Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 74(3) 2020 201–208

Fixing the Sample-Size in Direct and Randomized Response Surveys

Arijit Chaudhuri[1] and Aritra Sen[2]

[1] *Indian Statistical Institute, Kolkata*
[2] *International Institute for Population Sciences, Mumbai*

Received 13 May 2020; Revised 05 September 2020; Accepted 04 October 2020

SUMMARY

In order to estimate a population total or mean by an unbiased estimator from a Simple Random Sampling With Replacement or a Simple Random Sampling Without Replacement incurring an estimation error not exceeding a pre-assigned fraction of the estimand parameter with a high probability, we apply Chebyshev's inequality to see a reasonable solution, to fix the sample size, available for a Direct Response (DR) survey but extending this to Randomized Response (RR) survey to cover a sensitive feature we find an absurd solution. The same exercise is extended by applying Chebyshev's inequality to permit unequal probability sampling with suitably different estimators.

Keywords: Chebyshev's inequality; Finite population; Randomized response surveys; Sample-size; Sensitive issues.

1. INTRODUCTION

To fix the number of units to choose in a sample from a finite survey population is a classical problem. Distinguished sampling experts like Cochran (1953, 1963, 1977) and several others have prescribed requisite solutions chiefly demanding `Normality' in the distribution of the standardized difference between the estimate and the estimand parameter it seeks to estimate.

Chaudhuri (2010, 2014, 2018, 2020) resorted to taking advantage of the Chebyshev's inequality in tackling this issue. Chaudhuri & Dutta (2018, 2019) have pursued further with two distinct approaches (i) when an explicit formula for the variance of an estimator for a finite population total or mean is classically available or (ii) when a variance formula is hard to come by but an unbiased estimator for the variance is explicitly at hand.

But an insurmountable situation emerges in case one intends to extend to an investigation covering Randomized Response (RR) techniques (RRT) encountering stigmatizing characteristics of interest. Details follow in the next two sections.

2. THEORY FOR DIRECT RESPONSE (DR) SURVEYS

Suppose $U = (1, \cdots, i, \cdots, N)$ denotes a finite population of a known number of N units with $\underline{Y} = (y_1, \cdots, y_i, \cdots, y_N)$ a vector of unknown values $y_i, i \in U$ for a real variable y with a total $Y = \sum_1^N y_i$ and mean $\bar{Y} = \frac{Y}{N}$.

Let s be a sample of $n (2 \le n < N)$ units to be chosen from U with the selection-probability $p(s)$ according to a design p.

Let $t = t(s, \underline{Y})$ be employed as a statistic to unbiasedly estimate Y having a variance $V(t)$. Then, Chebyshev's inequality states that

$$\text{Prob}\left[|t - Y| \le \lambda\sqrt{V(t)}\right] \ge 1 - \frac{1}{\lambda^2}; \tag{2.1}$$

here $\lambda > 1$ is a positive constant.

Suppose, we demand, taking $f (0 < f < 1)$ and $\alpha (0 < \alpha < 1)$ as two suitably chosen constants so that

Corresponding author: Aritra Sen
E-mail address: aritrasen679@gmail.com

$$\text{Prob}[|t - Y| \leq fY] \geq 1 - \alpha \tag{2.2}$$

Combining (2.1) with (2.2) let us take

$$\alpha = \frac{1}{\lambda^2} \text{ and } fY = \lambda\sqrt{V(t)} \tag{2.3}$$

If a suitable formula for $V(t)$ involving N and n be at hand writing $\text{CV} = 100\frac{\sqrt{V(t)}}{Y}$ for the Coefficient of Variation (CV) of t, then, noting

$$\alpha = \frac{V(t)}{f^2 Y^2} = \frac{\text{CV}^2}{100^2 f^2} \tag{2.4}$$

suitable sample-size n may be tabulated against given values of α, f, CV and N. This is discussed explicitly by Chaudhuri (2010, 2014, 2018) illustrating the cases of Simple Random Sampling with Replacement(SRSWR) and Simple Random Sampling without Replacement (SRSWOR) in details.

For more general sampling schemes and designs with suitable estimators since $V(t)$ may not be easily expressible Chaudhuri & Dutta (2018), instead recommend postulating a simple model

$$y_i = \beta x_i + \epsilon_i, i \in U \tag{2.5}$$

with x_i known for $i \in U$ and β as an unknown constant and ϵ_i's as independent random variables with mean $E_m(\epsilon_i) = 0 \,\forall\, i \in U$ and variances $V_m(\epsilon_i) = \sigma^2 x_i^g$ with an unknown $\sigma > 0$ and g $(0 \leq g \leq 2)$. Then, they recommend replacing α in (2.4) by

$$\alpha_m = \frac{E_m V(t)}{f^2 E_m(Y^2)} \tag{2.6}$$

Since α given in (2.4) is difficult, rather impossible to work out when general sampling strategies are employed, Chaudhuri & Dutta (2018) approximated α by α_m in (2.6), though almost nothing may be said about how close is α to α_m or how far they are away from each other. But such an approximation is rather often resorted to in practical sampling exercises. We are encouraged to try this on perusing the above publication by them in 2018.

Chaudhuri & Dutta (2018) have elaborated on illustrating well-known PPSWR & Hansen-Hurwitz (1943) estimator t_{HH}, IPPS & Horvitz-Thompson (1952) estimator t_{HT} and Rao, Hartley & Cochran strategy (1962); here PPSWR means probability proportional to size with replacements, IPPS means sampling with inclusion-probability proportional to size and strategy means a combination of a sampling design and an estimator based on a sample chosen according to that design. They showed the sampling fractions $\frac{n}{N}$ coming out quite reasonable in each case with numerical illustrations.

Chaudhuri (2018A) observed the extension of these to cover sensitive issues applying Randomized Response (RR) techniques when arose baffling problems as is reported in Section 3 in details.

3. SAMPLE-SIZE FIXATION IN RR SURVEYS

Let us, for simplicity, illustrate Warner's (1965) RR technique for a sensitive 'qualitative' characteristic A and the well-known 'quantitative' characteristic respectively quoting from Chaudhuri (2011) and Chaudhuri & Christofides (2013) below, starting with

$$y_i = \begin{cases} 1, \textit{if i bears A}, & \textit{a sensitive attribute} \\ 0, \textit{if i bears } A^C, & \textit{the compliment of A} \end{cases}$$

Let a person i sampled be approached with a box of cards in proportions $p : (1-p)$ with $p \neq 1/2$ marked A and A^C with a request to truthfully respond.

$$I_i = \begin{cases} 1, \textit{if a randomly chosen card from the box matches his or her true feature A or } A^C \\ 0, \textit{if it does not match} \end{cases}$$

Its expectation and variance respectively are

$$E_R(I_i) = p y_i + (1-p)(1-y_i)$$

$$= (1-p) + (2p-1)y_i$$

implying $r_i = \dfrac{I_i - (1-p)}{(2p-1)}$ having $E_R(r_i) = y_i$,

$V_R(I_i) = p(1-p)$ since $I_i^2 = I_i, y_i^2 = y_i$

$$V_R(r_i) = \frac{p(1-p)}{(2p-1)^2} = V_i, \text{ say } \forall\, i.$$

To consider the case of a sensitive quantitative characteristic, a sampled person i is approached with 2 boxes respectively containing cards marked with real numbers $a_1, \cdots, a_L$ and $b_1, \cdots, b_K$

with $\mu = \dfrac{1}{L}\sum_{j=1}^{L} a_j \neq 0, \sigma^2 = \dfrac{1}{L-1}\sum_{1}^{L}(a_j - \mu)^2$

and $\nu = \dfrac{1}{K}\sum_{k=1}^{K} b_k$ and $\psi^2 = \dfrac{1}{K-1}\sum_{1}^{K}(b_k - \nu)^2$

with a request to take independently and randomly from both the boxes just one card each say marked a_j and b_k and report the randomized response

$$I_i = a_j y_i + b_k$$

Consequently, the yielded mean and variance are

$$E_R(I_i) = \mu y_i + \nu$$

$$implying\ r_i = \frac{I_i - \nu}{\mu}$$

$$having\ E_R(r_i) = y_i$$

$$and\ V_R(r_i) = y_i^2 \frac{\sigma^2}{\mu^2} + \frac{\psi^2}{\mu^2} = V_i, \text{ say,}$$

$$\text{with } \frac{r_i^2 \frac{\sigma^2}{\mu^2} + \frac{\psi^2}{\mu^2}}{1 + \frac{\sigma^2}{\mu^2}} = \frac{r_i^2 \sigma^2 + \psi^2}{\sigma^2 + \mu^2}$$

as an unbiased estimator for V_i above. Later we shall write

$$C = \frac{\sigma^2}{\mu^2}, D = \frac{\psi^2}{\mu^2}.$$

In these respective qualitative and quantitative cases with similar notations unbiased estimators for $Y = \sum_1^N y_i$ from RR's based on samples turn out as

$$e = t \Big|_{\substack{y_i = r_i \\ i \in s}} \text{ with } E(e) = E_p E_R(e) = E_p(t) = Y$$

and variance as

$$V(e) = E_p V_R(e) + V_p E_R(e)$$
$$= E_p V_R(e) + V_p(t);$$

here E_p, V_p denote design-based expectation and variance, E_R, V_R the RR-based counter parts and E, V, the overall expectation and variance, generically, with usual linear estimator t. The condition $E_R(e)$ customarily equals t and hence e may be treated as unbiased for Y in the sense $E(e) = E_p E_R(e) = E_R E_p(e) = Y$. So, in order to appropriately fix the sample-size, parallelly with direct response surveys, in the case of RR surveys for both qualitative and quantitative the same approach utilizing Chebyshev's inequality should apply with no palpable difficulty at all.

Taking various alternative strategies (p, t) and (p, e) and writing

$$V(e) = E_p V_R(e) + V_p(t) = I + II, \text{ say,}$$

we shall soon illustrate our findings that numerically I is much larger than II resulting in absurd choices for n using Chebyshev's inequality treating e and $V(e) = I + II$ compared to genial findings with Chebyshev's inequality to find n treating t and $V(t) = II$. The illustrations follow:

To choose n using Chebyshev's inequality in tackling stigmatizing features by RR surveys also the model (2.5) will be honored requiring, instead of α_m in (2.6) its version

$$\alpha_{RRm} = \frac{E_m V(e)}{f^2 E_m(Y^2)} \tag{3.1}$$

and also,

$$\alpha_R = \frac{EV(e)}{f^2(Y^2)} \tag{3.2}$$

and instead of $I = E_p V_R(e)$ and $II = V_p(t)$ their model expectations respectively

$$I' = E_m(I)\ \ and\ \ II' = E_m(II).$$

Here, let us illustrate the case of SRSWOR with $t = N(\bar{y})$, that is, the expansion estimator for Y in case of a DR and $e = N(\bar{r})$, with $\bar{r} = \frac{1}{n}\sum_{i \in s} r_i$ where $r_i = \frac{I_i - (1-p)}{(2p-1)}$, in case Warner's (1965) Randomized Response Technique (RRT) is employed. Then,

$$V(t) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2;$$

$$where\ S^2 = \frac{1}{N-1}\sum_1^N \left(y_i - \bar{Y}\right)^2$$

$$\text{and, } V(e) = V(t) + \frac{N^2}{n}\frac{p(1-p)}{(2p-1)^2}$$

Then (3.1) and (3.2) respectively yield, on writing $CV = 100\frac{S}{\bar{Y}}$, with $\bar{Y} = \frac{Y}{N}$,

$$n = \frac{N}{1 + N\alpha f^2 (\frac{100}{CV})^2} \tag{3.3}$$

and,

$$n = \frac{\frac{N(CV)^2}{f^2(100)^2} + \frac{Np(1-p)}{(2p-1)^2 f^2 (\bar{Y})^2}}{N\alpha + \frac{(CV)^2}{f^2(100)^2}} \tag{3.4}$$

So, let us tabulate as below, writing $n(DR)$ for 'n' based on Direct Response and $n(RR)$ as 'n' based on Randomized Response.

Table (Preliminary)

N	f	α	CV	p	$\overline{Y}$	$n(DR)$	$n(RR)$
80	0.1	0.05	5%	-	-	8	-
100	0.1	0.05	10%	-	-	17	-
80	0.1	0.05	5%	0.45	0.25	-	745416
100	0.1	0.05	10&	0.45	0.25	-	660017

N.B.: p and $\overline{Y}$ values are needed for RR only.

Comment: The findings $n(DR)$ seem reasonable while $n(RR)$ look preposterous.

Next, we use our modelled x, y values for all illustrated sampling schemes and estimators. Considering specific cases let us see the following findings for arbitrarily chosen $N,\ x_i,\ i \in U$.

3.1 SRSWR

DR considering $E_m(II)$ we note

Table 3.1.1

N	f	α	σ^2	g	β	n
10	0.2	0.05	1	1.5	10	5
12	0.3	0.05	1	2	10	9
15	0.3	0.05	1	1.5	10	8
20	0.3	0.05	1	1.5	10	13

N.B. Fraction n/N seems reasonable.

Considering $[E_m(I) + E_m(II)]/f^2 E_m(Y^2)$ in RR:

RR: Qualitative case: $y_i = 1$ or 0

Table 3.1.2

N	α	f	β	σ^2	g	p	n
10	.05	0.1	5	1	1.5	0.45	885
12	.05	0.1	5	1	1.5	0.52	1676
15	.05	0.1	5	1	1.5	0.49	4754
20	.05	0.1	5	1	1.5	0.56	1071

RR: Quantitative case: a_j, b_k arbitrary.

Table 3.1.3

N	α	f	C	D	β	σ^2	g	n
10	0.05	0.1	0.317	0.18	5	1	1.5	22900
12	0.05	0.1	0.317	0.18	5	1	1.5	24453
15	0.05	0.1	0.317	0.18	5	1	1.5	30844
20	0.05	0.1	0.317	0.18	5	1	1.5	43157

N.B. In both qualitative & quantitative cases n comes out absurd.

While constructing Tables 1–3, we noted that $E_m(I)$ far exceeds $E_m(II)$ in magnitude. This led us to treat $E_m(I) + E_m(II)$ as close to $E_m(I)$, with $E_m(II)$ viewed as negligible vis-a-vis $E_m(I)$. So, we considered $E_m(I)/f^2 E_m(Y^2)$ for both the qualitative and the quantitative cases as above separately to venture to present respectively the Tables $3.1.2'$ – $3.1.3'$ below.

Table $3.1.2'$

(RR Surveys for Qualitative case with data mostly as in Table 3.1.2)

N	α	f	β	σ^2	g	p	n
10	0.05	0.1	5	1	1.5	0.45	124
12	0.05	0.1	5	1	1.5	0.52	867
15	0.05	0.1	5	1	1.5	0.49	3905
20	0.05	0.1	5	1	1.5	0.56	101

Table $3.1.3'$

(RR Surveys for Quantitative case with data mostly as in Table 3.1.3)

N	α	f	C	D	β	σ^2	g	n
10	0.05	0.1	0.317	0.18	5	1	1.5	18127
12	0.05	0.1	0.317	0.18	5	1	1.5	22101
15	0.05	0.1	0.317	0.18	5	1	1.5	27434
20	0.05	0.1	0.317	0.18	5	1	1.5	36433

N.B. From both Table $3.1.2'$ and Table $3.1.3'$ the value of emerging n is absurd.

When other sampling designs were considered, similar results were seen as well.

3.2 PPSWR due to Hansen-Hurwitz

DR considering $E_m(II)$:

Table 3.2.1

N	f	α	σ^2	g	β	n
10	0.2	0.05	1	1.5	5	3
12	0.2	0.05	1	2	10	5
15	0.1	0.05	1	1.5	10	3
20	0.3	0.05	1	2	5	8

N.B. Fraction n/N seems reasonable.

Considering $[E_m(I) + E_m(II)]/f^2 E_m(Y^2)$

RR: Qualitative case: $y_i = 1$ or 0

Table 3.2.2

N	α	f	β	σ^2	g	p	n
10	.05	0.1	5	1	1.5	0.45	23219
12	.05	0.1	5	1	1.5	0.52	211702
15	.05	0.1	5	1	1.5	0.49	149969
20	.05	0.1	5	1	1.5	0.56	80392

RR: Quantitative case: a_j, b_k arbitrary

Table 3.2.3

N	α	f	C	D	β	σ^2	g	n
10	0.05	0.1	0.317	0.18	5	1	1.5	17777882
12	0.05	0.1	0.317	0.18	5	1	1.5	3102865
15	0.05	0.1	0.317	0.18	5	1	1.5	6073874
20	0.05	0.1	0.317	0.18	5	1	1.5	14063416

N.B. In both qualitative & quantitative cases n comes out absurd.

Now considering $E_m(I)/f^2E_m(Y^2)$ for both the qualitative and the quantitative cases as above separately we venture to present respectively the Tables 3.2.2$'$ – 3.2.3$'$ below.

Table 3.2.2$'$

(RR Surveys for Qualitative case with data mostly as in Table 3.2.2)

N	α	f	β	σ^2	g	p	n
10	0.05	0.1	5	1	1.5	0.45	32743
12	0.05	0.1	5	1	1.5	0.52	55645
15	0.05	0.1	5	1	1.5	0.49	97152
20	0.05	0.1	5	1	1.5	0.56	90794

Table 3.2.3$'$

(RR Surveys for Quantitative case with data mostly as in Table 3.2.3)

N	α	f	C	D	β	σ^2	g	n
10	0.05	0.1	0.317	0.18	5	1	1.5	1754129
12	0.05	0.1	0.317	0.18	5	1	1.5	3046945
15	0.05	0.1	0.317	0.18	5	1	1.5	5962327
20	0.05	0.1	0.317	0.18	5	1	1.5	13800438

N.B. Similar absurd values of 'n' can be seen here as well.

3.3 Hartley-Rao (1962) sampling scheme and Horvitz-Thompson Estimator:

DR considering $E_m(II)$:

Table 3.3.1

N	f	α	σ^2	g	β	n
10	0.1	0.05	1	1.5	10	3
12	0.2	0.05	1	2	10	5
15	0.1	0.05	1	2	10	8
20	0.1	0.05	1	1.5	10	7

N.B. Fraction n/N seems reasonable.

Considering $[E_m(I) + E_m(II)]/f^2E_m(Y^2)$ RR: Qualitative case: $y_i = 1$ or 0

Table 3.3.2

N	α	f	β	σ^2	g	p	n
10	.05	0.1	5	1	1.5	0.45	53
12	.05	0.1	5	1	1.5	0.52	2117
15	.05	0.1	5	1	1.5	0.49	2163
20	.05	0.1	5	1	1.5	0.56	82

RR: Quantitative case: a_j, b_k arbitrary

Table 3.3.3

N	α	f	C	D	β	σ^2	g	n
10	**0.05**	**0.1**	**0.317**	**0.18**	**5**	**1**	**1.5**	**492**
12	0.05	0.1	0.317	0.18	5	1	1.5	310
15	0.05	0.1	0.317	0.18	5	1	1.5	745
20	0.05	0.1	0.317	0.18	5	1	1.5	982

N.B. In both qualitative & quantitative cases n comes out absurd.

Now considering $E_m(I)/f^2E_m(Y^2)$ for both the qualitative and the quantitative cases as above separately weventure to present respectively the Tables 3.3.2$'$ – 3.3.3$'$ below.

Table 3.3.2$'$

(RR Surveys for Qualitative case with data mostly as in Table 3.3.2)

N	α	f	β	σ^2	g	p	n
10	0.05	0.1	5	1	1.5	0.45	20053
12	0.05	0.1	5	1	1.5	0.52	53645
15	0.05	0.1	5	1	1.5	0.49	1520466
20	0.05	0.1	5	1	1.5	0.56	62984

Table 3.3.3′

RR Surveys for Quantitative case with data mostly as in Table 3.3.3

N	α	f	C	D	β	σ^2	g	n
10	0.05	0.1	0.317	0.18	5	1	1.5	61725
12	0.05	0.1	0.317	0.18	5	1	1.5	30469
15	0.05	0.1	0.317	0.18	5	1	1.5	146616
20	0.05	0.1	0.317	0.18	5	1	1.5	250917

N.B. Similar absurd values of 'n' can be seen here as well.

3.4 Rao-Hartley-Cochran Sampling Scheme:

(Ideally n= [N/Group size in RHC method of sampling]+1)

DR considering $E_m(II)$:

Table 3.4.1

N	f	α	σ^2	g	β	n
10	0.2	0.05	1	1.5	5	3
12	0.1	0.05	1	1.5	5	5
15	0.1	0.05	1	2	5	7
20	0.1	0.05	1	1.5	5	6

N.B. Fraction n/N seems reasonable.

Considering $[E_m(I) + E_m(II)]/f^2 E_m(Y^2)$

Calculating the sample size 'n' in this case involves extensive algebra and cumbersome procedures. Sample size derived from DR technique will therefore be used to check the validity of the assumed values of α when RRT is incorporated with DR technique.

N.B. $\hat{\alpha}$ is the observed value of α, defined by:

$$\hat{\alpha}_{RRm} = \frac{E_m V(e)}{f^2 E_m(Y^2)}$$

RR: Qualitative case: $y_i = 1$ or 0

Table 3.4.2

N	n	f	β	σ^2	g	p	$\hat{\alpha}$
10	3	0.2	5	1	1.5	0.45	0.3
12	5	0.1	5	1	1.5	0.52	5.9
15	7	0.1	5	1	1.5	0.49	17.75
20	6	0.1	5	1	1.5	0.56	0.6

RR: Quantitative case: a_j, b_K arbitrary

Table 3.4.3

N	n	f	C	D	β	σ^2	g	$\hat{\alpha}$
10	3	0.1	0.317	0.18	5	1	1.5	4.92
12	5	0.1	0.317	0.18	5	1	1.5	0.3
15	7	0.1	0.317	0.18	5	1	1.5	1.7
20	6	0.1	0.317	0.18	5	1	1.5	98.2

N.B. In both qualitative and quantitative cases the observed value of '$\hat{\alpha}$' shows huge contradiction to the assume value 'α'=0.05.

This leads us to conclude that (1) the sample-size, and may be the sampling design, should be specified by the consideration of the efficacy of the DR counter part of the estimator contemplated to be evaluated from the corresponding RR data and (2) then examine the estimated coefficient of variation of the RR-based estimator and reach conclusions about the RR-data utilization from the resulting findings.

In the following Section 4 we illustrate what may happen if we venture to adopt alternative sampling and estimation procedures.

4. FIXING SAMPLE-SIZE IN VARYING PROBABILITY SAMPLING FOR DR AND RR SURVEYS

To cover respectively the (i) qualitative and the (ii) quantitative cases we consider the following undernoted data sets each with 20 population units, with y and x arbitrarily taken; in the illustrative tables the 1^{st} 10, the first 12, the first 15 and all the 20 are respectively used to illustrate cases 10, 12, 15 and 20 as the population-sizes. $\text{CV(RR)} = 100\frac{\sqrt{E[V(e)]}}{e}$ and $\text{CV(DR)} = 100\frac{\sqrt{E[V(t)]}}{t}$.

Data Sets

A. Qualitative Case: $N = 20$

i	1	2	3	4	5	6	7	8	9	10
y_i	1	0	0	1	1	0	1	0	1	1
x_i	1	3	7	4	5	6	3	8	2	1

i	11	12	13	14	15	16	17	18	19	20
y_i	0	1	1	1	0	1	0	0	1	1
x_i	4	1	2	2	1	7	3	9	4	2

B. Quantitative Case: $N = 20$

i	1	2	3	4	5	6	7	8	9	10
y_i	260	310	370	350	490	445	390	500	490	414
x_i	25	29	35	32	45	42	37	48	46	44

i	11	12	13	14	15	16	17	18	19	20
y_i	368	468	295	349	302	480	442	340	265	480
x_i	45	37	49	30	36	22	40	55	23	47

The a_j and b_K values are not reported and the above y_i-values are used in the earlier Tables displayed.

4.1 PPSWR Sampling, Hansen-Hurwitz estimator

Table A. The sample-size by (2.6) and (3.5) and the estimated coefficients of variation (CV)

N	n	CV	CV (RR) Qualitative		CV (RR) Quantitative	
		(DR)	I	I+II	I	I+II
10	3	4.07	91.4	424.3	33.1	33.3
12	5	2.00	191.7	288.8	25.2	25.8
15	3	3.10	551.4	651.5	32.5	32.7
20	8	6.01	49.6	625.2	20.2	21.1

Comments: Sample-size determined for the DR surveys may be rightly used for the quantitative RR in the case PPSWR Hansen-Hurwitz strategy without much concern but not so for the qualitative case.

4.2 Hartley-Rao (1962) sampling scheme and Horvitz-Thompson Estimator

Table B. The sample-size, Estimated CV's

N	n	CV	CV (RR) Qualitative		CV (RR) Quantitative	
		(DR)	I	I+II	I	I+II
10	3	20.2	91.4	206.4	32.5	38.3
12	6	17.9	138.7	193.8	23.0	30.1
15	8	16.66	270.3	2004.2	20.0	26.0
20	7	12.02	43.0	124.9	21.6	24.7

Comments: The sample-size determined for the DR surveys may be rightly used for the quantitative RR case for this Horvitz-Thompson strategy without much concern but not so for the qualitative case.

4.3 Rao-Hartley-Cochran strategy

Table C. The sample-size, Estimated CV's

N	n	CV	CV (RR) Qualitative		CV (RR) Quantitative	
		(DR)	I	I+II	I	I+II
10	3	3.6	52.8	54.3	18.6	20.0
12	5	1.6	35.0	35.1	16.3	16.4
15	7	1.4	225.1	225.3	15.0	15.02
20	6	6.0	31.3	32.4	12.7	14.1

Comments: For the sample-size found from the DR survey works rather well for the quantitative RR case but only much worse for the qualitative case.

5. CONCLUSION

Chebyshev's inequality helps the fixing of a required high probability that an estimator may not deviate from the estimand parameter by a specified fraction of the latter. This appropriately with postulated modelling facilitates setting sample-sizes in Direct Response sample surveys. But extending this to cover stigmatizing characteristics by Randomized Response surveys places us in hazardous anomalies. The sample-size comes out to be absurd in RR cases. Thus, it is resolved that the sample-size should be fixed from the DR survey data and a conclusion should be reached from there while extending for the RR cases.

ACKNOWLEDGMENT

The authors are grateful to an anonymous referee whose recommendations have led to this improved version of our original submission.

REFERENCES

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

Chaudhuri, A. (2014). Modern Survey Sampling, Chapman & Hall, CRC, Taylor & Francis, Boca Raton, Florida, USA.

Chaudhuri, A. (2011). Randomized response and indirect questioning techniques in surveys. Chapman & Hall, CRC Press, Taylor & Francis Group, Boca Raton, FL. USA.

Chaudhuri, A. (2018). Survey Sampling. Taylor & Francis, CRC, Boca Raton, Florida, USA.

Chaudhuri, A. (2020). A Review on Issues of Settling the sample-size in surveys: Two approaches: Equal and varying probability sampling - Crises in sensitive Issues. Cal. Stat. Assoc. Bull. SAGE. https://doi.org/10.1177/000806832095238

Chaudhuri, A. and Christofides, T.C. (2013). Indirect questioning in Sample Surveys, Springer-Verlag, Berlin

Chaudhuri, A. and Dutta, T. (2018). Determining the size of a sample to take from a finite population. Statistics and Applications, (ISSN 2452-7395, on line), **16(1)** New Series, 37-44.

Chaudhuri, A. and Dutta, T. (2019). Ibid (ISSN 2452- 7395 (on line)), **17(1)**, New Series, 63-72.

Cochran, W.G. (1953, 1963, 1977). Sampling Techniques. 1st, 2nd, 3rd ED., John Wiley & Sons, N.Y. USA.

Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. Ann. Math. Stat., **14**, 333-362.

Hartley, H.O. and Ross, A. (1954). Unbiased Ratio Estimators, Nature, **174**, 270-271.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J.A.S.A.,**77**, 89-96.

Rao, J.N.K. Rao, Hartley, H.O. and Cochran, W.G. (1962). On a Simple Procedure of Unequal Probability Sampling Without Replacement. J.R.S.S., B, **24**, 482-491.

Warner, S.L. (1965). RR: A Survey Technique for Eliminating Evasive Answer bias. J.A.S.A., **60**, 63-69.