Volume : 2 | Issue : 12 | December 2013 • ISSN No 2277 - 8179

Research Paper

Statistics

KEYWORDS : Testing hypotheses, Fuzzy set, Membership function, One factor ANOVA, F-test

Analysis of Variance For Crisp Data with Membership Grades of Fuzzy Sets

P. Pandian	Department of Mathematics, School of Advanced Sciences, VIT University, Vellore-632 014, INDIA.
D. Kalpanapriya	Department of Mathematics, School of Advanced Sciences, VIT University, Vellore-632 014, INDIA.

ABSTRACT

One factor ANOVA technique for crisp data of populations is extended to with membership function of the fuzzy sets over the samples of the populations. The F-test statistic for the testing statistical hypothesis is given on the membership grades of the fuzzy sets. Decision rules for each proposed hypothesis test are provided. In the proposed hypothesis tests, h-level set, cut, fuzzy interval and the degrees of optimism and pessimism are not used. Numerical examples are provided for explaining the procedures of the proposed hypotheses tests. Decision makers may use the proposed one factor ANOVA tests in the real life issues based on the membership grades of fuzzy sets for taking appropriate decisions in a simple and effective manner.

1. Introduction

Analysis of variance (ANOVA) is one of the statistical techniques which is used to test the null hypothesis that three or more populations means are equal against the alternative hypothesis that they are not equal by using their samples information. The ANOVA technique was originally used in the analysis of agricultural research data. Due to the strength and versatility of the technique, the ANOVA is now used in all most all research areas especially social science research and managerial decision making. In traditional statistical testing [8], the observations of sample are crisp. However, in the real life, the data sometimes cannot be recorded or collected precisely. The statistical hypotheses testing under fuzzy environments have been studied by many authors using by the fuzzy set theory concepts introduced by Zadeh [13]. The application by using fuzzy sets theory to statistics has been widely studied in Buckley [2] and Viertl [9]. Grzegorzewski [5] and Watanabe and Imaizumi [10] proposed the fuzzy test for testing hypotheses with vague data, and the fuzzy test gave the acceptability of the null and alternative hypotheses. The statistical hypotheses testing for fuzzy data by proposing the notions of degrees of optimism and pessimism were proposed by Wu [11]. Arefi and Taheri [1] developed an approach to test fuzzy hypotheses upon fuzzy test statistic for vague data. A new approach to the problem of testing statistical hypotheses for fuzzy data using the relationship between confidence intervals and testing hypotheses is introduced by Chachi et al.[3]. Mikihiko Konishi et al. [7] proposed a method of ANOVA for the fuzzy interval data by using the concept of fuzzy sets. Hypothesis testing of one factor ANOVA model for fuzzy data was proposed by Wu [12] which is based on the h-level set and the notions of the degrees of optimism and pessimism. Kalpanapriya and Pandian [6] proposed a new test of hypothesis of one factor ANOVA for the fuzzy data without using h-level and the notions of degrees of pessimistic and optimistic.

In this paper, we propose a new one factor ANOVA test for a crisp population based on membership function (MF) of a fuzzy set defined on the crisp population. The F-test statistic is given for the testing hypothesis with the grades of the MF of the fuzzy set for each individual observation of the random sample of the population. Two types of one way ANOVA techniques (one way ANOVA technique with respect to one fuzzy set and one way ANOVA technique with respect to many number of fuzzy sets) are discussed. The decision rules for each hypothesis test for accepting or rejecting the null and alternative hypotheses are provided. In the proposed one way ANOVA techniques, the degrees of optimism and pessimism, h-level set, $\alpha-$cut and fuzzy interval are not used. The procedures of the proposed hypotheses tests are illustrated with numerical examples. In a real life problem based on the MF of fuzzy sets, decision makers may use the proposed one factor ANOVA technique for obtaining an appropriate decision in an acceptable manner.

2. Preliminaries

We need the following concepts related to fuzzy set and its MF which can be found in Chiang and Lin [3].

Let X be a crisp set and let $A \subseteq F$ be a fuzzy set where F is a fuzzy space. The fuzzy set A defined on a crisp set X with a MF μ_A, can be expressed as follows:

$$A = \{(x, \mu_A(x) / x \in X\},$$

where $\mu_A : X \to [0,1]$.

Now, when there are two fuzzy sets *A* and *B* in F where A and B are defined on a crisp set X, then *A* and *B* can be expressed as follows.

$$A = \{(x, \mu_A(x)) / x \in X\} \quad \text{and}$$

$$B = \{(x, \mu_B(x)) / x \in X\},$$

where μ_A and $\mu_B : X \to [0,1]$.

Now, when there is a fuzzy set A $\subseteq F$ where A is defined on two crisp sets X and also , on Y, then *A* can be expressed in two ways as follows.

$$A = \{(x, \mu_A(x)) / x \in X\},$$

where $\mu_A : X \to [0,1]$ and

$$A = \{(y, \mu_A(y)) / y \in Y\},$$

where $\mu_A : Y \to [0,1]$.

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 fuzzy set A where $A = \{(x, \mu_A(x)) / x \in X\}$

Then, the sample mean of the MF of the fuzzy A defined on X denoted by $\bar{\mu}_A(x)$ is defined as

$$\bar{\mu}_A(x) = \frac{1}{n}\left(\sum_{i=1}^{n} \mu_A(x_i)\right)$$

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 fuzzy set A where . $A = \{(x, \mu_A(x)) / x \in X\}$.

Then, the sample variance of the MF of the fuzzy A defined on X denoted by $S_A^2(x)$ is defined as

$$S_A^2(x) = \frac{1}{n=1}\left(\sum_{i=1}^{n} (\mu_A(x_i) - \bar{\mu}_A(x)\ ^2\right)$$

and the sample standard deviation , $S_A = \sqrt{S_A^2}$.

3. One -factor ANOVA

In this section, we propose one way ANOVA which tests the differences among the means of populations by studying the amount of variation within each of the samples relative to the amount of variation between the samples based on the MGs of a fuzzy set over the samples. Samples under consideration in ANOVA are assumed to be drawn from normal populations of equal variances. In statistical analysis, ANOVA technique is used when the analysis involves only one factor between different subjects with more than two levels.

3.1. One way ANOVA for s classes with respect to a fuzzy set

Consider a sample of size N of a given random variable X drawn from a normal population with variance σ^2 which is subdivided into s classes according to some factor of classification. Let A be a fuzzy set defined on X with MF $\mu_A(x)$.

The main objective is to test the null hypothesis that the factor of classification has no effect on the variables with respect to the fuzzy set A (or) there is no difference between various classes with respect to fuzzy set A (or) the classes are homogeneous with respect to fuzzy set A.

Now, let $\mu_i(A)$ be the mean of i th population class of the MGs of the fuzzy set A. The testing hypotheses are given below:

Null hypothesis,

$H_0 : \mu_1(A) = \mu_2(A) = ... = \mu_s(A)$ against

Alternative hypothesis,

H_A : not all μ_i(A)'s are equal

Let x_j be the MGs of the fuzzy set A of the j th member of the i th class which contains n_i values, $\bar{x}_i(A)$ be the mean of the MGs of the fuzzy set A over the i th class and $\bar{x}(A)$ be the general mean of the MGs of the fuzzy set A over all the N ($= \sum_i n_i$) grades.

Now, the total variation with respect to A, $Q(A) = \sum_i \sum_j (x_j(A) - \bar{x}(A))^2$

$$= \sum_i \sum_j (x_j(A) - \bar{x}_i(A))^2 + \sum_i n_i(\bar{x}_i(A) - \bar{x}(A))^2$$

$$= Q_2(A) + Q_1(A)$$

where $Q_1(A) = \sum_i n_i(\bar{x}_i(A) - \bar{x}(A))^2$ is the variation between classes and

$Q_2(A) = \sum_i \sum_j (x_j(A) - \bar{x}_i(A))^2$

is the variation within classes (or) the residual variation.

Now, since i th class is a sample of size n_i from the population with the variance σ^2, we can conclude that $M_2(A) = \frac{Q_2(A)}{N-s}$ is an unbiased estimator of σ^2 with degrees of freedom

$N-s$ and since the entire group is a sample of size N from the population with variance σ^2, we can conclude that

$M_1(A) = \frac{Q_1(A)}{s-1}$ is an unbiased estimator of σ^2 with degrees of freedom $s-1$.

Now, since sampled population is normal, the estimates $M_1(A)$ and $M_2(A)$ are independent and hence the ratio $\frac{M_1(A)}{M_2(A)}$ follows a F-distribution with $(s-1, N-s)$ degrees of freedom (or) the ratio follows a F-distribution with $\frac{M_2(A)}{M_1(A)}$ $(N-s, s-1)$ degrees of freedom. Choosing the ratio which is greater than one. Then, we employ the $F-$test for testing the hypotheses.

The above results are displayed in the form a table , known as the one-factor ANOVA table as given below:

	Source of variation	
	Between Classes	Within Classes
Sum of Squares	$Q_1(A)$	$Q_2(A)$
Degrees of freedom	$s-1$	$N-s$
Mean Square	$M_1(A) = \frac{Q_1(A)}{s-1}$	$M_2(A) = \frac{Q_2(A)}{N-s}$
F-value	$F = \frac{M_1(A)}{M_2(A)}$ (or) $F = \frac{M_2(A)}{M_1(A)}$	

The decision rules in F-test to accept or reject the null hypothesis and alternative hypothesis at α level of significance are given below:

(i) If $M_2(A) < M_1(A)$ and $F = \frac{M_1(A)}{M_2(A)} < F_0$

where F_0 is the table value of F for $(s-1, N-s)$ degrees of freedom at α level, the null hypothesis is accepted. Otherwise, the alternative hypothesis is accepted.

(ii) If $M_1(A) < M_2(A)$ and $F = \frac{M_2(A)}{M_1(A)} < F_0$

where F_0 is the table value of F for $(N-s, s-1)$ degrees of freedom at α level, the null hypothesis is accepted. Otherwise, the alternative hypothesis is accepted.

Note : For easy computing the values of $Q(A)$, $Q_1(A)$ and $Q_2(A)$, we use the following formulae:

$$Q(A) = \sum_i \sum_j x_j^2(A) - \frac{T^2}{N},$$

where $T = \sum_i \sum_j x_j(A)$;

$$Q_1(A) = \sum_i \left(\frac{T_i^2}{n_i}\right) - \frac{T^2}{N}$$ where $T_i = \sum_j x_j(A)$ and

$$Q_2(A) = Q(A) - Q_1(A).$$

Example 3.1. Let X = {Students in Tamil University}. Let P = {Students in VIT University}, Q = {Students in Anna University}, R = {Students in Annamalai University} and S = {Students in Thiruvalluvar University} be four populations. Let us define the fuzzy set A over the crisp populations P, Q, R and S, A ={comfort ability}.

Now, we test that all universities are the same with respect to comfort ability at 5% level of significance.

Let $S_1 = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$

be the sample of size 8 taken from the population P,

$S_2 = (y_1, y_2, y_3, y_4, y_5, y_6, y_7)$

be the sample of size 7 taken from the population Q,

$S_3 = (u_1, u_2, u_3, u_4, u_5, u_6, u_7)$

be the sample of size 7 taken from the population R and $S_4 = (v_1, v_2, v_3, v_4, v_5, v_6)$ be the sample of size 6 taken from the population S, then we collect information and obtain the MGs of these two samples concerning fuzzy set A as given below:

$\mu_P(A)$	$\mu_Q(A)$	$\mu_R(A)$	$\mu_S(A)$
0.91	0.74	0.71	0.75
0.85	0.76	0.64	0.79
0.82	0.62	0.58	0.84
0.79	0.79	0.63	0.63
0.89	0.86	0.72	0.57
0.76	0.65	0.80	0.48
0.81	0.58	0.69	-
0.78	-	-	-

Null hypothesis :

$$H_0 : \bar{\mu}_A(S_1) = \bar{\mu}_A(S_2) = \bar{\mu}_A(S_3) = \bar{\mu}_A(S_4)$$

Alternative hypothesis :

$$H_a : \bar{\mu}_A(S_1) \neq \bar{\mu}_A(S_2) \neq \bar{\mu}_A(S_3) \neq \bar{\mu}_A(S_4)$$

Now,

$$Q(A) = \sum_i \sum_j x_{ij}^2(A) - \frac{T^2}{N} = 17.054 - 14.921$$

$$= 0.2131$$

where $T = \sum_i \sum_j x_{ij}(A) = 20.44$;

$$Q_1(A) = \sum_i \left(\frac{T_i^2}{n_i} \right) - \frac{T^2}{N} = 15.031 - 14.92$$

$$= 0.111$$

where $T_i = \sum_j x_{ij}(A)$ and

$$Q_2(A) = Q(A) - Q_1(A) = 2.132 - 0.111 = 2.021$$

Now, the one-factor ANOVA table is given below:

	Source of Variation	
	Between Samples	Within Samples
Sum of Squares	0.111	2.021
Degrees of Freedom	3	24
Mean square	0.037	0.084
F-value	F = 2.27	

Now, $F < T$, where T (= 8.64) is the table value of F at 5% level of significance with (24,3) degrees of freedom. Therefore, the null hypothesis H_0 is accepted. The students in all the four universities felt comfortable.

3.2. One way ANOVA for one class with respect to many fuzzy sets

Consider a sample of size N of a given random variable X drawn from a normal population with variance σ^2. Let S be a sample of X and let A_i be a fuzzy set defined on X with MF $\mu_{A_i}(x)$, i=1,2,...,s.

The main objective is to test the null hypothesis that the factor of classification has no effect on the variables with respect to the fuzzy sets (or) there is no difference between various classes with respect to fuzzy sets (or) the classes are homogeneous with respect to fuzzy sets.

Now , let $\mu(A_i)$ be the mean of the MGs of the fuzzy set A_i. The testing hypotheses are given below:

Null hypothesis,

$H_0 : \mu(A_1) = \mu(A_2) = ... = \mu(A_s)$ against

the alternative hypothesis,

H_A : not all $\mu(A_i)$'s are equal

Let $x_j(S)$ be the MGs of the fuzzy set A_i over the j th member of the sample S which contains n_i values, $\bar{x}_i(S)$ be the mean of the MGs of the fuzzy set A_i over the sample S and $\bar{x}(S)$ be the general mean of the MGs of all the fuzzy sets over the sample S.

Let $N = \sum_i n_i$.

Now, the total variation over the sample S, $Q(S) = \sum_i \sum_j (x_j(S) - \bar{x}(S))^2$

$$= \sum_i \sum_j (x_j(S) - \bar{x}_i(S))^2 + \sum_i n_i(\bar{x}_i(S) - \bar{x}(S))^2$$

$$= Q_2(S) + Q_1(S)$$

where $Q_1(S) = \sum_i n_i(\bar{x}_i(S) - \bar{x}(S))^2$ is the variation between classes and

$$Q_2(S) = \sum_i \sum_j (x_j(S) - \bar{x}_i(S))^2$$

is the variation within classes (or) the residual variation.

tNow, since i th class is a sample of size n_i from the population with the variance σ^2, we can conclude that

$$M_2(S) = \frac{Q_2(S)}{N-s}$$

is an unbiased estimator of σ^2 with degrees of freedom $N-s$ and since the entire group is a sample of size N from the population with variance σ^2, we can conclude that

$$M_1(S) = \frac{Q_1(S)}{s-1}$$

is an unbiased estimator of σ^2 with degrees of freedom $s-1$.

Now, since sampled population is normal, the estimates $M_1(S)$ and $M_2(S)$ are independent and hence the ratio $\frac{M_1(S)}{M_2(S)}$ follows a F-distribution with $(s-1, N-s)$ degrees of freedom (or) the ratio $\frac{M_2(S)}{M_1(S)}$ follows a F-distribution with $(N-s, s-1)$ degrees of freedom. Choosing the ratio which is greater than one. Then, we employ the $F-$test for testing the hypotheses.

The above results are displayed in the form a table , known as the one-factor ANOVA table which is given below:

	Source of variation	
	Between Classes	Within Classes
Sum of Squares	$Q_1(S)$	$Q_2(S)$
Degrees of freedom	S-1	N-2
Mean Square	$M_1(S) = \frac{Q_1(S)}{s-1}$	$M_2(S) = \frac{Q_2(S)}{N-s}$

F-value	$F = \frac{M_1(S)}{M_2(S)}$ (or) $F = \frac{M_2(S)}{M_1(S)}$

The decision rules in F-test to accept or reject the null hypothesis and alternative hypothesis at α level of significance are given below:

(i) If $M_2(S) < M_1(S)$ and $F = \frac{M_1(S)}{M_2(S)} < F_0$ where F_0 is the table value of F for $(s-1, N-s)$ degrees of freedom at α level, the null hypothesis is accepted, Otherwise, alternative hypothesis is accepted.

(ii) If $M_2(S) > M_1(S)$ and $F = \frac{M_2(S)}{M_1(S)} < F_0$ where F_0 is the table value of F for $(N-s, s-1)$ degrees of freedom at α level, the null hypothesis is accepted, Otherwise, alternative hypothesis is accepted.

Note : For easy computing the values of $Q(S)$, $Q_1(S)$ and $Q_2(S)$, we use the following formulae:

$$Q(S) = \sum_i \sum_j x_{ij}^2(S) - \frac{T^2}{N}$$

where $T = \sum_i \sum_j x_{ij}(S)$;

$$Q_1(S) = \sum_i \left(\frac{T_i^2}{n_i} \right) - \frac{T^2}{N}$$

where $T_i = \sum_j x_{ij}(S)$

and $Q_2(S) = Q(S) - Q_1(S)$.

Example 3.2: Let X = { Doctors in a city} be the population . Let us define the fuzzy sets over the crisp set, A = { Compassionate}, B = { Contentment}, C = { Dedication} and

D ={ Forthright}.

Now, we test the hypothesis that the doctors in the city do not differ with these four attributes at 5% level of significance.

Let $S = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})$ be the sample of size 10 taken randomly. We collect information and obtain the MGs of the sample concerning fuzzy sets A, B, C and D as given below:

Doctors	μ_A	μ_B	μ_C	μ_D
I	0.91	0.74	0.72	0.69
II	0.85	0.76	0.78	0.76
III	0.82	0.62	0.65	0.79
IV	0.79	0.79	0.74	0.72
V	0.89	0.72	0.84	0.89
VI	0.76	0.82	0.78	0.91
VII	0.81	0.58	0.64	0.73
VIII	0.78	0.72	0.79	0.85
IX	0.86	0.67	0.69	0.78
X	1.0	0.45	0.82	0.97

Now, the null hypothesis

$$H_0 : \bar{\mu}_A(S) = \bar{\mu}_B(S) = \bar{\mu}_C(S) = \bar{\mu}_D(S)$$

and the alternative hypothesis

$$H_A : \bar{\mu}_A(S) \neq \bar{\mu}_B(S) \neq \bar{\mu}_C(S) \neq \bar{\mu}_D(S)$$

Now, $Q(S) = \sum_i \sum_j x_{ij}^2(S) - \frac{T^2}{N}$

$= 24.36 - 23.84 = 0.52$

where $T = \sum_i \sum_j x_{ij}(S) = 30.88$;

$Q_1(S) = \sum_i \left(\frac{T_i^2}{n_i} \right) - \frac{T^2}{N} = 24.076 - 23.84$

$= 0.24$

where $T_i = \sum_j x_{ij}(S)$ and

$Q_2(S) = Q(S) - Q_1(S) = 0.52 - 0.24 = 0.28$.

Now, the one factor ANOVA table is given below:

	Source of variation	
	Between Classes	Within Classes
Sum of Squares	0.24	0.28
Degrees of freedom	3	36
Mean Square	0.08	0.0078
F-value	F =10.26	

Now, since $F > T$, where T (= 2.84) is the table value of F at 5% level of significance with (3,36) degrees of freedom. Therefore, the null hypothesis H_0 is rejected and H_A is accepted , that is, the doctors in the city differ with these four attributes.

4. Conclusion

In this paper, we propose two types of one way ANOVA techniques for crisp populations based on the MF of fuzzy sets to test the statistical hypotheses which is completely exceptional from the conventional statistics. In the proposed hypotheses tests, the differences of significance between classes with respect to fuzzy set / fuzzy sets are studied. For taking decision on the hypotheses, we provide the decision rules for accepting or rejecting the null and alternative hypotheses. We can easily observe that this type of test of hypothesis is a characteristic or attribute based test on the population. In the proposed tests, fuzzy interval, the degrees of optimism and pessimism, h-level set and $\alpha -$ cut are not used. In the near future, the proposed tests of hypotheses are extended to tests of hypotheses for two factors ANOVA and random design of experiments. The proposed tests of hypotheses tests can help decision makers for handling real life problems to test statistical hypotheses based on one or more attributes and providing an appropriate decision in a satisfaction manner.

REFERENCE

[1] 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. | | [2] J.J. Buckley, Fuzzy Statistics, Springer | Verlag, New York, 2005. | | [3] 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 | | [4] D-A. Chiang and N.P. Lin , Correlation of fuzzy sets, Fuzzy Sets and Systems, 2(1999), 221-226. | | [5] P. Grzegorzewski, Testing statistical hypotheses with vague data, Fuzzy Set. Syst., | 112(2000), 501–510. | | [6] D. Kalpanapriya and P. Pandian , Fuzzy Hypothesis Testing Of Anova Model with Fuzzy Data, International Journal of Modern Engineering Research (IJMER), 2 (2012), 2951-2956. | | [7] Mikihiko Konishi, Tetsuji Okuda and Kiyoji Asai, Analysis of variance based on fuzzy interval data using moment correction method, International Journal of Innovative | Computing, Information and Control, | 2(2006), 83-99. | | [8] Paul Newbold, William L.Carlson and Betty Thorne, Statistics for Business and Economics, Pearson Education, New Delhi, 2007. | | [9] R. Viertl, Statistical methods for fuzzy data, John Wiley and Sons, Chichester, 2011. | | [10] N. Watanabe and T. Imaizumi, A fuzzy statistical test of fuzzy hypotheses, Fuzzy Sets and Systems, 53 (1993) , 167–178. | | [11] H.C. Wu, Statistical hypotheses testing for fuzzy data, Information Sciences,175 (2005), 30–56. | | [12] H .C. Wu, Analysis of variance for fuzzy data, International Journal of Systems Science, 38(2007), 235–246. | | [13] L.A. Zadeh, Fuzzy sets, Inform. Contr., 8 (1965), 338–353, 1965. |