International Journal of Fuzzy Logic and Intelligent Systems, vol. 4, no. 1, December 2004 pp. 97-100

Multivariate Analysis of Variance for Fuzzy Data

Man-Ki Kang* · Sung-Il Han**

*** Department of Information Statistics, Dong Eui University, Busan, 616-714, KOREA**

**** School of Engineering, Osaka Electrical and Communication College, JAPAN**

Abstract

We propose some properties of fuzzy multivariate analysis of variance by fuzzy vector operation with agreement index. We deals fuzzy null hypotheses and fuzzy alternative hypothesis and define the agreement index for the grades of the judgements that the hypothesis is rejection or acceptance. Finally, we provide an example to evaluate the judgements.

Key Words : fuzzy hypotheses testing, multivariate analysis of variance, agreement index, fuzzy null hypotheses, fuzzy alternative hypothesis, Wilks lambda.

1. Introduction

Our primary propose of statistical test is fuzzy multivariate analysis of variance.

The generalization for simple hypotheses is given by Watanabe and Imaizumi([8]). In Gizegorzewski([2]), the fuzzy hypotheses testing was considered that the data(observation) are vague data and the hypotheses are fuzzy. Also, Kang, Choi and Han([3],[4],[5]) was suggested some estimations of fuzzy variance components for fixed effect modeled with fuzzy data.

We propose some properties of fuzzy multivariate analysis of variance(MANOVA) by fuzzy vector operation with agreement index for the fuzzy hypotheses and provide an example to evaluate the judgements with fuzzy data.

2. Some distance of fuzzy number

We denote by fuzzy number in ε_N^p

$$A=(a_1,\ a_2,\ \cdots,\ a_p) \tag{2.1}$$

where $a_i(i=1,\cdots,p)$ are projection of A to axis $X_i(i=1,\cdots,p)$, fuzzy number in R, respectively.

Definition 2.1. The $\delta-$level set of fuzzy number in ε_N^p is define by

$$[A]^\delta=\left\{(x_1,\cdots,x_p)\in R^p: (x_1,\cdots,x_p)\in \prod_{i=1}^p[a_i]^\delta\right\} \tag{2.2}$$

where notation $\prod$ is the Cartesian product of sets.

Definition 2.2. Let A and B in ε_N^p, for all $\delta\in(0,1]$,

$$A=B \Leftrightarrow [A]^\delta=[B]^\delta \tag{2.3}$$

$$[A*_pB]^\delta=\prod_{i=0}^p[a_i*b_i]^\delta \tag{2.4}$$

where $*_p$ is operation in ε_N^p and $*$ is operation in ε_N.

Let $\prod_{i=0}^p[a_i]^\delta,\ 0<\delta\le 1$, be a given family of nonempty areas.

If

$$\prod_{i=0}^p[a_i]^{\delta_2}\subset \prod_{i=0}^p[a_i]^{\delta_1} \text{ for } 0<\delta_1<\delta_2<1 \tag{2.5}$$

and

$$\prod_{i=0}^p \lim_{k\to\infty}[a_i]^{\delta_k}=\prod_{i=0}^p[a_i]^\delta \tag{2.6}$$

then the family

$\prod_{i=0}^p[a_i]^\delta,\ 0<\delta\le 1$ represents the $\delta-$level sets of a fuzzy number $A\in\varepsilon_N^p$, where (δ_k) is a nondecreasing sequence converging to $\delta\in(0,1]$.

Conversely, if $\prod_{i=0}^p[a_i]^\delta,\ 0<\delta\le 1$ are the $\delta-$level sets of a fuzzy number in R^n, then conditions (2.5) and (2.6) are true.

We define the metric d_∞ on ε_N^p.

Definition 2.3. Let $A,\ B\in\varepsilon_N^p$,

$$\begin{aligned} d_\infty(A,B) &= sup\{d_H([A]^\delta,[B]^\delta): \delta\in(0,1]\} \\ &= sup\left\{d_H\left(\prod_{i=1}^p[a_i]^\delta, \prod_{i=1}^p[b_i]^\delta\right): \delta\in(0,1]\right\} \\ &= sup\left\{\sqrt{\sum_{i=1}^p d_H([a_i]^\delta,[b_i]^\delta))^2}: \delta\in(0,1]\right\} \end{aligned} \tag{2.7}$$

where d_H is Hausdorff distance.

Manuscrip received Mar. 16, 2004; revised May. 21, 2004.
The work was supported by the Dong Eui Research Foundation Grant in 2004.

3. Fuzzy statistical test

Let X be a random sample fuzzy vector from some sample space Ω and $\{P_\theta, \theta \in \Theta\}$ be a family of fuzzy probability distributions, where θ is a parameter vector and Θ is a parameter space. For each $\psi \in \Theta$, we can consider a family of hypotheses $\{(H_0(\psi), H_1(\psi)) | \psi \in \Theta\}$ and introduce the fuzzy hypothesis as a fuzzy subset.

Definition 3.1. The fuzzy hypothesis H_f is a fuzzy subset of $\{(H_0(\psi), H_1(\psi)) | \psi \in \Theta\}$ with fuzzy hypothesis membership function $\chi_{H_f}(H_0(\psi), H_1(\psi))$.

We set whit simplicity

$$\chi_{H_f}(\psi) \equiv \chi_{H_f}(H_0(\psi), H_1(\psi)) \tag{3.1}$$

and assume normality and convexity.

The fuzzy null hypothesis can be defined as follows.

Definition 3.2. The fuzzy null hypothesis $H_{f.0}$ is a fuzzy subset of Θ with a membership function $\chi_H(\psi)$. The fuzzy alternative hypothesis $H_{f.1}$ is a fuzzy subset of Θ defined by the equation

$$H_{f.1} = H_{f.0}{}^c \cap [\{\bigcup_{\delta \in [0,1]} \delta(\bigcup_{\{\psi \chi_H(\psi) \geq 0\}} \Theta_{K,\psi})\}] \tag{3.2}$$

where $\delta(\cdot)$ stands for the fuzzy set whose membership function of the set.

Definition 3.3. If we consider upper bound function T then we present this function by a fuzzy subset $T \subset R$. Let us consider number a fuzzy number $K \subset R$, which we call the agreement index of K whit regard to T, the ratio being defined in the following way:

$$I(K, T) = (area\, K \cap T)/(area\ T) \in [0.1]. \tag{3.3}$$

Using membership function $K(\alpha, \psi)$ of critical region where α is significance level, we also define the fuzzy hypothesis membership function χ_{R_α} on $\{0, 1\}$ for acceptance or rejection as follows.

Definition 3.4. If we have symmetric membership function χ_T with fuzzy number statistics $T = \langle \min(l, r), c, \min(l, r) \rangle$, then we define the real-valued function R_α on Θ as in Definition 3.1. The maximum grade membership function of acceptance or rejection is

$$\begin{aligned} \chi_{R_\alpha}(0) &= \sup_\psi \{\text{area}(\chi_T(\psi) \cap \chi_{K(\alpha,\psi)^c}) / \text{area}\, \chi_T(\psi)\} \\ \chi_{R_\alpha}(1) &= 1 - \chi_{R_\alpha}(0) \end{aligned} \tag{3.4}$$

Let R_α denote the fuzzy subset of an entire set $\{0, 1\}$ defined by χ_{R_α}, since $\{0, 1\}$ corresponds {"accept", "reject"}, the value $\chi_{R_\alpha}(1)$ and $\chi_{R_\alpha}(0)$ are equal to the grades of the judgements that the hypothesis is rejected or not rejected.

Now, we show the multivariate statistical properties of our testing method.

4. Fuzzy MANOVA

More then two populations need to be compared. The random samples, collected from each of a populations are arranged as

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_a \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n_1} \\ X_{21} & X_{22} & \cdots & X_{2n_2} \\ & \cdots\cdots\cdots & & \\ X_{a1} & X_{a2} & \cdots & X_{an_a} \end{pmatrix} \tag{4.1}$$

where $X_{ij} = (\widetilde{x_{ij1}}, \widetilde{x_{ij1}}, \cdots, \widetilde{x_{ijp}})'$ $i = 1, 2, \cdots, a,$ $j = 1, 2, \cdots, n_i$ is fuzzy random vector.

We denote $X_{ij} = (\widetilde{x_{ij1}}, \widetilde{x_{ij1}}, \cdots, \widetilde{x_{ijp}})'$ by the fuzzy number with $A = (a_1, a_2, \cdots, a_p)$ in ε_N^p as (2.1).

Fuzzy multivariete analysis of variance(MANOVA) model for comparing a population mean vectors is given by

$$X_{ij} = \mu \oplus \alpha_i \oplus \varepsilon_{ij}, \quad i = 1, 2, \cdots, a \tag{4.2}$$

or

$$X_{ij} = \begin{pmatrix} \widetilde{x_{ij1}} \\ \widetilde{x_{ij2}} \\ \cdots \\ \widetilde{x_{ijp}} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdots \\ \mu_p \end{pmatrix} \oplus \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \cdots \\ \alpha_{ip} \end{pmatrix} \oplus \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \cdots \\ \varepsilon_{ijp} \end{pmatrix}$$

and $j = 1, 2, \cdots, n_i$, where ε_{ij} are independent $N_p(\tilde{0}, \Sigma)$ variables.

Here the parameter vector μ is an overall fuzzy mean(level) and α_i represents the i-th treatment fuzzy effect with $\sum_{i=1}^{a} n_i \alpha_i = \tilde{0}$.

The errors for the components of X_{ij} are correlated, but the fuzzy variance-covariance matrix Σ is the same for the populations.

A vector of observations may be decomposed as suggested by the model.

Thus

$$X_{ij} = \overline{X} \oplus (X_i \ominus \overline{X}) \oplus (X_{ij} \ominus \overline{X}_i) \tag{4.3}$$

where X_{ij} is fuzzy observation vector, $\overline{X}$ is overall sample fuzzy mean v.s. $\widehat{\mu}$, $(\overline{X}_i \ominus \overline{X})$ is estimated treatment fuzzy effect v.s. $\widehat{\alpha}_i$, $(X_{ij} \ominus X_i)$ is fuzzy residual v.s. $\widehat{\varepsilon}_{ij}$.

The decomposition in (4.3) leads to the multivariate analog of the univariate sum of squares. First we note that the cross-product can be written as

$$\begin{aligned} &(X_{ij} \ominus \overline{X})(X_{ij} \ominus \overline{X})' \\ &= [(X_{ij} \ominus \overline{X}_i) \oplus (X_i \ominus \overline{X})][(X_{ij} \ominus \overline{X}_i) \oplus (X_i \ominus \overline{X})]' \end{aligned} \tag{4.4}$$

The sum of over j of the middle two expressions is the fuzzy zero matrix become $\sum_{j=1}^{n_i}(X_{ij}\ominus\overline{X})=\tilde{0}$. Next, summing the cross-product over i and j yields

$$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(X_{ij}\ominus\overline{X})(X_{ij}\ominus\overline{X})' \quad ;T$$
$$=\sum_{i=1}^{a}n_i(X_i\ominus\overline{X})(X_i\ominus\overline{X})' \quad ;B \qquad (4.5)$$
$$\oplus\sum_{i=1}^{a}\sum_{j=1}^{n_i}(X_{ij}\ominus\overline{X}_i)(X_{ij}\ominus\overline{X}_i)' \quad ;W$$

it's means that (Total sum of squares and cross product) = (trement(Between) sum of squares and cross product) + (residual(Within) sum of squares and cross product).

The hypothesis of no treatment effects

$$H_{f.0}: \alpha_1\simeq\alpha_2\simeq,\cdots,\simeq\alpha_a\simeq\tilde{0} \qquad (4.6)$$

is tested by considering the relative size of treatment and residual fuzzy sums of squares and cross-product.

The degree of freedom correspond involving Wishart densities.

One test of $H_{f.0}: \alpha_1\simeq\alpha_2\simeq,\cdots,\simeq\alpha_a\simeq\tilde{0}$ involves generalized variances. We reject $H_{f.0}$ if the ratio of generalized variance

$$\Lambda^*=\frac{|W|}{|B+W|} \qquad (4.7)$$

is too small. The quantity Λ^* proposed originally by Wilks. The exact distribution of Λ^* can be derived for the special cases of $p\geq 1$, $a=2$ with

$$\left(\frac{\sum n_i-p-1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right)\sim F(p,\sum n_i-p-1) \qquad (4.8)$$

Bartlett has shown that if H_0 is true and $\sum n_i=n$ is large,

$$-\left(n-1-\frac{(p+a)}{2}\right)\ln\Lambda^* \qquad (4.9)$$

has approximately a chi-square distribution with $p(a-1)$ degree of freedom. Consequently, for n large, we reject H_0 at significance level α if

$$-\left(n-1-\frac{(p+a)}{2}\right)\ln\Lambda^*>\chi^2(\alpha;p(a-1)). \qquad (4.10)$$

5. An example

Suppose an additional variable is observed. Arranging the observation pairs X_{ij} in rows, the random fuzzy number data are

$$X=\begin{pmatrix} \begin{bmatrix}(8.9,9.1)\\(2.9,3.1)\end{bmatrix} & \begin{bmatrix}(5.9,6.1)\\(1.9,2.1)\end{bmatrix} & \begin{bmatrix}(8.9,9.1)\\(6.9,7.1)\end{bmatrix} \\ \begin{bmatrix}(-0.1,0.1)\\(3.9,4.1)\end{bmatrix} & \begin{bmatrix}(1.9,2.1)\\(-0.1,0.1)\end{bmatrix} & \\ \begin{bmatrix}(2.9,3.1)\\(7.9,8.1)\end{bmatrix} & \begin{bmatrix}(0.9,1.1)\\(8.9,9.1)\end{bmatrix} & \begin{bmatrix}(1.9,2.1)\\(6.9,7.1)\end{bmatrix}\end{pmatrix}$$

Thus we have

$$B=\begin{pmatrix}(71.17,84.44) & (-17.81,-6.43)\\(-17.81,-6.43) & (43.97,52.37)\end{pmatrix},$$
$$W=\begin{pmatrix}(7.04,13.52) & (-3.08,4.2)\\(-3.08,4.2) & (19.44,29.12)\end{pmatrix}$$

and

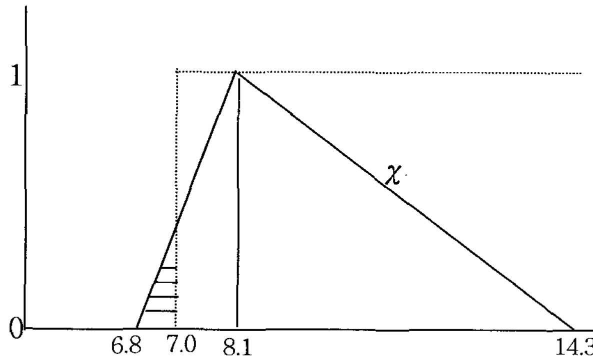
$$\Lambda^*=\frac{|W|}{|B+W|}=(0.0149,0.0509).$$

Since $p=2$ and $a=2$, an fuzzy test of $H_{f.0}: \alpha_1\simeq\alpha_2\simeq\tilde{0}$ versus $H_{f.1}$: at least one $\alpha_i\neq\tilde{0}$ is available. To carry out the test, we compare the fuzzy statistic

$$\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right)\frac{\Sigma n_i-a-1}{a-1}=(6.85,14.36)$$

with a percentage point of an $F-$distribution $F(0.01;4,8)=7.01$.

Finally, we have the rejection degree $\chi_{R_\alpha}(1)=1-0.98$ for the hypothesis by agreement index in Fig 1. For $\delta=1$, the F-test statistic is 8.19.

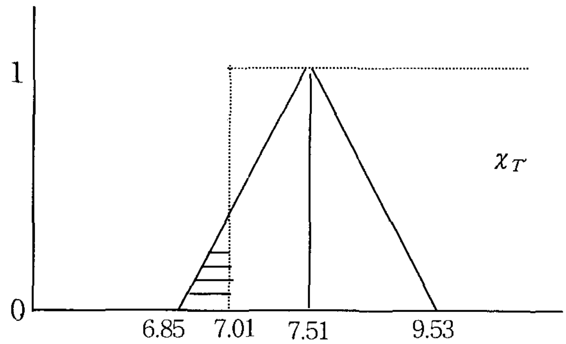


[Fig. 1]

Other hand, we consider symmetric F-test statistics LR-fuzzy number for $\alpha=0$ by Definition 4.3 as

$$T=\langle\min(8.19-6.85,14.36-8.19),8.19,\min(8.19-6.85,14.36-8.19)\rangle=\langle 1.34,8.19,1.34\rangle$$

then we have the rejection degree $\chi_{R_\alpha}(1)=1-0.936$ by Fig 2.



[Fig. 2]

References

[1] S. Fruhwirth-Schnatter, On statistical inference for fuzzy data with application to descriptive statistics, *Fuzzy Sets and Systems*, 50(1992), 143-165.
[2] P. X. Gizegorzewski, Testing Hypotheses with vague data, *Fuzzy Sets and Systems*. 112 (2000), pp.501-510.
[3] M. K. Kang and G. T. Choi , Estimation Variance Components for Fuzzy Data, *Proceeding of Korea Fuzzy Logic and Intelligent System Society Fall Conference*, Vol. 12, Num. 1, (2002).
[4] M. K. Kang, G. T. Choi and C. E. Lee, On Statistical for Fuzzy Hypotheses with Fuzzy Data, *Proceeding of Korea Fuzzy Logic and Intelligent System Society Fall Conference*, Vol. 10, Num. 2, (2000).
[5] M. K. Kang, C. E. Lee and S. I. Han, Fuzzy Hypotheses Testing for Hybrid numbers by Agreement Index, *Far East J. Theo. Stat.* 10(1) (2003), 1-9.
[6] A. Kaufmann, M. M. Gupta, *Introduction to Fuzzy Arithmetic*, Van Nostrand Reinhold, New York, 1991.
[7] D. Pill and K. Peter, *Metric Space of Fuzzy Sets*, World Scientific, Singapore, 1994.
[8] N. Watanabe, T. Imaizumi, A Fuzzy Statistical Test of Fuzzy Hypotheses, *Fuzzy Sets and Systems*, 53 (1993), pp.167-178.

Man Ki Kang
Man Ki Kang received the M.S. and Ph.D. in department of Mathematics form the University of Dong A in 1981 and 1990, respectively.
He is currently a professor in the college of natural sciences, Dongeui University.

His recent research interests include statistical fuzzy data processing and fuzzy pattern recognition.
Phone : 051-890-1482
E-mail : mkkang@deu.ac.kr

Sung Il Han
Sung Il Han received the M.S. and Ph.D. in school of Information Mathematics form the University of Osaka in 1996 and 2000, respectively.
He is currently a worker in the of School of Engineering, Osaka Electrical and Communication College .
His recent research interest statistical fuzzy data processing.
E-mail : hanstat@ybb.ne.jp