Int. J. Anal. Appl. (2022), 20:70

International Journal of Analysis and Applications

Testing the Difference of Means of Populations With Respect to Intuitionistic Fuzzy Sets

Sasiwimon Iwsakul[1], Khamika Urawong[1], Thammarat Panityakul[1], Ronnason Chinram[1], Pattarawan Singavananda[2,*]

[1] *Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand*

[2] *Program in Mathematics, Faculty of Science and Technology, Songkhla Rajabhat University, Songkhla 90000, Thailand*

[*] *Corresponding author:* pattarawan.pe@skru.ac.th

Abstract. In this study, the tests of statistical hypotheses with crisp data using small samples are extended to with the membership function and the non-membership function of the intuitionistic fuzzy set. The test procedure of statistical hypotheses for means of two normally distributed populations with respect to any intuitionistic fuzzy set is proposed.

1. Introduction

A concept of fuzzy set was introduced by Zadeh [7] in 1965. A fuzzy set is a class of objects with a continuum of membership grades which are allocated to each object a grade of membership ranging unit close from zero to one. Later, the original of fuzzy set theory were generalized to other fuzzy sets. One popular generalized fuzzy sets proposed by Atanassov [1] starts by the specification of membership and non-membership functions. It is called an intuitionistic fuzzy set (IFS).

Hypothesis testing is one of the major aspects of data analysis. The observations of the sample are crisp and a statistical test leads to the binary conclusion. In contrast, the data occasionally cannot be collected precisely. The data sometimes collected in the form of statistical hypotheses testing under fuzzy environments. Casals, Gil and Gil [2] approach to the problem of testing statistical hypotheses with fuzzy information. The statistical hypotheses testing for fuzzy data by proposing the notions of

Received: Nov. 5, 2022.

2010 *Mathematics Subject Classification.* 62F03, 03E72.

Key words and phrases. tests of statistical hypotheses; populations; intuitionistic fuzzy sets; means; variances.

https://doi.org/10.28924/2291-8639-20-2022-70

ISSN: 2291-8639

degrees of optimism and pessimism was investigated by Wu [5]. Moreover, Wu [6] proposed some approaches to construct fuzzy confidence intervals for the unknown fuzzy parameter. Kalpanapriya and Pandian [3] proposed tests of hypothesis for means of populations using imprecise samples. The proposed test is extended to statistical hypotheses testing for fuzzy data. Moreover, Kalpanapriya and Pandian [4] investigated the method of tests of statistical hypotheses with respect to a fuzzy set.

In this paper, we propose statistical hypothesis tests using small sample (or samples) based on the IFS. We focus not only on testing of significance for the difference between two independent population means but also between two dependent population means with respect to an IFS.

2. Basic Concepts

This section outlines some of the most important definitions relevant to the study.

2.1. **Intuitionistic fuzzy sets.**

Definition 2.1. Let S be a nonempty set. A *fuzzy set* (FS) $\mathcal{A}$ drawn from a set S is defined as

$$\mathcal{A} = \{< x, \mu_A(x) >| \; x \in S\}$$

where $\mu_A : S \to [0, 1]$ is the membership function of the fuzzy set $\mathcal{A}$.

Definition 2.2. Let S be a nonempty set. An *intuitionistic fuzzy set* (IFS) $\mathcal{A}$ in a set S is an object having the form

$$\mathcal{A} = \{< x, \mu_A(x), \nu_A(x) >| \; x \in S\}$$

where the functions $\mu_A : S \to [0, 1]$ and $\nu_A : S \to [0, 1]$ defined respectively, the degree of membership and degree of non-membership of the element x to the set $\mathcal{A}$, which is a subset of S, and for every element $x \in S$, $0 \leq \mu_A(x) + \nu_A(x) \leq 1$.

For any FS $\mathcal{A} = \{< x, \mu_A(x) >| \; x \in S\}$, we can define $\nu_A(x) = 1 - \mu_A(x)$ and we have $\{< x, \mu_A(x), \nu_A(x) >| \; x \in S\}$ is an IFS. Then every FS can be considered as an IFS.

2.2. **One-sample t-test.** Let $x_1, x_2, ..., x_n$ be a random small sample of size n (where $n < 30$) from a normally distributed population. Now, the mean value, denoted by $\overline{x}$ and the sample standard deviation, denoted by s of the above small sample are given by

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}.$$

Note that s^2 is called a sample variance.

In testing the null hypothesis that the population mean μ is equal to a specified value μ_o, one uses the statistic

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}.$$

We will use that the degrees of freedom used in this test is $\nu = n - 1$.

3. Main Results

In this section, we propose the following two types of tests of statistical hypotheses:

(3.1) Test for the difference between the means of two independent populations using their small sample size with respect to an IFS.

(3.2) Test for the difference between the means of two dependent populations using their small sample size with respect to an IFS.

Definition 3.1. Let $\{x_1, x_2, \ldots, x_n\}$ be a random sample of size n from a crisp set X with membership grades and non-membership grades $\mu_A(x_i)$ and $\nu_A(x_i)$, respectively of an IFS $\mathcal{A}$. The average of membership grades and non-membership grades of an IFS $\mathcal{A}$ over the random sample or the sample mean of the membership grades and non-membership grades of the IFS $\mathcal{A}$ denoted by $\overline{\mu}_A(X)$ and $\overline{\nu}_A(X)$ respectively defined as follows:

$$\overline{\mu}_A(X) = \frac{\sum_{i=1}^{n} \mu_A(x_i)}{n} \text{ and } \overline{\nu}_A(X) = \frac{\sum_{i=1}^{n} \nu_A(x_i)}{n}.$$

Definition 3.2. Let $\{x_1, x_2, \ldots, x_n\}$ be a random sample of size n from a crisp set X with membership grades and non-membership grades $\mu_A(x_i)$ and $\nu_A(x_i)$, respectively of an IFS $\mathcal{A}$. The variance of membership grades and non-membership grades of an IFS $\mathcal{A}$ over the random sample or the sample variance of the membership grades and non-membership grades of the IFS $\mathcal{A}$ denoted by $(S_\mu(X))^2$ and $(S_\nu^2(X))^2$ respectively defined as follows:

$$(S_\mu(X))^2 = \frac{\sum_{i=1}^{n} (\mu_A(x_i) - \overline{\mu}_A(X))^2}{n-1}$$

and

$$(S_\nu(X))^2 = \frac{\sum_{i=1}^{n} (\nu_A(x_i) - \overline{\nu}_A(X))^2}{n-1}.$$

Next, we will present the methods of two main problems.

3.1. **Testing of significance for the difference between two independent population means with respect to an IFS.** Let X and Y be two crisp populations and $\mathcal{A}$ be an IFS defined on X and Y. Let $\{x_1, x_2, \ldots, x_m\}$ be a linguistic random sample of X with membership grades and non-membership grades $\mu_A(x_i)$ and $\nu_A(x_i)$, respectively where $i = 1, 2, \ldots, m$. Let $\{y_1, y_2, \ldots, y_n\}$ be another linguistic random sample of Y with membership grades and non-membership grades $\mu_A(y_j)$ and $\nu_A(y_j)$, respectively where $j = 1, 2, \ldots, n$. Suppose that $\mu_A(x_i)$, $\nu_A(x_i)$, $\mu_A(y_j)$ and $\nu_A(y_j)$ are normally distributed. Based on the samples, we test that the mean of the population X with respect to $\mathcal{A}$, $\overline{\mu}(A, X)$ and the mean of the population Y with respect to $\mathcal{A}$, $\overline{\mu}(A, Y)$ are the same.

Now, we have the null hypothesis H_0: $\overline{\mu}(A, X) = \overline{\mu}(A, Y)$.

If both populations have equal standard deviations with respect to $\mathcal{A}$, we use the test statistic for testing the null hypothesis,

$$t_\mu = \frac{\overline{\mu}_A(X) - \overline{\mu}_A(Y)}{S_\mu\sqrt{1/n + 1/m}}$$

and

$$t_\nu = \frac{\overline{\nu}_A(X) - \overline{\nu}_A(Y)}{S_\nu\sqrt{1/n + 1/m}}$$

where

$$S_\mu = \sqrt{\frac{(m-1)(S_\mu(X))^2 + (n-1)(S_\mu(Y))^2}{m+n-2}}$$

and

$$S_\nu = \sqrt{\frac{(m-1)(S_\nu(X))^2 + (n-1)(S_\nu(Y))^2}{m+n-2}}.$$

If both populations standard deviations with respect to $\mathcal{A}$ are not the same, we use the test statistic for testing the null hypothesis,

$$t_\mu = \frac{\overline{\mu}_A(X) - \overline{\mu}_A(Y)}{\sqrt{(S_\mu(X))^2/n + (S_\mu(Y))^2/m}}$$

and

$$t_\nu = \frac{\overline{\nu}_A(X) - \overline{\nu}_A(Y)}{\sqrt{(S_\nu(X))^2/n + (S_\nu(Y))^2/m}}.$$

We let $t = \max\{|t_\mu|, |t_\nu|\}$. Now, the degree of freedom used in this test is $df = m + n - 2$ and we let $t_{\alpha,d}$ denote the critical value of t for df degrees of freedom at the level of significance α. Now, for the level of significance α, the critical region of the alternative hypothesis, H_A is given below:

Alternative Hypothesis	Critical Region
$\overline{\mu}(A, X) > \overline{\mu}(A, Y)$(upper-tailed test)	$t \geq t_{\alpha,df}$
$\overline{\mu}(A, X) < \overline{\mu}(A, Y)$(lower-tailed test)	$t \leq -t_{\alpha,df}$
$\overline{\mu}(A, X) \neq \overline{\mu}(A, Y)$(two-tailed test)	$\lvert t\rvert \geq t_{\alpha/2,df}$

If $|t| \leq |t_{\alpha,df}|$ (one-tailed test), we cannot reject the null hypothesis. There is insufficient evidence to conclude that the $\overline{\mu}(A, X)$ is greater than $\overline{\mu}(A, Y)$(the means of populations with respect to $\mathcal{A}$) for (upper-tailed test) or $\overline{\mu}(A, X)$ is less than $\overline{\mu}(A, Y)$ (for lower-tailed test) at α level. Otherwise, the null hypothesis $\overline{\mu}(A, X) \leq \overline{\mu}(A, Y)$ (for upper-tailed test) or $\overline{\mu}(A, X) \geq \overline{\mu}(A, Y)$ (for lower- tailed test)) is rejected.

If $|t| \leq t_{\alpha/2,df}$ (two-tailed test), there is not enough evidence to conclude that the difference between $\overline{\mu}(A, X)$ and $\overline{\mu}(A, Y)$ at α level is significant. Therefore, the null hypothesis is not rejected. Otherwise, the null hypothesis is rejected, that is, the population means of the membership and non-membership functions of $\mathcal{A}$ are different.

Now, we get the following example.

Example 3.1. Let X and Y be two populations where X be the sets of all students in Prince of Songkla University and Y be the sets of all students in Songkhla Rajabhat University. We let the IFS $\mathcal{A}$, where μ and ν be the membership and non-membership functions of students' satisfaction toward online learning, defined on X and Y. It is assumed that $\mu_A(x_i)$, $\nu_A(x_i)$, $\mu_A(y_j)$ and $\nu_A(y_j)$ are normally distributed. Now, we are going to test that the $\mathcal{A}$ in X is better than the $\mathcal{A}$ in Y, that is, $\overline{\mu}(A, X) > \overline{\mu}(A, Y)$.

Let $S_1 = \{x_1, x_2, x_3, x_4, x_5\}$ be the sample of size five taken from students in Prince of Songkla University (the population X) and $S_2 = \{y_1, y_2, y_3, y_4, y_5, y_6\}$ be the sample of size six taken from students in Songkhla Rajabhat University (the population Y).

Then, the membership grades and the non-membership grades of the given two samples based on their information concerning the IFS $\mathcal{A}$ are given below.

Now, with the help of the numerical example given below, the procedure of the above said testing of hypothesis is explained.

	x_1	x_2	x_3	x_4	x_5
$\mu_A(x_i)$	0.7	0.8	0.9	0.8	0.7
$\nu_A(x_i)$	0.2	0.2	0	0.1	0.3

and

	y_1	y_2	y_3	y_4	y_5	y_6
$\mu_A(y_i)$	0.8	0.8	0.7	0.8	0.7	0.6
$\nu_A(y_i)$	0.2	0.1	0.1	0.1	0.2	0.1

Our hypotheses are $H_0 : \overline{\mu}(A, X) \leq \overline{\mu}(A, Y)$ and $H_1 : \overline{\mu}(A, X) > \overline{\mu}(A, Y)$.
Now, we have

$$\overline{\mu}_A(X) = 0.78, \overline{\mu}_A(Y) = 0.73, \overline{\nu}_A(x) = 0.16 \text{ and } \overline{\nu}_A(Y) = 0.13.$$

Then $S_\mu(X) = 0.08, S_\mu(Y) = 0.08, S_\nu(X) = 0.11, S_\nu(X) = 0.05$.
Since $\sigma_\mu^2 = \sigma_\nu^2$, S_μ and S_ν are computed as follows:

$$S_\mu = \sqrt{\frac{(m-1)(S_\mu(X))^2 + (n-1)(S_\mu(Y))^2}{m+n-2}} = 0.08$$

and

$$S_\nu = \sqrt{\frac{(m-1)(S_\nu(X))^2 + (n-1)(S_\nu(Y))^2}{m+n-2}} = 0.09.$$

Therefore, the tests of statistic are

$$t_\mu = \frac{\overline{\mu}_A(X) - \overline{\mu}_A(Y)}{\sqrt{(S_\mu(X))^2/n + (S_\mu(Y))^2/m}} = 0.93$$

and

$$t_\nu = \frac{\overline{\nu}_A(X) - \overline{\nu}_A(Y)}{\sqrt{(S_\nu(X))^2/n + (S_\nu(Y))^2/m}} = 0.52.$$

Then $t = \max\{|t_\mu|, |t_\nu|\} = \max\{|0.93|, |0.52|\} = 0.93$. The critical value of $t_{0.05,9}$ is 1.83. Since $0.93 < 1.83$, we cannot reject the null hypothesis. There is insufficient evidence to conclude that the $\mathcal{A}$ in X is better than the $\mathcal{A}$ in Y.

3.2. **Testing of significance for the difference between two dependent population means with respect to an IFS.** Let X and Y be two crisp populations and $\mathcal{A}$ be an IFS defined on X and Y. Let $\{x_1, x_2, \ldots, x_n\}$ be a linguistic random sample of X with membership and non-membership grades measured at the first time point $\mu_A(x_i)$ and $\nu_A(x_i)$, respectively where $i = 1, 2, \ldots, m$. Let $\{y_1, y_2, \ldots, y_n\}$ be another linguistic random sample of Y with membership and non-membership grades measured at the second time point $\mu_A(y_i)$ and $\nu_A(y_i)$, respectively where $i = 1, 2, \ldots, m$. Suppose that $\mu_A(x_i)$, $\nu_A(x_i)$, $\mu_A(y_i)$ and $\nu_A(y_i)$ are normally distributed.

The mean difference in the population X and Y with respect to A, $\overline{\mu}_d(A)$. If our objective is to test whether the mean differences in the population X and Y with respect to A , $\overline{\mu}_d(A)$ are different. Then the null hypothesis is H_0: $\overline{\mu}_d(A) = 0$. The difference between the two observations on each pair can be calculated by $D(\mu_i) = \mu_A(x_i) - \mu_A(y_i)$ and $D(\nu_i) = \nu_A(x_i) - \nu_A(y_i)$. The sample means of the differences are denoted by $\overline{D}(\mu)$ and $\overline{D}(\nu)$, respectively. The sample means of the differences can be defined as follows:

$$\overline{D}(\mu) = \sum_{i=1}^{n} \frac{D(\mu_i)}{n}$$

and

$$\overline{D}(\nu) = \sum_{i=1}^{n} \frac{D(\nu_i)}{n}.$$

The sample variances are given by

$$(S_\mu(d))^2 = \frac{\sum_{i=1}^{n} (D(\mu_i) - \overline{D}(\mu))^2}{n-1}$$

and

$$(S_\nu(d))^2 = \frac{\sum_{i=1}^{n} (D(\nu_i) - \overline{D}(\nu))^2}{n-1}.$$

The test statistics are calculated as:

$$t_\mu = \frac{\overline{D}(\mu) - 0}{\frac{(S_\mu(d))}{\sqrt{n}}}$$

and

$$t_{\nu} = \frac{\overline{D}(\nu) - 0}{\frac{(S_{\nu}(d))}{\sqrt{n}}}.$$

Then $t = \max\{|t_{\mu}|, |t_{\nu}|\}$. The degree of freedom is $df = n - 1$. Hence, the critical value is $t_{\alpha,df}$ where α is significance level of the test. The critical region of the alternative hypothesis, H_A is given below:

Alternative Hypothesis	Critical Region
$\overline{\mu}_d(A) > 0$(upper-tailed test)	$t \geq t_{\alpha,df}$
$\overline{\mu}_d(A) < 0$(lower-tailed test)	$t \leq -t_{\alpha,df}$
$\overline{\mu}_d(A) \neq 0$(two-tailed test)	$\|t\| \geq t_{\alpha/2,df}$

If $|t| \leq |t_{\alpha,df}|$ (one-tailed test), we fail to reject the null hypothesis. There is insufficient evidence to conclude that $\overline{\mu}_d(A) > 0$ (the mean difference in the populations with respect to $\mathcal{A}$ is greater than 0) for (upper-tailed test) or $\overline{\mu}_d(A) < 0$ (the mean difference in the populations with respect to $\mathcal{A}$ is less than 0) for (lower-tailed test) at α level. Otherwise, the null hypothesis $\overline{\mu}_d(A) \leq 0$ (for upper-tailed test) or $\overline{\mu}_d(A) \geq 0$ (for lower-tailed test) is rejected.

If $|t| \leq t_{\alpha/2,df}$ (two-tailed test), there is not enough evidence to conclude that there is a difference in the mean populations with respect to $\mathcal{A}$ at α level . Therefore, the null hypothesis is failed to reject. Otherwise, the null hypothesis is rejected, that is, the mean difference in the populations with respect to $\mathcal{A}$ is not equal to zero.

Example 3.2. Suppose that we assign a homework to students in Prince of Songkla University. Let X and Y be two populations where X be the sets of all assignment scores evaluated by lecturer 1 and Y be the sets of all assignment scores evaluated by lecturer 2. We let the IFS $\mathcal{A}$, where μ and ν be the membership and non-membership functions of lecturers' satisfaction toward assignment, defined on X and Y. It is assumed that $\mu_A(x_i)$, $\nu_A(x_i)$, $\mu_A(y_i)$ and $\nu_A(y_i)$ are normally distributed. Now, we perform the test, that is, the mean differences in the population X and Y with respect to A , $\overline{\mu}_d(A)$ are different. Then the null hypothesis is H_0: $\overline{\mu}_d(A) = 0$. We randomly select a sample of five students.

Let $S_1 = \{x_1, x_2, x_3, x_4, x_5\}$ be the size five taken from assignment scores evaluated by lecturer 1 (the population X) and $S_2 = \{y_1, y_2, y_3, y_4, y_5\}$ be the sample of size five taken from assignment scores evaluated by lecturer 2 (the population Y). Then, the membership grades and the non-membership grades of the given two samples based on their information concerning the IFS $\mathcal{A}$ are given below.

Student	1	2	3	4	5
	x_1	x_2	x_3	x_4	x_5
$\mu_A(x_i)$	0.7	0.8	0.6	0.8	0.9
$\nu_A(x_i)$	0.3	0.1	0.2	0.1	0
	y_1	y_2	y_3	y_4	y_5
$\mu_A(y_i)$	0.7	0.9	0.7	0.8	0.7
$\nu_A(y_i)$	0.2	0.1	0.1	0.1	0

It is assumed that $D(\mu_i) = \mu_A(x_i) - \mu_A(y_i)$ and $D(\nu_i) = \nu_A(x_i) - \nu_A(y_i)$. Hence, $\overline{D}(\mu) = 0$, $\overline{D}(\nu) = 0.04$, $(S_\mu(d)) = 0.12$ and $(S_\nu(d)) = 0.05$.
The test statistics are

$$t_\mu = \frac{\overline{D}(\mu) - 0}{\frac{(S_\mu(d))}{\sqrt{n}}} = 0$$

and

$$t_\nu = \frac{\overline{D}(\nu) - 0}{\frac{(S_\nu(d))}{\sqrt{n}}} = 1.633.$$

Therefore $t = \max\{|t_\mu|, |t_\nu|\} = 1.633$. The critical value of $t_{0.025,4}$ is 2.77. Since $1.63 < 2.77$, we cannot reject the null hypothesis. There is insufficient evidence to conclude that the mean differences in the population X and Y with respect to A , $\overline{\mu}_d(A)$ are different.

4. Conclusion

This article proposes two types of tests of statistical hypotheses based on the membership and non-membership functions of intuitionistic fuzzy sets which are completely different from conventional statistical hypothesis testing. In the proposed tests of hypotheses, the differences of means of the populations are investigated with the help of intuitionistic fuzzy sets. The rules for making decisions regarding the hypotheses are provided. Each proposed statistical hypothesis test is a characteristic or attribute based test on the population. The proposed statistical hypotheses tests can assist decision makers for selecting an appropriate decision with satisfaction.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] K.T. Atanassov, Intuitionistic Fuzzy Sets, Fuzzy Sets Syst. 20 (1986), 87-96. https://doi.org/10.1016/s0165-0114(86)80034-3.

[2] M.R. Casals, M.A. Gil, P. Gil, The Fuzzy Decision Problem: An Approach to the Problem of Testing Statistical Hypotheses With Fuzzy Information, Eur. J. Oper. Res. 27 (1986), 371-382. https://doi.org/10.1016/0377-2217(86)90333-4.

[3] D. Kalpanapriya, P. Pandian, Statistical Hypotheses Testing With Imprecise Data, Appl. Math. Sci. 6 (2012), 5285-5292.

[4] D. Kalpanapriya, P. Pandian, Tests of Statistical Hypotheses With Respect to a Fuzzy Set, Modern Appl. Sci. 8 (2014), 25-35. https://doi.org/10.5539/mas.v8n1p25.

[5] H.C. Wu, Statistical Hypotheses Testing for Fuzzy Data, Inform. Sci. 175 (2005), 30-56. https://doi.org/10.1016/j.ins.2003.12.009.

[6] H.C. Wu, Statistical Confidence Intervals for Fuzzy Data, Expert Syst. Appl. 36 (2009), 2670-2676. https://doi.org/10.1016/j.eswa.2008.01.022.

[7] L.A. Zadeh, Fuzzy Sets, Inform. Control. 8 (1965), 338-353. https://doi.org/10.1016/s0019-9958(65)90241-x.