from one. Thus F is a plausible test statistic for the hypothesis of equal coefficients of variation.

4.2 The distribution of the test statistic

For inference it is essential that F is approximately F-distributed. We assume that this is the case because F is a quotient between two χ^2 approximations divided by their degrees of freedom. We can, however, not take it for granted, and will therefore investigate the properties of F analytically. We shall compare the distribution of F with the distribution of an F-distributed random variable X with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. The comparison shall be made under the assumptions that the measurements are normally distributed and that the null hypothesis of equal coefficients of variation is correct. We shall see that all moments of F are close to the moments of X if only the coefficient of variation is sufficiently small.

Let W_1 and W_2 denote independent χ^2 distributed random variables divided by their degrees of freedom, let $X = W_1/W_2$, and let Z_1 and Z_2 denote independent standardized normal random variables. The distributions of the sample averages m_i and the standard deviations $c_i m_i$ equals the distributions of $\mu_i + n_i^{-1/2} Z_i \mu_i \gamma$ and $W_i^{1/2} \mu_i \gamma$ respectively. Thus the distribution of c_i^2 equals the distribution of

$$W_i \gamma^2 \left(1 + \frac{Z_i \gamma}{\sqrt{n_i}}\right)^{-2}, \quad i = 1, 2, \tag{15}$$

which inserted in (14) gives

$$F \stackrel{d}{=} \frac{\frac{1}{W_2}\left(1 + \frac{Z_2\gamma}{\sqrt{n_2}}\right)^2 + \frac{n_2 - 1}{n_2}\gamma^2}{\frac{1}{W_1}\left(1 + \frac{Z_1\gamma}{\sqrt{n_1}}\right)^2 + \frac{n_1 - 1}{n_1}\gamma^2},$$

where d denotes equality in distribution. By a Taylor series expansion of F to a power of r, about $\gamma = 0$,

$$\begin{aligned} E[F^r] &= E[X^r] + \frac{1}{2}E\left[2r(r-1)X^{r-1}\left(\frac{Z_2}{\sqrt{n_2}} - \frac{Z_1}{\sqrt{n_1}}\right)\right. \\ &\quad \left. + 2rX^r\left(3\frac{Z_1^2}{n_1} - 4\frac{Z_1 Z_2}{\sqrt{n_1 n_2}} + \frac{Z_2^2}{n_2} - \frac{(n_1-1)W_1}{n_1} + \frac{(n_2-1)W_2}{n_2}\right)\right]\gamma^2 \\ &\quad + O(\gamma^3), \end{aligned}$$

where $X = W_1/W_2$ is an F-distributed random variable with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. From the formula for the r:th moment of X (Kotz and Johnson, 1983) we notice that

$$E[X^{r-1}] = \frac{(n_1-1)(n_2-2r-1)}{(n_2-1)(n_1+2r-3)}\,E[X^r], \quad n_2 > 2r+1.$$

Furthermore, by the formula for the r:th moment of a χ^2 distributed random variable with $n-1$ degrees of freedom (Kotz and Johnson, 1982), it can be shown that

$$E\left[\frac{W_1^{r+1}}{W_2^r}\right] = \frac{n_1+2r-1}{n_1-1}\,E[X^r],$$

since W_1 and W_2 are independent. As a result, the r:th moment of F is, in a neighbourhood of $\gamma = 0$,

$$E[F^r] = E[X^r] + 2rE[X^r]\left(\frac{2-r}{n_1} - \frac{r}{n_2}\right)\gamma^2 + O(\gamma^3), \quad n_2 > 2r+1.$$

We conclude that the moments are similar when the coefficient of variation is small, especially if the sample sizes are equal or large.

We also want to compare the distribution of F with the F-distribution (i.e., the distribution of X). Since F is a ratio of two independent χ^2 approximations it is, however, more convenient to compare the logarithm of F with the logarithm of X. This means that we shall compare the distribution of the logarithm of F with Fisher's z distribution, since originally Fisher (1924) did not define the F-distribution but the z distribution, which is the distribution of $(\log X)/2$. Write $\log F$ as

$$\log F = \log c_1^2\left(1 + \frac{n_1-1}{n_1}c_1^2\right)^{-1} - \log c_2^2\left(1 + \frac{n_2-1}{n_2}c_2^2\right)^{-1}. \qquad (16)$$

The first term in (16) is by (15)

$$\begin{aligned}
&\log c_1^2 \left(1 + \frac{n_1 - 1}{n_1} c_1^2\right)^{-1} \\
&\quad \stackrel{d}{=} \log W_1 + \log \gamma^2 - \log \left(1 + 2\frac{Z_1\gamma}{\sqrt{n_1}} + \frac{Z_1^2\gamma^2}{n_1} + \frac{n_1 - 1}{n_1} W_1\gamma^2\right) \qquad (17)
\end{aligned}$$

Expansion of the last term in (17) yields

$$\begin{aligned}
&\log \left(1 + 2\frac{Z_1\gamma}{\sqrt{n_1}} + \frac{Z_1^2\gamma^2}{n_1} + \frac{n_1 - 1}{n_1} W_1\gamma^2\right) \\
&= 2\frac{Z_1\gamma}{\sqrt{n_1}} + \frac{Z_1^2\gamma^2}{n_1} + \frac{n_1 - 1}{n_1} W_1\gamma^2 - \frac{1}{2}\left(2\frac{Z_1\gamma}{\sqrt{n_1}} + \frac{Z_1^2\gamma^2}{n_1} + \frac{n_1 - 1}{n_1} W_1\gamma^2\right)^2 + \cdots \\
&= 2\frac{Z_1\gamma}{\sqrt{n_1}} + \frac{n_1 - 1}{n_1} W_1\gamma^2 + O_p\left(\max\{\frac{\gamma^2}{n_1}, \gamma^4\}\right)
\end{aligned}$$

where O_p denotes order in probability (Azzalini, 1996). The corresponding calculations can of course be made also for the second term in (16). Now let U_1 and U_2 be independent χ^2 distributed random variables with $n_1 - 1$ and $n_2 - 1$ degrees of freedom respectively, and let Z be an independent standardized normal random variable. Then $\log F$ can be written

$$\log F \stackrel{d}{=} \log X + 2\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, Z\gamma + \left(\frac{1}{n_1} U_1 - \frac{1}{n_2} U_2\right)\gamma^2 + R(n_1, n_2, \gamma),$$

where $R(n_1, n_2, \gamma) = O_p(\max\{n_1^{-1}\gamma^2, n_2^{-1}\gamma^2, \gamma^4\})$. Note that the distribution of $\log F$ consequently equals the distribution of $\log X + O_p(\max\{n_1^{-1/2}\gamma, n_2^{-1/2}\gamma, \gamma^2\})$. We conclude that the distribution of $\log F$ and $\log X$ are similar especially if the coefficient of variation is small or the sample sizes are large.

4.3 A generalized approximate F-test

In applications there are often many independent samples from populations with a common coefficient of variation γ. In a recent article Tian (2005) addresses the problem of making inference about γ in this situation. Tian suggests a repeated sampling method for calculating a generalized probability value as defined by Tsui and Weerahandi (1989). An easy calculated alternative is obtained in the following way. Let $y_{jk} = \mu_j + e_{jk}$, where e_{jk} are independently

distributed $N(0, \gamma^2\mu_j^2)$ with $0 < \mu_j$ and $0 < \gamma < 1/3, j = 1, 2, \ldots, r$ and $k = 1, 2, \ldots, n_j$. Then, by (3), with $\theta_j = (n_j - 1)/n_j$,

$$\frac{\sum_{j=1}^{r}(n_j - 1)c_j^2/(1 + \theta_j c_j^2)}{\gamma^2/(1 + \gamma^2)} \tag{18}$$

is approximately χ^2 distributed with $\sum_j n_j - r$ degrees of freedom. Thus (18) can be used as an approximately χ^2 distributed test statistic for the hypothesis that the common coefficient of variation equals γ.

We shall also derive a useful extension of the approximate F-test. Let $y_{ijk} = \mu_{ij} + e_{ijk}$, where e_{ijk} are independently distributed $N(0, \gamma_i^2\mu_{ij}^2)$ with $0 < \mu_{ij}$ and $0 < \gamma_i < 1/3, i = 1, 2; j = 1, 2, \ldots, r_i$ and $k = 1, 2, \ldots, n_{ij}$. If the hypothesis $H_0 : \gamma_1 = \gamma_2$ is true, then by (18)

$$G = \frac{\left(\sum_{j=1}^{r_2} n_{2j} - r_2\right) \sum_{j=1}^{r_1} \frac{(n_{1j}-1)c_{1j}^2}{1+\theta_{1j}c_{1j}^2}}{\left(\sum_{j=1}^{r_1} n_{1j} - r_1\right) \sum_{j=1}^{r_2} \frac{(n_{2j}-1)c_{2j}^2}{1+\theta_{2j}c_{2j}^2}}, \tag{19}$$

with $\theta_{ij} = (n_{ij} - 1)/n_{ij}$, is approximately F-distributed with $\sum_j n_{1j} - r_1$ and $\sum_j n_{2j} - r_2$ degrees of freedom.

5 An immunoassay example

Brunnée et al. (1996) compares two methods for measuring concentration of specific IgE antibodies in blood samples. A new system, ELItest, was compared with the established Pharmacia CAP system (PCS). Among other things the variations between and within assays were studied. Specific IgE for the allergens mite, cat and birch was measured for 3 sera with very different levels of concentration. The *intra* assay coefficients of variation were calculated on 8 measurements performed on the same day, and the *inter* assay coefficients of variation were calculated on 10 measurements made on different days. Brunnée et al. (1996) perform no hypothesis tests of the coefficients of variation. This is very representative for studies of precision in diagnostic measuring instruments. Usually no tests are performed, since there is no well-known method for doing it.

The reported *intra* assay coefficients of variation are given in Table 1 together with calculated approximate F-tests (14). No differences are significant at level 5%. Observe that this is also true for the third sample of allergen mite, although the estimate of the coefficient of variation in ELItest (18.6%)

Table 1: The approximate F-test (14) applied to intra assay coefficients of variation (CV) reported by Brunnée et al. (1996)

Allergen	ELItest CV (%) ($n = 8$)	PCS CV (%) ($n = 8$)	F	P-value
Mite	6.6	9.5	0.485	0.360
Mite	3.3	4.8	0.473	0.345
Mite	18.6	8.3	4.904	0.052
Cat	6.9	10.0	0.478	0.352
Cat	4.5	5.5	0.670	0.610
Cat	4.2	4.6	0.834	0.817
Birch	4.7	9.2	0.262	0.099
Birch	3.8	5.4	0.496	0.375
Birch	4.8	8.2	0.344	0.182

is more than twice as large as the estimate of the coefficient of variation in Pharmacia CAP System (8.3%). The result is however close to the border of being significant (p-value 0.052), and it is notable that all other samples show smaller coefficients of variation in ELItest than in Pharmacia CAP System.

If we assume that each method has a constant intra assay coefficient of variation we can apply the generalized approximate F-test given in (19). The hypothesis of equal intra assay coefficients of variation is not rejected, because $G = 1.046$ with 63 degrees of freedom in the numerator and 63 degrees of freedom in the denominator (P-value 0.8597). However, this result is to large extent dependent on the third sample of allergen mite. If the estimate of the coefficient of variation in ELItest (18.6%) is considered to be an outlier, maybe because of suspected errors in the performance of the assay, and accordingly excluded from the calculation of the hypothesis test the result is clearly significant. Then $G = 2.285$ with 63 degrees of freedom in the numerator and 56 degrees of freedom in the denominator (P-value 0.0020).

Suppose that we require that the intra assay coefficient of variation is smaller than 10%. Consider the hypothesis that the common intra assay coefficient of variation is 10% in Pharmacia CAP System. The test statistic (18) equals 36.16, which shall be compared with a χ^2 distribution with 63 degrees of freedom. In conclusion the intra assay coefficient of variation is significantly smaller than 10% (P-value 0.0026).

Table 2 includes the *inter* assay coefficients of variation as reported by

Table 2: The approximate F-test (14) applied to inter assay coefficients of variation (CV) reported by Brunnée et al. (1996)

Allergen	ELItest CV (%) ($n = 10$)	PCS CV (%) ($n = 10$)	F	P-value
Mite	20.1	11.7	2.883	0.131
Mite	16.5	10.1	2.629	0.166
Mite				
Cat	26.9	10.3	6.465	0.010
Cat	13.9			
Cat				
Birch	32.6	15.6	4.073	0.048
Birch	16.5	12.7	1.671	0.456
Birch	17.4	8.0	4.632	0.032

Brunnée et al.(1996) and the corresponding results of the proposed approximate F-test given in (14). Due to missing values, only 6 comparisons can be made. Differences are significant at level 5% in 3 cases, all of advantage to the established system.

Note that the test statistic G given in (19) shall not be applied to the inter assay coefficients of variation. There are 7 estimates of the inter assay coefficient of variation in ELItest, but they are not independent since they are based on the same 10 days. Neither the 6 estimates of the inter assay coefficient of variation in Pharmacia CAP System are independent.

6 A simulation study

6.1 Objective

We shall by Monte Carlo technique investigate the significance levels and powers of the tests reported in Section 3 and the approximate F-test (14) introduced in Section 4.1.

6.2 Methods

The following tests were included in the study: the approximate F-test (14), the likelihood ratio test (6), Miller's test (8), Bennett's test (7), Doornbos and Dijkstra's test (11), the Wald test (9), the score test (10), the naive test

(13) and the log test (12). In each simulation two samples with n_1 and n_2 observations respectively were randomly generated 20 000 times in Release 13 of MATLAB (The Mathworks Inc., Natick, MA, USA). The observations belonged to normal distributions with expected values 100 and 1000, and with coefficients of variation γ_1 and γ_2 respectively. The tests were performed with significance level 5% against the alternative hypothesis of unequal coefficients of variation, *i.e.* the tests were two-sided. With the various ?2-tests the null hypothesis was rejected when the test statistic was larger than the 95th percentile of the χ^2 distribution. When using F-tests the null hypothesis was rejected if the test statistic was smaller than the 2.5th percentile or larger than the 97.5th percentile of the F-distribution. With Miller's test the null hypothesis was rejected when the test statistic was smaller than the 2.5th percentile or larger than the 97.5th percentile of the standard normal distribution.

Four cases were studied. The type I errors of the tests were investigated in Case 1-3, and the powers of the tests were investigated in Case 4. The first case had a small coefficient of variation (5%) and equal sample sizes. The second case had instead a large coefficient of variation (25%), and still equal sample sizes. The third case had large coefficients of variation but unequal sample sizes, since n_1 was fixed to 4. In the fourth case one coefficient of variation was 5% and the other 10%, and the sample sizes were equal. The size, n_2, of the second sample varied from 2 to 20 in all cases. Thus 19 simulations were made per case.

6.3 Results

All the tables and figures are attached in the end of the paper.

The results of the simulations according to Case 1 is reported in Table 3 and illustrated in Figure 1. The figure shows that three tests performed well with regards to type I error: the approximate F-test, the naive test and the log test all showed relative frequencies of rejections close to the significance level 5%. Miller's test, Bennett's test and the Wald test worked well when the sample sizes were not very small. The likelihood ratio test, Doornbos and Dijkstra's test and the score test required large sample sizes.

The results of the simulations according to Case 2 is reported in Table 4 and illustrated in Figure 2. In this case the coefficient of variation was large (25%). The approximate F-test showed nevertheless almost correct probability of type I error (5%). The naive test rejected the null hypothesis with a probability somewhat larger than 5%. The log test, interestingly, did not work in a proper way. Miller's test, Bennett's test and the Wald test behaved well when the

sample sizes were not very small. The likelihood ratio test, Doornbos and Dijkstra's test and the score test required large sample sizes.

The results of the simulations according to Case 3 is reported in Table 5 and illustrated in Figure 3. In this case, with unequal sample sizes and at least one small sample size ($n_1 = 4$) in combination with a large coefficient of variation, the approximate F-test was the only test that showed nearly correct probability of type I error (5%). The Wald test, which showed good performance in Case 1 and Case 2, did not perform well in this case. Neither did the likelihood ratio test nor Doornbos and Dijkstra's test. The log test had too large relative frequency of rejected hypotheses, and the score test had too small. Miller's test, Bennett's test and the naive test worked better, but not as good as the approximate F-test.

The results of the simulations according to Case 4 is reported in Table 6 and illustrated in Figure 4. For all tests the powers increased with the number of observations and reached a level of app. 80% when the sample sizes were 20. The likelihood ratio test showed large power for small sample sizes, but it also rejected the null hypothesis when it was true, see Figure 1. The score test and Doornbos and Dijkstra's test never rejected the hypothesis of equal coefficients of variation when the sample sizes were small. Miller's test and the Wald test had very small powers when $n_1 = n_2 = 2$, otherwise they worked similar as the approximate F-test, Bennett's test, the naive test and the log test.

6.4 Conclusions

The likelihood ratio test, the Wald test, Doornbos and Dijkstra's test and the log test all showed poor performance with regard to type I error in at least one of Case 1-3. For this reason they are not recommended for use. The results of the score test were not as good as the results of the other tests, neither considering type I error nor considering power. The naive test worked similar as the approximate F-test, but had too large probability of type I error when the coefficient of variation was large. Three tests performed well: the approximate F-test, Miller's test and Bennett's test. Miller's test did however not work properly when the sample sizes were very small, and Bennett's test rejected the true null hypotheses too often. The approximate F-test was the only test that showed almost correct probability of type I error when the sample sizes were small.

7 Discussion

Warren (1982) writes: "While workers in many fields recognize the imprecision in a sample mean, and will now routinely compute a standard error, or a confidence interval, for the mean, many of these same workers will treat the sample coefficient of variation as if it were an absolute quantity. Inferences based on this measure of variability may then be questionable. Nevertheless, it should be possible to persuade such workers that, as with the sample mean, some measure of precision should be attached to the sample coefficient of variation." Though many years have passed since Warren made this reflection the situation has not changed. Researchers still lack standard methods for expressing the precision in estimated coefficients of variation. The purpose of this article has been to explore tests that have been suggested but are seldom used, and to contribute to the knowledge about how to make valid statistical inference.

For the hypothesis of equal coefficients of variation we have proposed a new easily calculated test statistic F, which is approximately F-distributed. We have shown that all moments of F are close to the moments of an F-distributed random variable if the unknown common coefficient of variation is sufficiently small. We have also proved that the logarithm of F in distribution equals the logarithm of an F-distributed random variable plus some error variables that are in probability of small orders.

We have made a simulation study that is unique and important since many of the tests have never been compared with each other. The study revealed that several proposed tests have erroneous type I errors when the sample sizes are small. The likelihood ratio test, the Wald test, the score test and Doornbos and Dijkstra's test shall not be used unless the sample sizes are large. One of the most interesting results of the simulation study is that a variance test carried out on log values, i.e., the "log test", performs badly when the coefficient of variation is not small. This is a key result since statisticians often use the logarithmic transformation when the standard deviation is proportional to the average. The proposed approximate F-test was the only test that showed almost correct probability of type I error when the sample sizes were small.

Unlike several tests the proposed approximate F-test is easily generalized to a situation with many independent estimates of the coefficients of variation. We have made the appropriate extension and introduced the generalized approximate F-test. In this test estimates based on many observations are more important than estimates based on few observations. Each estimate is, after a transformation, simply weighted by its degrees of freedom. This method also manages the problem considered by Tian (2005) of testing an *a priori* common

coefficient of variation.

The coefficient of variation is the predominant measure of dispersion in diagnostic research. The measurements are often assumed to be normally distributed. We have studied a real example with immunoassay data in which the precisions of two diagnostic instruments were compared. The random variations in the blood sample concentrations were measured by coefficients of variation, but no statistical tests were performed in the original article. We have presented a method for analysing the data. Our basis has been that medical researchers often do right when calculating coefficients of variation, but are in need of statistical tools for evaluation, exactly as Warren (1982) pointed out.

Acknowledgements

The author thanks Prof. Dietrich von Rosen for many ideas and Associate Professor Jun Yu for editing the final version with great competence. The Centre of Biostochastics, Swedish University of Agricultural Sciences, and Pharmacia Diagnostics AB financed the research.

References

Azzalini, A. (1996). *Statistical Inference – Based on the Likelihood.* Chapman and Hall: London.

Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**: 307-322.

Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika*, **78**: 557-563.

Bennett, B.M. (1976). On an approximate test for homogeneity of coefficients of variation. Contributions to Applied Statistics dedicated to A. Linder, ed. by Ziegler, W.J. *Experientia Supplementum*, **22**: 169-171.

Bennett, B.M. (1977). LR tests for homogeneity of coefficients of variation in repeated samples. *Sankhyā, Series B*, **39**: 400-405.

Bhoj, D.S. and Ahsanullah, M. (1993). Testing equality of coefficients of variation of two populations. *Biometrical Journal*, **35**: 355-359.

Brunnée, T., Seeberger, A., Kleine-Tebbe, J. and Kunkel, G. (1996). Comparison between two automated systems to determine specific IgE: CAP and ELItest. *Clinical and Experimental Allergy*, **26**: 1420-1427.

DeSilva, B., Smith, W., Weiner, R., Kelley, M., Smolec, J., Lee, B., Khan, M., Tacey, R., Hill, H. and Celniker, A. (2003). Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharmaceutical Research*, **20**: 1885-1900.

Doornbos, R. and Dijkstra, J.B. (1983). A multi sample test for the equality of coefficients of variation in normal populations. *Communications in Statistics - Simulation and Computation*, **12**: 147-158.

Feltz, C.J. and Miller, G.E. (1996). An asymptotic test for the equality of coefficients of variation from k populations. *Statistics in Medicince*, **15**: 647-658.

Fieller, E.C. (1932). A numerical test of the adequacy of A.T. McKay's approximation. *Journal of the Royal Statistical Society*, **95**: 699-702.

Fisher, R.A. (1924). On a distribution yielding the error functions of several well known statistics. *Proceedings of the International Congress of Mathematics, Toronto*, **2**: 805-813.

Food and Drug Administration. *Guidance for Industry. Statistical Approaches to Establishing Bioequivalence.* Center for Drug Evaluation and Research (CDER). Maryland, 2001.

Gerig, T.M. and Sen, A.R. (1980). MLE in Two normal samples with equal but unknown population coefficients of variation. *Journal of the American Statistical Association*, **75**: 704-708.

Gupta, R.C. and Ma, S. (1996). Testing the equality of coefficients of variation in k normal populations. *Communications in Statistics - Theory and Methods*, **25**: 115-132.

Iglewicz, B. and Myers, R.H. (1970). Comparison of approximations to the percentage points of the sample coefficient of variation. *Technometrics*, **12**: 166-169.

Kotz, S. and Johnson, N.L. (1983). *Encyclopedia of Statistical Sciences, Vol. 3.* Wiley: New York.

Kotz, S. and Johnson, N.L. (1982). *Encyclopedia of Statistical Sciences, Vol. 1.* Wiley: New York.

Lohrding, R.K. (1975). A two sample test of equality of coefficients of variation or relative errors. *Journal of Statistical Computation and Simulation*, **4**: 31-36.

McKay, A.T. (1932). Distribution of the coefficient of variation and the extended t-distribution. *Journal of the Royal Statistical Society*, **95**: 695-698.

Miller, E.G. and Karson, M.J. (1977). Testing equality of two coefficients of variation. *American Statistical Association: Proceedings of the Business and Economics Section*, Part **I**: 278-283.

Miller, G.E. (1991). Asymptotic test statistics for coefficients of variation. *Communications in Statistics - Theory and Methods*, **20**: 3351-3363.

Miller, G.E. and Feltz, C.J. (1997). Asymptotic inference for coefficients of variation. *Communications in Statistics - Theory and Methods*, **26**: 715-726.

Owen, D.B. (1968). A survey of properties and applications of the noncentral t-distribution. *Technometrics*, **10**: 445-478.

Pearson, E.S. (1932). Comparison of A.T. McKay's approximation with experimental sampling results. *Journal of the Royal Statistical Society*, **95**: 703-704.

Pearson, K. (1896). Mathematical contributions to the theory of evolution III: Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London, Series A*, **187**: 253-318.

Pitman, E.J.G. (1939). Tests of hypotheses concerning location and scale parameters. *Biometrika*, **31**: 200-215.

Rao, K.A. and Vidja, R. (1992). On the performance of a test for coefficient of variation. *Calcutta Statistical Association Bulletin*, **42**: 87-95.

Schimmerl-Metz, S.M., Metz, V.M., Totterman, S.M.S., Mann, F.A. and Gilula, L.A. (1999). Radiologic measurement of the scapholunate joint: Implications of biologic variation in scapholunate joint morphology. *The Journal of Hand Surgery*, **24A**: 1237-1244.

Shafer, N.J. and Sullivan, J.A. (1986). A simulation study of a test for the equality of the coefficients of variation. *Communications in Statistics - Simulation and Computation* , **15**: 681-695.

Shao, J. (2003). *Mathematical Statistics*. 2nd ed. Springer: New York.

Tian, L. (2005). Inferences on the common coefficient of variation. *Statistics in Medicine*, **24**:2213-2220. DOI:10.1002/sim.2088

Tsui, K.W. and Weerahandi, S. (1989). Generalized P-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of American Statistical Association*, **84**:602-607.

Umphrey, G.J. (1983). A comment on McKay's approximation for the coefficient of variation. *Communications in Statistics - Simulation and Computation*, **12**: 629-635.

Vangel, M.G. (1996). Confidence intervals for a normal coefficient of variation. *The American Statistician*, **15**: 21-26.

Warren, W.G. (1982). On the adequacy of the chi-squared approximation for the coefficient of variation. *Communications in Statistics - Simulation and Computation*, **11**:659-666.

Wong, A.C.M. and Wu, J. (2002). Small sample asymptotic inference for the coefficient of variation: normal and nonnormal models. *Journal of Statistical Planning and Inference*, **104**: 73-82.

Tables and Figures

The significance level is 5% in all cases. In tables 3-6, following notation is used: F = F-test (14), R = Likelihood ratio test (6), M = Miller's test (8), B = Bennett's test (7), D = Doornbos and Dijkstra's test (11), W = Wald test (9), S = Score test (10), N = Naive test (13), L = Log test (12).

Table 3: Case 1: Pr(Type I error) in percentages when $\gamma_1 = \gamma_2 = 0.05$.

n_1	n_2	F	R	M	B	D	W	S	N	L
2	2	4.56	24.28	1.00	8.85	0.00	1.97	0.00	4.56	4.57
3	3	5.12	15.30	6.35	7.65	-	6.55	0.00	5.14	5.17
4	4	4.92	11.17	6.17	6.72	0.00	6.34	0.31	4.93	5.01
5	5	5.15	9.83	6.23	6.51	0.02	6.31	2.19	5.19	5.23
6	6	4.66	8.46	5.71	5.84	0.43	5.76	2.73	4.69	4.73
7	7	5.05	8.15	5.84	5.98	1.05	5.90	3.64	5.13	5.18
8	8	5.02	7.69	5.72	5.82	1.61	5.77	3.77	5.04	5.12
9	9	4.83	7.12	5.38	5.42	1.78	5.39	3.84	4.86	5.00
10	10	4.78	6.80	5.28	5.35	2.10	5.31	3.94	4.81	4.95
11	11	4.95	6.85	5.53	5.59	2.45	5.56	4.32	4.99	5.18
12	12	5.04	6.65	5.53	5.56	2.60	5.53	4.40	5.07	5.25
13	13	4.91	6.51	5.42	5.49	2.88	5.46	4.39	4.97	5.01
14	14	5.11	6.60	5.57	5.60	3.21	5.59	4.61	5.17	5.29
15	15	4.98	6.27	5.30	5.33	3.23	5.33	4.53	5.03	5.13
16	16	4.93	5.99	5.18	5.21	3.31	5.20	4.55	4.98	5.09
17	17	4.87	5.92	5.21	5.21	3.24	5.21	4.47	4.90	5.07
18	18	5.23	6.30	5.55	5.58	3.56	5.58	4.86	5.28	5.37
19	19	4.94	5.99	5.29	5.30	3.61	5.29	4.61	5.00	5.31
20	20	5.16	5.97	5.42	5.42	3.96	5.42	4.87	5.20	5.34

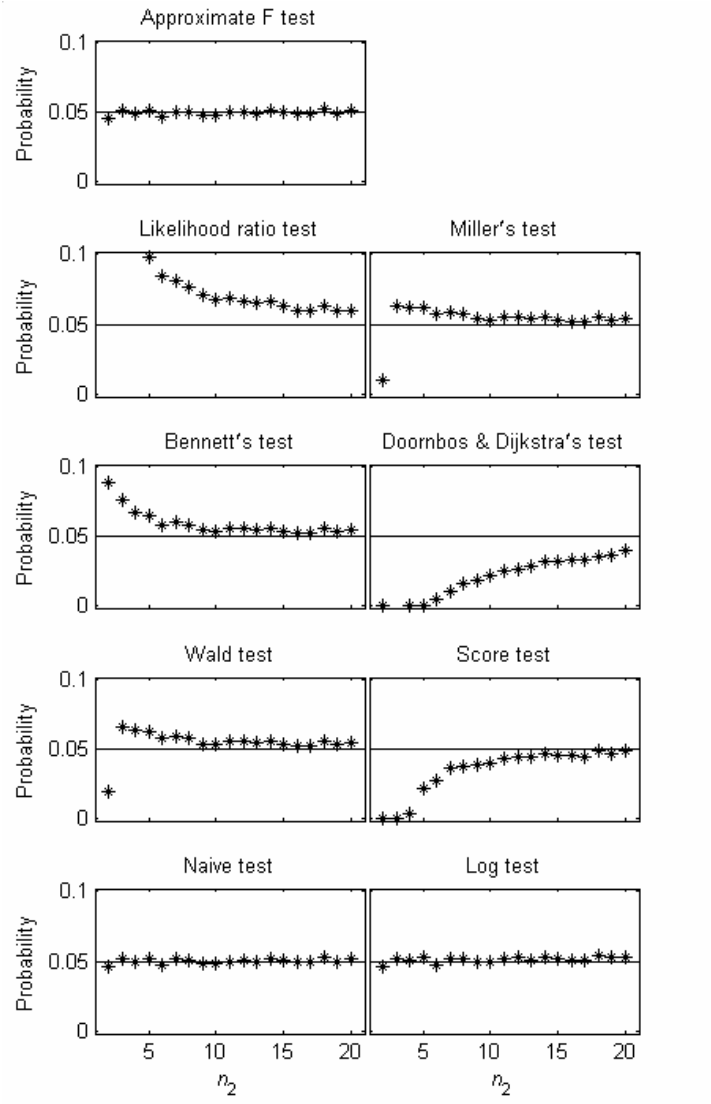

Figure 1: Case 1. Probability of type I error when $\gamma_1 = \gamma_2 = 0.05$ and $n_1 = n_2$.

Table 4: Case 2: Pr(Type I error) in percentages when $\gamma_1 = \gamma_2 = 0.25$.

n_1	n_2	F	R	M	B	D	W	S	N	L
2	2	5.02	24.69	0.32	9.18	0.00	0.37	0.00	5.20	5.34
3	3	4.91	14.78	5.50	7.57	-	3.59	0.00	5.38	6.01
4	4	5.16	11.64	5.89	7.02	0.00	4.31	0.37	5.75	7.13
5	5	5.21	10.02	5.77	6.53	0.07	4.62	2.31	5.96	7.95
6	6	5.07	9.00	5.66	6.21	0.53	4.65	3.17	6.01	8.76
7	7	4.96	8.29	5.46	5.97	0.96	4.58	3.54	5.95	9.25
8	8	4.87	7.49	5.21	5.63	1.43	4.43	3.70	5.84	9.66
9	9	5.05	7.46	5.48	5.84	2.14	4.66	4.07	6.21	10.42
10	10	5.04	7.15	5.37	5.72	2.33	4.71	4.20	6.15	10.92
11	11	5.48	7.19	5.82	6.03	2.87	5.23	4.76	6.61	11.33
12	12	4.96	6.37	5.22	5.42	2.73	4.75	4.40	5.97	11.30
13	13	4.93	6.52	5.22	5.48	2.92	4.69	4.37	6.17	11.53
14	14	4.96	6.38	5.17	5.31	3.14	4.74	4.50	6.18	11.79
15	15	5.24	6.70	5.48	5.69	3.41	5.03	4.75	6.63	12.69
16	16	4.81	6.13	5.03	5.15	3.30	4.64	4.46	6.11	12.58
17	17	5.35	6.39	5.56	5.64	3.64	5.14	4.94	6.51	13.20
18	18	5.03	6.26	5.29	5.51	3.52	4.86	4.70	6.47	13.39
19	19	4.84	5.74	4.99	5.09	3.54	4.70	4.52	6.06	13.08
20	20	4.99	5.88	5.15	5.24	3.87	4.85	4.73	6.23	13.55

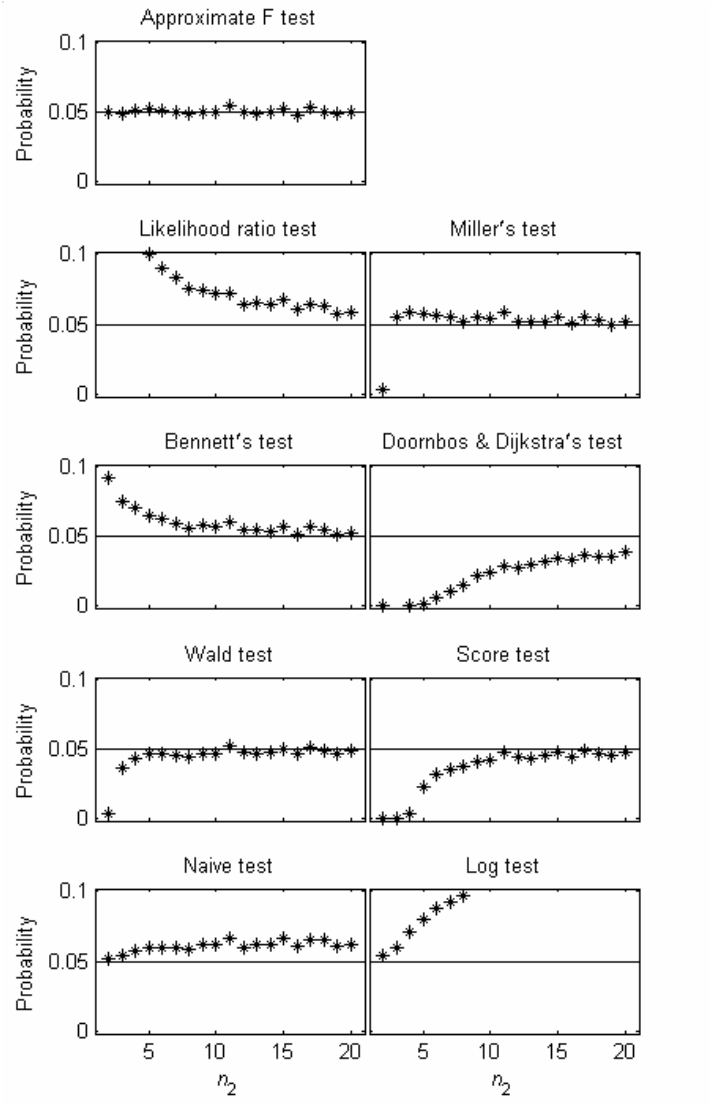

Figure 2: Case 2. Probability of type I error when $\gamma_1 = \gamma_2 = 0.25$ and $n_1 = n_2$.

Table 5: Case 3: Pr(Type I error) in percentages when $\gamma_1 = \gamma_2 = 0.25$.

n_1	n_2	F	R	M	B	D	W	S	N	L
4	2	5.17	21.56	3.49	8.36	0.00	20.45	2.24	5.55	6.00
4	3	5.29	13.95	5.48	7.59	-	6.49	1.35	5.85	6.84
4	4	5.11	11.55	5.77	6.92	0.00	4.22	0.47	5.69	7.10
4	5	5.31	11.02	5.90	6.98	0.00	5.50	1.82	6.09	7.64
4	6	5.15	10.38	5.38	6.53	0.00	6.91	2.54	5.95	7.77
4	7	5.20	10.68	5.18	6.71	0.01	9.22	2.62	6.13	8.19
4	8	4.87	10.02	4.79	6.33	0.05	10.50	2.50	5.67	7.78
4	9	5.03	10.48	4.67	6.27	0.14	12.12	2.74	5.85	8.09
4	10	4.86	10.60	4.43	6.19	0.21	13.51	2.81	5.81	8.18
4	11	5.19	11.00	4.53	6.29	0.27	14.77	2.97	6.12	8.10
4	12	5.18	10.58	4.56	6.46	0.44	15.34	2.86	6.24	8.38
4	13	5.27	10.84	4.37	6.56	0.48	16.38	3.08	6.26	8.75
4	14	5.05	10.84	4.20	6.29	0.43	17.10	2.92	6.15	8.46
4	15	5.04	10.58	4.08	6.06	0.61	17.75	2.75	6.02	8.40
4	16	4.91	11.09	3.93	6.12	0.63	18.82	2.69	5.94	8.13
4	17	4.98	11.08	4.00	6.15	0.68	19.11	2.90	6.11	8.18
4	18	4.79	10.72	3.73	5.88	0.78	19.18	2.62	5.71	7.92
4	19	5.24	11.23	4.01	6.34	0.87	19.87	2.89	6.24	8.43
4	20	5.10	11.38	3.86	6.30	1.07	20.28	2.68	6.12	8.35

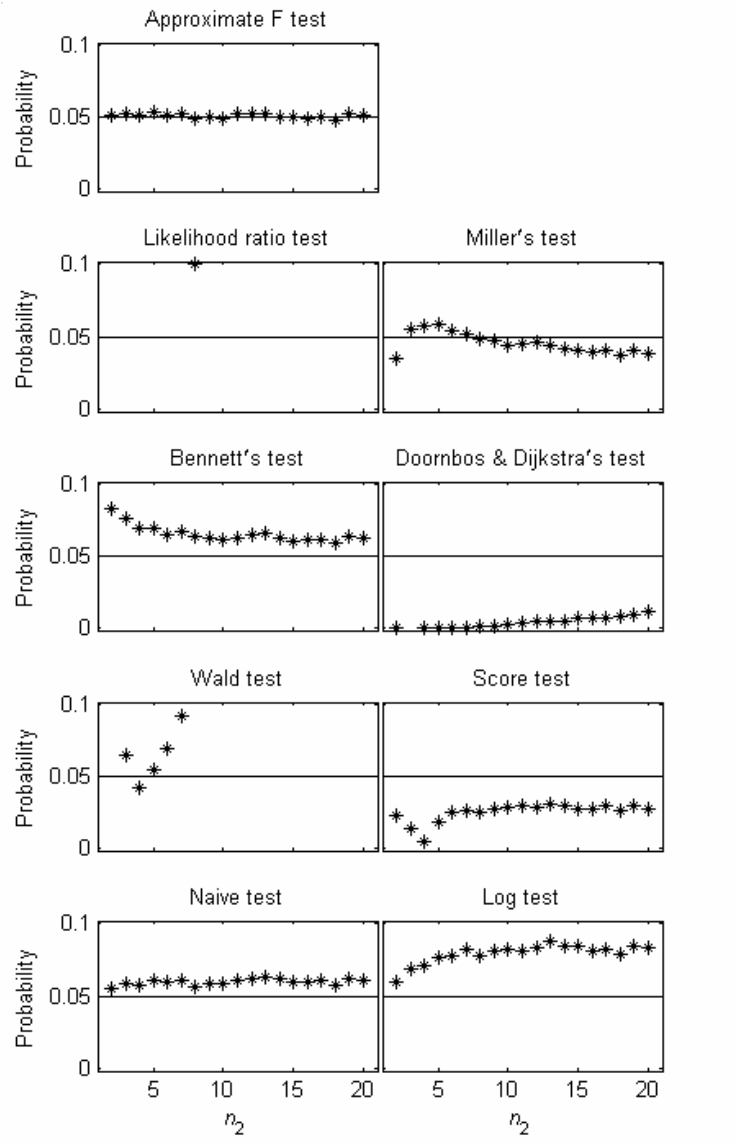

Figure 3: Case 3. Probability of type I error when $\gamma_1 = \gamma_2 = 0.25$ and $n_1 = 4$.

Table 6: Case 4: Power in percentages when $\gamma_1 = 0.05$ and $\gamma_2 = 0.10$.

n_1	n_2	F	R	M	B	D	W	S	N	L
2	2	6.15	30.44	1.25	11.77	0.00	1.91	0.00	6.17	6.18
3	3	9.72	26.19	12.03	14.56	-	11.99	0.00	9.82	9.95
4	4	15.09	29.05	18.04	19.34	0.00	17.92	1.43	15.30	15.68
5	5	20.52	32.27	23.38	24.31	0.28	23.32	11.32	20.84	21.22
6	6	26.65	37.40	29.71	30.38	4.72	29.63	19.42	26.98	27.68
7	7	32.51	41.86	35.14	35.59	12.30	35.03	26.77	32.88	33.43
8	8	38.76	47.13	41.21	41.56	20.38	41.14	34.15	39.15	39.82
9	9	44.07	51.27	46.19	46.62	28.07	46.14	40.14	44.40	45.04
10	10	49.47	55.97	51.51	51.76	34.83	51.47	46.07	49.85	50.47
11	11	54.16	59.92	55.96	56.13	41.80	55.94	51.36	54.54	55.12
12	12	57.96	63.13	59.54	59.73	47.12	59.50	55.50	58.37	58.77
13	13	63.40	68.02	64.93	65.05	53.97	64.90	61.58	63.74	64.44
14	14	67.23	71.28	68.53	68.65	58.74	68.51	65.52	67.57	68.29
15	15	69.55	73.20	70.73	70.81	62.58	70.70	68.18	69.89	70.47
16	16	73.44	76.36	74.30	74.39	67.03	74.27	72.15	73.69	74.10
17	17	76.04	78.92	76.94	77.06	70.64	76.94	74.90	76.37	76.90
18	18	78.72	81.13	79.43	79.49	74.03	79.41	77.58	78.96	79.48
19	19	81.09	83.26	81.82	81.88	77.13	81.80	80.19	81.39	81.79
20	20	83.68	85.59	84.37	84.42	80.02	84.33	82.98	83.95	84.26

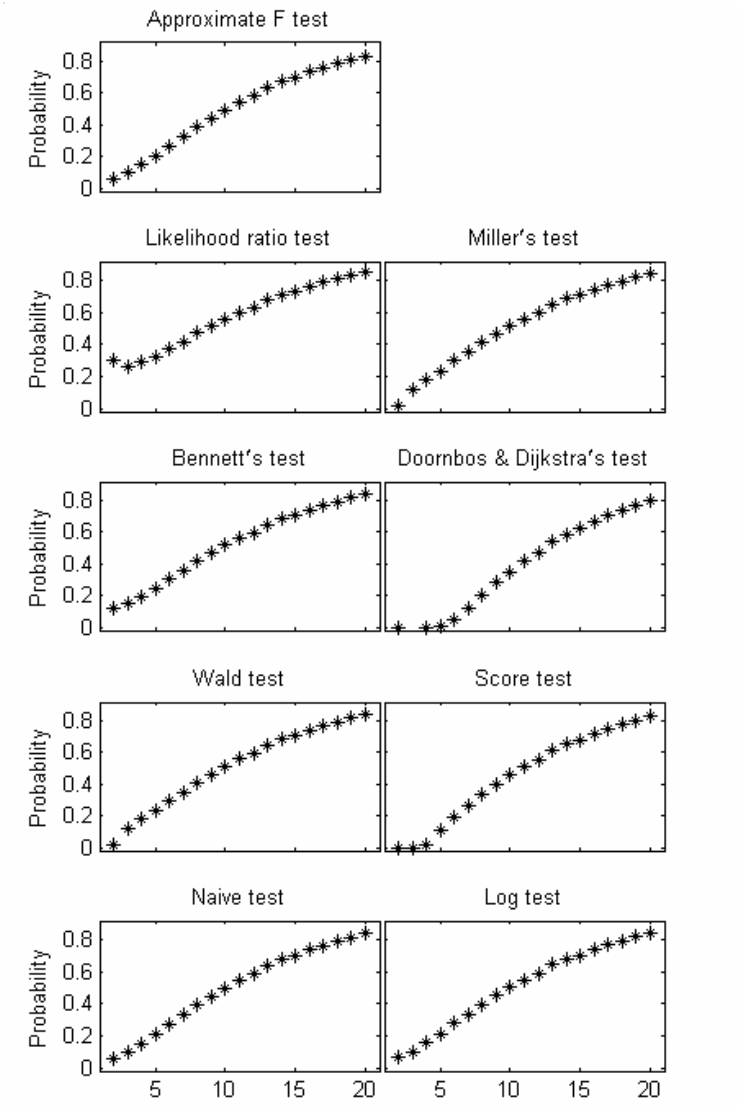

Figure 4: Case 4. Power when $\gamma_1 = 0.05, \gamma_2 = 0.10$ and $n_1 = n_2$.