

가구의 주거 및 사회경제적 특성과 출산 간의 패턴 분석: 의사결정나무모형 기법을 이용하여

이정화¹ · 김진우² · 최근호^{3*}

¹근로복지공단 근로복지연구원 책임연구원

²한밭대학교 경영회계학과 부교수

³한밭대학교 경영회계학과 조교수

The Relationship between Childbirth, Housing and Socio-Economic Factors: The Pattern Analysis using Decision Tree

Jeonghwa Lee¹ · Gunwoo Kim² · Keunho Choi^{3*}

¹Research Fellow, Labor Welfare Research Institute, Seoul 07254, Korea

²Associate Professor, Department of Business Administration, Hanbat National University, Daejeon 34158, Korea

³Assistant Professor, Department of Business Administration, Hanbat National University, Daejeon 34158, Korea

[요 약]

본 연구는 주거와 출산 간의 관계를 짚어보고 출산·양육 정책에 대한 의미 있는 결과와 시사점을 제공하고자 한다. 주거와 관련한 특징에 대해 다각화된 접근을 시도하였으며, 데이터 마이닝 기법 중 의사결정나무 기법을 통한 패턴 분석에 근거하여 출산에 영향을 주는 여러 요인들 간의 관계를 구조화하였다. 재정패널조사의 5~10차년도 데이터를 분석한 결과 전체 71개 변인 중 중요도가 높은 상위 20개 변인과 7개의 패턴을 도출하였다. 이러한 결과는 출산이 단순히 가구의 의사결정이 아니라 전체적인 사회의 환경 및 제도와 밀접한 관련을 맺는다는 점을 보여준다.

[Abstract]

This study aims to identify the relationship between housing and childbirth, and to contribute to childbirth and child care policies. This study attempted to diversify factors affecting childbirth, and to structure their relationships to childbirth based on pattern analysis using a decision tree. To this end, this study analyzed the panel data given by Korea Institute of Public Finance(KIPF). As a result, this study identified the top 20 important factors affecting the childbirth and derived seven meaningful patterns. These patterns let us know that although the decision making for childbirth depends heavily on the family formation plan of parents, the socio-economic factors including the housing also affect the decision significantly.

색인어 : 출산, 가족 형성, 주거 특성, 데이터 마이닝, 의사결정나무

Key word : Data mining, Decision Tree, childbirth, Family Formation, Housing

<http://dx.doi.org/10.9728/dcs.2020.21.2.327>



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 10 December 2019; **Revised** 01 January 2020

Accepted 25 February 2020

***Corresponding Author; Keunho Choi**

Tel: +82-42-821-1468

E-mail: keunho@hanbat.ac.kr

I. 연구의 배경 및 목적

주거는 현대 사회의 가족 형성(family formation)과 생활의 중요한 근간을 구성한다[1]. 한국 사회에서의 주거 상황은 심각한 사회 문제로 대두되고 있다. 2018년 주택보유가구의 비율은 56.2%로 절반을 조금 넘어선 수준이며[2], 수도권은 주택매매가는 최근 5년 간 지속적으로 상승(주택가격지수: 2015년 96.1 → 2019년 103.0)¹⁾하여 왔다[3]. 이러한 사회적 현실 속에서, 최근 정부의 주거 정책은 출산·양육 정책의 중요한 수단으로 읽힌다. 국토교통부의 저소득 다자녀가구 지원 정책, 2017년 내놓은 주거복지로드맵('17년) 등 다양한 정책들을 통하여, 한국 사회의 심각한 주거 문제와 혼인·출산 간의 연쇄적 고리에 대한 정부의 고민을 엿볼 수 있다. 이처럼 저출산 현상에 대한 정책적 대응 방안은, 주거 정책 등 거시적이고 구조적인 문제로부터 출산을 유도하는 구조적 접근(infra approach)으로 변화하고 있다[4].

주거는 가족의 경제적 안정성(economic security)[5]을 나타내는 동시에 사회경제적 지위를 대리하는 개념으로도 해석된다. Ben-Shahar & Warszawski(2016), Arundel(2017) 등 주거의 구매력(housing affordability)과 주거불평등(housing inequality)에 관한 다수의 연구들에서도 이와 같은 주거의 의미, 그리고 주거를 통해서 드러나는 사회경제적 지위의 특성을 강조하였다[5],[6]. 개인(가구)는 생애주기 단계별로 달라지는 경제적 제약이나 주거 공간의 위치·크기 등에 관한 욕구와 선호에 직면하게 되므로, 주거의 선택 역시 생애주기의 영향을 받을 수밖에 없기 때문이다[7]. 또한 가구의 주거 상황을 통하여 드러나는 가구의 사회경제적 수준은 부부가 자녀의 출산을 결정하는 과정에 복합적으로 작용한다. 주택매매가격의 상승은 출산율에 대하여 부정적인 영향을 미칠 수 있으며, 전세가격의 상승 역시 생계부담을 가중시켜 출산을 연기하는 등의 출산율 하락으로 귀결될 위험을 초래한다[8],[9]. 그 외 주거 면적, 지역, 주거 형태, 주거소유권, 주거지 이동 등 주거 상황과 출산 간의 관계는 다양한 차원에서 형성된다[1],[4],[10]-[13].

본 연구는 이러한 배경을 토대로 하여 주거와 출산 간의 관계를 짚어보고 출산·양육 정책에 대한 의미 있는 결과와 시사점을 제공하고자 한다. 주거의 특성에 대해 다각화된 접근을 시도하였으며, 데이터 마이닝 기법 중 의사결정나무 기법을 통한 패턴 분석에 근거하여 출산에 영향을 미치는 여러 요인들 간의 관계를 구조화했다는 점에서 차별된 시각을 제시하고자 한다.

1) 전국 기준 종합 매매가격지수의 평균값이며, 각 연도별 10월 기준 통계자료임.

II. 본 론

2.1. 출산과 주거

흥미롭게도 주거의 소유와 출산은 상반된 연구 결과를 보인다. Feijten & Mulder(2002), Mulder(2006)에서도 설명하듯 분명 주거의 점유는 출산을 증진시키는 효과가 있으며, 임신과 같은 구체적인 가족계획이 자가(自家) 마련을 유도하기도 한다[11],[14]. 하지만 이들 연구에 따르면 가구는 출산·양육에서 소요되는 다양한 비용을 고려하여 출산에 대한 의사결정을 내리고자 하므로, 주거를 소유하기 위한 비용은 가족형성에 따르는 비용과의 갈등을 불러일으킬 수 있다.

한편 주택매매가격이나 주택전세가격의 상승은 가구의 생계부담을 가중시켜 출산을 연기시키는 원인이 될 수 있다[8]. 또한 박진백·이재희(2016)에서는 출산에 대한 주택 가격의 영향력이 주거를 소유하는지에 따라 각기 상이한 방향으로 작용한다[9]. Mizutani(2015)에서도 주택 용자가 없는 가구의 경우 주택 가격의 상승이 출산과 유의한 양의 관계를 형성하였다[15].

한편 Kulu & Vikat(2007)에 의하면 가구는 출산을 위하여 적합한 환경을 갖추는 거주지로 이동하며, 이러한 가구에서 출산율이 증가하는 관계를 보였다[12]. 이는 개별 가구가 출산이라는 가구의 상황(household situation)에 의하여 주거 상황(housing situation)을 조정한다는 설명과도 일치한다[11]. 또한 Kulu & Washbrook(2014)에서는 출산·양육에 친화적인 환경일수록 출산에 대하여 긍정적인 효과를 발휘하였다[13].

한국의 사례로 천현숙 외(2012)에 의하면 주거의 상황이 안정적이거나 양호할수록 출산율과 양(+)의 관계를 맺는다[10]. 그러나 젊은 연령층일수록 주거지 마련의 어려움을 겪으며[16], 신혼기에 독립적인 경제력을 갖추지 못하며 부모의 자산에 대한 의존도가 높다[3]. 즉 젊은 세대에서 나타나는, 주거 자산에 의한 불평등(Arundel, 2017)이 나타나는 현실이다[5].

이러한 연구들은 공통적으로 주거지 마련을 위한 비용이 가구의 경제적 안정성을 위협하는 상황을 지적한다. 특히 최근의 매매 및 전·월세 추이는 주거생활을 비롯한 개별 가구경제 간의 큰 격차를 야기하는 원인으로 진단된다.

2.2. 출산과 사회경제적 특성 및 가족정책

주거는 가구의 사회경제적 특성과도 밀접한 관련을 맺으므로 가구의 소득 수준에 따라 출산하는 자녀의 수가 다르고[17], 둘째 자녀를 출산하는 시기가 가구경제와 관련이 있는가 하면[18], 첫째 자녀에게 지출하는 비용(특히 교육비)이 둘째 자녀의 출산에 영향을 미친다는 결

과도 있다[19]. 다소간의 차이가 있으나 선행연구들은 대체적으로 가구의 출산이 가구소득으로 대표되는 양육 자원의 상한선, 그리고 제한적인 자원 내에서 자녀(들)에게 배분할 수 있는 자원의 양과 질의 축에서 이루어지는 선택이라는 점에 동의한다[19],[20]. 때문에 첫째 출산과 그 이후에 이루어지는 출산의 배경은 동일하지 않으며, 가구의 소득이나 지출 수준은 이를 형성하는 주요한 축으로 기능한다는 공통된 의견을 확인할 수 있다.

또한 교육수준이나 노동시장 참여가 출산에 미치는 영향은 성별에 따라 다른 의미로 풀이된다. 이에 여성의 교육수준과 직종[22], 여성의 종사상 지위[23] 등 노동시장 참여의 특성에 초점을 두기도 한다. 특히 여성에게 출산과 양육은 여성의 노동시장 참여에 따른 기회비용이자[24] 인적자본을 제약하는 요인[25]이라는 점에서 남성의 경우와 차이가 있다.

한국의 출산·양육 정책과 관련한 연구를 살펴보면 자녀가 있는 가구는 더욱 보편적인 양육 정책에 대한 욕구를 갖고 있으며[26], 출산·양육 관련 공공정책은 출산의 지에 긍정적인 요인으로 기능할 수 있다[27],[28]. 하지만 수당의 지급 수준과 효과에 대해서 회의적인 시각도 여전히 남아있는데, 고제이·고경표(2017)에서 나타나는 바와 같이 영유아 자녀를 둔 부모는 자녀의 수나 가구소득 수준에 따라 지원금의 차등이 필요하다는 인식을 가질 여지가 있으므로[29] 향후 점진적인 정책의 개선이 수반되어야 할 것이다. 현재 지방자치단체는 지역사회의 현안과 특성을 고려하여 다양한 정책들을 시행하는데, 이종하·황진영(2018)에서 나타나듯 각 지역사회 특성이나 인구구조 등으로 인하여 그 실효성은 지역 간의 편차가 있다[21].

이상의 선행연구들은 대부분 회귀분석 등의 통계적 기법을 사용하여 주거 또는 사회경제적 특성과 관련된 개별 요인들이 출산에 미치는 영향을 분석하였다. 하지만, 본 연구는 기존 연구들에서 거의 사용되어 오지 않았던 데이터 마이닝을 분석 기법으로 사용하여, 주거 및 사회경제적 특성과 관련된 요인들이 출산에 미치는 개별 영향력뿐만 아니라, 요인들 간의 상호작용이 출산에 미치는 영향력도 함께 입체적으로 다루었다는 점에서 기존 연구들과 차별점이 있다.

III. 연구방법

3.1. 분석 프레임워크 및 분석 방법

본 연구에서 모델 구축을 위해 사용한 데이터 마이닝 기법은 분석 데이터로부터 의미 있는 패턴이나 지식을 추출하기 위한 목적의 작업이다[30],[31].

본 연구에서는 이 데이터 마이닝 기법 중 특히 의사결

정나무 알고리즘을 사용하였는데, 이 알고리즘은 독립변수와 목표변수 간의 관계뿐만 아니라 독립변수들 간의 복합적인 연관 관계를 탐색할 수 있으며, 다양한 변수들을 사용할 수 있고 예측 결과에 대한 설명력이 높은 장점을 갖고 있다. 본 연구는 단순히 예측 정확도가 높은 모델을 개발하는 것이 아니라, 원인을 설명해 줄 수 있어 정책 수립에 도움을 줄 수 있는 모델을 만드는 것으로 하고 있기 때문에 설명력이 높은 의사결정나무 알고리즘은 본 연구의 주제에 적합한 방법이다.

1) 분석 데이터 및 데이터 전처리

본고는 한국조세연구원에서 제공하는 재정패널 데이터 중 모든 차수의 데이터가 포함된 통합 데이터를 기반으로 본 연구에 적합한 형태의 분석 데이터를 구성하였다. 분석 데이터를 이용하여 결측치를 포함한 관측치 제거 및 새로운 변수 생성 등의 데이터 전처리 작업을 수행하였고, 데이터 마이닝 모델 구축 시 균형 잡힌 학습이 실행될 수 있도록 목표변수의 두 값(출산·미출산)의 비율을 균등하게 맞춰주는 데이터 균형화 작업을 수행하였다. 이와 같이 분석데이터를 구축한 후 Gain Ratio(GR) 알고리즘을 이용하여 변수의 중요도 평가를 실시하고, 다음으로 의사결정나무(Decision Tree) 알고리즘을 분석 데이터에 적용하여 출산 여부 예측하는 모델을 구축하고 이로부터 주거 등 가구의 경제적 특성과 출산간의 패턴을 도출하였다. 이 때 변수의 중요도 평가는 각각의 독립변수가 목표변수인 출산 여부를 예측하는데 얼마나 중요한 역할을 하는지를 보여주며, 의사결정나무 모형은 각 노드를 통하여 변수 간 결합하는 특정한 상황에서의 출산을 나타낸다.

분석데이터는 재정패널조사의 5차년도(2012년)부터 10차년도(2017년)까지의 패널데이터를 사용하였으며, 분석의 대상은 부부 중 1인이 가구주이며 혼인상태를 유지하고 동일한 가구에서 동거하는 가구이다. 데이터 처리 과정에서 반영한 사항을 정리하면 다음과 같다.

첫째, 본 연구에서는 목표변수인 출산 여부를 생성하기 위하여 ‘작년 한 해 출산 경험 가구원 유무’ 변수를 사용하였다. 이 때 임신기간 등을 고려하여 독립변수와 목표변수(출산) 간에 1년의 시간차를 두었으며, 목표변수는 재정패널조사에서 6차년도 조사부터 수집된 자료이므로 분석 자료의 시간적 범위는 5~10차년도로 정의하였다.

둘째, 본 연구의 분석 단위는 개별 가구가 아닌 개별 가구의 특성이므로 동일한 가구일지라도 가구의 특성은 1회 이상 관측될 수 있다.

셋째, 개별 가구의 범위는 첫 번째 가구원(가구원①)이 가구주이며 두 번째 가구원(가구원②)이 가구주의 배우자라고 응답한 가구이다. 또한 일반적인 출산 연령을 고려하여, 부부가 모두 만 45세 이하인 가구로 대상을 한정

하였다.

넷째, 패널데이터 상에서 가구원별 독립변수 값은 가구별로 응답한 사람에 따라 해당 변수가 지칭하는 대상이 달라질 수 있다. 이에 첫 번째 가구원(가구원①)에 대한 독립변수 값은 남성 배우자, 두 번째 가구원(가구원②)에 대한 값은 여성 배우자의 값으로 모두 변환하여 가구원 번호별 분석의 대상을 통일하였다.

다섯째, 분석 자료 상 목표변수의 두 값이 미출산 4,734건, 출산 232건으로 편향된 상태로 확인되었는데, 목표변수의 특정 값을 지닌 관측치의 수가 다른 값을 지닌 관측치의 수보다 상대적으로 많은 경우에는 편향된 모델을 만들 우려가 있기 때문에, 모델 구축 전 목표변수 값의 비율을 서로 동일하게 맞추어 주기 위한 데이터 균형화 작업을 거쳤다. 본고에서는 분석 단위가 가구의 특성인 점을 고려하여, 관측치의 수가 적은 목표변수 값인 출산 232건에 맞추어 미출산 232건을 무작위로 표집하는 과소표집(under-sampling) 방식을 택하였으며, 그 결과 총 464건으로 이루어진 데이터셋을 구축하였다.

표 1. 주요 변인의 기술통계량

Table 1. Descriptive Statistics of Major Variables

Variable	Pre-Balancing (n=4,966) (mean/std)	Post-Balancing (n=464) (mean/std)
Age of Couple	Male = 41.22/5.38 Female = 38.51/4.86	Male = 38.62/5.71 Female = 35.94/5.22
Number of Children aged from 6 to 17	1.15/0.92	0.73/0.88
Number of household members	3.79/0.83	3.37/0.91
Residential Acquisition Period(year)	7.08/4.59 (n=2,766)	6.53/4.50 (n=225)
Current Income (10,000KRW, month)	445.44/240.32	429.36/279.81
Total Household Expenditure (10,000KRW, year)	5,067.92 /2,617.59	4,673.14/2,520.63

2) 변수 중요도 평가

변수의 중요도 평가를 통해서 각각의 독립변수들이 목표변수인 출산 여부의 예측에서 어느 정도의 중요성을 지니는지 추정할 수 있다. 여기에서 변수의 중요도란 해당 변수의 값을 기준으로 관측치를 분류하였을 때, 동일하게 분류된 관측치들이 가지고 있는 목표변수의 값이 얼마나 동질적인지를 나타낸다. 즉, 중요도가 높은 독립변수일수록 목표변수의 값을 더욱 정확하게 예측할 수 있게 된다. 본 연구에서는 이러한 변수의 중요도를 평가할 수 있는 여러 알고리즘들 중 많은 연구들에서 사용되어 온 Gain Ratio(GR) 알고리즘을 이용하여 변수의 중요도를 평가하였다.

3) 모델 구축 및 패턴 도출

이 연구에서는 데이터 마이닝 기법 중 한 가지인 의사결정나무 알고리즘을 사용하여 가구의 출산 여부를 예측하기 위한 모델을 구축하였다. 의사결정나무 알고리즘은 모델을 구축하는 시간이 빠르고 정확도가 높아 많은 연구들에서 사용되어 온 알고리즘이다. 이는 또한 생성된 패턴을 이해하기 쉬운 형태로 보여주기 때문에 설명력이 뛰어난 장점을 갖고 있다.

본 연구에서는 여러 의사결정나무 알고리즘 중 가장 널리 사용되어 온 Gain Ratio(GR)을 분할 기준으로 사용하여 집합의 순수도를 계산하는 C4.5 알고리즘을 이용하였다. 전체 데이터를 학습데이터 70%(325건)와 검증데이터 30%(139건)의 비율로 분할하였으며, 기존의 데이터 마이닝 연구들에서 널리 사용되어 온 Weka ver.3.8.2를 데이터 마이닝 툴로 사용하여 분석을 수행하였다.

3.2. 연구모형 및 변수의 구성

이상의 과정을 통하여 총 71개의 변수를 초기 독립변수로 투입하였다. 변수의 영역은 크게 주거 특성, 가구경제, 가구원 특성, 가족생활 관련 공공정책으로 구분한다.

주거 특성은 크게 현재의 주거 특성과 주거 관련 지출로 구성된다. 전자는 자가 여부인 거주 형태, 단독주택 등 주택의 형태, 현재 거주하는 주택의 취득 기간과 공시매매가, 거주 지역 및 주거 환경에 대한 만족도 변인을 포함하였다. 후자에서는 주거비 금액과 가구에서 소득·지출 대비 부담하는 수준, 전·월세 보증금과 월세, 이사와 주거지 매입의 이력과 그에 소요된 비용(가격)의 변수를 투입하였다.

가구 경제에 대해서는 소득·지출·자산을 다각도로 살펴볼 수 있는 요인들과 상속 또는 증여 관련 변인을 통계하였다. 현재 가구의 규모나 부부의 자녀계획도 출산 결정의 중요한 요인이므로 전체 가구원과 연령대를 고려한 자녀의 현황, 그리고 부부(가구원①, 가구원②)의 특성과 연령 차이를 통하여 가구원 구성을 살펴보고자 하였다.

가족생활과 관련한 공공정책으로는 가사도우미 서비스, 출산·양육과 관련한 지원금, 기타 바우처 등을 포함하는 정부지원(현물)의 수급 현황을 분석 변수로 투입하였다.

IV. 분석 결과

4.1. 변수 중요도 평가 결과

<표 2>는 독립변수의 중요도 평가 결과로서 중요도가 높은 상위 20개 변수를 보여준다. 분석 결과에서는 출산에 중요한 영향을 미치는 요인들 파악할 수 있으며, 가구원 수를 중심으로 하는 ‘가구원 특성’과 함께 ‘가구경제’

및 ‘주거 특성’에 속하는 요인들이 도출되었다.

표 2. 변수 중요도 평가 결과 (상위 20개 변수)
Table 2. Importance Evaluation of Independent Variables(Top 20)

Rank	Name	Rank	Name
1	Number of Children aged from 6 to 17	11	Real Estate Ownership: Land(y/n)
2	Number of household members	12	Housing Price(KRW)
3	Number of children	13	Household expenditure on communication(KRW)
4	Residential Acquisition Period(year)	14	Real Estate(KRW)
5	Real Estate Ownership: Building	15	Tenancy Deposit(KRW)
6	Household expenditure on Education(KRW)	16	National Healthcare Expenditure(KRW)
7	Household Members Transferred Asset(y/n)	17	Total deposit from Real Estate Asset(KRW)
8	Direct Household Expenditures(KRW)	18	Number of Children aged under 6
9	Current Income(KRW)	19	Transportation Expenditure(KRW)
10	Total Household Expenditure(KRW)	20	Housing Expenditure(% of total household expenditure)

연구에서 사용한 독립변수 중 가구의 출산 여부를 예측하는 데 가장 중요한 변수는 ‘가구원 수 - 6세 이상 18세 미만 자녀’로 나타났고, ‘총 가구원 수’와 ‘총 자녀 수’를 포함한 상위 3개 변수가 모두 가구원 수와 관련된 변수로 확인된다. 즉 자녀를 출산하는 데에는 현재 자녀가 있는지 여부와 해당 자녀의 나이가 많은 영향을 미치는 관계임을 해석할 수 있다.

중요도 상위 20개 변수 중 주거 특성과 관련한 변수로는 자가(自家)의 가구에 해당하는 주거지 취득 기간과 주택 가격, 거주하는 주택의 전세금 또는 월세 보증금, 전세금 및 월세의 보증금 총액, 주거비 비율이 확인되었다. 이는 가구의 출산을 결정짓는 요소에서 주거의 특성과 환경이 가구의 출산에 영향을 미치는 결과를 보여주고 있다. 또한, 전체 가구 지출에서 주거비가 차지하는 비율인 ‘지출 대비 주거비(%)’ 변수는 중요도가 20위로 나타났고, 전체 가구 소득에서 주거비가 차지하는 비율인 ‘소득 대비 주거비(%)’ 변수는 중요도가 60위로 나타나, 전체 가구 소득에서 주거비가 차지하는 비율보다는 전체 가구 지출에서 주거비가 차지하는 비율이 가구의 출산 여부에 더 많은 영향을 주는 것으로 나타났다. 또한 중요도 상위 20개 변수 중 경제적 특성과 관련된 변수를 살펴보면, 가구의 경제적 상황과 관련한 변수의 중요도 평가 결과에서는 가구 소비(교육비, 직접소비지출, 가구총지출, 통신

비, 건강보험료, 교통비), 소득(경상소득), 부동산(주택 외 건물 및 토지, 부동산 자산 가치, 부동산 관련 보증금 총액), 증여한 가구원의 유무의 변인들이 확인되었다. 즉 가구의 소득보다는 가구의 자산과 지출이 가구의 출산에서 더욱 중요한 역할을 수행하는 결과로 해석할 수 있다. 상기 변수들은 가구원 구성, 주거, 가구 소득·소비, 자산의 4가지 유형으로 재분류 된다.

4.2. 가구의 출산 여부 예측모델 구축 결과

전체 분석 데이터 중 70%인 325건을 학습데이터로 이용하여 의사결정나무 알고리즘을 학습시키고, 나머지 30%인 139건을 검증데이터로 이용하여 구축된 모델의 정확도를 평가한 결과는 <표 3>과 같다.

우선 전체 예측 관측치 중 모델에 의해 예측된 목표변수 값과 실제 목표변수 값이 일치하는 관측치의 비율인 적중률(hit ratio)은 75.54%로 나타났다.

정밀도(Precision)는 모델에 의해 예측된 목표변수 값별로 예측된 목표변수 값과 실제 목표변수 값이 일치하는 비율을 나타내는데, 모델에 의해 목표변수 값이 출산이라고 예측된 관측치의 83.56%가 실제로 목표변수 값이 출산인 관측치였으며, 목표변수 값이 미출산이라고 예측된 관측치의 66.67%가 실제로 목표변수 값이 미출산인 관측치였던 것으로 나타났다.

재현율(Recall)은 실제 목표변수 값별로 모델에 의해 예측된 목표변수 값과 실제 목표변수 값이 일치하는 비율을 의미한다. 실제 목표변수 값이 출산인 관측치의 73.49%를 출산이라고 정확히 예측하며, 실제 목표변수 값이 미출산인 관측치의 78.57%를 미출산이라고 정확히 예측하는 것으로 나타났다.

표 3. 가구의 출산 여부 예측모델의 혼동행렬(confusion matrix)

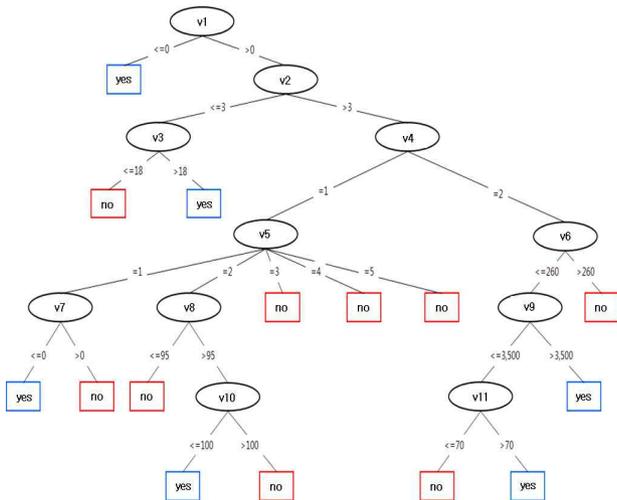
Table 3. Confusion Matrix

Measure	Childbirth	
	Yes	No
Hit ratio	75.54%	
Precision	83.56%	66.67%
Recall	73.49%	78.57%

<그림 1> 및 <표 4>는 의사결정나무 알고리즘에 의해 도출된 가구의 출산 여부 분류 패턴 중 나뭇잎 노드(leaf node)의 관측치가 5건 이상인 패턴들을 보여주고 있다.

이 연구에서는 총 7개의 패턴(미출산 관련 패턴 4개, 출산 관련 패턴 3개)이 도출되었으며, 이 패턴들은 총 11가지의 변수로 구성되어 있다. 가장 상위에 위치한 노드는 자녀를 비롯한 가구원 구성(‘6세 이상 18세 미만 자녀 유무’와 ‘가구원 수’)이며, 그 외 출산·양육 관련 지원금(월) 및 출산·양육 관련 지원금 수급 여부, 주거 특성(주

거지 형태), 경상소득, 식비, 의료비, 전·월세 보증금, 가구 구입비가 의사결정나무를 구성하는 중요 노드로 나타났다.



- v1: No. of children aged 6 to 18
- v2: No. of household members
- v3: childbirth and children allowance(amount)
- v4: childbirth and children allowance(y/n)
- v5: housing type
- v6: household income
- v7: rent fee
- v8: food expenditure
- v9: housing deposit
- v10: medical expenditure
- v11: furniture expenditure

그림 1. 도출된 의사결정나무
Fig. 1. Derived Decision Tree

표 4. 패턴 분석결과
Table 4. Derived Patterns (KRW., unit: 10,000)

No.	Pattern	Childbirth	Prob.
1	v1>0 & v2>3 & v4=1(Y) & v5=1(dwelling) & v7<=0	yes	80.00
2	v1<=0	yes	79.84
3	v1>0 & v2>3 & v4=1(Y) & v5=2(apartment) & v8 & v10<=100(KRW)	yes	71.43
4	v1>0 & v2>3 & v4=2(N) & v6>260(KRW)	no	98.04
5	v1>0 & v2>3 & v4=1(Y) & v5=2(apartment) & v8<=95(KRW)	no	92.31
6	v1>0 & v2>3 & v4=2(N) & v6<=260 & v9<=3,500(krw) & v11<=70(KRW)	no	90.91
7	v1>0 & v2<=3 & v3<=18(KRW)	no	77.14

출산에 대한 패턴은 다음과 같다. 첫째, 출산을 할 확률이 가장 큰 경우는 가구 내에 6세 이상 18세 미만의 자녀가 있으며, 전체 가구원 수가 3명을 초과하고, 출산·양육 관련 지원금을 지급하고 있으며, 단독주택에 거주하고, 월세금은 지출하지 않는 경우이며 출산 확률은 80.00%로 나타났다(패턴 1).

둘째, 학령기 또는 청소년기 자녀(6세 이상 18세 미만)가 없는 경우이며 출산 확률은 79.84%로 나타났다(패턴 2). 즉 가구의 특성만으로 설명하기 어려운 출산의 이면

에는 출산을 결정하는 가장 중요한 동기로써 부부의 자녀계획이 우선순위에 자리함을 설명하는 결과이다.

셋째, 학령기 또는 청소년기 자녀(6세 이상 18세 미만)가 1명 이상 있고, 가구원 수가 4명 이상, 출산·양육 관련 지원금을 지급하고, 아파트에 살고, 식비를 월 95만원을 초과하여 지출하고, 의료비는 월 100만원 이하로 지출하는 경우이며 출산 확률은 71.43%로 나타났다(패턴 3).

미출산에 대한 패턴을 살펴보면 다음과 같다. 첫째, 출산을 하지 않을 확률이 가장 큰 경우는 가구 내에 학령기 또는 청소년기 자녀(6세 이상 18세 미만)가 1명 이상 있고, 전체 가구원 수가 4명 이상이며, 출산·양육 관련 지원금을 지급하지 않고, 경상소득(근로소득, 사업소득, 재산소득, 이전소득)이 월 260만원을 초과하는 경우이며 미출산 확률은 98.04%로 나타났다(패턴 4).

둘째, 학령기 또는 청소년기 자녀(6세 이상 18세 미만)가 1명 이상 있고, 전체 가구원 수가 4명 이상이며, 출산·양육 관련 지원금을 지급하고, 아파트에 거주하며, 식비는 월 95만원 미만인 경우이며 미출산 확률은 92.31%로 나타났다(패턴 5).

셋째, 학령기 또는 청소년기 자녀(6세 이상 18세 미만)가 1명 이상 있고, 전체 가구원 수가 4명 이상이며, 출산·양육 관련 지원금을 지급하지 않고, 경상소득(근로소득, 사업소득, 재산소득, 이전소득)이 월 260만원 이하이고, 전·월세 보증금이 3,500만원 이하이며, 가구 및 가전제품 구입비가 연간 70만원 이하인 경우이며 미출산 확률은 90.91%로 나타났다(패턴 6).

넷째, 학령기 또는 청소년기 자녀(6세 이상 18세 미만)가 1명 이상 있고, 전체 가구원 수가 3명 이하, 출산·양육 관련 지원금을 월 18만원 이하로 지급하는 경우이며 미출산 확률은 77.14%로 나타났다(7번 패턴).

패턴 결과에 의하면 본인 소유의 주거지에서 거주하는 가구라면 필수 생활비를 우선적으로 고려하지만, 그렇지 않을 경우 월세금(패턴 1)이라는 고정지출비용의 부담이나 전·월세 보증금(패턴 6)과 같이 안정적인 주거를 위한 지출 수준이 중요한 역할을 하는 경향을 보인다. 이 중 후자에는 보증금을 부담하지 않는 자가(自家)이거나 보증금 수준이 3,500만원보다 낮은 주거에서 생활하는 가구가 혼합되어 있겠으나, 주거 면적이 좁거나 월세금 지출 수준이 높은 주거 상황으로 해석된다. 또한 경상소득이 비교적 낮은 가구에서는 주거에 투입되는 비용이 중요하게 고려됨을 유추할 수 있다.

한편 패턴 2를 제외한 패턴 결과를 살펴보면, 현재 자녀가 1명 이상 있는 상태에서 추가로 출산하는 가구는 월세금이라는 주거 관련 고정 지출이 없거나(패턴 1) 비교적 생활비(식비)가 높고 의료비 지출 수준이 낮아(패턴 3) 경제적 여건이 양호한 상황으로 유추할 수 있다. 반면 생활비(식비) 지출이 비교적 적고(패턴 5), 가구소득 수준이 낮은 등 경제적 여건이 넉넉하지 않은 경우(패턴 6)

에는 미출산으로 연결되고 있다. 또한 경상소득이 높은 가구에서도 미출산(패턴 4)의 결과로 이어져, 선행연구와 같이 소득수준이 높을수록 오히려 자녀에게 양질의 자원을 집중시키고자 가족규모를 늘리지 않는 경향을 엿보게 된다. 이처럼 개별 가구의 특성에는 다소 차이가 있겠으나, 가구의 소득·자산과 생활비 역시 가구의 출산과 연관된 요인임을 알 수 있다. 이 중 소비의 경우 필수생활비인 식비·의료비와 가구에서 으레 지출하게 되는 비용인 가구구입비가 도출되어, 가구의 지출 여건이 소득과 별개로 중요한 조건임을 나타낸다.

가족정책 중에서도 ‘출산·양육 관련 지원금’으로 묶이는 정책의 경우, 가구원 구성의 뒤를 이어 출산을 결정짓는 중요한 요인으로 확인되었다. 매달 수급하는 지원금 수준 역시 출산과 관련이 있는 것으로 나타났으며(패턴 7), 관측치가 적어 기술하지 않았으나 양육비 수준이 비교적 높은 경우(>18만원) 출산으로 이어지는 패턴도 확인할 수 있었다.

V. 결론

출산에 대하여 부부의 연령이나 건강을 포함하는 생물학적 요인, 출산에 대한 문화 또는 가치관 등이 중요한 비중을 차지한다는 점에는 이견이 없을 것이다. 때문에 주거를 비롯한 사회경제적 요인은 출산의 결정적인 조건이 되기 어려우나, 출산이 가능한 환경을 구성하고 가구의 출산의지를 실현하는 과정에서 중요한 배경을 구성한다. 이에 정부는 주택 정책, 양육비 지원 등 다양한 영역의 출산장려 정책을 수립하여 산적한 과제들을 해결하고자 노력하고 있다.

본 연구는 이 같은 사회적 문제의 인지와 정책적 필요에서 출발하여, 주거 상황을 비롯한 경제적 요인들이 우리나라 가구의 출산에 미치는 영향에 대해 살펴보았다. 본고는 가구의 주거 및 다양한 경제적 상태에 관하여 장기간 구축한 재정패널 데이터를 사용하였고, 기존 연구에서 많이 다루지지 않았던 접근법을 통하여 출산과 관련된 한국사회의 모습을 간접적으로 진단하고자 하였다.

이 연구의 결과를 정리하면 다음과 같다. 첫째, 변수 중요도 평가와 의사결정나무 모형의 분석 결과를 통하여 주거 및 경제적 요인의 중요성을 읽을 수 있다. 이러한 패턴들 간의 조합은, 출산이 단순히 개별 가구 수준의 의사결정이 아니라 전체적인 사회의 환경 및 제도와 밀접한 관련을 맺는다는 기존 연구들의 공통된 견해 [10],[19]를 뒷받침한다. 한 걸음 나아가 가구의 사회경제적 특성과 공공정책은 무자녀 가구의 출산 여부 자체를 결정하지는 않지만, 둘째 이상의 자녀의 출산 여부 또는 시기에 있어서는 중요한 조건으로 인식됨을 나타낸다

는 점에서도 여러 연구들의 결과와 일치한다.

둘째, 주거의 특성이 출산에 대해서 갖는 의미를 발견할 수 있다. 중요도 평가와 패턴 분석에서 주거와 관련된 요인들이 반복적으로 검증되었는데, 예상과 달리 주거지의 거주 형태(자가 및 비자가)는 유의한 요인으로 도출되지 않았다. 본 연구에서는 표면적으로 집을 소유했는가 하는 사실보다, 주거 매입을 위한 지출이 발생하였던 시기가 얼마나 지났으며 주거지의 가치가 어떠한 수준이었는가 하는 등, 자가 소유 여부를 구성하는 상황이 더욱 중요함을 보여주고 있다. 주거 마련이 녹록치 않은 한국의 현실을 고려할 때, 가구에서 집을 소유하더라도 실질적으로 그 효과를 얻을 수 없는 현실을 반영하는 결과로 읽힌다.

셋째, 기존 자녀의 연령대와 자녀의 수, 출산하고자 하는 자녀의 출생순위에 따라서는 출산여부와 관련 있는 변인들의 조합이 달라진다. 흥미로운 결과는, ‘6세 이상 18세 미만의 자녀 수’가 가장 중요한 변수로 확인된다는 점이다. 학령기 자녀일 경우 비학령기 자녀에 비해 교육비 등 지출이 급격히 늘어 경제적 부담이 가중된 상황이 주요 원인일 것으로 해석된다. 이는 첫째 자녀 이후의 출산이 소득계층 수준 간의 유의한 관계를 실증하였던 다수의 연구결과들과 동일한 맥락이다.

한편 가구의 생애주기 및 가족생활과 관련된 정책 변인의 결과는, 출산 또는 양육 관련한 정책의 수혜 경험이 있거나 가능성이 있고 그에 대한 지지도가 높을 경우 출산을 증진한다는 선행연구 결과를 뒷받침한다. 변수별 중요도를 평가한 결과에서는 해당 변수가 검증되지 않았으나, 대체로 자녀가 1명 이상 있는 가구의 경우 출산·양육 관련 지원금이 출산의 결정에 중요한 역할을 하는 것으로 해석된다. 다만 해당 변수에는 여러 정책들이 혼합되어 있어 그 효과를 면밀히 설명하기에는 제약이 있다.

상기의 결과들을 통하여 주거가 출산과 맞닿아 있으며, 출산·양육 지원 정책의 효과적인 수단으로 활용될 수 있음을 다시 한 번 확인하였다. 주거란 가구의 생활의 기반을 제공하는 장소이기도 하지만, 가구의 사회경제적 계층을 대변하는 동시에 이를 구성하는 요인들과 결합함으로써 출산에 영향을 미치고 있었다. 이 연구에서는 주거와 관련한 특성들을 구체적으로 다루었고, 가구·개인 특성과의 조합된 연관관계로 이루어진 패턴을 도출하였다는 점에서 선행연구의 공백을 메우고자 하였다. 또한 기존 연구들은 주로 개별 요인들이 출산에 미치는 전반적인 방향성 등에 대해 다루고 있으나, 이러한 방향성은 가구가 처해있는 상황에 따라 달라질 수 있다. 본 연구는 한 발짝 더 나아가 가구의 상황을 종합적으로 고려하고, 출산에 대하여 다면적 요인들의 상호작용에 의한 결과물임을 입체적으로 보여주었다.

주거와 출산·양육은 각기 그 특성 상 정책의 개입만으로 해결하기 어려운 속성을 갖는다는 점에서 정책의 효

과에는 일정수준 한계가 있다. 현재 다수의 정부부처와 지방자치단체에서 출산·양육을 포함하는 가족정책을 시행하고 있는데, 여러 정책수단들 간의 관계를 정리하여 보다 일관성 있는 정책의 흐름을 확보할 필요가 있을 것이다. 특히 지역의 특성에 따라 지자체별로 가족정책의 주된 대상이 되는 가구의 생애주기가 다르기 때문에 가구별로 출산·양육 관련 정책 수혜의 편차가 있을 것이다.

이상의 결과와 시사점에도 불구하고 연구의 제약과 한계점을 정리하면 다음과 같다. 무엇보다도 본 연구에 적합한 대상자를 추출하는 과정에서, 많은 표본을 소실하여 다차원적인 분석이 어려웠다는 아쉬움이 있다. 또한 개인의 태도 및 가치관 역시 중요하게 고려해야 할 요인이었으나, 분석데이터 상에서 대부분 최근의 조사 차수부터 수집된 자료이므로 사용하기 어려웠다는 점에서 한계가 있었다. 향후 그 밖의 한계점들도 보완하여 더욱 견고한 연구 성과를 도출하고자 한다.

※본 논문은 「2018 재정패널 학술대회」에서 발표한 연구를 수정·보완하였음

참고문헌

- [1] C. H. Mulder and M. Wagner, “The connections between family formation and first-time home ownership in the context of West Germany and the Netherlands”. *European Journal of Population*, Vol. 17, No. 2, pp. 137-164, 2001.
- [2] KOSTAT, *Housing Occupation Statistics by Administrative District*. Available: http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_10H0402&conn_path=I2 (accessed: 19. 12. 12.)
- [3] KAB, 2019. *National Housing Price Index*. Available: http://kosis.kr/statHtml/statHtml.do?orgId=408&tblId=DT_40803_N0001&conn_path=I2 (accessed: 19. 11. 13.)
- [4] S. L. Lee and J. H. Lee. *Association between Housing and Fertility among Newly Married Couples*. KIHASA. 2017.
- [5] R. Arundel, “Equity inequity: Housing wealth inequality, inter and intra-generational divergences, and the rise of private landlordism”. *Housing, Theory and Society*, Vol 34, No 2, pp. 176-200, 2017.
- [6] D. Ben-Shahar, and Warszawski, J. “Inequality in housing affordability: Measurement and estimation”. *Urban Studies*, Vol 53, No. 6, pp. 1178-1202, 2016.
- [7] M. Chudnovskaya, “Housing context and childbearing in Sweden: a cohort study”, *Housing Studies*, pp.1-20, 2018.
- [8] M. Y. Kim and J. Y. Hwang, “Housing Price and the Level and Timing of Fertility in Korea: An Empirical Analysis of 16 Cities and Provinces”, *Health and Social Welfare Review*, Vol. 36, No. 1, pp.118-142, March 2016.
- [9] J. B. Park and J. H. Lee, “Housing Price and Birth Rate under Economic Fluctuation: Evidence From 19 OECD Countries”, *Korean Journal of Child Care and Education Policy*, Vol. 10, No. 3, pp.51-69, October 2016.
- [10] H. S. Chon, Y. P. Kim, H. N. Jung, H. S. Kim, S. J. Ha, J. B. Kim, Y. J. Yoon, M. J. Oh, and T. H. Kim, *Housing and Urban Policy Strategies Responding to Low Fertility*. KRIHS. 2012.
- [11] P. Feijten and C. H. Mulder, “The timing of household events and housing events in the Netherlands: a longitudinal perspective”, *Housing Studies*, Vol. 17, No. 5, pp.773-792, 2002.
- [12] H. Kulu and A. Vikat, “Fertility differences by housing type: The effect of housing conditions or of selective moves?”, *Demographic Research*, Vol. 17, pp.775-802, 2007.
- [13] H. Kulu and E. Washbrook, “Residential context, migration and fertility in a modern urban society”. *Advances in Life Course Research*, Vol. 21, pp.168-182, 2014.
- [14] C. H. Mulder, “Home-ownership and family formation”. *Journal of Housing and the Built Environment*, Vol 21, No. 3, pp.281-298, 2006.
- [15] N. Mizutani. “The effects of housing wealth on fertility decisions: Evidence from Japan”. *Economics Bulletin*, Vol. 35, No. 4, pp. 2710-2724, 2015.
- [16] M. C. Kim, G. Y. Kim, H. S. Chon, M. A. Kang, Y. S. Lee, and S. S. Bae, *Housing Policy Reflecting Characteristics of Housing Consumption Regarding the Life-cycle Stage*. KRIHS. 2016.
- [17] S. S. Kim and H. Y. Baek, “The Effects of Household’s Economic Status on the Childbirth”, *Korea Social Policy Review*, Vol. 21, No. 3, pp.129-157, July 2014.
- [18] D. E. Park and G. S. Yoo, “The Delaying & Advancing Factors of Marriage and Fertility Timing Among Married Women”, *Korean Journal of Family Welfare*, Vol. 16, No. 2, pp.29-65. June 2011.
- [19] E. H. Jung and Y. S. Choi, “The Factors Associated with the Birth Plan for Second Child and Second Birth for Married Women in Korea”, *Health and Social Science*, Vol. 33, No. 1, pp.5-34, March 2013.
- [20] A. L. Booth and H. J. Kee, “Birth order matters: the effect of family size and birth order on educational attainment”. *Journal of Population Economics*, Vol. 22, No. 2, pp. 367-397, 2009.

- [21] J. H. Lee and J. Y. Hwang. “The Effectiveness of Childbirth Support Policy in 16 Metropolitan Governments”. *Health and Social Welfare Review*, Vol. 38, No. 3, pp.555-579, September 2018.
- [22] K. S. Eun, “Korean Married Women's Work Sequence and Fertility”, *Korea Journal of Population Studies*, Vol. 41, No. 1, pp.79-108, March 2018.
- [23] S. Y. Kong, “The Influence of Demographic and Socioeconomic Characteristics on Fertility of Married Women after the 1997 Economic Crisis”, *Health and Social Science*, Vol. 19, No. 1, pp119-149, April 2006.
- [24] E. J. Kim, “A Study on the Risk of Potential Economic Loss Due to Marriage”, *The Women's Studies*, Vol. 92 No. 1, pp. 7~30, Spring 2017.
- [25] L. Ferguson, “Gender, Work, and the Sexual Division of Labor”, in *The Oxford handbook of gender and politics*, Oxford University Press, ch.13, pp.340-360, 2013.
- [26] J. H. Yum, S. E. Chung, and Y. S. Joo, Young Sun, “Family Planning Experience and Service Needs among Families with Three Children”, *Health and Social Welfare Review*, Vol. 33, No. 1, Spring 2013.
- [27] J. Y. Seo and H. G. Kim, “Factors Influencing Married Women's Childbearing Willingness Based on Number of Children-Ever-Born”, *Korea Journal of Population Studies*, Vol. 38, No. 3, pp.1-24, September 2015.
- [28] H. J. Song and S. J. Woo, “A Macro Analysis of the Effect of Child-care Policy on Fertility and Female Labor Force Participation in Korea”, *The Journal of Korean Public Policy*, Vol. 17, No. 1, pp.3-36, March 2015.
- [29] J. Y. Ko, and K. P., Ko, *Issues and challenges in implementing the Korea child benefit*. KIHASA, 2017.
- [30] M. J. Berry and G. S. Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2004.
- [31] I. H. Witten and E. Frank, *Data mining: Practical machine learning tools and techniques*, San Francisco: Morgan Kaufmann Publishers. 2005.



이정화(Jeonghwa Lee)

2015년: 성균관대학교 대학원 (행정학 박사)

2016년~현 재: 근로복지공단 근로복지연구원

※관심분야: 사회보험, 복지국가, 여성학 등



김건우(Gunwoo Kim)

2010년: 고려대학교 대학원 (경영학 박사)

2011년~현 재: 한밭대학교 경영회계학과

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



최근호(Keunho Choi)

2013년: 고려대학교 대학원 (경영학 박사)

2018년~현 재: 한밭대학교 경영회계학과

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