Statistical Modeling of the Ratio of the Kurtoses Using Two Separate Datasets

Mohammad Reza Mahmoudi, Bui Anh Tuan and Kim-Hung Pho

April 12, 2022

Statistical modeling of the Ratio of the Kurtoses using Two Separate Datasets

Mohammad Reza Mahmoudi [1], Bui Anh Tuan [2], Kim-Hung Pho [3]

[1] Department of Statistics, Faculty of Science, Fasa University, Fasa, Fars, Iran

[2] Department of Mathematics Education, Teachers College, Can Tho University, Vietnam

[3] Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam; phokimhung@tdtu.edu.vn

Abstract. The ratio of the kurtoses of two separate datasets is an interested parameter in many studies. In this paper, the asymptotic distribution for the ratio of the sample kurtoses of two separate datasets is applied to construct the confidence interval and also to hypothesis test for the ratio of the kurtoses of two separate populations. The ability of the introduced technique is studied through simulation study and real example.

AMS Subject Classification: 62H20; 62F12; 62F05; 62F04.

Key words: Kurtosis, Ratio, Hypothesis Testing.

1. Introduction

Based on the literature, to describe a dataset (random variable), three main characteristics contained central tendencies, dispersion tendencies and shape tendencies, are used. A central tendency (or measure of central tendency) is a central or typical value for a random variable that describes the way in which the random variable is clustered around a central value. It may also be called a center or location of the distribution of the random variable. The most common measures of central tendency are the mean, the median and the mode. Measures of dispersion like the range, variance and standard deviation tell us about the spread of the values of a random variable. It may also be called a scale of the distribution of the random variable. The shape tendencies such as skewness and kurtosis describe the distribution (or pattern) of the random variable.

The Pearson's kurtosis and Fisher's kurtosis were respectively defined by β_2 and $\beta_2 - 3$, such that $\beta_2 = \frac{\mu_4}{\mu_2^2}$, where μ_4 and μ_2 are the fourth and the second central moments of population, respectively [1]. Fisher's kurtosis is a measure about flatness of a symmetric distribution versus an equal variance normal distribution. The distribution with small values ($\beta_2<3$) and the distribution with large values ($\beta_2>3$) are respectively called platykurtic and leptokurtic. Also a distribution with $\beta_2=3$ is named mesokurtic. Moors considered the kurtosis as a measure to evaluate the dispersion's rate in 1-sigma bounds [2]. Balanda and MacGillivray studied and review the consepts of kurtosis [3]. DeCarlo showed that the points in the tails have more effects on kurtosis rather than the central points [4]. Also, it was concluded that leptokurtic and platykurtic distributions have fat and thin tails, respectively. Since the skewed datasets are leptokurtic, thus the concept of kurtosis is commonly meaningful for symmetric or approximately symmetric datasets [5].

The ratio of the kurtoses of two random variables (or separate datasets) is an interested parameter in practice. The ratio of the kurtoses is usually more meaningful than the difference of the kurtoses, specially when the values are small. It issues from this fact that the difference of two small values is also small and has no meaningful description. It should be noted that when the ratio of the kurtoses is one, it's equivalent to the equality of the kurtoses. Therefore the results of this research can be specially applied to test the equality of the kurtoses of two independent random variables (or populations).

In this paper, the asymptotic distribution for the ratio of the sample kurtoses (The Pearson's kurtoses) of two separate datasets is applied to construct the confidence interval (CI) and also to hypothesis test for the ratio of the kurtoses of two separate populations. The ability of the introduced technique is studied through simulation study. An approach similar to the references [6-12] will be applied. It should be noted that Coeurjolly et al. [13] introduced *asympTest* R package to implement large sample tests and confidence intervals. They considered two sample mean and variance tests (differences and ratios). The given test statistics were all expressed in the same form as the Student t-test. This contribution also filled the gap of a robust (to non-normality) alternative to the chi-square single variance test for large samples, since no such procedure was implemented in standard statistical software.

2. Methodology

Assume X and Y are two independent random variables with finite i^{th} central moments

$$\mu_{iX} = E(X-\mu_X)^i, \mu_{iY} = E(Y-\mu_Y)^i, i \in \{2,3,\dots\}.$$

Let $X_1,\dots,X_m$, and $Y_1,\dots,Y_n$, be random samples of sizes *m* and *n*, from X and Y. The aim is to study the parameter $K = \frac{K_Y}{K_X}$, such that K_Y and $K_X \neq 0$ are respectively the kurtoses for Y and X.

Let

$$m_{iX} = \frac{1}{m}\sum_{k=1}^{m}(x_k-\bar{x})^i, \qquad m_{jY} = \frac{1}{n}\sum_{k=1}^{n}(y_k-\bar{y})^j, i,j \in \{2,3,\dots\}.$$

As we know the parameters K_X and K_Y can be consistently estimated by $\widehat{K}_X$ and $\widehat{K}_Y$ [14], where

$$\widehat{K}_X = \frac{m_{4X}}{(m_{2X})^2}, \widehat{K}_Y = \frac{m_{4Y}}{(m_{2Y})^2}.$$

Consequently, $\widehat{K} = \frac{\widehat{K}_Y}{\widehat{K}_X}$ is a rational estimator for the parameter K.

Without loss of generality, let $m = n$ (for $n \neq m$, replace n by $n^* = min(m, n)$ in the following) and $\mu_X = \mu_Y = 0$ (for $\mu_X \neq 0$, or $\mu_Y \neq 0$, assume $X - \mu_X$ and $Y - \mu_Y$).

Lemma 1: If the previous assumptions are satisfied, then

$$\sqrt{n}\left(\widehat{K}_X - K_X\right) \xrightarrow{\mathcal{L}} N(0, \delta_X^2) \ , \qquad as \ \ n \to \infty \ \ ,$$

such that

$$\delta_X^2 = \frac{1}{\mu_{2x}^6}\left[4\mu_{4X}^3 - \mu_{2X}^2\mu_{4X}^2 + 16\mu_{2X}\mu_{3X}^2\mu_{4X} - 4\mu_{2X}\mu_{4X}\mu_{6X} + 16\mu_{2X}^3\mu_{3X}^2 - 8\mu_{2X}^2\mu_{3X}\mu_{5X} + \mu_{2X}^2\mu_{8X}\right].$$

Proof: As an application of the Central Limit Theorem (CLT) given in [14], for the vector $\boldsymbol{M}_n = (\bar{X}, \overline{X^2}, \overline{X^3}, \overline{X^4})^T$, it can be concluded that

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

where $\boldsymbol{\mu} = (0, \mu_{2X}, \mu_{3X}, \mu_{4X})^T$,

$$\Sigma = \begin{pmatrix} \mu_{2X} & \mu_{3X} & \mu_{4X} & \mu_{5X} \\ & \mu_{4X} - \mu_{2X}^2 & \mu_{5X} - \mu_{2X}\mu_{3X} & \mu_{6X} - \mu_{2X}\mu_{4X} \\ & & \mu_{6X} - \mu_{3X}^2 & \mu_{7X} - \mu_{3X}\mu_{4X} \\ & & & \mu_{8X} - \mu_{4X}^2 \end{pmatrix}.$$

Now, the Cramer's Theorem [14] can be used to gain the asymptotic distribution of the function

$$g(\boldsymbol{M}_n) = \frac{\overline{X^4} - 4\bar{X}\overline{X^3} + 6\bar{X}^2\overline{X^2} - 3\bar{X}^4}{\left(\overline{X^2} - \bar{X}^2\right)^2}.$$

Let $g: \mathbb{R}^3 \to \mathbb{R}$ as $(x_1, x_2, x_3, x_4) = \frac{x_4 - 4x_1x_3 + 6x_1^2x_2 - 3x_1^4}{\left(x_2 - x_1^2\right)^2}$.

Consequently,

$$\nabla g(0,\mu_{2X},\mu_{3X},\mu_{4X}) = \left(-\frac{4\mu_{3X}}{\mu_{2X}^2}, -\frac{2\mu_{4X}}{\mu_{2X}^3}, 0, \frac{1}{\mu_{2X}^2}\right),$$

and

$$\nabla g(0,\mu_{2X},\mu_{3X},\mu_{4X})\Sigma\big(\nabla g(0,\mu_{2X},\mu_{3X},\mu_{4X})\big)^T = \delta_X^2 ,$$

where ∇ is gradient function.

Because of continuity of ∇g in neighbourhood of $(0,\mu_{2X},\mu_{3X},\mu_{4X})$, as an application of Cramer's Rule [14], it can be concluded that

$$\sqrt{n}\big(g(\boldsymbol{M}_n) - g(0,\mu_{2X},\mu_{3X},\mu_{4X})\big) = \sqrt{n}\big(\widehat{K}_X - K_X\big) \xrightarrow{\mathcal{L}} N(0,\delta_X^2)\ as\ n \to \infty . \blacksquare$$

Next theorem, that is the main part of this work, deals with the limiting distribution of $\widehat{K}$. This theorem will be applied to inference on K.

Theorem 1: If the assumptions of Lemma 1 are satisfied, then

$$\sqrt{n}\big(\widehat{K} - K\big) \xrightarrow{\mathcal{L}} N(0,\lambda^2) , \qquad as\ n \to \infty\ ,$$

such that

$$\lambda^2 = \frac{1}{K^2}\left(\left(\frac{\delta_X}{K_X}\right)^2 + \left(\frac{\delta_Y}{K_Y}\right)^2\right),$$

and

$$\delta_Y^2 = \frac{1}{\mu_{2Y}^6}\left[4\mu_{4Y}^3 - \mu_{2Y}^2\mu_{4Y}^2 + 16\mu_{2Y}\mu_{3Y}^2\mu_{4Y} - 4\mu_{2Y}\mu_{4Y}\mu_{6Y} + 16\mu_{2Y}^3\mu_{3Y}^2 - 8\mu_{2Y}^2\mu_{3Y}\mu_{5Y} + \mu_{2Y}^2\mu_{8Y}\right].$$

Proof: From Lemma 1,

$$\sqrt{n}\left(\widehat{K}_X - K_X\right) \xrightarrow{\mathcal{L}} N(0, \delta_X^2) \ , \qquad as\ n \to \infty \ ,$$

and

$$\sqrt{n}\left(\widehat{K}_Y - K_Y\right) \xrightarrow{\mathcal{L}} N(0, \delta_Y^2), \qquad as\ n \to \infty.$$

Independency of samples and applying Slutsky's Theorem [14] intend to

$$\sqrt{n}\left[\begin{pmatrix}\widehat{K}_X \\ \widehat{K}_Y\end{pmatrix} - \begin{pmatrix}K_X \\ K_Y\end{pmatrix}\right] \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \begin{pmatrix}\delta_X^2 & 0 \\ 0 & \delta_Y^2\end{pmatrix}\right).$$

Let $f: \mathbb{R}^2 \to \mathbb{R}$as $f(x_1, x_2) = \frac{x_2}{x_1}$. Consequently,

$$\nabla f(x_1, x_2) = \left(-\frac{x_2}{{x_1}^2}, \frac{1}{x_1}\right),$$

and

$$\nabla \mathrm{f}(K_X, K_Y)\Sigma\left(\nabla \mathrm{f}(K_X, K_Y)\right)^{\mathrm{T}} = \lambda^2 \ .$$

Because of continuity of ∇f in neighbourhood of (K_X, K_Y), as an application of Cramer's Rule, it can be concluded that

$$\sqrt{n}\left(f\left(\widehat{K}_X, \widehat{K}_Y\right) - f(K_X, K_Y)\right) = \sqrt{n}\left(\widehat{K} - K\right) \xrightarrow{\mathcal{L}} N(0, \lambda^2) \ , as\ n \to \infty \ . \quad \blacksquare$$

Therefore

$$T_n = \sqrt{n}\left(\frac{\widehat{K} - K}{\lambda}\right) = \sqrt{n}\left(\left(\frac{\delta_X}{K_X}\right)^2 + \left(\frac{\delta_Y}{K_Y}\right)^2\right)^{-1/2}\left(\frac{\widehat{K}}{K} - 1\right) \xrightarrow{\mathcal{L}} N(0,1) \ , \qquad as\ n \to \infty \, . \quad (1)$$

This asymptotic will be applied to provide an asymptotic CI and hypothesis testing about the parameter K.

2.1. Constructing CI for the parameter K

As can be seen, the left side of T_n is related to the unknown parameters $\delta_X, \delta_Y, K_X$and K_Y and consequently T_n is not a pivotal quantity to construct CI for the parameter K.

Theorem 2: If the assumptions of Lemma 1 are satisfied, then

$$T_n^* = \sqrt{n}\left(\left(\frac{S_X}{\widehat{K}_X}\right)^2 + \left(\frac{S_Y}{\widehat{K}_Y}\right)^2\right)^{-1/2}\left(\frac{\widehat{K}}{K} - 1\right) \xrightarrow{\mathcal{L}} N(0,1) \;, \qquad as\; n \to \infty \;, (2)$$

such that

$$\hat{\delta}_X^2 = \frac{1}{\mathrm{m}_{2\mathrm{X}}^6}[4\mathrm{m}_{4\mathrm{X}}^3 - \mathrm{m}_{2\mathrm{X}}^2\mathrm{m}_{4\mathrm{X}}^2 + 16m_{2X}\mathrm{m}_{3\mathrm{X}}^2 m_{4X} - 4m_{2X}m_{4X}m_{6X} + 16\mathrm{m}_{2\mathrm{X}}^3\mathrm{m}_{3\mathrm{X}}^2 - 8\mathrm{m}_{2\mathrm{X}}^2 m_{3X}m_{5X} + \mathrm{m}_{2\mathrm{X}}^2 m_{8X}].$$

and

$$\hat{\delta}_Y^2 = \frac{1}{\mathrm{m}_{2\mathrm{Y}}^6}[4\mathrm{m}_{4\mathrm{Y}}^3 - \mathrm{m}_{2\mathrm{Y}}^2\mathrm{m}_{4\mathrm{Y}}^2 + 16m_{2Y}\mathrm{m}_{3\mathrm{Y}}^2 m_{4Y} - 4m_{2Y}m_{4Y}m_{6Y} + 16\mathrm{m}_{2\mathrm{Y}}^3\mathrm{m}_{3\mathrm{Y}}^2 - 8\mathrm{m}_{2\mathrm{Y}}^2 m_{3Y}m_{5Y} + \mathrm{m}_{2\mathrm{Y}}^2 m_{8Y}].$$

Proof: As an application of the Weak Law of Large Numbers (WLLN), it can be concluded that

$$\bar{X} \xrightarrow{p} \mu_X \;,\; \bar{Y} \xrightarrow{p} \mu_Y, m_{iX} \xrightarrow{p} \mu_{iX} \;,\; m_{iY} \xrightarrow{p} \mu_{iY},\; i \in \{2,3,\dots\}, \quad \text{as } n \to \infty.$$

Slutsky's Theorem gives, $\sqrt{n}\left(\left(\frac{S_X}{\widehat{K}_X}\right)^2+\left(\frac{S_Y}{\widehat{K}_Y}\right)^2\right)^{-1/2}\xrightarrow{p}\sqrt{n}\left(\left(\frac{\delta_X}{K_X}\right)^2+\left(\frac{\delta_Y}{K_Y}\right)^2\right)^{-1/2}$, as $n\to\infty$. An application of Theorem 1 completes the proof.

Therefore, T_n^* is a pivotal quantity to construct CI for the parameter K, that is as following:

$$\left(\frac{\widehat{K}}{1+\frac{Z_{\alpha/2}}{\sqrt{n}}\left(\left(\frac{S_X}{\widehat{K}_X}\right)^2+\left(\frac{S_Y}{\widehat{K}_Y}\right)^2\right)^{1/2}},\frac{\widehat{K}}{1-\frac{Z_{\alpha/2}}{\sqrt{n}}\left(\left(\frac{S_X}{\widehat{K}_X}\right)^2+\left(\frac{S_Y}{\widehat{K}_Y}\right)^2\right)^{1/2}}\right). \qquad (3)$$

2.2. Test of hypothesis for the parameter K

Test of hypothesis about the parameter K is very important in real world problems. The case $K=1$ is equivalent to the equality of the kurtoses K_X and K_Y. Generally, the test statistic to test the null hypothesis H_0: $K=K_0$, can be computed by

$$T_0=\sqrt{n}\left(\left(\frac{S_X}{\widehat{K}_X}\right)^2+\left(\frac{S_Y}{\widehat{K}_Y}\right)^2\right)^{-1/2}\left(\frac{\widehat{K}}{K_0}-1\right). \qquad (4)$$

By applying similar technique that was used in Theorem 2, under the null hypothesis H_0, the asymptotic distribution of test statistic T_0 is standard normal.

3. Simulation Study

This section is devoted to study the ability of introduced approach. In the following, different datasets from different distributions and values of (m,n) and K are produced. For each parameter setting, we estimate the coverage probability (that is defined as the percent of the 10000 runs for which the CI contains the true value of K). Also, the Q–Q plots for the values of the test statistic T_0 are plotted. The normality test (Shapiro-Wilk) also is used to investigate the normality for the values of the test statistic T_0. Table 1 reports the values of coverage probability for simulated datasets.

Table 1: The values of coverage probability for simulated datasets

$Distribution$	(K_X, K_Y)	(m, n)					
		$(50, 100)$	$(75, 100)$	$(100, 200)$	$(200, 300)$	$(500, 700)$	$(700, 1000)$
Poisson	(1,1)	0.9427	0.9451	0.9486	0.9500	0.9508	0.9508
	(1,2)	0.9416	0.9461	0.9480	0.9501	0.9509	0.9509
	(2,4)	0.9422	0.9452	0.9474	0.9502	0.9510	0.9509
	(2,5)	0.9427	0.9468	0.9486	0.9502	0.9510	0.9509
Exponential	(1,1)	0.9404	0.9449	0.9477	0.9502	0.9505	0.9507
	(1,2)	0.9418	0.9444	0.9478	0.9502	0.9506	0.9507
	(2,4)	0.9418	0.9441	0.9473	0.9503	0.9506	0.9512
	(2,5)	0.9430	0.9456	0.9490	0.9501	0.9505	0.9510
Binomial	(1,1)	0.9412	0.9454	0.9496	0.9504	0.9505	0.9507
	(1,2)	0.9426	0.9435	0.9498	0.9502	0.9509	0.9507
	(2,4)	0.9400	0.9433	0.9492	0.9503	0.9509	0.9512
	(2,5)	0.9415	0.9462	0.9479	0.9503	0.9509	0.9507

As table 1 indicates, the values of coverage probability are very close to the considered size ($1-\alpha=$0.95), specially when the sample sizes grow, and consequently the introduced approach controlled the type I error. Therefore, proposed CI in Formula (3) is an asymptotic CI for the parameter K. The Figure 1 and Table 2 indicate the p–p plots versus standard normal distribution and the Shapiro-Wilk's normality test for the values of the test statistic T_0.

Figure 1: The normal p–p plots for the values of the test statistic T_0

First Column:

Up: $(K_X,K_Y) = (1,1)$ and $(m,n) = (50,100)$; Down: $(K_X,K_Y) = (1,2)$ and $(m,n) = (75,100)$

Second Column:

Up: $(K_X,K_Y) = (1,2)$ and $(m,n) = (100,200)$; Down: $(K_X,K_Y) = (2,3)$ and $(m,n) = (200,300)$

Third Column:

Up: $(K_X,K_Y) = (2,3)$ and $(m,n) = (500,700)$; Down: $(K_X,K_Y) = (3,5)$ and $(m,n) = (700,1000)$

Table 2: Shapiro-Wilk's normality test P-Value for the values of the test statistic T_0

$Distribution$	(K_X,K_Y)	(m,n)					
		$(50,100)$	$(75,100)$	$(100,200)$	$(200,300)$	$(500,700)$	$(700,1000)$
Poisson	$(1,1)$	0.493	0.581	0.602	0.727	0.834	0.939
	$(1,2)$	0.499	0.501	0.620	0.713	0.871	0.940
	$(2,4)$	0.434	0.561	0.612	0.749	0.806	0.905
	$(2,5)$	0.495	0.574	0.693	0.751	0.821	0.934
Exponential	$(1,1)$	0.484	0.524	0.649	0.711	0.865	0.915
	$(1,2)$	0.456	0.565	0.665	0.777	0.826	0.907
	$(2,4)$	0.460	0.557	0.623	0.746	0.893	0.947
	$(2,5)$	0.488	0.556	0.603	0.726	0.827	0.913
Binomial	$(1,1)$	0.495	0.597	0.670	0.756	0.837	0.943
	$(1,2)$	0.408	0.588	0.606	0.774	0.839	0.926
	$(2,4)$	0.430	0.565	0.671	0.748	0.806	0.912

	(2,5)	0.454	0.581	0.608	0.702	0.833	0.943

Figure 1 indicates that the points in p-p plots are close to the straight line. Also Table 2 shows that the P-Values of Shapiro-Wilk are more than 0.05. Consequently, the asymptotic results are quite acceptable in all simulated datasets and introduced technique acted well to establish CI and test the hypotheses for the ratio of the kurtoses in two separate populations.

4. Real Data

This section is devoted to illustrate the ability of the introduced approach in practical cases. The dataset includes information about the scale of hand, foot and mouth disease outbreaks and the humidity in two Malaysian states (Melaka and Sarawak) from January 2010 to December 2017 (Nelson et al. [15]). The information sources include Malaysian Administrative Modernisation and Management Planning Unit (MAMPU), National Centre for Biotechnology Information (NCBI) and National Oceanic and Atmospheric Administration (NOAA). Tables 3 and 4, respectively, summarized the dataset and the descriptive statistics about of them. As it can be seen, the kurtoses of the humidity in Melaka and Sarawak are respectively 4.058 and 3.098. These values are 9.348 and 4.738 for the scale of hand, foot and mouth disease outbreaks in Melaka and Sarawak, respectively.

Table 3: The scale of hand, foot and mouth disease outbreaks and the humidity in two Malaysian states (Melaka and Sarawak) from January 2010 to December 2017

Month/ Year	**Humidity (%)**		**Scale of Disease Outbreaks**		**Month/ Year**	**Humidity (%)**		**Scale of Disease Outbreaks**	
	Melaka	**Sarawak**	**Melaka**	**Sarawak**		**Melaka**	**Sarawak**	**Melaka**	**Sarawak**

Jan 10	73.3	74.9	11	450	**Jan 14**	71.9	72.9	47	368
Feb 10	74.3	76.2	14	704	**Feb 14**	72.6	73.7	83	677
Mar 10	75.2	75.8	11	1050	**Mar 14**	73.7	74.9	89	1331
Apr 10	76.6	76.3	18	543	**Apr 14**	76.6	75.4	65	1208
May 10	77.4	77	42	381	**May 14**	77.2	76.7	91	1535
Jun 10	76	75.3	26	348	**Jun 14**	77.3	76.8	43	858
Jul 10	75.2	75.1	35	349	**Jul 14**	76.8	75.1	33	669
Aug 10	75.3	74.5	35	425	**Aug 14**	76.1	74.4	26	665
Sep 10	75.3	74.4	17	162	**Sep 14**	75.5	73.9	46	628
Oct 10	75.2	74.5	22	274	**Oct 14**	75.9	74.3	64	455
Nov 10	74.9	74.7	10	174	**Nov 14**	75.9	74.8	122	599
Dec 10	74.2	74.1	4	232	**Dec 14**	75.3	74.4	83	488
Jan 11	73	73.9	3	6	**Jan 15**	72.8	73.5	54	1348
Feb 11	73.5	74.2	4	34	**Feb 15**	72.7	73.8	57	2264
Mar 11	74.7	74.3	14	233	**Mar 15**	75.2	74.3	80	2329
Apr 11	75.5	75.4	6	128	**Apr 15**	77.7	76.1	52	1348
May 11	76.5	75.7	25	109	**May 15**	77.3	76	32	849
Jun 11	76.7	75.1	21	125	**Jun 15**	77.8	75.7	24	312
Jul 11	76	74.8	22	126	**Jul 15**	76.7	75.2	10	100
Aug 11	75.7	75.3	52	277	**Aug 15**	77.3	75.7	37	74
Sep 11	75.8	75.1	13	259	**Sep 15**	76.7	75.7	33	38
Oct 11	75.6	75.3	33	404	**Oct 15**	77	75.9	30	36
Nov 11	75.8	75.2	22	483	**Nov 15**	76.9	76.1	23	35
Dec 11	74.9	75.4	53	763	**Dec 15**	76.7	76.6	11	16
Jan 12	74.4	74.8	13	256	**Jan 16**	77	76.8	25	70
Feb 12	75.4	74.7	63	818	**Feb 16**	75	76.3	39	52
Mar 12	75.8	75.1	129	2698	**Mar 16**	75.4	77	62	70
Apr 12	76.4	75.5	104	1263	**Apr 16**	77.6	76.8	161	240
May 12	76.6	76	210	1060	**May 16**	77.7	76.1	290	393
Jun 12	76.5	74.8	107	1837	**Jun 16**	76.6	75.2	224	707
Jul 12	75.9	75	66	1317	**Jul 16**	76.1	74.9	274	1235
Aug 12	75.9	75.1	44	1080	**Aug 16**	76.4	74.9	403	1313
Sep 12	76.1	74.9	23	1301	**Sep 16**	75.9	74.4	205	892
Oct 12	75.7	75.1	7	764	**Oct 16**	75.8	75	201	623
Nov 12	76	74.9	14	613	**Nov 16**	76.4	75.3	177	561
Dec 12	75.6	75	12	404	**Dec 16**	75.6	75.4	141	473
Jan 13	74.7	74.9	18	683	**Jan 17**	75.1	75.3	94	823
Feb 13	74.8	75.1	41	1273	**Feb 17**	74.8	74.1	142	1363
Mar 13	76.6	75.5	60	2105	**Mar 17**	75.6	74.8	167	1831
Apr 13	77.4	76	41	1136	**Apr 17**	76.4	75.4	145	1310
May 13	77.2	76.4	37	795	**May 17**	76.7	76.3	118	525

Jun 13	75.8	74.8	42	606	**Jun 17**	75.9	75.3	94	388
Jul 13	75.4	73.6	16	289	**Jul 17**	75.8	74.7	103	393
Aug 13	75.4	74	13	217	**Aug 17**	75.5	74.6	69	191
Sep 13	75.3	74.7	31	271	**Sep 17**	75.3	75.3	61	121
Oct 13	75.4	74.6	22	328	**Oct 17**	75.8	74.9	43	83
Nov 13	75.8	74.6	18	521	**Nov 17**	75.8	75.6	34	113
Dec 13	74.6	75.2	21	299	**Dec 17**	74.5	75.3	48	88

Table 4: The descriptive statistics about the scale of hand, foot and mouth disease outbreaks and the humidity in two Malaysian states (Melaka and Sarawak) from January 2010 to December 2017

	State	N	Mean	Std. Deviation	Skewness	Kurtosis
Humidity (%)	Melaka	96	75.7	1.2	-0.823	4.058
	Sarawak	96	75.1	0.8	0.147	3.098
Scale of hand, foot and mouth disease outbreaks	Melaka	96	64.8	69.9	2.282	9.347
	Sarawak	96	646.5	573.6	1.352	4.738

Then the introduced technique is used to establish 95% confidence intervals for the ratio of the kurtoses of the humidity and the ratio of the kurtoses of the scale of hand, foot and mouth disease outbreaks in Melaka respect to Sarawak. Table 5 reports the lower and upper bounds of computed confidence intervals. As it can be seen, the interval (0.941, 2.031) is a 95% confidence interval for the ratio of the kurtoses of the humidity in Melaka respect to Sarawak. Since this interval contains the value 1, hence the hypothesis of the equality of the kurtoses can’t be rejected. Also, the interval (1.166, 6.006) is a 95% confidence interval for the ratio of the kurtoses of the scale of hand, foot and mouth disease outbreaks in Melaka respect to Sarawak. Since this the lower bound is more than 1, it can be concluded that the kurtosis of the scale of hand, foot and mouth disease outbreaks in Melaka is significantly more than the kurtosis of the scale of hand, foot and mouth disease outbreaks in Sarawak. For future works investigation on the further application domains, e.g., [16-38] are suggested.

Table 5: The lower and upper bounds of computed confidence intervals for the ratio of the kurtoses of the humidity and the ratio of the kurtoses of the scale of hand, foot and mouth disease outbreaks in Melaka respect to Sarawak

	Ratio	Lower Bound	Upper Bound
Humidity (%)	1.310	0.941	2.031
Scale of hand, foot and mouth disease outbreaks	1.973	1.166	6.006

5. Conclusion

Kurtosis is a simple but useful statistical tool to make comparisons about independent populations. In real world applications, the researchers may intend to study the equality of the kurtoses in two separate populations to understand the structure of the data. Due to possible small differences of two small kurtoses and no strong interpretation, the ratio of kurtoses is more accurate than the difference of kurtoses. In this study, we proposed the asymptotic distribution, derived the asymptotic confidence interval and established hypothesis testing for the ratio of the kurtoses in two separate populations. The results indicated the coverage probabilities are very close to the considered level, specially when sample sizes were increased, and consequently the proposed method controlled the type I error. Shapiro-Wilk's normality test and Q-Q plot also verified the normality of proposed test statistic. The results verified that the asymptotic approximations were satisfied for all simulated datasets and the introduced technique acted well in constructing confidence interval and performing test of hypothesis.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] K. Pearson, “Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A Rejoinder,” *Biometrika* 4, no. 1/2 (June 1905), 169-212, http://doi.org/10.2307/2331536.

[2] J. J. A. Moors, “The meaning of kurtosis: Darlington reexamined,” *Am Stat* 40, no. 4 (1986), 283-284, http://doi.org/10.1080/00031305.1986.10475415.

[3] K. P. Balanda, H. L. Macgillivray, “Kurtosis: A critical review,” *Am Stat* 40, no. 2 (May 1988), 111-119, http://doi.org/10.2307/2684482.

[4] L. T. DeCarlo, “On the meaning and use of kurtosis,” *Psychol Methods* 2, 1997, 292-307, http://doi.org/10.1037/1082-989X.2.3.292.

[5] K. D. Hopkins, D. L. Weeks, “Tests for normality and measures of skewness and kurtosis: Their place in research reporting,” *Educ Psychol Meas* 50, no. 4 (Decebmer 1990), 717-729, https://doi.org/10.1177/0013164490504001.

[6] H. Haghbin, M. R. Mahmoudi, Z. Shishebor, (2011). “Large Sample Inference on the Ratio of Two Independent Binomial Proportions,” *J Math Ext* 5, no. 1 (2010), 87- 95.

[7] M. R. Mahmoudi, J. Behboodian, M. Maleki, “Inference on the Ratio of Means in Two Independent Populations,” *J Stat Theory and Appl* 16, no. 3 (September 2017), 366-374, https://doi.org/10.2991/jsta.2017.16.3.8.

[8] M. R. Mahmoudi, M. Mahmoodi, “Inferrence on the Ratio of Variances of Two Independent Populations,” *J Math Ext* 7, no. 2 (2014), 83-91.

[9] M. R. Mahmoudi, M. Mahmoodi, “Inferrence on the Ratio of Correlations of Two Independent Populations,” *J Math Ext* 7, no. 4 (2014), 71-82.

[10] M. R. Mahmouudi, M. Maleki, A. Pak, “Testing the Difference between Two Independent Time Series Models,” *Iran J Sci Technol A* 41, no. 3 (September 2017), 665-669, https://doi.org/10.1007/s40995-017-0288-8

[11] M. R. Mahmoudi, M. Mahmoudi, E. Nahavandi, “Testing the Difference between Two Independent Regression Models,” *Commun Stat Theory Methods* 45, no. 21 (August 2016), 6284-6289, https://doi.org/10.1080/03610926.2014.960584.

[12] M. R. Mahmoudi, R. Nasirzadeh, M. Mohammadi, “On the Ratio of Two Independent Skewnesses,” *Commun Stat Theory Methods* 48, no. 7 (May 2019), 1721-1727, https://doi.org/10.1080/03610926.2018.1438627.

[13] J. F. Coeurjolly, R. Drouilhet, P. Lafaye De Micheaux, J. F. Robineau, “asympTest: A Simple R Package for Classical Parametric Statistical Tests and Confidence Intervals in Large Samples,” *R J* 1, no. 2 (December 2009): 26-30, https://doi.org/10.32614/RJ-2009-015.

[14] Thomas S. Ferguson, *A Course in Large Sample Theory* (London: Chapman & Hall, 1996). https://doi.org/10.1201/9781315136288.

[15] B. R. Nelson, H. A. Edinur, M. T. Abdullah, “Compendium of hand, foot and mouth disease data in Malaysia from years 2010 to 2017,” *Data Brief* 24, 103868 (March 2019), https://doi.org/10.1016/j.dib.2019.103868.

[16] Samadianfard, Saeed, et al. "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm." Energy Reports 6 (2020): 1147-1159.

[17] Taherei Ghazvinei, Pezhman, et al. "Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network." Engineering Applications of Computational Fluid Mechanics 12.1 (2018): 738-749.

[18] Qasem, Sultan Noman, et al. "Estimating daily dew point temperature using machine learning algorithms." Water 11.3 (2019): 582.

[19] Mosavi, Amir, and Atieh Vaezipour. "Reactive search optimization; application to multiobjective optimization problems." Applied Mathematics 3.10A (2012): 1572-1582.

[20] Shabani, Sevda, et al. "Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis." Atmosphere 11.1 (2020): 66.

[21] Ghalandari, Mohammad, et al. "Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 892-904.

[22] Mosavi, Amir. "Multiple criteria decision-making preprocessing using data mining tools." arXiv preprint arXiv:1004.3258 (2010).

[23] Karballaeezadeh, Nader, et al. "Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road)." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 188-198.

[24] Asadi, Esmaeil, et al. "Groundwater quality assessment for sustainable drinking and irrigation." Sustainability 12.1 (2019): 177.

[25] Mosavi, Amir, and Abdullah Bahmani. "Energy consumption prediction using machine learning; a review." (2019).

[26] Dineva, Adrienn, et al. "Review of soft computing models in design and control of rotating electrical machines." Energies 12.6 (2019): 1049.

[27] Mosavi, Amir, and Timon Rabczuk. "Learning and intelligent optimization for material design innovation." In International Conference on Learning and Intelligent Optimization, pp. 358-363. Springer, Cham, 2017.

[28] Torabi, Mehrnoosh, et al. "A hybrid machine learning approach for daily prediction of solar radiation." International Conference on Global Research and Education. Springer, Cham, 2018.

[29] Mosavi, Amirhosein, et al. "Comprehensive review of deep reinforcement learning methods and applications in economics." Mathematics 8.10 (2020): 1640.

[30] Ahmadi, Mohammad Hossein, et al. "Evaluation of electrical efficiency of photovoltaic thermal solar collector." Engineering Applications of Computational Fluid Mechanics 14.1 (2020): 545-565.

[31] Ghalandari, Mohammad, et al. "Flutter speed estimation using presented differential quadrature method formulation." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 804-810.

[32] Ijadi Maghsoodi, Abteen, et al. "Renewable energy technology selection problem using integrated h-swara-multimoora approach." Sustainability 10.12 (2018): 4481.

[33] Mohammadzadeh S, Danial, et al. "Prediction of compression index of fine-grained soils using a gene expression programming model." Infrastructures 4.2 (2019): 26.

[34] Sadeghzadeh, Milad, et al. "Prediction of thermo-physical properties of TiO2-Al2O3/water nanoparticles by using artificial neural network." Nanomaterials 10.4 (2020): 697.

[35] Choubin, Bahram, et al. "Earth fissure hazard prediction using machine learning models." Environmental research 179 (2019): 108770.

[36] Emadi, Mostafa, et al. "Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran." Remote Sensing 12.14 (2020): 2234.

[37] Shamshirband, Shahaboddin, et al. "Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases." Mathematics 7.10 (2019): 965.

[38] Salcedo-Sanz, Sancho, et al. "Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources." Information Fusion 63 (2020): 256-272.