Frequently Asked Questions

1. What are the principal steps in Stratified Random Sampling?

Answer:
Principal Steps in stratified sampling are:

1) Identify and define the population.

2) Determine the desired sample size.

3) Identify the variable and subgroups (strata) for which you want to guarantee appropriate, equal representation.

4) Classify all members of the population as members of one identified subgroup.

5) Randomly select, using a table of random numbers) an "*appropriate*" number of individuals from each of the subgroups. Appropriate means *an equal number of individuals.*

2. What are the properties of Stratified random sampling?

Answer:
Stratified random sampling refers to a sampling method that has the following properties.

- The population consists of N elements.
- The population is divided into h groups, called **strata**.
- Each element of the population can be assigned to one, and only one, stratum.
- The number of observations within each stratum N_h is known, and

  $N = N_1 + N_2 + N_3 + ... + N_h + \ldots + N_k$.

- The researcher obtains a probability sample from each stratum

3. Outline the structure of distribution of population observations under Stratified Random sampling?

Answer:
The population consists of N elementary sampling units that are grouped exclusively and exhaustively into k strata such that stratum 1 contains N_1 elementary sampling units, stratum 2 contains N_2 sampling units,..., and stratum h contains N_h elementary sampling units . Hence the population of size N is subdivided into k strata of sizes $N_1, N_2, \ldots, N_h, \ldots, N_k$.

Let Y_{hi} represents the value of the variable of interest for the i^{th} elementary unit within stratum h and the elementary units in any particular stratum h are labelled from 1 to N.

or

Y_{hi} is the ith population unit of the population from the h^{th} stratum.

Let the population units be:

Stratum No.	Observations							Stratum size	Stratum mean
1	Y_{11}	Y_{12}	….	…..	Y_{1i}	….	Y_{1N1}	N_1	$\overline{Y_1}$
2	Y_{21}	Y_{22}	….	…..	Y_{2i}	….	Y_{2N2}	N_2	$\overline{Y_2}$
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
h	Y_{h1}	Y_{h2}	….	…..	Y_{hi}	….	Y_{hNh}	N_h	$\overline{Y_h}$
.	.	.	.	.	.	.	.	.	.
k	Y_{k1}	Y_{k2}	….	…..	Y_{ki}	….	Y_{kN1}	N_k	$\overline{Y_k}$
								$N=\sum_{h=1}^{k} N_h$	

4. Briefly explain the distribution of sample observations under Stratified Random sampling?

Answer:
Suppose we take a sample of size 'n ' which is subdivided into various strata having sizes n_1, n_2,….n_h,…,n_k and

let y_{hi} represents the value of the variable of interest for the ith elementary sample unit within stratum h and the elementary units in any particular stratum h are labelled from 1 to n .

i.e., y_{hi} is the i^{th} sample unit of the population from the h^{th} stratum.

Then the sample units are

Stratum No.	Observations							Stratum samplesize	Stratum samplemean
1	y_{11}	y_{12}	….	…..	y_{1i}	….	y_{1n1}	n_1	$\overline{y_1}$
2	y_{21}	y_{22}	….	…..	y_{2i}	….	y_{2n2}	n_2	$\overline{y_2}$
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
h	y_{h1}	y_{h2}	….	…..	y_{hi}	….	y_{hnh}	n_h	$\overline{y_h}$
.	.	.	.	.	.	.	.	.	.
k	y_{k1}	y_{k2}	….	…..	y_{ki}	….	y_{kn1}	n_k	$\bar{y}$
								$n=\sum_{h=1}^{k} n_h$	

5. List the parameters involved in Stratified random sampling.

Answer:

Population mean of the h[th] stratum is $\overline{Y_h} = \frac{\sum_{i=1}^{Nh} Y_{hi}}{N_h}$

Which implies $N_h \overline{Y_h} = \sum_{i=1}^{Nh} Y_{hi}$

Population mean square in the h[th] stratum is

$$S_h^2 = \frac{\sum_{i=1}^{Nh}(Y_{hi} - \overline{Y_h})^2}{N_h - 1}$$

Similarly we have population variance in the h[th] stratum

$$\sigma_h^2 = \frac{\sum_{i=1}^{Nh}(Y_{hi} - \overline{Y_h})^2}{N_h}$$

Therefore, we have $(N_h - 1)\, S_h^2 = N_h \sigma_h^2$

6. What is the sample estimators used in Stratified Random sampling?

Answer:

A sample mean of the h^th stratum is $\overline{y_h} = \frac{\sum_{i=1}^{nh} y_{hi}}{n_h}$

Which implies $n_h\, \overline{y_h} = \sum_{i=1}^{nh} y_{hi}$

Sample mean square in the h^th stratum is:

$$s_h^2 = \frac{\sum_{i=1}^{nh}(y_{hi} - \overline{y_h})^2}{n_h - 1}$$

7. What do you mean by stratum weight and how it is related to the population mean?

Answer:
Under Stratified Random sampling the population mean (Mean of the entire population)is given by

$$\overline{Y} = \frac{\sum_{h=1}^{k}\sum_{i=1}^{Nh} Y_{hi}}{N}$$

which implies

$$\overline{Y} = \frac{\sum_{h=1}^{k} N_h \overline{Y}_h}{N} = \sum_{h=1}^{k} W_h \overline{Y}_h$$ Where $W_h = N_h/N$ is the stratum weight

Therefore, $W_h = N_h/N$, the ratio of the stratum size to the population size is known as a Stratum weight. Population mean can be expressed interms of stratum weight.

8. What do you mean by $\overline{y_{st}}$? Is it equivalent to the sample mean $\bar{y}$?

Answer:
For the population mean per unit the estimate used in Stratified Sampling is:

$$\overline{y_{st}} = \frac{\sum\limits_{h=1}^{k} N_h \overline{y_h}}{N} = \sum_{h=1}^{k} W_h \overline{y_h}$$

Where, $N = N_1+N_2+\ldots\ldots.+N_k$

The estimate $\overline{y_{st}}$ (st means stratified) is not in general the same as the sample mean. The sample mean

$$\bar{y} = \frac{\sum\limits_{h=1}^{k} n_h \bar{y}_h}{n} = \sum_{h=1}^{k} w_h \overline{y_h}$$

The difference is that in $\overline{y_{st}}$ the estimates from the individual strata receive their correct weights N_h/N. It is evident that $\bar{y}$ coincides with $\overline{y_{st}}$ provided that in every stratum

$\frac{n_h}{n} = \frac{N_h}{N}$ or $\frac{n_h}{N_h} = \frac{n}{N}$ or $f_h = f$. This means that the sampling fraction is the same in all strata.

9. Under Stratified random sampling with replacement prove that an unbiased estimator of population mean is given by $\bar{y}_{st} = \sum\limits_{h=1}^{k} W_h \bar{y}_h$

Answer:
Let us divide the entire population of size N into k strata such that each strata consists of $N_1, N_2, \ldots, N_k$ number of population observations. Let us draw a sample of size n from each and every stratum of the population using SRSWR.

The number of sample observations selected are $n_1, n_2, \ldots, n_k$ from k strata.

Then we know that

$$\overline{Y} = \sum_{h=1}^{k} W_h \overline{Y_h} \qquad \bar{y} = \sum_{h=1}^{k} w_h \overline{y_h} \qquad \bar{y}_{st} = \sum_{h=1}^{k} W_h \bar{y}_h$$

and

we have to prove $E(\bar{y}_{st}) = \bar{Y}$ and $E(\bar{y}) \neq \bar{Y}$.

Consider $E(\bar{y}_{st}) = E(\sum_{h=1}^{k} W_h \bar{y}_h)$

$$= \sum_{h=1}^{k} W_h E(\bar{y}_h) \text{ -----------------(*)}$$

Consider E ($\bar{y}_h$)= E ($\dfrac{\sum_{i=1}^{nh} y_{hi}}{n_h}$) = $\dfrac{\sum_{i=1}^{n} E(y_{hi})}{n_h}$ ——(2)

$$E(y_{h_i}) = \sum_i y_{h_i} p_i$$

Consider $E(y_{hi})$, since each y_{hi} takes values Y_{hi} with probability $1/N_h$ we have $E(y_{hi}) = \sum_{i=1}^{Nh} Y_{h_i} \bullet \frac{1}{N_h} = \bar{Y}_h$

By substituting the above in equation (2), we get

$$E(\bar{y}_h) = \frac{\sum_{i=1}^{nh} \bar{Y}_h}{n_h} = \frac{n_h \bar{Y}_h}{n_h} = \bar{Y}_h$$

$E(\bar{y}_h) = \bar{Y}_h$ Under SRSWR

By substituting the above equation in (*) we get

Therefore, $E(\bar{y}_{st}) = \sum_{h=1}^{k} W_h \bar{Y}_h = \bar{Y}$.

Therefore $\bar{y}_{st} = \sum_{h=1}^{k} W_h \bar{y}_h$ is an unbiased estimator of population mean, $\bar{Y}$

$$\therefore \hat{\bar{Y}} = \bar{y}_{st}$$

10. Under Stratified sampling prove that a stratified sample mean is not unbiased for the population mean.

Answer:
To show that $E(\bar{y}) \neq \bar{Y}$

Consider $E(\bar{y}) = E(\sum_{h=1}^{k} w_h \bar{y}_h)$

$$= \sum_{h=1}^{k} w_h E(\overline{y}_h)$$

$$= \sum_{h=1}^{k} w_h \overline{Y}_h \quad \text{Under SRSWR}$$

$$\neq \overline{\mathrm{Y}}$$

Therefore, $\mathrm{E}(\overline{\mathrm{y}}) \neq \overline{\mathrm{Y}}$

11. An unbiased estimator of population total is given by $N\overline{y}_{st}$ under stratified Random Sampling using SRSWR

Answer:

We have to prove that $\hat{Y} = N\overline{y}_{st}$.

The population total under stratified random sampling is given by

$$Y = \sum_h \sum_i Y_{hi} = N\overline{Y} \qquad [\boxtimes\ \overline{Y} = \sum\sum Y_{hi} / N]$$

Unbiased estimator of population total $\hat{\mathrm{Y}} = \mathrm{N}\,\hat{\overline{\mathrm{Y}}} = \mathrm{N}\overline{\mathrm{y}}_{\mathrm{st}}$

$$[E(\overline{y}_{st)} = \overline{Y} \Rightarrow \hat{\overline{Y}} = \overline{y}_{st}]$$

12. Under Stratified random sampling without replacement an unbiased estimator of population mean is also given by, $\overline{y}_{st} = \sum_{h=1}^{k} W_h \overline{y}_h$

Answer:

Let us divide the entire population of size N into k strata such that each strata consists of N1,N2,…,Nk number of population observations. Let us draw a sample of size n from each and every stratum of the population using SRSWOR.

The number of sample observations selected are n1,n2,…,nk from k strata.

we have to prove $E(\overline{y}_{st}) = \overline{Y}$ and $E(\overline{y}) \neq \overline{Y}$.

Consider $E(\overline{y}_{st}) = E(\sum_{h=1}^{k} W_h \overline{y}_h)$

$$= \sum_{h=1}^{k} \mathrm{W_h E(\overline{y}_h)} \qquad \text{-----------------}(\ *)$$

Consider E ($\overline{y}_h$)= $\overline{Y}_h$ under SRSWOR also

Already we proved under Simple Random sampling Without Replacement that sample mean is an unbiased estimator of the population mean. Since a random sample of size n_h is drawn from the h^{th} stratum using SRSWOR sample mean of the h^{th} stratum must be equal to the population mean of the h^{th} stratum

By substituting the above equation in (*) we get,

Therefore, $E(\overline{y}_{st}) = \sum_{h=1}^{k} W_h \overline{Y}_h = \overline{Y}$.

Therefore $\overline{y}_{st} = \sum_{h=1}^{k} W_h \overline{y}_h$ is an unbiased estimator of population mean, $\overline{Y}$ under stratified Random sampling with SRSWOR

$\therefore \hat{\overline{Y}} = \overline{y}_{st}$

13. Obtain an unbiased estimator of the population total Under Stratified Sampling with SRSWOR.

Answer:

The population total under stratified random sampling is given by

$Y = N\overline{Y}$ [$\because \overline{Y} = \sum\sum Y_{hi}/N$]

Unbiased estimator of population total $\hat{Y} = N\hat{\overline{Y}} = N\overline{y}_{st}$

Under Stratified Sampling with SRSWR we have

$[E(\overline{y}_{st}) = \overline{Y} \Rightarrow \hat{\overline{Y}} = \overline{y}_{st}]$

Hence an unbiased estimator of the population total Under Stratified Sampling with SRSWOR is $N\overline{y}_{st}$

14. Which of the following is more efficient?

a) $\overline{y}_{st}$ Obtained using stratified sampling with SRSWR

b) $\overline{y}_{st}$ Obtained using stratified sampling with SRSWOR

Answer:

Since SRSWOR is more efficient than SRSWR sample mean obtained for h^{th} stratum under stratified sampling with SRSWOR has least variance as compared to Stratified sampling with SRSWR. Hence $\overline{y}_{st}$ Obtained using stratified sampling with SRSWOR is more efficient than stratified sampling with SRSWR

15. In a survey on the area under crop a total of 186 villages in a district was divided into 4 strata according to the area of the villages. From each stratum SR Samples are selected and the area under the crop in the selected villages noted. The following is the data from the survey. Obtain an estimate of the total area under the crop in the district.

Stratum No.	Stratum size	Sample size	Area under the crop in the sample villages
1	72	8	14,12,8,11,12,10,13,16
2	53	5	27,20,21,22,30
3	35	4	36,47,52,61
4	26	3	92,105,82

Answer:

An estimate of the total area under the crop in the district is given by $N\overline{y}_{st}$

Where $\overline{y}_{st} = \sum_{h=1}^{k} W_h \overline{y}_h$

Stratum No.	Stratum size(N_h)	Sample size(n_h)	Area under the crop in the sample villages(y_{hi})	Wh = Nh/N	$\overline{y}_h$	Wh $\overline{y}_h$
1	72	8	14,12,8,11,12,10, 13,16	0.3871	12	4.6452
2	53	5	27,20,21,22,30	0.2850	24	6.84
3	35	4	36,47,52,61	0.1882	49	9.2218
4	26	3	92,105,82	0.1398	93	13.0014
	N= 186					33.7084

$$\bar{y}_{st} = \sum_{h=1}^{k} W_h \bar{y}_h = 33.7084$$

Hence, an estimate of the total area under the crop in the district is $N\bar{y}_{st}$

= 186X 33.7084 = 6269.76 ~ 6270.