

RESEARCH

Open Access



Data analysis for vague contingency data

Muhammad Aslam¹ and Faten S. Alamri^{2*}

*Correspondence:
Fsalamri@pnu.edu.sa

¹ Department of Statistics,
Faculty of Science, King
Abdulaziz University,
Jeddah 21551, Saudi Arabia
² Department of Mathematical
Sciences, College of Science,
Princess Nourah bint
Abdulrahman University, P.O.
Box 84428, Riyadh 11671, Saudi
Arabia

Abstract

The existing Fisher's exact test has been widely applied for investigating whether the difference between the observed frequencies is significant or not. The existing Fisher's exact test can be applied only when the observed frequencies are in determinate form and has no vagues information. In practice, due to the complicity in the production process, it is not always possible to have observed frequencies in determinate form. Therefore, the use of the existing Fisher's exact test may mislead the industrial engineers. The paper presents the modification of Fisher's exact test using neutrosophic statistics. The operational process, simulation study, and application using the production data will be given in the paper. From the analysis of industrial data, it can be concluded that the proposed Fisher's exact test performs well than the existing Fisher's exact test.

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

Introduction

Fisher's exact test using classical statistics has been applied for investigating whether the observed frequencies from dichotomous distributions are associated with each other or independent from each other. Fisher's exact test using classical statistics is usually applied for 2×2 contingency table. The main aim of Fisher's exact test is to test the null hypothesis that observed frequencies dichotomous distributions are associated vs. the alternative hypothesis that observed frequencies dichotomous distributions are independent. According to Kanji [1], the test statistic ($\sum p$) of Fisher's exact test is calculated and compared with the specified level of significance (which is the probability of rejecting the null hypothesis when it is true) and the null hypothesis is rejected when the calculated value of the test statistic is less than the level of significance, otherwise, the null hypothesis is not rejected. Chen [2] differentiate between the chi-square test and Fisher's exact test for 2×2 contingency table. Choi et al. [3] discussed the foundations and inference of 2×2 contingency table. Zhong et al. [4] discussed the application of the test for biological data. Ma and Mao [5] discussed the application of this test for scanning dependency. More information on Fisher's exact test can be seen in [6–8].

Fuzzy-logic has the application where uncertainty is found in the data. To analyze the uncertain data, the statistical tests using classical statistics cannot be applied. The information about two measures (true and false) can be obtained from the fuzzy-based



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

analysis. The logic having more information in uncertainty is known as “neutrosophic logic” is introduced by [9]. Smarandache [10] discussed that the neutrosophic logic has an edge over the interval data analysis and fuzzy logic. Basha et al. [11] and Das et al. [12] discussed the applications of neutrosophic logic. Based on the idea of neutrosophic numbers, the idea of neutrosophic statistics was given by [13] and further investigated by [14, 15]. Neutrosophic statistics was found to be more informative and more efficient than classical statistics by [12, 16, 17].

The operational process of Fisher’s exact test using classical statistics is designed to analyze only the determinate or exact observed frequencies. The existing Fisher’s exact test cannot be applied when the observed frequencies are in intervals. By exploring the literature and best of the author’s knowledge, no efforts have been made to design Fisher’s exact test using neutrosophic statistics. In this paper, we will extend Fisher’s exact test using neutrosophic statistics. The test statistic of Fisher’s exact test will be modified to analyze the neutrosophic numbers. The power of Fisher’s exact test will be discussed and application will be given using the industrial data. It is expected that Fisher’s exact test under neutrosophic statistics will be more efficient than the existing Fisher’s exact tests in terms of the power of the test, information and flexibility.

The proposed fisher’s exact test

The exiting Fisher’s exact test under classical statistics is applied to investigate whether the difference between observed frequencies is significant or not. The existing Fisher’s exact test cannot be applied if the observed frequency is interval rather than the exact number. To overcome this issue, it is necessary to modify Fisher’s exact test using neutrosophic statistics so that an investigation about the difference in frequency can be done in the presence of interval, fuzzy, imprecise and indeterminate data. Similar to Fisher’s exact test under classical statistics, the proposed Fisher’s exact test under neutrosophic statistics will be applied using a 2×2 contingency table. Let $a_N = a_L + a_U I_N$; $I_N \in [I_L, I_U]$, $b_N = b_L + b_U I_N$; $I_N \in [I_L, I_U]$, $c_N = c_L + c_U I_N$; $I_N \in [I_L, I_U]$, and $d_N = d_L + d_U I_N$; $I_N \in [I_L, I_U]$ be neutrosophic observed frequencies. Note here that the first values of observed frequency denote the determinate values, $a_U I_N, b_U I_N, c_U I_N, d_U I_N$ are indeterminate observed frequencies and $I_N \in [I_L, I_U]$ is a measure of indeterminacy associated with observed frequencies. These measures can be calculated from the imprecise data as (upper value-lower value)/upper value. Suppose that $N_N = N_L + N_U I_N$; $I_N \in [I_L, I_U]$ be the total observed frequency. A 2×2 contingency table to carry out Fisher’s exact test under the idea of neutrosophy is presented in Table 1 as follows, see [18, 19] for more details. The neutrosophic test statistic $\sum p_N \in [\sum p_L, \sum p_U]$ for Fisher’s exact test is defined as

Table 1 A 2×2 contingency table

	Class 1	Class 2	Total
Sample 1	$a_L + a_U I_N$	$b_L + b_U I_N$	$(a_L + b_L) + (a_U + b_U) I_N$
Sample 2	$c_L + c_U I_N$	$d_L + d_U I_N$	$(c_L + d_L) + (c_U + d_U) I_N$
Total	$(a_L + b_L) + (a_U + c_U) I_N$	$(b_L + d_L) + (b_U + d_U) I_N$	$N_L + N_U I_N$

$$\sum p_N = \sum p_L + \sum p_U I_{p_N}; I_{p_N} \in [I_{p_L}, I_{p_U}] \quad (1)$$

where the first part $\sum p_L$ denotes the statistic of Fisher's exact test under classical statistics, the second part $\sum p_U I_{p_N}$ denote the indeterminate part and $I_{p_N} \in [I_{p_L}, I_{p_U}]$ is the uncertainty measure associated with the proposed test statistic. The proposed test statistic $\sum p_N \in [\sum p_L, \sum p_U]$ reduces to the existing test statistic $\sum p_L$ when $I_{p_L}=0$. By following [1], the test statistic of the proposed test can be written as

$$\begin{aligned} \sum p_N = & \frac{(a_L + b_L)!(c_L + d_L)!(a_L + c_L)!(b_L + d_L)!}{N_L!} \sum_i \frac{1}{a_{iL}!b_{iL}!c_{iL}!d_{iL}!} \\ & + \frac{(a_U + b_U)!(c_U + d_U)!(a_U + c_U)!(b_U + d_U)!}{N_U!} \\ & \sum_i \frac{1}{a_{iU}!b_{iU}!c_{iU}!d_{iU}!} I_{p_N}; I_{p_N} \in [I_{p_L}, I_{p_U}] \end{aligned} \quad (2)$$

The proposed test statistic $\sum p_N \in [\sum p_L, \sum p_U]$ can be expressed as

$$\sum p_N \in \left\{ \frac{(a_L + b_L)!(c_L + d_L)!(a_L + c_L)!(b_L + d_L)!}{N_L!} \sum_i \frac{1}{a_{iL}!b_{iL}!c_{iL}!d_{iL}!}, \frac{(a_U + b_U)!(c_U + d_U)!(a_U + c_U)!(b_U + d_U)!}{N_U!} \sum_i \frac{1}{a_{iU}!b_{iU}!c_{iU}!d_{iU}!} \right\} \quad (3)$$

As mentioned in [1] “the summation is over all possible 2×2 schemes with a cell frequency equal to or smaller than the smallest experimental frequency (keeping the row and column totals fixed as above)”.

The computed value of $\sum p_N \in [\sum p_L, \sum p_U]$ is compared with the pre-specified level of significance α . The null hypothesis of independence between sample and class is rejected if $\sum p_N \in [\sum p_L, \sum p_U] < \alpha$, otherwise, the alternative hypothesis that sample

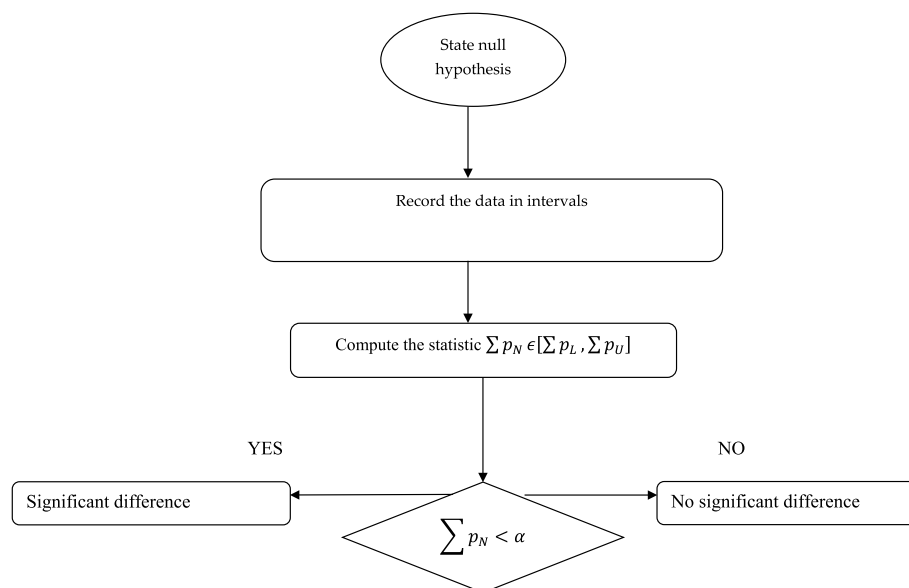


Fig. 1 The procedure of Fisher's exact test under classical statistics

Table 2 A 2 × 2 contingency table of machines and production

Machines	A ₁	A ₂	Total
Production time	[1, 1]	[1, 1]	[2, 2]
Number of defective	[10, 15]	[26, 32]	[36, 47]
Total	[11, 16]	[27, 33]	[38, 49]

Table 3 1st combination of the original data

[0,0]	[2,2]	[2,2]
[11,16]	[25,31]	[36,47]
[11,16]	[27,33]	[38,49]

Note that the bold numbers are showing the totals

and class are not independent is not rejected. The operational procedure of the proposed Fisher's exact test under classical statistics is discussed in Fig. 1.

Application using industrial data

In this section, the application of the proposed test is given using the information obtained from the manufacturing industry. Two machines A_1 and A_2 work for an hour and produced defective items in intervals. To explain the process of the proposed test, a 2×2 contingency table is extracted from [20] and the data is shown in Table 2. Industrial engineers are interested to investigate there is a significant difference between the performance of machines A_1 and A_2 . As mentioned before, the neutrosophic-based tests have the ability to analyze the interval-based data more effectively than the tests using classical statistics.

The neutrosophic test statistic is derived by computing all conceivable combinations utilizing the hypergeometric distribution, as outlined in Table 3. The minimum value among these combinations is identified and compared against all other combinations to ascertain those below this minimum. It's important to emphasize that these combinations are carefully selected to ensure that both the row and column totals remain consistent with those presented in Table 2.

Based on the possible combinations in Tables 2, 3, $\sum p_N$ is calculated as

$$\sum p_N = \frac{2!36!11!27!}{38!} \left\{ \frac{1}{10!26!} + \frac{1}{2!11!25!} \right\} + \frac{2!47!16!33!}{49!} \left\{ \frac{1}{15!32!} + \frac{1}{2!16!31!} \right\} I_{p_N}; I_{p_N} \in [I_{p_L}, I_{p_U}]$$

The simplified neutrosophic form of $\sum p_N \in [0.9218, 0.8980]$ is given as $\sum p_N = 0.9218 - 0.8980 I_{p_N}; I_{p_N} \in [0, 0.0267]$. Suppose that $\alpha=0.05$. The calculated values of $\sum p_N \in [0.9218, 0.8980]$ will be compared with 0.05. By comparing the values of statistic $\sum p_N \in [0.9218, 0.8980]$ with 0.05, the values of statistic $\sum p_N \in [0.9218, 0.8980]$ is greater than 0.05, therefore, the industrial engineers do not reject the null hypothesis H_0 of no difference between the performance of machines A_1 and A_2 . Figure 2 depicts the operational procedure of the proposed Fisher's exact test for the production data.

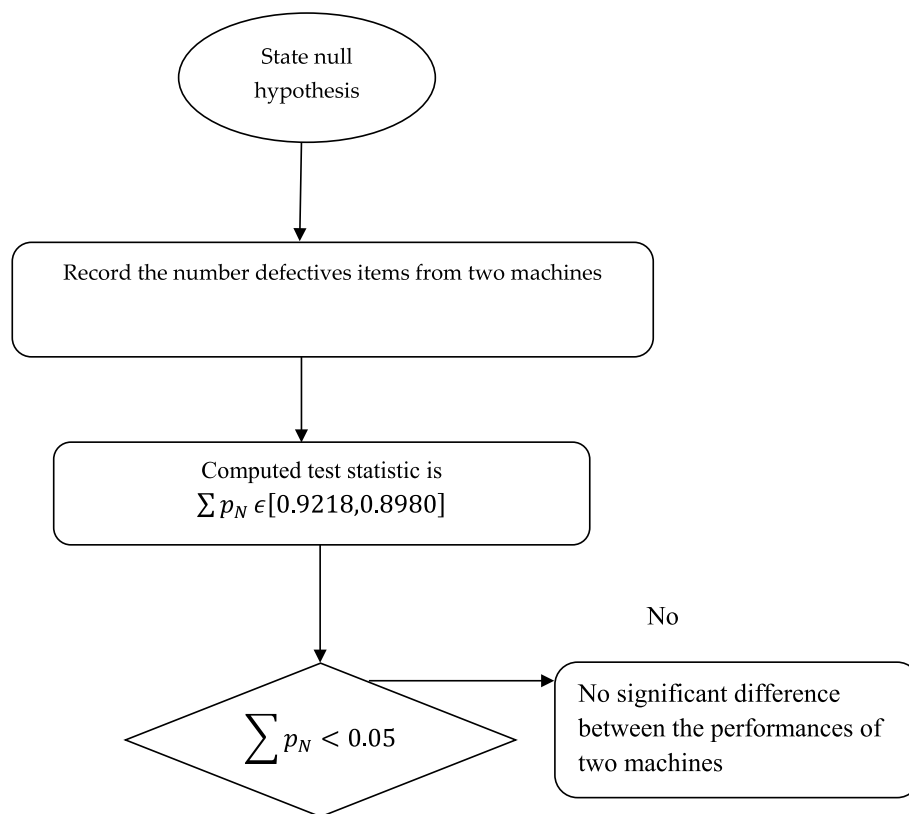


Fig. 2 The procedure of Fisher's exact test for production data

Advantages based on industrial data

The proposed Fisher's exact test using neutrosophic statistics is a generalization of several tests. Now, the efficiency of the proposed Fisher's exact test under neutrosophic statistics will be compared with Fisher's exact test using Fisher's exact test under classical statistics, interval-statistics, and Fisher's exact test using fuzzy logic in terms of information and adequacy. To compare the efficiency of various tests, the neutrosophic statistic $\sum p_N \in [\sum p_L, \sum p_U]$ obtained for the production data will be considered. The neutrosophic form of the statistic from the data is given as: $\sum p_N = 0.9218 - 0.8980I_{p_N}; I_{p_N} \in [0, 0.0267]$. Note that the first value $\sum p_L = 0.9218$ presents Fisher's exact test under classical statistics and $0.8980I_{p_N}$ is an indeterminate part, and $I_{p_N} \in [0, 0.0267]$ is a measure of indeterminacy associated with $\sum p_N \in [\sum p_L, \sum p_U]$. The proposed statistic $\sum p_N \in [\sum p_L, \sum p_U]$ reduces to Fisher's exact test under classical statistics when $I_{p_N} = 0$. By comparing the proposed Fisher's exact test under neutrosophic statistics with Fisher's exact test under classical statistics, it can be seen that the proposed Fisher's exact test under neutrosophic statistics provide the values of statistic $\sum p_N \in [\sum p_L, \sum p_U]$ in indeterminate interval with the measure of indeterminacy. For example, for testing the null hypothesis at a level of significance $\alpha = 0.05$, the proposed Fisher's exact test under neutrosophic is explained as: the probability of accepting the null hypothesis is 0.95, the probability of committing an error is 0.05 and the measure of indeterminacy is 0.0267. From

the comparison, it is clear that the proposed Fisher's exact test under neutrosophic is more efficient and more informative than Fisher's exact test using classical statistics. Now, the efficiency of the proposed Fisher's exact test under neutrosophic statistics will be compared with Fisher's exact test using interval-statistics. The statistic $\sum p_N \in [\sum p_L, \sum p_U]$ using interval-statistics only capture the data inside the interval. The statistic $\sum p_N \in [\sum p_L, \sum p_U]$ using interval statistic tells that the values of the test statistic may vary from 0.9218 to 0.8980. Similarly, Fisher's exact test using fuzzy-logic gives information about the measure of truth that is 0.95, and the measure of falseness that is 0.05. Like the interval-statistics, it tells that the statistic $\sum p_N \in [\sum p_L, \sum p_U]$ may change from 0.9218 to 0.8980 under uncertain environment. From the analysis, it is concluded that the proposed Fisher's exact test under neutrosophic statistics has an edge over the three Fisher's exact tests. Therefore, the use of the proposed Fisher's exact test under neutrosophic statistics in the production industry will give more information and facilitate the decision-makers in the presence of an indeterminate environment.

Simulation study

To see whether the measure of indeterminacy $I_{p_N} \in [I_{p_L}, I_{p_U}]$ affects the decision about the null hypothesis or not. To study this affect, various intervals values of $\sum p_N \in [\sum p_L, \sum p_U]$ are considered in Table 4. The neutrosophic forms of $\sum p_N \in [\sum p_L, \sum p_U]$ for the selected values of $\sum p_N \in [\sum p_L, \sum p_U]$, the measure of indeterminacy $I_{p_N} \in [I_{p_L}, I_{p_U}]$, and the decision about the null hypothesis are also reported in Table 4. From Table 4, it can be seen that $\sum p_N \in [\sum p_L, \sum p_U]$ increases,

Table 4 Effect of indeterminacy

$\sum p_N \in [\sum p_L, \sum p_U]$	$\sum p_L + \sum p_U I_{p_N}; I_{p_N} \in [I_{p_L}, I_{p_U}]$	Decision
$\sum p_N \in [0.04, 0.01]$	$0.04 - 0.01 I_{p_N}; I_{p_N} \in [0, 3]$	Reject H_0
$\sum p_N \in [0.09, 0.05]$	$0.09 - 0.05 I_{p_N}; I_{p_N} \in [0, 0.80]$	Do not reject H_0
$\sum p_N \in [0.14, 0.10]$	$0.14 - 0.10 I_{p_N}; I_{p_N} \in [0, 0.40]$	Do not reject H_0
$\sum p_N \in [0.19, 0.15]$	$0.19 - 0.15 I_{p_N}; I_{p_N} \in [0, 0.27]$	Do not reject H_0
$\sum p_N \in [0.24, 0.20]$	$0.24 - 0.20 I_{p_N}; I_{p_N} \in [0, 0.20]$	Do not reject H_0
$\sum p_N \in [0.29, 0.25]$	$0.29 - 0.25 I_{p_N}; I_{p_N} \in [0, 0.16]$	Do not reject H_0
$\sum p_N \in [0.34, 0.30]$	$0.34 - 0.30 I_{p_N}; I_{p_N} \in [0, 0.13]$	Do not reject H_0
$\sum p_N \in [0.39, 0.35]$	$0.39 - 0.35 I_{p_N}; I_{p_N} \in [0, 0.11]$	Do not reject H_0
$\sum p_N \in [0.44, 0.40]$	$0.44 - 0.40 I_{p_N}; I_{p_N} \in [0, 0.10]$	Do not reject H_0
$\sum p_N \in [0.49, 0.45]$	$0.49 - 0.45 I_{p_N}; I_{p_N} \in [0, 0.09]$	Do not reject H_0
$\sum p_N \in [0.54, 0.50]$	$0.54 - 0.50 I_{p_N}; I_{p_N} \in [0, 0.08]$	Do not reject H_0
$\sum p_N \in [0.59, 0.55]$	$0.59 - 0.55 I_{p_N}; I_{p_N} \in [0, 0.07]$	Do not reject H_0
$\sum p_N \in [0.64, 0.60]$	$0.64 - 0.60 I_{p_N}; I_{p_N} \in [0, 0.07]$	Do not reject H_0
$\sum p_N \in [0.69, 0.65]$	$0.69 - 0.65 I_{p_N}; I_{p_N} \in [0, 0.06]$	Do not reject H_0
$\sum p_N \in [0.74, 0.70]$	$0.74 - 0.70 I_{p_N}; I_{p_N} \in [0, 0.06]$	Do not reject H_0
$\sum p_N \in [0.79, 0.75]$	$0.79 - 0.75 I_{p_N}; I_{p_N} \in [0, 0.05]$	Do not reject H_0
$\sum p_N \in [0.84, 0.80]$	$0.84 - 0.80 I_{p_N}; I_{p_N} \in [0, 0.05]$	Do not reject H_0
$\sum p_N \in [0.89, 0.85]$	$0.89 - 0.85 I_{p_N}; I_{p_N} \in [0, 0.05]$	Do not reject H_0
$\sum p_N \in [0.94, 0.90]$	$0.94 - 0.90 I_{p_N}; I_{p_N} \in [0, 0.04]$	Do not reject H_0
$\sum p_N \in [0.99, 0.95]$	$0.99 - 0.95 I_{p_N}; I_{p_N} \in [0, 0.04]$	Do not reject H_0

the values of $I_{p_N} \in [I_{p_L}, I_{p_U}]$ decreases. For example, when $\sum p_N \in [0.01, 0.04]$, the values of I_{p_N} is $I_{p_N} \in [0, 3]$. When $\sum p_N \in [0.95, 0.99]$, the values of I_{p_N} is $I_{p_N} \in [0, 0.04]$. In addition, it can be noted when $I_{p_N} \in [0, 0.80]$ are fewer, although the values of $\sum p_N \in [\sum p_L, \sum p_U]$ changes but no effect on the null hypothesis when comparing with $\alpha=0.05$. We note that the larger values of I_{p_N} affect the decision about the null hypothesis. For example, when $I_{p_N} \in [0, 3]$, the decision about the null hypothesis is changed from “Do not reject H_0 ” to “reject H_0 ”. From the study, it is clear that the larger values of the measure of uncertainty/indeterminacy affect the decision about the null hypothesis. Therefore, industrial engineers should be very careful in making decisions in the presence of uncertainty.

Sensitivity analysis

The sensitivity of the proposed Fisher’s exact test under neutrosophic statistics will be discussed now. The values of $I_{p_N} \in [I_{p_L}, I_{p_U}]$ are shown in Table 4. From Table 4, it can be seen that when $\sum p_N \in [\sum p_L, \sum p_U]$ changes from $[0.79, 0.75]$ to $[0.89, 0.85]$, the values of the measure of indeterminacy remain the same that is 0.05. When $\sum p_N \in [\sum p_L, \sum p_U]$ changes from $[0.94, 0.90]$ to $[0.99, 0.95]$, the values of the measure of indeterminacy remain the same that is 0.04. Similarly, there is not much change in $I_{p_N} \in [I_{p_L}, I_{p_U}]$ when $\sum p_N \in [\sum p_L, \sum p_U]$ changes from $[0.34, 0.30]$ to $[0.44, 0.40]$. This analysis shows that the change in the statistic $\sum p_N \in [\sum p_L, \sum p_U]$ change the values of $I_{p_N} \in [I_{p_L}, I_{p_U}]$ but it does not affect the decision about the null hypothesis. From the analysis, it is concluded that the proposed test is sensitive for the higher values of $I_{p_N} \in [I_{p_L}, I_{p_U}]$.

Power of the test

This section presents the discussion on the power of Fisher’s exact test under neutrosophic statistics. Suppose that α and β be the probability of rejecting $H_0|$ true and the probability of accepting $H_0|$ false. The power of the test is denoted by $(1 - \beta)$.

By following Nosakhare and Bright [21], the steps used to calculate β are given as.

Step-1: Generate a set of 10,000 random samples of the test statistic $\sum p_N$

Step-2: Compare the values of $\sum p_N$ with the level of significance and record whether the null hypothesis H_0 is rejected or accepted.

Step-3: Determine the values of β (Type II error rate) by the ratio of the number of erroneous conclusion to the total number of replications.

Table 5 The values of power of the tests

α	$(1 - \beta)$
0.01	[0.9895, 0.9418]
0.02	[0.9789, 0.8994]
0.04	[0.9589, 0.8284]
0.05	[0.951, 0.801]
0.08	[0.9183, 0.7184]
0.10	[0.8978, 0.6705]
0.15	[0.8527, 0.569]
0.20	[0.8036, 0.4878]

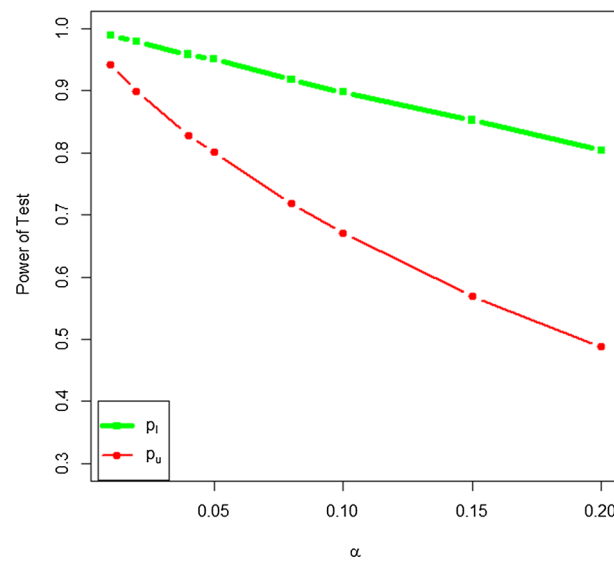


Fig. 3 The power curve of Fisher's exact test under neutrosophic statistics

The values of $(1 - \beta)$ for various values of α are shown in Table 5. The power curve for the proposed Fisher's exact test under neutrosophic statistics is shown in Fig. 3. As mentioned earlier, the proposed Fisher's exact test under neutrosophic statistics reduces to Fisher's exact test under classical statistics when no uncertainty is found. The lower line in Fig. 3 shows the power of the indeterminate part and the upper curve shows the power of the test for the determinate part. Overall, Fig. 3 shows the power of the test for Fisher's exact test under neutrosophic statistics. From Fig. 3, it is clear that as the values of α increases, the power of Fisher's exact test under neutrosophic statistics decreases. For example, when $\alpha = 0.01$, the power of the test ranges from 0.9857 to 1. When $\alpha = 0.10$, the power of the test ranges from 0.8978 to 0.6705. The first value 0.9878 presents the power of the test under classical statistics. It is clear that the power of the test reduces from 0.8978 to 0.6705; therefore, the use of the existing test under classical statistics may mislead decision-makers. In a nutshell, it is concluded that in the case of neutrosophy, the power of the test is in indeterminate intervals rather than the exact values. This study shows that the proposed test is more flexible than the existing Fisher's exact test.

Effect of indeterminacy on level of significance

Now, we will discuss the effect of indeterminacy on the level of significance. To study this effect, various values of specified values of levels of significance are considered to assess the effect of uncertainty. Suppose that α_0 denote the pre-defined/specified level of significance and let $\hat{\alpha}_N \in [\hat{\alpha}_L, \hat{\alpha}_U]$ denote the computed level of significance from the following simulation process.

Step-1: Generate a set of 10,000 random samples of the test statistic $\sum p_N$

Step-2: Compare the values of $\sum p_N$ with the level of significance and record whether the null hypothesis H_0 is rejected or accepted.

Step-3: Determine the values of $\hat{\alpha}_N \in [\hat{\alpha}_L, \hat{\alpha}_U]$ (Type I error rate) by the ratio of the number of rejection conclusions to the total number of replications.

Table 6 The computed values of α_N

α_0	$\hat{\alpha}_N \in [\hat{\alpha}_L, \hat{\alpha}_U]$
0.01	[0.01, 0.05]
0.02	[0.02, 0.10]
0.04	[0.04, 0.17]
0.05	[0.05, 0.20]
0.08	[0.08, 0.28]
0.10	[0.09, 0.33]
0.15	[0.15, 0.43]
0.20	[0.20, 0.52]

By implementing the above simulation process, the values of $\hat{\alpha}_N \in [\hat{\alpha}_L, \hat{\alpha}_U]$ are placed in Table 6. From Table 6, it can be noted that lower values of $\hat{\alpha}_N$ are the same as the α_0 . But it is worth noting that the upper values of $\hat{\alpha}_N$ are larger than α_0 . In addition, it can be noted that as α_0 , there is an increase in $\hat{\alpha}_N$. From the study, it is clear while implementing the test under uncertainty, the level of significance may change from α_0 . For example, when $\alpha_0=0.05$, the computed $\hat{\alpha}_N$ is $\hat{\alpha}_N \in [0.05, 0.20]$. We can see that level of significance changes from 0.05 to 0.20 which can affect the decision related to the null hypothesis.

Concluding remarks

In this paper, Fisher's exact test under neutrosophic statistics was presented. The design of the proposed Fisher's exact test under an indeterminate environment was given. The operational procedure was explained with the help of industrial data. The proposed Fisher's exact test was a generalization of the existing Fisher's exact test under classical statistics. Based on the analysis and the simulation studies, it is concluded that the proposed test efficiently indicates a change in the power of the test and the level of significance when the test is implanted in the presence of imprecise data. The use of the proposed test is quite adequate to be applied in the uncertain environment as compared to the existing test. Based on the analysis and simulation studies, the application of the proposed Fisher's exact test is recommended in the industry where the production data is ambiguous, imprecise, and or in-intervals. For future research, other statistical properties of the proposed Fisher's exact test under neutrosophic statistics can be studied. Another fruitful area of the research may be the extension of the proposed Fisher's exact test using other sampling schemes.

Acknowledgements

Thanks to the editor and reviewers for their valuable comments to improve the quality and presentation of the paper. This research was funded by Princess Nourah bint Abdulrahman University and Researchers Supporting Project number (PNURSP2023R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

M.A and F.S.A wrote the paper.

Funding

This research was funded by Princess Nourah bint Abdulrahman University and Researchers Supporting Project number (PNURSP2023R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

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 competing interests regarding the paper.

Received: 11 March 2023 Accepted: 14 August 2023

Published online: 23 August 2023

References

1. Kanji GK. 100 Statistical tests. Newcastle upon Tyne: Sage; 2006.
2. Chen Y-P. Do the chi-square test and fisher's exact test agree in determining extreme for 2×2 tables? *Am Stat*. 2011;65(4):239–45.
3. Choi L, Blume JD, Dupont WD. Elucidating the foundations of statistical inference with 2×2 tables. *PLoS ONE*. 2015;10(4):e0121263.
4. Zhong H, Song M. A fast exact functional test for directional association and cancer biology applications. *IEEE/ACM Trans Comput Biol Bioinf*. 2018;16(3):818–26.
5. Ma L, Mao J. Fisher exact scanning for dependency. *J Am Stat Assoc*. 2019;114(525):245–58.
6. West LJ, Hankin RK. Exact tests for two-way contingency tables with structural zeros. *J Stat Softw*. 2008;28:1–19.
7. Bolboacă SD, et al. Pearson-fisher chi-square statistic revisited. *Information*. 2011;2(3):528–45.
8. Ludbrook J. Analysing 2×2 contingency tables: which test is best? *Clin Exp Pharmacol Physiol*. 2013;40(3):177–80.
9. Smarandache F. Neutrosophy. Ann Arbor: Neutrosophic probability, set, and logic, proquest information & learning; 1998.
10. Smarandache F. Introduction to neutrosophic measure, neutrosophic integral, and neutrosophic probability. Infinite Study; 2013
11. Basha SH, et al. Hybrid intelligent model for classifying chest X-ray images of COVID-19 patients using genetic algorithm and neutrosophic logic. *Soft Comput*. 2021. <https://doi.org/10.1007/s00500-021-06103-7>.
12. Das R, Mukherjee A, Tripathy BC. Application of neutrosophic similarity measures in Covid-19. *Ann Data Sci*. 2021. <https://doi.org/10.1007/s40745-021-00363-8>.
13. Smarandache F. Introduction to neutrosophic statistics. HITEC City: Infinite Study; 2014.
14. 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.
15. Chen J, et al. Expressions of rock joint roughness coefficient using neutrosophic interval statistical numbers. *Symmetry*. 2017;9(7):123.
16. 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>.
17. Saeed M, et al. An application of neutrosophic hypersoft mapping to diagnose hepatitis and propose appropriate treatment. *IEEE Access*. 2021;9:70455–71.
18. Aslam M, Arif OH. Test of association in the presence of complex environment. *Complexity*. 2020. <https://doi.org/10.1155/2020/2935435>.
19. Aslam M. Chi-square test under indeterminacy: an application using pulse count data. *BMC Med Res Methodol*. 2021;21(1):1–5.
20. Parthiban S, Gajivaradhan P. A comparative study of chi-square goodness-of-fit under fuzzy environments. *Int Knowl Shar Platf*. 2020;6(2):2224.
21. Nosakhare UH, Bright AF. Statistical analysis of strength of W/S test of normality against non-normal distribution using monte carlo simulation. *Am J Theor Appl Stat*. 2017;6(5–1):62–5.

Publisher's Note

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

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com