1. Introduction

Welcome to the series of E-learning modules on Practical - Stratified Random Sampling. In this module, we are going to cover the basic problems on stratified random sampling, allocation of sample sizes, comparison of the technique with that of SRSWOR and the estimates of the standard errors.

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

- Obtain an estimate of population mean, total, variance of the estimators and its Standard error under stratified random sampling
- Understand proportional and Neyman's allocation techniques
- Compare stratified Neyman's and proportional allocations with SRSWOR

Problem 1:
In a survey of area under the crop, a total of 186 villages in a district were divided into 4 strata according to the area of the villages. From each stratum, SRS were selected and the area under the crop in the selected villages is noted.

Figure 1

Stratum no.	N_h	Sample size: n_h	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

In this table, 4 strata are given in the first column. Number of villages in each stratum is given in the second column and they are denoted Nh. The number of sample villages drawn from each stratum is given in the third column as a sample size and is denoted by nh. The areas of each village of the corresponding stratum are given in the last column.
Obtain an estimate of the total area under the crop in the district and estimate the standard error of the estimate.

Solution:
Wh is calculated as the ratio of Nh by N.
Where, N is summation Nh which is equal to one hundred and eighty six
And n is summation nh which is equal to twenty.
sh square is obtained as (summation yhi square minus nh into y bar h square) divided by (nh minus 1)

Some of the required computations are given in the following table.

Figure 2

N_h	n_h	y_{hi}	W_h	$\overline{y}_h$	$W_h \overline{y}_h$
72	8	14,12,8,11,12,10, 13,16	0.387097	12	4.645161
53	5	27,20,21,22,30	0.284946	24	6.83871
35	4	36,47,52,61	0.188172	49	9.22043
26	3	92,105,82	0.139785	93	13
186					**33.7043**

y bar h is obtained as the sum of the observations yhi given in the third column of the table divided by corresponding nh
In the sixth column, a product of Wh and y bar h is computed and the total is obtained, which is equal to thirty three point seven zero four three.

Figure 3

Stratum no.	$\sum y_{hi}^2$	s_h^2	$\frac{W_h^2 S_h^2}{n_h}$	$\frac{W_h^2 S_h^2}{N_h}$
1	1194	6	0.112383	0.012487
2	2954	18.5	0.300419	0.028341
3	9930	108.6667	0.961937	0.109936
4	26213	133	0.866266	0.099954
Total			2.241005	0.250718

In the second column, the sum of the squared yhi's are obtained and using this, the sh squares are obtained.
In the next column, values obtained by dividing (Wh square sh square by nh) and the total is obtained in the last row as (two point two four one zero zero five).
In the last column, values obtained by dividing (Wh square sh square by Nh) and the total is obtained in the last row as (point two five zero seven one eight).

Variance of estimated population total Y cap is equal to (N square) into (summation Wh square into sh square by nh) minus (summation Wh square into sh square by Nh)
Which is equal to (one hundred and eighty six whole square) into (two point two four one four minus zero point two five zero eight) which is equal to sixty eight thousand eight hundred and sixty six point seven nine seven six
Y bar st is equal to (summation Wh into y bar h) which is equal to (thirty three point seven zero four three)
Estimated population total Y cap is equal to (N into y bar st)
Which is equal to (one hundred and eighty six into thirty three point seven zero six zero)

which is equal to (six thousand two hundred and sixty nine point three one six zero)
Estimate of standard error of Y cap is equal to (square root of variance of Y cap) which is equal to (two hundred and sixty two point four two four eight)

2. Illustration 2

Problem 2:

A sample survey is to undertake a certain limit. The mean annual income of farmers in a certain area is given. The farms are satisfied according to their principal products given below for a sample of 12 farms. Compute sample size for each stratum under:

a) Neyman's Allocation
b) Proportional Allocation

Compare the precision of these methods with that of SRSWOR.

Figure 4

Types of farms	No. of farms	$\overline{y}_h$ Mean	σ_h S.D
Sheep	161	10946	2236
Wheat	195	6402	2614
Deicing	274	2228	606
Others	382	1458	230

In the first column of the table, types of the farms are given. In the second column, the number of farms in the respective types is given and the average annual income of farmers is given in the third column. The last column represents the standard deviation of the income.

Solution:

Figure 5

Types of farms	No. of farms N_h	Mean $\overline{y}_h$	S.D σ_h	s_h	$N_h S_h$	$W_h = \frac{N_h}{N}$
Sheep	161	10946	2236	2242.9766	36119.23	0.1591
Wheat	195	6402	2614	2620.7285	511042.05	0.1927
Deicing	274	2228	606	607.1089	166347.83	0.2708
Others	382	1458	230	230.3016	87975.21	0.3775
Total	N=1012					

In the table presented here, mean is denoted by y bar h and standard deviation by sigma h.
In the fifth column, Sh is calculated as sigma h into square root of (Nh by Nh minus one)
Sixth column gives the values of Nh into Sh
In the seventh column, Wh is calculated as Nh by N, where N is equal to summation Nh which is equal to one thousand and twelve

Figure 6

Types of farms	$W_h S_h$	$W_h S_h^2$	$W_h \bar{y}_h$	$(N_h - 1)s_h^2$	$N_h(\bar{y}_h - \bar{Y})^2$
Sheep	356.87	800423.19	1741.5086	804951044.5	7482104214
Wheat	505.014	1323505.5	1233.6654	1332434267	1007551698
Deicing	164.4	99811.79	603.3424	100622672.1	990089177.3
Others	86.93	20022.15	550.395	20207793.07	2725099264

In this table, for each strata, Wh into Sh values are entered in the second column and Wh into Sh square are given in the third column.
In the fourth column, product of wh and y bar h are obtained.
In the next two columns, values of (Nh minus 1) into (Sh square) and Nh into (y bar h minus Y bar whole square) are obtained.

Under proportional allocation, sample sizes for each stratum can be calculated using the formula
nh prop is equal to Nh into n by N
Under Neyman's allocation or optimum allocation, sample sizes are calculated as
Nh Neyman is equal to (Nh into Sh into n) divided by (summation h runs from 1 to k, Nh into Sh)
The table below gives the values of nh proportional allocation and nh Neyman's allocation obtained using the above formulae.

Figure 7

Types of farms	Sheep	Wheat	Deicing	Others
$n_{h\ prop}$	2	2	3	5
n_{hNey}	4	5	2	1

Estimate of population mean Y bar is y bar st is equal to (summation Wh into y bar h) which is equal to (four thousand one hundred twenty eight point nine one one four)
Variance of y bar st under proportional allocation is equal to (summation, h runs from 1 to 4

Wh into Sh square by n) minus (summation, h runs from 1 to 4 Wh into Sh square by N)
Which is equal to (one lakh eighty six thousand nine hundred and eighty point two two seven five) minus (two thousand two hundred and seventeen point one five six eight) which is equal to (one lakh eighty four thousand seven hundred and sixty three point zero seven zero seven)

Variance of y bar st under Neyman's allocation is equal to (summation, h runs from 1 to 4 Wh into Sh whole square by n) minus (summation, h runs from 1 to 4 Wh into Sh square by N)
Which is equal to (one lakh three thousand two hundred and seventy point eight two one nine) minus (two thousand two hundred and seventeen point one five six eight) which is equal to (one lakh one thousand and fifty three point six six five one)
Variance of y bar under SRSWOR is equal to (N minus n) into S square by N into n
Where, S square is equal to (1 by N minus 1) into summation (Nh minus 1) into Sh square plus summation Nh into (y bar h minus Y bar whole square)
S square is equal to fourteen lakhs three hundred and five thousand six hundred and ninety seven point nine six

Variance of y bar under SRSWOR is equal to (one thousand and twelve minus twelve) into (fourteen lakhs three hundred and five thousand six hundred and ninety seven point nine six) divided by (one thousand and twelve into twelve)
Which is equal to one lakh one seventy eight thousand and five point three nine
Gain in efficiency of Proportional Allocation with that of SRSWOR is (variance of y bar under WOR) minus (variance of y bar st under Proportional allocation) divided by (variance of y bar st under Proportional allocation) which is equal to (five point three seven five)
Gain in efficiency of Neyman's Allocation with that of SRSWOR is
(variance of y bar under WOR) minus (variance of y bar st under Neyman's allocation) divided by (variance of y bar st under Neyman's allocation) which is equal to (ten point six six)

3. Illustration 3

Problem 3:
An advertising firm, which is interested in determining how much to emphasize television advertising in a certain county, decides to conduct a sample survey to estimate the average number of hours each week that the households within that county watch television.

The county has two towns, A and B, and a rural area C. Town A is built around a factory and most households contain factory workers with school-aged children. Town B contains mainly retirees and the rural area C is mainly farmers.
There are 155 households in town A, 62 in town B and 93 in the rural area C. The firm decides to select 20 households from Town A, 8 households from Town B and 12 households from the rural area C.

The results are given in the following table.

Figure 8

Town A	35, 43, 36, 39, 28, 28, 29, 25, 38, 27,26, 32, 29, 40, 35, 41, 37, 31, 45, 34	$N_1 = 155$
Town B	27, 15, 4, 41, 49, 25, 10, 30	$N_2 = 62$
Rural Area C	8, 14, 12, 15, 30, 32, 21, 20, 34, 7, 11, 24	$N_3 = 93$

Estimate average number of hours each week that households within that county watch television and estimate the Standard error of the estimate.
Solution:
The values required for the computation of estimates of the population parameters and standard error is given in this table.

Figure 9

Area	N_h	n_h	$W_h = \frac{N_h}{N}$	Mean $\overline{y}_h$	$W_h \overline{y}_h$	s_h^2
Town A	155	20	0.5	33.90	16.95	6
Town B	62	8	0.2	25.12	5.024	18.5
Rural Area C	93	12	0.3	19	5.7	108.6667
Total	N=310				27.674	

Wh is calculated as the ratio of Nh by N where N is summation Nh which is equal to three hundred and ten
y bar h is obtained as the sum of the observations yhi given in the problem for the respective regions divided by corresponding nh
In the next column, a product of Wh and y bar h is computed and the total is obtained, which is equal to (twenty seven point six seven four) and then the values of sh square are obtained.

Therefore, average number of hours each week that household within that county watch television is (y bar st is equal to summation Wh into y bar h) which is equal to (twenty seven point six seven four) where (Wh is equal to Nh by N) approximately equal to twenty eight hours

To estimate the Standard Error of the sample mean:
Standard error of ybar st is equal to the square root of variance of y bar st
Where,
Variance of y bar st is equal to (summation Wh square into sh square by nh) minus (summation Wh square into sh square by Nh)

Figure 10

Area	$\frac{W_h^2 S_h^2}{n_h}$	$\frac{W_h^2 S_h^2}{N_h}$
Town A	0.112383	0.012487
Town B	0.300419	0.028341
Rural Area C	0.961937	0.109936
Total	1.374739	0.150764

In the second column, values are obtained by dividing (Wh square sh square by nh) and the total is obtained in the last row as (one point three seven four seven three nine).
In the last column, values obtained by dividing (Wh square sh square by Nh) and the total is obtained in the last row as (point one five zero seven six four).
Variance of y bar st is equal to (summation Wh square into sh square by nh) minus (summation Wh square into sh square by Nh)
Which is equal to (one point three seven four seven three nine) minus (zero point one five zero seven six four) which is equal to (one point two two three nine seven five)
Standard error of ybar st is equal to the square root of variance of y bar st
Which is equal to (square root of one point two two three nine seven five) which is equal to (one point one zero six three three three)

4. Illustration 4

Problem 4:
A stratified sample of size two hundred is to be taken from a population with 5 strata as shown below. Write down the sample sizes under proportional and optimum allocation and also compare the precision of these 2 estimates.

Figure 11

Stratum	1	2	3	4	5
Size	250	620	407	393	230
S_h	14	17	12	7	5

Solution
In this table, corresponding to each stratum, its size and the corresponding Sh values are given.

Figure 12

Stratum	N_h	s_h	$N_h s_h$	$W_h = \frac{N_h}{N}$	$W_h S_h^2$	$W_h S_h$
1	250	14	11900	0.34	66.64	4.76
2	620	17	105400	0.248	71.672	4.216
3	407	12	4884	0.1628	23.443	1.9536
4	393	7	2751	0.1572	7.7028	1.1004
5	230	5	1150	0.092	2.3	.46
Total	**2500**		**31225**		**171.7578**	**45.67128**

Here for the given values of Nh and Sh, the product of NhSh, the values of Wh which is equal to Nh by N, products Wh Sh square and Wh Sh are obtained and the corresponding totals are given in the last row of the table.

Under proportional allocation, sample sizes for each stratum can be calculated using the formula
nh prop is equal to (Nh into n by N)
Under Neyman's allocation or optimum allocation, sample sizes are calculated as:
Nh Optimum is equal to (Nh into Sh into n) divided by (summation, h runs from 1 to k, Nh into Sh)
The following table gives the values of sample sizes nh under proportional and optimum

allocations for each stratum using the formulae given above.

Figure 13

Areas	1	2	3	4	5
$n_{h_{prop}}$	68	50	33	31	18
n_{hOA}	76	68	31	18	7

To compare the precision of the estimates:
Variance of y bar st under proportional allocation is equal to (summation, h runs from 1 to 5 Wh into Sh square by n) minus (summation, h runs from 1 to 5 Wh into Sh square by N) Which is equal to (one hundred and seventy one point seven five seven eight by two hundred) minus (one hundred and seventy one point seven five seven eight by two thousand five hundred) which is equal to (zero point seven nine zero zero eight six)
Variance of y bar st under Optimum allocation is equal to (summation, h runs from 1 to 5 Wh into Sh whole square by n) minus (summation, h runs from 1 to 5 Wh into Sh square by N) Which is equal to (zero point seven eight zero zero zero zero five) minus (zero point zero six eight seven zero three one two) which is equal to (zero point seven one two nine seven)

From the above values, one can observe that variance of the estimate under proportional allocation is greater than that of optimum allocation.
Therefore, variance of y bar st under Optimum allocation is less than variance of y bar st under proportional allocation.
Hence, optimum allocation is more précised than proportional allocation.

5. Illustration 5

Problem 5:
A sample of 30 students is to be drawn from a population consisting of three hundred students belonging to 2 colleges A and B. The means and Standard deviations of their marks are given below. Determine the sample size under proportional allocation. Obtain the standard error of the estimate of the population mean and compare its efficiency with SRSWOR.

Figure 14

Colleges	Total No. of students N_h	Means	S.D
College A	200	30	10
College B	100	60	40

Solution:

Figure 15

Colleges	Total No. of students N_h	n_{hOA}
College A	200	20
College B	100	10
Total	N=300	n=30

In this table, total number of students of college A and college B is represented by Nh.
In the second and third columns, average values and standard deviation of marks are given respectively.
The sample size under Proportional Allocation can be obtained as follows:
nh proportional is equal to (Nh into n by N) which is equal to (Nh into thirty divided by three hundred)
Sample sizes obtained under proportional allocation using the above formula for the two colleges A and B are twenty and ten respectively.

Figure 16

Colleges	S_h	$W_h=\frac{N_h}{N}$	$W_h S_h^2$	$(N_h-1)s_h^2$	$N_h(\bar{y}_h-\bar{Y})^2$
College A	100.5025	0.6667	67.0050	19999.9975	45000
College B	1616.1616	0.3333	538.66	159999.9984	22500
Total			**605.665**	**179999.9959**	**67500**

In the table presented here, the second column gives Sh values which are obtained as (sigma h into square root of Nh by Nh minus 1)
Next column gives the values of wh which is equal to Nh by N
In the next columns, (Nh minus 1) into Sh square and Nh into (y bar h minus Y bar whole square) values are computed.

Standard error of the estimate of the population mean is
Standard error of y bar st is equal to square root of variance of y bar st
Since the sample observations are drawn using proportional allocation, let us obtain variance of y bar st using proportional allocation technique which is given as follows:
Variance of y bar st under proportional allocation is equal to (summation, h runs from 1 to 5 Wh into Sh square by n) minus (summation, h runs from 1 to 5 Wh into Sh square by N)
Which is equal to (eighteen point one six six nine three)
Standard error of y bar st is equal to (square root of variance of y bar st) which is equal to (four point two six two six two)
Variance of y bar under SRSWOR is equal to (N minus n into S square by N into n)
Where S square is equal to (1 by N minus 1) into (summation Nh minus 1 into Sh square) plus (summation Nh into y bar h minus Y bar whole square)
Which is equal to (eight hundred and twenty seven point seven five nine)
By substituting the given values and the value of S square in the formula of variance of y bar WOR we get the variance as (twenty four point eight one zero zero three)
Gain in efficiency of proportional allocation with that of SRSWOR is
(Variance of y bar under WOR) minus (variance of y bar st under proportional allocation) divided by (variance of y bar st under proportional allocation) which is equal to (zero point three six six six) which is equal to (thirty six point six six percent).

Here's a summary of our learning in this session, where we understood:

- The stratified random sampling technique to estimate the parameters of the population with the standard error
- The methods of allocation of sample sizes
- The comparison of stratified sampling proportional and optimum allocation techniques with that of SRSWOR