SAMPLE DESIGN FOR WVS 2011 IN ARMENIA

1. BASE APPROACH AND METHODOLOGY

A complete list of Armenian Households (HH) provided to CRRC by the "Electricity Networks of Arme (ENA) company was used as a sampling frame. It contains all households/electricity users in Arme having paid for electricity for the period of December 2010 - February 2011 (701,370 households total), by their residence: regions (marzes) and communities of the country. The selection of samp frame is determined mainly by the fact, that the alternative frame – list of HHs recorded during the l census in Armenia (2001) is out-dated and needs actualization (block listing). It assumes considera time, efforts and additional expenses. Besides, the National Statistical Service (NSS) is not willing share the HH lists with other organizations referring to confidentiality. Instead, the complete list Electricity User Households is up to date and more reliable. It also allows selecting respondents/members of the households to be interviewed, using the adequate well-known methods obtain the age and gender structure of respondents which is close to the national composition[1].

To design the sample of respondents to be surveyed within the WVS 2011 in Armenia **the stratified stage cluster sample using PPS** approach was employed:

.

(a) <u>Stratification</u>: The approach of stratification of households (and, correspondingly, th members: respondents) by the regions/marzes of the country was applied to design the samp The stratification is being considered as preferable option, as it allows ensuring representation all heterogeneity of objective social, economic, cultural and other characteristics of the samp units located in different geographic areas/regions of the country. At the same time, it ensu quite internal homogeneity of the aforementioned characteristics within each stratum. Therefo all 701,370 households in the sample frame were divided into 11 strata by the regional criter (Capital - Yerevan and 10 regions/marzes). At the same time, in each stratum we have second level of stratification made by urban/rural criterion, in order to obtain the urban-ru proportions of the households at the regional and country levels. The proportionate stratificat (PPS) was applied to the sample, which means that (a) the total number of households (a correspondingly: respondents within the households) in each stratum in the sample proportional to the general distribution of households (in the sample frame) by strata, and (b) urban-rural proportion of the households in each stratum in the sample is proportional to general distribution of the households by this criterion. The same principle of the P stratification was applied also within the capital: Yerevan. There are some 7 districts in Yerev which were determined as sub-strata within Yerevan and the total number of households in ea district of Yerevan in the sample is proportional to the general distribution of the households these sub-strata. Thus, in total there are 28 separate strata and sub-strata, of which selection of primary sampling units (PSUs) was carried out at first.

---

[1] The age-gender structure of the population is based on the census data and afterward demograp changes recorded by the NSS is available at www.armstat.am (as of July 1st 20 http://armstat.am/en/?nid=81&id=1263).

**Table 1. The distribution of the electricity user households in the sample frame/ENA 2(
database by strata (Marz and Rural/Urban area)**

| Marz | ENA data | | | | | | NSS official data*: | |
| | Number of households | | | Composition, % of total | | | distribution of de facto population by regions, % in total | The differen between EN h/h distribut and NSS population distribution percentage points |
	Rural	Urban	Total	Rural	Urban	Total	Total	
Aragatsotn	20,457	7,742	28,199	2.9	1.1	4.0	4.2	
Ararat	35,976	15,936	51,912	5.1	2.3	7.4	8.3	
Armavir	31,518	21,086	52,604	4.5	3.0	7.5	8.9	
Gegharkunik	27,755	16,783	44,538	4.0	2.4	6.4	7.3	
Kotayk	27,004	35,271	62,275	3.9	5.0	8.9	8.5	
Lori	23,126	41,006	64,132	3.3	5.8	9.1	8.7	
Shirak	20,358	37,366	57,724	2.9	5.3	8.2	8.3	
Syunik	9,840	22,768	32,608	1.4	3.2	4.6	4.8	
Tavush	17,983	11,905	29,888	2.6	1.7	4.3	1.6	
Vayots Dzor	7,154	4,944	12,098	1.0	0.7	1.7	4.0	
Yerevan, including:	…	265,361	265,361	0.0	37.8	37.8	35.3	
Erebuni	…	*30,769*	*30,769*	…	*4.4*	*4.4*	…	…
Arabkir	…	*56,054*	*56,054*	…	*8.0*	*8.0*	…	…
Saritagh	…	*8,107*	*8,107*	…	*1.2*	*1.2*	…	…
Kentron	…	*22,802*	*22,802*	…	*3.3*	*3.3*	…	…
Mashtots	…	*67,762*	*67,762*	…	*9.7*	*9.7*	…	…
Shengavit	…	*34,315*	*34,315*	…	*4.9*	*4.9*	…	…
Nor Nork	…	*45,552*	*45,552*	…	*6.5*	*6.5*	…	…
Total	221,171	480,168	701,339	31.5	68.5	100	100.0	

**There are two different official estimations of the population number, provided by the NSS (National Statis
Service): de-jure and de-facto population, which are based on the census data. The first one is: 3,266 thousand
the second: 3,035 thousand. The estimations of de facto population obtained from the different surveys are e
smaller. Thus, there are different bases for comparison with the data in ENA. Nevertheless, the distribution
households by strata/regions in ENA database is very close to the regional distribution of de facto population. Th
we definitely can use the regional proportions in ENA for our sampling issues.*

(b) Clusters (PSUs). Two-stage cluster sampling method was applied to the survey sample desi
At the first stage of sampling procedure, approximately equal sized (100 households±
clusters (PSUs) of households, which were formed based on the sample frame were selec
using SRS (simple random sampling) method in each stratum and sub-stratum.

Table 2. The distribution of the clusters (PSUs) in the sample frame/ENA 2011 database by strata

Marz	Number of clusters (of about 100 HH)			Composition, % of total		
	Rural	Urban	Total	Rural	Urban	To
Aragatsotn	205	78	283	2.9	1.1	
Ararat	360	159	519	5.1	2.3	
Armavir	315	211	526	4.5	3.0	
Gegharkunik	278	168	446	4.0	2.4	
Kotayk	270	353	623	3.8	5.0	
Lori	231	410	641	3.3	5.8	
Shirak	204	374	578	2.9	5.3	
Tavush	180	119	299	1.4	3.3	
Syunik	98	228	326	2.6	1.7	
Vayots Dzor	72	49	121	1.0	0.7	
Yerevan,	--	2,654	2,654	--	37.8	
including districts:						
Erebuni	--	_308_	_308_	--	_4.4_	
Arabkir	--	_561_	_561_	--	_8.0_	
Sari tagh	--	_81_	_81_	--	_1.2_	
Kentron	--	_228_	_228_	--	_3.3_	
Mashtots	--	_678_	_678_	--	_9.7_	
Shengavit	--	_343_	_343_	--	_4.9_	
Nor Nork	--	_456_	_456_	--	_6.5_	
Total	2,213	4,804	7,017	31.5	68.5	1(

(c) Households (SSUs). At the second stage of the sampling procedure the secondary samp units, i.e. households (SSUs) were selected in each already selected cluster, using the S method. In each selected PSU 20 SSUs is decided to interview. The households (SSUs) to interviewed within each PSU were selected from the list of electricity users, using the SRS.

(d) Respondents (FSUs). In each selected household (SSU) the respondent/FSU (h/h membe the age group of 18-85) was selected using the recent birthday method.

(e) No quotas were applied in the sampling methodology, because the distribution of population age officially published by NSS is based on 2001 census data and covers de-jure populati The real age distribution of de facto population is not estimated. We understand that using age quotas based on the NSS data will not be reliable. Instead of it, the random selec methods should give an age distribution of population, close to the reality.

2. DETERMINING THE SAMPLE SIZE

At the first stage of sampling, the number of FSUs in the sample, i.e. the sample size has to determined and then PSUs and SSUs have to be chosen. The following mechanism for sample s calculation was applied. At first, the initial sample size was estimated using the following formula:

$$n(1) = t^2 P(100-P)/v^2$$

*(where "v" is the degree of precision or the margin of error: **v=5%, P=50, t=1.96**, assuming that the confidence interval is 95%.)*

Then, the adjustment of the initial sample size was performed, taking into account the finite populat[ion] correction (*fpc*) term and using the number of households in the country in the ENA (unive[rse] population) by the following formula:

$$n(2) = Nn(1)/(N+n(1))$$

(where N is number of universe population, and n(1) is the sample size).

As a result of above-described mechanism, the effective sample size (n(2)) is estimated at 3[84] households/FSUs, in case of SRS design.

Taking into account that the two stage cluster sample design is supposed to be applied, the sample s[ize] is assumed to be corrected by the design effect factor (DEF). The estimated DEF for different survey[s in] the country varies from 1.5 to 3. Taking the DEF estimated at 2.8, which is close to the DEF obtai[ned] from some recent surveys, the sample size is estimated at:

$$n = n(2) \, xDEF=384x2.8=1,075$$

Thus, at the first stage of sampling design, the number of SSUs in the sample was determined. In ca[se] of SRS the quantity of 384 sampling units will ensure the representativeness of the sample at [the] country level (assuming 5% degree of precision and 95% confidence interval). Thus, taking into acco[unt] the DEF, it will be 384*2.8=1,075. Taking into account that the same quantity of 20 respondents has [to] be interviewed in each cluster, the final sample size will be a number, which is the closest to (n) an[d] divisible by 20: 1,100. Thus, the final sample size will be:

$$n = 1,100$$

Each stratum in the sample has to be represented according to its proportion in the total numbe[r of] households in the sample frame. These proportions allow defining the quantity of SSUs in each strat[um] summing up at 1,100 households in total (see the Table 3 below).

Table 3. The distribution of the households/SSUs in the sample by strata

Marz	Number of HHs			Composition, % of total		
	Rural	Urban	Total	Rural	Urban	Total
Aragatsotn	20	20	40	1.8	1.8	3.6
Ararat	60	20	80	5.5	1.8	7.3
Armavir	40	40	80	3.6	3.6	7.3
Gegharkunik	40	20	60	3.6	1.8	5.5
Kotayk	40	60	100	3.6	5.5	9.1
Lori	40	60	100	3.6	5.5	9.1
Shirak	40	60	100	3.6	5.5	9.1
Syunik	20	40	60	1.8	3.6	5.5
Tavush	20	20	40	1.8	1.8	3.6
Vayots Dzor	20	0	20	1.8	0.0	1.8
Yerevan, including districts:	--	420	420		38.2	38.2
Erebuni	--	*40*	*40*		*3.6*	*3.6*
Arabkir	--	*80*	*80*		*7.3*	*7.3*
Sari tagh	--	*20*	*20*		*1.8*	*1.8*
Kentron	--	*40*	*40*		*3.6*	*3.6*
Mashtots	--	*100*	*100*		*9.1*	*9.1*
Shengavit	--	*60*	*60*		*5.5*	*5.5*
Nor Nork	--	*80*	*80*		*7.3*	*7.3*
Total	340	760	1,100	30.9	69.1	100.0

Supposing 20 households in each PSU (the quite common standard for surveys in the country), in t[otal]
55 PSUs have to be selected for the sample at first, using the SRS. The numbers of the selec[ted]
clusters/PSUs will be proportionate to their numbers in each stratum (Table 4). Nevertheless, there [are]
some differences in the proportions of clusters and households in the sample frame and in the sam[ple]
(see Table 2 and Table 4), thus some weights could be applied to the clusters in the survey database [in]
order to obtain their original weights in the sample frame.

Table 4. The distribution of the clusters/PSUs in the sample by strata

Marz	Number of clusters in the sample			Composition, % of total		
	Rural	Urban	Total	Rural	Urban	Total
Aragatsotn	**1**	**1**	**2**	**1.8**	**1.8**	**3.6**
Ararat	**3**	**1**	**4**	**5.5**	**1.8**	**7.3**
Armavir	**2**	**2**	**4**	**3.6**	**3.6**	**7.3**
Gegharkunik	**2**	**1**	**3**	**3.6**	**1.8**	**5.5**
Kotayk	**2**	**3**	**5**	**3.6**	**5.5**	**9.1**
Lori	**2**	**3**	**5**	**3.6**	**5.5**	**9.1**
Shirak	**2**	**3**	**5**	**3.6**	**5.5**	**9.1**
Syunik	**1**	**2**	**3**	**1.8**	**3.6**	**5.5**
Tavush	**1**	**1**	**2**	**1.8**	**1.8**	**3.6**
Vayots Dzor	**1**	**0**	**1**	**1.8**	**0.0**	**1.8**
Yerevan, including districts:		**21**	**21**		**38.2**	**38.2**
Erebuni		*2*	*2*		*3.6*	*3.6*
Arabkir		*4*	*4*		*7.3*	*7.3*
Sari tagh		*1*	*1*		*1.8*	*1.8*
Kentron		*2*	*2*		*3.6*	*3.6*
Mashtots		*5*	*5*		*9.1*	*9.1*
Shengavit		*3*	*3*		*5.5*	*5.5*
Nor Nork		*4*	*4*		*7.3*	*7.3*
Total	17	38	**55**	30.9	69.1	**100.0**

3. WEIGHTING OF THE DATA

The weighting of the data was carried out after the database was created. The data weighting was
aimed at the correction of following issues:

(a) At first, in the sampling methodology it was supposed that the equal size clusterization ([?]
households in each) will be applied to the sample frame in each stratum (including sub-strata)[.]
practice, the approximately (±10 households) equal size clusters were created in order to av[oid]
including of more than one settlement/community in the same cluster. In the all general sam[ple]
frame comprising of 7,017 clusters 224 not equal sized clusters were created and only 5 of th[em]
were randomly selected for the sample. In these clusters the selected househo[lds]
representativness is not equal to the others. Thus, the weighting was carried out to equa[lize]
them, using the following formula.

$$W\_1_{ij}=100/n_{ij},$$

where n – is the number of the households in the cluster in the sample

i –number of strata, j – number of substrata

(b) The proportions of the selected households by strata (including sub-strata) in the sample slight differ from the same proportions in the sample frame (as a result of taking the same quantities households in each selected cluster). In order to come up with the original proportions households at two levels of strata, the appropriate weights were applied, using the follow formula:

$$W\_2_{ij}=q\_1_{ij}/q\_2_{ij}$$

where q_1 is the proportion of households in the sample frame (general population)

q_2 is the proportion of households in the sample, i –number of strata, j – number of substrata

(c) Although the selection of the respondents was carried out using the recent birthday method, gender distribution of the respondents is significantly deviating compared to official statistics sample level and at the two levels of stratification. Partially it can be explained by the refusals absence of male population in the course of fieldwork visits. In order to come up with the gen distribution of respondents corresponding to the national average, the appropriate weights w applied at strata and sub-strata levels, using the following formula:

$$W\_6_{ij}=q\_4_{ij}/q\_3_{ij}$$

where q_3 is the gender proportion of population in the country (general population)

q_4 is the gender proportion of respondents in the sample, i –number of strata, j – number of substrata

(d) IN order to summarize the correction of these issues a composite weight was calculated for WVS database using the following formula[2]:

$$W\_7=W\_1*W\_2*W\_6$$

4. WEIGHTING OF THE DATA

After weighting of the data a variable for spreading the data (in case of the need) on the total 18-85 group population was created (spread_1). The population numbers for the population in the mention age group obtained from the official statistics (estimations for 2010 based on 2001 census) w considered as a basis. Then these data were adjusted based on the ENA database regional/str proportions, taking into account that the ENA database is more updated and realistic.

---

[2] It should be mentioned, that no weights were applied to the age distribution of the responden because the official data is not reliable and the survey may yield more realistic picture. The official d are the estimations of NSS based on 2001 census data and refer to the de jure population, while surveys deal with the de facto population. There are the estimates of de facto population in Arme (NSS) at country and regional levels (also based on the 2001 census data) but the age distribution of not available.