LECTURE NOTES

ON

SAMPLING THEORY AND SURVEY METHODOLOGY I

[STA 331]

TOPIC ONE

INTRODUCTION TO SAMPLING THEORY

1.1. Introduction

To begin this course on sampling theory, we shall begin by highlighting some concepts with their definitions which are paramount in understanding the ideas behind the course.

(a) Sampling: Sampling is the scientific method of choosing part of a whole as a representative in order to make valid judgment about the whole.

(b) Sample survey: This is the collection and examination of data from a sample in order to make inference about the whole.

(c) Census: This is the total enumeration of the whole. It is the opposite of a sample survey.

(d) Population: This is the entire frame of reference under investigation. For example the total number of people in a country and the total number of houses in a community. A population can be finite or infinite depending on if the number of units making it up is finite or infinite respectively.

(e) Sample: This is a subset of the population.

(f) Sampling Unit: This is an object chosen in the sample survey upon which an investigation is performed. The sampling unit carries some quantity or possesses some attribute which is of interest to the investigator.

(g) Sampling scheme: This is a method of selecting a sample from a population.

(h) Sampling design: This is the collection of all possible samples together with their probabilities of selection.

(i) Estimator: This is any real-valued function defined on the sample values for estimating an unknown parameter. The value of an estimator for a given sample is called an estimate.

(j) Population parameter: This is a function defined on the population values.

(k) Precision: This is a measure of how close an estimate is to its average value over all possible samples.

(l) **Sampling error:** This is the error that occurs as a result of using a sample to make inference about the population

(m) **Enumeration Area (EA):** This is a small area, generally used as a sampling unit with well-defined and identifiable boundaries carved out of larger area of land.

(n) **Efficiency:** This is a measure of how well an estimator minimizes the error in estimating a population parameter better than that of another estimator.

1.2. Benefits of a Sample Survey

-It saves time and cost.

-When units are susceptible to being damaged, it is more economical to use a sample survey.

-Analysis of data that comes from a sample survey is quicker to perform than that of a census.

-Sample survey tends to have more coverage and less behavioral errors than a census.

1.3. Kinds of Sampling

There exist two main kinds of sampling: **non-probability (non-random) sampling** and **probability (random) sampling**. In non-probability sampling, each sampling unit in the population is not selected with any known probability and hence statistical inference cannot be made objectively about the population in this kind of sampling. Examples of non-probability sampling include:

-**Judgment or purposive sampling**: when the investigators pick sampling units they judge to be representative of the population.

-**Haphazard sampling:** when sampling units are chosen as they come along or make themselves available.

-**Quota sampling**: when specified number of sampling units is assigned to specific groups based on age, gender, location etc.

-**Capture-recapture:** when the need is to estimate the size of closed and mobile units of a population which remains constant in size within a short period of time.

In probability random sampling, every unit in the population has a known non-zero probability of being selected in the sample. This kind of sampling is our focus in this course since it has great implication in making statistical inference about the population.

1.4. Steps in Planning a Sample Survey

-define the objective of the survey

-determine the population to study

-determine the frame or map of the population

-decide on the method of collection of data

-organization and supervision of field work

-carry out a pre-test and a pilot survey

-decide on the precision level to be achieved for all parameters to be estimated

-chose a sampling design

-analyze the survey data

-write a report on the result of the survey.

1.5. Work to Do

i. Thoroughly discuss the rationale behind sample survey
ii. Suppose you are an investigator and you wish to carry out a sample survey study on the compliance of students to the non-pharmaceutical protocols for the control of COVID-19 on campus, discuss how you might possibly design, organize and present findings from the study.
iii. What are the possible shortcomings of carrying out a sample survey instead of total enumeration of a target population.

iv. Distinguish between a probability sampling and non-probability sampling and write down with explanations the types of non-probability sampling you know.

References

Kish, L. (1965): Survey Sampling. *John Wiley, New York*.

TOPIC TWO

EQUAL PROBABILITY SAMPLING

1.1.Introduction

In Topic One, we saw that there exist two major kinds of sampling: non-probability and probability sampling. While in non-probability sampling all items in the population have no known probability of being selected, in probability sampling such non-zero probability exist. Probability sampling itself is further divided into two types: equal probability sampling and unequal probability sampling. In equal probability sampling, all items in the population have the same non-zero probability of being selected while in unequal probability sampling, the probability of a given item being selected differs across the population. In this Course, we shall discuss two major types of equal probability sampling namely: Simple Random Sampling (SRS) and Systematic Sampling (SS). Sampling using SRS can be done with or without replacement.

1.2.Simple Random Sampling

The procedure for selecting a sample using the SRS routine is as follows:

-number the population serially from 1 to N where N is the size of the population

-Select random numbers which lies between 1 and N (1 and N inclusive) and draw a given number out of the population corresponding to the random number. This process is repeated till the required sample is obtained. This is called the *lottery method*.

If after each draw the selected number is removed from the population before the next selection is made, we refer to the selection method as Simple Random Sampling without Replacement (SRSWOR), generally referred to as SRS. If the selected number is replaced in the population before the next draw is made we would have Simple Random Sampling with Replacement (SRSWR).

1.2.1. Probability of selection

Suppose n is the size of the sample to be drawn from the population of size N, the number of ways of selecting n distinct units out of N such units without replacement is $\binom{N}{n}$. Also, any one of these $\binom{N}{n}$ possible samples of size n has the same probability $1/\binom{N}{n}$ of being selected. Suppose sampling is done Without Replacement (WOR), if the i^{th} unit, U_i, is selected at the first draw, we have $(n-1)$ units left to be selected from the remaining $(N-1)$ population units and this can be done in $\binom{N-1}{n-1}$ possible ways, which is the number of samples that contain any specific unit U_i. Thus the probability that any specific unit is included in the sample is $\binom{N-1}{n-1}/\binom{N}{n} = n/N$. It follows that the following holds:

- The probability of selecting , U_i at the second draw is
$$\frac{N-1}{N} \times \frac{1}{N-1} = \frac{1}{N}$$
- The probability of selecting , U_i at the third draw is
$$\frac{N-2}{N} \times \frac{1}{N-2} = \frac{1}{N}$$
- The probability of selecting , U_i at the r^{th} draw is
$$\frac{N-r+1}{N} \times \frac{1}{N-r+1} = \frac{1}{N}.$$

In SRSWR there are N^n possible samples of size$n$, each of which can be chosen with probability $1/N^n$ considering the order of appearance of the sample. Since successive draws are independent of one another, the probability of selecting U_i at any of the draws is $1/N$.