Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120121

Ratio of Coefficients of Variation for Comparing the Dispersions of Several Independent Populations

Abbas Bahrampour[1], *Zeynab Avazzadeh*[2,*] *and Mohamad Reza Mahmoudi*[3]

[1]Modeling in Health Research Center, Institute for Futures Studies in Health, Department of Biostatistics and Epidemiology, Faculty of Health, Kerman University of Medical Sciences, Kerman, Iran
[2]Department of Biostatistics and Epidemiology, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran
[3]Department of Statistics, Faculty of Science, Fasa University, Fasa, Iran

Received: 12 Jun. 2022, Revised: 22 Jul. 2022, Accepted: 8 Aug. 2022
Published online: 1 Jan. 2023

Abstract: The coefficient of variation (CV) is an important and useful statistical tool for comparing several populations. In cases where there are multiple populations with different means and variances, the ratio of the coefficients of variation (CVs) is a good way to compare the dispersion of the populations. Because of the possible minor differences between multiple CVs and the lack of a robust interpretation, the ratio of CVs is more accurate than the difference of CVs. When a statistical analysis consists of some simultaneous statistical tests such as equality of several CVs, multiple testing is useful. As an example, the multiple testing about the ratio of CVs is performed to evaluate and compare scale of hand, foot and mouth disease (SHFMD) of 79911 patients between January 2010 and December 2017 in the three Malaysian provinces.

Keywords: Ratio, Coefficient of Variation, HFMD, Test of Hypothesis, Scale of hand, foot and mouth disease(SHFMD), Multiple Testing, Malaysia.

1 Introduction

One of the most typical infectious diseases is hand, foot, and mouth disease (HFMD), which is caused by viruses named category of enteroviruses and the disease's symptoms appear as fever or illness feeling generally. Almost after two days, several smooth spots or even discoloured bumps might blister on the hands, feet and mouth and sometimes groin and buttocks [1]. Normally symptoms become visible after three to six days after exposure by the virus and remove on its own about six-seven days. It is possible fingernails and toenails are lost a few weeks later, but they will regrow by time [2]. The HFMD is specific for human beings and doesn't involve other animals. These viruses are transmitted by close contact person by person. They can be spread by coughing and the faces of an infected people. Normally, specialists diagnose HFMD based on symptoms and occasionally, a throat or stool sample can be helpful to detect the virus [3].

According to the literature, there are three fundamental discriptive criteria whose names are central tendencies, shape tendencies, and dispersion tendencies for a data set. In other words, these three criteria can be used to summarize datasets. One of the well-known tools related to dispersion tendency is the coefficient of variation (CV). CV is obtained by dividing the mean into population's standard deviation, $CV = \sigma/\mu$, and is applicable and suitable statistic which can be benefit to evaluate relative variability. Normally the mentioned statistic is used to compare the dispersion of several groups of data gathered with various units of measurement [4]. CV as a without-dimension parameter is used in diffrent sciences such as agriculture, biology, engineering, finance, medicine, and many others in order to evaluat reliability and variability [5,6,7]. Usually relation between standard deviation and the mean (level of measurement) is vital for researchers. So, it is clear that CV is widely used to evaluate dispersion. Understanding the structure and shape of data is researchers's favorite, therefore, they calculate CVs in order to compare the dispersions of populations. If the means or variances of the two populations are the same, ANOVA and Levens tests can be used to examine the equality of the CVs of populations [8]. In applican, it is possible that researchers need to compare two independent populations's parameters; for instance, proportions [9], means [10], variances [11], correlations [12] and skewnesses [13]. Most comparative tests to date have

* Corresponding author e-mail: zeynab.evazzadeh@gmail.com

used the difference between sample coefficients. However, the use of statistics based on the difference between sample coefficients is not very reliable because the existing distance between two coefficients may be small and imperceptible, and therefore does not allow a clear and unambiguous interpretation. On the other hand, if we use the ratio between two CVs, we obtain a better result. For example, suppose that the CVs of two populations are 0.1 and 0.01, respectively. Their difference is a small amount of 0.09, but their ratio is a relatively large amount of 10 [14]. Yue and Baleanu [15] used asymptotic methods to examine the hypothesis about the ratio of CVs and to construct the confidence interval about it.

The main motivation of this pape is to investigate and compare the scale of HFMD (SHMFD) amongst three provinces of Malaysia. To compare the means and the variances of SHFMD of three regions ANOVA, Scheffe post-hoc technique and Leven's test were used. Then, using the CVS ratio index, the CVs of SHFMD are compared.

2 Patients and Methods

The main objective of this study is to analyse a dataset containing one feature, SHFMD outbreaks from 2010 to 2017 for three provinces in Malaysia (Sarawak, Melaka, and Pulaupinang). Details of the dataset can be found in [16,17]. Various statistical methods such as ANOVA, Scheffe post-hoc technique, Leven test, and CVS ratio index were used. R software is also used in this paper. Only the Stat package is used and the programming is described in all sections. The full description of the two cases is given below.

2.1 Data

The data used in this study are extracted from the "Compendium of hand, foot and mouth disease data in Malaysia from years 2010 to 2017. Data in Brief" under the supervision of the Malaysian Administrative Modernisation and Management Planning Unit, National Centre for Biotechnology Information and National Oceanic and Atmospheric Administration. The dataset includes information about SHFMD in three Malisaian provinceses (Melaka, Sarawak and PulauPinang), starting from January 2010 to December 2017 [16,17].

2.2 Statistical Methods

2.2.1 ANOVA

ANOVA method can be employed to study the difference between the means of several populations. Here, ANOVA method is used to examine the significant difference between SHFMD in Melaka, Sarawak and PulauPinang provinces. If the ANOVA test is significant, different methods, such as the Scheffe method, can be applied for pairwise comparisons.

2.2.2 Leven's Test

Levene's Test is one of the inferential methods for comparing the variances in several independent populations. This test is employed to compare the variances of SHFMD in three provinceses.

2.2.3 CVs Ratio Test

If there are several populations, CV ratio is an appropriate choice for comparing dispersion of them. Due to the possible minor differences of several CVs and the lack of solid interpretation, the CV ratio is more accurate than the CVs difference. In this paper, we use the method given by Yue and Baleanu [15] to compare the CVs. The outline of their method is as following: Suppose that X and Y are two uncorrelated variables, with non-zero means μ_X and μ_Y, respectively. Also, Assume that $X_1,\dots X_m$ and $Y_1,\dots Y_n$ are two uncorrelated samples X and Y respectively. As mentioned earlier in the introduction, the CVs ratio parameter is as following:

$$\gamma = \frac{CV_Y}{CV_X},$$

where CV_X is the CV of the variable X and CV_Y is the CV of the variable Y. Therefore, it is clear that

$$\widehat{\gamma} = \frac{\widehat{CV}_Y}{\widehat{CV}_X},$$

can logically estimate the parameter [15,17] . Yue and Baleanu [15] showed

$$\widehat{\gamma} - \frac{\widehat{\lambda}}{\sqrt{n}} z_{\frac{\alpha}{2}}, \widehat{\gamma} + \frac{\widehat{\lambda}}{\sqrt{n}} z_{\frac{\alpha}{2}}, \tag{1}$$

is a $100(1-\alpha)\%$ confidence interval for the parameter γ, where

$$\lambda^2 = \frac{1}{\mathrm{CV}_X^2} \left(\gamma^2 \delta_X^2 + \delta_Y^2\right).$$

Moreover, they showed that the test statistic

$$T_0 = \sqrt{n} \left(\frac{\widehat{\gamma} - \gamma_0}{\lambda^*} \right),$$

is a suitable statistic to test $H_0: \ \gamma = \gamma_0$, where

$$\lambda^{*2} = \frac{1}{\widehat{\mathrm{CV}}_X^2} \left(\gamma_0^2 \widehat{\delta}_X^2 + \widehat{\delta}_Y^2 \right).$$

Yue and Baleanu [15] proved that the asymptotic distribution of T_0 is standard normal.

2.2.4 Multiple testing

When we wish to investigate many hypotheses simultaneously, we are facing with multiple testing problem. Many times, statistical analysis can consist of several simultaneous statistical tests and any of them can produce a discovery from a similar or dependent dataset. In this situation, multiple testing would be very helpful. The confidence level is usually stated for each test only, but it is more beneficial to have a confidence level for the whole family of simultaneous tests. Gaining the wrong results from multiple comparisons can be missleading intensly in real calculation.

Instance of issue, suppose that a test is performed at 5% significancy level and its null hypothesis is correct, the chance of rejecting the null hypothesis is only 5%. While, if 100 tests are performed and all their null hypotheses are correct, the Type I errors is 0.05. It can be proven that in case of statistical independence of the tests, the probability of at least one incorrect rejection is 99.4%. As mentioned earlier, it can be then generalized the results to the equality of CVs of several populations using different methods of multiple tests like Bonferroni, Q-value and FDR. These tests are simple and useful methods based on p-values and meantime do not require a normality condition. In the Bonferroni method, if we have tests, then the p-values are arranged based on the larger-sized order $(p_1, \ldots, p_M)$ and multiplied by the number of tests, i.e., M. Now the null hypothesis of number is rejected if only if $M^* p_i < \alpha$. That is, α is divided by M, and in fact, we use α/M in the Bonferroni method. The local FDR (LFDR) is the probability that the hypothesis comes from the null at a specific value of the statistic. A means for controling the positive false discovery rate (pFDR) can be provided by the Q-value. Whereas the p-value obtains the expected false positive rate gained by rejecting the null hypothesis for any result with an equal or smaller p-value, the Q-value gives the expected FDR received by rejecting the null hypothesis an equal or smaller Q-value [18,19,20,21,22,23,24].

To examine the equality of the CVs of populations, following test the null hypothesis should be tested

$$H_0: \ \gamma_2 = \gamma_3 = \ldots = \gamma_k = 1$$

where

$$\gamma = \frac{\mathrm{CV}_1}{\mathrm{CV}_i}, \qquad i = 2, \ldots, k.$$

3 Results

3.1 Descriptive Statistics

Based on what was mentioned earlier, the current study is devoted to the comparison of SHFMD from three provinces (Melaka, Sarawak and Pulaupinang) in Malaysia. Table 3.1 shows the descriptive statistics of the dataset, which include the means, standard deviations, and CVs. It can be seen that Sarawak and PulauPinang have the minimum and maximum of SHFMD. The means of SHFMD in these regions are 6220, 62061, and 11630, respectively. The standard deviations of SHFMD in these cities are 69.924, 573.597, and 93.470, respectively. Finally, it is found that the CVs of SHFMD in these regions are 1.079, 0.887, and 0.771, respectively.

Table 1 The descriptive statistics about SHFMD in three regions (Melaka, Sarawak, PulauPinang).

Province	Number of Patients	Mean	Std. Deviation	CV
Melaka	6220	64.791	69.924	1.079
Sarawak	62061	646.468	573.597	0.887
PulauPinang	11630	121.145	93.470	0.771

3.2 Variance's Comparison

Here, Leven's test is used to compare the variances of SHFMD in the studied regions. Table 2 summarizes the results. It can be seen that the variances of SHFMD are significantly different in all three regions ($p < 0.05$). The results presented in Table 2 demonstrate that the dispersion of SHFMD in Melaka is significantly less than in PulauPinang ($p < 0.05$). Moreover, the dispersion of SHFMD in PulauPinang is significantly less than in Sarawak ($p < 0.05$).

Table 2 Leven's Tests for SHFMD in three areas (Melaka, Sarawak and PulauPinang)

Province	Standard Deviation SHFMD
Melaka	69.924^a
Sarawak	573.597^c
PulauPinang	93.470^b

The cities with different letters have significant difference (a: low; c: high).

3.3 Mean's Comparison

To compare the mean values of SHFMD in the provinces, a ANOVA test is used. Table 3 summarizes the results. It can be seen that the p-value of ANOVA is less than 0.05 and consequently the hypothesis of equality of the means of SHFMD in the three regions is rejected. The results indicate that the means of the SHFMD are significantly different in all three provinces. The mean of SHFMD in Melaka is significantly less than in PulauPinang ($p < 0.05$). Moreover, the mean of SHFMD in PulauPinang is significantly less than in Sarawak ($p < 0.05$). Fig. 1(a) shows the error bar graph of the humidity variable, while Fig. 1(b) displays the error bar graph of the scale of HFMD variable.

Table 3 ANOVA and Scheffe test results for the SHFMD in three areas (Melaka, Sarawak, and Pulaupinang)

Province	Mean
Melaka	64.791^a
Sarawak	646.468^c
PulauPinang	121.145^b
P-value of ANOVA	< 0.001

The cities with different letters have significant difference (a: low; c: high).

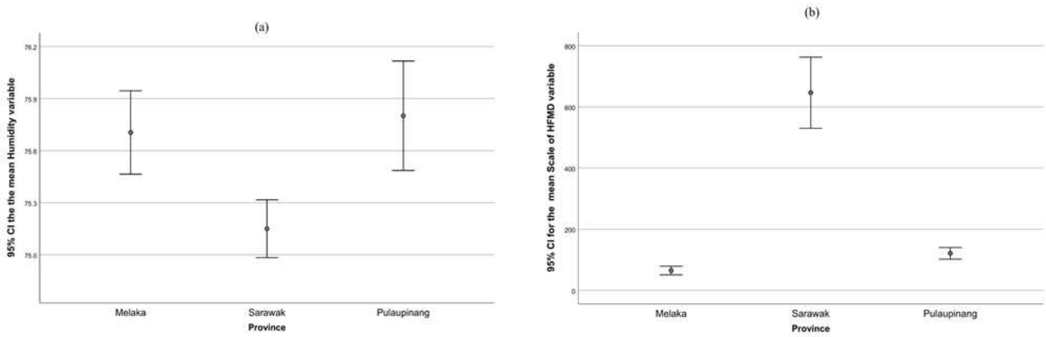

Fig. 1 (a) the error bar graph of the humidity variable (b) the error bar graph of the scale of HFMD variable.

3.4 CV Comparison

In this paper, the method described by Yue and Baleanu [15] is applied to compare the CVs of SHFMD in the provinces. Table 4 summarizes the results. It can be easily observed that all Q-values and LFDR values are less than 0.05. Therefore, the null hypothesis $H_0: \ \gamma_2 = \gamma_3 = 1$ is rejected when the magnitude is 0.05. Therefore, the CVs of SHFMD are significantly different. In addition, Table 4 shows that the ratio of SHFMD in Melaka and Sarawak is 1.216 and the confidence interval $[1.198, 1.499]$ and P-value make it clear that the coefficient of variation in these two areas is not the same and the coefficient of variation in Melaka is higher than that in Sarawak. Also, the ratio of SHFMD in Melaka to Pulaupinang is 1.399. Considering the confidence interval of $[1.191, 2.933]$ and the P value, it can be concluded that the coefficient of variation in Melaka is much higher than Pulaupinang.

The results in Table 5 indicate that the CVs of SHFMD are significantly different in all three provinces. The CV of SHFMD in PulauPinang is significantly lower than in Sarawak ($p < 0.05$). In addition, the CV of SHFMD in Sarawak is significantly lower than in Melaka ($p < 0.05$).

Table 4 Comparison of the CVs of SHFMD in three areas (Melaka, Sarawak, and Pulaupinang).

	Ratio	Test Statistic	p-value	LFDR	Q-value	Lower bound	Upper bound
Melaka respect to Sarawak	1.216	17.345	< 0.001	< 0.001	< 0.001	1.198	1.499
Melaka respect to PulauPinang	1.399	27.337	< 0.001	< 0.001	< 0.001	1.191	2.933

Table 5 Pairwise comparison of CVs of SHFMD in three areas (Melaka, Sarawak, and Pulaupinang).

Province	Mean
Melaka	1.079^c
Sarawak	0.887^b
PulauPinang	0.771^a

The cities with different letters have significant difference (a: low; c: high).

4 Discussion

The aim of this study, which was conducted in three areas in Melaka, Sarawak, and Pulaupinang in Malaysia, was to compare the prevalence of SHFMD in these areas using the "ratio of the coefficient of variation" statistic. The data used in this study were from independent populations. To better understand the structure of the data and to examine the dispersion of SHFMD values, the use of the ratio of the coefficient of variation was proposed. The analysis showed that the ratio of the coefficients of variation is better than the difference between the coefficients of variation, which means that the ratio of the coefficients of variation is better and more accurate than the difference between the coefficients of variation. In this study, the method introduced by Yue and Baleanu [15] with a large sample size is used. Using the methods of multiple testing and based on the hypotheses presented and the statistics given in the article of Yue and Baleanu [15], the ratio of the coefficients of variation of Melaka to Sarawak and Melaka to Pulaupinang was determined. As it turned out, the assumption of equality of coefficients of variation in all three provinces with respect to SHFMD was rejected. One of the reasons for the difference between the SHFMD of Sarawak and Pulaupinang regions and that of Melaka may be the higher population size of Melaka region. The higher the population size, the greater the likelihood that people will become infected with the virus. Meanwhile, direct contact is the main route of transmission for HFMD disease. One of the reasons for the difference in SHFMD could also be the difference in the quality of services in hospitals in these areas. The most important point affecting the study is a large sample. Accordingly, the choice of this research method, i.e., using the "ratio of coefficients of variation," is not appropriate in cases with a small sample size.

5 Conclusion

The HFMD as a relatively intense viral disease has been known and has affected many people around the world. The method of comparing studied in this paper, can be useful and usable in the implementation of treatment plans in different areas or medical centers. One of the most accessible and popular statistical techniques for comparing independent populations in most research areas is the coefficient of variation. In many cases, it is feasible there are several populations with non-identical means and variances whilst their CV are the same. In other words, if researchers intend to understand the data structures in several different populations, they can examine equivalence of CVs in mentioned populations. Additionally, due to several CVs' possible minor differences and the lack of solid interpretation, the CVs ratio is more accurate than the CVs difference. These findings can be exploited in any situation where predictions of outcomes are needed. The obtained results in the study has ability to run with caution to other areas with a high sample size. The executive officials of the health system can use current paper's results which is based on comparison the coefficient of variation of SHFMD. They can identify the places, where the disease is more in order to prevent the spread of the disease by intervening.

For future work, we suggest considering a different disease dataset to compare regions for other epidemic diseases that have many utilities. Additionally, the number of regions could be increased to more than three cases.

Acknowledgement

The authors are grateful for the anonymous referee for the careful checking of the details and for the helpful comments that improved this paper.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] G.L. Repass, W.C. Palmer, F.F. Stancampiano Hand, foot, and mouth disease: Identifying and managing an acute viral syndrome. *Cleve Clin J Med.*, **81** (9), 53743 (2014).
[2] Y.F. Chan, I.C. Sam, K.L. Wee, S. Abubakar, Enterovirus 71 in Malaysia: a decade later. *Neurol. Asia.*, **16**(1), 1-15 (2011).
[3] M. H. Ooi, S.C. Wong, P. Lewthwaite, M.J. Cardosa, T. Solomon, Clinical features, diagnosis, and management of enterovirus 71. *Lancet Neurol.*, **9** (11), 1097-1105 (2010).
[4] C. Long, H. Xu, Q. Shen, X. Zhang, B. Fan, C. Wang, B. Zeng, Z. Li, X. Li, H. Li, Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT. *Eur. J. Radiol.*, **126**, 108961 (2020).
[5] W.W. Daniel, C.L. Cross, *Nonparametric and distribution-free statistics. Daniel WW. Biostatistics: a foundation for analysis in the health sciences*. 6th ed. New York: John Wiley & Sons Inc. (1995).
[6] Q. Meng, L. Yan, Y. Chen, Q. Zhang, Generation of Numerical Models of Anisotropic Columnar Jointed Rock Mass Using Modified Centroidal Voronoi Diagrams. *Symmetry,* **10**(11), 618 (2018).

[7] M. Aslam, M.S. Aldosari, Inspection Strategy under Indeterminacy Based on Neutrosophic Coefficient of Variation. *Symmetry,* **11**(2), 193 (2019).
[8] M. Iglesias-Caamaño, J. Carballo-López, T. Álvarez-Yates, A. Cuba-Dorado, O. García-García, Intrasession Reliability of the Tests to Determine Lateral Asymmetry and Performance in Volleyball Players. *Symmetry*, **10**(9), 16 (2018).
[9] J. Forkman, Estimator and Tests for Common Coefficients of Variation in Normal Distributions. *Commun. Stat.*, **38**(2) 233-51 (2009).
[10] H. Haghbin, M. R. Mahmoudi, Z. Shishebor, Large Sample Inference on the Ratio of Two Independent Binomial Proportions. *J. Math. Ext.*, **5**(1), 87- 95 (2011).
[11] M. R. Mahmoudi, J. Behboodian, M. Maleki, Inference on the Ratio of Means in Two Independent Populations. *J. Stat. Theory Appl.*, **16**(3), 366-374 (2017).
[12] M. R. Mahmoudi, M.M. Mahmoodi, Inferrence on the Ratio of Variances of Two Independent Populations. *J. Math. Ext.*, **7**(2), 83-91 (2014).
[13] M. R. Mahmoodi, M. Inferrence on the Ratio of Correlations of Two Independent Populations. *J. Math. Ext.*, **7**(4), 71-82 (2014).
[14] M. R. Mahmoudi, R. Nasirzadeh, M. Mohammadi, On the Ratio of Two Independent Skewnesses. *Commun. Stat. Theory Methods*, **48**(7), 1721-1727 (2019).
[15] Ferguson, Thomas S. *A Course in Large Sample Theory.* Chapman & Hall: London, UK, (1996).
[16] Z. Yue, D. Baleanu, Inference about the Ratio of the Coefficients of Variation of Two Independent Symmetric or Asymmetric Populations. *Symmetry*, **11**(6), (2019).
[17] 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 Br.*, **24**, 103868 (2019).
[18] M.R. Mahmoudi, B.A. Tuan, K.H. Pho, On kurtoses of two symmetric or asymmetric populations. *J. Comput. Appl. Math.*, **391**, 113370 (2021).
[19] A.R. Nematollahi, A.R. Soltani, M. R. Mahmoudi, Periodically correlated modeling by means of the periodograms asymptotic distributions. *Stat. Pap.*, **58**(4), 126778 (2017)
[20] S. Midway, M. Robertson, S. Flinn, M. Kaller Comparing multiple comparisons: practical guidance for choosing the best multiple comparisons test. *PeerJ*, **8**, e10387 (2020).
[21] M.R. Mahmoudi, On Comparing Two Dependent Linear and Nonlinear Regression Models, *J. Test Eval.*, 47(1), 449-58 (2018).
[22] Y. Benjamini , Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**(1), 125-133 (1995).
[23] M. Kutner, C. Nachtsheim, J. Neter, W. Li, *Applied Linear Statistical Models*, 5th edn. McGraw-Hill Irwin, Newyork, (2005).
[24] J. D. Storey, The positive false discovery rate: a Bayesian interpretation and the Q-value. *Ann. Stat.*, **31** (6), 2013-2035 (2003).