Title:

Youth Hubs: Do They Strengthen Competition and Survival, or a Rebound in Birth Rates?

[Abstract]

This study aims to analyze the impact of the size and movement of the youth population (aged 20 to 39) on birth rates in South Korea. Based on data from two periods, 2012-2014 and 2021-2023, the analysis focuses on 204 regions, excluding Seoul, and examines the proportion of the youth population and birth rates. Using K-Means clustering, five groups were identified, and the performance was evaluated using Silhouette Score, DBI, and CHI metrics. Regression analysis revealed that while the youth population proportion positively influenced birth rates, the effect has decreased over time. Additionally, mapping the clustering results over both periods allowed for a clearer understanding of the relationship between regional population shifts and birth rates. This study provides essential data for formulating policy responses to address South Korea's declining birth rates.

[Background]

South Korea's youth population, aged 20 to 39, plays a crucial role in driving socio-economic changes, and their movement and population proportion have a direct impact on birth rates. South Korea's low birth rate problem is often referred to as a "birth rate shock," as the country has a lower fertility rate than even some war-torn nations, making it a global issue of significant concern. Understanding the age structure and regional characteristics of the population is critical in addressing this issue. Most previous studies have approached the problem from a perspective of balanced regional development, arguing that youth population concentration negatively affects birth rates in non-metropolitan areas.

However, this study proposes a different perspective, suggesting that youth population concentration could actually be an appropriate solution for improving birth rates. In contrast to the traditional view of balanced regional development, this research raises the question: should the declining youth population be evenly distributed across 229 regions, or should it be concentrated in specific areas to boost birth rates?

By utilizing data from Statistics Korea, this study leverages the strength of tracking youth population movement and concentration and their impact on regional birth rates. South Korea's precise population movement data allows for tracking not only inter-regional migration but also relocations within the smallest administrative districts. This enables a detailed analysis of population dynamics and supports evidence-based decision-making. Therefore, this study aims to utilize these data to conduct an in-depth analysis of the relationship between the proportion of youth populations and birth rates in different regions.

[Research Objective]

The objective of this study is to analyze the impact of youth population concentration on birth rates and to present a new perspective on birth rates that challenges the traditional view of balanced regional development. Unlike previous discussions that argue for the even distribution of the youth population across regions, this study tests the hypothesis that

concentrating youth in certain areas could actually increase birth rates. Through this, the study aims to demonstrate the crucial role that youth population movement and distribution play in addressing the low birth rate problem and to derive policy implications.

[Research Methodology]

This study aims to investigate the relationship between the size and movement of the youth population (aged 20-39) and birth rates in South Korea. Data from Statistics Korea, including the Resident Registration Population Status, Population Trends Survey, and Domestic Population Migration Statistics, were utilized. The analysis covered 204 administrative regions at the city/county/district level, excluding Seoul. The study was conducted for two periods: 2012-2014 and 2021-2023, examining changes in the proportion of the youth population and birth rates in each region. Variables used in the analysis include the proportion of the population aged 0-4, the in-migration rate of young adults aged 20-39, the out-migration rate, net migration rate, birth rate, youth population proportion, and total population.

[Statistical Analysis]

This study employed clustering techniques to identify patterns between the youth population (aged 20-39) and birth rates in each region. Several clustering algorithms were compared, and K-means clustering was found to have the best performance. The optimal number of clusters was determined to be five, based on the Silhouette Score and inertia values. Following this, a regression analysis was performed based on the main characteristics of each cluster, using Python 3.12.5 for the analysis. Finally, QGIS 3.34.1 was used to map the geographic visualization of each cluster.

[Results]

1. Cluster Performance Evaluation and Comparison Across Two Periods

1) Cluster Performance Evaluation

K-means clustering was used to group the regions into five clusters, and the performance of each clustering was evaluated as follows. Although the number of observations (n = 204) was relatively small, the Silhouette Score was reasonable. DBI and CHI evaluations indicated that the cohesion and separation of clusters were appropriately achieved for both periods.

Metric	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
2012~2014	0.390	0.868	272.07
2021~2023	0.396	0.824	291.99

2) Cluster Similarity Across Two Periods

The heatmap below illustrates the similarity between clusters from 2012-2014 and 2021-2023. It visually represents how clusters have changed between the two periods. Higher values indicate that regions within the same cluster during both periods remained in the same cluster. The Adjusted Rand Index (ARI) is 0.364, indicating a moderate level of agreement between

the clustering results of the two periods. This suggests that as time progressed, changes in the youth population proportion and birth rates in certain regions caused shifts in cluster composition.

2. Changes in Birth Rates According to the Youth Population (Aged 20-39)

The results of the regression analysis showed a significant positive correlation between the proportion of the youth population and birth rates. Specifically, regions with a higher proportion of youth showed an increase in birth rates. Both models exhibited R^2 values exceeding 50%, indicating statistically significant explanatory power.

- 2012-2014 Regression Equation: $y = 34.32x - 0.59$ ($R^2 = 0.54$, p-value < 0.0001)
  This equation indicates that when the youth population proportion increases by 0.1, the birth rate increases by approximately 3.432 births on average.
- 2021-2023 Regression Equation: $y = 18.31x + 0.35$ ($R^2 = 0.52$, p-value < 0.0001)
  This equation shows that when the youth population proportion increases by 0.1, the birth rate increases by approximately 1.831 births. Compared to the previous period, the influence of the youth population proportion on birth rates has diminished.

3. Scatter Plots and Geographic Visualization of Clusters Across Two Periods

1) Comparison of Cluster Characteristics Across Two Periods

The scatter plots below compare the average characteristics of clusters from 2012-2014 and 2021-2023. The comparison reveals that all clusters experienced a decline in the proportion of the youth population and birth rates over time. This suggests a simultaneous intensification of population aging in South Korea.

2) Geographic Visualization of Clusters Across Two Periods

This mapping visually represents the clusters from both periods. Despite the absence of geographic variables in the clustering analysis, the results reflect certain regional characteristics.

[Future Research Directions]

Future research will focus on identifying the infrastructure available in youth-dense areas, with a particular focus on regions that consistently fall into Cluster 4. The study will also include a separate analysis of Seoul's demographic trends and investigate whether insufficient infrastructure development in youth-concentrated areas has contributed to declining birth rates in recent years.

[Expected Academic Contributions]

This study challenges the current discourse on "balanced regional development" as a solution to low birth rates. Instead, it emphasizes that youth populations may need to concentrate in certain areas, with policies shaped around the types of infrastructure that attract and retain young people. South Korea's comprehensive demographic data allows for evidence-based decisions, and this research lays the foundation for future policy adjustments aimed at increasing birth rates.

Figure 1. Elbow Method for Optimal Cluster Counts and Silhouette Scores for Different Cluster Counts

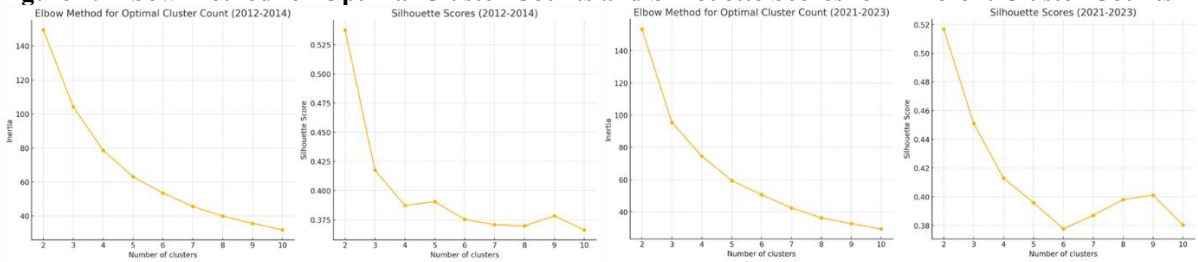


Figure 2. Heatmap Showing Cluster Similarity Between 2012-2014 and 2021-2023

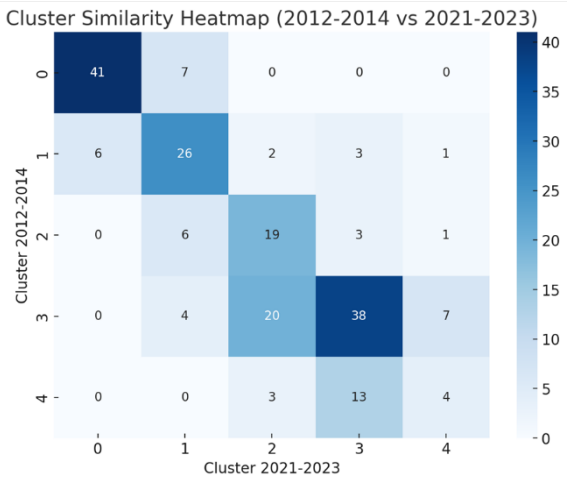


Figure 3. Regression Analysis of 20-39 Population Rate vs. Birth Rate (2012-2014 vs. 2021-2023)

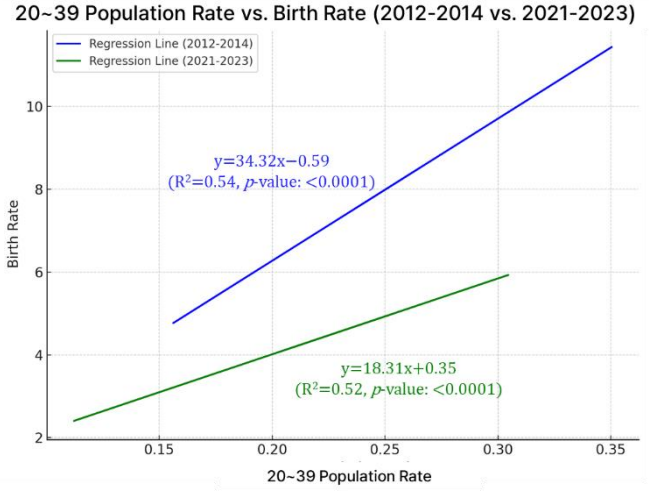
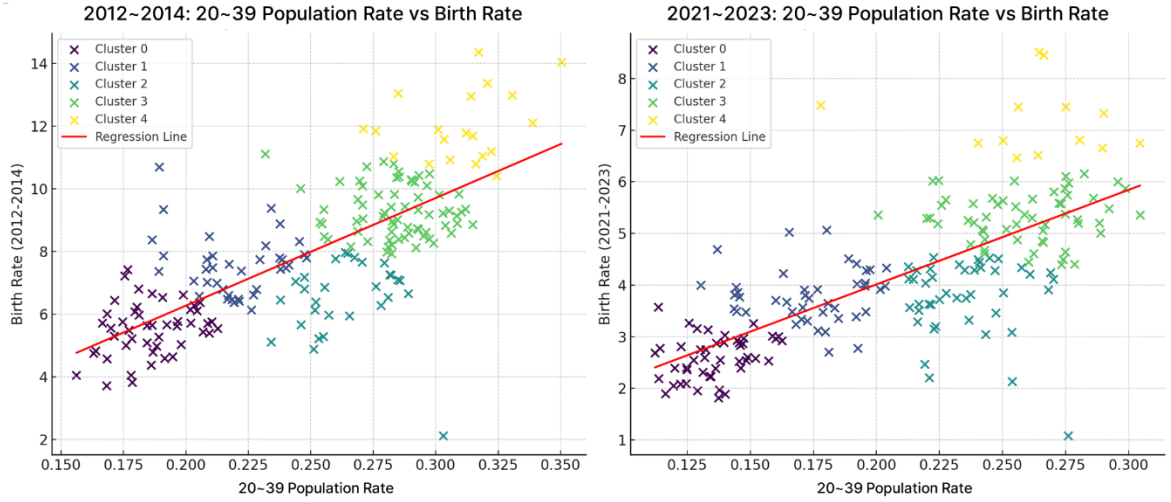


Figure 4. Scatter Plot of 20-39 Population Rate vs. Birth Rate by Cluster (2012-2014 vs. 2021-2023)



y=34.32x−0.59 (R^2=0.54, *p*-value: <0.0001)

y=18.31x+0.35, (R^2=0.52, *p*-value: <0.0001)

Figure 5. Cluster Means Scatter Plot (2012-2014 vs. 2021-2023)

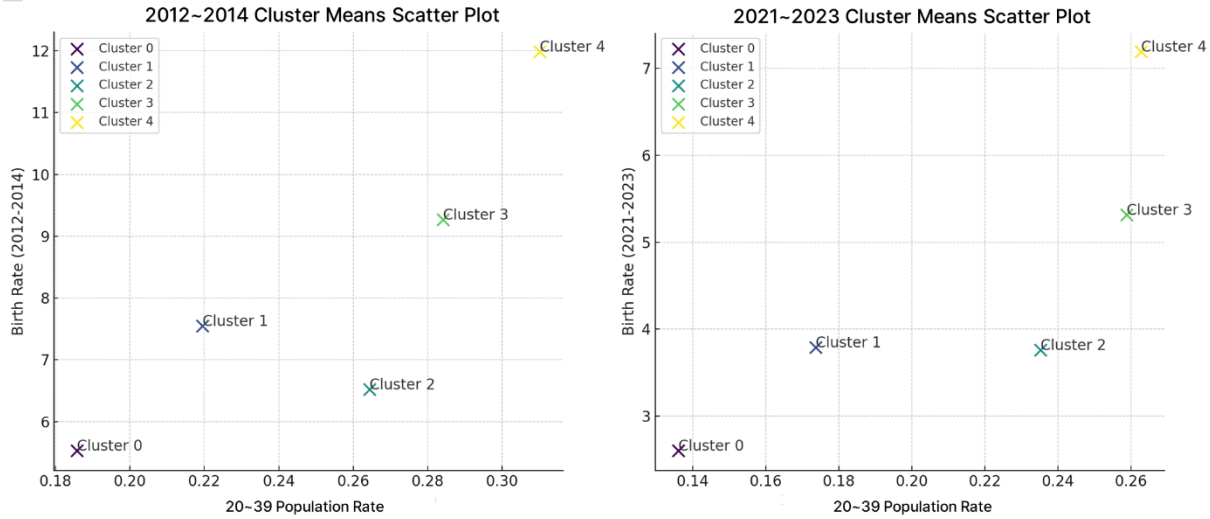


Figure 6. Geographic Visualization of Clusters (2021-2024 vs. 2021-2023)