

1.3 Sample Design

Background:

The HIES 2016 deviates from the sampling design used in the previous round of HIES 2010 in several ways. The objectives of HIES 2016 have changed significantly from HIES 2010. In HIES 2010, sample was designed to provide reliable annual estimates at division level with urban & rural break down. But in HIES 2016, the sample was designed to achieve three objectives:

- i) reliable annual estimates at 64 district level
- ii) reliable quarterly estimates at the national level and
- iii) reliable annual estimates at the division level for urban and rural areas.

To achieve this multiple objectives, BBS needed to change the sample design of HIES 2016 significantly from HIES 2010. The first significant change was to increase the sample size to almost four times compare to HIES 2010. This substantial increase in the sample size also forced to use a new sampling frame instead of the previous Integrated Multi Purpose Sample (IMPS). The IMPS is a master sample updated after each Census of Population and Housing. This IMPS was used as sampling frame for the selection of Primary Sampling Units (PSUs) for HIES 2010 and also for other surveys in BBS.

Sampling frame:

The frame used in the selection of Primary Sample Units (PSUs) for HIES 2016 was based on the Census of Population and Housing 2011. PSUs for HIES 2016 are the Enumeration Areas (EAs) used for the Census of Population and Housing in Bangladesh. Each EA is a cluster of 110 households on an average. The sampling frame for the selection of PSUs consists in the list of all EAs covering people residing in dwelling households (non-institutional households) in Bangladesh.

Stratification:

In the sample design of HIES 2016, two different levels of Stratification were followed:

- i) As of HIES 2016, Bangladesh had eight administrative divisions. These were Barishal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur and Sylhet. First of all, these 8 divisions of the country were stratified by 3 basic localities viz. Rural, Urban and City Corporation. Thus, there should have been $8 \times 3 = 24$ strata. But as the sampling frame (Population Census 2011) does not contain Rangpur city corporation and other two city corporations viz. Barishal & Sylhet are not much different from urban characteristics of these two city corporations, BBS included only the four main city corporations (Dhaka, Chattogram, Khulna and Rajshahi) in the city corporation locality. This brought the number of main strata to 20 (8 rural divisions + 8 urban divisions + 4 main city corporations).

ii) Secondly, as the PSUs of HIES 2016 will be allocated at district (zila) level, the sample was implicitly sub-stratified at the district level. Since there are a total of 64 districts in Bangladesh, the sample design includes a total of 132 sub-strata: (64 rural, 64 urban and 4 city corporations).

Sample Size:

As the survey needs to provide district (zila) level estimates, it is obvious, the sample size needs to be much higher than the previous HIES 2010. Sample size can be determined using the prevalence rate of the main indicator (poverty rate) or the coefficient of variation of per capita consumption expenditure or household consumption expenditure which are the core indicators of HIES and each one can be treated as target variable for determining the sample size. For our purpose, mean household consumption expenditure was treated as target variable.

The following formula was used to find the sample size for each district.

$$n = \left(\frac{Z_{\alpha/2} * CV_{SR5}(\bar{Y})}{r(\bar{Y})} \right)^2 * DEFF \quad (1)$$

where n is the minimum sample size required for allocation to each district in order to achieve a certain level in the accuracy statistic $r(\bar{Y})$ associated with the targeted variable \bar{Y} ; $CV_{SR5}(\bar{Y})$ is the coefficient of variation of the targeted variable estimated under the assumption of simple random sampling; $DEFF$ is the design effect of the target variable; and $Z_{\alpha/2}$ is the critical value of a standard normal distribution with $\alpha\%$ confidence level.

In the case of the HIES 2016, $r(\bar{Y})$ is the 10 percent RSE desired for the mean total household expenditure estimated at the district level; $DEFF$ is the average design effect of the target variable across all districts; $CV_{SR5}(\bar{Y})$ is the coefficient of variation of total household expenditure estimated at the national level; and $Z_{\alpha/2}$ is set at 1.96, which corresponds to a significance level of 5 percent.

Substituting all values in equation (1), we find that the sample size needed is 715 households for each district. However, for practical consideration and to facilitate field work and survey implementation management, 720 households were allocated to each district. More specifically, 720 households is divisible by 36 (number of PSUs ultimately allocated to each district), which allows easy allocation of 20 households per PSU. Also, two enumeration teams can easily cover the 36 PSUs in each district over 12 months without having to move to a different district. Lastly, the number 36 PSUs, allocated to each district is a multiple of 4, which allows dividing into quarters.

Sample Allocation:

As one of our goals here is to estimate and compare Zila level means, equal allocation of PSUs to Zila may be a better choice. That is 36 PSU's will be assigned to each zila. Secondly these 36 PSUs will be allocated across rural, urban and city corporation sub-strata using modified Neyman's allocation technique which not only takes into account both the size and variability (standard deviation) of interest but also uses square root of design effect (deft) to modify the standard deviation. Although, the sample design is not supposed to provide Zila level estimates by rural, urban or city corporation, the Neyman's allocation taking into account the variability of the locality (rural/urban/city corporation) will greatly improve the precision of estimates at Zila as well as aggregate (National or Divisional) level.

Sample selection:

The HIES 2016 followed a stratified two stage cluster sampling design. At the first stage, a total of 36 PSUs (EAs) was drawn from each Zila (Domain) applying PPS systematic sampling technique, number of households in each PSU being the measure of size. These 36 PSUs were selected independently from rural, urban and city corporation sub-stratum. Therefore, in total, there will be $64 \times 36 = 2,304$ sample PSUs for the survey. Enumeration Area, a cluster of around 110 households of population census 2011, was treated as PSU for this sample design. The sampling frame for this purpose was developed from the population census 2011 data. A file containing all the EAs of the population census 2011 was created. This file contains all the unique geographic codes from division down to EA and also locality code (rural, urban and city corporation). In order to select the sample PSUs independently by stratum and Zila, the sampling frame was properly sorted by stratum and geo-codes. Then, at the first stage, the required number of PSUs as shown in table-1 was selected using probability proportional to size (PPS) systematic sampling, size measure being the number of households in each PSU. After selection of the PSUs, a complete household listing in these selected PSUs was done in the field. Subsequently, this was computerized and used to draw the 20 households along with 5 reserved households from each of the selected PSUs at the second stage. Thus, total sample size for the survey stands at $2304 \times 20 = 46,080$ households.

Sampling weights and probability of selection:

Sampling probability was computed separately for each sampling stage and for each PSU within a Domain (Zila).

In the case of a two-stage, stratified clustered design, such as HIES 2016, the probability of being selected into the sample is a function of: i) the probability of a PSU being selected in the first stage and ii) the probability of a household being selected within each PSU in the second stage. This can be calculated as follows:

$$p_{hij} = p_1 * p_2 = \frac{k_h m_{hi}}{N_h} * \frac{m_{hi}}{m'_{hi}} \quad (2)$$

where p_{hij} is the probability of household j , in stratum h , and PSU i to be included in the sample; p_1 is the probability of the PSU to be selected in the first stage; p_2 is the probability of a household being selected in the second stage; k_h is the number of PSUs selected in stratum h ; m_{hi} is the number of households selected in PSU hi ; and N_h is the total number of households in stratum h .

Table 1: Allocation of PSUs for Rural/Urban/City Corporation, HIES 2016

Zila name	Rural	Urban	City corporation	Total
Bagerhat	26	10	0	36
Bandarban	32	4	0	36
Barguna	25	11	0	36
Barishal	26	10	0	36
Bhola	29	7	0	36
Bogura	30	6	0	36
Brahmanbaria	29	7	0	36
Chandpur	29	7	0	36
Chapai Nababganj	32	4	0	36
Chattogram	4	4	28	36
Chuadanga	27	9	0	36
Cumilla	31	5	0	36
Cox's Bazar	30	6	0	36
Dhaka	4	4	28	36
Dinajpur	32	4	0	36
Faridpur	28	8	0	36
Feni	26	10	0	36
Gaibandha	29	7	0	36
Gazipur	4	32	0	36
Gopalganj	24	12	0	36
Habiganj	25	11	0	36
Jamalpur	28	8	0	36

Zila name	Rural	Urban	City corporation	Total
Jashore	24	12	0	36
Jhalokati	19	17	0	36
Jhenaidah	15	21	0	36
Joypurhat	29	7	0	36
Khagrachhari	32	4	0	36
Khulna	10	4	22	36
Kishoregonj	32	4	0	36
Kurigram	29	7	0	36
Kushtia	28	8	0	36
Lakshmipur	32	4	0	36
Lalmonirhat	32	4	0	36
Madaripur	32	4	0	36
Magura	20	16	0	36
Manikganj	28	8	0	36
Maulvibazar	29	7	0	36
Meherpur	32	4	0	36
Munshiganj	22	14	0	36
Mymensingh	20	16	0	36
Naogaon	32	4	0	36
Narail	27	9	0	36
Narayanganj	4	32	0	36
Narsingdi	20	16	0	36
Natore	30	6	0	36
Netrakona	28	8	0	36
Nilphamari	26	10	0	36
Noakhali	31	5	0	36
Pabna	25	11	0	36
Panchagarh	25	11	0	36
Patuakhali	32	4	0	36
Pirojpur	31	5	0	36
Rajbari	19	17	0	36
Rajshahi	4	4	28	36

Zila name	Rural	Urban	City corporation	Total
Rangamati	32	4	0	36
Rangpur	25	11	0	36
Satkhira	27	9	0	36
Shariatpur	24	12	0	36
Sherpur	25	11	0	36
Sirajganj	29	7	0	36
Sunamganj	24	12	0	36
Sylhet	29	7	0	36
Tangail	20	16	0	36
Thakurgaon	21	15	0	36
Total	1,605	593	106	2,304

Table 1a: Number of Sample PSUs, Households and Population covered in HIES 2016 by Residential Division

Division	Number of sample PSUs, household and population		
	National	Rural	Urban
Number of sample PSUs			
Barishal	216	162	54
Chattogram	396	308	88
Dhaka	468	261	207
Khulna	360	236	124
Mymensingh	144	101	43
Rajshahi	288	211	77
Rangpur	288	219	69
Sylhet	144	107	37
Total	2304	1605	699

Number of sample PSUs			
Barishal	4320	3240	1080
Chattogram	7916	6156	1760
Dhaka	9360	5220	4140
Khulna	7200	4720	2480
Mymensingh	2880	2020	860
Rajshahi	5760	4220	1540
Rangpur	5760	4380	1380
Sylhet	2880	2140	740
Total	46076	32096	13980
Number of sample PSUs			
Barishal	17893	13462	4431
Chattogram	34681	27075	7606
Dhaka	37340	21546	15794
Khulna	26960	17478	9482
Mymensingh	11140	7771	3369
Rajshahi	21478	15548	5930
Rangpur	22472	17004	5468
Sylhet	14112	10551	3561
Total	186076	130435	55641