Statistical Hypothesis Test Under Fuzzy Observations by Euler Centroid Method

S. Parthiban[1] A. Sridhar[2] S. Vinoth[3] T. L. Yookesh[4, *]

[1, 3, 4]Department of Mathematics, Vignan's Foundation for Science, Technology & Research, AP 522213, India.

[2]PG and Research Department of Mathematics, Government Thirumagal Mills College, Gudiyattam, Tamil Nadu, India.

Article Info
Page Number: 1968-1976
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract
A statistical hypothesis test is an indispensable and powerful tool for making better decisions in various fields. When it comes to tasks involving large populations, it is not always possible to make accurate sample observations. That is, real-time observations may be accurate or imprecise by their nature. If the observed samples are imprecise, the corresponding samples can be manipulated with fuzzy numbers; more generally trapezoidal fuzzy numbers. Moreover, using a new ranking method derived from the lifespan of Euler centroid these fuzzy samples are fuzzy and a relevant statistical procedure was followed to test the hypothesis and obtain better results.

Keywords: Test of hypothesis, trapezoidal fuzzy numbers, Euler's line, ranking method, expected interval, value of fuzzy numbers.

Introduction

The fuzzy compilation theory has been applied in many areas related to vague and uncertain data. These areas include basic reasoning, decision-making, optimization, and control. Most of our traditional descriptive and inferential statistical tools depend on data clarity (accuracy), measurements, random variables, assumptions, etc. Type yes or no instead of plus or minus. However, there are many situations where the above assumptions are unrealistic, so we need new tools to characterize and analyze the problem. With the introduction of fuzzy set theory, several branches of mathematics have been explored recently. But probabilities and statistics have received more attention in this regard due to their random nature. Mathematical statistics do not have methods for analyzing problems where random variables are fuzzy. With this in mind, a new, simple technique for testing hypotheses in widespread environments is proposed. Here, the observed data are trapezoidal fuzzy numbers (TrFN).

TrFN have many advantages over linear and non-linear membership functions. First, TrFN form a regular class of fuzzy numbers with a linear membership function. This blurred speech class completely enhances the much-discussed triangular blurred speech class, suggesting its common property. In this way, TrFN find many applications in the linear uncertainty model in utility specialized technical and scientific problems, including fuzzy - linear systems, traffic problems, and classification problems etc.

The use of the fuzzy set hypothesis for measurements is described in Manton et al., Buckley and Viertl [6, 7]. Arnold [5] proposed confusion with the conventional statistical hypothesis and tested the suspicion as Type I and II errors under ambiguous conditions. Saade et al., analyzed the experiment with binary hypotheses and studied the probability functions of fuzzy in decision-making using a fuzzy version of Bayes' rule. Grzegorzewski, Watanabe, and Imaizumi [10, 11, 20] proposed a fuzzy test to test the hypothesis of fuzzy data, and it revealed the adequacy of erroneous and selective theories. Wu [21, 22, 23] recommended testing theories of fact to confuse information by suggesting the terms trust and negativity. Viertl [18, 19] looked at some strategies for creating safety margins and statistical tests for fuzzy data. Arefi and Taheri [4] have found a way to tackle fuzzy test theories to get fuzzy test information out of vague data.

In this article, we propose a new fuzzy statistical hypothesis test that uses Student's t-test for a single mean. With the proposed technique, the observed data are inherently fuzzy as trapezoidal fuzzy numbers. And these fuzzy numbers are removed from their ranking function by a unique process by finding the best balancing points using the Euler line. The proposed technique has been illustrated by an example.

Preliminaries

Definition-1

A fuzzy set $\widetilde{A}$ of a universal X is defined by its membership function $\mu_{\widetilde{A}}: X \to [0,1]$ and we write$\widetilde{A} = \{(x_1, \mu_{\widetilde{A}}(x_1)): x_1 \in X\}$.

Definition-2

A fuzzy set $\widetilde{A}$ is called normal fuzzy set if there exists an element (member) 'x_1' such that $\mu_{\widetilde{A}}(x_1) = 1$.

Definition-3

A generalized fuzzy number $\widetilde{A}$ is described as any fuzzy subset of the real line , whose membership function $\mu_{\widetilde{A}}(x_1)$ satisfies the following conditions:

i. $\mu_{\widetilde{A}}(x_1)$ is a continuous mapping from to the closed interval$[0, \omega], 0 \leq \omega \leq 1$,
ii. $\mu_{\widetilde{A}}(x_1) = 0$, for all $x_1 \in (-\infty, a_1]$,

iii. $\mu_L(x_1) = L(x_1)$ is strictly increasing on$[a_1, b_1]$,
iv. $\mu_{\tilde{A}}(x_1) = \omega$, for all $[b_1, c_1]$
v. $\mu_R(x_1) = R(x_1)$is strictly decreasing on $[c_1, d_1]$,
vi. $\mu_{\tilde{A}}(x_1) = 0$, for all $x_1 \in [d_1, \infty)$

where a_1, b_1, c_1, d_1 are real numbers such that $a_1 < b_1 \le c_1 < d_1$.

Definition-4

A fuzzy number $\widetilde{A} = (a_1, b_1, c_1, d_1; 1)$ is said to be a Normalized Trapezoidal Fuzzy Number if its membership function is given by

$$\begin{cases} 0 & ; x_1 < a_1 \\ \dfrac{x_1 - a_1}{b_1 - a_1} & ; a_1 < x_1 \le b_1 \\ 1 & ; b_1 < x_1 < c_1 \\ \dfrac{d_1 - x_1}{d_1 - c_1} & ; c_1 \le x_1 < d_1 \\ 0 & ; x_1 > d_1 \end{cases}$$

Where $a_1 \le b_1 \le c_1 \le d_1$.

Centroid of TrFNs by Euler Line Method

Euler line:

In any triangle, the centroid, circumcenter and orthocenter are always on a straight line called the Euler line, that is, they are linear points. Moreover, a triangle may determine many points and those points may or may not within or on boundaries of the triangle. The circumcenter, the centroid and the orthocenter are some of such points. In this work, the circumcenter, the centroid and the orthocenter are taken as the primary referral points of any triangle.

For an equilateral triangle, these points are concentric and they do not represent a line. But, if the triangle is not equilateral, then the points are distinct and they determine lines. This line contains the centroid, the circumcenter and the orthocenter of the triangle called a Euler line.

A cartesian representation of the Euler line can be found by any two of the three points noted above. Let us consider a GTrFN $\breve{T} = (t_1, t_2, t_3, t_4; w)$. This GTrFN is partitioned into three plane areas such as triangle APB, rectangle PQCB and triangle QCD. Let the centroids of those three plane figures be G_1, G_2, and G_3 respectively. These centroids form a triangle and the proposed Euler line passes through the centroid of the centroids of those three planes and therefore this Euler line can be taken as the better balancing line of the trapezoid.

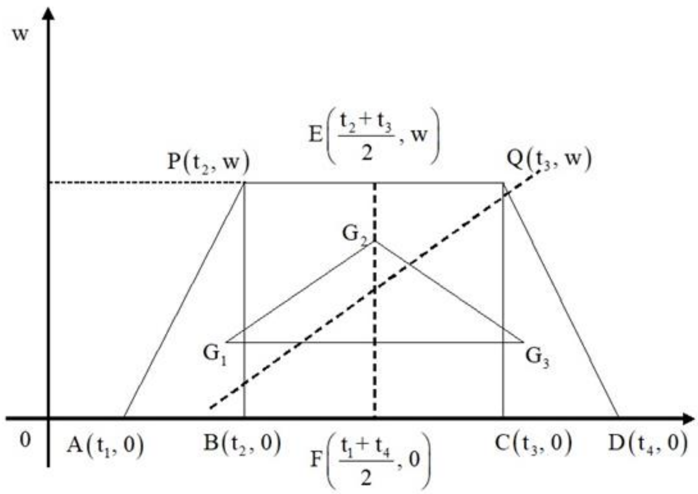

Figure: Euler line of the Centroids

Consider a GTrFN $\widetilde{T} = (t_1, t_2, t_3, t_4; w)$. The centroid of the three plane figures are

$G_1 = \left(\frac{t_1+2t_2}{3}\right)$; $G_2 = (\frac{t_2+t_3}{2})$ and $G_3 = (\frac{2t_3+t_4}{3})$ respectively. Equation of the line joining G_1 and G_3 is $y = \frac{w}{3}$ and G_2 does not lie on the line $\overline{G_1G_3}$. Hence G_1, G_2 and G_3 are non-collinear and they form a triangle.

Euler line of GTrFN $\widetilde{T} = (t_1, t_2, t_3, t_4; w)$:

$$\bar{y}_0 + \left\{\frac{3(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + w^2}{2w(t_2 + t_3 - t_1 - t_4)}\right\} x_0$$
$$= \frac{(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2)(7t_2 + 7t_3 + 2t_1 + 2t_4) + w^2}{12w(t_2 + t_3 - t_1 - t_4)} --- (1)$$

Now the equation of the line $\overline{EF}$ is,

$$\bar{y}_0 - \frac{2w\bar{x}_0}{t_2 + t_3 - t_1 - t_4} = \frac{-(t_1 + t_4)w}{t_2 + t_3 - t_1 - t_4} --- (2)$$

Now, the point of intersection of (1) and (2) is denoted by

$$I_{\widetilde{T}_G}(x_i, y_i) = \left\{\begin{array}{c} \frac{(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2)(7t_2 + 7t_3 + 2t_1 + 2t_4) + w^2(8t_1 + 7t_2 + 7t_3 + 8t_4}{6[3(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + 5w^2]}, \\ \frac{7w}{3} \cdot \frac{(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + w^2}{3(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + 5w^2} \end{array}\right\}$$

--- (3)

Hence, the point of intersection referred as the more balancing point for a normalized trapezoidal fuzzy number (NTrFN) $\widetilde{\mathrm{T}} = (\mathrm{t}_1, \mathrm{t}_2, \mathrm{t}_3, \mathrm{t}_4; 1)$ will be

$$\mathrm{I}_{\widetilde{\mathrm{T}}_{\mathrm{N}}}(\mathrm{x_i}, \mathrm{y_i}) = \left\{ \begin{array}{c} \dfrac{(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2)(7t_2 + 7t_3 + 2t_1 + 2t_4) + (8t_1 + 7t_2 + 7t_3 + 8t_4)}{6[3(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + 5]}, \\ \dfrac{7}{3} \cdot \dfrac{(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + 1}{3(2t_1 + t_2 - 3t_3)(t_3 + 2t_4 - 3t_2) + 5} \end{array} \right\}$$

--- (4)

Now, the rank of the NTrFN $\widetilde{\mathrm{T}} = (\mathrm{t}_1, \mathrm{t}_2, \mathrm{t}_3, \mathrm{t}_4; 1)$ it displays the set of all fuzzy numbers $(\mathrm{x_i}, \mathrm{y_i})$ defined as the geometric mean.

That is, $\mathrm{R}(\widetilde{\mathrm{T}}) = (\mathrm{x}^2 + \mathrm{y}^2)^{1/2} - - - (5)$

Using equation (4), the obtained fuzzy samples are transformed to real numbers and the test of hypothesis will be applied to arrive results accordingly.

Numerical example:

Hindustan Copper Limited, a public sector enterprise produces copper products in the market like copper cathodes, wire bars and rods etc. To test the tensile strength of copper wire from a specific line, ten samples of copper wire with the following tensile strength (in kg. Unit) are taken from a large area. Due to the manufacturing congestion, the observed samples are around the exact samples having four approximations and they are recorded in terms of trapezoidal fuzzy numbers as follows:

(570, 572, 578, 579), (569, 572, 575, 578), (568, 570, 573, 576), (565, 568, 570, 574), (570, 572, 575, 578), (574, 578, 580, 582), (566, 570, 573, 576), (568, 572, 575, 578), (563, 567, 569, 573), (540, 545, 548, 550)

But it is assumed from the company authority that the mean breaking strength of the lot is 578 kg. weight. From this fuzzy observation, a statistical hypothesis test will be applied by using the proposed unique defuzzification method and thereby a reliable result will be concluded.

The null hypothesis $\widetilde{\mathrm{H}}_0$: The products meet the mean breaking strength

That is, $\widetilde{\mathrm{H}}_0: \mu = 578$ kg. weight

The alternative hypothesis $\widetilde{\mathrm{H}}_{\mathrm{A}}$: The products do not meet the mean breaking strength

That is, $\widetilde{\mathrm{H}}_{\mathrm{A}}: \mu \neq 578$ kg. weight (Two tailed test)

Now, the better balancing spot of the observed fuzzy numbers can be found by using equation (4) by letting them as NTrFNs.

Let $\widetilde{T}_1 = (570, 572, 578, 579; 1)$; here $t_1 = 570$, $t_2 = 572$, $t_3 = 578$ and $t_4 = 579$. Substituting these values in equation (4) we get,

$$I_{\widetilde{T}_1} = \left\{ \frac{\big(2(570)+572-3(578)\big)\big(578+2(579)-3(572)\big)\big(7(572)+7(578)+2(570)+2(579)\big)+\big(8(570)+7+7(572)+7(578)+8(579)\big)}{6\big[3\big(2(570)+572-3(578)\big)\big(578+2(579)-3(572)\big)+5\big]}, \frac{7}{3}.\frac{\big(2(570)+572-3(578)\big)\big(578+2(579)-3(572)\big)+1}{3\big(2(570)+572-3(578)\big)\big(578+2(579)-3(572)\big)+5} \right\}$$

That is, $I_{\widetilde{T}_1} = (574.8895, 0.7790)$

Similarly, for the remaining fuzzy data, $I_{\widetilde{T}_m}; m = 2, 3, \dots, 10$, the balancing spot have been obtained and tabulated below:

$I_{\widetilde{T}_m}; m = 1, 2, 3, \dots, 10$	
$I_{\widetilde{T}_1} = (574.8895, 0.7790)$	$I_{\widetilde{T}_6} = (578.7815, 0.7815)$
$I_{\widetilde{T}_2} = (573.5000, 0.7801)$	$I_{\widetilde{T}_7} = (571.3899, 0.7798)$
$I_{\widetilde{T}_3} = (571.6098, 0.7805)$	$I_{\widetilde{T}_8} = (573.3899, 0.7798)$
$I_{\widetilde{T}_4} = (569.1096, 0.7809)$	$I_{\widetilde{T}_9} = (568.0000, 0.7804)$
$I_{\widetilde{T}_5} = (573.6098, 0.7805)$	$I_{\widetilde{T}_{10}} = (546.1698, 0.7799)$

And the corresponding rank of the NTrFNs at their balancing spot are obtained by calculating their geometric mean described in equation (5).

That is, for $I_{\widetilde{T}_1} = (574.8895,\ 0.7790)$,

$$R_{\widetilde{T}_1} = [(574.8895)^2 + (0.7790)^2]^{1/2}$$

$$R_{\widetilde{T}_1} = 574.8900$$

The ranks of the remaining data R_{T_i}; $i = 2, 3, \dots, 10$ have been calculated and tabulated below:

$R_{\tilde{T}_i}$; i = 1, 2, 3, …,10	
$R_{\tilde{T}_1} = 574.8900$	$R_{\tilde{T}_6} = 578.7821$
$R_{\tilde{T}_2} = 573.5005$	$R_{\tilde{T}_7} = 571.3904$
$R_{\tilde{T}_3} = 571.6103$	$R_{\tilde{T}_8} = 573.3904$
$R_{\tilde{T}_4} = 569.1101$	$R_{\tilde{T}_9} = 568.0005$
$R_{\tilde{T}_5} = 573.6103$	$R_{\tilde{T}_{10}} = 546.0455$

Test of hypothesis:

The sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{10} R_{\tilde{T}_i} = 570.0455$

And the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{10}(R_{\tilde{T}_i} - \bar{x}) = 79.4406 \Rightarrow s = 8.9129$

Now, the test statistic is, $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{570.0455-578}{\left(\frac{8.9129}{\sqrt{10}}\right)} = -2.8222 \Rightarrow |t| = 2.8222$

Here, the tabulated value of ‘t’ at 5% level of significance with n – 1 = 10 – 1 = 9 degrees of freedom is $t_{\propto} = 2.262$.

Conclusion:

Here, the calculated value of ‘t’ is greater than the tabulated value of ‘t’. That is, $|t| > t_{\propto} \Rightarrow$ the null hypothesis $\tilde{H}_0$ is rejected. Therefore, the mean value of the breaking strength of the copper wires 578 kg. unit is not reliable.

Even though better decision can be arrived by using the method of ranking function obtained from the balancing spot of fuzzy numbers by using Euler line, it is not an ultimate solution to all problems. Moreover, it has shown the way for the test of hypotheses involving fuzzy observations such as TrFNs through their defuzzified forms such as rank, center of gravity, better balancing point etc. and of course it needs further refinement and research to arrive better result.

Reference:

1. S. Abbasbandy and M. Amirfakhrian, The nearest approximation of a fuzzy quantity in parametric form, Appl. Math. Comput., 172 (2006) 624-632.
2. S. Abbasbandy and M. Amirfakhrian, The nearest trapezoidal form of generalized left right fuzzy number, Internat. J. Approx. Reason., 43 (2006) 166-178.
3. S. Abbasbandy and T. Hajjari, Weighted trapezoidal approximation-preserving cores of a fuzzy number, Comput, and Maths. with Applications, 59 (2010) 3066-3077.
4. M. Arefi and S. M. Taheri, Testing fuzzy hypotheses using fuzzy data based on fuzzy test statistic, Journal of Uncertain Systems, 5 (2011) 45-61.
5. B. F. Arnold, Testing fuzzy hypotheses with crisp data, Fuzzy Sets and Systems, 94 (1998) 323-333.
6. J. J. Buckley, Fuzzy statistics, Springer-Verlag, New York, 2005.
7. J. Chachi, S. M. Taheri and R. Viertl, Testing statistical hypotheses based on fuzzy confidence intervals, Forschungsbericht SM-2012-2, Technische Universitat Wien, Austria, 2012.
8. D. Dubois and H. Prade, Operations on fuzzy numbers, Int. J. Syst. Sci., 9 (1978) 613-626.
9. P. Gajivaradhan and S. Parthiban, Two sample statistical hypothesis test for trapezoidal fuzzy interval data, Int. J. of Applied Mathematics and Statistical Sciences. Vol. 04, Issue: 05, Aug.-Sep. 2015, pp. 11-24.
10. P. Grzegorzewski, Testing statistical hypotheses with vague data, Fuzzy Sets and Systems, 112 (2000) 501-510.
11. P. Grzgorzewski and E. Mrowka, Trapezoidal approximation of fuzzy numbers, Fuzzy Sets and Systems, 153 (2005) 115-135.
12. S. C. Gupta and V. K. Kapoor, Fundamentals of mathematical statistics, Sultan Chand & Sons, New Delhi, India.
13. R. R. Hocking, Methods and applications of linear models: regression and the analysis of variance, New York: John Wiley & Sons, 1996.
14. George J. Klir and Bo Yuan, Fuzzy sets and fuzzy logic, Theory and Applications, Prentice-Hall, New Jersey, 2008.
15. Michael J. Ross, The Euler Line's Cartesian Formula (27-06-2018).
16. Parthiban, S. and Gopinathan, P., Statistical Hypothesis on Industrial Applications through Ranks from COG of TrFNs, International Journal of Recent Technology and Engineering (IJRTE), Vol. 6, Issue: 1S4, (June 2019), pp. 1116-1118.
17. Parthiban, S. and Gajivaradhan, P., Statistical hypothesis test in three factor ANOVA model under fuzzy environments using trapezoidal fuzzy numbers, Bulletin of Mathematical Sciences and Applications, Vol. 14, (2016), pp. 23-42.
18. R. Viertl, Univariate statistical analysis with fuzzy data, Computational Statistics and Data Analysis, 51 (2006) 33-147.
19. R. Viertl, Statistical methods for fuzzy data, John Wiley and Sons, Chichester, 2011.
20. N. Watanabe and T. Imaizumi, A fuzzy statistical test of fuzzy hypotheses, Fuzzy Sets and Systems, 53 (1993) 167-178.
21. H. C. Wu, Analysis of variance for fuzzy data, International Journal of Systems Science, 38 (2007) 235-246.
22. H. C. Wu, Statistical confidence intervals for fuzzy data, Expert Systems with Applications, 36 (2009) 2670-2676.
23. H. C. Wu, Statistical hypotheses testing for fuzzy data, Information Sciences, 175 (2005) 30-56.
24. L. A. Zadeh, Fuzzy sets, Information and Control, 8 (1965) 338-353.

25. T. P. Latchoumi, G. Kalusuraman, J. F. Banu, T. L. Yookesh, T. P. Ezhilarasi and K. Balamurugan, Enhancement in manufacturing systems using Grey-Fuzzy and LK-SVM approach, 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), (2021) 72-78.