Research Article

Test of Association in the Presence of Complex Environment

Muhammad Aslam and Osama H. Arif

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

Correspondence should be addressed to Muhammad Aslam; aslam_ravian@hotmail.com

Received 7 April 2020; Revised 13 May 2020; Accepted 14 May 2020; Published 28 May 2020

Academic Editor: Quanmin Zhu

Copyright © 2020 Muhammad Aslam and Osama H. Arif. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new test of independence under neutrosophic statistics for testing the association between two criteria of classification is presented in this paper. The necessary contingency tables for the neutrosophic population and the neutrosophic sample are presented. The test statistic of the proposed test is introduced under neutrosophic statistics. A real example from education is selected to explain the proposed test. From the real example, it is concluded that the proposed test of independence is more informative, flexible, and suitable to be applied under uncertainty as compared to the existing test under classical statistics.

1. Introduction

One of the important uses of the chi-square distribution is in the test of independence. In this test, the association between two categorical variables is tested using the statistic with asymptotic chi-squared distribution. The test of independence is applied to test the null hypothesis that classification according to two criteria is independent versus the alternative hypothesis that classification according to two criteria is associated. The Pearson correlation is applied when the variables under study are quantities, while the test of independence is applied to see the association between qualitative variables. McHugh [1] discussed the application of the chi-square test for medical science. Singhal and Rana [2] mentioned some applications of the chi-square test. Benhamou and Melot [3] presented the graphical interpretation of the test. Lin et al. [4] introduced bootstrap for the test and applied in the biopharmaceutical industry. Kroonenberg and Verbeek [5] discussed some technical issues of the contingency table. Dutton et al. [6, 7] discussed the application of the test in education. More applications of the test can be seen in [8–17].

The chi-square test of independence is designed under the assumption that all observed frequencies in the contingency table are in determined form. For example, an education expert can make only the approximate estimation. In this situation, the test using the fuzzy approach is applied instead of the test under classical statistics. Runkler [18] presented the chi-square test using the fuzzy logic. Parthiban and Gajivaradhan [19] discussed the application of the fuzzy-based chi-square test in the environment. Lin et al. [20] proposed this test using membership functions. Taheri et al. [21] worked on fuzzy-based contingency tables. Alevizos et al. [22] discussed the application for education interval data. Anetzakis et al. [23] applied this fuzzy approach-based test for invasive species data.

A neutrosophic logic which is the extension of the fuzzy logic is introduced by Smarandache [24]. The neutrosophic logic provides additional information that is called the measure of indeterminacy. This logic is more efficient than the fuzzy logic and interval-based analysis, see [25]. Several authors worked on neutrosophic and provided the applications, see, for example, [26–36]. Smarandache [37] introduced the neutrosophic statistics that can be applied for data having indeterminate observations using the idea of the neutrosophic logic. The neutrosophic data are expressed in neutrosophic numbers \( X_N = X_L + X_U I_N; I_N \in [I_L, I_U] \), where \( X_L \) and \( X_U I_N \) are determinate and indeterminate parts, respectively. Note that \( I_N \in [I_L, I_U] \) represents the indeterminacy interval. Suppose that \( I_{NC} \in [0, 2] \) and
\( X_N = 2 + 3 I_N \); the neutrosophic data will be in the form of [2, 8], for more details, the reader may refer to [38–40] which mentioned the applications of neutrosophic numbers. Aslam et al. [41–44] introduced some statistical tests under neutrosophic statistics.

A lot of literature studies on the chi-square test of independence using the classical statistics and fuzzy-based approach are available. The existing tests are unable to provide information about the measure of indeterminacy when the data are obtained from the complex system. The existing test under classical statistics can be improved using the idea of neutrosophic statistics. By exploring the literature and to the best of our knowledge, there is no work on designing the chi-square test of independence under neutrosophic statistics. In this paper, we will introduce the chi-square test of independence under indeterminacy. We will introduce the contingency tables and chi-square statistic under indeterminacy. The application of the proposed test will be given using real data from the education. We expect that the proposed test will be the best alternative of the existing tests under an uncertainty environment. The rest of the paper is organized as follows: the proposed test of independence will be discussed in Section 2. The application and comparative studies will be discussed in Sections 3 and 4. A simulation study is given in Section 5, and some concluding remarks are given in the last section.

### 2. Proposed Test of Independence

One of the most important of the applications of the chi-square distribution is to test either the two criteria of classification are independent or not. As mentioned earlier, the existing test of independence can be applied only when the observations in the \( r \times c \) contingency table are determined, where \( r \) shows the rows and \( c \) shows the columns of the contingency table. In this section, we propose the test of independence under the neutrosophic statistics. The classification according to two criteria of the neutrosophic population is given in Table 1. The classification according to two criteria of the neutrosophic sample is given in Table 2. The main aim of introducing the new test of independence under the neutrosophic statistic is to test the neutrosophic null hypothesis \( H_{ON} \), that is, in the neutrosophic population, two criteria are independent versus the alternative hypothesis \( H_{IN} \) that, in the neutrosophic population, two criteria are associated.

The necessary steps to evaluate the proposed test of independence under the neutrosophic statistic are explained as follows:

- **Step 1:** state the null hypothesis \( H_{ON} \) that two criteria are independent versus the alternative hypothesis that two criteria are associated.
- **Step 2:** specify the level of significance \( \alpha \).
- **Step 3:** compute the neutrosophic values of the expected frequency \( E_N = \left( \frac{(O_N - E_N)^2}{E_N} \right) \).
- **Step 4:** the proposed statistic \( \chi^2_N \in [\chi^2_L, \chi^2_U] \) is given by \( \chi^2_N = \sum \left( \frac{(O_N - E_N)^2}{E_N} \right) \); \( \chi^2_N \in [\chi^2_L, \chi^2_U] \), \( E_N \in [E_L, E_U] \).

The neutrosophic form of \( \chi^2_N \in [\chi^2_L, \chi^2_U] \) is given by

\[
\chi^2_N = \chi^2_L + \chi^2_U I_N; \quad I_N \in [I_L, I_U].
\]

Note here that the proposed statistic \( \chi^2_N \in [\chi^2_L, \chi^2_U] \) is the extension of the existing statistic \( \chi^2 \) used for the test of independence. The proposed test reduces to test under classical statistics when \( I_N = 0 \).

**Step 5:** select the critical value, say \( \chi^2_c \), from the chi-square table at the degree of freedom \((r - 1) \times (c - 1)\) at the level of significance \( \alpha \).

**Step 6:** reject \( H_{ON} \) if \( \chi^2_N \in [\chi^2_L, \chi^2_U] \) is larger than the critical value. It means \( H_{ON} \) will be rejected if the values of both \( \chi^2_L \) and \( \chi^2_U \) are larger than \( \chi^2_c \). In some cases, it may happen that the critical value \( \chi^2_c \) is between the value of \( \chi^2_L \) and the value of \( \chi^2_U \). In this situation, according to Smarandache [37], \( H_{ON} \) will be rejected if \( \chi^2_U > \chi^2_c \); otherwise, do not reject it.

### 3. Application of the Proposed Test

In this section, we will discuss the application of the proposed test using real data from education. The education expert is interested to investigate the association between the “student’s group-community” and the “department.” The details about the real data can be seen in [45]. According to Alevizos et al. [22], “consider three groups of candidate-students who are preparing to enter in a university, where each group represents a different geographical community, indicated as Student-1, Student-2, and Student-3. In order to succeed in a university, the students should choose to participate at exactly one from a choice of three directions-departments of the university indicated as Mathematics, Physics, and Literature.” As mentioned earlier, an education expert is interested to test the null hypothesis that “student’s group-community” and “department” are independent versus the alternative hypothesis that “student’s group-community” and “department” are associated. The data of the two criteria are taken from [22] and reported in Table 3. From Table 3, it is quite clear that the observations of “student’s group-community” and “department” are presented in the neutrosophic interval. Therefore, the use of the test of independence under classical statistics may mislead the education expert. In addition, the existing test is unable to provide the probability of indeterminacy associated with the test of independence. Therefore, the education expert is interested to use the proposed test for testing the association between “student’s group-community” and “department.” The proposed test for the data is stated as follows:

- **Step 1:** state the null hypothesis \( H_{ON} \) that “student’s group-community” and “department” are independent versus \( H_{IN} \) that “student’s group-community” and “department” are associated.
Step 2: let $a = 0.05$ for this test.

Step 3: the neutrosophic values of the expected frequency $\bar{E}_{N} \in [I_{L}, I_{U}]$ are shown in Table 4.

Step 4: the calculation of the statistic $\chi_{N}^{2} \in [x_{L}^{2}, x_{U}^{2}]$ is shown as follows:

$$\chi_{N}^{2} = \sum \frac{(O_{N} - E_{N})^{2}}{E_{N}} \in [4.66, 13.43]; \quad \chi_{N}^{2} \in [4.66, 13.43].$$

The neutrosophic form of $\chi_{N}^{2} \in [x_{L}^{2}, x_{U}^{2}]$ is given by

$$x_{N}^{2} = 4.66 + 13.43; \quad I_{N} \in [0, 0.65].$$

Step 5: the critical value from the chi-square table at the degree of freedom $2 \times 2$ at $a = 0.05$ is 9.49.

Step 6: according to Smarandache [37], the null hypothesis is rejected if $\chi_{U}^{2} > 9.49$. Therefore, the null hypothesis that "student’s group-community” and “department” are independent is rejected.

### 4. Comparative Study Based on Real Data

As mentioned earlier, the proposed test of independence is the generalization of the existing test under classical statistics. For the comparison, the same level of all parameters is used for both tests. First, we compare both tests in terms of values of test statistics. Then, we will compare both tests in terms of probabilities.

The neutrosophic form of the statistic $\chi_{N}^{2} \in [x_{L}^{2}, x_{U}^{2}]$ is $\chi_{N}^{2} = 4.66 + 13.43; I_{N} \in [0, 0.65]$. Note here that the proposed neutrosophic form reduces to the statistic under
classical statistics when $I_L = 0$. Therefore, the value of $\chi^2_L = 4.66$ is corresponding to the statistic under classical statistics. From this result, it can be seen that the proposed test statistic values lie in the indeterminacy interval $\chi^2_N \in [4.66, 13.43]$. On the contrary, the existing test under classical statistics provides only the determined value of the test statistic which is not suitable in the indeterminacy environment. From this comparison, it is concluded that the proposed test is quite affective, adequate, and effective to be applied in uncertainty as compared to the test of independence under classical statistics.

Now, we will compare both tests in terms of probabilities related to the null hypothesis. For the real data, the level of significance is 5%. The existing test tells us that the chance of accepting the null hypothesis $H_{0N}$ is 0.95, and the probability of rejecting the null hypothesis is 0.05. On the contrary, the proposed test tells that the chance of accepting the null hypothesis $H_{0N}$ is 0.95, the probability of rejecting the null hypothesis is 0.05, and the probability of indeterminacy is 0.65. From this comparison, it can be concluded that the proposed test provides the additional probability of indeterminacy that cannot be obtained from the existing test.
under classical statistics. Therefore, the proposed test is efficient than the existing test in providing information about the probabilities associated with the test of independence.

5. Simulation Study

Now, we discuss the importance of the proposed test in terms of the power of the test using simulation data. We will compare the proposed test and the existing test in terms of the power of the test. Let $\beta$ denote the type-II error which is defined as the probability of rejecting $H_{0N}$ when $H_{0N}$ is true. On the contrary, the power of the test $(1-\beta)$ is defined as the probability of rejecting $H_{0N}$ when $H_{0N}$ is false. The values of $\beta$ and $(1-\beta)$ for both tests are shown in Table 5. The following algorithm is used to construct Table 1:

Step 1: for the $r-1 \times c-1$ contingency table, specify $df$.

Step 2: compute the values of the statistic $\chi^2 \in [\chi^2_L, \chi^2_U]$ and compare it with $\chi^2$. 

Step 3: compute $\beta$ by dividing the total number of accepting $H_{0N}$ to the total number of replications, say 100.

From Table 1, it is clear that, for $df = 1$ and $df = 2$, the power of the existing test is between the indeterminacy of the proposed test. For $df > 2$, the power of the existing test is decreasing as $df$ is increased. The power curves of both tests are shown in Figure 1. Figure 1 indicates that the proposed test has higher values of the power of the test as compared to the existing test. From this study, it can be concluded that the proposed test is better than the existing test in terms of the power of the test.

6. Concluding Remarks

A new test of independence under neutrosophic statistics for testing the association between two criteria of classification is introduced in this paper. The necessary contingency tables for the neutrosophic population and the neutrosophic sample were presented. The test statistic of the proposed test is introduced under neutrosophic statistics. The application of the proposed test was given using the data from the education department. The comparative study shows the efficiency of the proposed test over the existing test. We recommend applying the proposed test for association testing in the presence of indeterminacy. The proposed test can be applied for big data as future research. The development of software for the proposed test is also a fruitful area of research.

Data Availability

The data are given in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are thankful to Dr. Nasrullah Khan (UVAS, Pakistan) for his valuable suggestions. This work was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

References


