Note on Sample Design and Estimation Procedure of NSS 71st Round

1. Introduction

1.1 The National Sample Survey (NSS), set up by the Government of India in 1950 to collect socio-economic data employing scientific sampling methods, started its seventy first round from 1st January 2014 and will continue up to 30th June 2014.

1.2 **Subject Coverage**: The 71st round (January 2014 – June 2014) of NSS is devoted to the subject of Social Consumption and earmarked for surveys on 'Health' and 'Education'. The last survey on health was conducted in 60th round of NSS (January 2004 - June 2004) and the same on education was conducted during 64th round of NSS (July 2007 - June 2008).

2. Outline of Survey Programme

2.1 **Geographical coverage**: This survey covers the whole of the Indian Union.

2.2 **Period of survey and work programme**: The period of survey is of six months duration starting on 1st January 2014 and ending on 30th June 2014.

2.3 **Sub-rounds:** The survey period of this round is divided into two sub-rounds of three months' duration each as follows:

sub-round 1 : January - March 2014
sub-round 2 : April - June 2014

In each of these two sub-rounds equal number of sample villages/ blocks (FSUs) is allotted for survey with a view to ensuring uniform spread of sample FSUs over the entire survey period. Attempt has been made to survey each of the FSUs during the sub-round to which it is allotted. *Because of the arduous field conditions, this restriction is not strictly enforced in Andaman and Nicobar Islands, Lakshadweep, Leh (Ladakh) and Kargil districts of Jammu & Kashmir and rural areas of Arunachal Pradesh and Nagaland.*

2.4 **Schedules of enquiry**: During this round, the following schedules of enquiry are being canvassed:

Schedule 0.0 : List of Households
Schedule 25.0 : Social consumption: Health
Schedule 25.2 : Social consumption: Education

2.5 **Participation of States:** In this round all the States and Union Territories except Andaman & Nicobar Islands, Chandigarh, Dadra & Nagar Haveli and Lakshadweep are participating. The following is the matching pattern of the participating States/ UTs.

State/UT	Extent of matching
Nagaland (U)	triple
Andhra Pradesh, Jammu & Kashmir , Manipur	double
Maharashtra (U)	one and half
Remaining States/ UTs	equal

3. **Sample Design**

3.1 **Outline of sample design:** A stratified multi-stage design has been adopted for the 71st round survey. The first stage units (FSU) are the census villages (Panchayat wards in case of Kerala) in the rural sector and Urban Frame Survey (UFS) blocks in the urban sector. The ultimate stage units (USU) are households in both the sectors. In case of large FSUs, one intermediate stage of sampling is the selection of two hamlet-groups (hgs)/ sub-blocks (sbs) from each rural/ urban FSU.

3.2 **Sampling Frame for First Stage Units:** *For the rural sector,* the list of 2011 census villages (henceforth the term óvillageô would mean Panchayat wards for Kerala) constitutes the sampling frame. In case of Kerala, due to the non-availability of Panchayat wards based on census 2011, the available list of Panchayat wards based on census 2001 is used as the rural frame. For the urban sector, the latest updated list of UFS blocks (phase 2007-12) is considered as the sampling frame.

3.3 **Stratification**: Stratum has been formed at district level. Within each district of a State/UT, generally speaking, two basic strata have been formed: (i) rural stratum comprising of all rural areas of the district and (ii) urban stratum comprising of all the urban areas of the district. However, within the urban areas of a district, if there are one or more towns with population 1 lakh or more as per Census 2011, each of them formed a separate basic stratum and the remaining urban areas of the district has been considered as another basic stratum.

3.3.1 **Special stratum in the rural sector**: There are some villages in Nagaland and Andaman & Nicobar Islands which reamins difficult to access. As in earlier rounds, a special stratum has been formed at State/UT level comprising these villages in the two State/UTs.

3.4 **Sub-stratification:**

3.4.1 **Rural sector**: If ór*ô* be the sample size allocated for a rural stratum, the number of sub-strata formed was ór/2ô. The villages within a district as per frame have been first arranged in ascending order of population. Then sub-strata 1 to ór/2ô have been demarcated in such a way that each sub-stratum comprised a group of villages of the arranged frame and had more or less equal population.

3.4.2 **Urban sector:** If áuô be the sample size allocated for an urban stratum, the number of sub-strata formed was áu/2ô. For all strata, if u/2 >1, implying formation of 2 or more sub-strata, all the UFS blocks within the stratum have been first arranged in ascending order of total number of households in the UFS Blocks as per UFS phase 2007-12. Then sub-strata 1 to áu/2ô have been demarcated in such a way that each sub-stratum had more or less equal number of households.

3.5 **Total sample size (FSUs):** 8300 FSUs have been allocated for the central sample at all-India level. For the state sample, there are 9274 FSUs allocated for all-India. State wise allocation of sample FSUs is given in Table 1.

3.6 **Allocation of total sample to States and UTs:** The total number of sample FSUs heve been allocated to the States and UTs in proportion to population as per *Census 2011* subject to a minimum sample allocation to each State/ UT. While doing so, the resource availability in terms of number of field investigators has been kept in view.

3.7 **Allocation of State/ UT level sample to rural and urban sectors**: State/UT level sample size has been allocated between two sectors in proportion to population as per *Census 2011* with double weightage to urban sector subject to the restriction that urban sample size for bigger states like Maharashtra, Tamil Nadu etc. do not exceed the rural sample size. A minimum of 16 FSUs (minimum 8 each for rural and urban sector separately) is allocated to each State/ UT.

3.8 **Allocation to strata:** Within each sector of a State/ UT, the respective sample size has been allocated to the different strata in proportion to the population as per Census 2011. Stratum level allocation has been adjusted to multiples of 2 with a minimum sample size of 2.

For special strata in the rural areas of Nagaland and A & N Islands, 4 FSUs has been allocated to each.

3.9 **Allocation to sub-strata:** Allocation for each sub-stratum has been 2 in both rural and urban sectors.

3.10 **Selection of FSUs:**

For the rural sector, from each stratum/sub-stratum, required number of sample villages have been selected by Probability Proportional to Size With Replacement (PPSWR), size being the population of the village as per Census 2011.

For the urban sector, from each stratum/sub-stratum, FSUs have been selected by Probability Proportional to Size With Replacement (PPSWR), size being the number of households of the UFS Blocks.

Both rural and urban samples have been drawn in the form of two independent sub-samples and equal number of samples has been allocated among the two sub rounds.

3.11 **Selection of hamlet-groups/ sub-blocks - important steps**

3.11.1 **Criterion for hamlet-group/ sub-block formation:** After identification of the boundaries of the FSU, it is to be determined whether listing will be done in the whole sample FSU or not. In case the approximate present population of the selected FSU is found to be 1200 or more, it will be divided into a suitable number (say, D) of "hamlet-groups" in the rural sector and "sub-blocks" in the urban sector by more or less equalising the population as stated below.

approximate present population of the sample FSU		no. of hgs/sbs to be formed
less than 1200	(no hamlet-groups/sub-blocks)	1
1200 to 1799		3
1800 to 2399		4
2400 to 2999		5
3000 to 3599		6
.....and so on		-

For rural areas of Himachal Pradesh, Sikkim, Uttarakhand (except four districts Dehradun, Nainital, Hardwar and Udham Singh Nagar), Poonch, Rajouri, Udhampur, Reasi, Doda, Kistwar, Ramban, Leh (Ladakh), Kargil districts of Jammu and Kashmir and Idukki district of Kerala, the number of hamlet-groups will be formed as follows:

approximate present population of the sample village		no. of hgs to be formed
less than 600	(no hamlet-groups)	1
600 to 899		3
900 to 1199		4
1200 to 1499		5
1500 to 1799		6
.é ..and so on		-

3.11.2 **Formation and selection of hamlet-groups/ sub-blocks:** In case hamlet-groups/ sub-blocks are to be formed in the sample FSU, the same should be done by more or less equalizing population. Note that while doing so, it is to be ensured that the hamlet-groups/ sub-blocks formed are clearly identifiable in terms of physical landmarks.

Two hamlet-groups (hg)/ sub-blocks (sb) will be selected from a large FSU wherever hamlet-groups/ sub-blocks have been formed in the following manner ï one hg/ sb with maximum percentage share of population will always be selected and termed as hg/ sb 1; one more hg/ sb will be selected from the remaining hgs/ sbs by simple random sampling (SRS) and termed as hg/ sb 2. Listing and selection of the households will be done independently in the two selected hamlet-groups/ sub-blocks. The FSUs without hg/ sb formation will be treated as sample hg/ sb number 1.

3.12 **Formation of second stage strata and allocation of households**:

SSS	composition of SSS within a sample FSU	number of households to be surveyed	
		FSU without hg/sb formation	FSU with hg/sb formation (for each hg/sb)
Schedule 25.0: Social Consumption: Health			
SSS 1	households having at least one child of age less than 1 year	2	1
SSS 2	from the remaining, households with at least one member (including deceased former member) hospitalised during last 365 days	4	2
SSS 3	other households	2	1
Schedule 25.2: Social Consumption: Education			
SSS 1	households with at least one student receiving technical/professional education	2	1
SSS 2	from the remaining, households having at least one student receiving general education	4	2
SSS 3	other households	2	1

3.13 **Selection of households:** From each SSS, for both the schedules, the sample households are selected by SRSWOR.

4. **Estimation Procedure**

4.1 **Notations:**

s = subscript for s-th stratum

t = subscript for t-th sub-stratum

m = subscript for sub-sample (m =1, 2)

i = subscript for i-th FSU [village (panchayat ward)/ block]

d = subscript for a hamlet-group/ sub-block (d = 1, 2)

j = subscript for j-th second stage stratum in an FSU/ hg/sb [j = 1, 2 or 3]

k = subscript for k-th sample household under a particular second stage stratum within an FSU/ hg/sb

D = total number of hgôs/ sbôs formed in the sample FSU

D* = 0 if D = 1

= (D ï 1) for FSUs with D > 1

Z = total size of a rural/urban sub-stratum (= sum of sizes for all the FSUs of a sub-stratum)

z = size of sample village/UFS block used for selection.

n = number of sample FSUs surveyed including óuninhabitatedô and ózero casesô but excluding casualty for a particular sub-sample and sub-stratum.

H = total number of households listed in a second-stage stratum of an FSU / hamlet-group or sub-block of sample FSU

h = number of households surveyed in a second-stage stratum of an FSU / hamlet-group or sub-block of sample FSU

x, y = observed value of characteristics *x*, *y* under estimation

$\check{X}$, $\check{Y}$ = estimate of population total X, Y for the characteristics *x*, *y*

Under the above symbols,

$y_{stmidjk}$ = observed value of the characteristic y for the k-th household in the j-th second stage stratum of the d-th hg/ sb (d = 1, 2) of the i-th FSU belonging to the m-th sub-sample for the t-th sub-stratum of s-th stratum.

However, for ease of understanding, a few symbols have been suppressed in following paragraphs where they are obvious.

4.2 **Formulae for Estimation of Aggregates for a particular sub-sample and stratum × sub-stratum:**

4.2.1 **Schedule 0.0:**

4.2.1.1 **Rural/Urban:**

(i) For estimating the number of households in a stratum × sub-stratum possessing a characteristic:

$$\hat{Y} = \frac{Z}{n} \sum_{i=1}^{n} \frac{1}{z_i} \left[y_{i1} + D_i^* \times y_{i2} \right]$$

where y_{i1} , y_{i2} are the total number of households possessing the characteristic *y* in hgôs 1 & 2 of the i-th FSU respectively.

(ii) For estimating the number of villages in a stratum × sub-stratum possessing a characteristic:

$$\hat{Y} = \frac{Z}{n} \sum_{i=1}^{n} \frac{1}{z_i} y_i$$

where y_i is taken as 1 for sample villages possessing the characteristic and 0 otherwise.

4.2.2 **Schedules 25.0 & 25.2:**

4.2.2.1 **Rural/ Urban:**

(i) For j-th second-stage stratum of a stratum × sub-stratum:

$$\hat{Y}_j = \frac{Z}{n_j} \sum_{i=1}^{n_j} \frac{1}{z_i} \left[\frac{H_{i1j}}{h_{i1j}} \sum_{k=1}^{h_{i1j}} y_{i1jk} + D_i^* \times \frac{H_{i2j}}{h_{i2j}} \sum_{k=1}^{h_{i2j}} y_{i2jk} \right]$$

(ii) For all second-stage strata combined:

$$\hat{Y} = \sum_j \hat{Y}_j$$

4.3 **Overall Estimate for Aggregates for a sub-stratum**:

Overall estimate for aggregates for a sub-stratum ($\hat{Y}_{st}$) based on two sub-samples in a sub-stratum is obtained as:

$$\hat{Y}_{st} = \frac{1}{2}\sum_{m=1}^{2}\hat{Y}_{stm}$$

4.4 **Overall Estimate for Aggregates for a stratum**:

Overall estimate for a stratum ($\hat{Y}_s$) will be obtained as

$$\hat{Y}_s = \sum_t \hat{Y}_{st}$$

4.5 **Overall Estimate of Aggregates at State/UT/all-India level**:

The overall estimate $\hat{Y}$ at the State/ UT/ all-India level is obtained by summing the stratum estimates $\hat{Y}_s$ over all strata belonging to the State/ UT/ all-India.

4.6 **Estimates of Ratios**:

Let $\hat{Y}$ and $\hat{X}$ be the overall estimates of the aggregates Y and X for two characteristics *y* and *x* respectively at the State/ UT/ all-India level.

Then the combined ratio estimate $(\hat{R})$ of the ratio $\left(R = \frac{Y}{X}\right)$ will be obtained as $\hat{R} = \frac{\hat{Y}}{\hat{X}}$.

4.7 **Estimates of Error**: The estimated variances of the above estimates will be as follows:

4.7.1 **For aggregate** $\hat{Y}$:

$$\hat{Var}(\hat{Y}) = \sum_s \hat{Var}(\hat{Y}_s) = \sum_s \sum_t \hat{Var}(\hat{Y}_{st})$$ where $\hat{Var}(\hat{Y}_{st})$ is given by

$\hat{Var}\left(\hat{Y}_{st}\right) = \frac{1}{4}\left(\hat{Y}_{st1} - \hat{Y}_{st2}\right)^2$, where $\hat{Y}_{st1}$ and $\hat{Y}_{st2}$ are the estimates for sub-sample 1 and sub-sample 2 respectively for stratum ‘s’ and sub-stratum ‘t’.

4.7.2 **For ratio** $\hat{R}$:

$$\hat{MSE}(\hat{R}) = \frac{1}{4\hat{X}^2}\sum_s \sum_t \left[\left(\hat{Y}_{st1} - \hat{Y}_{st2}\right)^2 + \hat{R}^2\left(\hat{X}_{st1} - \hat{X}_{st2}\right)^2 - 2\hat{R}\left(\hat{Y}_{st1} - \hat{Y}_{st2}\right)\left(\hat{X}_{st1} - \hat{X}_{st2}\right)\right]$$

4.7.3 **Estimates of Relative Standard Error (RSE)**:

$$\hat{RSE}\left(\hat{Y}\right) = \frac{\sqrt{\hat{Var}\left(\hat{Y}\right)}}{\hat{Y}} \times 100$$

$$\hat{RSE}\left(\hat{R}\right) = \frac{\sqrt{\hat{MSE}\left(\hat{R}\right)}}{\hat{R}} \times 100$$

5. **Multipliers**:

The formulae for multipliers at stratum/sub-stratum/second-stage stratum level for a sub-sample and schedule type are given below:

sch type	sector	formula for multipliers	
		hg / sb 1	hg / sb 2
0.0	rural/urban	$\frac{Z_{st}}{n_{stm}} \times \frac{1}{z_{stmi}}$	$\frac{Z_{st}}{n_{stm}} \times \frac{1}{z_{stmi}} \times D^{*}_{stmi}$
25.0/ 25.2	rural/urban	$\frac{Z_{st}}{n_{stmj}} \times \frac{1}{z_{stmi}} \times \frac{H_{stmi1j}}{h_{stmi1j}}$	$\frac{Z_{st}}{n_{stmj}} \times \frac{1}{z_{stmi}} \times D^{*}_{stmi} \times \frac{H_{stmi2j}}{h_{stmi2j}}$
	(j = 1, 2, 3)		

Note:

(i) For estimating any characteristic for any domain not specifically considered in sample design, indicator variable may be used.

(ii) Multipliers have to be computed on the basis of information available in the listing schedule irrespective of any misclassification observed between the listing schedule and detailed enquiry schedule.

(iii) For estimating number of villages possessing a characteristic, D^{*}_{stmi} = 0 in the relevant multipliers and there will be only one multiplier for the village.

6. **Treatment for zero cases, casualty cases etc.**:

6.1 While counting the number of FSUs surveyed (n_{sm} or n_{stm}) in a stratum/sub-stratum, all the FSUs with survey codes 1 to 6 in schedule 0.0 will be considered. In addition, if no SSU is available in the frame for a particular schedule then also that FSU will be treated as surveyed in respect of that schedule. However, if the SSUs of a particular schedule type are available in the frame of the FSU but none of these could be surveyed then that FSU has to be treated as casualty and it will not be treated as surveyed in respect of that schedule.

6.2 *Casualty cases*: FSUs with survey code 7 as per schedule 0.0 are treated as casualties. In addition to this, an FSU, although surveyed, may have to be treated as casualty for a particular schedule type and a particular *second stage stratum* as given in the following para:

6.2.1 FSUs with survey codes 1 or 4 as per schedule 0.0 having number of households in the frame of j-th second stage stratum greater than 0 but number of households surveyed according to data file, considering both hg/sb together, as nil (i.e. $H_{i1j} + H_{i2j} > 0$ but $h_{i1j} + h_{i2j} = 0$) will be taken as casualties for j-th second stage stratum.

All the FSUs with survey codes 1 to 6 as per schedule 0.0 minus the number of casualties as identified above will be taken as the number of surveyed FSUs (n_{stmj}) for that (stratum/sub-stratum) ×(second stage stratum).

When casualty for j-th second stage stratum occurs for a particular hg/sb but not for the other hg/sb, the FSU will not be treated as casualty but some adjustments in the value of H for the other hg/sb will be done as follows:

(i) Suppose for hg/sb 1, $H_{i1j} > 0$ but $h_{i1j} = 0$ while for hg/sb 2, $H_{i2j} > 0$ and $h_{i2j} > 0$. In that case $D_i^* \times H_{i2j}$ will be replaced by $(H_{i1j} + D_i^* \times H_{i2j})$ in the formula for multiplier of hg/sb 2.

(ii) Suppose for hg/sb 1, $H_{i1j} > 0$ and $h_{i1j} > 0$ while for hg/sb 2, $H_{i2j} > 0$ but $h_{i2j} = 0$. In that case H_{i1j} will be replaced by $(H_{i1j} + D_i^* \times H_{i2j})$ in the formula for multiplier of hg/sb 1.

It may be noted that n_{smj} or n_{stmj} would be same for hg/sb 1 & 2 of an FSU.

7. **Treatment in cases of void second-stage strata/sub-strata /strata at FSU or household level**

7.1 A stratum/sub-stratum may be void because of the casualty of all the FSUs belonging to the stratum/sub-stratum. This may occur in one sub-sample or in both the sub-samples. If it relates to only one sub-sample, then estimate for the void stratum/sub-stratum may be replaced with the estimate as obtained from the other sub-sample for the same stratum/sub-stratum.

7.2 When a stratum/sub-stratum is void in both the sub-samples, the following procedure is recommended:

Case(I): Stratum/Sub-stratum void cases at FSU levels (i.e. all FSUs having survey code 7):

(i) If a rural/urban sub-stratum is void then it may be merged with the other sub-stratum of the stratum.

(ii) If a rural/urban stratum (district) is void due to all FSUs being casualty, it may be excluded from the coverage of the survey. The state level estimates will be based on the estimates of districts for which estimates are available and remarks to that effect may be added in appropriate places.

Case (II): Stratum/Sub-stratum void case at second stage stratum level (i.e. all the FSUs are casualties for a particular second stage stratum):

An FSU may be a casualty for a particular *second stage stratum* although survey code is not 7. If all the FSUs of a stratum/sub-stratum become casualties in this manner for a particular *second stage stratum*, the stratum/sub-stratum will become void. In such cases, sub-strata will be merged with other sub-strata for all the second stage strata as in *Case (I) above.*

However, if whole district/stratum becomes void in this manner for a particular second stage stratum, adjustment for this type of stratum void case may be done according to the following guidelines.

The adjustment will be made involving other strata/sub-strata (within NSS region) of the State/U.T. Suppose A, B, C and D are the four strata in the State/UT/Region and stratum C is void for j-th *second stage stratum*. If $\check{Y}_{aj}$, $\check{Y}_{bj}$ and $\check{Y}_{dj}$ are the aggregate estimates for the strata/sub-strata A, B and D respectively, then the estimate $\check{Y}_{cj}$ for stratum/sub-stratum C may be obtained as $\left(\frac{\check{Y}_{aj}+\check{Y}_{bj}+\check{Y}_{dj}}{Z_a+Z_b+Z_d}\times Z_c\right)$ where Z_a, Z_b, Z_c and Z_d are the sizes of strata A, B, C and D respectively.

8. **Reference to the values of Z_{st}, n_{st}, n_s, z_{sti}, D_{sti}, D^*_{sti}, D_{si}, D^*_{si}, H_{sti1j}, h_{sti1j}, H_{sti2j}, h_{sti2j}:**

(a) Values of $\mathbf{Z_{st}}$ and allotted $\mathbf{n_{st}}$ for the whole round are given in appendix Table 2 for rural sector and in Table 3 for urban sector.

(b) $\mathbf{n_{st}}$ should not be taken from the tables. The value of $\mathbf{n_{stm}}$ for each sub-sample is to be obtained following the guidelines given in para 6 above. It includes uninhibited and zero cases but excludes casualty cases.

(c) The value of $\mathbf{z_{sti}}$ for the samples selected by PPS is to be taken from the column of sample list under the heading ôframe populationô for rural samples and ôblock sizeô i.e. total number of households in each UFS block for urban samples.

(d) Value of $\mathbf{D_{sti}}$ is to be taken from item 16 of block 1, sch 0.0. $\mathbf{D^*_{sti}}$ is to be calculated from the value of $\mathbf{D_{sti}}$.

(e) Values of $\mathbf{H_{sti1j}}$, $\mathbf{H_{sti2j}}$ are to be taken from col.(5), block 6 of sch 0.0 for respective hg/sb and second-stage stratum.

(f) The value of $\mathbf{h_{sti1j}}$ and $\mathbf{h_{sti2j}}$ should not be taken from col (9), block 6 of sch.0.0. The figures should be obtained by counting the number of households in the data file excluding the casualty households.

****