

# PRAM and its influence of multivariate analyses

Tomokazu Fujino

*Graduate school of natural science and technology, Okayama University*

*3-1-1 Tsushima-naka*

*Okayama, Japan*

*fujino@ems.okayama-u.ac.jp*

Tomoyuki Tarumi

*Dept. of Environmental & Mathematical Sciences, Okayama University*

*3-1-1 Tsushima-naka*

*Okayama, Japan*

*t2@ems.okayama-u.ac.jp*

## 1. Introduction

PRAM(Post Randomization Method) was developed by Kooiman et al.(1997). It is a method of the statistical disclosure control(SDC) for microdata. The procedure is that the score of the categorical variable was changed into another score according to a certain structure of probability. The receiver of the microdata can get PRAMed microdata together with the structure of transition probability and estimate original contingency tables with them.

Now let  $A$  be a categorical variable with  $K$  categories, and  $A^*$  be its perturbed variable. Then PRAM matrix  $P_A$  that is the structure of probability is defined by

$$p_{kl} = P(A^* = l | A = k). \quad (1)$$

Then  $P_A$  consists of a  $K \times K$  Markov matrix. When  $T_A$  and  $T_{A^*}$  are frequency tables of  $A$  and  $A^*$  respectively,  $T_{A^*}(l)$  has the sum of the binomial distribution  $B(T_A(k), p_{kl}), k = 1, \dots, K$  given  $T_A$ . Thus we have

$$E(T_{A^*} | T_A) = P_A^t T_A. \quad (2)$$

Therefore the analyst(the receiver of microdata) can derive the unbiased estimator of  $T_A$  as follows.

$$\hat{T}_A = (P_A^{-1})^t T_{A^*}. \quad (3)$$

This is called the moment estimator of  $T_A$ .

In general, the microdata has numerical variables. Now let  $Y$  be a numerical variable and  $D_A$  be a dummy variable for  $A$ . Then

$$E(Z_{A^*} | Z_A) = P_A^t Z_A, \quad (4)$$

where  $Z_A = D_A^t Y$ . Therefore the analyst can derive the unbiased estimator of  $Z_A$  just like  $T_A$  as follows.

$$\hat{Z}_A = (P_A^{-1})^t Z_{A^*}. \quad (5)$$

## 2. Multivariate Analysis using PRAMed data and PRAM matrix

An advantage that the analyst receives the PRAMed data and PRAM matrix is to analyze the data without restriction. Kooiman et al.(1997) suggests that the analyst can perform the regression analysis and discriminant analysis using the moment estimator. In addition we think that the analyst can also perform the correspondence analysis. But when we perform such analyses using the microdata in practice, there are many problems in the results. Table 1 is the result of the regression analysis using original data and moment estimators of PRAMed data.

**Table 1. results of regression analysis using original data and PRAMed data**

	category	frequency	category score	estimated frequency	category score
A	1	5	-13.17	7.25	-9.69
	2	20	-3.32	19.75	-3.73
	3	16	6.98	17.25	7.58
	4	3	6.89	-0.25	-52.27

## REFERENCES

- Kooiman,P., Willenborg,L.C.R.J., and Gouweleeuw,J.M.(1997), PRAM: a Method for Disclosure Limitation of Microdata, Research paper no.9705, Voorburg/Heerlen: Statistics Netherlands.
- Gouweleeuw,J.M.,Kooiman,P., Willenborg,L.C.R.J., and De Wolf,P.-P.(1998), Post Randomisation for Statistical Disclosure Control: Theory and Implementation, *Journal of Official Statistics*, 14, pp 463-478
- A.D.L. van den Hout(1999), *The Analysis of Data Perturbed by PRAM*, Delft University Press, Delft
- Leon Willenborg and Ton de Waal(2000), *Elements of Statistical Disclosure Control*, Springer, New York

## RÉSUMÉ

PRAM est une méthode de la commande statistique de révélation pour des microdonnées. Nous étudions son influence des analyses multivariées, quand PRAM est appliqué aux microdonnées.