1. Introduction

Welcome to the series of E-learning modules on Allocation- Proportional , Neyman or Optimum Allocation for fixed precision. In this module we are going cover the allocation sample sizes, Two methods of allocation- Proportional, Neyman or Optimum Allocation, Derivation of the formulae for the sample size using two techniques.

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

- Allocation of sample sizes
- Methods of allocation of sample sizes under Stratified Random Sampling
- Derivation of the formula for Proportional Allocation
- Derivation of the formula for Neyman or Optimum Allocation for a fixed precision

In stratified sampling, the population of interest can be divided into *'k'* non-overlapping sub-populations or strata of size 'Nh' according to a stratification variable *'Z'*, where 'h' is equal to 1,2, etc to k.

The stratification variable is either discrete or has to be recoded into a discrete variable with as many unique values as the desired number of strata.

The values of 'Z' are denoted by 'Zh'.

The total sample size n is then allocated to the strata, so that 'n is equal to summation, 'h' runs from '1 to k', 'nh'.

Samples of size 'nh' are drawn within each of the 'k' strata.

There are 4 choices to make when planning a stratified sample:

- What variable to use for stratification?
- How are the stratum boundaries defined?
- By what method should the total sample size be allocated to the strata?
- What sample design is used to draw the samples within the strata?

The first two questions have already been discussed in the earlier topics.

In this topic let us concentrate on the third question.

By what method should the total sample size be allocated to the strata?

In a practical situation, the total sample size is normally decided by a single consideration viz., the budget available for a survey.

However, the allocation of sample size to different strata is made by a statistician.

Here, it is important to remember that the precision of estimators largely depends on the allocation plans.

In fact, in order to increase the efficiency of estimators, it is imperative to choose a proper allocation plan.

The precision and cost of a stratified design is influenced by the way that sample elements are allocated to strata.

To obtain efficient results the allocation of sample sizes ‘nh’, where ‘h’ equal to 1,2, etc k, that is the number of units to be selected from the ‘h^{th}’stratum, and the total sample size ‘n’ which is equal to n-1 plus n-2 plus etc plus nk being given, is done in the following ways:

Three considerations which can affect the choice of allocation In the process of allocation of sample sizes:

1. The strata sizes i.e., the values of Nh, where h=1,2,…,k
2. Variability within a stratum, and
3. The cost of observing a sampling unit within various strata

2. Allocation of Sample Sizes

ALLOCATION OF SAMPLE SIZES

Total number of samples ‘n’ that we draw from different strata should be in such a way that summation , h runs from 1 to k, nh is equal to n. The selection of the sample sizes nh from each stratum can be done using the 2 methods:

One, Proportional Allocation

Two, Optimum or Neyman’s Allocation

Proportional allocation

First approach is proportionate stratification.

In this approach, the sample size of each stratum is proportionate to the population size of the stratum. Strata sample sizes are determined by the equation ‘Nh’ is equal to ‘Nh by N’ into n

where ‘nh’ is the sample size for stratum ’*h’*, ‘N_h’ is the population size for stratum ’*h’*, ‘N’ is total population size, and ‘n’ is total sample size.

First, ‘n’ can be allocated to the strata proportionally to ‘Nh’ so that the total number of observations drawn is equal to the required sample size.

One advantage of proportional allocation is that the inclusion probabilities are constant.

Generally, the inclusion probability of the ‘j^{th}’ element in stratum ‘h’ can be expressed as:

P h j is equal to nh divided by Nh.

Since, in proportional allocation, the stratum sample sizes ‘nh’ are by definition proportional to ‘Nh’, the above ratio is constant.

Proportionate allocation uses a sampling fraction in each of the strata that is proportional to that of the total population.

In this, the items are selected from each stratum in the same proportion as they exist in the population.

In other words:

The allocation of sample sizes is termed as proportional if the sample fraction that is the ratio of the sample size to the population size remains the same in all the strata.

For instance, if a population consists of 60 percent in the male stratum and 40 percent in the female stratum, then the relative size of the two samples, that is 3 males, 2 females should reflect this proportion.

3. Optimum Allocation or Neyman's Allocation

Optimum allocation or Neyman's allocation The second approach to stratified sampling is Optimum Allocation or Disproportionate Stratification, which can be a better choice if sample elements are assigned correctly to strata.

To take advantage of disproportionate stratification, researchers need to answer some questions like:

- Given a fixed budget, how should the samples be allocated to get the most precision from a stratified sample?
- Given a fixed sample size, how should sample be allocated to get the most precision from a stratified sample?
- Given a fixed budget, what is the most precision that I can get from a stratified sample?
- Given a fixed sample size, what is the most precision that I can get from a stratified sample?
- What is the smallest sample size that will provide a given level of survey precision?
- What is the minimum cost to achieve a given level of survey precision?
- Given a particular sample allocation plan, what level of precision can I expect?

And so on…

In case of Optimum Allocation, each stratum is proportionate to the standard deviation of the distribution of the variable.

Larger samples are taken in the strata with the greatest variability to generate the least possible sampling variance.

Proportional allocation:

This method was proposed by Bowley and has its motivation in the argument that samples are distributed to different strata in proportion to strata sizes.

That is, larger strata should get a larger share of allocation while the smaller strata are allocated smaller number of units.

This method is simple to use and numerous estimates can be made with greater degree of precision by this method.

However, it does not take into account an important aspect associated with stratified sampling, namely, the variability within strata.

In this method the sample allocation to the h^{th}-stratum is made by ‘nh’ is equal to ‘n into Nh’ divided by ‘N’.

If the sample size in the h^{th} stratum is nh proportional to Nh, then the sample is said to have been selected under Proportional allocation.

Finding ‘n_h ‘’using Proportional Allocation:

Under Proportional allocation

we have nh is proportional to Nh

which implies nh is equal to k into Nh. Call this as (1) where k is the constant of proportionality.

k is equal to nh divided by Nh. Call this as (2)

Adding summation on both sides of equation 1, we get:

Summation, h runs from 1 to k, nh is equal to k into summation h runs from 1 to k, Nh

Which implies n is equal to k into N

Which implies k is equal to n by N. Call this as (3)

By comparing equation 2 and equation 3 we get

n divided by N is equal to nh divided by Nh

Therefore, Nh is equal to Nh into n by N.

Optimum allocation or Neyman’s allocation:

This allocation method is given by Neyman.

Here, the basic idea is that, for population with larger variability, sample sizes have to be large.

That is, we should take larger allocation sample sizes for strata with higher variability.

If the sample size in the h^{th} stratum is directly proportional to the product of the population size in the h^{th} stratum and the population root mean square in the h^{th} stratum is,

nh is proportional to Nh into Sh .

Then, the sample is said to have been selected under Optimum allocation or Neyman’s allocation.

Also, as we want that larger strata should have a higher allocation.

So, to improve the precision of estimates, an important criterion of allocating the sample sizes should be to minimize the variance of stratified sample mean for a fixed total sample size ‘n’ and for a fixed cost ‘C’.

To find n_h using Optimum Allocation:

We can find '*nh*' under Optimum Allocation using the following two procedures:

N is equal to summation, h runs from 1 to k, nh.

1) By minimizing the variance of Stratified random sampling for a fixed sample size , that is summation nh is equal to n

2) By minimizing the variance of Stratified random sampling for a fixed cost, say 'c'.

Neyman Allocation:

How to Maximize Precision, Given a Stratified Sample With a Fixed Sample Size

Sometimes, researchers want to find the sample allocation plan that provides the most precision, given a fixed sample size. The solution to this problem is a special case of optimal allocation, called '*Neyman Allocation'*.

The equation for Neyman Allocation can be derived from the equation for optimal allocation by assuming that the direct cost to sample an individual element is equal across strata.

Based on Neyman allocation, the best sample size for stratum '*h'* would be:

nh is equal to n into (Nh into Sh) divided by (summation, h runs from 1 to k, Nh into Sh.

where 'nh' is the sample size for stratum '*h'*, 'n' is total sample size, 'Nh' is the population size for stratum '*h'*, and 'Sh' is the Root mean square deviation of stratum '*h'*.

4. Finding n_h for a Fixed Sample Size

1) *Finding n_h for a fixed sample size:*

Theorem:

In stratified random sampling variance is minimum for a fixed sample size, say:

summation nh is equal to n, if nh is proportional to Nh into Sh

Proof:

We know that

Variance of (y bar st) is equal to '1 by N square' into summation, 'h runs from 1 to k', 'Nh' into 'Nh minus nh' into 'Sh square' by 'nh'.

We have to minimize 'variance of y bar st' subject to the condition that summation 'nh is equal to n'

This is equivalent to the minimizing of:

'Phi' is equal to 'variance of y bar st' plus 'lambda' into 'summation h, runs from 1 to k, nh minus n'

where lambda is known as the Lagrange's multiplier.

Phi is equal to 1 by N square into summation, h runs from 1 to k, Nh into 'Nh minus nh into 'Sh square by nh' plus lambda times 'summation, h runs from 1 to k, nh minus n'.

Which is equal to '1 by N square' into 'summation h runs from 1 to k', 'Nh square into Sh square' by 'nh' minus '1 by N square' into summation, h runs from 1 to k, 'Nh into Sh square' plus 'lambda times summation, h runs from 1 to k, nh' minus 'lambda into n'

Therefore differentiating phi with respect to nh and equating to zero we get:

'Delta phi' by 'deta nh' is equal to 'zero', which implies 'minus Nh square' into 'Sh square' by 'N square into nh square' plus 'lambda' is equal to 'zero'.

Which implies lambda is equal to 'Nh square into Sh square' by 'N square into nh square'.

Which implies nh square is equal to 1 by N square into Nh square into Sh square by lambda

Which implies nh is equal to 1 by N into Nh into Sh by square root of lambda. Call this as equation (1)

By taking summation over equation 1 we get

Summation nh is equal to summation Nh into Sh by N into square root of lambda

Which implies n is equal to 1 by N into summation Nh into Sh by square root of lambda

Which implies square root of lambda is equal to 1 by N into summation Nh into Sh by n. Call this as equation (2)

By substituting equation 2 in equation 1 we get

nh is equal to 1 by N into Nh into Sh divided by summation Nh into Sh by N into n

Which implies nh is equal to (n by summation Nh into Sh) whole multiplied by Nh into Sh

Which implies nh is equal to k into Nh into Sh

Which implies k is equal to n by summation, h runs from 1 to k, Nh into Sh

Hence nh is proportional to Nh into Sh

Therefore under Optimum Allocation an expression for the sample size from h^{th} stratum is given by nh is equal to n into Nh into Sh divided by summation, h runs from 1 to k, Nh into Sh

5. Finding n_h for a Fixed Cost

2) Finding n_h for a fixed cost:

Theorem:

Under Stratified random sampling variance is minimum for a fixed cost, say

'c knot' plus 'summation ch into nh' is equal to 'C' if 'nh' is proportional to 'Nh into Sh'

where 'C' stands for the overall budget, 'C knot' for the fixed overhead cost, and 'Ch' is the average cost of observing the study variable for each unit selected in the sample from the hth stratum.

Then, an optimum allocation is given by that value of nh for which C is minimum. And, using standard techniques from calculus, one can see that such a value exists.

Proof:

We know that:

Variance of (y bar st) is equal to '1 by N square' into summation, h runs from 1 to k, Nh into (Nh minus nh) into 'Sh square by nh'

We have to minimize variance of y bar st subject to the condition that 'C knot' plus 'summation ch into nh' is equal to 'C'

This is equivalent to the minimizing of:

Phi is equal to 'variance of y bar st' plus lambda into 'C knot' plus 'summation h, runs from 1 to k, ch into nh' minus 'C'

Where, lambda is known as the Lagrange's multiplier.

Phi is equal to '1 by N square' into summation, h runs from 1 to k, Nh into 'Nh minus nh' into 'Sh square by nh' plus lambda times ('C knot' plus summation h, runs from 1 to k, 'ch into nh' minus C)

Which is equal to 1 by N square into summation h runs from 1 to k, 'Nh square into Sh square by nh' minus '1 by N square' into summation, h runs from 1 to k, 'Nh into Sh square' plus lambda into 'c knot' plus 'lambda times summation, h runs from 1 to k, ch into nh' minus lambda into c.

Therefore, differentiating phi with respect to nh and equating to zero we get:

'Delta phi' by 'delta nh' is equal to zero, which implies 'minus Nh square into Sh square' by 'N square' into 'nh square' plus 'lambda into ch' is equal to zero.

Which implies 'lambda into ch' is equal to 'Nh square into Sh square' by 'N square into nh square'

Which implies nh is equal to '1 by N' into 'Nh into Sh' by 'square root of lambda into ch' .

Call this as equation (1)

By taking summation over equation 1 we get:

Summation nh is equal to summation 'Nh into Sh' by 'N into square root of lambda into ch'

Which implies square root of lambda is equal to '1 by N' into summation 'Nh into Sh' by 'n into square root of ch'. Call this as equation (2)

By substituting equation 2 in equation 1 we get:

nh is equal to '1 by N' into 'Nh into Sh' divided by 'summation Nh into Sh into root ch' by 'N into n into root ch'.

Which implies nh is equal to 'n into Nh into Sh' divided by 'summation Nh into Sh'

Which implies nh is equal to 'k into Nh into Sh'

Which implies k is equal to n by summation , h runs from 1 to k, 'Nh into Sh'

Where, nh is proportional to Nh into Sh

Therefore under Optimum or Neyman Allocation an expression for the sample size from h^{th} stratum is given by:

nh, which is equal to n into Nh into Sh divided by summation Nh into Sh

It may be noted out that proportional allocation approach is simple and, if all one knows about each stratum is, the number of items in that stratum, it is generally also the preferred procedure.

In disproportional sampling, different strata are sampled at different rates.

As a general rule, when variability among observations within a stratum is high, one samples that stratum at a higher rate than for strata with less internal variation.

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

- Illustrated allocation of sample sizes under Stratified Random Sampling
- Discussed two types of allocation of sample sizes
- Derived the formula for the sample size under Proportional allocation
- Derived the formula for the sample size under Neyman or Optimum allocation for fixed sample size and fixed cost