

# Approaches to Sample Designing for Public Educational Institutions in Ukraine

Vashchaieva Nataliia

Institute for Demography and Social Research, Ukraine  
e-mail: [n\\_vashchaieva@ukr.net](mailto:n_vashchaieva@ukr.net)

## Abstract

In the article questions of development of effective sample design for survey of general educational institutions and pupils in Ukraine are considered. Thus, the external information of the state social and demographic statistics, the administrative data is taken into account, necessary territorial scope and level of the data reliability is provided.

The basic information source which allow to estimate real condition of the education system, to develop effective policy of its reforming and to estimate efficiency of corresponding actions is the statistical reporting and the Ministry of Education and sciences administrative information. However, the data from these sources it is not enough to estimate quality of education and school results level. In such conditions the role of special surveys are grows. Among them sample survey which allow to receive reliable results at the minimal expenses are allocated. At use of sample survey in sphere of social researches especially actual there is a problem of development of effective design of sample. It should take into account features of population, provide a necessary level of the data reliability.

At carrying out of survey in system of secondary education of Ukraine as population, depending on the survey purpose, can be population of establishments or population of pupils. Also these two populations can simultaneously be investigated.

For construction of effective sample design important in details to represent population structure. Let's consider the basic characteristics of schools population.

General secondary schools make the basic part of schools. The share of usual schools by the beginning of 2004-2005 made 88,7 % of the state day time schools general number, a share of pupils in them - 86,8 %. The share of the new type schools (for example, gymnasiums and lyceums) made 11,3 %, a share of pupils - 13,2 %. Almost all such elite schools are located in urban areas, while about 70 percent of general secondary schools are rural. Thus essential distinctions of the specified characteristics on regions were observed. In urban areas is 31,2 % of usual schools and 65,2 % of pupils, and in a rural areas - 68,8 % of institutions and 34,8

% of pupils (fig. 1). Thus, in a rural areas there is a most part of schools, but the basic part of pupils is concentrated in urban areas.

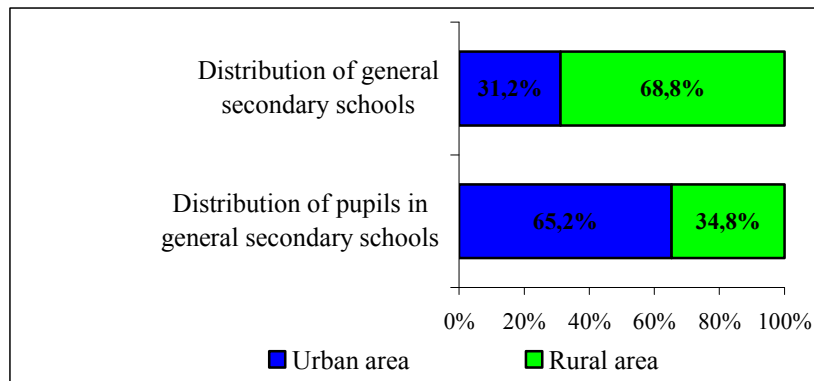


Fig. 1. Distribution of general secondary schools and pupils in them by area

Prevailing part general secondary schools belongs to primary and upper-secondary school - 58,3 % on the beginning 2004/05 – figure 2.

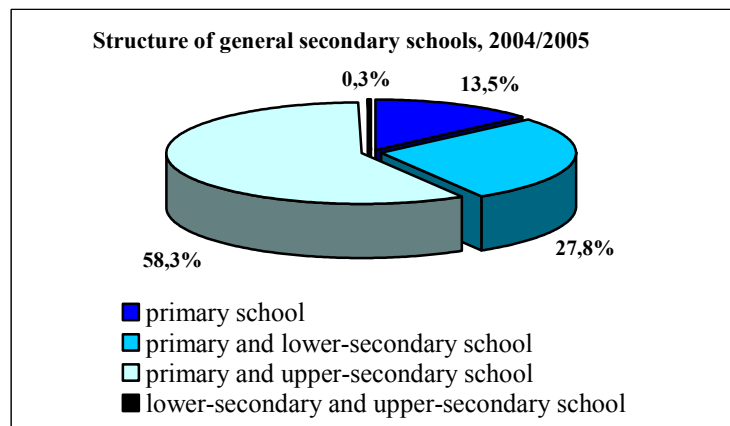


Fig. 2. Structure of general secondary schools, beginning of 2004/2005.

On an example of primary and upper-secondary schools it is visible, that the share of schools and a share of pupils on area types was distributed disproportionate (figure 3, 4).

Besides, population of schools is non-uniform by language of instruction. As a whole across Ukraine establishments with the Ukrainian language of instruction prevail, but in regional distribution there are essential differences. For example, in the Lviv region 98,8 % of pupils are trained in the Ukrainian language, and in Crimea - 96,5 % in Russian.

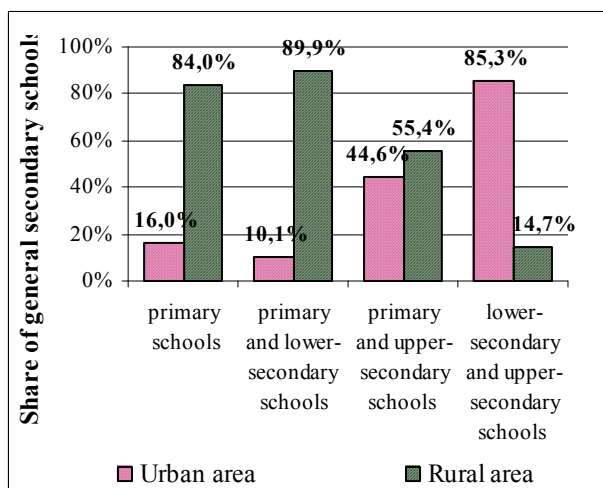


Fig. 3. Distribution of general secondary schools by area, 2004/05 н.р.

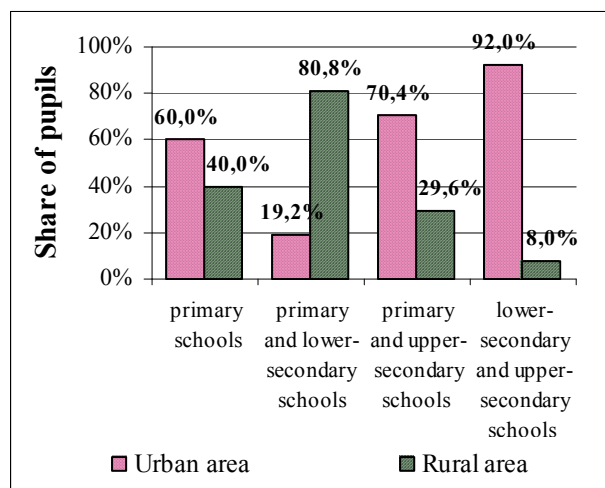


Fig. 4. Distribution of pupils of general secondary schools by area, 2004/05 н.р..

So, population is non-uniform under such characteristics as region, type of area, school type, language of instruction. Therefore, depending on the program of survey sample, the effective sample design should take into account features of general secondary schools population.

It is important to note the following. At survey of separate aspects of education quality there can be expedient an association of the received information with the data of other surveys. For example, estimations of population incomes and expenditures, poverty, so living conditions too influence of training conditions. It demands the account sample designs of the other surveys at the survey organization in the education system. Household incomes and expenditures are measured in Ukraine at survey of household living conditions.

The sample design created for concrete survey, depends on the purpose of survey, requirements to the data reliability, a way of the information gathering, the financial resources, existing information base. At carrying out of survey in the education system on the basis of questioning by interviewers of institution workers and pupils sample of usual schools is expedient for forming as probabilistic, territorial, stratified.

From the previous surveys proceeds, that at stratification the sample is expedient for distributing by regions. Among regions to allocate type of settlement - cities and a countryside that will provide adequate representation in sample of all territories. The variant of stratification when in regions allocate the big cities with a population more than 100000 person, small cities with the population less than 100000 person and administrative areas in a rural areas is possible. Within the limits of area's type the list of schools on each settlement included in sample is formed, and strata are allocated. For example, by school's type (general secondary schools and new type), by a language of instruction (Ukrainian or Russian).

Depending on the survey purpose the sample size can be distributed on strata by the certain quantitative attribute. For example, it is proportional to number of pupils if it is necessary to survey pupils, or it is proportional to number of institutions if institutions are surveyed.

If are surveyed both pupils and institutions, the researcher should solve how to distribute sample by strata. Thus it is necessary to take into account, that it can influence on design-effect. The design-effect characterizes the variance ratio of the certain parameter estimate to real sample design to a variance of the same parameter estimate at simple random sample (C-E. Särndal, 1992 and Kish, 1995). For an estimate of influence effect of sample distribution on the strata, generated by number of pupils, on variance size of selective parameters estimate calculations have been carried out with an assumption, that the variance of each parameter within the limits of strata is identical.

Under the above-stated conditions and without the amendment on set completeness the ratio of variance at disproportionate sample distribution to a variance at proportional distribution can be submitted as:

$$eff = \frac{V_{nprop}(\bar{y})}{V_{prop}(\bar{y})} = \frac{\sigma_w^2 \cdot \sum_{h=1}^H \frac{W_h^2}{n_{h(nprop)}}}{\frac{\sigma_w^2}{n}} = \frac{\sigma_w^2 \cdot \frac{1}{n^2} \cdot \sum_{h=1}^H \frac{n_{h(prop)}^2}{n_{h(nprop)}}}{\frac{\sigma_w^2}{n}} = \frac{\sum_{h=1}^H \frac{n_{h(prop)}^2}{n_{h(nprop)}}}{n}, \quad (1)$$

where  $eff$  - influence effect of a way of sample distribution by strata on sample variance;

$V_{nprop}(\bar{y})$  - sample variance at disproportionate sample distribution;

$V_{prop}(\bar{y})$  - sample variance at proportionate sample distribution;

$\sigma_w^2$  - variance of  $y$  - attribute on sample units;

$W_h = \frac{N_{h(prop)}}{N} \approx \frac{n_{h(prop)}}{n}$  -  $h$  - weight of strata  $h = 1, 2, 3, \dots, H$  - total strata number;

$n_{h(nprop)}$  - sample size at disproportionate distribution;

$n_{h(prop)}$  - sample size at proportionate distribution;

$n$  - total sample size.

Suppose, that the total sample size for survey of pupils school results level is made by 7200 person which are distributed on settlement on 7 regions of Ukraine proportionally to number of pupils. Provided that the variance of estimates of results level within the stratum is identical, at distribution of sample proportionally to number of schools  $eff$  makes 1,4 (see (1) and table 1).

It is necessary to note, that the value of  $eff$  can be and less than 1 if to compare variances at proportional and optimum layout.

**Table 1 - Distribution of sample size of pupils by strata (according to social statistics by the beginning of 2004/2005)**

Region	Area type	Number of stratum	Weight of stratum, %	Proportionally number of pupils, pers	Proportionally number of schools, pers	$n_{h(prop)}^2$
						$n_{h(nprop)}$
Autonomous Republic of Crimea	urban	1	6,9	498	290	855,8
Donetsk region		2	19,3	1387	955	2013,5
Kyiv region		3	6,4	464	246	875,4
Lviv region		4	9,6	694	388	1240,8
Odesa region		5	8,2	588	317	1089,1
Kharkiv region		6	11,0	791	457	1369,0
Kyiv city		7	12,8	919	536	1573,6
Autonomous Republic of Crimea	rural	8	4,1	296	394	221,9
Donetsk region		9	2,3	163	377	70,1
Kyiv region		10	4,1	298	646	137,2
Lviv region		11	6,7	480	1257	183,2
Odesa region		12	5,6	404	727	224,8
Kharkiv region		13	3,1	220	611	79,3
<b>Total</b>	-	-	<b>100,0</b>	<b>7200</b>	<b>7200</b>	<b>9933,6</b>

The sample design realized in survey causes using for construction of base weight for each stratum such components:

- probability of the Primary Sample Units selection (urban settlements (city, town, settlements of city type) and rural councils),  $P_{1i}$ ;
- probability of the Secondary Sample Units selection,  $P_{2i}$  (schools);
- probability of classes selection,  $P_{3i}$

The inverse number of product of these three probabilities is the base weight of the pupil. The general probability of selection for all pupils in the certain selected class is identical and equal to the total probability of selection of class as in each class all pupils are surveyed.

Thus, at three-stage formation of sample for survey of education quality the  $p_i$  is determined:

$$p_i = P_{1i} \cdot P_{2i} \cdot P_{3i}, \quad (2)$$

The selection probability of city settlements and rural councils  $P_{1i}$  are determined on each region separately:

$$P_{1i} = a \cdot \frac{M_{psu}}{M}, \quad (3)$$

where  $a$  - number of urban settlements or rural councils, that are selected in the region;  $M_{psu}$  - primary sampling unit size: for urban settlements – number of population, for rural

council – number of households;  $M$  - total population size of urban settlements in urban area of region or total number of households in rural area of region.

The institutions are selected at the second step of selection from the selected city settlements and rural councils with the probability proportional to the total number of pupils. Selection of schools is carried out in each strata separately.

Selection probabilities of schools within the limits of the selected urban settlements and rural councils  $P_{2i}$  are determined separately for general secondary schools and schools of new type:

$$P_{2i} = b \cdot \frac{N_{ssu}}{N}, \quad (4)$$

where  $b$  - number of the certain school's type which is selected in urban settlements or in rural councils;  $N_{ssu}$  - secondary sample unit size - total number of pupils in school;  $N$  - total number of pupils of the school's certain type in the selected urban settlement or in rural council.

If a survey unit - pupils after selection of schools classes are selected. In these classes all pupils are surveyed.

Probabilities of class selection within the limits of the selected schools  $P_{3i}$  are determined under the formula:

$$P_{3i} = \frac{Q_{tsu}}{Q}, \quad (5)$$

where  $Q_{tsu}$  - tertiary sample unit size - total number of pupils in class;  $Q$  - total number of pupils in all classes of school (which represent this class).

For distribution of survey results on population the system of the statistical weights, intended pays off to take into account:

- total probabilities of schools and classes selection;
- actual levels of survey units participation in survey;
- coordination of survey results with the existing data from other sources, in particular the data of the statistical reporting.

Final weight of the  $i$ -pupil calculate as product of the pupil base weight and the certain weight coefficient (reweighting coefficient) under the formula:

$$w_i = w_{Bi} \cdot k, \quad (6)$$

where  $w_i$  – final weight of the pupil;  $w_{Bi}$  – base weight of the pupil;  $k$  – weight coefficient.

In probability sample each pupil has the total probability to get in sample. Thus, this pupil representing  $1/p_i$  of pupils of population. Accordingly at calculations for population it is necessary to take into account  $1/p_i$  times. The value  $1/p_i$  is the pupil base weight for distribution of survey results on population:

$$w_{Bi} = 1/p_i. \quad (7)$$

For a correcting of base weights with the purpose of increase of representing level survey results it is expedient to construct system of cells on such classification variables: region, type of area, pupil sex.

The total number of pupils in population in the certain cell  $N_q$  is determined according to social and demographic statistics about a population size corresponding sex and age groups.

Coefficients for base weights correction  $k_q$  calculate on each  $q$ -cell under the formula:

$$k_q = \frac{N_q}{\sum_{i=1}^{L_q} w_{Bi} \cdot \lambda_{qi}}, \quad (8)$$

$$\text{where } \lambda_{qi} = \begin{cases} 1, & \text{if } i \in L_q \\ 0, & \text{if } i \notin L_q \end{cases},$$

$L_q$  – set of the surveyed pupils of  $q$ -cells,  $N_q$  – set of pupils of  $q$ -cells by the external data.

The final statistical weight for  $i$ -th pupil  $w_i$  calculates under the formula:

$$w_i = w_{Bi} \cdot k_q, \text{ for } i \in L_q. \quad (9)$$

Described sample design was applied for survey of general secondary schools and their pupils at the organization of survey of school results in different regions of Ukraine in 2005. By results of this survey the design - effect for an average level of school results, for example, for Kyiv made 2,0, and on countryside for the Kyiv region - 2,2, for the Kharkiv region - 4,0, that is sufficient for regional level. Therefore it is possible to count, that the sample design is effective. This design takes into account the external information, reflects the basic features of population, provides a sufficient level of reliability of the survey data and reduces expenses for formation of sample. Results of the research testify, that such sample design can be used at the organization of sample survey in Ukraine education system.

It is necessary to notice that during last few years sample surveys are more and more used in Ukraine for measuring of student's knowledge level, qualification of teachers, level of material support of schools and overall quality of education. But in organization and conducting of such surveys there are some problems basic from which are:

- absence of actual sample frames of educational institutions;
- absence of clear determinations of types of educational establishments, in addition, not enough concerted with international classifications;
- absence of information about a residence of students in rural schools and schools of cities.

It follows to notice also, that on the whole not enough attention is given to determination of reliability of information and influence on it of sample design. In addition, indicators of quality which characterize relevance, timeliness, accessibility, comparability and coherence of information are not measured practically.

## **References**

Carl-Erik Särndal, Bengt Swensson, Jan Wretman (1992) *Model Assisted Survey Sampling*. Springer, New York.

Kish L. (1995) *Survey sampling*. Wiley, New York.