THE USE OF CLUSTER ANALYSIS FOR DEMOGRAPHIC POLICY DEVELOPMENT: EVIDENCE FROM RUSSIA

Oksana Shubat
Ural Federal University
620002, Ekaterinburg, Russia
Email: o.m.shubat@urfu.ru

Anna Bagirova
Ural Federal University
620002, Ekaterinburg, Russia
Email: a.p.bagirova@urfu.ru

Abilova Makhabat
Magnitogorsk State Technical University
455000, Magnitogorsk, Russia
Email: abilova.mahabat@yandex.ru

Anton Ivlev
Magnitogorsk State Technical University
455000, Magnitogorsk, Russia
Email: ivlevanton@bk.ru

KEYWORDS

Cluster analysis, Demographic policy, Fertility, Birth rate, Russian regions
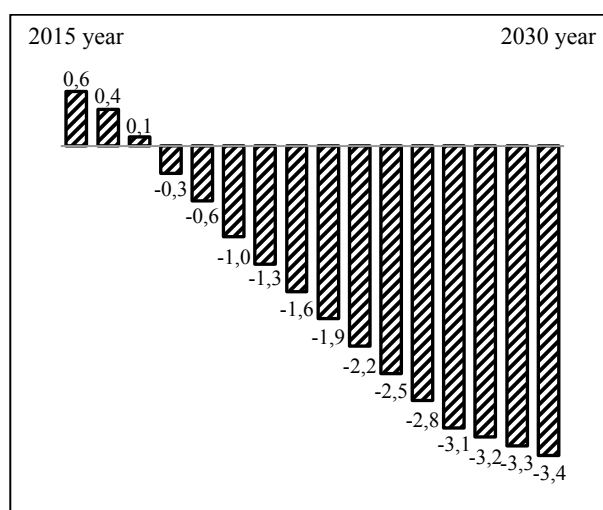
ABSTRACT

Russia has been experiencing a demographic crisis since the 1990s. The most obvious manifestations include an excess of mortality over fertility rates, population decline and an ageing population. The last 20 years have seen considerable activity to come up with new demographic policy measures to mitigate these adverse trends, with single solutions developed for all regions in Russia. This paper presents the results of a study where cluster analysis was applied to enable the identification of groups of regions with significant differences in the dynamics of socio-demographic indicators. We used hierarchical cluster analysis to classify and group Russian regions on the basis of social and economic development indices for 2002 and 2008. The validity of the profiling was confirmed using parametric and non-parametric tests. The analysis identified three clusters of Russian regions. These clusters have significant differences in socio-demographic indicators and the associated dynamics. The results of our analysis identified 'growth points' for each cluster: the fertility correlates that should be factored into the development of effective demographic policy measures.

INTRODUCTION

Since the early 1990s, Russia has been grappling with a demographic crisis. It is most evident through a mortality rate that exceeds the birth rate, population decline and an ageing population. According to forecasts by official Russian statistics body Rosstat, between 2015 and 2030, the rate of natural increase in Russia will fall from 0.6‰ to -3.4‰ (Figure 1).

There are several key reasons for low fertility in Russia. These include: a prevailing social norm of having few children, a lack of stability guarantees, structural factor influences (a declining proportion of women of a reproductive age), and an extended period of time when the authorities paid no attention to this problem (Klupt 2008).

Figures 1: Rate of natural increase in Russia (‰): medium variant of projection by official Russian statistics (Births, deaths and natural increase 2014)

The end of the 20th century and the start of the 21st century saw an uptick in the development of new demographic policy measures aimed at mitigating these negative trends. However a single set of measures aimed at boosting fertility was proposed for different Russian regions. The most innovative was the introduction of the so-called 'maternity capital' – a one-off lump-sum amount paid for the birth of the second child. When it was introduced in 2007, this payment amounted to $9,755; in today's terms (to account for the change in the rouble/dollar exchange rate), this is equal to $7,215. Under Russian law, this money can be spent on improving living conditions, on the child's education or on the mother's retirement savings. Other noteworthy fertility incentives include 140-day fully paid maternity leave; part-paid maternity leave until the child is 18 months old; and unpaid maternity leave until the child turns 3. According to a number of studies, the introduction of lump sum maternity payments had an

impact on Russian birth rates, most of all in regions with a poor standard of living. In many parts of Russia, this money was enough to buy a flat, whereas in Moscow it would not get much further than a few square feet of living space.

Notably, there are important historical differences between the various parts of Russia as regards socio-economic parameters and overall standards of living. Thus the 2013 birth rate in Russia was 13.2‰, but the recorded minimum was 8.8‰ in Leningrad Region and 26.1‰ in Tyva Republic. The average gross regional product for 2013 was $11,930 per capita, with a minimum of $2,804 in the Republic of Chechnya and a maximum of $45,075 in Tyumen Region. Clearly, the use of a single set of measures to address demographic problems in different parts of Russia cannot be effective. The applied measures should be tailored to match the needs of groups of regions that are in similar demographic situations and have common demographic dynamics. We believe that the statistical methodology of cluster analysis could be an appropriate instrument for doing this.

Cluster analysis is hardly used today in the development of differentiated demographic policies. Instead, there is a preference for purportedly 'one size fits all' mechanisms to boost fertility, despite the fact that such measures do not exist. Every country and every region should come up with a set of measures that reflect the unique combination of economic, political, socio-cultural, psychological and religious specifics of the local population. Unfortunately, the particular nature of these factors and their possible effects on the impact of demographic policy are not taken into consideration in Russia today. We believe that this could relate to a number of factors: 1) insufficient attention from the authorities to the results of demographic research in Russia; 2) variability between Russian regions, considerable differentiation in their economic and socio-cultural statuses; 3) the absence of the requisite analytical skills among officials implementing demographic policies around Russia.

Notably, in a broader context, cluster analysis is often used for the segmentation of regions. For example, O. Simpach and J. Langhamrova used it to study the impact of ICT growth on households and municipalities in regions of the Czech Republic (Simpach and Langhamrova 2014). A. Repkin identified clusters of Asian countries on the basis of economic indicators (Repkin 2012); F. Kronthaler identified clusters of German regions on the basis of economic capability (Kronthaler 2005). O. Simpach applied cluster analysis to demographic development indicators to segment municipalities (Simpach 2013). The results of cluster analysis are used today to develop "cluster strategies", which "have become a popular economic development approach among state and local policymakers and economic development practitioners" (Cortright 2006).

Our research aims to identify groups of Russian regions that share common demographic characteristics and dynamics. They would thus benefit from the

implementation of similar demographic policy measures that are nevertheless tailored to the needs of each individual regional cluster.

DATA AND METHODS

1. We performed a hierarchical clustering procedure to identify groups of regions characterised by similar fertility trends. In carrying out this analysis, we undertook activities typical for this type of statistical work: the selection and transformation of input variables; the selection of distance measures and linkage rules; the selection of the method of clustering; the selection of the number of clusters; the profiling of clusters and interpretation of the attained results.

2. We used hierarchical cluster analysis in our research. We used squared Euclidean distance as the distance measure and Ward's method to gauge distance between clusters. The decision on the number of identified groups of regions was taken on the basis of:
   ✓ Graphical representation of the clustering process (we examined a dendrogram);
   ✓ A coefficient showing the distance between the linked clusters;
   ✓ An evaluation of the between-group and within-group variability;
   ✓ Cluster size (we tracked the number of regions that form a single cluster to ensure that each group contained a sufficient number of regions).

3. To assess the stability and validity of the cluster solution, we performed several iterations of the clustering procedure, using different measures of distance between objects. Moreover, we applied partitioning methods of clustering (k-means procedure). For most Russian regions, their allocation into homogenous groups coincided. Some differences in cluster composition did not skew the profile of each identified group of regions. The shared characteristics and relationships identified in the course of the analysis did not change when different distance measures were used.

4. We used both the clustering variables and other variables that describe the socio-economic development in a region to interpret the clusters themselves. We examined cluster centroids, calculated the clustering variables' average values for all objects in a certain group of regions and executed tests of significance in difference between two means. We applied one-way analysis of variance (ANOVA) to evaluate differences between the means and Levene's test to assess the homogeneity of variances. To test the assumption of normality, we used the Shapiro-Wilks test.

5. To carry out cluster analysis, we selected five variables that provide some indication of the demographic situation in a region. The first four variables are well-known demographic indicators, presented in official Russian statistics by region:
   1) Birth rate;
   2) Perinatal mortality rate;
   3) Infant mortality rate;
   4) Under-five mortality rate.

As an additional clustering variable, we introduced a calculated value: pregnancy rate. This is the number of completed or terminated pregnancies per 1,000 women aged 15-49. We believe that this variable enables us to assess the relative level of reproductive activity within a particular community. Thus we believe that the coefficient of pregnancy rate, along with other demographic indicators could be used as an informational basis for developing effective demographic policy.

Another variable that also characterises the demographic situation is the number of abortions per 100 births. The Russian State Statistics Service provides this data both at a country and a regional level, enabling us to use it in our analysis. However we did not use this variable for our clustering, as it clearly correlates to other variables in our study, including pregnancy rate. At the same time, we used the number of abortions per 100 births for subsequent interpretation of the identified clusters.

6. Throughout the clustering, we did not use the variables themselves but rather their indices: values that describe changes in the demographic variables. We believe that studying the dynamics of demographic processes, rather than their static values, enables a more profound understanding of the demographic situation in a region. Similarly, the identification of regional groups that have similar trends in the development of demographic processes is more justified than basing the groups on attained (transpired) demographic indicators. In our view, developing state fertility support measures that account for the dynamics of population replacement enables greater effectiveness. We also note that since all of the input variables we drew on in the clustering were indices, we did not need to standardize the variables beforehand.

7. To carry out the hierarchical clustering procedure, we used variables that characterise the demographic situation in Russian regions between 2002 and 2008. This was not an accidental choice: active state support and the incentivizing of fertility started in Russia in 2000. In response to dramatic depopulation challenges, the government developed a new Demographic Policy Concept and began its implementation in late 2000-early 2001. Thus we felt it would be quite fitting to analyse the possible results of this new demographic policy from 2002. The subsequent years – up to and including 2008 – can be described as a period of relative macroeconomic stability. However the financial crisis that hit most intensely in early 2009 had a negative impact on reproductive indicators in Russia. We believe that these "fertility shocks" should be excluded from the sample in our analysis and studied as a separate phenomenon. Thus our analysis focused on data for 2002-2008.

8. For the purposes of the cluster analysis, our study sample included all Russian regions that had complete data for all input variables. We note that the time between 2003 and 2008 saw a great deal of political and economic activity around the administrative consolidation of neighbouring territorial units with tight economic links. As a result of this process, the number of Russian constituent federal parts went from 89 to 83. For our hierarchical clustering procedure, we only selected regions that were not affected by the consolidation processes. Thus 78 Russian regions were included in the cluster analysis.

RESULTS

In performing the cluster analysis, we settled on a three-cluster solution. The first cluster included 37 Russian regions; there were 14 regions in the second and 27 in the third (Figure 2). The evaluation of the cluster centroid confirmed the appropriateness of these three groups: the mean and median values of the cluster variables differed significantly between the identified clusters and, in most cases, when compared to the nationwide levels (table 1). Levene's test confirmed the validity of the one-way ANOVA (table 2). The Shapiro-Wilks test confirmed that each of the levels of the variables is normally distributed (table 3). Results of the one-way ANOVA are presented in Table 4.

Table 1: Statistical characteristics for the cluster variables in 2008

		Cluster 1	Cluster 2	Cluster 3
Birth rate	Mean	11.9	12.3	13.3
	Median	11.5	11.8	12.5
Perinatal mortality rate	Mean	7.6	12.6	8.9
	Median	7.1	12.0	8.5
Infant mortality rate	Mean	8.5	9.2	8.5
	Median	8.1	8.9	8.2
Under-five mortality rate	Mean	8.3	12.5	8.4
	Median	8.1	11.7	7.9
Pregnancy rate	Mean	83.2	87.5	75.6
	Median	79.9	82.3	75.0

Table 2: Test of homogeneity of variances

Variables	Levene Statistic	df1	df2	Sig.
Birth rate	0.770	2	75	0.467
Perinatal mortality rate	0.045	2	75	0.956
Infant mortality rate	2.301	2	75	0.107
Under-five mortality rate	2.049	2	75	0.136
Pregnancy rate	1.367	2	75	0.261

Table 3: Test of normality

Variables	Clusters	Shapiro-Wilk		
		Statistic	df	Sig.
Birth rate	1	0.966	37	0.319
	2	0.958	27	0.331
	3	0.919	14	0.213
Perinatal mortality rate	1	0. 970	37	0.413
	2	0. 964	27	0.457
	3	0.905	14	0.135
Infant mortality rate	1	0.958	37	0.170
	2	0.950	27	0.220
	3	0.902	14	0.120
Under-five mortality rate	1	0.943	37	0.059
	2	0.975	27	0.742
	3	0.879	14	0.056
Pregnancy rate	1	0.976	37	0.581
	2	0.967	27	0.520
	3	0.932	14	0.328

Table 4: ANOVA

	Between Groups	Within Groups	Total
Birth rate			
Sum of Squares	31.399	105.228	136.627
df	2	75	77
Mean Square	15.7	1.403	
F	11.19		
Sig.	0.000		
Perinatal mortality rate			
Sum of Squares	275.274	46.521	321.795
df	2	75	77
Mean Square	137.637	0.62	
F	221.893		
Sig.	0.000		
Infant mortality rate			
Sum of Squares	61.616	14.029	75.645
df	2	75	77
Mean Square	30.808	0.187	
F	164.699		
Sig.	0.000		
Under-five mortality rate			
Sum of Squares	160.622	33.262	193.884
df	2	75	77
Mean Square	80.311	0.443	
F	181.09		
Sig.	0.000		
Pregnancy rate			
Sum of Squares	2870.706	8268.406	11139.112
df	2	75	77
Mean Square	1435.353	110.245	
F	13.02		
Sig.	0.000		

The results of the clustering and the subsequent interpretation showed that the clusters differ in:

1) The values of the evaluated demographic variables;
2) The nature of the dynamics of the evaluated demographic variables;
3) The level of economic development.

Cluster 1 – "Low fertility amid low economic activity"
This is the largest cluster comprising 37 regions. It covers approximately 44% of the territory of Russia and accounts for around half of the country's population. Its most outstanding demographic characteristics include:

– for the entire period of the study, the fertility rate in the regions of this cluster was below that of the other two clusters and lower than the national average;

– the pregnancy rate for the entire period was below the national average;

– the number of abortions per 100 births was high and exceeded the national average.

Moreover, regions in the first cluster also shared the following characteristics of the index dynamics:

– a growth in the fertility rate (in 2002 this was 9.7‰, rising to 11.9‰ in 2008);

– a decline in the number of pregnancies (88.7‰ in 2002 to 84.2‰ in 2008);

– a decline in the number of abortions (143 abortions per 100 births in 2002; 89 abortions/100 births in 2008);

– a certain lag behind national dynamics (for example, the total fertility rate in regions of this cluster for the studied period grew by 22%, while the increase across all of Russia was 25%; the decline in abortions per 100 births for this cluster fell by 38% compared to 42% in Russia overall).

Interpretation of this cluster through variables that reflect the level of economic development showed that this cluster includes a dominant share of people employed in the economy, and also produces the greatest volume of manufacturing. This cluster is also characterised by the largest share of commissioned residential buildings (almost half of the total quantity for Russia). However, despite the significant size of the first cluster, a number of economic indicators were not particularly high. These included the value of fixed assets, volume of natural resource production, import and export turnover, contribution to GNP and gross national income. It can be said that regions of the first cluster are not economically active as regards a number of key economic factors.

Cluster 2 – "Cautious" fertility amid high economic activity"
This cluster comprises 14 Russian regions, which cover a quarter of the total area of the country. The total population accounts for one-fifth of the total population. The key characteristics of this cluster include:

– the highest pregnancy rate across the clusters;

– the lowest number of abortions per 100 births across the clusters;

–   the highest perinatal and infant mortality rates across the clusters;

–   a higher level of under-five mortality compared to other clusters.

Other characteristics of the demographic situation in regions of this cluster were not generally outstanding and were quite similar to pan-Russian trends. Thus the total fertility rate for this cluster for the entire study period virtually matched the national rate; there was an overall tendency towards a decrease in mortality rates, in line with the overall trend for Russia.

Interpretation of the cluster through variables related to economic development showed that this cluster did not have a significant share of people employed in the economy (around one-fifth of the total active population). At the same time, the cluster leads on contribution to GNP (the total gross regional product for this cluster was approximately 42% of GNP). Regions of this cluster also account for a majority stake in natural resource production industries (around 60% of total Russian volumes), substantial fixed asset value, and export and import turnover. Moreover, the share of investments in fixed assets in this group of regions was almost one-third of the figure for Russia. Thus as far as key economic indicators go, this cluster can be seen as relatively successful and extremely vigorous as regards economic activity. The economy in this cluster is largely geared towards innovative development.

Cluster 3 – "High fertility amid economic passivity"
This cluster includes 27 Russian regions and comprises around one-third of the Russian population. Its key characteristics include:

–   a fertility rate that exceeded that of other clusters and the national average for the duration of the study;

–   a pregnancy rate that was below the other clusters and below the national average for the duration of the study.

Moreover, regions in this group were also characterised by the following demographic dynamics:

–   the highest growth in fertility rates among the clusters;

–   an insignificant decline in the pregnancy rate;

–   a significant decrease in the number of abortions: from 111 abortions per 100 births in 2002 to 67 abortions per 100 births in 2008.

Other demographic characteristics in this cluster were not particularly remarkable and were very close to overarching trends for Russia.

Interpretation of this cluster through economic development variables showed that this group accounts for approximately one-third of people employed in the economy. Most of the analysed economic indicators had the lowest values in this cluster. Thus the share of the gross regional product was approximately one-fifth of GNP. Natural resource production volumes, fixed asset value, retail trade turnover, and import and export volumes were similarly low compared to the other

clusters. At the same time, as regards a number of other economic factors, the third cluster was an undisputed leader. This concerns agriculture (43% of the total volume for Russia), plant products (45%) and animal products (41%). On the whole, the third cluster is characterised by a certain economic passivity and even depression against a backdrop of considerable agricultural activity.

DISCUSSION

The results of our cluster analysis show that regional specifics should be factored into fertility research. This is hardly surprising, given that there is no single theory for stimulating fertility in the modern world. On the contrary, there are a great many diverse approaches for mitigating low fertility rates being implemented around the world, with varying degrees of success. Moreover, the replication of one country's (or one region's) successful strategy in another location is unlikely to be effective. The specific nature of the demographic, economic and political trends of a country or region, the nuances of local social and moral norms and processes should be taken into account in the development of fertility incentive frameworks.

However in Russia, maternity capital payments for the birth of the second child remain the key measure for stimulating fertility. We are inclined to criticise this measure for the following reasons. Firstly, a two-child family should not be the priority for Russia at this point in time. Between January and May 2015, the natural decrease was several times higher than in previous years. The economic crisis is likely to have exacerbated this adverse dynamic. Undoubtedly, it will continue to have a delayed negative impact in the future as well. The situation is made worse by changing age structures. According to Rosstat, by 2025, the number of women of childbearing age in Russia will fall by 7.2%, and by 2030 – by 10.3%. The share of women of fertile age in the population will decrease from the current 21.6% today to 19.2% in 2030. In this context, state policy should focus on increasing the number of three and four-child families, rather than two-child ones. Secondly, rewarding the quantity of children without regard for their "quality" will undoubtedly lead to deterioration in the quality of the population: a less healthy nation with lower levels of educational and cultural standing. Thirdly, financial incentives for the mere act of childbirth without accounting for subsequent activities around children's upbringing and socialisation suggests that the state is simply interested in having more citizens, irrespective of their development and that state policy does not view parenting as a multi-faceted process that is difficult to carry out well.

The results of our analysis have allowed us to suggest measures that would be effective in boosting fertility and parenting in each cluster. For instance, the first cluster would benefit from better socio-economic conditions for childbirth and parenting at a regional level. In particular, this should focus on improving

organisational and medical facilities for families, like those envisaged in the Russian state Demographic Policy Concept through 2025 (Demographic Policy Concept 2007):

1) Improved accessibility and quality of free medical services for women for pregnancy and childbirth, and for newborn children;

2) Improved material and human resourcing for motherhood and childhood services;

3) The development of high-tech medical assistance for women during pregnancy and childbirth, and for newborn children;

4) The introduction of integrated measures for further reduction in the number of abortions.

Undoubtedly, these measures are important for all territories, but the results of our analysis showed that they are of the highest priority for regions in the first cluster.

Let us furnish another example of the use of our analysis for regions of the second cluster. To stimulate fertility here, we propose introducing regional tax concessions for parents and creating an integrated set of special legislative conditions to support parenting at the regional and municipal levels. The innovation-focused socio-economic development of these regions will drive the evolution of skills like logical and abstract thinking, an analytical mind-set, the ability to filter and process large volumes of information and an aptitude for creativity. This will lead to more complex, differentiated and diverse work, in turn prompting changes in the structure and quality of workers and expectations about their professional competencies and traits like accountability, self-motivation, creativity and overall effort at work. We believe that these are the very regions that would benefit from a system of training for parents aimed at developing special competencies: the skills required to operate, build relationships and work in an innovation-centric economic. Parents should be taught how to help their children be creative, develop analytical skills, cope with multitasking and so on: that is, to establish the very personal qualities that will be sought after in regions of the second cluster in the near future.

Thus the results of our cluster analysis across Russian regions helped us to identify high-priority instruments for stimulating fertility and parenting, which are shaped to the needs of each type of region. Applying these as part of regional strategies enables:

1) Creating a tailored approach to mitigate mismatches between actual and forecast population numbers and composition in a particular region;

2) Developing and testing the legislative, economic and methodological bases for using a set of instruments to stimulate fertility with a defined set of regions;

3) Targeting the demographic situation in different regions with account of the identified specifics;

4) Creating conditions for the growth of birth rates in different types of regions through the use of different types of incentive measures

CONCLUSIONS

Our research showed that countries composed of many constituent parts with a high degree of variability as regards socio-economic and demographic development and the nature of the dynamics of key socio-economic indicators require demographic policies that are differentiated by region type. The regional groups should be identified on the basis of static key indicators as well as the nature of the dynamics of key indicators. Cluster analysis is an effective analytical instrument for carrying out regional segmentation in order to develop tailored demographic policy measures. Indeed, on the basis of analysis that highlighted key problems in each cluster of regions, we proposed a redistribution of resources allocated to implementing demographic policy measures. We tried to identify the measures that would not only allow regions to improve fertility in the future, but also raise the qualitative indicators of the population.

We expect to extend our research by further differentiating the regions within each cluster. Thus, there is scope to identify sub-clusters with indicators above and below median values – both for input variables and other variables related to the assessment of the socio-economic situation in the regions. This will allow ascertaining the most challenging Russian territories in each segment, which require priority measures for stimulating fertility and supporting parenting. These most disadvantaged of regions could become pilot projects for focusing state policy in the short term.

ACKNOWLEDGMENTS

REFERENCES

Births, deaths and natural increase. 2014. Demographic Projections by the Federal State Statistics Service. Rosstat, Moscow.

Cortright, J. 2006. *Making Sense of Clusters: Regional Competitiveness and Economic Development*. The Brookings Institution, Washington, D.C.

Demographic Policy Concept of the Russian Federation until 2025. 2007. Approved with Presidential Decree from 9 October 2007 No. 1351.

Klupt, M.A. 2008. *Demographics of regions of the world*. Piter, Saint Petersburg.

Kronthaler, F. 2005. "Economic Capability of East German Regions: Results of a Cluster Analysis". *Regional Studies*, Vol 39, Issue 6, 739-750.

Repkine, A. 2012. "How Similar Are the East Asian Economies? A Cluster Analysis Perspective on

Economic Cooperation in the Region". *Journal of International and Area Studies*, Vol 19, Issue 1, 27-44.

Simpach, O. 2013. "Application of Cluster Analysis on the Demographic Development of Municipalities in the Districts of Liberecky Region". In *Conference Proceedings of the 7th International Days of Statistics and Economics (*Prague, Chezh Republic, Sept. 19-21). Melandrium, 1390-1399.

Simpach, O. and Langhamrova, J. 2014. "The Impact of ICT Growth on Households and Municipalities in the Czech NUTS-3 Regions: the Application of Cluster Analysis". *Schriftenreihe Informatik*, Vol 43: IDIMT-2014: Networking Societies - Cooperation and Conflict, 63-70.
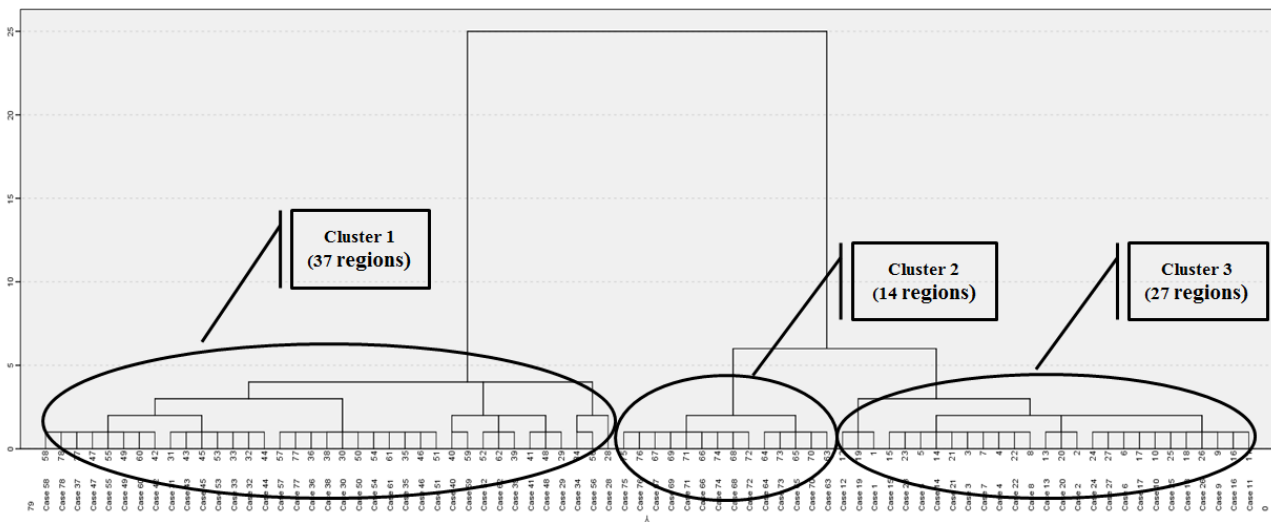
AUTHOR BIOGRAPHIES

OKSANA SHUBAT is an Associate Professor of Economics at Ural Federal University (Russia). She has received her PhD in Accounting and Statistics in 2009. Her research interests include demographic processes, demographic dynamics and its impact on human resources development and the development of human capital (especially at the household-level). Her email address is: o.m.shubat@urfu.ru and her Web-page can be found at http://urfu.ru/ru/about/personal-pages/O.M.Shubat/

ANNA BAGIROVA is a professor of economics and sociology at Ural Federal University (Russia). Her research interests include demographical processes and their determinants. She also explores issues of labour economics and sociology of labour. She is a doctoral supervisor and a member of International Sociological Association. Her email address is: a.p.bagirova@urfu.ru and her Web-page can be found at http://urfu.ru/ru/about/personal-pages/a.p.bagirova/

MAKHABAT ABILOVA is an Associate Professor of Economics at Magnitogorsk State Technical University (Russia). She has received her PhD in Economics in 2010. Her research interests include the analysis of human capital and statistical methods of demography. Her email address is: abilova.mahabat@yandex.ru

ANTON IVLEV is an Associate Professor of Economics at Magnitogorsk State Technical University (Russia). He has received his PhD in Economics in 2011. He explores issues of labour economics and financial model of education. His email address is: ivlevanton@bk.ru

Figures 2: Dendrogram using Ward Linkage