Coefficients of Variation – an Approximate F-Test

F. Johannes Forkman

Institutionen för biometri och teknik
Department of Biometry and Engineering

SLU
Uppsala 2005

Licentiatavhandling 003
Licentiate thesis 003

ISSN 1652-3261
ISBN 91-576-6886-8

Till pappa

Abstract

Forkman, F.J. 2005. *Coefficients of Variation – an Approximate F-Test*.
Licentiate thesis.
ISSN 1652-3261. ISBN 91-576-6886-8.

Basic inferential methods for analysing coefficients of variation in normally distributed data are studied. The assumptions of normally distributed observations and a constant coefficient of variation are discussed and motivated especially for immunoassay data. An approximate F-test for comparing two coefficients of variation is introduced. All moments of the proposed test statistic are shown to be approximately equal to the moments of an F distribution. It is proved that 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. The approximate F-test is compared with eight other tests in a simulation study. The new test turns out to perform well, also in case of small sample sizes. A generalized version of the approximate F-test is defined for the case that there are several estimates of each coefficient of variation, calculated with different averages. The test is based on a χ^2 approximation given 1932 by A. T. McKay. It is proved that McKay's approximation is noncentral beta distributed.

Keywords: coefficient of variation, normal distribution, confidence interval, hypothesis test, McKay's approximation.

Author's address: Johannes Forkman, Department of Biometry and Engineering, SLU, P.O. Box 7032, SE-750 07 Uppsala, Sweden. E-mail: Johannes.Forkman@bt.slu.se.

Sammanfattning

Variationskoefficienten är standardavvikelsen dividerad med medelvärdet. Detta spridningsmått används på många håll i praktiken när man är intresserad av hur stor variationen är i relation till nivån på observationerna. Ändå analyseras variationskoefficienter sällan statistiskt. Syftet med denna avhandling är att redogöra för metoder för analys av variationskoefficienter i normalfördelade material, särskilt metoder för att beräkna konfidensintervall och för att pröva hypotesen att två variationskoefficienter är lika, samt att introducera några nya idéer.

När standardavvikelsen är proportionell mot medelvärdet är det vanligt att logaritmera observationerna och anta att de logaritmerade observationerna är normalfördelade. I denna avhandling argumenteras för alternativet att istället anta normalfördelning i ursprunglig skala och analysera observationerna utan att först logaritmera dem. Speciellt visas för immunokemiska metoder att det kan vara rimligt att anta att mätvärdena är normalfördelade i ursprunglig skala och att variationskoefficienten är konstant.

Ett nytt test för att pröva hypotesen att två variationskoefficienter är lika föreslås. Teststatistikan jämförs med F-fördelningen. Alla teststatistikans moment är approximativt lika momenten för en F-fördelad variabel. Logaritmen av teststatistikan är i fördelning är lika med logaritmen av en F-fördelad variabel plus några feltermer som i sannolikhet är av små ordningar.

Teststatistikan är baserad på A. T. McKays χ^2-approximation från 1932. Många har visat numeriskt och analytiskt att McKays approximation verkligen är approximativt χ^2-fördelad. I denna avhandling visas att McKays approximation är ickecentralt betafördelad.

Det nya testet jämförs med åtta andra test i en simuleringsstudie. I studien är det nya F-testet det enda som ger nästan korrekt typ I-fel när antalet observationer är litet. Alternativet att logaritmera observationerna före analys fungerar i studien dåligt när variationskoefficienten är stor.

Det nya testet går, till skillnad från många andra test, att generalisera till fallet att det finns många oberoende skattningar av variationskoefficienterna. Testet generaliseras och tillämpas på ett verkligt exempel från diagnostisk forskning.

Contents

1. Introduction

The coefficient of variation is the standard deviation divided by the mean. Pearson (1896) defined the term and applied it when comparing various measurements on females with corresponding measurements on males. Ever since then it has been used as a measure of dispersion in many biological applications. 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.

When laboratory analytical procedures are employed 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.

When clinical trials are studied 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, according to Schafer & Sullivan (1986), be similar or equal in the two groups.

Despite the large number of applications the properties of the coefficient of variation is seldom discussed in statistical textbooks. As a consequence there is among practitioners often an inadequate knowledge on how to make basic statistical inference concerning the measure. We shall in this thesis investigate different solutions and recommend statistical methods. We are motivated by the gap between theory and practice. In theory the variance is the primary measure of dispersion, but in practice it is, in some fields of interest, the coefficient of variation. The objective is to collect and report existing information about how to construct confidence intervals and hypothesis tests for the coefficient of variation and add some new ideas. This investigation shall be a basis for future research.

A popular technique when the standard deviation is proportional to the average is to apply the logarithmic transformation and perform the analysis on log values.

The assumption of lognormally distributed data thus introduced is however not always reasonable. We discuss this in Chapter 2 and provide especially for immunoassay applications a rational for assuming a constant coefficient of variation in combination with normally distributed data. In Chapter 3 we give an overview of basic inference for the coefficient of variation when the distribution is normal with positive expected value. We explain how to make exact confidence intervals for coefficients of variation, based on the noncentral t distribution, or how to calculate approximate confidence intervals with readily used formulas. We engage especially in the comparison of two coefficients of variation and describe several tests that could be of interest to the practitioner. In Chapter 4 we introduce a new hypothesis test, which is an approximate F-test that we develop specifically for the hypothesis of equal coefficients of variation. The test is easily generalized to the case that there are several estimates per coefficient of variation. We show by a real data example from diagnostic research how the approximate F-test and its generalized version can be applied. In Chapter 5 we study the properties of the approximate F-test theoretically. We show that all moments of the test statistic is close to the moments of an F distributed random variable if the coefficient of variation is sufficiently small. We also prove that the distribution of the logarithm of the test statistic approximately equals the distribution of the logarithm of an F distributed random variable. In Chapter 6 we compare the new test with other tests by a simulation study. The new test turns out to perform well in comparison with other tests. Since the new test is built on a χ^2 approximation given by McKay (1932) we study in Chapter 7 the distribution of this approximation and prove that it is actually noncentral beta distributed. Finally the thesis ends with a discussion in Chapter 8.

2. The statistical model

We intend to model positive observations by assuming that the observations are normally distributed and the coefficient of variation is constant. Before continuing with issues about how to analyse data from such a model we shall discuss the relevance of the model itself.

2.1. Assumptions

At first it may seem contradictory to assume normally distributed observations when the data are genuinely positive. Nevertheless this assumption is often made in practice. Introductory textbooks such as Kleinbaum, Kupper & Muller (1988), explain how positive variables, *e.g.* blood pressure, waiting time or wavelength, can be modelled with assumptions of normal distributions. Such analyses are very informative despite the fact that the model admits negative measurements when only positive values can be obtained. Of course it is assumed that the probability of negative values is very small and negligible. In our setting it is reasonable to require that the coefficient of variation is smaller than 1/3, because the probability

of values deviating from the expected value more than 3 standard deviations is small when the data is normally distributed. With this requirement it is also unlikely that the average by chance is negative or close to zero. We can thus be confident that the sample coefficient of variation, *i.e.* the estimated standard deviation divided by the sample mean, is positive and that it does not explode because of division by zero. In conclusion we can assume a normal distribution provided that the coefficient of variation is smaller than 1/3. The assumption of normal distribution is appropriate if we believe that the data are approximately normally distributed. There is however no need to believe that the model is the actual truth.

There is a strong tradition among statisticians to use the logarithmic transformation when the standard deviation is proportional to the mean. The standard deviation in log values is approximately equal to the coefficient of variation. 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 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. There may be many ways to deal with this problem. In this thesis we investigate the possibility to stick to the assumption of normally distributed additive error and not transform the data into log scale. For comparison we will however include the log method in the example in Section 3.2 and in the simulation study in Chapter 6.

2.2. Application to immunoassays

The model we shall discuss is widely used for analysing immunoassay data. In diagnostic research the coefficient of variation is the predominant measure of dispersion. The data is often approximately normally distributed and the coefficient of variation is often approximately constant over the measuring range. This reality can be explained in the following way.

In immunoassays the amount of some particular substance is measured in a blood sample. Let C denote the concentration of the particular molecules in the sample, and let V denote the volume of the sample. The number N of molecules in

the sample equals VC. The response obtained in an immunoassay could be fluorescence or radioactivity, but if the immunoassay has good performance we can assume that the response is proportional to N and without loss of generality assume that the response equals N. This holds if the calibration curve, *i.e*. the relationship between response and concentration, is linear.

There are many sources of errors in immunoassays. One important source of error is the random variation in the pipetted volume of the sample. It is reasonable to assume that the pipetted volume is not constantly equal to V, but perhaps normally distributed with expected value μ_V and standard deviation σ_V. Then, since we still have $N = VC$,

$$Var(N) = C^2 \sigma_V^2 .$$

However, the result of an immunoassay is not an estimate of the number of the particular molecules in the sample but an estimate of the concentration. This estimate is obtained by division of the response N by the expected volume μ_V (since the true volume V is unknown). As a result the expected value of the estimate of the concentration is

$$\mathbf{E}\left(\frac{N}{\mu_V} \right) = \mathbf{E}\left(\frac{VC}{\mu_V} \right) = \frac{C}{\mu_V} \mathbf{E}(V) = C$$

and the variance in the estimate of the concentration is

$$Var\left(\frac{N}{\mu_V} \right) = C^2 \frac{\sigma_V^2}{\mu_V^2} .$$

Consequently the coefficient of variation in N / μ_V, *i.e.* the coefficient of variation in the estimate of the concentration, is constant and equals $\gamma_V = \sigma_V / \mu_V$. Therefore a model with normally distributed measurement errors and constant coefficient of variation could be adequate for analysing immunoassay data.

As already indicated, in reality immunoassays are not as simple as suggested by the arguments above. There are many sources of errors, not only the variation in the pipetted volume of the sample, but also for example pipetted volumes of reagents, temperature variations, variations in the solid phase on which the specific antibody is attached *etc*. The resulting variance is thus a sum of many variance components, many of which reasonably can be assumed to be approximately normal. Therefore it is not surprising that in immunoassays the final response is often approximately normally distributed.

Many authors have pointed out that the response of an immunoassay can be regarded as a mixture of a Poisson distribution and other variance components. Rodbard *et al*. (1976) writes ‘If there were no experimental errors in the pipetting

of any of the reagents, nor in the separation of the bound and free fractions, such that the response variable was subject only to the counting error caused by the random radioactive decay process for the isotope, than one should have a true Poisson variance for the observed (raw) counts.' Raab (1981) states 'The variation of the response includes the Poisson error of the radioactive counts (usually less than 50 per cent of the total), as well as random errors for each of a series of steps which consist of the addition of reagents, incubation and separation of the bound radioactivity.' They both discuss the radioimmunoassay, in which the response is radioactive counts. However, also if the response is fluorescence we may argue that the response, *i.e.* essentially the number N of particular molecules in the sample, is Poisson distributed. Even if the volume V was always pipetted without variation the number of molecules in the sample would vary between samples. Since each molecule has a small probability to be included in the sample, it is reasonable to assume that the total number N included in the sample is Poisson distributed. Still there is no contradiction between this assumption and the assumption of normally distributed data, because N is usually very large and the Poisson distribution is then well approximated by the normal distribution.

To study this argument a little bit more carefully, assume that N is Poisson distributed with expected value VC. Assume also, as before, that V is normally distributed with expected value μ_V and standard deviation σ_V. Since N is large, N is approximately normally distributed with expected value VC and variance VC. We can write

$$N \overset{d}{\approx} VC + Z\sqrt{VC}$$

where Z is a standardized normal variable independent on V, and d denotes equality in distribution. The expected value of N is

$$\mathbf{E}(N) \approx C\ \mathbf{E}(V) = C\mu_V\,,$$

and the variance is

$$Var(N) \approx \mathbf{E}((VC + Z\sqrt{VC})^2) - (\mathbf{E}(VC + Z\sqrt{VC}))^2$$

$$= \mathbf{E}(V^2C^2 + 2ZVC\sqrt{VC} + Z^2VC) - (\mathbf{E}(VC) + \mathbf{E}(Z\sqrt{VC}))^2$$

$$= \mathbf{E}(V^2C^2) + \mathbf{E}(VC) - (\mathbf{E}(VC))^2$$

$$= C^2\ Var(V) + C\ \mathbf{E}(V)\,.$$

The coefficient of variation in the estimate N / μ_V of the concentration is equal to the coefficient of variation in N. This is because the coefficient of variation is not

affected if the observations are multiplied or divided by a constant. In conclusion, the coefficient of variation in the estimate N / μ_V of the concentration is

$$\frac{\sqrt{C^2\sigma_V^2 + C\mu_V}}{C\mu_V} = \sqrt{\gamma_V^2 + \frac{1}{C\mu_V}} = \sqrt{\gamma_V^2 + \frac{1}{\mathbf{E}(N)}}$$

where $\gamma_V = \sigma_V / \mu_V$, is the coefficient of variation in the sample volume. Note that the squared coefficient of variation in the Poisson error is $1/\mathbf{E}(N)$. The total coefficient of variation is thus the square root of the sum of two variance components. When N is large the coefficient of variation in the estimate of the concentration is approximately constant and equal to the coefficient of variation in the sample volume. Thus once again we conclude that a model with constant coefficient of variation may be appropriate for analysing immunoassay data.

3. Review of inference for coefficients of variation

We shall survey basic statistical methods for inference on the coefficient of variation when the data is normally distributed. Several methods have been proposed for calculating approximate confidence intervals for the coefficient of variation. Usually it is however possible to calculate an exact confidence interval that is finite. A small example will indicate that some of the approximate confidence intervals are better than others. Also various test statistics have been proposed for the hypothesis that two coefficients of variation are equal. The most well known are included in this chapter. Their performances are later investigated in a simulation study in Chapter 6.

3.1. Point estimator

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

$$m = \frac{1}{n}\sum_{j=1}^{n} y_j \,, \quad c = \frac{1}{m}\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(y_j - m)^2} \,. \tag{3.1}$$

Usually c is used as a point estimate of γ. Since the density of m is positive in a neighbourhood of zero the expected value of c does not exist. In applications this is seldom a problem. In many situations, *e.g.* when measuring length or concentration, the measurements can only take positive values, but they may nevertheless be approximately normally distributed. If the probability of a

negative sample coefficient of variation is negligible the expected value of c is, according to Reh & Scheffler (1996),

$$\mathbf{E}(c) \approx \gamma\left(1 - \frac{1}{4n} + \frac{\gamma^2}{n}\right) \approx \gamma\left(1 - \frac{1}{4n}\right),$$

where the second approximate equality (≈) holds if γ is small, which is usually the case. We note that the expected value of $s = cm$ is (Lynch & Walsh, 1997)

$$\mathbf{E}(s) \approx \sigma\left(1 - \frac{1}{4n}\right).$$

The bias of the coefficient of variation is thus of the same magnitude as the bias of the standard deviation.

3.2. Confidence intervals

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 $\tau = n^{1/2}/\gamma$. Owen (1968) discusses this and other applications of the noncentral t distribution. 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 $\Pr(t < n^{1/2}/c \mid \tau = \tau_1) = \alpha/2$ and $\Pr(t > n^{1/2}/c \mid \tau = \tau_2) = \alpha/2$ 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 is commonly the case in practice. The exact finite confidence interval is

$$\left[\frac{\sqrt{n}}{\tau_1}, \frac{\sqrt{n}}{\tau_2}\right]. \tag{3.2}$$

If the percentiles of the noncentral t distribution are not available there are several ways to calculate approximate confidence intervals. When the sample size is not too small asymptotic normality can be used. Miller (1991) shows that c is asymptotically normally distributed with expected value γ and variance

$$\frac{\gamma^2(1+2\gamma^2)}{2(n-1)}. \tag{3.3}$$

Miller & Feltz (1997) suggest that γ^2 in (3.3) are estimated by c^2, and that an approximate 100(1 - α)% confidence interval for γ is calculated as

$$\left[c - z\sqrt{\frac{c^2(1+2c^2)}{2(n-1)}} \quad , \quad c + z\sqrt{\frac{c^2(1+2c^2)}{2(n-1)}} \right], \tag{3.4}$$

where z is the 100(1 – α/2):th percentile of the standard normal distribution. This confidence interval is symmetric around c. An unsymmetrical interval, more likely to perform well also for smaller sample sizes, is obtained if only the second γ^2 in (3.3) is estimated by c^2. Hence

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

is approximately distributed as a standard normal distribution. The corresponding 100(1 - α)% confidence interval for γ is

$$\left[\frac{c}{1+z\sqrt{(1+2c^2)/(2(n-1))}} \quad , \quad \frac{c}{1-z\sqrt{(1+2c^2)/(2(n-1))}} \right]. \tag{3.5}$$

Graf *et al.* (1987) suggest, according to Reh & Scheffler (1996), this approximate confidence interval. Hald (1952) gives another approximate confidence interval based on asymptotic normality.

McKay (1932) shows that if γ is small, *i.e.* less than 1/3, and if

$$\theta = \frac{n-1}{n}$$

then

$$(n-1)\frac{c^2/(1+\theta\, c^2)}{\gamma^2/(1+\gamma^2)} \tag{3.6}$$

is approximately χ^2 distributed with n - 1 degrees of freedom. As explained in Chapter 2 the condition $\gamma < 1/3$ is in practice often reasonable since it makes negative observations unlikely. Since (3.6) is an approximate pivotal quantity it

can be used for calculating an approximate confidence interval (Shao, 2003). This interval can be written

$$\left[\frac{c}{\sqrt{\frac{u_1}{n-1} + c^2\left(\frac{u_1}{n} - 1\right)}} \quad , \quad \frac{c}{\sqrt{\frac{u_2}{n-1} + c^2\left(\frac{u_2}{n} - 1\right)}} \right], \qquad (3.7)$$

where $u_1 = \chi^2_{n\text{-}1,1\text{-}\alpha/2}$ denote the (100(1 - α/2)):th percentile of a χ^2 distribution with n - 1 degrees of freedom, and where $u_2 = \chi^2_{n\text{-}1,\alpha/2}$ denote the (100(α/2)):th percentile of a χ^2 distribution with n - 1 degrees of freedom Of course, as David (1949) points out, we might expect (3.6) to be approximately χ^2 distributed with n - 1 degrees of freedom also if θ = 1. Vangel (1996) derives an optimal choice of θ for calculating quantiles. He finds that

$$\theta = \frac{n-1}{n}\left(\frac{2}{\chi^2_{n-1,\alpha}} + 1\right)$$

is suitable for calculating the (100α):th percentile of the sample coefficient of variation. The confidence interval based on this approximation can be written

$$\left[\frac{c}{\sqrt{\frac{u_1}{n-1} + c^2\left(\frac{u_1+2}{n} - 1\right)}} \quad , \quad \frac{c}{\sqrt{\frac{u_2}{n-1} + c^2\left(\frac{u_2+2}{n} - 1\right)}} \right]. \qquad (3.8)$$

This interval is accurate even for small sample sizes.

Another approximate method, developed by Wong & 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.

It is also possible to use the logarithmic approach discussed in Chapter 2. According to this approach the logarithmic transformation is applied to all measurements and the coefficient of variation in original scale is estimated by the standard deviation in log scale. The usual confidence interval for a standard deviation calculated on log values is thus an approximate confidence interval for the coefficient of variation in the original values. This confidence interval is

$$\left[\sqrt{\frac{(n-1)s^2_{\log y}}{\chi^2_{n-1,1-\alpha/2}}} \quad , \quad \sqrt{\frac{(n-1)s^2_{\log y}}{\chi^2_{n-1,\alpha/2}}} \right] \tag{3.9}$$

where $s_{\log y}$ is the standard deviation calculated on log values.

In the following example we also include the 'naive' interval

$$\left[\sqrt{\frac{(n-1)c^2}{\chi^2_{n-1,1-\alpha/2}}} \quad , \quad \sqrt{\frac{(n-1)c^2}{\chi^2_{n-1,\alpha/2}}} \right]. \tag{3.10}$$

This interval is obtained if the coefficient of variation is treated as if it was a proper standard deviation, *i.e.* if the limits of the usual confidence interval for σ^2 are divided by the average m. Vangel (1996) compares analytically the errors in this naive approximation with the error in the McKay approximation and conclude that the naive approximation is 'substantially less accurate.'

As an example we calculate the presented confidence intervals on the tensile strength data given by Vangel (1996). We obtain the exact confidence interval using the function *tnonct* in Release 9.1 of the SAS System (SAS Institute Inc., Cary, NC, USA). The data, presented in Table 3.1, consists of measurements on five specimens of a composite material.

Table 3.1. *Tensile strength data*

Specimen	Tensile strength (1000 psi)
1	326
2	302
3	307
4	299
5	329
Mean (1000 psi)	312.6
Coefficient of variation	0.0446

The calculated confidence intervals are given in Table 3.2.

Table 3.2. *95% confidence intervals for the coefficient of variation in the tensile strength data*

Method (Formula)	Confidence interval
Exact (3.2)	[0.0267 , 0.1287]
Miller & Feltz (3.4)	[0.0136 , 0.0756]
Graaf *et al.* (3.5)	[0.0263 , 0.1459]
McKay (3.7)	[0.0267 , 0.1291]
Vangel (3.8)	[0.0267 , 0.1287]
Log (3.9)	[0.0266 , 0.1274]
Naive (3.10)	[0.0267 , 0.1281]

In this dataset the estimated coefficient of variation is not very large (4.46%). Most of the approximate confidence intervals are similar to the exact confidence interval. The symmetric Miller & Feltz confidence interval does however not perform well. The confidence interval suggested by Graf *et al.* (1987), based on the same normal approximation, is much better. The McKay approximation seems to be very accurate, and the modification due to Vangel (1996) is successful. The method of logarithmic transformation works fine in this example but is not as good as the McKay approximation. The naive method, finally, performs well, but the interval is a little bit too narrow since it ignores the variation in the estimate of the average.

3.3. Tests for equality of two coefficients of variation

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

3.3.1. Likelihood ratio test

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

$$\hat{\mu}_1 = \frac{n_1 m_1 \hat{\mu}_2}{(n_1 + n_2)\hat{\mu}_2 - n_2 m_2},$$

$$\hat{\mu}_2 = -\frac{q}{2p} + \sqrt{\frac{q^2}{4p^2} - \frac{r}{p}}$$

and

$$\hat{\gamma} = \sqrt{\frac{\frac{n_2 - 1}{n_2} c_2^2 m_2^2 + m_2^2 - m_2 \hat{\mu}_2}{\hat{\mu}_2}}, \qquad (3.11)$$

respectively, where m_i denotes the average response in sample i, and where

$$p = (n_1 + n_2)c_1^2 + n_2,$$

$$q = -(2n_2c_1^2 + (2n_2 - n_1))m_2$$

and

$$r = \frac{(n_2^2(c_1^2+1) - n_1^2(c_2^2+1))m_2^2}{n_1 + n_2}.$$

The likelihood ratio test statistic can be written

$$R = -2\log\lambda = n_1 \log \frac{\left(\hat{\gamma}\,\hat{\mu}_1\right)^2}{\left(\frac{n_1-1}{n_1}\right)c_1^2 m_1^2} + n_2 \log \frac{\left(\hat{\gamma}\,\hat{\mu}_2\right)^2}{\left(\frac{n_2-1}{n_2}\right)c_2^2 m_2^2}, \quad (3.12)$$

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

3.3.2 Bennett's test

Bennett (1976) utilise the approximation (3.6) given by McKay (1932) and applies a test according to Pitman (1939) of the hypothesis of equal scale parameters of gamma variables. Shafer & Sullivan (1986) note that Bennett by mistake uses a variance with devisor n - 1 where McKay in his article uses a variance with devisor n. For this reason they modify Bennett's test correspondingly. The modified Bennett's test statistic is

$$B = (n_1 + n_2 - 2)\log\left(\frac{\frac{n_1\theta_1c_1^2}{1+\theta_1c_1^2} + \frac{n_2\theta_2c_2^2}{1+\theta_2c_2^2}}{n_1+n_2-2}\right)$$

$$-(n_1-1)\log\left(\frac{\frac{n_1\theta_1c_1^2}{1+\theta_1c_1^2}}{n_1-1}\right) - (n_2-1)\log\left(\frac{\frac{n_2\theta_2c_2^2}{1+\theta_2c_2^2}}{n_2-1}\right), \quad (3.13)$$

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.

3.3.3. 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, γ_W, of the sample coefficients of variation:

$$\gamma_W = \frac{(n_1 - 1)c_1 + (n_2 - 1)c_2}{n_1 + n_2 - 2}.$$

This estimate is employed in the calculation of a test statistic

$$M = \frac{c_1 - c_2}{\sqrt{\frac{\gamma_W^2}{2(n_1 - 1)} + \frac{\gamma_W^4}{n_1 - 1} + \frac{\gamma_W^2}{2(n_2 - 1)} + \frac{\gamma_W^4}{n_2 - 1}}}, \qquad (3.14)$$

which is compared with a standard normal distribution.

3.3.4. Wald test

The Wald statistic given by Rao & Vidya (1992) for the case of equal sample sizes is modified to the general case of unequal sample sizes by Gupta & Ma (1996). The statistic

$$W = \frac{(c_1 - c_2)^2}{\frac{c_1^2}{2n_1} + \frac{c_1^4}{n_1} + \frac{c_2^2}{2n_2} + \frac{c_2^4}{n_2}} \qquad (3.15)$$

is approximately χ^2 distributed with 1 degree of freedom. This test statistic is obviously closely related to Miller's statistic (3.14). Bhoj & Ahsanullah (1993) give a third statistic on the same theme, but only for the case of equal sample sizes.

3.3.5. Score test

Gupta & Ma (1996) derive the score test, based on the maximum likelihood estimate (3.11). Its explicit value is given by

$$S = \left(\frac{1}{2} \left(\hat{\gamma} \right)^2 + \left(\hat{\gamma} \right)^4 \right) \left(\frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} \right), \qquad (3.16)$$

where

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

The test statistic (3.16) shall be compared with a χ^2 distribution with 1 degree of freedom.

3.3.6. Doornbos & Dijkstra's test

Doornbos & Dijkstra (1983) develop a test based on the distribution of the inverse of the sample coefficient of variation. Let $b_i = 1 / c_i$, and let b_W denote a weighted average of b_1 and b_2:

$$b_W = \frac{n_1 b_1 + n_2 b_2}{n_1 + n_2}.$$

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 & Dijkstra estimate the expectation of T by

$$\hat{\mathbf{E}}(T) = \frac{n_2(n_1 - 1)}{(n_1 + n_2)(n_1 - 3)} + \frac{n_1(n_2 - 1)}{(n_1 + n_2)(n_2 - 3)}$$

$$+ \frac{1}{c_p^2}\left[\frac{n_1 n_2(n_1 - 1)}{(n_1 + n_2)(n_1 - 3)} + \frac{n_1 n_2(n_2 - 1)}{(n_1 + n_2)(n_2 - 3)}\right.$$

$$\left. + \frac{1}{n_1 + n_2}(n_1^2 e_1^2 + n_2^2 e_2^2 - (n_1 e_1 + n_2 e_2)^2)\right]$$

where c_p is an estimate of the common population coefficient of variation γ:

$$c_p^2 = \frac{\frac{n_1(n_1 - 1)}{n_1 - 3} + \frac{n_2(n_2 - 1)}{n_2 - 3}}{n_1 b_1^2 + n_2 b_2^2 - \left(\frac{n_1 - 1}{n_1 - 3} + \frac{n_2 - 1}{n_2 - 3}\right)},$$

and

$$e_i = \left(\frac{n_i - 1}{2}\right)^{\frac{1}{2}} \frac{\Gamma\left[\frac{1}{2}(n_i - 2)\right]}{\Gamma\left[\frac{1}{2}(n_i - 1)\right]}.$$

The test statistic

$$D = \frac{T}{\hat{\mathbf{E}}(T)} \tag{3.17}$$

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

3.3.7 Log test

A test based on the logarithmic approach discussed in Chapter 2 can be made in the following way. Take the logarithm of all observations and calculate the standard deviation s_1 (log y) in sample 1 and the standard deviation s_2 (log y) in sample 2. Then compare

$$L = \frac{s_1^2(\log y)}{s_2^2(\log y)} \tag{3.18}$$

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

3.3.8. Naive test

With the 'naive' test we mean comparing the sample coefficients of variation by an F-test in the same way as proper standard deviations are compared. Thus

$$N = \frac{c_1^2}{c_2^2} \tag{3.19}$$

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

4. An approximate F-test

We shall now develop an approximate F-test for the hypothesis of equality of two coefficients of variation. In our search for a constructive statistical method 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 (3.19). This test does not take into account the variation in the estimated averages in the denominators of the coefficients of variation. Therefore it is natural to suppose that it is better to use the ratio between two McKay transformations.

4.1. The idea of the approximate F-test

Let again $y_{ij} = \mu_i + e_{ij}$, where e_{ij} are independently distributed $N(0,(\gamma_i \mu_i)^2)$, $i = 1, 2$ and $j = 1, 2, \ldots, n_i$, with positive population coefficients of variation γ_i and positive expected values μ_i. Assume also that $\gamma_i < 1/3$. Then we know from (3.6) that

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

is approximately χ^2 distributed with n_i - 1 degrees of freedom when $\theta = (n_i - 1) / n_i$. Consequently we can, if H_0: $\gamma_1 = \gamma_2$ is true, anticipate

$$F = \frac{(c_1^2 \big/ (1 + \theta(n_1)c_1^2)) \big/ (\gamma_1^2 \big/ (1 + \gamma_1^2))}{(c_2^2 \big/ (1 + \theta(n_2)c_2^2)) \big/ (\gamma_2^2 \big/ (1 + \gamma_2^2))} = \frac{c_1^2 \big/ (1 + \theta(n_1)c_1^2)}{c_2^2 \big/ (1 + \theta(n_2)c_2^2)} \qquad (4.1)$$

to be approximately F distributed with n_1 - 1 and n_2 - 1 degrees of freedom. In (4.1) θ is a function of n_i, which not necessarily equals $(n_i - 1) / n_i$. Since $c^2/(1 + \theta\, c^2)$ is an increasing function of c, 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.

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 (Chapter 5). We shall also perform a simulation study (Chapter 6).

At first we shall investigate the possibility that $(n_i - 1) / n_i$ is not the best choice of $\theta(n_i)$ for the F-test. This issue is motivated because David (1949) noted that (3.6) is approximately χ^2 distributed also if $\theta = 1$, and Vangel (1996) showed that

$$\theta = \frac{n-1}{n}\left(\frac{2}{\chi^2_{n-1,\alpha}} + 1\right)$$

is an optimal choice for calculating confidence intervals (see Section 3.2). Note also that $\theta(n_i) = 0$ in (4.1) gives the 'naive' F-test (3.19). According to this test the coefficients of variation are analysed as if they were standard deviations. No account is made for the variation in the averages.

We shall look for a function $\theta_i = \theta(n_i)$ such that F is approximately F distributed when the null hypothesis is true. This search will be made by comparing the moments of F with the moments of an F distributed random variable. Primarily we want the first moments to be as similar as possible when the sample sizes, *i.e.* n_1 and n_2, are equal. It is a difficulty that the moments depend on the unknown coefficients of variation γ_1 and γ_2, which under the null hypothesis are equal to the unknown common coefficient of variation γ. We shall make a Taylor series expansion of F about $\gamma = 0$, because in applications γ is often small (see Chapter 2 for a discussion of the assumption that $\gamma < 1/3$).

Finally we require that θ in the numerator of (4.1) is the same function as θ in the denominator of (4.1). Different functions would result in different inferential conclusions depending on which sample was considered as sample 1 and which sample was considered as sample 2. When the functions are the same, however, the inference is not influenced, because the 100(1 - $\alpha/2$):th percentile of the F distribution with v_1 and v_2 degrees of freedom is equal to the inverse of the 100($\alpha/2$):th percentile of the F distribution with v_2 and v_1 degrees of freedom.

4.2. Development of the approximate F-test

Let W denote a χ^2 distributed random variable divided by its degrees of freedom, and let Z denote a standardized normal random variable. Then, for the average m given in (3.1),

$$m \overset{d}{=} \mu + Z\frac{\sigma}{\sqrt{n}}$$

and

$$s = \sqrt{\frac{1}{(n-1)}\sum_{j=1}^{n}(y_j - m)^2} \overset{d}{=} \sigma\sqrt{W},$$

where $\sigma = \gamma\mu$ and d denotes equality in distribution. With this notation we can write the squared sample coefficient of variation

$$c^2 = \frac{W\gamma^2}{\left(1 + \frac{Z\gamma}{\sqrt{n}}\right)^2},$$

which turns (4.1) into

$$F \stackrel{d}{=} \frac{\dfrac{1}{W_2}\left(1+\dfrac{Z_2\gamma}{\sqrt{n_2}}\right)^2 + \theta_2\gamma^2}{\dfrac{1}{W_1}\left(1+\dfrac{Z_1\gamma}{\sqrt{n_1}}\right)^2 + \theta_1\gamma^2},$$

where we have written θ_1 and θ_2 instead of $\theta(n_1)$ and $\theta(n_2)$ respectively. Note that W_1, W_2, Z_1 and Z_2 are independent. Also recall that W_1 / W_2 is F distributed with $n_1 - 1$ and n_2 - 1 degrees of freedom.

Since γ is often small in applications it makes sense to expand F in a Taylor series about $\gamma = 0$. Thus, in a neighbourhood of $\gamma = 0$,

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

$$+ \frac{W_1}{W_2}\left(3\frac{Z_1^2}{n_1} - 4\frac{Z_1Z_2}{\sqrt{n_1n_2}} + \frac{Z_2^2}{n_2} - \theta_1W_1 + \theta_2W_2\right)\gamma^2$$

$$+ \frac{W_1}{W_2}\left(-4\frac{Z_1^3}{n_1\sqrt{n_1}} + 6\frac{Z_1^2Z_2}{n_1\sqrt{n_2}} - 2\frac{Z_1Z_2^2}{\sqrt{n_1}n_2} + \left(4\frac{\theta_1W_1}{\sqrt{n_1}} - 2\frac{\theta_2W_2}{\sqrt{n_1}}\right)Z_1 - 2\frac{\theta_1W_1Z_2}{\sqrt{n_2}}\right)\gamma^3$$

$$+ \frac{W_1}{W_2}\left(5\frac{Z_1^4}{n_1^2} - 8\frac{Z_1^3Z_2}{n_1\sqrt{n_1n_2}} + 3\frac{Z_1^2Z_2^2}{n_1n_2} + \left(3\frac{\theta_2W_2}{n_1} - 10\frac{\theta_1W_1}{n_1}\right)Z_1^2 + 8\frac{\theta_1W_1Z_1Z_2}{\sqrt{n_1n_2}}\right.$$

$$\left. - \frac{\theta_1W_1Z_2^2}{n_2} + \theta_1W_1(\theta_1W_1 - \theta_2W_2)\right)\gamma^4 + O(\gamma^5). \quad (4.2)$$

Since $\mathbf{E}(Z_1) = \mathbf{E}(Z_2) = \mathbf{E}(Z_1Z_2) = \mathbf{E}(Z_1^3) = 0$ and $\mathbf{E}(Z_1^2) = \mathbf{E}(Z_2^2) = 1$, the expectation of F, in a neighbourhood of $\gamma = 0$, can be written

$$\mathbf{E}(F) = \mathbf{E}\left[\mathbf{E}(F \mid W_1, W_2)\right]$$

$$= \mathbf{E}\left(\frac{W_1}{W_2}\right) + \gamma^2\mathbf{E}\left[\frac{W_1}{W_2}\left(\frac{3}{n_1} + \frac{1}{n_2} - \theta_1W_1 + \theta_2W_2\right)\right] + O(\gamma^4),$$

where $\mathbf{E}(F \mid W_1, W_2)$ denotes the conditional expectation of F given W_1 and W_2.

Furthermore, since the expectation of an F distribution is $\mathbf{E}(W_1 / W_2) = (n_2 - 1) / (n_2 - 3)$ when $n_2 > 3$, and since $\mathbf{E}(W_1^2) = (n_1 + 1)(n_1 - 1)$ we have

$$\mathbf{E}\left[\left(\frac{3}{n_1}+\frac{1}{n_2}\right)\frac{W_1}{W_2}-\theta_1\frac{W_1^2}{W_2}+\theta_2 W_1\right]$$

$$=\left(\frac{3}{n_1}+\frac{1}{n_2}\right)\frac{n_2-1}{n_2-3}-\theta_1\frac{(n_2-1)(n_1+1)}{(n_2-3)(n_1-1)}+\theta_2$$

when $n_2 > 3$. If we now let

$$\theta_i = \frac{n_i - 1}{n_i}, \quad i = 1, 2,$$

the expectation of F is, in a neighbourhood of $\gamma = 0$,

$$\mathbf{E}(F) = \mathbf{E}\left(\frac{W_1}{W_2}\right) + 2\frac{n_2-1}{n_2-3}\left(\frac{1}{n_1}-\frac{1}{n_2}\right)\gamma^2 + O(\gamma^4), \qquad n_2 > 3.$$

Obviously the expectation of F is then close to the expectation of an F distribution. This is true especially in the balanced case, *i.e.* when $n_1 = n_2 = n$, but also otherwise since γ is small. In the balanced case we get from (4.2), since $\mathbf{E}(Z_1^4) = 3$ and $\mathbf{E}(W_1^3) = (n_1 + 1)(n_1 + 3) / (n_1 - 1)^2$, that

$$\mathbf{E}(F) = \mathbf{E}\left(\frac{W_1}{W_2}\right) - 2\frac{n^2-3n+2}{n^2(n-3)}\gamma^4 + O(\gamma^5).$$

It is proved in Chapter 5 that the r:th moment of F is

$$\mathbf{E}(F^r) = \mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right] + 2r\,\mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right]\left(\frac{2-r}{n_1}-\frac{r}{n_2}\right)\gamma^2 + O(\gamma^3), \quad (4.3)$$

provided that $n_2 > 2r + 1$. We conclude that the original choice of θ, suggested by McKay (1932), is indeed appropriate also for the F statistic. Thus we introduce the following test statistic for testing equality of two coefficients of variation.

Definition 4.1. Let there be two samples. Let y_{ij} denote the j:th observation in sample i. Let n_i denote the number of observations, m_i the average and c_i the coefficient of variation in the i:th sample:

$$m_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}\ , \qquad c_i = \frac{1}{m_i}\sqrt{\frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - m_i)^2}\ , \quad i = 1, 2.$$

The statistic $\boldsymbol{F}$ is defined as

$$F = \frac{c_1^2\left(1 + \dfrac{n_2 - 1}{n_2}c_2^2\right)}{c_2^2\left(1 + \dfrac{n_1 - 1}{n_1}c_1^2\right)}. \quad \blacklozenge \tag{4.4}$$

When $n_2 < 4$, the expected value is infinitely large for F as well as for W_1 / W_2. The expected value of F approximately equals the expected value of an F distribution with n_1 - 1 and n_2 - 1 degrees of freedom, even if the sample sizes are small. In case of large sample sizes, F is a quotient between two accurate approximations of χ^2 random variables divided by their degrees of freedom. We can thus anticipate F to work well as a test statistic for the null hypothesis. As mentioned earlier the conformity with the F distribution is studied theoretically in Chapter 5 and by simulation in Chapter 6.

4.3. A generalized approximate F-test

To generalize the ideas of Section 4.1 and 4.2 assume that $y_{ijk} = \mu_{ij} + e_{ijk}$, where e_{ijk} are independently distributed $N(0,(\gamma_i\,\mu_{ij})^2)$ with $0 < \mu_{ij}$ and $0 < \gamma_i < 1/3$, $i = 1, 2$, $j = 1, 2, \ldots, r_i$; $k = 1, 2, \ldots, n_{ij}$. Thus we now assume that we have r_i independent estimates c_{ij} of γ_i instead of only one. The estimates are independent, because the distribution of c_{ij} does not depend on μ_{ij}. Then, with $\theta_{ij} = (n_{ij} - 1) / n_{ij}$,

$$\frac{\sum_{j=1}^{r_i}(n_{ij} - 1)c_{ij}^2 \big/ (1 + \theta_{ij}c_{ij}^2)}{\gamma_i^{\,2} \big/ (1 + \gamma_i^{\,2})}$$

is approximately χ^2 distributed with $\Sigma_j n_{ij} - r_i$ degrees of freedom. If H_0: $\gamma_1 = \gamma_2$ is true,

$$\frac{\left(\sum_{j=1}^{r_2} n_{2j} - r_2\right) \sum_{j=1}^{r_1} \frac{(n_{1j}-1)c_{1j}^2}{1+\theta_{1j}c_{1j}^2}}{\left(\sum_{j=1}^{r_1} n_{1j} - r_1\right) \sum_{j=1}^{r_2} \frac{(n_{2j}-1)c_{2j}^2}{1+\theta_{2j}c_{2j}^2}} \tag{4.5}$$

is approximately F distributed with $\Sigma_j n_{1j} - r_1$ and $\Sigma_j n_{2j} - r_2$ degrees of freedom. This motivates the following definition.

Definition 4.2. Let there be two sets of samples. Let y_{ijk} denote the k:th observation in the j:th sample in set i. Let r_i denote the number of samples in the i:th set. Let n_{ij} denote the number of observations, m_{ij} the average and c_{ij} the coefficient of variation in the j:th sample in the i:th set:

$$m_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}\,, \quad c_{ij} = \frac{1}{m_{ij}} \sqrt{\frac{1}{n_{ij}-1} \sum_{k=1}^{n_{ij}} (y_{ijk} - m_{ij})^2}\,, \quad i = 1, 2.$$

The statistic $\boldsymbol{G}$ is defined as

$$G = \frac{\left(\sum_{j=1}^{r_2} n_{2j} - r_2\right) \sum_{j=1}^{r_1} \frac{(n_{1j}-1)c_{1j}^2}{1+\theta_{1j}c_{1j}^2}}{\left(\sum_{j=1}^{r_1} n_{1j} - r_1\right) \sum_{j=1}^{r_2} \frac{(n_{2j}-1)c_{2j}^2}{1+\theta_{2j}c_{2j}^2}}\,. \quad \blacklozenge \tag{4.6}$$

4.4. An immunoassay example

Brunnée *et al.* (1996) compares two methods for measuring concentration of specific IgE antibodies in blood samples. A new system, ELItest, is compared with the established Pharmacia CAP system (PCS). Among other things the variations between and within assays are studied. The specific IgE for the allergens mite, cat and birch is measured for 3 sera with very different levels of concentration. The *inter* assay coefficients of variation are calculated on 10 measurements made on different days, and the *intra* assay coefficients of variation are calculated on 8 measurements performed on the same day. 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 5.1 together with calculated approximate F-tests (4.4). 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 coefficients of variation for ELItest (18.6 %) is more

than twice as large as the estimate of the coefficient of variation for 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.

Table 5.1. *The approximate F-test based on (4.4) applied to intra assay coefficients of variation (CV) reported by Brunnée* et al. *(1996)*

Allergen	ELItest CV (%) (n = 8)	PCS CV (%) (n = 8)	F	Probability 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

If we assume that each method has constant intra assay coefficients of variation we can apply the generalized test statistic given in (4.6). 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 (probability 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 for 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 (probability value 0.0020).

Table 5.2. *The approximate F-test based on (4.4) applied to inter assay coefficients of variation (CV) reported by Brunnée* et al. *(1996)*

Allergen	ELItest CV (%) (n = 10)	PCS CV (%) (n = 10)	F	Probability 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

Table 5.2 includes the *inter* assay coefficients of variation as reported by Brunnée *et al.* (1996) and the corresponding results of the proposed approximate

F-test given in (4.4). 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.

Under assumption that each method has a constant inter coefficient of variation the generalized test statistic G given in (4.6) equals 3.265. It shall be compared with an F distribution with 63 degrees of freedom in the numerator and 54 degrees of freedom in the denominator. In conclusion the inter coefficient of variation is significantly larger in ELItest than in Pharmacia CAP System (probability value 0.00002).

5. Properties of the approximate F-test

In this chapter we shall theoretically study the properties of the statistic F given in (4.4). 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. We shall also express the random variable $\log F$ as a sum of $\log X$ and some error variables that are in probability of small orders.

To begin with we give a lemma that is useful when comparing the moments of F and X. We already know the lemma from the development of the approximate F-test in the beginning of Section 4.2, but state it here in a complete formulation with regard to the statistic F defined not until the end of the same section, in Definition 4.1.

Lemma 5.1. Let $y_{ij} = \mu_i + e_{ij}$, where e_{ij} are independently distributed $N(0,(\gamma \mu_i)^2)$, $i = 1, 2$; $j = 1, 2, \ldots, n_i$, with positive population coefficient of variation γ and positive expected values μ_i. Let W_1 and W_2 denote independent χ^2 distributed random variables divided by their degrees of freedom, and let Z_1 and Z_2 denote independent standardized normal random variables. Then the distribution of F, as defined in Definition 4.1, equals the distribution of

$$\frac{\frac{1}{W_2}\left(1+\frac{Z_2\gamma}{\sqrt{n_2}}\right)^2+\frac{n_2-1}{n_2}\gamma^2}{\frac{1}{W_1}\left(1+\frac{Z_1\gamma}{\sqrt{n_1}}\right)^2+\frac{n_1-1}{n_1}\gamma^2}.$$

Proof. The averages m_i, given in Definition 4.1, equals

$$\mu_i + Z_i \frac{\sigma_i}{\sqrt{n_i}}, \quad i = 1, 2,$$

in distribution, and the standard deviation $c_i\, m_i$ equals

$$\gamma\mu_i\sqrt{W_i}, \quad i = 1, 2,$$

in distribution. Thus the distribution of c_i^2 equals the distribution of

$$\frac{W_i\gamma^2}{\left(1+\frac{Z_i\gamma}{\sqrt{n_i}}\right)^2}, \qquad i = 1, 2, \tag{5.1}$$

which inserted in (4.4) makes the lemma. ♦

The following theorem provides approximate differences between the moments of the statistic F, as defined by Definition 4.1, and the moments of an F distributed random variable. We conclude that the moments are similar when the coefficient of variation is small, especially if the sample sizes are equal or large.

Theorem 5.1. Let $y_{ij} = \mu_i + e_{ij}$, where e_{ij} are independently distributed $N(0\ ,(\gamma\,\mu_i)^2)$, $i = 1, 2$ and $j = 1, 2, \ldots, n_i$, with positive population coefficient of variation γ and positive expected values μ_i. Let X be distributed $F(n_1 - 1, n_2 - 1)$. Then the r:th moment of F is

$$\mathbf{E}(F^r) = \mathbf{E}(X^r) + 2r\ \mathbf{E}(X^r)\left(\frac{2-r}{n_1} - \frac{r}{n_2}\right)\gamma^2 + O(\gamma^3), \tag{5.2}$$

provided that $n_2 > 2r + 1$.

Proof. According to Lemma 5.1

$$F^r \overset{d}{=} \left(\frac{\frac{1}{W_2}\left(1+\frac{Z_2\gamma}{\sqrt{n_2}}\right)^2 + \theta_2\gamma^2}{\frac{1}{W_1}\left(1+\frac{Z_1\gamma}{\sqrt{n_1}}\right)^2 + \theta_1\gamma^2}\right)^r, \qquad r = 1, 2, 3, \ldots$$

where d denotes equality in distribution, and $\theta_i = (n_i - 1) / n_i$, $i = 1, 2$. By a Taylor series expansion about $\gamma = 0$,

$$\mathbf{E}(F^r) = \mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right] + \frac{1}{2}\mathbf{E}\left[2r\,(r-1)\left(\frac{W_1}{W_2}\right)^{r-1}\left(\frac{Z_2}{\sqrt{n_2}} - \frac{Z_1}{\sqrt{n_1}}\right)\right.$$

$$\left. + 2r\left(\frac{W_1}{W_2}\right)^r\left(3\frac{Z_1^2}{n_1} - 4\frac{Z_1 Z_2}{\sqrt{n_1 n_2}} + \frac{Z_2^2}{n_2} - \theta_1 W_1 + \theta_2 W_2\right)\right]\gamma^2 + O(\gamma^3).$$

The r:th moment of an F distribution with n_1 - 1 and n_2 - 1 degrees of freedom is (Kotz & Johnson, 1983)

$$\mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right] = \left(\frac{n_2 - 1}{n_1 - 1}\right)^r \frac{\Gamma\left(\frac{n_1 - 1}{2} + r\right)\,\Gamma\left(\frac{n_2 - 1}{2} - r\right)}{\Gamma\left(\frac{n_1 - 1}{2}\right)\,\Gamma\left(\frac{n_2 - 1}{2}\right)},\ n_2 > 2r + 1.$$

It is noted that the $(r - 1)$:th moment can be written

$$\mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^{r-1}\right] = \frac{(n_1 - 1)(n_2 - 2r - 1)}{(n_2 - 1)(n_1 + 2r - 3)}\mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right],\ n_2 > 2r + 1.$$

The r:th moment of a χ^2 distribution with n - 1 degrees of freedom is (Kotz & Johnson, 1982)

$$\mathbf{E}\left[((n-1)W)^r\right] = 2^r \frac{\Gamma\left(\frac{n-1}{2} + r\right)}{\Gamma\left(\frac{n-1}{2}\right)}. \tag{5.3}$$

Moreover, because W_1 and W_2 are independent,

$$\mathbf{E}\left(\frac{1}{W_2^r}\right) = \frac{\mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right]}{\mathbf{E}(W_1^r)} = \left(\frac{n_2 - 1}{2}\right)^r \frac{\Gamma\left(\frac{n_2 - 1}{2} - r\right)}{\Gamma\left(\frac{n_2 - 1}{2}\right)}$$

and

$$\mathbf{E}\left(\frac{W_1^{r+1}}{W_2^r}\right) = \mathbf{E}\left(W_1^{r+1}\right)\ \mathbf{E}\left(\frac{1}{W_2^r}\right)$$

$$= \frac{2(n_2-1)^r\ \ \Gamma\left(\frac{n_1+1}{2}+r\right)\ \Gamma\left(\frac{n_2-1}{2}-r\right)}{(n_1-1)^{r+1}\ \ \Gamma\left(\frac{n_1-1}{2}\right)\ \Gamma\left(\frac{n_2-1}{2}\right)}$$

$$= \frac{n_1+2r-1}{n_1-1}\ \ \mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right].$$

As a result, the r:th moment of F can be written

$$\mathbf{E}(F^r) = \mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right]$$

$$+ r\left(\frac{3}{n_1}+\frac{1}{n_2}-\theta_1\frac{n_1+2r-1}{n_1-1}+\theta_2\frac{n_2-2r-1}{n_2-1}\right)\ \mathbf{E}\left[\left(\frac{W_1}{W_2}\right)^r\right]\ \gamma^2$$

$$+ O(\gamma^3),\ \ n_2 > 2r+1,$$

in a neighbourhood of $\gamma = 0$. Finally $\theta_i = (n_i - 1) / n_i$, $i = 1, 2$, produces (5.2). ♦

We do not only want to compare the moments of F with the moments of an F distributed random variable X. We also want to compare F with X in a more straightforward way. Since F is a ratio of two independent χ^2 approximations it is 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 $(\log X) / 2$.

Before comparing the distributions we need to make clear the concept 'probability of order'. Recall that if $\{r_n\}$ and $\{s_n\}$ are sequences of real numbers, then r_n is said to be of order $O(s_n)$ if

$$\lim_{n\to\infty} \left|\frac{r_n}{s_n}\right| < \infty .$$

The order in probability is an extension of this concept. It can be defined in the following way (see Azzalini, 1996).

Definition 5.1. Let $\{X_n\}$ be a sequence of random variables and let $\{r_n\}$ be a sequence of positive real numbers. Then we say that X_n is **in probability of order $O(r_n)$**, written $X_n = O_p(r_n)$, if for all $\varepsilon > 0$ there exists a real number M_ε such that

$$\Pr\left(\left|\frac{X_n}{r_n}\right| > M_\varepsilon\right) < \varepsilon$$

for all n greater than N_ε. ♦

We will also use two theorems from the book by Azzalini (1996). We state them here for a quick reference, but refer to Azzalini (1996) for the proofs. According to the first theorem it suffices to check the second moments to determine the order in probability of a sequence of random variables.

Theorem 5.2. Let $\{X_n\}$ be a sequence of random variables with $\mathbf{E}(X_n^2) = r_n^2 < \infty$, and $\{s_n\}$ a sequence of positive reals. Then, if r_n^2 is $O(s_n^2)$,

$$X_n = O_p(s_n) . \quad ♦$$

The second theorem from Azzalini (1996) states that a product of random variables is in probability of the same order as the product of the order in probabilities of the random variables. Furthermore, a sum of random variables is in probability of the same order as the largest order in probability of the random variables.

Theorem 5.3. Let $\{r_n\}$ and $\{s_n\}$ be two sequences of positive real numbers and $\{X_n\}$, $\{Y_n\}$ be two sequences of random variables. Then, if $X_n = O_p(r_n)$ and $Y_n = O_p(s_n)$,

i) $X_n Y_n = O_p(r_n s_n)$

ii) $X_n + Y_n = O_p(\max(r_n, s_n))$. ♦

We are now ready to compare the distribution of the logarithm of F with the distribution of the logarithm of an F distributed random variable X.

Theorem 5.4. Let $y_{ij} = \mu_i + e_{ij}$, where e_{ij} are independently distributed N(0 , $(\gamma \mu_i)^2$), $i = 1, 2$ and $j = 1, 2, \ldots, n_i$, with positive population coefficient of variation γ and positive expected values μ_i. Let X be F($n_1 - 1$, $n_2 - 1$) distributed, let Z be N(0, 1) distributed, and let U_i be χ^2 distributed with $n_i - 1$ degrees of freedom, $i = 1, 2$. Let X, Z, U_1 and U_2 be independent. Then

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

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 theorem implies that the distribution of log F equals the distribution of $\log X + O_p(\max(n_1^{-1/2} \gamma , n_2^{-1/2} \gamma , \gamma^2))$.

Proof. Write log F as

$$\log F = \log c_1^2\left(1+\frac{n_1-1}{n_1}c_1^2\right)^{-1} - \log c_2^2\left(1+\frac{n_2-1}{n_2}c_2^2\right)^{-1}. \qquad (5.4)$$

The first term is by (5.1)

$$\log c_1^2\left(1+\frac{n_1-1}{n_1}c_1^2\right)^{-1} \overset{d}{=} \log\frac{W_1\gamma^2}{\left(1+\frac{Z_1\gamma}{\sqrt{n_1}}\right)^2} - \log\left(1+\frac{n_1-1}{n_1}\frac{W_1\gamma^2}{\left(1+\frac{Z_1\gamma}{\sqrt{n_1}}\right)^2}\right)$$

$$= \log W_1\gamma^2 - \log\left(\left(1+\frac{Z_1\gamma}{\sqrt{n_1}}\right)^2 + \frac{n_1-1}{n_1}W_1\gamma^2\right)$$

$$= \log W_1 + \log\gamma^2 - \log\left(1+2\frac{Z_1\gamma}{\sqrt{n_1}} + \frac{Z_1^2\gamma^2}{n_1} + \frac{n_1-1}{n_1}W_1\gamma^2\right),$$

where W_1 denotes a χ^2 distributed random variable divided by its degrees of freedom, and Z_1 denotes a standardized normal random variable. Expansion of the last term yields

$$\log\left(1+2\frac{Z_1\gamma}{\sqrt{n_1}}+\frac{Z_1^2\gamma^2}{n_1}+\frac{n_1-1}{n_1}W_1\gamma^2\right)$$

$$=2\frac{Z_1\gamma}{\sqrt{n_1}}+\frac{Z_1^2\gamma^2}{n_1}+\frac{n_1-1}{n_1}W_1\gamma^2-\frac{1}{2}\left(2\frac{Z_1\gamma}{\sqrt{n_1}}+\frac{Z_1^2\gamma^2}{n_1}+\frac{n_1-1}{n_1}W_1\gamma^2\right)^2$$

$$+\frac{1}{3}\left(2\frac{Z_1\gamma}{\sqrt{n_1}}+\frac{Z_1^2\gamma^2}{n_1}+\frac{n_1-1}{n_1}W_1\gamma^2\right)^3\ \ldots \qquad (5.5)$$

which consists of terms of the form

$$\frac{Z_1^k\gamma^k}{n_1^{k/2}},\quad k=1,2,\ldots \qquad (5.6)$$

and

$$\left(\frac{n_1-1}{n_1}\right)^k W_1^k\gamma^{2k},\quad k=1,2,\ldots \qquad (5.7)$$

and of terms that are a product of the two forms. Now use Theorem 5.2. For the terms on the form (5.6), since

$$\mathbf{E}\left(\frac{Z_1^{2k}}{n_1^k}\right)=\frac{\mathbf{E}(Z_1^{2k})}{n_1^k}=O(n_1^{-k})$$

it is concluded that

$$\frac{Z_1^k}{n_1^{k/2}}=O_p(n_1^{-k/2}),\quad k=1,2,\ldots$$

For the terms on the form (5.7)

$$\mathbf{E}\left(\left(\frac{n_1-1}{n_1}\right)^{2k}W_1^{2k}\right)=\frac{\mathbf{E}\left[((n_1-1)W_1)^{2k}\right]}{n_1^{2k}}=\frac{2^{2k}\ \Gamma\left(\frac{n_1-1}{2}+2k\right)}{n_1^{2k}\ \Gamma\left(\frac{n_1-1}{n_1}\right)}=O(1),$$

by (5.3), and consequently

$$\left(\frac{n_1-1}{n_1}\right)^k W_1^k = O_p(1).$$

After collecting terms it is, by Theorem 5.3, possible to write (5.5) as

$$\log\left(1+2\frac{Z_1\gamma}{\sqrt{n_1}}+\frac{Z_1^2\gamma^2}{n_1}+\frac{n_1-1}{n_1}W_1\gamma^2\right)$$

$$=2\frac{Z_1\gamma}{\sqrt{n_1}}+\frac{n_1-1}{n_1}W_1\gamma^2+O_p\left(\max\left(\frac{\gamma^2}{n_1},\gamma^4\right)\right).$$

The corresponding calculations can of course be made also for the second term in (5.4). Then $\log F$ can be written

$$\log F \overset{d}{=} \log W_1 - \log W_2 + 2\frac{Z_1\gamma}{\sqrt{n_1}} - 2\frac{Z_2\gamma}{\sqrt{n_2}} + \frac{n_1-1}{n_1}W_1\gamma^2 - \frac{n_2-1}{n_2}W_2\gamma^2$$

$$+O_p\left(\max\left(\frac{\gamma^2}{n_1},\frac{\gamma^2}{n_2}\gamma^4\right)\right)$$

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

where $R(n_1, n_2, \gamma) = O_p(\max(n_1^{-1}\gamma^2, n_2^{-1}\gamma^2, \gamma^4))$. ♦

6. A simulation study

In this chapter we investigate, by Monte Carlo technique, the significance levels and powers of the tests described in Chapter 3 and the approximate F-test proposed in Chapter 4.

6.1. Methods

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 χ^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 when 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.

Five cases were studied, as summarized in Table 6.0. The type I errors of the tests were investigated in Case 1–3, and the powers of the tests were investigated in Case 4 and 5. 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%), but still equal sample sizes. The third case had large coefficients of variation but unequal sample sizes (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. In the fifth case the sample sizes were unequal (n_1 was fixed to 4), and a larger discrepancy between the coefficients of variation was studied, 5% vs. 15%, since it is hard to obtain a good power when one of the sample sizes is small.

Table 6.0. *The cases investigated in the simulation study*

Case	Objective	γ_1	γ_2	Sample sizes	Table	Figure
1	Type I error	0.05	0.05	$n_1 = n_2$	6.1	6.1
2	Type I error	0.25	0.25	$n_1 = n_2$	6.2	6.2
3	Type I error	0.25	0.25	$n_1 = 4$	6.3	6.3
4	Power	0.05	0.10	$n_1 = n_2$	6.4	6.4
5	Power	0.05	0.15	$n_1 = 4$		
		0.15	0.05	$n_1 = 4$	6.5	6.5

The size, n_2, of the second sample varied from 2 to 20. Thus 19 simulations were made per case. In Case 5, however, 2 simulations were made per value of n_2, *i.e.* in total 38 simulations. This was because the tests had one power when the smaller coefficient of variation was measured with the smaller sample size, and another power when the smaller coefficient of variation was measured with the larger sample size. For this reason, per value of n_2, 20 000 samples were simulated according to the first situation, and 20 000 samples were simulated according to the second situation. The average powers were calculated and reported.

The following tests were included in the study: the approximate F-test (4.4), the likelihood ratio test (3.12), Miller's test (3.14), Bennett's test (3.13), Doornbos

and Dijkstra's test (3.17), the Wald test (3.15), the score test (3.16), the naive test (3.19) and the log test (3.18).

6.2. Results

We look at one case at a time.

6.2.1. Case 1: $\gamma_1 = \gamma_2 = 5\%$ *and* $n_1 = n_2$

The results of the simulations according to Case 1 is reported in Table 6.1 and illustrated in Figure 6.1. 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.

Table 6.1. *Case 1. Pr(Type I error) in percentages when* $\gamma_1 = \gamma_2 = 0.05$*. Significance level: 5%.* F = *F-test (4.4),* R = *Likelihood ratio test (3.12),* M = *Miller's test (3.14),* B = *Bennett's test (3.13),* D = *Doornbos & Dijkstra's test (3.17),* W = *Wald test (3.15),* S = *Score test (3.16),* N = *Naive test (3.19),* L = *Log test (3.18)*

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

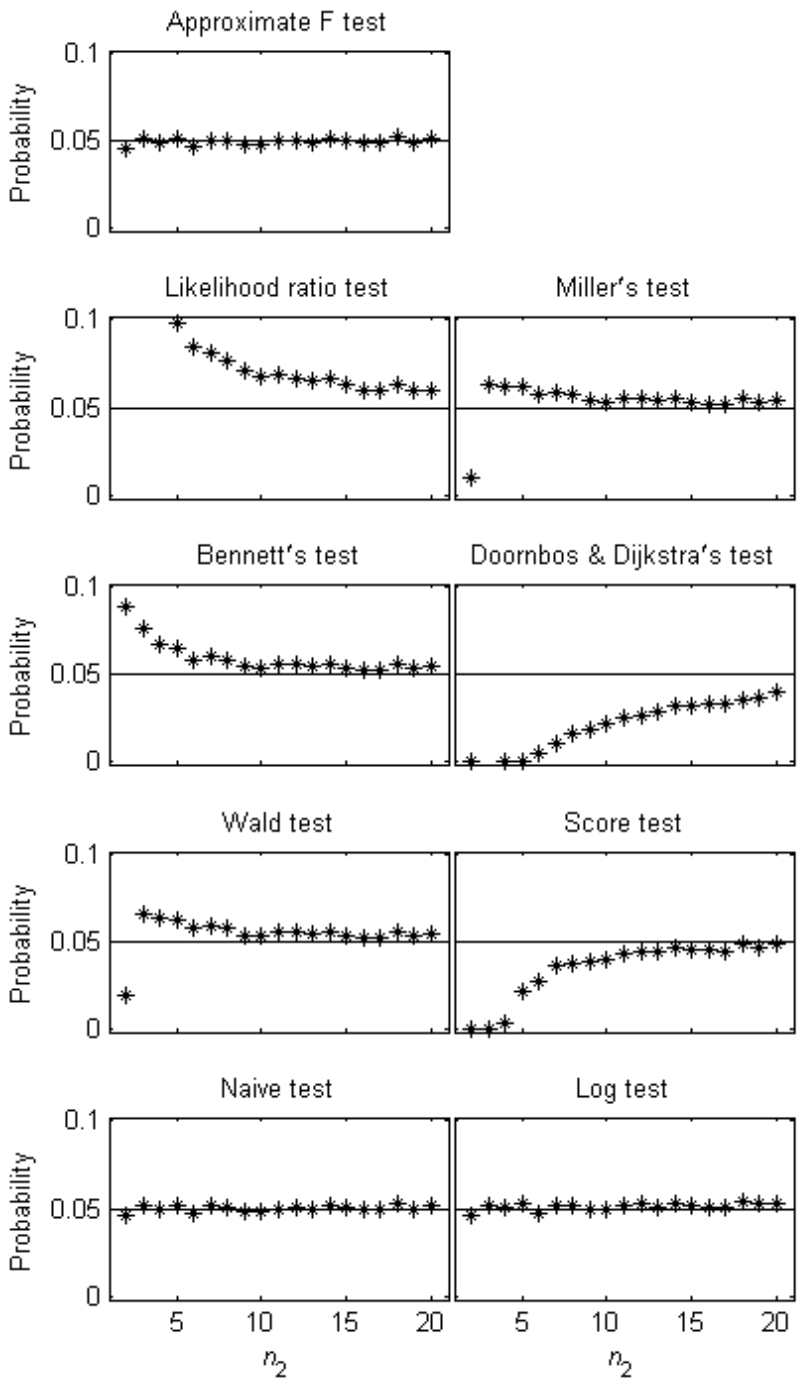

Fig 6.1. Case 1. Probability of type I error when $\gamma_1 = \gamma_2 = 5\%$ and $n_1 = n_2$. Significance level 5%.

6.2.2. Case 2: $\gamma_1 = \gamma_2 = 25\%$ and $n_1 = n_2$

The results of the simulations according to Case 2 is reported in Table 6.2 and illustrated in Figure 6.2. 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.

Table 6.2. *Case 2. Pr(Type I error) in percentages when $\gamma_1 = \gamma_2 = 0.25$. Significance level: 5%.* F = *F-test (4.4),* R = *Likelihood ratio test (3.12),* M = *Miller's test (3.14),* B = *Bennett's test (3.13),* D = *Doornbos & Dijkstra's test (3.17),* W = *Wald test (3.15),* S = *Score test (3.16),* N = *Naive test (3.19),* L = *Log test (3.18)*

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

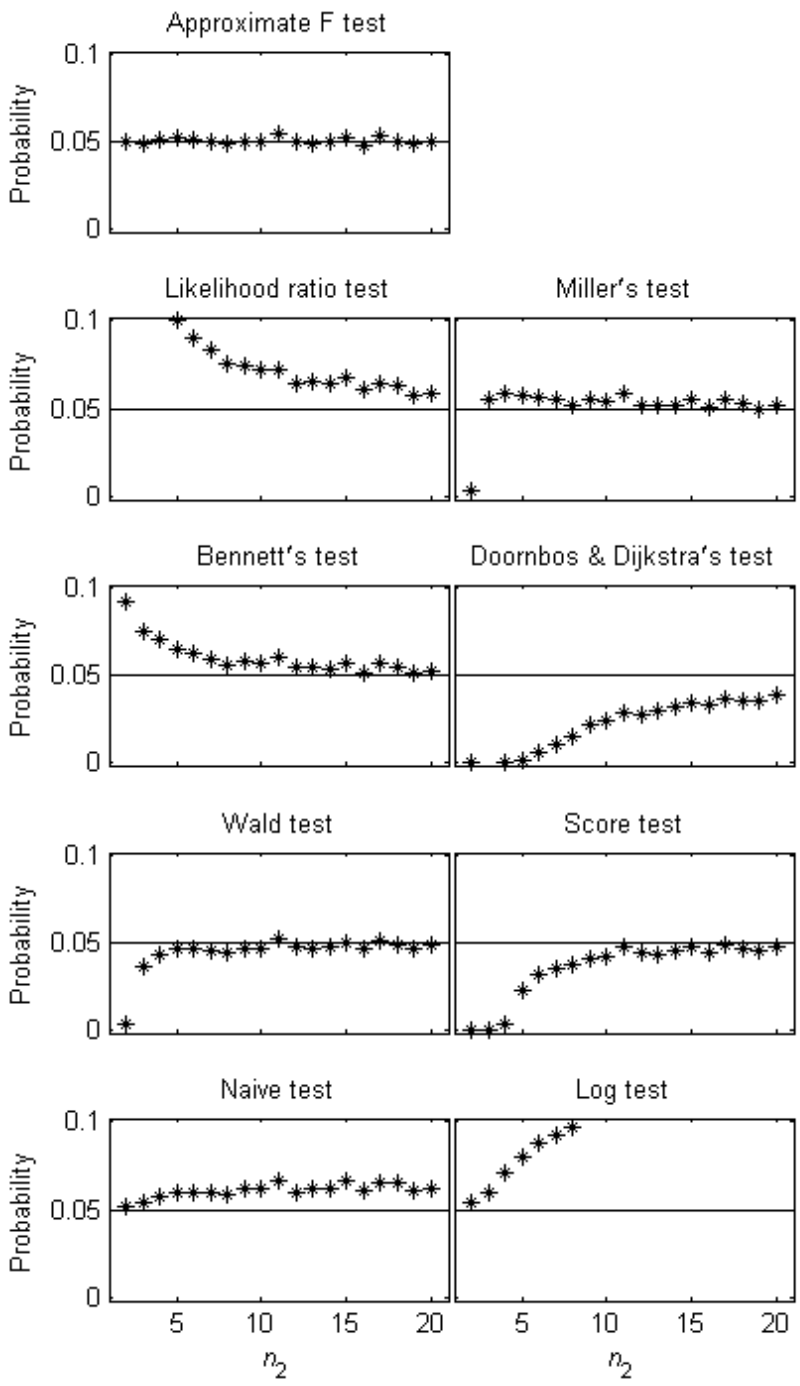

Fig 6.2. Case 2. Probability of type I error when $\gamma_1 = \gamma_2 = 25\%$ and $n_1 = n_2$. Significance level 5%.

6.2.3. Case 3: $\gamma_1 = \gamma_2 = 25\%$ and $n_1 = 4$

The results of the simulations according to Case 3 is reported in Table 6.3 and illustrated in Figure 6.3. 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.

Table 6.3. *Case 3. Pr(Type I error) in percentages when $\gamma_1 = \gamma_2 = 0.25$. Significance level: 5%.* F = *F-test (4.4),* R = *Likelihood ratio test (3.12),* M = *Miller's test (3.14),* B = *Bennett's test (3.13),* D = *Doornbos & Dijkstra's test (3.17),* W = *Wald test (3.15),* S = *Score test (3.16),* N = *Naive test (3.19),* L = *Log test (3.18)*

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

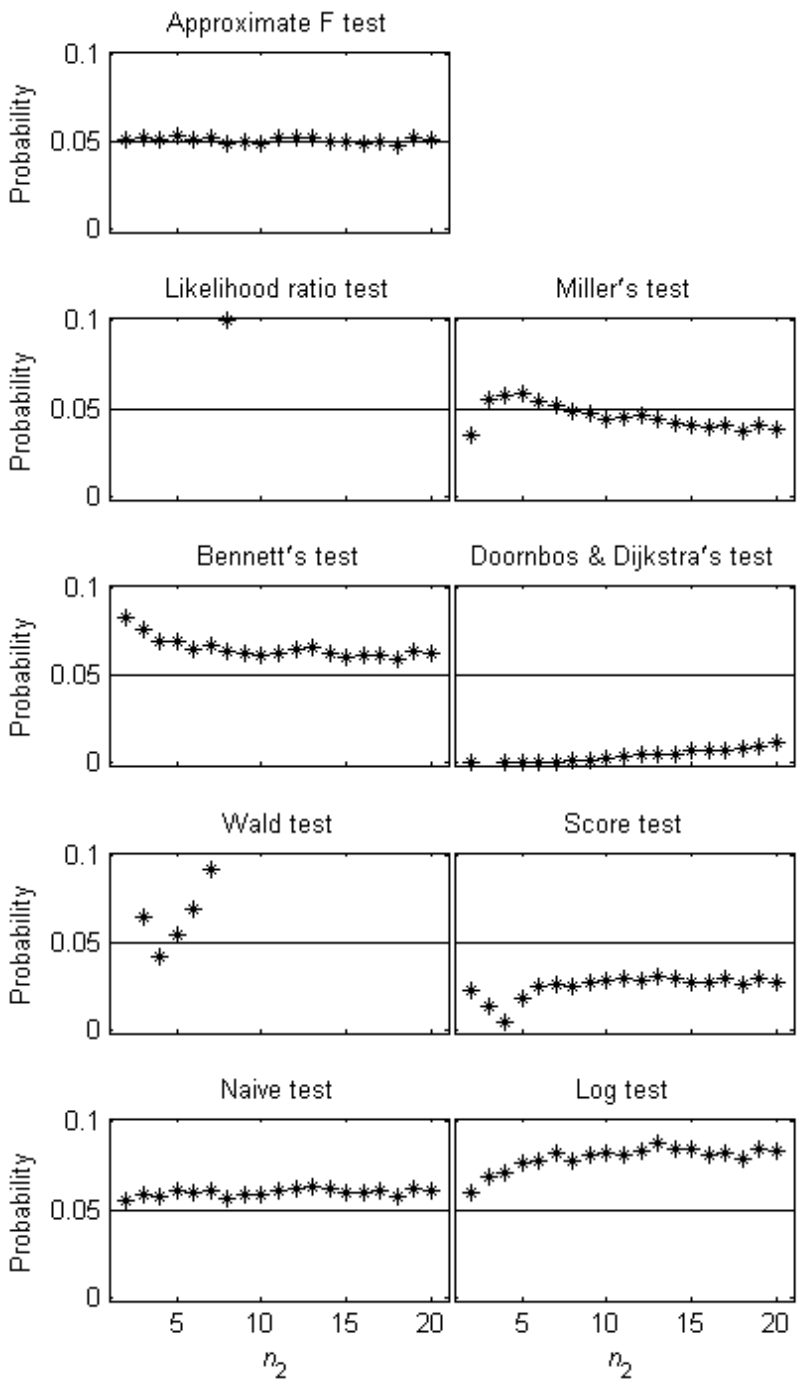

Fig 6.3. Case 3. Probability of type I error when $\gamma_1 = \gamma_2 = 25\%$ and $n_1 = 4$. Significance level 5%.

6.2.4. Case 4: $\gamma_1 = 5\%$, $\gamma_2 = 10\%$ and $n_1 = n_2$

The results of the simulations according to Case 4 is reported in Table 6.4 and illustrated in Figure 6.4. 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, cf. Figure 6.1. 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.

Table 6.4. *Case 4. Power in percentages when $\gamma_1 = 0.05$ and $\gamma_2 = 0.10$. Significance level: 5%.* F = *F-test (4.4),* R = *Likelihood ratio test (3.12),* M = *Miller's test (3.14),* B = *Bennett's test (3.13),* D = *Doornbos & Dijkstra's test (3.17),* W = *Wald test (3.15),* S = *Score test (3.16),* N = *Naive test (3.19),* L = *Log test (3.18)*

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

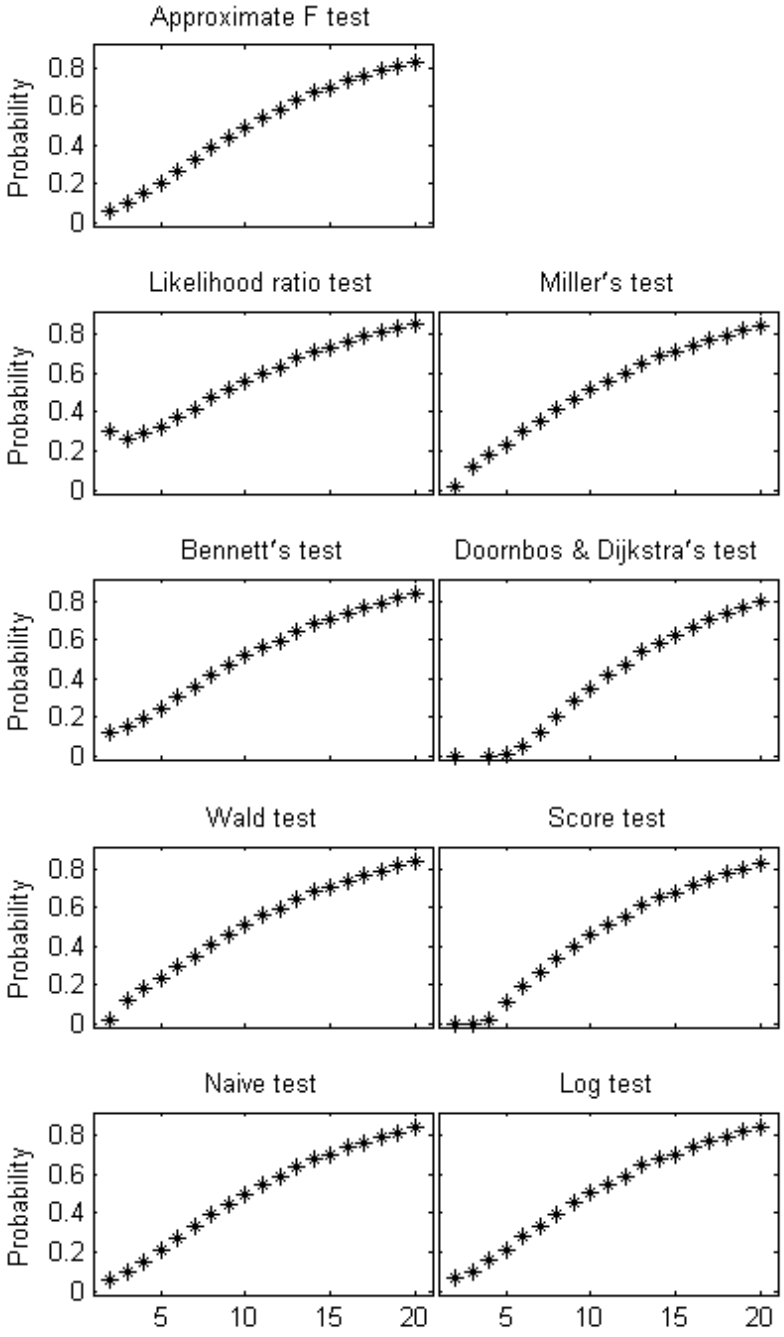

Fig 6.4. Case 4. Power when γ_1 = 5% and γ_2 = 10% and $n_1 = n_2$. Significance level 5%.

6.2.5. Case 5: one γ is 5% and the other γ is 15%, $n_1 = 4$

The results of the simulations according to Case 5 is reported in Table 6.5 and illustrated in Figure 6.5. The likelihood ratio test showed the largest power, but it should be clear from the studies of type I error (Case 1–3) that this test does not behave in a proper way when the sample sizes are small. The approximate F-test, Miller's test, Bennett's test, the Wald test, the naive test and the log test all gave similar power patterns, but Miller's test reached a smaller power and Bennett's test reached a larger. Doornbos and Dijkstra's test was not successful at all, and the power of the score test was comparatively small.

Table 6.5. *Case 5. Average power in percentages when one γ is 0.05 and the other γ is 0.15. Significance level: 5%.* F = *F-test (4.4),* R = *Likelihood ratio test (3.12),* M = *Miller's test (3.14),* B = *Bennett's test (3.13),* D = *Doornbos & Dijkstra's test (3.17),* W = *Wald test (3.15),* S = *Score test (3.16),* N = *Naive test (3.19),* L = *Log test (3.18)*

n_1	n_2	*F*	*R*	*M*	*B*	*D*	*W*	*S*	*N*	*L*
4	2	17.21	42.76	15.51	22.27	0.00	32.64	11.82	17.41	17.57
4	3	25.00	46.68	27.27	31.83	-	28.97	9.75	25.41	25.97
4	4	32.78	51.98	37.30	39.39	0.00	36.44	3.99	33.42	34.06
4	5	39.65	57.05	43.46	45.74	0.00	43.25	18.31	40.25	40.87
4	6	43.81	60.27	46.19	49.68	0.02	47.57	25.16	44.36	45.23
4	7	46.31	62.25	47.27	52.01	0.30	49.97	28.32	46.80	47.63
4	8	49.23	64.84	49.10	54.98	0.78	52.46	30.90	49.78	50.59
4	9	49.76	65.80	48.71	55.65	1.35	53.31	31.95	50.29	51.23
4	10	51.04	67.15	48.93	57.02	2.09	54.52	32.79	51.60	52.34
4	11	52.49	68.67	49.58	58.83	2.83	55.46	33.86	53.03	53.97
4	12	53.12	69.41	49.79	59.62	3.96	55.93	34.39	53.63	54.50
4	13	53.64	70.19	49.79	60.04	4.47	56.02	35.06	54.11	54.97
4	14	54.35	70.79	49.91	60.69	5.39	56.82	35.51	54.84	55.71
4	15	54.93	71.40	49.98	61.17	6.23	56.84	35.71	55.39	56.27
4	16	54.85	71.82	49.64	61.61	6.92	56.99	35.89	55.38	56.41
4	17	55.44	72.47	49.87	62.06	7.77	57.42	36.17	55.93	56.84
4	18	55.23	72.55	49.27	62.01	8.17	57.17	36.47	55.72	56.74
4	19	55.74	73.12	49.49	62.84	9.10	57.66	36.45	56.23	57.30
4	20	55.97	73.42	49.58	63.03	9.72	57.53	36.77	56.45	57.40

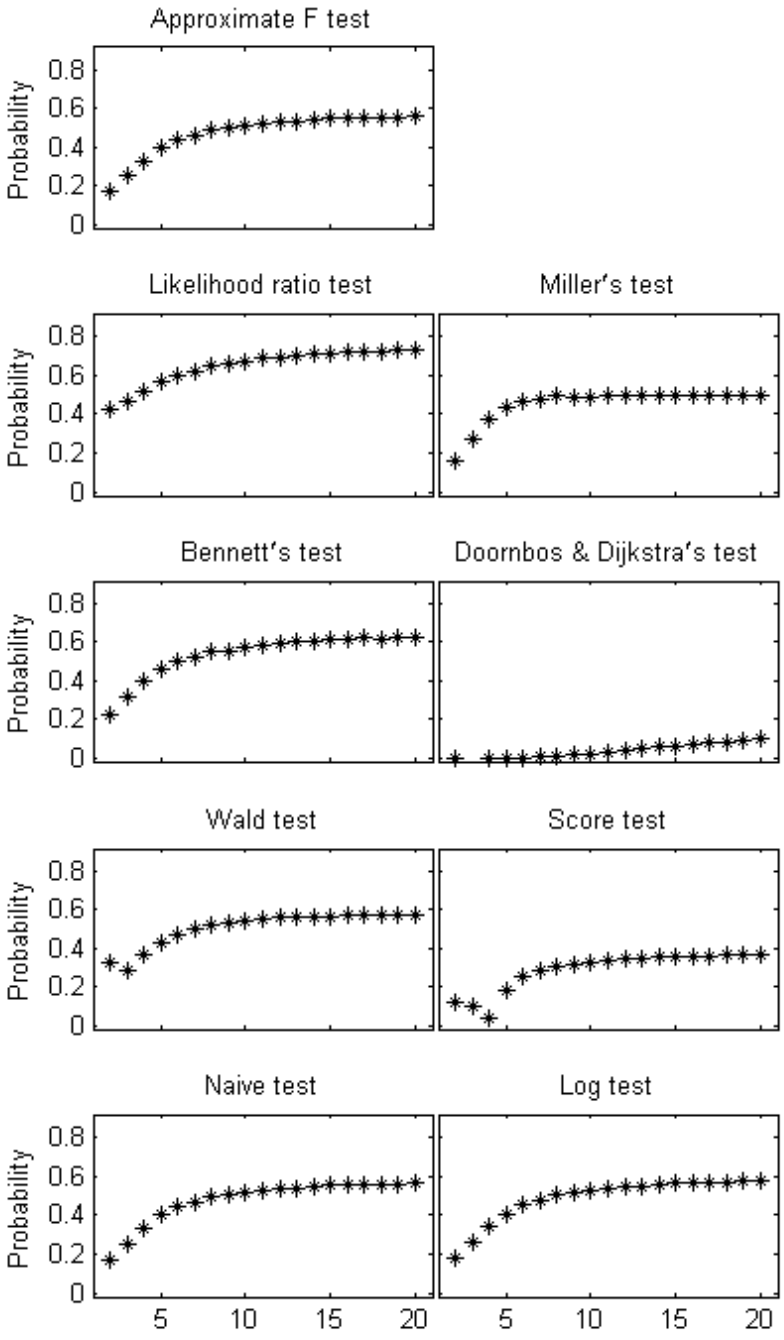

Fig 6.5. Case 5. Average power when one γ is 5% and the other γ is 15%, n_1 = 4. Significance level 5%.

6.3. 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 was 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 it also reached a smaller power. Bennett's test was more powerful than the approximate F-test, but it also 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. Based on the given simulations we conclude that the approximate F-test is recommendable.

7. The distribution of McKay's approximation

As indicated by the example in Section 3.2 McKay's approximation works well for constructing confidence intervals for the coefficient of variation. The proposed statistic F given in (4.4) is a quotient between two McKay approximations. We have seen that F is approximately F distributed when the null hypothesis of equal coefficients of variation is true (Chapter 5), and that the F-test works well in comparison with other tests (Chapter 6). These results make us interested in McKay's approximation. We define it in the following way.

Definition 7.1. Let y_j, $j = 1, 2, \ldots, n$ be n independent observations from a normal distribution with expected value μ and variance σ^2. Let γ denote the population coefficient of variation, *i.e.* $\gamma = \sigma / \mu$, and let c denote the sample coefficient of variation, *i.e.*

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

McKay's approximation $\boldsymbol{K}$ is defined as

$$K = \frac{(n-1)\;c^2\;(1+\gamma^2)}{\left(1+\dfrac{n-1}{n}c^2\right)\gamma^2}. \quad \blacklozenge$$

McKay (1932) obtains the result that K is approximately χ^2 distributed with n - 1 degrees of freedom by transforming the sample coefficient of variation c to t by

$$t^2 = \frac{(n-1)\ c^2}{\left(1+\frac{n-1}{n}c^2\right)\ \gamma^2}$$

and then expressing the density of t as a contour integral in the complex plane. McKay deforms the path of the counter integral to the path of steepest descent, which passes through the saddle point $z = n$ of the integral (*cf.* de Bruijn, 1970). In the next step McKay makes an approximation in order to solve the integral. This makes it possible to study the approximate density of $K = t^2 (1 + \gamma^2)$. Finally McKay notes that K is approximately χ^2 distributed with $n - 1$ degrees of freedom if γ is small.

McKay (1932) does not theoretically express the size of the error in the approximation. For this reason Fieller (1932), in immediate connection to McKay's article, investigates the approximation numerically and concludes that it is 'quite adequate for any practical purpose.' Also Pearson (1932) examines the new approximation and finds it 'very satisfactory.' Later Iglewicz & Myers (1970) study the usefulness of McKay's approximation for calculating quantiles of the distribution of the sample coefficient of variation c when the underlying distribution is normal. They compare results according to the approximation with exact results and find that the approximation is accurate. Umphrey (1983) corrects a similar study made by Warren (1982) and concludes that McKay's approximation is adequate. Vangel (1996) analytically shows that the error in the approximation is small when the population coefficient of variation γ is small.

It is thus well documented that McKay's approximation K, as defined by Definition 7.1, is indeed approximately χ^2 distributed with $n - 1$ degrees of freedom. However we shall soon see that K is actually noncentral beta distributed. We shall prove it in two ways. The first utilises the following lemma.

Lemma 7.1. Let s denote the standard deviation, *i.e.* $s = cm$. Then

$$\frac{c^2}{1+\frac{n-1}{n}c^2} = \frac{ns^2}{\sum_{j=1}^{n} y_j^2}.$$

Proof.

$$\frac{c^2}{1+\frac{n-1}{n}c^2} = \frac{\frac{s^2}{m^2}}{1+\frac{(n-1)s^2}{nm^2}} = \frac{ns^2}{nm^2+(n-1)s^2}$$

$$= \frac{ns^2}{nm^2+\left(\sum_{j=1}^{n} y_j^2 - 2m\sum_{j=1}^{n} y_j + nm^2\right)} = \frac{ns^2}{\sum_{j=1}^{n} y_j^2}. \quad \blacklozenge$$

Lemma 7.1. is of special interest to us since it gives another representation of the test statistic F given in (4.4):

$$F = \frac{n_1 s_1^2 \sum_{j=1}^{n_2} y_{2j}^2}{n_2 s_2^2 \sum_{j=1}^{n_1} y_{1j}^2}.$$

Recall that if U_1 and U_2 are independent χ^2 distributed variables with v_1 and v_2 degrees of freedom respectively, then

$$V = \frac{U_1}{U_1+U_2} \tag{7.1}$$

is Beta($v_1/2$, $v_1/2$). When U_1 is instead a random variable with a noncentral χ^2 distribution with v_1 degrees of freedom and noncentrality parameter λ the distribution of the ratio (7.1) is according to Johnson & Kotz (1970) the noncentral beta distribution with noncentrality parameter λ.

The other side of the coin is the distribution of

$$1-V = \frac{U_2}{U_1+U_2},$$

which is sometimes also called noncentral beta, *e.g.* by Hodges (1955) and Seber (1963). We shall use the following definition which include both cases and which is also given by Johnson & Kotz (1970).

Definition 7.2. Let U_1 and U_2 be independently χ^2 distributed random variables with v_1 and v_2 degrees of freedom respectively and with noncentrality parameters λ_1 and λ_2 respectively. The **doubly noncentral beta distribution** with parameters $v_1/2$, $v_2/2$, λ_1 and λ_2, denoted Beta($v_1/2$, $v_2/2$, λ_1 , λ_2) is defined as the distribution of

$$V = \frac{U_1}{U_1 + U_2} . \quad \blacklozenge$$

We are now ready to state that McKay's χ^2-approximation is doubly noncentral beta distributed.

Theorem 7.1. The distribution of McKay's approximation K, as defined by Definition 7.1, is

$$\frac{(1+\gamma^2)n}{\gamma^2} \ \text{Beta}\left(\frac{n-1}{2} \ , \ \frac{1}{2} \ , \ 0 \ , \ \frac{n}{\gamma^2} \right). \quad \blacklozenge \qquad (7.2)$$

We shall prove the theorem in two ways. The first proof makes use of Lemma 7.1.

Proof 1. By Lemma 7.1,

$$\frac{(n-1)c^2}{1+\dfrac{n-1}{n}c^2} = \frac{n\sum_{j=1}^{n}(y_j - m)^2}{\sum_{j=1}^{n} y_j^2} = \frac{n\sum_{j=1}^{n}(y_j - m)^2}{\sum_{j=1}^{n}(y_j - m)^2 + \sum_{j=1}^{n} m^2} = \frac{nU_1}{U_1 + U_2},$$

where $U_1 = \Sigma(y_j - m)^2/\sigma^2$ and $U_2 = \Sigma m^2/\sigma^2$. Here U_1 is χ^2 distributed with $n - 1$ degrees of freedom. The average m is normally distributed with expected value μ and variance σ^2/n. Consequently nm^2/σ^2, *i.e.* U_2, is χ^2 distributed with 1 degree of freedom and noncentrality parameter $n\mu^2/\sigma^2 = n/\gamma^2$ (see *e.g.* Shao, 2003). Since the sums of squares $\Sigma(y_j - m)^2$ and Σm^2 are independent it follows from Definition 7.2 that McKay's approximation K is distributed as (7.2). $\blacklozenge$

The second proof utilise that $n^{1/2}/c$ is t distributed with $n - 1$ degrees of freedom and noncentrality parameter $n^{1/2}/\gamma$. We show this fact in the beginning of the proof.

Proof 2. The noncentral t distribution with $n - 1$ degrees of freedom and noncentrality parameter λ is the distribution of

$$T = \frac{V_1}{\sqrt{\frac{V_2}{n-1}}}, \tag{7.3}$$

where V_1 is N(λ, 1) and V_2 is χ^2 distributed with $n - 1$ degrees of freedom (see *e.g.* Shao, 2003). Since m is distributed as

$$\mathrm{N}\left(\mu \ , \ \frac{\sigma^2}{n}\right) = \frac{\sigma}{\sqrt{n}} \ \mathrm{N}\left(\frac{\mu\sqrt{n}}{\sigma} \ , \ 1\right) = \frac{\sigma}{\sqrt{n}} \ \mathrm{N}\left(\frac{\sqrt{n}}{\gamma} \ , \ 1\right) = \frac{\sigma}{\sqrt{n}} \ V_1$$

with $\lambda = n^{1/2}/\gamma$, and since $s^2 = c^2m^2$ is distributed as $\sigma^2 V_2/(n - 1)$ it follows that

$$\frac{\sqrt{n}}{c} = \frac{m\sqrt{n}}{s} \overset{d}{=} \frac{\frac{\sigma}{\sqrt{n}} V_1 \sqrt{n}}{\sqrt{\frac{\sigma^2 V_2}{n-1}}} = \frac{V_1}{\sqrt{\frac{V_2}{n-1}}} = T \, .$$

In conclusion $n^{1/2}/c$ is t distributed with $n - 1$ degrees of freedom and noncentrality parameter $\lambda = n^{1/2}/\gamma$. Thus we can write

$$\frac{(n-1)c^2}{1 + \frac{n-1}{n}c^2} = \frac{(n-1)n}{\frac{n}{c^2} + (n-1)} \overset{d}{=} \frac{(n-1)n}{T^2 + n - 1} \, .$$

Thus, with T from (7.3),

$$\frac{(n-1)c^2}{1 + \frac{n-1}{n}c^2} = \frac{(n-1)n}{\frac{V_1^2(n-1)}{V_2} + n - 1} = \frac{nV_2}{V_1^2 + V_2} = \frac{nU_1}{U_2 + U_1} \, ,$$

where $U_1 = V_2$ is χ^2 distributed with $n - 1$ degrees of freedom, and $U_2 = V_1^2$ is χ^2 distributed with 1 degree of freedom and noncentrality parameter $\lambda^2 = n/\gamma^2$. Then, by Definition 7.2, McKay's approximation K is distributed as (7.2). ♦

8. 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 thesis has been to explore confidence intervals and tests that have been suggested but are seldom used, and to contribute to the knowledge about how to make valid statistical inference. The present work shall be the basis for future research and development of methods for analysing normally distributed measurements with constant coefficient of variation.

A sceptic may claim that researchers do wrong when they calculate coefficients of variation and at the same time assume normally distributed data. The sceptic may think that models that are not easy to analyse should be avoided. For this reason we have discussed the adequacy of the model. We have especially discussed immunoassay data and derived a rational for assuming that the measurements are approximately normally distributed and that the coefficient of variation is approximately constant. This is central, because the coefficient of variation is the predominant measure of dispersion in diagnostic research. Our presumption has been that researchers do right, but are in need of statistical tools for analysing their estimated coefficients of variation, exactly as Warren (1982) pointed out.

8.1. Conclusions

With the advanced computer programs of today it is easy to calculate an exact confidence interval for the coefficient of variation, based on the noncentral t distribution. Otherwise we recommend the approximate confidence interval suggested by Vangel (1996). This confidence interval is built on the χ^2-approximation found by McKay (1932). We have, as we believe for the first time, shown that McKay's approximation is noncentral beta distributed.

For the hypothesis of equal coefficients of variation we have proposed a new test statistic F, which is approximately F distributed. The test statistic F is simply a quotient between two McKay approximations. It is thus easy to calculate. 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. For this reason we recommend the approximate F-test for comparing two sample coefficients of variation.

The proposed approximate F-test is, unlike several tests, 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 possibility of weighing results is a valuable feature of the F-statistic.

8.2. Future research

This thesis is the starting point for further research on the analysis of coefficients of variation. We have focused on two fundamental inferential problems: that of constructing a confidence interval and that of comparing two coefficients of variation. There are many other problems to study.

One basic question is how to make point estimates. Usually the unknown coefficient of variation is estimated by the coefficient of variation in the sample, but not much has been written about the properties of this estimator. Could there be estimators that are better in some respect? In applications there are often many estimates of a common coefficient of variation and a need for pooling the estimates into one single estimate. How shall this calculation be made when the estimates are based on different numbers of observations? It is possible to calculate a weighted average with number of observations or with degrees of freedom as weights. Another possibility, suggested by the generalized F-test introduced in this thesis, is to weight the coefficients of variation after transformation according to McKay's approximation.

We have in Chapter 5 studied the properties of the approximate F-test. The corresponding studies of the generalized F-test, as defined in Definition 4.2, remain to be made. The agreement between the moments of the test statistic F and the moments of an F-distributed random variable could perhaps be utilised to make an asymptotic expansion of the density of F. The observation that McKay's

approximation is noncentral beta distributed make it possible to take advantage of results for this distribution in the study of how McKay's approximation can be used for analysing coefficients of variation. The poor performance of the log test, observed in the simulation study when the coefficient of variation was large, calls for a theoretical explanation. The approximate F-test is limited to the case of comparing two coefficients of variation. We would like to generalize to a test for the comparison of several coefficients of variation.

In immunoassay data there are often many variance components. There could be variation *e.g.* between batches of reagents, between days, between laboratories, between instruments and between positions within instruments. How shall such complex models be analysed when many blood samples with very different levels of concentration are included and the researchers are interested in the coefficients of variation?

We have argued for a model with the standard deviation proportional to the expected value, but more sophisticated models are sometimes appropriate for analysing immunoassay data. It is often assumed (see *e.g.* O'Connell, Belanger & Haaland, 1993) that the standard deviation is linearly increasing not with the expected value, but with the expected value to a power of an additional variance parameter, *i.e.*

$$\sigma = \gamma \mu^{\phi}, \tag{8.1}$$

where σ is the standard deviation, μ is the expected value and γ and ϕ are variance parameters. In this thesis we have only discussed the case that $\phi = 1$. A future work could involve models in which both γ and ϕ are estimated.

We have studied differences in coefficients of variation between two groups. In the future we would like to model the coefficient of variation, or the variance parameters in (8.1), by explanatory variables. A closely related problem of large interest is the problem of modelling the expected values when the coefficient of variation is assumed to be constant or when the standard deviation is assumed to depend on the expected values according to (8.1).

Acknowledgments

All through the development of this thesis, from my first idea to the final result, my supervisor Dietrich von Rosen has been encouraging and optimistic. His continuous support and much advice have made work easy. I thank you, Dietrich, for good collaboration.

I thank all PhD students and other colleagues at the Department of Biometry and Engineering for friendship and care. I thank all my friends at Pharmacia

Diagnostics AB and Statisticon AB for being interested in my research and for accepting that my mind sometimes has been elsewhere. Especially I thank Lars Söderström and Katarina Hedman at Pharmacia Diagnostics AB. Lars Söderström initially made me interested in the research problem, supported the idea of postgraduate studies and provided the contact with SLU. Katarina Hedman made it possible for me to begin the research, and her believe in me has been of great value.

I thank Rolf Sundberg at Stockholm University for his critical comments on a draft version of this thesis. He made me see possibilities of significant improvements of the presentation.

Jenny, I love you for understanding that this was important to me. Viola and Vilhelm – you are the best.

The Centre of Biostochastics, SLU and Pharmacia Diagnostics AB financed the research.

References

Azzalini, A. 1996. *Statistical Inference. Based on the likelihood.* Chapman & 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 39,* Series B, 400-405.

Bhoj, D.S. & Ahsanullah, M. 1993. Testing Equality of Coefficients of Variation of Two Populations. *Biometrical Journal 35*, 355-359.

de Bruijn, N.G. 1970. *Asymptotic Methods in Analysis*. 3rd edition. North-Holland Publ. Co. Amsterdam.

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

David, F.N. 1949. Note on the Application of Fisher's k-statistic. *Biometrika 36*, 383-393.

DeSilva, B., Smith, W., Weiner, R., Kelley, M., Smolec, J., Lee, B., Khan, M., Tacey, R., Hill, H. & 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. & 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.

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

Gerig, T.M. & Sen, A.R. 1980. MLE in Two Normal Samples With Equal but Unknown Population Coefficients of Variation. *Journal of the American Statistical Association 75*, Theory and Methods Section, 704-708.

Graf, U., Henning, H.J., Stange, K. & Wilrich, P.T. 1987. *Formeln und Tabellen der angewandten mathematischen Statistik*. Springer. Berlin.

Gupta, R.C. & Ma, S. 1996. Testing the Equality of Coefficients of Variation in k Normal Populations. *Communications in Statistics - Theory and Methods 25*, 115-132.

Hald, A. 1952. *Statistical Theory with Engineering Applications.* Wiley. New York.

Hodges, J.L. 1955. On the Noncentral Beta-Distribution. *The Annals of Mathematical Statistics 26*, 648-653.

Iglewicz, B. & Myers, R.H. 1970. Comparison of Approximations to the Percentage Points of the Sample Coefficient of Variation. *Technometrics 12*, 166-169.

Johnson, N.L. & Kotz, S. 1970. *Distributions in Statistics: Continuous Univariate Distributions-2*. Wiley. New York.

Kleinbaum, D.G, Kupper, L.L. & Muller, K.E. 1988. *Applied Regression Analysis and Other Multivariable Methods*. 2nd edition. PWS-KENT Publishing Company. Boston, Massachusetts.

Kotz, S. & Johnson, N.L. 1982. *Encyclopedia of Statistical Sciences,* Vol. 1. Wiley. New York.

Kotz, S. & Johnson, N.L. 1983. *Encyclopedia of Statistical Sciences,* Vol. 3. 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.

Lynch M. & Walsh B. 1997. *Genetics and Analysis of Quantitative Traits*. Sinauer. Sunderland.

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, G.E. 1991. Asymptotic Test Statistics for Coefficients of Variation. *Communications in Statistics - Theory and Methods 20*, 3351-3363.

Miller, G.E. & Feltz, C.J. 1997. Asymptotic Inference for Coefficients of Variation. *Communications in Statistics - Theory and Methods 26*, 715-726.

Miller, E.G. & 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.

O'Connell, M., Belanger, B.A. & Haaland, P.D. 1993. Calibration and assay development using the four-parameter logistic model. *Chemometrics and Intelligent Laboratory Systems 20*, 97-114.

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.

Raab, G.M. 1981. Estimation of a Variance Function, with Application to Immunoassay. *Applied Statistics 30*, 32-40.

Rao, K.A. & Vidja, R. 1992. On the Performance of a Test for Coefficient of Variation. *Calcutta Statistical Association Bulletin 42*, 87-95.

Reh, W. & Scheffler, B. 1996. Significance Tests and Confidence Intervals for Coefficients of Variation. *Computational Statistics and Data Analysis 22*, 449-452.

Rodbard, D., Lenox, R.H., Wray, H.L. & Ramseth, D. 1976. Statistical Characterization of the Random Errors in the Radioimmunoassay Dose-Response Variable. *Clinical Chemistry 22*, 350-358.

Schimmerl-Metz, S.M., Metz, V.M., Totterman, S.M.S., Mann, F.A. & 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.

Seber, G.A.F. 1963. The non-central chi-squared and beta distributions. *Biometrika 50*, 542-544.

Shafer, N.J. & 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.

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. & 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.

Earlier reports in this series

002 Sundberg, C. 2004. Food waste composting – effects of heat, acids and size.