Thailand Statistician
July 2021; 19(3): 511-521
http://statassoc.or.th
Contributed paper

On Using the Exact Sampling Distribution of Multivariate Coefficient of Variation

David Sam Jayakumar*, Sulthan Ayyup Kan and Samuel Wilfred
Jamal Institute of Management, Jamal Mohamed College, Bharathidasan University, Trichy, Tamil Nadu, India.
*Corresponding author; e-mail: samjaya77@gmail.com

Received: 22 August 2019
Revised: 9 November 2019
Accepted: 10 February 2020

Abstract

The objective of this manuscript is to propose the exact sampling distribution of sample multivariate coefficient of variation from the normal population. The authors identified its relationship between sample mcv and non-central F-ratio and derived the density function in terms of confluent hyper geometric function by using Jacobian method of one dimensional transformation k^{th} order moment and they proved the p-variate sample $(cv)_p$ is the biased estimator of the true or population $(cv)_p$ Moreover, the shape of the density function of sample multivariate coefficient of variation is also visualized and the authors computed the critical points of sample $(cv)_p$ at 5% and 1% significance level for different sample sizes. Using the Iris plants database in the pattern recognition literature and has shown a numerical study for testing the significance of the sample uni-variate, bi-variate, tri-variate and multivariate coefficient of variation.

Keywords: Sample multivariate coefficient of variation, normal population, non-central F-ratio, confluent hyper-geometric function, biased estimator.

1. Introduction

The coefficient of variation is the widely used measures of variation, especially quantifying the consistency of random variable and its distribution. Hendricks and Robey (1936) made an attempt to extent its use in biometry by proposing a sampling distribution. Similarly, Sharma and Krishna (1994) developed the asymptotic sampling distribution of the inverse of the coefficient of variation where the distribution is used for making statistical inference about the population coefficient of variation or inverse CV without making an assumption about the population distribution. Likewise, Hürlimann (1995) proposed a uniform approximation to the sampling distribution of the coefficient of variation proposed by Hendricks and Robey. Later, Albrecher and Teugels (2007) proposed an asymptotic of the sample coefficient of variation and the sample dispersion which are examples of widely used measures of variation. The Multivariate extension of the coefficient of variation was studied by many (refer: Reyment (1960), Van Valen (1974) Voinov and Nikulin (1996), and Albert and Zhang (2010)) and they gave different structure of multivariate coefficient of variation according to the nature of application. The authors of this paper proposed the exact sampling distribution of sample multivariate coefficient of variation from the normal population and the properties of this sampling distribution is discussed with numerical example in the subsequent sections.

2. Sample Multivariate Coefficient of Variation and Non-Central F-ratio

If X is p-variate random vector of order $p\times 1$ follows multivariate normal distribution $X \sim MN(\mu,\Sigma)$ with non-zero mean (μ) and Variance $(\Sigma),$ then the true or population multivariate coefficient of variation (CV) of the p-variate random vector is given by Voinov and Nikulin (1996, p.68) as

$$(CV)_p = \frac{1}{\sqrt{\mu^T \Sigma^{-1} \mu}}. \tag{1}$$

From (1) the population multivariate coefficient of Variation used by the statisticians to know the consistency of the multivariate data and it may call as multivariate inverse signal to noise ratio. Many disciplines such as renewal, queuing, reliability theory utilizes coefficient of variation to determine the nature of probability distribution based on its variance. It can be logically extended to the p-variate case of the coefficient of variation, If the $(CV)_p < 1,$ then the multivariate distribution of the multivariate data is said to be low-variance distribution and if $(CV)_p > 1,$ then it is said to be high-variance. In order to give the inference about the population $(CV)_p,$ the computation of multivariate sample coefficient of variation is inevitable. The multivariate sample coefficient of variation (CV) of a p-variate sample random vector from the normal population is given as

$$(cv)_p = \frac{1}{\sqrt{\hat{\mu}^T \hat{\Sigma}^{-1} \hat{\mu}}}. \tag{2}$$

Where from (2), $\hat{\mu}$ and $\hat{\Sigma}$ are the maximum likelihood estimates of the population mean vector and variance-co-variance matrix respectively. Based on the sample multivariate coefficient of variation, the authors derived the relationship between the $(cv)_p$ and the non-central F-ratio and it is given as follows: Without loss of generality, rewrite (2) as

$$(cv)_p = \frac{1}{\sqrt{\bar{X}^T \hat{\Sigma}^{-1} \bar{X}}}. \tag{3}$$

From (3), we know the maximum likelihood estimates $\bar{X}$ is the unbiased estimate of $\mu,$ that is $E(\bar{X}) = \mu,$ but $\hat{\Sigma}$ is the biased estimate of $\Sigma,$ that is $E(\hat{\Sigma}) \neq \Sigma,$ then the unbiased estimate of the $\hat{\Sigma}$ is given as

$$S = \frac{n}{n-1}\hat{\Sigma}. \tag{4}$$

Rewrite (4) in terms of maximum likelihood estimate of $\hat{\Sigma}$ and substitute in (4), we get

$$(cv)_p = \frac{1}{\sqrt{\bar{X}^T \left(\frac{n-1}{n} S\right)^{-1} \bar{X}}}. \tag{5}$$

From (5), it can be modified by using the population mean vector (μ) as

$$(cv)_p = \sqrt{\frac{n-1}{n\left(\bar{X}-\mu+\mu\right)^T S^{-1}\left(\bar{X}-\mu+\mu\right)}}. \tag{6}$$

Further by using multivariate version of the central theorem if $\bar{X} \sim MN(\mu, \Sigma/n),$ then $(\Sigma/n)^{-1/2}(\bar{X}-\mu) \sim SMN(o, I_p)$ standard multivariate normal distribution. Therefore rewrite (6) in terms of the pre and post-multiplication of the quantity $(\Sigma/n)^{1/2}$ and it's inverse, then it is given as

$$(cv)_p = \sqrt{\frac{n-1}{n\left((\Sigma/n)^{1/2}(\Sigma/n)^{-1/2}(\bar{X}-\mu)+(\Sigma/n)^{1/2}(\Sigma/n)^{-1/2}\mu\right)^T}}$$

$$\times\sqrt{\frac{1}{S^{-1}\left((\Sigma/n)^{1/2}(\Sigma/n)^{-1/2}(\bar{X}-\mu)+(\Sigma/n)^{1/2}(\Sigma/n)^{-1/2}\mu\right)}}. \tag{7}$$

Using the fact from the multivariate central limit theorem, the term $Z=(\Sigma/n)^{-1/2}(\bar{X}-\mu) \sim SMN(o, I_p),$ and then (7) can be rearranged as

$$(cv)_p = \sqrt{\frac{n-1}{n\left(Z+(\Sigma/n)^{-1/2}\mu\right)^T (\Sigma/n)^{1/2} S^{-1} (\Sigma/n)^{1/2}\left(Z+(\Sigma/n)^{-1/2}\mu\right)}}, \tag{8}$$

Further (8) can be arranged in terms of the inversion of dispersion matrix as

$$(cv)_p = \sqrt{\frac{n-1}{n\left(Z+(\Sigma/n)^{-1/2}\mu\right)^T \left((\Sigma/n)^{-1/2} S (\Sigma/n)^{-1/2}\right)^{-1}\left(Z+(\Sigma/n)^{-1/2}\mu\right)}}. \tag{9}$$

From (9), Wishart (1928) proved if S is the unbiased estimate of $\Sigma,$ then $(n-1)S \sim W_p\left(n-1, \Sigma/n\right)$ follows p-variate Wishart distribution with $n-1$ degrees of freedom and scale parameter $\Sigma/n.$ Moreover, he proved $(\Sigma/n)^{-1/2} S (\Sigma/n)^{-1/2} \sim SW_p(n-1, \mathrm{I}_p)$ follows p-variate standard Wishart distribution and the quantity $Z+(\Sigma/n)^{-1/2}\mu \sim MN\left((\Sigma/n)^{-1/2}\mu, \mathrm{I}_p\right)$ follows multivariate normal distribution with mean $(\Sigma/n)^{-1/2}\mu$ and unit variance matrix $(\mathrm{I}_p).$ Then from (6)

$$n\left(Z+(\Sigma/n)^{-1/2}\mu\right)^T \left((\Sigma/n)^{-1/2} S (\Sigma/n)^{-1/2}\right)^{-1}\left(Z+(\Sigma/n)^{-1/2}\mu\right)$$

$$\sim \frac{(n-1)p}{n-p} F_{\left(p, n-p, \left(\left((\Sigma/n)^{-1/2}\mu\right)^T \left((\Sigma/n)^{-1/2}\mu\right)\right)\right)}. \tag{10}$$

From (10), it follows non-central F-distribution with $(p, n-p)$ degrees of freedom and non-centrality parameter $\left(\left((\Sigma/n)^{-1/2}\mu\right)^T \left((\Sigma/n)^{-1/2}\mu\right)\right).$ Therefore, from (9), the sample multivariate coefficient of variation can be written in terms of the non-central F-ratio as

$$(cv)_p = \sqrt{\frac{n-1}{\frac{(n-1)p}{n-p} F_{\left(p, n-p, \left(\left((\Sigma/n)^{-\frac{1}{2}}\mu\right)^T \left((\Sigma/n)^{-\frac{1}{2}}\mu\right)\right)^2\right)}}}. \tag{11}$$

From (11), the non-centrality parameter $\left(\left((\Sigma/n)^{-1/2}\mu\right)^T \left((\Sigma/n)^{-1/2}\mu\right)\right)$ can be rearranged in terms of the population multivariate coefficient of variation based on (1) and it is given as

$$(cv)_p = \sqrt{\frac{n-p}{p} F_{\left(n-p,p,\left(\sqrt{n}/(CV)_p\right)^2\right)}}. \tag{12}$$

From (12) if $p=1$, then the p-variate multivariate coefficient of variation was reduced into univariate coefficient of variation and it's having relationship with non-central F-ratio follows non-central F-distribution with $(n-1,1)$ degrees of freedom with non-centrality parameter $\left(\sqrt{n}/CV\right)^2$. Based on the identified relationship from (12), the authors derived the exact sampling distribution of the sample multivariate coefficient of variation and it is discussed in the next section.

3. Exact Sampling Distribution of Sample Multivariate Coefficient of Variation

Using the technique of one-dimensional Jacobian of transformation, the probability density function of the non-central F-distribution with $(n-p,p)$ degrees of freedom with a non-centrality parameter $\left(\sqrt{n}/(CV)_p\right)^2$ was transformed into density function of sample multivariate coefficient of variation $(cv)_p$ and it is given in (13) as

$$f\left((cv)_p\right) = f(F')|J|, \tag{13}$$

$$f\left((cv)_p\right) = f(F')\left|\frac{dF'}{d(cv)_p}\right|. \tag{14}$$

The non-central F-distribution was first introduced by Fisher (1924), and Aroian (1941) and its density function is given by

$$f\left(F';n-p,p,\left(\sqrt{n}/(CV)_p\right)^2\right) = e^{-\left(\sqrt{n}/(CV)_p\right)^2/2} g\left(F';n-p,p\right)$$

$${}_1F_1\left(n/2;(n-p)/2;\frac{\left(\sqrt{n}/(CV)_p\right)^2\left((n-p)/p\right)F'}{2\left(1+\frac{n-p}{p}F'\right)}\right),$$

$$0 \le F' < \infty,\ \ n>p,\ \ p,n,(CV)_p > 0. \tag{15}$$

Then from (12), compute and substitute the first derivative $dF'/d(cv)_p = 2p(cv)_p/n-p$ with respect to sample multivariate coefficient of variation $(cv)_p$ in (14) and also substitute (12), (15) in (14), we get

$$f\left((cv)_p;n,p,\left(\sqrt{n}/(CV)_p\right)^2\right) = \frac{2pe^{-\left(\sqrt{n}/(CV)_p\right)^2/2}}{n-p}(cv)_p\, g\left(\frac{2p}{n-p}\left((cv)_p\right)^2;n-p,p\right)$$

$${}_1F_1\left(n/2;(n-p)/2;\frac{\left(\sqrt{n}/(CV)_p\right)^2\left((cv)_p\right)^2}{\left(1+2\left((cv)_p\right)^2\right)}\right)$$

$$0 \le (cv)_p < \infty,\ \ n>p,\ \ p,n,(CV)_p > 0. \tag{16}$$

From (16), it is the density function of sample multivariate coefficient of variation $(cv)_p$ from the normal population and it involves the confluent hypergeometric function ${}_1F_1(\cdot)$, the auxiliary function of multivariate coefficient of variation $g(\cdot)$ with a normalizing constant namely (p,n) are two shape parameters and a scale parameter $(CV)_p$, where n is the sample size, p is the no. of dimensions and $(CV)_p$ is the population multivariate coefficient of variation. Moreover, the authors derived the k^{th} order moments of the distribution of sample multivariate coefficient of variation and it is given as follows.

$$E\left((cv)_p\right)^k = \int_0^\infty \left((cv)_p\right)^k f\left((cv)_p\right) d(cv)_p$$

$$E\left((cv)_p\right)^k = \int_0^\infty \left((cv)_p\right)^k \frac{2pe^{-\left(\sqrt{n}/(CV)_p\right)^2/2}}{n-p}(cv)_p\, g\left(\frac{2p}{n-p}\left((cv)_p\right)^2; n-p, p\right)$$

$$ {}_1F_1\left(n/2; (n-p)/2; \frac{\left(\sqrt{n}/(CV)_p\right)^2\left((cv)_p\right)^2}{\left(1+2\left((cv)_p\right)^2\right)}\right) d(cv)_p$$

$$E\left((cv)_p\right)^k = \frac{B\left((n-p+k)/2, n/2\right)}{B\left((n-p)/2, p/2\right)} e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}$$

$$ {}_1F_1\left((n-p+k)/2; (n-p)/2; \left(\sqrt{n}/(CV)_p\right)^2/2\right). \tag{17}$$

From (17), it is the k^{th} moment of the $(cv)_p$ and the first four moments of the distribution are given as in (18), (19), (20) and (21) as follows.

If $k=1$,

$$E\left((cv)_p\right) = \frac{B\left(\left(n-p+1\right)/2, n/2\right)}{B\left(\left(n-p\right)/2, p/2\right)} e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}$$

$$ {}_1F_1\left(\left(n-p+1\right)/2; \left(n-p\right)/2; \left(\sqrt{n}/\left(CV\right)_p\right)^2/2\right), \tag{18}$$

$k=2$,

$$E\left((cv)_p\right)^2 = \frac{B\left((n-p+2)/2, n/2\right)}{B\left((n-p)/2, p/2\right)} e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}$$

$$ {}_1F_1\left((n-p+2)/2; (n-p)/2; \left(\sqrt{n}/(CV)_p\right)^2/2\right), \tag{19}$$

$k=3$,

$$E\left((cv)_p\right)^3 = \frac{B\left((n-p+3)/2, n/2\right)}{B\left((n-p)/2, p/2\right)} e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}$$

$$ {}_1F_1\left((n-p+3)/2; (n-p)/2; \left(\sqrt{n}/(CV)_p\right)^2/2\right), \tag{20}$$

$k = 4,$

$$E\left((cv)_p\right)^4 = \frac{B\left((n-p+4)/2, n/2\right)}{B\left((n-p)/2, p/2\right)} e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}$$
$$ {}_1F_1\left((n-p+4)/2;(n-p)/2;\left(\sqrt{n}/(CV)_p\right)^2/2\right).$$

Therefore, we already know

$$V\left((cv)_p\right) = E\left((cv)_p\right)^2 - \left(E(cv)_p\right)^2. \tag{21}$$

Substitute (18) and (19) in (21), we get the variance of $(cv)_p$ as

$$V\left((cv)_p\right) =$$
$$\frac{e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}}{B\left((n-p)/2, p/2\right)}\Bigg[B\left((n-p+2)/2, n/2\right) {}_1F_1\left((n-p+2)/2;(n-p)/2;\left(\sqrt{n}/(CV)_p\right)^2/2\right)$$
$$-\frac{e^{-\left(\sqrt{n}/(CV)_p\right)^2/2}}{B\left((n-p)/2, p/2\right)}\left(B\left((n-p+1)/2, n/2\right) {}_1F_1\left((n-p+1)/2;(n-p)/2;\left(\sqrt{n}/(CV)_p\right)^2/2\right)\right)^2\Bigg].$$

From (18), the sample multivariate coefficient of variation $(cv)_p$ is the biased estimator of population multivariate coefficient of variation $(CV)_p$ because $E\left((cv)_p\right) \neq (CV)_p$. The following simulation graph from Figure 1 shows the shape of the density function of the distribution of sample uni-variate, bi-variate, tri-variate and multivariate coefficient of variation $(cv)_p$ for different values of $\left(p, n, (CV)_p\right)$.

Moreover, the authors also derived the critical points of the sample multivariate coefficient of variation by utilising the relationship in (12) for different values of $\left(p, n, (CV)_p\right)$ and the significance probability is given as $P\left((cv)_p > \left((cv)_p\right)_{p,n,(CV)_p}(\alpha)\right) = \alpha$ and the critical points are visualized in the following Tables 1 and 2.

4. Numerical Results and Discussion

In this section, the authors have shown a numerical study for testing the significance of the sample multivariate coefficient of variation. The sample multivariate coefficient of variation is calculated for the best known Iris plants database in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda and Hart 1973) The data set contains 3 classes of iris plant namely Iris Setosa, Iris Versicolour, Iris Virginica which of each contains 50 instances and the attribute was identified by Fisher as Sepal length (in cm), Sepal width (in cm), petal length (in cm) and petal width (in cm) respectively. The variables in the data set analyzed with the help of Maple 18 trial version (2018). The results of evaluating the bi-variate, tri-variate and multivariate coefficient of variation were presented and discussed with the help of the following table.

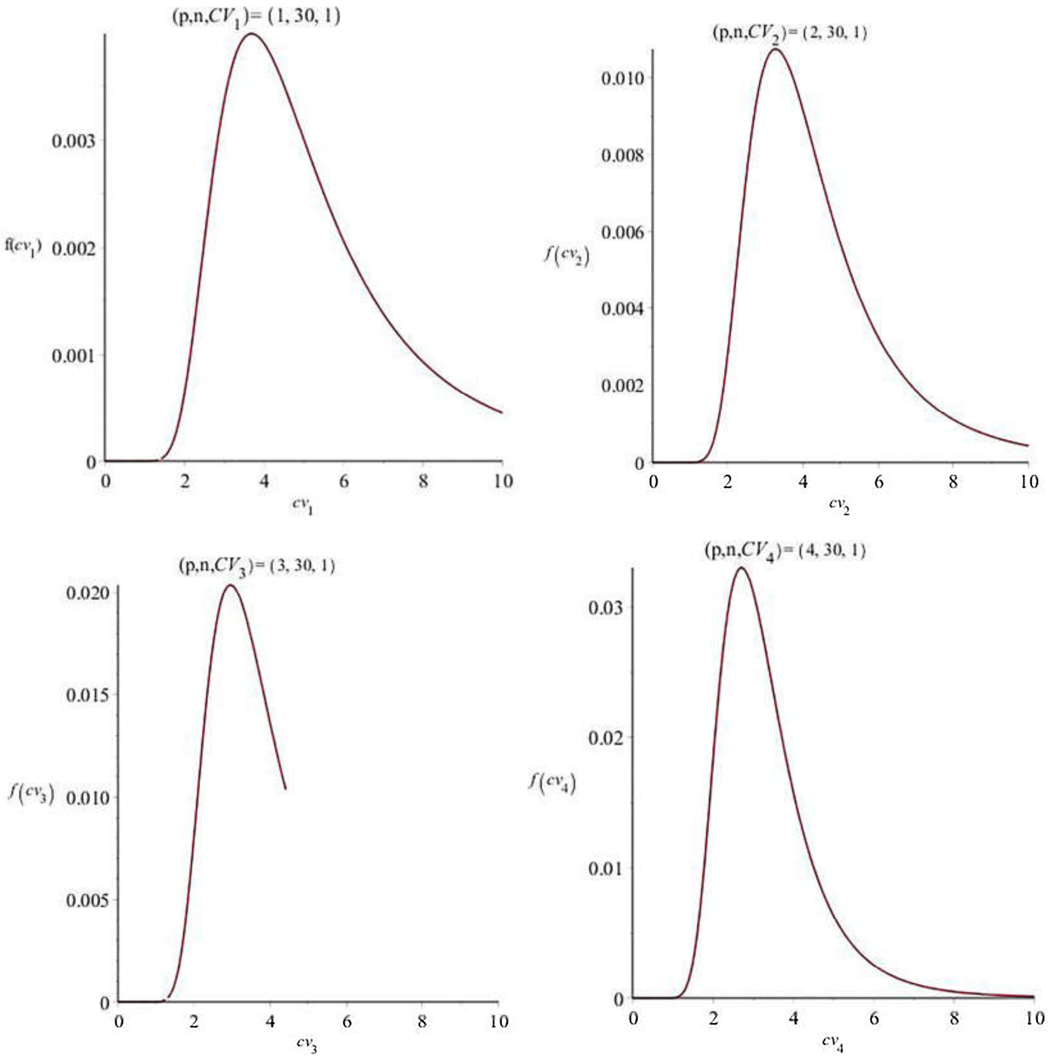

Figure 1 Probability curves of the sample multivariate coefficient of variation

Table 3 visualizes the result of the two-tail test of significance of the sample multivariate coefficient of variation for the 4 variables in Iris Plants Database. As far as uni-variate, bi-variate, tri-variate and multivariate coefficient of variation of the above said are variables less than 1 and for many $(CV)_p$ are close to zero. This shows the variables are highly consistent individually as well as jointly. The sample coefficient of variation of the variables are less than the critical '$(CV)_p$' at 5%, 1% level of significance, hence accept the null hypothesis $H_0 : (CV)_p = 1$ and reject the alternative hypothesis $H_1 : (CV)_p \neq 1$. This shows that the above-mentioned variables have low variation, highly consistent and the uni-variate, bi-variate, tri variate and multivariate distribution of these variables are having low variance.

Table 1 Significant two-tail percentage points of $(cv)_p$ at 5% level

$$P\left((cv)_p > \left((cv)_p\right)_{p,n,(CV)_p}(0.05)\right) = 0.05 \text{ when } (CV)_p = 1$$

n	p									
	1	2	3	4	5	6	7	8	9	10
2	0.21052	0.00000	-	-	-	-	-	-	-	-
3	0.64302	0.21454	0.00000	-	-	-	-	-	-	-
4	0.94947	0.54994	0.25727	0.00000	-	-	-	-	-	-
5	1.18795	0.78434	0.53326	0.30950	0.00000	-	-	-	-	-
6	1.38817	0.96948	0.72592	0.53659	0.35850	0.00000	-	-	-	-
7	1.56368	1.12638	0.88086	0.69986	0.54653	0.40003	0.00000	-	-	-
8	1.72168	1.26469	1.01368	0.83378	0.68771	0.55849	0.43444	0.00000	-	-
9	1.86651	1.38965	1.13161	0.94992	0.80579	0.68253	0.57072	0.46317	0.00000	-
10	2.00095	1.50447	1.23869	1.05382	0.90932	0.78813	0.68116	0.58254	0.48755	0.00000
11	2.12696	1.61124	1.33742	1.14863	1.00259	0.88166	0.77665	0.68197	0.59370	0.50857
12	2.24593	1.71143	1.42946	1.23637	1.08813	0.96649	0.86200	0.76909	0.68406	0.60413
13	2.35893	1.80612	1.51601	1.31839	1.16758	1.04464	0.93988	0.84761	0.76416	0.68691
14	2.46678	1.89613	1.59793	1.39569	1.24206	1.11748	1.01196	0.91966	0.83687	0.76102
15	2.57010	1.98208	1.67590	1.46898	1.31240	1.18595	1.07934	0.98660	0.90392	0.82873
16	2.66944	2.06447	1.75042	1.53882	1.37921	1.25075	1.14285	1.04938	0.96645	0.89146
17	2.76522	2.14371	1.82192	1.60566	1.44297	1.31240	1.20307	1.10868	1.02527	0.95016
18	2.85779	2.22013	1.89074	1.66985	1.50406	1.37132	1.26047	1.16503	1.08096	1.00551
19	2.94746	2.29402	1.95715	1.73169	1.56279	1.42785	1.31540	1.21882	1.13397	1.05804
20	3.03448	2.36560	2.02139	1.79140	1.61941	1.48224	1.36816	1.27038	1.18465	1.10813
21	3.11908	2.43509	2.08366	1.84920	1.67413	1.53473	1.41898	1.31994	1.23328	1.15608
22	3.20144	2.50266	2.14413	1.90526	1.72713	1.58550	1.46806	1.36774	1.28009	1.20216
23	3.28174	2.56845	2.20295	1.95973	1.77857	1.63470	1.51557	1.41394	1.32527	1.24655
24	3.36012	2.63260	2.26024	2.01272	1.82857	1.68248	1.56165	1.45869	1.36898	1.28944
25	3.43671	2.69523	2.31612	2.06437	1.87724	1.72895	1.60642	1.50212	1.41135	1.33097
26	3.51164	2.75644	2.37069	2.11475	1.92469	1.77421	1.64998	1.54434	1.45250	1.37125
27	3.58500	2.81632	2.42403	2.16397	1.97100	1.81835	1.69243	1.58545	1.49252	1.41040
28	3.65688	2.87495	2.47623	2.21210	2.01626	1.86144	1.73384	1.62552	1.53151	1.44849
29	3.72739	2.93241	2.52735	2.25921	2.06052	1.90357	1.77430	1.66464	1.56954	1.48562
30	3.79658	2.98877	2.57746	2.30535	2.10386	1.94479	1.81386	1.70287	1.60667	1.52185
50	4.98262	3.95069	3.42934	3.08688	2.83503	2.63754	2.47610	2.34019	2.22322	2.12083
70	5.93625	4.72056	4.10833	3.70730	3.41317	3.18315	2.99559	2.83810	2.70292	2.58489
90	6.75661	5.38142	4.69005	4.23789	3.90673	3.64810	3.43752	3.26093	3.10956	2.97757
110	7.48762	5.96956	5.20720	4.70908	4.34459	4.06018	3.82880	3.63493	3.46889	3.32423
130	8.15335	6.50475	5.67744	5.13724	4.74220	4.43414	4.18366	3.97390	3.79435	3.63802
150	8.76869	6.99913	6.11161	5.53236	5.10897	4.77893	4.51069	4.28616	4.09404	3.92682
170	9.34358	7.46083	6.51692	5.90110	5.45111	5.10047	4.81557	4.57717	4.37324	4.19580

Table 2 Significant two-tail percentage points of $\left(cv\right)_p$ at 1% level

$$P\left((cv)_p > \left((cv)_p\right)_{p,n,(CV)_p}(0.01)\right) = 0.01 \text{ when } (CV)_p = 1$$

n	p									
	1	2	3	4	5	6	7	8	9	10
2	0.04267	0.00000	-	-	-	-	-	-	-	-
3	0.29517	0.04473	0.00000	-	-	-	-	-	-	-
4	0.54698	0.26344	0.05765	0.00000	-	-	-	-	-	-
5	0.75030	0.47102	0.26830	0.07954	0.00000	-	-	-	-	-
6	0.91941	0.63919	0.44981	0.28505	0.11138	0.00000	-	-	-	-
7	1.06564	0.78022	0.59594	0.44628	0.30657	0.15175	0.00000	-	-	-
8	1.19574	0.90288	0.71927	0.57591	0.45041	0.32946	0.19521	0.00000	-	-
9	1.31385	1.01244	0.82721	0.68609	0.56675	0.45814	0.35193	0.23621	0.00000	-
10	1.42266	1.11216	0.92407	0.78313	0.66646	0.56344	0.46752	0.37314	0.27247	0.00000
11	1.52400	1.20418	1.01252	0.87062	0.75485	0.65450	0.56349	0.47755	0.39278	0.30396
12	1.61920	1.28998	1.09436	0.95082	0.83492	0.73575	0.64726	0.56557	0.48770	0.41084
13	1.70922	1.37065	1.17082	1.02522	0.90859	0.80970	0.72247	0.64309	0.56888	0.49768
14	1.79483	1.44698	1.24282	1.09489	0.97712	0.87798	0.79124	0.71310	0.64098	0.57294
15	1.87660	1.51960	1.31102	1.16060	1.04144	0.94167	0.85495	0.77740	0.70646	0.64030
16	1.95499	1.58898	1.37597	1.22295	1.10221	1.00158	0.91454	0.83716	0.76685	0.70180
17	2.03040	1.65552	1.43808	1.28238	1.15995	1.05829	0.97071	0.89320	0.82316	0.75874
18	2.10314	1.71954	1.49768	1.33927	1.21507	1.11224	1.02397	0.94614	0.87609	0.81198
19	2.17346	1.78129	1.55506	1.39390	1.26787	1.16381	1.07472	0.99641	0.92618	0.86215
20	2.24160	1.84101	1.61043	1.44653	1.31863	1.21326	1.12327	1.04439	0.97384	0.90972
21	2.30774	1.89887	1.66399	1.49735	1.36756	1.26084	1.16989	1.09035	1.01938	0.95505
22	2.37205	1.95504	1.71592	1.54654	1.41484	1.30675	1.21479	1.13451	1.06305	0.99842
23	2.43467	2.00966	1.76633	1.59425	1.46063	1.35113	1.25813	1.17709	1.10506	1.04006
24	2.49573	2.06285	1.81537	1.64059	1.50506	1.39414	1.30007	1.21822	1.14559	1.08015
25	2.55534	2.11471	1.86314	1.68567	1.54824	1.43590	1.34074	1.25805	1.18478	1.11887
26	2.61359	2.16533	1.90972	1.72960	1.59026	1.47649	1.38023	1.29669	1.22276	1.15633
27	2.67058	2.21482	1.95520	1.77246	1.63122	1.51602	1.41866	1.33423	1.25961	1.19265
28	2.72638	2.26322	1.99966	1.81431	1.67119	1.55457	1.45608	1.37078	1.29545	1.22792
29	2.78107	2.31062	2.04317	1.85523	1.71025	1.59219	1.49259	1.40639	1.33035	1.26224
30	2.83470	2.35707	2.08577	1.89528	1.74844	1.62896	1.52824	1.44115	1.36437	1.29567
50	3.74988	3.14565	2.80570	2.56906	2.38813	2.24212	2.12003	2.01532	1.92380	1.84260
70	4.48211	3.77314	3.37581	3.10015	2.89007	2.72106	2.58016	2.45968	2.35470	2.26183
90	5.11052	4.31031	3.86277	3.55285	3.31705	3.12765	2.97002	2.83544	2.71835	2.61494
110	5.66973	4.78761	4.29491	3.95410	3.69507	3.48723	3.31440	3.16700	3.03887	2.92582
130	6.17854	5.22148	4.68739	4.31824	4.03788	3.81307	3.62626	3.46704	3.32872	3.20676
150	6.64852	5.62197	5.04947	4.65400	4.35380	4.11320	3.91337	3.74312	3.59530	3.46501
170	7.08742	5.99577	5.38728	4.96712	4.64830	4.39288	4.18081	4.00021	3.84344	3.70532

Table 3 Test of significance based on sample multivariate coefficient of variation

Variable name	Variables	Dimensions (p)	$(cv)_p$	Critical '$(cv)_p$' 5% sig.	Critical '$(cv)_p$' 1% sig.
Sepal length	X_1	1	0.14171	8.76869	6.64852
Sepal Width	X_2		0.14256		
Petal length	X_3		0.46974		
Petal Width	X_4		0.63555		
Sepal length, Sepal Width	X_1, X_2	2	0.06808	6.99913	5.62197
Sepal length, Petal length	X_1, X_3		0.07611		
Sepal length, Petal Width	X_1, X_4		0.04600		
Sepal width, Petal length	X_2, X_3		0.05038		
Sepal Width, Petal Width	X_2, X_4		0.10104		
Petal length, Petal Width	X_3, X_4		0.14047		
Sepal length, Sepal Width, Petal length	X_1, X_2, X_3	3	0.07351	6.11161	5.04947
Sepal length, Sepal Width, Petal Width	X_1, X_2, X_4		0.05864		
Sepal length, Petal length, Petal Width	X_1, X_3, X_4		0.07854		
Sepal Width, Petal length, Petal Width	X_2, X_3, X_4		0.062292		
Sepal length, Sepal Width, Petal length,	X_1, X_2, X_3, X_4	4	0.087961	5.53236	4.65400

$n = 150$, *p-value <0.01 & p-value <0.05 under $H_0 : (CV)_p = 1$

5. Conclusions

From the previous sections, the authors derived the exact sampling distribution of sample multivariate coefficient of variation $(cv)_p$ and the authors proved $(cv)_p$ is the biased estimator of the true or population $(CV)_p$. The numerical study is shown with a view of testing the significance of the sample multivariate coefficient of variation $(cv)_p$ from which the authors got an insight to check the consistency and the variability of the distribution of the variables. The authors believed that the proposed sampling distribution helps to test the significance of sample $(cv)_p$ from the normal population irrespective of any sample size. Finally, the sampling distribution of the difference and ratios of the two-sample multivariate coefficient of variation can also be derived and the authors left it for future research.

Acknowledgements

The authors are grateful to anonymous referees and the editor for their valuable comments and suggestions which lead us to improve the final version of the manuscript.

References

Albert A, Zhang l. A novel definition of the multivariate coefficient of variation. Biom J. 2010; 52(5): 667-675.

Albrecher H, Teugels JL. Asymptotic analysis of measures of variation. Theor Prob Math Stat. 2007; 74: 1-10.

Aroian LA. A study of R. A. Fisher's Z distribution and the related F distribution. Ann Math Stat. 1941; 12(4): 429-448.

Duda RO, Hart PE. Pattern classification and scene analysis. Hoboken: John Wiley & Sons; 1973.

Fisher RA. On a distribution yielding the error functions of several well-known statistics. Proceedings of the International Congress of Mathematics, Toronto, 1924; 2: 805-813.

Hendricks WA, Robey KW. The sampling distribution of the coefficient of variation. Ann Math Stat. 1936; 7(3): 129-132.

Hürlimann W. A uniform approximation to the sampling distribution of the coefficient of variation. Stat Prob Lett. 1995; 24(3), 263-268.

Reyment R. Studies on Nigerian upper cretaceous and lower territory ostracoda. P.1, Senonian and maastrichtian ostracoda. Stockholm Contributions in Geology. 1960; 7: 1-238.

Sharma KK, Krishna H. Asymptotic sampling distribution of inverse coefficient of variation and its applications. IEEE Trans Reliab. 1994; 43(4): 630-633.

Valen LV. Multivariate structural statistics in natural history. J Theor Biol. 1974; 45(1): 235-247.

Voinov VG, Nikulin MS. Unbiased estimators and their applications, vol. 2, Multivariate Case. Dordrecht: Springer Netherlands; 1996.

Wishart J. The generalised product moment distribution in samples from a normal multivariate population. Biometrika. 1928; 20A (1/2): 32-52.

Maple 18 Trial version [Computer software]. 2018 [Waterloo, Canada; Mar. 21, 2018]; Available from:https://www.maplesoft.com/company/news/releases/2018/2018-03-21-Maple2018-Release.aspx