Neutrosophic Sets and Systems, Vol. 58, 2023
University of New Mexico

Normality Testing under Neutrosophic Statistics

Rahul Thakur[1], S.C.Malik[2*], N. Nandal[3]

[1] Department of Statistics, Maharishi Dayanand University, Rohtak, Haryana, India (124001); thakurrahul3394@gmail.com

[2] Department of Statistics, Maharishi Dayanand University, Rohtak, Haryana, India (124001) ;sc_malik@rediffmail.com

[3] Department of Statistics, Maharishi Dayanand University, Rohtak, Haryana, India (124001) ;nsinghnandal@gmail.com

* Correspondence: sc_malik@rediffmail.com

Abstract

The normal distribution has been extensively been used in the statistical inference and decision making problems. One of the main assumptions of the existing tests in classical statistics is that the data should be from normally distributed population.Shapiro- Wilk test is widely used to test the normality of data when the observations are precise or exact in nature. But in many areas including agriculture, engineering and reliability, the data may be in interval form, indeterminate form or uncertain. In such cases, the existed Classical Shapiro- Wilk test fails to test the normality of data. In this paper, we proposed the Shapiro- Wilk test under neutrosophic environment to check whether the uncertain data is from neutrosophic normal distribution or not. The hypothesis testing process is executed on the observations based on lifetime of batteries. The comparative analysis has been done with the existed Shapiro- Wilk test.The comparison shows that the proposed test is efficacious, appropriate and well-suited to be applied in scenarios involving indeterminacy.

Keywords: Neutrosophic Statistics, Shapiro- Wilk Test, Neutrosophic Normal Distribution, Classical Statistics, Hypothesis Testing.

1. Introduction

Now a day, inferential statistics have been used commonly in all fields of the research to test the hypotheses and making predictions on the bases of data. The statistical tests in inferential statistics have common assumption about data that it should be taken from a population following a specific distribution. It is a crucial assumption for the selection of relevant test. The distribution from which the sample has been taken is always unknown in advance. In classical statistics, there are two ways to check the distribution of the data. First is graphical approach in which the graphs are formed on the bases of given data. Another strategy that yields more trustworthy and superior results is “goodness of fit”. These tests used the “cumulative distribution function” of the fitted or underlying distribution. For the testing of the assumption that the data follows the generalized Pareto distribution or not, Anderson- Darling test proposed by Arshad et al. [1]. Various tests, most notably "goodness-of-fit," are used to determine whether the sample has been taken from a specific distribution. The most commonly used “goodness-of-fit” tests are; Jarque–Bera [2-3], Kolmogorov–Smirnov [4-5], Lilliefors [6], “Pearson's chi-square” [7], “Cramèr–von Mises” [8-9], D'Agostino–Pearson [10] and Anderson–Darling [11-12]. Aslam [13] proposed a new test of “goodness of fit” under the presence of neutrosophic parameters. Ahsan-ul-Haq [14] discussed the Cramèr–von Mises test under uncertainty. Smarandache [15] proposed a generalisation of fuzzy logic known as neutrosophic logic. The advantages of neutrosophic logic over fuzzy logic and interval-based analysis were shown by Smarandache and Khalid [16].The more literature, related articles and books on the neutrosophic statistics can be view in [17-18]. Shapiro [19] proposed the Shapiro–Wilk test for checking the assumption of normality. Nornadiah and Yap [20] proved that the Shapiro–Wilk test is most powerful normality test as comparison to the Kolmogorov–Smirnov test, Lilliefors test and Anderson–Darling test.Jeyaraman et al. [21] defined the neutrosophic norms and made some finding regarding the respective categories. Uma and Nandhitha[22] determined the Quick Switching System using the Neutrosophic Poisson distribution and compared with the Fuzzy Poisson distribution by means of Operating Characteristic (OC) curves. Dey and Ray[23] investigated some properties of the redefined neutrosophic composite relations. Utilizing the principles of neutrosophic science, Jdid and Smarandache [24]reformulated the Lagrangian multiplier technique originally designed for nonlinear models constrained by equality.

Under uncertainty, the existed Classical Shapiro- Wilk test is failed to test the normality assumption. Here, we proposed the Shapiro-Wilk test under neutrosophic environment to check whether the uncertain data is from neutrosophic normal distribution or not. We discussed the hypothesis testing procedure on lifetime of batteries. The comparative analysis has been done with the existed Shapiro-Wilk test.

2. Preliminaries

Let us consider $A_N = A_l + A_u D_N$ are the neutrosophic numbers, such that $D_N \in [D_l, D_u]$ is an indeterminacy interval, follows that neutrosophic normal distribution (NND) [17-18]with the neutrosophic mean $\mu_N = \mu_l + \mu_u D_N$; $D_N \in [D_l, D_u]$ and neutrosophic variance $\sigma_N^2 = \sigma_l^2 + \sigma_u^2 D_N$;$D_N \in [D_l, D_u]$. The "probability density function" of the NND is given by

$$f_N(A_N) = \frac{1}{\sigma_N\sqrt{2\pi}} exp\left\{-\frac{(A_N-\mu_N)^2}{2\sigma_N^2}\right\} ; \mu_N \in [\mu_l, \mu_u] , \sigma_N^2 \in [\sigma_l^2, \sigma_u^2], D_N \in [D_l, D_u] \quad (1)$$

The NND given by equation (1) is the generalized version of existed normal distribution. NND will reduce to the classical normal distribution if the $D_l = 0$.

3. Shapiro- Wilk Test Under Neutrosophic Statistics (NSW)

The Shapiro- Wilk (SW) test is a frequently used test in classical statistics to access the normality of the data. Here, we extend the SW test under the neutrosophic environment. We'll go over the algorithm for determining whether the given data follows the NND or not. The neutrosophic parameters of NND are assumed to be unknown and estimated from the given neutrosophic data in order to develop the proposed test. When compared to the traditional SW test, the proposed test will produce the results in terms of interval of indeterminacy. The assumptions of the NSW test are that the data should consist neutrosophic observations and observations should be independent of each other. The null and alternative hypothesis for the NSW test is:

H_{0N}: The data has been taken from the NND.

H_{aN}: The data has not been from the NND.

The intended application procedure for the neutrosophic Shapiro-Wilks test is outlined as follows:

Step 1: Find the mean of neutrosophic observations $\boldsymbol{A_{(i)N}}$ where $\boldsymbol{i}$ =1, 2,…, n i.e.

$$\overline{A_{(i)N}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{A_{(i)N}} \tag{2}$$

Step 2: Find the Neutrosophic Sum of Squares (NSS) by subtracting the neutrosophic mean value from neutrosophic observations then squaring and summing the obtained neutrosophic values as given below:

$$NSS = \sum_{i=1}^{n}\left(A_{(i)N} - \overline{A_{iN}}\right)^2 \tag{3}$$

Step 3: Now calculated g_N as given below

$$g_N = \left(\sum_{i=1}^{n} w_i\left(A_{(n+1-i)N} - A_{(i)N}\right)\right)^2 \tag{4}$$

where w_i weights taken from Shapiro-Wilk Table [19].

$A_{(i)N}$ is the ith smallest neutrosophic number or ith neutrosophic order statistic. The median value is not included if n is odd.

Step 4: Compute the value of test statistic for NSW is given by

$$W_N = \frac{\left(\sum_{i=1}^{n} w_i (A_{(n+1-i)N} - A_{(i)N})\right)^2}{\sum_{i=1}^{n} \left(A_{(i)N} - \overline{A_{\iota N}}\right)^2} \; ; \; W_N \in [W_l, W_u] \; and D_N \in [D_l, D_u] \quad (5)$$

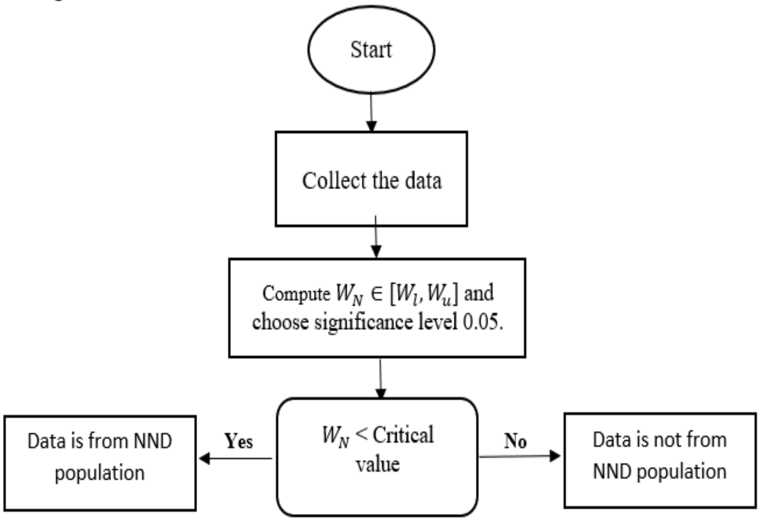

Figure 1. Testing procedure of NSW test

From the Shapiro-Wilks table, we choose the critical value corresponding to significance level α. The null hypothesis, which assumes that the data conforms to theNND, is considered valid if the calculated Shapiro-Wilk test statistic (W_N) lies within the range of critical values. Otherwise, we can conclude that the data do not conforms the NND

4. Application

In this section, we will delve into the implementation of the suggested test by utilizing data of life-span of batteries in which the life time is observed for twenty-three batteries. This data set was utilized by [13]. In practice, the failure time of the batteries cannot be measured precisely becauseit's tough to know exactly when they'll stop working.But it can be measured in neutrosophic form. The lifetime in 100h of twenty-three batteries are given below:

Table 1. The failure time of 23 batteries (in 100h)

2.9,3.99	5.24,7.2	6.56,9.02	7.14,9.82	11.6,15.96	12.14,16.69	12.65,17.4	13.24,18.21
13.67,18.79	13.88,19.09	15.64,21.51	17.05,23.45	17.4,23.93	17.8,24.48	19.01,26.14	19.34,26.59
23.13,31.81	23.34,32.09	26.07,35.84	30.29,41.65	43.97,60.46	48.09,66.13	73.48,98.04	

The engineers make some claim that the average failure time of the batteries is somewhere 3000 hours. For this they need to access whether the data has been taken from normally distributed population or not. From table 1, the data is in interval form that is indeterministic in nature. Hence, it is not possible to use the SW Test under classical statistics. Therefore, we use here proposed NSW to check the normality of data.

H_{0N}: The sample of batteries came from the neutrosophic Normal distribution.

H_{aN}: The sample of batteries do not come from the neutrosophic Normal distribution.

Step 1:the mean of neutrosophic observations given in table 1 is

$$\overline{A_{(i)N}} = \frac{1}{n}\sum_{i=1}^{n} A_{(i)N} = \frac{1}{23}(\,473.63,647.66)$$

Step 2: By equation (3), theNeutrosophic Sum of Squares (NSS)is given by

$$\sum_{i=1}^{n}\left(A_{(i)N} - \overline{A_{iN}}\right)^2 = (6912.268,11460.13)$$

Step 3: The value of sample size (n) = 23 which is odd. From Shapiro-Wilk Table [19], the weightsw_i values corresponding to n =23 are 0.4542, 0.3126, 0.2563, 0.2139, 0.1787, 0.1480, 0.1201, 0.0941, 0.0696, 0.0459, 0.0228, 0.0000.

$$\left(\sum_{i=1}^{n} w_i (A_{(n+1-i)N} - A_{(i)N})\right)^2 = (3736.699,9183.037)$$

Step 4: The value of test statistic for NSW is given by

$$W_N = \frac{(3736.699,9183.037)}{(6912.268,11460.13)}$$

$$\mathrm{W_N} = (0.540589,0.801303) \text{ where } W_l = 0.540589, W_u = 0.801303$$

We found that the critical value corresponding to n = 23 and significance level 0.01 is (0.881,0.881). Since $W_N < (0.881,0.881)$, for the lifetime of batteries data. Hence it can be concluded that the lifetime of batteries follows the NND.

5. Comparative Study & Discussion

In this analysis, we compared the effectiveness of the provided NSW test with the traditional SW test. When working with data characterized by imprecise, uncertain, or ambiguous observations, the suggested test demonstrates greater efficiency by delivering outcomes in the form of indeterminacy.The obtained value of NSWtest statistic is $W_N \in (0.5406, 0.8013)$. It can be written as $W_N = 0.5406 + 0.8013D_N$; $D_N \in [0, 0.3254]$. The critical value of neutrosophic Shapiro-Wilk Test is (0.881,0.881) at 0.01significance level. Here the value 0.5406 represents the value of Classical Shapiro-Wilk Test when the $D_N = 0$ .For the significance level 0.01, "the probability of the rejecting the null hypothesis when it is true is 0.01 and probability of accepting the null hypothesis when it is true will be 0.99 and measure of indeterminacy is 0.3254". Hence it can be said that the proposed Shapiro-Wilk Test of normality under neutrosophic statistics gives the test statistic value with the measure indeterminacy (D_N) while the existed Shapiro-Wilk Test of normality under classical statistics fails to provide any information about the measure of indeterminacy. Therefore, the proposed test of normality under uncertainty or neutrosophic statistic is more effective than the existed Shapiro-Wilk Test and Classical Shapiro-Wilk Test becomes special case of the suggested Shapiro-Wilk Test under neutrosophic environment when the indeterminacy $(D_N) = 0$.

6. Conclusion

In this paper, we suggested the Shapiro-Wilk Test under the presence of neutrosophic data. We obtained the value of test statistic of the proposed test and decision rule. We take the data of the lifetime of batteries in neutrosophic form and used the proposed model to check whether the data is taken from neutrosophic normally distribution population or not. From the obtained value of test statistic, we conclude that the lifetime of batteries follows the neutrosophic normal distribution and data can be used for further analysis under neutrosophic statistical inferences where it is necessary that the obtained data should be from neutrosophic normally distributed population. From the comparative study, the proposed test of normality under uncertainty or neutrosophic statistic is more effective than the existed Shapiro-Wilk Test and Classical Shapiro-Wilk Test becomes special case of the proposed Shapiro-Wilk Test under neutrosophic statistics when the indeterminacy $(D_N) = 0$. In future, the proposed test can be applied on some other data to check the neutrosophic normality.

Acknowledgements: "The authors are deeply thankful to the editor and the reviewers for their useful suggestions which really enable us for improving this manuscript."

Funding: "This research work has no external Funding."

Conflicts of Interest: "The authors declare no Conflict of interest."

References

1. Arshad, M; Rasool, M; Ahmad, M. Anderson Darling and Modified Anderson darling tests for generalized pareto distribution. *Journal of Applied Sciences.***2003**, 3(2):85–88. doi: 10.3923/jas.2003.85.88
2. Jarque, C.M.; Bera, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic. Leters.***1980**, 6, 255–259, doi:10.1016/0165-1765(80)90024-5.
3. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. International Statistical Review, 1987, 55, 163–172, doi:10.2307/1403192.
4. Kolmogorov, A. Sulla determinazioneempirica di unalegge di distribuzione. *Giornaledell'IstitutoItalianodegliAttuari***1933**, 4, 83–91.

5. Smirnov, N. Table for estimating the goodness of fit of empirical distributions. The *Annals of Mathematical Statistics.***1948**, 19, 279–281, doi:10.1214/aoms/1177730256.

6. Lilliefors, H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown.*Journal of the American Statistical Association*, **1967**, 62, 399–402, doi:10.2307/2283970.

7. Pearson, K. Contribution to the mathematical theory of evolution. II. Skew variation in homogenous material. Philosphical Transactions of the Royal Society,**1895**, 91, 343–414, doi:10.1098/rsta.1895.0010

8. Cramér, H. On the composition of elementary errors. *Scand. Actuar. J.***1928**, 1, 13–74, doi:10.1080/03461238.1928.10416862.

9. Von Mises, R.E. Wahrscheinlichkeit, Statistik und Wahrheit; Julius Springer: Berlin, Germany, **1928**, 10. Shapiro

10. D'Agostino, R.B.; Belanger, A.; D'Agostino, R.B., Jr. A suggestion for using powerful and informative tests of normality. *Am. Stat.***1990**, 44, 316–321, doi:10.2307/2684359.

11. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Ann. Math. Stat*. **1952**, 23, 193–212, doi:10.1214/aoms/1177729437. 6.

12. Anderson, T.W.; Darling, D.A. A Test of Goodness-of-Fit. *J. Am. Stat. Assoc.***1954**, 49, 765–769, doi:10.2307/2281537.

13. Aslam M. A new goodness of fit test in the presence of uncertain parameters. *Complex & Intelligent Systems.***2021**, 7(1):359-65.

14. Ahsan-ul-Haq M.A new Cramèr–von Mises Goodness-of-fit test under Uncertainty. *Neutrosophic Sets and Systems.***2022**,49; 262-268.

15. Smarandache, F. Neutrosophy. In: Neutrosophic probability, set, and logic, proquest information & learning. *Ann Arbor.***1998**,105, 118–123.

16. Smarandache, F; Khalid, H.E. Neutrosophic precalculus and neutrosophic calculus,**2015**, infinite study

17. Smarandache, F. Definitions derived from neutrosophic, **2003**, Infinite Study. 19.

18. Smarandache, F. Introduction to Neutrosophic Statistics, Sitech and Education Publisher, Craiova. Romania-Educational Publisher, Columbus, Ohio, **2014**, USA, 123.

19. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika***1965**, 52, 591–611, doi:10.1093/biomet/52.3-4.591.

20. Nornadiah, M.R.;Yap, B.W. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics.***2011**,2(1); 21-33.
21. Jeyaraman, M.; Iswariya, S.; and Pandiselvi, R. Generalized Double Statistical Convergence Sequences on Ideals in Neutrosophic Normed Spaces.*Neutrosophic Systems with Applications*. **2023**, 8, 50–60. (doi: https://doi.org/10.5281/zenodo.8222997)
22. Uma G.;Nandhitha S. An Investigative Study on Quick Switching System using Fuzzy and Neutrosophic Poisson Distribution.*Neutrosophic Systems with Applications*. **2023**, 7, 61–70. (doi: https://doi.org/10.5281/zenodo.8218361)
23. Dey S.; Ray G.C., Properties of Redefined Neutrosophic Composite Relation.*Neutrosophic Systems with Applications*. 2023 7: 1–12. (Doi: https://doi.org/10.5281/zenodo.8207037)
24. Jdid F.;Smarandache F. Lagrange Multipliers and Neutrosophic Nonlinear Programming Problems Constrained by Equality Constraints.*Neutrosophic Systems with Applications*.**2023**, 6, 25–31. (doi: https://doi.org/10.5281/zenodo.8215851)

Received: June 1, 2023. Accepted: Sep 28, 2023