STATISTICAL ANALYSIS OF THE MAIN CAUSES OF REGIONAL DEVELOPMENT PROCESSES

Prof. Begalov Bakhodir Abdusalomovich

Director of Statistics Agency under the President of the Republic of Uzbekistan

Email: b.begalov@stat.uz Phone: (+99871) 230-80-06

Dr. Hurriyat Khudoykulova Berdiyeva

Senior lecturer at the National

University of Uzbekistan named after Mirzo Ulugbek

Email:hxb-2004@mail.ru

Phone: (+99890) 008-50-95

Abstract

The paper deals with the potential use of statistical methods in defining a framework for regional socio-economic development processes. The spatial context of the study is Uzbekistan. The paper uses factor, cluster analysis, and comparative analysis. The experience of Uzbekistan in the application of these methods to the grouping of local entities by homogeneity is particularly relevant for the purposes of comparison. The topicality of the research is determined by the need to define an effective territorial or spatial framework for Uzbekistan's regional socio-economic development processes, and to ensure that the framework will facilitate access to other countries, programs, and projects.

JEL classification**:** R11, R14, C49

Keywords: Principal component analysis, cluster analysis, composite index, regional socio-economic development.

1.

2. INTRODUCTION

Geographical location, climatic conditions, demographic situation, development history and other factors have a significant impact on the economic development of the region. However, poverty in the economic development of the regions is likely to increase if there is high inequality and there is no change in income. This has a negative impact on the socio-economic development of the country.

Therefore, each country strives to pursue a regional policy aimed at improving living standards in developing countries, that is, creating conditions and

opportunities for their socio-economic development, thus contributing to improving the quality and level of human potential.

To analyze the socio-economic development of the regions, radically improve the forms and methods of organizing work, the quality and living standards of the population, and other important tasks. A major objective of the development programs launched in Uzbekistan is to bring balanced regional development. To achieve this goal, the economic planning in the country has traditionally been focused on the need to provide special support to disadvantaged areas.

Although the country remains on course to achieve its socio-economic development goals, related challenges such as inequality and regional disparities persist. There were several causes leading to the distortion of balance in all regions specifically, the rate of development in the initial years of the market economy, the attractiveness of the place to investors, the economic-geographic location of the area, the condition of infrastructure, and features of innovation.

3. LITERATURE REVIEW

The fact that the subject of variations in regional development in recent years is considered a topical issue is of great interest to geographic and Economic Sciences, this situation can be seen in the rapid increase in the number of publications and literature on this topic.

The chief present-day problem of socio-economic development, in geographical-economic terms, is growing spatial inequality when viewed from a regional approach. In recent years, regional disparities have become of great interest to geographical and economic sciences, as manifested by a fast-growing number of publications on the subject. The empirical-methodological research conducted by Alois Kutscherauer (2010), Jaba E(2009), Ivana Bakaric (2005), Domański (2005), Henley (2005), Kosfeld et al. (2006), Rodriguez-Lopez et al. (2009), Czyż& Hauke (2011), Otsuka et.al (2016), Pratap Kumar (2019), by Agnieszka (2019) and others conducted effective scientific research within this topic. According to the results of such studies, it was found that regional policy measures will have a positive effect if they are implemented only in regional units with homogeneous social and economic development trends in order to have a desired and equal effect in a specific area.

4. RESEARCH METHODOLOGY AND DATA

To ensure the effectiveness of regional policy and rapid socio-economic development of the regions, it is first necessary to assess the current socio-economic status of the regions. Assessing the possibility of socio-economic development of regions means determining the system of appropriate indicators and assessing the current state of socio-economic development in the region with the help of identified indicators.

Assessment of the state of socio-economic development of regions according to individual indicators - independent description of their socio-economic development and forecasting of future development is one of the possibilities. Although this method is considered one of the optimal methods, it is not suitable for rapid and comparative assessment of the state of development [63, 78, 80]. Because it is difficult to determine the direction according to many indicators, and the level of accuracy of the analysis will be low. In addition, the results of the analysis of individual indicators do not express relationships and dependencies. In short, a number of indicators analyzed separately do not provide an integrated and clear representation of reality.

Socio-economic development is a multi-dimensional process that cannot be fully assessed by a single indicator. This requires the construction of a composite index (rating) of socio-economic development based on the optimal combination of various development indicators.

Steps in the construction of composite indicators[3].

- *Theoretical framework.* A theoretical framework should be developed to provide the basis for the selection and combination of single indicators into a meaningful composite indicator under a fitness-for-purpose principle.

- *Data selection.* Indicators should be selected on the basis of their analytical soundness, measurability, country coverage, relevance to the phenomenon being measured and relationship to each other. The use of proxy variables should be considered when data are scarce.

- *Imputation of missing data.* Consideration should be given to different approaches for imputing missing values. Extreme values should be examined as they can become unintended benchmarks.

- *Multivariate analysis.* An exploratory analysis should investigate the overall structure of the indicators, assess the suitability of the data set and explain the methodological choices, *e.g.* weighting, aggregation.

- *Normalisation.* Indicators should be normalised to render them

comparable. Attention needs to be paid to extreme values as they may influence subsequent steps in the process of building a composite indicator. Skewed data should also be identified and accounted for.

– *Weighting and aggregation.* Indicators should be aggregated and weighted according to the underlying theoretical framework. Correlation and compensability issues among indicators need to be considered and either be corrected for or treated as features of the phenomenon that need to be be retained in the analysis.

– *Robustness and sensitivity.* Analysis should be undertaken to assess the robustness of the composite indicator in terms of, *e.g.,* the mechanism for including or excluding single indicators, the normalisation scheme, the imputation of missing data, the choice of weights and the aggregation method.

– *Back to the real data.* Composite indicators should be transparent and fit to be decomposed into their underlying indicators or values.

– *Links to other variables.* Attempts should be made to correlate the composite indicator with other published indicators, as well as to identify linkages through regressions.

– *Presentation and Visualisation.* Composite indicators can be visualised or presented in a number of different ways, which can influence their interpretation.

The analysis of regional development should contain all indicators the monitoring and evaluation of which is of great importance for determination of the development potential. It regards mainly indicators characterizing[13]:

– Overall description of a region

– Economic situation in a region

– Human potential in a region

– Social level

– Infrastructure

– Environment

– Industry

– Services

– Countryside and agriculture

The study measures and compares the levels of socio-economic development of different regional administrative divisions of Uzbekistan (i.e., 12 regions (Andijan, Bukhara, Jizzakh, Kashkadarya, Navoi, Namangan, Samarkand, Surkhandarya, Syrdarya, Tashkent, Fergana, Khorezm), one autonomous

republic (Republic of Karakalpakstan) and one independent city-the capital of Uzbekistan (Tashkent city)) based on socio-economic indicators. The data used in this analysis have been taken from the State Committee of the Republic of Uzbekistan on statistics. The analysis was carried out for the years 2018-2020, these being the years for which we had the latest regional statistical data.

The following methods of multivariate statistical analysis were used to analyze the socio-economic development of the regions of the country:

- A reduction of excessive number of variables, let us say a compression of information in a smaller number of immeasurable hypothetical magnitudes with as small loss of information as possible, which is contained in the monitored features. In this group of methods

the most frequently used approaches of factor analysis are the component analysis and common factor analysis. Since the results of factor analysis will be used as clustering variables (to which end factor scores must also be calculated), it is recommended to use the principal component analysis[6].

– An object typology – is an organization, let us say a hierarchical settlement in relatively identical groups, whereas neither their number is known nor defined precisely. At the same time, a sequence of these groups can be set according to the chosen criteria. For this task, it is very suitable to use cluster analysis.

According to the principal component features and the component index value we use this formula to calculate the composite score[8]:

$$CI = \sum_{i=1}^{n} (PCV_i * PCC_i) \qquad (1)$$

where:

CI – composite index,

PCV_i – *i*-th principal components variance contribution rate,

PCC_i – *i*-th principal component coefficients,

n – number of components.

Assign each of the composite *Z*-scores to a score in the range 0 to 1 by mapping to the cumulative normal distribution.

The inter-district variations are grouped into four categories of less developed region, moderate developed region, developed region and highly developed region on the basis of natural break (Jenks) method.

We use the non-hierarchical "*k*-means" clustering method to divide the objects being grouped into clusters. The main advantage of this cluster method is that it is more convenient to group the objects into clusters according to the measured

characteristic. Most clustering methods are based on distance or similarity matrices calculated from the underlying data rather than from the original data. Here we shall concentrate on distance measures, of which the most common is the Euclidean distance given by[2]:

$$d_{ij} = \sqrt{\sum_{l=1}^{n} (x_{il} - x_{jl})^2} \qquad (2)$$

where d_{ij} is the Euclidean distance for two individuals i and j,

each measured on n variables, $x_{il}, x_{jl}, l = 1, \ldots, n$.

Statistical data processing was conducted using SPSS software.

5. RESULTS AND DISCUSSIONS

5.1. Construction of composite index by Principal components analysis

Principal components analysis (PCA) is justified by the data set dimension (12 characteristics for the 14 regional administrative divisions), all the 12 variables being quantitative continuous. Using PCA the dimensionality of data is reduced by creating principal components from the original variables in the context of this study, principal components analysis is used to explore the original data set and to select the appropriate variables used to identify a regional profile of economic development in Uzbekistan.

To verify the adequacy of data for a factorial analysis, Bartlett's Test of Sphericity (to test the null hypothesis that the variables in the correlation matrix of the population are uncorrelated), and the indicator MSA (Measure of Sampling Adequacy) of Kaiser-Meyer-Olkin (to evaluate in which degree each variable may be predicted by all the other variables) were used. The results obtained by data processing with SPSS are presented in Table 1.

The significance level associated with Barlett's test of sphericity, *Sig* 0.00, is smaller than 0.05 (conventional value), which means the null hypothesis of variables' uncorrelation is rejected.

Table 1. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.68
Bartlett's Test of Sphericity	Approx. Chi-Square	188.73
	df	55
	Sig.	0.000

Source: Author's calculations.

Therefore one can conclude that the considered variables are adequate for a PCA. The value of the indicator MSA of KMO (0.68), greater than 0.5, also indicates the suitability of the considered data for factor analysis [6].

Table 2 represents the varimax rotated factor structure and the majority of the variables under study have been appropriately focused on the structure exposed by this factor matrix. The socio-economic communalities value varied from 0.56 for the Export per capita to 0.94 for real income per capita. Other remaining social and economic indicators are suitably represented in the form of two extracted factors. Two factors meet not only the eigenvalue criterion but also the variance proportion criterion. In social sciences, the lowest limit of acceptability is 60 percent of variance accounted by obtained factors [4]. This solution accounts for 79.9 percent of the total variance.

Table 2. Rotated Component Matrix

	Component		Communalities
	F_1	F_2	
Poverty rate (percentage)	-0.79	-0.49	0.86
Real income per capita (in thousand soums)	0.89	0.38	0.94
Industrial products per capita(in thousand soums)	0.83	-0.12	0.70
Services per capita (thousand soums)	0.64	0.70	0.90
Export per capita (in USD dollars)	0.64	0.39	0.56
Number of enterprises (per 1000 people)	0.86	0.42	0.91
Unemployment rate (percentage)	-0.81	-0.52	0.92
Share of households with gas supply (percentage)	0.79	0.21	0.67
Percentage of household having vehicles	0.35	0.71	0.62
Persons employed in trade. transportation and storage. housing and food services as percentage of total employment	-0.05	0.94	0.88
Persons employed in agriculture. hunting and fisheries as percentage of total	-0.43	-0.80	0.83
Eigenvalue	*7.34*	*1.45*	*-*
Percent of total variance	*66.74*	*13.20*	*79.9*

Source: Database of the State committee of Uzbekistan on statistics. author's calculations.

The first factor has a high positive factor loading on variables: per capita income, per capita volume of industrial products, per capita export volume, number of enterprises, and share of households with gas supply. This means that it positively correlates to the respective characteristics of local government units. The first factor has a high negative factor loading on the unemployment rate and poverty rate.

The second factor has a high positive factor loading on the services per capita, the share of people employed in trade, catering, preparation, and supply about the total employed population, and the percentage of households having vehicles. A negative factor loading on persons employed in agriculture. hunting and fisheries as a percentage of the total.

By using the method of the composite index (CI) the level of development in terms of socio-economic development in Uzbekistan has been evaluated (table 3).

Table 3: Composite index value matrix of socio-economic indicators

Composite index	Level of development	Name of regional administrative divisions
0.26 -0.36	Very low	Surkhandarya(0.26), Kashkadarya(0.29), Namangan(0.33) Samarkand(0.30), Fergana(0.34).
0.36 -0.52	Low	Andijan(0.36), Rep.of Karakalpakstan(0.39), Khorezm(0.42), Jizzakh(0.46).
0.52-0.90	Moderate	Bukhara(0.52), Syrdarya(0.58), Tashkent(0.64).
0.90-0.95	High	Navoi(0.90).
0.95-1	Very high	Tashkent city (0.95).

Source: Author's calculations.

The relation between the value of composite score and the level of development is direct for instance regional administrative divisions. With the greater value of composite score have the advanced level of development and the state having lesser value recognize the lower level of improvement.

5.2.   Classification of Regions Using Cluster Analysis

In order to carry out the analysis first as the main hypothesis we focus on determining the characteristics of individual territorial units and dividing them into homogeneous groups using the methods of factor and cluster analysis. The factor analysis method summarizes highly correlated indicators from a set of real variables into a relatively small set of socio-economic indicators. Cluster analysis is used to divide similar regions into groups according to the level of socio-economic development. This makes it possible to implement regional policy

measures or monitor their consequences.

In social sciences, cluster analysis was recognized as the most suitable method of classifying units into groups of similar characteristics. Serving as input variables for cluster analysis were factor scores. Since the factor analysis resulted in two factors, for each observed region two factor scores will be calculated. For grouping objects into clusters. non-hierarchical clustering method, the "k-means" method was used. The main argument in favour of this clustering method is that this method of grouping objects into clusters is more suitable when grouping units (objects) on which specific characteristics were measured, and not

when grouping characteristics, i.e. variables [6]. The decision on the number of clusters is based on the analysis of variance (ANOVA) (table 4).

Table 4. ANOVA Results for Two Proposed Clusters

Factor	Means between clusters	s.s	Means within clusters	s.s	F ratio	p-value
Factor 1	2,439	5	0,101	8	24,167	0,000
Factor 2	2,397	5	0,127	8	18,856	0,000

At the given significance level of 5 percent and empirical significance level of 0.000 for factor one and 0.0000 for factor two, hypothesis H1 is accepted, i.e. we may say that the means between the six proposed clusters differ significantly (see Table 4 ANOVA results for six clusters). The results indicating grouping of regions into six different clusters are significant.

As this is confirmed by the significance test in ANOVA, the same can be verified by looking at the graph of means. i.e. by factor analysis identified dimensions of socio-economic development and means of an individual cluster (see Figure 1).
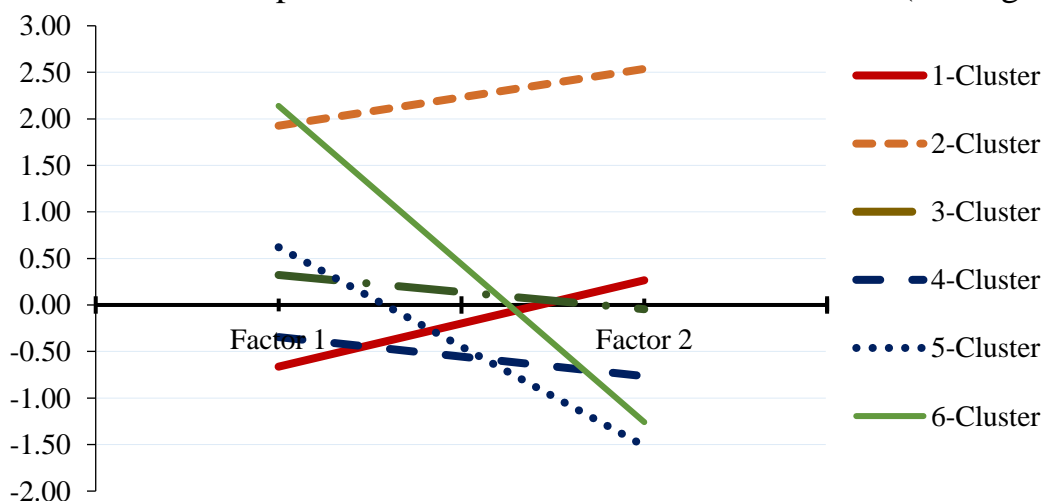

Figure 1. Means of each Cluster

According to the results of the cluster analysis Khorezm, Andijan, Fergana, Namangan, Samarkand, Kashkadarya regions and the Republic of Karakalpakstan belong to Cluster 1; Tashkent city to Cluster 2; Tashkent and Bukhara regions to Cluster 3; Jizzakh and Surkhandarya regions to Cluster 4; Syrdarya region to Cluster 5; Navoi region was classified as the regions belonging to cluster 6.

The conclusion is that in the regions included in cluster 1, according to the characteristics of factor 1, the income of the population, the number of enterprises and the volume of exports are low, the level of gas supply is low, and the industrial network is underdeveloped; the unemployment rate and high poverty rate are determined. According to the average correlation with Factor 2, Cluster 1 areas have a moderate share of the service sector, in particular, the share of people employed in trade, catering, preparation and supply, the share of households with a car, and specialization in agriculture expressed. This cluster includes 7 regions, which are Kashkadarya, Fergana, Khorezm, Andijan, Namangan, Samarkand regions and the Republic of Karakalpakstan.

Cluster 2 is characterized by a strong positive relationship with Factors 1 and 2. Regions belonging to this cluster have high incomes, large number of enterprises, high export volumes, high level of gas supply and high development of the industrial sector, and low levels of unemployment and poverty. It is also stated that the percentage of people employed in the service sector, trade, catering, preparation and supply, and the percentage of households with a car is high. This cluster includes only the city of Tashkent, the capital of the Republic of Uzbekistan.

Cluster 3 is expressed in average relationships with all factors, and we can evaluate the regions included in this cluster as having average development characteristics in terms of all indicators. They are the Tashkent and Bukhara regions.

Areas belonging to Cluster 4 although expressed in higher values than groups of Cluster 1 according to Factor 1 and groups of Clusters 5 and 6 according to Factor 2 have lower development levels according to general indicators. Jizzakh and Surkhandarya regions are included in this cluster.

Areas belonging to clusters 5 and 6 are represented in the places after cluster 2 (Tashkent city) in relation to factor 1. At the same time, according to factor 2, it differs sharply from the indicators of cluster 2 (Tashkent city). This means that the regions belonging to these cluster groups have Factor 1 characteristics of high

industrial employment. high level of economic development and low unemployment. At the same time. compared to other cluster members, the indicators of Factor 2 are the lowest in Factor 2 due to the high percentage of those employed in agriculture. hunting and fishing in relation to the total employed population and the low percentage of those employed in trade, catering, preparation and supply in relation to the total employed population. The indicator is recorded. Syrdarya region is included in Cluster 5 and the Navoi region in Cluster 6.

CONCLUSION

In the study, we have measured the development levels of different regional administrative divisions of Uzbekistan applying the composite index based upon an optimum combination of selected socio-economic development indicators. The association between developments of different sectors of the economy is assessed and the regional administrative divisions are ranked precisely according to their levels of socio-economic development. The level of development is assessed in overall socio-economic fields. All 14 regional administrative divisions have been included in the study and classified into four development categories according to the values of the composite indices.

Based on the results obtained through the cluster analysis of the state of development of the regions, we can say that by establishing branches of large-scale production enterprises and concerns, as well as universities and other public places in Kashkadarya, Fergana, Khorezm, Andijan, Namangan, Samarkand regions and the Republic of Karakalpakstan it is possible to form a demand for social and industrial infrastructure facilities, trade and catering facilities, etc. Also, there is a need to pay serious attention to the development of agriculture in these regions and to thoroughly study the factors affecting the development of the sector. Taking advantage of the competitive advantage of trade and catering establishments in the Syrdarya region, attracting investments to the sector, increasing the quality and speed of services, creating more opportunities for the development of agriculture in the Surkhondarya and Jizzakh regions, establishing industrial enterprises that process agricultural products, and the social development of the regions - can give an effective and quick result to the development of the economic situation.

REFERENCES:

1. Alois Kutscherauer. Regional disparities. Disparities in country regional development-concept. theory. identification and assessment–Ostrava 2010. Pp.120

2. Czyż. T., Hauke. J. (2011). "Evolution of regional disparities in Poland". QuaestionesGeographicae. Vol. 30 No.2. pp. 35–48. available at: https://cyberleninka.org/article/n/1430483

3. Elza Jurun, Snježana Pivac, Cluster and Multicriterial Comparative Regional Analysis – Case Study of Croatian Counties // Croatian Operational Research Review (CRORR). Vol. 1. 2010.pp.1-11

4. Hair. F. J., E. R. Anderson and L. R. Tahtam. 1987. Multivariate Data Analysis. NewYork: Macmillan Publishing Company. P.368

5. Henley. A.. (2005). "On regional growth convergence in Great Britain". Regional Studies. Vol.39 No.9. pp.1245–1260

6. Ivana Rasic Bakaric .Uncovering Regional Disparities –the Use of Factor and Cluster Analysis. // Economic Trends and Economic Policy-2005.P.24

7. Jaba E.. Balan C., Ionescu A. The Evaluation Of The Regional Profile Of The Economic Development In Romania. Analele Stiinłifice Ale Universităłii Alexandru Ioan Cuza" DIN IASI Tomul LVI StiinŃe Economice 2009

8. Jayanta Das. Tapash Mandal. Regional disparity in the level of development in West Bengal: A geographical analysis. // Review of Research. volume 7 issue 10. july – 2018. 1-17 p

9. Kosfeld R. Eckey H.F. &Dreger C., (2006). "Regional productivity and income convergence in Unified Germany 1992–2000". Regional Studies. Vol. 40 No.7. pp.755–767

10. Marija Burinskiene, Vitalija Rudzkiene. Comparison of spatial-temporal regional development and sustainable development strategy in Lithuania. //International Journal of Strategic Property Management.-2004. Pp. 150

11. Partha Dasgupta And Martin Weale . On Measuring the Quality of Life. World Development. Vol. 20 No. 1. 1992. 119-131p

12. Sabine L.. Brian S. E. . A Handbook of Statistical Analyses using SPSS. 2004 by Chapman & Hall/CRC Press LLC. pp.130.

13. Svatošová L.. Methodological starting points of regional development Analyses Agric. Econ. – Czech. 51. 2005 (2). p. 64–68.

14. Hurriyat Khudoykulova. Initial Statistical Analysis of Regional Development: Case Study Uzbekistan, Available at SSRN 4325271, 2021.

15. Hurriyat Khudoykulova. Formation Of Rating Indicators Based On The Results Of The Assessment Of The Level Of Socio-Economic Development Of Regions- Экономика: анализы и прогнозы, 2020, 93-105 p.

16. Handbook on Constructing Composite Indicators: Methodology and User Guide – ISBN 978-92-64-04345-9 - © OECD 2008

17. stat.uz- official website of the State Committee of the Republic of Uzbekistan on statistics.