Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.2304.2567

2023, 42(4) 01-07

Capturing the role of rural consumers in sensitive price indicator through geometric consistency check on urban and rural market prices

Humaira Faraz *, Fahim Raees, Mirza Mahmood Baig

Mathematics Department, NED University of Engineering and Technology

* Corresponding author: Humaira Faraz, Emai: humairafraz7@gmail.com

Received: 11 October 2022, Accepted: 25 September 2023, Published: 01 October 2023

KEYWORDS

Geometric Variation

Market Price

Non-Parametric Data

Price Consistency

Q-Q Plot

Shapiro-Wilk Test

ABSTRACT

This paper focuses on capturing price movement and price stability in rural and urban markets. Pakistan collects prices from both markets, but only for the computation of the Consumer Price Index (CPI), which is calculated on a monthly basis. It is proposed to include rural markets by widening the scope of the Sensitive Price Indicator (SPI), computed by the Pakistan Bureau of Statistics (PBS). The inconsistency in the prices of rural markets is observed on the basis of different measures, including descriptive and inferential statistics. Therefore, it is suggested to include rural markets as well in the SPI computation. Large numbers of head counts make transactions in rural areas, which may be taken into account for capturing the accurate weekly consumption pattern of the consumers. Prices of all SPI items were taken from urban and rural markets and checked for normality through the Kolmogorov-Smirnovand Shapiro-Wilk tests. Further, the normality was checked graphically by the Q-Q plot (Quantile-Quantile plot) and histogram. Due to the skewness of the data, non-parametric methods like the geometric coefficient of variation and the coefficient of mean deviation from the median have been applied to check the consistency of the prices in rural markets. The prices of rural markets for further framing of policy to widen the scope of capturing the movement of prices.

1. Introduction

Inflation is a powerful indicator used all over the world to find the changes in the purchasing power of the consumers over a period of time, which could be month over month or year over year. Mishchenko, V. et al. (2018), analyzed deeply in the paper how inflation affects the economic growth of the country [1]. This effect can be positive or negative for the country's economy and monetary policies. Inflation plays a key role in understanding and planning the economy of a country.

According to the World Bank's Pakistan report [2], Pakistan is one of the developing countries, with a population of nearly 220.9 million people. Despite having many natural resources, the per capita income, which is used to measure the average income of a person for an area to calculate the standard of living and the quality of lifestyle of that country's population, does not support them in raising their living standards.

According to the Borgen Project report about poverty in Pakistan, almost 31.3% of Pakistan's population in 2018 lived below the poverty line, which was expected to jump to 40% in a couple of years, which is very high and alarming. Therefore, a small rise in the prices of day-to-day commodities badly affects the budget of a layman.

As per Pakistan Census (2017) conducted by the Pakistan Bureau of Statistics, 36.44% of the population lives in urban areas, while 63.56% lives in rural areas [3]. Therefore, the consumption pattern of the rural population is equally important to cover in SPI as it affects the demand and supply of many producer products.

For the further growth of the economy, constant monitoring of the hikes in the prices of commodities is required. Like other countries, Pakistan also calculates inflation on a regular basis. Among many, the two most important indices formed to calculate inflation are the Consumer Price Index (CPI) and the Sensitive Price Indicator (SPI) by the Pakistan Bureau of Statistics (PBS). SPI is calculated on a weekly basis for 51 items, which include 32 food items. But unlike CPI, it is only calculated by covering urban markets and completely neglecting all rural markets or rural areas. Since a large number of the population in Pakistan lives in rural areas, it is important even for the concerned departments to control the prices in all the administrative units.

Urban and rural areas are defined differently in different regions. Beynon and Craley et al. (2016), explain in the paper that mostly, planners and policymakers define the location. And sometimes this bifurcation is based on the physical dynamics of the locality [4].

In another paper, Alvi (2018) defined that in Pakistan, the two sectors (i.e., urban and rural) are categorized on the basis of population size and growth rate. It was observed that during the last 57 years, the growth rate of urban areas varied from 4.6% to 3.1%, and the growth rate of rural areas ranged from 3.1% to 1.2% [5].

Different countries use different policies to define urban, rural, and sometimes sub rural areas. Therefore, even the United Nations' (UN) sustainable development goals' indicator cannot be used to compare urban and rural areas in different countries. To ease international comparison, Dijkstra, L. et al. (2020), in their report to the World Bank blog, develop new global definitions of cities, towns, sub towns, semi dense areas, and rural areas. The minimum population size for urban areas is different in different countries. Some use the government's administrative decision to define the rural boundaries, while others divide the region into urban and rural areas according to the infrastructure and employment in that area [6].

In this paper, we will check and discuss the fluctuation of the prices of essential commodities in both rural and urban areas. For this purpose, surveys are conducted in the largest city of Pakistan, Karachi. The prices of essential items from rural and urban markets in Karachi have been chosen. By using different statistical tools, which are defined in the methodology sections, consistency in the prices is observed. It was observed that the prices from the rural markets should be collected on a regular basis to monitor the hikes in them, which could badly affect the less privileged or those under the poverty line, especially those living in rural areas.

2. Methodology

To check the consistency of the prices, a survey was conducted in the urban and rural markets of Karachi. The survey form consists of 51 essential items that are used for the calculation of SPI. These items are selected from the basket of goods. It was preferred to pick up the most crowded markets. Five shops from each market were selected on the basis of the maximum number of consumers flowing into those shops. The averages of those eleven markets were taken and further categorised into two average series, namely urban market average prices and rural market average prices.

There are many statistical tools used to determine the consistency of the data. Consistent means that there is less variation in the price of essential items. The purpose of this paper is to study and suggest the inclusion of rural market commodity prices in the calculation of the weekly price index calculated by PBS. To use the appropriate statistical tool, the most fundamental step is to check the normality of the data. Statistical tools or methods for normal data are different from those for non-normal data. The Z-statistic of skewness, the z-statistic of kurtosis, the Kolmogorov-Smirnov test, the Shapiro-Wilk test, the histogram, the Q-Q Plot (quantile-quantile) and the de-trended Q-Q plot are used to check the normality of the two-price series.

2.1 Descriptive Statistics

Descriptive statistics always play a vital role in understanding the data. It summarizes the nature of the data by particularly finding its total count, central value, and dispersion. Skewness and kurtosis values, which are mentioned in Table 1, are used in Table 2 for further investigation of the normality of the data.

The descriptive statistics in Table 1 show that both urban and rural markets follow a skewness trend, which indicates that the data is not normal. For the non-normal data, our focus parameter is the median. Both markets are positively skewed, which indicates that the majority of the prices among the 51 items are accumulated before the median price.

The values are calculated by the given formulas.

$$\text{mean} = \frac{\sum(x)}{n},$$

$$\text{skewness} = \frac{\sum(x-\bar{x})^3}{ns^3}, \text{kurtosis} = \frac{\sum(x-\bar{x})^4}{ns^4}$$

$$SE_{skewness} = \sqrt{\frac{6\times n\times(n-1)}{(n-2)\times(n+1)\times(n+3)}},$$

$$SE_{kurtosus} = 2 \times SE_{skewness} \times \sqrt{\frac{n^2-1}{(n-3)\times(n+5)}}$$

Here n = total number of items, s=standard deviation, SE= standard error

Table 1

The summarized descriptive statistics table of urban and rural market average price

	No. of items (n)	Mean price	Median price	Skewness value	$SE_{skewness}$	Kurtosis value	$SE_{kurtosis}$
Urban market avg. prices	51	307.08	165.64	1.980	0.333	3.253	0.656
Rural market avg. prices	51	296.84	154	2.048	0.333	3.624	0.656

2.2 Skewness and Kurtosis Significance

Different tools are used to check the normality of the data or the distribution formed by the data. Skewness and kurtosis are easy and informative ways to explain the behaviour of the distribution. Aslam, M. (2021), and Bonato, M. et al. (2022), even used skewness and kurtosis for forecasting techniques [7, 8].

Table 2 shows the significant z-values of skewness and kurtosis for both urban and rural market average prices. As mentioned by Ghasemi, A., and Zahediasl, S. (2012), the statistical significance z-value of the skewness is obtained by dividing the skewness value by the standard error of the skewness (i.e., column (a) / column (b) in table 2). If that ratio is larger than 1.96 or smaller than -1.96, it is stated that the effect is significantly different from zero at a P less than 0.05. In small samples, this value is sufficient to establish the normality of the data. In this paper, z-distribution is used to compare the trends, which are derived from z-skewness [9]. Table 2 shows that the z-significant value of skewness for the urban market is 5.946, and for the rural market it is 6.150. These two values are beyond 1.96, which means that there is no symmetry in urban market prices as well as in rural market prices. The null hypothesis of no skews is rejected. Thus, both series are non-normal data.

Similarly, to define the z-statistical significant kurtosis value, take the ratio of the kurtosis value to the standard error of the kurtosis. If the result is not between -1.96 and +1.96, then reject the null hypothesis that the data has no kurtosis. Table 2 shows that the kurtosis z-significant for urban market prices is 4.9588 and for rural market prices it is 5.524 (i.e., column (c) /column (d)). Both values are higher than 1.96. So we reject the claim that the series has no kurtosis. As the values of z-kurtosis and z-skewness are not between +1.96 and -1.96, that means the data is not normally distributed [9].

Formulas used in the table 2 are given below:

$Z_{Skewness}$= (Skewness-0) / $SE_{Skewness}$

$Z_{Kurtosis}$= (Kurtosis-0) / $SE_{Kurtosi}$

Table 2

Significant values of skewness and kurtosis to find the normality of the data

	Skewness			Kurtosis		
	Statistic (a)	Standard Error (b)	Z-Significant (a/b)	Statistic (c)	Standard Error (d)	Z-Significant (c/d)
Urban market avg. prices	1.980	0.333	5.946	3.253	0.656	4.9588
Rural market avg. prices	2.048	0.333	6.150	3.624	0.656	5.524

2.3 Kolmogorov-Smirnov and Shapiro-Wilk Test

Among many, the Kolmogorov-Smirnov and Shapiro-Wilk tests are very useful to check the normality of the data. As proposed by Demir, S. (2022), when skewness and kurtosis statistics are not approaching zero, other normality tests are good to use depending on the size of the data [11]. Normality tests like Kolmogorov-Smirnov and Shapiro-Wilk are applied to urban and rural market price data, which clearly indicate that the data is non-normal. Table 3 shows the significant values of Kolmogorov-Smirnov and Shapiro-Wilk are zero. And to reject the null hypothesis of normality, the significant value of the test should be P 0.05. Ahad et al. (2011), explain in detail that the Kolmogorov-Smirnov test needs a larger sample size (i.e., n $\geq$ 77) to measure the normality of the data, but the Shapiro-Wilk test is always preferred to check the normality of the data even for a small sample size [12].

Table 3

Kolmogorov-Smirnov and Shapiro-Wilk significant values to check the normality of urban and rural market average prices.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Degree of freedom	Significant	Statistic	Degree of freedom	Significant
Urban market avg. prices	0.253	51	0.000	0.721	51	0.000
Rural market avg. prices	0.272	51	0.000	0.713	51	0.000

The above results are obtained by statistical software (SPSS). It was found in Table 3 that the significant values obtained from the test are zero (i.e., P ≤0.05). Therefore, the null hypothesis of normality is rejected.

2.4 Q-Q and De-trended Plots

Graphical methods are always a quick way to understand the trend at a glance. Further normality of the data can be seen through the graphical method known as the Q-Q Plot (Quantile-Quantile Plot). It can be seen in Fig. 3a and 3b. This plot compares the quantiles of the series with each other. As explained by RoduandKafadar (2021) and Yang andBerdine (2021) about Q-Q plots, if the data is lying on a diagonal line$y = x$, it is considered normal. It can be seen in both the Q-Q plot of rural and urban market average prices (Fig. 3a and 3b) that more prices of commodities are accumulated at the beginning of the plot and lie above the diagonal line. Laterally, below the line but not on that straight line, which clearly indicated that the data is not normal, or in other words, both series are non-parametric [13, 14], SPSS also shows the de-trended Q-Q plot (Figs. 3c and 3d), which explains in a single glance the deviation of the data from the horizontal line, or in other words, how much the data is deviated from a normally distributed distribution. So basically, that horizontal line in the de-trended Q-Q plot represents the normal distribution.

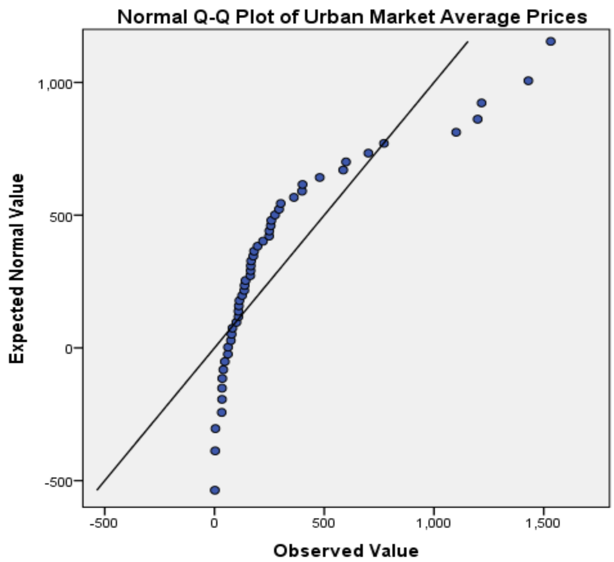

Fig. 3a. Q-Q plot of urban market prices

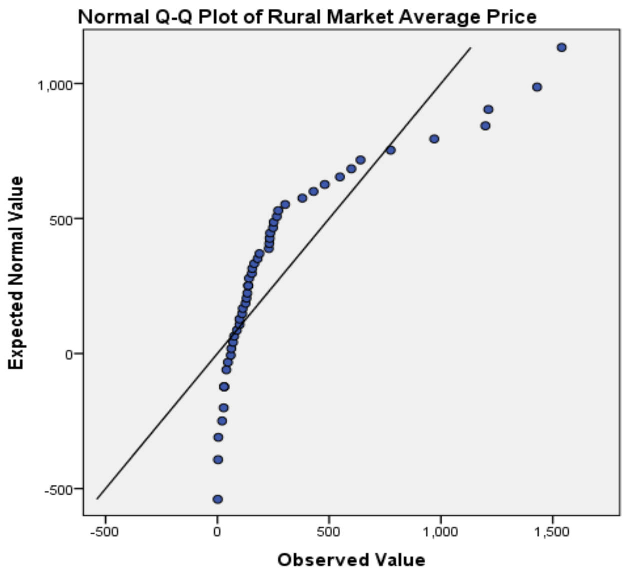

Fig. 3b. Q-Q plot of rural market prices

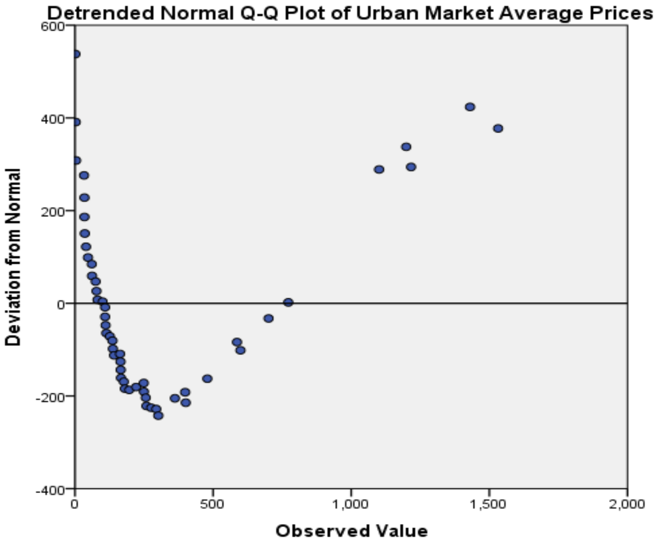

Fig. 3c. De-trended Q-Q plot along market prices

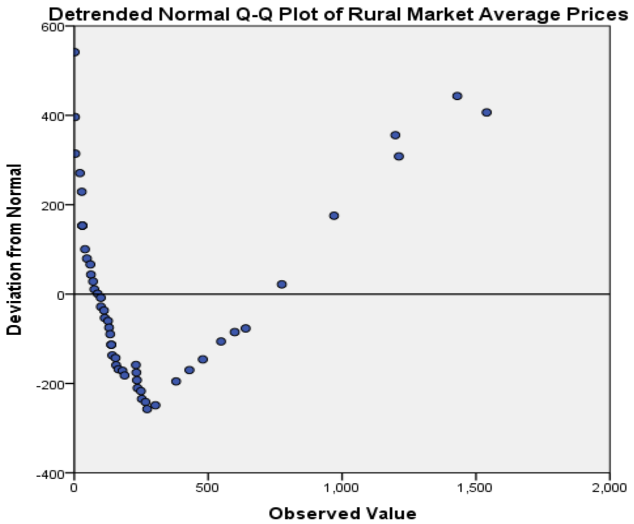

Fig. 3d. De-trended Q-Q plot along zero for rural zero for urban market price

2.5 Coefficient of Variation

To check the consistency or stability of prices in two different markets, different relative measures of dispersion can be used. All the basic tests of relativity, like the coefficient of range, the coefficient of quartile deviation, the coefficient of mean deviation, the coefficient of median deviation, and the coefficient of variance, are well explained by Astana [15].

It was observed that the data is non-normal, so to find consistency in the two sets of data, the preferred measure is the coefficient of mean deviation from the median. For the non-parametric data, the estimated central value is the median. Roenfeldt, K. (2018) explains that when the data is skewed, the geometric mean and geometric coefficient of variation are used. The geometric mean is best in cases where influences from extreme values are observed [16]. The geometric mean always falls in the centre of the data, while the arithmetic mean lies more towards the higher or extreme values.

Geometric Mean

$$\frac{\log(x_1)+\log(x_2)+\log(x_3)+\cdots\ldots\log(x_{n-1})+\log(x_n)}{n} \quad (1)$$

Geometric SD

$$(\sigma) = \sqrt{\frac{\sum(\ln X_i - \ln \mu_g)^2}{n}} \quad (2)$$

Geometric CoV

$$\sqrt{e^{(\ln(10))^2\times\sigma^2}-1}\times 100\% = \sqrt{10^{\ln(10)\times\sigma^2}-1}\times 100\% \quad (3)$$

Here n = total number of items, δ = standard deviation, μ_g = geometric mean

Generally, the coefficient of variation (CoV) formula is the ratio between standard deviation to mean, and it is expressed in percentages (i.e., $\mathrm{CoV} = \frac{\delta}{\bar{x}} \times 100$). Canchola et al. (2017), in his paper explained in detail in their paper that for the log series, this CoV is observed to produce an incorrect result. Therefore, for skewed data, the geometric coefficient of variation is calculated by the formula mentioned in Eq. (3) [17]. Roenfeldt, K. (2018), and Troon, B. (2021) also endorsed that the geometric mean gives the best centre value for non-normal or skewed data, especially when the data is of small size. [16, 18].

3. Findings and Results

By using the defined methodology and formulas, the same result is obtained by using two different formulas for the geometric coefficient of variation.

Result 1

There is a clear difference between the arithmetic mean value and the geometric mean value, as shown in Table 4. Extreme values have a significant impact on the arithmetic mean, whereas the geometric mean ranges close to the median. The coefficient of mean deviation about the median for rural markets is greater than that for urban markets (i.e., 144.8861 > 136.0797), which shows that rural market prices are more inconsistent as compared to urban market prices. (See Shechtman (2013) [19]. A similar result with different values is obtained in the geometric coefficient of variation (i.e., 7.47E+20 > 5.15052E+19).

Arthimatic Mean: $$\frac{\sum_{i=1}^{n} x}{n} \quad (4)$$

$$\text{Coefficient of Mean Deviation about Median} = \frac{\text{mean deviation about median}}{\text{median}} \times 100 \quad (5)$$

$$\text{Mean Deviation about Median: } \frac{\sum_{i=1}^{n}|x-\hat{x}|}{n} \quad (6)$$

Table 4

Important statistical finding to calculate the coefficient of mean deviation about median and geometric coefficient of variation

Statistics	Urban Market Average Price	Rural Markets Averages Price
Mean	307.08	296.84
Geometric Mean	148.61508	138.2626
Median	165.64	154
Standard Deviation	375.3692	371.7101
Geometric Standard Deviation	4.137813	4.257969
Coefficient of Mean Deviation about Median	136.0797	144.8861
Geometric Coefficient of Variation	5.15052E+19	7.47E+20

Result 2

Researchers have also used various approaches to study the geometric coefficient of variation. Another researcher, Humphire, L. (2010), gets the CoV by taking the power of the reciprocal of the geometric mean on the geometric standard deviation. This strategy followed the same underlying structure logic as CoV for the arithmetic mean [20]. The formula that the researcher derived is given in Eq. 7.

All its results are shown in Table 5.

$$\text{CoV(g)} = (\text{Geometric Standard Deviation})^{\frac{1}{\text{Geometric Mean}}} \quad (7)$$

Table 5

Finding of Thinking Applied approach by using Eq. (7) to calculate the coefficient of geometric mean

	Urban Market Average Price	Rural Market Average Price
Geometric Mean	148.61508	138.2626
Geometric Standard Deviation	4.137813	4.257969
Geometric Coefficient Of Variation (*Thinking Applied)*	1.009602	1.010534

By using Eq. (7), the values of the geometric CoV are smaller, as shown in Table 5, and also different from the ones obtained in Table 4. But the interpretation is the same that the geometric CoV of rural market prices has a higher value than the geometric CoV calculated for urban markets. So this method also shows that rural prices are consistent.

4. Conclusion

In this paper, we focused on the variation in the prices of essential items commodities particularly in the rural market of Karachi. According to the PBS 2017 census, 63.56% of the population resides in urban areas. Therefore, markets in rural places have a lot of people buying and selling things, which should be captured to get a true picture of how consumers spend their money each week. For that, two statistical methods have been used. Those are the coefficients of mean deviation through the median and the geometric coefficient of variation. These statistical tools were chosen after finding that the two series, the rural market average price series and the urban market average price series, are non-parametric or non-normal. The Kolmogorov-Smirnov and Shapiro-Wilk tests helped to identify that. Further, Q-Q plots verify the non-normality of the data series.

If we look at the geometric CoV obtained in Table 4, the result shows that rural market prices value is 7.47E+20, which is greater than urban market prices' value of 5.15052E+19. This could be better understood if we take the ratio of these two values of geometric CoV (i.e., 7.47E+20 / 7.1705E+19). The result is14.5019, which is approximately 14:1. This translates to a 14-unit change in rural areas and a 1-unit change in urban areas. So rural prices are not stable and are definitely affecting the inflation rates of a country. This also suggests that fluctuations in price hikes are far greater in rural areas than urban ones. Therefore, the weekly monitoring of prices in rural markets is equally important as it is for urban market prices.

5. Acknowledgment

Author thanks Pakistan Bureau of Statistics for the identification of urban and ruler's consumer markets.

6. References

[1] V. Mishchenko, S. Naumenkova, S. Mishchenko and V. Ivanov, "Inflation and economic growth: The search for a compromise for the Central Bank's monetary policy", Banks and Bank systems, (13, Iss. 2), 153-163, 2018.

[2] World Bank Group, "World development report", World Bank Publications, 2022.

[3] Pakistan Bureau of Statistics, "Provisional summary results of 6th population and housing census", 2017.

[4] M. J. Beynon, A. Crawley and M. Munday, "Measuring and understanding the differences between urban and rural areas", Environment and Planning B: Planning and Design, 43(6), 1136-1154, 2016.

[5] M. H. Alvi, "Difference in the population size between rural and urban areas of Pakistan", 2018.

[6] L. Dijkstra, E. Hamilton, S. Lall and S. Wahba, "How do we define cities, towns, and rural areas", Sustainable Cities, 2020.

[7] M. Aslam, "A study on skewness and kurtosis estimators of wind speed distribution under indeterminacy", Theoretical and Applied Climatology, 143(3), 1227-1234, 2021.

[8] M. Bonato, O. Cepni, R. Gupta and C. Pierdzioch, "Forecasting realized volatility of international REITs: The role of realized skewness and realized kurtosis", Journal of Forecasting, 41(2), 303-315, 2022.

[9] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: a guide for non-statisticians", International Journal Of Endocrinology And Metabolism, 10(2), 486, 2012.

[10] A. A. Razek, B. N. Gadelhak, I. A. El Zahabey, G. A. E. A. Elrazzak and B. Mowafey, "Diffusion-weighted imaging with histogram analysis of the apparent diffusion coefficient maps in the diagnosis of parotid tumours", International Journal of Oral and Maxillofacial Surgery, 51(2), 166-174, 2022.

[11] S. DEMİR, "Comparison of normality tests in terms of sample sizes under different skewness and kurtosis coefficients", International Journal of Assessment Tools in Education, 9(2), 397-409, 2022.

[12] N. A. Ahad, T. S. Yin, A. R. Othman and C. R. Yaacob, "Sensitivity of normality tests to non-normal data", SainsMalaysiana, 40(6), 637-641, 2011.

[13] J. Rodu and K. Kafadar, "The qq boxplot", Journal of Computational and Graphical Statistics, 1-21, 2021.

[14] S. Yang and G. Berdine, "Normality tests", The Southwest Respiratory and Critical Care Chronicles, 9(37), 87-90, 2021.

[15] A. Asthana, measures of dispersion.

[16] K. Roenfeldt, "Better than average: calculating geometric means using SAS", Henry. M. Foundation for the Advancement of Military Medicine, 2018.

[17] J. A. Canchola, S. Tang, P. Hemyari, E. Paxinos and E. Marins, "Correct use of percent coefficient of variation (% CV) formula for log-transformed data", MOJ Proteomics Bioinform, 6(4), 316-7, 2017.

[18] B. Troon, "Empirical comparison of relative precision of geometric measure of variation about the mean and standard deviation", 2021.

[19] O. Shechtman, "The coefficient of variation as an index of measurement reliability: In Methods of clinical epidemiology", pp. 39-49, Springer, Berlin, Heidelberg, 2013.

[20] L. Humphire, "The geometric coefficient of variation", 2010.

