Significance of Price Markets in Karachi and Cost Reduction Factor in Computing Price Index for Food Items

Humaira Faraz*, Fahim Raees **, Mirza Mahmood Baig ***

NED University of Engineering and Technology

* PhD scholar, Department of Mathematics, NED University of Engineering and Technology, Karachi.

**Assistant Professor, Department of Mathematics, NED University of Engineering and Technology, Karachi.

***Professor, Department of Mathematics, NED University of Engineering and Technology, Karachi.

Abstract

In this paper the effort has been made to highlight the same price pattern observed through sample survey in order to exclude the markets which are behaving in the same direction. Findings of this paper may benefit the government in terms of reduction of its human resource and budgetary as well.

Before the application of any statistical technique it is important to check the nature of the data. For this purpose different normality check in applied. Graphically by Q-Q plot (Quantile-Quantile plot). Further Kolmogorov-Smirnov and Shapiro-Wilk test is applied after checking z-skewness and z-kurtosis. Due to skewness in the data, non-parametric method Kruskal-Wallis H test is applied to check whether there is any significant difference among the prices of food items or not, collected from different markets of Karachi city.

Key Words: Non-Parametric data; Q-Q Plot; Kruskal-Wallis H Test; Market Prices.

1. Introduction

Price statistics including Consumer Price Index (CPI), Sensitive Price Indicator (SPI) and Wholesale Price Indicator (WPI) that are computed not only to gauge the inflation of the country but also to frame the short term and long term future policies by keeping in view the directions of economy. As mentioned by Mishchenko, V. et al., (2018), its effect can be positive or negative on the different policies. It is very important to understand that the data collected for price index computation is very sensitive and essential.

It always plays a vital role in the planning and development of any country's economy.

Pakistan is also a developing county and computes the price statistics for the same purpose as discussed by Khan, A., & Gill,R. (2010). Presently this paper focuses on the essential food items

which are mainly and regularly consumed by almost each and every individual in the biggest city of Karachi, Pakistan. Fifty essential food items chosen and prices of the same were collected through personal survey from nine major markets of the city. The criteria of choosing cities was based on the flow of the consumers to markets to buy commodities.

Price patterns was to be compared among different markets in order to observe the price change in various markets. On the basis of the findings of this paper, if the same price pattern in the city is seen then Government may be benefitted to reduce the number of markets which are surveyed on regular basis which includes huge cost and engagement of human resources as well. Pakistan Bureau of Statistics (PBS) mentioned in "Methodology of Price Statistics" that price data collection for the computation of price index is gathered from overall 68 markets of Pakistan which includes 11 markets only from Karachi, by PBS (2019).

2. Methodology

As discussed earlier, this paper is focused to find out whether the amount spend on the collection of prices for the calculation of price indices by the government of Pakistan is needed or not?

For that purpose, a survey is carried out in nine different markets of Karachi. A survey form is particularly listing 50 most used food items that includes vegetables, fruits and essential food items grocery as needed by every income group. Food items are most purchased commodities so their prices collection is very sensitive matter. So prices of food items are collected from these markets. To check the requirement of the number of markets, it is needed to check are those markets prices significantly different from each other's. The normality of the data is checked through Q-Q plot, significant values of skewness and kurtosis, Kolmogorov-Smirnov and Shapiro-Wilk test. As the data found to be non-parametric and skewed therefore to check that is there any statistical difference among all the mean prices of independent markets Kruskal-Wallis H test is applied which is a generalized form of Mann- Whitney U test. Kruskal-Wallis test is the alternate form of one way ANOVA test which is used in case on data is not normal, same mentioned by Johnson, R. W. (2022).

2.1. Descriptive Statistics

Descriptive analysis always help to understand the data. Table 1 is explaining the nature of the data collected from all the nine markets. There is a clear difference in mean and median prices in all the nine markets. It can be clearly observed through table 1 that minimum and maximum prices in all the nine markets is almost same. Fisher, M. J., & Marshall, A. P. (2009), standard deviation and interquartile are important statistical tool to understand the spread of the data. Greater the value of standard deviation mean more is the spread of the distribution. The Interquartile range is considered best if the data is found skewed. It involves 50% of the data, so greater values indicate that the central portion of the data is spread out further.

Table 1: The summarized descriptive statistics table of 50 food items average prices in urban markets.

	No. of items (n)	Mean price	Median price	Standard Deviation	Minimum	Maximum	Inter Quartile Range
Market 1	50	210.22	128	272.49	34.0	1540.0	188.5
Market 2	50	226.61	138	302.96	33.0	1540.0	177.5
Market 3	50	209.94	133	270.25	37.0	1540.0	188.0
Market 4	50	212.26	127	270.69	30.0	1540.0	178.5
Market 5	50	230.8	135	304.50	40.0	1540.0	198.0
Market 6	50	220.3	153	268.18	30.0	1540.0	204.0
Market 7	50	204.67	120	266.74	26.0	1540.0	190.0
Market 8	50	235.12	140	305.92	40.0	1540.0	201.5
Market 9	50	202.95	125	261.87	30.0	1540.0	180.0

2.2. Test for Normality

Skewness and Kurtosis Significance

Skewness and kurtosis values are basic statistic to check the normality of the data that is mentioned in table 2. Skewness and kurtosis are simple way to explain the behavior of the data. Aslam, M. (2021) and Bonato, M. et al., (2022), used skewness and kurtosis techniques for the forecasting procedure.

These statistics are computed through SPSS software. However, the formula used to calculate these values are given below. Demir, S. (2022) mentioned that normal values of skewness and kurtosis statistics should be zero to claim that the data is normally distributed.

The values are calculated by the given formulas.

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

$$SE_{skewness} = \sqrt{\frac{6\times n\times(n-1)}{(n-2)\times(n+1)\times(n+3)}}, \quad SE_{kurtosus} = 2 \times SE_{skewness} \times \sqrt{\frac{n^2-1}{(n-3)\times(n+5)}}$$

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

Formulas used in the table 2 are given below

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

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

Table 2: Skewness and Kurtosis values of 50 food items average prices in Karachi markets.

		Skewness			Kurtosis		
MARKETS	No. of items	Skewness Statistic	$SE_{skewness}$	**Z-skewness**	Kurtosis Statistic	$SE_{kurtosis}$	**Z-kurtosis**
	(n)	(a)	(b)	**(a/b)**	(c)	(d)	**(c/d)**
Market 1	50	3.441	0.337	**10.21**	13.561	0.662	**20.48**

Market 2	50	3.095	0.337	**9.18**	9.870	0.662	**14.91**
Market 3	50	3.430	0.337	**10.18**	13.607	0.662	**20.55**
Market 4	50	3.469	0.337	**10.29**	13.759	0.662	**20.78**
Market 5	50	3.077	0.337	**9.13**	9.706	0.662	**14.66**
Market 6	50	3.338	0.337	**9.91**	13.278	0.662	**20.06**
Market 7	50	3.461	0.337	**10.27**	14.056	0.662	**21.23**
Market 8	50	3.073	0.337	**9.12**	9.635	0.662	**14.55**
Market 9	50	3.518	0.337	**10.44**	14.770	0.662	**22.31**

Table 2 shows the significant z-skewness and z-kurtosis for 50 food items prices obtained from nine markets. As mentioned by Ghasemi, A., & Zahediasl, S. (2012), the statistical significance z-value of the skewness is obtained by divide the skewness value with the standard error of the skewness (i.e. column (a) /column (b) in table 3). Similarly kurtosis z-significant value is obtained by dividing kurtosis statistic with kurtosis standard error. (i.e. column (c) /column (d) in table 3).

Ghasemi, A., & Zahediasl, S. (2012), explain that if ratio of z-skewness or z-kurtosis is larger to 1.96 or smaller to -1.96, it is stated that the effect is significantly different from zero at P less than 0.05. In that case reject the null hypothesis that data has no skewness or no kurtosis.

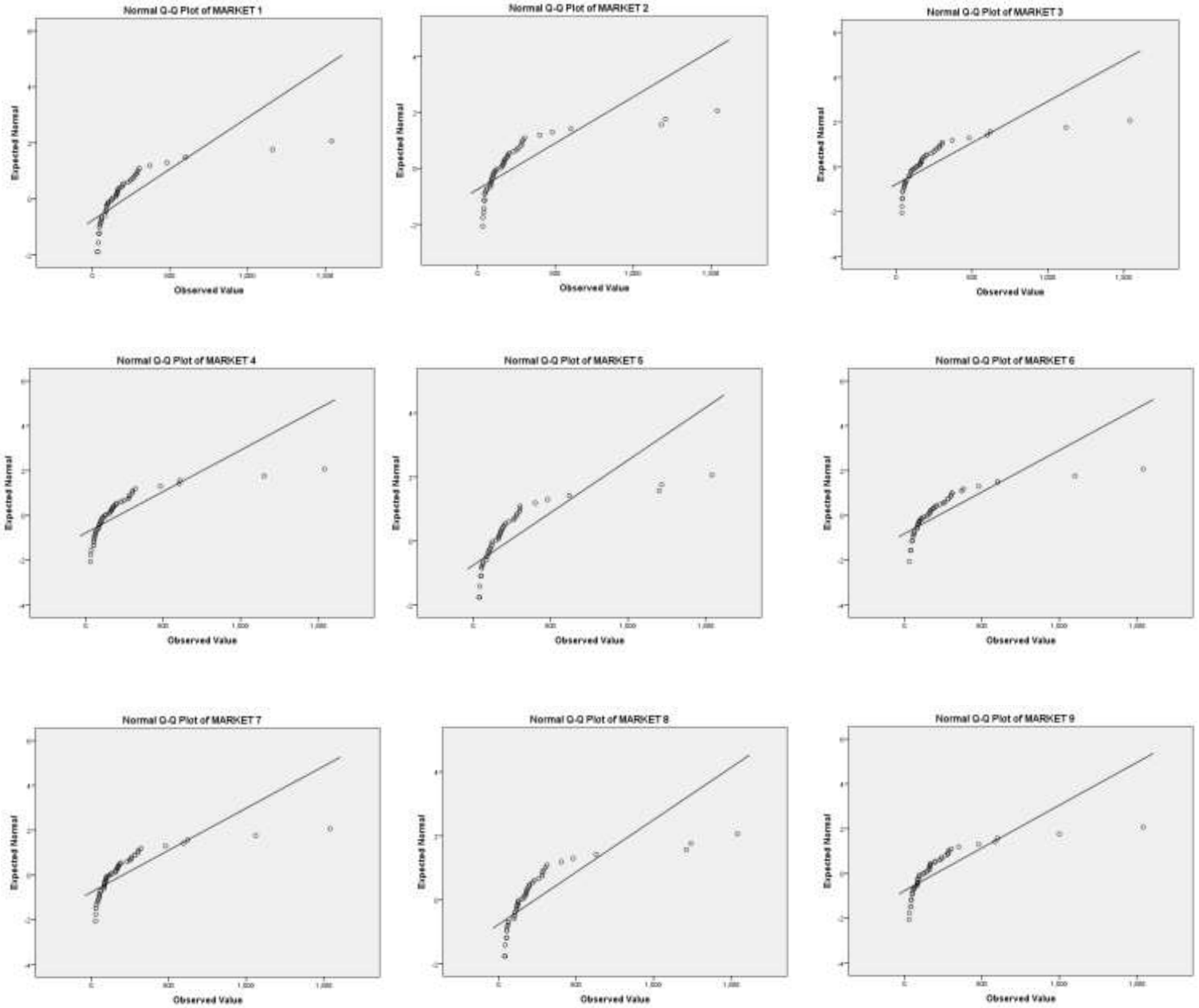
Table 3 shows that all the z-skewness and z-kurtosis values of eleven markets are quite larger than 1.96. Therefore clearly indicating that the obtained data is not normal.

Q-Q Plots

Q-Q Plot (Quantile-Quantile Plot) is the basic graphical method to determine whether the data is normal or not. Through graph it is easy to see that if the data is gathered along a diagonal line $y = x$, than it is considered as normal, Rodu & Kafadar (2021) and Yang & Berdine (2021).

All nine Q-Q plot shows almost same trend. Data observed are not accumulated along the diagonal line$y = x$. Clearly indicating that the observed prices of food items are not normally distributed.

Figure 1. : Q-Q plot of 50 food items average prices from 11 markets in Karachi.



Kolmogorov-Smirnov and Shapiro-Wilk Test

Kolmogorov-Smirnov and Shapiro-Wilk test is considered as the basic statistical tool that is used to find the normality of the data. Ahad et al., (2011) explains that Kolmogorov-Smirnov test results are good if data is lager then 77 observations otherwise for smaller sample size Shapiro-Wilk test is always preferred to check the normality of the data.

Table 3 shows the significant value of Kolmogorov-Smirnov and Shapiro-Wilk are zero which are obtained through SPSS. To reject the null hypothesis of normality the significant value of test should be $P \leq 0.05$. Therefore, through the result it can be stated that the observed values are non-normal.

Table 3: Kolmogorov-Smirnov and Shapiro-Wilk significant values to check the normality of 50 food items average prices in Karachi markets.

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Degree of freedom	Significant	Statistic	Degree of freedom	Significant
Market 1	0.259	50	0.000	0.588	50	0.000
Market 2	0.280	50	0.000	0.581	50	0.000
Market 3	0.261	50	0.000	0.593	50	0.000
Market 4	0.250	50	0.000	0.585	50	0.000
Market 5	0.286	50	0.000	0.581	50	0.000
Market 6	0.239	50	0.000	0.621	50	0.000
Market 7	0.251	50	0.000	0.595	50	0.000
Market 8	0.281	50	0.000	0.581	50	0.000
Market 9	0.254	50	0.000	0.594	50	0.000

2.3. Kruskal-Wallis Test

The Kruskal-Wallis test is a non-parametric test that is used to find the significant difference among the median values of observed from independent sample groups. McKight, P., & Najab, J. (2010), mention that for normal data one way ANOVA test is used that is a parametric test. Similarly for non-parametric data, the Kruskal-Wallis H test that is a generalized form of Mann-Whitney U (Wilcoxon rank) test. Table 4 is showing the results of obtained by SPSS.

Formula and detail working is mention by Ostertagova, E., et al, (2014),

Formula used to calculate Kruskal-Wallis test,

$H = \frac{12}{N(N+1)}(\sum \frac{R_i^2}{n_i}) - 3(N+1)$; Where

$n_i = the\ total\ number\ of\ obseration\ in\ the\ ith\ sample.$

$R = the\ rank\ sum\ of\ the\ ith\ sample.$

$N = the\ total\ number\ of\ sample\ points.$

Table 4a: **Kruskal-Wallis Test** values to check the normality of 50 food items average price Karachi markets.

Prices of Markets	Mean Rank
1	219.58
2	219.05
3	223.88
4	232.67
5	234.35
6	228.38
7	205.75
8	237.28
9	214.13

Table 4b: **Kruskal-Wallis Test** result through SPSS

Test Statistic	Chi-square	Degree of freedom (df)	Asymptotic. significance
	2.514	8	**0.961**

3. Results

Kruskal-Wallis Analysis

i) The null hypothesis : There is no significant difference among the median prices of all the nine markets

The alternative hypothesis: The nine markets are significantly different. (Or at least one of the median rating is different from other)

ii) Critical region is 5% or p value is less than 0.05

iii) $\chi^2(\ calculated) = 2.51$

$\chi^2(\ tabulated\ at\ 0.05, 8) = 15.507$

iv) From the result it can be concluded that as $\chi^2(\ tabulated) > \chi^2(\ calculated)$

i.e. 15.507 > 2.51 there null hypothesis is accepted that there is no significant difference among the median prices in the markets.

v) Asymptotic significant value i.e. p=0.961 which is also greater then p = 0.05, thus clear indication to accept null hypothesis.

4. Conclusion

On the basis of different statistical tests, it may be concluded that prices of all markets in Karachi city are significantly same which clearly indicates that the reduction of number of markets, surveyed for the computation of price indices , may be beneficial for the Government in terms of human resources and budget side as well.

All normality test conducted graphically or statistically give the same result for all the nine markets taken during the survey. Kruskal-Wallis H test, clearly indicated that all the markets are

significantly same. The asymptotic significant p value is very high i.e. 0.961. In other words it can be claim that till 96.1 % chances that median prices of 11 markets will be same. Only 3.9% chances that prices among the markets will be different which again raise the question that out of nine how many markets have different prices with each other's.

References

Ahad, N. A., Yin, T. S., Othman, A. R., & Yaacob, C. R. (2011). Sensitivity of normality tests to non-normal data. *Sains Malaysiana*, *40*(6), 637-641

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

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

DEMİR, S. (2022). 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.

Fisher, M. J., & Marshall, A. P. (2009). Understanding descriptive statistics. *Australian critical care*, *22*(2), 93-97.

Johnson, R. W. (2022). Alternate Forms of the One-Way ANOVA F and Kruskal–Wallis Test Statistics. *Journal of Statistics and Data Science Education*, *30*(1), 82-85.

Khan, R. E. A., & Gill, A. R. (2010). Determinants of inflation: A case of Pakistan (1970-2007). *Journal of economics*, *1*(1), 45-51.

McKight, P. E., & Najab, J. (2010). Kruskal-wallis test. *The corsini encyclopedia of psychology*, 1-1.

Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. In *Applied Mechanics and Materials* (Vol. 611, pp. 115-120). Trans Tech Publications Ltd.

PBS (2019). "Methodology of Price Statistics". https://www.pbs.gov.pk/

Rodu, J., & Kafadar, K. (2021). The qq boxplot. *Journal of Computational and Graphical Statistics*, (just-accepted), 1-21.

Yang, S., & Berdine, G. (2021). Normality tests. *The Southwest Respiratory and Critical Care Chronicles*, *9*(37), 87-90.

AUTHORS

First Author – Humaira Faraz, PhD. Scholar in Department of Mathematics, NED University of Engineering and Technology, Karachi, Pakistan humairafraz7@gmail.com

Second Author – **Fahim Raees**, Ph.D. in Computational Fluid Dynamics, Assistant Professor NED University of Engineering and Technology, Karachi, Pakistan, fahimned@neduet.edu.pk

Third Author- Mirza Mehmood Baig, Ph.D. in Computer Science, Professor NED University of Engineering and Technology, Karachi, Pakistan, baig@neduet.edu.pk

Correspondence Author – **Humaira Faraz**, humairafraz7@gmail.com, +92-333-2376277