Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 75(3) 2021 215–220

Rationalizing Sample-Size Allocation in Stratified Sampling

Arijit Chaudhuri[1] and Chandrima Chakraborty[2]

[1] *Indian Statistical Institute, Kolkata*
[2] *Swami Vivekananda University, Kolkata*

Received 09 September 2021; Accepted 07 March 2022

SUMMARY

Applying Chebyshev's Inequality one may rationally choose an appropriate sample-size in Simple Random Sampling Without Replacement (SRSWOR). This should give us a rule to hit upon the sizes n_h of SRSWOR's to be independently drawn from the respective h^{th} stratum $(h=1,2,3,...,H)$ in Stratified SRSWOR Sampling. This gives us the total sample-size $n=\sum_{h=1}^{H} n_h$ in this. The standard equal, proportional and Neyman's allocation rules for which arbitrarily chosen n is allocated to strata may be compared against the Chebyshev's rule above.

Keywords: Chebyshev's Inequality; Sample-size Allocation; Stratified Sampling.

1. INTRODUCTION

No rule is cited for specifying in traditionally standard stratum-wise sample-size allocation rules in Stratified Simple Random Sampling Without Replacement (SRSWOR) the total sample size n to be drawn from a finite population of size $N(>n)$. But recently, Chaudhuri and Dutta (2018) and Chaudhuri and Sen (2020) have presented specification rule based on Chebyshev's Inequality to prescribe size of an SRSWOR to be rationally chosen from a given finite population.

Our prescription here is to first fix the stratum-wise sizes of the SRSWOR's and take their aggregate across the strata as the total size of the sample to be chosen in a scientific manner.

In Section 2, we present the details. In Section 3, we compare the performance of this procedure vis-à-vis the equal, proportional and Neyman's optimal sample-size allocation rules yielding variances of the standard unbiased estimator of the population mean from a stratified SRSWOR sample. In Section 4, we comment on the relative performances and state our conclusions in Section 5.

2. STRATUM-WISE SPECIFICATION OF SAMPLE-SIZES OF SRSWOR'S FROM STRATA OF SIZES $N_h\,(h=1,2,3,\ldots,H)$ OF A POPULATION OF SIZE $N=\sum_{h=1}^{H} N_h$

Suppose we intend to independently choose SRSWOR's of suitable sizes, n_h from strata of sizes N_h with strata means $\overline{Y}_h=\frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$ so that the stratum-wise sample means $\overline{y}_h$ are so accurate that

$$Prob\left[\left|\overline{y}_h-\overline{Y}_h\right|\le f_h\,\overline{Y}_h\right]\ge 1-\alpha \tag{2.1}$$

for every $h(=1,2,3,\ldots,H)$ so that f_h is a fixed positive proper fraction $(0<f_h<1\,\forall\, h=1,2,3,\ldots,H)$ and α is a fixed small positive proper fraction $0<\alpha<1$.

Chebyshev's inequality in this situation gives us

$$Prob\left[\left|\overline{y}_h-\overline{Y}_h\right|\le \lambda\sqrt{V\left(\overline{y}_h\right)}\right]\ge 1-\frac{1}{\lambda^2} \tag{2.2}$$

Corresponding author: Chandrima Chakraborty
E-mail address: chandrimachakraborty1997@gmail.com

for a positive number λ so that $1-\frac{1}{\lambda^2}>0$ and the variance of $\overline{y_h}$ is

$$V\left(\overline{y_h}\right)=\left(\frac{N_h-n_h}{N_h n_h}\right)S_h^2 \tag{2.3}$$

writing $S_h^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_{hi}-\overline{Y_h}\right)^2$ and y_{hi} is

the value of a real variable y for the unit $i(\,i=1,2,3,\ldots,N_h)$ in the h^{th} stratum ($h=1,2,3,\ldots,H$). Also we should write

$$CV_h=100\frac{S_h}{\overline{Y_h}} \tag{2.4}$$

which is the co-efficient of variation of the values y_{hi} , $i=1,2,3,\ldots,N_h$ for each particular stratum with $h=1,2,3,\ldots,H$.

Using (2.1)-(2.4) and taking $\alpha=\frac{1}{\lambda^2}$ and

$$\lambda S_h\sqrt{\left(\frac{1}{n_h}-\frac{1}{N_h}\right)}=f_h\overline{Y_h}$$

it follows that we may take

$$n_h=\frac{N_h}{1+\alpha N_h\frac{f_h^2\left(100\right)^2}{CV_h^2}} \tag{2.5}$$

as the "Chebyshev inequality based" specification for an appropriate sample-size n_h to be drawn from respective strata of sizes N_h independently across strata $h=1,2,3,\ldots,H$. Also, let $n=\sum_{h=1}^{H}n_h$ which is thus prescribed as the total sample size to be taken from the finite population of size $N=\sum N_h$.

3. EQUAL, PROPORTIONAL AND NEYMAN'S OPTIMAL 'SAMPLE-SIZE ALLOCATION RULES'

From Cochran (1977) and Chaudhuri (2010) we know the following Allocation rules:

(i) Equal Allocation:

$$n_h=\frac{n}{H}\text{, for every } h=1,2,3,\ldots,H$$

(ii) Proportional Allocation:

$$n_h=n\frac{N_h}{N}\text{, for every } h=1,2,3,\ldots,H$$

(iii) Neyman's Optimal Allocation:

$$n_h=n\frac{N_hS_h}{\sum_{h=1}^{H}N_hS_h}\text{, for every } h=1,2,3,\ldots,H$$

A comparison of n_h by (2.5) versus n_h by (i), (ii) and (iii):

For (i) and (ii) we need little background materials. For (2.5), α, f_h and CV_h are to be appropriately specified though not much probe into basic raw data is needed. But since $CV_h=100\frac{S_h}{\overline{Y_h}}$, for intuitively taken CV_h values, we need $\overline{Y_h}$ and S_h values to apply rule (iii). Neyman's rule is the most complex.

For the population mean $\overline{Y}=\frac{1}{N}\sum_{h=1}^{H}N_h\overline{Y_h}$ the standard unbiased estimator $\overline{y_{st}}=\frac{1}{N}\sum_{h=1}^{H}N_h\overline{y_h}$ based on a stratified SRSWOR the variance is $V\left(\overline{y_{st}}\right)=\frac{1}{N^2}\sum_{h=1}^{H}N_h^2V\left(\overline{y_h}\right)$

$$=\frac{1}{N^2}\sum_{h=1}^{H}\frac{N_h^2S_h^2}{n_h}-\frac{1}{N^2}\sum_{h=1}^{H}N_hS_h^2 \tag{3.1}$$

For the n_h-values determined by (2.5), (i)-(iii) respectively we present below the tabulated values of (3.1) for the respective Allocation rules with arbitrary choices of the parameters involved.

Table 3.1. Details for Comparison of the 4 Allocation Rules

Stratum Number (h)	Stratum Size N_h	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	20	0.05	0.11	7	6	500.45	35.03	1227.10	9	6	3
2	22	0.05	0.12	8	6	250.27	20.02	400.80	9	7	2
3	30	0.05	0.13	11	10	1035.02	113.85	12961.82	9	9	15
4	40	0.05	0.10	10	13	829.82	82.98	6885.68	9	13	15

Table 3.1A. Based on the results in this table, we find the following:

$V\left(\overline{y_{st}}\right)$ by Chebyshev's Allocation	$V\left(\overline{y_{st}}\right)$ by Equal Allocation	$V\left(\overline{y_{st}}\right)$ by Proportional Allocation	$V\left(\overline{y_{st}}\right)$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
114.04	151.37	124.01	85.71

Table 3.2 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	20	0.05	0.03	2	6	500.45	10.01	100.20	9	6	4
2	22	0.05	0.03	2	6	250.27	5.01	25.10	9	7	2
3	30	0.05	0.025	2	9	829.82	16.60	275.56	9	9	10
4	40	0.05	0.02	2	13	1035.02	20.70	428.49	9	12	17

Table 3.2A Showing the variances:

$V\left(\overline{y_{st}}\right)$ by Chebyshev's Allocation	$V\left(\overline{y_{st}}\right)$ by Equal Allocation	$V\left(\overline{y_{st}}\right)$ by Proportional Allocation	$V\left(\overline{y_{st}}\right)$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
4.87	6.51	5.19	4.25

Table 3.3 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	24	0.05	0.11	8.50	8	1510.11	128.36	16476.29	9	7	5
2	20	0.05	0.12	7.65	6	1103.47	84.42	7126.74	9	6	3
3	16	0.05	0.13	6.80	4	942.31	64.08	4106.25	9	5	2
4	60	0.05	0.09	10.75	19	2400.23	258.02	66574.32	9	19	27

Table 3.3A Showing the variances:

$V\left(\overline{y_{st}}\right)$ by Chebyshev's Allocation	$V\left(\overline{y_{st}}\right)$ by Equal Allocation	$V\left(\overline{y_{st}}\right)$ by Proportional Allocation	$V\left(\overline{y_{st}}\right)$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
690.29	1633.31	698.41	531.42

Table 3.4 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	10	0.05	0.10	5	3	198.56	9.93	98.60	8	3	2
2	30	0.05	0.08	7	10	231.33	16.19	262.12	8	9	10
3	25	0.05	0.09	6	7	187.42	11.25	126.56	8	8	6
4	35	0.05	0.10	9	11	201.11	18.10	327.61	8	11	13

Table 3.4A Showing the variances:

$V(\overline{y_{st}})$ by Chebyshev's Allocation	$V(\overline{y_{st}})$ by Equal Allocation	$V(\overline{y_{st}})$ by Proportional Allocation	$V(\overline{y_{st}})$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
5.11	6.72	5.23	4.9

Table 3.5 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	15	0.05	0.16	8	4	260.3	20.82	433.47	7	5	2
2	23	0.05	0.13	10	8	470.8	47.08	2216.53	7	7	8
3	33	0.05	0.11	9	10	750.5	67.55	4563	7	10	16
4	19	0.05	0.10	6	5	200.2	12.01	144.24	7	6	2

Table 3.5A Showing the variances:

$V(\overline{y_{st}})$ by Chebyshev's Allocation	$V(\overline{y_{st}})$ by Equal Allocation	$V(\overline{y_{st}})$ by Proportional Allocation	$V(\overline{y_{st}})$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
57.72	84.94	59.49	39.65

Table 3.6 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	12	0.05	0.11	5	3	400.01	20	400	6	3	2
2	15	0.05	0.15	8	4	680.20	54.42	2961.54	6	4	6
3	25	0.05	0.16	11	7	251.34	27.65	764.52	6	7	5
4	30	0.05	0.18	15	9	300.56	45.08	2032.21	6	8	10

Table 3.6A Showing the variances:

$V(\overline{y_{st}})$ by Chebyshev's Allocation	$V(\overline{y_{st}})$ by Equal Allocation	$V(\overline{y_{st}})$ by Proportional Allocation	$V(\overline{y_{st}})$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
48.78	55.89	52.55	42.98

Table 3.7 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	11	0.05	0.20	10	3	1500.01	150	22500	7	3	2
2	17	0.05	0.19	11	5	1723	189.53	35921.62	7	5	3
3	26	0.05	0.16	12	8	2600.34	312.04	97368.96	7	8	8
4	36	0.05	0.14	13	12	3600	468	219024	7	11	16

Table 3.7A Showing the variances:

$V(\overline{y_{st}})$ by Chebyshev's Allocation	$V(\overline{y_{st}})$ by Equal Allocation	$V(\overline{y_{st}})$ by Proportional Allocation	$V(\overline{y_{st}})$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
2912.52	5006.31	3178	2409.34

Table 3.8 Continuation of Table 3.1 with revised materials:

Stratum Number (n)	Stratum Size (N_h)	α	f_h	CV_h	n_h by Chebyshev's rule	$\overline{Y_h}$	S_h	S_h^2	n_h by Equal rule	n_h by Proportional rule	n_h by Neyman's Optimal rule
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	10	0.05	0.10	5	3	60	3	9	6	3	2
2	17	0.05	0.12	8	6	77	6.16	37.95	6	5	6
3	23	0.05	0.11	7	6	67	4.69	22	6	7	6
4	25	0.05	0.09	6	7	85	5.10	26.01	6	7	8

Table 3.8A Showing the variances:

$V(\overline{y_{st}})$ by Chebyshev's Allocation	$V(\overline{y_{st}})$ by Equal Allocation	$V(\overline{y_{st}})$ by Proportional Allocation	$V(\overline{y_{st}})$ by Neyman's Optimal Allocation
(1)	(2)	(3)	(4)
0.80	0.84	0.83	0.77

4. COMPARISON OF VARIANCES OF ESTIMATED POPULATION MEANS BY THE 4 ALLOCATION RULES:

From the Tables 3.1A through Tables 3.8A we clearly see that $V(\overline{y_{st}})$ by Neyman's Optimal Allocation $< V(\overline{y_{st}})$ by Chebyshev's Allocation $< V(\overline{y_{st}})$ by Proportional Allocation $< V(\overline{y_{st}})$ by Equal Allocation in each of the 8 cases illustrated.

It matches intuition because equal allocation uses no facts, or intuition but is quite casual, proportional allocation is intuitive but no effort is made to use any additional data, in Chebyshev's allocation only arbitrary magnitudes of co-efficient of variation CV_h are used, while in Neyman's allocation strata variances are utilized on deriving them from arbitrary CV_h's and strata means $\overline{Y_h}$'s also arbitrarily assigned.

But in each case the total sample-size n is used which is technically derived only by applying Chebyshev's allocation rule with a definite criterion for controlling estimation error in estimating each stratum mean in a conscious way specifying f_h values and an α.

5. CONCLUDING REMARKS

(i) In stratified sampling how to fix the total size of a sample to be drawn from a finite population is nowhere in the literature stated by any expert.

(ii) In drawing an SRSWOR from a population it is rational to demand the absolute magnitude of error in estimating the population mean by the sample mean not to exceed a prescribed positive fraction of the population mean with a high probability $\geq(1-\alpha)$.

The well-known Chebyshev's inequality may then be utilized to achieve this and this yields a rule to prescribe the sample-size tabulated in terms of population size N, f, α and co-efficient of variation (CV) Chaudhuri and Sen (2020) have given the resulting rule and also Chaudhuri (2010) and Chaudhuri and Dutta (2018).

This may be employed as Chebyshev's rule to assign the size of every SRSWOR to be drawn from respective strata and independently so giving the total sample-size to be prescribed to draw from the population.

(iii) The same total sample-size may be then rationally prescribed for the stratified SRSWOR and then size-allocation for the various strata for respective allocation rules in the literature.

(iv) The variance of the usual unbiased estimator for the population mean should of course be the least for the Neyman's optimal allocation rule. The Chebyshev's rule helps only to rationally choose the total sample-size needed to implement the well-known allocation rules in the literature.

REFERENCES

Chaudhuri, Arijit (2010). Essentials of Survey Sampling. Prentice Hall of India, Delhi, India.

Chaudhuri, Arijit and Dutta, Tathagata (2018). Determining the size of a sample to take from a Finite Population, Statistics and Applications, (ISSN 2452-7395, online), 16(1)New Series, 37-44.

Chaudhuri, Arijit and Sen, Aritra (2020). Fixing the Sample-Size in Direct and Randomized Response Surveys. JISAS, **74(3)**, 201-208.

Cochran, W.G. (1977). Sampling Techniques. John Wiley and Sons, N.Y. USA.