

실험자료 분석 시 무선효과 구조 선택을 위한 우도비 검정에서의 영가설 분포의 모양

김 용 기 이 우 열[†]

충북대학교 심리학과

실험자료 분석을 위해 다층모형을 사용하는 경우, 무선효과 구조 설정에 따라 가설 검정의 결과가 달라질 수 있다. 선행연구들에 따르면, 무선효과 구조를 선택하는 방법 중 우도비 검정(likelihood-ratio test; LRT)은 분산의 모수 공간 제약으로 인해 영가설 분포의 모양이 표준적인 카이제곱 분포와 일치하지 않는다. 본 연구에서는 몬테카를로 시뮬레이션을 통해 무선효과 구조를 검정하기 위한 LRT 영가설 분포의 모양, 1종 오류 비율, 수렴 실패율에 대해 알아 보았다. 본 연구를 진행함에 앞서, 고정효과를 검정하기 위한 LRT 영가설 분포의 모양을 확인함으로써 본 연구방법의 타당성을 확보하였다. 실험설계와 참가자 수, 문항 수를 조작하여 시뮬레이션 실험을 수행한 결과, 무선효과 구조를 검정하기 위한 LRT 영가설 분포의 평균, 분산, 95분위수와 99분위수는 모든 조건에서 표준적인 카이제곱 분포보다 작았다. 표준적인 카이제곱 분포를 사용하여 LRT를 수행했을 때 1종 오류 비율 또한 유의수준인 .05보다 작았다. 실험설계가 복잡하고 무선효과 구조에 상관 구조를 포함한 모형에서 추정 과정에서 수렴에 실패가 빈번하게 발생하였다. 경험적 연구자들이 실험조건 효과 검정을 위해 다층모형을 사용할 때 모형의 무선효과 구조 선택 시 주의할 점을 논의하였다.

주요어 : 다층모형, 우도비 검정, 무선효과 구조, 수렴 실패율, 영가설 분포

[†] 교신저자: 이우열, 충북대학교 심리학과 부교수, (28644) 충북 청주시 서원구 충대로 1, 충북대학교 사회과학대학 심리학과, E-mail: wooyeollee@chungbuk.ac.kr



Copyright © 2025, The Korean Psychological Association. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial Licenses(<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

심리학 연구에서 참가자가 군집에 내재된 자료나 종단연구로부터 수집된 자료, 또는 참가자가 실험조건의 수준에 걸쳐 반복 측정된 자료(참가자 내 실험설계)는 흔하게 수집된다. 다층모형은 이처럼 위계적인 구조를 가진 자료를 분석하기 위해 사용될 수 있다. 실제로, 다층모형은 심리학 연구 자료를 분석하기 위해 점차 널리 사용되고 있다(Eckardt et al., 2021; Hoffman & Walters, 2022; Luo et al., 2021; Meteyard & Davies, 2020).

다층모형은 분석 모형의 선택, 추정 방법의 선택, 가설 검정 방법의 선택 등의 각 절차에서 연구자의 의사결정이 필요하다. 각 절차에서 내리는 연구자의 선택이 연구의 주요 가설을 검정하는 데 영향을 준다. 따라서 경험적 연구자가 적절한 의사결정을 내리기 위한 다층모형의 절차에 대한 연구가 필요하다. 본 연구에서는 다층모형의 절차 중 하나인 무선효과 구조 선택에 초점을 맞추어, 실험자료를 분석하기 위해 다층모형을 사용하는 경우에서 무선효과 구조를 선택하는 방법을 조사하였다.

다층모형에서 무선효과의 구조

다층모형은 선형회귀 모형의 확장된 형태로 고정효과와 무선효과를 포함한 통계 모형이다. 개별 참가자를 상위 수준(between-level)으로, 참가자의 반복된 반응들을 하위 수준(within-level)으로 가정할 때, 고정효과는 일반적으로 자료 수집 단위에서 상위 수준 개체들의 평균적인 효과를 가리킨다. 무선효과는 각 개체들이 갖는 고유한 효과이며 특정 확률분포를 따른다고 가정된다. 예를 들어, 심리학 실험에서 참가자 내 설계로부터 수집된 자료는 종속변인의 개별 반응이 참가자 또는 문항에 내재된

구조를 이루고 있다. 이 자료에 다층모형을 적용했을 때, 고정효과는 예측변인 또는 실험조건의 각 수준에서 평균적인 참가자와 문항에서 기대되는 효과를 가리킨다. 개별 참가자 또는 문항이 종속변인에 미치는 효과는 각각 참가자 무선효과와 문항 무선효과에 해당한다.

선행연구에서 심리학 실험 자료 분석을 위해서 다층모형을 사용하는 경우, 모형에 참가자 무선효과와 문항 무선효과를 모두 포함하는 것이 제안되었다(Baayen et al., 2008; Barr, 2013; Judd et al., 2012). 그 이유는 참가자가 종속변인에 영향을 미치는 것과 마찬가지로, 개별 문항 역시 종속변인에 영향을 미칠 수 있기 때문이다. 문항효과(item effect)는 실험조건의 각 수준의 효과를 측정하기 위해 포함된 문항들이 종속변인에 미치는 영향을 가리킨다. 언어 심리학 분야에서는 문항효과가 실험조건 효과에 대한 가설 검정 결과에 영향을 미친다는 것이 오래전부터 알려져 왔다(Clark, 1973). 문항효과를 무선효과로 포함하는 것은 마치 참가자가 모집단에서 표집되는 것과 같은 논리로 문항이 문항 모집단(item population)에서 표집된다는 가정에 기초한다. 참가자 무선효과와 문항 무선효과를 모형에 포함할 때 관행적으로 이 둘이 독립적임을 가정하는데, 그 이유는 전형적인 참가자 내/문항 내 설계(within-participant and within-item design)에서 참가자와 문항이 종속변인의 개별 반응에 교차하여 걸려있기 때문이다. 이러한 다층모형을 교차 분류 다층모형(cross-classified multilevel model)이라 한다(Rasbash & Goldstein, 1994).

다층모형은 무선효과 구조에 따라 크게 무선절편 모형(random-intercept only model)과 무선기울기 모형(random slope model)으로 나누어진다¹⁾(Snijders & Bosker, 2011). 여기서 무선효과

구조란 각 무선효과들이 갖는 공분산 구조를 의미한다. 무선절편 모형은 상위 수준 개체가 오직 모형의 절편에 대한 무선효과만 갖는 모형이다. 이때 무선효과는 상위 수준 개체의 전체적인 반응 경향성을 의미한다. 무선기울기 모형은 상위 수준의 개체가 모형의 절편뿐 아니라 기울기에 대해서도 무선효과를 갖는 모형이다. 이때 무선효과는 개별 예측변인 효과에 대한 개인차를 의미한다. 무선기울기 모형은 두 개 이상의 무선효과가 다변량 확률분포를 따른다고 가정한다. 무선기울기 모형의 무선효과 구조가 다변량 정규분포를 따른다고 가정될 때, 연구자는 무선효과의 분산-공분산 구조(variance-covariance structure)를 설정할 수 있다. R의 lme4 패키지(Bates et al., 2015a)의 lmer 함수는 무선기울기 모형에 대해 무선효과 사이의 모든 상관계수를 추정하거나(이하, 상관 구조[correlated structure]), 무선효과 사이의 모든 상관계수를 0으로 고정하고 무선효과의 분산만을 추정할 수 있다(이하, 영상관 구조[zero-correlated structure]).

예를 들어, 세 개의 수준을 갖는 독립변인 한 개가 있고 문항과 참가자가 독립변인의 모든 수준에 걸쳐 반복되는 실험설계에서, 다층 모형과 그 분산-공분산 구조는 다음과 같다. 문장 읽기 시간에 대한 글씨체의 효과를 알아보기 위한 실험 연구에서, 한 집단의 참가자가 명조체, 고딕체, 굴림체 조건 각각에 대해

문장 읽기 과제를 수행하였다고 가정해 보자. 각 참가자는 실험조건에 걸쳐 글씨체를 달리 한 같은 문장들에 반복해서 응답하게 된다. 문항으로 사용한 문장은 어떤 문제 은행에서 연구자가 임의로 선택한 것으로 볼 수 있기 때문에, 문항 모집단에서 추출된 것으로 간주할 수 있다.

이때 각 실험조건에서의 평균 반응시간이 고정효과에 해당한다. 기저선 수준(reference level)인 명조체 조건에서의 평균 반응시간이 절편이고, 명조체 조건의 평균 반응시간(즉, 절편)과 다른 글씨체 조건의 평균 반응시간의 차이값이 기울기가 된다. 문항 무선효과는 개별 문항의 효과, 즉 실험조건 및 참가자에 따른 각 문항의 평균 반응시간과 전체 평균 반응시간의 차이값이다. 참가자 무선효과는 고정효과 대비 각 참가자의 효과, 즉 참가자의 실험조건별 평균 반응시간에서 그에 대응되는 고정효과를 뺀 값이다.

무선효과의 분산은 개별 참가자와 문항이 가진 효과의 변산성을 의미한다. 참가자 무선절편의 분산이 크다면, 참가자에 따라서 명조체 조건에서의 평균 반응시간의 차이가 큼을 의미한다. 그리고 무선효과의 공분산은 무선효과 사이의 선형적 관계의 강도를 의미한다. 만약 참가자 무선절편과 고딕체 효과와의 공분산이 크다면, 명조체 조건에서 평균 반응시간이 큰 참가자가 고딕체와의 반응시간 차이 또한 큼을 의미한다. 만약 무선효과 구조로 영상관 구조를 설정한다면, 연구자는 무선효과 사이의 관계를 가정하지 않음을 의미한다.

위에 서술한 바를 다음과 같이 수식으로 나타낼 수 있다. 식 (1)은 참가자 $j(= 1, 2, \dots, J)$ 가 실험조건의 수준 $k(=1, 2, 3)$ 에 제시된 문항 $i(= 1, 2, \dots, I)$ 에 반응한 종속변인 Y_{ijk} 를 예

1) 익명의 심사위원이 지적한 것처럼, 일부 문헌에서는 다층모형의 종류로 무조건 모형(unconditional model 또는 null model)을 포함한다(예를 들어, Huang, 2022). 이 모형에는 예측변인이 포함되어 있지 않다. 독립변인의 효과 검정을 주요 목표로 하는 실험 연구에서는 무조건 모형이 거의 사용되지 않으므로 본 원고에서는 이 모형을 다루지 않는다.

측하기 위해 참가자 무선효과에 무선기울기를 포함하고 문항 무선효과에 무선절편이 포함된 다층모형의 수식을 나타낸다.

$$Y_{ijk} = \beta_0 + P_{0j} + (\beta_1 + P_{1j})X_{1ijk} + (\beta_2 + P_{2j})X_{2ijk} + S_i + e_{ijk} \quad (1)$$

식 (1)은 종속변인 Y_{ijk} 가 고정효과 $\beta_0, \beta_1, \beta_2$, 참가자 무선효과 P_{0j}, P_{1j}, P_{2j} , 문항 무선효과 S_i , 잔차 e_{ijk} 의 선형조합으로 예측된다는 것을 보여준다. 고정효과는 독립변인에 대한 코딩스킴(coding scheme)에 따라 해석이 달라진다. 더미코딩(dummy coding)을 사용한다면 β_0 는 기저선 수준의 종속변인 평균을, β_1, β_2 는 기저선 수준과 비교되는 수준 사이의 평균 차이를 나타낸다. 문항 무선효과 S_i 와 잔차 e_{ijk} 는 각각 식 (2)와 식 (3)과 같이 평균이 0인 정규분포를 따른다고 가정된다.

$$S_i \sim N(0, \omega^2) \quad (2)$$

$$e_{ijk} \sim N(0, \sigma^2) \quad (3)$$

참가자 무선효과 P_{0j}, P_{1j}, P_{2j} 는 식 (4)와 같이 각 효과의 평균이 0인 다변량 정규분포를 따른다고 가정된다.

$$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \end{bmatrix} \sim MVN \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & \\ \tau_{10} & \tau_1^2 & \\ \tau_{20} & \tau_{21} & \tau_2^2 \end{bmatrix} \right) \quad (4)$$

참가자 무선효과와 분산-공분산 구조의 모수 중 $\tau_0^2, \tau_1^2, \tau_2^2$ 는 분산을, $\tau_{10}, \tau_{20}, \tau_{21}$ 은 공분산을 나타낸다. 앞에서 언급한 무선효과 구

조는 식 (4)의 공분산 행렬의 설정에 따라 달라진다. 상관 구조는 여섯 개의 분산, 공분산 모수를 제약 없이 모두 추정한다. 영상관 구조는 공분산 모수를 모두 0으로 제약하고 분산 모수인 $\tau_0^2, \tau_1^2, \tau_2^2$ 만을 추정한다. 그리고 무선절편 구조는 τ_0^2 를 제외한 나머지 모수를 모두 0으로 제약한 것과 같다.

무선효과 구조를 결정하는 방법

경험적 연구자의 연구문제가 고정효과에 관한 것이어도, 다층모형을 사용할 때 분석 모형의 무선효과 구조를 결정하는 것은 중요하다. 무선효과 구조가 고정효과의 표준오차 추정과 가설 검정 결과에 영향을 미칠 수 있기 때문이다(Gurka et al., 2011; Kwok et al., 2007; Luo & Kwok, 2009). 분석 모형의 무선효과가 실제보다 지나치게 단순하면 표준오차는 작게 추정되고, 고정효과에 대한 가설 검정에서 1종 오류의 가능성이 증가하게 된다(Baayen et al., 2008; Judd et al., 2012). 반대로 무선효과 구조가 지나치게 복잡하면 검정력이 감소할 수 있다(Shin & Lee, 2022).

분석 모형의 무선효과 구조를 설정할 때, 모형의 적합성(fit)과 간명성(parsimony) 사이의 균형이 연구자에게 요구된다. Barr 등(2013)은 주어진 실험설계에서 가능한 한 최대한으로 복잡한 무선효과 구조를 포함하는 모형의 사용을 제안하였다. 그러나 이 연구자들이 근거로 제시한 실험설계는 현실 세계에 비해 지나치게 단순하다는 문제가 있다(Bates et al., 2015b). 무선효과 구조를 최대한 설정하면 실험설계가 복잡해질수록 추정해야 할 모수의 수가 급격히 증가한다. 예를 들어, 2×2 반복측정 요인설계에서 참가자 무선효과 구조를 최대한 설

정하면, 네 개의 참가자 무선효과에 대해 네 개의 분산과 여섯 개의 공분산 모수를 갖는다(Lee, 2020). 모형이 실제보다 지나치게 복잡하면 추정 과정에서 수렴 실패가 빈번히 발생하고(Park et al., 2020), 고정효과의 검정력이 낮아진다(Shin & Lee, 2022).

모형이 지나치게 복잡해지는 것을 방지하기 위해, 무선효과 구조가 다른 여러 모형을 비교함으로써 결정된 최적의 모형을 분석 모형으로 선택할 수 있다. 이때 최적의 모형을 선택하는 지표는 추정 방법에 따라 달라진다. 예를 들어, 베이지안 추정을 사용했을 때는 베이즈 요인(bayes factor; Jeffreys, 1961)과 DIC(deviance information criterion; Spiegelhalter et al., 2002) 같은 방법이 사용된다. 심리학 연구 맥락에서는 자료에 다층모형을 적합하기 위해 최대우도 추정이 빈번히 사용된다(Luo et al., 2021). 최대우도 추정법에서는 무선효과 구조를 결정하기 위해 AIC(Akaike information criterion; Akaike, 1974), BIC(Bayesian information criterion; Schwarz, 1978)과 같은 정보기준(information criteria)을 참고하거나, 우도비 검정(likelihood ratio test, 이하 LRT; Wilks, 1938)과 같은 가설 검정 방법이 사용될 수 있다.

Park 등(2020)은 심리학 실험연구 맥락에서 다층모형의 무선효과 구조를 선택하기 위해 사용되는 방법의 수행을 시뮬레이션 방법으로 조사하였다. 연구자들은 1, 2, 3개의 예측변인을 시뮬레이션 조건으로 설정하고 각 조건에서 무선절편 모형, 상관 구조 모형, 영상관 구조 모형으로부터 자료를 생성하였다. 교차한 총 아홉 개의 조건에서 생성된 자료에 무선절편 모형, 상관 구조 모형, 영상관 구조 모형을 적합하여 AIC와 BIC가 자료를 생성하는 데 사용된 참 모형을 선택하는 비율을 조사하였다.

그 결과, 참 모형의 무선효과 구조가 단순한 무선절편 구조, 영상관 구조일 때 AIC와 BIC는 적어도 85% 이상의 비율로 참 모형을 선택하였다. 반면 참 모형이 상관 구조일 때, AIC와 BIC 모두 실제보다 간단한 무선효과 구조 모형을 선택하는 비율이 참 모형이 무선절편 구조, 영상관 구조일 때보다 증가하였다. 특히 BIC에서 예측변인의 수가 증가할수록 그러한 경향성이 두드러졌다. 이 때문에 예측변인의 수가 3개인 조건에서 BIC가 참 모형인 상관 구조 모형을 올바르게 선택하는 비율은 0%였다.

무선효과 구조를 결정하기 위한 LRT

LRT는 서로 포함관계인 두 모형을 비교할 때 사용된다. LRT는 간단한 모형을 영가설 모형으로, 복잡한 모형을 연구 모형으로 설정하는 영가설 유의성 검정 방법이다. 참 모형이 영가설 모형일 때 두 모형에서 산출된 이탈도(deviance)의 차이가 검정통계량이 된다. LRT는 검정통계량이 두 모형의 모수의 수 차이를 자유도로 갖는 카이제곱(χ^2) 분포를 따른다고 가정하고 통계적 유의성을 검정한다.

무선절편 구조, 영상관 구조, 상관 구조는 서로 포함관계이므로 무선효과 구조를 선택하기 위해 LRT를 사용할 수 있다. 그러나 무선효과 구조에 대한 가설을 검정하기 위한 LRT 검정통계량이 모수의 수 차이를 자유도로 갖는 표준적인 카이제곱 분포를 따르지 않는다는 연구들이 존재한다(Self & Liang, 1987; Stram & Lee, 1994). 그 이유는 무선효과 구조에 대한 가설을 검정할 때 검정 대상이 되는 모수가 분산인 것과 관련이 있다. 검정하려는 모수가 분산일 때는 모수 공간(parameter space)이 실수

범위 전체가 아니라 0부터 양의 무한대의 범위로 제한된다. 이 때문에 분산 모수를 포함한 LRT 영가설 분포는 표준적인 카이제곱 분포보다 0쪽으로 끌어당겨진 모양을 갖는다. 이 경우, 실제로는 연구 모형이 참 모형에 비해 적합도가 통계적으로 유의하게 높지만, 0쪽으로 끌어당겨진 표준적인 카이제곱 분포로 인해 연구자가 이탈도 차이의 통계적 유의성을 발견하지 못할 수 있다. 이로 인해 연구자가 무선효과 구조를 잘못 설정하게 되면, 고정효과의 가설 검정시 1종 오류 가능성이 증가할 수 있다.

이러한 현상에 대응하기 위해 몇몇 연구자들은 무선효과의 분산 및 공분산 모수 검정을 위한 영가설 분포로 표준적인 카이제곱 분포를 사용하는 것 대신, 자유도가 서로 다른 두 카이제곱 분포의 혼합분포(mixture distribution)를 사용하는 것을 제안하였다(Molenberghs & Verbeke, 2007; Stram & Lee, 1994). 혼합분포는 두 개 이상의 확률분포가 섞여 있는 분포를 가리킨다. 검정 대상인 모수에는 분산, 공분산 모수가 섞여 있고, 그중 분산 모수에만 모수 공간 제약이 있으므로, 영가설 분포의 자유도는 전체 모수의 수 차이와 공분산 모수의 수 차이를 갖는 카이제곱 분포를 혼합하여 생성할 수 있다는 논리에 기반한다. 예를 들어, 하나의 공분산과 하나의 분산을 검정하기 위해서는 $\chi^2(1)$ 과 $\chi^2(2)$ 를 0.5:0.5로 섞은 혼합분포가 영가설 분포로 제안된다. Shapiro(1985)는 이를 확장하여 검정해야 할 분산 모수의 수가 k 일 때, 자유도가 (전체 모수의 수 차이)-(공분산의 수 차이)부터 전체 모수의 수 차이까지의 카이제곱 분포의 혼합비율이 이항분포 $\binom{k}{x} 2^{-k}$ 를 따름을 제안하였다. 그러나 시뮬레이션을

통해 LRT 영가설 분포와 혼합 카이제곱 분포를 비교한 결과, 두 분포는 유사하지 않았다(Crainiceanu & Ruppert, 2004).

심리학 실험자료 맥락에서 무선효과 구조 선택 과정에서의 논의사항

LRT를 통해 다층모형의 무선효과를 결정하는 방법에 관한 선행연구들은 다음과 같은 이유로 심리학 연구 맥락에 적용하기엔 충분하지 못할 가능성이 있다. 첫째, 선행연구들(Molenberghs & Verbeke, 2007; Self & Liang, 1987; Stram & Lee, 1994)은 심리학에서 자주 사용되는 다층모형의 형태에 대해 다루지 않았다. 심리학에서 다층모형은 교차 무선효과를 포함하는 형태로 널리 쓰이고 있다. Baayen 등(2008)은 하나의 실험조건 안에서 다수의 참가자가 다수의 실험자극에 걸쳐 반응한 종속변인을 수집한 상황에서, 모형에 참가자 무선효과와 자극 무선효과를 모두 포함할 것을 제안하였다. 이렇게 참가자로부터 같은 특질을 측정하는 여러 변인을 수집하는 것은 심리학 연구에서 매우 흔하게 발견된다. 그러나 선행연구들은 이러한 교차 무선효과에 대해서 고려하지 않았다. 교차 무선효과를 지닌 다층모형에 대한 선행연구(Baayen et al., 2008; Matuschek et al., 2017)들은 무선효과 구조를 결정하기 위해 LRT를 사용할 것을 언급하였지만, LRT 영가설 분포가 표준적인 카이제곱 분포와 어떻게 다르고 이것이 무선효과 구조 결정에 대한 의사결정에 어떤 영향을 미치는지 서술하지 않았다.

둘째, 선행연구들은(Molenberghs & Verbeke, 2007; Self & Liang, 1987; Stram & Lee, 1994) 복잡한 무선효과 구조에 대해 고려하지 않았다.

심리학 실험자료에서 무선효과 구조는 선행연구에서 조사된 것보다 더 쉽게 복잡해진다. 범주형 예측변인이 무선기술기 모형에 포함될 때, 범주의 수가 증가할수록 무선효과의 개수도 증가한다. 게다가 요인설계처럼 모형에 복수의 범주형 예측변인이 포함되고 상호작용효과도 고려된다면, 무선효과의 개수는 기하급수적으로 증가한다. Stoel 등(2006)은 잠재성장모형의 맥락에서 공분산 구조의 가설검정을 위한 LRT 영가설 분포에 대해 논의하였다. 연구자들은 시뮬레이션 방법으로 영가설 분포를 생성할 것을 제안하였고, 무선효과 구조를 4×4 공분산 행렬까지 확장하여 자신들이 제안한 방법을 적용하였다. 그러나 세 수준 이상의 독립변인을 포함하거나 세 개 이상의 독립변인을 포함한 실험설계에서는 공분산 행렬의 크기가 이보다 더 쉽게 커질 수 있다.

셋째, 선행연구들은(Molenberghs & Verbeke, 2007; Self & Liang, 1987; Stram & Lee, 1994) 무선효과 구조가 복잡해짐에 따라 발생할 수 있는 다층모형의 수렴 실패 증가에 대해 알아보지 않았다. Park 등(2020)은 심리학 실험연구 맥락에 적용되는 다층모형의 경우에, 무선효과 구조가 복잡해지면 수렴 실패율이 증가할 가능성을 제기하였다. 해당 시뮬레이션에서 가장 복잡한 모형(3개의 예측변인과 상관 구조를 포함)에서는 생성 모형과 자료에 적합한 모형이 일치하는 경우에도 0.045의 비율로 수렴 실패가 발생하였다. 이와 같은 수렴 실패율의 증가는 LRT 영가설 분포의 모양을 왜곡할 가능성이 있다. 만약 수렴에 실패한 시행에서의 검정통계량의 분포가 나머지 시행에서의 다른다면, 수렴에 실패한 시행과 성공한 시행으로 이루어진 전체 시행에서의 영가설 분포는 왜곡될 수 있다.

심리학에서 무선효과 구조 선택에서의 관행

통계학에서 무선효과 구조 검정을 위한 LRT의 수행을 조사한 연구들이 존재하는 것에 비해, 심리학에서는 이 주제에 대한 논의가 거의 이루어지고 있지 않다. Meteyard와 Davies(2020)는 2013-2016년에 출판된 심리학, 언어학, 신경과학 논문들을 대상으로 다층모형 사용 시의 관행에 대해 조사하였다. 무선효과 구조 선택에서의 관행 가운데 Barr 등(2013)이 제안한 최대 모형(상관 구조 모형)을 사용한 연구($n=86$)를 제외하면, LRT를 사용한 연구의 수가 가장 많았다($n=25$). 그러나 연구자들은 이들 연구에서 사용된 LRT 영가설 분포를 별도로 보고하지 않았고, 영가설 분포 설정에 대한 제안 또한 언급하지 않았다. 이는 경험적 연구자들 사이에서 LRT 영가설 분포가 표준적인 카이제곱 분포와 달라질 수 있음이 널리 알려지지 않았을 가능성을 보여준다.

현재 다층모형을 위해 널리 사용되고 있는 통계 프로그램은 무선효과 구조를 비교하기 위한 LRT에 앞서 언급한 영가설 분포의 특징을 고려하지 않고 있다. R의 lme4 패키지(Bates et al., 2015a)에서 anova 함수를 사용하여 두 모형을 비교할 때는 모수의 종류와 상관없이 표준적인 카이제곱 분포를 영가설 분포로 사용하여 p 값을 산출한다. 이는 무선효과 구조에 대한 가설을 검정하기 위해 jamovi의 GAMLj3 모듈(Gallucci, 2024)에서 LRT를 사용하는 경우에도 마찬가지이다. 그러므로 심리학 연구자들은 통계 프로그램에서 LRT로 무선효과 구조를 결정하려 할 때, LRT 영가설 분포가 표준적인 카이제곱 분포와 다르다는 것을 고려하지 못한 채 분석을 수행했을 수 있다. 최근

몇몇 R 패키지에서는 다층모형에서 무선효과 구조에 대한 가설을 검정하기 위한 함수를 제공하고 있지만(예: RLRsim 패키지[Scheipl & Bolker, 2022], varTestnlme 패키지[Baey & Khun, 2023]), 심리학 연구자들이 활용하기에는 기능이 제한적이다. 예를 들어, RLRsim 패키지에서는 exactLRT, exactRLRT 함수 모두 단 하나의 분산 모수에 대해서만 검정할 수 있다. 그리고 varTestnlme 패키지의 varCompTest 함수는 교차분류 다층모형을 위해서는 사용할 수 없다.

연구 문제와 가설

선행연구를 종합해 보면 무선효과 구조를 결정하기 위한 LRT 영가설 분포가 표준적인 카이제곱 분포를 따르지 않는다는 것을 알 수 있다. 이 때문에 실험자료를 분석하고자 하는 경험적 연구자는 무선효과 구조를 잘못 선택하게 될 수 있다. 그러나 심리학 연구의 맥락에서 LRT 영가설 분포의 형태를 조사한 연구는 아직 수행되지 않았다. 따라서 본 연구는 다층모형에서 무선효과 구조 선택을 위한 LRT 영가설 분포의 모양을 조사하고, 분포의 모양이 무선효과 구조에 대한 의사결정에 어떤 영향을 미치는지 알아보기 위해 수행되었다. 연구 문제를 알아보기 위해 몬테카를로 시뮬레이션 방법을 사용하여 영가설 분포를 생성한 다음, 이를 표준적인 카이제곱 분포와 비교하였다. 또한, 유의수준 0.05에서의 1종 오류 비율을 구하여 LRT의 수행을 평가하였다. 연구의 타당성을 확보하기 위해 본 연구에 앞서 예비연구에서 고정효과를 검정하는 LRT의 영가설 분포를 확인하고, 이를 표준적인 카이제곱 분포와 비교하였다.

본 연구에서는 무선효과 구조의 복잡성을 조작하여 무선효과의 개수가 증가할 때, LRT 영가설 분포가 어떻게 변화하는지에 관심을 두었다. 또한, Park 등(2020)은 자료의 실제 무선효과 구조에 비해 모형의 무선효과 구조가 더 복잡할 때 수렴 실패율이 높아진다고 보고하였다. LRT 영가설 분포를 생성할 때, 자료 생성 모형인 참 모형의 무선효과 구조가 경쟁 모형보다 언제나 더 단순하기 때문에 수렴 실패율이 높아질 수 있다. 그러므로 본 연구에서도 수렴 실패율을 조사하였다. Meteyard와 Davies(2020)는 표본의 크기가 작으면 수렴에 문제가 생길 가능성이 크다고 예측하였다. 작은 표본에서 자료에 모형을 적합하기 위한 가능한 해(solution)를 찾는 것이 어려울 수 있기 때문이다. 본 연구에서는 수렴 실패율이 자료의 크기에 영향을 받는지 알아보기 위해 참가자 수와 문항 수를 조작하였다.

본 연구의 가설은 다음과 같다. 첫째, 모든 실험설계에서, 무선효과 구조를 검정하는 LRT 영가설 분포는 자유도가 모수의 수 차이인 카이제곱 분포에 비해 평균, 분산, 95분위수, 99분위수가 더 작을 것이다. 둘째, 관행적으로 사용하는 표준적인 카이제곱 분포에 근거하여 영가설 검정을 할 경우, 1종 오류 비율은 유의수준에 비해 더 작을 것이다. 셋째, 실험설계가 복잡할수록 복잡한 무선효과 구조의 모형을 적합할 때의 수렴 실패율이 증가할 것이다. 넷째, 수렴 실패율은 참가자 수와 문항 수가 커질수록 줄어들 것이다.

방 법

본 연구는 무선효과 구조를 선택하기 위한

LRT 영가설 분포의 모양(평균, 분산, 95분위수, 99분위수)을 확인하고, 이 분포가 가설 검정에 어떠한 영향을 미치는지(1종 오류 비율) 알아보기 위해 수행되었다. 또한, 실험설계의 복잡성과 참가자 수, 문항 수에 따른 수렴 실패율의 변화를 조사하였다. 본 연구를 위한 자료를 생성하는 방법으로는 몬테카를로 시뮬레이션이 사용되었다.

본 연구의 절차는 다음과 같다. 첫째, 참 모형과 경쟁 모형을 설정한다. 둘째, 참 모형으로부터 자료를 생성한다. 셋째, 생성된 자료에 참 모형과 경쟁 모형을 각각 적합하고 두 모형의 로그우도(log-likelihood)로부터 LRT 검정통계량인 이탈도(deviance)의 차이를 구한다. 넷째, 앞의 세 절차를 충분한 수만큼 반복하여 수집된 검정통계량의 분포를 확인한다.

한편, 본 연구방법의 타당성을 확보하기 위해 예비 연구가 수행되었다. 시뮬레이션 방법으로 생성된 표집분포에 대해 이론적으로 기대되는 분포가 존재할 때, 생성된 표집분포의 모양이 이론적인 분포와 일치하는지 확인하였다. 고정효과를 검정하기 위한 LRT 영가설 분포의 모양은 모든 실험설계에서 표준적인 카이제곱 분포와 유사하였다. 예비 연구의 절차와 결과는 보충자료 1에서 제시되었다.

예비 연구와 본 연구에서 자료를 생성하고 모형을 적합하는 과정에 R(4.4.0 버전; R Core Team, 2024)의 lme4 패키지(Bates et al., 2015a)의 lmer 함수가 사용되었다. 그리고 자료 생성, 모형 적합, 자료 분석을 위한 R 코드가 보충자료 2에 제시되었다. 보충자료 1과 2는 교신저자의 OSF 저장소(<https://osf.io/4t5sp/>)에서 내려받을 수 있다.

참 모형과 자료 생성

본 연구의 참 모형은 상호작용효과를 포함한 실험설계의 모든 효과가 고정효과로 포함되고, 참가자 무선효과와 문항 무선효과가 모두 무선절편 구조인 다층모형이었다(부록 1 참조). 자료 생성을 위한 실험설계는 참가자와 문항이 실험조건의 수준에 걸쳐 반복되는 반복측정 설계였다. 실험조건의 수준은 더미코딩을 사용하여 선행식으로 표현되었다. 본 연구의 참 모형은 다음의 구성요소로 이루어졌다. 첫째, 고정효과로는 절편만이 포함되었다. 즉, 실험조건의 수준에 걸친 평균 반응만을 고정효과로 포함하고, 모든 독립변인의 효과(기울기)는 0으로 설정하였다. 둘째, 참가자 무선효과 구조는 독립변인 수준의 효과에 대한 무선기울기를 포함하고, 무선효과 사이의 상관 구조 없이 분산만을 모수로 포함하였다. 셋째, 문항 무선효과는 무선절편 구조를 가정하였다.

참 모형의 모수는 다음과 같이 설정되었다. 각각의 고정효과()는 0으로, 모든 참가자 무선효과의 분산(σ_k^2 ; k =실험조건 수준과 상호작용효과의 수)은 0.5로, 문항 무선효과의 분산(σ^2)은 0.5로, 잔차의 분산(σ^2)은 1로 각각 설정되었다. 참가자 수와 문항 수를 30으로 사용한 전형적인 반복측정 요인분석에서 중간 크기의 효과크기(Cohen's $d=0.5$)를 갖는 고정효과가 0.8의 검정력으로 탐지될 수 있다고 알려져 있다(Westfall et al., 2014). 그러므로 참가자와 문항의 수는 시뮬레이션 조건에 따라 각각 30 및 60으로 설정되었다.

경쟁 모형과 시뮬레이션 조건

참 모형과 함께 참가자 무선효과의 구조만 다른 두 종류(영상관 구조, 상관 구조)의 경쟁 모형을 무선키울기 모형으로 설정하였다. 영상관 구조는 참가자 무선효과의 분산만을 고려한 무선효과 구조이다. 상관 구조는 참가자 무선효과의 분산뿐만 아니라 참가자 무선효과 간의 공분산까지 고려한 무선효과 구조이다. 시뮬레이션의 조건에 따른 참 모형과 경쟁 모형의 수식은 부록 1에, 참 모형과 경쟁 모형의 무선효과 구조 및 잔차 분포는 부록 2에 제시되었다.

조작된 시뮬레이션 조건은 실험설계, 참가자 수, 문항 수였다. 실험설계는 ‘단일변인 두 수준’, ‘단일변인 세 수준’, ‘2×2 요인설계’, ‘2×3 요인설계’의 네 수준으로 조작되었다. 참가자와 문항이 모든 실험조건의 수준에서 반복되었다. 참가자 수와 문항 수의 수준은 30과 60의 두 수준으로 각각 설정되었다. 시뮬레이션 조건의 수준들은 완전히 교차하여 총 시뮬레이션 조건의 수는 4(실험설계)×2(참가자 수)×2(문항 수)=16이었다.

자료분석의 방법 및 준거

참 모형으로부터 시뮬레이션 조건별로 5,000개의 자료를 생성한 다음, 해당 자료에 참 모형과 영상관 구조 모형, 상관 구조 모형을 적합하였다. lmer 함수로 설정된 조건별 분석 모형은 부록 3에 제시되어 있다. 추정을 위해 기본 설정인 nloptwrap 최적화 알고리즘과 전체 정보 최대우도 방법(full information maximum likelihood, FIML)을 사용하였다. 자료에 모형이 적합된 후 참 모형과 영상관 구조 모형 사이의 이탈도, 참 모형과 상관 구조 모형 사이의 이탈도가 각각 계산되었다.

본 연구에서 LRT의 수행은 영가설 분포의 모양, 1종 오류 비율, 수렴 실패율의 측면에서 평가되었다. 평가는 다음과 같이 이루어졌다. 계산된 5,000개의 검정통계량으로 영가설 분포를 생성한 다음에 분포의 평균, 분산, 95분위수와 99분위수를 산출하였다. 이후 각 조건의 자유도에 따른 표준적인 카이제곱 분포(표 1)와 생성된 영가설 분포의 모양을 비교하였다.

다음으로 5,000번의 시행 중 검정통계량이 표준적인 카이제곱 분포의 95분위수보다 큰

표 1. 본 연구의 시뮬레이션에서 표준적인 카이제곱 분포의 모양

경쟁모형	실험설계	자유도(df)	평균(μ)	분산(σ^2)	95분위수	99분위수
영상관 구조	단일변인 두 수준	1	1	2	3.84	6.63
	단일변인 세 수준	2	2	4	5.99	9.21
	2×2 요인설계	3	3	6	7.81	11.34
	2×3 요인설계	5	5	10	11.07	15.09
상관 구조	단일변인 두 수준	2	2	4	5.99	9.21
	단일변인 세 수준	5	5	10	11.07	15.09
	2×2 요인설계	9	9	18	16.92	21.67
	2×3 요인설계	14	14	28	23.68	29.14

시행의 비율인 탐지율을 계산하였다. 참 모형의 고정효과는 모두 0이므로 탐지율은 1종 오류 비율을 나타낸다. 1종 오류 비율은 유의수준에 해당하는 0.05에 가까울수록 좋은 지표임을 나타낸다.

마지막으로 수렴 실패율은 R의 performance 패키지(Lüdtke et al., 2021)의 check_convergence 함수를 통해 확인하였다. 이 함수는 추정 과정에서 최종적으로 산출된 추정값의 상대 기울기(relative gradient) 벡터를 통해 수렴 여부를 판단한다. 5,000번의 시행 가운데 상대 기울기의 최댓값이 0.001보다 큰 시행의 비율을 수렴 실패율로 정의하였다.

추가적으로, 다음 두 가지를 조사하였다. 첫째, 수렴에 성공한 시행만으로 영가설 분포를 생성하고 그 모양을 관찰하였다. 둘째, Shapiro(1985)의 공식에 따른 혼합 카이제곱 분포를 생성된 영가설 분포의 모양과 비교하였다.

결 과

시뮬레이션 조건에 따른 수렴 실패율은 표 2에 제시되었다. 경쟁 모형의 참가자 무선효과가 영상관 구조일 때는 실험설계가 복잡해

표 2. 본 연구의 시뮬레이션에서 LRT 영가설 분포의 수렴 실패율

실험설계	참가자 수, 문항 수	모형의 종류		
		참 모형	경쟁 모형 1 (영상관 구조)	경쟁 모형 2 (상관 구조)
단일변인 두 수준	30명, 30문항	0.000	0.001	0.003
	30명, 60문항	0.000	0.002	0.005
	60명, 30문항	0.000	0.001	0.007
	60명, 60문항	0.000	0.001	0.007
단일변인 세 수준	30명, 30문항	0.000	0.009	0.142
	30명, 60문항	0.000	0.004	0.139
	60명, 30문항	0.000	0.009	0.151
	60명, 60문항	0.000	0.009	0.159
2×2 요인설계	30명, 30문항	0.000	0.008	0.288
	30명, 60문항	0.000	0.007	0.287
	60명, 30문항	0.000	0.010	0.315
	60명, 60문항	0.000	0.007	0.317
2×3 요인설계	30명, 30문항	0.000	0.015	0.558
	30명, 60문항	0.000	0.013	0.552
	60명, 30문항	0.000	0.014	0.623
	60명, 60문항	0.000	0.016	0.591

지더라도 최대 0.016의 낮은 수렴 실패율을 보였다. 반면 경쟁 모형이 참가자 무선효과 구조로 상관 구조를 포함할 때는 실험설계가 복잡해짐에 따라 수렴 실패율이 증가하였다. 단일변인 두 수준의 시뮬레이션 조건에서는 수렴 실패율이 최대 0.007이었다. 그러나 2×3 요인설계에서의 수렴 실패율의 범위는 [0.558, 0.623]이었다. 예측되는 결과와는 반대로, 참가자 수의 증가는 수렴 실패율을 증가시키는 경향을 보였다. 문항 수의 증가는 수렴 실패율에 영향을 주지 못하였다.

무선효과 구조를 검정하기 위한 LRT 평가설 분포의 모양과 분포의 평균, 분산, 95분위수, 99분위수는 그림 1과 표 3에 각각 제시되었다. 시뮬레이션으로 생성된 LRT 분포의 평균, 분산, 95분위수, 99분위수는 자유도가 참 모형과 경쟁 모형의 차이인 표준적인 카이제곱 분포에 비해 작았다. 모양에 대한 통계량은 경쟁 모형의 종류, 실험설계의 복잡성, 참

가자 수, 문항 수에 걸쳐 일관적으로 작았다. 수렴에 성공한 시행만으로 생성한 평가설 분포의 모양(그림 1)은 수렴에 실패한 시행도 포함된 전체 시행으로 생성한 평가설 분포의 모양과 크게 달라지지 않았다(그림 2). 수렴실패 시행을 포함했을 때와 포함하지 않았을 때 사이의 95분위수 차이값과 99분위수 차이값은 모든 시뮬레이션 조건을 통틀어 0.793을 넘지 않았다.

표준적인 카이제곱 분포를 사용한 1종 오류 비율도 모든 조건에서 유의수준인 .05보다 낮았다(표 4). 시뮬레이션 조건에 걸쳐 관찰된 1종 오류 비율의 범위는 0.002에서 0.025였다. 모양에 대한 통계량과 마찬가지로, 1종 오류 비율에 미치는 경쟁 모형의 종류와 시뮬레이션 조건의 영향은 찾기 어려웠다.

또한 Shapiro(1985)에 의해 제안된 혼합 카이제곱 분포와 비교했을 때, 생성된 평가설 분포는 오직 영상관 구조 모형과 비교할 때만

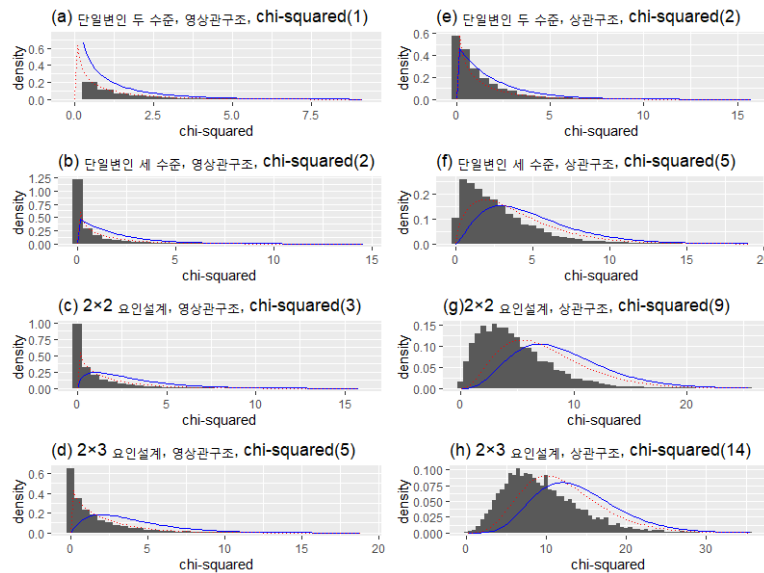


그림 1. 본 연구의 시뮬레이션에서 LRT 평가설 분포의 모양(참가자 수=30, 문항 수=30). 실선(파란색)은 표준적 카이제곱 분포, 점선(빨간색)은 Shapiro(1985)에 의해 제안된 혼합 카이제곱 분포를 각각 나타냄.

표 3. 본 연구의 시뮬레이션에서 LRT 영가설 분포의 모양

실험설계	참가자 수, 문항 수	경쟁 모형							
		영상관 구조				상관 구조			
		평균(μ)	분산(σ^2)	95분위수	99분위수	평균(μ)	분산(σ^2)	95분위수	99분위수
단일변인 두 수준	30명, 30문항	0.358	0.795	2.053	4.538	1.267	2.553	4.443	7.399
	30명, 60문항	0.357	0.787	2.159	4.193	1.319	2.669	4.647	7.512
	60명, 30문항	0.400	1.094	2.254	5.138	1.302	2.718	4.505	7.618
	60명, 60문항	0.400	0.927	2.220	4.678	1.346	2.745	4.720	7.593
단일변인 세 수준	30명, 30문항	0.670	1.712	3.302	6.217	2.827	6.124	7.575	11.668
	30명, 60문항	0.648	1.515	3.101	6.037	2.796	5.701	7.456	10.903
	60명, 30문항	0.754	1.777	3.491	6.060	2.994	6.153	7.887	11.271
	60명, 60문항	0.708	1.680	3.333	5.882	2.933	6.289	7.787	11.696
2×2 요인설계	30명, 30문항	0.951	2.327	4.036	7.119	4.772	10.071	10.822	14.979
	30명, 60문항	0.927	2.107	3.892	6.661	4.705	10.165	10.806	14.883
	60명, 30문항	1.029	2.486	4.239	7.406	4.891	10.060	10.928	14.436
	60명, 60문항	1.013	2.413	4.215	7.077	4.966	11.106	11.262	16.134
2×3 요인설계	30명, 30문항	1.511	3.946	5.548	8.869	9.445	22.186	18.241	23.662
	30명, 60문항	1.473	3.593	5.356	8.395	9.352	20.209	17.575	22.370
	60명, 30문항	1.615	3.855	5.627	8.705	9.988	22.352	18.552	23.818
	60명, 60문항	1.595	3.813	5.551	8.432	10.068	23.806	19.126	24.632

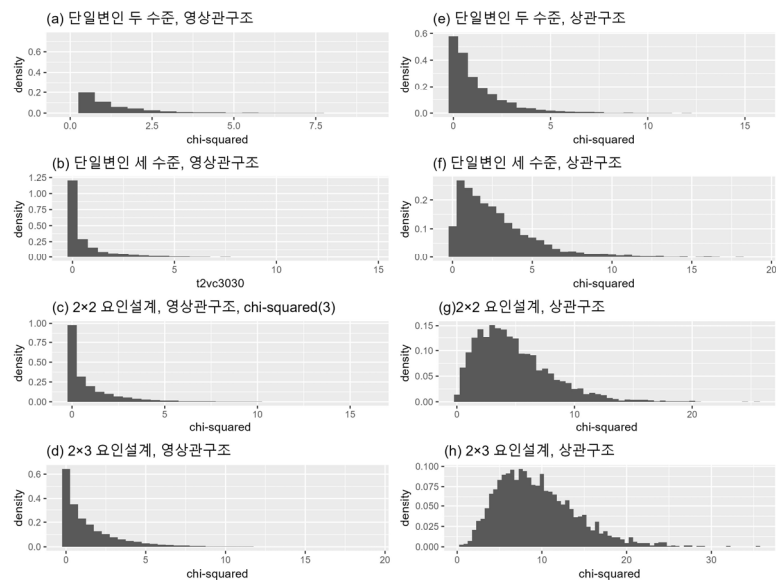


그림 2. 수렴에 성공한 시행만으로 생성한 LRT 영가설 분포(참가자 수=30, 문항 수=30)

표 4. 본 연구의 시뮬레이션에서 LRT 영가설 분포의 1종 오류 비율

실험설계	경쟁 모형	참가자 수, 문항 수			
		30명, 30문항	30명, 60문항	60명, 30문항	60명, 60문항
단일변인 두 수준	영상관 구조	0.015	0.013	0.020	0.016
	상관 구조	0.023	0.024	0.023	0.025
단일변인 세 수준	영상관 구조	0.012	0.101	0.011	0.009
	상관 구조	0.013	0.009	0.011	0.014
2×2 요인설계	영상관 구조	0.006	0.005	0.009	0.007
	상관 구조	0.004	0.004	0.004	0.007
2×3 요인설계	영상관 구조	0.004	0.003	0.003	0.002
	상관 구조	0.010	0.007	0.011	0.013

혼합 카이제곱 분포에 가까웠다(그림 1(a)-(d)). 상관 구조 모형과 비교할 때의 영가설 분포는 혼합 카이제곱 분포보다도 더 0에 치우친 모양을 보였다(그림 1(e)-(h)).

논 의

본 연구는 무선효과 구조를 선택하기 위한 LRT의 영가설 분포 모양을 조사하였다. 몬테 카를로 시뮬레이션을 수행한 결과, 분산, 공분산 모수에 대한 가설을 검정하기 위한 LRT 영가설 분포의 평균과 분산은 모수의 수 차이를 자유도로 갖는 카이제곱 분포보다 작았다. 이 결과는 여러 실험설계에서 일관적으로 관찰되었고 분산만을 검정했을 때뿐만 아니라 분산과 공분산을 함께 검정했을 때도 관찰되었다. LRT 영가설 분포의 1종 오류 비율은 유의수준에 비해 작았다. 참가자 수와 문항 수는 LRT 영가설 분포에 영향을 주지 못했다. 본 연구는 다층모형에서 무선효과 구조 검정을 위한 LRT 영가설 분포를 논의했던 선행연구들과 일

치하는 결과를 보여주었다. 그러나 본 연구는 기존 문헌들에 비해 공분산 구조의 차원을 증가시켜 조사하였다. 또한, 시뮬레이션 방법을 사용하여 LRT 영가설 분포를 관찰하였고 참가자 수와 문항 수를 조작했다는 점에서 기존 문헌들과 구별된다.

추정 과정에서의 수렴 실패는 LRT 영가설 분포의 모양에 직접적인 영향을 주지 않았다. 수렴 실패율은 실험설계가 복잡해지고 무선효과 구조에 분산과 공분산이 모두 포함되어 있을 때 높았다. 그러나 무선효과 구조에 분산만이 포함된 영상관 구조에서는 복잡한 실험설계에서도 수렴 실패율이 높지 않았다. 또한 수렴에 성공한 시행만을 가지고 영가설 분포를 생성했을 때도, 그 영가설 분포의 모양은 변화하지 않았다.

본 연구의 경쟁 모형은 모두 무선기울기 모형이었으므로 복잡한 실험설계일수록 과모수화(overparameterization)가 더 강하게 발생했을 것이다. 과모수화란 모형이 자료의 정보량에 비해 더 많은 모수를 갖는 상태이다(Green et al., 1999). 본 연구는 과모수화와 수렴 실패율

의 관계에 대해 두 가지를 시사한다. 첫째, 과모수화된 모수의 종류는 수렴 실패율에 영향을 미친다. 단일변인 세 수준의 상관 구조 모형과 2×3 요인설계의 영상관 구조 모형 간의 모수 수 차이는 다섯 개로 동일했지만 두 조건에서 수렴 실패율은 0.15 대 0.015 정도로 큰 차이를 보였다. 단일변인 두 수준의 상관 구조 모형과 단일변인 세 수준의 영상관 구조 모형 간의 모수 수 차이는 두 개로 동일하였다. 그러나 두 조건의 수렴 실패율이 모두 0.01 미만이었으므로 조건 사이에 수렴 실패율의 차이가 존재한다고 말하기 어렵다. 추가되는 모수가 공분산인 경우엔 수렴 실패율이 증가하지만, 분산 및 평균을 개별적으로 추정하는 경우엔 모수의 수가 많더라도 수렴 실패율이 급격하게 높아지지 않는 것으로 보인다. 둘째, 추정해야 할 공분산 행렬의 차원이 클 때 참가자 수와 문항 수를 증가시키는 것은 수렴 실패율을 감소시키지 않는다. 상관행렬은 모수 공간의 경계에 제한이 존재한다(Rousseeuw & Molenberghs, 1994). 차원이 커질수록 모수 공간이 갖는 범위의 제약은 커진다. 예를 들어, 변인의 수가 세 개일 때 $[-1, 1]$ 사이에서 무선적으로 생성된 삼각행렬 중 61.7%만이 가능한 상관행렬이었다. 변인의 수가 네 개, 다섯 개로 늘어난 경우, 가능한 상관행렬의 범위가 18.2%, 2.2%로 줄어들었다(Numpacharoen & Atsawarungruangkit, 2012; Table 2). 이 같은 상관행렬의 특징이 패턴이 없는 다차원 공분산 구조를 추정하기 어렵게 만들었을 수 있다.

본 연구는 선행연구(Molenberghs & Verbeke, 2007; Shapiro, 1985; Stram & Lee, 1994)에서 제안된 혼합 카이제곱 분포가 다층모형의 공분산 구조를 검정하기 위한 LRT 영가설 분포로

사용하기에 적절하지 않을 가능성을 보여준다. 제안된 혼합 카이제곱 분포는 무선절편 모형과 영상관 구조 모형이 비교될 때에만 생성된 영가설 분포의 모양과 유사하였다. 이때는 오직 분산 모수만이 검정된다. 반면 검정하려는 모수에 분산과 공분산 모수가 함께 포함될 때는 혼합 카이제곱 분포가 영가설 분포를 충분히 교정하지 못했다. Stoel 등(2006)은 혼합 분포의 비율을 설정하기 위해 시뮬레이션 방법을 사용할 것을 제안하였다. 그러나 본 연구에서 생성된 영가설 분포는 공분산 모수의 차이만을 자유도로 갖는 카이제곱 분포에 비해서도 왼쪽으로 끌어당겨진 모양을 보였다. 이 분포는 가장 자유도가 작은 카이제곱 분포가 1, 나머지 분포가 모두 0의 혼합비율을 가진 가장 극단적인 분포라고 볼 수 있으므로, 모든 혼합 카이제곱 분포는 이 분포보다 오른쪽에 있다. 본 연구는 공분산을 포함한 무선효과 구조를 검정할 때 혼합 카이제곱 분포가 생성된 영가설 분포와 일치하지 않는 이유를 제공하지 않는다. 이에 대한 이론적 설명을 위해서 추후 연구가 수행될 필요가 있다.

본 연구는 무선효과 구조의 가설 검정을 위한 LRT의 영가설 분포가 표준적인 카이제곱 분포보다 왼쪽에 위치해 있음을 보였다. 그러나 이 결과만으로 LRT가 무선효과 구조를 실제보다 단순하게 추정한다고 보기 어렵다. 실제 분산과 공분산의 크기가 0이 아닐 때 통계적 유의성을 얼마나 민감하게 탐지할 수 있는지는 1종 오류 비율과 다른 문제이기 때문이다. LRT의 무선효과 구조 추정에 대해 정확하게 알기 위해서는 검정력을 조사해야 한다. LRT의 검정력을 조사하기 위해서는 분산, 공분산 모수에 대한 효과크기가 필요하다. 고정효과를 탐지하기 위한 검정력을 계산할 때,

Cohen's d 와 같은 효과크기가 사용될 수 있다 (Arend & Schäfer, 2019). 그러나 분산-공분산 행렬의 크기가 3×3 이상일 때 효과크기가 어떻게 정의될 수 있는지 알려진 바가 없다. LRT의 검정력을 조사하기 위한 효과크기의 정의에 대해서는 추후 연구문제로 남겨둔다.

무선효과 구조를 선택하기 위한 LRT의 검정력을 간접적으로 추론하기 위해 Shin과 Lee (2022)를 참고할 수 있다. 이 연구의 목적은 2×2 참가자 내 요인설계를 위한 다층모형에서 분석 모형이 고정효과에 대한 가설 검정에 미치는 영향을 조사하는 것이었다. 무선효과를 최대한 포함한 모형(상관 구조 모형)과 LRT를 통해 선택된 무선효과 구조를 포함한 모형의 수행이 비교되었다. 본 연구에서처럼 시뮬레이션 조건은 $\tau^2=1$, $\sigma^2=0.5$ 로 설정되었다. τ^2 의 크기는 무선효과 분산, 공분산에 비례하여 조작되었다. $\tau^2=0.25$ 이상일 때는 상관 구조 모형과 무선허면 모형의 고정효과 검정 결과가 일치하였다. 이는 LRT가 상관 구조 모형과 무선허면 모형 중 상관 구조 모형을 100%에 가까운 비율로 선택하였음을 의미한다. $\tau^2=0.09$ 조건에서는 동일한 결과를 얻지 못했지만 80% 정도의 높은 비율로 상관 구조 모형을 선택하였다. 선행연구의 결과는 무선효과의 분산이 잔차 분산의 크기에 비해 상대적으로 작은 상황에서도 무선효과의 분산, 공분산을 탐지하기 위한 LRT의 검정력이 나쁘지 않음을 함의한다. 그러나 검정력을 이야기하기 위해서는 검정력을 직접 조사한 추가 연구가 필요할 것이다.

본 연구는 전형적인 심리학 실험설계에서 모형 비교를 위한 LRT의 영가설 분포 모양을 시뮬레이션하여 구현하였다. 고정효과와 무선효과 구조에 대한 가설 검정 시 영가설 분포

의 모양에 차이가 있음을 보였다는 점에서 의의가 있다. 그러나 생성 모형이 현실 세계를 얼마나 잘 반영하는지 알 수 없다는 한계점이 존재한다. 본 연구의 생성 모형은 완전한 무선허면 모형으로, 모든 무선효과와 잔차 분산이 정규분포를 따른다. 이 때문에 이론적인 상황에서 결과를 해석할 수밖에 없다.

위와 같은 한계점에도 불구하고 본 연구는 다층모형을 사용하는 심리학 연구자에게 다음과 같은 함의점을 제공한다. 첫째, 다층모형에서 무선효과 구조를 선택하기 위해 LRT를 사용할 경우, 영가설 분포로 자유도를 모수의 수 차이로 사용하는 카이제곱 분포 대신 대안적인 분포를 사용해야 한다. 가능하면 공분산 구조를 검정하는 도구(예: RLRsim 패키지 [Scheipl & Bolker, 2022])로 무선효과 구조를 선택하기를 권장한다. 만약 표준적인 카이제곱 분포를 영가설 분포로 설정해야 하는 경우, p 값은 참고적인 지표로만 사용해야 한다. 단순히 유의수준과 비교하여 의사결정을 내리는 것은 지나치게 보수적인 방법일 수 있다. 둘째, 무선효과 구조의 선택을 위해 LRT와 같은 가설검정 방법뿐 아니라 무선효과의 분산, 공분산의 크기, 그리고 현실 세계에서의 무선효과의 해석 가능성을 모두 고려해야 한다 (Snijder & Bosker, 2012). 가능한 무선효과의 분산이 모두 0이라는 LRT에서의 영가설 모형은 현실 세계에서 종종 지나친 제약일 수 있다. 셋째, 성공적인 모수 추정을 위해 공분산 구조를 필요한 만큼만 설정하여 무선효과 구조의 차원을 줄이거나 복잡성을 줄이는 것이 바람직하다(Bates et al., 2015b). 복잡한 실험설계로 인해 모형에 포함된 무선효과의 수가 많을 때는, 무선효과의 상관 구조를 밝히는 것이 연구의 주목적이 아니라면, 상관 구조를 설정

하지 않는 것이 수렴 실패 가능성을 줄일 수 있다. 베이지안 추정을 통해서도 수렴 실패 가능성을 줄일 수 있다(Eager & Roy, 2017). 복잡한 자료에 다층모형을 적합할 때, ML 추정을 사용하는 lme4는 14%의 수렴 실패율을 보였다. 그러나 베이지안 추정 프로그램인 Rstan을 사용했을 때, 수렴 실패율은 3%였다. 베이지안 추정은 ML 추정과 다르게, 자료에 대한 연구자의 가정인 사전 분포(prior distribution)를 설정함으로써 모수를 추정하는 데에 도움을 줄 수 있다.

참고문헌

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
<https://dx.doi.org/10.1109/TAC.1974.1100705>
- Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, 24(1), 1-19.
<https://doi.org/10.1037/met0000195>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.
<https://doi.org/10.1016/j.jml.2007.12.005>
- Baey, C., & Kuhn, E. (2020). varTestnlme: An R package for variance components testing in linear and nonlinear mixed-effects models. *arXiv preprint arXiv:2007.04791*.
<https://doi.org/10.48550/arXiv.2007.04791>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
<https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015a). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1).
<https://doi.org/10.18637/jss.v067.i01>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015b). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
<https://doi.org/10.48550/arXiv.1506.04967>
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, 12(4), 335-339.
[https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3)
- Crainiceanu, C. M., & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1), 165-185.
<https://doi.org/10.1111/j.1467-9868.2004.00438.x>
- Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint arXiv: 1701.04858*. <https://doi.org/10.48550/arXiv.1701.04858>
- Eckardt, R., Yammarino, F. J., Dionne, S. D., & Spain, S. M. (2021). Multilevel methods and statistics: The next frontier. *Organizational Research Methods*, 24(2), 187-218.
<https://doi.org/10.1177/1094428120959827>
- Gallucci, M. (2024). *GAMLj: General Analyses for Linear Models (jamovi module)*.

- <https://gamlj.github.io/>
- Green, S., Marquis, J., Hershberger, S., Thompson, M., & McCollam, K. (1999). The overparameterized analysis of variance model. *Psychological Methods*, 4, 214-233. <https://doi.org/10.1037/1082-989X.4.2.214>.
- Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, 30(22), 2696-2707. <https://doi.org/10.1002/sim.4293>
- Hoffman, L., & Walters, R. W. (2022). Catching up on multilevel modeling. *Annual Review of Psychology*, 73(1), 659-689. <https://doi.org/10.1146/annurev-psych-020821-103525>
- Huang, F. L. (2022). *Practical multilevel modeling using R*. SAGE Publications.
- Jeffreys, H. (1961). *Theory of Probability*. (3rd ed.). Oxford University Press.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54-69. <https://doi.org/10.1037/a0028347>
- Kwok, O., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research*, 42(3), 557-592. <https://doi.org/10.1080/00273170701540537>
- Lee, W. (2020). Linear mixed-effects model for analyzing experimental data. *Korean Journal of Cognitive and Biological Psychology*, 32(2), 197-211.
- <http://dx.doi.org/10.22172/cogbio.2020.32.2.006>
- Lüdtke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60). <https://doi.org/10.21105/joss.03139>
- Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, 44(2), 182-212. <https://doi.org/10.1080/00273170902794214>
- Luo, W., Li, H., Baek, E., Chen, S., Lam, K. H., & Semma, B. (2021). Reporting practice in multilevel modeling: A revisit after 10 Years. *Review of Educational Research*, 91(3), 311-355. <https://doi.org/10.3102/0034654321991229>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305-315. <http://dx.doi.org/10.1016/j.jml.2017.01.001>
- Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. <https://doi.org/10.1016/j.jml.2020.104092>
- Molenberghs, G., & Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter Space. *The American Statistician*, 61(1), 22-27. <https://doi.org/10.1198/000313007x171322>
- Numpacharoen, K., & Atsawarungruangkit, A. (2012). Generating correlation matrices based

- on the boundaries of their coefficients. *PLoS ONE*, 7(11), e48902.
<https://doi.org/10.1371/journal.pone.0048902>
- Park, J., Cardwell, R., & Yu, H.-T. (2020). Specifying the random effect structure in linear mixed effect models for analyzing psycholinguistic data. *Methodology*, 16(2), 92-111. <https://doi.org/10.5964/meth.2809>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, 19(4), 337-350.
- Rousseeuw, P. J., & Molenberghs, G. (1994). The Shape of correlation matrices. *The American Statistician*, 48(4), 276-279.
<https://doi.org/10.1080/00031305.1994.10476079>
- Scheipl, F., & Bolker, B. (2022). *RLRsim: Exact (Restricted) Likelihood Ratio tests for mixed and additive models*.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
<http://dx.doi.org/10.1214/aos/1176344136>
- Self, S. G., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398), 605-610.
<https://doi.org/10.2307/2289471>
- Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, 72, 133-144.
<https://doi.org/10.1093/biomet/72.1.133>
- Shin, Y. C., & Lee, W. (2022). Comparison of procedures for detecting the experimental condition effect in a linear mixed-effects model. *Korean Journal of Psychology: General*, 41(5), 437-460.
<http://dx.doi.org/10.22257/kjp.2022.12.41.5.437>
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. (2nd ed.). Sage. <https://uk.sagepub.com/en-gb/eur/multilevel-analysis/book234191>
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002), Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64: 583-639.
<https://doi.org/10.1111/1467-9868.00353>
- Stoel, R. D., Garre, F. G., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, 11(4), 439-455.
<https://doi.org/10.1037/1082-989X.11.4.439>
- Stram, D. O., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50(4), 1171-1177.
<https://doi.org/10.2307/2533455>
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020-

한국심리학회지: 일반

2045. <https://doi.org/10.1037/xge0000014>

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60-62.

<https://www.jstor.org/stable/2957648>

1차원고접수 : 2024. 08. 17

2차원고접수 : 2024. 11. 13

최종게재결정 : 2025. 03. 03

Shape of the Null Distribution in the Likelihood Ratio Test for Selecting Random Effects Structures in Experimental Data Analysis

Yonggi Kim

Wooyeol Lee

Department of Psychology, Chungbuk National University

When using a multilevel model to analyze experimental data, the results of hypothesis testing can vary depending on the random effects structure. Previous studies have shown that the likelihood ratio test (LRT) for selecting the random effect structure does not match the shape of the null distribution with the theoretical chi-squared distribution due to the parameter space constraints of the variance. In this study, we investigate the shape, the type 1 error rate, and convergence failure rate of null distribution of LRT to select random effects structures through Monte Carlo simulations. Before proceeding with this study, the validity of the methodology was secured by checking the shape of null distribution of LRT for testing fixed effects. We conducted simulations by manipulating the experimental design, number of participants, and number of items, and found that the mean, variance, 95th and 99th percentiles of null distribution of LRT for selecting random effects structures were smaller than the theoretical chi-squared distribution under all conditions. When LRT was performed using the theoretical chi-squared distribution, the type 1 error rate was smaller than .05 significance level. The convergence failure occurred frequently in models with complex experimental designs and correlation structure in the structure of the random effects. The caveats of choosing the model's random effects structure were discussed for empirical researchers when using multilevel models to test for experimental condition effects.

Key words : multilevel model, likelihood-ratio test, random effects structures, convergence failure rate, null distribution

부 록

부록 1. 참 모형과 경쟁 모형

실험설계	참 모형	경쟁 모형 1 (영상관 구조)	경쟁 모형 2 (상관 구조)
단일변인 두 수준 ²⁾		$Y_{ijk} = \beta_0 + P_{0j} + (\beta_1 + P_{1j})X_{1ijk} + S_i + e_{ijk}$	
단일변인 세 수준 ³⁾		$Y_{ijk} = \beta_0 + P_{0j} + (\beta_1 + P_{1j})X_{1ijk} + (\beta_2 + P_{2j})X_{2ijk} + S_i + e_{ijk}$	
2×2 요인설계 ⁴⁾		$Y_{ijk} = \beta_0 + P_{0j} + (\beta_1 + P_{1j})X_{1ijk} + (\beta_2 + P_{2j})X_{2ijk} + (\beta_3 + P_{3j})X_{1ijk}X_{2ijk} + S_i + e_{ijk}$	
2×3 요인설계 ⁵⁾		$Y_{ijk} = \beta_0 + P_{0j} + (\beta_1 + P_{1j})X_{1ijk} + (\beta_2 + P_{2j})X_{2ijk} + (\beta_3 + P_{3j})X_{3ijk} + (\beta_4 + P_{4j})X_{1ijk}X_{2ijk} + (\beta_5 + P_{5j})X_{1ijk}X_{3ijk} + S_i + e_{ijk}$	

- 2) X_1 : 독립변인의 두 번째 수준을 위한 지표변인
 3) X_1 : 독립변인의 두 번째 수준을 위한 지표변인, X_2 : 독립변인의 세 번째 수준을 위한 지표변인
 4) X_1 : 첫 번째 독립변인의 두 번째 수준을 위한 지표변인, X_2 : 두 번째 독립변인의 두 번째 수준을 위한 지표변인
 5) X_1 : 첫 번째 독립변인의 두 번째 수준을 위한 지표변인, X_2 : 두 번째 독립변인의 첫 번째 수준을 위한 지표변인, X_3 : 두 번째 독립변인의 두 번째 수준을 위한 지표변인

부록 2. 참 모형과 경쟁 모형의 무선효과 구조와 잔차 분포

실험설계	분포		
	참 모형	경쟁 모형 1 (영상관 구조)	경쟁 모형 2 (상관 구조)
단일변인 두 수준	$\begin{bmatrix} P_{0j} \\ P_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \\ & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \\ & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \\ & \tau_{10}^2 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$
단일변인 세 수준	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & \\ & 0 & 0 \\ & 0 & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & \\ & 0 & \tau_1^2 \\ & 0 & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & \\ & \tau_{10}^2 & \tau_1^2 \\ & \tau_{20}^2 & \tau_{21}^2 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$
2×2 요인설계	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ P_{3j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & \\ & 0 & 0 & \\ & 0 & 0 & 0 \\ & 0 & 0 & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ P_{3j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & \\ & 0 & \tau_1^2 & \\ & 0 & 0 & \tau_2^2 \\ & 0 & 0 & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ P_{3j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & \\ & \tau_{10}^2 & \tau_1^2 & \\ & \tau_{20}^2 & \tau_{21}^2 & \tau_2^2 \\ & \tau_{30}^2 & \tau_{31}^2 & \tau_{32}^2 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$
2×3 요인설계	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ P_{3j} \\ P_{4j} \\ P_{5j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & & & \\ & 0 & 0 & & & \\ & 0 & 0 & 0 & & \\ & 0 & 0 & 0 & 0 & \\ & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ P_{3j} \\ P_{4j} \\ P_{5j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & & & \\ & 0 & \tau_1^2 & & & \\ & 0 & 0 & \tau_2^2 & & \\ & 0 & 0 & 0 & \tau_3^2 & \\ & 0 & 0 & 0 & 0 & \tau_4^2 \\ & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$	$\begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ P_{3j} \\ P_{4j} \\ P_{5j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & & & \\ & \tau_{10}^2 & \tau_1^2 & & & \\ & \tau_{20}^2 & \tau_{21}^2 & \tau_2^2 & & \\ & \tau_{30}^2 & \tau_{31}^2 & \tau_{32}^2 & \tau_3^2 & \\ & \tau_{40}^2 & \tau_{41}^2 & \tau_{42}^2 & \tau_{43}^2 & \tau_4^2 \\ & \tau_{50}^2 & \tau_{51}^2 & \tau_{52}^2 & \tau_{53}^2 & \tau_{54}^2 \end{bmatrix}\right),$ $S_i \sim N(0, \omega^2), e_{ij} \sim N(0, \sigma^2)$

부록 3. 본 연구의 시뮬레이션에서 lmer 함수로 적합된 조건별 분석 모형 설정

모형	실험설계			
	단일변인 두 수준	단일변인 세 수준	2×2 요인설계	2×3 요인설계
참 모형	$y \sim 1 + X1 + (1 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + (1 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + X1:X2 + (1 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + X3 + X1:X2 + X1:X3 + (1 \text{Participant}) + (1 \text{Item})$
경쟁 모형 1 (영상관 구조)	$y \sim 1 + X1 + (1 + X1 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + (1 + X1 + X2 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + X1:X2 + (1 + X1 + X2 + X1:X2 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + X3 + X1:X2 + X1:X3 + (1 + X1 + X2 + X1:X2 + X1:X3 \text{Participant}) + (1 \text{Item})$
경쟁 모형 2 (상관 구조)	$y \sim 1 + X1 + (1 + X1 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + (1 + X1 + X2 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + X1:X2 + (1 + X1 + X2 + X1:X2 \text{Participant}) + (1 \text{Item})$	$y \sim 1 + X1 + X2 + X3 + X1:X2 + X1:X3 + (1 + X1 + X2 + X1:X2 + X1:X3 \text{Participant}) + (1 \text{Item})$