



# Analysis and Allocation of Cancer-Related Genes Using Vague DNA Sequence Data

Muhammad Aslam<sup>\*†</sup> and Mohammed Albassam

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

To test the equality of several independent multinomial distributions, the chi-square test for count data is applied. The existing test can be applied when complete information about the data is available. The complex process, such as DNA count, the existing test under classical statistics may mislead. To overcome the issue, the modification of the chi-square test for multinomial distribution under neutrosophic statistics is presented in this paper. The modified form of the chi-square test statistic under indeterminacy/uncertainty is presented and applied using the DNA count data. From the DNA count data analysis, simulation, and comparative studies, the proposed test is found to be informative, springy, and good as compared with the existing tests.

**Keywords:** multinomial distribution, chi-square test, classical statistics, neutrosophy, DNA data

## OPEN ACCESS

### Edited by:

Xuekui Zhang,  
University of Victoria, Canada

### Reviewed by:

Florentin Smarandache,  
University of New Mexico,  
United States  
Surapati Pramanik,  
Nandalal Ghosh B. T. College, India

### \*Correspondence:

Muhammad Aslam  
aslam\_ravian@hotmail.com

### †ORCID ID:

Muhammad Aslam  
<https://orcid.org/0000-0003-0644-1950>

### Specialty section:

This article was submitted to  
Statistical Genetics and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 19 January 2022

**Accepted:** 17 February 2022

**Published:** 19 April 2022

### Citation:

Aslam M and Albassam M (2022)  
Analysis and Allocation of Cancer-  
Related Genes Using Vague DNA  
Sequence Data.  
Front. Genet. 13:858005.  
doi: 10.3389/fgene.2022.858005

## INTRODUCTION

Without statistical analysis, it is not possible to check the significance of variables under study. For testing the significance of variables, statistical tests are applied in a variety of fields (Ali & Bhaskar, 2016 and Greenland et al., 2016). The chi-square test for multinomial distribution is applied for testing whether the allocation of objects to different groups is equally likely or not. This test is applied for testing the null hypothesis that allocation of objects to different groups is equal vs. the alternative hypothesis that allocation of objects to different groups is unequal. The test statistic is computed from the data, and the null hypothesis is accepted if the values of the statistic fall within the acceptance region. Cohen, Kolassa, & Sackrowitz (2006) use the test for equality of multinomial distributions. Chafai & Concordet (2009) study confidence intervals for multinomial distribution in the case of small samples. Turner, Deng, & Houle (2020) use the statistical tests for head and face data. Shin, Yamamoto, Brady, Lee, & Haynes (2019) and Mollan et al. (2019) discuss the applications of statistical tests.

Statistical methods are widely used in analyzing and testing the significance of DNA data. A rich literature of statistical methods analyzing DNA data is available. Goldman (1993a) applies statistical tests using DNA data. Buldyrev et al. (1998) and Kugiumtzis & Provata (2004) analyze DNA data using statistical physics. Yoshida, Kobayashi, Futagami, & Fujikoshi (1999) use statistical analysis for DNA data. Pai, Mathew, & Anindya (2021) work on prediction using DNA data. Yao, Jin, & Lee (2018) improve the statistical analysis for genetic data. Gunasekaran et al. (2021) analyze DNA data using hybrid models. Halla-aho and Lähdesmäki (2021) use statistical analysis for DNA cancer data. More applications of the statistical techniques for DNA data can be seen in Goldman (1993b), Keinduangjun, Piamsa-nga, & Poovorawan (2005), Rodriguez et al. (2012), and Pai et al. (2021).

Fuzzy-based statistical tests are applied when the data in hand has vague or incomplete information. Viertl (2006) mentions that “statistical data are frequently not precise numbers but more or less non-precise also called fuzzy. Measurements of continuous variables are always fuzzy to

a certain degree.” Several studies using fuzzy-based multinomial distribution are available in the literature. Amirzadeh, Mashinchi, & Yaghoobi (2008) study multinomial distribution using fuzzy logic. Mashuri & Ahsan (2018) work on a fuzz-based chart using multinomial distribution. More information for fuzzy-based multinomial distribution can be seen in Amirzadeh et al. (2008) and Hrafnkelsson, Oddsson, & Unnthorsson (2016).

Smarandache (2013) discusses that neutrosophic logic is more efficient than interval- and fuzzy-based analysis. Neutrosophic statistics are applied to analyze the data having neutrosophic numbers; see F Smarandache (2014). Interval statistics use interval data to capture the data in the interval only and are silent about the measure of indeterminacy. On the other hand, fuzzy-based analysis only gives information about the measure of truth and of falseness. Neutrosophic statistics become classical statistics when no indeterminate information is found in the data. Chen et al. (2017a,b) introduced the methods to deal with the neutrosophic data. Later on, Sherwani et al. (2021), Aslam (2021), and Albassam, Khan, & Aslam (2021) introduced statistical tests under neutrosophic statistics.

The chi-square test for multinomial distribution available in the literature can be applied when full information about data is given. Complex processes or processes under uncertainty do not possess the full information about the data or level of significance. Therefore, there is a gap in the design of the chi-square test for multinomial distribution under neutrosophic statistics. Therefore, in this study, the chi-square test for multinomial distribution using neutrosophic statistics is introduced the first time according to the best of the author’s knowledge. The application of the proposed test is given with the aid of DNA cancer data. It is expected the proposed test will be more competent than the existing tests in terms of springy, deftness, and goodness.

## METHODS

The existing test for the equality of multinomial distribution can only be utilized when no vague information is presented. To overcome this issue, modification of the existing test is necessary. In this section, modification of the existing test under classical statistics is presented under neutrosophic statistics. With the expectation that the proposed test for the equality of multinomial distribution performs better for testing the null hypothesis under an uncertain environment. The main objective of the paper is to introduce the test for the equality of  $h_N$  independent neutrosophic multinomial distributions. Let  $Y_{1jN}, Y_{2jN} \dots Y_{kjN}$  ( $j = 1, 2, \dots, h_N$ ) present the neutrosophic frequencies for the neutrosophic events  $A_{1N}, A_{2N} \dots A_{kN}$ . Let  $p_{ijN} = P(A_{iN})$ ;  $i_N = 1, 2, \dots, k_N$ ;  $j_N = 1, 2, \dots, h_N$ . The neutrosophic form of  $p_{ijN} \in [p_{ijL}, p_{ijU}]$  is expressed as

$$p_{ijN} = p_{ijL} + p_{ijU} I_{p_{ijN}}; I_{p_{ijN}} \in [I_{p_{ijL}}, I_{p_{ijU}}] \tag{1}$$

where  $p_{ijL}$  presents the determined part, and  $p_{ijU} I_{p_{ijN}}$  presents the indeterminate part and  $I_{p_{ijN}} \in [I_{p_{ijL}}, I_{p_{ijU}}]$  is the measure of indeterminacy. The alternative expression of Eq. 1 can be given as

$$p_{ijN} = (1 + I_{p_{ijN}}) p_{ij}; I_{p_{ijN}} \in [I_{p_{ijL}}, I_{p_{ijU}}] \tag{2}$$

The  $j_{th}$  experiment is carried out  $n_{jN}$  times under the assumption that  $n_{jN}$  instances are independent. The modified form of the test statistic  $Q_N \in [Q_L, Q_U]$  is expressed as follows:

$$Q_N = Q_L + Q_U I_{Q_N}; I_{Q_N} \in [I_{Q_L}, I_{Q_U}] \tag{3}$$

where

$$Q_N = \sum_{j=1}^{h_N} \sum_{i=1}^{k_N} \frac{(Y_{ijN} - n_{jN} p_{ijN})^2}{n_{jN} p_{ijN}}$$

The proposed statistic  $Q_N \in [Q_L, Q_U]$  can be written as

$$Q_N = \sum_{j=1}^{h_L} \sum_{i=1}^{k_L} \frac{(Y_{ijL} - n_{jL} p_{ijL})^2}{n_{jL} p_{ijL}} + \sum_{j=1}^{h_U} \times \sum_{i=1}^{k_U} \frac{(Y_{ijU} - n_{jU} p_{ijU})^2}{n_{jU} p_{ijU}} I_{Q_N}; I_{Q_N} \in [I_{Q_L}, I_{Q_U}] \tag{4}$$

The simplified form of statistic can be written as

$$Q_N = (1 + I_{Q_N}) \sum_{j=1}^{h_N} \sum_{i=1}^{k_N} \frac{(Y_{ijN} - n_{jN} p_{ijN})^2}{n_{jN} p_{ijN}}; I_{Q_N} \in [I_{Q_L}, I_{Q_U}] \tag{5}$$

Note that the proposed test  $Q_N \in [Q_L, Q_U]$  is a generalization of the test under classic statistics. The proposed test  $Q_N \in [Q_L, Q_U]$  reduces to the classic test under classic statistics when  $I_{Q_L} = 0$ . The proposed test is also a generalization of the tests under interval statistics and fuzzy-based logic. The proposed test  $Q_N \in [Q_L, Q_U]$  follows the neutrosophic chi-square distribution with  $h_N (k_N - 1)$  degree of freedom. The proposed test  $Q_N \in [Q_L, Q_U]$  is applied to test the following null hypothesis:

$$H_{0N}: p_{i1} = p_{i2} = \dots = p_{ih_N} = p_{iN}, \quad i = 1, 2, 3, \dots, k_N \tag{6}$$

Under the null hypothesis, we estimate  $k_N - 1$  probabilities from

$$\hat{p}_{iN} = \frac{\sum_{j=1}^{h_L} Y_{ijL}}{\sum_{j=1}^{h_L} n_{jL}} + \frac{\sum_{j=1}^{h_U} Y_{ijU}}{\sum_{j=1}^{h_U} n_{jU}} I_{\hat{p}_{iN}}; I_{\hat{p}_{iN}} \in [I_{\hat{p}_{iL}}, I_{\hat{p}_{iU}}] \tag{7}$$

The statistic  $Q_N \in [Q_L, Q_U]$  based on  $\hat{p}_{iN} \in [\hat{p}_{iL}, \hat{p}_{iU}]$  is expressed as

$$Q_N = \sum_{j=1}^{h_L} \sum_{i=1}^{k_L} \frac{(Y_{ijL} - n_{jL} \hat{p}_{ijL})^2}{n_{jL} \hat{p}_{ijL}} + \sum_{j=1}^{h_U} \times \sum_{i=1}^{k_U} \frac{(Y_{ijU} - n_{jU} \hat{p}_{ijU})^2}{n_{jU} \hat{p}_{ijU}} I_{Q_N}; I_{Q_N} \in [I_{Q_L}, I_{Q_U}] \tag{8}$$

The simplified form of statistic can be written as

$$Q_N = (1 + I_{Q_N}) \sum_{j=1}^{h_N} \sum_{i=1}^{k_N} \frac{(Y_{ijN} - n_{jN} \hat{p}_{ijN})^2}{n_{jN} \hat{p}_{ijN}}; I_{Q_N} \in [I_{Q_L}, I_{Q_U}] \tag{9}$$

Note that  $Q_N \in [Q_L, Q_U]$  based on  $\hat{p}_{iN} \in [\hat{p}_{iL}, \hat{p}_{iU}]$  follows the neutrosophic chi-square distribution with  $(h_N - 1)(k_N - 1)$  degree of freedom.

**TABLE 1 |** The counts of nucleotide data.

Category	1	2	3	4	Total
Nucleotide	A	C	G	T	
Count Group 1	38,514	24,631	25,685	38,249	127,079
Count Group 2	38,550	24,635	25,700	38,288	127,173

## APPLICATION

In this section, the application of the proposed test is given using DNA sequence data. The data is related to the cancer-related gene BRCA 2. According to <https://medlineplus.gov/genetics/gene/brca2/#:~:text=Mutations%20in%20the%20BRCA2%20gene,one%20generation%20to%20the%20next> “Mutations in the BRCA2 gene are associated with an increased risk of breast cancer in both men and women, as well as several other types of cancer. These mutations are present in every cell in the body and can be passed from one generation to the next.” By following <https://www.math.mcgill.ca/~dstephens/OldCourses/204-2007/Handouts/Math204-ChiSquareWithResults.pdf>, the counts of nucleotide (A, C, G, T) having two counting groups are reported in **Table 1**. Note here that, in **Table 1**, the data given in “Count Group 1” is selected from the given reference, and the data given in “Count Group 2” is generated by simulation. The DNA sequence is a complex process, and there may be uncertainty/indeterminacy in counts; see Yurov, Vorsanova, & Iourov (2011). In the presence of uncertainty/indeterminacy in counts, the proposed test can be applied more effectively than the existing test under classic statistics. Suppose that there is 5% uncertainty/indeterminacy in counts of the numbers of nucleotides (A, C, G, T) in the DNA sequence of the cancer-related gene BRCA 2. Based on the information and data given in **Table 1**, the proposed test statistic is calculated as follows:

$$\sum_{j=1}^4 \sum_{i=1}^4 \frac{(Y_{ijL} - n_{jL} \hat{P}_{ijL})^2}{n_{jL} \hat{P}_{ijL}} = 0.000365921 + 0.002051303 + \dots + 0.000748132 = 0.00664$$

The statistic  $Q_N \epsilon [Q_L, Q_U]$  in neutrosophic form can be expressed as follows:

$$Q_N = 0.00664 + 0.00664 I_{Q_N}; I_{Q_N} \epsilon [0, 0.05]$$

The simplified form of statistic can be written as

$$Q_N = (1 + 0.05)0.00664 = 0.00697; I_{Q_N} \epsilon [0, 0.05]$$

The proposed test DNA count data is implemented in the following steps.

**Step 1:** State the null hypothesis  $H_0$ : The allocation of DNA count is equally likely vs. the alternative hypothesis  $H_1$ : The allocation of DNA count is unequal.

**Step 2:** The level of significance  $\alpha = 0.05$  and the tabulated value from Kanji (2006) is 9.35.

**Step 3:** Compute the value of statistic  $Q_N = 0.00697$  and compare it with the tabulated value.

**Step 4:** As the computed value of  $Q_N$  is less than 9.35,  $H_0$  is accepted.

Based on the analysis, it can be concluded that there is no evidence to suspect unequal allocation of counts of nucleotide (A, C, G, T).

## SIMULATION STUDY

A simulation study is performed to assess the effect of indeterminacy  $I_{Q_N}$  in counts of the numbers of nucleotides (A, C, G, T) in the DNA sequence of the cancer-related gene BRCA 2 on the statistic  $Q_N$ . To see the effect of  $I_{Q_N}$  on the statistic  $Q_N$ , various values of  $I_{Q_N}$  are considered. Using the neutrosophic form obtained for the DNA count data, the values of statistic  $Q_N$  are shown in **Table 2**. From **Table 2**, it can be noted that, as the value indeterminacy  $I_{Q_N}$  increases, the values of  $Q_N$  also increase. The decision about  $H_0$  at various values of  $I_{Q_N}$  is also shown in **Table 2**. From **Table 2**, although the values of statistic  $Q_N$  increase as  $I_{Q_N}$  increases, but it does not change the decision about the acceptance  $H_0$ .

## COMPARATIVE STUDIES

The springy, deftness, and goodness of the proposed test over the tests under interval statistics, the fuzzy-based approach, and classic statistics is shown in this section. The efficiency of the proposed test is shown in terms of the measure of indeterminacy, springyness, deftness, and goodness. The neutrosophic form of the statistic  $Q_N \epsilon [Q_L, Q_U]$  is expressed as follows:

$$Q_N = 0.00664 + 0.00664 I_{Q_N}; I_{Q_N} \epsilon [0, 0.05]$$

The abovementioned neutrosophic form is based on two types of information. The first part, 0.00664, gives information about the determinate part, and the second part,  $0.00664 I_{Q_N}$ , gives information about the indeterminate part. The proposed statistic  $Q_N \epsilon [Q_L, Q_U]$  reduces to the test under classic statistics when  $I_{Q_N} = 0$ . Therefore, it can be analyzed that the existing test under classic statistics gives only information about the determinate part. On the other hand, the proposed test gives information about the indeterminacy additionally as compared with the test using classic statistics. Therefore, the proposed test is more bendable than the existing test under classic statistics. The interval statistics only utilize the information given in the interval. In simple words, the interval statistics capture the information between intervals. Now comparing the results of the proposed test under the test statistic under interval statistics, it can be seen that the proposed test is more explanatory than the test using interval statistics as earlier it did not give any information about the measure of indeterminacy. Therefore, the proposed test is also more efficient than the test using the interval-based statistic. The test statistic using fuzzy logic can be considered measures of truth and falseness. The neutrosophic statistics use the set analysis and can be used for any type of set. The proposed statistic  $Q_N \epsilon [Q_L, Q_U]$  gives three types of information. The

**TABLE 2 |** The effect of Indeterminacy on  $Q_N$ .

$I_{Q_N}$	$Q_N$	Decision about $H_0$	$I_{Q_N}$	$Q_N$	Decision about $H_0$
(0, 0)	(0.00664, 0.00664)	Do not reject $H_0$	(0, 0.1)	(0.00664, 0.007304)	Do not reject $H_0$
(0, 0.01)	(0.00664, 0.006706)	Do not reject $H_0$	(0, 0.2)	(0.00664, 0.007968)	Do not reject $H_0$
(0, 0.02)	(0.00664, 0.006773)	Do not reject $H_0$	(0, 0.3)	(0.00664, 0.008632)	Do not reject $H_0$
(0, 0.03)	(0.00664, 0.006839)	Do not reject $H_0$	(0, 0.4)	(0.00664, 0.009296)	Do not reject $H_0$
(0, 0.04)	(0.00664, 0.006906)	Do not reject $H_0$	(0, 0.5)	(0.00664, 0.00996)	Do not reject $H_0$
(0, 0.05)	(0.00664, 0.006972)	Do not reject $H_0$	(0, 0.6)	(0.00664, 0.010624)	Do not reject $H_0$
(0, 0.06)	(0.00664, 0.007038)	Do not reject $H_0$	(0, 0.7)	(0.00664, 0.011288)	Do not reject $H_0$
(0, 0.07)	(0.00664, 0.007105)	Do not reject $H_0$	(0, 0.8)	(0.00664, 0.011952)	Do not reject $H_0$
(0, 0.08)	(0.00664, 0.007171)	Do not reject $H_0$	(0, 0.9)	(0.00664, 0.012616)	Do not reject $H_0$
(0, 0.09)	(0.00664, 0.007238)	Do not reject $H_0$	(0, 1)	(0.00664, 0.01328)	Do not reject $H_0$

proposed test states that the chance of accepting  $H_0$  is 0.95 (a measure of truth), the chance of committing a type-I error is 0.05 (a measure of falseness), and the measure of indeterminacy associated with the test is 0.05. From the study, it is concluded that the proposed test is also a generalization of the test using fuzzy logic. Therefore, the proposed test is more informative than the three existing tests.

## CONCLUDING REMARKS

The modification of the existing test for the equality of multinomial distribution under neutrosophic statistics is introduced in the paper. The proposed test is the generalization of several existing tests under interval statistics, fuzzy-based, and classic statistics. The modification of the test statistic is presented in the presence of indeterminacy. The simulation and comparative studies show that the proposed test is adequate and effective to apply in the presence of uncertainty. The application of the proposed test for DNA count data also shows its efficiency. The proposed test can be applied for testing the allocation of count is equally likely or not in medical science, engineering, and political science. More properties of the proposed test can be studied in future research. The proposed test using a double sampling scheme is another fruitful area for future research.

## REFERENCES

- Albassam, M., Khan, N., and Aslam, M. (2021). Neutrosophic D'Agostino Test of Normality: An Application to Water Data. *J. Mathematics* 2021, 1. doi:10.1155/2021/5582102
- Ali, Z., and Bhaskar, S. (2016). Basic Statistical Tools in Research and Data Analysis. *Indian J. Anaesth.* 60 (9), 662. doi:10.4103/0019-5049.190623
- Amirzadeh, V., Mashinchi, M., and Yaghoobi, M. A. (2008). Construction of Control Charts Using Fuzzy Multinomial Quality. *J. Mathematics Stat.* 4 (1), 26–31. doi:10.3844/jmssp.2008.26.31
- Aslam, M. (2021). Neutrosophic Statistical Test for Counts in Climatology. *Scientific Rep.* 11 (1), 1–5. doi:10.1038/s41598-021-97344-x
- Buldyrev, S. V., Dokholyan, N. V., Goldberger, A. L., Havlin, S., Peng, C. K., Stanley, H. E., et al. (1998). Analysis of DNA Sequences Using Methods of Statistical Physics. *Physica A: Stat. Mech. Its Appl.* 249 (1-4), 430–438. doi:10.1016/s0378-4371(97)00503-7

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

The paper was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, thank DSR for their financial and technical support.

## ACKNOWLEDGMENTS

The authors are deeply thankful to the editor and reviewers for their valuable comments to improve the quality of the paper. The authors, therefore, thank DSR for their financial and technical support.

- Chafai, D., and Concordet, D. (2009). Confidence Regions for the Multinomial Parameter with Small Sample Size. *J. Am. Stat. Assoc.* 104 (487), 1071–1079. doi:10.1198/jasa.2009.tm08152
- Chen, J., Ye, J., and Du, S. (2017a). Scale Effect and Anisotropy Analyzed for Neutrosophic Numbers of Rock Joint Roughness Coefficient Based on Neutrosophic Statistics. *Symmetry* 9 (10), 208. doi:10.3390/sym9100208
- Chen, J., Ye, J., Du, S., and Yong, R. (2017b). Expressions of Rock Joint Roughness Coefficient Using Neutrosophic Interval Statistical Numbers. *Symmetry* 9 (7), 123. doi:10.3390/sym9070123
- Cohen, A., Kolassa, J., and Sackrowitz, H. (2006). A Test for Equality of Multinomial Distributions vs Increasing Convex Order Institute of Mathematical Statistics. *Recent Dev. Nonparametric Inference Probab.* 1, 156–163. doi:10.1214/074921706000000662
- Goldman, N. (1993a). Simple Diagnostic Statistical Tests of Models for DNA Substitution. *J. Mol. Evol.* 37 (6), 650–661. doi:10.1007/BF00182751
- Goldman, N. (1993b). Statistical Tests of Models of DNA Substitution. *J. Mol. Evol.* 36 (2), 182–198. doi:10.1007/bf00166252

- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (2016). Statistical Tests, P Values, Confidence Intervals, and Power: a Guide to Misinterpretations. *Eur. J. Epidemiol.* 31 (4), 337–350. doi:10.1007/s10654-016-0149-3
- Gunasekaran, H., Ramalakshmi, K., Rex Macedo Arokiaraj, A., Deepa Kanmani, S., Venkatesan, C., Suresh Gnana Dhas, C. J. C., et al. (2021). Analysis of DNA Sequence Classification Using CNN and Hybrid Models. *Comput. Math. Methods Med.* 2021, 1835056. doi:10.1155/2021/1835056
- Halla-aho, V., and Lähdesmäki, H. (2021). *Probabilistic Modeling Methods for Cell-Free DNA Methylation Based Cancer Classification (bioRxiv Preprint)*. doi:10.1101/2021.06.18.444402
- Hrafnkelsson, B., Oddsson, G., and Unnthorsson, R. (2016). A Method for Estimating Annual Energy Production Using Monte Carlo Wind Speed Simulation. *Energies* 9 (4), 286. doi:10.3390/en9040286
- Kanji, G. K. (2006). *100 Statistical Tests*. United Kingdom: Sheffield Hallam University.
- Keinduangjun, J., Piamsa-nga, P., and Poovorawan, Y. (2005). “DNA Sequence Identification by Statistics-Based Models,” in *Paper Presented at the International Conference on Fuzzy Systems and Knowledge Discovery*. 1.
- Kugiumtzis, D., and Provata, A. (2004). Statistical Analysis of Gene and Intergenic DNA Sequences. *Physica A: Stat. Mech. Its Appl.* 342 (3–4), 623–638. doi:10.1016/j.physa.2004.05.070
- Mashuri, M., and Ahsan, M. (2018). *Performance Fuzzy Multinomial Control Chart. Paper Presented at the Journal of Physics: Conference Series*.
- Mollan, K. R., Trumble, I. M., Reifeis, S. A., Ferrer, O., Bay, C. P., Baldoni, P. L., et al. (2019). *Exact Power of the Rank-Sum Test for a Continuous Variable. arXiv Preprint arXiv:1901.04597*. Available at: <https://arxiv.org/abs/1901.04597>.
- Pai, S. S., Mathew, A. R., and Anindya, R. (2021). *A Comparative Analysis of Computational Tools for the Prediction of Epigenetic DNA Methylation from Long-Read Sequencing Data*. doi:10.1101/2021.04.24.441281
- Rodriguez, B. A., Frankhouser, D., Murphy, M., Trimarchi, M., Tam, H.-H., and Curfman, J. (2012). Methods for High-Throughput MethylCap-Seq Data Analysis. *BMC Genomics* 13 (6), 1–11. doi:10.1186/1471-2164-13-s6-s14
- Sherwani, R. A. K., Shakeel, H., Saleem, M., Awan, W. B., Aslam, M., and Farooq, M. (2021). A New Neutrosophic Sign Test: An Application to COVID-19 Data. *PLoS One* 16 (8), e0255671. doi:10.1371/journal.pone.0255671
- Shin, D., Yamamoto, Y., Brady, M. P., Lee, S., and Haynes, J. A. (2019). Modern Data Analytics Approach to Predict Creep of High-Temperature Alloys. *Acta Materialia* 168, 321–330. doi:10.1016/j.actamat.2019.02.017
- Smarandache, F. (2013). *Introduction to Neutrosophic Measure, Neutrosophic Integral, and Neutrosophic Probability: Sitech – Education*.
- Smarandache, F. (2014). *Introduction to Neutrosophic Statistics, Sitech and Education Publisher, Craiova. Romania-Educational Publ. Columbus, Ohio USA* 123, 1.
- Turner, D. P., Deng, H., and Houle, T. T. (2020). Statistical Hypothesis Testing: Overview and Application. *Headache: J. Head Face Pain* 60 (2), 302–308. doi:10.1111/head.13706
- Viertl, R. (2006). Univariate Statistical Analysis with Fuzzy Data. *Comput. Stat. Data Anal.* 51 (1), 133–147. doi:10.1016/j.csda.2006.04.002
- Yao, Y., Jin, Z., and Lee, J. H. (2018). An Improved Statistical Model for Taxonomic Assignment of Metagenomics. *BMC Genet.* 19 (1), 98–11. doi:10.1186/s12863-018-0680-1
- Yoshida, K., Kobayashi, M., Futagami, K., and Fujikoshi, Y. (1999). “Statistical Analysis of DNA Sequencing Data (1): Accuracy Test of DNA Data by Partial Re-Sequencing,” in *Paper Presented at the Nucleic Acids Symposium Series*, 1.
- Yurov, Y. B., Vorsanova, S. G., and Iourov, I. Y. (2011). The DNA Replication Stress Hypothesis of Alzheimer’s Disease. *Scientific World J.* 11, 2602–2612. doi:10.1100/2011/625690

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Aslam and Albassam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.