

다문화 청소년의 비행 경험에 대한 예측 변수 탐색 모형 비교: 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 모형을 중심으로*

유재효, 조영일†

동국대학교 경찰행정학과

본 연구는 다문화 청소년의 비행에 영향을 미치는 요인을 탐색하고 적합한 분석 모형을 결정하기 위하여 모형 비교를 통한 모형 선택 과정을 제시하였다. 한국 다문화 청소년 패널 자료를 활용하여 부모감독, 비행친구, 이중문화수용태도가 비행에 미치는 영향을 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 회귀분석을 통해 비교하였다. 분석 결과, 비행 변수는 정규성 및 등분산성 가정을 충족하지 않았으며, 영(zero)이 과도하게 많고 과산포(overdispersion) 현상이 확인되었다. 모형 간 정보지수(AIC, BIC)와 LRT, Wald 검정 결과, 영과잉 음이항 회귀 모형이 가장 적합한 것으로 나타났다. 특히, 과산포를 고려한 영과잉 음이항 모형이 예측력 및 적합도에서 우수성을 보였다. 본 연구는 비행 변수가 평균과 분산이 다르고 영의 비율이 높은 경우, 영과잉 음이항 회귀모형의 적용이 타당함을 시사한다.

주요어: 모형 비교, 비행, 가산자료, 과산포, 포아송, 영과잉 음이항

* 이 논문은 2026년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A3A2A02089039).

† 교신저자(Corresponding Author) : 조영일, 동국대학교(서울) 경찰행정학과 교수, 서울특별시 중구 필동로 1길 30, E-mail : yicho@dongguk.edu

유재효, 동국대학교 경찰행정학과 법심리학전공 석사과정생(제1저자)

■ 최초투고일 : 2025년 8월 7일 ■ 심사마감일 : 2026년 1월 2일 ■ 게재확정일 : 2026년 1월 15일

1. 서론

범죄와 같이 발생 가능성이 작은 경우, 빈도 분포에서 '0'이 과대하게 포함되는 경우가 많이 발생한다. 이렇게 '0'이 과대하게 포함된 경우, 연속 자료의 분석에 자주 사용되는 선형회귀분석을 사용하게 되면, 모수 등의 추정에 있어 편향이 발생할 수 있다(이휘인, 2016; Coxe, West, & Aiken, 2009). 선형회귀분석에서는 예측 변수에 대한 가정이 없기 때문에 가산자료를 포함한 다양한 척도의 수준을 가진 변수가 사용 가능하다. 하지만 선형회귀분석에서 투입되는 준거 변수에 관해서는 다양한 가정이 존재한다. 선형회귀분석의 가정이 위반되는 경우에는 회귀분석의 결과를 신뢰할 수 없는 문제가 발생한다. 구체적으로, 선형회귀분석의 준거 변수에 대한 정규성 가정은 오차항이 평균이 0이고, 분산이 일정한 정규분포를 따른다는 것이다. 특히 오차항이 정규분포를 따르지 않으면 선형회귀분석에서 실시한 t 검정과 F 검정의 결과를 신뢰할 수 없다(김수영, 2019). 오차항의 정규성이 위배되면 추정오차가 커지고, 이에 따라 t 값은 작아지며 p 값은 증가하게 된다. 결과적으로, 오차가 정규 분포를 따르지 않을 경우, 회귀계수에 대한 t 통계량과 F 통계량이 t 분포나 F 분포를 따르지 않을 수 있다(Grajales, 2013). 이는 결국 회귀계수에 대한 유의성 검정의 해석이 왜곡되어 가설검증의 결과를 신뢰할 수 없게 된다(Grajales, 2013).

본 연구는 다문화 청소년의 비행 변수를 사례로 한다. 특히 범죄학 연구에서 빈번히 활용되는 가산형 비행 변수를 분석할 때 자료의 분포적 특성을 충분히 고려하지 않을 경우, 발생할 수 있는 방법론적 문제를 검토하는 데 목적이 있다. 이에서론에서는 비행 변수를 중심으로 기존 연구의 분석 관행과 그 한계를 살펴보고, 이에 대한 대안

적 접근으로 모형을 탐색하려 한다. 범죄학에서 주로 다루는 청소년의 '비행 횟수'는 가산자료로서, 희귀한 사건의 발생을 측정하는 변수로 정적으로 편포되어 있다. 또한, 사건의 발생 횟수를 측정한 것이기 때문에 연속형 변수가 아니며 음수값이 존재할 수 없다. 이러한 가산자료에 선형회귀분석을 적용할 경우 오차항의 정규성 가정을 만족시키기 어렵다. 정규성 가정의 위반은 추정량의 효율성을 저하시킬 수 있으며, 궁극적으로 가설검정 결과의 신뢰도를 떨어뜨린다(Grajales, 2013).

2021년부터 2025년까지 출판된 국내 다문화청소년 비행 관련 논문 21편을 검토한 결과, 5편을 제외한 대부분의 연구가 비행변수를 12가지(또는 10가지) 비행행동에 대한 '예/아니오' 응답을 합산한 가산변수로 코딩하여 사용하고 있었다. 그러나 이러한 변수의 특성에도 불구하고, 분석방법은 주로 선형회귀분석 혹은 그 확장형을 중심으로 이루어졌다. 구체적으로, 전체 21편 중 9편이 매개효과분석, 조절된 매개효과분석, 자기회귀교차지연모형 등 선형회귀모형의 확장방식을 적용하였고, 나머지 12편 중 3편은 잠재성장모형을 사용하였다. 또한 3편은 로지스틱 회귀분석, 1편은 그 확장형인 생존분석을 사용하였다. 그 외에 구조방정식, 랜덤포레스트, 일반화추정방정식, 문헌리뷰, 비연속생존분석을 사용한 연구가 각각 1편씩 존재하였다. 이러한 결과를 볼 때, 국내에서 비행변수의 특성을 고려하여 진행한 선행연구는 많이 존재하지 않으며, 잠재성장모형과 구조방정식도 데이터의 포아송 가정과 영과잉 구조를 간과하였다. 또한 로지스틱 회귀분석의 경우 가산변수를 이분형으로 간주하여 데이터가 가지는 정보를 온전히 활용하지 못하는 한계가 존재한다. 그러므로 다문화 청소년의 비행 결과에 대한 정확한 검증을 위해서는 비행 변수의 분포적 특성을 고려해야 한다. 앞에서 논의된 연구들에서 관찰된 문제점을

해결하기 위해서는 분석하려 하는 자료의 특성을 확인하고 이에 맞는 분석 방식을 선택해야 한다. 자료의 특성을 잘 반영하는 모형을 적용하는 것은 결과의 신뢰성과 유용성 측면에서 매우 중요하기 때문이다(나우열, 이현숙, 2018).

선형회귀분석에서 사용하는 준거 변수의 또 다른 형태는 연속형 자료이다. 그러나 범죄학을 포함한 다양한 사회과학 분야에서 다루는 청소년의 비행은 대부분 가산 변수의 형태로 조사자료에 포함된다. 일정 기간 동안 비행 경험을 묻는 비행 변수의 분포는 '비행을 전혀 하지 않는다'라고 응답하는 집단이 상대적으로 가장 많다. 이로 인하여 비행 횟수를 물어보는 질문에 대한 응답은 '0'이 가장 많은 값을 가지며 정적으로 편포된 형태를 가진다. 그러므로 '0' 값이 많고 가산자료 형태인 비행 변수를 단순히 선형회귀분석을 사용하면 예측 변수를 통해 비행 발생을 제대로 예측하지 못할 수 있다.

한국의 사회과학 분야의 연구에서는 주로 비행 변수를 합산하여 연속형으로 간주하고 중다회귀 분석을 사용하거나(신재현, 2014; 이원기, 유효은, 이창환, 2019) 혹은 비행집단과 그렇지 않은 집단을 나누어 이항 로지스틱 회귀분석을 사용한다(박준호, 성용은, 2021; 박지수, 김하영, 한운선, 2021; 서봉언, 2019; 한용우, 이창배, 2025). 특히 다문화 청소년 비행에 관하여 포아송 분포(Poisson distribution)와 음이항 분포(Negative binomial distribution)를 적용하여 해석한 연구는 드물다. 나우열과 이현숙(2018)의 연구에서는 영과잉 가산 자료의 분포 특성을 고려하여 포아송 및 음이항 성장모형 및 잠재계층 성장모형을 비교하여 청소년의 가출 빈도의 중단적 변화를 분석하였다. 이길재와 이정미(2014)의 연구에서는 학교 폭력 발생 건수를 준거 변수로 하여 학교폭력의 발생 여부 및 발생 빈도에 영향을 미치는 요인을 허들

모형(Hurdle Model)과 '0' 절단 포아송 회귀모형(Zero Truncated Poisson Model)을 적용하여 분석을 진행하였다. 이휘인(2016)의 연구에서는 무단결석 빈도에 영향을 주는 요인들을 영과잉 포아송((Zero-inflated poisson), 영과잉 음이항(Zero-inflated negative binomial), 허들 포아송(Hurdle poisson), 허들 음이항(Hurdle negative binomial) 분포를 사용하여 분석하였다. 그러나 기존의 연구들은 대체로 중단적 변화를 설명하거나 특정 사건의 발생 요인을 내용적으로 분석하는 것에 초점을 두었기 때문에 횡단 가산 자료에 적합한 분석 모형을 선택하는 연구는 제한적이다. 나우열과 이현숙(2018)의 연구에서는 변화양상을 포착하기 위한 성장모형을 사용하였으나 횡단 자료에 대한 분석에는 적용이 어렵다. 장미희와 박창기(2012)의 연구는 노인 학대의 위험요인을 로지스틱, 음이항, 영과잉 음이항 모형을 통해 분석하였으나 각 모형 간 적합성 비교 및 선택의 근거에 대한 논의는 부족하였다. Gardner, Mulvey, Shaw (1995)의 연구에서는 포아송, 과산포 포아송, 음이항 모형의 차이를 논의하였지만, 영과잉 자료를 고려한 모형의 적용 가능성은 다루지 않았다.

이에 본 연구는 이러한 선행연구의 한계를 보완하고자, 선형회귀, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 모형을 비교 분석하여 본 데이터의 특성에 가장 적합한 모형을 경험적으로 검증하고자 한다. 본 연구는 다양한 가산자료 분석 모형을 체계적으로 비교함으로써 범죄학 및 사회과학의 연구에서 자료의 분포 특성을 반영한 모형 선택의 실증적 근거를 제시한다는 점에서 의의가 있다.

교육부에 따르면 2022년 다문화학생의 학업중단율은 고등학교 2.05%로 전체 학생의 학업중단율보다 소폭 높은 경향을 보인다(심예은, 2024). 2020년에 다문화 청소년의 학업중단 사유 중 약

34.9%가 학교 부적응을 그 이유로 들었다(안병수, 박유민, 2022). 이러한 통계 결과는 한국에서 다문화 청소년은 소수 집단이며 향후 사회에서 부적응 가능성을 내포한다. 이는 주목해야 할 사회문제이다. 학교에서 적응을 하지 못하는 다문화 청소년은 학교 밖 청소년이 되고 학교 밖 청소년은 특히 범죄와 비행에 취약하다. 윤정과 곽대경의 연구(2024)에서는 학교 밖 청소년의 비행 피해 경험이 사회적 차별 경험을 매개하여 공격적이고 비행적인 행동으로 이어질 가능성을 제시하였다. 이를 바탕으로 볼 때, 사회적 소수자이자 차별받을 가능성이 크고 학교 밖 청소년이 될 위험성이 큰 다문화 청소년은 비행에 취약하다고 볼 수 있다. 그러므로 다문화 청소년의 비행은 지속적으로 관심을 가져야 한다. 이에 본 연구는 다문화 청소년을 대상으로 비행 행태에 대해 탐구해보려 한다. 본 연구에서는 한국 청소년 비행 연구에서 가장 많이 사용되는 『다문화청소년 패널조사2기 데이터(MAPS 2)』를 사용하였다. 본 자료는 한국 청소년 정책 연구원에서 발간하는 패널 자료로써 아동 청소년 패널과 동일한 문항을 사용하여 청소년의 비행 행동을 측정한다. 그러므로 다수의 한국 청소년 비행 연구는 본 문항을 사용하여 연구가 진행되었다. 따라서 본 연구가 다양한 연구에서 적용될 수 있게 하고자 다문화 청소년 패널을 사용하였다.

1) 다문화 청소년의 비행

본 연구에서 다문화 청소년은 교육부 다문화 학생의 정의에 근거하는 대상자로서 국제결혼 가정자녀, 중도입국청소년, 외국인자녀를 포괄하여 정의한다(한국청소년정책연구원, 2024). 다문화 청소년의 비행이 단일한 문화만을 경험하는 일반 청소년의 비행과 동일한 양상을 보이는데 대해

서는 많은 논쟁이 있어 왔다. 박성훈과 전영실의 연구(2017)에서는 지위 비행과 심각한 비행 모두에게서 친구의 비행 성향은 두 집단에서 공통적으로 유의한 예측 변수였으나, 다문화 청소년의 경우 부모 감독이 추가적으로 유의미한 예측 변수로 확인되었다. 선행 연구에 따라 다문화 청소년이라도 일반 청소년과 마찬가지로 가족 유대, 친구 관계, 긴장 경험 등 일반적인 요인에 따라서 비행의 수준이 달라질 수 있다(박성훈, 전영실, 2017). 손진영과 배미경의 연구(2015)의 연구에서는 다문화 가정 청소년과 일반 청소년의 비행 경험 수준에 미치는 요인들은 비슷하다고 제안하였다. 두 개의 집단이 청소년기의 특성과 발달 과정에서 겪는 어려움에는 별다른 차이가 없기 때문이라고 주장하였다(손진영, 배미경, 2015).

다양한 선행 연구들을 토대로 다문화 청소년의 비행은 일반 청소년의 비행과 유사한 양상을 보인다는 것을 확인할 수 있으며, 다문화 청소년도 일반 청소년들과 같이 학교, 친구, 가족과 관련된 변인이 비행에 중요한 영향을 가짐을 추론할 수 있다. 요약하자면, 다문화 청소년의 비행에 관한 선행 연구에서 상대적으로 주요하게 강조되는 요인은 친구요인과 부모감독 요인, 그리고 다문화 수용성 요인이었다.

(1) 부모감독과 비행

부모감독은 양육 행동에서 중요한 하위 기능으로 언급되며 자녀가 어디에서 누구와 무엇을 얼마나 하고 있는지에 관해 얼마나 알고 있는지를 의미한다(이상균, 2008; Hirschi, 1969). 부모감독은 Reckless의 견제이론(Containment Theory)을 바탕으로 비행에 영향을 미치는 주요 변수로 추정된다.

견제이론에 따르면(Reckless, 1961, 1967) 청소년은 사회 규범에 벗어난 행위를 저지를 가능성이

존재하는 잠재적 비행자라고 가정한다(Thompson & Dodder, 1983). 그는 비행을 억제하는 요인으로 ‘외적 견제’와 ‘내적 견제’를 구분하여 제시하였다. ‘내적 견제’는 자기 통제를 통해 기대되는 규범을 따르는 개인의 능력을 의미한다(Thompson & Dodder, 1983). 반면 ‘외적 견제’는 사회가 개인의 행동을 허용 가능한 범위 내에 가두는 능력을 의미하며, 규칙의 내면화, 의미 있는 역할의 제공, 집단의 강화 3가지 요소로 구성된다(Reckless, 1967; Thompson & Dodder, 1983). ‘외적 견제’는 우리 사회가 제공하는 가족과 지역사회 등의 핵집단(nuclear group)이 중심이 된다(이윤호, 이승욱, 2024). 부모 감독의 변수는 ‘외적 견제’로서 청소년이 비행하지 않도록 억제하는 사회적 요인이라고 판단된다. 이에 청소년 비행에 영향을 미치는 주요 원인이라고 판단하여 예측 변수로 설정하였다.

(2) 비행친구와 비행

Sutherland의 차별적 접촉이론(Differential Association theory)은 범죄행위가 친밀한 집단 내 다른 사람과의 상호작용에서의 학습을 통해 이루어진다고 제시한다(Sutherland, Cressey, & Luckenbill, 1992). 이 이론은 집단적이고 모방적인 면이 많다는 청소년 비행의 특성이나 또래집단을 중시하고 환경에 민감한 이들의 특성을 반영하여 비행을 설명하고 있다(이윤호, 이승욱, 2024).

Sutherland는 범죄 행동이 학습된 행동이라고 강조하면서 비행을 하는 사람들과 접촉한 사람들은 그 특성을 습득하게 될 가능성이 높다고 주장했다(Warr, 1993). 실증적 연구에서도 청소년 비행에 비행친구와의 접촉이 중요한 예측 변수임을 많은 연구에서 보고하였다(노성호, 2006; 이상균, 2008; 이성식, 2007; Agnew, 1991; Brendgen, Vitaro, & Bukowski, 2000). 이를 근거로 청소년에게 비행친구의 존재는 비행을 예측하는 아주

강력한 요인이 된다. 이에 본 연구에서는 다문화 청소년의 비행을 예측하기 위한 예측 변수로 비행친구를 설정하였다.

(3) 이중문화수용태도와 비행

이중문화수용태도는 다문화 청소년만이 경험할 수 있는 속성이며 다문화 청소년의 이중문화 수용태도가 비행에 영향을 미치는지를 확인하기 위하여 예측 변수로 설정하였다. 일반 청소년과 달리 다문화 청소년은 이중문화에 대한 충돌 및 차이로 스트레스가 유발된다(박순희, 2009). 그러므로 다문화 청소년이 이중문화를 적절히 수용하여 균형을 이루며 학교에 잘 적응할 수 있도록 해야 한다(최홍숙, 2024).

조영선과 조영일(2019)의 연구에서는 이중문화수용태도 수준이 높을수록 지위·폭력·재산 비행을 경험하지 않을 확률이 증가한다는 결과를 제시하였다. 또한 Hirschi의 사회유대 이론(Social bond theory)에 따르면 관습적인 사회에 대한 유대와 결속이 약화하는 만큼 일탈할 자유는 강화됨을 강조한다(이윤호, 이승욱, 2024). 즉 사회에 대한 유대와 결속이 강할수록 반대로 비행 행동은 억제된다. 그러므로 이중문화수용태도가 높을수록 학교 적응을 잘하게 되며, 높은 수준의 학교 적응은 사회유대와 결속을 증가시켜 비행을 억제하는 통제 요인이 된다. 선순환은 궁극적으로 다문화 청소년의 비행 예방에 도움이 되기 때문에 비행의 예측 변수로서 이중문화수용태도를 설정하였다.

2) 가산자료에서의 분석 방법

(1) 가산자료

가산자료는 연속형 데이터와는 다르게 ‘0’ 이하의 수를 가질 수 없고 자연수만 존재한다. 사건의

발생 수, 물건의 수, 시행 횟수 등의 자료들은 그 예시로 들 수 있다. 그러므로 선형회귀분석에서 주로 사용하는 연속형 자료와는 다른 방식으로 접근해야 한다. 일반적으로 가산자료에서도 평균이 10보다 큰 경우에는 큰 어려움 없이 선형회귀분석으로 접근할 수 있다(Gardner, Mulvey, & Shaw, 1995). 그러나 사건의 발생 확률이 매우 희박할 때에는 선형회귀분석은 표준오차와 유의성 검정에서 편향된 결과를 초래할 수 있어 적절하지 않은 분석 방법이 될 수 있다(Gardner et al., 1995). 본 연구에서 사용하는 비행 발생 횟수는 가산자료의 형태이면서도 발생 확률이 낮으므로 평균이 낮게 나타난다. 비행 횟수처럼 희박한 사건 발생 자료일 경우 선형회귀분석의 기본 가정을 충족하지 못할 수 있다. 기본 가정을 충족하지 못한 채로 선형회귀분석을 적용하면 가설검정에서 편향된 결과가 발생할 수 있다. 이러한 경우, 선형회귀분석이 아닌 다른 회귀분석 방식을 적용해야 한다. 가산자료 분석에서 주로 사용되는 자료 분포에 대한 가정은 포아송 분포가 있다(나우열, 이현숙, 2018).

(2) 포아송 분포

포아송 분포는 이항분포의 특정한 분포이다. 동일한 확률 p 를 가지는 독립적인 베르누이 시행들이 반복될 때 M 번의 시행에서 성공한 총 횟수 P 는 이항 분포를 따른다. 이항 분포에서는 M 이 매우 크고 p 가 매우 작으며 MP 가 일정할 때 포아송 분포를 따른다고 가정한다(Freund, 1971, pp. 81-93). M 과 p 의 곱은 일반적으로 λ (람다)로 부른다(Berk & MacDonald, 2008). 정규분포에서는 평균(μ)과 표준편차(σ) 2개의 파라미터를 가지는 것과 달리 λ_i 는 사건의 기댓값을 보여주는 단일 파라미터이며 i 는 각 사건을 나열하여 명명한 것이다. 각 사건의 λ_i 의 값은 다를 수 있으며,

포아송 분포에 따라 각 사건은 독립적이며, 랜덤하며, '0'과 '1'과 같은 이진 변수로 구성된다(Berk & MacDonald, 2008). 이에 따라 포아송 회귀분석(Poisson regression)은 분산과 평균이 동일하며 사건 간 독립성과 일정한 발생 확률을 가정한다. i 번째 개인에 대한 사건 발생 빈도가 k 일 확률에 대한 포아송 분포는 다음과 같이 정의될 수 있다(Liu, 2007). <수식 1>은 포아송 분포의 확률질량함수이다(Coxe, West, & Aiken, 2009).

$$P(Y_i = k) = \frac{\exp(-\lambda_i)\lambda_i^k}{k!} \quad \langle \text{수식 1} \rangle$$

이 식에서 Y_i 는 특정 사건에 대한 발생 빈도이며, '0' 이상의 정수 값을 가진다. λ_i 는 사건의 평균 발생횟수(산술평균)을 나타낸다(Coxe et al., 2009). 선형 회귀 분석을 적용하면 발생하는 주요 문제들인 오차의 분산이 일정하지 않고, 오차 분포가 정규분포를 따르지 않는 문제를 해결하기 위해 포아송 오차 구조를 가지는 GLM을 사용한다(Coxe et al., 2009). λ_i 의 체계적 변화를 설명하기 위해서 로그 선형모형을 활용하여 그 값을 로그로 변환하여 설명한다. X_i^T 는 사건 i 에 대한 예측 변수들의 선형결합이다. β 는 예측 변수들의 선형 결합에 대한 회귀계수 벡터를 의미한다. <수식 2>는 $\lambda_i = e^{X_i^T \beta}$ 로 표현 가능하며, 포아송 회귀계수를 해석할 때 지수화하는 이유이다. 이러한 이유로 선형회귀는 덧셈적으로 영향을 미치지만 포아송 회귀에서는 곱셈적 영향을 준다(Coxe et al., 2009). <수식 2>에는 포아송 회귀모형의 선형 예측식을 나타낸다. 이 수식은 일반화 선형모형(GLM)에서 자연로그를 링크함수로 사용하는 구조이며 다음과 같이 정의한다(Coxe et al., 2009).

$$\ln \lambda_i = X_i^T \beta \quad \langle \text{수식 2} \rangle$$

포아송 분포는 단일 파라미터 λ 를 통해 평균과 분산을 동시에 설명하는 분포로 이 값이 작을수록 '0'의 분포가 크며 긴 꼬리를 가지는 형태가 되며 값이 클수록 평균 근처로 몰리는 경향이 있다. 포아송 분포는 꼬리 부분이 짧은 형태이므로 데이터의 분포가 꼬리가 긴 형태를 띤다면 포아송 분포는 적절하지 않을 수 있다(Green, 2021). 이러한 현상은 과산포(over-dispersion)가 발생한다고 명명한다(Green, 2021). 과산포 발생은 데이터에 적합한 모형을 탐색할 때 중요한 문제이다. 포아송 분포에서 과산포는 개인 간 차이와 사건 간 의존성으로 인해 발생한다. 포아송 분포는 사건이 서로 독립적이고 평균과 분산이 동일하다는 가정하에 성립되므로 이 두 가정이 위배되면 과산포가 발생한다. 이 문제점의 대안으로 음이항 분포를 사용하며 포아송 분포에서 발생하는 과산포를 처리할 수 있다.

(3) 음이항 분포

범죄발생, 가출빈도, 청소년의 비행 발생률은 일반적인 사건 발생 확률에 비하여 희박하다. 이러한 이유로 평균과 분산이 동일하다는 포아송의 가정은 쉽게 성립하지 않는다. 이로 인하여 발생하는 과산포는 회귀계수의 표준오차를 편향되게 만들기 때문에 음이항 분포를 사용하여 과산포를 개선하고자 한다(정재풍, 최종후, 2014). 음이항 분포(negative binomial distribution)는 확률분포함수를 바탕으로 포아송 분포에 감마분포를 혼합하여 나타낸다. 음이항 분포의 감마분포는 기본적으로 포아송 분포에서 설명하지 못하는 이질성(Poisson heterogeneity)을 반영한다(Hilbe, 2011). 과산포를 설명하는 감마 분포 형태의 모수는 α (알파) 값이다. 그러므로 음이항 분포의 분산은 $Var(Y) = \mu + \alpha\mu^2$ 로 표현한다. α 가 0이면 음이항 분포의 평균과 분산은 같아지며 음이항

분포는 포아송 분포로 수렴된다(Coxe et al., 2009). 음이항 분포 식은 <수식 3>과 같다(Berk & MacDonald, 2008).

$$\ln(\mu_i) = X_i^T\beta + \epsilon_i = \ln(\lambda_i) + \ln(v_i) \quad \langle \text{수식 3} \rangle$$

위의 식에서 μ_i 는 i번 째 개인의 기댓값이며, v_i 는 과산포 요인을 반영하는 임의효과로 설명할 수 있다. ϵ_i 는 일반적인 일반선형회귀에서 교란항(disturbance term)과 유사하며 평균이 0이고 분산이 동일한 값을 가지며 서로 독립적으로 생성된다고 가정된다(Berk & MacDonald, 2008). 이는 오차항 또는 랜덤효과와 유사한 개념이다. ϵ_i 의 성질에 대한 추가적인 가정을 위해 수정한 포아송 분포 식은 다음과 같다(Berk & MacDonald, 2008). X_i 와 v_i 가 존재할 때, Y_i 의 분포가 여전히 포아송 분포 형태를 유지한다는 것을 의미한다. 실제 X_i 는 관측 가능한 값이며 따라서 Y_i 의 분포가 X_i 에 조건부로 설명 가능해야 하며 이를 위해서는 v_i 가 어떤 분포를 따르는지 명확히 해야 한다(Berk & MacDonald, 2008). Greene(2003)은 계산상 편의를 위해 $v_i = e^{\epsilon_i}$ 가 감마분포를 따른다고 가정한다. <수식 4>의 식은 Berk와 MacDonald(2008)의 논문에서 인용하였다.

$$f(Y_i = k | X_i, v_i) = \frac{\exp(-\lambda_i v_i)(\lambda_i v_i)^k}{k!} \quad \langle \text{수식 4} \rangle$$

이 식을 적용하여 음이항 분포의 확률밀도함수는 <수식 5>와 같이 표현된다(Cameron & Trivedi, 1998, 2013). α 값이 0이면 포아송 분포와 기댓값이 동일해진다(Coxe et al., 2009).

$$f(Y_i | X_i) = \frac{\Gamma(Y_i + 1/\alpha)}{\Gamma(Y_i + 1)\Gamma(1/\alpha)} \left(\frac{1/\alpha}{1/\alpha + \lambda_i}\right)^{1/\alpha} \left(\frac{\lambda_i}{1/\alpha + \lambda_i}\right)^{Y_i} \quad \langle \text{수식 5} \rangle$$

(4) 영과잉 모형

발생하기 어려운 사건일수록 자료에 ‘0’이 차지하는 비율이 많아지며, 이로 인해 가산자료에서는 분산이 평균보다 커지는 과산포 현상이 발생한다(나우열, 이현숙, 2018). 앞선 포아송과 음이항 분포도 ‘0’ 분포가 많은 가산자료 분석이 가능하지만 매우 많은 영과잉 자료는 다른 모형을 적용하여 분석하여야 한다(Green, 2021). 영과잉 자료일 경우, 포아송 모형과 음이항 모형보다는 허들 또는 영과잉 모형을 적용해야 한다. 허들 모형과 영과잉 모형은 모두 2 단계 구조이나 ‘0’을 어떻게 처리할 것인지에 차이점이 있다(Green, 2021). 영과잉 모형은 ‘0 응답 집단’을 전혀 비행을 하지 않는 집단과 아직 비행을 저지르지 않은 집단이 혼합되어 있다고 가정한다. 그러나 허들 모형에서는 ‘0 응답 집단’이 전혀 비행을 하지 않는 집단이라고 가정한다. 결론적으로 영과잉 모형은 ‘0’의 발생 가능성을 가지고 분석을 시행하는 반면, 허들 모형에서는 ‘0’을 제외한 양수 부분의 분석만 시행한다. 영과잉 포아송 분포 식은 Lambert(1992)에 의해 제안된 것으로 <수식 6>에 확률 식으로 정의한다. 여기서 p_i 는 항상 ‘0’ 집단에 속할 확률이며 $1-p_i$ 는 일반적인 포아송 분포에 따르는 집단에 속할 확률이다. 따라서 ‘0’ 집단에 속할 확률은 항상 ‘0’ 집단에 속할 확률과 포아송 분포에서 우연히 ‘0’이 나올 확률을 합한 것으로 설명할 수 있다.

$$\begin{aligned}
 Y_i = 0 & \quad p_i + (1-p_i)\exp(-\lambda_i) \\
 Y_i = k & \quad (1-p_i) \frac{\exp(-\lambda_i)(\lambda_i)^k}{k!}, \quad k = 1, 2, 3, \dots
 \end{aligned}
 \quad \langle \text{수식 6} \rangle$$

영과잉 음이항 분포 식은 <수식 7>과 같이 정의될 수 있다(Zhang, Lei, Cai, & Liu, 2012). 아래 식에서 α 는 과산포 모수이며 분산이 평균보

다 큰 정도를 조절하며 0에 가까워질 때 영과잉 음이항 모형은 영과잉 포아송 모형으로 수렴한다. 감마 함수 $\Gamma()$ 는 음이항 분포의 이산확률을 계산할 때 사용한다.

$$\begin{aligned}
 Y_i = 0 & \quad p_i + (1-p_i)\left(\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right)^{1/\alpha} \\
 Y_i = k & \quad (1-p_i) \frac{\Gamma(Y_i+1)\Gamma(1/\alpha)}{\Gamma(Y_i+1)\Gamma(1/\alpha)} \left(\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha)+\lambda_i}\right)^{Y_i}
 \end{aligned}
 \quad \langle \text{수식 7} \rangle$$

영과잉 모형은 준거변수의 두 가지 측면을 분석하기 위해 두 개의 하위 모델을 사용한다. 첫 번째로는 ‘Count’ 모델에 대한 분석으로, 준거 변수에서 ‘0’값을 제외한 데이터에서 계수값을 분석한다(Carroll, Rhew, & Larimer, 2020). 이는 예측 변수가 준거 변수에 미치는 영향을 추정할 결과이다. 두 번째는 ‘Zero’ 모델에 대한 분석으로, 준거 변수가 ‘0’인 값과 ‘0’이 아닌 값을 예측하는 logit모형이다(Carroll et al., 2020). ‘Count’ 모델에서는 포아송 또는 음이항 분포를 사용하여 실제 카운트가 발생할 수 있는 경우에만 적용한다. 회귀계수는 사건 발생률의 비율(Count Ratio 또는 Rate Ratio)로 해석한다. 즉 예측 변수가 한 단위 증가할 때, 사건 발생률이 몇 배로 변하는지로 해석해야 한다. 이와 다르게 ‘Zero’ 모델은 준거 변수가 ‘0’에 속하는지 아닌지를 예측하는 모델이다. 즉 비행을 하지 않은 집단(0 응답자)에 있을 확률로 분석하며 비행을 저지르지 않는 확률로 해석해야 한다. 회귀계수는 ‘0’값 대비 ‘0’이 아닌 값의 발생 확률에 대한 Odds Ratio(OR)로 해석되며 특정 예측 변수가 준거 변수가 ‘0’일 확률에 어떤 영향을 미치는지 해석한다(Carroll, Rhew, & Larimer, 2020).

2. 방법

1) 연구 문제 및 연구 모형

본 연구는 희박한 사건 발생 횟수 분석에 적합한 회귀모형을 찾기 위하여 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 회귀 모형을 비교 검증하는 것이다. 희박한 사건의 발생 횟수를 분석하기 위하여 다문화 청소년의 비행 횟수 자료를 예제로 활용하였다. 다문화 청소년의 비행 경험 자료는 영과잉 가산 횡단자료의 전형적인 분포 특성을 지니는 실증 사례로 볼 수 있다. 본 연구는 이러한 자료를 분석 사례로 설정하여, 영과잉 가산자료 분석 시 모형 선택에 따라 실증 결과의 해석이 어떻게 달라질 수 있는지를 예시적으로 제시하고자 한다. 설정한 연구 문제는 아래와 같다.

- 연구문제 1. 가산 자료 분석에 가장 적합한 회귀 분석 모형은 무엇인가?
- 연구문제 2. 영과잉 가산자료를 분석하기 위하여 적합한 회귀 모형은 무엇인가?

2) 연구 대상

본 연구는 한국청소년정책연구원에서 수행한 「다문화청소년 패널조사2기 데이터(MAPS 2)」를 활용하였다. 본 자료는 다문화 청소년의 발달을 종단적으로 추적함으로써 보다 객관적인 특성을 파악하고자 실시되었다. 2기 패널은 2019년부터 새로운 패널을 구축하여 현재 2023년도까지 총 5차례 조사가 진행되었으며, 가장 최근에 수집된 5차년도 패널 데이터를 사용하였다. 5차 조사 대상은 2023년 당시 중학교 2학년인 다문화 청소년 패널이며, 비행 경험을 묻는 질문에 응답한 청

소년을 대상으로 총 1,842명을 연구 대상으로 설정하였다.

3) 측정 변수

(1) 준거 변수: 비행

비행 변수는 지난 1년 동안 비행 경험을 묻는 문항을 합쳐 구성하였다. 다양한 유형의 비행 경험을 묻는 문항은 12개로 ‘담배를 피운 적이 있다’, ‘술을 마신 적이 있다’, ‘가출한 적이 있다’ 등과 같은 문항으로 구성되어 있다. 원 문항은 해당 경험에 대해 6개의 서열형 문항(1=‘전혀 없다’, 2=‘1년에 1~2번’, 3=‘한 달에 1번’, 4=‘한 달에 2~3번’, 5=‘1주일에 1번’, 6=‘1주일에 여러 번’)으로 측정하였다. 서열형 문항의 빈도분석 결과, ‘전혀 없다’ 응답을 제외한 대부분이 ‘1년에 1~2번’으로 응답하였다. 이에 이를 이분형 문항으로 재코딩(0=‘전혀 없다’, 1= ‘1년에 1번 경험하였다’)하였다. 총점의 범위는 0점부터 12점이며 점수가 높을수록 1년에 다양한 비행을 경험한 것으로 해석할 수 있다.

(2) 예측 변수

① 부모감독

부모감독 문항은 총 3개로 구성되어 있으며 허묘연(2000)에서 발췌하고 수정한 척도를 사용하였다. 문항은 ‘부모님(보호자)은 내가 방과 후에 어디에 가는지 알고 계신다’, ‘부모님(보호자)은 내가 시간을 어떻게 보내는지 알고 계신다’, ‘부모님(보호자)은 내가 외출할 경우 언제 들어올지 알고 계신다’ 3개의 문항으로 구성되었다. 원 문항은 4점 리커트 척도(1=‘전혀 그렇지 않다’, 4=‘매우 그렇다’)로 측정하였으며 3개의 문항의 평균값을 사용하였다. 1점부터 4점까지의 범위를 가지며 점수가

높을수록 부모감독 수준이 높음을 의미한다. 부모 감독의 내적 일치도(Cronbach's α)값은 .77(어머니)~.79(아버지)로 나타났다.

② 비행친구

MAPS 2기에서 비행친구 문항은 김순규(2001)의 문항을 수정하여 보완한 문항을 사용하였으며 문항은 다음과 같다. '내 친구 또는 가까운 선후배는 특별한 이유 없이 종종 결석을 한다', '내 친구 또는 가까운 선후배는 담배를 핀다' 등 총 6문항으로 구성되어 있으며 범주형 문항(1='그렇다' 2='그렇지 않다')으로 측정하였다. 이를 이분형 문항으로 재코딩(0='그렇지 않다', 1='그렇다')하여 6문항을 합산하였다. 비행친구 변수의 총점의 범위는 0에서 6까지이며, 점수가 높을수록 1년에 여러 유형의 비행을 하는 친구가 있음을 의미한다.

③ 이중문화수용태도

이중문화수용태도를 측정하기 위한 문항은 노충래와 홍진주(2006)에서 척도를 발췌하고 수정하여 사용하였다. 이 문항들은 문화적 동화에 초점을 맞추어 몽골 이주 노동자 자녀들의 한국과 몽골 문화와 미래 거주 및 진학 국가, 취학 희망 정도 등을 측정하기 위하여 4점 리커트 척도로 구성하였다(노충래, 홍진주, 2006). 본 설문에서 사용된 이중문화수용태도의 문항은 다음과 같다. '나는 한국문화(음악, 영화, 음식, 옷 등)를 즐기는 편이다', '나는 모국의 문화(음악, 영화, 음식, 옷 등)를 즐기는 편이다' 등 10개의 문항으로 구성되었다. 원문항은 4점 리커트 척도(1='전혀 그렇지 않다', 4='매우 그렇다')로 측정하였으며, 이중문화수용태도의 변수는 10개의 문항을 합산하여 평균값으로 설정하였다. 범위는 1에서 4까지의 값을 가지며 4점에 가까울수록 이중문화수용태도가 높음을 의미한다. 이중문화수용태도 문항의 내적일

치도(Cronbach's α)값은 .77로 나타났다.

4) 분석 방법

데이터를 분석하기 위해 SPSS 25.0과 R 4.4.2를 활용하였다. 첫 번째로 변수의 특성을 분석하기 위하여 빈도분석을 진행 후, 사용된 변수 간 관계를 살펴보기 위하여 상관분석과 선형회귀분석을 진행하였다. 두 번째로 오차의 정규성 가정을 만족하는지 확인하기 위하여 R 4.4.2의 기본함수를 활용하였다(R Core Team, 2024). 세 번째로 적합한 모형을 찾기 위하여 R 4.4.2에서 glm 함수를 사용하여 포아송 모델을 분석하였고, MASS 패키지(Venables & Ripley, 2002)의 glm.nb 함수를 통해 음이항 모델을 분석하였다. 그 다음으로 pscl 1.5.9 패키지(Jackman, 2024)를 사용하여 zeroinfl 함수(Zeileis, Kleiber, & Jackman, 2008)를 통해 영과잉 음이항 회귀분석을 진행하였다. 마지막으로 가장 적합한 두 모형의 적합도를 판단하기 위하여 R의 기본함수 stats 패키지를 사용하여 AIC값과 BIC값을 도출하였다. AIC와 BIC의 값이 작을수록 더 좋은 모형의 적합도를 가진다(이휘인, 2016; Hilbe, 2011). AIC 지표는 모형이 얼마나 데이터를 잘 설명하는지와 얼마나 단순한 모형이냐를 함께 고려하는 지표이다(Green, 2021). BIC는 일반적으로 AIC 값보다 두 번째 항(복잡도에 대한 패널티)의 값이 더 크다(Hilbe, 2011). AIC와 BIC는 모두 같은 분포 유형과 동일한 관측치 수를 가진 비중첩모델(non-nested models)간 비교를 위해 고안되었다(Hilbe, 2011). 굳이 AIC 값과 BIC 값을 쓰지 않아도 포아송과 음이항은 서로 중첩된 모형이므로 likelihood ratio test(LRT)를 사용하여 모형 비교가 가능하며, 영과잉 포아송 모형과 영과잉 음이항 모형 역시 LRT를 사용하여 모형 비교가 가능하다(Hilbe,

2011). 그러나 본 연구에서는 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 4가지 모형을 동시에 비교하려 한다. 4개의 모형 모두 중첩이 되는 것은 아니므로 LRT를 사용하여 판단하기에는 무리가 있으므로 AIC 값과 BIC 값을 사용하여 가장 적합한 모형을 찾으려고 한다. 마지막으로 영과잉 포아송 모형과 영과잉 음이항 모형 비교를 위하여 R의 기본 함수를 사용하여 LRT(Likelihood ratio test) 검정 통계량과 Wald 검정 통계량을 도출하였다. Wald의 통계량은 추정된 과산포 계수를 그 표준오차로 나눈 비율로 정의된다(Zulkifli, Ismail, & Razali, 2011). 우도비 통계량은 $T=2(\ln L_1 - \ln L_2)$ 로 정의되며, T 값은 자유도 1의 카이제곱 분포를 따른다.

3. 결 과

1) 인구통계학적 특성

연구 대상의 인구통계학적 특성을 살펴보기 위하여 빈도분석을 실시하였으며 그 결과를 <표 1>에 제시하였다. 남성이 949명(51.5%)로 여성($n=893$, 48.5%)보다 많았으며, 나이는 14세가 가장 많은 분포를 차지하고 있다($n=1,728$, 93.8%).

<표 1> 표본의 성별 및 연령 빈도분석
N=1,842

변수	값	빈도(%)
성별	남성	949(51.5)
	여성	893(48.5)
나이	13	21(1.1)
	14	1,728(93.8)
	15	79(4.3)
	16	12(0.7)
	17	2(0.1)

다문화 청소년의 어머니와 아버지의 출생 국가 빈도표를 <표 2>에 제시하였다. 어머니의 경우, 베트남(34.7%), 중국(17.8%), 조선족(14.5%) 순으로 나타났으며, 아버지는 한국(83.3%)이 가장 많이 차지함을 확인할 수 있다.

<표 2> 다문화 청소년의 어머니와 아버지의 출생 국가 빈도표

	어머니 출생국가 빈도(%)	아버지 출생국가 빈도(%)
나라	빈도(%)	빈도(%)
한국	32(1.7)	1,402(83.3)
중국(한족, 기타민족)	327(17.8)	28(1.7)
중국(조선족)	268(14.5)	127(7.5)
베트남	639(34.7)	19(1.1)
필리핀	170(9.2)	4(0.2)
일본	107(5.8)	4(0.2)
태국	41(2.2)	0
기타	258(14.0)	99(5.9)
전체	1,842(100)	1,683(100)

2) 비행 빈도분석

준거 변수인 비행의 분포를 확인하기 위하여 빈도분석을 실시하였으며 그 결과를 <표 3>에 제시하였다. 비행 변수의 경우 '0회' 응답이 가장 많았다($n=1,650$, 89.6%). 비행횟수가 '0회'인 경우를 제외하면 한 해에 한 가지 유형의 비행을 저지른 경우가 가장 많았다($n=130$, 7.1%). 준거 변수인 비행은 연속형 변수가 아닌 가산자료 형태로 나타난다. 즉 비행 변수는 음수가 존재하지 않으며 연속형이 아닌 '0' 이상의 값으로 정적 편포된 형태를 보인다. 비행변수의 평균값은 0.17이며 분산은 0.39로 나타난다. 응답값 '0'을 제외한 192개의 평균은 1.61이며 분산은 1.42이다. 이를 통해 비행 변수의 분포는 두 개의 이질적인 집단이 혼합되어 있음을 예측할 수 있다.

<표 3> 비행 변수의 빈도분석

N=1,842

비행유형별 횟수	0	1	2	3	4	5	6	7
빈도	1,650(89.6)	130(7.1)	36(2.0)	14(0.8)	3(0.2)	5(0.3)	0(0)	4(0.2)

3) 비행, 부모감독, 비행친구, 이중문화수용 태도의 평균, 표준편차 및 상관분석

비행, 부모감독, 비행친구, 이중문화수용태도의 분포를 살펴보기 위하여 평균과 표준편차를 제시하였으며, 변수 간 관계를 파악하기 위하여 상관분석을 실시하였다(<표 4> 참조). 분석 결과, 준거 변수의 평균 값은 ‘0’에 매우 근접하였으며 표준편차도 크지 않아 ‘0’ 주위로 값이 집중되어 있는 분포를 보여준다($M=0.17$, $SD=0.62$). 비행은 부모감독($r=-.14$, $p<.001$)과 이중문화수용태도($r=-.06$, $p<.01$)와 부적상관관계가 통계적으로 유의하였다. 이는 부모감독과 이중문화수용태도가 높은 다문화 청소년은 적은 횟수의 비행을 경험하고 있는 것을 의미한다.

이와는 반대로 비행친구($r=.31$, $p<.001$)는 정적 상관관계가 통계적으로 유의하였다. 이는 다양한 유형의 비행을 많이 저지르는 비행친구를 가진 다문화 청소년은 많은 유형의 비행을 경험함을

시사한다.

4) 비행과 부모감독, 비행친구, 이중문화수용 태도의 선형회귀분석

<표 5>에 독립변수가 비행에 미치는 영향을 파악하기 위하여 선형회귀분석을 시행한 결과를 제시하였다. 비행을 예측하는 부모감독, 비행친구, 이중문화수용태도를 포함한 회귀모형은 통계적으로 유의하게 나타났다($F(1838,3)=75.39$, $p<.001$). 부모감독($b=-.10$, $p<.001$)과 비행친구($b=.14$, $p<.001$)가 비행에 미치는 영향이 통계적으로 유의한 것으로 나타났으나, 이중문화수용태도는 통계적으로 유의하지 않았다. 부모감독과 비행친구가 1 단위 증가할 때 다문화 청소년의 비행 경험은 증가하는 것으로 해석할 수 있다. 비행의 변화량의 11%는 부모감독, 비행친구, 이중문화수용태도에 의해서 설명할 수 있다($R^2=.11$).

<표 4> 비행, 부모감독, 비행친구, 이중문화수용태도 간의 범위, 평균, 표준편차, 상관분석

N=1,842

	비행	부모감독	비행친구	이중문화수용태도
비행	1			
부모감독	-.14 ^{***}	1		
비행친구	.31 ^{***}	-.10 ^{***}	1	
이중문화수용태도	-.06 ^{**}	.27 ^{***}	0.003	1
범위	0-12	1-4	0-6	1-4
M	0.17	3.20	0.55	2.87
SD	0.62	0.57	1.31	0.40

* $p<.05$ ** $p<.01$ *** $p<.001$

<표 5> 부모감독, 비행친구, 이중문화수용태도와 비행의 선형회귀분석

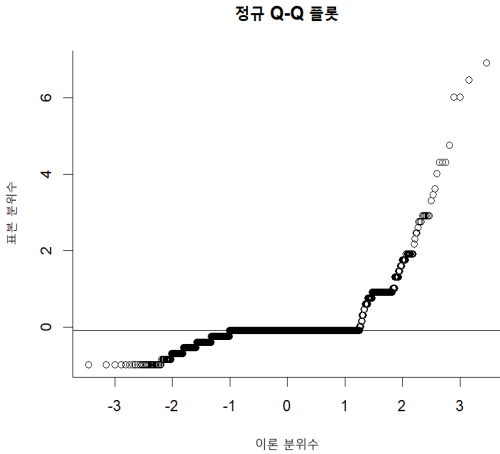
N=1,841

	B	SE	β	t	F	R ²
(상수)	0.59***	0.11		5.17	75.39***	.11
부모감독	-0.10***	0.03	-0.09	-4.09		
비행친구	0.14***	0.01	0.30	13.67		
이중문화수용태도	-0.06	0.04	-0.04	-1.66		

* p<.05 ** p<.01 *** p<.001

5) 비행과 비행친구 간 오차의 정규성 검증

다음으로 잔차가 선형회귀분석의 기본가정을 만족하는지를 살펴보기 위하여 Q-Q Plot을 <그림 1>에 제시하였다. 예측 변수는 비행친구이며, 준거 변수는 비행횟수이다. <그림 1>에는 점들이 계단형이며 수평선 형태가 나타난다.



<그림 1> 행 친구와 비행 간 오차의 정규 Q-Q플롯

이는 잔차가 정규성 가정을 만족하지 못함을 의미하며, 더 나아가 데이터가 '0'에 집중되어 있어 '0'이 많은 형태의 분포임을 보여준다. 오차의 정규성이 만족하지 않는 경우, 선형회귀분석이 아닌 음이항, 포아송 회귀분석을 고려해야 한다. 다음으로 통계적

으로 정규성을 검증하기 위하여 Shapiro-Wilk 검정을 실시하였다. 그 결과, 검정통계량은 0.44이며 p값은 .001보다 작아 영가설을 기각하게 되며, 즉 정규성 가정이 위배됨이 입증되었다.

6) 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 회귀 모형 비교

<표 6>에는 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 모형을 적용해 분석한 회귀분석 결과를 제시하였다. 포아송, 음이항, 영과잉 모형의 회귀계수 B값은 모두 로그 링크(Log link)를 사용하여 추정된 값으로, 각 예측변수가 1단위 증가할 때 비행 횟수의 기댓값(λ)이 e^B 배로 변함을 의미한다. 결과를 해석하면, 동일한 변수를 투입하였을 때, 비행 빈도에 영향을 미치는 예측 변수는 상이하게 나타났다. 먼저 포아송 회귀분석에서는 이중문화수용태도가 비행에 부적으로 영향을 준다($b=-0.30, p<.05$). 약 0.74배이므로 비행이 26% 감소함을 의미한다. 그러나 포아송을 제외한 음이항, 영과잉 포아송, 영과잉 음이항 모델에서는 그 효과가 통계적으로 유의하지 않게 나타났다. 포아송 모형이 데이터의 과산포를 고려하지 않아 불안정한 회귀계수가 도출되는 것을 실증적으로 확인한 결과이다. 반면 영과잉 포아송 모형에서 부모감독 변수가 비행에 유의한 영향을 미치지 않는 것으로 확인되었다. 이는 선형회귀분

<표 6> 선형, 포아송, 음이항, 영과잉 모델의 회귀계수 및 모델 적용

	선형		포아송		음이항		영과잉 포아송		영과잉 음이항	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
(상수)	0.59 ^{***}	0.11	0.22	0.42	0.06	0.63	1.03	0.56	0.83	0.75
부모감독	-0.10 ^{***}	0.03	-0.53 ^{***}	0.09	-0.46 ^{***}	0.14	-0.18	0.15	-0.41 ^{**}	0.16
비행친구	0.14 ^{***}	0.01	0.40 ^{***}	0.02	0.47 ^{***}	0.04	0.18 ^{***}	0.05	0.21 ^{***}	0.06
이중문화 수용태도	-0.06	0.04	-0.30 [*]	0.14	-0.34	0.21	-0.32	0.17	-0.33	0.23
<i>AIC</i>	3279.91		1676.80		1512.70		1524.92		1484.35	
<i>BIC</i>	3307.48		1698.90		1540.24		1569.04		1533.99	
θ	-		-		.29		-		.52	

* $p < .05$ ** $p < .01$ *** $p < .001$

note. 제시된 표에서 영과잉 모델은 Count model만 제시하였다.

석, 포아송, 음이항, 영과잉 음이항 모델에서 부모 감독 변수의 회귀계수가 유의하게 나타난 것과는 다른 결과이다.

모형의 적합도를 확인하기 위하여 제시된 AIC 값과 BIC 값을 고려하였을 때, 영과잉 음이항 모형이 4개의 모형 중 값이 가장 작아 데이터를 가장 잘 설명한다고 할 수 있다(AIC=1484.35, BIC=1533.99). 음이항 모형에만 제시되는 θ 값은 작을 수록 과산포가 큰 것으로 해석된다(Hilbe, 2011). 음이항 모형에서는 과산포($\theta=.29$)가 존재하는 것을 확인할 수 있다. 영과잉 음이항에서도 음이항 모형보다는 과산포 수준이 낮아졌음을 나타냈지만 그래도 과산포가 존재함을 시사한다($\theta=.52$). AIC 값, BIC 값, θ 값을 모두 고려하였을 때 결론적으로 영과잉 음이항 모형이 본 데이터에 가장 적합하였다.

영과잉 포아송 모형과 영과잉 음이항 모델이 실제 차이가 존재하는지를 검증하기 위해 LRT와 Wald 검정을 실시하였다. 영과잉 포아송 모형은 영과잉 음이항 모형의 제한된 경우로 해석될 수 있으며, 몇몇 선행 연구에서 두 모형 간 비교를 위해 LRT를 적용하였다(Zulkifli et al., 2011). 우도비 검정 외에도 모형 간 과산포 여부를 검증하기 위해 추가적으로 Wald 검정을 진행하였다. 영

과잉 음이항 모형이 영과잉 포아송 모형보다 더 적합한지에 평가하기 위해 단측 우도비 검정을 실시하였다. 우도비 통계량(LRT)과 Wald 통계량을 계산한 결과, 우도비 통계량은 42.56($p < .001$), Wald 통계량은 4.52($p < .001$)로 영과잉 음이항이 영과잉 포아송보다 더 우수한 모형임을 시사한다.

5개의 회귀계수를 비교한 결과 주요 예측변수의 회귀계수 방향과 통계적 유의성은 비슷한 패턴을 보인다. 그럼에도 불구하고 자료의 분포적 특성을 고려하지 않고 선형회귀분석을 적용해 분석한다면 회귀계수의 통계적 유의성은 비슷하게 나타날지 몰라도 계수의 크기, 표준오차, 신뢰구간 등이 왜곡될 수 있다(이휘인, 2016; Cox et al., 2009). 평균과 분산의 차이가 큰 경우, 나타나는 과산포 문제를 무시하고 포아송 모형을 적용하면 회귀계수의 추정량의 표준오차가 편향되는 현상이 발생한다(정재풍, 최중후, 2014). 즉 실제 변동성을 반영하지 못하게 되면 표준오차가 작아지고 실제로 유의하지 않은 효과가 유의하다고 잘못 판단될 위험이 증가한다. 또한 비행을 전혀 하지 않는 집단과 비행 집단을 혼합하여 포아송 또는 음이항 회귀분석을 실시하게 되면 두 집단에서의 효과를 혼동하여 보게 되는 결과를 산출하게 되며 해석이 왜곡될 위험이 있다. 따라서 모

형의 선택은 단순히 적합도 지표만을 두고 비교하는 것이 아니라 자료의 분포 특성과 이론적 배경을 함께 고려해야 한다.

7) 영과잉 음이항 회귀분석

영과잉 음이항 모델을 적용하여 회귀분석 결과를 <표 7>에 제시하였다. 영과잉 음이항 회귀분석 결과, ‘Zero’ 모델에서는 다문화 청소년이 비행 여부에 영향을 주는 변수가 없음을 확인할 수 있다. 이는 부모감독, 비행친구, 이중문화수용태도가 비행 여부에 미치는 영향은 확인할 수 없음을 의미한다. ‘Count’ 모델에서는 부모감독($B=-0.41, p=.009$)과 비행친구($B=0.21, p<0.001$)는 비행 빈도에 영향을 미치는 것으로 나타났다. 이는 부모감독 수준이 증가할수록 비행하는 빈도는 33.95% 감소하며, 비행친구의 수준이 증가할수록 비행하는 빈도는 23.90% 증가함을 시사한다. ‘Zero’ 모델에서는 logit 모형을 통해 0(없음)과 1(있음)에 속할 확률을 예측한다. 즉 비행을 하지 않음(‘0’)에 속할 확률이 얼마나 변하는지로 해석해야 한다. 모든 요인이 유의하지 않았지만 예를 들어 부모감독이 1단위 증가하면 비행을 전혀 하지 않을

(‘0’집단) 확률이 1.15배 증가하는 것으로 해석한다. 이를 바탕으로 앞선 결과를 해석한다면 다문화 청소년의 비행여부(비행을 하나 하지 않느냐)를 예측하는 요인으로 부모감독, 비행친구, 이중문화수용태도는 비행 진입 여부를 설명하는 데 뚜렷한 관련성을 보이지 않았음을 시사한다.

‘Count’ 모델에서는 ‘0’ 빈도 수를 제외한 상태에서 결과변수를 연속형으로 간주한 채 선형회귀분석을 실시하여 각 예측변수가 결과변수에 미치는 영향을 확인하였다. 이미 비행집단에 포함된 다문화 청소년의 비행 빈도 수를 증가시키는 요인이 무엇인지를 확인하고자 한다. 그 결과 부모감독이 증가할수록 비행 빈도가 감소하며 이는 이미 비행집단에 있는 청소년들에게 부모감독이 비행정도를 억제하는 요인으로 작용하는 것을 의미한다. 반대로 비행친구가 증가할수록 비행 빈도가 증가하며 이는 비행집단에 있는 청소년들에게 비행친구가 비행의 정도를 심화시키는 강화요인으로 작용하는 것을 의미한다. 이러한 결과는 비행 여부(진입 단계)와 비행 빈도(심화단계)가 서로 다른 요인에 의해 설명될 수 있음을 보여준다. 즉, 부모감독과 비행친구는 비행의 발생 자체보다는 이미 비행을 한 청소년의 비행 빈도와 더 밀접

<표 7> 영과잉 음이항 회귀분석

	Zero model				Count model			
	<i>B</i>	<i>SE</i>	OR	%	<i>B</i>	<i>SE</i>	IRR	%
(Intercept)	-0.10	1.72	0.90		0.83	0.75	2.29	
부모감독	0.14	0.37	1.15	14.96	-0.41**	0.16	0.66	-33.95
비행친구	-5.30	24.32	-0.005	-99.50	0.21***	0.06	1.24	23.90
이중문화수용태도	0.13	0.51	1.14	14.42	-0.33	0.23	0.72	-28.26
Log(θ)			-		-0.66***	0.23		
	θ				0.52			
	Log-likelihood(<i>df</i>)				-733.2(9)			
	LR $\chi^2(6)$				167.09 (p<.001)			

* $p<.05$ ** $p<.01$ *** $p<.001$

한 관련을 가지는 것으로 해석된다. 따라서 비행의 예방을 위해서는 비행 진입 이전 단계에서 다른 요인을 탐색할 필요가 있으며, 이미 비행을 경험한 청소년의 경우에는 부모감독 강화와 비행친구와의 관계 조절이 비행의 심화를 완화하는 데 기여할 수 있음을 시사한다.

4. 논 의

본 연구는 다문화 청소년의 비행 경험 변수를 사례로 하여 가산형 자료를 분석할 때 자료의 특성을 충분히 고려하지 않을 경우 발생할 수 있는 방법론적 문제를 검토하였다. 국내 다문화 청소년 비행에 관한 연구의 대부분은 한국 청소년 정책 연구원에서 제공되는 한국 다문화 청소년 패널 자료를 중심으로 이루어지고 있다. 다문화 청소년 패널 자료에서 제공하고 있는 비행 척도는 12가지 비행 경험을 묻는 문항으로 구성되어 있어 가산자료의 형태로 분석에 사용된다. 그러나 다수의 국내 다문화 청소년 비행에 관한 연구는 비행 변수를 연속형으로 간주하고 선형회귀분석을 적용하였다. 이 경우, 변수의 특성에 맞지 않는 분석모형을 적용한 연구는 신뢰성을 보장할 수 없다는 문제점이 발생한다. 이에 본 연구에서는 데이터가 가진 특성에 대해 분석한 후, 어떠한 분포와 분석 모형이 가장 적합한지에 대해 논의하고자 하였다. 한국 다문화 청소년 패널 자료를 사용하여 중다회귀분석을 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 회귀분석을 적용해 비교하여 차이가 있는지 검증하였다. 주요 분석 결과는 다음과 같다.

첫 번째로 비행변수의 분포를 확인한 결과, 대다수의 응답은 '비행을 경험한 적 없다(=0)'이며 정적으로 편포된 형태가 나타난다. 이로 인하여 비

행 변수의 평균이 '0'에 근접하고 평균보다 분산이 더 큰 특성을 보인다($M=0.17, s^2=0.39$). 가산자료의 평균이 매우 낮을 경우, 선형회귀분석을 적용하면 표준오차와 유의성 검정에서 편향된 결과를 초래할 수 있어 적절하지 않은 분석 방법이 될 수 있다(Gardner et al., 1995). 비행친구와 비행 간의 관계에서 오차의 정규성 가정을 확인한 결과, 정규성 가정을 만족하지 못하였다. 사회과학의 범죄학 분야에서 범죄 횡수, 비행 횡수 등과 같은 가산자료를 사용할 때에는 분석 방식에 대한 고찰이 선행되어야 한다. 데이터 특성을 고려하지 않은 채 분석된 결과는 그 신뢰성을 보장할 수 없으며(Grajales, 2013) 이러한 연구를 근거로 진행되는 정책은 그 효과성을 발휘할 수 없다.

두 번째로, 동일한 변수를 투입하여 선형, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 회귀분석을 각각 실시하였다. 다섯 가지 모형의 정보지수를 비교한 결과, 영과잉 음이항 회귀분석이 AIC 값과 BIC 값이 가장 낮아 본 자료에 가장 적합한 모형으로 판단하였다. 영과잉 포아송과 영과잉 음이항 모형의 예측값의 실증적 비교를 위하여 LRT와 Wald 검정을 실시한 결과, 영과잉 음이항 모형이 영과잉 포아송 모형보다 더 우수한 적합성을 보이는 것으로 나타났다. AIC 및 BIC 값과 LRT와 Wald 검정 모두 영과잉 음이항 모형이 분석에 가장 적합한 것으로 나타났다.

세 번째로, 음이항과 영과잉 음이항 모형에서는 모두 과산포가 존재하는 것으로 확인되었다. 과산포는 자료 내에서 '0'값이 과도하게 많아 평균과 분산이 동일하다는 가정이 충족하지 않을 때 나타난다(나우열, 이현숙, 2018). 이와 같은 경우 음이항 회귀분석을 적용함으로써, 과산포 문제를 해결할 수 있다. 그러나 음이항 모형을 적용한 후에도 '0'값의 비율이 여전히 과도하게 많아 모형 적합에 어려움이 존재하였다. 이러한 영과잉 상태를

고려하여 영과잉 음이항 모형을 적용하였다. 이는 단순히 음이항 회귀모형을 적용할 경우 발생하는 회귀계수 추정치 왜곡과 예측력이 감소하는 문제를 해결하기 위함이다. 그 결과, 음이항 모형에서의 과산포 계수가 영과잉 음이항에서 개선되는 것을 실증적으로 확인하였다. 본 결과를 통해 평균과 분산이 동일하다는 가정을 충족하지 못하며, '0'이 과도하게 많은 경우 영과잉 음이항 모형이 가장 적합함을 확인하였다.

본 결과를 통해 영과잉 가산자료의 경우, 선형 회귀분석이 아닌 영과잉 포아송 또는 영과잉 음이항 회귀분석을 사용해야 하며, 선형회귀분석에서 도출된 결과에 대해 주의를 기울여야함을 시사한다. 오차가 정규분포를 따를 경우, 선형회귀분석은 모든 불편추정량 중 가장 효율적이다(White & MacDonald, 1980). 그러나 오차가 정규분포를 따르지 않을 경우에는, 선형회귀분석은 선형 불편추정량이라는 제한된 범위에서만 가장 효율적일 수 있다(Wooldridge, 2009). 본 연구에서 선형회귀분석의 결과, 회귀계수와 유의성 검증에 문제는 발견되지 않았으나, 선형회귀분석의 기본가정을 만족하지 못하였다, 또한 AIC 값이 영과잉 음이항 모형과 비교하여 상대적으로 커, 적합한 모형이 아님을 시사한다. 그러므로 사회과학 연구자들은 선형회귀분석을 적용하기 전에 데이터의 특성에 맞는 모형에 대한 고찰을 선행한 뒤 연구를 진행해야 한다.

마지막으로 영과잉 음이항 회귀분석을 적용하여 본 연구의 주요 변수들이 다문화 청소년의 비행에 미치는 영향을 분석하였다. 영과잉 구조 중 'Zero' 모델에서는 비행을 경험하지 않는 집단에 대해 분석하였으며, 세 변수 모두 유의미한 영향을 미치지 않는 것으로 확인되었다. 즉 부모감독, 비행친구, 이중문화수용태도는 다문화 청소년의 비행을 할지 여부를 결정하는 직접적인 요인으로

작용하지 않았다. 반면 'Count' 모델에서는 비행을 경험한 집단에서 비행에 영향을 미치는 요소를 분석하였다. 그 결과 부모감독과 비행친구 변수가 비행에 유의한 영향을 미치는 것으로 나타났다. 이는 비행을 경험한 집단 내에서 부모감독은 보호 요인으로, 비행친구는 위험 요인으로 기능함을 시사한다. 이는 다문화 비행 청소년에게 부모감독과 비행친구는 중요한 역할을 하는 것을 내포한다. 특히 비행이 발생한 이후, 부모감독과 또래 요인이 비행의 지속 또는 심화에 중요한 영향을 미친다는 점에서 실천적 함의를 갖는다. 따라서 다문화 청소년의 비행이 심각한 수준으로 진행되지 않기 위해서는 부모감독 수준을 증가시키고 비행친구의 접촉을 차단하는 정책의 필요성을 강조한다. 이러한 결과를 학교 현장에서 적용할 경우, 다문화 청소년 전체 집단 내에서 비행 청소년과 일반 청소년을 이질적인 집단으로 구분하여 접근할 필요가 있다. 일반 청소년의 비행을 차단하는 요인에 대해서는 추가적인 연구가 이루어져야 한다. 반면 비행 청소년의 경우 학교 실무자들이 비행 청소년의 부모와의 관계와 친구 관계에 집중하며 진중한 개입이 필요함을 시사한다.

또한 부모감독과 비행친구가 다문화 청소년의 비행에 유의미한 영향을 주는 결과는 다문화 청소년의 비행 양상이 일반 청소년과 유사하게 나타남을 시사한다. 이를 바탕으로 청소년의 비행 예방 정책은 다문화 청소년과 일반 청소년을 분리하는 정책이 아닌 청소년 전체를 통합한 정책이어야 함을 주장한다. 추가적으로 앞서 중요하게 다루었던 다문화 청소년의 이중문화수용태도는 비행 여부와 비행 수준에 영향을 미치지 않는 것으로 나타났다. 이는 선행연구와 일치하는 결과가 아니므로 이에 대한 추가적인 연구가 필요하다. 이러한 논의는 다문화 청소년의 비행 행동의 특성을 설명하기 위한 목적이 아닌, 비행 변수가 지

년 계량적 분석 결과를 어떻게 해석하고 확장하는 것이 좋을지를 보여주는 것에 목적이 있다.

본 연구는 비행 데이터에 적합한 모형을 찾고 분석을 논의하였으나 몇 가지 한계점이 존재한다. 첫 번째로 5가지의 모형만 비교하였던 것을 한계로 들 수 있다. 범점화, 허들 모형에 대한 분석 역시 추가로 진행되어야 한다. 두 번째로 연구에서 더 많은 비행 예측 변수들을 투입하지 못한 한계가 존재한다. 모형 비교의 정확도 향상을 위하여 주요 변수만을 설정하여 모형을 간소화하였다. 그리하여 비행에 영향을 미치는 다른 변수들을 통제하지 못한 것을 한계로 들 수 있다. 향후 많은 변수를 투입하여 영과잉 음이항 모델을 사용하여 분석한다면 기존에 제시하였던 결과와는 다른 결과를 도출할 수 있을 것으로 예상된다. 마지막으로 본 문항을 모두 활용하지 못했다는 한계점이 존재한다. 각각의 유형의 비행에 대해 묻는 문항에 대한 응답 값은 ‘전혀없다’, ‘1년에 1-2번’, ‘한달에 1번’, ‘한달에 2~3번’, ‘일주일에 1번’, ‘일주일에 여러 번’으로 구성되어 있어 정확한 횟수를 측정하기에는 무리가 있다. 그럼에도 불구하고 ‘전혀없다’와 ‘1년에 1-2번’을 제외한 응답은 각 문항에서 13번, 11번, 9번, 15번, 43번, 3번, 7번, 1번, 3번, 6번 정도로 그 값이 전체(N=1,842) 값에 비하여 작음을 확인할 수 있다. 만약 각 응답을 추정치로 간주하여 가산변수로 환산한다면 비행 횟수 분포는 ‘0 회’(비행을 전혀 하지 않은 집단)과 ‘1보다 큰 횟수’(비행 집단)으로 확연하게 나누어 구분될 것이다. 즉 하나의 표본에 이질적인 2개의 집단이 서로 혼합되어 나타난다. 모든 응답치를 반영하여 적합한 모형이 영과잉 음이항 모형인지에 대해서는 추후 연구를 통해 확인 해 봐야 할 것이다. 그럼에도 불구하고 두 이질적인 집단이 혼합되어 있다는 점, 응답이 ‘0’인 집단이 매우 많다는 점, 그리고 분포가 꼬리가 길어진다는

점을 모두 고려하였을 때 영과잉 음이항 모형도 적합할 가능성이 높다고 생각된다. 이 부분에 대해서는 추후 연구가 진행되어야 할 것이다.

본 연구는 분석을 수행하기에 앞서, 데이터의 분포 및 특성을 면밀히 검토하고 그에 적합한 통계 분석 방법을 선택하는 것이 연구 결과의 타당성과 신뢰성을 확보하는 데 있어 얼마나 중요한지를 보여준다. 특히 이 연구에서는 기존 선행연구들이 주로 종단적 변화 분석 또는 특정 사건의 발생 요인 규명에 초점을 두어 횡단 가산자료의 모형 적합성에 대한 검증이 부족했다는 한계를 보완하고자 하였다. 이를 위해 선형회귀, 포아송, 음이항, 영과잉 포아송, 영과잉 음이항 모형을 비교·분석함으로써 가산 자료의 분포 특성에 가장 적합한 분석모형을 경험적으로 검토하였다.

그 결과, 본 연구는 포아송 계열 모형 간의 적합성 차이를 실증적으로 확인함으로써, 자료의 분포 특성이 모형 선택에 미치는 영향을 구체적으로 제시하였다. 이는 Gardner 등(1995)과 같은 기존 연구들이 다루지 않았던 영과잉 자료 상황에서의 모형 적용 가능성을 실제 데이터로 검증했다는 점에서 학문적 의의가 있다. 또한 본 연구는 나우열과 이현숙(2018), 이길재와 이정미(2014), 이휘인(2016) 등의 연구가 보여준 가산자료 분석의 확장 가능성을 범죄학 및 사회과학 맥락으로 확장하였다는 점에서도 의미가 있다. 즉, 본 연구는 하나의 분석 방법에 의존하지 않고 다양한 모형을 비교하고 검토하는 과정을 거쳤다. 이를 통해 연구 데이터의 분포 특성을 고려한 합리적 모형 선택의 중요성을 실증적으로 보여주었다. 이러한 점에서 본 연구는 향후 사회과학 및 범죄학 분야에서 가산 자료를 다루는 연구들이 자료의 분포 특성을 명확히 파악하고, 이에 적합한 통계모형을 선택하는 실증적 기준을 마련하는 데 기초자료로 활용될 수 있을 것이다.

마지막으로 본 예제 데이터가 다문화 청소년이 보편적 청소년 정책 안에서 통합적 지원 전략을 대상임을 고려할 때, 다문화 청소년의 비행은 일 마련할 필요성을 제기한다. 이러한 결과는 다문화 반 청소년과 동일한 비행원인을 가지는 것으로 나 청소년의 비행에 대한 선입견을 완화하고, 사회통 타났다. 이는 다문화 청소년을 ‘특수한 위험집단’ 합적 관점에서 청소년 정책을 재구조화하는 데 기 으로 간주해온 통념을 재검토하게 하며, 향후에는 초자료로 활용될 수 있을 것이다.

참 고 문 헌

- 김도선 (2024). 다문화가정 중기청소년의 비행에 미치는 영향요인 검증: 상호작용이론을 중심으로. <한국범죄학> 18(2), 25-47.
- 김수민 (2024). 다문화 청소년의 사회적 위축이 비행에 미치는 영향: 부모의 방임과 우울의 조절된 매개효과를 중심으로. <한국공안행정학회보>, 33(2), 85-118.
- 김수영 (2019). <사회과학통계의 기본>. 서울: 학지사.
- 김태현 · 박철형 · 남중오 (2018). 어가의 고용량 결정요인 분석. <자원 · 환경경제연구>, 27(3), 545-567.
- 김희찬 (2023). 다문화청소년 무단결석과 비행경험과의 관계에 관한 연구. <청소년학연구>, 30(3), 319-341.
- 나우열 · 이현숙 (2018). 영과잉 가산 자료의 분포 특성을 고려한 포아송 및 음이항 성장모형 및 잠재계층성장모형 비교: 청소년 가출빈도의 종단적 변화 분석을 중심으로. <교육평가연구>, 31(4), 913-939.
- 노성호 (2006). 비행친구와 비행행동의 인과성에 대한 검증. <형사정책연구>, 297-332.
- 문성원 · 이경은 (2025). 집단 괴롭힘 피해 경험이 다문화 청소년 비행에 미치는 영향: 우울을 통한 가족지지의 조절된 매개효과를 중심으로. <복지와 문화다양성연구>, 7(1), 167-200. 10.51155/KAMW.2025.7.1.167
- 박선숙 (2022). 다문화 청소년의 부모양육태도가 비행에 미치는 영향: 심리적 부적응의 매개효과를 중심으로. <교정복지연구>, 79, 1-31.
- 박성훈 · 전영실 (2017). 다문화가정 청소년의 비행특성에 관한 연구: 일반적인 비행요인과 다문화 특성요인의 상호작용을 중심으로. <한국범죄학>, 11(3), 55-76.
- 박순희 (2009). 다문화가족 아동의 특성과 사회적응. <한국아동복지학>, (29), 125-154.
- 박여주 (2022). 다문화가정 부모의 한국 생활 적응도가 자녀들의 비행과 심리적 반응에 미치는 영향에 관한 연구 탐구적 모델을 중심으로. <한국범죄학>, 16(2), 127-144.
- 박완경 · 김세봄 (2023). 다문화청소년의 학교적응이 학업중단의도에 미치는 영향: 우울 및 비행의 직렬다중매개효과를 중심으로. <청소년복지연구>, 25(3), 59-84.
- 박준호 · 성용은 (2021). 다문화가정 청소년의 문화변수준이 비행에 미치는 영향: 사회적 지지의 조절효과 분석. <한국치안행정논집>, 18(3), 73-95.
- 박지수 · 김하영 · 한윤선 (2021). 다문화청소년의 비행유형별 예측요인: 긴장이론, 사회유대이론 및 사회학습이론의 적용. <청소년복지연구>, 23(1), 27-53.
- 서봉언 (2019). 다문화 청소년의 비행경험 실태 및 비행 참여요인 탐색. <현대사회와다문화>, 9(2), 13-150.

- 송진영 · 배미경 (2015). 다문화가정청소년의 비행경험수준에 영향을 미치는 위험요인. <교정복지연구>, (39)29-58.
- 신예지 · 전성은 · 강영신 (2023). 다문화청소년의 우울 및 비행의 발달과정: 환경적 요인과 개인적 요인의 영향을 중심으로. <다문화교육연구>, 16(2), 33-56.
- 신재현 (2014). 다문화 청소년의 비행에 미치는 가정환경적요인 비교연구. <한국치안행정논집>, 11(2), 1-20.
- 심예은 (2024.3.4). 다문화학생, 학업중단을 뚜렷... 세심한 사회적 관심 필요, 케이헬스뉴스.
<https://k-health.com/news/articleView.html?idxno=70426>
- 안병수 · 박유빈 (2022.2.22). [단독] “교내 차별 너무 힘들어요”... 학업 중단 30%가 ‘부적응’ [학교밖 떠도는 다문화 청소년들]. 세계일보. <https://www.segye.com/newsView/20220221514814>
- 유창민 (2020). 다문화가정 청소년의 이중문화수용태도 변화 유형. <다문화사회연구>, 13(3), 73-107
- 윤정 · 광대경 (2025). 학교 밖 청소년의 비행 피해 경험이 비행 가해 경험에 미치는 영향: 사회적 차별의 매개효과를 중심으로. <한국민간경비학회보>, 24(2), 109-134.
- 윤홍주 (2023). 다문화청소년의 최초 비행요인에 대한 비연속시간 생존분석. <학습자중심교과교육연구>, 23(14), 617-627.
- 이길재 · 이정미 (2014). Hurdle Model 을 활용한 학교폭력 영향요인 분석. <교육행정학연구>, 32(1), 53-74.
- 이나라 · 조은경 (2023). 다문화 청소년의 비행 예측 요인 탐색: 랜덤포레스트를 활용하여. <한국경찰학회보>, 25(4), 61-86.
- 이상균 (2008). 청소년 비행행동에 대한 부모양육행동과 비행친구집단간의 조절된 매개효과. <한국아동복지학>, (27), 121-151.
- 이성식 (2007). 낮은 자기통제력과 비행친구의 청소년비행에의 영향: 청소년패널자료의 분석. <한국청소년연구>, 18(2), 159-181.
- 이승현 · 황성현 (2021). 다문화청소년 비행에 대한 범죄이론의 상대적인 영향력 검증. <한국공안행정학회보>, 9(2), 187-210.
- 이영주 (2009). 다문화가족 청소년의 심리사회적 적응에 영향을 미치는 위험요인에 관한 연구. <한국가족복지학> 14(2), 103-119.
- 이원기 · 유효은 · 이창한 (2019). 다문화청소년의 비행 영향요인에 관한 연구. <한국경찰학회보>, 21(1), 56-90.
- 이운호 · 이승욱 (2024). <범죄학 제5판> 서울: 박영사.
- 이종형 · 이준배 (2023). Analyzing dynamic joint trajectories between strain and delinquency in multicultural adolescents: Application of joint trajectory modeling. <한국사회복지조사연구>, 78, 59-88.
- 이택호 · 서연정 (2024). 청소년 비행 국내 연구동향 분석: 키워드 네트워크 및 토픽모델링을 활용하여. <청소년학연구>, 31(4), 445-474.
- 이휘인 (2016). 고등학생 무단결석 발생과 빈도에 영향을 미치는 변인 비교. <한국청소년연구>, 27(4), 201-229.
- 임혜빈 · 황정운 · 주혜린 (2024). 다문화 청소년의 비행 행위 영향요인에 관한 연구: 친구의 비행 행위와 의지할 수 있는 어른을 중심으로. <국정관리연구>, 19(1), 33-69.
- 장미희 · 박창기 (2012). 영과잉 가산자료(Zero-inflated Count Data) 분석 방법을 이용한 지역사회 거주 노인의 노인 학대 발생과 심각성에 미치는 위험요인 분석. <Journal of Korean Academy of Nursing>, 42(6), 819-832.

- 정재풍 · 최종후 (2014). 교통사고건수에 대한 포아송 회귀와 음이항 회귀모형 적합. <Journal of the Korean Data Analysis Society>, 16(1), 165-172.
- 정재훈 · 황성현 · 박희균 (2021). 다문화청소년 비행 및 범죄와 범죄친구와의 관계: 잠재성장모형을 중심으로. <한국경찰학회보>, 23(3), 1-22.
- 조강현 (2021). 다문화청소년의 문화적응스트레스가 비행에 미치는 영향. 서울대학교 대학원 박사학위논문.
- 조영선 · 조영일 (2019). 다문화청소년의 비행집단에 대한 종단적 변화 유형 및 심리사회적 특성. <인문사회 2>, 10(5), 325-340.
- 지영섭 (2022). 다문화청소년의 최초 비행 발생시점에 대한 영향요인 검증. <학교사회복지>, 60, 67-99
- 최홍숙 (2024). 다문화가정 자녀의 이중문화수용이 학교 적응에 미치는 영향: 사회적 지지의 조절효과를 중심으로 <현대사회와다문화>, 14(4), 127-154.
- 한국청소년정책연구원 (2024). [유저가이드] MAPS 2기패널 1~5차 조사 유저가이드 [PDF]. 한국청소년정책연구원. <https://www.nypi.re.kr/archive>
- 한영우 · 이창배 (2025). 다문화 청소년의 문화적응 스트레스가 비행에 미치는 영향: 내재화 문제의 매개효과. <한국치안행정논집>, 193-212.
- 홍성초 · 조제성 (2021). 다문화 청소년의 비행 유형별 범죄이론 영향요인 연구. <한국범죄심리연구>, 17(1)209-221.
- Achim Zeileis, Christian Kleiber, Simon Jackman (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8). URL <http://www.jstatsoft.org/v27/i08/>
- Agnew, R. (1991). A longitudinal test of social control theory and delinquency. *Journal of Research in Crime and Delinquency*, 28(2), 126-156.
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike* (pp. 199-213). New York, NY: Springer New York.
- Berk, R., & MacDonald, J. M. (2008). Overdispersion and Poisson regression. *Journal of Quantitative Criminology*, 24, 269-284.
- Brendgen, M., Vitaro, F., & M. Bukowski, W. (2000). Deviant friends and early adolescents' emotional and behavioral adjustment. *Journal of Research on Adolescence*, 10(2), 173-189.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (No. 53). Cambridge: Cambridge university press.
- Carroll, H. A., Rhew, I., & Larimer, M. E. (2020). Moderation of relation between psychological risk factors and alcohol use by sex. *Women & Health*, 60(3), 300-313.
- Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121-136.
- Deutsch, A. R., Crockett, L. J., Wolff, J. M., & Russell, S. T. (2012). Parent and peer pathways to adolescent delinquency: Variations by ethnicity and neighborhood context. *Journal of Youth and Adolescence*, 41, 1078-1094.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson,

- overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118(3), 392.
- Grajales, C. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research & Evaluation* [an open access journal].
- Green, J. A. (2021). Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine*, 9(1), 436-455.
- Greene, W. H. (2003). *Econometric Analysis*. 5th Edition, Prentice Hall, Upper Saddle River.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge: Cambridge University Press.
- Hirschi, T., & Stark, R. (1969). Hellfire and delinquency. *Social Problems*, 17(2), 202-213.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.
- R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.2). R Foundation for Statistical Computing. <https://www.R-project.org/>
- Sileshi, G. W. (2015). The relative standard error as an easy index for checking the reliability of regression coefficients. *August. DOI*, 10.
- Simon Jackman (2024). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. Sydney, Australia. R package version 1.5.9. URL. <https://github.com/atahk/pscl/>
- Sutherland, E. H., Cressey, D. R., & Luckenbill, D. F. (1992). *Principles of Criminology*. Lanham, MD: Altamira Press.
- Thompson, W. E., & Dodder, R. A. (1983). Juvenile delinquency explained? A test of containment theory. *Youth & Society*, 15(2), 171-194.
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. New York: Springer.
- Warr, M. (1993). Age, peers, and delinquency. *Criminology*, 31(1), 17-40.
- White, H., & MacDonald, G. M. (1980). Some large-sample tests for nonnormality in the linear regression model. *Journal of the American Statistical Association*, 75(369), 16-28.
- Wooldridge, J. M. (2016). *Introductory econometrics a modern approach*. Mason, OH: South-Western cengage learning.
- Zhang, X., Lei, Y., Cai, D., & Liu, F. (2012). Predicting tree recruitment with negative binomial mixture models. *Forest Ecology and Management*, 270, 209-215.
- Zulkifli, M., Ismail, N., & Razali, A. M. (2011, July 21-23). Zero-inflated Poisson versus zero-inflated negative binomial: Application to theft insurance data. In *Proceedings of the 7th IMT-GT International Conference on Mathematics, Statistics and Its Applications (ICMSA 2011)* (pp. 511-524). Bangkok, Thailand.

Model Comparison for Predicting Delinquency in Multicultural Adolescents: Focusing on Linear, Poisson, Negative Binomial, Zero-Inflated Poisson, and Zero-Inflated Negative Binomial Models

Jae Hyo Yu, Young-Il Cho

Department of Police Administration Forensic Psychology major, Dongguk University

This study aimed to identify the factors influencing delinquency among multicultural adolescents and to determine the most appropriate statistical model for analyzing these variables. Using data from the Multicultural Adolescents Panel Study in Korea, this study examined the effects of parental supervision, delinquent peers, and bicultural acceptance attitudes on delinquency by applying and comparing five regression models: linear, Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial regressions. The delinquency variable violated the assumptions of normality and homoscedasticity and exhibited overdispersion and excessive zeros. Among the five models, the zero-inflated negative binomial regression showed the best fit based on AIC, BIC, likelihood ratio test, and Wald test results. Furthermore, the zero-inflated negative binomial model addressed overdispersion more effectively than the negative binomial model. These findings suggest that the zero-inflated negative binomial regression model is most suitable for analyzing delinquency among multicultural adolescents when the outcome variable has unequal means and variance and a high proportion of zeros.

Keywords: Model Comparison, Delinquency, Count Data, Overdispersion, Poisson, Zero-inflated Negative Binomial