

# An Application of Multi-variable Probability Proportional to Size Sampling for Multi-Level Use in China

Zhang Guohong

*Food and Agricultural Statistics Center, National Bureau of Statistics*

*No.75 Yuetan South Street*

*Beijing-100826, P.R.C*

*e-mail: guohong\_zhang@yahoo.com*

Zhou Wei

*Rural Socio-economic Survey Organization*

*No.75 Yuetan South Street*

*Beijing-100826, P.R.C*

## 1. Background

(1) In the transition period from a planned economy to a more market-oriented economy, decision-makers are more and more relying on statistics to make their decisions. Not only the central government but also the governments at provincial, prefecture and county levels need data to head and monitor the economic development. Under this circumstance, the needs for statistics at various levels should be integrated in the sampling design.

(2) Present agricultural statistical system is a mixed outcome from the planned and market economies, which needs to be reformed to adapt to the present reality. Efficiency and effective sampling surveys should completely replace the complete reporting system.

(3) In order to integrate multi-variable and multi-level use, new sampling design is under study in a Sino-American joint project, collaborated with NASS/USDA. We are exploring an efficient way to collect agricultural data under Chinese conditions.

## 2. Multi-variable Probability proportional to Size (MPPS) Method for Multi-level Use

### 2.1 Multi-variable Probability Proportional to Size Sampling

(1) Defining probability of selection for each village,  $p_i$

$$p_i = \min \left[ 1, \max \left[ n_1 \frac{x_{1,i}^{3/4}}{\sum_{i=1}^N x_{1,i}^{3/4}}, K, n_k \frac{x_{k,i}^{3/4}}{\sum_{i=1}^N x_{k,i}^{3/4}} \right] \right]$$

where  $x_{k,i}$  is the size measure of the  $k^{\text{th}}$  item of interest for the  $i^{\text{th}}$  village;  $n_k$  is the desired sample size for the  $k^{\text{th}}$  item;  $K$  is the total number of items of interest in the survey and  $N$  is the total number of villages in the sampling population.

(2) Estimating for the population total and variance of estimates

$$\hat{Y}_k = \sum_{j=1}^N x_{k,j} \frac{\sum_{i=1}^n w_i \hat{Y}_{k,i}}{\sum_{i=1}^n w_i x_{k,i}} \quad \text{where } x_{k,i} \text{ is control data, and } \hat{Y}_{k,i} \text{ is survey data}$$

$$V(\hat{Y}_k) = \frac{\left( \sum_{j=1}^N x_{k,j} \right)^2}{\left( \sum_{i=1}^n w_i x_{k,i} \right)^2} \sum_{i=1}^n w_i^2 e_{k,i}^2 \quad \text{where } e_{k,i} = \hat{Y}_{k,i} - x_{k,i} \frac{\sum_{l=1}^n w_l \hat{Y}_{k,l}}{\sum_{l=1}^n w_l x_{k,l}} \quad \text{and } w_i = \frac{1}{p_i}$$

## 2.2 MPPS applied for multi-level purpose

- (1) Assigning permanent random numbers (PRN) to the sampling population
- (2) Defining probability of selection for each village,  $p_i$
- (3) Selecting samples with Poisson sampling
- (4) Estimating for the population total and variance of estimates
- (5) Calculating the standard error (RMSE) of estimator

## 3. Pilot Test in Guangdong province

(1) We determined 9 indicators involved in MPPS sampling, which could be divided into three categories in terms of their priorities. Under the Expected C.V for province, prefecture and county levels, the total sample size for province, prefectures and counties in Guangdong province is 1008, 2982 and 5889 respectively based on the 1997 agricultural census frame. The sampling ratio is approximately 4.22%, 12.49% and 24.67% respectively in three levels.

(2) We selected three prefectures in Guangdong province to do the Pilot test, namely Guangzhou, Zhongshan and Huizhou. In Feb.2001, We summarized the survey data which collected from sampling villages in these three prefectures. Generally speaking, Estimations for three levels are reasonable and acceptable, a better precision in terms of C.V achieved in provincial and prefecture levels compared with that in county levels. In addition, survey data compared with control data have a large gap, that means, list frame has changed quite a lot since 1997 the agricultural census year.

## 4. Further study to be done

- (1) To define the variables used in the survey and up-date the frame
- (2) Based on the new frame, to select sample villages, from which data can be derived for the purpose of various levels
- (3) Study and test the ways to select households/pilots in the sample villages and also collect data
- (4) To prepare a complete program for the MPPS sampling for multi-level use to expand all over China

## REFERENCE

1. John F. Amrhein and Jeffrey T. Bailey, Sampling Villages for Multipurpose Surveys, Proceedings of the Joint IASS/IAOS Conference, September 1998.