Proceedings of KFIS Spring Conference 2004 Volume 14, Number 1

벡터 연산에 의한 퍼지 다변량 분산분석

Fuzzy Multivariate Analysis of Variance by Fuzzy Vector Operation

강만기 * , 최규탁 **, 한성일 ***

* 동의대학교 자연과학대학 정보통계학과 교수

(mkkang@deu.ac.kr 051-890-1482)

**경남정보대학 산업시스템경영학과 교수

***오사카전기통신대학 공학부 강의교수

Abstract

We propose some properties of fuzzy multonomial analysis of variance by fuzzy vector operation with agreement index.

Keywords : fuzzy hypotheses testing, multonomial analysis of variance, agreement index, fuzzy vector, agreement index.

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 modeld with fuzzy data.

We propose some porperties of fuzzy multivariate analysis of variance by fuzzy vector operation with agreement index.

2. Some distance of fuzzy number

We denote by fuzzy number in ε_N^n

$$A=(a_1, a_2, \cdots, a_n) \qquad (2.1)$$

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

Definition 2.1. The γ-level set of fuzzy number in ε_N^n is define by

$$[A]^{\gamma}=\{(x_1,\cdots,x_n)\in R^n: (x_1,\cdots,x_n)\in\prod_{i=1}^{n}[a_i]^{\gamma}\} \quad (2.2)$$

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

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

$$A=B \Leftrightarrow [A]^{\gamma}=[B]^{\gamma} \quad (2.3)$$

$$[A*_nB]^{\gamma}=\prod_{i=0}^{n}[a_i*b_i]^{\gamma} \quad (2.4)$$

where $*_n$ is operation in ε_N^n and $*$ is operation in ε_N.

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

If

$$\prod_{i=0}^{n}[a_i]^{\gamma_2}\subset\prod_{i=0}^{n}[a_i]^{\gamma_1} \text{ for } 0<\gamma_1<\gamma_2<1 \quad (2.5)$$

and

$$\prod_{i=0}^{n}\lim_{k\to\infty}[a_i]^{\gamma_k}=\prod_{i=0}^{n}[a_i]^{\gamma} \quad (2.6)$$

then the family

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

Conversely, if $\prod_{i=0}^{n}[a_i]^{\gamma},\ 0<\gamma\le 1$ are the γ-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^n.

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

$$d_\infty(A,B)=sup\{d_H([A]^{\gamma},[B]^{\gamma}):\gamma\in(0,1]\}$$

$$=sup\left\{d_H(\prod_{i=1}^{n}[a_i^{\gamma}],\prod_{i=1}^{n}[b_i]^{\gamma}):\gamma\in(0,1]\right\}$$

$$=sup\left\{\sqrt{\sum_{i=1}^{n}d_H([a_i^{\gamma}],[b_i]^{\gamma}))^2}:\gamma\in(0,1]\right\} \quad (4.7)$$

where d_H is Hausdorff distance.

3. A fuzzy statistical test

Let $\tilde{x}$ be a random sample from space Ω and $\{P_\theta,\theta\in\Theta\}$ be a family of fuzzy probability, where θ is a parameter and Θ is a parameter space. For each $\psi\in\Theta$, we 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

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

and assume normality an 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

$$H_{f.1}=H_{f.0}{}^c\cap[\{\bigcup_{\gamma\in(0,1]}\gamma(\bigcup_{\{\psi|\chi_H(\psi)\ge 0\}}\Theta_{K,\psi})\}] \quad (3.2)$$

where $\gamma(M)$ stands for the fuzzy set whose membership function of a set M.

The first term on the right hand characteristic function of a set M.

Definition 3.3. Let us consider a fuzzy number $M\subset R$, which we call the agreement index of M whit regard to H, the ratio being defined in the following way:

$$I(M,H) = (area\, M \cap H)/(area\, M) \in [0.1]. \quad (3.3)$$

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

Definition 3.4. 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

$$\chi_{R_\alpha}(0) = \sup_\phi \{\text{area}(\chi_{H_1}(\phi) \cap \overline{K(\alpha,\phi)})/\text{area}\,\chi_{H_1}(\phi)\}$$

$$\chi_{R_\alpha}(1) = 1 - \chi_{R_\alpha}(0) \quad (4.4)$$

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

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

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \ i=1,2,\cdots,a \quad (4.1)$$

and $j=1,2,\cdots,n_i$,

where $X_{ij} = (\widetilde{x_{ij1}}, \widetilde{x_{ij1}}, \cdots, \widetilde{x_{ijp}})'$ is fuzzy vector and 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 ith treatment fuzzy effect with $\sum_{i=1}^{n} n_i \alpha_i = \tilde{0}$.

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

$$X_{ij} = \overline{X} \oplus (X_i \ominus \overline{X}) \oplus (X_{ij} \ominus \overline{X_i}), \quad (4.2)$$

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

$$(X_{ij} \ominus \overline{X})(X_{ij} \ominus \overline{X})'$$

$$= [(X_{ij} \ominus \overline{X_i}) \oplus (X_i \ominus \overline{X})][(X_{ij} \ominus \overline{X_i}) \oplus (X_i \ominus \overline{X})]' \quad (4.3)$$

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

$$\sum_{i=1}^{a} \sum_{j=1}^{n_i} (X_{ij} \ominus \overline{X})(X_{ij} \ominus \overline{X})' \quad ;T$$

$$= \sum_{i=1}^{a} n_i (X_i \ominus \overline{X})(X_i \ominus \overline{X})' \quad ;B$$

$$\oplus \sum_{i=1}^{a} \sum_{j=1}^{n_i} (X_{ij} \ominus \overline{X_i})(X_{ij} \ominus \overline{X_i})' \; ;W \quad (4.4)$$

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

$$H_{f.0}: \alpha_1 \simeq \alpha_2 \simeq, \cdots, \simeq \alpha_a \simeq \tilde{0} \quad (4.5)$$

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

$$\Lambda^* = \frac{|W|}{|B+W|} \qquad (4.6)$$

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

$$\left(\frac{\sum n_i - p - 1}{p}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F(p, \sum n_i - p - 1) . (4.7)$$

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

$$-\left(n-1-\frac{(p+a)}{2}\right)\ln \Lambda^* \qquad (4.8)$$

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

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

5. Example

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

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

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

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

with a percentage point of an F-distribution $F(0.01; 4, 8) = 7.01$. Finally, we have the reject degree $\chi_{R_a}(1) = 1 - 0.98$ for the hypothesis by agreement index.

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 , Estimarion 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, *Far East J. Theo. Stat,* 53 (1993), pp.167-178.