CHAPTER I

INTRODUCTION

1.1 Background

Information on population can be gathered through censuses, involving complete enumeration of all people living in the country, systematic registration of certain vital events or sample surveys of a cross-section of the population or of special groups.

Population censuses have a long history dating back to at least 3000 BC in ancient Babylonia, China and Egypt. However, the sample surveys are a much more recent development. The enumeration of a population by sampling methods was proposed by Laplace in 1783 but only came into widespread use in the last 60 years.

The development of the techniques of sample surveys was initiated in the 1930s. A paper on 'the representative method' by Neyman in 1934 represented a pioneering work in research and development in survey sampling (Krewski, 1981). Since then, considerable advances have been made in all aspects of survey methodology, particularly in sampling methods. In 1940s, a great deal of efforts was devoted to the development of efficient sampling techniques for use in nationwide surveys. Subsequently, sampling has become widely accepted and many sampling methods had been developed to enable efficient and economic samples to be drawn in various practical settings. The most commonly used probability sampling methods are simple random sampling, systematic sampling, stratified random sampling, probability proportional to size sampling, cluster sampling and multistage sampling. Multistage sampling is particularly popular in nationwide surveys as it generally reduces travelling cost. In multistage sampling, the primary

sampling units are frequently selected with probability proportional to some measure of size.

Sample surveys are now widely accepted as a means of providing statistical data on an extensive range of subjects for both research and administrative purposes. Numerous surveys have been conducted to develop and refine research hypotheses for a better understanding of a population in many disciplines such as demography, political science, economics, education, sociology and public health. Governments make considerable use of surveys to gather information on the conditions of their populations in terms of education, nutrition, health, employment and unemployment, income and expenditure, housing condition, travel patterns and other many other aspects.

Besides household surveys, various Government's agencies also conduct surveys of organizations such as hospitals, factories, industries, schools, retail outlets, companies and farms. Opinion polls gained considerable attentions especially in developed countries to keep track of the popularity of political leaders and their parties and to measure public reactions on a variety of current issues. Market researchers carry out surveys to identify markets for products, to understand how the products are used and to determine consumer reactions. Market research is an important tool in decision making in today's competitive business environment and sample surveys are the overwhelming choice of researchers for collecting primary data.

1.2 Significance of the Study

Survey research has been one of the important tools employed to provide a better understanding of a wide variety of characteristics of various target populations. The sample survey is the major tool and preferable data collection method over complete enumeration when cost and speed of information gathering are the major concerns.

Collecting data from only a part of the population to provide the estimates that are sufficiently precise is clearly less costly than a complete enumeration. A sample inquiry can also be conducted and processed more speedily, leading to more timely reporting. Furthermore, by concentrating the resources on only a part of the population, the quality of the data collected may be superior to that of a complete enumeration. As a result, a sample survey may in fact produce more accurate results than a census. For these reasons, unless the population is small, sampling is almost always used.

In studying any groups - people, plants, stores and institutions - sampling offers a practical compromise between certainty (actual knowledge) and expediency (use of a single observation or a guess). In general, certainty is too expensive and a single observation or guess is too unreliable. Generalizations regarding the whole group can come from the study of a small part that is properly selected.

Since sample survey deals with methods of selecting and observing a part of the population in order to make inferences regarding the whole population, it is important to use appropriate sampling methodologies to produce quality and accurate sample. Rigorous probability sampling is essential in providing a scientific

basis for estimating sampling error and generalization to the survey population (WFS, 1975).

With a probability sample, each element has a known non-zero chance of being included in the sample. An essential requirement for any form of probability sample is the existence of a sampling frame from which the elements can be selected. In a simple case, when a list of all the population is available, the frame may be the list. When there is no list, the frame is some equivalent procedure for identifying the population elements.

The choice of a sampling techniques hinges on the availability of sampling frame which differs from country to country as well as from discipline to discipline. Consideration on cost, time, the level of precision required and fieldwork operation also influences the techniques employed. Therefore, a variety of probability sampling techniques have been developed to provide efficient practical sampling designs. In practice, most large-scale surveys use complex designs with a combination of these techniques.

The simple random sampling technique is the basis for sample selection. However, as most of the surveys involve face-to-face interviews with respondents, this method of selecting a sample is only suitable for small population which are not too disperse. In practical sample surveys, especially in market research of general population, the sponsors and researchers always have multiple objectives with limited resources to cover a big population. Therefore, complex sampling designs rather than simple random sampling are often used in large-scale surveys to reduce travelling cost and to facilitate organization of field operations.

There are many methods available for variance estimation of complex designs. Besides the classical linearization and replication methods that depend heavily on theoretical work and calculation, various re-sampling procedures for variance estimation in sample surveys have also been developed.

Although there have been important advances in the development of survey theories and techniques, the application of some of these advances is not particularly widespread. The main concern of survey statisticians is to provide estimates that are reliable. However, they are often questioned on this point by the main users who may not be statisticians. Besides, the used of these information in repeated sampling seem to be merely a hypothetical situation since most the surveys may not be repeated for several years. Often, little attention is given to strict probability sampling. Furthermore, little or no attempt is made to compute standard errors consistent with the sampling designs. Rather, statements of precision will be made based on a simple random sampling even though a complex design was employed. Consequently, the actual mean square errors are significantly underestimated in many instances (Namboodiri, 1978).

While the techniques for variance estimation exist for complex survey data, they are rather cumbersome to implement. Also, the necessary processing resources may not be available. The software currently available is severely limited (Rennolls, 1991). To date, very few statistical software packages have specific facilities, which take into account the unique aspects of sampling designs.

Sampling statisticians asserted that it is important to compute and present the sampling errors in all cases irrespective of the designs together with the survey results (Kish, 1965; UN, 1989 pp145). Information on sampling errors permits the readers to construct confidence intervals and to make inferences according to their

needs. The design effects and rate of homogeneity should be presented along with survey results to facilitate the computation and presentation of sampling errors (Namboodiri, 1978).

1.3 Objectives of the Study

The main focus of this study is on estimation of the efficiency of the stratified two-stage cluster sampling used in the Media Index Survey (MIS) conducted in Peninsular Malaysia. The main objectives are:-

i. To estimate the sampling errors for selected statistics according to geographical domains and subclasses of the population.

ii. To estimate the design effects by geographical domains and subclasses.

iii. To estimate the intra-class correlation (rho) by geographical domains and subclasses.

iv. To evaluate the suitability of the stratified two-stage cluster sampling in Media Index Survey, with respect to the estimates of selected study variables.