Aslam *Journal of Big Data* (2023) 10:24
https://doi.org/10.1186/s40537-023-00700-z

Journal of Big Data

RESEARCH Open Access

Data analysis for sequential contingencies under uncertainty

Muhammad Aslam*

*Correspondence:
aslam_ravian@hotmail.com

Department of Statistics, Faculty of Science, King Abdulaziz University, Jeddah 21551, Saudi Arabia

Abstract

The existing Z-test for comparing sequential contingencies under classical statistics can be implemented only in the presence of certain frequencies, and the level of significance. The existing Z-test for comparing sequential contingencies cannot be applied when uncertainty/indeterminacy is found in observed frequencies, and the level of significance. To apply frequencies, Z-test for comparing sequential contingencies under indeterminate environment its modification under neutrosophic statistics will be given in this paper. The decision procedure of the Z-test for comparing sequential contingencies under neutrosophic statistics will be given with the help of an example selected from the psychology field. From the comparison, the proposed Z-test for comparing sequential contingencies was found to be more effective and more informative than the Z-test for comparing sequential contingencies.

Keywords: Classical statistics, Neutrosophic statistics, The power of the test, Simulation, Physiology data

Introduction

The Z-test has been applied to investigate whether the means of two underlying populations are different. Usually, Z-test is applied when the population variance is assumed to be known and the sample size is larger than 30. The Z-test is applied to test the null hypothesis that two population means are equal vs. the alternative hypothesis that two population means are not equal. The null hypothesis is rejected if the calculated Z-test value is greater than the tabulated value at a specific level of significance. Z-test for comparing sequential contingencies has been applied for investigating whether there is a significant difference in sequential connection across the groups in 2×2 contingency table [8]. This type of test utilizes the idea of logit function and logit transformation. The logit function is based on a quantile function that is associated with the standard logistic distribution and has many applications in data transformation and data analysis, see [https://en.wikipedia.org/wiki/Logit]. According to Holland [7] "the logit transformation is the log of the odds ratio, that is, the log of the proportion divided by one minus the proportion". The Z-test for comparing sequential contingencies used the logit transformation for testing the null hypothesis of independence vs. the alternative hypothesis that two characteristics are associated. The efficiency of the test can be evaluated using the power of the test. The power of the test is stated as the chance of rejecting the null hypothesis when it is false. The high power of the test indicates a high probability of perceiving a true effect. Kanji [8] applied the Z-test for comparing sequential contingencies in assessing the behavior of spouse behavior. Meeker [13] proposed the sequential test for 2×2 contingency table. Rayalu et al. [14] discussed the application of the test in pharmacy. Amiri and Modarres [1] discusses the advantages of contingency tables. Lee et al. [11] presented the associated test for small size sequencing data. More applications of the test can be seen [12].

Neutrosophic statistics is an extension of classical statistics and is applied to analyze the data having neutrosophic numbers, see [16]. Neutrosophic statistics has been applied to analyze and interpret indeterminate data. Chen et al. [4, 5] discussed the applications of neutrosophic statistics. In practice, the decision-makers get more indeterminate data than determinate data due to the complex process; therefore, neutrosophic statistics get attention due to the application for indeterminate data. Duan et al. [6], Khan et al. [9, 10] proposed the neutrosophic exponential distribution, gamma distribution, and Rayleigh distribution, respectively. More applications of neutrosophic statistical tests can be seen in Aslam et al. [2, 3] and Sherwani [15].

The existing Z-test for comparing sequential contingencies under classical statistics cannot be applied when uncertainty/indeterminacy is found during the implementation of the test. By exploring the literature and according to the best of the author's knowledge, there is no work on Z-test for comparing sequential contingencies using neutrosophic statistics. In this paper, the modification of the Z-test for comparing sequential contingencies using neutrosophic statistics will be presented. The application of the proposed Z-test for comparing sequential contingencies will be given with the help of an example. It is expected that the proposed Z-test for comparing sequential contingencies under neutrosophic statistics will perform better than the existing Z-test for comparing sequential contingencies in terms of information.

Methodology

The existing Z-test for comparing sequential contingencies using classical statistics is only applied when the decision-makers are certain about parameters, level of significance, and observations. In practice, particularly in the testing of a hypothesis, the uncertainty about frequency and or level of significance is always presented. The existing Z-test for comparing sequential contingencies using classical statistics cannot be applied in uncertain situations. This section presents the modification of the Z-test for comparing sequential contingencies under neutrosophic statistics. The methodology of the proposed Z-test for comparing sequential contingencies under neutrosophic statistics is explained as follows: under neutrosophy, let W_{tN} be person's antecedent behaviors and can assume one of the following values [8].

$$W_{tN} = \begin{cases} 1 \; \textit{for negative effect} \\ 0 \; \textit{for positive effect} \end{cases} \quad (1)$$

Let H_{tN+1} be spouse's consequent behaviors and can assume one of the following values

$$H_{tN+1} = \begin{cases} 1 \; for\ negative\ effect \\ 0 \; for\ positive\ effect \end{cases} \quad (2)$$

The logit transformation has been applied to investigate the association in contingency tables and used to find the marginal total using sensitive or insensitive row totals. The neutrosophic logit transformation can be expressed as follows

$$logit(P_N) = logit(P_L) + logit(P_U)I_{P_N};\, I_{P_N} \epsilon \left[I_{P_L}, I_{P_U}\right] \quad (3)$$

Note here that $logit(P_L)$ presents the lower logit transformation (determinate logit transformation), and $logit(P_U)I_{P_N}$ present the upper logit transformation (indeterminate logit transformation) and I_{P_N} presents the measure of indeterminacy associated with neutrosophic logit transformation.

The neutrosophic logit transformation can be written as

$$logit(P_N) = log_e\left(\frac{P_L}{1-P_L}\right) + log_e\left(\frac{P_U}{1-P_U}\right)I_{P_N};\, I_{P_N} \epsilon \left[I_{P_L}, I_{P_U}\right] \quad (4)$$

Note here that the first value on the right side of Eq. (4) presents the determinate part; the second value presents the indeterminate part and $I_{P_N} \epsilon \left[I_{P_L}, I_{P_U}\right]$ is the measure of uncertainty associated with neutrosophic logit transformation. The neutrosophic statistic $\beta_{iN} \epsilon [\beta_{iL}, \beta_{iU}]$ is based on the logarithm of the "odds ratio" and given by [8]

$$\beta_{iN} = \beta_{iL} + \beta_{iU}I_{\beta_N};\, I_{\beta_N} \epsilon \left[I_{\beta_L}, I_{\beta_U}\right] \quad (5)$$

where β_{iL} is determinate and is defined by

$$\beta_{iL} = logit\left[P_r\left(H_{tL+k} = 1|W_{tL} = 1\right)\right] - logit\left[P_r\left(H_{tL+k} = 1|W_{tL} = 0\right)\right]$$

and $\beta_{iU}I_{\beta_N}$ is indeterminate part, where I_{β_N} is measure of indeterminacy and β_{iU} is given by

$$\beta_{iU} = logit\left[P_r\left(H_{tU+k} = 1|W_{tU} = 1\right)\right] - logit\left[P_r\left(H_{tU+k} = 1|W_{tU} = 0\right)\right]$$

Let us have a 2×2 contingency table having neutrosophy (Table 1).

Note that the first values in each cell show the determinate part, the second values show the indeterminate part, and I_{a_N} shows the indeterminacy for the first cell. By following [8], the values of $\beta_{i1N} \epsilon [\beta_{i1L}, \beta_{i1U}]$ for antecedent behaviors can be computed as

Table 1 The neutrosophic 2×2 contingency table

	Antecedent behaviors W_{tN+1}		**Non-antecedent behaviors W_{tN+1}**	
H_{tN}	**1**	**0**	**1**	**0**
1	$a_L + a_U I_{a_N}$	$b_L + b_U I_{b_N}$	$a_L + a_U I_{a_N}$	$b_L + b_U I_{b_N}$
0	$c_L + c_U I_{c_N}$	$d_L + d_U I_{d_N}$	$c_L + c_U I_{c_N}$	$d_L + d_U I_{d_N}$

$$\beta_{i1N} = \log\left(\frac{a_L d_L}{b_L c_L}\right) + \log\left(\left(\frac{a_U d_U}{b_U c_U}\right)\right) I_{\beta_{1N}};\ I_{\beta_{1N}} \epsilon \left[I_{\beta_{1L}}, I_{\beta_{1U}}\right];\ i = 1, 2 \quad (6)$$

where $\log\left(\frac{a_L d_L}{b_L c_L}\right)$ and $\log\left(\left(\frac{a_U d_U}{b_U c_U}\right)\right) I_{\beta_{1N}}$ present the determinate and indeterminate parts, respectively and $I_{\beta_{1N}}$ is the measure of indeterminacy.

The values of $\beta_{i2N} \epsilon [\beta_{i2L}, \beta_{i2U}]$ for non-antecedent behaviors can be computed as

$$\beta_{i2N} = \log\left(\frac{a_L d_L}{b_L c_L}\right) + \log\left(\left(\frac{a_U d_U}{b_U c_U}\right)\right) I_{\beta_{2N}};\ I_{\beta_{2N}} \epsilon \left[I_{\beta_{2L}}, I_{\beta_{2U}}\right];\ i = 1, 2 \quad (7)$$

where $\log\left(\frac{a_L d_L}{b_L c_L}\right)$ and $\log\left(\left(\frac{a_U d_U}{b_U c_U}\right)\right) I_{\beta_{2N}}$ present the determinate and indeterminate parts, respectively and $I_{\beta_{2N}}$ is measure of the indeterminacy.

The following test statistic $Z_N \epsilon [Z_L, Z_U]$ is the extension of the test statistics proposed by [8] and will be applied for testing whether $\beta_{iN} \epsilon [\beta_{iL}, \beta_{iU}]$ is different across groups

$$Z_N = \frac{(\beta_{i1L} - \beta_{i2L})}{\sqrt{\sum\left(\frac{1}{f_{iL}}\right)}} + \frac{(\beta_{i1U} - \beta_{i2U})}{\sqrt{\sum\left(\frac{1}{f_{iU}}\right)}} I_{\beta_N};\ I_{\beta_N} \epsilon \left[I_{\beta_L}, I_{\beta_U}\right] \quad (8)$$

where $f_{iN} \epsilon [f_{iL}, f_{iU}]$ be the ith cell frequency, $Z_N \epsilon [Z_L, Z_U]$ be a neutrosophic standard normal distribution and I_{β_N} is the measure of indeterminacy associated with $Z_N \epsilon [Z_L, Z_U]$. Suppose that the decision-makers are uncertain about the level of significance. Let $\alpha_N = \alpha_L + \alpha_U I_{\alpha_N};\ I_{\alpha_N} \epsilon \left[I_{\alpha_L}, I_{\alpha_U}\right]$ be the neutrosophic form of the level of significance. Note here that α_L presents the level of significance when decision-makers are uncertain about it, the second part $\alpha_U I_{\alpha_N}$ denotes the indeterminate part and $I_{\alpha_N} \epsilon \left[I_{\alpha_L}, I_{\alpha_U}\right]$ is the measure of indeterminacy associated with the level of significance. The value of $Z_N \epsilon [Z_L, Z_U]$ and computed and compared with the tabulated value $Z_{CN} \epsilon [Z_{CL}, Z_{CU}]$ at $\alpha_N \epsilon [\alpha_L, \alpha_U]$ level of significance. The null hypothesis $H_0 : \beta_{iN} \epsilon [\beta_{iL}, \beta_{iU}]$ is not differ cross the groups vs. the alternative hypothesis $H_1 : \beta_{iN} \epsilon [\beta_{iL}, \beta_{iU}]$ is differ significantly across the groups.

Application of the proposed test

In this section, the application of the proposed Z-test for comparing sequential contingencies will be given using the data of spouse's behavior for couples in financial distress. According to [8] "A social researcher wishes to test a hypothesis concerning the behavior of adult couples. She compares a man's behavior with a consequent spouse's behavior for couples in financial distress and for those not in financial distresses". The data is selected from [8] and reported in Table 2. Kanji [8] presented the Z-test for comparing

Table 2 The spouse's behavior data [8]

		Distressed couples W_{tN+1}		Non-distressed couples W_{tN+1}	
	H_{tN}	**1**	**0**	**1**	**0**
Financial distress	1	76	100	80	63
Not in financial distress	0	79	200	43	39

sequential contingencies when certainty is presented during the implementation of the test. Suppose that decision-makers are uncertain about the level of significance with the measure of indeterminacy $I_{\alpha_N}\epsilon[0, 0.50]$. Let $\alpha_L = 0.05$ and $I_{\alpha_U} = 0.50$ which yield $\alpha_N = 0.05 + 0.10I_{\alpha_N}; I_{\alpha_N}\epsilon[0, 0.50]$. From [8], the calculated value of $Z_N\epsilon[1.493, 1.493]$. The tabulated values at $\alpha_N\epsilon[0.05, 0.10]$ are [1.96, 1.64]. By comparing $Z_N\epsilon[1.493, 1.493]$ with the tabulated values [1.96, 1.64], it can be seen that when the decision-makers are uncertain about the level of significance, according to [8] "She concludes that there is insufficient evidence to suggest financial distress affects couples' behavior in the way she hypothesizes". On the other hand, by comparing $Z_N\epsilon[1.493, 1.493]$ with the indeterminate level of significance, the same conclusion can be obtained. But the calculated values of $Z_N\epsilon[1.493, 1.493]$ close to the tabulated value 1.64. From the analysis it can be seen that although the null hypothesis is not rejected in uncertain environment, the values of $Z_N\epsilon[1.493, 1.493]$ close to the tabulated value so the decision-makers should be careful while making decisions about the null hypothesis.

Application for uncertain frequency

Now, the example for uncertain frequency will be given. Suppose that there is indeterminacy/uncertainty in frequencies. Table 3 shows some frequencies in intervals rather than the exact value. Therefore, the proposed test can be applied when the frequencies are presented in the interval.

The calculated values of $\beta_{i1L}, \beta_{i1U}, \beta_{i2L}$ and β_{i2U} for this data are given as

$$\beta_{i1L} = 0.6041, \beta_{i2U} = 0.654, \beta_{i2L} = 0.1027, \beta_{i2U} = 0.141$$

The statistic $Z_N\epsilon[Z_L, Z_U]$ for this data is given as

$$Z_N = 1.4504 + 1.493I_{\beta_N}; I_{\beta_N}\epsilon[0, 0.0285]$$

Let $\alpha = 0.05$ and tabulated value is 1.96. By comparing $Z_N\epsilon[1.4504, 1.493]$ with tabulated value 1.96, it can be concluded that there is an insufficient indication to advise financial distress affects couples' behavior with the degree of uncertainty that is 0.0285.

Simulation study

This section discusses the effect of an uncertain level of significance on the decision about the null hypothesis. To see the effect, various values of the measure of indeterminacy I_{α_U} are considered. We consider the determinate values of α_L are as: 0.001, 0.0026, 0.02, 0.05, 0.20 and the measure of indeterminacy I_{α_U}= 0.5, 0.74, 0.5614, 0.5 and 0.3711. The neutrosophic form of $\alpha_N\epsilon[\alpha_L, \alpha_U]$ for these measures of indeterminacy are shown in

Table 3 The spouse's behavior data with indeterminacy

		Distressed couples W_{tN+1}		Non- distressed couples W_{tN+1}	
	H_{tN}	1	0	1	0
Financial distress	1	[73,76]	100	[79,80]	63
Not in financial distress	0	79	[198,200]	43	[38,39]

Table 4. From Table 4, it can be seen that there is no effect on the decision about the null hypothesis when $I_{\alpha_U} > 0.50$. It is important to note that when $I_{\alpha_U}= 0.3711$, the decision about the null hypothesis has been changed from "do not reject H_0" to "reject H_0". From Table 4, it can be noted when the determinate value of $\alpha_L > 0.10$, the decision about H_0 is changed. Therefore, the increase in α_L may affect the decision about the null hypothesis.

Sensitivity analysis

To study the sensitivity when the level of significance is uncertain, the results were given in Table 4 will be utilized. From Table 4, it can be noted that when I_{α_U} is from 0.50 to 0.74, the decision about the acceptance of the null hypothesis should be remain the same. When $I_{\alpha_U} = 0.3711$, the decision about the null hypothesis is changed from acceptance to the rejection of the null hypothesis. From Table 4, it is quite clear that the decision about the null hypothesis does effected when there is a change in the level of significance. Therefore, the proposed method when the level of significance is uncertain is not much sensitive. The proposed method is sensitive for higher values of the level of significance.

Advantages

The proposed Z-test for comparing sequential contingencies using neutrosophic statistics is a generalization of the Z-test for comparing sequential contingencies using classical statistics. We will compare the efficiency of the proposed Z-test for comparing sequential contingencies using neutrosophic statistics with Z-test for comparing sequential contingencies using classical statistics in terms of level of significance. As mentioned earlier, α_L denotes a certain level of significance. The neutrosophic form of level of significance for the spouse's behavior data is $\alpha_N = 0.05 + 0.10I_{\alpha_N}$; $I_{\alpha_N} \epsilon [0, 0.50]$. The proposed neutrosophic form of the level of significance reduces to α_L when I_{α_L}=0. The second part $0.10I_{\alpha_N}$ denotes the indeterminate level of significance and $I_{\alpha_U}= 0.50$ is the specified measure of indeterminacy. From the study, it can be seen that the proposed test uses the level of significance in interval rather than the exact level of significance. For example, in implementing of the proposed test Z-test for comparing sequential contingencies, the level of significance can be from 0.05 to 0.10 with the measure of indeterminacy 0.50. On the other hand, the existing Z-test for comparing sequential contingencies using classical statistics gives only information about the determinate level of significance. Based on the study, it is concluded that the proposed test is quite flexible in using the level of significance as compared to the existing Z-test for comparing sequential contingencies using classical statistics.

Table 4 Effect of level of significance in decision-making

$I_{\alpha_N} \epsilon [I_{\alpha_L}, I_{\alpha_U}]$	$\alpha_N = \alpha_L + \alpha_U I_{\alpha_N}$	$Z_N \epsilon [Z_L, Z_U]$	$Z_{CN} \epsilon [Z_{CL}, Z_{CU}]$	Decision about H_0
[0, 0.5000]	$0.001 + 0.002I_{\alpha_N}$	$Z_N \epsilon [1.493, 1.493]$	[3.29,3.09]	Do not reject H_0
[0, 0.7400]	$0.0026 + 0.01I_{\alpha_N}$	$Z_N \epsilon [1.493, 1.493]$	[3.00,2.58]	Do not reject H_0
[0, 0.5614]	$0.02 + 0.0456I_{\alpha_N}$	$Z_N \epsilon [1.493, 1.493]$	[2.33,2.00]	Do not reject H_0
[0, 0.5000]	$0.05 + 0.10I_{\alpha_N}$	$Z_N \epsilon [1.493, 1.493]$	[1.96,1.64]	Do not reject H_0
[0, 0.3711]	$0.20 + 0.318I_{\alpha_N}$	$Z_N \epsilon [1.493, 1.493]$	[1.28,1.00]	Reject H_0

Table 5 Power of the test of two tests

α	Proposed test	Existing test
0.01	[0.9888,0.9912]	0.989
0.02	[0.9805,0.9854]	0.9825
0.025	[0.9696,0.9818]	0.9757
0.05	[0.944,0.9637]	0.9479
0.075	[0.9194,0.9353]	0.9234
0.1	[0.8896,0.8968]	0.8956

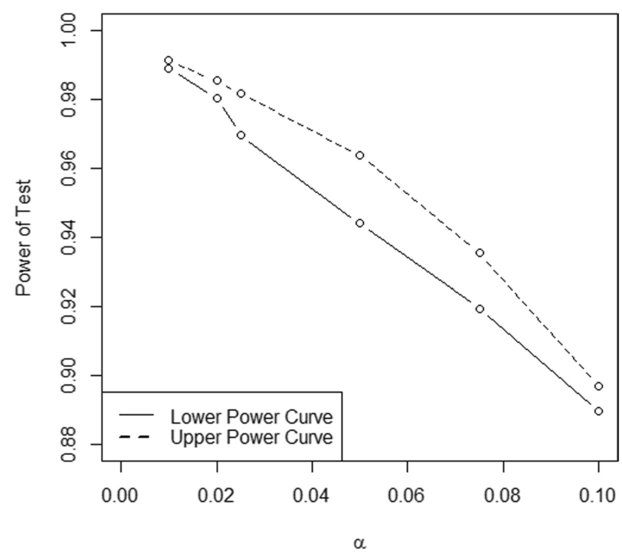

Fig. 1 The power curves of test

Power of the test

This section discusses the power of the proposed Z-test for comparing sequential contingencies. Suppose that α be the type-I error that means the probability of committing the error of accepting the null hypothesis when it is true and $(1-\beta)$ be the power of the test, where β is the probability of accepting when it is false. To study the power of the test, various values of α are considered. The power of the test for the proposed Z-test for comparing sequential contingencies under neutrosophic statistics and Z-test for comparing sequential contingencies using classical statistics are shown in Table 5. The power curve of the proposed test is also shown in Fig. 1. From Fig. 1, it can be seen that the power of the proposed sequential contingencies under neutrosophic statistics is in an indeterminate environment. The lower curve presents the power of the sequential contingencies under classical statistics. In Fig. 1 and Table 5, it is clear that the power of the test decreases as α values increase.

Discussion

The Z-test for comparing sequential contingencies under neutrosophic statistics reduces to Z-test for comparing sequential contingencies under classical statistics when no ambiguity is found during the implementation of the test. The proposed test performs better than the existing Z-test for comparing sequential contingencies. The proposed test has some limitations that it can be applied only when the decision-makers are uncertain about the level of significance or in frequency. The proposed Z-test for comparing sequential contingencies under neutrosophic statistics can be applied when the logit transformation can be done and 2×2 contingency table is available.

Concluding remarks

The Z-test for comparing sequential contingencies under neutrosophic statistics was introduced in the paper. The proposed Z-test for comparing sequential contingencies under neutrosophic statistics was a generalization of the existing Z-test for comparing sequential contingencies under classical statistics. The testing procedure of the proposed Z-test for comparing sequential contingencies was explained with the help of an example. The study showed that the proposed test can be applied in decision-making in an indeterminate environment. The proposed Z-test for comparing sequential contingencies under neutrosophic can be applied in phycology, medical science, political science, and industry when uncertainty is presented in frequency and level of significance. The proposed Z-test for comparing sequential contingencies using big data can be studied as future research.

Acknowledgements
The author is deeply thankful to the editor and reviewers for their valuable suggestions to improve the quality and presentation of the paper.

Author contributions
MA wrote the paper. The author read and approved the final manuscript.

Funding
None.

Availability of data and materials
The data is given in the paper.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
No conflict of interest regarding the paper.

Received: 24 August 2022 Accepted: 9 February 2023
Published online: 17 February 2023

References

1. Amiri S, Modarres R. Comparison of tests of contingency tables. J Biopharm Stat. 2017;27(5):784–96.
2. Aslam M, Arif OH, Sherwani RAK. New diagnosis test under the neutrosophic statistics: an application to diabetic patients. BioMed Res Int. 2020. https://doi.org/10.1155/2020/2086185.
3. Aslam M, Sherwani RAK, Saleem M. Vague data analysis using neutrosophic Jarque-Bera test. PLoS ONE. 2021;16(12):e0260689.
4. Chen J, Ye J, Du S. Scale effect and anisotropy analyzed for neutrosophic numbers of rock joint roughness coefficient based on neutrosophic statistics. Symmetry. 2017;9(10):208.
5. Chen J, Ye J, Du S, Yong R. Expressions of rock joint roughness coefficient using neutrosophic interval statistical numbers. Symmetry. 2017;9(7):123.
6. Duan W-Q, Khan Z, Gulistan M, Khurshid A. Neutrosophic exponential distribution: modeling and applications for complex data analysis. Complexity. 2021. https://doi.org/10.1155/2021/5970613.
7. Holland S. Data analysis in the geosciences. Ternary diagrams developed in the R. 2019.

8. Kanji GK. 100 statistical tests. London: Sage; 2006.
9. Khan Z, Al-Bossly A, Almazah M, Alduais FS. On statistical development of neutrosophic gamma distribution with applications to complex data analysis. Complexity. 2021. https://doi.org/10.1155/2021/3701236.
10. Khan Z, Gulistan M, Kausar N, Park C. Neutrosophic Rayleigh model with some basic characteristics and engineering applications. IEEE Access. 2021;9:71277–83.
11. Lee J, Lee S, Jang J-Y, Park T. Exact association test for small size sequencing data. BMC Med Genomics. 2018;11(2):21–31.
12. McComas JJ, Moore T, Dahl N, Hartman E, Hoch J, Symons F. Calculating contingencies in natural environments: issues in the application of sequential analysis. J Appl Behav Anal. 2009;42(2):413–23.
13. Meeker WQ. Sequential tests of independence for 2 × 2 contingency tables. Biometrika. 1978;65(1):85–90.
14. Rayalu GM, Sankar JR, Felix A. Testing sequential connections in contingency tables using coefficient of contingency in pharmaceutical statistics. Res J Pharm Technol. 2016;9(11):1902–4.
15. Sherwani RAK, Shakeel H, Saleem M, Awan WB, Aslam M, Farooq MJPo. A new neutrosophic sign test: an application to COVID-19 data. PLoS ONE. 2021;16(8):e0255671.
16. Smarandache F. Introduction to neutrosophic statistics, Sitech and Education Publisher, Craiova. Columbus: Romania-Educational Publisher; 2014. p. 123.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.