MICRO-ENTITIES SAMPLE SURVEY DESIGN IN BELARUS (PRELIMINARY VERSION)

Natallia Bokun,

Helsinki, 2015

SMALL BUSINESS IN BELARUS

According to the legislation **small enterprise (SE)** is the organization with the number of employees 100 persons and less, **micro-entities** – 15 persons and less

SMALL BUSINESS IN BELARUS

Main indicators of micro and small entities

Indicators	2000	2005	2010	2012	2013	Of which	
						micro-entities	SE
Micro- and small entities	28310	32243	74246	82612	91596	79800	11796
Average number of employees, persons	385829	456180	784944	806556	846699	384960	461739
Nominal gross average monthly wages and salaries, rubles	48053	346329	1031658	3270303	4603.3	3634.4	5356.6
Revenues from sales of goods, products, services, rubles	4126.6	31768.4	126388	413319.2	477278.4	176346.0	300932.5

SMALL BUSINESS IN BELARUS

Main indicators of micro and small entities

Indicators	2000	2005	2010	2012	2013	Of which	
						micro-entities	SE
Percentage share of loss-making entities	21.9	26.9	21.0	18.6	20.6	20.8	19.0
Profitability of sales, %	4.7	3.7	5.3	8.3	6.9	8.2	6.2
Volume of foreign trade in goods, bln USD	…	6972.9	20279.4	26314.0	22341.4	6068.8	16272.5
- exports	…	2118.8	9820.5	13776.0	9758.4	903.2	8855.2
- imports	…	4854.1	10458.9	12538.0	12583.0	5165.7	7417.3

SAMPLE SURVEY IN THE SMALL BUSINESS STATISTICS

Since 2005 and until 2008 sample surveys of small enterprises spent quarterly. Survey objects were SE. Sample frame was the file of SE. The territorial one-stage stratified sample was used. But in 2008 quarter survey was cancelled. Due to this reason only annual continuous small enterprises survey is conducted.

In 2014 National Statistical Committee of the Republic of Belarus and Department of Statistics (BSEU) began to develop the methodology and specialized software for micro-entities sample survey. In November 2014 a test sample survey was conducted; since 2015 MSS is provided on a regular basis. The main survey aims are:

- to obtain summary statistics on the average number of employees, the volume of manufactured products, the revenues from sales of goods, products, works, services and other indicators of micro-entities by economic activity and regions;
- evaluation of gross output and gross value added by micro-entities;
- frequency of the results: annual.

SAMPLING PROBLEMS

- *Non-responses*. The population of micro- and small enterprises is extremely dynamical: the creation of new entities, liquidations, changes in kinds of activity and size of the enterprises are taken place constantly. Sampling frame is based on the data of the previous year of complete survey, and not responded enterprises can be included into the sample
- *Atypical units (outliers)*, i.e. presence in the frame of atypical units
- *Samples in small domains*. Construction samples of small enterprises by economic activity and regions, in some cases is connected with partition of survey population into the small groups and sample fractions become unacceptably high (50 – 60 %)
- *Problems of compromise* between the accuracy requirement for various groups caused by stratification and restrictions on sample size

SAMPLING PROBLEMS

- *Estimation*. The problem of estimation still persists when the univariate stratified sample with admissible standard error and sampling fraction is built. Weights, raising factors allow to estimate precisely enough values of the parameter which was used for sample selection, but other estimates which number can rich 10-30 are of a low quality
- In the case of multivariate sample, the error for some group of indicators will be in admissible limits (to 10 %), but will be considerably above comparing with the case of univariate sample
- *The problems of the software* are caused by the necessity of integration of the sample survey programs in the general system of collection and processing of statistical data
- Specific problems are met designing the *multivariate sample*: complexity of a choice of an optimal way of multivariate selection, complexity of a choice of a leading indicator (variable), absence of the standard estimation methods
- The problems of non-responses and atypical units may be solved within traditional univariate sample (the change of general population structure, separate files of atypical enterprises, use of weighting or replacement procedures). Multivariate sampling and different weighting schemes are used to handle remaining problems

UNIVARIATE SAMPLING

- Simple random sample, provides objectivity of selection, makes a basis of the majority of more difficult methods of selection. Possibilities of its application are in the pure state limited: absence of uniform distribution of units on set, use at studying only those phenomena which development submits to the normal law of distribution

- Stratified random (typical) sample reflects variability of the stratification variable in the observed population, allows to build rather homogeneous selection groups, allows to reduce sample size keeping necessary accuracy, gives benefits when selection problems in different domains of the population strongly differ

- Systematic samples are convenient for planning and selecting, sometimes they yield more exact results, than stratified; accuracy of systematic sample may be low if there is an unexpected periodicity in the frame. The use of systematic selection can be recommended, if stratification is planned very poorly, and in the case of two-stage selection, independent systematic selection in separate strata can be used

MULTIVARIATE SAMPLE: MAIN DEFINITIONS, PRINCIPLES, INVESTIGATION AREAS

- Multivariate sampling (MS) - is a kind of sampling design at which random and systematic sampling of units is carried out, taking into account the features of several quantitative and qualitative variables

Sampling can be made:

- from the typified frame, which is formed by a combination of qualitative and quantitative variables taking into account their structural features (on the basis of combinational tables)
- from a multiple frame which elements are organized in two or more frames
- by composite, or multi-dimensional variable
- on specially developed procedures of selection (lattice sampling, sample using cluster analysis)

In MS designs each population element is characterized by specific collection of variables (indicators); randomly selected unit simultaneously is the representative of some indicators

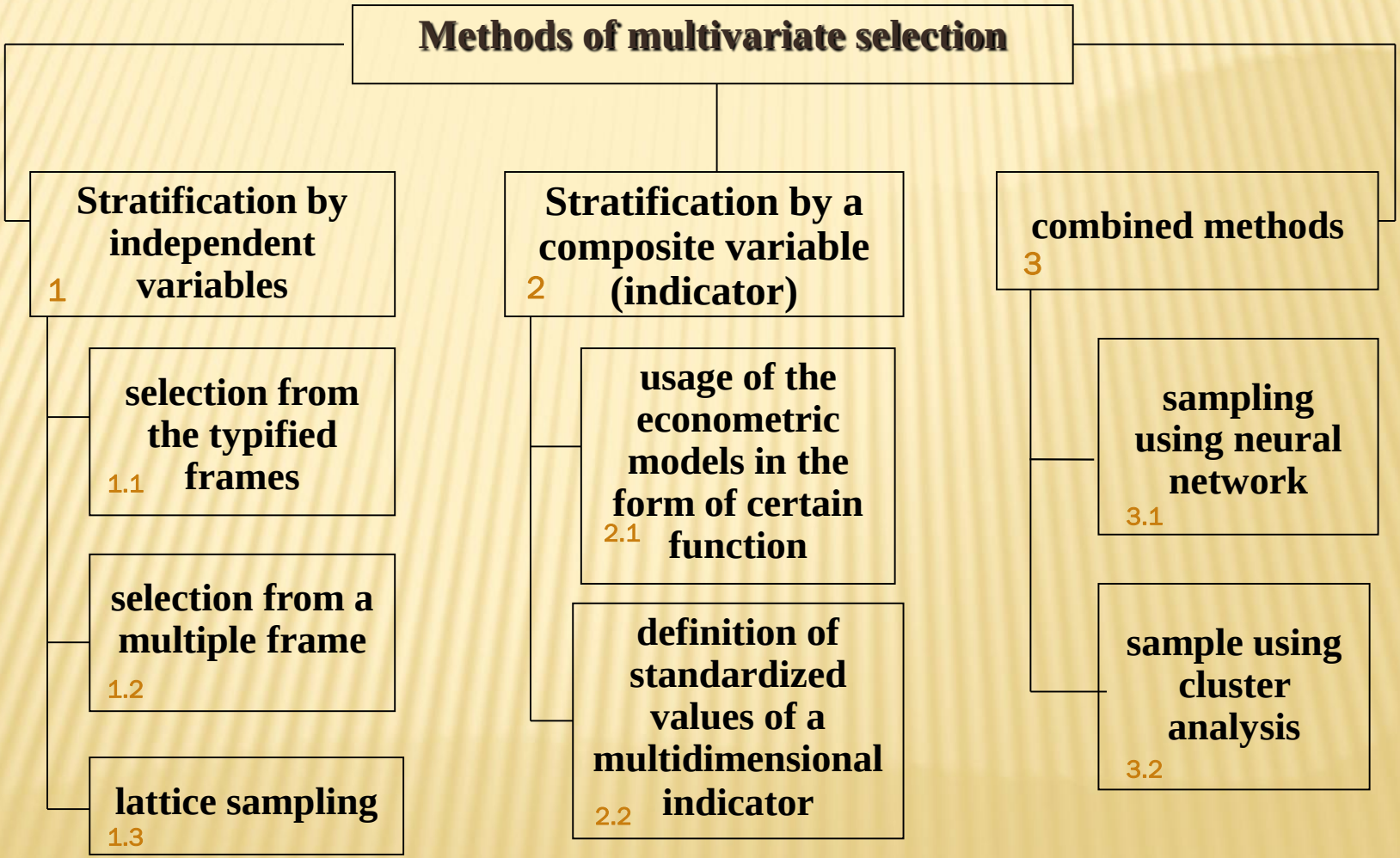
MAIN PRINCIPLES

- Classical selection schemes are not efficient
- Methods of optimization, combinatorial, cluster analysis, the combinational tables, multidimensional indicators are used
- Representative sampling
- Probability sampling

Areas to use:

- Territorial Households' samples
- Small business
- Branch samples of enterprises

METHODS OF MULTIVARIATE SELECTION



1. STRATIFICATION BY INDEPENDENT VARIABLES

Variables are considered absolutely independent, stratification on each of them is done independently; in final groups all received independent bounds are considered

Shortages: considerable amount of small groups of weak fullness, complicated and time-consuming procedure of modeling

1.1. Selection from the typified frames

(RSE / Gosmosstat in Russia, 1991; 1996-1997 – HH; Rosstat: small business, employment)

Construction of one or several combinational tables on a population, the numbers of population elements are allocated to the cells of combinational tables; the probability sample, multistage or cluster sample of units from the typified frames is selected

COMBINATIONAL GROUPING OF THE ENTERPRISES BY FOUR PARAMETERS (X_1, X_2, X_3, X_4)

Groups by X_2, %	Groups by X_1, % / Groups by X_3, mln. byr / Groups by X_4, persons	to 15		15-20	
		to 100	100 and more	to 100	100 and more
to 50	to 1000	№ 33*	№ 11, 13*	№ 24	№ 21
	1000-3000	№ 4, 12, 15	№ 7, 16*	№ 17, 19, 22, 23*	№ 18, 20
50-100	to 1000	№ 3, 10	№ 8, 9*	№ 31	№ 27, 28*, 29, 30, 32
	1000-3000	№ 1, 2, 5, 6	№ 14*	№ 25*	№ 26

$N = 33$; $N_r = 16$; $d_s = 50\%$; $n_r = 0{,}5 \cdot 16 = 8$; $n = 18$

1.2. Selection from two-dimensional or multiple frame

(Jessan R., 1985; Cochran W., 1976; Mahalanobis G. CH., 1958)

Population elements are allocated in 2 or more frames; the frames can be full or incomplete; a special case – the typified frames (analogue of different frames – grouping of units by different variables; SE and farms)

1.3. Lattice sampling

Selection is carried out under the "lattice" scheme: if there is a square with the side p, divided on p^2 unit squares, sample of size p is selected in such a way that one unit from each row and one from each column is being selected. The grouping by 2, 3 and more variables is possible

1	2	3	4
A			
	A		
		A	
			A

Before randomization

$r = 1; n = 4$

4	2	1	3
		A	
	A		
			A
A			

After randomization

2. SELECTION BY A COMPOSITE VARIABLE

One-dimensional stratification is carried out by a composite variable

Shortages: impossibility to consider simultaneously numerical and attributive parameters, 3 applicability to rather homogenous variables, conventionality of a multivariate indicator, individual indicators are not well represented

2.1. Econometrical models

The multivariate variable is used as an independent variable of the econometrical models (Rosstat: Survey on the distribution of the number of employees by the wage size, since 2001)

2.2. Standardized multidimensional variable:

$\overline{P_{ij}} = \frac{\sum P_{ij}}{k}$, where $\overline{P_{ij}}$ – multidimensional value for i – th unit;

P_{ij} – standardized value of j–th variable component for i – th unit;

k – total number of variables.

Ways of standardization:

$P_{ij1} = {}^{y_{ij}}/_{\overline{y}_j}$; $P_{ij2} = {}^{(y_{ij}-\overline{y}_j)}/_{\sigma_{yj}}$; $P_{ij3} = {}^{y_{ij}}/_{y'_j}$, where y'_j – etalon (normal) value;

$P_{ij4} = {}^{y_{ij}}/_{y_{jmax}}$; $P_{ij5} = {}^{(y_{ij}-\overline{y}_j)}/_{(y_{jmax}-y_{jmin})}$, etc.

3. COMBINED MULTIVARIATE SAMPLING

The multidimensional approaches described in items 1 and 2 of this presentation are combined

3.1. Sampling using neural network

(Krosnoyarsk regional committee of state statistics 1999 (SE), Stepanov S., 2004)

Observed population is presented as a structural model of groups – random variables – abstract observation units, for which quantitative (number employed, the income, production quantity etc.) and attributive variables (the branch, a pattern of ownership etc.) are assigned; elements of the model – neurons; criteria of formation of neuron – objective indicators of elements and preferences of the statistician; an initial database – statistical register; selection method – imitating modeling

Shortages: potential inadequacy of studied data to real situation, complexity of integration with existing system of the statistical data processing

3.2. Sampling using cluster analysis

(Belarus, SRI, Belstat: SE, labour statistics, retail trade, small business since 2005-2006, 2014 – 2015

1) surveyed population is partitioned using cluster analysis (agglomerative hierarchical, iterative method of k-means) on homogeneous groups;

2) in each received group the basic (leading) variable is determined, and subsequent casual or systematic selection of units in sample is carried out;

3) if for the leading indicator the coefficient of variation exceeds 50%, additional stratification inside cluster is used.

4) for each indicator the standard sample error is calculated; if the error exceeds admissible bounds, three methods of its reduction may be applied:

   - increasing the sample size in cluster;

   - additional stratification of the enterprises in cluster by a leading variable;

   - repetition of the process of clustering, possibly using the same method of clustering as earlier, but with larger number of steps, or using of an iterative method with the preliminary number of clusters r> l.

OPTIMIZING SAMPLING MODEL: EXPERIENCE OF USE

Optimizing sampling model:

- combination of univariate and multivariate sampling;
- choice of the way of selection for carrying out of multivariate sample survey depends on the size of population, number and character of the investigated parameters;
- is offered a collection of multivariate methods:
  - selection from the typified frames;
  - stratification by a composite variable;
  - modeling the sample by cluster analysis.

MICRO-ENTITIES SAMPLE SURVEY

- Sampling frames are:

1) micro-entities, represented the state statistical reports on the financial results for basic years (report 1-MP (micro));
2) set of the private farms. The first file is 80 thousands units, sample fraction depends on a character of the initial information, namely: the size of total population, kind of economic activity, region. The second array includes more than 2 thousands farms; it is observed completely

SAMPLING FRAMES

Micro-entities population. Belarus 2011-2013

Regions	Reports 1-MP (micro) respondents	Respon-ses	Non-responses			Private farms	Total number of micro-entities	Non-response rate, %
			total	of which				
				didn't work	other reasons (liquidated, not found)			
A	1	2	3	4	5	6	7=2+4 +6	$8=\frac{1+6-7}{1+6}\cdot 100$
2011								
Brest region	8064	6323	1741	1223	518	409	7955	6,1
Vitebsk region	7568	5746	1822	1254	568	309	7309	7,2
Gomel region	8150	6138	2012	1210	802	394	7742	9,4
Grodno region	6724	5288	1436	932	504	320	6540	7,1
Minsk	34955	24942	10013	3171	6842	–	28113	19,6
Minsk region	13779	10110	3669	2308	1361	453	12871	9,6
Mogilev region	7107	5246	1861	939	922	281	6466	12,5
Total	**86347**	**63793**	**22554**	**11037**	**11517**	**2166**	**76996**	**13,0**
2012								
Brest region	8894	6677	2217	1625	592	553	8855	6,3
Vitebsk region	8142	5896	2246	1545	701	369	7810	8,2
Gomel region	8892	6585	2307	1219	1088	428	8232	11,7
Grodno region	7389	5561	1828	991	837	375	6927	10,7
Minsk	38656	27205	11451	4389	7062	–	31594	18,3
Minsk region	16322	11385	4937	2989	1948	673	15047	11,5
Mogilev region	7639	5445	2194	1428	766	344	7217	9,6
Total	**95934**	**68754**	**27180**	**14186**	**12994**	**2742**	**85682**	**12,7**

SAMPLING FRAMES

Micro-entities population. Belarus 2011-2013

Regions	Reports 1-MP (micro) respondents	Respon-ses	Non-responses			Private farms	Total number of micro-entities	Non-response rate, %
			total	of which				
				didn't work	other reasons (liquidated, not found)			
A	1	2	3	4	5	6	7=2+4 +6	$8=\frac{1+6-7}{1+6}\cdot 100$
2013								
Brest region	9862	7681	2181	1249	932	580	9510	8,9
Vitebsk region	9035	6761	2274	1651	623	375	8787	6,6
Gomel region	9485	7500	1985	1057	928	438	8995	9,3
Grodno region	8111	6675	1436	755	681	375	7805	8,0
Minsk	39227	28501	10726	6473	4253	–	34974	10,8
Minsk region	19365	14220	5145	3653	1492	714	18587	7,4
Mogilev region	8359	6255	2104	1618	486	328	8201	5,6
Total	**103444**	**77593**	**25851**	**16456**	**9395**	**2810**	**96859**	**8,8**

SAMPLE DESIGN

- The combination of univariate and multivariate (multidimensional) sample is used.

- To receive optimal sample size for *i*-th kind of activity and *j*-th region the next algorithm is used:

1. The set of observed variables is allocated (for example, the wages fund, average number of employees, volume of production, revenues, profitability). Average, total values, variability of indicators are calculated.
2. Statistician chooses sampling method: univariate or multivariate. Univariate stratified samples with simple, proportional and optimal allocation are most often used.
3. It should be executed one of three conditions for applying multidimensional sampling:

- variation coefficient is more than 100 %;
- survey objects are non-uniform on many variables;
- the small size of total population (top limit – 30 – 40 units).

Otherwise univariate sampling should be used: random selection without allocation, simple random sample, proportional and optimal allocation.

SAMPLE DESIGN

4. It is expediently to use univariate stratified sample, total population is divided by rather homogenous groups. Then different variants of the sample size are executed (minimal is 0.05N, maximal is 0.8N). Minimal error is a main criteria of the determination of sample size.

   The choice of an optimum way of selection for carrying out of particular survey depends on a survey object and character of the auxiliary information, namely: degrees of uniformity, the sample size, presence of the natural isolated groups, availability of the correlated auxiliary information. It is expedient to approve several sampling designs for the same survey and to choose that from them which gives more precise and unbiased and unbiased estimates.

SAMPLE DESIGN

5. It is expediently to use multidimensional sample, selection is carried out by cluster analysis:

- total population is partitioned using cluster analysis on homogenous groups to k-variables, i.e. clustering;
- in each received group the leading (basic) variable is determined and subsequent random selection of units is performed.

Optimal sample population is chosen for each cluster, where standard errors of k-variables are criterias of productivity. If the error exceeds admissible bounds, three methods of its reduction may be applied: a) increasing sample population in cluster; b) additional stratification of the enterprises in cluster to a leading variable; c) repetition of clustering, but with larger number of steps.

Sample population is formed once in three-four years, i.e. fixed sample (yearly) is used.

STATISTICAL WEIGHTING

- To extrapolate sample data on the total population methods have been used:

1) Traditional group raising factors (weights), HT

2) Calibration (GREG-estimators)
3) Calibration (SYN – estimator)

HT-ESTIMATORS

The methodology of weighting for *univariate stratified sample* is based on the assignment for each enterprise corresponding statistical weight (k_{ijl}):

$$k_{ijl} = \frac{N_{ijl}}{n_{ijl}}, \qquad (1)$$

where k_{ijl} – individual weight for each enterprise of *l*-th group of *i*-th kind of activity (3 digit for NACE) in *j*-th region;

N_{ijl} – the size of *l*-th group of *i*-th kind of activity in *j*-th region in total population;

n_{ijl} – the size of sample group;

l – number of groups by observed variable value ($l = \overline{1, c}$).

HT-ESTIMATORS

Individual weights are equal within each group of micro-entities, calculated by region, kind of activity, observed indicator.

Individual weights, determined for *multidimensional sample*:

$$k_{ijr} = \frac{N_{ijr}}{n_{ijr}}, \quad (2)$$

$$k_{ijrh} = \frac{N_{ijrh}}{n_{ijrh}}, \quad (3)$$

where k_{ijr} – weight for *r*-th cluster of enterprises;

k_{ijrh} – weight for *h*-th group of *r*-th cluster;

r– number of clusters in *i*-th branch of *j*-th region ($r = \overline{1,l}$);

h – number of groups in *r*-th cluster ($h = \overline{1,s}$).

HT-ESTIMATORS

$$X_{Эj} = \sum_{r=1}^{l} \frac{N_r}{n_r} \cdot x_{sri}; \quad (4)$$

$$X_{Эj} = \sum_{r=1}^{l} \sum_{n=1}^{s} \frac{N_r}{n_r} \cdot x_{srih}, \quad (5)$$

where $X_{Эj}$ - the estimated value of j -th region;

x_{srj} - sample value in r -th cluster of enterprises;

x_{srjh} - sample value of m -th parameter in h -th group of r -th cluster;

$r = \overline{1, l}$ - number of clusters, defined on k to parameters;

$h = \overline{1, s}$ - number of groups in cluster, allocated to a leading variable;

$m = \overline{1, k}$ - number of parameters of an sampling unit;

nr - number of enterprises in r -th sampling cluster;

Nr - number of enterprises in r- th general cluster.

CALIBRATION (GREG-ESTIMATORS)

$$\hat{X}_{j_{GREG}} = \hat{X}_{j_{HT}} + \hat{\beta} \cdot \left(Y - \hat{Y}_{HT}\right),$$

$$\hat{X}_{j_{GREG}} = \sum_{i=1}^{k} k_i x_i + \hat{\beta} \cdot \left(Y - \sum k_i y_i\right),$$

$$\hat{\beta} = \frac{\sum q_i k_i x_i y_i}{\sum q_i k_i y_i^2},$$

where q_i - scale coefficients, selected by the researcher;

y_i – average number of employees.

CALIBRATION (SYN-ESTIMATOR)

$$\hat{X}_{j_{SYN}} = \sum \hat{x}_{ij},$$

где $\hat{x}_{ij} = f(y_{ij}\hat{\beta}); \hat{x}_{ij} = \sum \hat{\beta} \cdot y_i,$

$$\hat{\beta} = \left(\sum \frac{y_i y_i'}{\sigma_{yi}^2 k_i} \right)^{-1} \cdot \sum \frac{y_i x_i}{\sigma_{yi}^2 k_i},$$

$k_i = w_i$ – weight of HT-estimator

INDICATORS OF THE SAMPLE SET BY THE ACTIVITY "TEXTILE AND CLOTHING MANUFACTURE." BREST REGION. PROPORTION ALLOCATION

Indicator	groups (X), mln. rubles	Size		Sampling fraction $d = \frac{n}{N}$	Total value		Variance		Variation coefficient, %	
		total population, N	sample, n		total population, X	sample, Σx_i	total population, σ^2_x	sample, S^2_x	total population, V_x	sample, v_x
Revenue (X)	до 50	31	7	0,225	712	133	252,4	174,6	69,2	69,5
	50-200	54	11	0,20	6640	1293	1742,9	1561,2	33,95	33,62
	200-1000	65	15	0,20	34435	6578	40959,2	30380,5	44,1	39,8
	1000-2000	14	3	0,21	18353	4640	39401,9	27671,3	15,14	10,8
	2000 и более	6	2	0,33	28575	11472	7417021,6	13286025	57,2	63,5
	Итого	180	38	0,21	88715	24116	1007191	2311616	203,6	239,6

INITIAL AND CALCULATION DATA FOR GREG - AND SYN-ESTIMATORS

№	Revenue, x_i	Number of employees, y_i	Basic weight, k	$k \cdot x_i$	$k \cdot x_i y_i$	$k \cdot y_i^2$	$k \cdot y_i$	y_i^2	x_i^2
1	9381	8	3	28143	225144	192	24	64	88003161
2	24	2	4,4285714	106,2857	212,5714	17,71429	8,857143	4	576
3	626	21	5	3130	65730	2205	105	441	391876
4	390	12	5	1950	23400	720	60	144	152100
5	411	12	5	2055	24660	720	60	144	168921
6	109	3	4,9090909	535,0909	1605,273	44,18182	14,72727	9	11881
7	476	14	5	2380	33320	980	70	196	226576
8	7	0	4,4285714	31	0	0	0	0	49
9	150	4	4,9090909	736,3636	2945,455	78,54545	19,63636	16	22500
10	1612	13	4,6666666	7522,667	97794,67	788,6667	60,66667	169	2598544
11	219	10	5	1095	10950	500	50	100	47961
12	465	11	5	2325	25575	605	55	121	216225
13	165	5	4,9090909	810	4050	122,7273	24,54545	25	27225
14	662	7	5,0000000	3310	23170	245	35	49	438244
15	86	1	4,9090909	422,1818	422,1818	4,909091	4,909091	1	7396
16	284	13	5,0000000	1420	18460	845	65	169	80656
17	67	2	4,9090909	328,9091	657,8182	19,63636	9,818182	4	4489
18	525	11	5	2625	28875	605	55	121	275625
19	2091	8	3	6273	50184	192	24	64	4372281
20	377	7	5	1885	13195	245	35	49	142129

INITIAL AND CALCULATION DATA FOR GREG - AND SYN-ESTIMATORS

№ п/п	Revenue, x_i	Number of employees, y_i	Basic weigh, k	$k{\cdot}x_i$	$k{\cdot}x_i{\cdot}y_i$	$k \cdot y_i^2$	$k{\cdot}y_i$	y_i^2	x_i^2
21	245	15	5	1225	18375	1125	75	225	60025
22	1317	14	4,6666666	6146	86044	914,6667	65,33333	196	1734489
23	882	20	5	4410	88200	2000	100	400	777924
24	246	7	5	1230	8610	245	35	49	60516
25	193	5	4,9090909	947,4545	4737,273	122,7273	24,54545	25	37249
26	1711	14	4,6666666	7984,667	111785,3	914,6667	65,33333	196	2927521
27	118	8	4,9090909	579,2727	4634,182	314,1818	39,27273	64	13924
28	436	5	5	2180	10900	125	25	25	190096
29	28	3	4,4285714	124	372	39,85714	13,28571	9	784
30	114	6	4,9090909	559,6364	3357,818	176,7273	29,45455	36	12996
31	3	0	4,4285714	13,28571	0	0	0	0	9
32	67	7	4,9090909	328,9091	2302,364	240,5455	34,36364	49	4489
33	334	7	5	1670	11690	245	35	49	111556
34	81	1	4,9090909	397,6364	397,6364	4,909091	4,909091	1	6561
35	41	2	4,4285714	181,5714	363,1429	17,71429	8,857143	4	1681
36	25	2	4,4285714	110,7143	221,4286	17,71429	8,857143	4	625
37	5	1	4,4285714	22,14286	22,14286	4,428571	4,428571	1	25
38	143	4	4,9090909	702	2808	78,54545	19,63636	16	20449
Total	**24116**	**285**		**95895,79**	**1005171**	**15717,06**	**1369,437**	**3239**	**103149334**

HT-estimator:

$$\hat{X}_{HT} = \sum_{i=1}^{h} k_i x_i = \frac{31}{7} \cdot 133 + \frac{54}{11} \cdot 1239 + \frac{75}{15} \cdot 6578 + \frac{14}{3} \cdot 4640 + \frac{6}{2} \cdot 11472 = 95895{,}79 \ m\ln \ rubles$$

$$Er_{HT} = \left| X - \hat{X}_{HT} \right| = \left| 88715 - 95895{,}79 \right| = 7180{,}8 \ m\ln \ rubles, \ or \ 8{,}1\%$$

GREG-estimator:

$$\hat{X}_{GREG} = \sum_{i=1}^{h} k_i x_i + \frac{\sum_{i=1}^{h} k_i x_i y_i}{\sum_{i=1}^{h} k_i y_i^2} \cdot (Y - \sum_{i=1}^{h} k_i y_i) = 95895{,}79 + \frac{1005171}{15717{,}06}(1270 - 1369{,}4) =$$

$$= 89538{,}75 \ m\ln \ rubles$$

$$Er_{GREG_i} = \left| X_i - \hat{X}_{GREG_i} \right| = \left| 88715 - 89538{,}75 \right| = 823{,}75 \ m\ln \ rubles, \ or \ 0{,}93\%$$

SYN-оценка:

$$\hat{X}_{SYN} = \sum_{i=1}^{h} k_i y_i \cdot \frac{\sum_{i=1}^{h} k_i x_i y_i}{\sum_{i=1}^{h} k_i y_i^2} = 1369 \cdot \frac{1005171}{15717{,}06} = 87553{,}2 \ m\ln \ rubles$$

$$Er_{SYN} = \left| X - \hat{X}_{SYN} \right| = \left| 88715 - 87553{,}2 \right| = 1161{,}8 \ m\ln \ rubles, \ or \ 1{,}3\%$$

INDICATORS OF THE MICRO-ENTITIES SAMPLE POPULATION. BELARUS. 2014

Indicators	N	n	Total value		Sampling fraction	Error, %
			total population	sample		
Brest region						
revenue	7076	1474	10 434 551	3 829 516	20,83	0,62
volume of production	7076	1474	4 444 599	1 582 473	20,83	1,21
average number of employees	7076	1474	33 076	8 243	20,83	2,29
Vitebsk region						
revenue	6291	1575	8 583 785	4 143 654	25,04	0,74
volume of production	6291	1575	3 392 467	1 613 969	25,04	1,06
average number of employees	6291	1575	27 205	8 620	25,04	2,67
Gomel region						
revenue	6861	1904	8 862 320	4 596 506	27,75	0,41
volume of production	6861	1904	3 642 661	1 790 068	27,75	0,66
average number of employees	6861	1904	28 973	10 505	27,75	2,52
Grodno region						
revenue	6102	1786	9 274 248	5 405 327	29,27	0,55
volume of production	6102	1786	3 523 414	1 972 308	29,27	0,67
average number of employees	6102	1786	27 343	10 773	29,27	1,75

INDICATORS OF THE MICRO-ENTITIES SAMPLE POPULATION. BELARUS. 2014

Indicators	N	n	Total value		Sampling fraction	Error, %
			total population	sample		
Minsk						
revenue	25431	3403	73 747 281	27 954 534	13,38	0,42
volume of production	25431	3403	24 793 886	8 684 264	13,38	0,69
average number of employees	25431	3403	115 092	20 382	13,38	1,63
Minsk region						
revenue	12995	2633	21 932 541	10 534 756	20,26	0,35
volume of production	12995	2633	9 290 420	4 356 673	20,26	2,17
average number of employees	12995	2633	54 249	15 128	20,26	2,07
Mogilev region						
revenue	5799	1717	8 231 268	4 499 110	29,61	0,45
volume of production	5799	1717	3 328 475	1 847 117	29,61	0,59
average number of employees	5799	1717	25 435	9 852	29,61	2,97
Belarus						
revenue	**70555**	**14492**	**141 065 994**	**60 963 403**	**20,54**	**0,45**
volume of production	**70555**	**14492**	**52 415 922**	**21 846 872**	**20,54**	**1,01**
average number of employees	**70555**	**14492**	**311 373**	**83 503**	**20,54**	**2,07**

DONOR POOL

1. In 2011 – 2013 average nonresponses rate of micro-entities is 13-14%. These data based on annual continuous small enterprises survey (report 1-MP (micro)). Besides, it is necessary to take into account potential nonresponses on donor pool. Therefore donor pool size is 0,17n.
2. Allocation of donor pool by regions: proportional the number of nonresponses in previous year (report 1-MP (micro)); by Minsk – 18%, by regions – 7-12%.
3. Selection of donor entities in region:

$$n_{pm}=d_m \cdot n_{p,}$$

where d_m – the share of region in total number of donors;

$$N_m^*=N_m - n_m,$$

where N_m^* – population, from which donors are selected.

CONCLUDING REMARKS

In 2014 – 2015 the results of trial calculations testing the first version of methodological and software sampling have shown:

1) main difficulties are associated with the use of different weighting schemes, necessary estimation of the whole variables, splitting of the same population on the smaller groups, little subsamples

2) sampling fraction – 15-20 %. As to branch sampling fraction is depending on the number of enterprises and the degree of accuracy on a leading variable: a relative sampling error by regions less than 2-3 %, by kinds of activity less than 4-5 %, and by small branches less than 8 – 9 %. Total sampling fraction – 21%, by Minsk – 13%, by kinds of activity – trade, transport – 7-10%, construction – 5-10%, agriculture, some branches of industry – 30-40%

CONCLUDING REMARKS

3) the use of combination univariate and multidimensional samples, different weighting methods will provide very reliable information over larger number of variables of micro-entities survey, conducted in Belarus: employment, wages fund, profit, output, revenues and others

4) to improve the representativeness by region weighting procedure can be complicated by usage of auxiliary calibration estimators. Besides, it is important to take into account the necessity of annual sample updating. The creation of new entities, liquidations, changes in kinds of activity and size of enterprises are taken place constantly

REFERENCES

1. Bokun, N., Chernyshova, T (1997). Methods of sample surveys. Ministry of statistics and analysis of Belarus, Research Institute of Statistics.

2. Bokun, N. (2008): Ispolzovanie mnogomernoj vyborki dlia izmerenia neoficialnoj ekonomiki // Voprosy statistiki . – 2008. - №2. – P. 30 – 37

3. Bokun, N., (2010). Problems of multidimensional samples in retail trade. Questions of statistics 3, 52-60.

4. Bokun, N. Micro-entities sample survey: problems of design, formation and usage / N. Bokun |// Workshop of Baltic – Nordic – Ukrainian network on Survey Statistics, Tallinn, Estonia, August 25-28, 2014. – P. 25-31.

5. Cochran, W.G. (1953, 1963): Sampling Techniques. New York, John Wiley and Sons, Inc.

6. Jessan, R. (1985): Methods of statistical surveys. New York

7. Särndal, C.-E., Swensson, B., Wretman, J. (2003). Model assisted survey sampling. Springer Verlag.

8. Stepanov, S. V. (2004): Nejronnaja interpretacia vyborki dlia strukturnogo statisticheskogo modelirovania sovokupnostej predprijatij // Voprosy statistiki. – 2004. - № 6.

THANK YOU VERY MUCH!