Journal of Applied Science and Engineering Innovation, Vol.9, No.1, 2022, pp. 15-18
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

The Establishment and Information Processing Analysis of A Kind of Multi-information Statistical Inference Optimization Decision Model

Xiao Xiaonan [1,2]

Xiamen University Tan Kahkee College, Xiamen 363105, China.

Abstract*:* In recent years, comprehensive multi information big data statistical analysis has become one of the most challenging frontier emerging technology fields at present and in the future. Based on a brand-new thinking and new thinking, this paper comprehensively uses modern information decision-making methods such as multi-index Fuzzy dynamic cluster analysis, big data analysis and intelligent calculation to study the establishment of a type of disease diagnosis and optimization model. It provides an ideal mathematical processing method and method for the effective realization of automatic control and information management of such problems.

Keywords. **S**tatistic inference; Information fusion; Sampling design and investigation; System optimization; Fuzzy dynamic cluster analysis; Big data; Statistical analysis

INTRODUCTION

Sampling design and investigation play a significant role in statistic inference. In order to overcome the defects in traditional method of sampling design and investigation, the article implements multi-index Fuzzy dynamic cluster analysis and offers a unique way of information statistics and packet optimization. In this way the internal information in the sample will be sufficiently explored, thus providing a unique method in statistical analysis and system optimization that will improve the reliability of overall sample grouping and reduce the sampling error on a maximal scale.

THE DEFECTS AND OPTIMIZATION OF THE TRADITIONAL SAMPLING INVESTIGATION METHODOLOGY

Sampling investigation is not only the method to collect statistical data but the one to estimate and judge the appearance of the population scientifically. In addition, it is of great importance in the statistics analysis and statistics pre-decision study. What's more, it is applied extensively to the modern enterprise management with preferential plan, investment decision, efficiency evaluation and various fields of scientific technology [Du, *et. al.*, 2005]. While there still exists a great many factors in the investigation and improvement on ies and methods [Xiao, 2010], in terms of the traditional sampling investigation method, it is to randomly choose part of the entire objects for investigation. Moreover, based upon the acquired data, we could make somewhat reliable estimation and judgment on the quantitative feature of the entire objects so that all of the studied objects can be recognized [He, 2009]. However, there are several problems in sampling investigation. For one thing, in the sampling project and the signs of sampling indicators we may have difficulties in taking into account various factors thoroughly which may have an impact on the sampling target. As a result, the research may lose a few helpful statistic information, especially the very complex information such as gray model, fuzzy model, fuzzy and random model, fuzzy and gray model and so on. It may lead to a short of sufficient scientific reliability in the investigation. For another thing, when the interclass unit is adjusted by interclass density, the packet will be not accurate and too sketchy for the reason that it has used the similar method which seems to have made a strong assumption [Shi, *et. al.*, 2008]. To solve such a statistics methodology problem quickly, the article will apply the multi-index Fuzzy dynamic cluster analysis to conduct the sampling optimization classification study.

In the sampling investigation, the multi-index Fuzzy dynamic cluster analysis is introduced into sampling and a way of information statistics to group and optimize is given so that it can overcome the defects many fuzzy and complex factors of sampling being unreasonably grouped in the traditional stratified sampling information processing [Qiu, *et. al.*, 2009]. Combining qualitative sampling analysis with quantitative sampling analysis organically and closely, it offers a new methodology about statistics analysis and optimization for improving the reliability of population sample packet and reducing the sample error furthest.

THE STATISTICS SAMPLING METHOD BASED ON MULTI-INDEX FUZZY DYNAMIC CLUSTER ANALYSIS

Because the objective things may contain gray nature and fuzziness in most cases, applying the multi-index Fuzzy dynamic cluster analysis to classify and investigate the sampling makes the grouping more practically.

Supposing U={u1，u2，…，un} as the unit to be grouped, every sample has m statistic indicators (flag values). xik stands for the number k flag value of the number i sample. What is the result of applying the multi-index Fuzzy dynamic cluster analysis to the above? The key is to make the statistic indicators to be the preferential choice. In other words, statistic indicators should have a clear and real meaning as well as the stronger discrimination, representation and broader meaning.

After choosing the statistic indicator, the multi-index Fuzzy dynamic cluster analysis of the population sampling can be classified as the following three steps:

Firstly, standardize the statistic indicators of each classified sample so that they can be analyzed and compared reasonably. And the formula of the standardized value is:

$$x'_{ik} = \frac{x^0_{ik} - \bar{x}_i}{s_i}$$

（i=1，2，…，n；k=1，2，…，m） （1）

In it, x^0_{ik} is the primary data, $\bar{x}_i$ is the average of the primary data, so

$$\bar{X}_i = \frac{1}{m}\sum_{k=1}^{m} x^0_{ik}$$

Si is the deviation of the primary data, so

$$s_i = \sqrt{\frac{1}{m}\sum_{k=1}^{m}(x^0_{ik} - \bar{x}_i)^2}\,.$$

Reducing the standardized value to a closed interval [0，1], we can calculate the extreme, standardized value in terms of the formula:

$$x_{ik} = \frac{x'_{ik} - \min(x'_{ik})}{\max(x'_{ik}) - \min(x'_{ik})}$$

(*i*=1，2，…，*n;k=1，2，…，m*) （2）

Secondly, calculate the statistic which is used to measure the similarity between the two classified samples rij(i，j，=1，2…，n), and then create the similar relationship based on the population U R =（rij）n×n .

rij can be calculated following the formula：

$$r_{ij} = \frac{\sum_{k=1}^{m}\left|X_{ik} - \bar{X}_i\right|\left|X_{jk} - \bar{X}_j\right|}{\sqrt{\sum_{k=1}^{m}(X_{ik} - \bar{X}_i)^2}\sqrt{\sum_{k=1}^{m}(X_{jk} - \bar{X}_j)^2}}$$

In the formula, .

$$\bar{X}_i = \frac{1}{m}\sum_{k=1}^{m} X_{ik},\ \bar{X}_j = \frac{1}{m}\sum_{k=1}^{m} X_{jk}$$

There is another formula can be used for calculation according to the actual conditions. For instance:

$$r_{ij} = \frac{\left|\sum_{k=1}^{m} x_{ik}x_{jk}\right|}{\sqrt{(\sum_{k=1}^{m} x^2_{ik})\,(\sum_{k=1}^{m} x^2_{jk})}}; \quad (4)$$

$$r_{ij} = \frac{\sum_{k=1}^{m}\min(x_{ik}x_{jk})}{\frac{1}{2}\sum_{k=1}^{m}\sqrt{x_{ik}x_{jk}}}; \quad (5)$$

$$r_{ij} = \begin{cases} 1 & \text{while } i = j \\ 1 - C\sum_{k=1}^{m}\left|x_{ik} - x_{jk}\right| & \text{while } i \neq j \end{cases}$$

（6）

Specifically, C should be chosen properly, and make 0≤rij≤1 and so on.

Thirdly, cluster. Applying synthesis to transform the Fuzzy similar matrix into the Fuzzy equivalence matrix, enable $\underset{\sim}{R}^{2k} = \underset{\sim}{R}^{k} = \underset{\sim}{R}^{*}$. Then R* should be clustered and analyzed. Thereby, a realistic classification will be gained according to the arrangement of 0,1 among Rλ。

Generally, when the number of samples tends to be quite large, we can build the similar Fuzzy relationship among the samples and synthesize them for several times n terms of the characteristics of the samples to be clustered. Then, we can transform R into the Fuzzy equivalent matrix R* .What's more, all this can be completed by the electronic computer. It is not only accurate but also fast to adopting the electronic computer to the population sample with the multi-index Fuzzy dynamic cluster classification. For instance, if there are 100 samples and every sample has 10 indicators, it just needs 3 minutes or so from starting calculation to outputting the Fuzzy equivalent matrix. The accurate and instant level is what the traditional stratified sample cannot reach.

ESTABLISHMENT OF FUZZY DECISION OPTIMIZING DISCRIMINANT MODEL

Let us set $A_1, A_2, \ldots, A_n$ as several Fuzzy Subsets on Universe U. In the Universe U, for any element U0, if there is

$\mu_{A_i}(u_0) = max\{\mu_{A_1}(u_0), \mu_{A_2}(u_0), \dots, \mu_{A_n}(u_0)\}$

It is said that U_0 belongs to A_i relatively. Where $i \in \{1,2, \dots, n\}$, a mathematical model for identifying diseases can be established.

It is assumed that m symptoms S_1, S_2, …, S_m can be used to diagnose n diseases $A_1, A_2, \dots, A_n$, and each symptom is only selected in two states "occurrence and non-occurrence". Now "1" is used to indicate the occurrence of symptoms, and "0" is used to indicate the absence of symptoms. Then

$S_i = \begin{cases} 1 \text{ occurrence of symptoms} \\ 0 \text{ absence of symptoms} \end{cases}$ （i =1,2,…,n）

So at this time the universe U=$\{S_1, S_2, \dots, S_m \mid S_i = 0,1, i \le m\}$

Let us suppose that the elements in the Universe U corresponding to the symptoms of a typical case Ai are $U^{(i)} = (S_1^{(i)}, S_2^{(i)}, \dots, S_m^{(i)})$ i = 1,2,…,m, the membership function of Ai can be selected as

$$\mu_{A_i}(u) = cos\sqrt{\sum_{j=1}^{m}(S_j - S_j^{(i)})^2/m} \quad (i = 1,2, \dots, m) \qquad (7)$$

If there is an element u_0 = （$S_1^{(0)}$, $S_2^{(0)}$, …, $S_m^{(0)}$） in the domain U corresponding to the symptoms of a patient, then u_0 is substituted into formula （7）. Then we get $\mu_{A_1}(u_0), \mu_{A_2}(u_0), \dots, \mu_{A_m}(u_0)$ .. So we choose $U_{A_i}(u_0) = max\{U_{A_1}(u_0), U_{A_2}(u_0), \dots, U_{A_n}(u_0)\}$. According to the principle of maximum subordination, the patient's disease can be diagnosed as Ai.

According to the above mathematical model, the membership functions of typical cases A1 chronic hepatitis, A2 acute jaundice hepatitis, A3 cirrhosis、A4 hepatocellular carcinoma can be used to identify which kind of liver diseases a patient suffers from.

The development of modern decision-making science and the deepening of decision-making research, as well as the new problems raised in decision-making practice, have prompted the decision-making methods to be mathematically, modeled and computerized, which in turn requires the continuous improvement of mathematical processing means and logical procedures, and has promoted the development of mathematics. Fuzzy decision-making is a new one. The main contribution of Xing's frontier discipline is that it integrates Fuzziness with mathematical quantitative research. Its method is not to let mathematics abandon strictness to accommodate fuzziness, but to penetrate mathematical methods into the forbidden zone with fuzzy phenomena, so as to solve some scientific decision-making problems of complex large systems and fuzzy factors. It has opened up a new road, and its influence on decision-making science is far-reaching .

EXAMPLE DISCUSSION

Supposing X ~ N(μ, σ 2), σ 2 is generally known, x1, x2, ⋯, xn, is a random sample from X whose sample size is of n. Under the significance level α (0 < α < 1) the hypothesis test is

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0$$

From (4), we get that if only we acquire the two-sided confidence interval L out μ of the $1-\alpha$, then when μ0∈(R - L), it reject H0. Now drawing on X ~ N(μ, σ 2), we should select statistics

$$U = \frac{\bar{X} - \mu}{a/\sqrt{n}} \sim N(0,1)$$

($\bar{X}$ is the sample mean), then based on the sub-site definition of standard normal distribution α, we have

$$P\left\{\frac{\bar{X} - \mu}{a/\sqrt{n}} < U_{\alpha/2}\right\} = 1-\alpha \qquad (8)$$

namely

$$P\left\{\frac{\bar{X} - \mu}{a/\sqrt{n}} \ge U_{\alpha/2}\right\} = \alpha \qquad (9)$$

Moreover we find that the 1- α confidence interval of μ is:

$$L = \left(\bar{X} - \frac{\sigma}{\sqrt{n}}U_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}U_{\alpha/2}\right)$$

And the rejection region of H0is (the α significance level):

$$R - L = \left(-\infty, \bar{X} - \frac{\sigma}{\sqrt{n}}U_{\alpha/2}\right] \cup \left[\bar{X} + \frac{\sigma}{\sqrt{n}}U_{\alpha/2}, +\infty\right)$$

Obviously, the (9) and the

$$P\{rejected\ H_0 \mid H_0\ is\ true\} = \alpha$$

are equivalent.

Like the above double-sided test with respect to the mean μ, it is not difficult to use the similar method with which we get the one-sided confidence of $1-\alpha$ for the μ to solve the μ rejection region issue for one-sided test[9-10].

6 The prospect

The article uses the method of the statistic information sampling investigation with an emphasis of the multi-index Fuzzy dynamic cluster analysis in order to improve and develop the sampling investigation method. There are several advantages to group the population samples by the multi-index Fuzzy cluster analysis. First, the internal information will be discovered substantially. Second, many fuzzy and complex factors of the samples can be considered thoroughly. Third, it will overcome the defects of the approach of the traditional stratification sampling whose information is unilateral. Fourth, it may combine the qualitative sampling analysis and quantitative sampling analysis closely. Last but not

least, it offers a new way of statistics analysis for optimization in order to improving the reliability of population sample packet and minimizing the sampling errors.

This article uses interval estimation knowledge to solve the problem of rejection region in hypothesis test through the control of the probability of reducing error and getting the corrected answer and the neglect of the possible errors in the probability test. It inspires us. To accurately grasp the intrinsic link between the two mathematical concepts is the key to led to the new conclusion. And the organic link and distinction between interval estimation and hypothesis test affirm the dialectic relationship between mathematics [Yan, et. al. 2002].

REFERENCES

Dogru O,Duman O.Statistical approximation of Meyer-Konig and Zeller operators based on q-integers[J].Publ Math Debrecen,2006,68(1-2):199-214.

Du Dong, Statistics Information System [M]. Beijing: Chinese Statistics Press, 2006.

Du Zifang, Sampling Technique and Application [M]. Beijing : Tsinghua University Press, 2005.

He Weilian. Permanence for single species model with feedback control and finite continuous delays[J]. Journal of Fuzhou University(Natural science Edition), 2009, 37(4): 468-470.

Huang Tianyun. The restrained optimization pattern search method research progress[J]. Chinese Journal of Computers, 2008, 31(7): 1200-1251.

Lin R,Liu F.Fractional high order approximation methods for the nonlinear fractional ordinary differential equation[J].Nonlinear Analysis,200766(4):856-869.

Qiu J,Zhang L.F-intefere law genercetion and its feature recognition[J].Journal of Systems Engineering and Electronics,2009,20(4):777-783.

Shi Kaiquan, Yao Bingxue. Function S-rough sets and las identification[J].Science in China Series F:Information Science,2008,51(5):499-510.

Wang Yufang and Xiao Shitang, The Planning and Process of Statistics Sampling Investigation [M]. Beijing: Chinese Economy Press, 2005.

Wei W L,Ysn H.A method of transferring polyhedron between the intersection-form and the sum-form[J].Computers and Mathematics with Applications,2001(41):1327-1342.

Xiao Xiaonan. The optimal non-liner filtering and majorized algorithm of a kind of nonstationary stochastic transmission system[J]. Journal of Mathematical Study, 2010, 43(4): 342-351.

Yan H,Wei Q L.Determining compromise weights for group decision making[J].Journal of Operational Research Society,2002(53):680-687.