

## Decision Tree Analysis for Prediction Model of Poverty of The Older Population in South Korea

<sup>1</sup>Soochang Lee, <sup>2</sup>Daechan Kim\*

<sup>1</sup>Professor of Department of Police Administration, DaeKyeung University, Korea

<sup>2\*</sup>Professor of Byuckang Liberal Arts School, Kyungwoon University, Korea  
[leesc@ikw.ac.kr](mailto:leesc@ikw.ac.kr), [hjkdc1273@naver.com](mailto:hjkdc1273@naver.com)

### Abstract

*This study aims to investigate factors that affect elderly poverty based on a comprehensive and universal perspective, suggesting some alternatives for improving the poverty rate of the elderly. The comprehensive and universal approach to the poverty of the aged that this study attempts can give a better understanding of the elderly poverty beyond the contribution of the existing literature, with the research model including individual, family, labor, and income factors as the causes of old-age poverty from the comprehensive and universal perspective on the causes of poverty of the elderly. In addition, the study attempts to input variants of variables into the equation for the causes of elderly poverty by using panel data from the 8th Korean Retirement and Income Study. This study employs decision tree analysis to determine the cause of the poverty of the elderly using CHAID. The decision tree analysis shows that the most vital variable affecting elderly poverty is making income. For the poor elderly without earned income, public pensions, educational careers, and residential areas influence elderly poverty, but for the poor elderly with earned income, wage earners and gender are variables that affect poverty. This study suggests some alternatives to improve the poverty rate of the aged. The government should create a better working environment such as senior re-employment for old people to be able to participate in economic activities, improve public pension or social security for workers with unfavorable conditions for public security of old age, and give companies that create employment of the aged diverse incentives.*

**Keywords:** *Elderly Poverty, Decision Tree Analysis, Panel Data, Korean Retirement and Income Study*

## 1. INTRODUCTION

Population aging, one of the social problems that Korea is suffering from, is a structural problem that is likely to threaten Korean society. Aging must be a fatal threat to the Korean community. Furthermore, population aging threatens the Korean future in many sectors including social, political, economic, and cultural. The poverty of the elderly caused by the loss of income sources is recognized as an absolute threat to the quality of life of the aged.

As Korean society already entered an aging society in 2000, the government has come up with policies to solve social and national problems related to elderly poverty. The Korean government has introduced and implemented the basic old-age pension system to alleviate poverty for the elderly since 2007. There are various

---

Manuscript received: April 7, 2022 / revised: May 8, 2022 / accepted: June 2, 2022

Corresponding Author: [hjkdc1273@naver.com](mailto:hjkdc1273@naver.com)

Tel: +82-53-850-1253, Fax: +82-53-850-1180

Byuckang Liberal Arts School, Kyungwon University, Korea

Copyright©2022 by The International Promotion Agency of Culture Technology. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>)

recognition differences in the introduction and implementation of this system, and Korean society has even temporarily experienced conflict and confrontation as well.

The basic old-age pension system has rapidly emerged as an academic research subject. Previous studies have discussed various topics, including problems of the basic old-age pension system and suggestions for improving the system. Above all, the main research interest is whether the basic old-age pension system has the effect of reducing poverty for the elderly. It is said that the basic old-age pension system has the effect of reducing poverty of the elderly with controversy. The basic old-age pension contributes to poverty alleviation for the elderly, but there are still some shortcomings in terms of improving the quality of life with actual income increase. When it comes to the poverty of the elderly, it is meaningful and constructive to examine what causes old-age poverty, with policy effort reducing poverty. However, empirical research to investigate the causes of the poverty of the elderly has not been enough yet. Early research on elderly poverty was mainly analyzing the status of poverty [1, 2]. They contributed to showing the seriousness of elderly poverty but had a limitation in explaining the causes of elderly poverty. There were studies to examine the causes of elderly poverty scientifically and empirically [3], but with a limit of not reflecting the variant of variables over time based on cross-sectional data. To be beyond this limit, recently, there have been research models including more variables of the poverty of the elderly using longitudinal or panel data, but just focusing on investigating household and demographic characteristics as the causes of poverty. Studies have been conducted to examine more causes of poverty of the elderly, including labor factors in the research model. Some writers criticized that the analysis level of previous studies was microscopic, requiring middle-range or macroscopic analysis to investigate poverty reasons [4].

This study investigates factors that have an influence on elderly poverty in terms of comprehensive and universal perspectives that can give a better understanding of the reasons for elderly poverty to writers, including individual, family, labor, and income factors as the causes of old-age poverty. In addition, the study attempts to input variants of variables into the equation for the causes of elderly poverty by using panel data from the 8th Korean Retirement and Income Study.

## **2. THEORETICAL BACKGROUND FOR RESEARCH MODEL**

The study on the causes of poverty that the elderly suffer has not been conducted in diverse ways, including individual, social, institutional, and political problems. Despite not enough theoretical demonstration, we tried to describe the causes of poverty in the old population to set up the research model by looking over the relevant literature. In the earlier study, most researchers mentioned both the characteristics of the household of the elderly on the impact of poverty like non-engaged in economic activities, living alone, and being married [5] and the individual characteristics of the elderly like gender, age, education level, marital status, and health status [6]. Besides, it was very persuasive that the number of assets and types of a residential district can determine the poverty rate of the elderly [7].

In the research conducted in the Korean context for identifying the causes of the poverty of the elderly, there are no significant differences from the findings of studies done in the oversea context. Similarly, most studies found that the sociodemographic characteristics such as gender, age, education, marital status, and health status play a key role in explaining how the old put their lives in poverty after retirement [8]. In addition, the characteristics of households or families like city types of residence, household income, and types of labor correlate with the poverty of the elderly. The approach to the poverty of the elderly gave politicians the wrong perception that the poverty of the elderly is not a social problem but an individual one, with conflicts caused by the fallacy triggering the need for a new approach that focuses on an institutional arrangement for the stability of old age [9]. The new approach stresses the importance of making income through labor, taking a part-time or full-time job, except for earning from a public or private pension.

The implications of the studies mentioned above on what causes the poverty of the elderly showed that female is at a higher level of poverty than male, the elderly with low academic background and worse health condition face more hardships from the poverty, and living alone makes the rate of getting into the poverty higher than living together with their family. Furthermore, these findings of the studies found that the poverty rate of the aged living in urban cities gets worse than the ones living in rural ones, the less income from family labor accelerates the poverty of the elderly, and aged people with physical labor-oriented jobs are more likely to stay at the worse level of poverty. In addition, if they do not have any income obtained from real estate, pension, or financial investment, they are less likely to get out of a higher poverty rate.

### 3. VARIABLES AND RESEARCH METHOD

This study selects the causes of elderly poverty to be tested based on the review of previous studies. First, since elderly poverty is closely related to the personal characteristics of the elderly, this study includes gender, age (65 years or older), an academic career (secondary education or higher education), and health (bad or good). Second, the environment of the elderly households also affects elderly poverty, so we choose residential areas (large cities or small and medium-sized cities) and family composition (spouses, spouses with grandchildren, or single-person households) as characteristics of the old-aged households. Third, income from the working of the elderly households correlates with elderly poverty, this study inputs whether elderly families have any earned income or wage income into our model. Lastly, the income generated from work after retirement affects the poverty of the elderly, and the research model includes whether to receive a public pension, private pension, social security benefit, real estate income, and other income. Unlike independent variables, it does not seem easy to define the dependent variable of elderly poverty and how to measure it for the argument on the definition and measurement of elderly poverty. We seek median income using household equalized income adjusted by the equalization index of the total income of households and then define income equivalent to 40% of this median income as the poverty line.

**Table 1. Factors and Variables**

Factors	Variables
Individual Characteristics	<ul style="list-style-type: none"> <li>·Gender: male, female</li> <li>·Age: 65-70, 71-75, 76-80, over 81</li> <li>·Education: secondary, higher</li> <li>·Health: bad, good</li> </ul>
Family Characteristics	<ul style="list-style-type: none"> <li>·Residence: big, small-medium</li> <li>·Family: spouse, spouse with grandchildren, living alone</li> </ul>
Wage Earning	<ul style="list-style-type: none"> <li>·Earned wage: yes, no</li> <li>·Paid employee: yes, no</li> </ul>
Other Income	<ul style="list-style-type: none"> <li>·Public pension: yes, no</li> <li>·Private pension: yes, no</li> <li>·Social Security: yes, no</li> <li>·Income from property: yes, no</li> <li>·Income from other pension: yes, no</li> </ul>

The decision tree analysis is an analytical method widely used in the field of data mining, which creates a tree-based classification model that classifies cases or predicts values of a dependent (target) variable based on values of independent (predictor) variables [10]. The generation of a decision tree is an iterative procedure

that involves two processes: tree generation and tree pruning. The process of tree generation is to classify the initial dataset level by level until it can no longer be divided or no longer needs to be divided according to certain partitioning conditions, to generate the tree adequately. Specifically, in each classification, the model compares the differences of each branch obtained by using different independent variables as the classification variables and selects the independent variables that make the most considerable differences as the classification variables of the node. The above process may generate huge trees that require pruning to reduce the tree nodes, control the complexity of the tree, and measure the complexity by the number of leaf nodes of the tree [11]. According to the type of dependent variable, tree models can be divided into two categories: classification tree and regression tree. The commonly used algorithms include CRT (Classification and Regression Trees), CHAID (Chi-square Automatic Interaction Detector), and QUEST (Quick, Unbiased, Efficient, and Statistical Tree) [12, 13].

#### **4. DECISION TREE ANALYSIS**

This study obtained panel data from the 8th National Retirement and Income Study conducted by the National Pension Research Institute. The 8th survey collected data from 4,531 homes with over 50 aged people (7,343 individuals) nationwide. The study extracted 2,795 households (3,418 people) for the analysis by excluding homes with under 65 aged people and those with missing values after classifying elderly households. According to data, the minimum cost of living for middle-aged and older people was 1.17 million won per person, and 1.95 million won per couple, while the modest cost of living was 1.65 million won per person, and 2.68 million won per couple.

This study employs decision tree analysis to determine the cause of the poverty of the elderly using CHAID as one of the algorithms of the decision tree model. Since the target variable, old-aged poverty, was measured by a discrete scale, we conducted a chi-square test. The maximum number of calculation iterations for model estimation was set to 5,000 times to prevent overfitting. The maximum depth of trees was automatically set by the algorithm, minimum of the final node, 6. Since the study used a large sample, the numbers of upper and lower nodes were set to 100 and 20, respectively. In node segmentation, the significance was 5%, and Pearson chi-square statistics were used as the criteria.

From the analysis of the decision tree model using data from the 8th year, the decision tree model is excellent, with 83.5 percent of the overall accuracy of the classification. The value for estimating the risk of misclassification is 0.165, and the standard error is 0.006. We conduct a split-sample model validity test to check whether the model is overfitted by separating training and test data. As a result of random sampling with 70% of the training data and 30% of the test data among the total samples, it shows that the average risk of the training sample and standard error are respectively 0.166 and 0.022, and the average risk difference and the standard error are 0.009 and 0.012. As can be seen from Figure 1, the most important cause affecting elderly poverty is whether an earned income is or not. For the elderly without earned income (node "1"), whether they have public pension benefits or not influences elderly poverty. Among the elderly without earned income, 287 people do not receive public pension benefits. The educational career of the elderly influences whether the old aged get the public pension or not. Among the poor elderly not receiving public pension benefits, 259 are below secondary education. 133 people are receiving public pensions, and the poverty rate has differences depending on the area of residence, 20 in big cities and 113 in small and medium-sized ones, about more than six times. In the case of the elderly with earned income (Node "2"), whether they are wage earners or not affects elderly poverty, and 104 are not wage earners. Gender influences the degree of poverty, and 82 among the poor elderly with no wage are women, about four times higher than 22 men.

To sum up, the most important variable affecting elderly poverty is whether earned income is or not, classifying into node "1" for the elderly without earned income and node "2" for the elderly with earned

income. Node "1" shows that the causes of elderly poverty are whether to receive a public pension, education career, and residential area. Based on the results of this analysis, the degree of poverty gets worse if the elderly without earned income live in small and medium-sized cities without public pensions. On the other hand, node "2" shows that wage earners and gender affect elderly poverty, which means that the poverty rate grows in the case of non-wage earners and women with earned income.

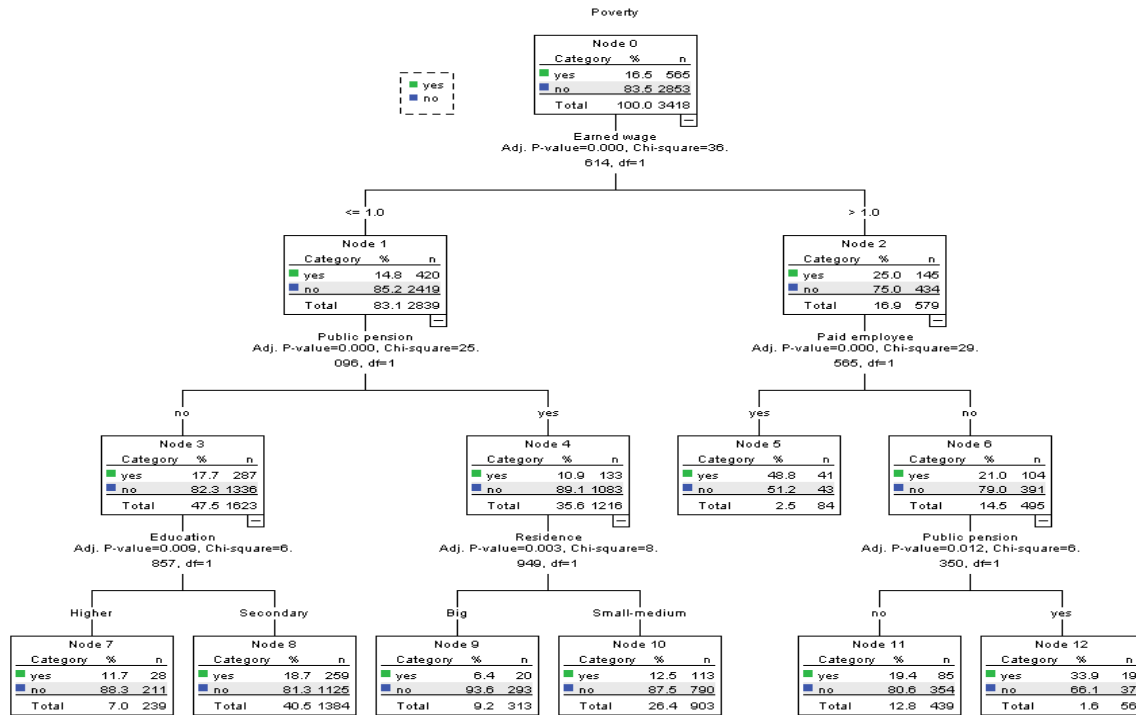


Figure 1. Decision Tree Analysis on the Elderly Poverty

## 5. CONCLUSION

The decision tree analysis shows that the most important variable affecting elderly poverty is an earned income. For the poor elderly without earned income, public pensions, educational careers, residential areas have large influences on elderly poverty, but for the poor elderly with earned income, wage earners and gender are variables affecting poverty.

The determinant affecting elderly poverty is whether they have earned income or not. The determinant affecting elderly poverty is whether they have earned income or not. Excluding the elderly from economic activities after retirement degrades the quality of life and makes it difficult to live a basic life because of the lack of earned income. To solve that, the government is implementing a basic old-age pension system and job creation for the elderly, but it still seems insufficient to solve the poverty of the elderly. Creating a working environment where healthy elderly can participate in economic activities is an urgent problem.

Whether or not to receive a public pension is also the cause of elderly poverty. For the acceptance of public pension, the number of elderly people receiving secondary education or higher is much higher than those below secondary education. Public pensions include a public official pension, a military pension, and a national pension. Except for the national one, it is evident that the rest is not related to an educational career. It must be a social and institutional problem that it is difficult for the elderly with a low educational career to have an opportunity to accept a public pension due to the limitation of job choice.

Residential areas, employment types such as wage-earning jobs, and gender are variables affecting elderly poverty. The elderly living in small and medium-sized cities is poorer than in large cities because it is not easy for the elderly living in small and medium-sized cities to get opportunities for economic activities. In addition, the fact that many old aged are working as non-wage earners proves that it is hard for the elderly to participate in economic activities. It is unlikely that the poverty of elderly women is worse than elderly men, which means there is a correlation between labor and physical conditions. To improve the elderly poverty rate, governments should expand and strengthen the senior employment promotion programs for the elderly living in rural areas or small-sized cities that can increase opportunities to participate in economic activities. In addition, to encourage companies that should preferentially hire elderly females with better wage conditions, the governments should provide various incentives such as tax cutbacks and subsidies or grants.

The poverty of the elderly must be one of the social problems that Korean society must deal with by utilizing various alternatives. That is why the whole community should concentrate on campaigns to solve this problem.

## REFERENCES

- [1] H. Jung & J. Jo & H. Oh & G. Byeon, & S. Moon, The 1998 Elderly Living Condition Survey, *Korean Institute for Health and Social Affairs*, 1998.
- [2] H. Choi & Y. Ryu, A Study on the Levels, Trends, and Composition of the Old-Age Poverty in Korea, *Journal of Korean Gerontological Society*, Vol. 23, No. 3, pp. 143-160, 2003.
- [3] B. Hong, Factors Influencing the Economic Status of the Elderly in Korea, *Korean Journal of Social Welfare*, Vol. 57, No. 4, pp. 275-290, 2005.
- [4] O. Choi, Factors Influencing Poverty of the Elderly: Utilizing the Panel Data Model, *Korean Journal of Social Welfare*, Vol. 59, No. 1, pp. 5-25, 2007.
- [5] D. K. McLaughlin & L. Jensen, Work History and US Elders' Transitions into Poverty, *The Gerontological Society of America*, vol. 40, No. 4, pp. 469-479, 2000.
- [6] K. Rupp, A. Strand & P. S. Davies, Poverty among Elderly women: Assessing SSI Options to Strengthen Social Security Reform, *Journal of Gerontology*, vol. 58, No. 6, pp. s359-s368, 2003, doi:10.1093/geronb/58.6.s359.
- [7] K. McGarry & R. F. Schoeni, Widow Poverty and Out-of-Pocket Medical Expenditures near the End of Life, *Journals of Gerontology*, Vol. 60, No. 3, pp. s160-s168, 2005, doi: 10.1093/geronb/60.3.s160.
- [8] J. S. Choi. Study on the Cause and Improvement Measures of Poverty in Seniors, *Korean Journal of Local Government & Administration Studies*, Vol. 23, No. 1, pp. 457-467, 2009.
- [9] O. G. Choi, Factors Influencing Poverty of the Elderly: Utilizing the Panel Data Model, *Korean Journal of Social Welfare*, Vol. 59, No. 1, pp. 5-25, 2007.
- [10] Y. Wenyue & Z. Suhong, Using Decision Tree Analysis to Identify the Determinants of Residents' CO<sub>2</sub> Emissions from Different Types of Trips: A Case Study of Guangzhou, China, *Journal of Cleaner Production*, Vol. 277, 2020. doi: .org/10.1016/j.jclepro.2020.124071.
- [11] W. Huang & Z. Wang, *Data Mining: R Language Practice*, Electronic Industry Press, Beijing, 2014.
- [12] B. P., Roe, H. Yang, J Liu, Y. Stancu, and G. McGregor, Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 543, Issues 2-3, pp. 577-584, 2005. doi: org/10.1016/j.nima.2004.12.018.
- [13] W. Zhang and Y. Zhong, *IBM SPSS Data Analysis and Mining the Essence of Practice Cases*, Tsinghua University Press, Beijing, 2013.