Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3, 1810-1818
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i3.5692

Study on predicting low birth rates through data analysis on the relationship between various economic indicators and birth rates

Bong-Hyun Kim[1*]
[1]Department of Computer Engineering, 377-3, Musimseo-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, 28674, Republic of Korea; bhkim@seowon.ac.kr (B.H.K.).

Abstract: In recent years, the phenomenon of declining birth rates among younger generations has been increasing due to various factors such as rising education costs, job concentration, climate change, and a preference for single living. To address this issue, the government has implemented policies that provide various incentives to encourage childbirth. However, without fundamental data analysis on birth rates, the effectiveness of these policies cannot be reliably assessed. Therefore, this paper analyzes trends in newborn births based on the population of individuals in their 20s and 30s. The analysis employs correlation techniques to examine the relationship between various economic indicators and birth rates. The data was analyzed for the interrelationship and impact among the number of marriages, number of newborns, birth rate, consumer prices, economic growth rate, and population. Through this approach, we identify key economic factors that significantly influence birth rates and explore predictions regarding low birth rates. Finally, based on the research findings, we provide insights into policies aimed at promoting childbirth.

Keywords: *Birth rate, Correlation analysis, Data prediction, Economic indicators regression analysis.*

1. Introduction

Demographic changes, including shifts in population size, age structure, and geographic distribution, are reshaping societies worldwide. These changes pose challenges and opportunities for economic growth, social cohesion, and policy-making.

The world's population is projected to increase from 7.7 billion to 9.7 billion by 2050. This growth is uneven, with significant increases in developing regions and stagnation or decline in more developed areas. Also, many countries are experiencing a rise in the elderly population due to increased life expectancy and declining birth rates. This shift affects labor markets, social security systems, and healthcare services. There is a continuing trend towards urbanization, with more people living in cities than in rural areas. Urban areas are becoming central to economic activity but also face challenges related to infrastructure, housing, and sustainability [1, 2].

In recent years, OECD countries have experienced rapid changes in their population structure due to aging and declining birth rates. The proportion of the elderly population is increasing, which leads to a decrease in the working population and an increase in social burden. A decrease in the working population has a negative impact on economic growth [3]. As the proportion of the young working population decreases, the economically active population decreases, which leads to a decrease in productivity and a slowdown in economic growth. In addition, aging causes economic pressure by increasing health care costs and increasing the burden on pension systems. In such an aging society, medical and welfare costs increase. This places a great burden on government budgets and increases the tax burden on the younger generation [4, 5].

The most serious reflection of the change in population structure is the low birth rate. Low birth

History: Received: 7 January 2025; Revised: 10 March 2025; Accepted: 11 March 2025; Published: 22 March 2025
* Correspondence: bhkim@seowon.ac.kr

rate refers to the phenomenon of population shrinking due to a decrease in birth rate. This causes significant changes in population structure and has various effects on society and economy [6]. There are various causes for low birth rate. There are economic causes such as increased childcare costs and employment insecurity, and there are also social causes such as increased female economic activity and delayed marriage and childbirth. In addition, the lack of effectiveness of government policies is also being raised as a cause [7].

This low birth rate is causing various problems. First, the low birth rate leads to a decrease in the working population in the long term, which hinders economic growth. In addition, the population decline causes a shrinkage in the consumer market, which has a negative impact on the overall economy. The low birth rate is also causing the problem of increased social costs. As the aging society accelerates, the low birth rate increases the social burden on the aging generation. In addition, as the elderly population increases, welfare costs increase, and the tax burden on the younger generation increases. In addition, the low birth rate and the population outflow to the city are causing the population decline in rural areas to worsen, resulting in the collapse of local communities [8-10].

Korean society is facing serious problems in all areas of society due to the ongoing low birth rate phenomenon. Due to the decrease in the number of births, everything from early childhood education and childcare facilities to universities is on the verge of extinction. The basic plan for low birth rate and aging society was implemented in 2006 to try to recover the birth rate, but the effect was minimal. The government invested a huge budget to implement policies related to low birth rates, but the birth rate continues to decline. Many OECD member countries have also experienced low birth rates, but among the countries that experienced the ultra-low birth rate phenomenon, 10 countries, excluding Korea, have escaped the ultra-low birth rate phenomenon, and Korea is the only one among the 37 member countries with a total fertility rate of less than 1. There is a perception that the low birth rate phenomenon is a global phenomenon, but the cases of countries that have recovered their birth rates show that it is a problem that can be overcome.

Therefore, in this paper, we analyzed the birth trend of newborns according to the population in their 20s and 30s, and conducted data analysis research on the relationship between various economic indicators and birth rate. To this end, we conducted correlation analysis by comprehensively linking consumer prices, marriage, economic growth, birth rate, and population. In addition, based on the research results, we derived perspectives on policies to encourage birth.

2. Related Works

2.1. Regression Analysis

A statistical method that shows the relationship between two or more variables. This method, usually expressed graphically, tests the relationship between the dependent variable and the independent variable. The independent variable changes depending on the dependent variable, and regression analysis attempts to find the answer to the most important factor in that change. Figure 1 shows an example of a regression analysis graph [11, 12].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

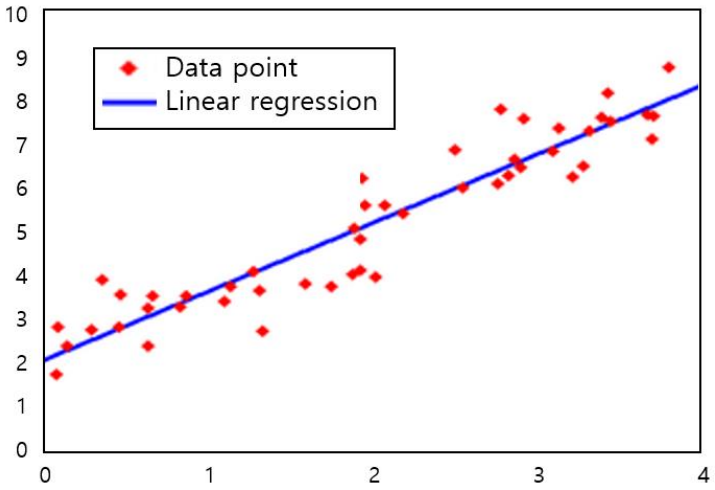

Figure 1.
Regression analysis grape.

2.2. Correlation Analysis

It is a statistical analysis method that measures the relationship between two variables. It is used to determine what kind of relationship or association exists between variables. Correlation analysis mainly focuses on measuring the linear relationship between two variables [13]. It measures the degree of correlation between two variables by calculating the correlation coefficient. When interpreting the results obtained through correlation analysis, you should consider not only the correlation coefficient value but also the significance test results. The significance test is a statistical procedure to determine whether the correlation occurred by chance or is actually meaningful [14].

2.3. Correlation Coefficient

It is the degree to which the degree of correlation is determined. It only indicates the degree of association between two variables, but does not explain the causal relationship. Whether there is a cause-and-effect causal relationship between two variables can be confirmed through regression analysis to determine the direction, degree, and mathematical model of the causal relationship. The correlation coefficient has a value from - to 1, with -1 indicating a perfect negative correlation, 0 indicating no correlation, and 1 indicating a perfect positive correlation. The closer the correlation coefficient value is to 0, the weaker the correlation between the two variables. Figure 2 shows the relationship diagram according to the correlation coefficient [15, 16].

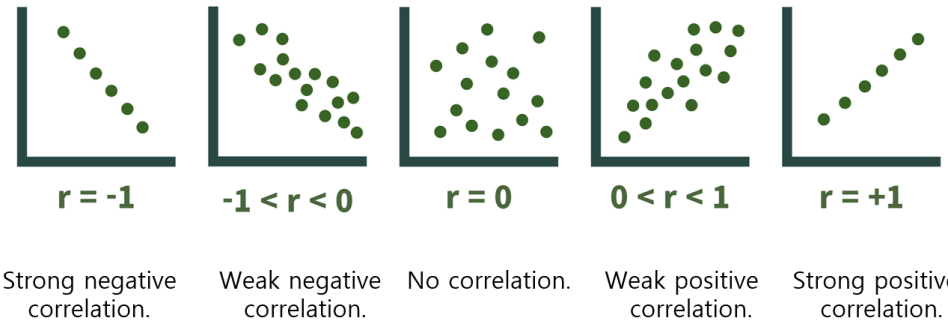

Figure 2.
Correlation coefficient relationship diagram.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

2.4. Logistic Regression Analysis

Logistic Regression is a statistical method used to predict a binary dependent variable. It is mainly used when the outcome variable is divided into two categories (e.g. success/failure, true/false, 0/1). This technique estimates the probability of occurrence of the dependent variable through a linear combination of independent variables. This type of statistical model (also known as a logit model) is often used in classification and predictive analytics. Since the outcome is a probability, the dependent variable is between 0 and 1. In logistic regression, the logit transformation is applied to the odds, which is the probability of success divided by the probability of failure.

This method tests different beta values over a number of iterations to optimize the log odds to best fit. All of these iterations produce a log-likelihood function, which logistic regression tries to maximize to find the best model parameter estimates. Once the optimal coefficients (or coefficients, if there are more than two independent variables) are found, the conditional probabilities for each observation can be computed, recorded, and summed to produce predicted probabilities. For binary classification, probabilities less than 0.5 predict 0, and probabilities greater than 0 predict 1. After computing a model, it is best practice to evaluate how well the model predicts the dependent variable, which is called goodness of fit [17].

There are three types of logistic regression models, which are defined based on a categorical response:

Table 1.
Logistic regression models.

Binary Logistic Regression	In this approach, the response or dependent variable is essentially binary. That is, there are only two possible outcomes (e.g., 0 or 1). Some popular examples include predicting whether an email is spam or not, or whether a tumor is malignant or not. This is the most commonly used approach within logistic regression, and more generally, one of the most common classifiers for binary classification [18].
Multinomial Logistic Regression	In this type of logistic regression model, the dependent variable has more than two possible outcomes. However, these values are not in any specified order. For example, a movie studio wants to predict which movie genres moviegoers are likely to see in order to market the movie more effectively. A multinomial logistic regression model can help the studio determine the strength of the influence that an individual's age, gender, and whether they are in a relationship have on the type of movie they prefer. The studio can then target its advertising campaigns for specific movies to people who are more likely to go see that movie [19].
Ordinal Logistic Regression	This type of logistic regression model is used when the response variable has more than three possible outcomes, but these values have a defined order. Examples of ordinal responses include a rating scale from A to F or a rating scale from 1 to 5 [20].

3. Data Analysis

In this paper, the data used for data analysis and prediction are population, birth rate, economic growth rate, marriage rate, and price index. Since all data on marriage rate exist from 1994, data from 1994 onwards were used for analysis. From the collected data, a data set was created with eight pieces of information, including consumer prices, number of marriages, birth rate, and economic growth rate. Each piece of information consists of 28 pieces of data, and a model was proposed to analyze them by classifying them as rising or falling (0, 1).

However, it is not easy to compare data for about 30 years at a glance. It is also difficult to compare with other data. Therefore, as shown in Table 2, a data set was created based on eight pieces of information, such as consumer prices, number of marriages, birth rate, and economic growth rate, from the collected data. Each piece of information consists of 28 pieces of data, and a model is proposed to analyze it by classifying it as rising or falling (0, 1).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

Table 2.
Data Set.

1	Consumer price
2	Number of marriages
3	Birth rate
4	Number of newborns
5	Economic growth rate
6	Population
7	Number of people in their 20s
8	Number of people in their 30s
9	Result (0,1)

Data analysis on independent variables was performed based on the constructed data set. Each independent variable data was visualized, and the results are shown in Figure 3. The trend of the independent variable is visualized as a decrease in the birth rate, but it is not possible to determine what factors are causing the population to decrease. Therefore, the next analysis step was to conduct a correlation analysis. Using correlation analysis, you can determine which factors have high weights and which factors have low weights. Therefore, you can determine which factors have high weights among the birth rate and each independent variable data.

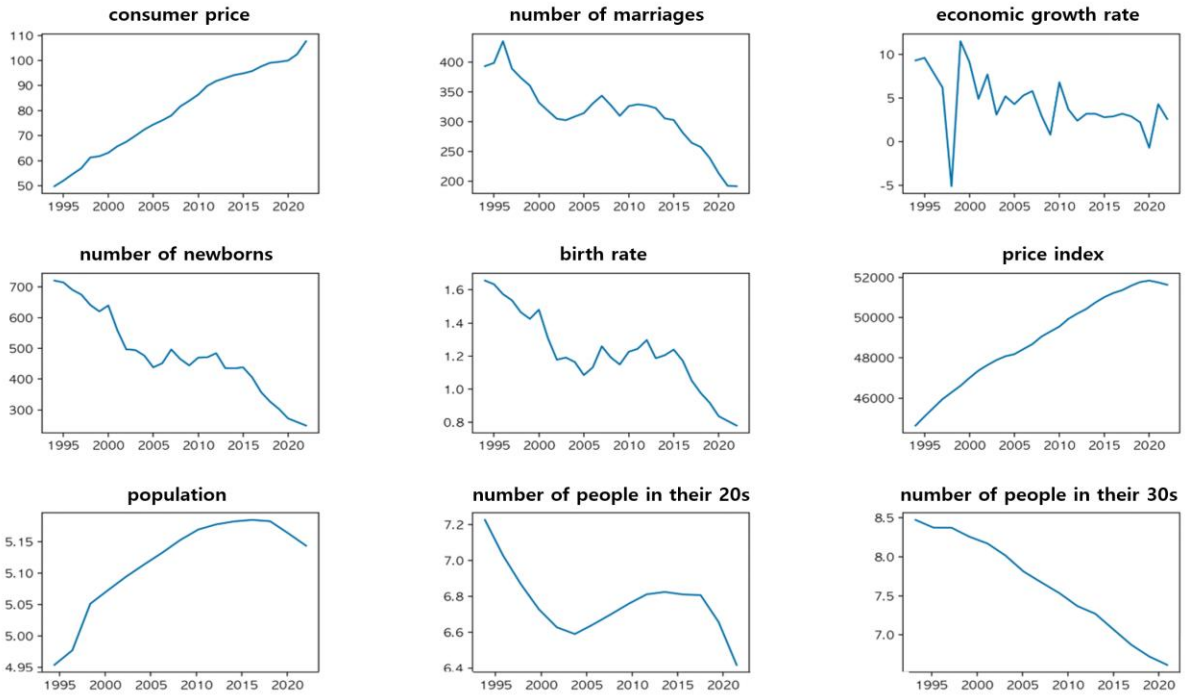

Figure 3.
Visualization of independent variables.

Correlation analysis is an important analytical tool in the field of data analysis science that allows us to discover important variables that other variables depend on. It plays an important role in building machine learning models, allowing us to better understand the degree to which a pair of variables are linearly related in data analysis. Therefore, in order to find out the correlation between dependent and independent variables, a correlation matrix can be created to show the interdependence. Figure 4 shows a heatmap graph derived through correlation analysis [21].

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

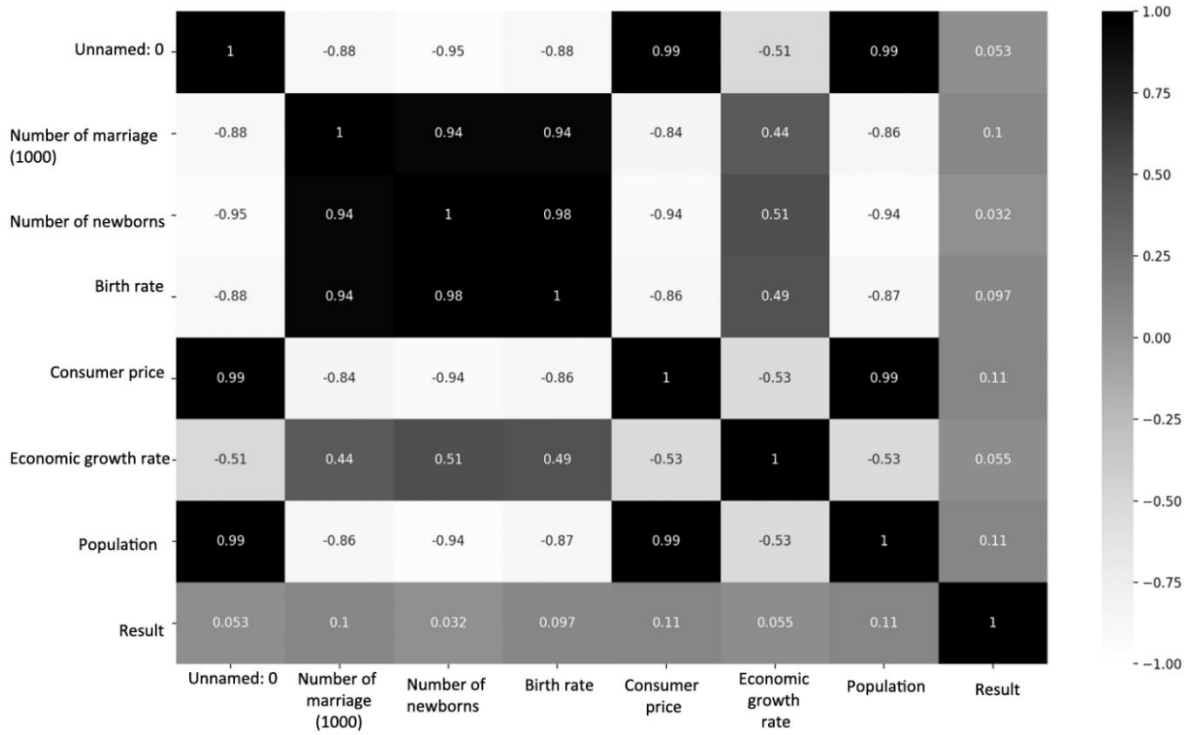

Figure 4.
Heatmap with correlation analysis.

The results of the correlation analysis showed that there was a negative correlation of 86% between the birth rate and consumer prices. In addition, there was a positive correlation of 49% and 94% between the economic growth rate and the number of marriages. In other words, it can be determined that the marriage rate has the greatest correlation. In addition, it can be inferred that the birth rate decreases as consumer prices rise. However, a large correlation coefficient does not necessarily mean that there is a causal relationship.

4. Data Prediction

For data prediction, we performed analysis using logistic regression analysis. The results of the logistic analysis are shown in Figure 5, and the test case was set to 30%.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

	num	number of marriages(1000)	number of newborns	birth rate	consumer price	economic growth rate	population	result
0	0	393.1	721.0	1.656	49.765	9.3	44642	0
1	1	398.5	715.0	1.634	51.995	9.6	45093	0
2	2	434.9	691.0	1.574	54.555	7.9	45525	0
3	3	389.0	675.4	1.537	56.977	6.2	45954	0
4	4	373.5	641.6	1.464	61.258	-5.1	46287	0
5	5	360.4	620.7	1.425	61.756	11.5	46617	0
6	6	332.1	640.1	1.480	63.151	9.1	47008	1
7	7	318.4	559.9	1.309	65.719	4.9	47370	0
8	8	304.9	496.9	1.178	67.534	7.7	47645	0
9	9	302.5	495.0	1.191	69.908	3.1	47892	1
10	10	308.6	477.0	1.164	72.418	5.2	48083	0
11	11	314.3	438.7	1.085	74.413	4.3	48185	0
12	12	330.6	451.8	1.132	76.081	5.3	48438	1
13	13	343.6	496.8	1.259	78.010	5.8	48684	1
14	14	327.7	465.9	1.192	81.656	3.0	49055	0
15	15	309.8	444.8	1.149	83.906	0.8	49308	0
16	16	326.1	470.2	1.226	86.373	6.8	49554	1
17	17	329.1	471.3	1.244	89.850	3.7	49937	1
18	18	327.1	484.6	1.297	91.815	2.4	50200	1
19	19	322.8	436.5	1.187	93.010	3.2	50429	0
20	20	305.5	435.4	1.205	94.196	3.2	50747	1
21	21	302.8	438.4	1.239	94.861	2.8	51015	1
22	22	281.6	406.2	1.172	95.783	2.9	51218	0
23	23	264.5	357.8	1.052	97.645	3.2	51362	0
24	24	257.6	326.8	0.977	99.086	2.9	51585	0
25	25	239.2	302.7	0.918	99.466	2.2	51765	0
26	26	213.5	272.3	0.837	100.000	-0.7	51836	0
27	27	192.5	260.6	0.808	102.500	4.3	51745	0
28	28	191.7	249.0	0.780	107.710	2.6	51628	0

Accuracy: 0.7777777777777778
Precision: 1.0
Recall: 0.5

Figure 5.
The results of the logistic analysis.

Accuracy indicates the proportion of correctly predicted samples among all predictions. The accuracy in the given data is approximately 0.778. This means that the model accurately predicted approximately 77.8% of the entire test data. Precision indicates the proportion of positive samples predicted among positive samples. The precision in the given data is 1.0. This means that all positive samples predicted by the model are actually positive. Recall indicates the proportion of positive samples predicted by the model correctly as positive among the samples that are actually positive. The recall in the given data is 0.5. This means that the model accurately predicted half of the actual positive samples as positive.

Accuracy, precision, and recall are indicators for evaluating the performance of the model from various aspects. If the accuracy is high, the model accurately performed the prediction overall, and if the precision is high, the proportion of samples predicted to be positive is high. Also, if the recall is high, the proportion of actual positive samples that the model accurately predicted as positive is high. The results of the analysis show that the accuracy is relatively high, but the difference between precision and recall is large. This means that the model tends to make many positive predictions but not accurately predict actual positive samples. When evaluating the performance of a model, accuracy, precision, and recall should be comprehensively considered.

The results of various analyses show that there is an 89% negative correlation between birth rate and consumer prices. In other words, the higher the consumer price index, the lower the birth rate. Of course, there may be additional reasons. This is because rising consumer prices can cause economic instability. It is also because economic burdens increase. When living expenses, housing prices, and education costs rise, people face economic burdens in living and educating. In addition, if the employment situation worsens along with economic instability, it can affect the decision to have children. When the unemployment rate rises, job opportunities are limited, household income decreases, and this increases the burden, which affects the birth rate decrease.

5. Conclusion

The low birth rate problem is a social phenomenon that is being experienced in various countries around the world, and it is especially prominent in developed countries and countries with economic development. This low birth rate trend is analyzed as the result of a complex interaction of economic, social, and cultural factors. In particular, Korea has the lowest birth rate in the world, with a total

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

fertility rate of only 0.78 as of 2023. This is analyzed as the result of a complex interaction of various factors such as the economic burden of childbirth and childrearing, housing issues, and career interruptions.

Declining birth rates lead to population decline and aging in the long term. This can lead to slow economic growth, labor shortages, and increased burdens on pension and welfare systems. In addition, a shrinking working-age population can lead to a decline in the vitality of the economy as a whole, with many effects on economic power, such as a shrinking consumer market and slowing innovation. Population decline can lead to a decline in the vitality of local communities, as well as a weakening of cultural and social solidarity.

Therefore, in this paper, we analyzed the birth trend of newborns according to the population in their 20s and 30s, and conducted data analysis research on the relationship between various economic indicators and birth rate. To this end, we conducted correlation analysis by comprehensively linking consumer prices, marriage, economic growth, birth rate, and population. In addition, based on the research results, we derived perspectives on policies to encourage birth.

As a result of the analysis, the correlation analysis between various independent variables led to the conclusion that the birth rate and consumer prices are related. In other words, the higher the consumer price, the lower the birth rate. When consumer prices rise, economic instability occurs, which increases the economic burden. When living expenses, housing prices, and education costs rise, people feel economic burdens such as food, clothing, and shelter as well as education and leisure. In addition, if the employment situation worsens along with economic instability, it can affect the decrease in the birth rate. For these reasons, rising consumer prices may be related to declining birth rates, but various factors in individual circumstances and countries must also be considered, and applying them comprehensively can produce accurate results.

Transparency:

The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] K. S. Hong, "Population aging and optimal taxation," *Journal of Budget Policy*, vol. 13, no. 1, pp. 1-48, 2024. https://doi.org/10.35525/NABO.2024.13.1.001

[2] I. S. Jang, "An effect of innovation city policy on population regrowth in depopulation area: Focusing on Gwangju-Jeonnam Innovation City in Korea," *Journal of Regional Policy Research*, vol. 33, no. 3, pp. 21-38, 2022. https://doi.org/10.22773/RPR.2022.33.2.21

[3] J. S. Han, "Spatial economics in the age of declining population," *Journal of Economic Geography Society of Korea*, vol. 23, no. 4, pp. 454-456, 2020. https://doi.org/10.23841/EGSK.2020.223.4.454

[4] H. J. Park, T. M. Lee, and H. S. Lim, "Analysis of topics related to population aging using natural language processing techniques," *Journal of Information Technology Service*, vol. 23, no. 1, pp. 55-79, 2024. https://doi.org/10.9716/KITS.2024.23.1.055

[5] C. H. Lee, "Policy direction to respond to low birthrate and aging society: a demographic perspective," *Health Welfare Policy Forum*, no. 7, pp. 50-64, 2018. https://doi.org/10.23062/2018.07.5

[6] S.-C. Lee, "Moderating effect of residential selection on the relationship between marriage and childbirth - From perspective of regional integration against local extinction," *The Journal of the Convergence on Culture Technology*, vol. 9, no. 4, pp. 51-60, 2023. https://doi.org/10.17703/JCCT.2023.9.4.51

[7] D. K. Yim, "Changes to the future working age population and the economically active population in declining populations: Focusing on Si(Cities)," *Korea Spatial Planning Review*, vol. 119, pp. 51-76, 2023. https://doi.org/10.15793/KSPR.2023.119..003

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692

[8] N. H. Hwang, "Monitoring of low birth rate and aging society policies - aging society sector," *Korea Institute for Health and Social Affairs*, 2023. https://doi.org/10.23060/KIHASA.F.2023.03

[9] S. E. Park, "Population aging and the current status and outlook of social security finance," *Health Welfare Policy Forum*, vol. 12, pp. 6-20, 2022. https://doi.org/10.23062/2022.12.2

[10] Y. W. Jo, Y. B. Choi, and C. Park, "Exploring spatial distribution of empty houses and vacant land due to population decrease in Mokpo," *Journal of Korean Society of Environmental Restoration Technology*, vol. 23, no. 2, pp. 33-47, 2020. https://doi.org/10.13087/KOSERT.2020.23.2.33

[11] S. Lee and A. Shin, "Analysis on variables influencing teacher burnout using LASSO regression," *Journal of Korean Teacher Education*, vol. 41, no. 2, pp. 25-54, 2024. https://doi.org/10.24211/TJKTE.2024.41.2.25

[12] B. J. Koo and D. H. Jang, "Spatial regression analysis of factors affecting the spatial accessibility of the public libraries in Busan," *Journal of Korean Society for Library and Information Science*, vol. 55, no. 4, pp. 67-87, 2021. https://doi.org/10.4275/KSLIS.2021.55.4.067

[13] S. Son, J. Lee, and K. Son, "A study on the verification of sales price factors in residential building development by using correlation analysis," *Korean Journal of Construction Engineering and Management*, vol. 25, no. 4, pp. 45-52, 2024. https://doi.org/10.6106/KJCEM.2024.25.4.045

[14] M. Yang, K. Park, and S. H. Choi, "Designing a customized product recommendation model through relevance analysis when using e-commerce," *Journal of the Korean Convergence Society*, vol. 13, no. 3, pp. 203-216, 2022. https://doi.org/10.15207/JKCS.2022.13.03.203

[15] C. Liu, "Multi-attribute decision-making method applying a novel correlation coefficient of interval-valued neutrosophic hesitant fuzzy sets," *Journal of Information Processing Systems*, vol. 14, no. 5, pp. 1215-1224, 2018. https://doi.org/10.3745/JIPS.04.0089

[16] P. Ejegwa and J. Awolola, "Real-life decision making based on a new correlation coefficient in Pythagorean fuzzy environment," *Ann Fuzzy Math Inform*, vol. 21, no. 1, pp. 51-67, 2021. https://doi.org/10.30948/AFMI.2021.21.1.51

[17] B. Kim, "A study on diabetes management system based on logistic regression and random forest," *International Journal of Advanced Smart Convergence*, vol. 13, no. 2, pp. 61-68, 2024. https://doi.org/10.7236/IJASC.2024.13.2.61

[18] J.-H. Ryu, J.-M. Hwang, S.-Y. Kim, H.-Y. Seo, and J.-H. Lee, "A proposal for partial automation preparation system of BIM-based energy conservation plan - Case study on automation process using BIM software and Excel VBA," *Journal of KIBIM*, vol. 12, no. 2, pp. 49-59, 2022. https://doi.org/10.13161/KIBIM.2022.12.2.049

[19] D. A. Kwak, "Change prediction of forestland area in South Korea using multinomial logistic regression model," *Journal of the Korean Association of Geographic Information Studies*, vol. 23, no. 4, pp. 42-51, 2020.

[20] K. Y. Kim and H. Y. Jung, "A study on the demand of taxi transfer discount using ordinal logistic model," *KSCE Journal of Civil and Environmental Engineering Research*, vol. 38, no. 5, pp. 685-692, 2018. https://doi.org/10.12652/KSCE.2018.38.5.0685

[21] Y. H. Park, J. H. Mun, J. S. Choi, and J. Y. Choi, "Recommendation system based on correlation analysis of user behavior data in online shopping mall environment," *KIPS Transactions on Computer and Communication Systems*, vol. 13, no. 1, pp. 10-20, 2024. https://doi.org/10.3745/KTCCS.2024.13.1.10

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 1810-1818, 2025
DOI: 10.55214/25768484.v9i3.5692