Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 10, Number 3 (2014), pp. 473-481

http://www.ripublication.com

Fuzzy Hypothesis Testing based on Linear Correlation

P. Pandian and D. Kalpanapriya

Department of Mathematics, School of Advanced Sciences,
VIT University, Vellore–632 014, INDIA
E-mail: [1]pandian61@rediffmail.com, [2]dkalpanapriya@vit.ac.in

Abstract

Tests of fuzzy hypotheses based on linear correlation using small sample data with membership grades are presented which are tests of characteristics or attributes. The t-test statistic with respect to a fuzzy set / fuzzy sets are defined using the membership grades. Also, the rules for making decisions about the fuzzy hypotheses are provided. Further, the linguistic data examples are given for understanding the proposed testing procedures. The proposed tests of hypotheses can help decision makers to analyze a linear relation between two attributes on a population or between two populations on an attribute.

Keywords: testing hypothesis, fuzzy set, membership grades, linear correlation, t-test, confidence limits.

1. Introduction

A statistical hypothesis test is an essential part of statistical inference for analyzing the population parameters using data. In conventional hypotheses testing (Devore [7]). the samples are crisp and the significance test leads to the binary decision. In real life situations, the sample data cannot be always recorded precisely. So, imprecise data sample may be got for testing hypotheses. Arnold [3] proposed various tests of statistical hypotheses with imprecise samples. Using the fuzzy data sample, the fuzzy tests of hypotheses were discussed in Kruse[11] , Viertl [12] , Assel S, Mohammad [4] , Baloui Jamkhaneh and Nadi Ghara [5] , Grzegorzewski [9], Wu [13] and Akbari [1]. Arefi and Taheri [2] studied statistical hypotheses for comparing means of the populations with vague data samples. Hui-Li Hsu and Berlin Wu [10] investigated an innovative approach to measure the correlation for variables when the data are intervals. The computation of fuzzy correlation coefficient with fuzzy data was developed by Yu-Ting Cheng and Chih-Ching Yang [14].

In this paper, we propose two types of statistical fuzzy hypothesis tests based on linear correlation using small samples data with their membership grades namely, (i) to test the linear relation between two populations with respect to a fuzzy set and (ii) to test the linear relation between two fuzzy sets defined on a population. The t-test statistic are defined on the membership grades of the fuzzy set over a random sample. The decision rules for accepting or rejecting the fuzzy hypotheses are provided. The optimistic and pessimistic approach, h-level set, α -cut and fuzzy interval are not used in the proposed tests. The procedures of the proposed tests of hypotheses are illustrated by means of linguistic examples. The proposed hypotheses testing can help decision makers to find a linear relationship between linguistic variables of real life problems

2. Preliminaries

We need the following concepts related to fuzzy set and its MF which can be found in George Klir and Bo Yuan [8].

Let *X* and *Y* be two crisp set and *F* be a fuzzy space. Let $\tilde{A}$ and $\tilde{B} \subseteq F$ be fuzzy sets .

Then, the fuzzy set $\tilde{A}$ defined on X with a membership function (MF) $\mu_{\tilde{A}}(x)$, can be represented as follows:

$\tilde{A} = \{(x, \mu_{\tilde{A}}(x) / x \in X\}$, where $\mu_{\tilde{A}} : X \to [0,1]$.

Then, the fuzzy sets $\tilde{A}$ and $\tilde{B}$ in F defined on *X* with MFs $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ respectively, can be represented as follows.

$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in X\}$ and $\tilde{B} = \{(x, \mu_{\tilde{B}}(x)) / x \in X\}$,

where $\mu_{\tilde{A}}$ and $\mu_{\tilde{B}} : X \to [0,1]$.

Then, $\tilde{A} \subseteq F$ defined on *X* and *Y* with MFs $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{A}}(y)$ respectively , can be expressed in two different ways as follows.

$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in X\}$ and $\tilde{A} = \{(y, \mu_{\tilde{A}}(y)) / y \in Y\}$

where $\mu_{\tilde{A}} : X \to [0,1]$ and $\mu_{\tilde{A}} : Y \to [0,1]$.

Let $\tilde{A}$ be a fuzzy set defined on X with MF $\mu_{\tilde{A}}(x)$. Then, the probability of a fuzzy set (Zadeh [16]) $\tilde{A}$ is given by

$$P(\tilde{A}) = \int_x \mu_{\tilde{A}}(\tilde{x}) dP = E(\mu_{\tilde{A}}(x)) \tag{1}$$

where *P* is the probability measure over *X*. From (1), we can say that if the probability measure of *X* is known, the probability of the occurrence of the fuzzy event $\tilde{A}$ is the mean of $\mu_{\tilde{A}}(x)$.

Now, we need the following definitions of the sample mean and the sample variance of the MGs of a fuzzy set which can be found in Chiang and Lin [6].

Definition 2.1: Let $(x_1, x_2, ..., x_n)$ be a random sample of size *n* from a crisp set *X* with the Membership grades (MGs) of a fuzzy set $\tilde{A}$ where $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in X\}$. Then, the sample mean of the MF of the fuzzy $\tilde{A}$ defined on *X* or the average MGs of fuzzy set $\tilde{A}$ over the random sample $(x_1, x_2, ..., x_n) \in X$ denoted by $\overline{\mu}_{\tilde{A}}(x)$ is given by

$$\overline{\mu}_{\tilde{A}}(x) = \frac{1}{n}\left(\sum_{i=1}^{n} \mu_{\tilde{A}}(x_i)\right).$$

Definition 2.2: Let $(x_1, x_2, ..., x_n)$ be a random sample of size *n* from a crisp set *X* with the MGs of a fuzzy set $\tilde{A}$ where $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) / x \in X\}$. Then, the sample variance of the MF of the fuzzy $\tilde{A}$ defined on *X* or the variance of the MGs of fuzzy set $\tilde{A}$ over the random sample $(x_1, x_2, ..., x_n) \in X$ denoted by $S^2_{\tilde{A}}(x)$ is given by

$$S^2_{\tilde{A}}(x) = \frac{1}{n-1}\left(\sum_{i=1}^{n} (\mu_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{A}}(x))^2\right)$$

and the sample standard deviation of the MF of the fuzzy $\tilde{A}$ defined on *X*, $S_{\tilde{A}}(x) = \sqrt{S^2_{\tilde{A}}(x)}$.

Definition 2.3. Let $\{x_1, x_2, ..., x_n\}$ be a sample of size *n* from a crisp set X with MGs of fuzzy sets $\tilde{A}$ and $\tilde{B}$ defined on X , $\mu_{\tilde{A}}(x_i)$ and $\mu_{\tilde{B}}(x_i)$, $i = 1, 2, ..., n$ respectively. Then, the correlation coefficient between the fuzzy sets $\tilde{A}$ and $\tilde{B}$, $r(\tilde{A}, \tilde{B})$ is defined as follows:

$$r(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^{n} (\mu_{\tilde{A}}(x_i) - \overline{\mu}_{\tilde{A}}(x))(\mu_{\tilde{B}}(x_i) - \overline{\mu}_{\tilde{B}}(x))}{(n-1) S_{\tilde{A}}(x) S_{\tilde{B}}(x)}.$$

3. Fuzzy Hypothesis Testing

In this section, two types of tests of statistical fuzzy hypothesis based on linear correlation , namely, (i) to test of fuzzy hypothesis of correlation between two populations using their samples with respect to a fuzzy set and (ii) to test of fuzzy hypothesis of correlation of a population using its sample with respect to two fuzzy sets are discussed. Both the tests are attribute / quality based study on crisp populations.

3.1. Fuzzy hypothesis testing with respect to a fuzzy set

Let *X* and *Y* be two crisp populations and $\tilde{A}$ be a fuzzy set defined on *X* and *Y*. Let $\{x_1, x_2, \ldots, x_m\}$ be a linguistic random sample of *X* with MGs $\mu_{\tilde{A}}(x_i)$, i=1,2,…,m and $\{y_1, y_2, \ldots, y_n\}$ be another linguistic random sample of *Y* with MGs $\mu_{\tilde{A}}(y_j)$, j=1,2,…,n and $r(\tilde{A})$ be the coefficient of correlation between the samples of the two populations with respect to the fuzzy set $\tilde{A}$.

Based on the samples, we test that the population *X* with respect to $\tilde{A}$ and the population *Y* with respect to $\tilde{A}$ at the level of significance α are linearly related. Let $\rho(\tilde{A})$ be the linear correlation between X and Y with respect to $\tilde{A}$.

Let the sample mean of the MF of *A* defined on *X* be $\overline{\mu}_{\tilde{A}}(x)$, the sample mean of the MF of $\tilde{A}$ defined on *Y* be $\overline{\mu}_{\tilde{A}}(y)$, the sample variance of the MF of $\tilde{A}$ defined on *X* be $S^2_{\tilde{A}}(x)$ and the sample variance of the MF of $\tilde{A}$ defined on *Y* be $S^2_{\tilde{A}}(y)$.

Null hypothesis , H_0: No linear relationship between the two populations with respect to the fuzzy set $\tilde{A}$, that is, $\rho(\tilde{A}) = 0$.

Alternative hypothesis,

$H_A : \rho(\tilde{A}) > 0$, positively correlated with respect to the fuzzy set $\tilde{A}$

$H_A : \rho(\tilde{A}) < 0$, negatively correlated with respect to the fuzzy set $\tilde{A}$

$H_A :\ \rho(\tilde{A}) \neq 0$, correlated with respect to the fuzzy set $\tilde{A}$

Now, the test statistic $t(\tilde{A}) = \dfrac{r(\tilde{A})\sqrt{(n-2)}}{\sqrt{1-r^2(\tilde{A})}}$ with the degrees of freedom $n-2$.

Now, the rejection region of the null hypothesis H_0 against the alternative hypothesis H_A for level α is given below:

Alternative hypothesis , H_A	Rejection region of H_0 for α level
$H_A : \rho(\tilde{A}) > 0$	$t(\tilde{A}) > t_{\alpha,n-2}$ (one tailed test)
$H_A : \rho(\tilde{A}) < 0$	$t(\tilde{A}) < -t_{\alpha,n-2}$ (one tailed test)
$H_A :\ \rho(\tilde{A}) \neq 0$	$\lvert t(\tilde{A}) \rvert > t_{\alpha/2,n-2}$ (two tailed test)

Now, the 100(1 –α)% confidence interval for the population correlation with respect to $\tilde{A}$, corresponding to the given samples is given below:

$$l(\tilde{A}) < \rho(\tilde{A}) < u(\tilde{A})$$

where $l(\tilde{A})=$ maximum $\left\{ r(\tilde{A}) - t_{\alpha/2,n-2}\left(\sqrt{\frac{1-r^2(\tilde{A})}{n-2}}\right),\ -1 \right\}$ and

$u(\tilde{A})=$ minimum $\left\{ r(\tilde{A}) + t_{\alpha/2,n-2}\left(\sqrt{\frac{1-r^2(\tilde{A})}{n-2}}\right),\ 1 \right\}$.

Now, the demonstration of the procedure of the above said fuzzy hypothesis test with help of the imprecise data example is given below:

EXAMPLE 3.1. A study were carried out through interviews with engineering managers responsible for both graduate (Stage1 engineers) and experienced professional engineers (Stage2 engineers) engaged in the most common mechanical engineering roles in the industries that employ the largest proportions of Indian mechanical engineers. Let X = {Stage 1 engineers} and Y = { Stage 2 engineers } be the two populations. Let us define the fuzzy set, $\tilde{A}$ over the crisp sets *X* and *Y* as $\tilde{A}$ = { Communication skill}. Now, we are going to test that the $\tilde{A}$ = {Communication skill } in *X* is linearly related to the $\tilde{A}$ = { Communication skill } in *Y*.

Let S_1=(Thanu(x_1),Ravi(x_2), Raj(x_3), Mani(x_4), Sekar(x_5), Priya(x_6), Jaya (x_7), Mahi(x_8)) be a random sample of size eight taken from the population *X* and S_2 =(Thanu(y_1),Ravi(y_2), Raj(y_3), Mani(y_4), Sekar(y_5), Priya(y_6), Jaya(y_7), Mahi (y_8)) be a random sample of size eight taken from the population *Y*.

Then, we collect information about the two samples and obtain the MGs of these two samples concerning the fuzzy set $\tilde{A}$ as follows.

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\mu_{\tilde{A}}(x_i)$	0.68	0.69	0.71	0.71	0.72	0.73	0.74	0.76

and

y_j	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
$\mu_{\tilde{A}}(y_j)$	0.71	0.72	0.68	0.61	0.76	0.76	0.73	0.75

Now, the sample average membership grade of $\tilde{A}$ and the sample variance of $\tilde{A}$ over these two samples are $\bar{\mu}_{\tilde{A}}(x) == 0.716$; $\bar{\mu}_{\tilde{A}}(y) = 0.713$; $S^2_{\tilde{A}}(x) = 0.00057$ and $S^2_{\tilde{A}}(y) = 0.002398$.

Now, the correlation coefficient, $r(\tilde{A})$ = 0.4608.

Now, the NH, $H_0 : \rho(\tilde{A}) = 0$ (There is no correlation between the stage 1 and stage 2 engineers with respect to the communication skills) against the AH, $H_A : \rho(\tilde{A}) \neq 0$.

Now, the test statistic, $t(\tilde{A}) = \dfrac{r(\tilde{A})\sqrt{(n-2)}}{\sqrt{1-r^2(\tilde{A})}} = \dfrac{0.4608\sqrt{6}}{1-0.21195} = 1.431$

Now, the decision table about the null hypothesis and confidence limits are given below

Test statistic value	Degrees of freedom Freedom	Level of significance (Two tailed test)	Table value	Decision about H_0
$t(\tilde{A}) = 1.431$	14	5%	2.145	Accepted
		1%	2.977	Accepted

and

Confidence level	Lower limit	Upper limit
95%	-0.31638	1
99%	-0.61783	1

3.2. Fuzzy hypothesis testing between two fuzzy sets

Let X be a crisp population and let $\tilde{A}$ and $\tilde{B}$ be two fuzzy sets defined on X. Let $\{x_1, x_2, ..., x_m\}$ and $\{y_1, y_2, ..., y_n\}$ be a linguistic random samples of X with MGs $\mu_{\tilde{A}}(x_i)$ and $\mu_{\tilde{B}}(x_i)$, i=1,2,…,m and $r(\tilde{A}, \tilde{B})$ be the coefficient of correlation between the samples of the population with respect to the fuzzy sets. Based on the samples, we test that the population X with respect to $\tilde{A}$ and the population X with respect to $\tilde{B}$ at the level of significance α are linearly related. Let the linear correlation of X with respect to A and with respect $\tilde{B}$ be $\rho(\tilde{A}, \tilde{B})$. Let the sample mean of the MF of $\tilde{A}$ and $\tilde{B}$ defined on X be $\overline{\mu}_{\tilde{A}}(x)$ and $\overline{\mu}_{\tilde{B}}(x)$, the sample variance of the MF of $\tilde{A}$ and $\tilde{B}$ defined on X be $S^2_{\tilde{A}}(x)$ and $S^2_{\tilde{B}}(x)$.

Null hypothesis, H_0: No linear relationship between two fuzzy sets $\tilde{A}$ and $\tilde{B}$ defined on the population X or the population with respect to the fuzzy sets $\tilde{A}$ and $\tilde{B}$, that is, $\rho(\tilde{A}, \tilde{B}) = 0$.

Alternative hypothesis,

H_A: $\rho(\tilde{A} < \tilde{B}) > 0$, positively correlated with respect to the fuzzy sets $\tilde{A}$ and $\tilde{B}$.

H_A: $\rho(\tilde{A} < \tilde{B}) < 0$, negatively correlated with respect to the fuzzy sets $\tilde{A}$ and $\tilde{B}$.

H_A: $\rho(\tilde{A} < \tilde{B}) \neq 0$, correlated with respect to the fuzzy sets $\tilde{A}$ an $\tilde{B}$.

Now, the test statistic $t(\tilde{A},\tilde{B}) = \dfrac{r(\tilde{A},\tilde{B})\sqrt{(n-2)}}{\sqrt{1-r^2(\tilde{A},\tilde{B})}}$ with the degrees of freedom $n-2$.

Now, the rejection region of the null hypothesis H_0 against the alternative hypothesis H_A for level α is given below:

Alternative hypothesis H_A	**Rejection region of** H_0 **for** α **level**
$H_A : \rho(\tilde{A},\tilde{B}) > 0$	$t(\tilde{A},\tilde{B}) > t_{\alpha,n-2}$ (one tailed test)
$H_A : \rho(\tilde{A},\tilde{B}) < 0$	$t(\tilde{A},\tilde{B}) < -t_{\alpha,n-2}$ (one tailed test)
$H_A : \rho(\tilde{A},\tilde{B}) \neq 0$	$\lvert t(\tilde{A},\tilde{B}) \rvert > t_{\alpha/2,n-2}$ (two tailed test)

Now, the 100(1 $-\alpha$)% confidence interval for the population correlation with respect to $\tilde{A}$ and $\tilde{B}$, corresponding to the given samples is given below:

$$l(\tilde{A},\tilde{B}) < \rho(\tilde{A},\tilde{B}) < u(\tilde{A},\tilde{B})$$

where $l(\tilde{A},\tilde{B}) =$ maximum $\left\{ r(\tilde{A},\tilde{B}) - t_{\alpha/2,n-2}\left(\sqrt{\dfrac{1-r^2(\tilde{A},\tilde{B})}{n-2}}\right),\ -1 \right\}$ and

$$u(\tilde{A},\tilde{B}) = \text{minimum}\left\{ r(\tilde{A}) + t_{\alpha/2,n-2}\left(\sqrt{\frac{1-r^2(\tilde{A})}{n-2}}\right),\ 1 \right\}.$$

Now, the procedure of the above said hypothesis test is illustrated with help of the imprecise data example given below:

EXAMPLE 3.2: Let X = {workers in the office} be a population. Let us define the fuzzy sets, $\tilde{A}$ and $\tilde{B}$ over the crisp set X as $\tilde{A}$ = {passion} and $\tilde{B}$ ={ loyalty}. Now, we are going to test that the $\tilde{A}$ in X is linearly related to the $\tilde{B}$ in X.

Let S =(Priya(x_1),Dass(x_2), Rahul(x_3), Metha(x_4), Manu(x_5), Ajit(x_6), Mahi (x_7)) be a random sample of size seven taken from the population X.

Then, we collect information about the two samples and obtain the MGs of these two samples concerning the fuzzy sets $\tilde{A}$ and $\tilde{B}$ as follows.

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7
$\mu_{\tilde{A}}(x_i)$	0.82	0.79	0.63	0.59	0.84	0.77	0.69
$\mu_{\tilde{B}}(x_i)$	0.72	0.68	0.75	0.64	0.81	0.64	0.86

Now, the sample average membership grade of $\tilde{A}$ and $\tilde{B}$ and the sample variance of $\tilde{A}$ and $\tilde{B}$ over these two samples are $\overline{\mu}_{\tilde{A}}(x) == 0.769$; $\overline{\mu}_{\tilde{B}}(x) = 0.747$; $S^2_{\tilde{A}}(x) = 0.007169$ and $S^2_{\tilde{B}}(x) = 0.007659$.

Now, the correlation coefficient , $r(\tilde{A},\tilde{B}) = 0.5604$.

Now the NH, $H_0 : \rho(\tilde{A},\tilde{B}) = 0$ against the AH, $H_A : \rho(\tilde{A},\tilde{B}) \neq 0$.

Now, the test statistic, $t(\tilde{A},\tilde{B}) = \dfrac{r(\tilde{A},\tilde{B})\sqrt{(n-2)}}{\sqrt{1-r^2(\tilde{A},\tilde{B})}} = \dfrac{0.5604\sqrt{5}}{\sqrt{1-0.68595}} = 1.513$.

Now, the decision table about the null hypothesis and confidence limits are given below:

Test statistic value	**Degrees of freedom Freedom**	**Level of significance (Two tailed test)**	**Table value**	**Decision about H_0**
$t(\tilde{A},\tilde{B}) = 1.513$	12	5%	2.179	Accepted
		1%	3.055	Accepted

and

Confidence level	**Lower limit**	**Upper limit**
95%	-0.24669	1
99%	-0.57115	1

4. Conclusion

In this paper, we study a linear relation between two fuzzy sets defined on two crisp populations / defined on a population using crisp data with their membership grades. This study differs from the conventional statistical hypothesis testing. The proposed tests of hypotheses are helped to analyze the linear relation between two fuzzy sets on a population and the linear relation between two populations with respect to a fuzzy set. We provide the rules for decisions taken about the fuzzy hypotheses. The proposed tests of hypotheses provide appropriate decisions in an acceptable manner to decision makers when they analyze to find a linear relation of linguistic hypotheses for real life problems.

References

[1] Akbari, M.G., 2012, “Statistical hypotheses testing in the fuzzy environment”, Journal of Uncertain Systems, 6, pp.186-199.

[2] Arefi, M, and Taheri, S.M., 2011, “Testing fuzzy hypotheses using fuzzy data based on fuzzy test statistic”, Journal of Uncertain Systems, 5, pp. 45-61.

[3] Arnold, B.F., 1998, "Testing fuzzy hypotheses with crisp data", Fuzzy Sets and Systems, 94, pp.323–333.
[4] Assel S, Mohammad, 2013, "A method for constructing fuzzy test Statistics with application", Mathematical Theory and Modeling , 3, pp. 24-29.
[5] Baloui Jamkhaneh, E. and Nadi Ghara, A., 2011, "Testing statistical hypotheses for compare fuzzy test statistic", Journal of Uncertain Systems, 5 , pp.45-61.
[6] Chiang, D.A., and Lin, N.P., 1999, "Correlation of fuzzy sets", Fuzzy Sets and Systems, 102, pp. 221-226.
[7] Devore, J.L., 2008, Probability and statistics for engineers, Seventh edition, Cengage Learning, New Delhi.
[8] George Klir and Bo Yuan, 2008, Fuzzy sets and fuzzy logic, Theory and Applications, Prentice-Hall, New Jersey.
[9] Grzegorzewski, P., 2000, "Testing statistical hypotheses with vague data", Fuzzy Sets and Systems, 112, pp. 501–510.
[10] Hui-Li Hsu and Berlin Wu , 2010, " An innovative approach on fuzzy correlation coefficient with interval data", International Journal of Innovative computing, Information and Control, 6, pp.1049-1058.
[11] Kruse, R., 1984, "Statistical estimation with linguistic data", Information Sciences, 33, pp.197–207.
[12] Viertl, R., 2009, Statistical methods for fuzzy data, John Wiley and Sons, Chichester .
[13] Wu, H.C., 2009, "Statistical confidence intervals for fuzzy data", Expert Systems with Applications, 36, pp. 2670-2676.
[14] Yu-Ting Cheng and Chih-Ching Yang , 2013, "The application of fuzzy correlation coefficient with fuzzy interval data", International Journal of Innovative Management, Information and Production, 4, pp. 65-71.
[15] Zadeh, L.A., 1968, "Probability measures of fuzzy events", J. Math. Anal. Appl., 23, pp. 421-427.