

Representativeness of the RLMS-HSE Sample

The RLMS sample is representative of the Russian Federation at the national-level. At the sub-national level, the Moscow and St. Petersburg samples are representative. No other sub-national portions of the sample are representative of their geographic or administrative areas. A more detailed discussion of the issues follows.

As is commonplace in national surveys, a few particularly populous cities were designated as "self-representing" (SR) areas. Moscow city, Moscow Oblast, and St. Petersburg city (with 25 million inhabitants) were included with certainty, i.e., they were put into the sample automatically.

Stratification was employed to increase the precision of estimates in non-self-representing (NSR) areas. At the time when the sample was created, there were no federal districts, and there were 11 geographic regions: Northern regions, Northwestern region, Central region, Volga-Viatsky region, Central Black Earth region, Volga basin region, North Caucasus region, Ural region, Western Siberian region, Eastern Siberian region and Far East regions. We combined Eastern Siberian region and Far East region because of the small population living in these regions.

Since it was necessary that the strata were approximately equal in terms of population size (the average size of the strata about 3.5 million people), in order to achieve this we had to change the boundaries of the original geographic regions (see Table 1).

Table 1

Ten adjusted Geographical Regions of the Russian Federation (Excluding 3 Self-Representing Areas)

NN	Regions	Number of strata
1	Northern regions and Kostroma oblast	2
2	Northwestern region	1
3	Central region minus Kostroma oblast	4
4	Volga-Viatsky region	3
5	Central Black Earth region minus south Voronezh oblast	2
6	Volga basin region minus Astrakhan and Penzen oblasts and Kalmyk	4
7	North Caucasus region, Astrakhan and south Voronezh oblast, Kalmyk	5
8	Ural region	6

9	Western Siberian region	4
10	Eastern Siberian region and Far East regions	4
	Total (excluding SR strata)	35

In principle, the data can be regarded as representative only for designed strata or their associations. Table 1. shows that the strata (and their associations) do not always coincide with the existing geographical regions or their associations.

A second problem associated with the presentation at the level of the region is that only one PSU be drawn from each stratum. Table 1 shows that the number of PSU in each adjusted geographical region varies from one to six. It is too small to provide adequate representation at the regional level.

Ideally, more than one PSU would be drawn within each stratum because, all else being equal, the greater the number of PSUs, the greater the precision of estimates. However, sampling always boils down to trade-offs between precision and cost. In this case, funds were not available for increasing the number of PSUs by a factor of two or more.

Best regards.

Mikhail Kosolapov

CEO Research Center "DEMOSCOPE"