Use of Adaptive Cluster Sampling for Identifying the Land Producing Non Hybrid Crops

Submitted by
Saiful Alam Chowdhury
Candidate for M.Phil. in Applied Statistics
Reg. No.: 447
Session: 2004-2005

Supervised by
Dr. Syed Shahadat Hossain
Professor of ISRT
University of Dhaka

A thesis submitted in partial fulfillment of the requirement for the degree of M.Phil. in Applied Statistics

Institute of Statistical Research and Training (ISRT)
University of Dhaka

June 2014

Supervisor's Declaration

I clarify that the thesis entitled "Use of Adaptive Cluster Sampling for Identifying the Land Producing Non Hybrid Crops" submitted as a partial requirement for the degree of M.Phil. in Applied Statistics is the result of Saiful Alam Chowdhury's own research, except where otherwise acknowledged, and that this thesis is whole or in part has not been submitted for an award, including higher degree, to any other university or institution.

(Syed Shahadat Hossain, Ph.D.)
Professor of ISRT
University of Dhaka

Dedicated

To

My Parents

Acknowledgement

First of all, I express my solemn gratitude to Almighty Allah, the merciful, who has given me the ability, strength and opportunity to perform this thesis work. I would like to convey my heartiest gratitude and thanks to my supervisor Dr. Syed Shahadat Hossain, Professor, Institute of Statistical Research and Training, University of Dhaka, for his excellent supervision, encouragement, suggestions and assistance for the completion of this thesis. Without his constant guidance, it would not have been possible for me to complete this thesis work.

I would like to pay my gratitude to Director, Institute of Statistical Research and Training, University of Dhaka, for giving me the permission to undertake the thesis and providing me the computer facilities of this institute. Also I would like to take this opportunity to express my appreciation of the help and support given to me by all the academic and administrative staff of the Institute of Statistical Research and Training, University of Dhaka.

I would like to express my cordial thanks to Mohammad Shahed Masud, Associate Professor and Mr. Paritosh Kumar Roy, Lecturer of Institute of Statistical Research and Training, University of Dhaka for their helpful suggestions and kind cooperation. I would also like to thank Mr. Mohammad Samsul Alam, then MS student of ISRT for providing help and support during the preparation of the thesis.

I would like to convey my warm thanks to the Principal, B N College Dhaka for his helpful suggestions and kind cooperation.

Thanks are due to my family, my wife's family and all of my relatives, friends and well-wishers who kept supporting and encouraging me throughout my study period. Most of all, I would like to thank my uncle (Sarwar Alam Chowdhury), aunty (Nurjahan Nazneen), elder sister, my wife and my little sons Musaab and Suhaib for their love and inspiration.

Saiful Alam Chowdhury

Abbreviation

ACS	Adaptive Cluster Sampling
AgL HH	Agricultural Labor Household
AOSN	Autonomous Ocean Sampling Network
APRB	Absolute Percentage Relative Bias
AS	Adaptive Sampling
ASAP	Adaptive Sampling and Prediction
ATS	Adaptive Two-phase Sampling
ATSS	Adaptive Two-stage Sequential Sampling
BBS	Bangladesh Bureau of Statistics
BSEC	Balanced Sampling Excluding Contiguous Units
CRS	Capture-Recapture Sampling
EA	Enumeration area
FOA	Food and Agriculture Organization
GDP	Gross Domestic Product
HB Boro	Hybrid Boro
HH	Household
LIS	Line Intercept Sampling
LTS	Line Transect Sampling
MC mean	Monte Carlo mean
MC se	Monte Carlo standard error
MSE	Mean Squared Error
NS	Network Sampling
PEC	Post Enumeration Check
PSU	Primary Sampling Units
RE	Relative Efficiency
RSS	Ranked Set Sampling
SALT	Selection at list time
SRS	Simple Random Sampling
StRS	Stratified Random Sampling
TSACS	Two Stage Adaptive Cluster Sampling
USS	Ultimate Sample Size
WOR	Without Replacement
WR	With Replacement

Abstract

Use of Adaptive Cluster Sampling for Identifying the Land Producing Non Hybrid Crops

Saiful Alam Chowdhury

Candidate for M.Phil. in Applied Statistics

For estimating rare and cluster characteristics the conventional sampling methods are hard very difficult to be used, instead an alternative method like adapative cluster sampling is thought to be appropriate for in such situations. Coverage of *Hybrid Boro* usage in Bangladesh is rare and as well as of cluster pattern, so an Adaptive cluster sampling may be appropriate in estimating proportion of *Hybrid Boro* use. The applicability of the adaptive cluster sampling for this purpose, that is why, is planned to be visualized by a simulation consisting re-sampling of the Agriculture census (2008) data.

Investigation of the suitability of the adaptive cluster sampling method for estimating the proportion of land producing non-hybrid crops is the prime objective of the study. Since only household (HH) level data are available (Agricultural Census, 2008), the proportion of HH producing non-hybrid (having the same sense of proportion of HH producing hybrid) was under investigation. The specific objectives of the study included to find an estimate of the proportion of HH cultivating *Hybrid Boro* using simple random sampling and adaptive cluster sampling methods. Also of the interest was to obtain bias and standard error of the estimators for each of the methods using extensive simulation studies and to compare them. The simulation study considered different small sample sizes, namely 100, 200, 300 and 500. The choice of sample size is made arbitrarily keeping in the sense that adaptive cluster sampling is more profitable for smaller sample sizes.

The Monte Carlo absolute percentage relative bias and Monte Carlo standard error of the estimators were calculated for each of the methods for each of the sample sizes. The major findings of the study compared in terms of

Monte Carlo absolute percentage relative bias and Monte Carlo standard error revealed that the estimator of the proportion of HH cultivating *Hybrid Boro* using adpative cluster sampling method has higher variance and lower bias than simple random sampling has for all the sample sizes considered in this study. The ultimate sample size realized by the application of the adaptive cluster sampling were also recorded and it is seen that the average ultimate sample size is about 10 to 20 percent higher than the initial sample sizes.

The most interesting finding of using an adaptive cluster sampling method was seen to be its strength of capturing more information. In estimation of proportion of HH cultivating *Hybrid Boro*, it has been revealed from the simulation that the adaptive cluster sampling method is way far better than the simple random sampling method in terms of chances of avoiding a bad sample containing very small number of targetted characteristics. For simple random sampling method such risk is higher for divisions with smaller true population proportion. The findings can be triangulated to the issue that the simple random sampling method may produce more bias than the adaptive cluster sampling method.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bangladesh is mainly an agricultural country. Most of the people are engaged and depends on agriculture even in this urbanization period. There are different types of short term and long term crops are cultivated in most of the cultivable land around the whole year. It is small but one of the most populated country in the world. Our land is too limited compared to our large population. The main food of Bangladeshis is rice. Different types of paddy are cultivated in different season in this country. Now-a-days our farmer uses bio-technically powered varieties of foreign seed named hybrid seed to meet the food demand of this large population. One of the mainly used varieties of hybrid paddy seed by Bangladeshi farmer is hybrid boro. The ratio of agriculture household cultivating hybrid paddy seed to that of cultivating non-hybrid seed is too small. Hybrid seed is harmful for land fertility. Because these bio-technically powered varieties of foreign seeds demand more fertilizers, pesticides and water. And these seeds are not developed to thrive in local climate conditions. So it is of huge interest for policy-maker of different field in multi-dimensional aspects to estimate the proportion of agriculture household who cultivates hybrid seed. The main tool for such information is statistically sound data.

1.1 Sample Survey

For collecting statistical primary data, the main two approaches are sample survey and census. In census each and every element of the population is enumerated. But in sample survey only a representative part of the population is considered. Census is often conducted after every ten years, since it requires large number of trained fieldworkers as well as high cost in terms of both money and time. Even if one have abundance of resources in money, time and manpower, it is not possible in most real life situation to conduct census like fish survey of a sea, animal survey of a forest, aerial survey of hill area etc. Also census deals with a huge number of data as a result takes a lot of time for compiling, editing, analyzing as well as publishing final report. It may also be noteworthy that in dealing with huge data size, censuses often are in danger of incurring more non-sampling errors. That's why sample survey is usually preferable to census survey. Sample survey may be done in every year since it does not deal with the whole population. In developed countries sample survey of different fields are conducted in every year. But in developing countries like Bangladesh it is not conducted too frequently, even not done in every two or three year. The main and important point in sample survey is representativeness. The more representative sample produce more precise estimate of the parameter i.e. the characteristics of the population. The procedure or technique or method of drawing sample from the population is called *sampling*.

1.2 Sampling

Sampling is a part of statistics, associated with the selection of individual observations, intended to obtain some knowledge about a population of concern, especially for the purpose of statistical inference. The sampling technique is now commonly used for obtaining information on various social and economic activities of society. Local government is making use of sampling to obtain information needed for planning. Market research is heavily depended on sampling approach. Business firms and industry are also use sampling for investigating the efficiency of internal operations. Scholars derive or suggest different types of sampling design depend on diverse situation. Depending on population characteristics we decide which sampling design will be used.

Appropriate selection of design gives more appropriate focus on population. So it is a very important as well as technical task to choose the sampling design. Different types of sampling design are simple random sampling, stratified sampling, systematic sampling, and cluster sampling etc. Usually these types of conventional sampling are used in agriculture research as well as various social, economic and market research.

1.2.1 Sampling for Rare Population

When the characteristics of interest in a sampling is rare, a conventional sampling plan may not be appropriate to use since the coverage of the sampling may fail to capture enough item satisfying the characteristics even with a moderately large sample size. Often in practical situations, researchers encounter such problem. The following examples may give an idea regarding the lack of appropriateness of conventional sampling methods in estimating rare characteristics.

1. Surveys of commercial species of shrimp in the Gulf of Alaska involve trawl sampling, in which a net is towed behind the research vessel at selected locations and estimates of shrimp abundance for the whole study region are made by expanding the amounts caught, taking into account the area swept by the net. The survey covers a vast area and sampling method is very time consuming. Biologists aboard the research vessel found that typically shrimp were concentrated in just a few areas, whereas tows in other areas would produce hardly any shrimp. Yet the regions of concentration changed from survey to survey due to the schooling movements of shrimp, and the concentration areas were not entirely predictable in advanced. The field biologists expressed the desire for a plan in which they could cut down on effort in the low abundance areas. For this reason, instead of conventional sampling design an adaptive sampling design was proposed for rare and cluster pattern population.

2. Spatial patterns of hardwood tree species in northern North America were examined by Roesch (1993), who found that whereas overall forest density appeared fairly random in distribution, the individual species of interest were apt to be rare and highly clustered in their distributions and therefore difficult to sample by using conventional sampling design.

For estimating the prevalence of insect infestation some of these species, Roesch therefore investigated an adaptive sampling in which once a tree of the species was found, an area of specified radius around it would be searched for additional individuals of the species.

3. Household surveys for estimating rare characteristics, such as drug use, infection with a rare disease, or specially consumer purchases, also present difficult sampling and estimation problems. A marketing survey to estimate purchase of disposable diapers in a region of New Zealand ran into difficulty when it was discovered that only 2% of household in the region had children less than two years of age. The distribution of such household was not known but was known to be clustered spatially. Danaher and King (1994) therefore investigated the use of an adaptive cluster sampling design in which a systematic sample of households was augmented with neighboring households in the sample had the rare characteristic.

4. Estimation of contribution of hatchery salmon to an ocean salmon fishery is crucial both for determining the economic effectiveness of hatchery program and to protect wild stocks of fish from ecological displacement. Because proportion may change during the course of fishing season, Geiger (1994) investigated the use of an adaptive Bayes design for estimating the proportion from fish otoliths.

In each of the above situations investigators are dealing with characteristics of populations that for one reason or another are inherently difficult for sampling and estimation. These studies deal with issues of great importance in terms of economics, health, environmental quality, or scientific understanding. In each of the above case the investigators are motivated to search for sampling and estimation methods that go beyond the conventional set of techniques while at the same time using the conventional methods to the greatest advantage possible.

For dealing with rare cases the available procedure are link-tracing designs such as snowball sampling, random walk methods, and network sampling along with adaptive allocation and adaptive cluster sampling. Among of these special sampling designs for rare characteristics adaptive sampling can be an option for our research. Adaptive cluster sampling can be used in situations where the population is rare, geographically located and have a cluster pattern.

1.2.2 *Hybrid Boro* Use in Bangladesh

The use of *Hybrid boro* although being used by farmers in Bangladesh, is still a rare scenario in context of the whole country. *Hybrid boro* use in Bangladesh is thought to have a cluster pattern with respect to both social context and land fertility context. The social context being the issue that farmers often choose seeds and crops for cultivation by learning it from their neighbours and relatives. So clearly, *Hybrid Boro* usage would form a geographically cluster pattern. The fertility context for this is quite evident because the fertility level of land of Bangladesh is definitely geographically sparse and land suitable for cultivation of *Hybrid Boro* would have a cluster pattern. It can thus be thought that the neighbouring households or neighbouring farmers are expected have similar characteristics regarding usage of *Hybrid Boro*.

As mentioned earlier, estimation of rare and cluster characteristics are difficult through conventional sampling, and for such situations, an alternative method like adaptive cluster sampling can be applied. Estimation of the coverage of *Hybrid Boro* usage in Bangladesh also would, therefore, need special sampling approach because it is rare and as well as of cluster pattern. Adaptive cluster sampling may be an appropriate choice in estimating the proportion of *Hybrid Boro* use. It can be observed from the literature that adaptive cluster sampling is so far not being considered by any of the surveys conducted in Bangladesh. Conducting such a survey is beyond the scope or objectives of this thesis. However, the applicability of the adaptive cluster sampling can be visualized by a simulation consisting re-sampling from an established population. The Agriculture census (2008) conducted by Bangladesh Bureau of Statistics (BBS) can be considered as a good candidate for such established population for estimating the proportion of *Hybrid Boro* usage. So a computer simulation can be conducted to demonstrate the appropriateness of the adaptive cluster sampling in estimating *Hybrid Boro* proportion in Bangladesh.

1.3 Objectives of the Study

The main objective of the study is to investigate whether the adaptive cluster sampling is more advantageous for estimating the proportion of land used for producing non-hybrid crops. The study objective has been analysed and matched with the appropriateness of the data. Since the data available (Agricultural Census, 2008) had information at the household (HH) level, the proportion of HH producing non-hybrid crops were under investigation. The challenge of this estimation is the rarity of the usage of hybrid crops, because failure to capture enough HH cultivating hybrid crops would cause failure of estimation of the proportion of the HH cultivating non-hybrid crops. Since the usage of *Hybrid Boro* is the only remarkable hybrid crops used in Bangladesh, we concentrate in estimating the proportion of HH cultivating *Hybrid Boro* crops in different division of Bangladesh. The specific objectives of the study are given in the following:

1. To find an estimate of the proportion of HH cultivating *Hybrid Boro* using simple random sampling.
2. To find an estimate of the proportion of HH cultivating *Hybrid Boro* using adaptive cluster sampling.
3. To find an estimate of the proportion of HH cultivating *Hybrid Boro* using two stage adaptive cluster sampling.
4. To obtain bias and standard error of the estimator for each of the methods using extensive simulation studies.
5. To obtain the ultimate sample size required for use of adaptive cluster sampling using simulation studies.
6. To compare each of the methods in terms of bias, standard error and coverage strength to see whether adaptive cluster sampling has any advantage over the simple random sampling.

1.4 Organization of the Study

This thesis is organized in a smooth arrangement of required background and derived results supported by interpretation and analysis. In Chapter 1, an introduction is given followed by a detailed literature review in Chapter 2. Chapters 3, 4 and 5 describe the underlying methodologies of sampling design, adaptive cluster sampling and simulation respectively. Chapter 6 introduces the data source Agricultural Census, 2008 conducted by BBS and Chapter 7 details the features and description of the data. Chapter 8 produces the theoretical development of adaptive cluster sampling usable for the estimation of the proportion of HH cultivating *Hybrid Boro* in individual division. Detailed description of the simulation study and its results along with interpretation is given in Chapter 9. Chapter 10 concludes the thesis with a discussion of the findings and overall remarks.

Chapter 2

Literature Review

For years, researchers in the behavioral and biomedical sciences have perceived that conventional sampling may not be the best way to collect data on rare and hidden populations. Even conventional sampling approaches that incorporate over sampling do not allow the investigator to take advantage of new information about potential sources of study participants. This may emerge as the investigator begins sampling and, through increased contact with the population, learns more about where it can be found. For these reasons, substantial researchers have turned to unconventional sampling procedures. Sampling designs that can redirect sampling effort during a survey in response to observed values are known as adaptive sampling (e.g. Thompson and Seber, 1995). This method is necessary because before the survey is in progress, investigators would not be able to know what social connections to follow or whom to include in the sample. The use of information gathered during the survey to inform sampling procedures is a key feature that distinguishes adaptive sampling and conventional sampling. In conventional sampling, the sampling design is based entirely on a priori information, and is fixed before the study begins. A researcher using a conventional sampling design would identify the universe of individuals eligible for sampling before any sampling was actually done, and would not add any eligible individuals discovered during the course of the study

Sudman (1972) remarks, ‘it may be efficient to screen a very large sample for a rare population and to use the screening information for future samples'. This attitude embodies much of the thinking about sampling from populations where with rare characteristics. Screening is not really a sampling procedure in its own right. It involves pre-sampling the population to make it easier to find the rare events.

Anderson (1986) explain that when a list or sampling frame exists for the rare population members, then there are no special sampling difficulties and the standard finite sampling methods can be used. With no such list however then special are needed to sample the rare population members. They considered general screening methods for finding the rare members.

Salehi and Seber (1995) suggested a method where initial units for the ACS method are drawn by a two-stage sampling method. In their method, the population is divided in to a number of sub units. At the first stage, a number of primary units are drawn using SRS method. From each of the selected primary units, pre-determined number of sub-units is drawn using a two stage sampling method, and then ACS method is used to select the neighboring sub-units of selected sub-units. They also suggested two types of unbiased estimators of the population mean and discussed their properties.

Thompson (1996) in his paper examined the behavioral characteristics in rare and hidden populations by using the adaptive sampling and graph sampling methods. Methods of adjusting for the non-sampling errors that arise in such studies are also discussed here for both adaptive and conventional designs

Thompson (1996a) also describes the ACS design in which the population is strip sampling or a systematic sampling method according to primary units and subsequent sampling is done in terms secondary units. He proposed several unbiased estimators for the population mean, the ACS design is more efficient than the conventional designs for aggregated population.

Thompson (1996b) described the ACS method based on order statistics. In this method, the value of the variable (say, y) of interest is observed for each of the selected primary units and they are ordered. The neighborhood units are added to the sample only for the units whose y-values exceeds the r^{th} order statistics of y (r is predetermined by the investigators). He used the theories of order statistics to obtain estimator of the population mean.

Christman (1997) considered the efficiencies of two sampling designs for estimating the mean of a fixed finite population. The first is adaptive cluster sampling (ACS) which is designed to adaptively increase sampling effort in the neighborhood of units whose observed value meets some predefined condition. The other is Balanced Sampling Excluding Contiguous Units (BSEC), a conventional design in which neighboring units are deliberately excluded from being sampled under the idea that they provide little new information to the sampling effort. They consider the effect of type of neighborhood, initial sample size, condition for adaptively sampling neighbors, and degree and extent of clustering in the population on the efficiency of ACS relative to BSEC and simple random sampling. Populations having different degrees of clustering are simulated using a modified Neyman Scott process. They compare the design-based variances of two estimators, a Horvitz Thompson-type estimator and a Hansen Hurwitz-type estimator. While the Horvitz Thompson type estimator can have the lowest variance under some of the situations explored, it is also the most sensitive to changes in the conditions. The efficiency of the estimator often comes at the cost of a large effective sample size. In general, ACS is more efficient when the population elements are rare and highly clustered although BSEC designs are generally more efficient for a wider array of combinations of conditions.

Dryver & Thompson (2000) in their article presented unbiased estimators which are easy to compute. The usual design-unbiased estimators in adaptive cluster sampling can be improved using the Rao-Blackwell method by conditioning on the minimal sufficient statistic. However, the resulting estimators are not commonly used because they are complicated to compute. In their paper an easy-to-compute unbiased estimators are presented. These estimators are obtained by conditioning on a statistic that is sufficient but not minimal.

Morrison (2001) describes an adaptive sampling procedure which was developed in order to make more efficient use of participant's time whilst still obtaining a reasonable degree of resolution in the proportional responses. The procedure focuses on boundaries and has some similarities with up-down methods.

Thompson & Collins (2002) in their paper introduces adaptive sampling designs to substance use of the researchers. Adaptive sampling is particularly useful when the population of interest is rare, unevenly distributed, hidden, or hard to reach. In conventional sampling, the sampling design is based entirely on a priori information, and is fixed before the study begins. By contrast, in adaptive sampling, the sampling design adapts based on observations made during the survey. In the present paper several adaptive sampling designs are discussed. Link-tracing designs such as snowball sampling, random walk methods, and network sampling are described, along with adaptive allocation and adaptive cluster sampling. It is stressed that special estimation procedures taking the sampling design into account are needed when adaptive sampling has been used. These procedures yield estimates that are considerably better than conventional estimates. For rare and clustered populations adaptive designs can give substantial gains in efficiency over conventional designs, and for hidden populations link-tracing and other. Adaptive procedures may provide the only practical way to obtain a sample large enough for the study objectives.

Fiorelli, Leonard, Bhatta and Paley (2005) describe co-operative control and adaptive sampling strategies and present results from sea trials with a fleet of autonomous under water gliders in Monterey Bay during August 2003. These sea trials were performed as part of the Autonomous Ocean Sampling Network (AOSN) II project. A central objective of the project is to bring robotic vehicles together with ocean models to improve their ability to observe and predict ocean processes. New cooperative control and adaptive sampling activities are underway as part of the Adaptive Sampling and Prediction (ASAP) project. Sea trials for this project will take place in Monterey Bay in 2006.

Noor, Ishwar and Vasudevan (2006) attempted to set the comparison between the unfamiliar Adaptive Cluster Sampling and Simple Random Sampling. In their paper they compared the estimates of population abundance per quadrate and population totals from SRS and ACS based on data collected from forest floor herpetofaunal. They compared the two sampling methods in terms of estimates of species richness, sampling of rare species, and insights to species habitat associations. For their research they sampled forest floor herpetofaunal communities in a monsoonal rainforest in South India for three consecutive years to evaluate the use of cluster sampling in estimating species composition and density. Their initial experimental design consisted of comprehensive random searches of multiple $25m^2$ quadrates (SRS) for animals. After their initial season they found that most quadrates had zero animals detected and, when encountered, animals were spatially aggregated. To increase sampling efficiency and derive more precise density estimates, they shifted to adaptive cluster sampling (ACS). They compared the relative sampling efficiencies of ACS to SRS and the ability of the 2 methods to detect rare species. Adaptive cluster sampling failed to yield the more precise density estimates as predicted by statistical theory. However, ACS yielded more individual and rare species detections. Their results suggest the ACS assumptions should be carefully evaluated prior to use because it may not be appropriate for all rare, spatially aggregated populations.

In animal or bird survey, it is important problem to detect all the elements in a selected unit of site during observation or the elements may be detected imperfectly. These lead to the non-sampling error. Thompson (1994) considered the case, where the elements may be detected imperfectly in ACS. They also proposed the estimators of population total and estimated variance for the case of constant detectabilities and extended it to unequal detectabilities and arbitrary variable.

M. Arabkhedri, F. S. Lai, I. Noor-Akma and M. K. Mohamad-Roslan (2010)apply adaptive cluster sampling for estimating total sediment load. Suspended sediment transport in river for a particular period is a timescale finite population. This population shows natural aggregation tendencies in sediment concentration particularly during floods. Adaptive cluster sampling (ACS) can be potentially conducted for sampling from this rare clustered population and estimating total load. To illustrate the performance of ACS in sediment estimation, a comparative study was carried out in the Gorgan-Rood River, Iran, with around a 5 year daily concentration record. The total sediment loads estimated by ACS were statistically compared to the observed load, estimations of selection at list time (SALT) and conventional sediment rating curve with and without correction factors. The results suggest that none of the sediment rating curves produced accurate estimates, while both ACS and SALT showed satisfactory results at a semi-weekly sampling frequency. The best estimation obtained by the rating curves did not show a percent error better than -40%; however, ACS and SALT underestimated the load at less than 5%. The results of this study suggest ACS could improve river monitoring programs.

W. Paul Sullivani, Bruce J. Morrison, F. William H. Beamish apply Adaptive Cluster Sampling for Estimating Density of Spatially Autocorrelated Larvae of the Sea Lamprey with Improved Precision. Adaptive cluster sampling (ACS) provides researchers with an alternative technique to estimate the abundance of rare or spatially clustered organisms, but its application in field investigations has been limited to relatively few studies. They used ACS to estimate parameters of a spatially autocorrelated population of larval sea lampreys, Petromyzon marinus, in Wilmot Creek, a Lake Ontario tributary. When compared with simple random sampling (SRS), ACS significantly increased catch per sample as well as the percentage of samples that contained larvae. Although ACS-generated samples are spatially biased, the use of established formulae enabled them to calculate unbiased estimators of mean larval density and variance. With ACS, variance was reduced, improving the precision around estimates of mean density, however; increased precision came at the price of greater sampling effort. When variance was adjusted for effort, ACS provided equal or greater efficiency than SRS in 33% of sampling events, with no apparent site or seasonal bias. Based on the knowledge that larval sea lampreys are spatially aggregated, they anticipated that ACS would result in higher precision for a greater proportion of sampling events. Nonetheless, they consider ACS to be a useful technique for evaluating larval sea lamprey populations and anticipate increased application for investigating other spatially over-dispersed species.

Hao Yu, Yan Jiao and Kevin Reid (2011) compared the performance of two traditional sampling designs with three adaptive sampling designs using simulated data based on fishery-independent surveys for yellow perch in Lake Erie. Traditionally, the fishery-independent survey has been conducted with a stratified random sampling design based on basin and depth strata; however, adaptive sampling designs are thought to be more suitable for surveying heterogeneous populations. A simulation study was conducted to compare these designs by examining the accuracy and precision of the estimators. Initially in the simulation study, they used bias, variance of the mean, and mean squared error (MSE) of the estimators to compare simple random sampling (SRS), stratified random sampling (StRS), and adaptive two-phase sampling (ATS). ATS was the best design according to these measurements. Then they compared ATS, adaptive cluster sampling (ACS), adaptive two-stage sequential sampling (ATSS), and the currently used stratified random sampling design. ATS performed better than the other two approaches and the current stratified random sampling design. They concluded that ATS is preferable for yellow perch fishery-independent surveys in Lake Erie. Simulation study is a preferred approach when they seek an appropriate sampling design or evaluate the current sampling design.

Chapter 3

Methodology: Sampling Design

Population may have different characteristics in nature so that selection procedures are also needed to be different for different situations. Considering the different natures of target population, sampling designs can be classified into different types. Based on availability of sampling frame we can divide the set of all possible sampling designs into two types:

1. Conventional Sampling Design
2. Non-Conventional Sampling Design

3.1 Conventional Sampling Design

In Conventional sampling design, selection of units does not depend on the value of the variable of interests. Here the size of the sample is predetermine and not change over time and situation. The basic assumption of this type of design is that the population of the variable of interest is finite and a complete data list is available i.e. the sampling frame is available. Most commonly used conventional sampling designs are Simple Random Sampling (SRS), Stratified Random Sampling, Systematic Random Sampling, Cluster Random Sampling etc. Cochran (1977) discussed different conventional sampling design.

3.1.1 Simple Random Sampling

One of the simplest sampling design is Simple Random Sapling (SRS). In this design we select n units from population of size N units and each unit has same chance of being selected. The selection procedure can utilize both with replacement (WR) and without replacement (WOR) criterion.

Let $Y_1, Y_2, \ldots, Y_N$ be the value of the population of size N and $y_1, y_2, \ldots, y_n$ be the value of the sample of size n, which are selected from this population. So an unbiased estimator of the population mean

$$\bar{Y} = \sum_{i=1}^{N} Y_i \ , \tag{3.1}$$

is given by the sample mean

$$\bar{y} = \sum_{i=1}^{n} y_i \ . \tag{3.2}$$

The variance of the sample mean is

$$V(\bar{y}) = \frac{N-n}{N}\frac{S^2}{n}, \text{For WOR sampling} \ . \tag{3.3}$$

$$V(\bar{y}) = \frac{\sigma^2}{n}, \text{For WR sampling} \ . \tag{3.4}$$

Where,

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \ . \tag{3.5}$$

and

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \ . \tag{3.6}$$

The unbiased estimate of the variance of $\bar{y}$ is

$$v(\bar{y}) = (1-f)\frac{s^2}{n}, \text{For WOR sampling} \ . \tag{3.7}$$

$$v(\bar{y}) \ = \ \frac{s^2}{n}, \text{For WR sampling} \ . \tag{3.8}$$

3.1.2 Stratified Random Sampling

Consider a population consisting of N units. In stratified sampling, first the population is divided into L sub-population, which are non-overlapping, so that $N_1 + N_2 + N_3 + \ldots + N_L = N$. Each of these sub-populations is called stratum. If $n_1, n_2, \ldots, n_L$ denote the sample size and is drawn from strata of size $N_1, N_2, \ldots, N_L$ respectively, then $n_1 + n_2 + \ldots + n_L = n$. If a random sampling procedure is used to selecting sample then it is called Stratified Random Sampling.

Let N_h denotes the total number of units and n_h denotes the number of units in the h^{th} stratum and y_{hi} is the value obtained from the i^{th} unit. So the population mean can be written as

$$\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h \ . \tag{3.9}$$

An estimator $\bar{y}_{st}$ for the population mean $\bar{Y}$ can be written as

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \ . \tag{3.10}$$

Now the variance of the estimate $\bar{y}_{st}$ is

$$V\left(\bar{y}_{st}\right) = \sum_{h=1}^{L} W^2 \frac{S_h^2}{n_h}(1 - f_h) \ . \tag{3.11}$$

And an unbiased estimate of this variance is

$$v\left(\bar{y}_{st}\right) = \frac{1}{N^2} N_h (N_h - n_h) \frac{s_h^2}{n_h} \ . \tag{3.12}$$

3.1.3 Systematic Random Sampling

Suppose N units in the population are numbered from 1 to N in some order. Now for selecting a sample of n units, it is to be selected a unit from first k units and then every k^{th} unit thereafter.

The mean of the i^{th} systematic sample is

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^{n-1} y_{ij} \ . \tag{3.13}$$

And the mean of k possible sample mean is

$$\bar{y}_{sys} = \frac{1}{k} \sum_{i=1}^{k} \bar{y}_i \ . \tag{3.14}$$

And its variance is

$$V\left(\bar{y}_{sys}\right) = \frac{N-1}{N} S^2 - \frac{k\left(n-1\right)}{N} S_{wsy}^2 \ . \tag{3.15}$$

3.1.4 Cluster Random Sampling

Suppose a population consists of N clusters each of M elements and a sample of n cluster are drawn from the population by simple random sampling. Let y_{ij} is the observed value for the j^{th} element within i^{th} selected cluster. Then the mean of the i^{th} cluster is

$$\bar{y}_i = \frac{1}{M} \sum_{j=1}^{M} y_{ij} \ . \tag{3.16}$$

And the sample mean of n cluster is

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \ . \tag{3.17}$$

The variance of sample mean is

$$V\left(\bar{y}_n\right) = \frac{1-f}{n} S_b^2, \text{For WOR sampling} \ . \tag{3.18}$$

$$V\left(\bar{y}_n\right) = \frac{S_b^2}{n}, \text{For WR sampling} \ . \tag{3.19}$$

$$v\left(\bar{y}_n\right) = \frac{1-f}{n} s_b^2, \text{For WOR sampling} \ . \tag{3.20}$$

$$v\left(\bar{y}_n\right) = \frac{s_b^2}{n}, \text{For WR sampling} \ . \tag{3.21}$$

3.2 Non-Conventional Sampling Design

When the population is rare and we have no complete data frame for sampling then conventional sampling design does not work properly. For example, in pollution monitoring, if we want to measure the pollutant concentration at a few sites or in ecological studies, if we want to know the abundance of animals or plants etc., conventional sampling designs are not useful. To deal with such situations non-conventional sampling procedures are used. Some well known non-conventional sampling designs are:

1. Line Intercept Sampling
2. Line Transect Sampling
3. Network Sampling
4. Ranked Set Sampling
5. Capture-Recapture Sampling
6. Adaptive Sampling

3.2.1 Line Intercept Sampling

In line intercept sampling (LIS) method, a sampling line, known as transect, which passed through study region and the objects intercepted by the transect line are included in the sample. A common procedure to choose a point is at random on some baseline and to run transect through this point perpendicular to the baseline. The baseline assumption of this sampling procedure is that transects completely crosses the study region. One or more transect line can be used for sampling.

The LIS method has been used extensively by wild life biologists, range manager ecologists and others to estimate the percent coverage by vegetation. LIS was introduced in forestry by Warren and Olser (1964), who coined the term "line-intercept sampling", which is used to describe the technique in the forest literature.

3.2.2 Line Transect Sampling

Line transect sampling (LTS) is widely used in wildlife population assessment e.g. to estimate abundance of marine mammal populations, bird and plant species as well as other object for which detectability depends on location relatively to the observer. In LTS an investigator move along with a selected line and note the location relative to the line and detect and record.

3.2.3 Network Sampling

Network sampling (NS) was introduced by Birnbaum and Sirken (1965) to estimate the number of people with a rare disease. In is also known as “multiplicity sampling”. Here first units are selected by SRS or by stratified random sampling and which are linked to any other of selected units, are included in the sample. An unit can be linked with one or more than one selected units. A variety of linking rule and sampling design that fall under the general category of network sampling have been investigated (Sirken and Levy, 7974; Sudman et al., 1988; Thompson, 1992). Snowball sampling has been applied to NS in two different ways. In one type (Kalton and Anderson, 1986) an initial sample of members of a rare population are asked to identify other members of the population, those so identified are asked to identify other members of the population, those so identified are asked to identify other members of the population and so on, for the purpose of obtaining a non probability sample or for constructing a frame from which to sample. In another type of snowball sampling (Goodman, 1961) individuals in the sample are asked to identify a fixed number of other individuals, who in turn are asked to identify other individuals, for a fixed number of waves, for the purpose of estimating the number of ‘mutual relationships’ or ‘social circles’ in the population.

3.2.4 Ranked Set Sampling

The basic premise for ranked set sampling (RSS) is an infinite population under study and the assumption that a set of sampling units drawn from the population can be ranked by certain means rather cheaply without the actual measurement of the variable of interest which is costly and/or time-consuming.

The idea of RSS was first proposed by McIntyre (1952). The original form of RSS conceived by McIntyre can be described as follows. First, a simple random sample of size k is drawn from the population and the k sampling units are ranked with respect to the variable of interest, say X, by judgement without actual measurement. Then the unit with rank 1 is identified and taken for the measurement of X. The remaining units of the sample are discarded. Next, another simple random sample of size k is drawn and the units of the sample are ranked by judgement, the unit with rank 2 is taken for the measurement of X and the remaining units are discarded. This process is continued until a simple random sample of size k is taken and ranked and the unit with rank k is taken for the measurement of X. This whole process is referred to as a cycle. The cycle then repeats m times and yields a ranked set sample of size $N = mk$.

The essence of RSS is conceptually similar to the classical stratified sampling. RSS can be considered as post-stratifying the sampling units according to their ranks in a sample. Although the mechanism is different from the stratified sampling, the effect is the same in that the population is divided into sub-populations such that the units within each sub-population are as homogeneous as possible. In fact, we can consider any mechanism, not necessarily ranking the units according to their X values, which can post-stratify the sampling units in such a way that it does not result in a random permutation of the units. The mechanism will then have similar effect to the ranking mechanism considered above.

3.2.5 Capture-Recapture Sampling

In capture-recapture sampling (CRS) to estimate the total number of individuals in a population, an initial sample is obtained and the individuals in that sample are marked or identified. A second sample is independently obtained and it is noted how many of the individuals in that sample are marked. If the second population is representative of the population as a whole, then the sample proportion of marked individuals should be about the same as the population proportion of marked individuals. From this relationship, the total number of individuals in the population can be estimated. CRS method have been used to estimate the abundance of animal population including birds, mammal, fish and other species to estimate the detectability of animals for other survey methods, and to estimate the survival and other population parameter. This method has also been used to estimate the abundance of elusive human population.

3.2.6 Adaptive Sampling

Most of the methods discussed in sampling theory are limited to sampling designs in which the selection of the samples can be done before the survey, so that none of the decisions about sampling depend in any way on what is observed as one gathers the data. A new method of sampling that makes use of the data gathered is called adaptive sampling (AS). For example, in doing a survey of a rare plant, a botanist may feel inclined to sample more intensively in an area where one individual is located to see if others occur in a clump. The primary purpose of AS design is to take advantage of spatial pattern in the population to obtain more precise measures of population abundance. In many situations, AS is much more efficient for a given amount of effort that the conventional random sampling designs discussed previously in this chapter. Thompson (1992) presents a summary of these methods. Detail description of this method is described in Chapter 4.

Chapter 4

Methodology: Adaptive Sampling

Adaptive sampling designs are those in which the procedure for selecting units to be included in the sample may depend on values of the variable of interest observed during the survey. For spatially clustered populations, additional observations may be added in the neighbouring vicinity whenever high abundance is encountered. In a survey sample of a rare, contagious disease, whenever an infected person is observed, close contacts of that person might be added to the sample. In a drug use study, sampling intensity could be adaptively increased in the neighbourhood of respondents with self-reported use.

4.1 Adaptive Cluster Sampling

When dealing with rare or hidden populations, it is often useful after locating a unit that meets a specified criterion to continue sampling in that region. One way of doing so is provided by adaptive cluster sampling. In spatial sampling, adaptive cluster sampling can provide efficient unbiased estimators for the abundance of rare clustered populations (Thompson and Seber 1996). So when a population possess the following behaviors ACS gives the better result about the population total:

1. Complete frame of units is not available.
2. Sampling is made on the basis of geographical region.
3. Population units have a pattern of clustering.
4. Population is rare.

Adaptive cluster sampling refers to design in which an initial set of units is selected by some probability sampling procedure, and, whenever the variable of interest of a selected unit satisfies a given criterion; additional units in the neighborhood of that unit are added to the sample.

4.1.1 Design of Adaptive Cluster Sampling

Thompson (1990) first introduced adaptive cluster sampling (ACS). ACS has a wide range of application in ecological and environmental studies when the population is rare and cluster tendency.

Consider a region where population is finite. In ACS design the whole study area is divided into N units. An initial set of n units is selected by simple sandom sampling (SRS). Let the variable of interest is y which satisfy a condition C (say). Here y is the number of elements in a unit, so y_i is the number of elements of observed value for the i^{th} selected unit and the condition C, may be defined as $y_i > 0$. Now if the i^{th} unit satisfies the condition C, the neighbourhoods are also included in the sample and so on. This process is terminated when a full cluster of unit is selected and cluster is added to the sample. Within such a cluster of unit is a sub-collection of unit, termed as a network. Any unit is not satisfying the condition but in the neighbourhood of one that termed as an edge unit. These edge units are not included in the sample.

4.1.2 Initial Random Sample Without Replacement

Let N be the number of sampling units in the population. An initial sample of size n_1 is selected by simple random sampling without replacement. So these $n_1$units are distinct. Let A_i be the network for sampling unit i and m , be the number of sampling units in A_i. Again, let y_i be the response of unit i and C be the condition that when satisfied, that sampling unit's network is added to the sample. Now if a_i be the total number of sampling units in networks of which sampling unit i is an edge unit and if unit i satisfies C, then $a_i = 0$. If unit i does not satisfy C, then $m_i = 1$. The probability that the unit i will be selected in the sample at any of the n_1 draw is (Thompson 1992)

$$p_i = \frac{m_i + a_i}{N} \ . \tag{4.1}$$

The probability that unit i is included in the sample is

$$\pi_i = 1 - \binom{N - m_i - a_i}{n_1} \times \left[\binom{N}{n_1}\right]^{-1} \ . \tag{4.2}$$

4.1.3 Estimator Based Upon Initial Intersection Probabilities

The modified Horvitz-Thompson estimator of population mean is (Thompson 1990)

$$\begin{aligned} \hat{\mu}_{HT} &= \frac{1}{N}\sum_{i=1}^{v}\frac{y_i}{\pi_i} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{y_i I_i}{\pi_i} \ . \end{aligned} \tag{4.3}$$

Where $y_1, y_2, \ldots, y_v$ are the y-values of the v distinct units in the final sample, and I_i takes the value 1 when unit i is included in the sample and 0 otherwise. Since all the m_i's are known for all units in the sample and some of a_i are not known. To overcome this problem we drop a_i from π_i and consider the "partial" inclusion probability

$$\pi_i' = 1 - \binom{N - m_1}{n_1} \times \left[\binom{N}{n_1}\right]^{-1} \ . \tag{4.4}$$

The unbiased estimator of population mean based on the initial intersection probabilities is (Thompson 1990)

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i I_i'}{\pi_i'} \ , \tag{4.5}$$

where I_i' takes value 1 (with propability π_i') if the initial sample intersects A_i, and 0 otherwise.

This can be rewritten in terms of the distinct networks as

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{K}\frac{{y_k}^* J_k}{\alpha_k} = \frac{1}{N}\sum_{k=1}^{\kappa}\frac{{y_k}^*}{\alpha_k} \ , \tag{4.6}$$

where ${y_k}^*$ is the sum of the y- values for the k^{th} network, K is the total number of distinct networks in the population, κ is the number of distinct networks in the sample and J_k is another indicator function that takes 1 (with probability α_k)if the initial sample intersects the k^{th} network and 0 otherwise.

If there are x_k sampling units in the k^{th} network then

$$\alpha_k = 1 - \binom{N-x_k}{n_1} \times \left[\binom{N}{n_1}\right]^{-1} \ . \tag{4.7}$$

Let p_{jk} to be the probability that j^{th} and k^{th} networks are not intersected then

$$p_{jk} = \binom{N-x_j-x_k}{n_1} \times \left[\binom{N}{n_1}\right]^{-1} \ . \tag{4.8}$$

So that the joint probability that networks j and k are both intersected is

$$\begin{aligned} \alpha_{jk} &= \alpha_j + \alpha_k - (1 - p_{jk}) \\ &= 1 - \left[\binom{N-x_j}{n_1} + \binom{N-x_k}{n_1} - \binom{N-x_j-x_k}{n_1}\right]\left[\binom{N}{n_1}\right]^{-1} \end{aligned} \tag{4.9}$$

Now the variance can be derived as

$$Var\left(\hat{\mu}\right) = \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K}\frac{{y_j}^*{y_k}^*\left(\alpha_{jk} - \alpha_j\alpha_k\right)}{\alpha_j\alpha_k}\right] \ . \tag{4.10}$$

And the unbiased estimator of the variance of $\hat{\mu}$ is (Thompson 1990)

$$\widehat{Var}\left(\hat{\mu}\right) = \frac{1}{N^2}\left[\sum_{j=1}^{\kappa}\sum_{k=1}^{\kappa}\frac{{y_j}^*{y_k}^*}{\alpha_j}\left(\frac{\alpha_{jk}}{\alpha_j\alpha_k}-1\right)\right] . \tag{4.11}$$

4.1.4 Estimator Using Number of Initial Intersections

Let f_i be the number of units in the initial sample that fall in the network A_i that includes unit i. Ignoring the edge units of clusters in the estimation process, f_i is the number of times that the i^{th} unit in the final sample appears in the estimator. An unbiased estimator of the population mean is

$$\tilde{\mu} = \frac{1}{n_1}\sum_{i=1}^{N}\frac{y_i f_i}{m_i} . \tag{4.12}$$

This can be rewritten in terms of the n_1 not necessarily distinct networks intersected by the initial sample,

$$\tilde{\mu} = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{m_i}\sum_{j\epsilon A_i} y_j = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i = \tilde{w} , \tag{4.13}$$

where w_i is the mean of the m_i observations in A_i.

This estimator has variance

$$Var\left(\tilde{\mu}\right) = \frac{N-n_1}{Nn_1(N-1)}\sum_{i=1}^{N}(w_i-\mu)^2 , \tag{4.14}$$

with unbiased estimator

$$\hat{Var}\left(\hat{\mu}\right) = \frac{N-n_1}{Nn_1(N-1)}\sum_{i=1}^{n_1}(w_i-\tilde{\mu})^2 . \tag{4.15}$$

4.1.5 Initial Random Sample With Replacement

If the initial random sample is selected by simple random sampling with replacement then repeated units can occur in the initial sample due either to repeat selection or selection of more than one unit in a cluster. Then the probability that the initial sample intersects the network k becomes

$$\alpha_k = 1-\left(1-\frac{x_k}{N}\right)^{n_1} , \tag{4.16}$$

and

$$\alpha_{jk} = 1 - \left[\left(1 - \frac{x_j}{N}\right)^{n_1} + \left(1 - \frac{x_k}{N}\right)^{n_1} - \left(1 - \frac{x_j + x_k}{N}\right)^{n_1} \right] . \qquad (4.17)$$

Chapter 5

Methodology: Simulation

A *simulation* is the imitation of the operation of real-world process or system over time. Whether done by hand or on a computer, simulation involves the generation of an artificial history of a system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system. Simulation has long been an important tool to researchers for studying a phenomena. Although simulation is often viewed as a 'method of last resort' to be employed when everything else has failed, recent advances in simulation methodologies, availability of software and technical developments have made simulation one of the most widely used and accepted tools in system analysis and operations research. Naylor defined simulation as "Simulation is a numerical technique for conducting experiments on a digital computer, which involves certain types of mathematical and logical models that describe the behaviour of business and economic system (or some component thereof) over extended periods of time."

5.1 Some Situations Where Simulation Can be Successfully Used

Some example situations where a simulation study can be appropriately used are given in the following:

- The observed system may be so complex that, it cannot be described in terms of a set of mathematical equations for which analytic solution are obtainable. Most economic system fall into this category. For example, it is virtually impossible to describe the operation of business farm , an industry or an economy in terms of a few simple equations. Simulation has been found to be an extremely effective tool for dealing with problems of this type. Another class of problems involving multiple channels that are either parallel or in series (or both).

- It may be either impossible or extremely expensive to obtain data from certain real world. Such process might involve ,for example the performance of large-scale rocket engines, the effect of proposed tax cuts on the economy, the effect of an advertising campaign on total sales. In these cases the simulated data are necessary to formulate hypotheses about the system.

- It may be either impossible or very costly to perform validating experiments on the mathematical models describing the system. In this case the simulation data can be used to test alternative hypothesis.

- Even though a mathematical model can be formulated to describe some system of interest, it may not be possible to obtain a solution to the model by straightforward analytic techniques. Again economic system and complex queuing problems provide examples of this type of difficulty. Although it may be conceptually possible to use a set of mathematical equations to describe the behavior of a dynamic system operating under conditions of uncertainty, present-day mathematics and computer technology are simply incapable of handling a problem of this magnitude.

In all these above cases simulation is the only practical tool for obtaining relevant answer.

5.2 Reasons for Simulation Analysis

The main reasons for resorting to simulation are:

- Simulation make it possible to study and experiment with the complex internal interactions of given system whether it be a firm, an industry, an economy or some subsystem of one of these.

- Operational gaming has been found to be an excellent means of stimulating interest and understanding on the part of the participant. In particularly useful in the orientation of the persons who are experienced in the subject of the game.

- Simulation can be used as a pedagogical device for teaching both students and practitioners basic skills in theoretical analysis, statistical analysis and decision making. Among the disciplines in which has been used successfully for this purpose are business administration, economics, medicine, and law.

- The experience of designing a computer simulation model may be more valuable than the actual simulation itself. The knowledge obtained in designing a simulation study frequently suggests changes in the system being simulated. The effects of these changes can then be tested via simulation before implementing them on the actual system.

- Simulation of complex system can yield valuable insight into which variable are more important than others in the system and how these variables interact.

- Simulations are sometimes valuable in that they afford a convenient way of breaking down a complicated system into subsystems, each of which may then be modeled by an analyst or team that is expert in that area.

- Simulation makes it possible to study dynamic systems in either real time, compressed time or expanded time.

- When new components are introduced into a system, simulation can be used to help foresee bottlenecks and other problems that may arise in the operation of the system.

- Detailed observation of the system being simulated can lead to a better understanding of the system and to suggestions for improving it, suggestions that otherwise would not be apparent.
- Simulation can be used to experience with new situations about which we have little or no information so as to prepare for what may happen.
- Simulation can be used to study the effects of certain informational, organizational and environmental changes on the operation of a system by making alterations in the model of the system and observing the effects of these alterations on the system's behavior.
- Simulation can serve as a "pre-service test" to try out new policies and decision rules for operating a system, before running the risk of experimenting on the real system.

Computer simulation also enables us to replicate an experiment. Replication means running an experiment with selected changes in parameters or operating conditions be made by the investigator. In addition, computer simulation often allows us to induce correlation between these random number sequences to improve the statistical analysis of the output of a simulation. In particular a negative correlation is desirable when the results of two replications are to be summed, whereas a positive correlation is preferred when the results are to be differenced, as in the components of experiments.

Simulation can be defined as a technique of performing *sampling experiments* on the model of the system. This general definition is often called simulation in a *wide sense*, whereas simulation in a *narrow sense* or *stochastic simulation* is defined as experimenting with the model over time. It includes sampling stochastic vitiates from probability distribution. Therefore stochastic simulation is actually a statistical sampling experiment with the model. This sampling involves all the problems of statistical design analysis. Because sampling from a particular distribution involves the use of random numbers, stochastic simulation is sometimes called *Monte Carlo simulation*.

5.3 Monte Carlo Method

Historically, the Monte Carlo method was considered to be a technique using random or pseudo-random numbers for solution of a model. Random numbers are essentially independent random variables uniformly distributed over the unit interval [0,1]. The term "Monte Carlo" was introduced by Von Neumann and Ulam during the World War II, as a code word for the secret work at Los Alamos. It was suggested by the gambling casinos at the city of Monte Carlo in Monaco. The Monte Carlo method was then applied to problem related to the atomic bomb. Shortly thereafter Monte Carlo methods were used to evaluate complex multidimensional integrals and to solve certain integral equations occurring in physics that were not amenable to analytic solution. The Monte Carlo method is used not only for solution of stochastic problem but also for solution of deterministic problems. Another field of application of Monte Carlo methods is sampling of random variates from probability distributions.

The Monte Carlo method is now the most powerful and commonly used technique for analysing complex problems. Applications can be found in many fields from radiation transport to river basin modeling. Recently the range of application has been broadening and the complexity and computational effort required has been increasing. Because realism is associated with more complex and extensive problem descriptions.

5.4 Comparison Between Monte Carlo Method and Simulation

The main differences between simulation and Monte Carlo method can be listed as in the following:

- The observations in the Monte Carlo method, as a rule, are independent. In simulation, however, we experiment with the model over time so, as a rule, the observations are serially correlated.
- In the Monte Carlo method time does not play as substantial a role as it does in stochastic simulation.
- In Monte Carlo method it is possible to express the response as a rather simple function of the stochastic input variates. In simulation the response is usually a very complicated one and can be expressed explicitly only by the computer program itself.

5.5 Some Definitions Used in this Study

5.5.1 Absolute Percentage Relative Bias

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from the density $f(x, \theta)$. An estimator $T = t(X_1, X_2, \ldots, X_n)$ is defined to be an unbiased estimator of θ if

$$E(T) = \theta \ . \tag{5.1}$$

An estimator that is not unbiased is said to bias and the difference

$$b(\theta) = E(T) - \theta \ , \tag{5.2}$$

is called the *bias* of the estimator.

Let $\hat{T}$ be a biased estimator of the population parameter θ, then absolute percentage relative bias (APRB) is defined as ,

$$APRB = \left| \frac{E(\hat{T}) - \theta}{\theta} \right| 100 \ . \tag{5.3}$$

5.5.2 Relative Efficiency

If the variance of the estimator is small, the distribution of the estimator will be better in that its value will be closer to the parameter value. This is the notion of efficiency. It can be treated as a relative term. The efficiency of an estimator depends on its variance. A measure of relative efficiency can be computed by taking the ratio of the variance of two estimator of interest.

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from the density $f(x, \theta)$ where θ is the unknown parameter. Suppose T_1 and T_2 are the two unbiased estimators of the same population parameter θ with variances $V(T_1)$ and $V(T_2)$ respectively. The relative efficiency of T_2 relative to T_1 is defined by the ratio.

$$RE = \frac{\frac{1}{V(T_2)}}{\frac{1}{V(T_1)}} = \frac{V(T_1)}{V(T_2)} \ . \tag{5.4}$$

Now if $RE < 1$, then $V(T_1)$ is less than $V(T_2)$, that is T_1 is better estimate than T_2 and if $RE > 1$, vice versa interpretation could be attained.

Chapter 6

Agriculture Census Data 2008

6.1 Bangladesh Bureau of Statistics

Bangladesh has a centralized official statistical system named Bangladesh Bureau of Statistics (BBS). BBS is the only national Statistical institution responsible for collecting, compiling and disseminating statistical data of all the sectors of the Bangladesh economy to meet and provide the data-needs of the users and other stake holders like national level planners and other agencies of the Govt. The role of the BBS in providing necessary statistics for preparing the various national plans and policies for the overall development of the country is very significant. BBS is also responsible for providing technical and administrative guidance in matters of all official statistical programmes and acts as the implementing agency of all programmes of official statistics of Bangladesh. Bangladesh is predominantly an agricultural country. The government has, therefore, accorded a highest priority to agricultural development. For determining policy for planning, policy formulation and for developing action program in agriculture, basic data regarding the structural and other characteristics of agriculture are essential. Data required for agricultural purpose are provided through the Agriculture wing of BBS. Again Bangladesh Bureau of Statistics conducts censuses through the Census Wing. It is one of the six wings of BBS constituting of three sections namely Population Census, Agriculture Census and Economic Census. The

2008 Census of Agriculture is the fourth Agriculture Census in Bangladesh. Previous Agriculture censuses were conducted in 1977, 1983-84, 1996. In our study, we use the data of fourth Agriculture Census i.e. the data of 2008 Agriculture Census as a secondary data.

6.2 Agricultural Census, Bangladesh

Despite steady progress towards industrialization, agriculture remains the most important sector in Bangladesh. About 21% of Gross Domestic Product (GDP) of the country comes from agriculture sector. Besides, it has indirect contribution to the overall growth of GDP. Many sectors included in broad service sector such as wholesale and retail trade, hotel and restaurants, transport and communication are strongly supported by the agriculture sector. This sector also provides employment for around 50% of the total labor force and seems to have managed to feed around 140 million people of the country. During the last decade, significant changes took place in agriculture sector which include, among others, new production structure, use of high yielding varieties supported by fertilizers, pesticides, mechanized cultivation, irrigation etc. All these changes have contributed much to the increased production of food grains. The development of agriculture sector is very much urgent for poverty reduction, food security and sustainable development of our country.

The importance of the agriculture sector has long been recognized by the Government of Bangladesh. Government has been launching a range of development projects and programmes in the agriculture sector in the line of National Agriculture Policy formulated in 1999. Also, an action plan is in place for achieving goals and objectives articulated in the National Agriculture Policy. This action plan will assist in building a sustainable food security system by achieving optimum growth in agriculture. Considering the changing situation at national and international levels, the preparation of a new National Agriculture Policy is in the process of finalization. Measures have been taken to strengthen the national agricultural research bodies so that they can carry out agriculture research in an effective and efficient manner. Timely and realistic statistics on structure and performance of agriculture sector is inevitable for sound agricultural development planning. Agriculture census is the basic source of information relating to the structure and performance of the agriculture sector. The 2008 Agriculture Census is the most recent one country-wide statistical operation to measure the overall performance of the agriculture sector. The census was designed to collect various information such as number of agriculture holdings, their area, tenancy, irrigation status, size of holdings, land ownership, land use, agriculture labor, number of poultry birds, livestock and many others.

6.3 Agriculture Census 2008

The 2008 Census of Agriculture is the fourth Census of Agriculture in Bangladesh. Prior to Agriculture Census, 2008, the last Agriculture Census was conducted in 1996. Previous censuses were conducted in 1977, 1983-84. Additionally, an Agriculture Sample Survey was conducted in 2005. It may be mentioned that though Agriculture Census, 2008 is the 4th Agriculture Census in the country-yet this census is the first of its kind in the sense that all dwelling households both in rural and urban area were canvassed simultaneously using the same questionnaire.

6.3.1 Objectives of Agriculture Census, 2008

The Agriculture Census, 2008 was envisioned with following objectives: The broad objective of the Agriculture Census, 2008 is to determine the structure and characteristics of agricultural holdings managed by dwelling households. The specific objectives of the Agriculture Census, 2008 are as follows:

- To determine the number of agricultural holdings, area of holdings, average size of holding etc.
- To determine the number and distribution of households engaged in agriculture.
- To determine the economic and employment size of the agriculture sector.
- To determine the number of wage labor by gender employed in agriculture sector.

- To determine the irrigated area under different crops.
- To determine the stock of livestock and poultry.
- To measures of the state and changes in attributes relating to the structure of agriculture such as the size and distribution of holding, tenure ship, size and type of farming, extent of agriculture resources etc.
- To provide benchmark data for improving current estimates of crop acreage, production and livestock resources.
- To form a basis for the formulation, development and implementation of the programme and policies for agricultural development of the country.

6.3.2 Concept and Definitions of Some Terms Used in Agriculture Census-2008

Statistical Unit:

The agricultural holding is normally considered as the statistical unit for agricultural census. An agricultural holding is defined as a techno-economic unit of agricultural production under a single management which is generally operated by a household or jointly by two or more households. In some limited cases, this economic unit is operated by institutions or establishments or government and semi-government organizations or other agencies or by a judicial person. In almost 100% cases, households and agricultural holdings has a one to one correspondence. The households operating agricultural land and keeping livestock or poultry birds can be used instead of the agricultural holdings. Again important information regarding households, agriculture labor households or households having no own land etc. could only be obtained if the households was treated as a statistical unit. In view of these considerations it was decided to treat the households (agriculture labor households or other kinds of households) as the statistical unit for the agricultural census.

Census Domain:

Urban Domain:

The Metropolitan Cities of Barisal, Chittagong, Dhaka, Khulna, Rajshahi and Sylhet and 58 other Municipalities located in the District Headquarters were included in the 'Urban domain' of the country in Agricultural Census 2008.

Rural Domain:

The rest of the country including Municipalities located at the Upazila Headquarter and other urban areas were included in the Rural domain. Small growth areas with urban characteristics adjacent to municipalities and metropolitan cities are also treated as rural areas in the 2008 Agriculture Census.

Household:

A household means a group of persons normally living together and eating in one mess (i.e. with common arrangement of cooking) with their dependants, relatives, servants etc. A household may be a one person household or a multi-person household. In other words, when a group of persons living together generally maintain a family or family like relations and take meals from the same dish is termed as a household. Popularly, it is described as "Khana". In some cases there may be more than one household in a single house or in one dwelling arrangement. Similarly, a household may have more than one house or structure or shed. The household must be distinguished from a family which consists of blood related members who may live in different places but members of the household must share the same kitchen and live together.

Holder:

The holder is the member of the household who exercises management control over the operations of the agricultural holding and takes the major decisions regarding the utilization of the available resources. He has technical and economic responsibility for the holding, which he may operate directly as owner or tenant or through a manager (hired person) to whom he has delegated the responsibility for day-to-day management of the work.

Agricultural Holding:

An agricultural holding is a techno-economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form, or size. Single management may be exercised by an individual household, jointly by two or more individuals or households or a juridical person such as a corporation, co-operative or government agency. A holding may consist of more than one parcel (fragment) located in one or more separate areas or mauzas or in more than one administrative unit or division provided that all the separate parcels or fragments form a part of the same technical unit under operational control of the same management.

The definition covers practically all holdings or virtually all households engaged in agricultural production and includes livestock with no agricultural land. So holdings may have no significant agricultural land area, e.g., poultry, hatcheries, holding keeping livestock for which land is not an indispensable element for their production. In the context of this report a holding and a household may be treated as same.

Agricultural labor Household:

Households whose major source of income was from agricultural labor force during the preceding years were considered as agricultural labor households. Agricultural labor was defined as labor exchanged for wages in cash or kind or both for agricultural activities on land operated by other households.

Farm Household:

The basic unit of enumeration in the Agricultural Census was the household. The household could constitute a farm holding which for census purpose, is a techno-economic unit of agricultural production comprising all livestock kept and all the land which is used wholly or partly for agricultural purposes and is operated under a single management by one person alone or with others, without regard to title, size or location. Households with less than 0.05 acres of cultivated area were treated as non-farm households. The area of land situated at different villages/mauzas but under the same operational control constitutes one farm household.

Classification of farm households

Farm households are broadly classified as (*a*) Small (*b*) Medium and (*c*) Large.

- Small farm household is a farm holding having an operated area of 0.05 to 2.49 acres of land.

- Medium farm household is a farm holding having an operated area of 2.50 to 7.49 acres of land.

- Large farm household is a farm holding having an operated area of 7.50 acres and above.

The cut-off point distinguishing farm households was fixed at 0.04 acre of cultivated area, and the households having cultivated area up to 0.04 acre were considered as non- farm households. Small cultivated area like 0.04 acre or less is generally utilized as kitchen garden. Vegetables are grown within the homestead area. Even the seeds of white gourd, water gourd, pump kin etc. are sown by the side of the structures and houses but keep spreading on and around the roofs and the structures. Considering all these factors, the minimum cultivated area of .05 acres was fixed for qualifying to be a farm household.

Classification of landless households

In this census different types of landlessness is defined as follows:

- **Landless Type-I:** households owning no own land whatsoever.

- **Landless Type-II:** households with homestead land but no cultivated land.

- **Landless Type-III:** households with homestead area and also cultivated area upto 0.05 acre (whether owned or taken from others).

- **Landless Type-IV:** households having homestead area and cultivated area 0.51 to 1.00 acre (owned or taken from others).

Owned Land:

Owned land means the area of land owned by the holder including the members of his family having a title to the land with the right to determine the nature and extent of its use and to transfer the same. Moreover, there might be some land over which the holder or any member of his households has owner-like possession. This type of land was included in the area of owned land. The land held by the holder in owner like possession, can be operated by him in the same way as owned land although the holder does not possess a title of ownership.

Chapter 7

Descriptives of the 2008 Agriculture census Data

7.1 Introduction

The 2008 Agriculture census was carried out in accordance with the board guidelines provided by the Food and Agriculture Organization (FAO). This census reflected the recommendations of FAO and other international organizations as it was vital for proper quality and objectiveness of statistical information of agriculture. The 2008 Agriculture Census was carried out on a full count basis throughout the country from 11th May, 2008 to 25th May, 2008. A structured questionnaire was designed to collect information on agriculture. The Questionnaire was pre-tested at the field level to test the relevance and appropriateness of the survey instruments. The questionnaire then modified and finalized on the basis of the findings available from the pre-testing. A comprehensive and operational training manual was developed for enumerators and supervisors. The enumerators and supervisors who were engaged in the data collection in the field level were given intensive and adequate training on the concepts, definitions, classification used in the census questionnaire. The enumerators were also trained on mock interview. Educated unemployed youths were hired for collecting data from the field level. They were intensively and adequately trained on the goal and objectives of the census, methodology and census instruments. Senior officials from BBS were deployed to supervise data collection at the field level. Enumerators conducted face to face interview using structured and pre-coded questionnaire.

Upon completion of the field work of the full count census, a Post Enumeration Check (PEC) was carried out in order to assess the quality of census data and to examine the extent of coverage of the census frame used. The sample for the PEC was drawn from the list of EAs of full count census. Out of 1, 55,551 full count census EAs, 200 EAs were selected systematically for conducting PEC survey. A separate schedule was specially designed for this purpose. The PEC was carried out during 29 June to 8 July, 2008.

We collected 2008 Agriculture Census data from the BBS through proper cannel which required submission an application to the Director General of BBS via Project Director, Agriculture Census 2008,BBS. This application was forwarded and recommended by director of ISRT, DU and my supervisor. In 2008 Agriculture census, data was collected for more than 100 variables under 24 main items in the questionnaire for each and every household. Some of the main items are geographical location code (such as division code, zila code, upazila code, union code, mouza code, enumerated area code etc.) of household (HH), age and sex-wise member in each HH, age and sex-wise population engaged in agriculture in each HH, Agri.labor HH, amount of land used in different purpose, cultivated different types of long term and short term local and hybrid crops etc. The bio-technically powered foreign seed considered in Agriculture census 2008 are HYV Aus, HYV aman, *Hybrid Boro*, HYV Boro, and Hybrid Maize. Among these powered seed here we consider only Hybrid Boro because now-a-days it is significantly cultivated by the Bangladeshi farmer compare to other foreign seed. Finally we consider case number, status of hybrid Boro cultivation and geographically identification code (formed by using division, zila, upazila, union, mouza and ea code) as variables in our research.

7.2 Summary of Findings

7.2.1 Number of Households by Urban, Rural and Divisions

The preliminary findings available from 2008 Agriculture Census show that there is a total of 28.67 households (dwelling households) in the country of which 25.36 million (88.45%) are in rural areas and only 3.31 million (11.55%) are in urban areas. Dhaka division has the highest number of households (32.91% of total households) followed by Rajshahi (26.69%) Division.

Division	Number of households('000)			Percent of households	
	Total	Rural	Urban	Urban	Rural
Barisal	1729	106	1623	6.13	93.87
Chittagong	4887	580	4307	11.86	88.14
Dhaka	9437	1776	7661	18.81	81.19
Khulna	3430	309	3121	9.08	90.92
Rajshahi	7654	464	7190	6.09	93.94
Syllhet	1533	80	1453	5.51	94.85
Bangladeh	28670	3315	25355	11.56	88.44

Table 7.1: Distribution of Households by Urban, Rural and Division.

The table 7.1 also shows that the highest number of urban households (1.77 million) is in Dhaka division followed by Chittagong division (0.58 million).

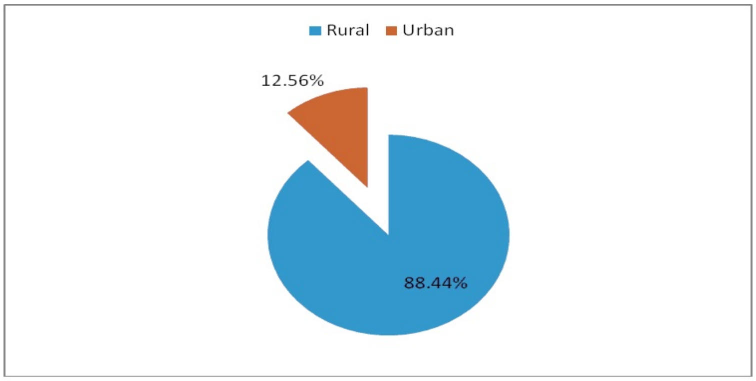

Figure 7.1: Distribution of Households by Urban and Rural.

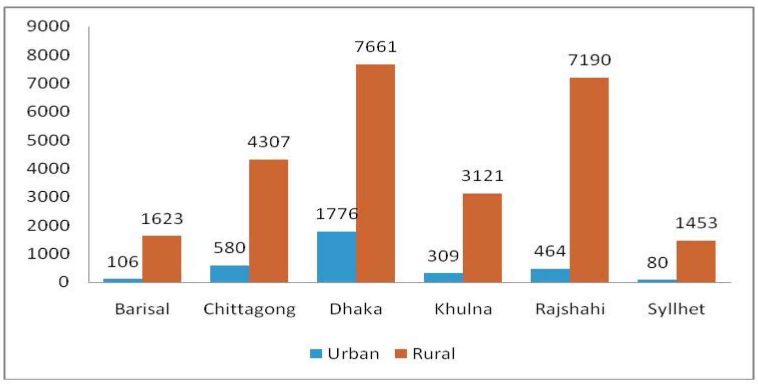

Figure 7.2: Distribution of Division wise Households by Urban and Rural.

7.2.2 Agriculture labor Households

Agriculture labor households are those households whose main source of income was wages/earnings from agriculture labor. The table 7.2 shows that agriculture labor households account for 31.13% of the total household in the country of which only 0.27% is in urban area while 30.86% is in rural areas. The table 7.2 also discloses the fact that Rajshahi Division accounts for the highest proportion (41.60%) of agriculture labor households followed by Khulna Division (38.07%). On the other hand, Dhaka Division reported the lowest proportion (23.12%) of agriculture labor households.

Division	Total HH('000)	No of agriculture labor HH(000)			% of agriculture labor HH		
		Total	Urban	Rural	Total	Urban	Rural
Barisal	1729	509	5	504	29.44	0.29	29.14
Chittagong	4887	1296	11	1285	26.52	0.22	26.29
Dhaka	9437	2182	26	2156	23.12	0.27	22.84
Khulna	3430	1306	13	1293	38.07	0.37	37.69
Rajshahi	7654	3184	22	3162	41.60	0.69	40.91
Syllhet	1533	451	2	449	29.41	0.13	29.28
Bangladeh	28670	8927	78	8849	31.13	0.27	30.86

Table 7.2: Distribution of Agriculture labor Households by Urban, Rural and Division.

7.2.3 Landless Households

Landless households are those households who don't own any type of land. The table 7.3 reveal that out of a total of 28.67 million households in the country, 4.48 million households (15.62%) are absolutely landless. Out of 15.62% landless households, 4.25% households are in urban area and 11.35% are in rural areas. The total number of households in urban area has been recorded at 3.31 million of which 1.22 million (36.85%) are absolutely landless. On the other hand, total number of households in rural areas has been recorded at 25.35 millions of which 3.26 million households (12.85%) are landless. Thus the data available from the Agriculture Census, 2008 disclose the fact that the percentage of landless households in urban areas is significantly higher than that of in rural areas.

The table 7.3 also shows that Dhaka Division has the highest (20.32%) landless households followed by Sylhet and Rajshahi Division with 15.32% and 14.50% landless households respectively. The lowest percentage (9.36%) of landless households was recorded for Barisal Division proceeded by Khulna

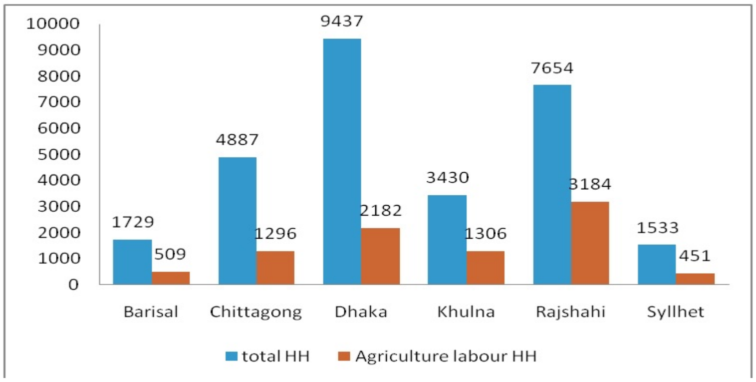

Figure 7.3: Distribution of Division wise Agriculture labor HH and total HH.

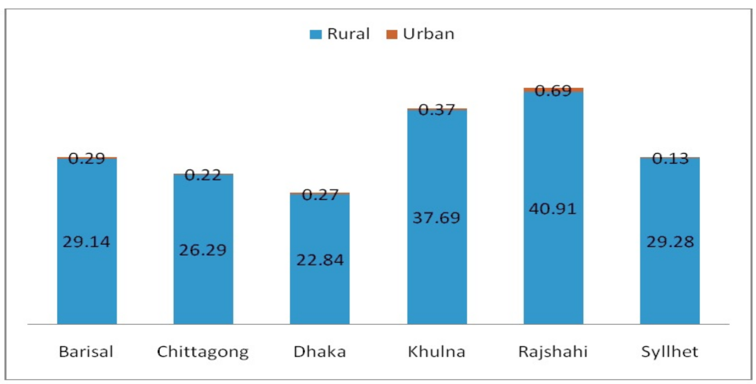

Figure 7.4: Distribution of Agriculture labor Households by Urban, Rural and Division.

Division (11.74%). For urban area, DhakaDivision has the highest landless households (8.36%) while Rajshahi Division stands at the bottom with 1.35% landless households. For rural, the highest percentage (13.76%) of landless households is observed for Sylhet Division and the lowest percentage (7.92%) is observed for Barisal Division.

7.2.4 Tenancy

Tenant households are those households who pay rent (either in cash or in kind) to use or occupancy land for cultivation or other purposes owned

Division	Total HH('000)	No of landless HH(000)			Percent of landless HH		
		Total	Urban	Rural	Total	Urban	Rural
Barisal	1729	162	25	137	9.36	1.44	7.92
Chittagong	4887	649	174	475	13.28	3.44	9.84
Dhaka	9437	1918	789	1129	20.32	8.36	11.96
Khulna	3430	403	104	298	11.74	3.03	8.71
Rajshahi	7654	1110	104	1006	14.50	1.35	13.15
Syllhet	1533	235	24	211	15.32	1.56	13.76
Bangladeh	28670	4477	1221	3256	15.62	4.25	11.35

Table 7.3: Distribution of Landless Households by Urban, Rural and Division.

by another. The table 7.4 show that 8.42 million households are tenant households which account for 29.36% of total households. Out of 29.36% tenant households, 1.60% is in urban areas and 27.76% is in rural areas.

7.2.5 Agriculture Farm Houehold

The table 7.5 shows that out of total 28.67 million households, the number of agriculture farm households (households operating 0.05 acres of cultivated area) has been recorded at 14.72 million which account for 51.33% of total households. Out of total 51.33%, only 1.15% agriculture farm is in urban area while 50.18% is in rural areas.

The highest percentage (65.12%) of agriculture farm has been recorded for Barisal Division followed by Khulna Division (59.09%) and Rajshahi Division (55.83%) respectively. The 2008 Agriculture Census also disclose the fact that out of total 3.31 million urban households, 0.33 million households (9.97%) are tenant households while out of 25.35 million rural households,

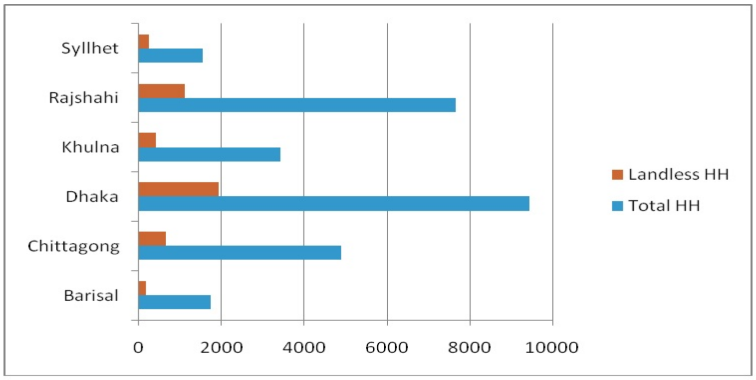

Figure 7.5: Distribution of Landless Households by Division.

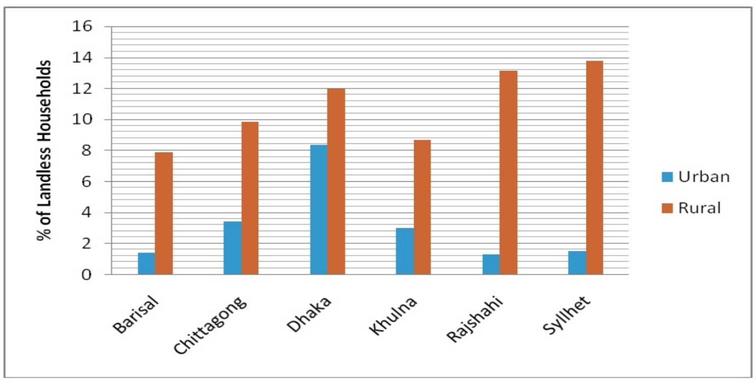

Figure 7.6: Distribution of Landless Households by Urban and Rural.

14.38 million households (56.72%) are tenant households.

7.2.6 *Hybrid Boro* Cultivating Farm Household by Division

There are different types of Boro seed is cultivated in Bangladesh like Local Boro, *Hybrid Boro*, HYV Boro etc. The bio-technically powered variety of Boro seed considered in Agriculture cencus 2008 are HYV Boro and *Hybrid Boro*. The table 7.6 shows that out of total 14.72 million of farm households, 2.77 million households (18.85%) cultivated *Hybrid Boro* during the

Division	Total HH('000)	No of tenant HH(000)			Percent of tenant HH		
		Total	Urban	Rural	Total	Urban	Rural
Barisal	1729	480	29	452	27.76	1.67	26.09
Chittagong	4887	1337	111	1227	27.36	2.27	25.09
Dhaka	9437	2386	133	2253	25.28	1.41	23.87
Khulna	3430	1200	86	1114	34.98	2.51	32.47
Rajshahi	7654	2639	95	2544	34.48	1.24	33.24
Syllhet	1533	375	6	369	24.46	0.39	24.07
Bangladeh	28670	8418	460	7958	29.36	1.60	27.76

Table 7.4: Distribution of Tenant Households by Urban, Rural and Division.

Boro season February-April, 2008). Rajshahi Division reported the highest percentage (27.22%) of households with *Hybrid Boro* cultivation and Barisal Division reported the lowest percentage (6.31%) of such households

7.2.7 Distribution of *Hybrid Boro* Cultivating Household and Total Household by Division

The concept of cultivating *Hybrid* variety of *Boro* Rice is relatively new. The table 7.7 shows that out of total 28688791 households, 2531644 households (8.82%) cultivated *Hybrid Boro* during the Boro season (February-April, 2008). Rajshahi Division reported the highest percentage (13.7%) of households with *Hybrid Boro* cultivation and Barisal Division reported the lowest percentage (3.6%)of such households.

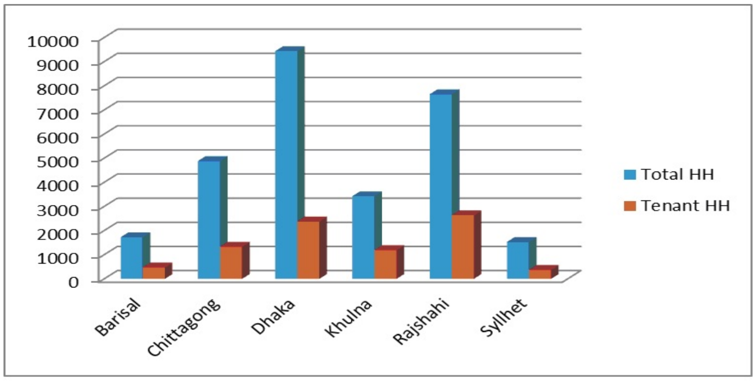

Figure 7.7: Distribution of Tenant Households by Division.

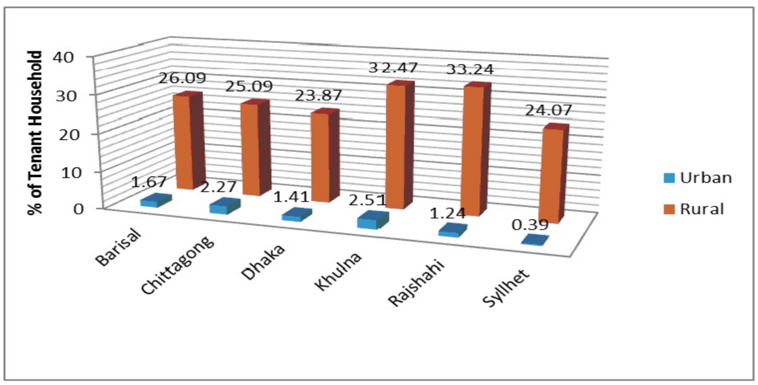

Figure 7.8: Distribution of Tenant Households by Urban and Rural.

7.3 Comparison

7.3.1 Introduction

This section presents a vibrant comparison among the data available from the 2008 Agriculture Census with the corresponding data available from previous censuses and surveys.

Division	Total HH('000)	No of Agriculture Farm HH(000)			% of Agriculture Farm HH		
		Total	Urban	Rural	Total	Urban	Rural
Barisal	1729	1126	26	1100	65.12	1.50	63.62
Chittagong	4887	2449	52	2397	50.11	1.06	49.05
Dhaka	9437	4060	125	3934	43.02	1.32	41.70
Khulna	3430	2027	45	1983	59.09	1.31	57.81
Rajshahi	7654	4273	70	4202	55.83	0.91	54.90
Syllhet	1533	782	12	770	51.01	0.78	50.23
Bangladesh	28670	14716	330	14387	51.33	1.15	50.18

Table 7.5: Distribution of Agriculture Farm Households by Urban, Rural and Division.

7.3.2 Comparison of Household Numbers

A comparative picture of the division wise number of dwelling households obtained in the Agriculture Census, 2008 with those of the Population Census, 2001 is given below:

The table 7.8 shows that the division wise percentage of dwelling households remains almost unchanged though the growth of households over the period is quite different for different divisions. At the national level it shows that there has been 15.4% increase in number of dwelling households over a period of 7 years. The growth rate of households is however the highest (18.5%) for Dhaka Division closely followed by Rajshahi Division (17.0%) and the lowest for Barisal Division (7.1%). The national growth in terms of compound rate has been estimated at 2% per annum.

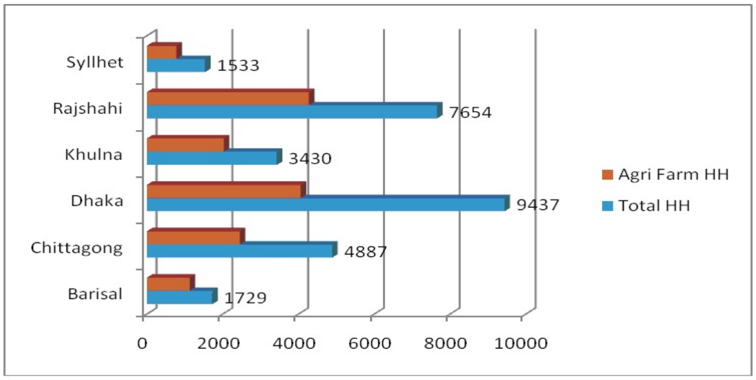

Figure 7.9: Distribution of Agriculture Farm Households from Total Households by Division.

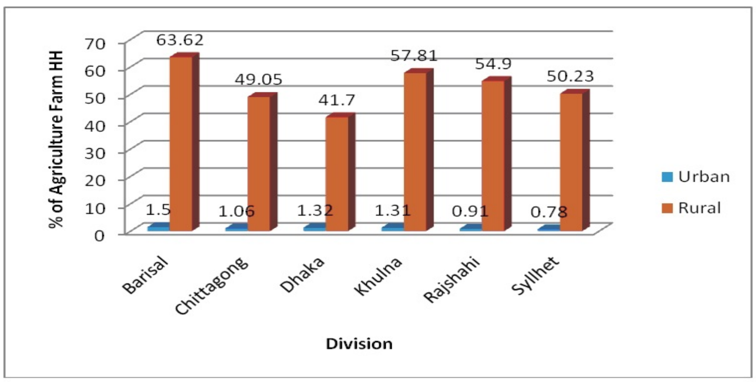

Figure 7.10: Distribution of Agriculture Farm Households by Rural and Urban.

7.3.3 Comparison of Farm Households (Rural Only)

One of the most important reasons for holding Agriculture Census is to measure the structural changes in the agriculture sector over time. The 1977 census of agriculture was actually large scale sample survey covering rural areas only. In 1983-84, the census on a full count basis was carried out in 3 phases viz. (1) 50% of the rural Unions excluding those of the Chittagong Hill Tracts were enumerated in 1983 (2) The rest of the Unions were enumerated in 1984 and (3) Using a separate questionnaire the urban municipal census was carried out in 1984 together with a census of Institutional Holdings. The results were thus published separately for rural and urban areas.

Division	Total Farm HH('000)	No of Boro Cultivating HH('000)			% of Boro Cultivating HH		
		Total	Urban	Rural	Total	Urban	Rural
Barisal	1126	71	2	70	6.31	0.18	6.13
Chittagong	2449	392	6	386	16.01	0.24	15.77
Dhaka	4060	591	20	571	14.56	0.49	14.07
Khulna	2027	453	7	446	22.33	0.33	22.00
Rajshahi	4273	1163	13	1150	27.22	0.30	26.92
Syllhet	782	104	1	103	13.32	0.13	13.17
Bangladesh	14716	2774	50	2725	18.85	0.34	18.51

Table 7.6: Distribution of Boro Cultivating Households by Urban, Rural and Division.

In 1996, it was originally planned to conduct Agricultural Census in the rural areas and then conduct census in Municipal areas using a modified questionnaire. But due to a devastating flood that engulfed about two-third of the country, the census in Municipal areas could not be completed.

Thus, in order to measure the structural changes in agriculture over time, there is no alternative but to keep the comparative study limited to rural areas only. However, the domain of the 'rural area' itself is also different in different censuses. For example, in 1983-84, there were only 79 Municipalities. Its number increased to 147 in 1996. In 2008 Agriculture Census, 6 Metropolitan cities of the Divisional Headquarters and 58 Municipalities located at 58 other District Headquarters, were included in 'urban area'. The rest of the country which included Municipalities at the Upazila Headquarters was defined as 'rural area'.

In 2008, for the first time, the census of agriculture in Bangladesh was conducted both in urban and rural areas simultaneously using the same questionnaire. However, for comparison with previous censuses, only the 'rural area' data of the censuses have been utilized, although the domain of the

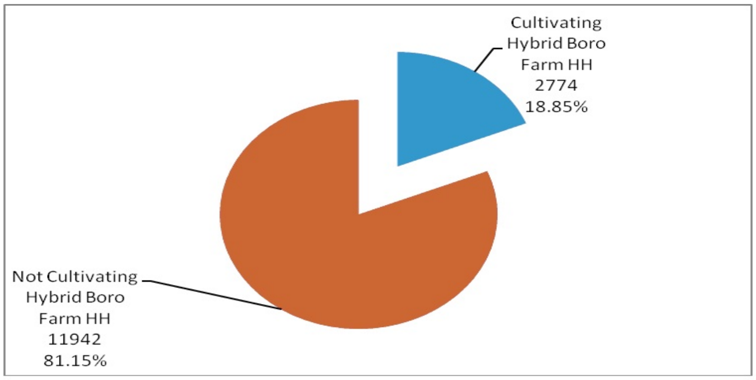

Figure 7.11: Distribution of Farm HH Cultivating *Hybrid Boro*.

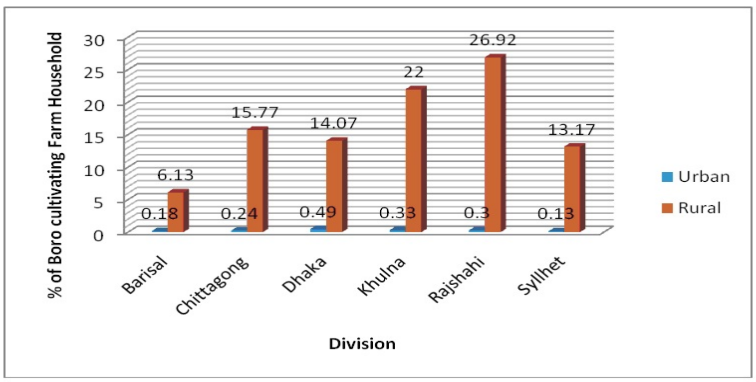

Figure 7.12: Distribution of Farm HH Cultivating *Hybrid Boro* by Rural and Urban.

'rural area' itself, as explained above, is also different in different censuses. Therefore, the users should bear in mind the limitation of the comparison.

The table 7.9 shows that though the absolute number of farm households is gradually increasing, yet the percent (as percent of total households) of farm households is gradually decreasing. A farm household is defined as a holding whose net cultivated area is 0.05 acre or more. In 1983-84, the percentage of farm households in the rural areas was 72.70%. It decreased to 66.18% and 56.74% in the year 1996 and 2008 respectively. The decreasing trend is also observed in every Division of Bangladesh although the rate of decrease is not uniform in all the Divisions. The causes behind the gradual decrease in percentage of farm households likely to be the rapid urbanization

Division	Total HH	No of *Hybrid Boro* cultivating HH	% of *Hybrid Boro* cultivating HH
Barisal	1731282	63086	3.6
Chittagong	4876160	352177	7.2
Dhaka	9456838	553901	5.9
Khulna	3435470	408711	11.9
Rajshahi	7663035	1052875	13.7
Syllhet	1526006	100894	6.6
Bangladesh	28688791	2531644	8.82

Table 7.7: Distribution of *Hybrid Boro* Cultivating Households by Division.

throughout the country and many of the people are switching over to the non-agriculture sector from agriculture sector.

7.3.4 Comparison of Agriculture labor Households

The table 7.10 presents the number of agriculture labor households obtained from Agriculture Censuses conducted in the year 1983-84, 1996 and 2008.

The table 7.10 shows that in the rural area there are about 8.85 million households (about 34.90%) whose main source of income comes from work as agriculture labor. Percentage of agriculture labor households is gradually decreasing. It has decreased to 34.90% in 2008 from 39.77% in 1983-84. This decreasing trend compared with 1996 census is observed in Barisal, Dhaka and Sylhet Divisions whereas it shows an increasing trend in Chiittagong, Khulna and Rajshahi Division, though the increase in Chittagong Division is marginal (about 1%). In terms of percentage of total households Rajshahi stands at the top with 43.98% agriculture labor households followed by Khulna Division (41.42%) and Dhaka Division stands at the bottom with

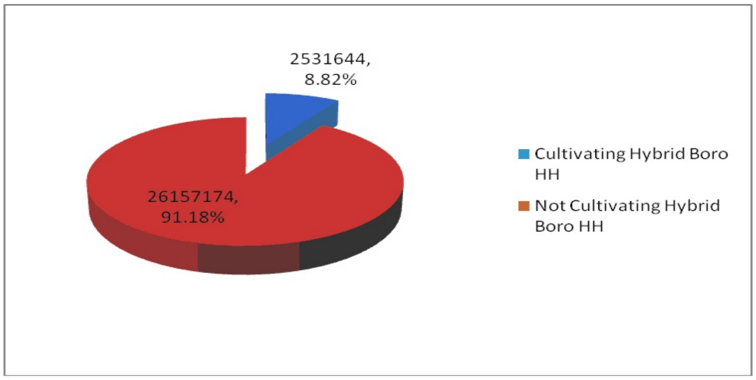

Figure 7.13: Distribution of Cultivating *Hybrid Boro* Households.

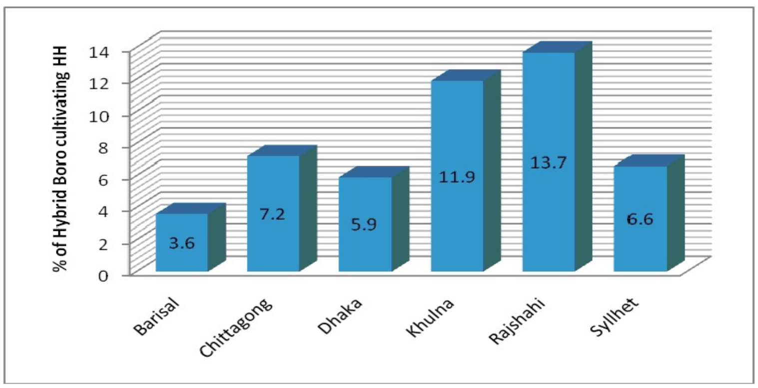

Figure 7.14: Distribution of *Hybrid Boro* Cultivating Households by Division.

28.14% agriculture labor households. During the previous census, Rajshahi Division was also at the top with 41.11% agriculture labor households and Chittagong Division reported the lowest with 28.86% agriculture labor households.

Division	Population census-2001		Agriculture census-2008		
	No of dwelling HH('000)	%	No of dwelling HH('000)	%	% increse
Barisal	1614	6.5	1729	6.0	7.1
Chittagong	4315	17.4	4887	17.0	13.3
Dhaka	7962	32.0	9437	32.9	18.5
Khulna	3060	12.3	3430	12.1	12.1
Rajshahi	6544	26.3	7654	26.7	17.0
Sylhet	1354	5.5	1533	5.3	13.2
Bangladesh	24850	100.0	28670	100.0	15.4

Table 7.8: A Comparative Position with the Population Census-2001.

7.3.5 Rural Landless Households of 1983-84, 1996 and 2008 Census of Agriculture by Division

The table 7.11 shows that absolute landlessness (households owning no land whatsoever) is steadily increasing over time. According to Agriculture Census, 2008 total number of absolute landless households in the 'Rural' area is about 3.26 million which is about 12.84% of total households in 'Rural' area. The current rate of landlessness is 12.84% in rural areas against 10.18% in 1996 and 8.67% in 1983-84. The increasing trend is observed in all the Divisions except in Barisal Division. A sharp increase in Dhaka and Chittagong Division could be partly attributed to the increasing trend in urbanization around metropolitan areas and partly to the possible in-migration of landless people from 4 other Divisions. In respect of Division, it is observed that Dhaka Division shows the highest percentage of landless household closely followed by Sylhet and Rajshahi Divisions. Barisal Division shows the lowest percentage (8.44%) of rural landless households whereas Sylhet Division was at the top in respect of landless households in the previous censuses.

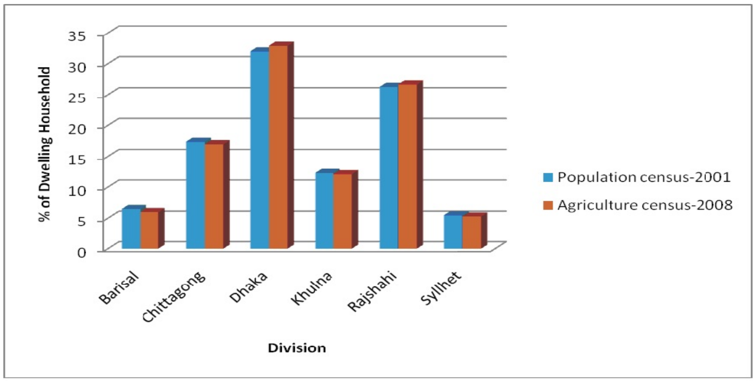

Figure 7.15: Distribution of Dwelling Households.

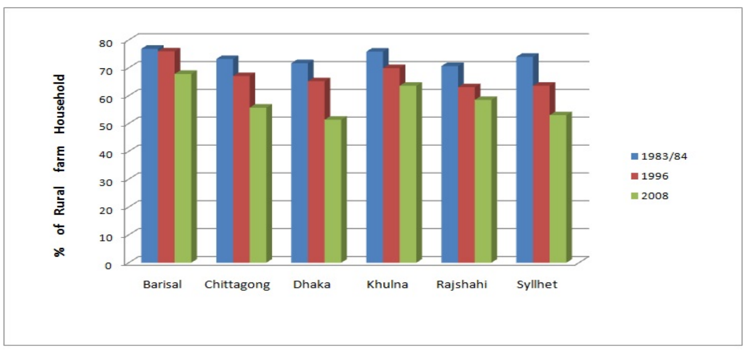

Figure 7.16: Distribution of Rural Farm Households.

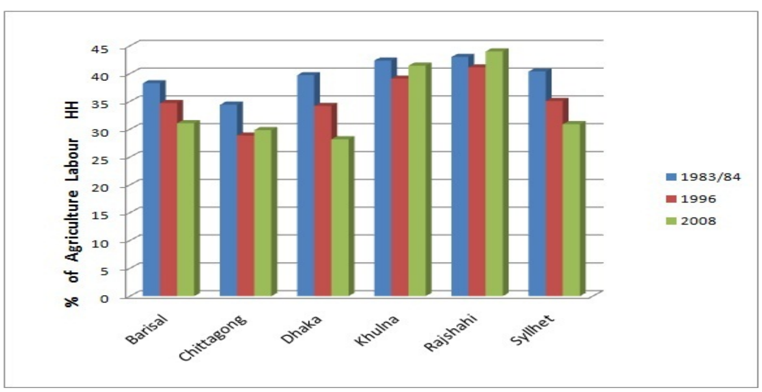

Figure 7.17: Distribution of Agriculture labor Households.

Division	1983/84-HH('000)			1996-HH('000)			2008-HH('000)		
	Tot	Farm	%	Tot	Farm	%	Tot	Farm	%
Barisal	1100	845	76.82	1352	1026	75.89	1623	1100	67.79
Chittagong	2686	1965	73.16	3174	2127	67.01	4307	2397	55.66
Dhaka	3969	2843	71.63	4939	3219	65.18	7661	3934	51.36
Khulna	1566	1187	75.80	2174	1519	69.87	3121	1983	63.53
Rajshahi	3588	2533	70.60	5084	3204	63.02	7190	4202	58.45
Sylhet	909	672	73.93	1106	703	63.56	1453	770	52.99
Bangladesh	13818	10045	72.70	17828	11798	66.18	25355	14387	56.74

Table 7.9: Rural Farm Households of 1983-84, 1996 and 2008 Census of Agriculture by Division.

Division	Agriculture labor households in Agriculture census of								
	1983/84-HH('000)			1996-HH('000)			2008-HH('000)		
	Tot HH	AgL HH	%	Tot HH	AgL HH	%	Tot HH	AgL HH	%
Barisal	1100	421	38.27	1352	469	34.69	1623	504	31.07
Chittagong	2686	924	34040	3174	916	28.86	4307	1285	29.84
Dhaka	3969	1576	39.71	4939	1688	34.18	7661	2156	28.18
Khulna	1566	663	42.34	2174	850	39.10	3121	1293	41.42
Rajshahi	3588	1542	42.98	5084	2090	41.11	7190	3162	43.98
Sylhet	909	367	40.37	1106	388	35.08	1453	449	30.90
Bangladesh	13818	5495	39.77	17828	6401	35.90	25355	8849	34.9

Table 7.10: Rural Agriculture labor Households of 1983-84, 1996 and 2008 Census of Agriculture by Division.

Division	Rural households with no own land in Agriculture census of								
	1983/84-HH('000)			1996-HH('000)			2008-HH('000)		
	Tot HH	no own Land	%	Tot HH	no own Land	%	Tot HH	no own Land	%
Barisal	1100	80	7.27	1352	123	9.10	1623	137	8.44
Chittagong	2686	130	4.84	3174	236	7.44	4307	475	11.04
Dhaka	3969	360	9.07	4939	454	9.19	7661	1129	14.73
Khulna	1566	130	8.30	2174	167	7.68	3121	298	9.56
Rajshahi	3588	390	10.87	5084	673	13.24	7190	1006	13.99
Sylhet	909	108	11.88	1106	161	14.56	1453	211	14.52
Bangladesh	13818	1198	8.67	17828	1815	10.18	25355	3256	12.84

Table 7.11: Rural Landless Households of 1983-84, 1996 and 2008 Census of Agriculture by Division.

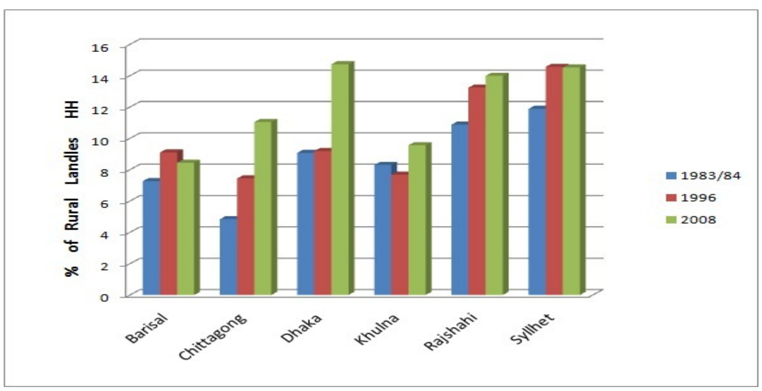

Figure 7.18: Distribution of Rural Landless Households.

Chapter 8

Adaptive Cluster Sampling For Estimating Hybrid Boro Usage

8.1 Introducton

The adaptive cluster sampling methodologies described in chapter 4 is modified and customized for use in estimating the Hybrid Boro usage in Bangladesh using the Agricultural census data as the population of interest. Let the variable of interest, the *cultivation status of hybrid boro*, be denoted by y where

$$y_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ HH cultivates hybrid boro} \\ 0 & \text{otherwise ,} \end{cases} \tag{8.1}$$

The variable *household id* indicates just the case number which is actually the serial number provided at the primary data collection stage. Households with successive case number can thus be assumed to be *adjacent HH*. A Monte Carlo Simulation is conducted (see Chapter 9) to see whether an adaptive sampling strategy could be more feasible to reveal information regarding using the *hybrid boro* by the HH. Two adaptive cluster sampling strategies appropriate for this scenario are described in the following sections.

8.2 Adaptive Cluster Sampling for *Hybrid Boro* Usage

The method developed by Thompson (1990) and described in section 4.1.1 is modified for the situation of finding coverage of *Hybrid Boro* usage in agriculture in Bangladesh. The stratification under division as done in the Agricultural Census is maintained, hence the sampling is conducted independently in each of the divisions. The procedure can be summarized as follows:

1. From the listing of the HH in the h^{th} division, an initial set of n_h HHs is selected by simple random sampling (SRS) without replacement.
2. If any of the selected HH satisfies the condition that $C : y_{ih} = 1$, the adjacent HH, i.e. the HH with the case number before and after the selected HH is also included in the sample.
3. The adjacent HH of these selected HHs are also checked for the condition and selected in the sample if satisfied.
4. The process is continued until a HH not satisfying the condition is found. As a result a network or cluster of HHs is selected from each of the HH selected in the initial sample.

8.2.1 Estimator of the Population Proportion

The ultimate sample is then utilized to obtain the value of the estimators described in equations (4.6), (4.11), (4.7) and (4.9). Note that, y being an indicator variable, the estimators of the population mean $\hat{\mu}$ would be the estimator of the proportion of HH cultivating *Hybrid Boro*. The formulae can be re-written as

$$\hat{P}_h = \frac{1}{N_h} \sum_{k=1}^{\kappa_h} \frac{y^*_{kh}}{\alpha_{kh}} \ , \tag{8.2}$$

where N_h and P_h are the population size and population proportion of HH cultivating *hybrid boro* in the h^{th} division, y^*_{kh} is the sum of the y- values for the k^{th} network in the h^{th} division, κ_h is the number of distinct networks in the sample from the h^{th} division, and α_{kh} is the inclusion probability of the k_h^{th} network only if the initial sample intersects it. If there are x_{kh} sampling units in the k^{th} network of the h^{th} division, then

$$\alpha_{kh} = 1 - \binom{N_h - x_{hk}}{n_h} \times \left[\binom{N_h}{n_h}\right]^{-1} . \tag{8.3}$$

Let p_{jkh} to be the probability that the j^{th} and k^{th} networks of the h^{th} division are not intersected, then

$$p_{jkh} = \binom{N_h - x_{jh} - x_{kh}}{n_h} \times \left[\binom{N_h}{n_h}\right]^{-1} . \tag{8.4}$$

So that the joint probability that networks j and k of the h^{th} division are both intersected is

$$\begin{aligned} \alpha_{jkh} &= \alpha_{jh} + \alpha_{kh} - (1 - p_{jkh}) \\ &= 1 - \left[\binom{N_h - x_{jh}}{n_h} + \binom{N_h - x_{kh}}{n_h} - \binom{N_h - x_{jh} - x_{kh}}{n_h}\right]\left[\binom{N_h}{n_h}\right]^{-1} \end{aligned} \tag{8.5}$$

Now the variance can be derived as

$$Var\left(\hat{P}_h\right) = \frac{1}{N_h^2}\left[\sum_{j=1}^{K_h}\sum_{k=1}^{K_h} \frac{y_{jh}^* y_{kh}^* \left(\alpha_{jkh} - \alpha_{jh}\alpha_{kh}\right)}{\alpha_{jh}\alpha_{kh}}\right] . \tag{8.6}$$

And the unbiased estimator of the variance of $\hat{P}_h$ is given by

$$\widehat{Var}\left(\hat{P}_h\right) = \frac{1}{N_h^2}\left[\sum_{j=1}^{\kappa_h}\sum_{k=1}^{\kappa_h} \frac{y_{jh}^* y_{kh}^*}{\alpha_{jh}} \left(\frac{\alpha_{jkh}}{\alpha_{jh}\alpha_{kh}} - 1\right)\right] , \tag{8.7}$$

where K_h denotes the total number of distinct networks in the h^{th} division. Let us also denote N and L as the total population size and number of strata respectively. The overall estimator of the population proportion P is thus given by

$$\hat{P} = \sum_{h=1}^{L} \frac{N_h}{N}\hat{P}_h , \tag{8.8}$$

with

$$Var\left(\hat{P}\right) = \sum_{h=1}^{L} \frac{N_h^2}{N^2} Var\left(\hat{P}_h\right) \ , \tag{8.9}$$

and

$$\widehat{Var}\left(\hat{P}\right) = \sum_{h=1}^{L} \frac{N_h^2}{N^2} \widehat{Var}\left(\hat{P}_h\right) \ . \tag{8.10}$$

8.3 Two Stage Adaptive Cluster Sampling for *Hybrid Boro* Usage

In Agricultural Census 2008, the sampling design adopted was a stratified two stage sampling where the primary sampling units (PSU) were the Enumeration areas (EA) each consisting of about 120 HH. Considering these first stage units, a two stage adaptive cluster sampling is also thought about. The two-stage adaptive cluster sampling method for finding coverage of *Hybrid Boro* usage is also designed in a similar way to the method described in section 8.2. The procedure can be summarized as follows:

1. From the listing of the PSUs in the h^{th} division, an initial set of n_{1h} PSUs is selected by simple random sampling (SRS).

2. From the listing of the HH in the m^{th} selected PSUs, an initial set of n_{2mh} HHs is selected by simple random sampling (SRS) without replacement.

3. If any of the selected HH satisfies the condition that $C : y_{ijh} = 1$, the adjacent HH, i.e. the HH with the case number before and after the selected HH is also included in the sample.

4. The adjacent HH of these selected HHs are also checked for the condition and selected in the sample if satisfied.

5. The process is continued until a HH not satisfying the condition is found. As a result a network or cluster of HHs is selected from each of the HH selected in the initial sample.

8.3.1 Estimator of the Population Proportion

The ultimate sample is then utilized to obtain the value of the estimators by modifying the equations (4.6), (4.11), (4.7) and (4.9). The formulae for estimation can be re-written as

$$\hat{P_{mh}} = \frac{1}{N_{mh}} \sum_{k=1}^{\kappa_{mh}} \frac{y^*_{kmh}}{\alpha_{kmh}} \ , \tag{8.11}$$

where N_{mh} and P_{mh} are the population size and population proportion of HH cultivating *hybrid boro* in the m^{th} selected PSU of the h^{th} division, y^*_{kmh} is the sum of the y- values for the k^{th} network in the m^{th} selected PSU of the h^{th} division, κ_{mh} is the number of distinct networks in the sample from the m^{th} selected PSU of the h^{th} division, and α_{kmh} is the inclusion probability of the k^{th}_{mh} network only if the initial sample intersects it. If there are x_{kmh} sampling units in the k^{th} network in the m^{th} selected PSU of the h^{th} division, then

$$\alpha_{kmh} = 1 - \binom{N_{mh} - x_{hk}}{n_{2mh}} \times \left[\binom{N_{mh}}{n_{2mh}}\right]^{-1} \ . \tag{8.12}$$

Let p_{jkmh} to be the probability that the j^{th} and k^{th} networks in the m^{th} selected PSU of the h^{th} division are not intersected, then

$$p_{jkmh} = \binom{N_{mh} - x_{jmh} - x_{kmh}}{n_{2mh}} \times \left[\binom{N_{mh}}{n_{2mh}}\right]^{-1} \ . \tag{8.13}$$

So that the joint probability that networks j and k in the m^{th} selected PSU of the h^{th} division are both intersected is

$$\begin{aligned} \alpha_{jkmh} &= \alpha_{jmh} + \alpha_{kmh} - (1 - p_{jkmh}) \\ &= 1 - \left[\binom{N_{mh} - x_{jmh}}{n_{2mh}} + \binom{N_h - x_{kmh}}{n_{2mh}} \right. \\ &\quad \left. - \binom{N_{mh} - x_{jmh} - x_{kmh}}{n_{2mh}}\right] \left[\binom{N_{mh}}{n_{2mh}}\right]^{-1} \ . \end{aligned} \tag{8.14}$$

Now the variance can be derived as

$$Var\left(\hat{P_{mh}}\right) = \frac{1}{N^2_{mh}} \left[\sum_{j=1}^{K_{mh}} \sum_{k=1}^{K_{mh}} \frac{y^*_{jmh} y^*_{kmh} \left(\alpha_{jkmh} - \alpha_{jmh}\alpha_{kmh}\right)}{\alpha_{jmh}\alpha_{kmh}}\right] \ . \tag{8.15}$$

And the unbiased estimator of the variance of $\hat{P}_h$ is given by

$$\widehat{Var}\left(\hat{P_{mh}}\right) = \frac{1}{N_{mh}^2}\left[\sum_{j=1}^{\kappa_{mh}}\sum_{k=1}^{\kappa_{mh}} \frac{y_{jmh}^* y_{kmh}^*}{\alpha_{jmh}}\left(\frac{\alpha_{jkmh}}{\alpha_{jmh}\alpha_{kmh}} - 1\right)\right] , \tag{8.16}$$

where K_{mh} denotes the total number of distinct networks in the h^{th} division. Let us also denote N and L as the total population size and number of strata respectively. The divisional estimator $(\hat{P}_h)$ and the overall estimator $(\hat{P})$ of the population proportion P is thus given by

$$\hat{P}_h = \sum_{l=1}^{n_{1h}} \frac{N_{mh}}{n_{2mh}} \hat{P_{mh}} , \tag{8.17}$$

and

$$\hat{P} = \sum_{h=1}^{L}\sum_{l=1}^{n_{1h}} \frac{N_{mh}}{n_{2mh}} \frac{N_{mh}}{N} \hat{P_{mh}} , \tag{8.18}$$

wheras the variance and estimated variance can be computed using two-stage sampling formulae.

Chapter 9

A Computer Simulation

9.1 Introducton

To see how the adaptive cluster sampling would perform in identifying the coverage of use of Hybrid seeds in agriculture of Bangladesh, a Monte-Carlo simulation is conducted where the Agricultural Census data is considered as to be the population. In our study, we are interested only on the *cultivation status of Hybrid Boro* out of large number of variables in agriculture census 2008. The variable *cultivation status of Hybrid Boro* is denoted by y and defined as in equation (8.1). The variable *household id* indicates just the case number which is actually the serial number provided at the primary data collection stage. Households with successive case number can thus be assumed to be *adjacent HH*. A Monte Carlo Simulation is conducted to see whether an adaptive sampling strategy could be more feasible to reveal information regarding using the *Hybrid Boro* by the HH. Using the methods described in sections 8.2 and 8.3 are utilized to draw samples and repeated iterations of the procedures are made. The equations (8.2), (8.8), (8.18), (8.17) and (8.18) are used to compute the estimated proportion at each of the iterations.

9.2 Simulation Set-ups

Considering the Agricultural census data as the population, sub-samples of different sizes are drawn from it using different sampling strategies and they are compared in respect of absolute percentage relative bias (APRB) and relative efficiencies (RE), see section 5.5.1 and 5.5.2. The different sample sizes considered are

1. 100,
2. 200,
3. 300, and
4. 500.

And the simulation is done for the following sampling methods:

1. Simple random sampling,
2. Single stage adaptive cluster sampling, and
3. Two stage adaptive cluster sampling.

For each method and each sample size, 500 iteration of sample was drawn and the Monte Carlo mean (MC mean) of the estimator of population proportion are computed using the following formulae:

$$\text{MC mean}\left(\hat{P}_h\right) = \frac{1}{k}\sum_{i=1}^{k}\hat{P}_{hi},\ (h = 1, 2, \ldots, L), \tag{9.1}$$

$$\text{MC mean}\left(\hat{P}\right) = \frac{1}{k}\sum_{i=1}^{k}\hat{P}_{i}, \tag{9.2}$$

where for each of the sampling methods, k stands for the number of iteration and $\hat{P}_{hi}$ denotes the estimated proportion in the i^{th}iteration in the h^{th} stratum. Similarly $\hat{P}_i$ denotes the estimated overall proportion in the i^{th}iteration. The Monte Carlo standard error (MC se) of the estimator can thus be computed using the formulae:

$$\text{MC se}\left(\hat{P}_h\right) = \frac{1}{k-1}\sum_{i=1}^{k}\left\{\hat{P}_{hi} - \text{MC mean}\left(\hat{P}_h\right)\right\}^2, \ (h = 1, 2, \ldots, L), \tag{9.3}$$

$$\text{MC se}\left(\hat{P}\right) = \frac{1}{k-1}\sum_{i=1}^{k}\left\{\hat{P}_i - \text{MC mean}\left(\hat{P}\right)\right\}^2 \tag{9.4}$$

The results of the simulation for different methods and different sample sizes are presented in Tables 9.1 through 9.12 and discussed in Sections 9.3 through 9.4.

9.3 Results for the SRS Estimators

Table 9.1 shows the MC mean, percentage absolute relative bias (APRB) and MC se of estimator of proportion of HH cultivating *Hybrid Boro* when sampling design is SRS.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0353	3.02%	0.01873
200	0.0368	1.1%	0.01331
300	0.0355	2.47%	0.01042
500	0.0368	1.1%	0.01042

Table 9.1: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using SRS for Barisal Division (True Proportion=0.036).

From the results of Table 9.1 it is observed that, APRB of SRS estimator is high when sample size 100. In addition, APRB is equal for both 200 and

500 sample sizes which is 1.1%. Also the APRB of SRS estimator is 2.47% when sample size is 300. Also it is clear that, MC se of SRS estimator is decreasing with the increase in sample size.

The same investigation as we did for barisal division is also made for Chittagong division through simulation study. The results we obtained for Chittagong division are shown in Table 9.2.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.072	0%	0.0272
200	0.0724	0.56%	0.0188
300	0.0722	0.28%	0.0157
500	0.0718	0.28%	0.0117

Table 9.2: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using SRS for Chittagong Division (True Proportion=0.072).

From Table 9.2 we observed that, APRB of SRS estimator is 0% when sample size is 100, APRB is equal when sample size is 300 and 500 which is 0.28%. We also observed that APRB is high when sample size is 200.

The results for Dhaka division are shown in Table 9.3.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0588	0.34%	0.0235
200	0.0597	1.19%	0.0176
300	0.0579	1.86%	0.0139
500	0.0580	1.69%	0.0109

Table 9.3: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using SRS for Dhaka Division (True Proportion=0.059).

From the results for Dhaka division in Table 9.3 it is observed that, although APRB is small when sample size is 100 but the MC se of the SRS estimator is high relative to the other sample sizes we considered in our investigation. It is also clear that, although MC mean of SRS estimator is almost same for all the sample sizes we considered but APRB differ arbitrarily.

The simulation results for same things of khulna division is shown in Table 9.4.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.1178	1.02%	0.0320
200	0.1194	0.34%	0.0241
300	0.1188	0.17%	0.0188
500	0.1195	0.42%	0.0145

Table 9.4: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using SRS for Khulna Division (True Proportion=0.119).

From the results of Khulna Division in Table 9.4 it is observed that, the APRB is lowest for sample size 300 and highest for sample size 100. The MC mean of SRS estimator is almost same for all sample sizes we considered. But APRB and MC se of SRS estimator both are decreasing with increasing the sample size we considered except the sample size 500.

The simulation result of same investigation for Rajshahi Division in SRS are shown in the Table 9.5.

Table 9.5 shows the Monte Carlo mean, APRB and Monte Carlo Standard error of estimator of proportion of HH cultivating *Hybrid Boro* when sampling design is SRS. From the results of Table 9.5 it is observed that, APRB of SRS estimator is high when sample size 100. In addition, APRB is equal for both 300 and 500 sample sizes which is 0.29%. Also the APRB of SRS estimator is 0.36% when sample size is 200. Also it is clear that, MC se of SRS estimator is decreasing with the increase in sample size. But same as earlier Divisions, MC mean of SRS estimator is almost same for all sample size we considered.

The Table 9.6 shows the same investigation of SRS simulation results for Syllhe Division. The Table 9.6 shows that, APRB of SRS estimator is high

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.1383	0.95%	0.0353
200	0.1375	0.36%	0.0264
300	0.1366	0.29%	0.0193
500	0.1374	0.29%	0.0155

Table 9.5: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using SRS for Rajshahi Division (True Proportion=0.137) .

when sample size 200 and it is smallest for sample size 300.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0671	1.67%	0.0243
200	0.0677	2.7%	0.0182
300	0.0666	0.91%	0.0147
500	0.0672	1.82%	0.0114

Table 9.6: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using SRS for Syllhet Division (True Proportion=0.066).

It is also clear from the Table 9.6 that, although MC mean of SRS estimator is almost same for all sample size we considered but APRB differ arbitrarily. In addition, MC se in SRS is decreasing when sample size increasing.

9.4 Results for the ACS Estimators

The results obtained from simulation study using adaptive cluster sampling design would be described with respect to different sample sizes. Results for the Barisal division are shown in Table 9.7.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0361	0.82%	0.0409
200	0.0368	1.1%	0.0392
300	0.0368	1.1%	0.0386
500	0.0367	0.82%	0.0376

Table 9.7: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using ACS for Barisal Division. (True Proportion=0.036)

The Table 9.7 shows that the highest APRB in ACS is 1.1% for sample size both 200 and 300. And the lowest APRB 0.82% is same for sample size 100 and 500 in Barisal division. It is also clear that, MC mean of sample proportion in ACS are allmost same for all sample size we considered in our study. Again, MC se of the estimator are decreasing as the sample size increasing. Figures 9.1 and 9.2 show a comparative display of the MC APRB and MC se of the ACS and SRS methods.

Figures 9.1 and 9.2 reveal that APRB of ACS estimator of the proportion of HH cultivating *hybrid boro* in Barisal division is smaller than that of the SRS method for all sample sizes. On the other hand, the MC se of ACS

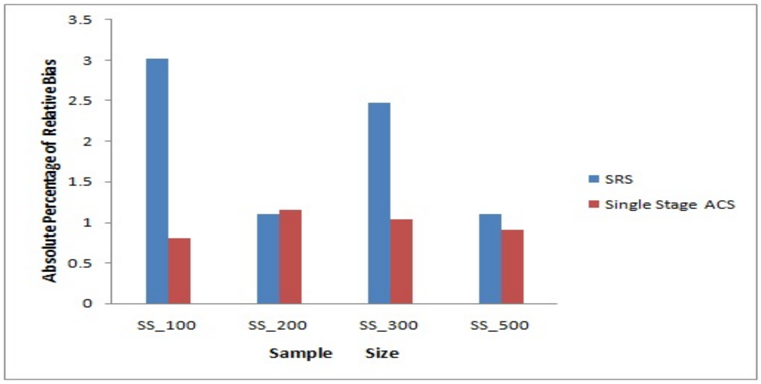

Figure 9.1: Comparison of APRB of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Barisal Division.

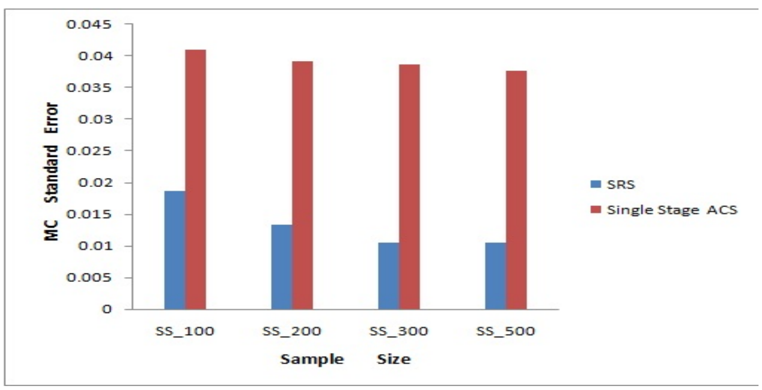

Figure 9.2: Comparison of MC se of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Barisal Division.

estimator of the proportion of HH cultivating *Hybrid Boro* in Barisal division is higher than that of the SRS estimator.

The Table 9.8 shows the MC mean, APRB, Mc se of estimator for proportion of HH cultivating *Hybrid Boro* using ACS method for Chittagong division.

From the Table 9.8 we see that APRB is highest for sample size 100 which is 1.11%. The Table 9.8 also shows that APRB is same for the sample size 200 and 300. The table 9.8 also display that both MC mean and MC se in ACS decrease as sample size increases. Figures 9.3 and 9.4 show a comparative

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0728	1.11%	0.0773
200	0.0725	0.69%	0.0747
300	0.0725	0.69%	0.0739
500	0.0723	0.42%	0.0732

Table 9.8: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using ACS for Chittagong Division (True Proportion=0.072).

display of the MC APRB and MC se of the ACS and SRS methods.

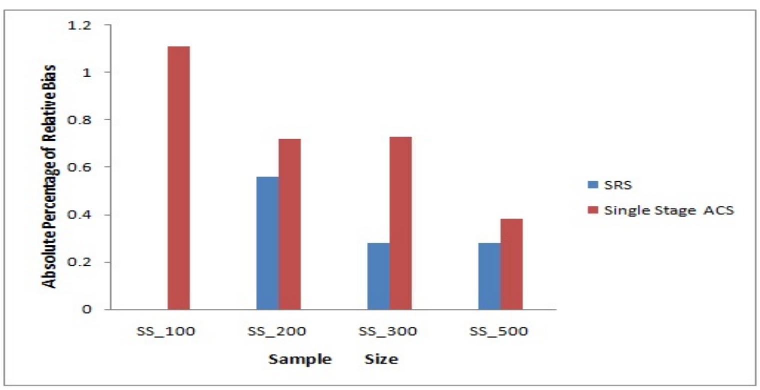

Figure 9.3: Comparison of APRB of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Chittagong Division.

It can be observed from the Figures 9.3 and 9.4 that estimator of the proportion of HH cultivating *hybrid boro* in Chittagong division has smaller APRB for ACS method than that for SRS method. The MC se of the ACS estimator of proportion of HH cultivating *hybrid boro* in Barisal division is

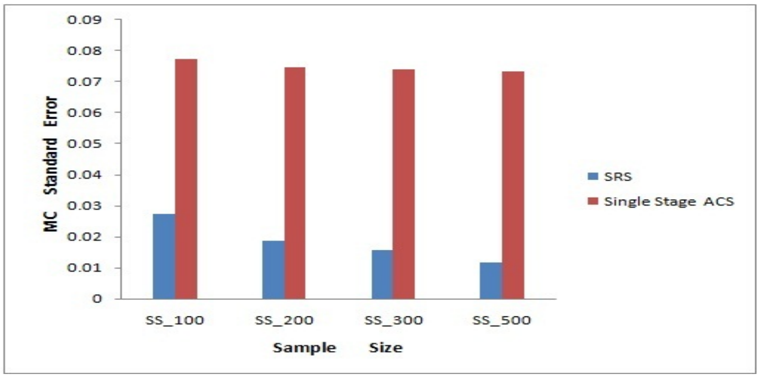

Figure 9.4: Comparison of MC se of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Chittagong Division.

higher than that of the SRS estimator for all sample sizes considered.

The simulation procedure is repeated for Dhaka division, results are shown in the table 9.9.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0582	1.36%	0.0627
200	0.0591	0.17%	0.0613
300	0.0585	0.85%	0.0599
500	0.0585	0.85%	0.0596

Table 9.9: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using ACS for Dhaka Division (True Proportion=0.059).

The Table 9.9 shows all of the MC mean are close to each other but MC mean 0.0591 for sample size 200 is slightly larger compare to that of other sample sizes. The table also disclose that, MC mean of single stage ACS estimator is almost same for all sample size we considered.

The table 9.10 shows highest APRB 0.50% in single stage ACS is for 100 sample size and that are lowest 0.17% for sample size 200 but for sample size 300 and 500 it produces the same APRB which is 0.25%. Like other divisions MC se is decreasing with sample size increasing in Khulna division. Figures 9.5 and 9.6 show a comparative display of the MC APRB and MC se of ACS and SRS methods.

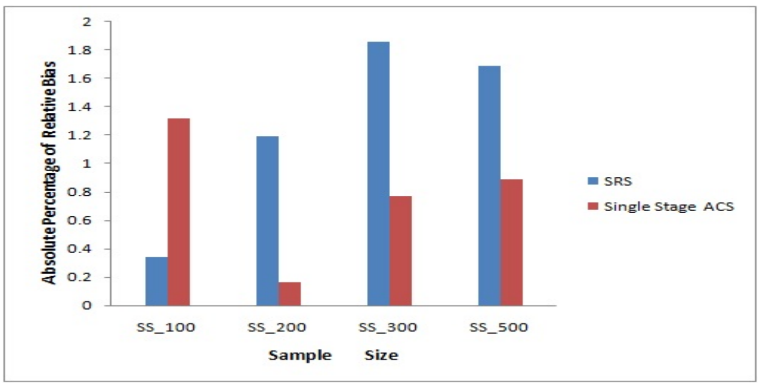

Figure 9.5: Comparison of APRB of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Dhaka Division.

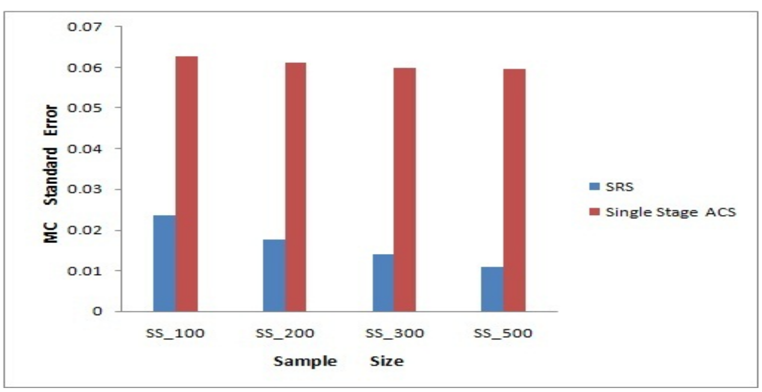

Figure 9.6: Comparison of MC se of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Dhaka Division.

Figures 9.5 and 9.6 show features similar to the other divisions, that is ACS method produces smaller APRB and higher MC se than the SRS method does for estimating the proportion of HH cultivating *hybrid boro* in Dhaka division.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.1184	0.50%	0.1229
200	0.1192	0.17%	0.1217
300	0.1187	0.25%	0.1203
500	0.1193	0.25%	0.1203

Table 9.10: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using ACS for Khulna Division (True Proportion=0.119).

The Table 9.10 also disclose that MC mean are almost same for all sample size we considered. The Table 9.11 display the ACS simulation result for Rajshahi division. Figures 9.7 and 9.8 show a comparative display of the MC APRB and MC se of ACS and SRS methods.

From Figures 9.7 and 9.8, it can be revealed that for Khulna division too, the APRB of the ACS estimator of proportion of HH cultivating *hybrid boro* is smaller than that of SRS estimator and the MC se of ACS method is higher than SRS method has.

From the Table 9.11 we see that highest APRB 0.88% for sample size 300 and lowest APRB 0.29% with sample size 500. The APRB 0.51% is same for both sample size100 and 200. MC se are same in almost all the sample size. Figures 9.9 and 9.10 show a comparative display of the MC APRB and MC se of ACS and SRS methods.

From Figures 9.9 and 9.10, Rajshahi division results were also found similar to those in other divisions.

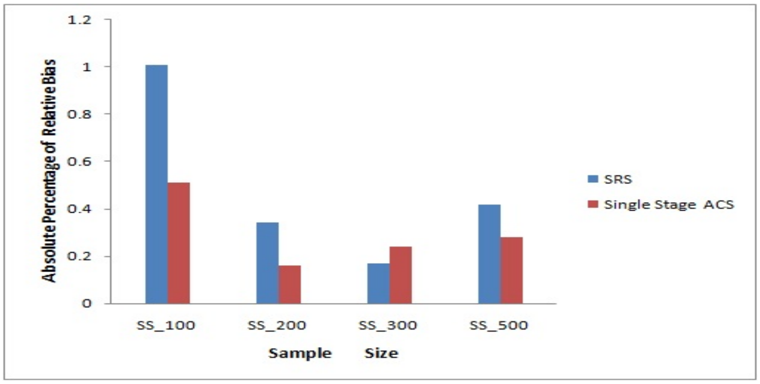

Figure 9.7: Comparison of APRB of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Khulna Division.

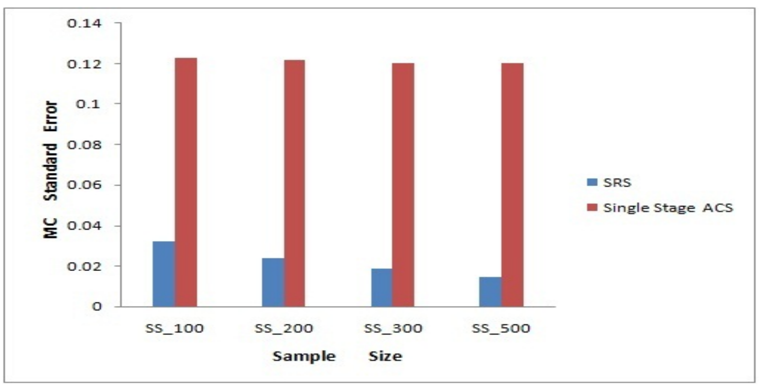

Figure 9.8: Comparison of MC se of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Khulna Division.

Figures 9.11 and 9.12 show a comparative display of the MC APRB and MC se of ACS and SRS methods.

Figures 9.11 and 9.12 reveal that APRB of ACS estimator of the proportion of HH cultivating *Hybrid Boro* in Barisal division is smaller than that of the SRS method for all sample sizes. On the other hand, the MC se of the ACS estimator of proportion of HH cultivating *Hybrid Boro* in Barisal division is higher than that of the SRS estimator.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.1363	0.51%	0.1405
200	0.1377	0.51%	0.1401
300	0.1382	0.88%	0.1398
500	0.1374	0.29%	0.1383

Table 9.11: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using ACS for Rajshahi Division (True Proportion=0.137).

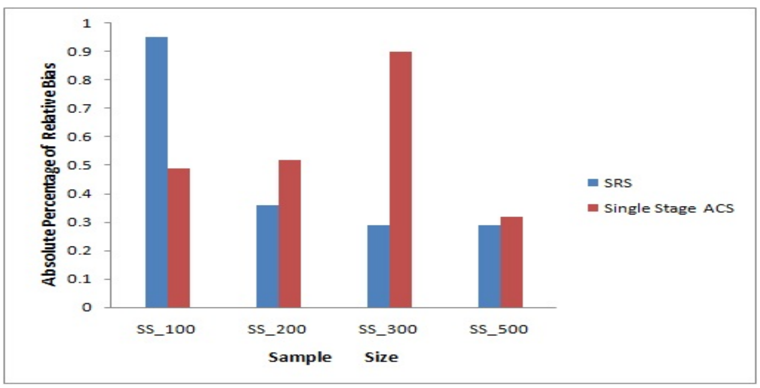

Figure 9.9: Comparison of APRB of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Rajshahi Division.

9.4.1 The Ultimate Sample Size in ACS Estimators

In adaptive cluster sampling (ACS) the ultimate sample size may vary from the initial sample size. If the ultimate sample is too higher than the initial sample size then it could reduce the applicability of ACS. It may be due to increase in the cost as well as time. From this perspective, we investigated

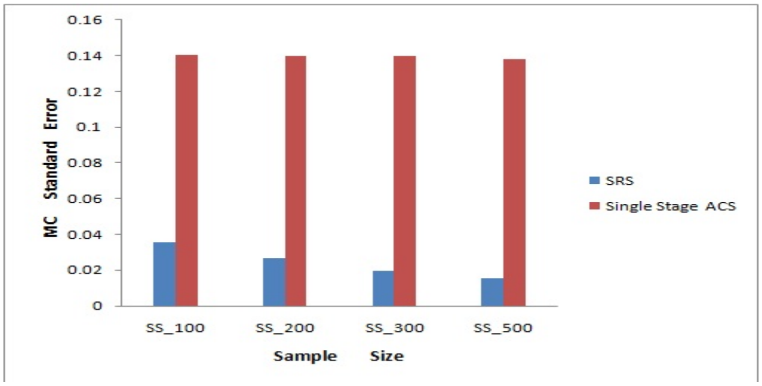

Figure 9.10: Comparison of MC se of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Rajshahi Division.

Sample size	MC mean of estimated proportion	APRB	MC se
100	0.0653	1.06%	0.0698
200	0.0655	0.76%	0.0678
300	0.0657	0.45%	0.0675
500	0.0659	0.15%	0.0670

Table 9.12: MC mean, APRB and MC se of Estimator of Proportion of HH Cultivating *Hybrid Boro* Using ACS for Syllhet Division(True Proportion=0.066) .

the ultimate sample size (USS) in ACS for each division separately as well as for different initial sample size. The results of our simulation study related to the USS in ACS are summarized in Table 9.13 through Table 9.18. Results of USS in ACS found in simulation study for Barisal division are shown in Table 9.13.

From the results shown in Table 9.13, it is seems that the USS in ACS

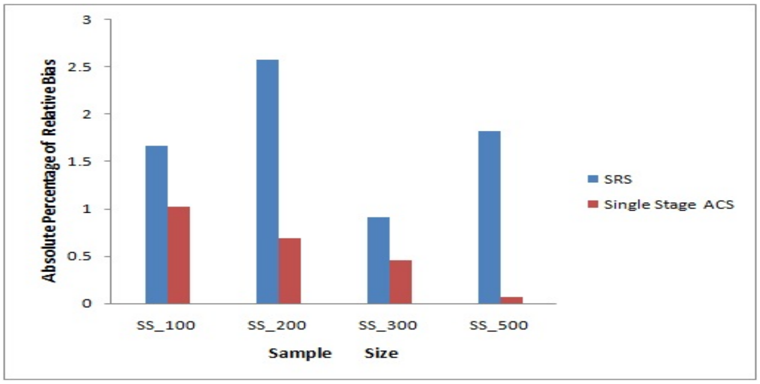

Figure 9.11: Comparison of APRB of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Syllhet Division.

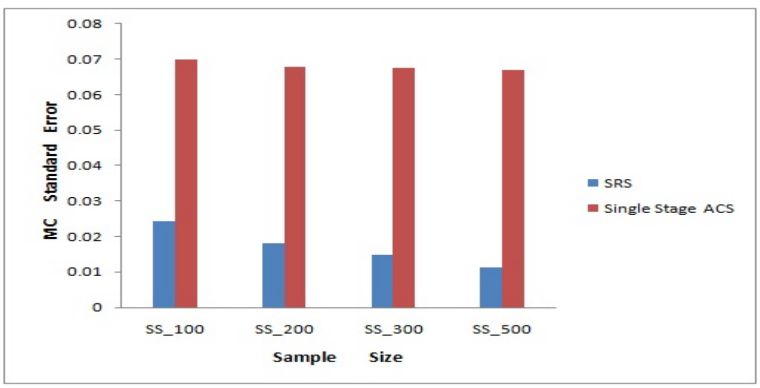

Figure 9.12: Comparison of MC se of Sample Proportion of HH Cultivating *Hybrid Boro* Between SRS and ACS Method Over Different Sample Size for Syllhet Division.

is not too large relative to the initial sample size. Specifically, in barisal division, when initial sample size was 100 then we found USS is 112. So it is clear that the difference between initial and ultimate sample size is not large enough. Empirically, we observed a relationship between initial sample size and USS. From our investigation, it can be says that the USS would be 12% higher than the initial sample size in Barisal division.

A similar investigation is also made for Chittagong division and the results obtained from the investigation are represented in Table 9.14.

Initial sample size	MC mean of estimated proportion	Ultimate sample size	
		mean	variance
100	0.0361	112	13167
200	0.0368	223	50793
300	0.0368	335	113855
500	0.0367	561	317369

Table 9.13: Initial Sample Size, Mean and Variance of Ultimate Sample Size in MC Simulation for Estimating Proportion of HH Cultivating *Hybrid Boro* Using ACS for Barisal Division.

Initial sample size	MC mean of estimated proportion	Ultimate sample size	
		mean	variance
100	0.0728	123	15904
200	0.0725	248	62461
300	0.0725	371	139584
500	0.0723	614	380395

Table 9.14: Initial Sample Size, Mean and Variance of Ultimate Sample Size in MC Simulation for Estimating Proportion of HH Cultivating *Hybrid Boro* Using ACS for Chittagong Division.

From results exhibit in Table 9.14, it is observed that the USS in ACS for Chittagong division also increased slightly. Although this increment is higher than the increment for Barisal division but not differ substantially.

Also, empirically we can say that, in Chittagong division USS would be 24% larger than the initial sample size.

The results of the simulation study related to USS in ACS for Dhaka division are shown in Table 9.15.

Initial sample size	MC mean of estimated proportion	Ultimate sample size	
		mean	variance
100	0.0582	121	115009
200	0.0591	246	61371
300	0.0585	366	135178
500	0.0585	607	371648

Table 9.15: Initial Sample Size, Mean and Variance of Ultimate Sample Size in MC Simulation for Estimating Proportion of HH Cultivating *Hybrid Boro* Using ACS for Dhaka Division.

Table 9.15 represents the USS in ACS for Dhaka division. Form these results we observed 20% increase in USS than initial sample size. Also this increment is almost close to the increment that we found for previous two divisions.

Table 9.16 represents the USS in ACS for Khulna division. Results of this table disclose that the sample size increase more than 50% in ultimate sample than the initial sample. Empirically we can state that the USS is 1.55 times than the initial sample size.

USS in ACS design is also determined for Rajshahi division through simulation. Results that found for Rajshahi division are shown in Table 9.17.

Table 9.17 also exhibit a relationship between initial sample size and USS in ACS. Empirically this relationship can be stated as, USS would be 52% higher than the initial sample size in Rajshahi division.

Finally we investigate the USS in ACS for Sylhet division. The results that we found in simulation study are summarized in Table 9.18.

Initial sample size	MC mean of estimated proportion	Ultimate sample size	
		mean	variance
100	0.1184	154	24807
200	0.1192	311	98667
300	0.1187	460	215202
500	0.1193	769	597837

Table 9.16: Initial Sample Size, Mean and Variance of Ultimate Sample Size in MC Simulation for Estimating Proportion of HH Cultivating *Hybrid Boro* Using ACS for Khulna Division.

Initial sample size	MC mean of estimated proportion	Ultimate sample size	
		mean	variance
100	0.1363	153	23979
200	0.1377	307	95545
300	0.1382	458	211964
500	0.1374	761	583438

Table 9.17: Initial Sample Size, Mean and Variance of Ultimate Sample Size in MC Simulation for Estimating Proportion of HH Cultivating *Hybrid Boro* Using ACS for Rajshahi Division.

Similarly as the other divisions, results of the Sylhet division in Table 9.18 we observed a empirical relation between initial sample size and USS in ACS. Based on the results of Sylhet division we conclude that the USS

Initial sample size	MC mean of estimated proportion	Ultimate sample size	
		mean	variance
100	0.0653	120	14559
200	0.0655	240	58301
300	0.0657	360	130600
500	0.0659	602	363978

Table 9.18: Initial Sample Size, Mean and Variance of Ultimate Sample Size in MC Simulation for Estimating Proportion of HH Cultivating *Hybrid Boro* Using ACS for Syllhet Division.

would be 1.2 times higher than the initial sample size.

9.4.2 The Coverage in ACS Estimators

To see how well adaptive cluster sampling (ACS) can select a good sample than simple random sampling (SRS) we conduct a simulation study using a real life census data as our population. Here we tried to investigate the advantage of ACS over SRS in sense of capturing the cases that contain the characteristic of interest where the proportion of that characteristic in population is too small. We conduct this study for each division separately. Also ACS design is examined at different sample sizes. In our study, we take 500 samples using both SRS and ACS designs and quantified the number of sample that have less than a pre-specified number of case contains the characteristic of interest. This investigation is made separately for each sample size and we considered different pre-specified values of case. The results of our investigation are summarized form Table 9.19 through Table 9.24.

Table 9.19 shows the results that we found for Barisal division. In this table, at each row, values are representing the number of samples that contains the pre-specified number of case having the characteristic of interest. Form this table it is observed that, ACS design has smaller values than SRS

No. Cases Contain HH with HB Boro Usage	Sample Size							
	100		200		300		500	
	SRS	ACS	SRS	ACS	SRS	ACS	SRS	ACS
<2	34	22	7	0	0	0	8	0
<3	52	15	17	0	20	0	4	0
<4	73	44	34	0	12	0	6	0
<5	83	82	57	6	8	0	7	0
<6	95	5	87	79	24	0	7	0
<7	100	27	111	41	55	31	13	0
<8	100	81	143	63	78	38	8	0
<9	100	35	169	113	108	97	17	0
<10	100	74	181	136	144	75	14	0

Table 9.19: Number of samples containing different number of HH cultivating *Hybrid Boro* using SRS and ACS for Barisal Division.

design at every case. That gives an indication that, if ACS is used instead of SRS than the risk would be minimum. Also it has seen that, with the increase in sample size the performance of ACS design increase rapidly. It is important to mention that, in our study the population proportion of HH that use *Hybrid Boro* was 0.036 in Barisal division. Based on the simulation results, it can be stated that, if a sample of size greater than 100 is selected by ACS design than the probability that the sample estimate would be less than 0.02 is zero. In contrast, SRS design has a high probability of getting a more deviated result.

The results that we obtained from simulation study for Chittagong division are shown in Table 9.20.

In Table 9.20 we observed that, samples selected by ACS design captured more HH with characteristic of interest than the samples selected by SRS design. It has seen that, a sample of size greater than 100 captured at least

No. Cases Contain HH with HB Boro Usage	Sample Size							
	100		200		300		500	
	SRS	ACS	SRS	ACS	SRS	ACS	SRS	ACS
<2	2	0	5	0	3	0	15	0
<3	10	1	4	0	8	0	4	0
<4	19	5	3	0	1	0	1	0
<5	34	1	5	0	5	0	11	0
<6	44	32	12	0	6	0	4	0
<7	64	15	8	0	7	0	2	0
<8	80	28	18	0	4	0	3	0
<9	90	49	26	0	13	0	3	0
<10	100	66	38	0	4	0	10	0

Table 9.20: Number of samples containing different number of HH cultivating *Hybrid Boro* using SRS and ACS for Chittagong Division.

more than 10 HH with our interested characteristic when the sampling design is ACS. On the other hand, if SRS is used then the numbers are not negligible even though the sample size is 500. This makes sense that, there is atleast a chance of getting bad sample. But this chance is too small when ACS is used to select the sample. It might be said that, also for Chittagong division the chance of getting deviated estimate is smaller for ACS design than the SRS design.

The similar investigation is also conducted for Dhaka division. Before going to the discussion of the results for Dhaka division it is important to note that the population proportion of HH with *Hybrid Boro* usage in Dhaka division was 0.059. Here our interest is to see which design between SRS and ACS will give a sample that would contain more HH with characteristic of interest. Simulation results for Dhaka division are summarized in Table 9.21.

No. Cases Contain HH with HB Boro Usage	Sample Size							
	100		200		300		500	
	SRS	ACS	SRS	ACS	SRS	ACS	SRS	ACS
<2	2	0	5	0	3	0	15	0
<3	10	1	4	0	8	0	4	0
<4	19	5	3	0	1	0	1	0
<5	34	1	5	0	5	0	11	0
<6	44	32	12	0	6	0	4	0
<7	64	15	8	0	7	0	2	0
<8	80	28	18	0	4	0	3	0
<9	90	49	26	0	13	0	3	0
<10	100	66	38	0	4	0	10	0

Table 9.21: Number of samples containing different number of HH cultivating *Hybrid Boro* using SRS and ACS for Dhaka Division.

Table 9.21 shows that, for sample size is 100, among the samples selected by SRS there is a good chance of getting sample that would give a poor estimate of the population characteristic. This is because, numbers of samples containing fewer cases with interested characteristics is seems to be significant. It is also clear that, the chance of getting sample which will produce a less deviated result is high when ACS is used. From our study results it is observed that, the performance of ACS design is significantly higher than the performance of SRS design when sample is 200 or more.

To see which design is better in capturing good sample in the sense of covering more items with characteristics of interest, we also studied the Khulna division through simulation study. The results found for Khulna division are given in Table 9.22.

No. Cases Contain HH with HB Boro Usage	Sample Size							
	100		200		300		500	
	SRS	ACS	SRS	ACS	SRS	ACS	SRS	ACS
<2	3	0	7	0	2	0	2	0
<3	1	0	8	0	3	0	8	0
<4	3	0	7	0	5	0	4	0
<5	4	0	18	0	2	0	10	0
<6	9	1	0	0	6	0	10	0
<7	11	5	12	0	1	0	15	0
<8	27	10	7	0	4	0	10	0
<9	31	5	17	0	8	0	10	0
<10	44	29	1	0	7	0	13	0

Table 9.22: Number of samples containing different number of HH cultivating *Hybrid Boro* using SRS and ACS for Khulna Division.

From the results displayed in Table 9.22, it can be said that the probability of getting a sample by ACS which will give an estimate that is greater than 0.04 is almost 1 when sample size is 100. But this probability is small when sample is selected by SRS. Since population proportion for Khulna division was 0.119, hence it would be desirable to get sample which will give an estimate that would be near or equal to 0.119. Since we observed that, samples selected by ACS have a high probability than samples selected by SRS to produce a less deviated result, hence we can be said that ACS would be better than SRS design.

Table 9.23 represents the results that we found in simulation study for Rajshahi division.

No. Cases Contain HH with HB Boro Usage	Sample Size							
	100		200		300		500	
	SRS	ACS	SRS	ACS	SRS	ACS	SRS	ACS
<2	6	0	3	0	2	0	1	0
<3	11	0	4	0	5	0	8	0
<4	8	0	1	0	2	0	3	0
<5	5	0	9	0	8	0	4	0
<6	5	0	4	0	8	0	13	0
<7	3	0	13	0	11	0	6	0
<8	1	0	11	0	1	0	4	0
<9	27	6	2	0	16	0	1	0
<10	22	15	24	0	5	0	19	0

Table 9.23: Number of samples containing different number of HH cultivating *Hybrid Boro* using SRS and ACS for Rajshahi Division.

From the results in Table 9.23, it can be observed that, ACS selects less number of samples that contains minimum numbers of HH with *Hybrid Boro* usage at every sample sizes. In contrast, SRS selects a good number of samples that contain few HH with *Hybrid Boro* usage. It is obvious to say that this feature of SRS is not desirable. Results for Rajshahi division also reveals that both ACS and SRS design behave as similar as they behave for previous divisions. So it is clear that, also for Rajshahi division ACS design gives more desirable samples than SRS design.

Finally we investigate those designs for Sylhet division and the results that we found in our investigation are summarized in Table 9.24.

No. Cases Contain HH with HB Boro Usage	Sample Size							
	100		200		300		500	
	SRS	ACS	SRS	ACS	SRS	ACS	SRS	ACS
<2	10	1	4	0	4	0	11	0
<3	13	4	1	0	3	0	5	0
<4	21	18	2	0	3	0	4	0
<5	40	31	13	0	10	0	2	0
<6	50	23	14	0	4	0	1	0
<7	66	22	17	0	2	0	14	0
<8	91	21	23	0	4	0	3	0
<9	95	20	42	0	14	0	5	0
<10	98	5	52	0	19	0	20	0

Table 9.24: Number of samples containing different number of HH cultivating *Hybrid Boro* using SRS and ACS for Sylhet Division.

Table 9.24 represents the number of samples that contain fewer cases than a pre-specified value. As similarly as previous divisions, ACS design is more likely than SRS to capture a sample that contain more cases with characteristic of interest. This is an evidence for better performance of ACS in capturing cases with rare characteristics.

Chapter 10

Discussion and Conclusion

10.1 Introduction

There is a factorial structure to be taken into consideration when discussing the results of this study. The factors are

1. The relative performances of the estimators of the population proportion obtained using the simple random sampling (SRS), adaptive cluster sampling (ACS) and two stage adaptive cluster sampling (TSACS) methods are discussed in this thesis. We considered different sample sizes to conduct the simulation study, in that way the effect of the sample sizes may also be observed.

2. For each of the methods, the way the estimators of the population proportion perform in terms of MC se and absolute percentage relative bias (APRB) were studied.

3. For each of the methods, the way the estimators of the population proportion perform when there are different levels of true proportion in the population was studied. The variation in the true population proportion in different divisions are taken under consideration in this purpose.

10.2 Comparison with Respect to Bias

The comparison of bias of ACS estimators of the proportion of HH cultivating *Hybrid Boro* in different divisions show that the APRB involved in the SRS estimator of the population proportion is, in allmost all cases, higher than that of the ACS method.

However, for TSACS, the content of the bias is also not very high, direct comparision of TSCS with ACS is not made because the sample size considered for the TSACS is taken to be different. This was mainly due to computation time required of TSACS. It can be concluded that ACS performs better than SRS in terms of bias. The content of bias seems to remain similar for all sample sizes we considered.

10.3 Comparison with Respect to Standard Error

Features observed by the MC se of the ACS are remarkable in the sense that the MC se of the SRS method are relatively smaller than that of ACS method with the same sample size. For estimation of the proportion of HH cultivating *Hybrid Boro*, the ratio of the MC se is not very high for any of the divisions, is about $3-4$ for most of the cases. For higher sample sizes the ratio seems to get lower for all the division. Figures 9.2 through 9.12 shows a comparative display of the MC se of the two methods.

This result may raise issues regarding the relative advantage of the ACS method, but it is well known that the effect of clustering increases the variance of the estimators even for the conventional designs. However the main reason for using ACS is not to increase precision, it is rather used to facilitate capturing rare characteristics. That is why even a three fold higher MC se is not deemed a real concern in case of estimation of the proportion of HH cultivating *Hybrid Boro* in different divisions. The main advantage intended to be gained by the use of the ACS method is more flexibility in capturing more information which is described in Section 9.4.2. The results for the different divisions can demonstrate the effect of true proportion of HH cultivating *Hybrid Boro* on the ratio of the MC se of the two methods. It can be observed that the MC se for division with smaller true proportion of HH cultivating *Hybrid Boro* has smaller MC se.

10.4 Comparison with Respect to Ultimate Sample Size

Another issue often argued regarding the use of the ACS method could be the ultimate sample size realized by the ACS method. If the clustering of the population characteristic of interest are too concentrated, there would be a lot of adaptation and as a result a huge number of units may be selected in the sample. The potential risk of this scenario could lead to a more expensive survey by accumulating a large sample size. For the estimation of proportion of HH cultivating *Hybrid Boro*, an extensive simulation study described in section 9.4.1 has revealed that the ultimate sample sizes are not too large in respect of the initial sample sizes for each of the divisions and each of the sample sizes. The ultimate sample sizes for Barisal and Sylhet division are found to be increased by lesser quantity than that of other divisions. This reveals the fact that rarer the characteristics of interest, smaller the increase in ultimate sample sizes. Yet again, the minimal increase in the sample size to constitute an ultimate sample size may be traded off with the minimization of risk of ending up with very small number of units containing the targeted characteristics (see Section 9.4.2).

10.5 Comparison with Respect to Coverage

The most impressive feature of the possible application of ACS method was seen to be its strength of capturing more information. In estimation of proportion of HH cultivating *Hybrid Boro*, it has been revealed from the simulation that the ACS method is way far better than the SRS method in terms of chances of getting a bad sample containing very small number of targeted characteristics. The discussion in Section 9.4.2 has clearly established the feature. In SRS method, for Barisal division, about 52 cases out of 500 simulations of a sample of size 100 were found to capture less than 3 HHs cultivating *Hybrid Boro* in the sample, whereas the number is only 15 in case of ACS method. This indicates that in case of estimation of proportion of HH cultivating *Hybrid Boro*, an SRS of size 100 may end up with 2, 1, or 0 cases satisfying the characteristics in almost 10% cases. Thus applicability of the method in practical situations may be, to some extent, risky in terms of probable wastage of resource. However for larger sample sizes or for lower true population proportion, such probabilities are smaller but the comparative edge of the ACS method remains superior in this regard.

10.6 Overall Comments

From the above summary discussion and the discussion in chapter 9, the following concluding remarks can be made:

1. The bias of the estimator of the proportion of HH cultivating *Hybrid Boro* is mostly higher for the SRS method for each of the division and each of the sample sizes considered.

2. For estimation of proportion of HH cultivating *Hybrid Boro* in each of the divisions of Bangladesh, the SRS method have a smaller MC se than the ACS method have. But this relatively higher standard error of ACS method can be counted as the general fact that the cluster sampling has higher standard error of the estimator than the SRS has. The initial sample size has a negative influence on the standard error for both the methods.

3. For each of the divisions, the ultimate sample sizes for ACS method were found to be not too much for estimation of proportion of HH cultivating *Hybrid Boro*.

4. The risk of getting samples with very small number of HH cultivating *Hybrid Boro* in the sample is smaller for ACS method whereas the SRS method seems to suffer badly in this regard for each division and each of the sample sizes considered in this thesis. For SRS method such risk is higher for divisions with smaller true population proportion. These findings can be triangulated to the issue that the SRS method produce more APRB than ACS to conclude that the bias is induced from capturing bad samples with no or very low number of HH cultivating *Hybrid Boro* in the sample by the SRS method.

10.7 Limitation of the Study

The simulations conducted under the study were time consuming due to multiple looping in the codes and as well as due to the huge size of data associated. Huge amount of computer time was demanded in each of the simulations, which also were disrupted due to unavailability of uninterrupted electricity.

10.8 Future Study

The simulations for the two stage adaptive cluster sampling (TSACS) may be conducted for similar data set provided the availability of high speed computers with uninterrupted electricity for handling such huge computations.

Bibliography

Basu, D. (1969) Role of the sufficiency and likelihood principles in sample survey theory. *Sankhya A*, **31**,441-454.

Blanke, D. (2006) Adaptive sampling schemes for density estimation. *Journal of Statistical Planning and Inference* , **136**,2898-2917.

Blackwell, D (1947) Conditional expectation and unbiased sequential estimation. *Journal of Mathematical Statistics*, **18**,105-110.

Bohn, L.L. and Wolfe, D.A. (1992) Nonparametric two-sample procedures for ranked set samples data. *Journal of the American Statistical Association* , **87**,552-561.

Brown, J. A., Manly, B. J. F. (1998) Restricted adaptive cluster sampling. *Environmental and Ecological Statistics*, **5**,49-63.

Brown, J. A. (1994) The application of adaptive cluster sampling to ecological studies. Statistics in Ecology and Environmental Monitoring, *Otago Conference Series No.2. Dunedin, New Zealand: University of Otago Press*,86-97.

Cassel, C.M., Sarndal, C.E. and Wretman, J.B. (1979) Prediction theory for finite populations when model based and design based principles are combined. *Scandinavian Journal of Statistics* , **6**,97-106.

Chaudhuri, A., Bose, M., and Ghosh, J.K (2004) An application of adaptive sampling to estimate highly localized population segments. *Australian and New Zealand Journal of Statistics* , **121**,175-189.

Cormack, R. M. (1988)Statistical challenges in the environmental sciences: A personal view. *Journal of the Royal Statistical Society* , **A 151**,201-210.

Danaher, P. J., and King, M. (1994) Estimating rare household characteristics using adaptive sampling. *The New Zealand Statistician* , **29**,14-23.

Dell, T.R. and Clutter, J.1. (1972). Ranked set sampling theory with order statistics background. *Biometrics* , **28**,545-553.

Francis, R. 1. C. C. (1984) An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research* , **18**,59-71.

G. A. F. (1986) A review of estimating animal abundance. *Biometrics* , **42**,267-292.

Gasaway, W. C., DuBois. S. D., Reed, D. J., and Harbo, S. J. (1986) Estimating Moose Population Parameters/rom Aerial Surveys. *Biological Papers of the University of Alaska (Institute of Arctic Biology) Number 22. Fairbanks: University of Alaska* .

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of Royal Statistical Society* , **B, 17**,269-278.

Godambe, V.P. (1982). Estimation in survey sampling: Robustness and optimality. *Journal of the American Statistical Association*, **77**,393-40.

Halls, L.K. and Dell, T.R. (1966). Trial of ranked set sampling for forage yields. *Forest Science*, **12**,22-26.

Hao Yu, Yan Jiao, Zhenming Su, Kevin Reid (2012) Performance comparison of traditional sampling designs and adaptive sampling designs for fishery-independent surveys: A simulation study *Fisheries Research*, **Volume 113, Issue 1, January 2012**,Pages 173-181.

Hu, J. and Su, Z. (2007) Adaptive resampling algorithms for estimating bootstrap distributions. *Journal of Statistical Planning and Inference* .

Khan, A. (2008) A combination of adaptive and line intercept sampling applicable in agricultural and environmental studies. *Journal of Statistics* , **15**,44-53.

Khan A.and Muttlak H. A. (2002) Adjusted two-stage adaptive cluster sampling. *Applied Mathematics and Computation* , **126**,83-95.

Khan A.and Muttlak H. A. (2000) Adjusted two-stage adaptive cluster sampling with ranked set sampling *Bangladesh Journal of Scientific Research*, **18**,311-323.

Khan, A. and Hossain, S. S. (2002) Repeated Application of Ranked Set

Sampling. *Journal of Statistical Studies* , **22**,41-45.

Khan, A. (1999) et. al., Adaptive cluster sampling for characteristics other than total number of elements. *Parisankhyan Samikkha* , **6**,19-32.

Kremers, W. K. (1987) Adaptive sampling to account for unknown variability among strata.

Munholland, P. L., and Borkowski, 1. 1 (1993a) Latin square sampling + I designs. Technical Report No. 1-11-93, Department of Mathematical Sciences, Montana State University, Bozeman.

Munholland, P. L., and Borkowski, 1. 1 (1993b) Adaptive Latin square sampling + I designs. Technical Report No. 3-23-93, Department of Mathematical Sciences, Montana State University, Bozeman.

Ramsey, F. L., and Sjamsoe'oed, R (1995) Habitat association in conjuction with adaptive cluster samples. *Environmental and Ecological Statistics*

Roesch, F. A., Jr. (1993) Adaptive cluster sampling for forest inventories. *Forest Science* , **39**,655-669.

Salehi, M. M., Seber, G. A. F. (2002) Unbiased estimators for restricted adaptive cluster sampling. *Aust. N. Z.J. Stat* , **44**,63-74.

Salehi, M. M., Seber, G. A. F. (1997) Two stage adaptive cluster sampling. *Biometrics* , **53**,959-970.

Salehi, M. M., Seber, G. A. F. (2004) A general inverse sampling scheme and its application to adaptive cluster sampling *Australian and New Zealand Journal of Statistics* , **46**,483-494.

Salehi, M.M. and Seber, G.A.F. (1997) Two stage adaptive cluster sampling *Biometrics*, **53**,959-970.

Seber, G. A. F. (1982) The Estimation of Animal Abundance, 2nd edition. London: Griffin.

Seber, G. A. F. (1992) A review of estimating animal abundance ll. *International Statistical Review* , **60**,129-166.

Thompson, S. K. (1998) Adaptive sampling. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 784-786.

Thompson, S. K. (1990) Adaptive cluster sampling. *Journal of the American Statistical Association* , **85**,1050-1059.

Thompson, S. K. (1991a) Adaptive cluster sampling: Designs with primary and secondary units *Biometrics* , **47**,1103-1115.

Thompson, S. K. (1991b) Stratified adaptive cluster sampling *Biometrika* , **78**,389-397.

Thompson, S. K. (1992) *Sampling* , New York: Wiley.

Thompson, S. K. (1992) *Sampling* , John Wiley and Sons.

Thompson, S. K. (1993) Multivariate aspects of adaptive cluster sampling. *In* G. P. Patil and C. R. Rao (Eds.), *Multivariate Environmental Statistics, pp.561-572 Australian and New Zealand Journal of Statistics* . New York: North Holland/Elsevier Science Publishers.

Thompson, S. K. (1994) Factors influencing the efficiency of adaptive cluster sampling. Technical Report 94-0301, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park.

Thompson, S. K. (1996) Adaptive cluster sampling based on order statistics *Environmetrics*, **7**,123-133.

Thompson, S.K. and Seber, G.A.F. (1996) *Adaptive Sampling* , New York: Wiley.

Thompson, S. K. and Ramsey, F. L. (1983) *Adaptive Sampling of Animal Populations* ,Tech- nical Report 82, Dept. of Statistics, Corvallis: Oregon State University.

Thompson, S. K. and Ramsey, F. L. (1987) Delectability functions in observing spatial point processes. *Biometrics* , **43**,355-362.

Thompson, S. K., Ramsey, E L., and Seber, G. A. E (1992) An adaptive procedure for sampling animal populations. *Biometrics* , **48**,1195-1199.

Thompson, S. K., and Seber, G. A. E (1994) Detectability in conventional and adaptive sampling. *Biometrics* , **50**,712-724.

Thompson, S.K. (1996). Adaptive cluster sampling based on order statistics. *Environmetrics*, **7**,123-133.

Appendix A

The questionnaire for the Agricultural Census 2008

Schedule-1 **Government of the Peoples' Republic of Bangladesh** **(Confidential)**

Agricultural Census-2008 **Bangladesh Bureau of Statistics** Form:Agc/08-01

1 Household serial No. [ ] [ ] [ ] 2 Name of Head of House ------------------------------ 3 **Male** 1 **Female** 2

Name of Father/Sp ------------------------------

4 Member of Household	Code	Male	Female
A) Less than 10 years ---	01		
B) 10 to less than 15 years ----------------------------------	02		
C) 15 years or more --	03		
D) Total no. of members --------------------------	04		
5 Population engaged in Agriculture work(Last 1 Year)	Code	Male	Female
A) Less than 10 years ---	05		
B) 10 to less than 15 years --------------------------------------	06		
C) 15 years or more ---	07		
D) Total No. of members engaged in Agri. work ----	08		
6 Agri. Labour Household (Tick) --------------------------	09	Yes - 1	No - 2
Land of Household	Code	Acre	Decimals
7 Total own land(Of all members) -------------------------	10		
8 Land given to others --------------------------------------	11		
9 Land taken from others -----------------------------------	12		
10 Total operated Area (7-8+9) -----------------------------	13		
11 Homestead land ---	14		
12 Area under bamboo clump -------------------------------	15		
13 Pond Area A) Single ownership -------	16		
B) Joint ownership ---------	17		
14 Net area land under temporary crops ------------------	18		
15 Land under parmanent/reserved land ------------------	19		
A) All trees --	20		
B) Flower cultivation ------------------------------------	21		
C) Tree & flower nursery --------------------------------	22		
16 Current fallow land --	23		
17 Net cultivated land (14+15+16) --------------------------	24		
18 Permanent fallow ---	25		
19 Land under pisciculture(Excluding pond) --------------	26		
20 Net Irrigated area --	27		
A) Driven by electricity -----------------------------------	28		
B) Driven by diesel ---------------------------------------	29		
C) Non-mechanical (Manual) --------------------------------	30		

24. Land under temporary crop

Name of crop	Code	Acre	Decimals
Local Aus	01		
HYV Aus	02		
Local Broadcast Aman	03		
Local Trnsplated Aman	04		
HYV Aman	05		
Local Boro	06		
Hybrid Boro	07		
HYV Boro	08		
Wheat	09		
Lacal Maize	10		
Hybrid Maize	11		
Gram	19		
Lentil	21		
Mung	23		
Mash Kalai	24		
Kheseri	25		
Mustard	29		
Ground Nut	33		
Jute	41		
Sugarcane	45		
Potato	48		
Sweet Potato	49		
Brinjal	52		
Sweet Pumpkin	53		
Water Pumpkin	58		
Tomato	59		
Bean	61		
Patal	62		
Cucumber	66		
Banana	79		
Water Melon	81		
Chilli	85		
Onion	86		
Garlic	87		
Turmeric	88		
Ginger	90		
Seed bed	96		
No. of crops	99		

21. No. of Livestock, Poultry & Losses

Kind	Code	Local	Hybrid	No.of death	Cause of death
Cow	31				
Baffalo	32				
Goat	33				
Sheep	34				
Cock/hen	35				
Duck	36				
Pigeon	37				

22 Loan Tak (Amount)

Purpose	Code	Institutional	Non-Institutional
Crop	38		
isicultur	39		
Livestock rearing	40		
Tree Plantation	41		
Nursery	42		
Others	43		
Total	44		

23 Ownership of Agri. Implements

Name of Implements	Code	Electricity	Diesel	Non mechanic
Deep tubewell	45			
Shallow tubewell	46			
Power pump	47			
Tractor	48			

Name of Implements	Code	Electricity	Diesel	Non mechan
Power tiller	49			
Rice thresher	50			
Seeder	51			
Other Agri. Implements	52			

*Cause of Death Code: Flood -1, Cyclone-2, Epidemic-3, Disease-4, Snake Bite-5 & Other-(