

Proposed Methodology of Area Frame Sampling using Unequal Probability Sampling technique for Crops' Surveys: A Case Study in Pakistan

Abstract:

The Crops' estimates have been great concerned by the Government of Punjab (Pakistan) all the times. For the purpose, Crop Reporting Service (CRS), Agriculture Department Punjab has been evolving, using and analyzing sampling technique, List Frame Sampling (LFS), for conducting surveys to gather information regarding crop acreage, cost of production, grain yield and production, and other agricultural items since 1978. In this paper, the current method (LFS) of crop area surveys in Pakistan has been illustrated and evaluated. However, the development of the rural economy in Pakistan brings new problems and challenges to this method. Based on this, a new approach of Area Frame Sampling (AFS) has been proposed, which will be more effective and fitting to cope with the existing problems of crops' estimates. The results show that the Area Frame Sampling method has many advantages over existing List Frame Sampling method. As most of the research papers focus on the method of Area Frame Sampling through GIS technique. But in this paper two-stage statistical sampling technique has been used to achieve the same objective in an efficient and economical way.

Keywords: List Frame Sampling (LFS), Area Frame Sampling (AFS), Probability Proportional to Size (PPS).

1. Introduction

Statistics deals with a sample data and Crop Reporting Service, Punjab being the largest network of statistics in Pakistan is extremely dependent on an appropriate sample of villages, which should cover all possible dimensions of population so that reliable estimates of area and yield of different crops can be obtained. For this purpose, different techniques of sample selection are used to select a representative sample from a population. In this section, a brief description of existing sampling frame and detailed description of proposed sampling frame and design have been discussed. It is important to note that we will not discuss sampling design of existing sampling frame rather a brief description of frame will be discussed.

1.1. List Frame Sampling

The sampling frame, which has been using in Crop Reporting Service Punjab since 1978, is List Frame Sampling. The current sample being exercised in CRS is the 4th one i.e. sample has been revised approximately after each decade. The first sample of 550 villages was selected in 1978 at district level which was only for wheat crop area. The second and third sample were also selected at district level in 1988 and 2004 respectively for five major crops (Wheat, Rice, Cotton, Maize, and Sugarcane) and the total number of villages in second and third sample was 1010 and 1100 respectively. In the first three samples District was used as a cluster but with the passage of time, increase of population and change in the cropping pattern; more and more districts and tehsils have been established. In the last 4th sample, total no of villages was same as third sample but this sample was selected on the basis of Tehsils of the Punjab to minimize the sampling error and to narrow down the data scale. Using List Frame of villages in the specific cluster (Districts in the first three samples and Tehsil in the fourth sample), Systematic Random Sampling design has been used to get a sample of required number of villages

An administrative breakup of the province for the better understanding of sampling of villages in the Punjab is given in this article (See Appendix). Administrative breakup shows that tehsil level villages do not cover the whole district in all dimensions of crops. It is important to highlight that in the running sample, Sampling Unit (SU) is a complete village whatever the area of the village. Mostly area of the sample villages vary from 500 to 4000 acres. But above 4000 acres, sub-sampling of 2500 acres is allowed i.e. a numerator (Crop Reporter) has to cover 500 to 3900 acres of a village completely and consequently causes an increase in the non-sampling error. It is also important to highlight that within a village distribution / varieties of crops and cropping pattern are almost same. So acreage and yield surveys of the whole village cause redundancy in the data because of homogeneity in agro parameters with in the village.

1.2. Area Frame Sampling

An Area Frame Sampling (AFS) is a special case of cluster sampling. The Sampling Units (SUs) are areas of land commonly called **Segments** or **Parcels**. The objective is to divide the entire available land area (Population area) into SUs and then to select a sample of such segments of required number using some suitable sampling technique. An Area Frame Sampling (AFS) is extremely suitable Sampling Frame for general purpose sampling and the best one in case of land

associated variables like crops, yield, orchards, forests, live stocks and geographical related surveys.

Area Frame Sampling has many advantages over List Frame Sampling. It has better representation of crops in the sample as compared to List Frame Sampling. As there is extensive mixed & scattered cropping pattern, especially in case of minor crops, in the Punjab, so List Frame Sampling does not cover such pattern rather Area Frame Sampling represents the area in the more appropriate way.

2. Proposed Methodology

As discussed earlier List Frame (LF) and Area Frame (AF) are two available frames for the selection of a sample of villages / segments for crops' surveys using suitable sampling techniques. There are two approaches remote sensing and statistical sampling techniques of sample selection using AFS. But here we discussed only statistical sampling techniques.

2.1. Segment's Frequency in a Union Council

There is only ONE segment selection from a village as there is a lot of homogeneity in cropping pattern and other agro based parameters with in a village so it is redundant to select more than one unit from a village. Each Union Council (UC) comprises a specific number of villages and selection of number of segments is subject to the total cropped area (acres) of the UC. The following rules are used to decide the number of segments to be selected from a UC,

Cropped area	No of Segment in a UC
Less than 5000	1
Less than 5000	2
Less than 5000	3
Above 15,000	4

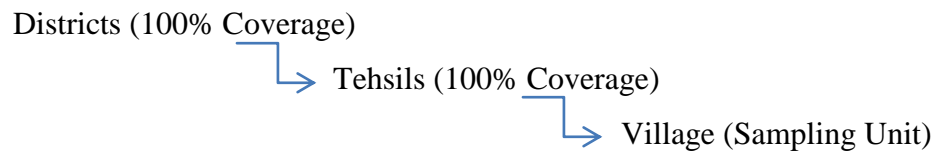
Also it is dependent on the available resources, staff strength and financial constraints of a particular district. By applying the above rules, sufficient number of segments (not more than 5 segments per Crop Reporter) is selected from each district of the Punjab.

2.2. Sampling Designs Using Area Frame

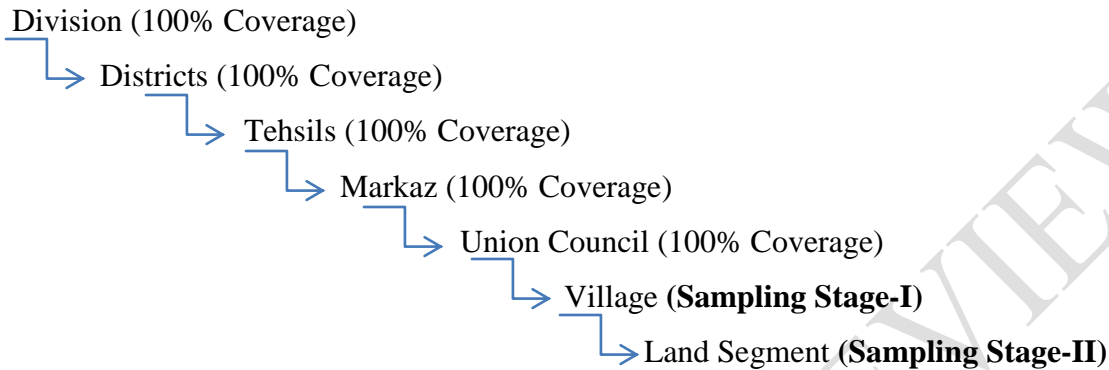
For better understanding and suitability of sampling design, the population's layers shown in List Frame Sampling is as

Division (100% Coverage)





On the other hand the population's layers in proposed Area Frame sampling technique are as



Two stage sampling techniques are being used in Area Frame Sampling. In this section two stage sampling techniques have been discussed. There is no sampling up to Union Council (UC) level i.e. representation of each UC is must in the sample. Sampling techniques have been used at village level and within a village at Land Segment level.

(a) Sampling Technique at Stage-I:

At Stage-I, a set of all villages of a Union Council (UC) is the population from which we have to select a sample of villages by using a suitable sampling technique. Each Union Council (UC) comprises 1 to maximum 25 villages with different areas of acreage. Some villages are significantly large and some are small. Also some villages have more area but less being used for cultivation i.e. comparatively small cropped area of a village. This fact leads to the need of implementing **Probability Proportional to Size Sampling (PPS) / Unequal Probability Sampling** i.e. villages have more cultivated area are more likely to be selected in the sample.

(b) Sampling Technique at Stage-II:

After selection of a village at Stage-I, the available land of village being used for cultivation is the population at Stage-II from which one Segment of size 150 acres is to be sampled or selected i.e. available total land of the village is broken in to all possible Segments (Primary Sampling Unit) and one Segment is selected by using the technique of Simple Random Sampling (SRS)

because there is a lot of homogeneity in all agro based parameters and cropping pattern with in a village. So SRS is the best choice as a sampling technique at Stage-II.

2.3 Implementing both Sampling Techniques

In this section, we briefly describe the introduction to Probability Proportional to Size sampling at Stage-I, Simple Random Sampling at Stage-II, sequence of steps in the selection of a sample and finally a case study of sample selection in the district Lodhran, province Punjab.

(a) Probability Proportional to Size Sampling (PPS) Stage-I

In Probability Proportional to Size sampling technique ith element of population has a specific probability of selection P_i in the sample subject to some particular measure of size Z_i such that ,

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

Where,

Z_i : Total cropped area of ith village of a U.C.

P_i : Probability of selection of ith village in the sample.

Obviously for all villages of a Union Council (UC),

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

The following steps are applied in selecting a sample using Probability Proportional to Size sampling (PPS) with in a Union Council (UC) for the selection of sampling units i.e. villages,

- Calculate P_i for each village of a UC.
- Multiply P_i by 100 to get a whole number. ($P_i \times 100$)
- Construct Cumulative Probability Totals for each village.
- Construct Ranges for each Cumulative Probability starting from '0' to one less concerned Cumulative Probability Total.
- All Ranges fall within 0 to 99.
- Use 2-Digit Fisher's Random No. to select a sample of required size.

(b) Simple Random Sampling (SRS) Stage-II

In Simple Random Sampling (SRS), each and every element of population has equal and known probability of selection in the sample i.e. no sampling unit is more or less favorable. It is the best technique (in terms of un-biasedness) in case of homogeneity in the variable of interest.

The following steps are applied in selecting a sample using Simple Random Sampling (SRS) with in land of a selected village for the selection of sampling unit i.e. segment,

- Split the area of a sample village in to a Grid of Squares (1 square = 25 acres) from North to South starting No.1 from North West-Corner of the village.
- A map of village’s land designed by Board of Revenue (BoR), Government of the Punjab, is already available for each village of the Punjab

A village with Squares distribution				N					
1	2	3	4	5	6	7	8		
16	15	14	13	12	11	10	9		
W	17	18	19	20	21	22	23	24	E
	32	31	30	29	28	27	26	25	
	33	34	35	36	37	38	39	40	
				S					

Fig: 1-A village with square distribution

- Select a number between 1 and Total Area of the sample village (in squares) X (say).
- Select a Fisher’s Random No. from 1 to X,

<u>Ran#</u>	<u>Village’s Area</u>
2-Digit	Max. up to 99 squares
3-Digit	Max. up to 999 squares
4-Digit	Max. up to 9999 squares

- The selected Random No. is the selected sample Square and at the NW-Corner of the selected Square, take adjacent **SIX SQUARES** as a sample Land Segment.

3. Experimental Area Frame Sample Selection

3.1. Sample Selection of Final Segment

In this section a complete example of selection of a Land Segment in the tehsil Karorpacka of District Lodhran, Punjab is discussed in detail. District Lodhran consists of three Tehsils. Details are shown in the Table 1,

Tehsils	No. of Markaz	No. of UCs	No. of Villages	Total Area (acres)	Total Cropped Area (acres)
Lodhran	2	24	146	267,053	221,868
Dunypur	2	14	192	239,847	210,626
Karorpacca	2	20	102	184,482	160,468
Total	6	58	440	691,382	592,962

Table 1: Detail of District Lodhran, Punjab

Taking an example of Tehsil Karorpacka, Markaz Karorpacka, this Markaz consists of 10 Union Councils (UCs) and 54 villages. Taking an example of Gai Mummar, one of the 10 UCs of the Markaz, this UC consists of 12 villages. As UC is our last population layer, we implement two-stage sampling on this UC level. Details of all villages of the UC, their total area (acres) and cropped area (acres) are shown in the Fig: 2.

District	Lodhran				
Tehsil	Karorpacka				
Markaz	Union Council	S.No	Village	Total Area	Cropped Area
Karorpacka	Gai Mummar	1	Gahi Mummar	792	679
		2	Jalil Pur	446	414
		3	Darwaish wala	265	237
		4	Mahaar	381	361
		5	Dhora Mahaar	252	224
		6	Choki Sobay Khan	439	345
		7	Jambooo Jat	871	770
		8	Chak Choki Kohna	443	411
		9	Choki Kohna	389	344
		10	Gulhar	509	464
		11	Kambooo	642	614
		12	Rindjada	3029	2663

Table 2: Villages, Total Area, Cropped Area of District Lodhran

As it is shown in the table 2, there are 12 villages in the UC, Gai Mummar with different cropped areas. The village, Rindjada, has the maximum cropped area i.e. 2663 acres and the most prominent village of the UC. It leads the need of Unequal Probability Sampling i.e. prominent villages of a UC must have higher probability of selection in the sample.

We have to select 4 Segments from this UC i.e. we have to select 4 villages (one Segment per village) subject to the size (cropped area) of the village. For this purpose Probability of Selection P_i is calculated for each village of the UC and implementing all steps mentioned in the Section VII-4-2 (a), a sample of 4 villages (mentioned in red color) is selected as shown in the table 2.

For selection of Random Numbers, Fisher's Random Numbers Table has been used and column selection is subject to the number of letters in the Tehsil under consideration. Like 'Karorpacka' has ten letters and column No.10 of 2-Digit Random Nos. has been used.

District	Lodhran									
Tehsil	Karorpacka									
Markaz	Union Council	S.No	Village	Total Area	Cropped Area	Pi	100xPi	Cumulative Totals	Range	Random No
Karorpacka	Gahi Mummar	1	Gahi Mummar	792	679	0.09	9	9	00-08	5
		2	Jalil Pur	445	414	0.06	6	15	09-14	
		3	Darwaish wala	265	237	0.03	3	18	15-17	
		4	Mahaar	381	361	0.05	5	23	18-22	
		5	Dhora Mahaar	252	224	0.03	3	26	23-24	
		6	Choki Sobay Khan	439	345	0.05	5	31	25-30	
		7	Jambo Jat	871	770	0.10	10	41	31-40	
		8	Chak Choki Kohna	443	411	0.05	5	46	41-45	
		9	Choki Kohna	389	344	0.05	5	51	46-50	
		10	Gulhar	509	464	0.06	6	57	51-56	55
		11	Kambo	642	614	0.08	8	65	57-64	58
		12	Rindjada	3029	2663	0.35	35	100	65-99	79

Fig: 2-Probability Selection of each village at UC (Stage-I)

Because of Probability Proportional to Size Sampling (PPS), villages have more cropped area have been selected in the sample. It is a completion of Stage-I sampling i.e. sampling of villages in a UC.

At the Stage-II, Land Segments (150 acres) from the selected villages (at the Stage-I) are selected. Implementing all steps mentioned in given below, a sample of a Land Segment is selected from each sample village.

Taking an example of first village Gahi Mummar, It is a village of 32 squares. Selecting a Random No. from 1 to 32, Random No. 15 is selected (using 10th column of Fisher Random Numbers Table) as shown in the Fig.3 Same process is repeated for other three villages Gulhar, Kamboo and Rindjada.

District		Lodhran						
Tehsil		Karorpacka						
Markaz	Union Council	S.No	Village	Total Area	Cropped Area	Random No	Total Area Squares	Selected Square No.
Karorpacka	Gahi Mummar	1	Gahi Mummar	792	679	5	32	15
		2	Jalil Pur	446	414		18	
		3	Darwaish wala	265	237		11	
		4	Mahaar	381	361		15	
		5	Dhora Mahaar	252	224		10	
		6	Choki Sobay Khan	439	345		18	
		7	Jamboo Jat	871	770		35	
		8	Chak Choki Kohna	443	411		18	
		9	Choki Kohna	389	344		16	
		10	Gulhar	509	464	55	20	18
		11	Kambooo	642	614	58	26	9
		12	Rindjada	3029	2663	79	121	66

Using 10th column of 2-Digit Fisher Random Nos Table, from 1 to 32, Ren# 15 is selected.

Select NW-Corner of the selected Square & take adjacent 6 (150 acre) Squares as sample Segment.

Using 10th column of 3-Digit Fisher Random Nos Table, from 1 to 121, Ren# 66 is selected.

Fig: 3- Probability Selection of each village at UC (Stage-II)

At the NW-Corner of square No.15, take upper & lower sides two adjacent squares as a sample **Land Segment** as shown in the Fig. 3.

		N							
		1	2	3	4	5	6	7	8
		16	15	14	13	12	11	10	9
W		17	18	19	20	21	22	23	24
		32	31	30	29	28	27	26	25
		33	34	35	36	37	38	39	40
		S							

NW-Corner of selected Square # 15, take Up & Down two adjacent Squares as a Segment

Locating a Segment in the sample village

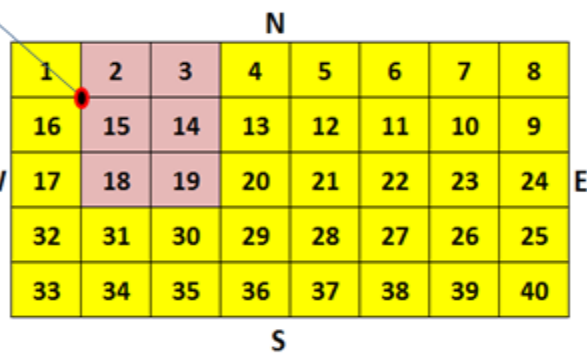


Fig: 4- Corner Selection Table

The following Fig. 4 shows different expected locations of a Land Segment in case of selection of different Random Nos. between 1 and 40 (e.g. Random No. 40, 9, 6, 37)

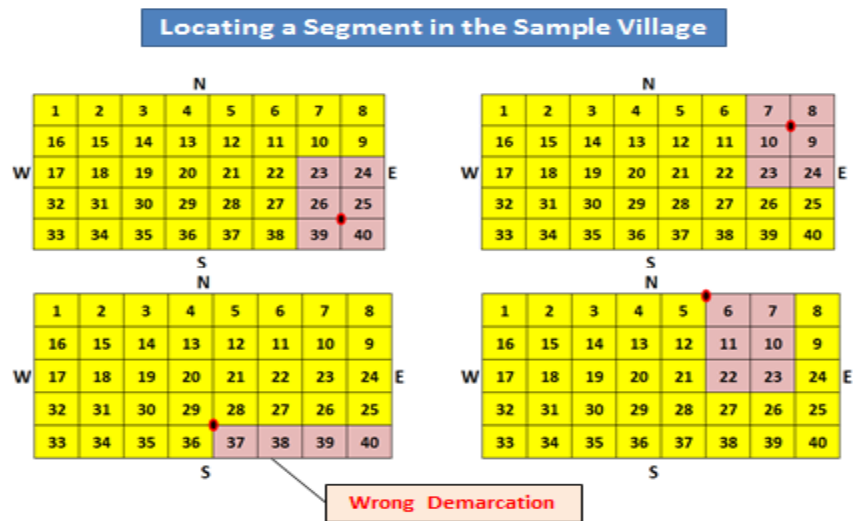


Fig: 5: Locating a Segment in the Sample Village

Segment of 6 squares is not allowed in a line as mentioned in the Fig.5 i.e. Segment in a rectangular shape is compulsory as mentioned in the other cases as shown in the said Figure. Sampling at both stages reveals that there is no discretion of researcher in the selection of the sample rather it is totally randomized and, obviously, returns an unbiased picture of population. The same methodology is repeated for each Union Council (UC) of the Punjab and ultimately a random sample of required size is selected on the provincial level using Area Frame List, which, obviously, returns a better, unbiased and more representative picture of population at the least cost.

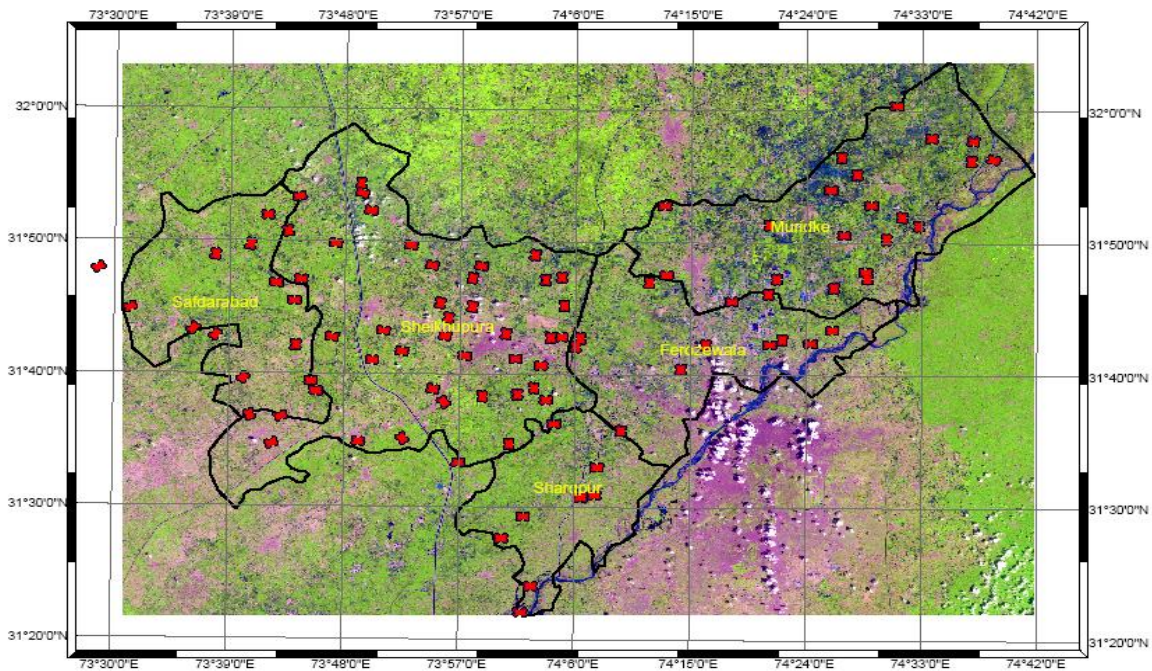


Fig: 6: Geo Tagged Location of all Segments in the district Sheikupura, Punjab

3.2 Yield Estimation in a Land Segment Using AFS

After area assessment of crops another important assessment is of yield of crops, which is included in the main responsibilities of Crop Reporting Service (CRS), Punjab.

As mentioned earlier, in each sample village six plots of size 15x20 ft² in three randomly selected fields are harvested for the yield estimation in case of List Frame Sampling. But in case of Area Frame Sampling, number of Land Segment is significantly higher than the number of sample villages using List Frame Sampling, so experimental plots of each size 6x8 ft² would be harvested for each crop as shown in the Fig.3-7.

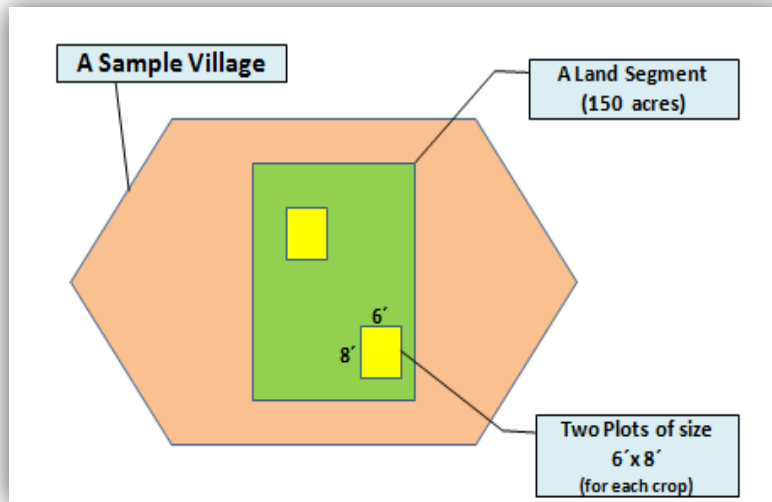


Fig. 7-Land Segment

Obviously, as compared to 15x20 ft² experimental plot, handling of 6x8 ft² plot is easier, gives least non-sampling error and the best results in terms of least Mean Square Error (MSE) among three plot sizes 6x8, 12x16, and 15x20.

4. Conclusions

Most of the research papers focus on the method of Area Frame Sampling through GIS technique i.e. the required land as a target population is divided in to a grid of Primary Sampling Units (PSUs) of equal sizes then by using Simple Random Sampling technique; a sample of required number of segments is selected. But in this paper, a part of GIS technique, two-stage statistical sampling technique has been used for the land of a Union Council (the smallest cluster of a district) subject to the condition that a Union Council has more cultivated area has more number of segments with in itself.

It would also be helpful in case of Crop Insurance Project based on the concept of AYII (Area Yield Index Insurance) to narrow down the unit Area of insurance from tehsil to Union Council to increase homogeneity in the average yield of a crop within the unit area of insurance.

References

1. Cochran, G. W. Sampling Techniques, 3rd ed. Harvard University: John Wiley & Sons; 1977
2. Pan Y, Wang M, Wei G, Wei F, Shi K, Li L, Sun G. Application of area frame sampling for agricultural statistics in China.
3. Cotter J, Davies C, Roberts R. Area frame design for agricultural surveys. National Agricultural Statistics Service 2009.
4. Gallego, F. J., Delincé, J., & Carfagna, E. (1994). Two stage area frame sampling on square segments for farm surveys. *Survey Methodology*, 20(2), 107-115.
5. Gallego, F. J. (1995). *Sampling frames of square segments*. Office for Official Publ. of the European Communities.
6. Cotter, J., Davies, C., Nealon, J., & Roberts, R. (2010). Area frame design for agricultural surveys. *Agricultural survey methods*, 169-192.
7. Fecso, R., Tortora, R. D., & Vogel, F. A. (1986). Sampling Frames for Agriculture in the. *Journal of Official Statistics*, 2(3), 279-292.
8. Delincé, J. (2001, June). A European approach to area frame survey. In *Proceedings of the Conference on Agricultural and Environmental Statistical Applications in Rome (CAESAR)*(Vol. 2, pp. 463-472).

APPENDIX

Demerits of List Frame Sampling

An administrative breakup of the province for the better understanding of sampling of villages in the Punjab is given below table. Total no of villages, cultivated area, sample area, proportion of sample area is 26275, 30.7 million (acre), 1.5 million (acre), and 4.9 % (Total cultivable area of Punjab) in Punjab respectively.

Unit	Count	Avg. No. of Villages
Division	9	2920
Districts	36	730
Tehsils	147	181
Markaz	354	74
Union Council	2797	10

UNDER PEER REVIEW