

## 한국 중고령자의 은퇴 결정요인 분석: 은퇴 패턴과 소요 기간을 중심으로

김수용<sup>1</sup> · 김건우<sup>2</sup> · 이정화<sup>3</sup> · 최근호<sup>4\*</sup><sup>1</sup>세림티에스지(주) 차장<sup>2</sup>한밭대학교 융합경영학과 부교수<sup>3</sup>근로복지공단 근로복지연구원 책임연구원<sup>4\*</sup>한밭대학교 융합경영학과 조교수

## The Analysis on Determinants of Retirement: Focusing on Patterns and a Period of the Retirement

Suyong Kim<sup>1</sup> · Gunwoo Kim<sup>2</sup> · Jeonghwa Lee<sup>3</sup> · Keunho Choi<sup>4\*</sup><sup>1</sup>Senior Manager, Selim Co. Ltd., Daejeon 34014, Korea<sup>2</sup>Associate Professor, Department of Business Administration, Hanbat National University, Daejeon 34158, Korea<sup>3</sup>Research Fellow, Labor Welfare Research Institute, Seoul 07254, Korea<sup>4\*</sup>Assistant Professor, Department of Business Administration, Hanbat National University, Daejeon 34158, Korea

### [요 약]

한국 사회에서 중·고령자의 은퇴는 정년제와 같이 공식적인 은퇴 연령을 맞이하여 이행되거나 고령·건강 등 개인의 사유로 노동시장을 완전히 이탈하는 경우 등의 유형으로 정리된다. 본 연구는 '고령화 연구 패널 조사(KLoSA)' 자료를 이용하여 55세 이상 중·고령자의 은퇴를 분석하였다. 의사결정나무모형을 활용하여 2년 이내의 은퇴에 관하여 분석한 후, Kaplan-Meier 방법을 통해 12년 이내 이행하는 은퇴를 분석하였다. 그 결과 은퇴를 결정하는 요인은 생물학적 노화와 고용 특성으로 검증되었으며, 후자는 고용 안정성과 전문성 등이 관련된 요인으로 나타났다. 본 연구의 결과는 안정적인 은퇴 계획 및 지원 정책 수립에 도움을 줄 것으로 기대된다.

### [Abstract]

The most of middle and old-aged people in Korea exit from the labor market as they reach the retirement age, or due to personal issues such as old age and health problems. This study analyzed the retirement of workers aged more than 55 years using the Korean Longitudinal Stud of Aging (KLoSA). The decision tree was first used to analyze the retirement within 2 years and then kaplan-meier was used to analyze the retirement timing within 12 years. As a result, the aging and the characteristics of employment were identified as important factors determining the retirement within 2 years, and the stability of employment and the professionalism were identified as important factors determining the retirement timing within 12 years. We expect our study would be helpful to establish stable plans and policies for the retirement.

**색인어** : 은퇴시기, 의사결정나무, 생존분석, Kaplan-meier, Cox regression**Key word** : Retirement Decision tree, Survival analysis, Kaplan-meier, Cox regression<http://dx.doi.org/10.9728/dcs.2021.22.3.527>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 16 December 2020; Revised 27 January 2021

Accepted 27 January 2021

**\*Corresponding Author; Keunho Choi**

Tel: +82-42-821-1468

E-mail: keunho@hanbat.ac.kr

## I. 서론

한국 사회의 은퇴 연령은 50대 후반에서 60대 초반에 분포한다<sup>1)</sup>. 중·고령자의 은퇴는 정년제와 같이 공식적인 은퇴 연령을 맞이하여 이행되거나 고령·건강 등 개인의 사유로 노동시장을 완전히 이탈하는 경우, 또는 자영업과 같이 은퇴 시점이 정해지지 않아 소득활동을 지속하거나 은퇴 후에도 소득활동을 재개하여 노동시장에 잔류하는 유형으로 정리된다<sup>3)</sup>[6]. 그러나 기대여명이 증가하며 길어진 노후는 노년기의 경제적 어려움과 불평등, 사회의 높은 부양비용을 초래한다<sup>7)</sup>[9]. 이는 중·고령자의 사회적 관계 및 자아실현으로 이어진다는 점에서도 의미가 있다<sup>10)</sup>, <sup>11)</sup>.

이에 완전한 은퇴보다는 점진적인 은퇴나 재취업을 이행하는 ‘지연된 은퇴(late retirement)’의 비율이 높아지고 있다. 고령화 세대의 연령이 증가할수록 은퇴연령 역시 증가하는 경향이 나타난다는 점을 참고할 수 있다<sup>12)</sup>. 실제로 우리나라 중·고령자의 노동시장 참여는 점차 증가하는 추세로, 55~59세의 고용률은 2000년 62.2%에서 2019년 72.8%로 상승하였다. 같은 시기에 60~64세는 53.0%에서 59.8%로, 65세 이상의 고용률은 29.4%에서 32.9%로 증가하였다<sup>13)</sup>. 이는 OECD 국가의 고용률(2019년 66.9%)보다 평균적으로 높은 수준이다<sup>3)</sup>. 또한 고령층(55~79세)은 평균적으로 73세에 은퇴하기를 희망하며, 여전히 일을 하기 원하는 이유는 ‘생활비에 보탬을 하기 위함(58.8%)’과 ‘일하는 즐거움(33.8%)’을 주요한 항목으로 꼽았다<sup>13)</sup>.

이처럼 은퇴와 맞물려 발생하는 다양한 사회적 요구에 대응하기 위한 정책이 수립되고 있으며, 일례로 정부는 50세 전후에서 70대 초반까지의 이른바 ‘5060세대’를 ‘신중년’으로 정의하고 이들의 은퇴 이후의 삶을 위한 중장기 로드맵을 개발한 바 있다<sup>4)</sup>. 이러한 사회 변화 속에서 중·고령층의 은퇴에 대한 연구는 개인으로 하여금 안정적인 은퇴 계획을 수립하도록 지원하고, 은퇴 연령 인구의 소득보장과 노후설계를 지원하는 정책 수립을 도울 수 있다.

본 연구에서는 ‘고령화 연구 패널 조사(KLoSA)’ 자료를 활용하여 은퇴 결정에 영향을 미치는 요인을 분석하였다. 의사결정나무와 생존 분석 기법을 이용하여 주요 은퇴 결정요인과 요인들 간 패턴, 요인들의 시간에 따른 효과 변화를 분석하고자 하였다. 그 결과 은퇴의 이행이 개인뿐만 아니라 노동시장 환경과의 상호작용에 의한 산물이라는 점을 확인할 수 있었다. 이를 통하여 은퇴의 배경을 이해할 뿐만 아니라 조기 은퇴 현상에 의

한 소득 감소, 나아가 사회 전체의 높은 부양비용 등 사회문제 해결에 유용한 자료를 제공하고자 하였다.

연구의 구성은 다음과 같다. 2장에서는 은퇴와 관련한 선행 연구를 고찰하고, 본 연구에서 사용된 의사결정나무와 생존 분석 방법론에 관하여 서술한 후 본 연구의 차이점에 대해 설명한다. 3장에서는 본 연구에서 사용한 자료와 사용 변수, 실험 설계 등 연구 방법을 설명하고, 4장에서는 실험 결과와 결과에 대한 해석을 다룬다. 마지막으로 5장에서는 본 연구의 요약과 더불어 시사점 및 한계점에 관하여 서술하였다.

## II. 이론적 배경

### 2-1 은퇴 관련 연구

본 장에서는 은퇴와 관련 선행연구로 중·고령자의 은퇴 결정 요인과 은퇴 시기, 그리고 이들의 재취업 내지는 노동시장 잔류에 대한 연구를 검토하였다.

[3]은 55세 이상 고령층의 일자리 변화와 은퇴의 경로를 유형화하여 정의하였다. 이에 따르면 고령자의 은퇴는 정년제 등에 따라 이행된 정규직 임금근로자의 은퇴, 자영업자와 같이 공식적인 은퇴정년이 없고 노후의 경제적 자원을 확보한 고령자의 은퇴, 자녀의 경제적 지원을 받는 상황에서의 은퇴, 자산이 적고 가족의 지원도 부재하여 취업과 비취업을 반복하는 은퇴로 나뉜다. 또 다른 연구로 [4]는 중·고령자의 은퇴 과정을 유형화하였다. 주변적 일자리를 가질 경우에 비정규직 고용의 지속 기간이 가장 길게 나타났으며 비임금근로자는 은퇴 후 다시 노동시장 참여를 재개하는 유형에서 높은 지속기간을 보였다.

한편 [14]는 65세 이상 고령자를 대상으로 은퇴와 재취업에 대해서 연구하였다. 그 결과 건강은 소득활동을 위한 중요한 요인이며 건강 상태가 취약할수록 은퇴 시기를 당기는 것으로 나타났다. 비슷한 연구로 [15]는 퇴직 경험이 있는 베이비붐 세대의 재취업을 연구하였다. 분석 결과 남성이고 건강 상태가 양호하며 미혼의 부양자녀가 많은 경우, 그리고 가구소득이 적고 임시·일용직인 경우 등의 특징을 갖는 경우에 재취업 의사가 높은 것으로 나타났다. 정규 근로시간보다 파트타임과 같은 단기 근로를 선호하며, 보다 전문적인 기술이나 지식을 갖고 있는 경우 은퇴 연령이 늦춰지는 경향을 보였다.

또한 [6]에 따르면 은퇴에는 정년퇴직을 비롯하여 비자발적 퇴직을 야기하는 ‘연령’이 주요한 변수로 작용하였다. 또한 자녀의 교육비나 건강 상태, 보유 자산도 유의한 요인으로 분석되었다. 마찬가지로 노동시장에 초점을 둔 [5]는 중·고령자의 재취업에 대해 과거 주된 일자리의 임금과 같은 시장적 유인보다 가구경제가 강한 유출요인으로 작용하는 것으로 분석하였다. 즉 이들의 재취업은 자발적인 측면보다 미흡한 노후소득의 불가피한 선택인 것으로 설명하였다. 비슷한 연구로 [16]은 소득 계층에 따라 중년층의 은퇴 의향이 어떠한 차이가 있는지 연구하였다. 그 결과 저소득 중년층일수록 건강상태에 관계없이 오

1) 중·고령자를 심층 조사한 국민노후보장패널의 분석 결과 은퇴 연령이 남성은 63.5세, 여성 55.2세이다<sup>1)</sup>. [2]는 평균 은퇴 연령을 63.13세로 분석하였다.

2) 15세 이상 경제활동인구의 고용률은 2000년 58.5%, 2019년 60.9%이다<sup>13)</sup>.

3) 자료: 대한민국 일자리 상황판([https://dashboard.jobs.go.kr/index/detail?pg\\_id=PDCT010101&data2=DCT010101&ct\\_type=run](https://dashboard.jobs.go.kr/index/detail?pg_id=PDCT010101&data2=DCT010101&ct_type=run))

4) 관계부처 합동(2017). 보도자료 “「신중년 인생 3모작」 기반구축 계획”

래 일하기 원하는 등 이들의 근로 동기는 경제적 필요와 강하게 연관되어 있었다. 또한 이들은 중년층은 금전적인 측면뿐만 아니라 노동시장 참여를 안정적으로 지속하기 어렵고 단편적인 소득활동을 이어갈 우려가 있으며, 건강 측면에서도 위험성이 높은 계층으로 연구되었다.

[17]은 지연 은퇴(delay retirement)의 동기를 계층 구조로 개념화하였다. 이에 따르면 은퇴를 지연시키는 6가지의 계층 구조 중 개인의 노동능력과 일을 지속할 수 있는 여건이 가장 일차적인 유인이며, 경제적 생활상이 두 번째의 단계에 위치하는 것으로 나타났다. 이후의 상위 단계에는 직업 및 사회적 역할에 대한 가치, 자율성, 사회적 기여에 대한 가치 등이 존재한다고 보았다.

선행연구를 종합하면 중고령자의 은퇴와 은퇴연령 후의 지속적인 소득활동에는 여러 요인들이 그 중에서도 개인의 연령과 건강, 그리고 경제적 필요가 중요하게 작용하며, 환경의 측면에서는 정년제와 노동시장 구조가 주요한 요인으로 이해된다.

## 2-2 의사결정나무

데이터 마이닝은 방대한 데이터로부터 유용하고 의미 있는 정보의 패턴이나 규칙을 발견하기 위해서 자동화되거나 반자동화된 도구를 이용하여 대량의 데이터를 탐색, 분석하는 일련의 과정[18] 또는, 통계 및 수학적 기술뿐만 아니라 패턴인식 기술들을 이용한 데이터 저장소에 저장된 대용량의 데이터를 조사함으로써 의미 있는 상관관계 및 패턴 등을 발견하는 과정[19]으로 정의할 수 있는데, 제조, 유통, 통신, 금융, 의료 분야 등 다양한 분야에서 의사를 결정하는 의사결정자들이 올바른 의사결정을 내리는 데 도움을 주고자 많이 사용되고 있다. 이러한 데이터 마이닝에는 군집 분석(clustering analysis), 분류분석(classification analysis), 연관성 분석(association rules analysis) 등으로 구분되며, 이 중 분류 분석은 관측대상을 사전에 정의한 목표변수의 범주로 분류하기 위한 패턴을 도출하는 기법으로 의사결정나무(decision tree) 기법이 대표적이다. 이 의사결정나무 기법은 서로 다른 목표 변수 값을 갖는 관측 대상들이 포함된 부모 노드(Parent node)를 유사한 형태의 관측대상들이 포함된 여러 개의 자식 노드(child node)들로 분할하는 과정을 반복적 수행함으로써 의사결정에 사용할 나무구조의 분류 패턴을 생성하는데, 그 방법은 다음과 같다.

우선적으로, 의사결정나무에서 사용할 각각의 독립변수를 대상으로 변수 값에 따라 부모 노드(parent node)에 포함된 관측대상을 여러 작은 자식 노드(child node)로 분할 한 후, 자식 노드(child node)의 평균 순수도가 가장 높게 분할된 독립변수를 해당하는 부모 노드(parent node)에서의 분할 변수로 선정하게 된다. Node의 순수도는 Node에 포함된 관측대상들이 가지는 목표변수 값의 다양성 정도를 의미하며, chi-square statistic, gini index, entropy index 등을 기준으로 측정할 수 있다. 이러한 분할 작업을 반복 수행하다가 분할된 자식 노드(child node)에 속하는 관측대상들의 수가 사용자 지정 수치보다 작거나, 그 자식 노드(child node)의 순수도가 사용자 지정 수치보다 크게 되

면 분할 작업을 멈추게 된다. 이렇게 생성된 분류패턴을 이해하기 쉬운 형태로 만들어지는 의사결정나무 기법은 그 설명력이 뛰어나다는 장점이 있다.

본 연구에서는 의사결정나무 기법의 여러 알고리즘 중 우수한 성능에 의해 기존 연구들에서 널리 사용되는 C4.5 알고리즘을 본 연구 분석에 활용하였는데, C4.5 알고리즘은 entropy index의 GR(gain ratio)를 분할 기준으로 사용하여 다음과 같이 정의된다.

$$GR = \frac{IG}{II} \quad (1)$$

$$IG(S,A) = Entropy(S) - \sum_{v \in value(A)} \frac{|\{x \in S \mid x_A = v\}|}{|S|} \times E(\{x \in S \mid x_A = v\}) \quad (2)$$

여기서,  $S$ 는 부모 노드(parent node)의 집합이고,  $x_A$ 는 관측대상  $x$ 가 독립변수  $A$ 에 대해 가지는 부분 집합의 값이며,  $value(A)$ 는 독립변수  $A$ 가 가지는 모든 가능한 값들의 집합이다. 독립변수  $A$ 에 대한  $IG$ (Information gain)는 위의 식과 같이  $Entropy(S)$ 의 항목으로 정의되는데,  $Entropy(S)$ 는 다음과 같이 정의된다.

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (3)$$

여기서,  $p_i$ 는 자식 노드(child node) 내에서 목표변수 값이  $i$ 인 관측대상이 차지하는 비율을 나타낸다.  $IG$ 를 분류 기준으로 사용하게 되면, ID code와 같이 다양한 값을 가지는 독립변수들이 높은  $IG$ 를 가지게 되어 중요한 독립변수로 선정되는 문제점을 지니고 있다.  $GR$ 은 이러한  $IG$ 의 문제점을 보완하기 위해  $IG$ 를  $II$ (Intrinsic Information)로 나눈 값으로 계산을 하는데,  $II$ 는 분할된 자식 노드(child node)의 수와 자식 노드(child node)들 내 존재하는 관측대상의 수(크기)를 고려하여 계산한 entropy 값이다. 본 연구에서는 목표변수를 은퇴로 선정하여, 은퇴에 대한 분류 패턴을 도출하고자 하였다.

## 2-3 생존분석

생존분석은 어떤 사건의 발생 확률을 시간이란 변수와 함께 생각하는 통계 분석 및 예측 기법으로, 종속변수가 관심 있는 사건의 발생 시간일 때, 사건의 발생률과 그에 미치는 요인을 찾아 그 영향을 분석하는 것을 말한다[20]. 생존분석에서는 분석 결과의 과대추정 기피 및 정확도 향상을 위해 연구종료 시점까지 발생하지 않은 사건의 중도절단자료(censored data)에 대해서도 함께 고려한다. 이에 본 연구에서는 비모수적 분석방법인 Kaplan-Meier 생존함수와 Cox 비례위험 모형을 사용한다.

Kaplan-Meier 생존함수는 각 개체의 생존 기간만 알면 표본 크기의 제약 없이 간단히 유도 가능하며, Cox의 비례위험 모형은 생존기간의 분포에 대한 가정 없이, 중도절단자료를 포함한 분석이 이루어져 광범위하게 사용되고 있다[21]. 생존분석에서의 주요 관심은 은퇴까지의 기간 T의 분포에 있다. 따라서 본 연구에서는 종속변수를 55세에서 은퇴 시점까지의 기간으로, 다음과 같이 정의한다.

$$\{(T_i, \delta_i), i = 1, \dots, n\} \tag{4}$$

여기서,  $T_i$ 는  $i$ 번째 관측대상의 은퇴까지의 기간을 나타내며,  $\delta_i$ 는 은퇴 여부를 나타낸다.

종속변수  $T$ 에 대한 함수로 생존함수  $S(t)$ 와 위험함수  $h(t)$ , 누적위험함수  $H(t)$  등이 있으며 다음과 같이 정의된다. 생존함수는 조사 당시 은퇴 여부의 확률인 생존율을, 위험함수는 임의의 시점  $t$ 에서 갑자기 은퇴할 확률인 순간위험률을 나타낸다.

생존함수  $S(t)$ 는 종속변수  $T$ 가 특정 시점  $t$ 보다 클 확률을 가리키며,  $T$ 의 누적분포함수를  $F(t)$ 라 하면 다음의 관계가 성립되는데, 이는  $t$ 시점까지 은퇴할 확률을 나타낸다.

$$S(t) = P(T > t) = 1 - F(t) \tag{5}$$

생존함수  $S(t) = P(T > t)$ 와 은퇴까지의 기간  $T$ 의 누적분포함수인  $F(t)$  사이에는  $S(t) = 1 - F(t)$  관계가 성립하므로 양변을  $t$ 에 관해 미분하면  $\frac{d}{dt}S(t) = -f(t)$ 인 관계를 얻을 수 있다. 여기서  $f(t)$ 는  $T$ 의 확률밀도함수이다.

위험함수  $h(t)$ 는 임의의 시점  $t$ 내 은퇴할 확률을 나타낸다.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{6}$$

$$= \frac{f(t)}{S(t)} = -\frac{d}{dt} \ln S(t)$$

누적위험 함수는 다음과 같다.

$$H(t) = \int_0^t h(u) du = -\ln S(t) \tag{7}$$

$$S(t) = e^{-H(t)}$$

대표적인 생존함수 추정법으로 누적한계추정법(Product-limit method)으로 불리는 Kaplan-Meier는 중도절단된 관측값을 포함하는 생존자료의 대표적인 기법으로 생존함수에 대한 추정량으로  $i$ 번째 구간에 은퇴할 확률의 추정값은 다음과 같다.

$$P_i = \frac{n_i - d_i}{n_i} \tag{8}$$

중도절단 및 은퇴가 발생하는 순서별 배열은  $t_{(1)} < t_{(2)} < \dots < t_{(n)}$ 라 하며, 은퇴 발생 기간  $t$ 에서의 Kaplan-Meier 추정량은 다음과 같다.

$$\hat{S}(t) = \prod_{t_{(i)} < t} \frac{n_i - d_i}{n_i} \tag{9}$$

여기서  $d_i$ 는 시간  $t_{(i)}$ 에서의 은퇴자 수를 나타내고,  $n_i$ 는 시간  $t_{(i)}$ 에 은퇴가 발생할 사람의 수를 나타낸다. Kaplan-Meier 추정량  $\hat{S}(t)$ 는 시간  $t_{(i)}$  바로 직전에 은퇴한 사람의 시간  $t_{(i)}$ 를 지나는 동안 은퇴할 조건부 확률의 곱으로 나타내며 추정량의 분산에 대한 추정엔 테일러 전개에 기초한 delta method라 불리는 방법으로부터 유도된 Greenwood 공식을 이용한다.

$$\widehat{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)} \tag{10}$$

위의 식들을 바탕으로 생존시간에 대한 함수를 추정할 후, Cox 비례위험 회귀모형을 이용하여 공변량과 위험함수와의 관계를 살펴볼 수 있으며, Cox 비례위험 회귀모형은 다음과 같이 정의된다.

$$h_i(t) = h_0(t) \exp(\beta_1 x_{i1} + \dots + \beta_p x_{ip}) \tag{11}$$

$$= h_0(t) \exp(\beta_p \cdot x_{ip})$$

여기서  $\exp(\beta_p)$ ,  $i = 1, \dots, p$ 는 공변량  $x_i$ 에 대한 효과이며 위험비(hazard ratio)로 해석되는데,  $h_0(t)$ 는 모든 공변량의 값이 0일 때의 위험률인 기저위험함수(baseline hazard function)를 나타낸다. 공변량이 위험률에 미치는 영향은 위험함수(hazard ratio)의 비로 설명할 수 있으며 다음과 같다.

$$\frac{h_i(t) \exp(\beta_p x_{ip})}{h_j(t) \exp(\beta_p x_{jp})} \tag{12}$$

$$= \exp\{\beta_1(x_{i1} - x_{j1}) + \dots + \beta_p(x_{ip} - x_{jp})\}$$

위의 식 중 좌변은 시간 길의 함수이고, 우변은 위험률의 비로 공변량의 효과가 시간과 관계없이 항상 일정하므로 비례위험모형(proportional hazard model)이라고 한다. 이때 공변량에 대한 비례위험 가정이 성립하는지 검토해 보아야 하는데 만약, 비례위험 가정이 위배되면 시간에 따라 공변량에 대한 효과가 변화한다는 것을 의미하므로 이를 위해서는 공변량과 시간의 교호작용을 모형에 추가하여 시간 가변 효과를 고려해야 한다.

본 연구에서는 생존 분석을 이용하여 은퇴하는데 소요되는 시간과 성별 경제 활동 기간 등을 분석하여 개인의 특성에 따른 은퇴 시기를 알아보려고 하였다.

### 2-3 기존 연구와의 차이점

본 연구의 목적은 빅데이터 분석을 이용하여 중·고령자의 은퇴 시기를 결정하는 요인에 대한 분석을 진행하였으며, 기존 선행 연구들과 다음과 같은 차이점을 보인다.

선행 연구들에서는 주로 취업자의 직업 변화와 특성에 의해 근로 생애 즉, 은퇴 시기와 은퇴에 영향을 미치는 요인들을 밝히는 데 초점을 두고 있으나, 본 연구에서는 데이터 마이닝 기법의 하나인 의사결정나무 기법과 통계적 기법인 생존분석 기법을 모두 활용하여 중·고령자의 은퇴 결정 요인과 요인들 간 패턴을 도출하고, 이와 더불어 시간이 흐름에 따른 은퇴 확률을 파악하고자 하였다.

## III. 연구방법

### 3-1 분석 데이터

본 연구의 분석에 이용된 데이터는 한국고용정보원에서 제공하는 ‘고령화 연구 패널 조사(KLoSA)’ 자료로서, 본 데이터는 2006년에서 2018년까지 대한민국 제주도를 제외한 지역에 거주하는 50세 이상의 중·고령자 개인에 대한 가족 정보, 구직 활동, 건강 상태, 은퇴 여부 등 다양한 내용을 포함하고 있다.

본 연구에서는 은퇴와 관련한 기존의 선행연구와 유사하게 1차년도 조사 당시 55세 이상이면서, 은퇴하지 않고 경제활동을 하는 사람으로 현재 노동 여부가 노동 중이며, 근로 형태가 임금 근로자인 자들을 분석 대상으로 하였다.

### 3-2 데이터 전처리 및 사용 변수

본 연구에서는 1차년도 자료를 기준으로 가까운 미래인 2년 이내의 은퇴 여부 변수를 타겟 변수로 사용하여 의사결정나무 모델을 생성하였다. 분석 모델은 데이터 균형화 작업을 진행하였고, 그 결과 202건의 데이터가 모델 구축을 위해 사용되었다.

또한 본 연구에서는 생존 분석을 수행하기 위해서 1차년도 조사 시점을 기준으로 최종 은퇴하기까지의 기간(년)을 종속변수로 생성하였으며, 연구에서 사용한 패널자료의 마지막 조사 차수인 6차년도 조사까지 은퇴하지 않은 경우는 중도 절단 데이터로 처리하였다. 그 결과, 총 496건의 데이터를 생존 분석에 사용하였다.

본 연구에서 사용한 독립변수는 선행연구 분석과 전문가 토의를 통해 선정하였으며, 최종 사용 독립변수는 <표 1>과 같다.

의사결정나무 분석은 많은 연구에서 널리 사용되는 데이터 마이닝 툴인 Weka ver.3.8.1을 이용하여 수행하였으며, 생존 분석은 통계 패키지인 SAS 9.4를 이용하여 수행하였다.

표 1. 최종 사용 변수

Table 1. Final Variables

NO	Variables
v1	Age
v2	Difficulty of daily activity due to pains
v3	Support for national health insurance in current job
v4	Need for computer skills in current job
v5	Total income of household
v6	Satisfaction on the quality of life
v7	Inconvenience in ADL
v8	Satisfaction on own health status
v9	Need for physical power in current job
v10	Diabetes / hyperglycemia
v11	Stability of employment
v12	Need for carrying heavy loads
v13	Number of unmarried children living together
v14	Sex
v15	Labor union
v16	Seniority system
v17	Heart disease
v18	Living parents
v19	Suitability of educational standards in current job
v20	Need for difficult tasks
v21	Satisfaction on current job
v22	Public pension

## IV. 연구 결과

### 4-1 의사결정나무 분석 결과

본 연구에서는 의사결정나무 분석 결과 도출된 많은 분류 패턴들 중 패턴의 신뢰도를 높이기 위해 최소 샘플 수가 5건을 초과하면서 해당 패턴의 정확도가 60% 이상인 분류 패턴들만을 주요 패턴으로 선정하였으며, 이러한 패턴들을 중점적으로 살펴보았다. <그림 1>은 2년 이내 은퇴 여부 모델의 주요 패턴들을 보여주며, <표 2>와 <표 3>은 각각 2년 이내 은퇴 여부 주요 패턴 중 은퇴 가능성이 높은 패턴과 미 은퇴 가능성이 높은 패턴을 보여준다.

<표 2>에서와 같이, 2년 이내에 은퇴할 가능성이 높은 주요 패턴을 살펴보면 다음과 같다.

첫 번째, 통증에 의해 일상 활동이 어렵고, 건강보험에 가입하지 않았으며, 가구 총소득이 2,040 미만인 경우 은퇴할 가능성이 높게 나타났다.

두 번째, 노동조합이 있는 직장에서 사무직 직종 종사자로 근무하고, 업무수행 시 컴퓨터 활용능력이 필요하며, 당뇨병 또는 고혈당을 진단받지 않고 통증으로 인한 일상 활동에 어려움이 없고, ADL(일상생활 동작)에 불편함이 없는 경우 은퇴 가능성이 높게 나타났다.



4-2 생존 분석 결과

생존 분석에서는 은퇴에 직접적으로 영향을 많이 주는 나이 변수와 함께 앞선 의사결정나무 분석 결과에서 은퇴 결정에 중요한 영향을 주는 요인으로 도출된 14개의 변수들에 대해 Kaplan-meier와 Cox regression을 수행하였다.

1) Kaplan-Meier 분석 결과

고령화 연구 패널 조사의 1차 자료 기준 55세 이상인 자들을 대상으로 은퇴 시점까지의 소요 기간을 비모수적 방법인 Kaplan-Meier 방법을 이용하여 분석한 결과는 <표 4>와 같다.

대상자들의 은퇴율이 50%가 되는 시점인 은퇴 기간의 중앙값은 8년으로 나타났으며, 전체 대상자 중 은퇴 비율은 72.4%로 나타났다.

분석에 사용된 독립변수별로 은퇴율과 은퇴 기간의 중앙값을 살펴보면, 예상과 유사하게 65세 이상인 경우 연령이 적은 집단에 비해 은퇴율이 높고 은퇴 소요 기간이 짧았으며, 여성의 경우 남성에 비해 은퇴율은 높았으나 은퇴 소요 기간은 비슷한 것으로 나타났다.

ADL 불편 여부의 경우 예상한 바와 달리 불편을 느끼는 집단이 그렇지 않은 집단에 비해 은퇴율이 약간 더 낮았으나, 은퇴까지의 소요 기간은 비슷하게 나타났다. 당뇨/고혈당 진단의 경우, 예상과 달리 당뇨/고혈당 진단을 받지 않은 집단이 당뇨/고혈당 진단을 받은 집단보다 은퇴율이 높았으나, 은퇴까지의 소요 기간은 길게 나타났다.

통증으로 일상생활이 어려운 집단의 경우 그렇지 않은 집단에 비해, 은퇴율이 높고, 은퇴까지의 소요 시간도 짧게 나타났고, 건강보험 가입의 경우, 지역 가입자 집단이 다른 집단에 비해 은퇴율이 가장 낮고, 은퇴까지의 소요 시간도 가장 길게 나타났다.

고용 안정성 여부의 경우 예상한 바와 유사하게, 매우 그렇다고 응답한 집단이 다른 집단에 비해 은퇴율이 월등히 낮고, 은퇴까지의 소요 시간도 가장 길게 나타났다. 업무 수행 시 육체적 힘 필요 여부의 경우 업무 수행 시 육체적 힘이 필요한 집단이 그렇지 않은 집단에 비해 은퇴율이 다소 높게 나타났다.

가구 총소득의 경우 4,500만 원 이상 7,000만 원 이하의 소득을 가지는 집단은 다른 집단에 비해 은퇴율이 가장 높고 은퇴까지의 소요 시간도 가장 짧게 나타난 반면, 5,700만원 이상의 집단은 은퇴율도 가장 낮고 은퇴까지의 소요 시간도 가장 길게 나타났다.

본인 건강 상태 만족도의 경우, 예상과는 달리 만족도가 높은 집단일수록 은퇴율이 높게 나타났으며, 80점 이상의 집단은 은퇴까지의 소요 기간이 6년으로 가장 짧게 나타났다. 이는 현재 일을 할 수 있을 만큼 건강한 사람들의 은퇴가 많다는 것을 보여준다. 삶의 질 만족도의 경우, 점수가 20점 미만으로 낮거나 반대로 80점 이상으로 높은 집단의 은퇴율이 다른 집단에 비해 높게 나타났다. 다만, 20점 미만의 집단은 은퇴까지의 소요 기간이 4년으로 짧은 반면, 80점 이상인 집단은 은퇴까지의 소요 기간이 8년으로 길게 나타나 상반된 결과를 보였다.

표 4. Kaplan-Meier 분석 결과

Table 4. The Results of Kaplan-Meier

Variables	Prob. of Retirement	Lead Time (Median)
Total	72.4	8
Age	55~59	69.0
	60~64	70.7
	>=65	80.2
Sex	Male	70.7
	Female	77.8
Inconvenience in ADL	No	72.4
	Yes	71.8
Diabetes / hyperglycemia	No	73.1
	Yes	67.2
Difficulty of daily activity due to pains	No	71.6
	Yes	77.5
Support for national health insurance in current job	No	76.3
	Self-employed	53.6
	Salaried employee	71.5
Labor union	No	71.8
	Yes	76.3
Seniority system	Agree	78.1
	Disagree	73.1
	Strongly agree	75.0
	Strongly disagree	70.4
Stability of employment	Agree	75.0
	Disagree	68.8
	Strongly agree	57.1
	Strongly disagree	78.8
Need for physical power in current job	Agree	72.9
	Disagree	71.1
	Strongly agree	73.2
Need for carrying heavy loads	Strongly disagree	71.4
	Agree	71.9
	Disagree	75.4
Need for computer skills in current job	Strongly agree	68.2
	Strongly disagree	68.3
	Agree	76.4
	Disagree	63.0
Total income of household	Strongly agree	70.0
	Strongly disagree	75.4
	<1,000	71.9
	1,000~3,000	72.2
Satisfaction on own health status	3,000~5,000	69.9
	5,000~7,000	84.2
	>=7,000	60.0
	<20	66.7
	20~50	70.5
Satisfaction on the quality of life	50~80	72.6
	>=80	73.4
	<20	75.0
	20~50	66.1
	50~80	72.7
	>=80	74.1

**표 5. Cox Regression 분석 결과**  
**Table 5. The Results of Cox Regression**

Variables		beta	SE	HR
Age(ref= >=65)	55~59	-0.498***	0.174	0.608
	60~64	-0.244	0.160	0.784
Sex(ref=Female)	Male	-0.125	0.147	0.883
Inconvenience in ADL(ref=Yes)	No	0.148	0.222	1.160
Diabetes / hyperglycemia(ref=Yes)	No	0.153	0.191	1.166
Difficulty of daily activity due to pains(ref=Yes)	No	-0.273	0.181	0.761
Support for national health insurance in current job (ref=Salaried employees)	No	0.061	0.169	1.063
	Self-employed	-0.518*	0.305	0.596
Labor union(ref=Yes)	No	-0.167	0.193	0.846
Seniority system (ref=Strongly disagree)	Agree	0.264	0.252	1.302
	Disagree	0.188	0.149	1.207
	Strongly agree	0.550	0.550	1.734
Stability of employment (ref=Strongly disagree)	Agree	-0.266	0.202	0.766
	Disagree	-0.350*	0.191	0.705
	Strongly agree	-1.067**	0.420	0.344
Need for physical power in current job (ref=Strongly disagree)	Agree	-0.149	0.343	0.862
	Disagree	-0.201	0.321	0.818
	Strongly agree	-0.139	0.363	0.870
Need for carrying heavy loads (ref=Strongly disagree)	Agree	0.080	0.329	1.084
	Disagree	0.209	0.297	1.233
	Strongly agree	0.035	0.372	1.036
Need for computer skills in current job (ref=Strongly disagree)	Agree	-0.069	0.230	0.933
	Disagree	-0.389**	0.164	0.678
	Strongly agree	-0.292	0.369	0.746
Total income of household (ref= >= 7,000)	< 1,000	0.017	0.472	1.017
	1,000~3,000	0.246	0.471	1.279
	3,000~5,000	0.284	0.479	1.328
	5,000~7,000	0.687	0.495	1.988
Satisfaction on own health status (ref= >= 80)	< 20	-0.058	0.445	0.944
	20~50	-0.156	0.210	0.856
	50~80	-0.058	0.143	0.944
Satisfaction on the quality of life(ref= >=80)	< 20	0.575	0.493	1.777
	20~50	0.146	0.233	1.157
	50~80	0.162	0.144	1.176

\*: p<0.1, \*\*: p<0.05, \*\*\*: p<0.01

**2) Cox Regression 분석 결과**

본 연구에서는 Cox Regression을 이용하여 주요 독립변수들이 은퇴율에 미치는 영향을 분석하였으며, 그 결과는 <표 5>와 같다. 각 독립변수별로 비례위험가정 만족 여부를 확인하였다. 특정 독립변수가 비례위험 가정을 만족하지 않는다는 것은 시간에 따라 해당 독립변수의 효과가 달라진다는 것을 의미한다.

분석 결과 비례위험 가정을 만족하는 변수는 ‘나이’, ‘건강보험 가입’, ‘고용 안정성 여부’, ‘업무수행 시 컴퓨터 활용능력 필요 여부’로 검증되었다.

구체적으로 살펴보면 ‘나이’에서 ‘55~59세’가 유의수준 범위 내에서 추정되었으며, 경과 시간에 상관없이 은퇴할 가능성은 ‘65세 이상’인 경우의 0.61배로 나타났다.

‘건강보험 가입’에서는 ‘지역 가입’의 경우 은퇴할 가능성이 ‘직장 가입’인 경우의 0.6배인 것으로 추정되었다.

‘고용 안정성 여부’별로 살펴보면, ‘전혀 그렇지 않다’를 준거변수로 설정하였을 때 ‘그렇지 않은 편이다’, ‘매우 그렇다’인 경우의 은퇴 가능성이 0.71배 및 0.34배로 나타났다.

‘업무수행 시 컴퓨터 활용 능력 필요 여부’의 경우 ‘그렇지 않은 편이다’에 응답한 사람의 은퇴할 가능성이 ‘전혀 그렇지 않다’인 경우의 0.68배로 검증되었다.

분석 결과를 정리하면 ‘55~59세’가 65세 이상의 노령층보다 은퇴할 확률이 낮은 결과를 보이는데, 이는 비교적 젊은 연령대는 중·고령자가 은퇴하지 않을 가능성을 높이며, 65세 이상의 노령층은 고령화되어 생물학적으로 은퇴를 이행하는 단계에 접어들었기 때문이다. 또 다른 요인은 일자리 특성과 관련되어 있다. 즉 자영업자 등의 정년이 정해지지 않은 일자리를 의미하는 ‘건강보험 가입(지역)’, 고용조건이 안정적인 또는 안정적이지 않은 일자리(‘고용 안정성 여부’)와 ‘업무수행 시 컴퓨터 활용 능력 필요 여부’ 요인으로 추측할 수 있는 사무직 등 화이트 칼라의 직종에 종사할 가능성이 낮은 경우이다.

**V. 결 론**

본 연구는 중·고령자를 장기적으로 추적한 패널데이터를 활용하여 일정한 시간이 흐른 후의 은퇴 양상을 분석하였다.

분석 결과를 종합하면 은퇴와 관련된 요인은 두 가지 영역으로 검증되었다. 첫 번째는 생물학적 요인에 해당하는 연령과 그에 따른 건강 내지는 업무 수행 능력이며, 두 번째는 고용 특성이다. 후자와 관련하여 중·고령자의 은퇴는 상반된 유형을 보인다. 한 가지는 고용조건이 안정적이거나, 사무직을 비롯하여 어느 정도의 전문성을 갖춘 직업이므로 사업체에서 규정하는 정년에 도달하여 은퇴를 이행한 경우이다. 다른 한 가지는 정년이 정해지지 않은 자영업·프리랜서 등의 일을 하거나, 비정규직 또는 비정형 일자리의 분포가 높은 육체적 노동자, 연공성이 중요시되지 않는 등 조직의 체제성이 약한 소규모 사업장에 종사함으로써 고용이 유지되지 않거나 오히려 은퇴 이행 시기가 늦추어진 유형이다.

본 연구의 결과는 은퇴와 관련한 정책 이슈들과 관련이 있다. 기대수명이 늘어남에 따라 정년 연령을 늦추어야 한다는 사회적 논의가 진행되고 있다. 이로 인하여 체계적인 노후 대비가 어려워지고 비자발적인 은퇴로 인한 사회적 관계의 단절을 경험할 수 있다. 또한 중·고령자가 취업 가능한 일자리는 대체로 단순노동과 비정형 일자리를 중심으로 구성된다. 때문에 고용유지가 불안정하고 임금 수준이 낮아 노년층의 빈곤과도 무관하지 않다. 이 연구는 중·고령자의 은퇴를 개인 차원에서만 머무르지 않고 노동시장 환경과도 연결지어 시사점을 도출하였다는 점에서 의의가 있다. 이는 직업훈련 및 평생교육 등의 인적자원 개발과 은퇴 관련 인식의 제고, 노후복지와 중·고령자 고용 정책으로도 방향을 넓힐 수 있을 것이다.

또한 본 연구는 은퇴를 결정하는 주요 요인들을 찾고, 요인들 간 패턴을 분석하여 은퇴 가능성이 높은 패턴과 은퇴하지 않을 가능성이 높은 패턴을 도출하였으며, 주요 변수의 변수값별로 은퇴 확률과 은퇴까지의 소요 기간, 그리고 시간에 따른 변수별 효과를 제시하였다는 점에서 학문적 시사점을 갖는다.

본 연구는 사용한 데이터의 특성상 분석에 사용한 데이터의 수가 많지 않다는 한계점이 있다. 하지만 이는 은퇴에 관련된 연구들이 특정한 연령대 및 조건의 고려해야 함에 따라 갖게 되는 공통적인 문제점이며, 향후 수집 가능한 데이터의 양이 증가하면 완화될 수 있을 것으로 여겨진다.

## 참고문헌

- [1] J. E. Hong, J. Kim, S. Lee, K. Kim, and J. Lee, *Study on Middle and Old-aged Female Workers' Economic Conditions; Focusing on Retirement Age*, Korean Women's Development Institute, Research Report, 2020.
- [2] S. Seok, "The Effect of Retirement on Health", *Journal of Labor Policy*, 11(1), pp. 81-102, 2011.
- [3] K. Park, "Labor Exit Process of Old Workers Aged 55 and over: Implication for Typology of Retirement", *Journal of Labor Policy*, 3(2), pp. 103-140, 2003.
- [4] K. Park, "A Study to Classify the Type of Retirement Process among the Mature-aged in Korea: Focusing on Diversity and Inequality", *Korean Journal of Social Welfare Studies*, 42(3), pp. 291-327, 2011.
- [5] H. Bang and I. Shin, "A Forced Choice: the dynamics of retirement and reemployment of old-aged workers in Korean", *Korean Journal of Sociology*, 45(1), pp. 73-108, 2011.
- [6] J. Son. "Determinants of Retirement and Retirement Satisfaction in Korea", *Journal of Labor Policy*, 10(2), pp. 125-153, 2010.
- [7] J. Curtis, and J. McMullin, "Dynamics of retirement income inequality in Canada", 1991-2011. *Journal of Population Ageing*, 12(1), pp. 51-68, 2019.
- [8] D. A. Hershey, K. Henkens, and H. P. van Dalen, "What drives retirement income worries in Europe? A multilevel analysis". *European journal of Ageing*, 7(4), pp. 301-311, 2010.
- [9] H. Jun, "Social security and retirement in fast-aging middle-income countries: Evidence from Korea". *The Journal of the Economics of Ageing*, 17, pp. 1-14, 2020.
- [10] K. Kim and S. Yang, "A Study on the Financial Status and Life Satisfaction of Retired Households", *Financial Planning Review*, 10(1), pp. 27-60, 2017.
- [11] F. Mazzonna, and F. Peracchi, "Unhealthy retirement?", *Journal of Human Resources*, 52(1), pp. 128-151, 2017.
- [12] D. Cho, "The Determinants of the Retirement of Aged People Across Cohorts", *Journal of Industrial Relations*, 24(1), pp. 47-66, 2014.
- [13] KOSTAT, Economically Active Population Survey, 2020.
- [14] R. Sewdas, A. De Wind, L. G. Van Der Zwaan, W. E. Van Der Borg, R. Steenbeek, A. J. Van Der Beek, and C. R. Boot, "Why older workers work beyond the retirement age: a qualitative study". *BMC public health*, 17(1), pp. 672, 2017.
- [15] S. Kim, S. Jang, J. Lee and K. Moon, "Factors affecting on the Willingness of Re-employment among Babyboomers", *Journal of Gerontological Social Welfare*, 67, pp. 107-132, 2015.
- [16] H. M. Hasselhorn, M. Ebener, and A. Vratzias, "Household income and retirement perspective among older workers in Germany: Findings from the lidA Cohort Study". *Journal of Occupational Health*, 62(1), pp.1-9, 2020.
- [17] U. Bratun, and J. Zurc, "The motives of people who delay retirement: An occupational perspective". *Scandinavian Journal of Occupational Therapy*, pp. 1-13, 2020.
- [18] M. Berry, and G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, Inc.; New York, 1997.
- [19] Gartner term, Gartner.com. .   
 <https://www.gartner.com/en/information-technology/glossary/data-mining>
- [20] Y. Kim, *Survival Analysis*, Free Academy, 2013.
- [21] K. Song and J. Ahn, *Survival Analysis using SPSS for windows*, Hannarae Publishing Co, 2006.



**김수용(Suyong Kim)**

2020년: 한밭대학교 창업경영대학원 빅데이터비즈니스학과 (경영학석사)

2013년~현 재: 세림티에스지(주)

※관심분야: 데이터웨어하우스, 텍스트마이닝, 워드임베딩, 딥러닝(RNN,LSTM,GAN), 데이터 네트워크 가시화 등



**김건우(Gunwoo Kim)**

2010년: 고려대학교 대학원 (경영학 박사)

2011년~현 재: 한밭대학교 융합경영학과

※관심분야: 비즈니스 온톨로지 모델, 빅데이터 분석, 핀테크 기술 및 전략 등



**이정화(Jeonghwa Lee)**

2015년: 성균관대학교 대학원 (행정학 박사)

2016년~현 재: 근로복지공단 근로복지연구원

※관심분야: 사회보험, 복지국가, 여성학, 공공정책, 노동시장 등



**최근호(Keunho Choi)**

2013년: 고려대학교 대학원 (경영학 박사)

2018년~현 재: 한밭대학교 융합경영학과

※관심분야: 추천시스템, 의료 빅데이터 분석, 딥러닝, 머신러닝, 데이터 마이닝 등