1. Introduction

Welcome to the series of E-learning modules on estimation of population mean, total and proportion. In this module, we are going to cover the basic concept of estimation, estimates of population mean, total and proportion under Simple Random Sampling with and without replacement.

By the end of this session, you will be able to:

- Explain the estimation of parameters
- Explain the estimation of population mean and total under Simple Random Sampling with Replacement (SRSWR)
- Estimation of population mean and total under Simple Random Sampling without replacement (SRSWOR)
- Discuss the sampling for attributes
- Explain the estimation of population proportion

Estimation is the process of determining a likely value for a variable in the survey population, which is based on the information collected from the sample.

Researchers are usually interested in looking at estimates of many statistics—totals, averages and proportions being the most frequent—for different variables.

For example, a sample survey could be used to produce any of the following statistics:

- Estimates for the proportion of smokers among all people aged fifteen to twenty four in the population
- The average earnings of men and women with a university degree
- The total number of cars possessed by the survey population

Underpinning the estimation process is the sampling weight of a unit, which indicates the number of units in the population (including the sampling weight) that are represented by the sampled unit. The sampling weight is the inverse of the unit's probability of selection.

Let us take an example where the city of Winnipeg has decided to award free one-year bus passes to the bus travellers as a way of promoting its services. A simple random sample of ten people is selected from the thirty passengers on a city bus.

As simple random sampling gives equal probability of selection to every member of the population (in this case, all passengers on the bus), each passenger has a chance of being selected out of three. This translates into a sampling weight of three for every selected unit. This means that each person in the sample represents three persons in the population—himself or herself, plus two other persons.

If each unit in a population has equal chance of being selected in the sample then, such a sample is known as Simple Random Sample and the technique is known as Simple Random Sampling (SRS). Under SRS we have two techniques:

a) SRS with replacement scheme

b) SRS without replacement scheme

2. Basic Idea of Sampling and Estimation

Let us discuss the Basic Idea of Sampling and Estimation.

One interesting and important fact to note is that in most useful sampling schemes, variability from sample to sample can be estimated using the single sample selected.

Using the collected sample, we can construct estimates for the parameter of the population of our interest. Usually, there are many ways to construct estimates. Thus, we need some guidelines to determine which estimates are desirable.

Some desirable properties for estimators are:

- Unbiased or nearly unbiased
- Have a low mean square error (MSE) or a low variance when the estimator is unbiased. [MSE measures how far the estimate is from the parameter of interest whereas, variance measures how far the estimate is from the mean of that estimate. Thus, when an estimator is unbiased, its MSE is the same as its variance.]
- Robust - so your answer does not fluctuate too much with respect to extreme values

Estimation using the simple random sampling method has been studied extensively. There are simple formulae to estimate the parameters when simple random sampling is used because it is a self-weighing design.

Estimation of Average household income in a country is important during tough economic times or assessment of taxes and property values.

We might want to:

- ➢ Estimate the total pyrite content of the rocks
- ➢ Estimate the total market size for electrical cars

In these situations, we have to go for the estimation of either population mean or total.

We present here the most common estimator for a population average (mean) total and proportion under simple random sampling.

3. SRSWR & SRSWOR

Let us discuss the Simple Random Sampling With Replacement.
Theorem 1: Under SRSWR sample, mean is an unbiased estimator of population mean i.e. to prove that the expected value of sample mean is equal to the population mean.

Proof: Let y one, y two, y three, up to y n be Simple Random sample with replacement sample of size n drawn from a population containing N units then, we know that population mean

Y bar is equal to summation Y i, i runs from one to N divided by N and the sample mean y bar is equal to summation yi, i runs from one to n divided by n.
Consider Expected value of sample mean y bar is equal to Expected value of (summation yi, i runs from one to n divided by n) which is equal to summation Expected value of yi, i runs from one to n divided by n. Let this equation be numbered as one.

Expected value of yi is equal to summation over i, yi multiplied by pi
Consider Expected value of y_i, since each y_i takes values Y_i with probability one by N we have Expected value of yi is equal to Summation Yi, i runs from one to N divided by N which is equal to population mean Y bar.
By substituting the above in equation one, we get
Expected value of sample mean is equal to summation population mean Y bar, i runs from one to n divided by n, which is equal to n multiplied by population mean Y bar divided by n, which is equal to population mean Y bar.
Hence, expected value of sample mean is equal to population mean Y bar. That is sample mean is unbiased estimator of population mean.

Theorem 2: Under Simple Random Sampling With Replacement an estimator of population total is population size N multiplied by sample mean y bar. Symbolically represented as: Y cap is equal to N multiplied by sample mean y bar.

Proof: Let y one, y two, y three up to yn be Simple Random Sample With Replacement sample of size n drawn from a population containing N units.
A population total Y is equal to summation Yi, where i runs from one to N which is equal to population size N multiplied by population mean Y bar.
Which implies Y cap is equal to N multiplied by Y bar cap, which is equal to N into sample mean y bar because Expected value of the sample mean y bar is equal to the population mean Y bar. Which implies Y bar cap is equal to sample mean.

Let us explain the Simple Random Sampling Without Replacement (SRSWOR).
Theorem 3: Under Simple Random Sampling Without Replacement sample mean is an unbiased estimator of the population mean i.e. to prove that Expected value of sample mean is equal to the population mean.

Proof: Let y one, y two, y three up to yn be Simple Random Sample Without Replacement sample of size n drawn from a population containing N units then we know that population

mean
Y bar is equal to summation Y i, i runs from one to N divided by N and the sample mean y bar is equal to summation yi, i runs from one to n divided by n.
Consider Expected value of sample mean y bar is equal to
Expected value of (summation yi, i runs from one to n divided by n)
Which is equal to summation Expected value of yi, i runs from one to n divided by n. let the equation be equation numbered as one.

Expected value of yi is equal to summation over i, yi multiplied by pi
Since the probability of selection of a unit under SRSWOR is also one by N, so the proof of SRSWR holds good for SRSWOR also.
Hence, Expected value of sample mean is equal to population mean Y bar. That is sample mean is unbiased estimator of population mean.

Theorem 4: An unbiased estimator of population total under SRSWOR is population size multiplied by the sample mean y bar. Symbolically represented as: Estimate of Y is equal to N into sample mean y bar.

Proof: Let y one, y two, y three up to yn be Simple Random Sample Without Replacement sample of size n drawn from a population containing N units.
A population total Y is equal to summation Yi, where i runs from one to N which is equal to population size N multiplied by population mean Y bar.
Which implies Y cap is equal to N multiplied by Y bar cap, which is equal to N into sample mean y bar because Expected value of the sample mean y bar is equal to the population mean Y bar. Which implies Y bar cap is equal to sample mean.

4. Simple Random Sampling of attributes

Let us discuss the Simple Random Sampling of attributes.
A qualitative characteristic, which cannot be measured quantitatively, is known as an attribute. For example: Honesty, beauty, intelligence etc. Quite often, we come across the situations where, it may not be possible to measure the characteristic under study but it may be possible to classify the whole population into various classes with respect to the attributes under study.

We consider the cases where the population is divided into two classes say C and C dash with respect to an attribute. Such a classification is termed as dichotomous classification. Hence, any sampling unit in the population may be placed in class C or C dash respectively, depending whether it possess or does not possess the given attribute. In the study of attributes, we are interested in the estimate of total number of proportion.

For example:

- Proportion of defective items in a large consignment of items
- Proportion of the literates or the breadwinners in a town
- Proportion of sales a particular product accounts for
- Proportion of the viewing public of a particular program
- Proportion of trees with a diameter of eleven inches or more

Notations:
Let us consider that a population with N units Y one, Y two … YN is classified into two disjoint and exhaustive classes C and C dash respectively with respect to a given attribute. Let the number of individuals in the classes C and C dash be A and A dash respectively such that A plus A dash is equal to N.
Then,
P equals to The Proportion of units possessing the given attribute equal to A divided by N.

Q is equal to the proportion of units, which does not possess the given attribute, which equals to A dash divided by N, which is equal to one minus P.
In statistical language, P and Q are the proportion of successes and failures respectively in the population.
Let us consider Simple Random Sample of size n from this population. Let 'a' be the number of units in the sample possessing the given attribute.
Then, p equals to Proportion of sampled units possessing the given attribute is equal to a divided by n.
and q is equal to Proportion of sampled units which do not possess the given attribute which is equal to one minus p.

With the i^{th} sampling unit, let us associate a variate Y_i (i equal to one, two, up to N) defined as follows:
Y_i is equal to one, if it belongs to the class C i.e., if it possess the given attribute.
And Y_i is equal to zero, if it belongs to the class C dash i.e. if it does not possess the given

attribute.
Similarly, let us associate a variable y_i (i equals to one, two, up to n) with the i^{th} sample unit defined as follows:

y_i is equal to one , if the i^{th} sampled unit possess the given attribute.
And Y_i is equal to zero, if the i^{th} sampled unit does not possess the given attribute.
Then, summation Y_i is equal to A, the number of units in the population possessing the given attribute.
Summation y_i is equal to a, the number of sample units possessing the given attribute.

Thus, population mean Y bar is equal to summation Yi, i runs from one to N divided by N, which is equal to A divided by N, which is equal to P. Call this equation as equation Number three.
And sample mean y bar is equal to summation yi, i runs from one to n divided by n which is equal to a divided by n which is equal to p. Call this equation as equation Number four.
Similarly, we have
Summation Y i square is equal to A which is equal to N multiplied by P and
Summation y i square is equal to a, which is equal to n multiplied by p.

5. **Estimation of Population Process**

Theorem 5: Sample proportion ‘p’ is an unbiased estimate of the population Proportion ‘P’
That is Expected value of sample proportion p is equal to P

Proof: we know that in simple Random sampling, the sample mean provides an unbiased estimate of the population mean. That is Expected value of sample mean y bar is equal to the population mean Y bar.
But, from Equations (three) and (four) we have,
Population mean Y bar is equal to P and sample mean y bar is equal to p.
Hence, Expected value of sample proportion p is equal to Population proportion P.

Corollary: We have expected value of sample proportion p is equal to Population proportion P.
Which implies expected value of N multiplied by sample proportion p is equal to N multiplied by Expected value of sample proportion p
Which is equal to N multiplied by population proportion P which is equal to A
Hence, an unbiased estimate of A which we denote by A cap is equal to N multiplied by sample proportion p.

In a Simple Random Sample, the estimate of the population mean is identical to the mean of the sample.

Estimate of population mean X bar cap is equal to summation x by n

Where,

x is equal to an observed value.
x bar cap is equal to estimate of the population mean.
Summation x is equal to sum of all observed x values in the sample.
n is equal to number of observations in the sample.

In a Simple Random Sample, the estimate of the population proportion is identical to the sample proportion, which is nothing but the mean of the sample.

Estimate of population proportion P cap is equal to summation yi by n
Where,

y is equal to an observation possessing an attribute.

P cap is equal to estimate of the population proportion.
Summation yi is equal to sum of all observed y values in the sample.
n is equal to number of observations in the sample.

A farmer randomly selects 10 eggs from a gross of 12 dozen eggs (144 eggs) and he carefully weighs each egg.

The following weights were recorded in grams:

0.75, 0.70, 0.55, 0.50, 0.60, 0.65, 0.75, 0.65, 0.75, 0.50

What is the mean weight of the gross of eggs?

Using this formula, we can determine the mean weight of all of the eggs.
Mean weight of the gross eggs is equal to
X bar cap is equal to summation x by n.
Equals to six point four by ten.
Equals to zero point six four grams.

For a Simple Random Sample, the estimation formula of a total for the population is:
X cap equals to N into summation x by n.
Where,
x is equal to an observed value.
Summation x is equal to sum of all observed x values in the sample.
n is equal to number of observations in the sample.
N is equal to total number of observations in the population.
x cap is equal to estimated population total.
It is just the estimate for the mean value multiplied by the number of units in the population.

A simple random sample of size hundred is drawn from a population of size three hundred
A fraction of students doing part time jobs during the college are:

Xi = 0 if she/ he is not doing the job
Xi = 1 if she/ he is doing the job

Student : one, two , three, four, ………………………….., ninety nine ,hundred
Xi : zero, one, one,………………………………..,one, one, one
Summation xi is equal to sixty five

Estimate the proportion of students doing a part time job during the college career.
P cap is equal to x bar is equal to summation xi by hundred
Equals to sixty five by hundred which is equal to zero point six five
Sixty five percent of the students had a part time job.

The sample mean for random sampling with or without replacement is an unbiased estimator of the population mean. Consequently, the sample proportion p for random sampling with or without replacement is an unbiased estimator of the population proportion.

Since the sample mean and sample proportion of simple random samples are unbiased estimators of the population mean and population proportion respectively, they would seem to be reasonable estimators of those parameters.
In fact, they are the most widely used estimators of the population mean and the population percentage.

Here's a summary of our learning in this session:

- Process of estimation of parameters
- Estimation of population mean and total under SRSWR
- Estimation of population mean and total under SRSWOR
- Sampling for attributes
- Estimation of population proportion