Frequently Asked Questions

1. What are the choices to make when planning a stratified sample?

A**nswer:**

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

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

2. Why allocation of sample sizes for the stratum is most important?

Answer:

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 n_h, $h=1,2,\ldots,k$ that is, the number of units to be selected from the h^{th} stratum , should be done using proper allocation techniques

3.What do you mean by Proportional Allocation?

Answer:

One approach of selecting a sample from the strata is proportional allocation. With proportionate stratification, the sample size of each stratum is proportionate to the population size of the stratum. Strata sample sizes are determined by the following equation

- $n_h = (N_h / N) * n$
- where n_h 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 N_h so that the total number of observations drawn is equal to the required sample size

  For instance, if the population consists of 60% in the male stratum and 40% in the female stratum, then the relative size of the two samples (three males, two females) should reflect this proportion.

4. Write a note on Optimum or Neyman Allocation

Answer:

Another approach of selecting a sample is Optimum allocationor Neyman Allocation (which can be a better choice (e.g., less cost, more precision) if sample elements are assigned correctly to strata.

Optimum allocation (or Disproportionate 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.

This allocation method is given by Neyman (1934). 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 i.e., $n_h \prec N_h S_h$ 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 (i.e., to reduce the variance), an important criterion of allocating the sample sizes should be to minimise the variance of stratified sample mean for a fixed total sample size, n and for a fixed cost C

5. What are the points to be considered under Optimum (Neyman) Allocation?

Answer:

One has to consider the following points under Optimum (Neyman) Allocation

- Given a fixed budget, how should sample 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.

6. Under Proportional Allocation derive an expression for the sample size from the h^{th} stratum.

Answer:

Under P.A. we have

$$n_h \prec N_h$$

$$n_h = kN_h \longrightarrow 1$$ where k is the constant of proportionality.

$$k = \frac{n_h}{N_h} \longrightarrow 2$$

Adding $\sum$ on both sides of eqn 1 we get

$$\sum n_h = k \sum N_h$$

$$\Rightarrow n = kN$$

$$\Rightarrow k = \frac{n}{N} \longrightarrow 3$$

By comparing equation 2 and equation 3 we get

$$\frac{n}{N} = \frac{n_h}{N_h}$$

$$n_h = \frac{N_h}{N} * n$$

7. In stratified random sampling prove that variance is minimum for a fixed sample size, say $\sum n_h = n$ if $n_h \prec N_h S_h$.

Answer:

We know that

$$V(\bar{y}_{st}) = (\frac{1}{N^2})\sum N_h (N_h - n_h)\frac{s_h^{\ 2}}{n_h}$$

We have to minimize $V(\bar{y}_{st})$ subject to the condition that $\sum n_h = n$

This is equivalent to the minimizing of $\phi = V(\bar{y}_{st}) + \lambda(\sum n_h - n)$

where λ is known as the lagrange's multiplier

$$\phi = \frac{1}{N^2}\sum N_h (N_h - n_h)\frac{S_h^{\ 2}}{n_h} + \lambda(\sum n_h - n)$$

$$= \frac{1}{N^2}\sum \frac{N_h^{\ 2} S_h^{\ 2}}{n_h} - \frac{1}{N^2}\sum N_h S_h^{\ 2} + \lambda \sum n_h - \lambda n$$

Therefore differentiating ϕ w.r.t n_h and equating to zer0 we get

$$\frac{\partial \phi}{\partial n_h} = 0 \Rightarrow \frac{-1}{N^2}\frac{N_h^{\ 2} S_h^{\ 2}}{n_h^{\ 2}} + \lambda = 0$$

$$\Rightarrow \lambda = \frac{1}{N^2}\frac{N_h^{\ 2} S_h^{\ 2}}{n_h^{\ 2}}$$

$$\Rightarrow n_h^{\ 2} = \frac{1}{N^2}\frac{N_h^{\ 2} S_h^{\ 2}}{\lambda}$$

$$\Rightarrow n_h = \frac{1}{N}\frac{N_h S_h}{\sqrt{\lambda}} \longrightarrow 1$$

By taking $\sum$ over equation 1 we get

$$\sum n_h = \frac{1}{N}\frac{\sum N_h S_h}{\sqrt{\lambda}}$$

$$\Rightarrow n = \frac{1}{N}\frac{\sum N_h S_h}{\sqrt{\lambda}}$$

$$\Rightarrow \sqrt{\lambda} = \frac{1}{N}\frac{\sum N_h S_h}{n} \longrightarrow 2$$

By substituting equation 2 in equation 1 we get

$$\Rightarrow n_h = \frac{1}{N}\frac{N_h S_h}{\frac{\sum N_h S_h}{Nn}}$$

$$\Rightarrow n_h = \frac{n}{\sum N_h S_h} \cdot N_h S_h$$

$$\Rightarrow n_h = k \cdot N_h S_h \quad \text{where} \quad \Rightarrow k = \frac{n}{\sum N_h S_h}$$

$$n_h \prec N_h S_h$$

Therefore under O.A. an expression for the sample size from h[th] stratum is given by $\Rightarrow n_h = \frac{n}{\sum N_h S_h} \cdot N_h S_h$

8. Under Stratified random sampling prove that variance is minimum for a fixed cost, say $c_0 + \sum c_h n_h = c$ if $n_h \prec N_h S_h$.

Answer:

We know that

$$V(\bar{y}_{st}) = (\frac{1}{N^2})\sum N_h(N_h - n_h)\frac{s_h^{\ 2}}{n_h}$$

We have to minimize $V(\bar{y}_{st})$ subject to the condition that $c_0 + \sum c_h n_h = c$. This is equivalent to the minimizing of $\phi = V(\bar{y}_{st}) + \lambda(c_0 + \sum c_h n_h - c)$

$$= \frac{1}{N^2}\sum N_h(N_h - n_h)\frac{S_h^{\ 2}}{n_h} + \lambda(c_0 + \sum c_h n_h - c)$$

$$= \frac{1}{N^2}\sum \frac{N_h^{\ 2} S_h^{\ 2}}{n_h} - \frac{1}{N^2}\sum N_h S_h^{\ 2} + \lambda c_0 + \lambda \sum c_h n_h - \lambda c$$

Therefore differentiating ϕ w.r.t n_h and equating to zer0 we get

$$\frac{\partial \phi}{\partial n_h} = 0 \Rightarrow \frac{-1}{N^2}\frac{N_h^{\ 2} S_h^{\ 2}}{n_h^{\ 2}} + \lambda c_h = 0$$

$$\Rightarrow \lambda c_h = \frac{1}{N^2}\frac{N_h^{\ 2} S_h^{\ 2}}{n_h^{\ 2}}$$

$$\Rightarrow n_h = \frac{1}{N}\frac{N_h^{\ 2} S_h^{\ 2}}{n_h\sqrt{\lambda c_h}} \longrightarrow 1$$

$$\Rightarrow \sqrt{\lambda} = \frac{1}{N}\frac{\sum N_h S_h}{n\sqrt{c_h}} \longrightarrow 2$$

By substituting equation 2 in equation 1 we get

$$\Rightarrow n_h = \frac{1}{N} \frac{N_h S_h}{\frac{\sum N_h S_h \sqrt{c_h}}{Nn\sqrt{c_h}}}$$

$$\Rightarrow n_h = \frac{n}{\sum N_h S_h} \cdot N_h S_h$$

$$\Rightarrow n_h = k \cdot N_h S_h \quad \text{where} \quad \Rightarrow k = \frac{n}{\sum N_h S_h}$$

$$n_h \prec N_h S_h$$

Therefore under O.A. an expression for the sample size from h^{th} stratum is given by $\Rightarrow n_h = \frac{n}{\sum N_h S_h} \cdot N_h S_h$.

9. What are the factors which affect the choice of allocation to the strata?

Answer:

The factors which can affect the choice of allocation.

(1) The strata sizes i.e., the values of N_h $(1 \le h \le k)$,

(2) Variability within a stratum, and

(3) The cost of observing a sampling unit within various strata

10. How do you allocate the sample observations to each stratum?

Answer:

In stratified sampling, the population of interest can be divided into k non-overlapping sub-populations or **strata** of size N_h (h = 1, ... , k) according to a stratification variable Z. 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 total required sample size n is then allocated to the strata, such that $n = n_1+n_2+\ldots+n_k$. Samples of size n_h are drawn from each of the k strata such that n_1 observations from stratum 1, n_2 observations from stratum2,……,n_k observation from stratum k are selected from k strata.

11. Under Proportional Allocation what is the probability that an element belong to the stratum h?

Answer:

The probability that the j^{th} element belong to the stratum h is expressed as

$$p_{hj} = n_h/N_h.$$

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

Since, in proportional allocation, the stratum sample sizes n_h are by definition proportional to N_h, the above ratio is constant

12. How the sampling fraction is used in Proportional Allocation?

Answer:

Sampling fraction is nothing but n/N. Proportional 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. 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.

13. Why Optimal allocation is preferred to Proportional Allocation?

Answer:

Data representing each subgroup are taken to be of equal importance if suspected variation among them warrants stratified sampling. If, on the other hand, the very variances vary so much, among subgroups, that the data need to be stratified by variance, there is no way to make the subgroup sample sizes proportional (at the same time) to the subgroups' sizes within the total population. (What is the most efficient way to partition sampling resources among groups that vary in both their means and their variances. The ideal sample allocation plan would provide the most precision for the least cost. Optimal allocation does just that. Optimum allocation not only considers the stratum size but also population mean square of each stratum. Also Optimum allocation gives more précised estimates than Proportional Allocation.

14. What are the merits and demerits of Optimal Allocation?

Answer:

Merits:

1) When larger variation is present in the subgroups optimal allocation is the suitable technique
2) Optimal allocation gives more précised estimates than other techniques of selection of samples.
3) Gives most precision for the least cost

Demerits:

1) computation procedure is not easy as Proportional Allocation
2) Optimum Allocation requires the knowledge of S_h , the population root mean square deviation which is usually unknown

15. Suppose that in a company there are the following staffs:
male, full time: 90, male, part time: 18, female, full time: 9
female, part time: 63
Total: 180
Select a sample of 40 staff, stratified according to the above categories using Proportional Allocation

Answer:

Under Stratified Proportional Allocation , the samples drawn from each stratum = $\frac{N_h n}{N}$

We have N = 180 with N1=90, N2=18, N3= 9 and N4=63, n=40

Number of male full time staff to be selected is 90X40/180 = 20
Number of male part time staff to be selected is 18X40/180 = 04
Number of female full time staff to be selected is 9X40/180 = 2
Number of female part time staff to be selected is 63X40/180 = 14