Sample Survey Practical
Stat-418

Practical #1

1. **The following data gives the number of households, number of women in reproductive age and number of children under age 5 ,in 10 villages in a country.**

Village	**Total number of households.**	**Total number of women at reproductive age(M_i)**	**Total number of children under age 5 (y_i)**
1	428	401	420
2	503	625	786
3	117	173	177
4	63	135	95
5	701	2704	3386
6	112	177	173
7	213	344	376
8	182	201	183
9	467	601	685
10	215	286	276

Select a simple random sample of 3 villages with equal probability without replacement and

- **i)** **Find an unbiased estimate of total number of children under age 5 in these 10 villages.**
- **ii)** **Estimate the child women ratio and its corresponding sampling variance.**
- **iii)** **Estimate the total number of children under age 5 using the estimated child women ratio.**
- **iv)** **Find the standard error of your estimate and compare the prcission of these methods of estimated total number of children and comment.**

Solution:

Here ith cluster total (y_i) is the number of children under age 5 in the ith village and total number of women at reproductive age in the ith village is the cluster size for the ith cluster (M_i)

Here N=10, n=3, we have to select a 3 two digit number from random number table .Suppose the selected numbers are 55, 77, 89. So we have to select a sample of 3 villages (consists of 5th, 7th, and 9th villages) with equal probability WOR and we construct the following table:

Table 1: Table for selection of a sample of size 3:

Selected village	Total number of women at reproductive age ($\boldsymbol{M_i}$)	Total number of children under age 5 ($\boldsymbol{y_i}$)
5	2704	3386
7	344	376
9	601	685
Total	$\sum_{i=1}^{n} M_i = 3649$	$\sum_{i=1}^{n} y_i = 4447$

i) Unbiased estimate of total number of children under age 5 in its village is

$$\hat{Y} = N\bar{y} = N\frac{\sum_{i=1}^{n} y_i}{n} = 10 * \frac{4447}{3}$$
$$= 14823.33 \approx 14283$$

ii) The child women ratio and its corresponding sampling variance is

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = \frac{4447}{3649} = 1.218$$

The estimated variance of child-women ratio is given by

$$MSE(\hat{R}) = \frac{1}{M_0^2} MSE(\hat{Y}_R)$$

$$where\ MSE(\hat{Y}_R) = \frac{N^2(1-f)}{n} * \frac{\sum_{i=1}^{N}(y_i - M_i\bar{\bar{Y}})^2}{N-1}$$

$$here\ \bar{\bar{Y}} = \frac{Y}{M_0} = \frac{6557}{5647} = 1.16$$

Table 2: Table for calculating sampling variance

Village	$\boldsymbol{M_i}$	$\boldsymbol{y_i}$	$\boldsymbol{M_i\bar{\bar{Y}}}$	$\boldsymbol{(y_i - M_i\bar{\bar{Y}})^2}$	$\boldsymbol{(y_i - \bar{Y})^2}$
1	401	420	465.16	2039.026	55554.49
2	625	786	725	3721	16978.09
3	173	177	200.68	560.742	229153.69
4	135	95	156.6	3794.56	314384.49
5	2704	3386	3136.64	62180.409	7454538.09
6	177	173	205.32	1044.582	23299.29
7	344	376	399.04	530.842	78232.09
8	201	183	233.16	2516.026	223445.29
9	601	685	677.16	147.866	858.49
10	286	276	331.76	30.9.178	144172.09
Total	$M_0 = 5647$	$Y = 6557$		$\sum_{i=1}^{N}(y_i - M_i\bar{\bar{Y}})^2$ =79644.633	$\sum_{i=1}^{N}(y_i - \bar{Y})^2$ =8750316.1

Now, $MSE(\hat{Y}_R) = \frac{N^2(1-f)}{n} * \frac{\sum_{i=1}^{N}(y_i - M_i\bar{\bar{Y}})^2}{N-1}$

$$= \frac{N^2(N-n)}{nN} * \frac{\sum_{i=1}^{N}(y_i - M_i\bar{\bar{Y}})^2}{N-1}$$

$$= \frac{N(N-n)}{n} * \frac{\sum_{i=1}^{N}(y_i - M_i\bar{\bar{Y}})^2}{N-1}$$

$$= \frac{10(10-3)}{3} * \frac{79644.63}{10-1}$$

$= 206486.0778$

$MSE(\hat{R}) = \frac{1}{M_0^2} MSE(\hat{Y}_R)$

$= \frac{1}{(5647)^2} * 206486.0778$

$= 0.00648$

iii) Estimating the total number of children under age 5 using the estimated child women ratio is given by,

$\hat{Y}_R = M_0\bar{\bar{Y}} = M_0R = 5647 * 1.218$

$= 6878.046 \approx 6878$

iv) $V(\hat{Y}) = \frac{N^2(1-f)}{n} * \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N-1}$, where $\bar{Y} = \frac{Y}{N}$

$= \frac{N(N-n)}{n} * \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N-1}$,

$= \frac{10(10-3)}{3} * \frac{8750316.1}{10-1}$

$= 22686.0047$

Now $SE(\hat{Y}) = \sqrt{V(\hat{Y})} = 4762.983$

$MSE(\hat{Y}_R) = M_0^2 * MSE(\hat{R})$

$= (5647)^2 * .00658$

$SE(\hat{Y}_R) = \sqrt{206638.1863}$

$= 454.575$

Precession: $\frac{1}{SE(\hat{Y})} = \frac{1}{4762.783} = .00021$

$\frac{1}{SE(\hat{Y}_R)} = \frac{1}{454.575} = .00219$

Comparison: Since the precession of the estimate obtained by ratio-to-size method is greater than that of unbiased estimate, we can conclude that ratio-to-size method will give better results for estimating such kind of data.

Practical # 2

1. An auditor wishes to sample bank accounts of the employees of various department of a University to investigate the amount of money of the employees over the last year. The university has departments with varying number of employees as shown in the following table:

Department	No. of employees	Amount
Bangla	16	161
English	21	409
Chemistry	28	256
Physics	19	281
Mathematics	32	261
Statistics	27	252
Management	11	65
Finance	13	96

i) Draw a sample of size 2 without replacement from these 8 departments and obtain unbiased estimate of total amount of money of all employees over the last year by-

a) Brewer's method

b) Murthy's method

c) Random group method by Rao-Hartley-Cochran

ii) Also find standard error of your estimates and hence compare the efficiency of these methods with pps WR sampling.

Solution:

Here, $M_i = number\ of\ employees$

$y_i = amount\ of\ money\ of\ the\ employees\ over\ last\ year$

i)a) Brewer's method:

<u>Sample Selection procedure:</u>

We select a sample of size 2 by Brewer's method by using the following procedure:

<u>For first sampling unit</u>

Table 1: Table for selecting 1st unit by Brewer's method:

Unit number	Mi	$P_i = \frac{M_i}{M_0}$	$\frac{P_i(1-P_i)}{1-2P_i}$	$\frac{P_i(1-P_i)}{D(1-2P_i)}$	Cumulative	range
1	16	0.0958	.01072	.0892	.0892	0-0.0822
2	21	0.1257	.1469	.1222	.2114	0.0892-0.2114

3	28	0.1677	.2099	.1747	.3860	0.2114-0.3860
4	19	0.1138	.1305	.1086	.4947	0.3860-0.4947
5	32	0.1916	.2511	.2089	.7036	0.4947-0.7036
6	27	0.1617	.2003	.1667	.8703	0.7036-0.8703
7	11	0.0659	.0708	.0589	.9292	0.8703-0.9292
8	13	0.0778	.0850	.0707	1	0.9292-1
	$M_0 = 167$		D=1.20186			

Selected unit: A random number is 0.072 so unit 1 is selected

<u>For 2nd sampling unit:</u>

Table 2: Table for selecting 2nd unit by Brewer's method:

Unit number	$\boldsymbol{Mi}$	$\boldsymbol{P_i = \frac{M_i}{M_0}}$	$\boldsymbol{P_i^{/} = \frac{P_i}{1 - P_j}}$	**Cumulative**	**range**
2	21	0.1257	.1390	0.1390	0-0.1390
3	28	0.1677	.1854	0.3245	0.1390-0.3245
4	19	0.1138	.1258	0.4503	0.3245-0.4503
5	32	0.1916	.2119	0.6622	0.4503-0.6622
6	27	0.1617	.1788	0.8410	0.6622-0.8410
7	11	0.0659	.0728	0.9139	0.8410-0.9139
8	13	0.0778	.0860	1	0.9139-1

Selected unit: A random number is 0.687 so unit 6 is selected.

Now by Brewer's method,

$$\hat{Y}_B = \sum_{i=1}^{2} \frac{y_i}{2P_i}$$

$$= \frac{1}{2}(\frac{y_1}{P_1} + \frac{y_6}{P_6})$$

$$= \frac{1}{2}\left(\frac{161}{0.958} + \frac{252}{0.1617}\right) = 1619.55374$$

b) Murthy's Method

For first sampling unit

Table 1: Table for selecting 1st unit by Murthy's method:

Unit number	Mi	$P_i = \frac{M_i}{M_0}$	Cumulative P_i	Range
1	16	0.0958	0.0958	0-0.0958
2	21	0.1257	0.2215	0.0958-0.2215
3	28	0.1677	0.3892	0.2215-0.3892
4	19	0.1138	0.5029	0.3892-0.5029
5	32	0.1916	0.6946	0.5029-0.6946
6	27	0.1617	0.8563	0.6949-08563
7	11	0.0659	0.9221	0.8563-0.9221
8	13	0.0778	1	0.9221-1
	$M_0 = 167$			

Selected unit: A random number 0.506 is drawn so unit 5 is selected

For 2nd sampling unit:

Table 2: Table for selecting 2nd unit by Murthyr's method:

Unit number	Mi	$P_i = \frac{M_i}{M_0}$	$P_i^/ = \frac{P_i}{1 - P_j}$	Cumulative $P_i^/$	range
1	16	0.0958	0.1185	0.1185	0-0.1185
2	21	0.1257	0.1555	0.2740	0.1185-0.2740
3	28	0.1677	0.2074	0.4814	0.2740-0.4814
4	19	0.1138	0.1407	0.6222	0.4814-0.6222
6	27	0.1617	0.2	0.8222	0.6222-0.8222
7	11	0.0659	0.0815	0.9037	0.8222-0.9037
8	13	0.0778	0.0963	1	0.9037-1

Selected unit: A random number is 0.884 so unit 7 is selected.

Unbiased estimate of population total in Murthy's method:

$$\hat{Y}_M = \frac{1}{2 - P_i - P_j}[\left(1 - P_j\right)\frac{y_i}{P_i} + (1 - P_i)\frac{y_j}{P_j}]$$

$$= \frac{1}{2 - P_5 - P_7}[(1 - P_7)\frac{y_5}{P_5} + (1 - P_5)\frac{y_7}{P_7}]$$

$$= \frac{1}{2 - 0.1916 - 0.0659}\left[(1 - 0.0659)\frac{261}{0.1916} + (1 - 0.1911)\frac{65}{0.0654}\right]$$

=1187.835

c)Random group method of Rao-Hartley-Cochran

Determination of number of units in groups:

Since, 8 is divisible by 2, so each of the 2 groups consists of 4 sampling units.

Table 1: Table for determining number of units in the groups.

units	Random numbers	Ranking	Grouping
1	39	5	2
2	14	3	1
3	30	4	1
4	64	7	2
5	42	6	2
6	80	8	2
7	00	1	1
8	05	2	1

For group 1:

Table 2: Table for selecting first unit from the first group (g=1):

Unit number	Mi	$z_{g_i} = \frac{M_i}{M_0}$	$Z_{g_i} = \frac{z_{g_i}}{Z_g}$	Cumulative Z_{g_i}	Range
2	21	0.1257	0.2877	0.2877	0-0.2877
3	28	0.1677	0.3836	0.6712	0.2877-0.672
7	11	0.0658	0.1507	0.8219	0.6712-0.8219
8	13	0.0778	0.1780	0.9999	0.8219-0.9999
		$Z_g = 0.43713$			

Selected unit: A random number 0.78 is drawn. So selected unit from 1st group is 7th unit.

For group 2:

Table 2: Table for selecting 2nd unit from the first group (g=2):

Unit number	Mi	$z_{g_i} = \frac{M_i}{M_0}$	$Z_{g_i} = \frac{z_{g_i}}{Z_g}$	Cumulative Z_{g_i}	Range
1	16	0.0958	0.1702	0.1702	0-0.1702

4	19	0.1138	0.2021	0.3723	0.1702-0.3723
5	32	0.1916	0.3404	0.7128	0.3723-0.7128
6	27	0.1617	0.2872	1	0.7128-1
		$Z_g = 0.56287$			

Selected unit: A random number 0.45 is drawn. So selected unit from 2nd group is 5th unit.

Now unbiased estimate of population total in Rao-Hartley-Cochran's method:

We know the unbiased estimate of population total in rao-hartley-Cochran's method:

$$\hat{Y}_{RHC} = \sum_{g=1}^{2} \hat{Y}_g$$

$$= \sum_{g=1}^{2} \frac{y_g}{\frac{z_{g_i}}{Z_g}}$$

$$= (\frac{y_7}{\frac{z_7}{Z_7}} + \frac{y_5}{\frac{z_5}{Z_5}})$$

$$= \left(\frac{65}{0.1507} + \frac{261}{0.3404}\right) = 1198.06551$$

Now, Sample selection procedure by pps WR method:

Here , N=8, $M_{max} = 32$

Table: for selecting sample of size 2 by pps method

1<i<8 **1<m<32**	**Remark**	**Selected unit**	$\boldsymbol{y_i}$	$\boldsymbol{M_i}$	$\bar{\boldsymbol{y}}_i = \frac{\boldsymbol{y_i}}{\boldsymbol{M_i}}$
(2,18)	$18<M_i$	2	409	21	19.4762
(8,14)	$14>M_i$	-	-	-	-
(4,18)	$18<M_i$	4	281	19	14.7895

so, the selected unit by pps method is 2 and 4 now,

$$\hat{Y}_{pps} = \frac{M_0}{n}\sum_{i=1}^{n}\bar{y}_i$$

$$= \frac{167}{2} * 34.2657 = 2861.1838$$

Estimation of standard error:

Brewer's method:

We know, $V(\hat{Y}_B) = \frac{1}{4}[\frac{2D(1-2P_i)(1-2P_j)}{(1-P_i-P_j)} - 1](\frac{y_i}{P_i} - \frac{y_i}{P_j})^2$

$$= \frac{1}{4}\left[\frac{2 * 1.20186 * (1 - 2 * .0958)(1 - 2 * .1617)}{(1 - .0958 - .1617)} - 1\right]\left(\frac{161}{0.0958} - \frac{252}{0.1617}\right)^2$$

$$= 2874.512214$$

Standard error of Brewer's method for population total, $S.E(\hat{Y}_B) = \sqrt{2874.512214}$

$$= 53.6145$$

Murthy's method:

We know, $V(\hat{Y}_M) = \frac{(1-P_i)(1-P_j)(1-P_i-P_j)}{(2-P_i-P_j)}\left(\frac{y_i}{P_i} - \frac{y_i}{P_j}\right)^2$

$$\frac{(1 - 0.1916)(1 - 0.0659)(1 - 0.1916 - 0.0659)}{(2 - 0.1916 - 0.0659)}\left(\frac{261}{0.1916} - \frac{65}{0.0659}\right)^2$$

$$= 45458.874$$

, $S.E.(\hat{Y}_M) = 213.2109$

Rao-Hartley and Cochran method:

We know,

$$V(\hat{Y}_{RHC}) = \frac{N^2 + K(n-K) - nN}{N^2(n-1) - K(n-K)}\sum_{g=1}^{2} Z_g(\frac{y_g}{z_g} - \hat{Y}_{RHC})^2$$

[here k=0 , cause we take the formula as N=nR+k=2*4+0=8]

$$=\frac{8^2+0-16}{8^2*1-1}\{0.1507*\left(\frac{65}{0.0658}-1198.0655\right)^2+0.3405*\left(\frac{6261}{0.1416}-1198.0655\right)^2$$

=46587165.44

$$S.E.(\hat{Y}_{RHC})=6825.479$$

PPS sampling

We know $V(\hat{Y}_{PPS})=\frac{{M_0}^2}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i-\bar{\bar{y}})^2$, where $\bar{\bar{y}}=\frac{\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}M_i}=\frac{(409+281)}{(21+19)}=17.25$

$$V(\hat{Y}_{PPS})=\frac{167^2}{2*1}\{(19.4762-17.25)^2+(14.7895-17.25)^2\}$$

=153529.3172

$S.E.(\hat{Y}_{PPS})=391.8282$

Efficiency of Brewer's method with sampling WR:

$$\frac{1/S.E.(\hat{Y}_B)}{1/S.E.(\hat{Y}_{PPS})}=\frac{S.E.(\hat{Y}_{PPS})}{S.E.(\hat{Y}_B)}=\frac{391.8282}{53.6145}=7.3083$$

Efficiency of Murthy's method with sampling WR:

$$\frac{1/S.E.(\hat{Y}_M)}{1/S.E.(\hat{Y}_{PPS})}=\frac{S.E.(\hat{Y}_{PPS})}{S.E.(\hat{Y}_M)}=\frac{391.8282}{213.2109}=1.8377$$

Efficiency of Rao=Hartley Cocharm method with sampling WR:

$$\frac{1/S.E.(\hat{Y}_{RHC})}{1/S.E.(\hat{Y}_{PPS})}=\frac{S.E.(\hat{Y}_{PPS})}{S.E.(\hat{Y}_{RHC})}=\frac{391.8282}{6825.479}=0.0574$$

Comment:

From the above efficiency measure we can conclude that, as we compare the three method With PPS method and find out that brewer's method give the highest efficiency (7.3083) than the other two methods. So according to our analysis brewer's method gives the most precise estimate for this data.