퍼지자료에 대한 분산성분 추정

Estimation variance components for fuzzy data

강만기 *, 최규탁 **
* 동의대학교 전산통계학과
** 경남정보대학, 산업시스템경영학과

Man-Ki Kang *, Gyu-Tag Choi
* Dept. of Computer Science and Statistics, Dongeui University
E-mail : mkkang@dongeui.ac.kr

Abstract

The observation fuzzy data in random effect and balanced designs for one way classification by using a the matrix formulation , we can estimate the fuzzy variance components for the ozone depletion example and test by the agreement index.

keyword : variance component, random effect, treatment effect, agreement index.

1. Introduction

In describing the random model for analysis of variance by fuzzy data, the situation envisaged is that having data grouped by class, those classes being considered a random sample from some population of class([10]). Thus the treatment effect if fuzzy random variable.

The theory of fuzzy random variable was development in recent years([2],[8],[9] and their references). The variance and covariance of fuzzy random variable if great importance in statistical analysis. The concept of the variance and covariance of fuzzy random variable and their properties are defined by Yuhn Feng etc([3],[6]). By using the concept of a fuzzy variance and covariance bilinear operator and statistical inference for fuzzy data to descriptive statistics([11]), we can estimate the fuzzy variance components.
For ozone depletion example, we can also test by agreement index([1],[4],[7],[7]).

In chapter 2 . we describe *statistical* inference for fuzzy data with constructing function from the α−cut representation for sum of squire. In chapter 3. we seek the variance components for given matrix formulation. We will show an example for ozone data and seek the fuzzy variance components, also test he hypotheses by agreement index.

2. Sum of squire for fuzzy vector valued data

The fuzzy vector will be discussed in a mathematical framework and will be applied to model of fuzzy data and fuzzy sample as fuzzy number.

Definition 1. A piecewise continuous function $\Psi_{a^*}: R^n \rightarrow [0,1]$ mapping R^n to the interval $[0,1]$ defines a fuzzy vector a^* in R^n if the family $C(a^*)_{\alpha \in (0,1]}$ derived from $\Psi_{a^*}(a)$ by

$$C(a^*)_\alpha = \{ a \in R^n : \Psi_{a^*}(a) \geq \alpha \}, \ \forall \alpha \in (0,1] \quad (2.1)$$

has the following properties:

$C(a^*)_\alpha$ is not empty for $\alpha = 1$,

$C(a^*)_\alpha$ is a simple connected, compact subset of R^n.

$F^*(R^n)$ denotes the set of all fuzzy vector in R^n.

Definition 2. Let $a^* \in F^*(R^n)$ be a fuzzy vector with characterizing function $\Psi_{a^*}(\cdot)$ and α-cut representing $\{C(a^*)_\alpha : \alpha \in (0,1]\}$ and let $g: R^n \rightarrow Y$, $Y \subseteq R$, be real valued continuous mapping. The fuzzy image $u^* = g(a^*)$ of the fuzzy vector a^* under the mapping $g(\cdot)$ is defined by the following characterizing function :

$$\Psi_{u^*}(u) = sup_{a \in X_u} \Psi_{a^*}(a),$$

$$X_u = \{ a \in R^n : g(a) = u \}, \ \forall u \in R, \quad (2.2)$$

with $sup_{a \in X_u} \Psi_{a^*}(a) := 0$ if X_u is empty.

The following theorem proves that this fuzzy image is fuzzy number. It describes how to the fuzziness of a fuzzy vector is propagated by a non-fuzzy function on a certain α-level. This theorem will be the basic tool for generalizing statistical inference to fuzzy data.

Theorem 1. Let $a^* \in F^*(R^n)$ be a fuzzy vector with characterizing function $\Psi_{a^*}(\cdot)$ and α-cut representation $\{C(a^*)_\alpha ; \alpha \in (0,1]\}$, and let $g: R^n \rightarrow Y$, $Y \subseteq R$, be a real valued continuous mapping. If the function defined by (2.2), The following holds:

(a) $u^* = g(a^*)$ fuzzy number

(b) The α-cut representation of u^* if given by

$$C(u^*)_\alpha = \left[min_{a \in C(a^*)_\alpha} g(a), \ max_{a \in C(a^*)_\alpha} g(a) \right], \ \forall \alpha \in (0,1]. \quad (2.3)$$

In descript statistics data $x = (x_1, x_2, \cdots, x_n)$ is described by some characteristic values such as the mean and sum of squares. Theses values can be extended to a fuzzy data $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$.

Corollary 1. Let $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$ be a fuzzy data. with characterizing function $\Psi_{x^*}(\cdot)$. The mean is fuzzy sample $\overline{x^*} \in F^*(R)$ with characterizing function $\Psi_{\overline{x^*}}(\cdot)$ given by

$$\Psi_{\overline{x^*}}(y) = sup_{x \in X_y} \Psi_{x^*}(x),$$

$$X_y = \left\{ x \in R^n : \frac{1}{x} \sum_{i=1}^{n} x_i = y \right\}, \ \forall y \in R. \quad (2.4)$$

The α-cut representation is given by

$$C(\overline{x^*})_\alpha = \left[min_{x \in C(\overline{x^*})_\alpha} \frac{1}{n} \sum_{i=1}^{n} x_i, \ max_{x \in C(\overline{x^*})_\alpha} \frac{1}{n} \sum_{i=1}^{n} x_i \right], \ \forall \alpha \in (0,1]. \quad (2.5)$$

Corollary 2. Let $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$ be a fuzzy data. with characterizing function $\Psi_{\overline{x^*}}(\cdot)$. The sum of squares are fuzzy number $(SS)^* \in F^*(R)$ with $supp(SS) \subseteq R^+$ and with characterizing function given by

$$\Psi_{(SS)^*}(y) = sup_{x \in X_y} \Psi_{x^*}(x),$$

$$X_y = \{x \in R^n : (x - C(\overline{x^*}))(x - C(\overline{x^*}))' = y\}. \quad (2.6)$$

The α-cut representation of $(SS)^*$ are given by

$$C(SS)^*_\alpha(y) = \left[min_{x \in C(\overline{x^*})_\alpha} g(x),\ max_{x \in C(\overline{x^*})_\alpha} g(x) \right],\ \forall \alpha \in (0,1] \quad (2.7)$$

where $g(x) = (x - C(\overline{x^*}))(x - C(\overline{x^*}))'$.

3. Matrix formulation for variance components

Using the matrix formulation, the observation fuzzy data in random effect and balanced designs for one way classification of n observation and a classes can be represented by the model

$$Y = \mu + X\beta + \varepsilon, \quad (3.1)$$

$$= (I_a \otimes 1_n)\mu + (I_a \otimes 1_n)\beta + \varepsilon \quad (3.2)$$

where μ is a fuzzy mean vector, the treatment effect β is a fuzzy random vector with fuzzy near $\tilde{0}$ and fuzzy variance $\widetilde{\sigma_\beta}$.

When $E(Y) = \theta$ and $Var(Y) = V$, we write

$$E(Y'AY) = tr(AV) + \theta' A \theta . \quad (3.3)$$

Now for Y of (3.2), we have $Y \sim (\theta, V)$ with $\theta = 1_{an}\mu = (1_a \otimes 1_n)\mu$

$$V = I_a \otimes (\sigma^2_\beta J_n + \sigma^2_e I_n) \quad (3.4)$$

where $\otimes$ is the Kronecker's product operator.

We can express SSA, using the $\overline{J_n}$ definition, as ,

$$SSA = Y'(\{_d\,\overline{J_n}\}^a_{i=1} - \overline{J_{an}})Y$$

$$= Y'(I_a \otimes \overline{J_n} - \overline{J_{an}}c)Y = Y'AY \quad (3.5)$$

where A is defined as

$$A = I_a \otimes \overline{J_n} - \overline{J_{an}} = I_a \otimes \overline{J_n} - \overline{J_a} \otimes \overline{J_n}$$

$$= (I_a - \overline{J_a}) \otimes \overline{J_n} \quad (3.6)$$

and $\overline{J_n} = \frac{1}{n} J_n$.

Hence, using (3.5) and (3.4) in (3.1) gives

$$E(SSA) = tr\{[(I_a - \overline{J_a}) \otimes \overline{J_n}] \cdot [(I_a \otimes (\sigma^2_\beta J_n + \sigma^2_e I_n)]\} + \mu(1'_a \otimes 1'_n) \cdot [(I_a - \overline{J_a}) \otimes \overline{J_n}](1_a \otimes 1_n)\mu$$

$$= tr[(I_a - \overline{J_a}) \otimes (\sigma^2_\beta J_n + \sigma^2_e J_n)] + 0. \quad (3.7)$$

The zero because

$$1_a'(I_a - \overline{J_a}) = 1_a' - 1_a' = 0$$

Thus we have

$$E(SSA) = [tr[(I_a - \overline{J_a})][tr(\sigma^2_\beta J_n + \sigma^2_e I_n)]$$

$$= (a-1)(n\sigma^2_\beta + \sigma^2_e) \quad (3.8)$$

The similar method

$$E(SSE) = a(n-1)\sigma^2_e \quad (3.9)$$

There yields the estimators

$$\widehat{\sigma^2_e} = \frac{SSE}{a(n-1)} = MSE \quad (3.10)$$

$$\widehat{\sigma^2_\beta} = \left(\frac{SSA}{a-1} - \widehat{\sigma^2_e}\right)/n = \frac{MSA - MSE}{n} . \quad (3.11)$$

4. Example

To illustrate the model the analysis we examine the data atmospheric ozone depletion for 7 observation and 3 latitude classes :

$$Y = \begin{bmatrix} (1.5(0.1\alpha+0.9),\ 1.5(-0.1\alpha+1.1)) \\ (0.3(0.1\alpha+0.9),\ 0.3(-0.1\alpha+1.1)) \\ (1.1(0.1\alpha+0.9),\ 1.1(-0.1\alpha+1.1)) \\ (1.7(0.1\alpha+0.9),\ 1.7(-0.1\alpha+1.1)) \\ (0.7(0.1\alpha+0.9),\ 0.7(-0.1\alpha+1.1)) \\ (1.8(0.1\alpha+0.9),\ 1.8(-0.1\alpha+1.1)) \\ (2.5(0.1\alpha+0.9),\ 2.5(-0.1\alpha+1.1)) \\ (3.9(0.1\alpha+0.9),\ 3.9(-0.1\alpha+1.1)) \\ (2.9(0.1\alpha+0.9),\ 2.9(-0.1\alpha+1.1)) \\ (0.6(0.1\alpha+0.9),\ 0.6(-0.1\alpha+1.1)) \\ (1.3(0.1\alpha+0.9),\ 1.3(-0.1\alpha+1.1)) \\ (3.0(0.1\alpha+0.9),\ 3.0(-0.1\alpha+1.1)) \\ (2.8(0.1\alpha+0.9),\ 2.8(-0.1\alpha+1.1)) \\ (1.8(0.1\alpha+0.9),\ 1.8(-0.1\alpha+1.1)) \\ (4.7(0.1\alpha+0.9),\ 4.7(-0.1\alpha+1.1)) \\ (3.2(0.1\alpha+0.9),\ 3.2(-0.1\alpha+1.1)) \\ (2.4(0.1\alpha+0.9),\ 2.4(-0.1\alpha+1.1)) \\ (4.7(0.1\alpha+0.9),\ 4.7(-0.1\alpha+1.1)) \\ (4.2(0.1\alpha+0.9),\ 4.2(-0.1\alpha+1.1)) \\ (3.8(0.1\alpha+0.9),\ 3.8(-0.1\alpha+1.1)) \\ (2.5(0.1\alpha+0.9),\ 2.5(-0.1\alpha+1.1)) \end{bmatrix}.$$

Thus we have

$$SSA = (0.00\alpha^2 + 10.22\alpha + 7.98,\ 5.18\alpha^2 - 27.16\alpha + 40.18)$$

$SSE = (1.96\alpha^2 + 10.09\alpha + 4.61,$
$5.93\alpha^2 - 28.14\alpha + 38.88),$

and [Table1], [Table 2].

α	SSA	SSE
0	(7.98, 80.18)	(4.61, 38.88)
0.5	(13.09, 27.86)	(10.15, 26.29)
1	(18.2, 18.2)	(16.67, 16.67)

[Table1]

α	$\widehat{\sigma}^2_\beta$	$\widehat{\sigma}^2_e$
0	(0.12, 2.83)	(0.26, 2.83)
0.5	(0.73, 1.91)	(0.56, 1.46)
1	(1.17, 1.17)	(0.93, 0.93)

[Table 2]

For the example, the hypotheses that

$$H_0 : \sigma^2_\beta = 0$$

against the alternative

$$H_1 : \sigma^2_\beta > 0$$

can be tested with

$$F = \frac{\widehat{\sigma}^2_e - 7\widehat{\sigma}^2_\beta}{\widehat{\sigma}^2_e}.$$

The ratio follows the central F-distribution with 2 and 18 degree of freedom under both the null and alternative hypotheses.

For the ozone depletion example, we can test as the following [Table 3] by agreement index ([7]) for $F(0.01;2,18) = 6.01$.

α	F-statistics	rejection retio
0	(0.51, 84.5)	0.077
0.5	(4.84, 26.48)	0.026
1	(9.81, 9.81)	1

[Table 3]

5. References

[1] Dubois, D. and Prade H. , Possibility Theroy, Plenum Press, New York, 1988.

[2] Feng, Y., Sums of independent fuzzy random variables, Fuzzy Set and Systems, 123, 11-18, 2001.

[3] Feng, Y., Hu, L. and Shu, H., The variance and covariance of fuzzy random variables and their applications, Fuzzy Set and Systems, 120, 487-497,2001,

[4] Gizegorzewski, P. X. , Testing hypotheses with vague data, Fuzzy Sets and Systems. 112, 501-510, 2000.

[5] Kang, M. K. , Choi, G. T. , Lee, C. E. , On statistical for fuzzy hypotheses with fuzzy data, Proceeding of Korea Fuzzy Logic and Intelligent System Society Fall Conference, Vol. 10, Num. 2, 2000.

[6] Kang, M. K. , Choi, J. R. , Bae, E. S. , Some properties of fuzzy sample correlation coefficient with fuzzy data, Far East J. Theo. Stat. 4(1), 165-177, 2000.

[7] Kang, M. K. , Jang, S. W., Kwon, O. J., Testing statistical hypotheses for fuzzy data with agreement index, Far East J. Theo. Stat. 5(2), 343-352, 2001.

[8] Li, F. and Wu. etc., Platform type fuzzy number a separability of fuzzy number space, Fuzzy Sets and Systems. 117, 347-353,2000.

[9] Pill, D. and Peter, K. , Metric space of fuzzy sets, World Scientific, 1994.

[10] Rao, P. S. R. S., Variance components estimations, Chapman & Hall, 1991.

[11] Sylbia F-S., On statistical inference for fuzzy data with applications to descriptive statistics, Fuzzy Set and Systems, 50, 143-165, 1992.