

SAMPLING STRATEGY FOR NOT-EVENLY SPREAD POPULATIONS

Muhammad Khalid Pervaiz

ABSTRACT The sampling strategy for not-evenly spread populations is discussed. The two stage stratified Random Sampling Design is employed by using "Density of an activity" as stratum weights. The first stage comprises a Village/Ward/Mahala – Primary Sampling Unit (PSU) – and the second stage comprises a household – Secondary Sampling Unit (SSU). The process is explained by taking a hypothetical example. The rule to estimate the sampled population total is developed.

1. INTRODUCTION

The problem under consideration is to develop a sampling strategy for not evenly spread populations. For example, one is interested to draw a sample among those under eighteen who are involved in carpet weaving industry. Obviously there are specific areas where carpet weaving is going on. The first step is then to divide the targeted area in to strata according to the density of the activity. Then a Two-Stage Stratified Random Sampling Design by using "Density of an activity" as stratum weights. The section 3 describes the procedure of selection of first stage a Village/Ward/Mahala – Primary Sampling Unit (PSU^s) — and the section 4 describes the procedure of selection of second stage Households – Secondary Sampling Unit (SSU^s). The section 5 & 6 provide the procedure (rule) of estimation of persons involved in an activity in each Tehsil and Stratum, respectively and ultimately the sampled population.

2. STRATUM-WISE NUMBER OF TEHSILS & VILLAGES/WARDS/ MAHALAS IN SAMPLED POPULATION

The sampled population is divided in to L-Stratum. There are N_h Tehsils and V_h village in each stratum. These are elaborated in Table 1. The suffixes h, k and j refer to the Stratum, Tehsil and Village, respectively.

TABLE-1

Stratum	N_h (Tehsils)	V_h (Villages)
1	$N_1 = \sum_k N_{1k}$	$V_1 = \sum_i V_{1i}$
2	$N_2 = \sum_k N_{2k}$	$V_2 = \sum_j V_{2j}$
3	$N_3 = \sum_k N_{3k}$	$V_3 = \sum_l V_{3l}$
.	.	.
.	.	.
.	.	.
L	$N_L = \sum_k N_{Lk}$	$V_L = \sum_j V_{Lj}$
	$\sum N_h$	$\sum V_h$

$\sum_h N_h$ = Total number of Tehsils in sampled population.

$\sum_h V_h$ = Total number of Villages in sampled population.

3. FIRST STAGE SAMPLE SELECTION OF VILLAGES/WARDS/MAHALAS FROM TEHSILS of SAMPLED POPULATION

3.1 Notation

The suffixes h, k, j and i refer to stratum, Tehsil, Village and Households, respectively.

$N_h = \sum_k N_{hk}$ Number of Tehsils

Z_h The stratum wise proportion of an Activity: $\sum Z_h = 1$

P_l Proportion of active Villages

δ Level of confidence

Z Confidence co-efficient

ϵ	Minimum acceptable error
n	$= Z^2 P_1 (1 - P_1) / \epsilon^2$ The Primary Sampling Units in Sample (PSU'S) i.e. Village/Wards/Mahalas Or
n	$> \frac{P_1 (1 - P_1)}{\delta \epsilon^2}$ (Weak Law of Large numbers)
$n = \sum_h \sum_k n_{hk}$	Total number of villages in a sample
$n_h = \sum_k n_{hk}$	Number of Sample villages in stratum h
n_{hk}	Number of Sample villages in stratum h and Tehsil k
$V_h = \sum_i V_{hi}$	Number of Villages
$V = \sum_h V_h$	Number of villages in Sampled Population
n/V	Ratio of the First Stage Sample to the total number of Village/Wards/Mahals in the sampled population

3.2 Stratum Wise Villages/Wards/Mahalas In A Sample By Using Density Of An Activity As Stratum Weights

By using the density of activity as stratum weights (Z_h) the PSU's are allocated to each stratum. The Table 2 elaborates the procedure.

TABLE. 2

Stratum	N_h	Z_h	V_h	$n_h = n.Z_h$
1	N_1	Z_1	V_1	$n_1 = n.Z_1$
2	N_2	Z_2	V_2	$n_2 = n.Z_2$
3	N_3	Z_3	V_3	$n_3 = n.Z_3$
.
.
.
L	N_L	Z_L	V_L	$n_L = n.Z_L$

Remarks: The n_h are selected from V_h by using simple random or systematic or any other suitable sampling technique.

4. SECOND STAGE SAMPLE SELECTION OF HOUSEHOLDS FROM VILLAGES/WARDS/MAHALAS OF FIRST STAGE SAMPLE

4.1 Notation

The suffixes h, k, j and l refer to Stratum, Tehsil, Village and Households, respectively.

$$M = \sum_h \sum_k \sum_j M_{hjk}$$

$$M_{hk} = \sum_l M_{hkl}$$

$$M_h = \sum_k M_{hk}$$

Number of households in stratum h

$$M = \sum_h M_h$$

Total Number of Households in sample

(The Secondary Sampling Units SSU'S)

P_{ll} The proportion of an active Households is sampled Villages

ϵ Minimum acceptable error

δ Level of Confidence

Z Confidence - Coefficient

$M = Z^2 P_{ll} (1 - P_{ll}) / \epsilon^2$ = the Secondary Sampling Units in Sample (SSU'S)
or

$M > \frac{P_{ll} (1 - P_{ll})}{\delta \epsilon^2}$ (Weak Law of Large numbers)

4.2 Stratum-Wise Households In A Sample By Using Density Of An Activity As Stratum Weights

Again density of activity is used to allocate the M_h households from the n_h villages in the sample. The Table 3 elaborates the procedure.

TABLE-3

Stratum	n_h	$M_h = MZ_h$
1	n_1	$M_1 = MZ_1$
2	n_2	$M_2 = MZ_2$
3	n_3	$M_3 = MZ_3$
.	.	.
.	.	.
.	.	.
L	n_L	$M_L = MZ_L$
	$\sum_h n_h = n$	$\sum_h M_h = M$

Remarks: M_h households are selected from households of n_h villages in the sample by simple random or systematic or any other suitable sampling technique.

5. ESTIMATION OF SAMPLED POPULATION TOTAL IN EACH TEHSIL

The number of persons involved in an activity in a Tehsil of sampled population are estimated by using the following procedure.

The suffixes are referred to stratum h , Tehsil k , village j and household i , respectively.

Y_{hkji} Number of active persons in j -th village and i -th Household in a sample

$Y_{hk} = \sum_j \sum_i Y_{hkji}$ Number of active persons in sample

M_{hkj} Number of active Households in the j -th village of sample

M_{hk} Number of active households in sample

M_{hk}^* Total number of Households in a sample

$R_{hk} = M_{hk} / M_{hk}^*$ Ratio of active Households to the total number of Households in sample

W_{hk}^* Total number of Households in sampled population

$W_{hk} = W_{hk}^* X R_{hk}$ Number of active Households in sampled population

$Y_{hk}^* = W_{hk} Y_{hk} / M_{hk}$ Estimated number of active persons involved in an activity in sampled population

6. ESTIMATION OF SAMPLED POPULATION TOTAL IN EACH STRATUM

The number of persons involved in an activity in the h-th stratum of the sampled population is estimated by using the following procedure.

The suffixes h and k represent the stratum and Tehsil, respectively.

Y_{hk}^* Estimated number of active persons involved in an activity in k-th Tehsil of the sampled population

$Y_h^* = \sum_k Y_{hk}^*$ Estimated number of active persons involved in an activity in h-th stratum of the sampled population

7. ESTIMATION OF SAMPLED POPULATION TOTAL

The suffixes h and k represent the stratum and Tehsil, respectively.

$Y^* = \sum_h Y_h^*$ Estimated number of active persons involved in an activity in the sampled population

ACKNOWLEDGEMENT

This paper was presented in the "First National Conference on Statistical Sciences-2003" organized by the Department of Statistics, Government College University, Lahore. I am thankful to Dr. Khalid Aftab, Vice Chancellor, GC University and Prof. Dr. Muhammad Akram Kashmiri, Dean, Faculty of Science, GC University who were source of inspiration behind the conference.

RECOMMENDED RESEARCH PLAN FOR NOT EVENLY SPREAD POPULATIONS

1. Literature Review of Sampling Strategy adopted in various surveys.
2. Comparative Study: merits and demerits of various strategies.
3. Critical view about hypotheses tested and estimators developed in previous studies.
4. Coverage of the population to be sampled.
5. Decision about the hypotheses to be tested and parameters to be estimated.

-
6. Preparation of the questionnaire.
 7. Developing of an improved sampling strategy.
 8. Developing of estimators of the parameters and their variances.
 9. Testing of hypotheses.
 10. Conclusions.