



**American Journal of Experimental Agriculture**  
1(2): 33-39, 2011

**SCIENCEDOMAIN international**

www.sciencedomain.org



# Determination of Probability of Selection for Two Measures of Sizes Using Unequal Probability Sampling in Agricultural Sample Surveys

Abdul Qayyum<sup>1\*</sup> and Muhammad Khalid Pervaiz<sup>1</sup>

<sup>1</sup>Department of Statistics, G. C. University, Lahore, Pakistan

Received 7<sup>th</sup> January 2011

Accepted 31<sup>st</sup> January 2011

Online Ready 10<sup>th</sup> February 2011

Research Article

## ABSTRACT

In this paper an effort has been made to design a methodology for the determination of probability of selection for the cases where more than one measure of sizes are involved, especially in the case of agricultural sample surveys. So far no effort has been made for the cases of this sort in the literature of unequal probability sampling / probability proportional to size sampling. In this paper, a method has been developed to merge two measures of sizes in to one, which has combined proportional affect of both measures of sizes. The method is also valid for two measures of sizes having different units of measurements. It is also an unprecedented effort. A comparative study, using Horvitz – Thompson estimator and taking examples from agricultural surveys have been carried out to assess the performance of the new methodology.

*Keywords: Thompson estimator, kharif season, probability proportional to size sampling, rabi season, unequal probability sampling;*

## 1. INTRODUCTION

In case of unequal probability sampling design, a sample with varying probabilities of selection for each unit is selected. The probability of selection  $P_i$  of  $i^{\text{th}}$  unit is proportional to its relevant measure of size  $Z_i$ , and importantly, they are directly allied. The variable  $Z$  is highly correlated with the variable of interest  $Y$ .

\*Corresponding author: Email: qayyum\_so@hotmail.co; qayyum.ravian@gmail.com

To find the probability of selection  $P_i$  of  $i^{\text{th}}$  unit, using  $Z_i$ , the following relation is used,

$$P_i = \frac{Z_i}{\sum_{i=1}^N Z_i}$$

There are population units and their respective probabilities of selection depending on their respective importance / measure of size.

There are many practical situations when it is required to select a sample using the technique of Probability Proportional to Size Sampling with consideration of more than one measure of sizes, i.e. there are two or more variables that are highly correlated with the main variable of interest. In such cases, there is a need of selection of a sample having proportional effect of both measures of sizes.

For instance, in an agricultural crop reporting survey, it may be desired to select a sample of villages using Probability Proportional to Size Sampling method giving consideration to two important measures of sizes like:

- i) 'Acreage of the village' as  $Z_1$  and 'Acreage of wheat in the village' as  $Z_2$ , i.e., the villages having more area as well as more wheat cultivation are more likely as compared to the villages having more area but less wheat cultivation or villages having less area but more wheat cultivation i.e., selection of sample villages should be proportional to  $Z_1$  &  $Z_2$ .
- ii) 'Wheat Acreage of the village' as  $Z_1$  in Rabi season and 'Acreage of cotton in the village' as  $Z_2$  in Kharif season, i.e., the villages having more wheat acreage as well as more cotton acreage are more likely as compared to the villages having more wheat acreage but less cotton acreage and vice versa.
- iii) 'Average per acre yield of wheat' as  $Z_1$  and 'Average per acre yield of cotton as  $Z_2$ .
- iv) To avoid the absolute values, 'proportion of wheat acreage of the village' as  $Z_1$  and 'proportion of cotton acreage of the village' as  $Z_2$  may be used.

Importantly, two types of surveys in both seasons (rabi & kharif) are conducted by the agriculture department, Punjab, in selected villages: one for estimation of acreage and second for estimation of yield of a crop, i.e., there are multiple variables of interest over the same selected sample.

## 2. PROPOSED METHODOLOGY

Let there is a population of size 'N' and two measures of sizes  $Z_{1i}$  and  $Z_{2i}$  for  $i^{\text{th}}$  unit of the population  $Y_i$ . Calculation of probability of selection  $P_{1i}$  and  $P_{2i}$  for each unit of population, independently for each measure of size as follows:

$$P_{1i} = \frac{Z_{1i}}{\sum_{i=1}^N Z_{1i}} \quad \& \quad P_{2i} = \frac{Z_{2i}}{\sum_{i=1}^N Z_{2i}}$$

As both measures of sizes are statistically independent, so the joint probability of selection can be calculated as:

$$(P_{1i} \cap P_{2i}) = P_{1i} \cdot P_{2i}$$

But the joint probability of selection is not collective exhaustive, which is the mandatory condition for the application of any kind of selection procedure in unequal probability selection design, i. e.,

$$\sum_{i=1}^N (P_{1i} \cap P_{2i}) < 1$$

To make the joint probability as collectively exhaustive, the following constant value is added,

$$\frac{(1 - \sum_{i=1}^N P_{1i} \cdot P_{2i})}{N}$$

to each joint probability, i.e.,

$$P_{Ci} = P_{1i} \cdot P_{2i} + \frac{(1 - \sum_{i=1}^N P_{1i} \cdot P_{2i})}{N}$$

Where  $P_{Ci}$  is the Combined Probability of selection of  $i^{\text{th}}$  unit of population in the sample proportional to the measure of sizes  $Z_{1i}$  and  $Z_{2i}$  such that,

$$\sum_{i=1}^N P_{Ci} = 1$$

Now against each unit of population  $Y_i$ , probability of selection  $P_{Ci}$  contains the combined effect of both measure of sizes  $Z_{1i}$  and  $Z_{2i}$ , proportional to their quantities. Now any conventional selection procedure can be implemented on the data to estimate the required parameter of interest.

Addition of a constant to each probability does not alter the variance in the variable of interest, as variance is independent of origin and scale. Similarly this idea can be extended for multiple measures of sizes.

### 3. RESULTS AND DISCUSSION

To check the validity of the proposed method, the well-known selection procedure (draw – by – draw) of Brewer (1963) and Horvitz – Thompson (1952) estimator for the estimation of population total is used.

The estimator is

$$Y_{HT} = \sum_{i=1}^n \frac{y_i}{\pi_i}$$

Where

$\pi_i$  is the probability of inclusion of  $i^{\text{th}}$  unit in the sample in all possible draws.

Probability of inclusion  $\pi_i$  and joint probability of inclusion  $\pi_{ij}$  of Brewer (1963) selection procedure are given below,

$$\pi_i = 2 P_i$$

$$\pi_{ij} = \frac{4P_i P_j}{K} \left[ \frac{1 - P_i - P_j}{(1 - 2P_i)(1 - 2P_j)} \right]$$

In case of one measure of size, coefficient of simple correlation between  $Y_i$  and  $Z_i$  is calculated. As this coefficient increases, the variance obtained by the Brewer (1963) procedure decreases and vice versa. In case of zero coefficient of correlation, the procedure approaches to simple random sampling.

But in case of two measures of sizes, partial coefficient of correlation is calculated between variable of interest and one measure of size removing the effect of other measure of size.

Two agricultural based experimental populations are considered for empirical study,

#### CASE- I:

$$Y_i : 7, 12, 14, 17, 20$$

$$Z_{1i} : 4, 7, 8, 9, 12$$

$$Z_{2i} : 1, 2, 4, 5, 8$$

Where

$Y_i$  : Total cultivated area of a village,

$Z_{1i}$  : Total cropped area under wheat in Rabi season (October to April).

$Z_{2i}$  : Total cropped area under cotton in Kharif season (May to Sept).

The objective is to determine the probability of selection of each village proportional to the combined cropped area of wheat and cotton i.e. a village having more cropped area of wheat and cotton, should have more probability of selection in the sample.

For the given population,

The coefficients of simple correlation are

$$r_{yz_1} = 0.98747 \quad ; \quad r_{yz_2} = 0.95902 \quad ; \quad r_{z_1z_2} = 0.97065$$

The coefficients of partial correlation are

$$r_{yz_1.z_2} = 0.83058 \quad ; \quad r_{yz_2.z_1} = 0.01402$$

Using the proposed method, the probability of selection for each village of population is as follows:

$Y_i$	$P_{Ci}$
7	0.15725
12	0.16975
14	0.19225
17	0.20850
20	0.27225

The maximum probability of selection 0.27225 is against the last village because it has maximum cropped area of wheat and cotton, though it also has maximum cultivated area.

Now applying Brewer (1963) selection procedure three times as for  $Y_i$  &  $P_{1i}$ ,  $Y_i$  &  $P_{2i}$  and  $Y_i$  &  $P_{Ci}$  independently, the following variances of the estimator are obtained,

<i>Between</i>	$V(Y_{HT})$
$Y_i$ & $P_{1i}$	3.43
$Y_i$ & $P_{2i}$	236.98
$Y_i$ & $P_{Ci}$	54.38

The variance between  $Y_i$  &  $P_{Ci}$  is 54.38, which lies in between other two variances rather near to the smaller variance, which depicts the goodness of the proposed method.

**CASE-II:**

$Y_i$ :	32,	18,	52,	16,	42,	48
$Z_{1i}$ :	3,	2,	5,	1,	4,	6
$Z_{2i}$ :	2,	4,	2,	5,	3,	9

Where

$Y_i$  : Total cultivated area of a village,

$Z_{1i}$  : Total cropped area under vegetables.

$Z_{2i}$  : Total cropped area under sugarcane.

Like the previous case, the objective is to determine the probability of selection of each village proportional to the combined cropped area of vegetables and sugarcane i.e. a village having more cropped area of vegetables and sugarcane, should have more probability of selection in the sample.

Similarly for the given population,

The coefficients of simple correlation are

$$r_{yz_1} = 0.95235 \quad ; \quad r_{yz_2} = 0.05625 \quad ; \quad r_{z_1z_2} = 0.30377$$

The coefficients of partial correlation are

$$r_{yz_1 \cdot z_2} = 0.98321 \quad ; \quad r_{yz_2 \cdot z_1} = - 0.80196$$

Using the new method, the probability of selection for each unit of population is as follows:

$Y_i$	$P_{Ci}$
32	0.147942
18	0.151742
52	0.155554
16	0.146030
42	0.159370
48	0.239362

The third village has total 52 cultivated area, but only 5 and 2 cropped area under vegetables and sugarcane, respectively and the last village has comparatively less cultivated area i.e. 48, but more cropped area under vegetables and sugarcane i.e. 6 & 9, respectively. Consequently, the village with more cultivated area has less probability of selection i.e. 0.155554 and less cultivated area has more probability of selection i.e. 0.239362, which indicates the utility of the proposed method.

By applying the Brewer's procedure, the table of variances is,

Between	$V(Y_{HT})$
$Y_i$ & $P_{1i}$	546.84
$Y_i$ & $P_{2i}$	11944.78
$Y_i$ & $P_{Ci}$	2326.19

Again variance for  $Y_i$  &  $P_{Ci}$  is smaller and closer to the smallest variance.

It is important to note that in both cases  $V(Y_{HT})$  for  $Y_i$  &  $P_{Ci}$  falls in between others two. Also in case of one measure of size  $V(Y_{HT})$  is associated with simple coefficient of correlation. But in case of two measures of sizes, it is associated with partial coefficient of correlation. So it is revealed that  $V(Y_{HT})$  in case of one measure of size is incomparable with  $V(Y_{HT})$  in case of two measures of sizes.

#### 4. CONCLUSION

The literature of unequal probability sampling has sufficient material for the selection of a sample using one measure of size. But use of two measures of sizes in unequal probability sampling is almost unprecedented effort. In Pakistan, for agricultural surveys, a sample of particular number of villages is selected giving priorities to more than one parameters or measure of sizes that led to this piece of work, as discussed in the introductory section. The selected sample of villages is used for both types of area and yield estimation surveys regarding all crops. So the condition of only one measure of size is not practical in agricultural surveys. It is worthy to highlight that the proposed method of probability of selection for two measures of sizes is independent of measuring units of each size. The same methodology can be easily extended for more than two measures of sizes and by applying the proposed technique; there is no more binding of usage of only one measure of size in unequal probability sampling.

#### REFERENCES

- Brewer, K. R. W. (1963a). A model of systematic sampling with unequal probabilities. *Aust. J. Stat.*, 5, 5 – 13.
- Brewer, K. R. W., Hanif, M. (1983) Sampling with unequal Probabilities. Lecture notes to Statistics, No. 15, Springer - Verlag.
- Hanif, M., Brewer, K.R.W. (1980). On Unequal Probability Sampling without replacement; a review. *Int. Stat. Rev.*, 48(3), 317 – 335.
- Horvitz, D. G., Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.*, 47, 663 – 685.