

Sample Design for Millennium Development Goals Survey 2006/2007

Sample design

A stratified two-stage sample design was used for this survey. A national sample of 25,000 housing units was decided in order to provide reasonably accurate estimate by district level, national level estimate by sector.

Sample frame

The list of census blocks (enumeration area with the no of housing units) which was prepared at the, Census of Population & Housing - 2001 data file.

Selection of primary sampling units (PSU's)

Primary sampling units are the census blocks which were selected using 'Probability proportion are to size with Systematic' sampling method (size measure is no of housing units in a census block)

Selection of secondary sampling units (SSU's)

Secondary sampling unit is a housing unit. From each selected Primary Sampling Unit, 10 housing units (SSU's) were selected using 'Systematic sampling Method'.

Estimation procedure

Let \hat{X}_{jk} be the estimate of any given characteristic for j^{th} district

This could be given by,

$$\begin{aligned}\hat{X}_{jk} &= \frac{1}{m_{jk(u)}} \sum_{h(u)=1}^{m_{jk(u)}} \frac{1}{P_{h(u)}} \frac{N_{h(u)}}{n_{h(u)}} \sum_{i(u)=1}^{n_{h(u)}} X_{hi(u)} \\ &+ \frac{1}{m_{jk(r)}} \sum_{h(r)=1}^{m_{jk(r)}} \frac{1}{P_{h(r)}} \frac{N_{h(r)}}{n_{h(r)}} \sum_{i(r)=1}^{n_{h(r)}} X_{hi(r)} \\ &+ \frac{1}{m_{jk(e)}} \sum_{h(e)=1}^{m_{jk(e)}} \frac{1}{P_{h(e)}} \frac{N_{h(e)}}{n_{h(e)}} \sum_{i(e)=1}^{n_{h(e)}} X_{hi(e)} + \frac{1}{m_{jk(ts)}} \sum_{h(ts)=1}^{m_{jk(ts)}} \frac{1}{P_{h(ts)}} \frac{N_{h(ts)}}{n_{h(ts)}} \sum_{i(ts)=1}^{n_{h(ts)}} X_{hi(ts)}\end{aligned}$$

Where

$m_{jk(u)}$ = Number of census blocks selected from the urban sector of the j^{th} district .

$P_{h(u)}$ = Selection probability of the h^{th} census block in the urban sector.

$$P_{h(u)} = \frac{S_{jh(u)}}{\sum_{h=1}^{M_{jh(u)}} S_{jh(u)}}$$

$S_{jh(u)}$ = Measure of size (number of housing units) of the h^{th} census block in the urban sector of the j^{th} district.

$M_{jh(u)}$ = Total number of census blocks in urban sector of the j^{th} district.

$N_{h(u)}$ = Total number of housing units listed in the h^{th} census block in the urban sector

$n_{h(u)}$ = Number of housing units selected from the h^{th} census block in the urban sector

$X_{hi(u)}$ = The observed value for the i^{th} sample household in the h^{th} census block in the urban sector.

$m_{jk(r)}$, $P_{h(r)}$, $S_{jh(r)}$, $M_{jh(r)}$, $N_{h(r)}$, $n_{h(r)}$, and $X_{hi(r)}$ are corresponding terms for the rural sector and $m_{jk(e)}$, $P_{h(e)}$, $S_{jh(e)}$, $M_{jh(e)}$, $N_{h(e)}$, $n_{h(e)}$, and $X_{hi(e)}$ are corresponding terms for the estate sector. The estimate for the total value of a characteristic for the country.

Sample weights

$$W_i = \frac{1}{m_{jk}} \times \frac{1}{p_{jk}} \times \frac{N_{h(c)}}{n_{h(c)}}$$

Adjustment for unit non-response

The occurrence of unit non-response was determined by examining the final result code recorded under control data section of the schedule. Based on the final result codes the households were grouped into the following categories, which were used as a basis for adjusting for the unit non-response.

Category and description	Result code
1. Schedule completed	1
2. Housing unit demolished or vacant	6,7,8
3. Unable to complete schedule, refusal, temporarily away etc.	2,3,4,5,9

a) Each PSU has 'Separate' adjustment factor

$$W_2 = \frac{\text{Category (1-9)} - \text{Category (6,7,8)}}{\text{Category (1)}}$$

Final weights for households

$$W_{\text{final HH}} = W_1 \times W_2$$

Millennium Development Goals Survey
Allocation of Primary Sampling units
by District & sector

District	Total	Urban	Rural	Estate
Colombo	272	190	80	2
Gampaha	207	56	149	2
Kalutara	122	22	93	7
Kandy	148	28	95	25
Matale	65	6	49	10
Nuwara eliya	107	7	30	70
Galle	104	18	83	3
Matara	90	12	75	3
Hambantota	66	7	57	2
Jaffna	80	20	60	
Mannar	65		65	
Vavunia	65		65	
Mulativu	65		65	
Killinochchi	65	8	57	
Batticaloe	65	25	40	
Ampara	80	20	60	
Trincomalee	65	7	58	
Kurunegala	145	6	137	2
Puttalam	92	10	80	2
Anuradhapur	90	8	80	2
Polonnaruwa	67		65	2
Badulla	119	8	75	36
Monaragala	65		63	2
Ratnapura	104	6	80	18
Kegalle	87	5	70	12
Total	2500	469	1831	200

Total no. of Secondary Sampling Units

$$= 2500 \times 10$$

$$= 25000$$