Center for Social and Marketing Researches 'Expert'

Sample Design

2011 Uzbekistan National Survey
of Adult Population

Tashkent - 2011

CONTENTS

A. REPORT ON SAMPLE DESIGN

Target population - All population in Uzbekistan aged 18 and over

Sample size - 1 500 adults

A.1. ADMINISTRATIVE-TERRITORIAL DIVISION OF UZBEKISTAN

Uzbekistan consists of 12 provinces, the Republic of Karakalpakstan and Tashkent city metropolitan area. Hereinafter, we will be referring to all these 14 units as provinces. The provinces in turn are comprised of several districts each. Totally, there are 168 districts in Uzbekistan. At last, each district incorporates a number of cities, small towns and villages. Among 233 cities and small towns in Uzbekistan, there are 76 cities which are subordinate directly to the province of location due to their particular importance.

The population of Uzbekistan amounted to 25 523 000 people, among them the urban part was 9 410 700 (37%), and the rural one was 16 112 300 (63%)[1].

A.2. SAMPLING FRAME

The sampling frame is the list composed of small territorial units (primary sampling units – PSUs) of the following types:

- MK ("Mahallia Komiteti") - town makhalla committee. Makhallas are the traditional neighborhood committees which have been revived (and in some urban areas created) by the government;

- SSG ("Selskiy Skhod Grazhdan") - village council. This type has been used for rural areas in all recent surveys.

Remarks

This sample design had been elaborated and approved by Dr. Lesly Kish (Michigan University) personally when he went our Center.

Before 1998 we were using four-stage sampling probability models. The first stage was the selection of districts, at the second, third and fourth stage population points, households and individuals were sampled respectively. It is known that reducing a number of sampling stages leads to more reliable data (with a less standard error and confidence interval). Therefore, we needed PSU to be much smaller than a district. It should have been a territorial unit within which we could make direct selection of households.

Besides MK and SSG, we investigated several else but rejected them. We tried to cover areas in Uzbekistan with diverse characteristics in order to get a reasonably complete picture of all the PSU options in sense of their availability, coverage, overlaps, foreign elements and other frame problems. Below are the rejected options:

- municipal list of customers served by the Electricity Board;
- postal districts, covered by the local post offices;

---

[1] "Statistical yearbook of the regions of Uzbekistan - 2002", Tashkent, 2003

- electoral districts;
- local passport offices;
- medical service zones.

Below is a brief description of the 2 types of units, which constitute the sampling frame.

MK

The MK has good up-to-date lists of households. The geographical boundaries of each MK are clearly defined, which excludes territorial overlap. Approximately once a year the makhalla committee representatives go round the households in order to revise the lists. The lists are also updated when members of the household apply to the MK for certificates. The MK is usually of an appropriate size, but in some cases we have to merge two or more MK in order to get a new unit with the overall number of households increased.

SSG

In the SSG the situation is similar to that in the MK. The SSG comes from the village soviets, which existed in the Soviet period. Traditionally, lists and accounting are quite detailed and accurate. For convenience, we use the word "SSG" instead of villages and rural makhallas. In the report MK and SSG are sometimes called "makhalla".

We have entire list of MK and SSGs updated annually.

A.3. GENERAL SAMPLING SCHEME

The sampling scheme has the following stages:

- proportionate stratification by population of provinces;
- for all provinces (include *Tashkent city* as urban stratum):
  - proportionate stratification by urban/rural population within provinces;
  - PPS-sampling of PSUs within urban/rural strata (see A.2 above for explanation of PSU types);
- sequential random sampling of households (Secondary Sampling Units - SSUs) in selected PSUs;
- The «latest birth day» (in this survey) or the Kish grid.

Thus, the sampling is three-stage stratified clustered sampling.

A.3.1. Proportionate stratification by population of provinces

There are 14 strata formed. All interviews (1 500) are proportionally allocated among the strata, according to the share of population in each stratum. Thereby, we attain the following advantages:

- the strata-provinces represent relatively homogeneous, in sense of the survey questions, groupings of units which makes the resulting sample more efficient, i.e. leads to reducing sampling error as compared with the entire unstratified sample;
- stratification ensures inclusion of all the provinces into the sample with certainty, not relying on chance;

- it permits to treat each province as a "population" in its own right, when data are required separately for provinces, and select a sample of the required size and design from each independently.

A.3.2. Proportionate stratification by urban/rural population

There are 27 strata formed on the base of the province strata. Each province is split into urban and rural. Interviews are proportionately allocated according to the share of the urban/rural population:

A.3.3. PPS-sampling of PSUs

After the urban/rural stratification the entire list of PSUs is composed within each stratum. A certain number of PSUs is sampled with PPS, that is, the probability to be sampled depends on their size (number of households, families or addresses – which is available). The PPS-sampling is carried out as follows:

- the list of PSUs is sorted by their size;

- cumulative column is calculated next to "Size of PSU";

- the sampling interval F is calculated as the total size of all PSUs in the list divided by a number of PSUs to be sampled;

- then the random start Rnd is fixed within the range of (1-F);

- beginning from Rnd and each time adding F, one systematically selects those PSUs, which cumulative numbers enclose the current number.

Adhering to PPS-sampling sample allocation, an approximately equal number of interviews is allocated among selected PSUs of the same stratum.

The number of PSUs to be sampled depends on a minimum number of interviews to be achieved per PSU. It also considers travel and other costs of data collection, supervision, control and follow-up aspects, as well as minimum effective workload per PSU. Below we introduce the number of primary selections (PS) within each stratum and averaged allocated sample size:

Totally, 75 PSUs are selected from the sampling frames (see **¡Error! No se encuentra el origen de la referencia.** for the complete list of the selected PSUs).

Remarks

The size of each PSU must be sufficient with regard to the following three terms:

1) it should be equal, at least, to the quota per PSU, since the next stage is the sampling of households *without replacement*;
2) possible frame imperfections should be considered (e.g., foreign elements);
3) a supplement for substitutions due to non-response should also be taken into account.

Taking into account the possibility of non-response and its rate in previous surveys, the PSU size is increased up to two times.

Sufficient PSU size: $N = n + n_f + n_r$, where n – the PSU quota; n_f – supplement size due to frame imperfections, or "bad" frame elements; n_r - supplement size due to non-response, or non-response elements.

Replacing n_f and n_r with the minimum expected frame quality rate $r_f = (N - n_f)/N$ and minimum expected response rate $r_r = n/(n + n_r)$, we obtain:

$$N = n + N(1 - r_f) + n(1 - r_r)/r_r = n/(r_f r_r).$$

To avoid selection problems during the fieldwork, before the PPS-sampling we have to enlarge those PSUs, which do not meet the terms (urban: $r_f = 0.5$, $r_r = 0.6$, $N \approx 100$; rural: $r_f = 0.5$, $r_r = 0.8$, $N \approx 75$). For this purpose, undersized PSUs of the same territorial unit are merged into a single PSU (see PSU No. 33 in the list of selected PSU **¡Error! No se encuentra el origen de la referencia.**).

A.3.4. Sequential random sampling of households

The root method will be used in this survey.

A.3.5. Sampling of respondents

The «latest birth day» will be used in the WVS survey

For applying and checking during the fieldwork, interviewers are provided with the Visit Registration Form enclosed with each questionnaire (see A.8).

A.4. The List of Selected PSUs

List pf PSUs and quotes of interviews

№	Region	Raion	Urban	Rural	PSUs' name	Quote per PSU
1	Каракалпакстан		г. Нукус		№14	20
2		Берунийский		Шимом	Чинобод	20
3		Берунийский		Шаббаз	Аскар	20
4		Муйнакский		Казакдарья	Кенес	20
5	Андижанская		г. Андижан		мах.Уйгурбобо	20
6			г. Андижан		мах.Навруз	20
7		Андижанский		Найман	Чавкон-дорё	20
8		Балыкчинский		Бустон	Нартепа м.ф.й.	20
9		Мархаматский		Шукурмерган	Юввош м. ф.й.	20
10		Алтынкульский		Сувюлдуз	Кораянток	20
11	Бухарская		г. Бухара		мах.Турки жонди	20
12		Бухарский		Кунжикалъа	Урта Новметан	20
13		Каганский		Каган	Навбахор	20
14		Вабкентский		Кумушкент	Ш. Бешработ	20
15		Гиждуванский		Сарвари	Сайдали	20
16	Джизакская		г. Джизак		мах.Г.Гулям	20
17		Дустликский		Навруз	Навруз	20
18		Мирзачульский		Богбан	Богбон	20
19	Кашкадарьинская		г. Карши		мах. Кургонча	20
20			г. Шахрисабз		мах.Оксарой	20
21		Каршинский		Ковчин	Ковчин к\к	20
22		Яккабагский		Кушчинор	к\к Олакуйлак	20
23		Каршинский		Чулибустон	к\к Чули Бустон	20
24		Камашинский		Ертепа	Хонтушди к\к	20
25	Навоийская		г. Навои		мах.Истиклол	20
26		Хатырчинский		Хонака	Навандак	20
27	Наманганская		г. Наманган		мах.Янги тонг	20

28		Янгикурганский	г.п. Янгикурган		мах. Ташкент	20
29		Туракурганский		Шаханд	Кошкирбузар	20
30		Касансайский		Косон	Озод	20
31		Уйчинский		Гайрат	Иттифок	20
32		Чустский		Варзин	Янгичак	20
33	Самаркандская		г. Самарканд		мах.Тошканди-2	20
34			г. Самарканд		мах.Самарканд	20
35			г.п. Кимегарлар		М.Зохидов	20
36		Пастдаргомский		Хамзаобод	Кукони	20
37		Пастдаргомский		Арабхона	Зангибойи	20
38		Иштыханский		Уртакишлок	Култепа	20
39		Иштыханский		Уртакишлок	Гузон	20
40		Самаркандский		Кайнама	Ипакчи	20
41	Сурхандарьинская		г. Кумкурган		мах.Беруний	20
42		Кумкурганский		Арисланбойли	Саробоб кишлоги (Арсланбойли)	20
43		Ангорский		Дустлик	Охунбобоев кишлоги (Охунбобоев)	20
44		Алтынсайский		Вахшивор кишлок фук. Йигини	Хайрондара кишлоги (Хайрондара)	20
45		Шерабадский		Сепилан	Пахтакор кишлоги (Бойбулок МФЙ)	20
46	Сирдарьинская	Гулистанский	г.п. Дехканабад		мах.Узбекистан	20
47		Гулистанский		Бешбулок	Бешбулок махалласи	20
48	Ташкентская		г. Ангрен		мах.Озодлик	20
49		Паркентский	г. Паркент		мах.Ойбек	20
50		Зангиатинский		Назарбек	Назарбек	20
51		Ахангаранский		Карахтой	Охунбобоева	20
52		Зангиатинский		Узгариш	Паст-Дархон	20
53		Куйичирчикский		Тошлок	Эркин	20
54		Кибрайский		Ункурган	Янгийул	20
55	Ферганская		г. Фергана		мах.Бобур	20
56			г. Маргилан		мах.Навруз	20

57		Узбекистанский	г. Яйпан		мах.Охунбобоев	20
58		Ахунбабаевский		Ахунбабаев	Янгиарик	20
59		Учкуприкский		Янгикишлок	Тургок	20
60		Алтыарыкский		Повулгон	Янгичек	20
61		Ферганский		Хонкиз	Ок олтин	20
62		Алтыарыкский		Алтиарик	Зархал	20
63		Фуркатский		Гунча	Гунча	20
64	Хорезмская		г. Ургенч		мах.8-Сон	20
65		Кошкупырский		Хасиян	Тегалак	20
66		Шаватский		Хитой	Хитой	20
67		Ханкинский		Навхас	Шодлик	20
68		Ханкинский		Тамадургодик	Навруз	20
69	г.Ташкент		г.Ташкент		мах.Обод	20
70	г.Ташкент		г.Ташкент		мах.Шухрат	20
71	г.Ташкент		г.Ташкент		мах.Намуна	20
72	г.Ташкент		г.Ташкент		мах. Белтепа	20
73	г.Ташкент		г.Ташкент		мах.Ибн Сино	20
74	г.Ташкент		г.Ташкент		мах.Байналминал	20
75	г.Ташкент		г.Ташкент		мах.Богишамол	20

Total Interviews   1500

A.5. ACHIEVED SAMPLE <mark>(AN EXAMPLE FROM PREVIOUS SURVEYS)</mark>

A.5.1. Comparison to census

Comparison to census is not valid because the followings:

- the last census was conducted in 1989;

- unregistered migration of labor migrants from Uzbekistan estimated on the level of 5,6 mln. people;

- Government declares about registered population only.

A.6. NON-RESPONSE

During the fieldwork, 766 cases of non-response were registered (non-eligible units are excluded from this count). The average response rate is about 66% (1 500 of 2 266 attempts). Generally, the non-response case was registered if an interviewer had made up to two failed callbacks. Below is listed the response rate by residence:

Table 1.   Response rate by urban/rural and provinces (example)

No	Residence	Response	Non-response	Response rate
	Urban	553	432	56.1%
	Rural	947	334	73.9%
	Provinces			
1	Karakalpakstan	91	20	82.0%
2	Andijan	134	41	76.6%
3	Buhara	87	20	81.3%
4	Jizzakh	61	11	84.7%
5	Kashkadarya	135	56	70.7%
6	Navoi	47	24	66.2%
7	Namangan	119	50	70.4%
8	Samarkand	164	110	59.9%
9	Surhandarya	108	28	79.4%
10	Syrdarya	39	11	78.0%
11	Tashkent	143	79	64.4%
12	Fergana	163	66	71.2%
13	Horezm	82	40	67.2%
14	Tashkent city	127	210	37.7%
	TOTAL in Uzbekistan	1 500	766	66.2%

* High non response rate can be explained by labor migration within country and outside of the country.

As one can see, the response rate in rural areas is higher than in urban areas. In Tashkent city very much high level of refusals is observed (response rate barely about 38%). This is caused mainly by the following factors:

a) rural residents are more willing to cooperate;
b) they are less active in sense of movement, therefore more reachable;
c) the theme of interview sets people on the alert;
d) population registration and register maintenance in cities are generally worse which leads to poor quality sampling frames.

The influence of first two factors is aligned lately because of a falling of a scale of living of people.

The table below showing the structure of non-response proves these assumptions.

Table 2.   Non-response structure by causes (example)

No	Cause of non-response	Urban		Rural		TOTAL	
		Freq.	%	Freq.	%	Freq.	%
3	Nobody at home during 3 visits	39	8.2%	20	5.1%	59	6.8%
4	Respondent was not at home by that time during 3 visits	13	2.7%	35	9%	48	5%
6	Emphatic refusal by respondent	9	1.9%	9	2.3%	18	2.1%
7	Household members refused contacting respondent	191	40.0%	122	31.2%	313	36.1%
8	Respondent was drunk			1	0.3%	1	0.1%
9	Respondent could not talk (sick, abnormal, very old, etc.)	7	1.5%	17	4.3%	24	2.8%
10	Not at home for a long time (long absence)	175	36.7%	151	38.6%	326	37.6%
11	Address was not found, does not exist	23	4.8%	25	6.4%	48	5.5%
12	Address is not residential	20	4.2%	11	2.8%	31	3.6%
	TOTAL	477	100.0%	391	100.0%	868	100.0%

In this table are included the causes of non-response owing to a non-eligible units (causes 11,12) and technique of sampling of households (cause 13, see also A.3.4).

Thus, 40% of all the causes in the urban areas is the "household members refused contacting respondent" (cause 7), as compared with the corresponding 31.2% in the rural areas. This cause has the most spread for urban people and the second at the prevalence for rural areas (about 31% of all causes of non-response), because the theme of interview (the internal politic, interethnic problem etc.) makes people mistrustful and situation with the criminality (especially in the cities) is very complicated.

Otherwise, cause 10 ("not at home for a long time") is second at the prevalence for urban areas (about 37%) and first for rural areas (about 39% of all non-response causes). This cause is spread for urban and rural people because they migrate in searches of earnings.

The similar reasons called cause 3 "nobody at home" and 4 "respondent was not at home by that time" (8.2% and 2,3% for urban and 5.1% and 3.6% for rural areas accordingly). Besides for these causes there is one more explanation – employment of urban population and "cotton campaign" for rural population.

The causes 6, 8, and 9 met not frequently. Therefore we may not make any conclusions.

The sampling frame quality is revealed by comparing the share of cause 11 "address was not found, does not exist"– 4.8% in the urban areas versus 6.4% in the rural. In the urban areas

2.8% of the non-response are "Address is not residential" (cause 12). In the rural areas this cause makes 4.2% of all causes of non-response. In most cases it originates from that a household, in order to get an additional land plot from a makhalla committee for running subsidiary economy, declares itself to be actually consisting of two households – parents' and a new, young one. Then the makhalla committee registers a new household and allocates a plot. However, this "household" continues living with the parents, making the new address not residential. Most urban cases are connected with fitting apartments for small offices, cafes, renting to foreigners, etc. More apartments in the cities are thrown (owners have left in searching of earnings).

A.7. FORMULAS

A.7.1. Probability of selection

As the stratification by provinces and urban/rural is proportionate, the sampling rate is equal across the strata.

The probability of selection of the αth PSU in the hth stratum is

$$P_{h\alpha} = \frac{a_h \cdot MOS_{h\alpha}}{\sum MOS},$$

where a_h is the number of primary selections in the hth stratum; $MOS_{h\alpha}$, the measure of size, is the expected number of households in the αth PSU; $\sum MOS$ is the sum of the expected household numbers across all PSUs in the stratum.

The probability of selection of any household in the αth PSU is uniform, for the βth household being equal to

$$P_{h\alpha\beta} = b_{h\alpha} / B_{h\alpha},$$

where $b_{h\alpha}$, the subsample size per PSU, is the number of interviews in the αth PSU; $B_{h\alpha}$ is the actual number of households in the αth PSU.

The combined probability of the two stages is the product of $P_{h\alpha}$ and $P_{h\alpha\beta}$:

$$P*_{h\alpha\beta} = \frac{a_h \cdot MOS_{h\alpha}}{\sum MOS} \cdot \frac{b_{h\alpha}}{B_{h\alpha}}.$$

In general, $MOS_{h\alpha} = B_{h\alpha}$ and $b_{h\alpha} = b_h = n_h / a_h$, where n_h is the number of interviews for the hth stratum. Noting, that $\overline{HS}_h = N_h / \sum MOS$, where N_h is the number of individuals in the hth stratum and $\overline{HS}_h$ is the stratum average household size, we obtain:

$$P*_{h\alpha\beta} = \frac{n_h}{N_h} \cdot \overline{HS}_h.$$

The probability of selection of the respondent in the βth household is inversely proportional to the number of eligible persons in it $HS'_{h\alpha\beta}$. Then the sampling fraction f is

$$f = \frac{n_h}{N_h} \cdot \frac{\overline{HS}_h}{HS'_{h\alpha\beta}} = f_h \cdot \frac{1}{ER_{h\alpha\beta}}, \text{ where } ER_{h\alpha\beta} = \frac{HS'_{h\alpha\beta}}{HS_h} \text{ is the eligibility rate.}$$

A.8. VISIT REGISTRATION FORM

Name of interviewer | No

Province | No

District | No

City/town | No

SSG | No

PSU (mahalla/village) | No

Address of HH/Name of HH head | No

Table of visits:

Visit No	Date of visit	Time of visit	Result of visit
1			
2			
3			

Result codes:

01 Interview completed
02 Interview started but not completed
03 Nobody at home
04 Respondent was not at home by that time
05 Respondent could not afford time, appointed another date
06 Emphatic refusal by respondent
07 Household members refused contacting respondent
08 Respondent was drunk
09 Respondent could not answer (sick, abnormal, very old, etc.)
10 Not at home for a long time

11 Address not found, does not exist
12 Address is not residential
Other (SPECIFY): _____

Attention!
RESPONDENT IS TO BE SELECTED ACCORDING TO THE KISH GRID
(see the reverse side)