World Applied Sciences Journal 10 (7): 853-857, 2010
ISSN 1818-4952

Two Approaches for Comparing the Coefficients of Variation of Several Normal Populations

[1]*A.A. Jafari and* [2]*J. Behboodian*

[1]Department of Statistics, Yazd University, Yazd, Iran
[1]Department of Statistics, Shiraz University, Shiraz 71454, Iran
[2]Department of Mathematics, Shiraz Islamic Azad University, Shiraz, Iran

Abstract: In this article, for comparing the coefficients of variation for several normal populations, we propose a novel approach by using the concepts of generalized test variable and generalized p-value and develop an approximation test based on conditional mean and conditional variance of this generalized approach. We compare these two tests with four existing tests via Monte Carlo simulation. Simulation studies show that the generalized approach and new approximation test have satisfactory type I error probabilities and these two tests are better than the existing tests. Finally, the tests are illustrated using two real examples.

Key words: Coefficient of variation . generalized *p*-value . monte carlo simulation . type i error

INTRODUCTION

The Coefficient of Variation (CV) of a random variable X, with mean $\mu \neq 0$ and standard deviation σ, is defined by the ratio σ/μ. This ratio is an important measure of variation and it is useful in medicine, biology, physics, finance and engineering, because it is free from measurement units and it can be used for comparing the variability of different populations. For describing the variation within the data, CV's more meaningful and useful than σ to compare the among several groups of observations. Several tests have been proposed for the equality of CV's from k normal populations: Bennett test (Bennett, 1976), Modified Bennett test (Shafer and Sullivan, 1986), modified Miller asymptotic test (Feltz and Miller, 1996), likelihood ratio, wald test and score test for equality of inverse coefficients of variation (Nairy and Rao, 2003).

Some Monte Carlo simulation studies are performed by Feltz and Miller (1996), Fung and Tsang (1998) and Nairy and Rao (2003) for comparing the sizes and powers. By evaluating these literature we observe that none of these tests performs satisfactorily in terms of type I error probability and especially when the sample sizes are unequal, type I error probabilities of these tests are larger than nominal level. Therefore, finding a test with type I error probabilities smaller than nominal level is more useful. The concepts of generalized test variable and generalized p-value are introduced by Tsui and Weerahandi (1989) and successfully are applied in developing hypothesis tests in situations where traditional approaches do not provide useful solutions. In this article, using these concepts, we obtain a method for testing the equality of CV's for k normal populations. Also a new approximation test is given for testing the equality of CV's by using this generalized approach. Our simulations studies show that type I error probabilities of these new approaches are smaller than nominal level.

This paper is organized as follows: In Section 2, a generalized approach test is given for comparing the CV's and using this generalized approach, a new approximation approach is proposed. In Section 3, a simulation study is performed for comparing the type I error probabilities of two new approaches with four existing tests. Also, Two real examples are provided.

TESTS FOR EQUALITY OF CV'S

First we give a Theorem that is applicable in this Section. Then we will introduce a generalized approach for testing the equality of CV's based on generalized test variable and generalized p-value. Using this generalized approach, a new approximation test is derived.

Theorem 1: Let $y = (y_1,\ldots,y_k)$ be a vector and $V = [\mathrm{diag}(v_1,\ldots,v_k)]$ be a $k\times k$ matrix. Then

$$(Hy)'\left[HVH'\right]^{-1}Hy = \sum_{i=1}^{k} v_i^{-1}y_i^2 - \frac{[\sum_{i=1}^{k} v_i^{-1}y_i]^2}{\sum_{i=1}^{k} v_i^{-1}} \qquad (1)$$

Corresponding Author: Dr. A.A. Jafari, Department of Statistics, Yazd University, Yazd, Iran

where H = [1:D], 1 = (1,…,1)′ and D = [diag(-1,…,-1)].

Proof: The proof is obvious. W

Let $X_{ij} \sim N(\mu_i, \sigma_i^2)$, i = 1,...,k, j = 1,…,n, be k independent random samples from normal populations and let $n = \sum_{i=1}^{k} n_i$ be the total sample size. The sample mean and sample variance of the ith population are

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$$

and

$$S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/(n_i - 1)$$

respectively. The CV for the ith population is defined as

$$\varphi_i = \frac{\sigma_i}{\mu_i}, i=1,\ldots,k$$

and the interested problem is hypothesis test

$$H_0: \varphi_1 = \ldots = \varphi_k = \varphi$$

where φ is the unknown common CV parameter. This test is equivalent to

$$H_0: H\theta = 0 \quad (1)$$

where

$$\theta = (\theta_1, \ldots, \theta_k)'$$

$$\theta_i = \frac{1}{\varphi_i} = \frac{\mu_i}{\sigma_i}, i=1,\ldots,k$$

A generalized approach: A generalized pivotal variable for θ_i based on ith sample is given by

$$T_{\theta_i} = \frac{U_i}{\sqrt{n_i - 1}} \frac{\bar{x}_i}{s_i} - \frac{Z_i}{\sqrt{n_i}} = \frac{U_i}{\sqrt{n_i - 1}} \hat{\theta}^*_{i(obs)} - \frac{Z_i}{\sqrt{n_i}} \quad (2)$$

where

$$\hat{\theta}^*_{i(obs)} = \frac{\bar{x}_i}{s_i}$$

$$U_i^2 = \frac{(n_i - 1)S_i^2}{\sigma_i^2} \sim \chi^2_{(n_i - 1)}, Z_i = \sqrt{n_i}\frac{\bar{X}_i - \mu_i}{\sigma_i} \sim N(0,1)$$

and $\bar{x}_i$ and s_i^2 are the observed values of $\bar{X}_i$ and S_i^2, respectively.

Remark 1: Tian (2005) proposed a generalized pivotal variable for the common coefficient variation parameter based on generalized pivotal variable in (3). For details see Tian (2005).

The observed value of T_{θ_i} is θ_i and distribution of T_{θ_i} does not depend on any unknown parameter. Hence T_{θ_i} is a generalized pivotal variable for θ_i. Therefore a generalized pivotal variable for Hθ is

$$T_{H\theta} = H(T_{\theta_1}, \ldots, T_{\theta_k})' = HT_\theta \quad (3)$$

where

$$T_\theta = (T_{\theta_1}, \ldots, T_{\theta_k})'$$

Consider

$$\bar{X} = (\bar{X}_1, \ldots, \bar{X}_k)$$

and

$$S^2 = (S_1^2, \ldots, S_k^2)$$

From (4), we see that the conditional expectation of $T_{H\theta}$ given $(\bar{X}, S^2) = (\bar{x}, s^2)$ is

$$\mu_T = E\left[T_{H\theta} \mid (\bar{x}, s^2)\right] = H\left(\mu_{T_1}, \ldots, \mu_{T_k}\right)' \quad (4)$$

where

$$\mu_{T_i} = E\left[T_{\theta_i} \mid (\bar{x}, s^2)\right] = \frac{\hat{\theta}^*_{i(obs)}}{\sqrt{n_i - 1}} E\left[U_i\right] = \frac{\sqrt{2}}{\sqrt{n_i - 1}} \frac{\Gamma(\frac{n_i}{2})}{\Gamma(\frac{n_i - 1}{2})} \hat{\theta}^*_{i(obs)} \quad (5)$$

Also the conditional covariance matrix of $T_{H\theta}$ given $(\bar{X}, S^2) = (\bar{x}, s^2)$ is

$$\Sigma_T = Cov\left[T_{H\theta} \mid (\bar{x}, s^2)\right] = H[diag\left(\sigma^2_{T_1}, \ldots, \sigma^2_{T_k}\right)]H' \quad (6)$$

where

$$\sigma^2_{T_i} = Var\left[T_{\theta_i} \mid (\bar{x}, s^2)\right] = \frac{\hat{\theta}^{*2}_{i(obs)}}{n_i - 1} Var\left[U_i\right] + \frac{1}{n_i} = \left(1 - \frac{2\Gamma^2(\frac{n_i}{2})}{(n_i - 1)\Gamma^2(\frac{n_i - 1}{2})}\right)\hat{\theta}^{*2}_{i(obs)} + \frac{1}{n_i} \quad (7)$$

Let $\tilde{T}$ denote the standard expression of $T_{H\theta}$ with

$$\tilde{T} = \Sigma_T^{-\frac{1}{2}}(T_{H\theta} - \mu_T)$$

where μ_T and Σ_T are given by (5) and (7), respectively. For given $(\bar{x}, s^2)$ the observed value of $\tilde{T}$ is

$$\tilde{t} = \Sigma_T^{-\frac{1}{2}}(H\theta - \mu_T)$$

and the distribution of $\tilde{T}$ is independent of any unknown parameter. Therefore, the distribution of $P\left(||\tilde{T}||^2 \geq ||\tilde{t}||^2\right)$ is independent of any unknown parameter. Let $q_{(||\tilde{T}||^2_\gamma)}$ be the 100γ th percentile of $||\tilde{T}||^2$; so we have

$$P\{||\tilde{T}||^2 \leq q_{(||\tilde{T}||^2_\gamma)}\} = \gamma$$

Since the observed value of $\tilde{T}$ is $H\theta$, the 100γ confidence region of $H\theta$ can be solved by the inequality

$$P\{(H\theta - \mu_T)'\Sigma_T^{-1}(H\theta - \mu_T) \leq q_{(||\tilde{T}||^2_\gamma)}\} = \gamma$$

Therefore, the generalized p-value for testing (2) can be given by

$$\begin{aligned} p &= P\{||\tilde{T}||^2 \geq ||\tilde{t}||^2 | H_0\} = P\{||\tilde{T}||^2 \geq ||\tilde{\mu}_0||^2\} \\ &= P\{(T_{H\theta} - \mu_T)'\Sigma_T^{-1}(T_{H\theta} - \mu_T) \geq \mu_T'\Sigma_T^{-1}\mu_T\} \\ &= P\{Q \geq \mu_T'\Sigma_T^{-1}\mu_T\} \end{aligned} \quad (8)$$

where

$$\tilde{\mu}_0 = \Sigma_T^{-\frac{1}{2}}(0 - \mu_T) = -\Sigma_T^{-\frac{1}{2}}\mu_T$$

and

$$Q = (T_{H\theta} - \mu_T)'\Sigma_T^{-1}(T_{H\theta} - \mu_T)$$

Remark 2: Based on Theorem 1, we have

$$\mu_T'\Sigma_T^{-1}\mu_T = \sum_{i=1}^{k}\sigma_{T_i}^{-2}\mu_{T_i}^2 - \frac{[\sum_{i=1}^{k}\sigma_{T_i}^{-2}\mu_{T_i}]^2}{\sum_{i=1}^{k}\sigma_{T_i}^{-2}} \quad (9)$$

and

$$\begin{aligned} Q &= (T_{H\theta} - \mu_T)'\Sigma_T^{-1}(T_{H\theta} - \mu_T) \\ &= \sum_{i=1}^{k}\sigma_{T_i}^{-2}(T_{\theta_i} - \mu_{T_i})^2 - \frac{[\sum_{i=1}^{k}\sigma_{T_i}^{-2}(T_{\theta_i} - \mu_{T_i})]^2}{\sum_{i=1}^{k}\sigma_{T_i}^{-2}} \end{aligned} \quad (10)$$

where μ_{T_i} and $\sigma_{T_i}^2$ are given in (6) and (8), respectively.

The generalized p-value in (9) rejects H_0 when p is less than the level α. The following algorithm is useful for computation of p.

Algorithm 1: For given $(n_1, \ldots, n_k)$ and $(\hat{\theta}^*_{1(obs)}, \ldots, \hat{\theta}^*_{k(obs)})$, For $J = 1, \ldots, M$ generate $Z_i \sim (0,1)$ and $U_i^2 \sim \chi^2_{(n_i - 1)}$, $i = 1, \ldots, k$. compute $T_j = T_{H\theta}$, $\mu_j = \mu_T$, $\Sigma_j = \Sigma_T$ in (4), (5) and (7), respectively. compute

$$||\tilde{T}||_j^2 = (T_j - \mu_j)'\Sigma_j^{-1}(T_j - \mu_j)$$

and

$$||\tilde{\mu}_0||_i^2 = \mu_i'\Sigma_i^{-1}\mu_i$$

Let $W_j = 1$ if $||\tilde{T}||_j^2 \geq ||\tilde{\mu}_{0j}||^2$, else $W_j = 0$.

$\frac{1}{M}\sum_{j=1}^{M} W_i$ is a simulated estimate of generalized p-value in (9) for hypothesis test in (2).

Our new approximation test: Here, we give a new approximation test for testing equality of CV's. Consider the following test statistic

$$AT = \sum_{i=1}^{k} v_i^{-1}y_i^2 - \frac{[\sum_{i=1}^{k} v_i^{-1}y_i]^2}{\sum_{i=1}^{k} v_i^{-1}}$$

where

$$y_i = \frac{\sqrt{2}}{\sqrt{n_i - 1}} \frac{\Gamma(\frac{n_i}{2})}{\Gamma(\frac{n_i - 1}{2})}\hat{\theta}_i^*$$

$$v_i = \left(1 - \frac{2\Gamma^2(\frac{n_i}{2})}{(n_i - 1)\Gamma^2(\frac{n_i - 1}{2})}\right)\hat{\theta}_i^{*2} + \frac{1}{n_i} \text{ and } \hat{\theta}_i^* = \frac{\bar{X}_i}{S_i}$$

Obtaining exact distribution of AT is very difficult and therefore, we propose an approximation distribution for it. Since the observed value of Q is equal to observed value of AT and also we can easily show that

$$E(Q|(\bar{x}, s^2)) = k - 1 \qquad Var(Q|(\bar{x}, s^2)) = 2(k - 1)$$

an approximation distribution for AT under H_0 is a chi-square distribution with k-1 degrees of freedom. Therefore, AT rejects H_0 if $AT > \chi^2_{\alpha,(k-1)}$.

NUMERICAL STUDY

We compare the two new tests with four existing tests via Monte Carlo simulation and all tests are illustrated using two real examples.

Table 1: Estimated type I error probabilities for tests at α=0.05 (k = 3)

$(n_1,\dots,n_k)$	GP	AT	BT	MBT	WT	MT
(5,5,5)	0.022	0.018	0.059	0.059	0.049	0.047
(5,6,7)	0.025	0.029	0.066	0.069	0.062	0.059
(5,10,15)	0.037	0.030	0.057	0.052	0.069	0.054
(5,5,20)	0.046	0.043	0.059	0.082	0.062	0.067
(5,20,30)	0.049	0.047	0.079	0.070	0.063	0.065
(5,15,15)	0.047	0.039	0.062	0.068	0.052	0.056
(7,7,7)	0.029	0.029	0.066	0.067	0.062	0.061
(7,8,9)	0.032	0.033	0.059	0.059	0.055	0.054
(7,7,20)	0.039	0.038	0.069	0.070	0.064	0.055
(7,20,20)	0.037	0.039	0.057	0.046	0.047	0.047
(10,10,10)	0.026	0.032	0.041	0.041	0.041	0.040
(10,11,12)	0.033	0.037	0.059	0.060	0.059	0.058
(10,15,15)	0.033	0.031	0.046	0.047	0.045	0.045
(10,15,20)	0.044	0.042	0.056	0.055	0.056	0.056
(10,10,20)	0.043	0.040	0.062	0.064	0.064	0.057
(10,20,30)	0.048	0.046	0.060	0.059	0.057	0.053
(20,20,20)	0.047	0.046	0.062	0.062	0.056	0.054
(20,21,22)	0.047	0.050	0.057	0.057	0.055	0.054
(20,20,25)	0.048	0.049	0.067	0.068	0.069	0.066
(20,25,25)	0.045	0.047	0.059	0.059	0.058	0.060
(20,30,30)	0.047	0.047	0.056	0.053	0.052	0.054
(30,30,30)	0.049	0.047	0.053	0.053	0.056	0.056
(30,31,32)	0.480	0.049	0.060	0.060	0.058	0.056
(30,35,40)	0.041	0.042	0.045	0.047	0.044	0.044
(30,40,40)	0.045	0.046	0.050	0.052	0.049	0.047

Table 2: Estimated type I error probabilities for tests at α=0.05 (k = 3)

$(n_1,\dots,n_k)$	GP	AT	BT	MBT	WT	MT
(5,5,5,5,5)	0.022	0.021	0.080	0.080	0.075	0.071
(5,5,5,10,30)	0.041	0.038	0.086	0.067	0.071	0.058
(5,5,5,30,30)	0.036	0.037	0.068	0.057	0.064	0.044
(5,6,7,8,9)	0.031	0.030	0.066	0.067	0.061	0.061
(5,10,15,20,25)	0.045	0.043	0.060	0.056	0.060	0.054
(5,10,10,20,20)	0.039	0.039	0.074	0.070	0.071	0.058
(5,30,30,30,30)	0.042	0.042	0.051	0.049	0.056	0.046
(7,7,7,7,7)	0.019	0.021	0.055	0.055	0.051	0.049
(7,8,9,10,11)	0.031	0.029	0.054	0.055	0.051	0.048
(7,7,7,30,30)	0.042	0.042	0.064	0.060	0.059	0.050
(7,15,15,30,30)	0.045	0.043	0.060	0.059	0.059	0.054
(7,30,30,30,30)	0.048	0.047	0.061	0.062	0.061	0.060
(10,10,10,10,10)	0.033	0.035	0.069	0.071	0.072	0.071
(10,11,12,13,14)	0.041	0.041	0.062	0.060	0.059	0.060
(10,10,10,20,20)	0.038	0.042	0.060	0.058	0.060	0.057
(10,10,30,30,30)	0.042	0.042	0.058	0.053	0.057	0.054
(10,20,20,30,30)	0.045	0.044	0.060	0.059	0.058	0.055
(20,20,20,20,20)	0.048	0.048	0.058	0.059	0.060	0.060
(20,21,21,23,24)	0.042	0.038	0.061	0.059	0.063	0.059
(20,20,20,30,30)	0.041	0.041	0.055	0.057	0.053	0.052
(20,30,30,30,30)	0.042	0.039	0.053	0.051	0.053	0.051
(30,30,30,30,30)	0.042	0.041	0.051	0.051	0.052	0.051
(30,31,32,32,34)	0.048	0.044	0.057	0.058	0.056	0.055
(30,30,30,40,40)	0.045	0.043	0.051	0.052	0.049	0.048
(30,40,40,40,40)	0.050	0.048	0.055	0.055	0.056	0.056

Simulation: A Monte Carlo simulation study is performed to compare the size of the tests to equality of CV's that are given in Section 2. For this propose we generate $\bar{x}_i \sim N(100,100)$ and $s_i^2 \sim 100\chi^2_{(n_i-1)}/(n_i-1)$, i = 1,...,k. Therefore, all populations have the same CV, φ_i = 0.1. We choose this value because in many of the agricultural experiments, CV is around 0.1 (Nairy and Rao, 2003). Also, for other values of CV, we found similar results.

Using $\bar{x}_i$ and s_i^2, i = 1,...,k we obtain generalized p-value (GP) by algorithm 1 with 5000. We repeat this N = 10000 times and number of the p-values that are less than α = 0.05. Also, we obtain the test statistics for new approximation test (AT), Bennett Test (BT), modified Bennett test (MBT), Wald Test (WT) and modified Miller test (MT) and for each test, we number of the cases that test statistics are greater than $\chi^2_{0.05,(k-1)}$. In fact, the nominal level is α = 0.05 and we estimate the type I error probabilities for tests. The results are given in Table 1 and 2 for k = 3 and 5, respectively.

We observe that Estimated type I error probabilities for the generalized p-value and the new approximation test always are smaller than α, but type I error probabilities of other tests are larger than nominal level. Therefore, the generalized approach test and the new approximation test are better than existing test for equality of CV's for several normal populations.

Two examples: These two real examples are proposed by Nairy and Rao (2003) and here, we illustrate for the given tests in Section 2. We note that the generalized p-values are obtained using Algorithm 1 with M = 10000. Also, the p-values for GP, AT, BT, MBT, WT and MT are given in Table 3.

Example 1: Over nine years from 1991-1999 for the State of Karnataka, India, This data is collected to catches of four kinds of fish. For these four kinds of fish the estimate of CV's, $\hat{\varphi}_i^*$, i = 1,2,3,4, are 0.3481, 1.0586, 0.3760,and 0.8880, respectively.

Table 3: p-values of tests for two real examples

	Test					
	GP	AT	BT	MBT	WT	MT
Example 1	0.0525	0.0539	0.0534	0.0439	0.0301	0.0609
Example 2	0.6963	0.6972	0.7064	0.7064	0.6277	0.6051

Example 2: This data refers to survival of patients from four hospital, which is a part of the data in Appendix D of the Fleming and Harrinton (1991). The sample size for these hospitals are 5, 4, 3 and 10. The estimate of CV's, $\hat{\varphi}_i^*$, i = 1,2,3,4, are 0.4937, 1.1224, 0.5852 and 0.6136, respectively.

REFERENCES

1. Bennett, B.M., 1976. On an approximate test for homogeneity of coefficients of variation. In Ziegler, W.J. (Ed.). Contribution to Applied Statis tics, Birkhauser Verlag, Basel and Stuttgart, pp: 169-171.
2. Feltz, C.J. and G.E. Miller, 1996. An asymptotic test for the equality of coeffcients of variation from k populations. Statist. Med., 15: 647-658.
3. Fleming, T.R. and D.P. Harrington, 1991. Counting processes and survival analysis. New York, John Wiley.
4. Fung, W.K. and T.S. Tsang, 1998. A simulation study comparing tests for the equality of coefficients of variation, Statist. Med., 17: 2003-2014.
5. Nairy, K.S. and K.A. Rao, 2003. Tests of coefficient of variation of normal population. Comm. Stat., Simul. Computat., 32: 641-661.
6. Rencher, A.C., 2000. Linear model in statistics. New York, John Wiley and Sons.
7. Shafer, N.J. and J.A. Sullivan, 1986. A simulation study of a test for the equality of the coefficients of variation. Comm. Stat., Simul. Computat., 15: 681-695.
8. Tian, L., 2005. Inferences on the common coefficient of variation. Statist. Med., 24: 2213-2220.
9. Tsui, K.W. and S. Weerahandi, 1989. Generalized p-values in s ignificance testing of hypothesis in the presence of nuisance parameters. J. Am. Statist. Assoc., 84: 602-607.
10. Weerahandi, S., 1995. Exact statistical methods for data analysis , Springer, New York.