Appendix 1

Sample Design

Introduction

Generally, the size of the sample is selected on the basis of the tolerance levels of accuracy and financial constraints. As a nationally representative and multi-topic household survey, Annual Household Survey (AHS) is primarily focused on studying two basic indicators, viz. household consumption aggregate and current labour force. Therefore, these study areas are given particular attention during the every steps of the sample design process.

This note summarizes the sampling procedures employed to carry out AHS-I. It describes the process of determining sample size. It also defines the allocation of PSUs, and illustrates the steps involved in selecting the Primary Sampling Units (PSUs), and demonstrates the method of calculating sampling weights to be used in analysis of the survey data.

Sample size

For the AHS 2012/13, the sample size is determined by an equation as given below:

$$x = \frac{z^2(r)(1-r)fk}{pne^2}$$

Where,

'z' = is the z - statistic;

'r' = represents a previous estimate of the key indicator (unemployment rate) to be measured by the survey;

'f ' = indicates the 'design-effect';

'k' = denotes an adjustment factor for non-response;

'p' = stands for the proportion of the total population accounted for by the target population upon which the estimate of r has been previously obtained, i.e., employed population;

'n' = denotes the number of individuals per household; and

'e' = symbolizes tolerance level to error in measuring r.

Assuming unemployment rate of the previous labour force survey as the key indicator, above

mentioned equation offered the sample size of about 2500 households which is essential for the sex-wise estimates for two domains at 95 % level of confidence. Then, to achieve the higher level of confidence by increasing the accuracy of the estimates, particularly for those indicators having long list of sub-items like educational attainment, industry or occupation, 500 more households is added there. Thus, 3000 household is considered statistically sufficient to measure consumption and labour force estimates at urban/rural level.

Master sample frame

When the sampling plans were being discussed at CBS, the report of Population and Housing Census 2011 was published. The brand new sample frame from the Census 2011 is being used for the first time for sampling of AHS. In this frame data were available for only 35,214 rural wards and 806 urban wards.

The samples were not selected directly from the frame of population census data. At the first stage, a master sample frame was created by taking 8,400 rural wards by taking 112 wards from each district. This was done to ensure geographical representativeness. The sampling selection was done by Probability Proportional to size (PPS) sampling where the square root of number of households in the ward was taken as the measure of size was. In case of the urban master sample frame, all 806 wards were included.

Sample allocation

The design is based on multi stages stratified sampling technique with equal PSUs or households distributed between urban and rural areas considering the heterogeneous labour force activities to provide a detailed picture of employment situation in the urban areas. So the prescribed 200 PSUs are divided equally in two parts, i.e., 100 PSUs each for urban and rural.

Sample selection

100 wards from urban master sample frame and 100 from rural master sample frame were selected using PPS method. At this stage, to ensure representativeness, the frames were sorted according to district and VDC/Municipality codes from East to West. In case of larger wards which contain more than 300 households, one Enumeration Area (EA) was selected from each of the wards by using PPS Sampling. For this, cartography maps were used to divide the ward into EAs.

At last, after the listing of all the households in the PSU (a ward or an EA), systematic sampling method was used for selecting 15 households.

Distribution of Selected PSU by Districts

The following table shows the distribution of selected PSUs by district.

		Selected PSU					Selected PSU		
S.N.	**District**	**Urban**	**Rural**	**Total**	**S.N.**	**District**	**Urban**	**Rural**	**Total**
1	Taplejung	0	0	0	39	Syangja	2	1	3
2	Panchthar	0	1	1	40	Kaski	7	1	8
3	Ilam	1	2	3	41	Manang	0	0	0
4	Jhapa	4	3	7	42	Mustang	0	0	0
5	Morang	4	4	8	43	Myagdi	0	1	1
6	Sunsari	5	2	7	44	Parbat	0	1	1
7	Dhankuta	1	1	2	45	Baglung	1	1	2
8	Terhathum	0	1	1	46	Gulmi	0	1	1
9	Sankhuwasabha	1	0	1	47	Palpa	1	2	3
10	Bhojpur	0	1	1	48	Nawalparasi	1	2	3
11	Solukhumbu	0	1	1	49	Rupandehi	4	3	7
12	Okhaldhunga	0	0	0	50	Kapilbastu	1	2	3
13	Khotang	0	1	1	51	Arghakhanchi	0	1	1
14	Udayapur	2	2	4	52	Pyuthan	0	1	1
15	Saptari	1	2	3	53	Rolpa	0	1	1
16	Siraha	2	3	5	54	Rukum	0	1	1
17	Dhanusa	2	2	4	55	Salyan	0	1	1
18	Mahottari	1	3	4	56	Dang	3	2	5
19	Sarlahi	1	3	4	57	Banke	2	2	4
20	Sindhuli	2	1	3	58	Bardiya	1	2	3
21	Ramechhap	0	1	1	59	Surkhet	2	1	3
22	Dolakha	1	1	2	60	Dailekh	1	1	2
23	Sindhupalchok	0	1	1	61	Jajarkot	0	1	1
24	Kavrepalanchok	2	2	4	62	Dolpa	0	0	0
25	Lalitpur	5	1	6	63	Jumla	0	1	1

		Selected PSU					Selected PSU		
S.N.	**District**	**Urban**	**Rural**	**Total**	**S.N.**	**District**	**Urban**	**Rural**	**Total**
26	Bhaktapur	4	1	5	64	Kalikot	0	0	0
27	Kathmandu	13	3	16	65	Mugu	0	0	0
28	Nuwakot	1	1	2	66	Humla	0	0	0
29	Rasuwa	0	1	1	67	Bajura	0	1	1
30	Dhading	0	1	1	68	Bajhang	0	1	1
31	Makwanpur	2	2	4	69	Achham	0	1	1
32	Rautahat	1	2	3	70	Doti	1	1	2
33	Bara	1	3	4	71	Kailali	3	2	5
34	Parsa	2	1	3	72	Kanchanpur	2	2	4
35	Chitawan	5	2	7	73	Dadeldhura	1	0	1
36	Gorkha	1	1	2	74	Baitadi	1	1	2
37	Lamjung	0	1	1	75	Darchula	0	1	1
38	Tanahu	1	2	3		Total	100	100	200

Calculation of sampling weight

The sampling weights for urban and rural domain is calculated separately as illustrated in the following table where,

$HH_i = number\ of\ Households\ in\ the\ i^{th}\ PSU$

$HH_{ij} = number\ of\ Households\ in\ the\ j^{th}\ EA\ of\ i^{th}\ PSU$

Description	**Urban Domain**	**Rural Domain**
Probability of a PSUi is selected in Master Sample frame (P_1)	1	$\frac{8400\times\sqrt{HH_i}}{\sum_{i=1}^{35215}\sqrt{HH_i}}$
A PSUi is selected in the sample (P_2)	$\frac{100\times\sqrt{HH_i}}{\sum_{i=1}^{806}\sqrt{HH_i}}$	$\frac{100\times\sqrt{HH_i}}{\sum_{i=1}^{8400}\sqrt{HH_i}}$
An EAj is selected from the selected PSUi (P_3)	$\frac{HH_{ij}}{HH_i}$	$\frac{HH_{ij}}{HH_i}$

Description	**Urban Domain**	**Rural Domain**
A Household is selected (P_4)	$\frac{15}{HH_{ij}}$	$\frac{15}{HH_{ij}}$
Weight of a HH in any domain=	$\frac{1}{P_1 \times P_2 \times P_3 \times P_4}$	$\frac{1}{P_1 \times P_2 \times P_3 \times P_4}$
Weight of a selected HH in the domain	$\frac{\sum_{i=1}^{806} \sqrt{HH_i} \times \sqrt{HH_i}}{1500}$	$\frac{\sum_{i=1}^{8400} \sqrt{HH_i} \times \sum_{i=1}^{35215} \sqrt{HH_i}}{8400 \times 15 \times 100}$
Actual Values	$\frac{25155.36 \times \sqrt{HH_i}}{1500}$	$\frac{102020.51 \times 366917.464}{8400 \times 15 \times 100}$ =2970.88

In rural domain, data collection was conducted only at 99 PSUs so that actual value of the weight of this domain equals 3000.89 after adjustment.