ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
STATE COMMITTEE OF THE RUSSIAN FEDERATION ON STATISTICS
GOVERNMENT OF ST. PETERSBURG AND LENINGRAD REGION
with the participation of FAO, UN/ECE, EUROSTAT

Sixth IWG.AGRI Seminar on Agricultural Statistics

Russian Federation, St.Petersburg, 29 June - 3 July 1998

THE SYSTEM FOR SAMPLING INQUIRIES CONCERNING THE DEVELOPMENT OF AGRICULTURE IN THE RUSSIAN FEDERATION

Prepared by Ms. N. Vasilevskaya, Statistical Research Institute, Goskomstat of Russia

.

THE SYSTEM FOR SAMPLING INQUIRIES CONCERNING THE DEVELOPMENT OF AGRICULTURE IN THE RUSSIAN FEDERATION

Ms. N. Vasilevskaya, Statistical Research Institute,
Goskomstat Rossii

With the emergence of new forms of ownership and the reorganization of agricultural production, the statistical authorities of the Russian Federation have been engaged since 1990 in creating a uniform methodology for the collection, processing and analysis of statistical information.

The changes in the way agriculture is organized have faced statisticians with the challenge of applying existing, and developing new methods of observing the processes involved.

With the transition to a market economy, the conditions of observation have changed. The main forms of inquiry are now censuses, registers and sample surveys.

The federal programme for the "Reform of Statistics in 1997-2000" provides for the adoption of a uniform system of sample inquiries concerning all categories of agricultural producer (agricultural enterprises, small enterprises, farmers' farms, private ancillary plots). That will provide the regional and federal statistical authorities with representative information.

Sampling inquiries concerning farmers' farms

Given the pace of creation of peasant (farmers') farms in 1992-1994 and the consequent sharp rise in the number of units to be observed, it has been decided that statistical observation of farmers' agricultural activity should be effected on a sample basis. During the period of unstable development of this form of farm, sample surveys have been made every other year, so that one year there has been complete reporting and the next year the information has been collected at the same periods by means of sample surveys. In 1994, sub-registers of farmers' farms were created in the regions to serve as a database for the design of a variety of samples. A sub-register is built on the principle of current reporting and is updated annually. Later the register may be updated every three or four years.

Before resolving the questions pertaining to the selection of the sample set of peasant farms a solution has to be found to the principal question, that of defining what the parent population, i.e. the population to be studied, will be: will it be the entire set of farmers' farms in the region, or the following two populations (arable farms and livestock farms) or a series of populations (by type of crop and type of livestock)? Then, once the parent population or populations has been defined, it has to be decided what indicator should be taken as the basis for the sampling, of what type the sample should be and what sample size is consistent with an acceptable level of precision for the main indicators.

In order to determine whether the sample should comprise one or several sets of peasant farms, it had to be determined what proportion of peasant farms produce all types of agricultural product. According to the register of farmers' farms in Belgorod oblast for 1996, not one of the total of more than 2,500 farms produced every sort of crop and livestock. Only 40 farms were simultaneously growing grain, sugar beet and sunflowers, while 942 farms were growing grain, 263 sugar beet and 143 fodder crops. The picture for livestock farming was similar: only 32 farmers' farms were simultaneously raising beef and dairy cattle, pigs and sheep.

The number of farms that simultaneously had beef and dairy cattle and pigs was somewhat higher: 183. As for the numbers of farms in the region with one type of livestock, there were 343 with cattle and 384 with pigs.

The design of stratified samples for above populations entails study of the variation within groups built for one of the indicators. The value of this approach soon becomes apparent because a breakdown into groups that are homogeneous in respect of one indicator will not lead to homogeneity in respect of other indicators. For example, groups which are homogeneous in respect of beef cattle (V = 0; 0; 13; 38; 37%) are heterogeneous in respect of pigs (V = 88; 61; 116; 107; 44%). That makes sampling virtually impossible when the number of farms in groups is small.

In order to obtain representative information on crop farming (sown areas and yields of the main crops) and livestock farming (numbers and productivity of main types of animal) on peasant farms it is helpful to construct single-purpose samples.

For crop farming by main crop groups (grain and legumes, sugar beet, sunflowers for seed, potatoes, fruit, fodder crops) and livestock farming by main types of animal (cattle, pigs, sheep and goats, poultry, horses, reindeer), the frames are in the form of lists of farms ranked by growth (in crop farming, by spring sown areas, form No. 1, “Farmer on 1 June of current year”; in livestock farming, by main types of animal, form No. 3, “Farmer on 1 January of current year”).

Initial study of the activity of peasant farms in a number of regions of Russia has confirmed the need for building samples by types of crop and livestock and the advisability of using stratified samples because of the considerable variation of all the indicators.

Since the software for sub-registers of farmers' farms at the regional level is constantly being improved and supplemented, but the process of register management is not at the same stage in every region, the decision has been taken to employ a simpler sampling procedure: systematic sampling from a list of farms ranked according to the indicator forming the basis for the sampling. Then the sample can be made manually in the event of error in the register and the data grossed up.

In 1997, software was developed and introduced in the regions for building sub-register-based systematic and stratified samples of peasant farms with a maximum (intervals) sampling error of ± 10% and for grossing up data from the systematic and stratified samples with an assessment of the data's representativeness (calculation of the sampling errors).

The building of a stratified sample with optimum allocation (proportional to the variation coefficient) and the grossing up of the stratified-sample data cut the cost of making sampling inquiries with an acceptable level of accuracy.

The main problem in writing the software was to find the best way of dividing the parent population into homogeneous typical groups without expert assistance.

Both building such a sample and grossing up the data entail a great deal of computation (grossing up for 12 sample populations and, within them, for numerous - as many as 20 - typical groups). Performance of these tasks would have been impossible without proper management of the sub-register, the availability of clear software and the presence in the regional statistical organs of highly qualified staff.

Grossing up of the data in the systematic and stratified samples to all types of crop and livestock is achieved by calculating means () for the indicators in the sample and multiplying them by the number of units, corrected for the date of the sample, in the parent population.
Sample network of private farms

In 1997, it was decided, in order to improve statistical observation of agricultural production in the private sector and ensure that information was comprehensive and effective, to create a sample population of rural private farms distinct from the current budget network.

Sample surveys procedure depends largely on the **basis for the sampling**. The general practice with the most commonly used methods of simple random and specified sampling is for the sampling units to be also the units that are studied. In this case, for each unit there is a corresponding value of the resulting characteristic or the factor characteristic. For these methods to be usable, the full sample frame has to be available in the form of lists, card files, registers or the like. No full information is currently available at the regional level on the population of private ancillary plots (approximately 17 million households).

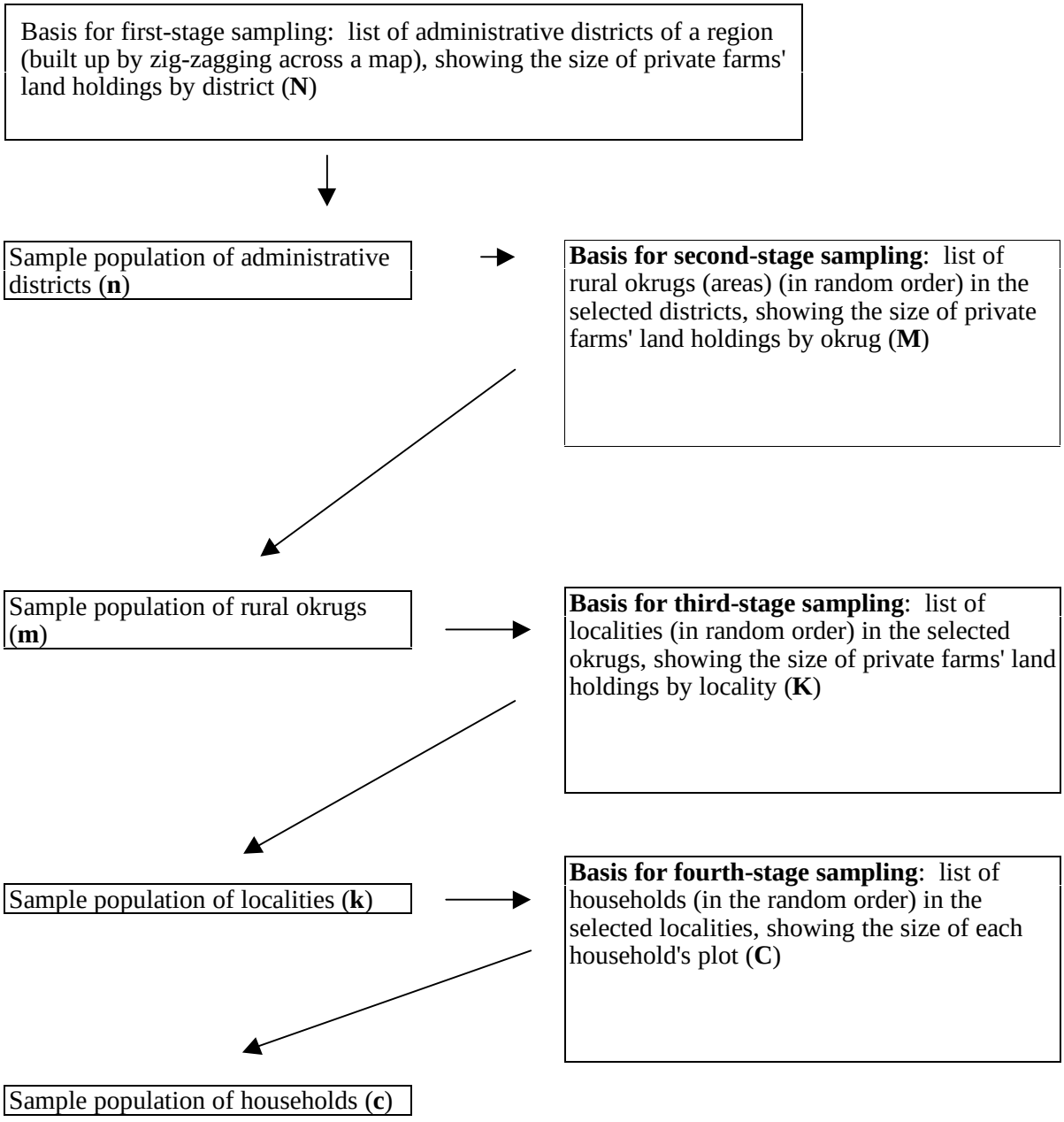
In such circumstances, the use of multi-stage sampling is advisable. With this form of sampling, the units to be observed are extracted only at the final stage, after several successive random sampling operations. The units in the first-stage sample (the primary units) differ from those in the second-stage sample (secondary units) and so on.

With multi-stage sampling, the sample frame is only needed in order to select the units for the following stage and is the lists of units selected at the previous stage.

For the formation of a sample network of private farms, multi-stage sampling offers the advantage that it is possible, using the small amount of information available in the regions and from rural authorities, to select a significant proportion of the localities the lists of households in which will be used to make up the final population. That greatly reduces survey costs.

Following the study of a number of normative documents and discussion of the problem with specialists from the Russian State Land Committee, it was decided to take as the basis for the sampling of private farms the **lists of land-tax payers**. Included in these lists are citizens who either own plots, have the use of plots or have inherited plots for their lifetimes. The first lists of land-tax payers were compiled as of 1 January 1996.

The method used to form the sample network of private farms was that of four-stage area random sampling, as follows:



For the area sampling, use is made of geographic or topographic maps. This ensures an equal chance of inclusion in the sample for all the types of soil and all the climate zones in the given region. The maps are used in all regions at the first stage (building of the sample population of administrative districts) and in some regions at the second (district-level) stage too (selection of rural okrugs).

The selection sequence is as follows: in the first stage, 25% of the administrative districts in the region; in the second stage, 15% of the rural okrugs in each selected administrative district; in the third stage, 10% of the localities in each selected rural okrug; in the fourth stage, fixing of the number of private farms on the basis of the chosen overall size of the sample (0.1% of the total number of private farms in the region).

Hence, the number of localities in which full lists of private farms will be used is 0.25 x 0.15 x 0.10 = 0.00375, i.e. the 0.38% of the total number of localities in the region.

The sampling interval at each stage is the ratio between the total land area given over to private farms at each stage and the size of the sample at that stage (the sampling interval at the first stage is equal to the area of land in farms in the region divided by the number of districts selected - see annex 1).

With multi-stage sampling, arriving at estimates of the parameters of the parent population is usually quite easy; the methodological problems arise in calculating the standard sampling errors.

In four-stage sampling, the population is assumed to comprise **N** first-stage units each containing **M** second-stage units, each of which contains **K** third-stage units and **C** fourth-stage units. The corresponding sample sizes are **n**, **m**, **k**, **c**. Let xijue be the value of the characteristic for observation of the e-th fourth-stage unit from the i-th third-stage unit selected from the j-th second-stage unit selected from the i-th first-stage unit.

The variance of the mean value of the characteristic of a fourth-stage unit in four-stage sampling may be written:

$$\delta^2_{\tilde{x}} = (\frac{N-n}{N}) \cdot \frac{\delta^2_1}{n} + (\frac{M-m}{M}) \cdot \frac{\delta^2_2}{m \cdot n} + (\frac{K-k}{K}) \cdot \frac{\delta^2_3}{n \cdot m \cdot k} + (\frac{C-c}{C}) \cdot \frac{\tilde{\delta}^2_4}{n \cdot m \cdot k \cdot c},$$

where: δ^2_1 is the variance for the mean value at the first stage

δ^2_2 is the variance for the mean value at the second stage

δ^2_3 is the variance for the mean value at the third stage

$\tilde{\delta}^2_4$ is the mean of the intra-class variances at the fourth stage.

Sample network of agricultural organizations

In 1999, the source of information for observation of selling prices for agricultural products will be a sample network of agricultural organizations (having more than 60 workers) which will be developed on the basis of the sub-register of agricultural producers.

Sample population of new enterprises

In view of the increase in the number of small enterprises (enterprises having less than 60 workers), methodology and an algorithm are being developed for building a sample population of small agricultural enterprises (broken down according to whether they are crop or livestock farms).

Annex 1

Building of the sample population of administrative districts in Penza oblast

(Sampling by snaking/zig-zagging over the entire area)
First stage
Sampling interval of = 8855; 0.5 interval = 4427

Code	List of districts in the oblast	Total land area allocated to households, ha	Sampling probability	Cumulative sum	Cumulative sum of intervals	Selected districts
1	Nikolsky	2 069	0.033	2 069	4 427	
2	Issinsky	1 886	0.030	3 955	4 427	
3	Narovchatsky	1 822	0.029	5 777	4 427	1
4	B.Demyanovsky	936	0.015	6 713	13 282	
5	Luninsky	2 581	0.042	9 294	13 282	
6	Mokshansky	2 422	0.039	11 716	13 282	
7	Vadinsky	2 471	0.040	14 187	13 282	1
8	Zemetchinsky	4 561	0.074	18 748	22 137	
9	Nizhnelomovsky	3 072	0.050	21 820	22 137	
10	Sosnovoborsky	2 104	0.034	23 924	22 137	1
11	Bessonovsky	3 786	0.061	27 710	30 992	
12	Pachelmsky	1 679	0.027	29 389	30 992	
13	Kuznetsky	3 160	0.051	32 549	30 992	
14	Gorodishchensky	2 964	0.048	35 513	39 847	
15	Penzensky	1 985	0.032	37 498	39 847	
16	Kamensky	2 421	0.039	39 919	39 847	1
17	Bashmakovsky	2 549	0.041	42 468	48 702	
18	Belinsky	3 957	0.064	46 425	48 702	
19	Neverkinsky	1 569	0.025	47 994	48 702	
20	Kameshkirsky	2 293	0.037	50 287	48 702	1
21	Shemsheisky	1 763	0.028	52 050	57 557	
22	Kalyshleisky	1 380	0.022	53 430	57 557	
23	Kondolsky	960	0.015	54 390	57 557	
24	Lopatinsky	1 327	0.021	55 717	57 557	
25	Tamalinsky	1 409	0.023	57 126	57 557	
26	Bekovsky	1 438	0.023	58 564	57 557	1
27	Serdobsky	2 208	0.036	60 772	66 412	
28	Maloserdobinsky	1 211	0.020	61 983	66 412	