

## 제21회 통계청 논문 공모전 수상작 발표

통계청(청장 이형일)은 『제21회 통계청 논문 공모전』 심사결과를 발표하였다. 2003년부터 매년 개최해 오고 있는 본 논문 공모전은 통계자료의 이용 저변을 확대하고 통계연구(분석) 결과를 공유하고 확산하는 것을 목적으로 하고 있다.

특히 올해에는 통계연구에 관심이 있는 국내외 대학·대학원에 재학 중인 학생들이 개인 또는 팀(2~3명)으로 경제, 사회, 데이터과학 등 다양한 분야에서 109편의 논문을 응모하였으며, 이 중에서 최종 54편의 논문이 제출되었다.

논문에 대한 심사기준은 연구주제의 창의성, 연구내용의 논리성과 일관성, 연구방법의 타당성, 연구자료의 활용성 등을 고려하였고, 관련 분야의 전문가 심사결과를 바탕으로 최종 심사위원회를 통해 총 7편(최우수 1편, 우수 2편, 장려 4편)의 수상작을 선정하였다.

심사결과, 최우수논문에는 ‘사회적 지지와 사회적 적응의 연관성: 다문화 청소년을 대상으로 한 종단적 잠재전이분석’(탁현삼·차준호, 한국외대 노어과 학사과정)이 선정되었다. 해당 연구는 다문화 청소년이 경험하는 사회적 지지와 사회적 적응 간의 연관성에 관하여 종단적 확장 모형을 통하여 국내 다문화 가정 청소년들의 사회적 지지와 사회적 적응의 연관성을 비교적 명확하게 설명하고 있으며, 이를 통해 다문화청소년이 성공적으로 정착하기 위해 필요한 지역사회와 정부의 인식과 대응방향을 제시하고 있다는 점에서 의의가 크다.

우수논문(2편)에는 ‘자동화 머신러닝(AutoML)과 SHAP 기법을 활용한 청소년의 다중시민성 유형 예측요인 연구’(정성경, 충남대 교육학과 박사과정) 및 ‘베이지안 가법 회귀 트리(BART) 및 심층 신경망을 기반으로 한 선제적 재활용률 분석 및 활용 방안’(김대영, 연세대 응용통계학과 학사과정)이 선정되었다.

이형일 통계청장은 제21회 통계청 공모전에 적극적으로 참가한 학생들에 대한 감사와 수상자에 대해 축하의 말씀을 전하면서, “인공지능과 데이터 과학의 시대를 맞이하여 데이터를 처리하고 활용하는 능력이 개인은 물론 기업과 국가의 경쟁력을 결정하고 있다. 이번 논문공모전이 미래의 데이터 전문가를 꿈꾸는 학생분들에게 그간의 데이터분석 실력을 선보이고 데이터가 가진 무궁한 활용가치를 탐색하는 도전의 장이 되었기를 기대한다”고 강조하였다.

시상은 9월 1일(금) 『제29회 통계의 날』 기념식에서 진행될 예정이다. 수상자(최우수상 1편, 우수상 2편, 장려상 4편)에게는 통계청장 표창과 함께 소정의 시상금이 지급되며, 수상논문은 『제13회 국가통계방법론 심포지엄』(9월 14일)에서 포스터로 발표될 예정이다.

### 【붙임】 2023년 『제21회 통계청 논문 공모전』 수상작

담당 부서	통계개발원 연구기획실	책임자	실 장	박상영 (042-366-7101)
		담당자	사무관	배준형 (042-366-7116)



등급	구분	논문제목	성명
최우수 (1)	학·석사 과정	사회적 지지와 사회적 적응의 연관성: 다문화 청소년을 대상으로 한 종단적 잠재전이분석	탁현삼 차준호
우수 (2)	박사 과정	자동화 머신러닝(AutoML)과 SHAP 기법을 활용한 청소년의 다중시민성 유형 예측요인 연구	정성경
	학·석사 과정	베이지안 가법 회귀 트리(BART) 및 심층 신경망을 기반으로 한 선제적 재활용률 분석 및 활용 방안 - 시차를 둔 영향요인 추출 및 예측 기반 정책 제안을 중심으로 -	김대영
장려 (4)	박사 과정	머신러닝을 활용한 임차인 주거불안 예측모형 개발 및 실증: 수도권 1인 가구를 중심으로	구한민
		국민연금 반환일시금 수급자 특성에 관한 실증연구	오종석
	학·석사 과정	다문화 청소년의 진로 결정성 변화 궤적 유형 분석 및 영향요인 검증	임예림
		인공지능을 활용한 질병확진자수 예측과 방역 정책 평가	김근한 서민경 안지혁

\* 수상논문은 통계개발원 홈페이지(sri.kostat.go.kr, 학술행사→통계청논문공모전)에서 열람 가능(10월 중)

# 수상작 요약문 소개

## [최우수상]

### 사회적 지지와 사회적 적응의 연관성: 다문화 청소년을 대상으로 한 종단적 잠재전이분석

탁현삼, 차준호 (한국외국어대학교 노어과, 학사과정)

최근까지 다문화 청소년이 경험하는 사회적 지지와 사회적 적응 간의 연관성에 관하여 종단적 확장 모형을 통한 실증연구는 잘 이루어지지 않았다. 이러한 시의성에 주목하여 기존의 변수중심적 접근법이 아닌 사람중심적 접근법인 잠재프로파일분석을 종단적 잠재전이분석으로 확장하여 다문화 초기 청소년들이 경험한 사회적 지지가 일정 기간이 지난 후 이들의 사회적 적응에 미치는 영향을 분석하였다. 사회적 지지를 크게 가족지지, 또래지지, 교사지지, 이웃지지로, 사회적 적응을 크게 소속감, 주관적 기대감, 삶의 만족도로 조작적 정의하였다. 연구 결과, 전기 청소년기(13세)에 높은 지지집단에 속한 다문화청소년은 높은 적응집단에 포함될 확률이 중간 지지집단, 낮은 지지집단에 비해 유의미하게 높았다. 또한, 공변인으로 처치한 전기 청소년기의 성별, 주관적 가정형편, 자아존중감, 우울은 잠재프로파일의 전이에 유의미한 영향을 끼쳤다.

특히 기존 연구와 다르게 성별에 따라 사회적 지지와 사회적 적응의 양상이 다르게 나타났다는 점에 주목하여, 분석결과를 토대로 시사점과 정책적 방향성을 제언하였다.

## [우수상]

### [1] 자동화 머신러닝(AutoML)과 SHAP기법을 활용한 청소년의 다중시민성 유형 예측요인 연구

정성경 (충남대학교 교육학과, 박사과정)

본 연구는 지구촌 시대에서 한국 청소년의 다중시민성 유형과 예측변인을 탐색하는 것을 목적으로 하며 「한국교육종단연구 2013」 데이터를 활용하였으며, 다중시민성 유형 탐색을 위해 잠재프로파일 분석을 활용하였다. 그리고 예측요인 탐색을 위해 자동화 머신러닝 알고리즘을 통해 가장 우수한 성능을 보이는 모델을 탐색하여 채택하였다. 머신러닝 결과의 설명가능한 해석을 위해 SHAP 기법을 활용하였으며 주요 분석결과는 다음과 같다. 첫째, 잠재프로파일 분석에서 청소년의 다중시민성은 ‘높은 다중시민성’, ‘중간 다중시민성’, ‘낮은 다중시민성’으로 분류되었다. 둘째, 자동화 머신러닝을 실시한 결과, 머신러닝 기법 중에서 Gradient boosting이 가장 높은 성능을 보였다. 머신러닝을 통해 청소년의 다중시민성에 대한 주요 변인들을 예측한 결과, 의사소통역량이 가장 중요한 변인으로 나타났으며, 지식정보처리역량, 문화예술활동 및 봉사활동 인식, 독립적 진로성숙도, 문화다양성 존중과 국제이슈를 다루는 교육경험, 성별, 교사와 또래 관계 등이 중요 변인으로 나타났다. 이러한 결과를 토대로 청소년을 위한 실효성 있는 시민교육을 위한 정책 제언과 시사점을 도출하였다.

## **[2] 베이지안 가법 회귀 트리(BART) 및 심층 신경망을 기반으로 한 선제적 재활용률 분석 및 활용 방안 - 시차를 둔 영향요인 추출 및 예측 기반 정책 제안을 중심으로 -**

**김대영 (연세대학교 응용통계학과, 학사과정)**

본 연구는 현 사회가 코로나-19 펜데믹을 거치며 직면하게 된 폐기물 차원상 탄소 중립 저해 현상에 집중하여, 이를 완화하기 위한 선제적 방안을 모색하는 것을 목적으로 설정하였다. 이에 환경부의 전국 폐기물 통계, 통계청 e-지방지표 통계 등을 기반으로 BART 모델과 DNN을 학습시킨 후, 각 모델을 전제로 특정 지역에서 발생할 재활용률을 약 1년 시차를 두고 예측하는 데 기여하는 주요 변수 및 변수별 주변 효과와 한계기여도를 추출, 이와 함께 재활용률 증진을 위한 선제적 개입 시 각 모델의 활용 방안을 제시하였다. 연구 결과 두 모형 모두에서 약 2%p의 오차 아래 유의한 재활용률 예측을 전개할 수 있음이 확인되었으며, 이때 공통적으로 자치구별 인당 공업지역 면적, 위탁급식영업체 수, 3인 이상 세대 수가 향후 재활용률 증감에 유의한 영향을 미치는 것으로 나타났다. 본 연구는 선제적 재활용률 예측을 위한 실질적 모델과 함께 시차를 둔 영향요인을 추출한다는 점에서 의의를 갖는다.

## **[장려상]**

### **[1] 머신러닝을 활용한 임차인 주거불안 예측모형 개발 및 실증: 수도권 1인 가구를 중심으로**

**구한민 (연세대학교 도시공학과, 석박사통합과정)**

본 연구는 기계학습법을 응용하여 수도권 1인가구 임차인의 주거불안정 예측모형을 개발하고 검증하였다. 본 연구의 주요 결과에 대한 요약과 결론은 다음과 같다. 첫째, 수도권 1인가구 임차인의 약 38%가 광범위한 의미에서의 주거불안정을 경험하고 있음을 확인하였다. 최근 수도권 1인가구의 변화 추이를 고려하였을 때, 이들은 머지않아 주거복지정책의 주요한 정책대상자들이 될 가능성이 높으므로 선제적인 대응책의 수립이 요구된다. 둘째, 앙상블 기법을 활용하여 어떠한 특성이 주거불안정 경험에 영향을 미치는지 식별하였다. 그 결과 주거불안정이란 주거환경에 대한 통제권을 쥐지 못하는 광범위한 상황으로 정의할 수 있으므로 정부·지자체는 주거복지정책을 수립하는 경우 주거불안정의 경험에 대한 지원의 범위를 광범위하게 확대할 필요가 있다. 셋째, 분류트리, 배깅, 랜덤 포레스트, 부스팅 등 네 가지의 분류모델 알고리즘을 활용하여 수도권 1인가구 임차인의 주거불안정 예측모형을 개발한 결과, 랜덤 포레스트 모델이 가장 우수한 성능을 가지고 있는 것으로 나타났다. 이를 향후 주거불안정을 예측하는 데에 활용하면 수도권 1인가구 임차인의 주거사각지대 해소에 기여할 수 있을 것으로 판단된다.

## [2] 국민연금 반환일시금 수급자 특성에 관한 실증연구

### 오종석 (서울대학교 경제학부, 박사과정)

현행 국민연금 제도 하에서는 그 취지가 무색하게, 매년 노령연금 신규 수급자의 절반보다 많은 18만 명 이상의 ‘반환일시금 수급자(국민연금 해약자)’가 발생하고 있다. 본 연구는 가입자 전수자료를 이용해 반환일시금 수급자들의 특성을 최초로 비교·분석하였다. 분석결과, 실가입개월수(기존에 국민연금을 납부한 기간)가 짧을수록 국민연금을 쉽게 포기하고 반환일시금을 수급받을 확률이 높아지는 것으로 나타났다. 한편 평균납입금액은 실가입개월수만큼 효과가 강력하지는 않았다. 또 국민연금의 가입기간 및 총등록기간이 길수록, 즉 국민연금 납입 의무가 있는 직장을 오래 다닐수록 반환일시금 수급을 선택할 확률이 감소하는 것으로 나타났다. 실가입기간 및 가입기간의 측면에서는, 국민연금을 충실히 납부하였거나 의무적으로 자동납부되는 직장을 다닌 사람들의 경우 반환일시금을 잘 선택하지 않게 된다는 사실을 발견하였다. 종합하면 국민연금을 보다 필요로 할, 상대적으로 열악한 노동조건 하에 놓여 있는 사람들이 오히려 소외되고 있는 것으로 판단된다. 이들의 국민연금 이탈을 방지하기 위해서는 대출지원 등이 조치가 필요하다고 사료된다.

## [3] 다문화 청소년의 진로 결정성 변화 궤적 유형 분석 및 영향요인 검증

### 임예림 (고려대학교 교육학과, 석사과정)

본 연구는 다문화 청소년 고등학생의 진로 결정성 변화 궤적에 따른 잠재계층 유형을 분류하고, 분류에 영향을 미치는 요인을 개인 특성, 가정, 학교 관점에서 종합적으로 검증하였다. 이를 위해 다문화 청소년 패널조사 자료를 활용하여 성장혼합모형 분석을 실시했다. 분석 결과, 진로 결정성 변화 궤적의 유형은 저수준 증가(6.1%), 중수준 증가(65.8%), 고수준 유지(28.0%)의 3개 계층으로 분류되었다. 다음으로 잠재계층 분류에 대한 영향요인의 효과를 검증한 결과, 개인 특성 요인으로는 성취동기 및 이중문화 수용태도, 가정 요인으로는 부모와의 대화 빈도, 학교 요인으로는 학업적응과 학교 진로교육 및 활동 만족도가 높을수록 진로 결정성 중수준 증가/고수준 유지 집단에 속할 확률이 높은 것으로 나타났다. 본 연구는 다문화 청소년의 진로 결정성이 시간에 따라 변화하며 같은 발달 시기에 있는 개인이라도 이질적인 진로 발달 양상을 보일 수 있음을 확인하고, 개인별 진로 발달 수준에 맞는 세분화된 진로 교육이 필요함을 도출했다. 나아가 영향요인 검증 결과에 기반하여 다문화 청소년의 진로 발달을 위해 가정, 학교, 기관에서 참고할 수 있는 교육적 함의를 제시하였다.

#### [4] 인공지능을 활용한 질병확진자수 예측과 방역 정책 평가

**김근한, 서민경, 안지혁 (계명대학교 기초과학부(수학전공), 학사과정)**

본 연구는 코로나19에 대응하기 위한 방역 정책의 효과를 분석하고 예측하기 위해 다양한 모델을 활용한다. 전염병 확산의 수학적 모델에 추가로 인공 신경망 모델을 도입하여 방역 정책에 따른 확진자 수 변화를 예측한다. 순환 신경망 모델과 물리 정보 기반 신경망을 활용하며, 방역 단계 변경 시점과 강도에 따른 전파율, 회복률, 사망률을 예측한다. 또한, 코로나19 변이 바이러스에 대한 예측과 해당 기간의 방역 정책 평가도 수행한다. 국내의 코로나19 발생 현황 데이터와 인구이동량 등의 다양한 변수를 활용하여 분석하며, 실제 전파 상황에 대한 예측과 정책 평가를 통해 코로나19 대응 방역 정책 수립에 기여하고자 한다.