아시아교육연구 21권 1호
Asian Journal of Education
2020, Vol. 21, No. 1, pp. 217-247.
https://doi.org/10.15753/aje.2020.03.21.1.217

비연속 분할함수 성장혼합모형에서 계층 분류과정의 공변인 접근법 비교*

이정민(李政珉)**
정혜원(鄭惠苑)***

논문 요약

본 모의실험 연구는 비연속 분할함수 성장혼합모형에서 계층 분류과정에서 적용한 공변인 접근법에 따라 공변인의 추정치에 어떠한 차이가 있는지 비교하는데 목적이 있다. 본 연구에서는 공변인 효과 추정시 주로 사용되는 1단계 접근법과 3단계 접근법 적용시 공변인의 추정치를 비교하였다. 모의실험을 위한 조건으로 공변인 간 상관계수(.20, .60), 계층 분류정도(1.3, 2.6), 계층 혼합비율(90:10, 70:30, 50:50), 표본크기(200, 500, 1000, 2000)를 설정하여 모의자료를 생성한 후, 두 가지 접근법을 활용한 결과의 평가 준거로 모수추정치의 편의, 상대적 편의, 평균제곱오차, 표준오차의 상대적 편의, 비수렴수와 비수렴 비율, 계층 간 교차이동 비율을 적용하였다. 모의실험 연구결과, 1단계 접근법과 3단계 접근법에 따른 공변인 효과의 모수추정치는 일부 조건을 제외하고 대부분의 모의실험 조건에서 기준치에 양호하게 추정되었다. 그러나 계층 간 거리가 좁고, 계층의 혼합비율이 하나의 계층으로 과도하게 분류되며 표본크기가 작은 경우 편의가 발생하였으며, 1단계 접근법과 3단계 접근법 모두에서 이상치의 결과를 보일 수 있는 점을 확인하였다. 본 모의실험 연구는 비연속 분할함수 성장혼합모형에 공변인 접근법에 따른 공변인 효과를 비교 검증하였다는 점에서 연구의 의의가 있으며, 연구결과를 바탕으로 시사점 및 후속연구를 제언하였다.

주요어 : 비연속 분할함수 성장혼합모형, 공변인 접근법, 1단계 접근법, 3단계 접근법

* 이 논문은 이정민의 박사학위논문 중 일부를 수정·보완한 것임.

** 제 1저자, 충남대학교 교육학과 박사 졸업
*** 교신저자, 충남대학교 교육학과 교수

Ⅰ. 서론

종단연구는 장기간 동안 반복적인 관찰과 조사에 따라 동일한 개인이나 집단을 대상으로 표집된 자료이기 때문에, 이들의 성장과정에서 나타나는 개인차뿐만 아니라 개인 내 변화도 파악할 수 있으며 집단의 특성에서 초래된 발달의 변화 및 그 변화와 관련된 변인들 간 인과관계 등을 분석하고자 할 때 유용하다(Heagerty, Liang, Zeger, & Diggle, 2002). 교육학 및 심리학의 사회과학 분야 연구에서 관심 변인은 시간이 흐름에 따라 어떠한 변화를 이루고, 그에 따른 성장궤적의 형태를 분석하여 개인 및 학습자를 보다 폭넓게 이해하는데 중요한 요소로 작용되고 있다. 특히 교육학 연구에서 학업성취와 같은 교육 자료들은 종단자료의 특성을 지니고 있으며(Gall, Gall, & Borg, 2003), 유치원과 초등, 중등, 고등 교육과 같이 학교급 단위로 이루어져 있기에 매학기 및 학년 단위로 교육과정이 나누어 진행되기 때문에 학생 간 지적·심정적·행동적 영역에서 종단적 변화의 특성이 내재되어 있다. 하지만, 각 국가들은 그들의 교육에 맞게 편성한 교육과정과 교육적 특성을 고려할 때 종단자료 구축 없이는 인간의 발달적 특성을 평가하고 분석하는데 어려움이 있다(Singer, & Willett, 2003). 이에 국내 기관에서는 서울교육종단연구(SELS), 한국교육종단연구(KELS) 등과 같은 패널자료가 축적되고 있다.

이러한 패널자료를 바탕으로 연구자들은 종단연구를 실시하고 있으며, 3회기 이상 반복 측정된 피험자 자료를 이용하여 분석하는 통계기법을 주로 활용하고 있다(여승수, 박소희, 2012). 이를 위해 단일 특성을 지닌 구인을 측정하고 변화궤적을 추적할 수 있는 잠재성장모형(Singer & Willett, 2003)과 개인 성장 특성의 이질성에 따라 분류되는 잠재계층 분석이 결합된 성장혼합모형(growth mixture model, GMM)이 활용되고 있다. 또한 성장혼합모형에서는 학기, 학년의 시작과 개입의 특정시점을 반영하는 분할함수 성장혼합모형으로 단계적이며 순차적으로 확장할 수 있으며(Kim, 2011), 개입시점에 따라 일정거리가 분절된 비연속 분할함수 성장모형(discontinuous piecewise growth mixture model, DPGMM)으로 모형화 할 수 있다. 이러한 모형은 유치원에서 초등학교, 초등학교에서 중학교, 중학교에서 고등학교에 이르는 학교급의 급격한 변화에 따른 교수·학습 내용 및 방법, 교육적 패러다임의 전환을 반영할 수 있다(허균, 2014). 이에 따라 학교 현장의 학교급, 매학기 및 학년 단위를 단계의 구분 기준으로 설정하고, 이를 반영할 수 있는 비연속 분할함수 성장혼합모형을 적용하여 이를 구체적으로 살펴볼 필요가 있다.

한편, 혼합모형에서는 잠재계층 소속 확률과 성장궤적에 관련된 공변인을 모형에 포함시킴으로써, 계층의 이질성 및 계층에 대한 추가적인 정보를 제공받는다(Muthén, 2004). 공변인은 소속 집단의 예측을 향상시킬 뿐만 아니라 잠재계층 분류를 용이하게 한다(Dayton & McReady, 1988). 이에 최근 혼합모형에서는 공변인 효과에 대한 편의와 관련하여 방법론적 문제가 대두되고 있으며(Nylund-Gibson & Masyn, 2016; Vermunt, 2010), 분석과정에서 공변인의 정보를 통합하는 것이

정확한 모수를 추정한다는 측면에서 보편적으로 받아들여지고 있지만(Li & Hser, 2011), 공변인의 투입 시점, 즉 잠재계층 분석과 동시에 공변인을 포함할 것인지, 잠재계층이 분류된 이후 공변인을 포함할 것인지에 대한 의견이 논의되고 있다(Asparouhov & Muthén, 2014a; Clark & Muthén, 2009; Wurpts & Geiser, 2014; Vermunt, 2010).

공변인의 투입 시점에 대한 논의는 '1단계 접근법(one-step approach)'과 '3단계 접근법(three-step approach)'을 통해 이루어지고 있다. 1단계 접근법은 공변인과 측정변인의 분석이 동일한 단계에서 분석되는 방법으로서, 잠재계층을 분류하는 동시에 공변인의 효과가 추정되는 것을 의미한다(Bolck, Croon, & Hagenaars, 2004). 반면에, 3단계 접근법은 1단계 접근법과 다르게 계층의 분류와 공변인 효과의 추정이 단계적으로 수행되며, 첫 번째 단계에서는 지정된 잠재계층의 수에 따라 잠재계층의 구조를 식별한 다음, 2단계에서는 사후계층소속확률(posterior class membership probability)에 의해 소속그룹에 개인을 할당한 후 계층의 오분류를 수정하고, 마지막 3단계에서 공변인과 분류된 잠재계층 간 관계가 평가된다(Asparouhov & Muthén, 2014a; Bakk, Tekle, & Vermunt, 2013; Bolck et al., 2004; Diallo & Lu, 2017). 1단계와 3단계 접근법 중 일부 연구자들은 1단계 접근법이 비편향적이고 효율적인 모수 추정과 관련이 있다고 하지만, Asparouhov와 Muthén(2014b), Vermunt(2010)를 비롯한 다른 연구자들은 3단계 접근법이 계층 분류의 왜곡이 발생하지 않기 때문에 더욱 신뢰할 수 있는 모수추정치가 산출된다고 주장한다. 그러나 이러한 선행연구는 두 접근법을 비교하기 위하여 잠재계층이나 성장혼합모형을 적용하였다. 특히 학교급 및 교육과정의 변화에 따라 분절된 개입시점을 반영한 비연속 분할함수 성장혼합모형을 적용하여 1단계 접근법과 3단계 접근법 간 차이를 비교한 연구는 전무한 실정이기에, 이들 모형을 적용하여 두 공변인 접근법이 어떠한 차이를 보이는지 구체적으로 살펴볼 필요가 있다.

이에 본 연구에서는 집단 간 이질적인 성장궤적을 갖는 비연속 분할함수 성장혼합모형의 계층 분류과정에서 1단계와 3단계 접근법에 따른 공변인 추정치의 양호도 검증을 위하여MonteCarlo 모의실험을 실시하고자 한다. 양호도 평가 준거로 편의, 상대적 편의, 평균제곱오차, 표준오차의 상대적 편의, 비수렴수와 비수렴 비율, 계층 간 교차이동 비율을 적용함으로서 어떠한 접근법이 모수에 가장 근접하는지에 대한 정보를 제공하고 한다. 이러한 연구의 목적을 달성하기 위해 제시된 연구문제는 다음과 같다.

연구문제 1. 비연속 분할함수 성장혼합모형에서 공변인 접근법(1단계, 3단계)에 따른 공변인의 양호도(편의, 상대적 편의, 평균제곱오차, 표준오차의 상대적 편의, 비수렴수와 비수렴 비율)은 어떠한가?

연구문제 2. 비연속 분할함수 성장혼합모형에서 공변인 접근법(1단계, 3단계)에 따른 계층 교차이동 비율은 어떠한가?

Ⅱ. 이론적 배경

1. 비연속 분함함수 성장혼합모형

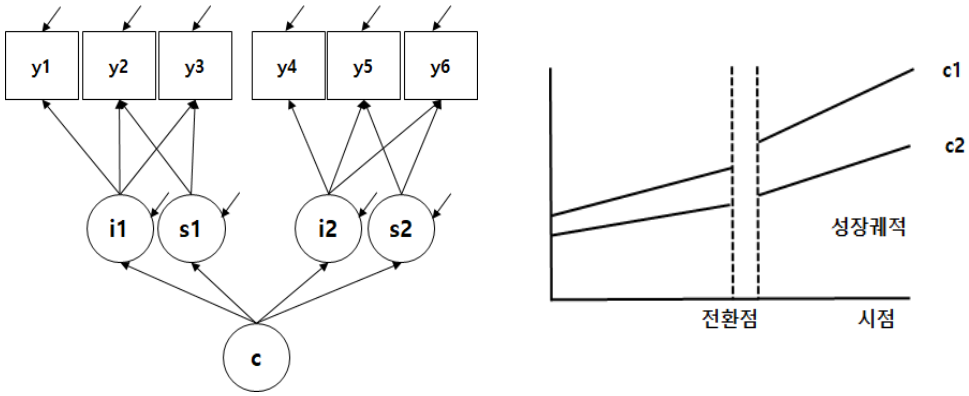
인간은 다양한 속성을 지니고 있으며, 하나의 단일요인으로는 판단하기 어려워 공통성을 지닌 하위집단을 파악하는 연구가 필요하다(Wang & Hanges, 2011). 실제로 관찰된 자료는 서로 다른 하위집단을 지니고 있으며, 이러한 이질적인 하위집단을 탐색하고 각 집단에 따른 성장유형을 확인하는 연구가 고려되고 있다(Kim, 2011). 이에 성장혼합모형은 잠재성장모형과 잠재계층이 결합되어 시간이 흐름에 따라 질적으로 다른 발달 양상을 지닌 집단이 구분하고 집단 간 성장궤적을 다르게 추정할 수 있다(Wang & Bodner, 2007).

성장혼합모형의 주요 목표는 시간 경과에 따른 변화를 반영하여 개인차를 이해하고 예측하는 것이며, 자료의 변화 유형을 토대로 계층을 분류하는 것이다(Berlin, Parra, & Williams, 2013). 이는 관측된 자료가 혼합 분포로 가정되는 유한혼합모형(finite mixture model)과 같이 k계층을 이룰 때 하나의 계층에서 속하는 개인은 그 계층의 특성을 나타내는 값으로 특징짓게 된다(Sitthisan, 2016). 즉, 성장혼합모형은 유한혼합모형을 바탕으로 수립된 분석기법으로서, 잠재계층의 구성원을 추론하여 개인을 하위집단에 할당될 확률에 따라 집단 간 성장궤적의 차이를 밝혀낸다(Nylund et al., 2007). 이는 각 집단의 개인이 집단별 평균성장곡선을 중심으로 변화할 수 있음을 의미한다(Muthén, 2004).

분할함수 성장혼합모형은 Kohli, Harring과 Zopluoglu(2016)에 의해 제안된 일종의 비선형 확률모형으로 간주하며, 분할함수 잠재성장모형과 같은 전환점(transition point)에서 절편과 기울기가 상이할 것으로 가정하며(Singer & Willett, 2003), 각 성장단계의 구체화로 전반적인 변화과정을 특정할 수 있을 뿐만 아니라 성장의 변화 형태를 반영할 수 있다(Chou et al., 2004). 이러한 변화지점, 즉 전환점은 여러 단계로 분리될 수 있다(Cudeck & Codd, 2012). 사건 발생 전환점을 기준으로 절편과 기울기 모두 상이하여 두 가지의 성장궤적을 추정하는 비연속 분할함수 성장모형을 바탕으로 전환점 이전의 절편 η_{0i}^{m}과 기울기 η_{1i}^{m}, 전환점 이후의 절편 η_{0i}^{s}과 기울기 η_{1i}^{s}을 추가하고, 본 연구에서 활용코자 하는 6시점 자료를 가정하여 6개의 측정시점과 시점 간 등간격의 비연속 분할함수 잠재성장모형의 행렬은 식(1)과 같이 표현할 수 있다(Singer & Willett, 2003).

$$\begin{bmatrix} y_{i1}^m \\ y_{i2}^m \\ y_{i3}^m \\ y_{i1}^s \\ y_{i2}^s \\ y_{i3}^s \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1\,0\,0\,0 \\ 1\,1\,0\,0 \\ 1\,2\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,1\,1 \\ 0\,0\,1\,2 \end{bmatrix} \times \begin{bmatrix} \eta_{0i}^m \\ \eta_{1i}^m \\ \eta_{0i}^s \\ \eta_{1i}^s \end{bmatrix} + \begin{bmatrix} \epsilon_{i1}^m \\ \epsilon_{i2}^m \\ \epsilon_{i3}^m \\ \epsilon_{i1}^s \\ \epsilon_{i2}^s \\ \epsilon_{i3}^s \end{bmatrix} \tag{1}$$

위의 식(1)과 같이 전환점 전후에 따라 절편과 기울기, 즉 성장요인의 불일치를 살펴보고(McCarthy, Piasecki, Fiore, & Baker, 2006), 이질적인 잠재집단을 식별하는 비연속 분할함수 성장혼합모형의 모형과 집단 간 성장궤적은 [그림 1]과 같이 표현된다.



[그림 1] 비연속 분할함수 성장혼합모형과 성장궤적

2. 공변인 접근법

비연속 분할함수 성장혼합모형에서 계층 분류과정의 공변인 접근법에 따라 공변인 효과의 차이를 검증하고자 한다. 이에 따라 공변인 효과를 검증하기 위하여 1단계 접근법과 3단계 접근법이 사용되며 이를 구체적으로 살펴보고자 한다.

1) 1단계 접근법

성장혼합모형에서 공변인 정보를 통합할 경우 통합하지 않은 잠재계층 분석보다 정확한 모수를 추정한다는 점이 더 보편적으로 받아들여져 왔지만, 공변인의 포함 시점에 대해서는 견해의 차이가 있다(Li & Hser, 2011). 즉, 잠재계층 분석과 동시에 공변인을 포함하는지, 잠재계층 분석 이후에

공변인을 포함하는지에 대한 것이다. 선행연구에서는 잠재요인과 잠재계층 소속에 공변인을 사용하는 보조변수가 성장혼합모형 분석에 포함되어 더 정확한 모수추정치와 잠재계층 할당을 얻는다고 보고하고 있다(Lubke & Muthén, 2007). 잠재계층의 수를 결정할 때 잠재계층에 공변인을 포함할지에 대해 제안하는 연구(Muthén, 2004), 계층분류 이후에만 공변인을 포함해야 한다는 연구가 이루어지고 있으며(Nylund-Gibson & Masyn, 2016), 이와 같은 시도는 계층 분류 시 공변인이 없는 성장혼합모형을 사용할 경우 방법론적으로 의문시 되는 것으로 이에 대한 방법론적 탐구가 요구된다(Bauer & Curran, 2003). 우선, 계층분류과정에서 공변인을 포함한 연구방법은 측정모형과 구조모형이 통합된 모형으로 이때 1단계 접근법이 적용된다.

$$f(y_i \mid x_i) = \sum_{k=1}^{K} p_i^k f(y_i \mid x_i, C_i = k) \tag{2}$$

위의 식(2)에서 y_i는 피험자 i의 관측 값, x는 개인 i의 공변인이며, 성장요인 η_i와 잠재계층 C_i에 영향을 미친다. 개인 i가 집단 k에 속하면 $C_i = 1$이고, 그렇지 않으면 0이 된다. 위의 식(2)에서 $p^k > 0$, $\sum_{k=1}^{K} p_i^k = 1$이며, $f(y_i \mid x_i, C_i = k)$는 집단 k에 대한 조건부 확률밀도 함수이다. 공변인을 포함한 비연속 분할함수 성장혼합모형의 측정모형과 구조모형 식은 다음과 같다.

$$\begin{aligned}
y_{ti} &= \sum_{k=1}^{K} p_i^k [\eta_{0i}^k + \eta_{1i}^k \lambda_{1t} + \eta_{2i}^k \lambda_{2t} + \eta_{3i}^k \lambda_{3t} + \epsilon_{ti}] \\
\eta_{0i}^k &= \mu_{00}^k + \mu_{01}^k x_{i1} + \zeta_{i0}^k \\
\eta_{1i}^k &= \mu_{10}^k + \mu_{11}^k x_{i1} + \zeta_{i1}^k \\
\eta_{2i}^k &= \mu_{20}^k + \mu_{21}^k x_{i1} + \zeta_{i2}^k \\
\eta_{3i}^k &= \mu_{30}^k + \mu_{31}^k x_{i1} + \zeta_{i3}^k \ , \qquad i = 1, 2 \ldots, N
\end{aligned} \tag{3}$$

잠재계층 k에서 η_{0i}^k는 절편, η_{1i}^k은 기울기이며, 이는 전환점 이전의 성장요인이다. η_{2i}^k는 절편, η_{3i}^k은 기울기이며, 전환점 이후의 성장요인이다. $\zeta_{i0}^k \sim \zeta_{i3}^k$은 k계층에서 성장요인의 잔차이다. $\mu_{00}^k \sim \mu_{30}^k$는 성장요인$\eta_i$의 절편이며, 공변인 x_i는 성장요인 η_i에 선형회귀계수이다. 잠재계층에서 예측변수의 영향력은 다항 로지스틱 회귀분석을 통하여 수행된다. 이는 식(4)와 같다.

$$p_i^k = \frac{\exp(a^k + b^k x_i)}{\sum_{k=1}^{K} \exp(a^k + b^k x_i)} \tag{4}$$

위의 식(4)에서 $a^K = 0$, $b^K = 0$이며, a^k는 로짓의 절편, b^k는 기울기를 의미한다. 모형의 추정에서 EM알고리즘에 기초한 강건한 최대우도법(maximum likelihood robust, MLR)을 활용한다.

2) 3단계 접근법

앞선 1단계 접근법과 달리 Vermunt(2010)에 의해 소개된 3단계 접근법은 잠재계층과 공변인을 분석할 때 잠재계층 분류 및 관련 분석을 단계별로 나누어 진행한다. 이러한 3단계 접근법은 잠재계층 분류과정에서 지표변인과 보조변인의 포함으로 인한 오분류 가능성을 최소화 할 수 있는 접근법이며, 1단계는 성장혼합모형을 추정하고 2단계에서는 최적 잠재계층 분류할당 변인을 생성하며 3단계에서는 잠재계층과 관련한 추가분석의 순서로 진행된다(안태용, 2018). 우선 분석을 위한 첫 번째 단계에서는 반복 측정을 야기하는 성장요인을 사용하여 잠재계층을 추정한 후 두 번째 단계에서 사후 확률을 이용하여 새로 관찰된 범주형 변인을 생성한다. 즉, 최적 잠재계층 분류 할당 변인을 이 단계에서 생성한다. 마지막으로 공변인은 두 번째 단계의 측정 오류를 고려하면서 세 번째 단계에서 관찰된 범주형 변인을 예측하고(Jung & Wickrama, 2008), 잠재계층과 결과변인 간 관계를 분석한다(Vermunt, 2010). 세 번째 단계에서 가장 높은 확률로 할당된 잠재계층 변인은 분류오류 확률을 고려하여 잠재계층의 지표변수로 사용되며, 모델 추정의 마지막 단계에 공변인이 투입되어 분석에 이용된다(Diallo, Morin, & Lu, 2017).

3단계 접근법은 잠재계층 수를 결정하는데 사용되는 측정모델이 예측변수, 즉 공변인의 영향을 받지 않는다는 이점을 지니고 있다. 이는 잠재계층 분류과정 중 공변인이 없는 무조건 모형에서 계층을 분류하는 추정과 같이 공변인과 관련된 잘못된 명세(misspecification)에 영향을 받지 않는다(Kim, Vermunt, Bakk, Jaki, & Van Horn, 2016). 이는 결국 3단계 접근법에서의 계층분류는 공변인이 없는 무조건 모형에서 계층분류와 동일한 결과를 나타난다는 것으로 이해할 수 있다. 3단계 접근법의 첫 번째 단계에서는 공변인 없이 잠재계층이 분류된다.

$$f(y_i) = \sum_{k=1}^{K} p_i^k f(y_i | C_i = k) \tag{5}$$

위의 식(5)는 1단계 방식을 적용하는 식(2)와 차이를 보인다. 특히 조건부 확률밀도에 의해 y_i는

피험자 i의 관측 값, 잠재집단 C_i는 k개의 집단을 갖는데, 이때 계층 혼합비율 $p_i^k > 0$과 $\sum_{k=1}^{K} p_i^k = 1$을 의미한다. 3단계 방법은 공변인을 포함하지 않으며, 측정모형과 구조모형 식은 다음과 같다.

$$
\begin{aligned}
y_{ti} &= \sum_{k=1}^{K} p_i^k [\eta_{0i}^k + \eta_{1i}^k \lambda_{1t} + \eta_{2i}^k \lambda_{2t} + \eta_{3i}^k \lambda_{3t} + \epsilon_{ti}] \\
\eta_{0i}^k &= \mu_{00}^k + \zeta_{i0}^k \\
\eta_{1i}^k &= \mu_{10}^k + \zeta_{i1}^k \\
\eta_{2i}^k &= \mu_{20}^k + \zeta_{i2}^k \\
\eta_{3i}^k &= \mu_{30}^k + \zeta_{i3}^k \ , \qquad i = 1, 2...., N
\end{aligned} \tag{6}
$$

또한, EM 알고리즘을 적용하기 위한 EM 알고리즘은 반복적인 최적화 과정을 사용하여 조건부 기댓값을 최대화하며, 잠재집단 소속의 조건부 확률은 다음과 같다.

$$
P(c_i = k \mid y_i) = \frac{p_i^k f(y_i \mid c_i = k)}{f(y_i)} \tag{7}
$$

EM 알고리즘은 E단계(expectation step)와 M단계(maximization step)의 두 단계로 구성되며, E단계에서 모든 모수의 시작값은 잠재집단 소속의 사후확률 추정치를 얻기 위해 수행된다. M단계에서 혼합비율, 성장 요인의 평균 및 공분산에 대한 모수추정치는 E단계로부터 사후확률을 최대화한다. E단계와 M단계는 수렴을 위해 특정 기준이 충족 될 때까지 반복 수행되며, 결과적으로 첫 단계에서 계층 내 공분산 구조의 모수는 다음 단계에 영향을 미칠 혼합모형의 추정치를 유도한다(Diallo & Lu, 2017).

3단계 접근법에서 두 번째 단계는 각 개인을 추정된 집단에 할당함으로써, 새로운 변수가 W가 생성된다. 또한 W에 대한 분류오류는 실제 값에 대한 조건부 사후 계층 소속으로부터 얻을 수 있다(Bakk, Oberski, & Vermunt, 2014; Vermunt, 2010). 분류오류는 다음과 같다.

$$
P(W = s \mid c = k) = \frac{\frac{1}{N} \sum_{i=1}^{N} P(c = k \mid y_i) P(W_i = s \mid y_i)}{P(c = k)} \tag{8}
$$

위의 식(8)에서 W는 s열(Row)을 갖는 가장 가능성이 높은 잠재계층 변인이고, c는 k열을 갖는 진정한 잠재계층 변인을 의미한다. $P(W_i = s|y_i)$는 사후소속 확률이 가장 큰 계층을 의미하는데, $P(C = s|y_i) > P(C = k|y_i)$로서 이때 $s \neq k$이며 다른 계층은 0이다.

식(8)에서와 같이, 두 번째 단계에서 나타나는 분류의 불확실성(uncertainty)은 계층 내 공분산 구조의 모수와 관련이 있으며, 분류오류의 크기는 3단계의 과정에 영향을 미칠 수 있는 계층 내 공분산 구조의 모수에 따라 변화할 수 있다(Bakk et al., 2014). 특히, $P(W = s \mid c = k)$는 실제값에 대한 추정된 값의 조건부 확률이며, 분류오류의 전체 비율은 다음과 같이 정의된다(Vermunt, 2010).

$$\sum_{k=1}^{K} P(c = k) \sum_{s \neq k} P(W = s \mid c = k) \tag{9}$$

마지막 단계에서는 잠재계층 변인 W는 분류오류를 고려하면서 다항 로지스틱 회귀분석을 실시하는데, 이는 공변인과 관련이 있다. 두 번째 단계에서 분류오류가 고정되고 공변인과 잠재계층 변인 c의 로지스틱 회귀분석이 이루어진다. 이러한 3단계 접근법은 공변인의 수가 증가하더라도 잠재계층 분류에 영향을 미치지 않기 때문에 잠재계층이 변하지 않는 장점을 지녔다(Vermunt, 2010). 3단계에서 수행되는 다항로지스틱 회귀분석을 위한 마지막 단계가 식(3)과 식(4)에 의해 수행된다.

3. 선행연구

성장혼합모형을 적용하여 분석할 때 계층 분류 과정에서 공변인을 통합하는 방법에 대한 비교연구가 수행되고 있다. 우선 Lubke와 Muthén(2007), Wurpts와 Geiser(2014)는 잠재계층 분석에서 공변인을 포함하는 분석, 즉 1단계 접근법을 수행해야 한다고 제안하고 있다. Wurpts와 Geiser(2014)는 잠재계층 분류 시 공변인의 크기가 계층분류의 확률 및 공변인 편의를 감소시키기 때문에 공변인을 포함하여 분석할 때 잠재계층 분석에 더 유리하다고 보고하였다. 또한, Clark와 Muthén(2009)은 1단계 접근법이 보조변수를 이용한 다른 접근법과 비교하여 공변인 효과를 추정하는 데 가장 우수하다고 하였다. Diallo와 Lu(2017)는 성장혼합모형에서 1단계 접근법이 올바르게 명세화되면 모수복구(parameter recovery)차원에서 3단계 접근법보다 더 우수하다고 보고하였다. 그러나 공변인을 투입하는 경우와 그렇지 않은 경우 모두 집단의 성격이 변화하지 않아야 하며, 잠재계층의 모수치나 소속확률이 변화하면 잠재계층 가정에 위배된다고 보았다(신택수, 2010). 이와 더불어 성장혼합모형에서 공변인을 추가하여 분석할 경우 계층의 평균과 작은 표본일 경우 계층 분류에 도움이 되지만, 표본이 크고 계층 간 거리가 넓을 때 이점이 발생되지 않는다(Hu, Leite, &

Gao, 2017). 이는 공변인과 계층분류를 하나의 단계로 수행하는 1단계 접근법에 대해 보조 변수의 영향력을 개별적이며 추가적인 단계로써 수행하기 위한 접근을 시도하는 연구로 이어지게 되었다(Collier & Leite, 2017). Marsh 등(2009)은 잠재계층 분석 시 공변인을 동시에 투입할 경우 잠재계층의 성격이 왜곡되는 경향을 보이며, 이에 대한 논의가 진행되어야 한다고 제안하고 있다.

한편, Vermunt(2010)는 표준오차 추정에서 공변인 효과 및 편의와 관련하여 1단계 접근법, 기존의 3단계 접근법과 새로운 3단계 접근법 등 여러 접근법의 견고성을 평가하기 위한 모의실험 연구를 수행하였다. 기존의 3단계 접근법은 공변량 없이 잠재집단의 수를 추정하고, 두 번째 단계에서 개인을 잠재집단에 할당한 후 잠재집단 간 공변인 효과를 추정하였다. 새로운 3단계 접근법은 두 번째 단계에서 기존의 3단계 접근법과 차이를 보이는데, 개인을 소속집단에 할당할 뿐만 아니라 잠재집단 오분류의 수정이 이루어진다. 그 결과 기존의 3단계 접근법은 표본크기가 크지 않을 경우 매우 낮은 분리조건, 즉 낮은 엔트로피에서 수행되지 못했고 모수치가 심각하게 하향 조정되었으나, 새로운 3단계 접근법은 모수 추정의 표준편차와 관련된 추정치들이 훨씬 더 효율적으로 추정되었다고 보고하였다. 또한 Asparouhov와 Muthén(2014a)은 기존의 3단계 접근법과 1단계 접근법보다 Vermunt(2010)가 제안한 새로운 3단계 접근법이 모수 추정에 안정적이었다고 보고하였다.

이와 더불어 Kim 등(2016)의 연구에서는 혼합회귀모형의 맥락에서 공변인을 포함하는 모의실험 연구에서 공변인의 직접효과는 1단계 접근법에서 이를 생략하면 상당한 편차가 발생되었지만, 3단계 접근법에서는 계층 열거과정에서 편차가 발생하였다 하더라도 무시할 만한 수준으로 나타났다. 그러나 결측치가 있는 종단자료에서는 1단계와 3단계 접근법의 분류 정확도 차이가 없었으며, 절편과 기울기, 분산, 공분산 등이 편의 없이 추정되었지만, 공변인의 효과 계수 추정치는 계층 간 거리가 좁고, 결측 비율이 높을 때 과소추정되었다(Cetin-Berber & Leite, 2018).

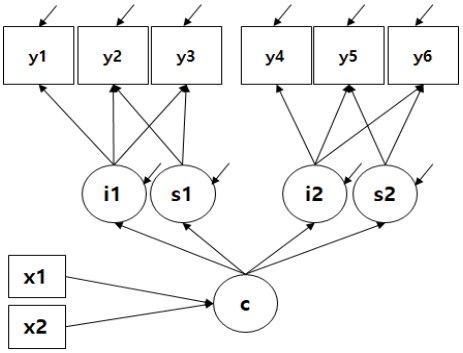
앞서 제시된 선행연구를 종합하여 보면, 계층분류와 공변인 효과의 추정에서 명세화가 잘 될수록, 계층 간 거리와 표본이 작을수록 1단계 접근법이 우수하지만, 계층 간 거리가 크고 표본이 클수록 3단계 접근법이 더 우수한 접근 방법으로 이해할 수 있다. 하지만, 이들의 연구는 잠재계층이나 성장혼합모형에 국한되어 있으며, 비연속 분할함수 성장혼합모형을 적용한 공변량 접근법의 모의실험 연구는 전무한 실정이다. 이에 본 연구에서는 비연속 분할함수 성장혼합모형을 적용하여 1단계 접근법과 3단계 접근법 간 차이를 비교하여 모의실험을 실시하고자 한다.

Ⅲ. 연구 방법

본 모의실험 연구는 비연속 분할함수 성장혼합모형의 계층분류과정에서 공변인 접근법에 따라 공변인 추정치에 어떠한 차이를 나타내는지 이를 검증하고자 한다. 모의실험연구를 수행하기 위한 조건, 분석절차 및 공변인 효과의 추정을 위해 적용되는 1단계 접근법과 3단계 접근법을 적용하고자 한다. 1단계 접근법과 3단계 접근법 간 공변인 추정치의 양호도를 평가하기 위한 준거를 설정하였다.

1. 모의실험 설계

본 모의실험 연구는 종단자료에서 이질적인 성장곡선에 따라 다수의 잠재계층이 도출되는 성장혼합모형에서 성장곡선이 연속적으로 이어지는 것이 아니라 분할시점에서 일정 거리가 분리되어 변화추정의 분할 시점을 제시하는 비연속 분할함수 성장혼합모형을 적용하고자 한다. 이러한 비연속 분할함수 성장혼합모형은 분절된 상황에서 절편과 기울기가 각각 다르게 추정되며, 본 모의실험에서는 5세~7세의 유치원, 1~3학년의 초등학교 저학년, 4~6학년의 초등학교 고학년, 1~3학년의 중·고등학교로 이루어지는 학교급 및 교육과정의 변화를 반영하여 측정시점을 분절된 상황 이전의 3시점, 이후의 3시점으로 총 6회로 설정하였다. 또한, 지표 변수에서 잔차 분산이 계층 간 불변이며, 잠재계층의 수는 2개로 설정하였다(Kohli et al., 2016; Li & Harring, 2017). 모집단 설정을 위하여 사용된 자료는 미국 교육국의 아동종단연구(The Early Childhood Longitudinal Study, Kindergarten Class of 2010-11, ECLS-K:2011)이며, 이는 2010년에 유치원에 입학한 아동을 대상으로 조사되었다. 본 연구에서는 이들의 검사자료에서 수직 동동화(vertical equating)된 수학 IRT 점수를 종속변인으로 설정하고자 하였으며, 모의실험을 위해 고정설계된 모수는 〈표 1〉과 같다. 연구모형은 [그림 2]와 같다.



[그림 2] 연구모형

〈표 1〉 비연속 분할함수 성장혼합모형의 공변인 추정치 모수

		계층1	계층2
고정설계 요소			
	성장모수		
(MD=1.3)	절편1 평균	50.8	36.9
	기울기1 평균	12.1	9.0
	절편2 평균	83.2	59.2
	기울기2 평균	7.6	6.9
(MD=2.6)	절편1 평균	65.0	36.9
	기울기1 평균	15.5	9.0
	절편2 평균	105.8	59.2
	기울기2 평균	8.3	6.9
	절편1 분산	119.9	
	기울기1 분산	15.5	
	절편2 분산	165.0	
	기울기2 분산	5.7	
공분산			
	절편1, 기울기1	7.2	
	절편2	115.0	
	기울기2	−9.3	
	절편2, 기울기1	33.0	
	기울기2	−15.5	
	기울기1, 기울기2	−3.2	
잔차분산			
	Y1	5.9	
	Y2	53.3	
	Y3	24.6	
	Y4	29.1	
	Y5	41.6	
	Y6	8.3	
공변인 회귀계수			
	범주형 변인(OR=2.2)	0.8	
	연속형 변인(OR=5.0)	1.6	

2. 모의실험 조건

모의실험을 위해 설정된 조건은 공변인 간 상관계수, 계층 분류정도, 계층 혼합비율, 표본크기의 4가지이다. 우선 공변인 효과를 추정하기 위하여 범주형 변인과 연속형 변인이 포함되며, 두 변인을 동시에 투입할 경우 두 변인 간 상관관계를 고려해야 하며, 두 공변인 간 상관계수가 .20, .30, .50와 같이 ‘낮은 상관’에서부터 ‘상관 관계가 있음’에 대해 선행 연구가 수행되었다(Diallo et al., 2017; Li & Harring, 2017). 이에 본 연구에서는 실제 자료에서 얻어진 공변인의 효과는 범주형

.8(OR=2.2), 연속형 1.6(OR=5.0)이며, 이들 변인 간 상관계수는 .20로 나타났다. 그러나 실제자료를 사용하여 분석한 선행연구에서의 공변인 간 상관은 .20~.60의 범위를 보이며(안영은, 박세진, 2019; 안태용, 2018; 이정민, 정혜원 2017), 모의실험 조건으로 선정된 .20의 낮은 상관과 비교하여 상대적으로 높은 .60의 상관을 보일 때 어떠한 차이를 보이는지 살펴보고자 한다. 이에 두 변인 간 상관계수 .60을 추가하여 상관계수를 .20과 .60으로 선정하였다. 둘째, 성장혼합모형의 추정 정확도는 하위집단이 얼마나 잘 분리되었는지에 따라 영향을 받는다(Lubke & Muthén, 2005). 이를 위해 자료가 공간으로부터 얼마나 벗어나고 있는지를 계산하는 마할라노비스 거리(Mahalanobis distance, MD)를 이용하였다. 마할라노비스 거리는 두 개의 하위집단 중심 사이의 통계적 거리를 측정하기 때문에 집단 간 이질성을 고려할 수 있다(Peugh & Fan, 2013). 이때 계층분리는 각 성장궤적의 잠재수준에서 발생되며 마할라노비스 거리는 잠재성장 요인, 즉 절편과 기울기를 변화시킨다(Depaoli, 2013). 마할라노비스 거리는 $\sqrt{(\mu_1-\mu_2)'\Psi^{-1}(\mu_1-\mu_2)}$ 이며, μ_1과 μ_2은 두 개의 잠재계층 중 계층1과 계층2에 대한 성장요인의 평균, Ψ^{-1}은 성장모수의 공분산 행렬의 역행렬이다(McLachlan & Peel, 2000). 선행연구에서는 계층 간 분류율을 0.5, 1.0, 1.5, 2.0(Tueller & Lubke, 2010), 1.3, 2.6, 3.9(Hu et al., 2017)으로 설정하여 분석하고 있으며, 본 연구에서는 실제 자료에서 얻어진 마할라노비스 거리는 1.3을 기준으로 배수인 2.6을 모의실험 조건으로 선정하였다. 셋째, 모의실험을 위한 표본크기 내 잠재계층 간 비율을 불균등과 균등의 형태로 구분하기 위하여 극단적인 수준 90:10과 불균등하지만 하나의 잠재계층이 다른 잠재계층보다 작은 비율 70:30, 두 개의 계층 간 균등한 비율로서 50:50의 혼합비율로 조정하였다(Hu et al, 2017; Kohli et al., 2015; Li & Harring, 2017). 넷째, 모의실험 연구를 위한 표본크기는 작은 표본크기의 200과 500(Hu et al., 2017; Vermunt, 2010), 중간 표본크기의 1000이 성장혼합모형에서 가장 보편적으로 수행되며(Clark & Muthén, 2009; Kamata, Kara, Patarapichayatham, & Lan, 2018; Li & Harring, 2017). 큰 표본크기로 2000이 수행되고 있다(Kamata et al., 2018; Hu et al., 2017). 이에 본 연구에서는 모의실험을 수행하기 위한 표본크기로 작은 표본에서부터 큰 표본까지 설정하였으며, 200과 500, 1000, 2000으로 설정하였다. 이와 같이 앞에서 제시된 모의실험 조건은 48가지(2×2×3×4)이며, 〈표 2〉와 같다.

〈표 2〉 모의실험 조건

공변인 간 상관계수	계층 분류정도(MD)	계층 혼합비율	표본크기
.20 .60	1.3 2.6	90:10 70:30 50:50	200 500 1000 2000

3. 모의자료 생성 및 분석 절차

Mplus 8.2 프로그램의 Monte Carlo 시뮬레이션의 3단계 접근법을 적용하여 모의실험을 위한 자료를 생성하였다. 자료생성을 위하여 모의실험의 반복(replications) 수행 횟수는 500회로 설정하였으며, 모의자료 생성과 분석 절차는 다음과 같다(Li & Harring, 2017; Nylund et al., 2007).

[모의자료 생성 절차]

모의실험을 위해 모집단과 잠재계층 성장모수 설정을 위하여 계층 분류정도(1.3, 2.6), 계층 혼합비율(90:10, 70:30, 50:50), 표본크기(200, 500, 1000, 2000), 공변인 간 상관계수(.20, .60)의 조건을 적용하였다. 첫째, 분할시점을 기준으로 두 개의 절편과 기울기의 성장 형태를 지닌 잠재계층을 〈표 1〉에 제시된 성장모수와 공분산을 기초로 모집단과 잠재계층 성장모수 설정하여 모의자료를 생성한다. 둘째, 범주형과 연속형 변인의 효과를 살펴보기 위하여 2.2와 5.0의 승산비와 두 변인 간 '낮은 상관'의 .20를 적용하며, Mplus 프로그램 상에서 3단계 접근법의 보조변수 명령어 R3STEP을 사용하여 범주형과 연속형 공변인을 설정하여 모의자료를 생성한다. 셋째, 계층 2의 절편과 기울기의 값을 고정하고, 계층 1의 절편과 기울기의 성장모수를 마할라노비스 거리를 통해 계산된 거리(1.3, 2.6)를 기준으로 계층 분류정도를 재지정하였으며, 〈표 1〉의 성장모수를 토대로 모의자료를 생성한다. 넷째, 모의자료 생성을 위해 앞서 제시된 조건과 잠재계층별 소속확률도 포함시킨다. 잠재계층 소속확률(90:10, 70:30, 50:50)과 표본크기(200, 500, 1000, 2000)가 함께 고려한다. 예를 들어 전체 모집단의 표본크기가 200이고, 소속확률이 90:10일 경우 계층 1은 전체의 90%, 계층 2는 나머지 10%가 소속된 것으로 정의하여 180과 20으로 할당될 수 있도록 자료 생성의 과정을 수행한다. 다섯째, 모의자료 생성을 위한 조건을 적용하여 500번 반복 수행한다.

[모의자료 분석 절차]

우선 공변인 효과를 추정하기 위하여 모의자료 생성 절차에 따라 생성된 자료를 1단계와 3단계 접근법을 적용하여 500번 반복하여 분석된 결과를 산출한다. 이때 비수렴된 결과는 제외하며, 부족한 자료 수 만큼 추가 생성하여 분석한다. 그 다음으로 비연속 분할함수 성장혼합모형의 계층 분류 과정에서 1단계와 3단계 접근법에 따라 다항로지스틱 회귀분석으로 추정된 범주형과 연속형 변인의 모수 값 추정치를 각각 산출한다. 이후 공변인 효과에 대한 양호도 평가를 위하여 모수 값과 1단계 접근법과 3단계 접근법을 적용하여 산출된 모수추정치를 통해 양호도 값을 산출한 후 평가한다. 이러한 절차가 완료되면 모의자료 생성 절차에서 공변인 간 상관계수를 .60으로 변경하여 모의자료 생성 및 분석절차의 과정을 반복하여 수행한다.

위의 모의자료 생성 절차는 Mplus 8.2 프로그램의 Monte Carlo 시뮬레이션을 이용하였으며, 모

의자료 분석 절차는 Mplus 8.2와 R 3.5.3프로그램에서 MplusAutomation(Ver. 0.7-3) 패키지를 연동하여 분석에 이용하였다.

4. 공변인 추정치의 양호도 평가 준거

모의실험 자료의 분석에서는 모수값 추정을 위해 표본 통계량을 사용하고자 할 경우 표본 통계량의 기댓값이 모수와 같다면 그 통계량은 모수에 대한 비편의 추정치가 발생된다(김주아, 강상진, 2006). 이에 모의실험을 통해 산출된 표집분포의 평균값과 모의자료를 생성할 때 설정한 모수값이 다르다는 것을 나타내는 것이 편의이며, 이는 모수의 추정 값과 실제 값의 차이에 대한 정보를 제공한다.

$$Bias = \bar{\hat{\theta}} - \theta = \frac{1}{n_r}\sum_{r=1}^{n_r}\hat{\theta}_r - \theta \tag{10}$$

위의 식에서 n_r는 500번 반복을 의미하며, $\bar{\hat{\theta}}$는 500번 반복 추정된 모수치의 평균이며, θ는 모의자료를 생성할 때 설정된 모수 값이다. 이 값이 '0'에 가까울수록 비편의된 추정치로 간주된다. 이와 더불어 Hoogland와 Boomsma(1998)가 제안한 상대적 편의는 추정치의 정확성을 평가함으로써 추정치의 평균이 모수 값과 얼마나 일치하는지 상대적으로 확인가능하며, 그 기준은 .05로 절댓값이 .05이상이면 편의된 추정치로 간주된다.

$$Relative\ bias = \frac{\bar{\hat{\theta}} - \theta}{\theta} \tag{11}$$

위의 식에서 $\bar{\hat{\theta}}$는 500번 반복 추정된 모수치의 평균이며, θ는 모의자료를 생성할 때 설정한 모수값이다. 또한, 본 모의실험 연구에서 계층분류의 접근방법을 비교하기 위하여 정확성 및 효율성을 동시에 고려하는 평균제곱오차(MSE)를 사용하고자 한다. 평균제곱오차는 실제 모집단에서 공변인의 추정지가 일관되고 정확하게 추정되는지를 함께 나타내는 지표이다(Muthén & Muthén, 2002). 이때 모수 값의 추정치가 모수로부터 얼마만큼 가까운 값인지를 측정하며 모수에 이웃할수록 평균제곱오차의 값은 작아진다. 이러한 평균제곱오차의 값은 작을수록 추정의 정확성이 높아짐을 의미하고, 아래의 식과 같이 추정치의 편의와 분산이 동시에 고려되며, 추정량 $\hat{\theta}$의 분산, 모수 θ와 추정

량 $E(\hat{\theta})$과의 편차제곱을 의미한다.

$$MSE = E[(\hat{\theta}-\theta)^2] = Var(\hat{\theta}) + (E(\hat{\theta})-\theta)^2 \tag{12}$$

이와 더불어 추정치의 표준오차의 상대적 편의는 추정치의 표준오차의 평균이 모수 값의 표준오차와 얼마나 일치하느냐를 상대적으로 확인 가능한 지표이며, 이때 0.1보다 작을 경우 편의가 거의 나타나지 않는다고 간주된다(Hoogland & Boomsma, 1998).

$$Bias(\widehat{se}_{\hat{\theta}}) = \sum_{r=1}^{n_r} \frac{\overline{\widehat{se}}_{\hat{\theta}_r} - \widehat{se}_{\theta_r}}{\widehat{se}_{\theta_r}} \tag{13}$$

위의 식에서 $\widehat{se}_{\theta_r}$는 모의자료를 생성할 때 설정한 모수 값의 추정된 표준오차이며, $\overline{\widehat{se}}_{\hat{\theta}_r}$는 모수 값의 추정된 표준오차의 평균을 의미한다. 한편, 모의실험에서 연구모형의 분석을 실행했을 경우 분석된 자료가 수렴되지 않음으로써 이상치의 결과로 인한 해석의 오류나 결과의 불안정성을 초래할 수 있다. 이는 수렴의 문제이며, 본 연구에서는 계층 분류과정에서 공변인 접근법의 안정성을 비수렴 비율로 살펴보고자 한다. 특히 비수렴 비율이 높을수록 잠재계층에서 공변인을 예측하는 분석에 적용할 경우 해당 접근법의 안정성이 떨어짐을 의미한다. 비수렴 비율의 식은 아래와 같다.

$$NCR = \frac{n_r - CR}{n_r} \tag{14}$$

위의 수식에서 n_r은 반복 수행된 모의실험 분석 수를 의미하여, CR은 해당 접근법을 사용하여 수렴된 분석의 수를 의미한다. 수렴의 문제와 더불어 혼합모형의 분석에서는 계층 분류과정에서 계층의 전환이 발생하게 되며, 이는 일관성 없는 계층 할당에서 비롯됨을 의미한다(Kohli et al., 2015). 이러한 계층 교차의 가능성을 확인하기 위하여 해당 접근법에서 할당된 k계층을 확인한 후 이와 대비되는 접근법에서 할당된 k계층 간 비교를 통해 계층 분류과정에서 계층의 전환이 발생하였는지 계층 교차 비율을 확인하였다.

Ⅳ. 연구 결과

1. 공변인 추정치의 편의

비연속 분할함수 성장혼합모형의 계층분류 과정에서 1단계 접근법과 3단계 접근법을 적용한 후 공변인의 모수추정치 편의를 살펴본 결과 〈표 3〉과 같다. 두 접근법의 편의 값은 상관계수가 .20의 경우 3단계 접근법이 '0'에 가까운 결과를 보였지만, 두 변인 간 상관계수가 .60에서는 계층 분류정도가 2.6일 때 3단계 접근법이 '0'에 가까운 결과를 나타냈다. 두 가지 접근법에 따라 공변인을 추정한 결과는 범주형 변인과 연속형 변인 일부를 제외하고 3단계 접근법을 적용하였을 경우 추정치가 모수에 더 근접한 것으로 나타났다.

〈표 3〉 공변인 추정치의 편의

MD	계층 혼합 비율	표본 크기	$r = .20$				$r = .60$			
			1단계		3단계		1단계		3단계	
			x1[1]	x2[2]	x1	x2	x1	x2	x1	x2
1.3	90:10	200	0.1486	0.2721	0.1378	−0.0948	−0.0642	0.2552	−0.0369	−0.1736
·	·	500	0.0535	0.0750	0.0085	−0.0627	0.0934	0.0541	0.0236	−0.0613
·	·	1000	0.0251	0.0291	−0.0058	−0.0359	0.0557	0.0195	0.0073	−0.0425
·	·	2000	0.0055	0.0180	−0.0030	−0.0146	0.0152	0.0126	0.0024	−0.0195
·	70:30	200	0.1068	0.1581	−0.0902	−0.1684	0.0316	0.0942	−0.0965	−0.2158
·	·	500	0.0453	0.0495	−0.0046	−0.0537	0.0455	0.0445	0.0018	−0.0746
·	·	1000	0.0181	0.0200	−0.0059	−0.0361	0.0201	0.0138	0.0007	−0.0386
·	·	2000	0.0064	0.0120	−0.0039	−0.0153	0.0027	0.0062	−0.0072	−0.0231
·	50:50	200	0.0101	0.0505	−0.1194	−0.2084	−0.0366	0.0684	−0.1174	−0.2104
·	·	500	0.0242	0.0377	−0.0398	−0.0750	0.0012	0.0476	−0.0376	−0.0816
·	·	1000	0.0111	0.0215	−0.0139	−0.0288	0.0056	0.0132	−0.0084	−0.0359
·	·	2000	0.0043	0.0122	−0.0087	−0.0149	−0.0016	0.0077	−0.0058	−0.0160
2.6	90:10	200	0.0745	0.0824	0.0710	0.0700	0.1022	0.0686	0.1138	0.0584
·	·	500	0.0263	0.0399	0.0218	0.0349	0.0477	0.0333	0.0422	0.0280
·	·	1000	0.0066	0.0188	0.0059	0.0151	0.0186	0.0134	0.0164	0.0107
·	·	2000	0.0062	0.0052	0.0066	0.0029	0.0150	0.0015	0.0151	−0.0002
·	70:30	200	0.0578	0.0614	0.0473	0.0531	0.0384	0.0539	0.0274	0.0408
·	·	500	0.0230	0.0251	0.0199	0.0181	0.0171	0.0177	0.0149	0.0126
·	·	1000	0.0082	0.0089	0.0082	0.0054	0.0062	0.0028	0.0059	0.0008
·	·	2000	0.0071	0.0044	0.0075	0.0029	0.0057	0.0001	0.0051	−0.0011
·	50:50	200	0.0113	0.0606	0.0059	0.0506	0.0177	0.0362	0.0120	0.0252
·	·	500	−0.0020	0.0208	−0.0048	0.0175	−0.0079	0.0224	−0.0114	0.0168
·	·	1000	−0.0028	0.0074	−0.0042	0.0040	−0.0008	0.0040	−0.0019	0.0001
·	·	2000	−0.0002	0.0079	−0.0009	0.0055	−0.0012	0.0044	−0.0019	0.0017

1) x1 : 범주형 변인
2) x2 : 연속형 변인

2. 공변인 추정치의 상대적 편의

〈표 4〉는 공변인 모수추정치의 상대적 편의 결과이다. 이때 평가 준거인 .05를 적용하였으며, 공변인 간 상관계수가 .20일 때 계층 혼합비율이 상이하고, 표본크기가 작을 때 1단계 접근법에서 범주형 변인이 상대적 편의가 발생하였다. 또한, 공변인 간 상관계수가 .60일 때 계층 혼합비율과 상관없이 작은 표본일 때 1단계 접근법에서 추정된 범주형 변인의 상대적 편의가 발생하였다. 범주형 변인과 연속형 변인 간 상관계수가 낮거나 높을 경우 모두 계층의 혼합비율이 하나의 집단으로 과도하게 할당되었을 경우 3단계 접근법과 다르게 1단계 접근법에서 추정된 공변인의 모수추정치는 작은 표본에서 이상 편의가 발생하였다.

〈표 4〉 공변인 추정치의 상대적 편의

MD	계층 혼합 비율	표본 크기	$r = .20$				$r = .60$			
			1단계		3단계		1단계		3단계	
			x1	x2	x1	x2	x1	x2	x1	x2
1.3	90:10	200	**0.1857**	**0.1701**	**0.1722**	**−0.0592**	**−0.0803**	**0.1595**	−0.0461	**−0.1085**
·	·	500	**0.0669**	0.0469	0.0106	−0.0392	**0.1168**	0.0338	0.0295	−0.0383
·	·	1000	0.0313	0.0182	−0.0072	−0.0224	0.0696	0.0122	0.0091	−0.0265
·	·	2000	0.0068	0.0113	−0.0037	−0.0091	0.0190	0.0079	0.0030	−0.0122
·	70:30	200	**0.1335**	**0.0988**	**−0.1127**	**−0.1053**	0.0395	**0.0589**	**−0.1206**	**−0.1349**
·	·	500	**0.0566**	0.0309	−0.0058	−0.0336	**0.0568**	0.0278	0.0023	−0.0466
·	·	1000	0.0226	0.0125	−0.0073	−0.0225	0.0252	0.0086	0.0009	−0.0241
·	·	2000	0.0080	0.0075	−0.0049	−0.0096	0.0034	0.0039	−0.0090	−0.0144
·	50:50	200	0.0127	0.0316	**−0.1492**	**−0.1303**	−0.0458	0.0427	**−0.1468**	**−0.1315**
·	·	500	0.0303	0.0236	−0.0498	−0.0469	0.0015	0.0298	−0.0470	**−0.0510**
·	·	1000	0.0139	0.0134	−0.0173	−0.0180	0.0070	0.0083	−0.0105	−0.0225
·	·	2000	0.0053	0.0076	−0.0108	−0.0093	−0.0020	0.0048	−0.0072	−0.0100
2.6	90:10	200	**0.0931**	**0.0515**	**0.0887**	0.0438	**0.1278**	0.0429	**0.1423**	0.0365
·	·	500	−0.0026	0.0130	−0.0060	0.0109	**0.0596**	0.0208	0.0527	0.0175
·	·	1000	0.0082	0.0117	0.0074	0.0095	0.0232	0.0084	0.0205	0.0067
·	·	2000	0.0077	0.0032	0.0082	0.0018	0.0187	0.0009	0.0189	−0.0001
·	70:30	200	0.0723	0.0384	**0.0591**	0.0332	0.0480	0.0337	0.0343	0.0255
·	·	500	−0.0026	0.0130	−0.0060	0.0109	0.0213	0.0111	0.0186	0.0079
·	·	1000	0.0103	0.0056	0.0103	0.0034	0.0077	0.0018	0.0073	0.0005
·	·	2000	0.0089	0.0027	0.0094	0.0018	0.0071	0.0001	0.0063	−0.0007
·	50:50	200	0.0141	0.0379	0.0074	0.0316	0.0221	0.0226	0.0150	0.0158
·	·	500	−0.0026	0.0130	−0.0060	0.0109	−0.0099	0.0140	−0.0142	0.0105
·	·	1000	−0.0035	0.0046	−0.0052	0.0025	−0.0010	0.0025	−0.0024	0.0000
·	·	2000	−0.0002	0.0050	−0.0011	0.0035	−0.0015	0.0028	−0.0024	0.0011

1) x1 : 범주형 변인
2) x2 : 연속형 변인

3. 공변인 추정치의 평균제곱오차

〈표 5〉는 공변인 모수추정치의 평균제곱오차의 결과이다. 1단계 접근법과 3단계 접근법을 적용한 결과 범주형 변인의 경우 계층 분류정도가 작은 경우 3단계 접근법에서 평균제곱오차 값이 작게 나타났지만, 계층 분류정도가 큰 경우 1단계 접근법에서 더 작은 값을 나타냈다. 하지만, 1단계 접근법과 3단계 접근법 간 평균제곱오차는 미미한 차이를 보였다. 연속형 변인의 경우 소표본을 제외하고, 모든 조건에서 1단계 접근법이 더 작은 평균제곱오차를 보였다.

〈표 5〉 공변인 추정치의 평균제곱오차

MD	계층 혼합 비율	표본 크기	$r = .20$				$r = .60$			
			1단계		3단계		1단계		3단계	
			x1	x2	x1	x2	x1	x2	x1	x2
1.3	90:10	200	1.2867	1.4938	3.4791	1.7127	1.0678	5.0522	3.2225	0.4637
·	·	500	0.2525	0.1441	0.2240	0.1290	0.4104	0.1304	0.3349	0.1615
·	·	1000	0.1064	0.0512	0.1095	0.0599	0.1612	0.0594	0.1677	0.0772
·	·	2000	0.0521	0.0247	0.0572	0.0318	0.0674	0.0277	0.0716	0.0386
·	70:30	200	1.0977	2.4538	0.2575	0.2499	0.7232	0.5240	0.3230	0.2910
·	·	500	0.1437	0.0745	0.1105	0.0745	0.1875	0.0867	0.1318	0.0843
·	·	1000	0.0614	0.0350	0.0512	0.0384	0.0757	0.0366	0.0576	0.0432
·	·	2000	0.0283	0.0166	0.0285	0.0179	0.0362	0.0177	0.0307	0.0200
·	50:50	200	0.4419	0.3402	0.2038	0.2042	0.4612	0.3955	0.2614	0.2591
·	·	500	0.1307	0.0723	0.1005	0.0678	0.1565	0.0788	0.0948	0.0707
·	·	1000	0.0597	0.0326	0.0487	0.0335	0.0611	0.0357	0.0437	0.0363
·	·	2000	0.0270	0.0147	0.0236	0.0153	0.0311	0.0184	0.0245	0.0204
2.6	90:10	200	0.3425	0.1426	0.3407	0.1468	0.4756	0.1429	0.5577	0.1394
·	·	500	0.1080	0.0508	0.1088	0.0507	0.1663	0.0499	0.1700	0.0531
·	·	1000	0.0524	0.0240	0.0545	0.0249	0.0737	0.0276	0.0766	0.0295
·	·	2000	0.0273	0.0121	0.0289	0.0124	0.0344	0.0125	0.0358	0.0132
·	70:30	200	0.1734	0.0772	0.1737	0.0757	0.2134	0.0940	0.2090	0.0902
·	·	500	0.0581	0.0285	0.0574	0.0285	0.0743	0.0328	0.0733	0.0333
·	·	1000	0.0281	0.0150	0.0287	0.0151	0.0315	0.0162	0.0324	0.0168
·	·	2000	0.0139	0.0071	0.0143	0.0075	0.0167	0.0083	0.0169	0.0087
·	50:50	200	0.1231	0.0760	0.1234	0.0745	0.1439	0.0880	0.1400	0.0894
·	·	500	0.0532	0.0257	0.0531	0.0268	0.0585	0.0280	0.0568	0.0293
·	·	1000	0.0259	0.0134	0.0261	0.0137	0.0257	0.0148	0.0253	0.0152
·	·	2000	0.0110	0.0061	0.0112	0.0063	0.0120	0.0075	0.0121	0.0076

1) x1 : 범주형 변인
2) x2 : 연속형 변인

4. 공변인 추정치의 표준오차의 상대적 편의

〈표 6〉은 공변인 모수추정치의 표준오차의 상대적 편의 결과이며, 평가준거로 0.1을 적용하였다. 1단계 접근법과 3단계 접근법 간 표준오차의 상대적 편의 값은 상관계수와 계층 분류정도가 작고, 계층 혼합비율이 하나의 계층으로 치우친 경우 작은 표본에서 연속형 변인에서 표준오차의 상대적 편의가 발생하는 것으로 나타났다. 이와 더불어 상관계수가 상대적으로 높은 조건일 때 1단계 접근법 모두 범주형 변인에서 표준오차의 상대적 편의가 나타났다. 또한, 계층 분류정도가 1.3일 때 모든 조건에서 3단계 접근법에서 더 작은 표준오차의 상대적 편의 추정치를 보였다.

〈표 6〉 공변인 추정치의 표준오차의 상대적 편의

MD	계층 혼합 비율	표본 크기	$r = .20$				$r = .60$			
			1단계		3단계		1단계		3단계	
			x1	x2	x1	x2	x1	x2	x1	x2
1.3	90:10	200	**0.2196**	**0.3687**	**0.1688**	**0.4041**	**0.2803**	**0.4149**	**0.1427**	**0.2362**
·	·	500	0.0661	**0.1328**	0.0465	0.0901	**0.2850**	0.0865	**0.1061**	0.0869
·	·	1000	0.0158	0.0315	0.0169	0.0356	0.0296	0.0267	0.0330	0.0342
·	·	2000	0.0055	0.0148	0.0074	0.0180	0.0105	0.0124	0.0133	0.0169
·	70:30	200	**0.2081**	**0.3136**	0.0390	0.1207	**0.1785**	**0.2184**	0.0498	0.1114
·	·	500	0.0357	0.0820	0.0124	0.0449	0.0421	0.0713	0.0154	0.0416
·	·	1000	0.0094	0.0238	0.0059	0.0226	0.0131	0.0237	0.0065	0.0204
·	·	2000	0.0034	0.0103	0.0028	0.0099	0.0050	0.0099	0.0030	0.0097
·	50:50	200	**0.1393**	**0.2168**	0.0179	0.0858	**0.1326**	**0.4109**	0.0219	0.1014
·	·	500	0.0395	0.0951	0.0089	0.0395	0.0499	0.0900	0.0066	0.0337
·	·	1000	0.0092	0.0248	0.0043	0.0196	0.0135	0.0252	0.0039	0.0175
·	·	2000	0.0033	0.0103	0.0021	0.0092	0.0046	0.0104	0.0020	0.0098
2.6	90:10	200	0.0240	0.0557	0.0235	0.0605	0.0422	0.0393	**0.1806**	0.0399
·	·	500	0.0058	0.0153	0.0061	0.0185	0.0124	0.0130	0.0128	0.0158
·	·	1000	0.0028	0.0072	0.0031	0.0093	0.0050	0.0067	0.0055	0.0083
·	·	2000	0.0014	0.0036	0.0015	0.0045	0.0026	0.0032	0.0028	0.0039
·	70:30	200	0.0058	0.0252	0.0051	0.0267	0.0094	0.0249	0.0075	0.0273
·	·	500	0.0016	0.0082	0.0017	0.0095	0.0023	0.0086	0.0024	0.0101
·	·	1000	0.0008	0.0042	0.0009	0.0048	0.0012	0.0044	0.0012	0.0051
·	·	2000	0.0004	0.0020	0.0004	0.0024	0.0006	0.0022	0.0006	0.0025
·	50:50	200	0.0037	0.0216	0.0035	0.0248	0.0040	0.0223	0.0033	0.0248
·	·	500	0.0011	0.0078	0.0012	0.0092	0.0012	0.0069	0.0012	0.0089
·	·	1000	0.0005	0.0041	0.0006	0.0048	0.0006	0.0039	0.0006	0.0048
·	·	2000	0.0002	0.0019	0.0003	0.0023	0.0003	0.0021	0.0003	0.0025

1) x1 : 범주형 변인
2) x2 : 연속형 변인

5. 공변인 추정치의 비수렴 비율

〈표 7〉은 공변인 모수추정치의 비수렴 비율 결과이다. 1단계 접근법과 3단계 접근법을 적용하였을 때 작은 표본에서 비수렴 비율이 약 60%로 매우 높게 나타났다. 하지만, 계층 분류정도와 계층 혼합비율이 높아지고, 표본크기가 커질수록 두 접근법 모두 비수렴 비율이 낮아졌다. 이러한 결과는 작은 표본일 경우 이상치의 결과로 해석의 오류의 문제가 발생될 수 있다. 특히, 계층 분류정도가 1.3이고, 모든 계층 혼합비율의 500에서 2000까지 1단계 접근법 보다 3단계 접근법을 이용한 분석이 더 높은 비수렴 비율을 나타냈다. 이러한 비수렴 비율은 두 접근법 모두 작을 표본일 때 매우 크게 나타났으며, 계층 혼합비율과 표본 크기가 커질수록 비수렴 비율이 점차 낮아지는 결과를 보였다. 이와 더불어 계층 분류정도가 2.6으로 높아지고 계층 혼합비율이 90:10일 때, 200에서 2000까지 조건에서 3단계 접근법보다 1단계 접근법을 이용한 분석이 더 높은 비수렴 비율을 보였지만, 두 접근법의 적용에 따른 비수렴 비율의 차이는 매우 미미하게 나타났다.

〈표 7〉 공변인 추정치의 비수렴수와 비수렴 비율

MD	계층 혼합 비율	표본 크기	$r = .20$				$r = .60$			
			1단계		3단계		1단계		3단계	
			n	NCR	n	NCR	n	NCR	n	NCR
1.3	90:10	200	307	0.614	321	0.642	327	0.654	308	0.616
·	·	500	163	0.326	165	0.330	154	0.308	158	0.316
·	·	1000	57	0.114	62	0.124	55	0.110	68	0.136
·	·	2000	13	0.026	17	0.034	12	0.024	16	0.032
·	70:30	200	325	0.650	330	0.660	330	0.660	326	0.652
·	·	500	158	0.316	176	0.352	171	0.342	176	0.352
·	·	1000	57	0.114	65	0.130	56	0.112	69	0.138
·	·	2000	13	0.026	12	0.024	9	0.018	16	0.032
·	50:50	200	320	0.640	332	0.664	340	0.680	332	0.664
·	·	500	164	0.328	173	0.346	169	0.338	176	0.352
·	·	1000	54	0.108	66	0.132	58	0.116	66	0.132
·	·	2000	16	0.032	16	0.032	15	0.030	16	0.032
2.6	90:10	200	287	0.574	269	0.538	279	0.558	261	0.522
·	·	500	144	0.288	143	0.286	143	0.286	140	0.280
·	·	1000	51	0.102	55	0.110	52	0.104	54	0.108
·	·	2000	11	0.022	10	0.020	10	0.020	9	0.018
·	70:30	200	289	0.578	270	0.540	290	0.580	277	0.554
·	·	500	146	0.292	140	0.280	153	0.306	147	0.294
·	·	1000	49	0.098	52	0.104	50	0.100	49	0.098
·	·	2000	7	0.014	10	0.020	10	0.020	10	0.020
·	50:50	200	290	0.580	273	0.546	288	0.576	273	0.546
·	·	500	158	0.316	149	0.298	154	0.308	148	0.296
·	·	1000	48	0.096	49	0.098	48	0.096	52	0.104
·	·	2000	8	0.016	12	0.024	8	0.016	9	0.018

6. 계층 교차이동 비율

〈표 8〉은 계층 교차이동 비율의 결과이다. 계층의 교차이동 비율을 3단계 기준으로 설정하였으며, 절편과 기울기가 큰 집단(계층 1)에서 절편과 기울기가 작은 집단(계층 2) 간 교차된 수와 비율이다. 우선 상관계수가 .20이며 계층 분류정도가 1.3일 때 계층 2에서 계층1로 교차하는 비율이 매우 크게 나타났다. 특히 계층 혼합비율이 90:10과 같이 극단적일 경우 모든 표본에서 약 20%의 계층 교차 비율을 보였다. 하지만 계층 혼합비율이 50:50으로 동등할 때 모든 표본에서 계층 교차 비율이 절반에 가까운 비율이 낮아졌다. 또한, 계층 분류정도가 2.6일 때 모든 조건에서는 계층 2에서 계층 1로, 계층 1에서 계층 2로 이동의 수가 1~5정도로 매우 작은 수가 이동하는 것으로 나타났으며, 이러한 계층 교차의 비율이 현저하게 낮아지는 것을 보였다.

한편, 상관계수가 .60이며 계층 분류정도가 1.3일 때 계층 2에서 계층1로 교차하는 비율이 매우 크게 나타났다. 특히 계층 혼합비율이 90:10과 같이 극단적일 경우 모든 표본에서 약 16%의 계층 교차 비율을 보였다. 하지만 계층 혼합비율이 50:50으로 동등할 때 모든 표본에서 계층1에서 계층2로 계층 2에서 계층 1로 계층이 교차하는 비율이 약 11%로 매우 낮아지는 결과를 보였다. 계층 분류정도가 2.6일 때 모든 조건에서는 계층 2에서 계층 1로, 계층 1에서 계층 2로 이동의 수가 1~5정도로 매우 작은 수가 이동하는 것으로 나타났으며, 이러한 계층 교차의 비율이 현저하게 낮아지는 것을 보였다. 이러한 결과에서와 같이 계층 교차이동 비율은 계층 분류정도와 계층 혼합비율에서 더 크게 나타날 수 있어 결과의 해석에 유의할 필요가 있다.

〈표 8〉 계층 교차이동 비율(3단계 기준)

MD	계층 혼합 비율	표본 크기	$r = .20$				$r = .60$			
			계층 교차수(n)		계층 교차비율(%)		계층 교차수(n)		계층 교차비율(%)	
			C1→C2	C2→C1	C1→C2	C2→C1	C1→C2	C2→C1	C1→C2	C2→C1
1.3	90:10	200	7.344	6.166	4.187	19.003	7.316	5.500	4.216	16.219
·	·	500	13.712	10.232	3.113	14.888	14.838	10.624	3.403	14.863
·	·	1000	25.884	18.362	2.946	14.131	28.252	19.454	3.244	14.158
·	·	2000	51.286	33.710	2.918	13.454	54.802	36.620	3.153	13.582
·	70:30	200	10.598	9.212	7.396	14.689	9.688	8.730	6.871	13.442
·	·	500	18.986	15.300	5.298	10.397	20.680	16.132	5.810	10.700
·	·	1000	35.354	29.566	4.960	10.113	37.852	30.678	5.368	10.214
·	·	2000	69.524	56.476	4.883	9.714	73.888	59.510	5.253	9.933
·	50:50	200	11.440	10.544	9.749	11.821	9.696	10.084	8.359	11.108
·	·	500	20.486	18.954	7.088	8.727	21.246	19.746	7.425	9.004
·	·	1000	37.210	34.776	6.489	8.044	38.376	36.412	6.735	8.365
·	·	2000	72.056	67.372	6.303	7.826	75.664	70.654	6.661	8.140

2.6	90:10	200	0.418	0.388	0.243	1.376	0.438	0.396	0.257	1.287
·	·	500	0.952	0.898	0.221	1.254	1.044	0.972	0.245	1.278
·	·	1000	1.880	1.796	0.219	1.260	1.970	1.810	0.232	1.192
·	·	2000	3.748	3.314	0.219	1.163	3.976	3.576	0.234	1.174
·	70:30	200	0.522	0.528	0.374	0.849	0.592	0.602	0.428	0.964
·	·	500	1.282	1.210	0.368	0.796	1.324	1.270	0.383	0.813
·	·	1000	2.552	2.270	0.365	0.749	2.646	2.398	0.383	0.772
·	·	2000	4.910	4.428	0.352	0.731	5.144	4.754	0.373	0.764
·	50:50	200	0.596	0.584	0.528	0.670	0.610	0.606	0.539	0.690
·	·	500	1.356	1.310	0.476	0.600	1.442	1.424	0.510	0.648
·	·	1000	2.674	2.596	0.471	0.598	2.822	2.770	0.500	0.634
·	·	2000	5.162	5.206	0.456	0.600	5.522	5.414	0.491	0.619

V. 논의 및 결론

본 연구는 비연속 분할함수 성장혼합모형에서 계층 분류과정에서 적용한 공변량 접근법에 따라 공변인의 추정치에 어떠한 차이가 있는지를 비교하는데 목적이 있다. 이에 공변량 접근법으로 사용되고 있는 1단계 접근법과 3단계 접근법에 따라 공변인 효과의 차이를 확인하기 위한 모의실험 연구를 수행하였다. 본 모의실험 연구결과를 고려하여 각 조건을 비교한 결과와 논의는 다음과 같다.

1단계 접근법과 3단계 접근법에서 공변인 간 관련성 검증을 위한 모의실험 결과 두 접근법이 모수추정치에 얼마나 근접하는지를 평가하는 양호도에서 차이가 있었다. 이러한 양호도를 평가하는 준거로 편의, 상대적 편의, 평균제곱오차, 표준오차의 상대적 편의, 비수렴수와 비수렴 비율이 사용되었으며 두 접근법에 따른 공변인 효과의 모수추정치는 일부 조건을 제외하고 모의실험 조건에서 양호하게 추정되었다. 양호도 평가 준거에 따른 평가 결과는 계층 간 거리가 좁고, 계층의 혼합비율이 하나의 계층으로 과도하게 분류되었을 때, 표본크기가 작을수록 1단계 접근법과 3단계 접근법에서 편의가 있었음을 본 연구에서 확인되었다.

우선, 1단계 접근법과 3단계 접근법에 따른 공변인 효과의 편의는 일부 조건을 제외하고 모수추정치에 근접한 결과를 보였다. 하지만, 일부조건, 즉 소표본이면서 극단적인 계층 혼합비율의 조건일 때 1단계 접근법에서 모수추정치가 과도하게 추정되었다. 범주형 변인의 조건에서는 계층 분류정도가 커질수록 3단계 접근법보다는 1단계 접근법을 적용했을 때 모수추정치의 편의 값이 작아지는 경향을 보였다. 연속형 변인 조건에서는 계층 분류정도가 커질수록 1단계 접근법보다 3단계 접근법에서 공변인 효과의 모수추정치가 더 작은 값을 보였고 편의가 나타나지 않았다. 이는 잠재집단 간 거리가 멀수록 3단계 접근법에서 모수추정치가 모수에 더 근접하게 추정되었으며, 연속형 변인은

잠재집단 간 거리에 따라 모수추정치가 더 정확해졌다는 Luke 와 Muthén(2005)의 연구결과와 일치했다. 이는 잠재집단의 통계적 거리가 멀어짐이 곧 두 집단의 이질성 강화를 의미하기 때문으로 추측된다.

한편, 상대적 편의를 절댓값 .05기준으로 적용한 결과(Hoogland & Boomsma, 1998), 소표본이면서 극단적인 계층의 혼합비율일 때 범주형 변인 조건에서 상대적 편의가 발생하였다. 이는 범주형 변인과 연속형 변인 간 상관이 낮거나 높은 경우 상대적 편의가 발생하였으며, 계층 혼합비율이 하나의 계층에 과도하게 집중되는 조건일 때 상대적 편의가 발생하는 것으로 이해할 수 있다. 한편, 평균제곱오차의 경우 소표본을 제외하고 모든 조건에서 1단계 접근법이 더 작은 평균제곱오차의 값이 나타났지만, 두 접근법 간 미미한 차이를 보였다. 이는 올바르게 명세화 될 경우 1단계 접근법이 3단계 접근법보다 더 모수에 근접한다는 선행연구 결과(Diallo & Lu, 2017)의 일부를 지지하는 결과이다. 즉 잠재계층의 성격이 변화한 상태에서 공변인의 효과를 추정할 경우 해석의 오류로 이어질 수 있으며, 이러한 이유로 인하여 1단계 접근법을 적용할 경우 왜곡된 경향이 나타날 수 있기에 주의가 필요하다(Marsh et al., 2009).

다음으로 표준오차의 상대적 편의에 0.1기준을 적용하여 살펴본 결과(Hoogland & Boomsma, 1998), 일부 조건에서 표준오차의 상대적 편의가 나타났으며, 소표본이며 계층 혼합비율이 극단적일 때 연속형 변인의 조건에서 표준오차의 상대적 편의가 발견되었다. 이와 같은 양호도 평가 준거 이외에도 이상치의 결과 해석으로부터 발생할 수 있는 문제를 비수렴율을 통해 확인하였다. 본 모의실험 연구에서 비수렴 비율은 표본이 작을수록 비수렴 비율이 높아지는 경향을 보였지만, 작은 표본에서 비수렴율은 계층 간 분류정도가 증가할수록 비수렴율이 감소하는 경향을 보였다. 또한 표본크기가 커지고, 공변인 간 상관계수가 높은 경우 비수렴율이 낮아졌다. 이처럼 소표본의 경우 비수렴 비율이 매우 높지만, 2000 이상의 대규모 자료에서는 비수렴율이 낮아지기 때문에 이를 고려한 해석이 필요하다. 특히 소표본의 경우 비수렴의 비율이 매우 높기 때문에 해석에 오류가 발생할 수 있다.

마지막으로 3단계 접근법의 계층이 1단계 접근법에서 결과를 예측할 경우 이들 두 단계에서 계층 교차이동이 빈번하게 발생하였다. 이러한 계층의 이동 비율은 극단적인 계층 혼합비율을 보였을 때, 계층 분류 기준에 상당한 영향력이 있음을 의미하며, 계층의 성격이 변화할 수 있음을 시사한다. 이와 더불어 범주형 변인과 연속형 변인 간 상관이 상대적으로 높은 경우 계층 분류정도 1.3, 2.6과 계층 혼합비율 90:10, 70:30 50:50과 같이 모든 조건에서 비수렴율이 낮아지는 것을 확인하였으며, 두 변인 간 상관계수가 비수렴율과 관련이 있음을 유추할 수 있다.

이상의 연구의 결과를 종합하면, 비연속 분할함수 성장혼합모형에서 공변인 효과를 추정하기 위해서는 1단계 접근법보다는 3단계 접근법이 적용되어야 할 것이다. 이는 잠재계층과 성장혼합모형을 적용하여 공변량 접근법을 살펴본 기존의 모의실험 연구와 동일한 결과이다. 특히 앞서 언급한 것과 같이 소표본인 경우 두 접근법 모두 공변인 효과가 이상치를 보일 수 있어 주의가 필요하다.

본 연구는 교육학 분야의 연구에서는 학업성취의 교육 자료들이 종단자료의 특성을 지니고 있으며 (Gall et al., 2003), 이들 자료는 하나의 성장 변화를 갖는 것이 아니라 유치원, 초등, 중등의 학교급 단위에 의해 종단적 변화 특성을 내재하고 있기에 이를 반영한 통계적 모형을 설정하였다. 또한, 이들은 동일한 특성을 지닌 하나의 집단이 아닌 이질적인 특성을 지닌 다수의 집단을 내포하고 있기 때문에 이러한 교육 자료들이 갖는 종단적 특성과 이질적 집단의 특성을 반영한 성장모형의 적용과 집단에 영향을 미치는 공변인을 검증하였다. 이에 다양한 모의실험 조건을 설정하여 모의실험 연구를 수행하였다는 점에서 의의가 있으며, 이를 바탕으로 후속연구를 위한 제언은 다음과 같다.

첫째, 본 연구에서는 공변인을 시간불변 공변인으로 투입하여 모의실험을 연구를 수행하였다. 하지만, 시간에 따라 공변인도 변화하기 때문에 공변인의 투입시점을 반영한 연구가 진행되어야 할 것이다. 둘째, 본 연구에서는 공변인 효과를 검증하기 위하여 범주형과 연속형 변인을 각각 하나의 변인만 투입하여 연구를 수행하였다는 점에서 제한점을 지니고 있다. 실제 많은 연구자들은 복수의 공변인을 투입하여 연구를 수행하고 있다. 이에 따라 후속연구에서는 이러한 현실을 반영하여 다수의 공변인을 포함한 연구를 수행함으로써, 이들 공변인과 잠재계층 간 관계의 변화를 살펴보는 구체적인 연구가 이루어져야 할 것이다. 셋째, 본 연구에서 공변인 접근법에 따라 분류된 잠재계층의 계층 교차 이동비율이 90:10과 같이 극단적인 계층 혼합비율로 전체 계층 분류의 약 15%정도 다른 계층으로 이동하는 것으로 나타났다. 이에 따라 후속연구에서는 계층 혼합비율을 다양하게 적용하여 이동된 계층의 특성을 파악하는 연구가 수행되어야 할 것이다.

참고문헌

김주아, 강상진(2006). 학교수, 학생수 및 집단내 상관 크기에 따른 다층모형의 고정효과와 집단내 상관 추정치의 정확도. **교육평가연구, 19**(3), 89-106.

신택수(2010). 종단프로파일분석과 군집분석을 이용한 잠재집단연구. **교육평가연구, 23**(3), 641-664.

안영은, 박세진(2019). 혁신학교와 일반학교 중학생의 전학년 학업성취도 변화 양상 비교: 경기도 혁신학교를 중심으로. **교육학연구, 57**(1), 81-104.

안태용(2018). 학교폭력 피해경험 잠재계층, 예측 및 결과변인 탐색: 성장혼합 모형 검증. **초등상담연구, 17**(1), 81-101.

여승수, 박소희(2012). 잠재성장모형(Latent Growth Modeling)분석의 활용: 교육과정중심측정 종단 자료를 중심으로. **아시아교육연구, 13**(4), 247-273.

이정민, 정혜원(2017). 성장혼합모형을 적용한 초기 청소년의 공동체의식 유형분류 및 영향요인 검증. **교육방법연구, 29**(2), 257-283.

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 329-341.

Asparouhov, T., & Muthén, B. (2014b). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes, 21*(2), 1-22.

Bakk, Z., Oberski, D. L. & Vermunt, J. K. (2014). Relating latent class assignments to external variables: Standard errors for correct inference. *Political Analysis, 22*(4), 520-540.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology, 43*(1), 272-311.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338-363.

Berlin, K. S., Parra, G. R., & Williams, N. A. (2013). An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology, 39*(2), 188-203.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*(1), 3-27.

Cetin-Berber, D. D., & Leite, W. L. (2018). A comparison of one-Step and three-step approaches for including covariates in the shared parameter growth mixture model. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(4), 588-599.

Chou, C. P., Yang, D., Pentz, M. A., & Hser, Y. I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis, 46*(2), 213-225.

Clark, S., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis.* Retrieved April 5, 2018, from http://www.statmodel.com/download/relatinglca.pdf.

Cudeck, R., & Codd, C. L. (2012). A template for describing individual differences in longitudinal data, with application to the connection between learning and ability. In J. R. Harring & G. R. Hancock (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 3-24). Charlotte, NC: Information Age.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association, 83*(401), 173-178.

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods, 18*(2), 186-219.

Diallo, T. M., & Lu, H. (2017). On the application of the three-step approach to growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(5), 714-732.

Diallo, T. M., Morin, A. J., & Lu, H. (2017). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods, 22*(1), 166-190.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Education research: An introduction* (7th Eds). Boston, MA: A & B Publications.

Heagerty, P., Liang, K. Y., Zeger, S., & Diggle, P. (2002). *Analysis of longitudinal data.* New York, NY: Oxford University Press.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*(3), 329-367.

Hu, J., Leite, W. L., & Gao, M. (2017). An evaluation of the use of covariates to assist in class

enumeration in linear growth mixture modeling. *Behavior Research Methods, 49*(3), 1179-1190.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kamata, A., Kara, Y., Patarapichayatham, C., & Lan, P. (2018). Evaluation of analysis approaches for latent class analysis with auxiliary linear growth model. *Frontiers in Psychology, 9*, 130.

Kim, M., Vermunt, J., Bakk, Z., Jaki, T., & Van Horn, M. L. (2016). Modeling predictors of latent classes in regression mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(4), 601-614.

Kim, S. Y. (2011). *Issues in using stage-sequential growth mixture models*. Unpublished doctoral dissertation, University of Wisconsin-Madison, USA.

Kohli, N., Harring, J. R., & Zopluoglu, C. (2016). A finite mixture of nonlinear random coefficient models for continuous repeated measures data. *Psychometrika, 81*(3), 851-880.

Kohli, N., Hughes, J., Wang, C., Zopluoglu, C., & Davison, M. L. (2015). Fitting a linear-linear piecewise growth mixture model with unknown knots: A comparison of two common approaches to inference. *Psychological Methods, 20*(2), 259-275.

Li, L., & Hser, Y. I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research, 46*(2), 266-302.

Li, M., & Harring, J. R. (2017). Investigating approaches to estimating covariate effects in growth mixture modeling: A simulation study. *Educational and Psychological Measurement, 77*(5), 766-791.

Lubke, G. H., & Muthén, B. O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*(1), 21-39.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14*(1), 26-47.

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person-and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(2), 191-225.

McCarthy, D. E., Piasecki, T. M., Fiore, M. C., & Baker, T. B. (2006). Life before and after

quitting smoking: An electronic diary study. *Journal of Abnormal Psychology, 115*(3), 454-466.

McLachlan, G., & Peel, D. (2000). Multivariate normal mixtures. Finite Mixture Models(pp. 81-116). New York, NY:Wiley.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599-620.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535-569.

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 782-797.

Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 616-639.

Singer, J. D., Willett, J. B. (2003). Modeling change using covariance structure analysis. In Commager, H. S.(Ed.), *Applied longitudinal data analysis: modeling change and event occurrence* (pp. 266-302). New York, NY: Oxford University Press.

Sitthisan, C. (2016). *The bayesian method of estimation for the number of latent classes in growth mixture models.* Unpublished doctoral dissertation, University of Northern Colorado, USA.

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling: A Multidisciplinary Journal, 17*(2), 165-192.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18*(4), 450-469.

Wang, M., & Bodner, T. E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods, 10*(4), 635-656.

Wang, M., & Hanges, P. J. (2011). Latent class procedures: Applications to organizational research. *Organizational Research Methods, 14*(1), 24-31.

Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in Psychology, 5*, 920.

* 논문접수 2020년 2월 10일 / 1차 심사 2020년 3월 10일 / 게재승인 2020년 3월 18일

* 이정민: 충남대학교 교육학과 교육평가 전공으로 박사학위를 취득하였으며, 주요연구 분야는 성장혼합모형, 다층모형, 종단자료 분석 등이다.
* E-mail: nori99@cnu.ac.kr

* 정혜원: 충남대학교 교육학과 교수로 재직 중이며, 주요연구 분야는 고급다층모형, 성장모형, 종단자료 연구 등이다.
* E-mail: chw7@cnu.ac.kr

Abstract

Comparing Covariate Approaches of the Classification Process in Discontinuous Piecewise Growth Mixture Model*

Lee, Jungmin**
Chung, Hyewon***

This paper was to investigate the covariate approaches of the classification process in discontinuous piecewise growth mixture model. Two covariate approaches were compared with the aid of a Monte Carlo simulation study for a one-step approach and a three-step approach. Simulation conditions for estimating covariate effects were outlined: covariate correlation coefficients (.20, .60), the class separation (1.3, 2.6), the mixing proportions (90:10, 70:30, 50:50), the sample size (200, 500, 1000, 2000). In the present simulation study, Outcome measures were determined by bias, relative bias, mean square error, bias of standard error, non-convergence rates, and class label switching across the 500 replications. Under these conditions, The results of the study showed that the parameter estimates of the covariate effects of the one-step and three-step approaches were good except for some conditions. In addition, the bias occurred when the distance of class was narrow, the mixing ratio was excessively classified into one class, and the sample size was small. In this case, we confirmed that the results of outliers can be seen in both one-step and three-step approach. Thus, this research was meaningful to compare one-step approach and three-step approach within the simulation study using discontinuous piecewise growth mixture model. Implications and following studies were suggested.

Key words: Discontinuous Growth Mixture Model, Covariate Approach, One-step Approach, Three-step Approach

* This paper is revised from a part of the doctoral dissertation of the author
** First author, Graduate Doctor, Chungnam National University
*** Corresponding author, Professor, Chungnam National University