MULTIDIMENSIONAL SAMPLING OF FARMS WITHIN

A successful Kazakh-German cooperation

New Challenges for Statistical Software – The Use of R in Official Statistics – Bucharest, 7 April 2016

Introduction

Twinning Project between The Republic of Kazakhstan and International Bank for Reconstruction and Development

One component of this twinning project named "Improving the quality and methods of conducting sample surveys" included theoretical and practical missions in Kazakhstan and Germany.

About the need for a software update.

Starting point

Sampling from Kazakh Statistical Agricultural Registry

Method existed and results of this approach were satisfactory

However, work process was cumbersome and not very handy

- Old software (Statistica on a stand-alone Windows95 computer)

- All workers involved in the development of the Statistica program left

- Checking the sample took many working hours

- Further processing was done with Microsoft Excel

Goals

- Reprograming in R

- Automatize the process

- Reduce the work load

- Capacity building on the usage of the statistical software R

…and thereby give the workers at the Committee of Statistics the possibility to change the method on their own

The Statistical Agricultural Registry (SAR)

≈ 180 000 agricultural units

- agricultural enterprises;
- country farms;
- households;
- gardening and country cooperatives.

Updates

- agricultural census every decade;
- twice a year by administrative sources;
- every month from statistical forms;
- annually, by three samples (livestock, crops);
- data of the tax authorities.

The Statistical Agricultural Registry (SAR)

16 oblasts

210 raions

The Statistical Agricultural Registry (SAR)

The registry includes about 300 variables

55 are frequently used

Sample size n

Sample sizes have been estimated and adopted after the statistical units have been surveyed over several years:

- For peasant and farming enterprises – 30 %;

- For subsistence farms – 5 %;

- For horticultural, market-gardening and dacha associations/cooperatives – 5 %.

Sampling method

Multivariate systematic PPS sampling

Farm i	Crops		
	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$
1	584	6	0
2	2450	17	0
3	6	6	6
4	22	22	11
5	13	13	13
6	10	10	9
⋮	⋮	⋮	⋮
Total (N=20)	6551	533	173

Sampling method

Farm i	Crops		
	$x_{1,i}^{0,75}$	$x_{2,i}^{0,75}$	$x_{3,i}^{0,75}$
1	118.8	3.6	0.0
2	**348.2**	8.3	0.0
3	**3.9**	3.9	3.9
4	**10.0**	10.0	6.0
5	**7.0**	7.0	7.0
6	**5.5**	5.5	5.0
⋮	⋮	⋮	⋮
Total (N=20)	1 226.8	209.3	87.9

Sampling method

$$p_i = n_1 \frac{X_{1,i}^f}{\sum_{i=1}^{N} X_{1,i}^f}$$

Probability of inclusion= $min\{1, max\{p_i, \ldots, p_k\}\}$

Sampling method

Farm i	Probability of inclusion with $n_1 = 8, n_2 = 5, n_3 = 4$					Cumulative	Inclusion in sample
	p_1	p_2	p_3	$max\{p_i\}$	$min\{1, max\{p_i\}\}$	$\sum_i min\{1, max\{p_i\}\}$	1 = „yes", 0 = „no"
1	0.77	0.09	0.00	0.77	0.77	0.77	0
2	2.27	0.20	0.00	2.27	1.00	1.77	1
3	0.03	0.09	0.18	0.18	0.18	1.95	0
4	0.07	0.24	0.27	0.27	0.27	2.23	1
5	0.05	0.17	0.32	0.32	0.32	2.54	0
6	0.04	0.13	0.23	0.23	0.23	2.77	0

Sampling method

Random start number $r \in [0; 1]$ is chosen and added to the cumulative value (in the example it is 0)

Because the probability of inclusion is at maximum 1, the number of included farms is not equal to the beforehand chosen n_i → sample size has to be adopted

Sampling method

"ratio adjustment":

ratio of the extrapolated sum $\hat{Y}$ to the sum of the crop in the registry

must be within the interval [0.7; 1.3]

Adoption and "ratio adjustment" could easily be programmed in R

Wrap-up

Existing method for sampling in agricultural statistics was reprogrammed in R and is used

Work load for workers at CS was massively reduced

R is one standard software tool now at the CS

Thank you for your attention!

Zhandos Kozbanov

Sven Schmiedel

Zha.Kozbanov@economy.gov.kz

Sven.Schmiedel@destatis.de

Meiram Seidazimov

M.Seidazimov@economy.gov.kz