



선형 혼합효과 모형(linear mixed-effects model)은 자료의 위계적 구조로 인해 발생하는 종속변인의 반응들 사이의 상관을 고려할 수 있는 통계모형이다(Gelman & Hill, 2007; Pinheiro & Bates, 2000; Raudenbush & Bryk, 2002; Snijders & Bosker, 2011). 심리학 연구에서는 연구 설계 또는 표집 방법 등의 이유로 자료가 위계적인 구조를 갖는 경우에 본 모형을 활용한다. 연구자에 따라 선형 혼합효과 모형을 다층모형(multilevel model) 또는 위계적 선형 모형(hierarchical linear model)이라고 부르기도 한다. 이 모형은 언어 심리학 분야에서 실험자료를 분석하는데 사용될 것이 제안되었고(Baayen, Davidson, & Bates, 2008), 그 이후 타 분야의 심리학 실험 연구를 통해 얻은 자료를 분석하는 데에도 널리 사용되고 있다(Judd, Westfall, & Kenny, 2012; Meteyard & Davies, 2020). 몇몇 연구들(Brown, 2021; Magezi, 2015; Singmann & Kellen, 2020)은 실험 연구자들을 위해 선형 혼합효과 모형을 사용하는 방법을 안내하였고, lme4 패키지(Bates, Mächler, Bolker, & Walker, 2015)와 같은 선형 혼합효과 모형 적용을 위한 R 패키지가 보급되어 실험 연구자들이 쉽게 모형을 사용할 수 있게 되었다. 반면, 이 모형이 연구자료의 특성에 따라 어떻게 사용되어야 하는지 연구한 논문은 상대적으로 그 수가 부족하다. 따라서, 본 연구에서는 실험조건 효과를 정확히 탐지하기 위해서 실험자료의 특성에 따라 선형 혼합효과 모형을 적용하는 절차를 조사하였다. 앞으로, 서론은 아래와 같이 구성되었다. 첫째, 실험자료 분석을 위한 선형 혼합효과 모형의 특징을 설명하고, 두 종류의 모형을 설정하고, 이 모형을 실제로 사용함에서의 연구문제를 논의하였다. 둘째, 두 종류의 모형선택 전략, 즉 최대모형 접근

과 모형비교 접근을 비교하고, 관련 문헌을 고찰하였다. 마지막으로, 본 연구에서의 연구 질문을 구체화하여 제시하였다.

#### 실험자료 분석을 위한 선형 혼합효과 모형

실험자료 분석 도구로서 선형 혼합효과 모형의 사용을 제안하는 논문들은 전통적 분석 방법인 Fisher(1921)의 변량분석이 심리학 실험자료의 특성을 불완전하게 반영한다는 한계를 극복하기 위한 대안적 방법으로 제안하였다. 흔히, 심리학 실험에서 종속변인은 한 실험조건 안에서 여러 문항(또는 자극)을 통해 측정된다. 그리고 참가자가 실험조건 모두에 반복해서 노출되는 참가자 내 설계(within-participant design) 실험을 통해 얻은 자료의 경우 각 문항의 난이도와 같은 특성(문항효과)과 참가자의 개인차가 동시에 종속변인의 변산성을 설명할 것이라고 가정할 수 있다.

반복측정 변량분석은 참가자 내 설계 실험을 통해 얻은 자료를 분석하는 방법으로, 종속변인의 전체 변량에서 참가자의 개인차를 포함하고 남은 잔차의 변량을 이용해 실험조건 효과의 효과를 검정(test)한다. 전통적인 반복측정 변량분석에서는 종속변인의 변산성을 설명하는 요인으로 참가자의 개인차와 잔차, 두 개의 요인만을 가정한다. 그러나 문항효과와 같이 분석에 포함되지 않았지만 종속변인을 설명할 수 있는 요인은 잔차들 사이의 상관을 유발할 수 있고, 이는 1종 오류가 유의수준보다 더 빈번하게 발생하는 결과를 낳는다(Scariano & Davenport, 1987). 선형 혼합효과 모형은 복수의 무선평과(random effect)를 모형에 포함하여 이론적으로 종속변인에 영향을 미치는 모든 요인의 효과를 고려하는 방법으로 그

한계점을 극복한다. 따라서 본 연구에 사용할 선형 혼합효과 모형은 참가자효과와 문항효과를 무선평과로 포함하는 “교차 무선평과를 가진 선형 혼합효과 모형(linear mixed-effects model with crossed random effects)”이다.

선형 혼합효과 모형을 실제 실험자료에 적용하여 모형의 특징을 설명한 예를 위해서 Lee(2020)를 참고할 수 있다. Lee(2020)에서 사용된 예제 연구의 목적은 한국어를 모국어로 사용하는 화자들이 담화(discourse) 안에서의 문장을 이해하는 데 영향을 미치는 요인들을 조사하는 것이었다. 종속변인으로서 검사문장을 읽고 이해하는데 걸린 반응시간이 측정되었고, 독립변인으로서 검사문장의 주어가 이전 문장들과 일치하는지(일치 대 불일치)와 검사문장에 주어가 생략되었는지(언급 대 생략)가 조작되었다. 실험 문항으로 사용된 문장들은 모든 실험조건에서 반복되었으며 각 실험조건에 맞게 특징이 수정되었다. 참가자 내 설계를 사용하여 모든 참가자는 모든 실험조건에서 종속변인이 측정되었다. 이와 같은 실험설계로부터 수집된 자료에서 연구자는 종속변인인 반응시간이 독립변인인 실험조건 이외에 참가자와 문항의 변산성에 의해 설명될 것이라고 가정할 수 있다. 선형 혼합효과 모형은 이처럼 종속변인을 설명할 것이라고 연구자가 가정하는 모든 요인을 하나의 모형에 포함할 수 있다.

**모형설정**

선형 혼합효과 모형은 하나의 연속형 종속변인을 고정효과(fixed effect), 무선평과(random effect), 그리고 잔차(residual)의 합으로 설명한다. 고정효과는 자료의 상위수준 단위(즉, 참가자와 문항)에서 평균적으로 갖는 독립변인

의 효과를 가리킨다. 본 논문에서는 실험조건 효과(experimental condition effect)와 상호교환 가능한 용어이다. 무선평과는 자료의 상위수준 단위의 구성원이 개별적으로 갖는 독립변인의 효과를 가리킨다. 그리고 잔차는 모형이 종속변인에 대해 예측하는 예측값과 종속변인의 차이로서, 각 잔차 사이의 관계는 서로 독립(independence)임이 가정된다.

선형 혼합효과 모형의 특징 중 하나는 연구자가 가정하는 무선평과의 구조에 따라서 하나의 자료에 복수의 모형을 설정할 수 있다는 점이다. 무선평과(random intercept-only model) 모형은 자료의 상위수준에서 하나의 무선평과만을 갖는 모형이고, 무선평과기 모형(random slope model)은 상위수준 단위에서 여러 개의 무선평과를 갖는 모형이다(Snijder & Bosker, 2011).

앞으로 본 논문 전체에 걸쳐 2 × 2 참가자 내 설계에서 논의를 진행한다. 두 개의 실험조건 A, B는 각각 두 개의 수준(각각 a=1과 2; b=1과 2)을 갖고 있다. 두 독립변인은 모두 참가자 내 변인으로서 J명의 참가자(j=1, ..., J)는 요인이 교차되는 모든 수준(즉, a=1/b=1, a=1/b=2, a=2/b=1, a=2/b=2)에서 종속변인이 측정된다. 각 수준에서 종속변인은 I개의 문항(i=1, ..., I)을 통해 측정된다. 모든 문항들은 모든 수준에서 반복되어 사용된다. 이 때, 실험조건 A의 a수준과 실험조건 B의 b수준에서 j번째 참가자와 i번째 문항으로부터 측정된 종속변인  $y_{abji}$  을 설명하는 무선평과 모형(M1)은 수식 (1)과 같이 표현할 수 있다.

$$y_{abji} = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2b} + \beta_3 x_{1a} x_{2b} + s_j + w_i + e_{abji} \tag{1}$$

수식에 표현된 기호는 각각 다음을 나타낸다. 우선, 더미 코딩된 변인  $x_{1a}$ ,  $x_{2b}$ ,  $x_{1a}x_{2b}$ 은 단순주효과와 상호작용효과를 위한 항을 나타낸다.  $x_{11} = 0$ 은 종속변인이 독립변인 A의 1수준에서,  $x_{12} = 1$ 은 A의 2수준에서,  $x_{21} = 0$ 은 독립변인 B의 1수준에서,  $x_{22} = 1$ 은 B의 2수준에서 측정되었음을 각각 나타낸다. 그리고 상호작용효과항  $x_{1a}x_{2b}$ 은 독립변인 A와 B가 모두 2수준일 때만 1로 코딩되고, 나머지 수준에서는 0으로 코딩된 더미변인이다. 고정효과  $\beta_0, \beta_1, \beta_2, \beta_3$ 은 평균적인 참가자와 문항이 갖는 절편, 단순주효과, 상호작용효과를 각각 나타낸다. 구체적으로,  $\beta_0$ 은 A, B가 모두 1수준에서의 종속변인의 평균,  $\beta_1$ 은 B가 1수준일 때의 A의 단순주효과,  $\beta_2$ 는 A가 1수준일 때의 B의 단순주효과,  $\beta_3$ 은 A와 B의 상호작용효과로 각각 해석한다. 무선 절편 모형에서 참가자 무선효과  $s_j$ 와 문항 무선효과  $w_i$ 는 개별 단위가 갖는 효과, 즉 고정효과로부터 떨어진 거리라고 해석한다.

두 무선효과는 평균을 0으로 하는 정규분포  $s_j \sim N(0, \tau^2)$ ,  $w_i \sim N(0, \omega^2)$ 를 각각 따른다고 가정한다. 모형에서 고정효과와 무선효과를 고려하고 남은 잔차인  $e_{abji}$ 는 반응  $y_{abji}$ 와 예측값  $\widehat{y_{abji}}$ 의 차이를 의미하며, 이 또한 정규분포  $e_{abji} \sim N(0, \sigma^2)$ 를 따른다고 가정한다.

선형 혼합효과 모형에서의 효과의 해석에 대한 이해를 돕기 위해 그림 1을 제시하였다. 그림 1은  $2 \times 2$  참가자 내 설계 실험에서의 가상적인 결과를 보여준다. 그래프의 네 점 ㉠, ㉡, ㉢, ㉣는 각각 실험조건인 네 수준에서의 평균을 나타낸다. 먼저, 절편에 대한 고정효과  $\beta_0$ 은 독립변인 A와 B의 수준이 1일 때의 평균값, 즉 ㉠을 나타낸다. 고정효과  $\beta_1$ 와  $\beta_2$ 는 독립변인 A와 B의 수준이 1에서 2로 변화했을 때의 평균차이, 즉 ㉠-㉡, ㉠-㉢을 각각 나타낸다. 그리고 상호작용효과에 대한 고정효과  $\beta_3$ 은 한 독립변인의 수준에 따른 다른 독립변인의 효과의 차이, 즉 (㉠-㉡)-(㉢-㉣)을 나타낸다. 참가자 무선효과  $s_j$ 는 개별

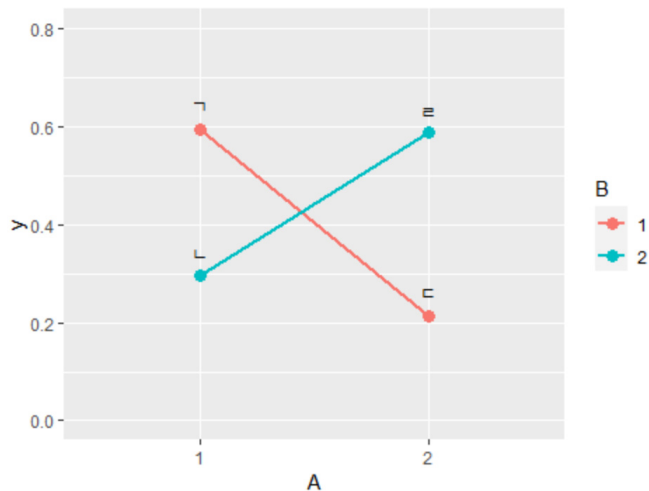


그림 1.  $2 \times 2$  참가자 내 설계 실험에서의 가상적인 결과

참가자의 반응경향성, 즉 종속변인이 평균으로부터 떨어진 거리를 의미한다. 예를 들어 어떤 참가자의 무선효과가 0.1이라면 이 참가자는 독립변인의 모든 수준에서 평균적인 참가자에 비해 0.1 큰 예측값을 가질 것임을 의미한다. 문항 무선효과  $w_i$  또한 참가자 무선효과와 마찬가지로 평균반응에서 떨어진 거리로서 해석한다.

무선기울기 모형(M2)은 수식 (2)와 같이 표현할 수 있다.

$$\begin{aligned}
 y_{jab} &= \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2b} + \beta_3 x_{1a} x_{2b} + s_{0j} \\
 &\quad + s_{1j} x_{1a} + s_{2j} x_{2b} + s_{3j} x_{1a} x_{2b} + w_i + e_{jab} \\
 &= (\beta_0 + s_{0j}) + (\beta_1 + s_{1j}) x_{1a} + (\beta_2 + s_{2j}) x_{2b} \\
 &\quad + (\beta_3 + s_{3j}) x_{1a} x_{2b} + w_i + e_{jab}
 \end{aligned} \tag{2}$$

무선절편 모형과의 차이는 참가자 무선효과가 모든 고정효과에 대응된 것이다. 4개의 참가자 무선효과  $[s_{0j}, s_{1j}, s_{2j}, s_{3j}]$ 는 고정효과  $[\beta_0, \beta_1, \beta_2, \beta_3]$ 에 대한 개인차로서, 실험조건 효과의 크기가 참가자마다 다를 때 이를 모형에 반영할 수 있다. 그림 1의 예시를 들어 설명하면, 무선절편 모형은 개별 참가자가 실험조건에 반응할 수 있다. 그림 1의 예시를 들어 설명하면, 무선절편 모형은 개별 참가자가 실험조건에 반응할 수 있다. 그림 1의 예시를 들어 설명하면, 무선절편 모형은 개별 참가자가 실험조건에 반응할 수 있다. 그림 1의 예시를 들어 설명하면, 무선절편 모형은 개별 참가자가 실험조건에 반응할 수 있다.

참가자 무선효과의 분포는 수식 (3)과 같이 다변량 정규분포로 설정할 수 있다.

$$\begin{bmatrix} s_{0j} \\ s_{1j} \\ s_{2j} \\ s_{3j} \end{bmatrix} \sim MVN \left[ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & \\ \tau_{10} & \tau_1^2 & & \\ \tau_{20} & \tau_{21} & \tau_2^2 & \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_3^2 \end{bmatrix} \right] \tag{3}$$

참가자 무선효과의 분산은  $\tau_0^2, \tau_1^2, \tau_2^2, \tau_3^2$ 이며 이 모수들은 고정효과  $[\beta_0, \beta_1, \beta_2, \beta_3]$ 에 대한 개인차의 변산성의 크기를 나타낸다. 각 무선효과들 사이의 상관관계의 방향과 강도를 나타내는 공분산은  $\tau_{10}, \tau_{20}, \tau_{30}, \tau_{21}, \tau_{31}, \tau_{32}$ 을 통해 고려한다. 예를 들어,  $\tau_{10} = Cov(s_{1j}, s_{0j})$ 가 0보다 크다면, 이는 단순 주효과의 무선효과  $s_{1j}$ 이 0보다 큰 참가자가 절편에서의 무선효과  $s_{0j}$  역시 0보다 클 가능성이 큼을 의미한다. 무선효과의 구조와 관련하여 한 가지 언급해야 할 점은 무선기울기 모형은 무선절편 모형과 서로 포함관계(nested structure)에 있으며, 모수의 수의 관점에서 복잡성이 추가된 모형이라는 것이다. 실험조건 효과와 관련된 참가자 무선효과에 변산성이 없다면, 즉 공분산구조 행렬에 포함된 모수 중  $\tau_0^2$ 를 제외한 9개의 모수를 0으로 고정시키면 무선절편 모형과 동일한 모형이 된다.

### 주요문제

선형 혼합효과 모형을 사용하는 데 있어 실험심리학자들이 해결해야 하는 질문 중 하나는 모형 적용을 위한 절차에서 해야 하는 일련의 선택들과 관련된 것이다. 연구자는 선형 혼합효과 모형을 사용하기 위해 i) 여러 모형 중 하나를 선택하여 분석해야 하며, ii) 모형의 모수추정방법, 그리고 iii) 가설검정 방법을 가능한 선택지 내에서 고를 수 있다. 그리고 각각의 선택에 따라서 실험조건 효과의 검정결

과가 달라질 수 있다. 이는 분석방법에 연구자의 개입 여지가 적고 통계 프로그램에 따라 결과가 달라지지 않는 변량분석에 익숙한 실험심리학자들이 선형 혼합효과 모형을 사용하는 것을 어렵게 만든다.

선형 혼합효과 모형의 설정은 어떤 기준으로 이루어져야 하는가? 핵심 기준은 실험조건 효과의 정확한 탐지일 것이다. 일반적으로, 통계모형은 수집된 자료를 생성한 참모형이 존재하며 해당 참모형이 분석을 위해 설정된 모형과 일치한다는 가정이 만족될 때 올바른 결론을 보장한다(LeCam, 1953; Wald, 1949). 즉, 이상적인 상황에서는 탐지하려는 효과의 추정값에 대한 기대값이 모수와 일치하고 (unbiased), 추정값의 표준편차 또한 편향이 존재하지 않는다. 반면, 자료를 생성한 참모형과 분석모형이 일치하지 않으면 추정값과 그 표준오차 중 적어도 하나에는 편향이 발생할 수 있다. 그 결과, 탐지하려는 효과를 이론적으로 기대되는 것보다 과도하게 민감한 기준으로 탐지하여 영가설을 기각하는 경우(1종 오류)나, 과도하게 엄격한 기준으로 탐지하여 영가설을 기각하지 않는 경우(2종 오류)가 발생할 수 있다.

따라서, 실험연구자들이 연구가설에 대해 타당한 결론을 내리기 위해서 심리학 실험자료 분석을 위한 선형 혼합효과 모형의 설정에 관한 연구가 필요하다. 특히, 변량분석과 비교하여 두 가지 측면에서의 연구가 필요하다. 하나는 실험자료의 특성에 따른 모형의 수행을 평가하는 것이다. 변량분석에서는 자료가 모형의 분포적 가정을 만족할 때와 그렇지 못했을 때 가설검정 결과가 어떻게 달라지는지와 가설검정 결과의 왜곡을 줄이기 위해 사용될 수 있는 대안적인 방법들이 알려져 있다

(Brown & Forsythe, 1974; Lix, Keselman, & Keselman, 1996; Tomarken & Serlin, 1986). 선형 혼합효과 모형 역시 분포적 가정에 의존한다 (Verbeke & Lessafre, 1996). 하지만, 실험자료 분석모형으로서 변량분석의 한계점을 지적한 연구의 수는 많으나(Baayen et al., 2008; Barr, Levy, Scheepers, & Tily, 2013; Raaijmakers, Schrinemakers, & Gremmen, 1999), 실험자료 특성에 따른 선형 혼합효과 모형의 가설검정 능력을 보인 연구는 상대적으로 그 수가 적다. 실험조건 효과의 유의성에 주로 관심을 갖는 실험연구자가 자료의 무선효과 구조에 대한 가정을 갖는 것은 언제나 가능한 일이 아니므로, 무선효과 구조<sup>1)</sup>의 관점에서 자료의 특성에 따른 선형 혼합효과 모형의 수행을 연구하는 것은 실험연구자에게 유용한 정보를 제공할 수 있을 것이다.

둘째로, 실험연구에서 관행적으로 사용되는 표본 수와 실험설계에서의 모형 수행에 대한 평가가 필요하다. 변량분석은 잔차의 변산성에 대한 가정이 단순하므로 실험설계가 복잡해지더라도 추정해야 할 모수의 수가 많이 증가하지 않는다. 반면, 선형 혼합효과 모형은 실험설계가 복잡해짐에 따라서 필요한 모수의 수도 급격히 증가할 수 있다. 따라서, 복잡한 실험설계를 고려하는 연구자는 모형의 복잡성과 간명성 사이에서 선택의 필요성이 커지고, 실험연구에서 사용되는 표본 수가 비실험연구에서보다 상대적으로 작다는 점을 고려하면 더욱 그렇다. 실험연구에서 관행적으로 사용

1) 본 원고에서 무선효과 구조(random effects structure)는 종속변인의 변산성을 설명하는 요인들의 공분산구조를 통칭하기 위해 사용된다(Barr et al., 2013). 참가자 무선효과의 공분산구조는 무선효과 구조의 부분집합으로 생각될 수 있다.

되는 참가자 수와 설계의 복잡성에 대한 증거를 제공하기 위해 현재 한국에서 출판된 실험 연구들의 참가자 수와 설계를 조사하였다. 2020년 9월 1일부터 2021년 7월 31일 사이에 출판된 ‘한국심리학회지: 일반’과 ‘한국심리학회지: 사회 및 성격’, 그리고 ‘한국심리학회지: 발달’에서 참가자 내 실험 설계가 적용된 실험 논문 18개를 조사한 결과 참가자 수의 중앙값은 36이었으며, 그중 13개의 논문(72.2%)이 10-50명 사이의 참가자를 사용하였다. 또한, 참가자 내 설계 실험을 수행한 논문 18개 중 독립변인의 수가 두 개인 연구는 9개(50%)이고, 세 개인 연구는 5개(27.77%)였다. 또한, 18개 중 7개(38.89%)의 연구에서 독립변인에 걸친 전체 수준의 수가 4개를 초과하였고 수준 수의 범위는 8에서 18이었다.

**모형선택 방법: 최대모형 접근과 모형비교 접근**

실험자료를 위한 선형 혼합효과 모형을 선택하는 방법으로 크게 두 가지 방법이 제안되었다. 첫째, 무선기울기 모형만을 사용하여 실험조건 효과를 검정하는 방법이다(최대모형 접근, Barr et al., 2013). 해당 접근의 무선기울기 모형은 연구 설계에 의해 가능한 최대로 복잡한 형태의 참가자 무선효과와 공분산 행렬을 포함한다. 이 방법은 실험설계에서 만들어지는 최고차원의 상호작용효과와 그 아래의 모든 검정 가능한 효과를 분석에 포함하는 변량분석의 관행과 일치한다. 둘째, 하나의 자료에 무선절편 모형과 무선기울기 모형을 각각 적합한 후 모형비교를 통해 선택된 모형을 실험조건 효과 검정에 사용하는 방법이다(모형비교 접근, Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). 이 방법은 모형의 비교를 통해

최적의 모형을 찾고 이를 통해 예측변인의 효과를 검정하는 회귀분석의 관행과 일치한다.

논리적으로, 모형비교 접근은 최대모형 접근에서 추가적인 절차를 수행하는 방법이므로 두 접근의 장단점을 비교하기 위해서는 그 절차를 통해 얻어지는 이득의 관점에서 살펴보아야 한다. 다음 조건들이 만족되는 상황에서 최대모형 접근과 모형비교 접근을 사용한 추론 결과의 차이는 없을 것이다. 첫째로는 실험을 통해 얻은 심리학 자료가 무선기울기 모형의 가정을 잘 만족하고, 둘째로는 자료의 크기가 충분히 크다는 것이다. 이 경우에는 모형비교 접근에서 무선절편 모형과 무선기울기 모형의 적합성을 비교하더라도 대부분 무선기울기 모형이 선택될 것이기 때문이다. 반면 다음의 경우에는 모형비교 접근이 최대모형 접근과 추론 결과가 일치하지 않을 수 있을 것이다. 첫째, 참가자 무선효과와 공분산구조가 실제로 무선절편 모형의 가정과 더 가까운 경우이다. 이 경우에는 모형비교 접근의 모형비교 과정에서 무선절편 모형이 선택될 확률이 더 높아지므로 두 접근 사이에서 실험조건 효과에 대한 추론결과가 정확히 일치하지 않을 것이다. 그리고 최대모형(무선기울기 모형)은 실제보다 지나치게 복잡해지므로 검정력에서 손해를 볼 것이다(Bates, Kliegl, Vasishth, & Baayen, 2015). 둘째, 참가자 무선효과와 공분산구조가 실제로 무선기울기 모형의 가정과 더 가깝지만, 표본 수와 비교하면 모형이 지나치게 복잡하여 무선기울기 모형을 사용했을 때 모수추정에 실패하는 경우이다. 이 경우, 최대모형 접근을 통한 가설검정은 이루어지지 않거나 신뢰롭지 않을 것이다(Maas & Hox, 2005). 모형비교 접근을 통한 가설검정 역시 모형비교 과정에서 무선절편 모

형이 선택되더라도 잘못된 모형설정으로 인한 부정확한 추론 결과를 예상할 수 있다(Rao, 1971).

몇몇 연구자들이 시뮬레이션 연구를 통해 실험자료를 위한 선형 혼합효과 모형의 추론 능력을 평가했다. Barr 등(2013)은 직접적으로 최대모형 접근과 모형비교 접근 사이에서 자료생성모형(data generating model)이 무선기울기 모형일 때 실험조건 효과를 탐지하는 비율을 비교했다. 그 결과, 모형비교 접근에서 1종 오류비율은 최대모형 접근에서보다 더 높았던 반면, 통계적 검정력에 있어서는 두 접근 사이에서 차이를 보이지 않았다. 반면, 동일한 연구질문에 대해 조사한 Matuschek 등(2017)은 최대모형 접근이 모형비교 접근에 비해 검정력이 낮을 수 있음을 보였다. 특히, 자료생성 모형(data-generating model)이 무선절편 모형이거나 무선기울기 모형이더라도 참가자 무선효과와 분산크기가 작을 때 그러한 결과가 나타났다. 두 연구 사이의 일치되지 않은 결과는 첫째, 모형비교를 위해 사용한 방법이 다르고 둘째, 시뮬레이션 조건의 수준 가운데 무선효과와 잔차의 분산을 조작한 강도가 다르기 때문으로 보인다. Barr 등(2013)과 Matuschek 등(2017)은 모두 50명 이하의 참가자 수를 사용하였지만, 시뮬레이션 조건으로 참가자 수를 조작하지는 않았다. 또 다른 선행연구인 Luke(2017)는 시뮬레이션 조건으로 참가자 수와 문항 수를 체계적으로 조작하여 무선기울기 모형의 수행을 조사하였다. 그 결과, 모수 추정 방법과 가설검정 방법에 무관하게 참가자 수와 문항 수가 모두 36 이상일 때 적절한 수준의 1종 오류비율이 산출되었다.

복잡한 실험설계에서의 무선기울기 모형의 수행은 어떠한가? 앞서 언급된 시뮬레이션 연

구들은 모두 두 수준을 포함한 하나의 실험조건을 가진 실험설계를 가정했다. 하지만 이와 같은 실험설계는 현재 수행되고 있는 경험적 연구의 설계에 비해 지나치게 단순하고, 무선기울기 모형이라고 할지라도 단순절편 모형과 비교하면 두 개의 모수( $\tau^2, \tau_{01}$ )만이 추가되어서 복잡성이 크게 증가하지 않는다. 따라서 선행연구에서 제안된 모형선택 전략을 반복측정 요인설계(repeated-measures factorial design)와 같은 복잡한 실험설계에 일반화하기에는 어려움이 있다. 몇몇 연구자들이 요인설계에서의 선형 혼합효과 모형의 수행을 보고하였다. Park, Cardwell과 Yu(2020)는  $2 \times 2$ 와  $2 \times 2 \times 2$  설계에서 무선효과와 절편을 달리하여 선형 혼합효과 모형의 수행을 평가했다. 두드러진 결과는 자료생성모형이 단순절편 모형과 같이 단순한 모형일 때 최대모형을 사용할 경우 높은 비율( $2 \times 2$ 와  $2 \times 2 \times 2$  요인설계에서 각각 33%와 95%)로 모형의 모수 정과정에서 수렴실패(convergence failure)가 나타난 것이다. 하지만 실험조건 효과의 추정값은 모든 자료생성모형의 무선효과와 구조와 상관없이 편향이 작았다. Lee(2020)는  $2 \times 2$  설계에서 선형혼합효과 모형의 수행을 변량분석의 관행에서의 수행과 비교하였다. 자료생성모형이 무선기울기 모형인 경우에 변량분석을 사용했을 때보다 검정력은 더 높은 반면 1종 오류비율은 더 낮은 결과를 보였다. 하지만, 두 논문 모두에서도 무선효과 구조의 복잡성에 따라서 최대모형 접근과 모형비교 접근에 따른 실험조건 효과의 검정결과를 직접적으로 조사하지는 않았다. 따라서 본 연구에서는 무선효과와 그 크기를 조작해서 모형설정 방법에 따른 추론 결과를 비교해보고자 한다.



### 연구목적

본 연구의 목적은 반복측정 요인설계로부터 수집된 실험자료를 분석하기 위해 선형 혼합효과 모형을 사용할 때 실험자료 특성에 따라서 모형선택 전략이 실험조건 효과 검정에 미치는 영향을 알아보는 것이다. 구체적으로, 두 가지 연구질문에 대해 조사하였다. 첫째, 자료생성모형의 무선효과 구조의 복잡성에 따른 두 종류의 선형 혼합효과 모형(즉, 무선절편 모형과 무선기울기 모형)의 상호작용 효과에 관한 모수복구(parameter recovery) 능력은 어떠한가? 둘째, 자료생성모형의 무선효과 구조의 복잡성에 따른 두 종류의 모형선택 전략(즉, 최대모형 접근과 모형비교 접근)이 상호작용 효과를 정확히 탐지하는 능력(탐지율, detection rate)은 어떠한가? 상호작용 효과를 관심 모수로 사용한 이유는 실험연구에서 요인설계를 사용하는 주요한 목적이 상호작용 효과를 탐지하는 것이기 때문이다(Maxwell & Delaney, 2017).

### 방 법

본 연구에서는 연구질문에 대한 답을 얻는데 필요한 경험적 증거를 수집하기 위해 몬테카를로 시뮬레이션을 사용하였다. 몬테카를로 시뮬레이션이란 특정 연구 상황에 맞게 난수생성기를 사용해서 자료를 충분한 수로 생성한 후, 각각의 자료에 분석을 수행한 결과를 조사하는 방법이다(Metropolis, 1987). 본 연구의 절차는 다음과 같다. 먼저, 자료생성모형과 시뮬레이션 조건을 설정하고 각 조건에서 자료를 생성한다. 다음으로, 평가할 분석도구를

사용해서 생성된 자료를 분석한다. 셋째로, 평가지표를 정의하고 개별 자료를 분석한 결과를 종합하여 평가지표를 계산한다. 마지막으로, 계산된 평가지표를 통해 분석도구의 성능을 평가한다.

### 자료생성모형과 시뮬레이션 조건

2 × 2 참가자 내 설계를 위한 선형 혼합효과 모형이 자료생성모형으로 사용되었다(표 1.a). 실험조건 효과를 탐지하는 능력을 평가하기 위해 고정효과의 크기를 조작하였고, 본 연구에서 관심 있는 자료의 특성인 무선효과 구조의 복잡성은 참가자 무선효과의 분산크기와 구조로 조작적으로 정의되었다. 따라서 i) 고정효과  $\beta_3$ 의 크기, ii) 참가자 무선효과의 분산에 대한 모수와 iii) 참가자 무선효과의 공분산구조가 시뮬레이션 조건으로 조작되었다(표 1.b).

고정효과  $\beta_3$ 는 0, 0.2, 0.5의 세 수준으로 설정하였다. 잔차의 분산은 1이므로 Cohen's  $d$  해석에 따르면 설정된 수준은 각각 '효과크기 없음, 작음, 중간'에 해당된다(Cohen, 1988).

참가자 무선효과의 분산  $\sigma^2$ 는 0.09(=0.3<sup>2</sup>), 0.25(=0.5<sup>2</sup>), 0.49(=0.7<sup>2</sup>) 세 수준으로 설정하였다. 참가자 무선효과의 분산이 작을수록 실험 참가자들 간 종속변인의 유사성이 크다는 의미이다. 잔차 분산의 약 50%에 해당하는 큰 크기의 참가자 무선효과 분산은 실제 실험자료를 반영한 것이다(Lee, 2020).

참가자 무선효과의 공분산구조는 복잡성이 증가하는 순서대로 무선절편(random intercept-only), 분산성분(variance component), 비구조(unstructured) 구조로 설정했다. 참가자 무선효과의 공분산구조와 모수는 표 1.c에 제시되었

다. 무선절편 수준은 자료생성모형이 무선절편 모형임을 나타낸다. 분산성분에서는 절편과 실험조건 효과들에 참가자의 개인차가 있지만, 각 실험조건 효과의 참가자 무선효과 사이의 상관계수는 0이었다. 즉, 한 참가자의 상호작용효과의 크기는 그 참가자의 단순주요

표 1. 본 연구의 자료생성모형과 시뮬레이션 조건

(a) 자료생성모형

---

모형

---


$$y_{jiab} = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2b} + \beta_3 x_{1a} x_{2b} + s_{0j} + s_{1j} x_{1a} + s_{2j} x_{2b} + s_{3j} x_{1a} x_{2b} + w_i + e_{jiab}$$

$s_{0j}, s_{1j}, s_{2j}, s_{3j}$ 의 분포는 표 1. (c) 참조,  
 $w_i \sim N(0, \omega^2)$ ,  
 $e_{abji} \sim N(0, \sigma^2)$ .

---

(b) 시뮬레이션 조건

조건	모수	수준	해석
조작된 조건	$\beta_3$	0, 0.2, 0.5	상호작용효과의 고정효과
	$\tau^2$	0.09, 0.25, 0.49	참가자 무선효과의 분산
	$\tau$ 행렬	표 1. (c) 참조	참가자 무선효과의 공분산구조
고정된 조건	$\beta_0, \beta_1, \beta_2$	0	절편, 단순주요효과의 고정효과
	$\omega^2$	0.49	문항 무선효과의 분산
	$\sigma^2$	1	잔차의 분산
	J	30	참가자 수
	I	10	실험조건의 수준별 문항 수

---

(c) 참가자 무선효과의 공분산구조

공분산구조	공분산구조 행렬	출처
무선절편	$\begin{bmatrix} s_{0j} \\ s_{1j} \\ s_{2j} \\ s_{3j} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 & & & \\ & 0 & 0 & \\ & 0 & 0 & 0 \\ & 0 & 0 & 0 \end{bmatrix} \right)$	Lee(2020)
분산성분	$\begin{bmatrix} s_{0j} \\ s_{1j} \\ s_{2j} \\ s_{3j} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 & & & \\ & 0 & 0.8\tau^2 & \\ & 0 & 0 & 0.64\tau^2 \\ & 0 & 0 & 0 & 0.512\tau^2 \end{bmatrix} \right)$	Kwok, O. M., West, S. G., & Green, S. B.(2007)
비구조	$\begin{bmatrix} s_{0j} \\ s_{1j} \\ s_{2j} \\ s_{3j} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^2 & & & \\ & 0.3\tau^2 & \tau^2 & \\ & 0.7\tau^2 & 0.7\tau^2 & \tau^2 \\ & -0.3\tau^2 & -0.7\tau^2 & -0.7\tau^2 & \tau^2 \end{bmatrix} \right)$	Lee(2020)

과의 크기를 예측하지 못했다. 비구조 수준에서는 실험조건 효과의 개인차가 체계적으로 공변하였다.

반면, 시뮬레이션 조건에 걸쳐 다음의 모수가 공통으로 사용되었다. 고정효과  $\beta_0, \beta_1, \beta_2$ 는 0으로, 문항 무선효과의 분산  $\omega^2$ 은  $0.49(=0.7^2)$ , 잔차의 분산  $\sigma^2$ 은 1로 고정했다. 서론에서 언급된 최근 1년 한국심리학회지에 출판된 실험연구에 대한 문헌조사 결과를 반영하여 참가자 수는 30명, 문항 수는 수준 당 10개로 설정했다.

종합하여, 총  $3 \times 3 \times 3 = 27$ 개의 조건으로부터 자료가 생성되었다. 실험조건마다 1,000개의 자료 세트가 생성되었다. 따라서, 총 27,000개의 시뮬레이션이 수행되었다.

**분석방법**

시뮬레이션을 통해 생성된 자료는 다음과 같은 순서로 분석되었다. 첫 번째 순서로, 시뮬레이션으로 생성된 모든 자료가 서론에 제시되었던 무선절편 모형(M1)과 무선기울기 모형(M2)에 적합되었다. 언급된 바와 같이 무선기울기 모형은 참가자 무선효과의 구조는 비구조로서, 참가자 무선효과의 모든 분산과 공분산이 모수로서 추정되었다. 선형 혼합효과모형의 적합을 위해 R의 lme4 패키지(Bates et al., 2015)의 lmer 함수를 사용하였고, 추

정방법으로 전체정보최대우도(full information maximum likelihood, 이하 FIML) 추정법을 사용하였다. 표 2에 lmer함수에서의 두 분석모형 설정을 제시하였다. lmer 함수의 기본설정인 제약최대우도(restricted maximum likelihood, REML) 추정법을 사용하면 두 모형이 서로 포함관계라고 할지라도 우도비 검정(likelihood ratio test, 이하 LRT)을 위한 검정통계량이 점근적으로 카이제곱 분포를 따른다는 가정이 충족되지 않는다(Gurka, 2006).

두 번째 순서로 무선절편 모형과 무선기울기 모형 사이에서 LRT를 통한 모형비교가 이루어졌다. LRT를 위해, 자유도를 두 모형의 모수의 수 차이로 하는 카이제곱분포를 영가설분포로 사용하는 LRT를 함께 수행하였다. 카이제곱분포를 사용할 때의 기각값은  $\chi_{9,.95}^2 = 16.92$ 였다. 무선절편 모형과 무선기울기 모형 사이의 편차가 기각값보다 더 클 경우 무선기울기 모형을 적용하며, 그렇지 않을 경우 무선절편 모형을 적용하여 자료를 분석했다.

세 번째 순서로 실험조건 효과에 대한 영가설 검정 결과가 무선기울기 모형(최대모형)과 모형비교 후 선택된 모형(선택모형)사이에서 비교되었다. 실험조건 효과는 Wald 검정으로 검정되었다. 추정값을 그 표준편차로 나눈 값의 절대값이 Z분포의 97.5 백분위값인 1.96을 초과하면 실험조건 효과가 통계적으로 유의하다고 판단하였다.

표 2. lmer 함수에서의 분석모형 설정

모형	lmer 설정
무선절편 모형	lmer(y~x1*x2+(1 participant)+(1 item), REML=F, data=dat)
무선기울기 모형	lmer(y~x1*x2+(x1*x2 participant)+(1 item), REML=F, data=dat)

평가지표

모수복구에 대한 평가는 무선절편 모형과 무선기울기 모형으로 추정된 실험조건 효과  $\beta_3$ 에 대해 이루어졌고, 추정값의 편향, 추정된 표준오차의 편향, 95% 복구비율을 통해 평가하였다. 평가지표를 계산하기 위한 통계량의 계산식을 표 3에 제시하였다. 추정값의 편향은 모형 생성 시의  $\beta_3$ 와 모수추정을 통해 얻은 추정값의 기댓값  $E(\hat{\beta}_3)$ 의 차이로 정의된다. 본 연구에서는 실험조건 당 생성된 1000개의 추정값의 평균과 모수의 차이(표 3.1)로 편향을 계산하였다. 표준오차의 편향을 평가하기 위해서 다음의 두 가지 통계량을 계산하였다. 첫 번째 통계량은 선형 혼합효과모형에서 산출한 추정값에 대한 표준오차의 평균을 구하는 것이다(표 3.2). 두 번째 방법은

실험조건 당 생성된 1000개의 추정값의 표준편차를 계산하는 것이다(표 3.3). 그리고 두 통계량의 차이(표 3.4)를 추정된 표준오차의 편향으로 정의하였다. 95% 복구비율은 1000개의 분석 가운데 추정값  $\hat{\beta}_3$ 과 추정값의 표준오차  $SE(\hat{\beta}_3)$ 를 사용해서 구한 95% 신뢰구간이 모수를 포함하는 분석의 비율(표 3.5)로 정의하였다.

탐지율에 대한 평가는 최대모형 접근과 모형비교 접근에서 산출된 결과를 기준으로 이루어졌다. 탐지율은 1000개의 시뮬레이션으로 생성된 자료에서 검정도구를 통해  $\beta_3$ 가 통계적으로 유의하다고 판단된 자료의 비율로 정의한다.  $\beta_3=0$ 인 조건에서의 탐지율은 1종 오류 비율,  $\beta_3>0$ 인 조건에서의 탐지율은 통계적 검정력을 나타낸다. 1종 오류 비율은 유의수준 0.05에 가까울수록, 검정력은 1에 가까울

표 3. 평가지표에 사용된 수식

	수식	의미
(1)	$Bias(\hat{\beta}_3) = \frac{\sum_{r=1}^{1000} \hat{\beta}_3^{(r)}}{1000} - \beta_3$	추정값의 편향
(2)	$SD(\hat{\beta}_3) = \frac{\sum_{r=1}^{1000} (\hat{\beta}_3^{(r)} - \bar{\beta}_3)^2}{1000 - 1}$	추정된 표준오차
(3)	$E(SE(\hat{\beta}_3)) = \frac{\sum_{r=1}^{1000} SE(\hat{\beta}_3^{(r)})}{1000}$	추정값의 표준편차
(4)	$Bias(SE(\hat{\beta}_3)) = E(SE(\hat{\beta}_3)) - SD(\hat{\beta}_3)$	표준오차의 편향
(5)	$CR = \frac{\sum_{r=1}^{1000} I^{(r)}}{1000},$ $I^{(r)} = \begin{cases} 1 & \text{만약 } \hat{\beta}_3^{(r)} - 1.96 \cdot SE(\hat{\beta}_3^{(r)}) < \beta_3 < \hat{\beta}_3^{(r)} + 1.96 \cdot SE(\hat{\beta}_3^{(r)}) \text{ 이면,} \\ 0 & \text{그밖의 경우.} \end{cases}$	95% 복구비율

수록 좋은 지표임을 나타낸다. 본 연구에서는 경험칙(rule of thumb)에 따라서 1종 오류비율은 0.03과 0.07 사이, 검정력은 0.8 이상일 때 좋은 검정도구라고 판단하였다. 추가적으로, 두 접근 사이에서 탐지율의 차이를 설명하기 위해서, 모형비교 접근의 모형비교 단계에서 무선기울기 모형이 선택된 비율을 계산하였다.

#### 예측되는 결과

##### 모수복구

복구비율은 자료생성모형과 분석모형이 일치할 때 설정된 신뢰구간과 일치할 것이다. 따라서 자료생성모형의 구조가 무선절편 모형일 때는 분석모형이 무선절편 모형이, 나머지 구조에서는 무선기울기 모형의 복구비율이 0.95에 더 가까울 것으로 예측된다. 반면, 모형이 잘못 설정된 경우 추정값 또는 그 표준오차에 편향으로 인해 복구비율은 설정된 신뢰수준과 같지 않을 것이다. 본 연구에서 추정법으로 사용한 FIML은 무선효과의 구조와 상관없이 같은 크기의 고정효과를 산출하므로 추정값의 편향의 크기는 두 분석모형 사이에 차이가 없을 것이다(Snijder & Bosker, 2011). 하지만 추정값의 표준오차는 실제보다 복잡한 모형을 사용할수록 더 크게 추정되는 반면 실제보다 더 단순한 모형을 사용할수록 더 작게 추정된다. 따라서 복구비율은 다음의 패턴으로 부정확해질 것으로 예측된다. 첫째, 자료생성모형이 무선절편 모형일 때 무선기울기 모형을 통한 복구비율은 0.95보다 클 것이다. 그리고 자료생성모형의 참가자 무선효과의 분산이 커질수록 복구비율은 더 커질 것이다. 둘째, 자료생성모형이 무선기울기 모형일 때 무선절편 모형을 분석모형으로 사용

한 복구비율은 0.95보다 작을 것이다. 그리고 자료생성모형의 무선효과 구조가 복잡하고 무선효과의 분산이 커질수록 복구비율은 더 작아질 것이다.

##### 탐지율

선형 혼합효과 모형을 이용한 추론의 정확성은 모수복구와 마찬가지로 자료생성모형과 추정에 사용된 모형의 무선효과 구조가 더 유사할수록 높을 것이다. 자료생성모형이 단순한 무선효과 구조(무선절편, 분산성분)인 조건에서는 모형비교 접근에서 최대모형 접근에서 보다 무선절편 모형이 선택되는 비율이 더 높을 것이다. 그러므로 1종 오류 비율이 최대모형 접근에 따를 때보다 더 유의수준에 근접한 결과를 나타낼 것이며 검정력 또한 높을 것이다. 해당 상황에서 두 접근 간 탐지율의 차이는 참가자 무선효과의 분산이 커질수록 커질 것으로 예상된다. 반면, 자료생성모형이 복잡한 무선효과 구조(비구조)인 조건에서는 모형비교 과정에서 무선기울기 모형이 더 많이 선택될 것이다. 그러므로 두 접근 간 탐지율의 차이가 작을 것이다. 두 접근 사이의 탐지율 차이는 참가자 무선효과의 분산이 커질수록 더 작아질 것이다.

#### 결 과

시뮬레이션 결과는 모수복구와 탐지율 순으로 보고하였다.

##### 모수복구

표 4에 시뮬레이션 조건에 따른 실험조건

효과 추정값의 편향이 제시되었다. 예상결과와 같이 본 연구의 모든 시뮬레이션 조건에서 추정값의 편향의 크기는 무선절편 모형과 무선기울기 모형 사이에서 일치하였다. 편향의 범위는 -0.006에서 0.009이었다. 이는 본 시뮬레이션 조건에서의 작은 효과크기인  $\beta_3 = 0.2$ 의 10%에 미치지 못하는 작은 크기의 상대적 편향에 해당한다(상대적 편향의 기준에 대한 예로는 Enders, Keller, & Levy, 2018을 참고).

반면, 표준오차의 편향에 대한 결과는 다음과 같이 요약될 수 있다(표 5). 무선절편 모형은 자료생성모형이 무선절편 모형일 때 최대 0.006의 작은 편향을 보였으나 자료생성모형이 무선절편 모형이 아닌 경우에는 일관적으로 표준오차를 과소추정하였다. 편향의 크기는 참가자 무선효과의 분산이 커짐에 따라서 증가하는 패턴을 보였다. 대조적으로, 무선기울기 모형은 대부분의 시뮬레이션 조건에서 0.010 미만의 편향을 보였다. 다만, 자료생성

모형이 무선절편 모형이고, 참가자 무선효과의 분산이 0.49인 조건에서는 편향의 크기가 0.012로 실제보다 과대추정되었다.

95% 신뢰구간을 이용한 모수복구 비율은 표준오차의 편향의 패턴과 일치하였다(표 6). 무선절편 모형은 자료생성 모형이 무선절편 모형일 때 95%에 가까운 모수복구 비율을 나타냈다(범위: 0.941-0.958). 반면, 자료생성 모형이 무선절편 모형이 아닌 경우에는 일관적으로 95% 미만의 복구비율을 보였다. 참가자 무선효과의 공분산구조가 복잡하고 참가자 무선효과의 분산이 커질수록 복구비율은 감소하였다. 복구비율이 가장 낮은 시뮬레이션 조건은 자료생성모형의 참가자 무선효과의 공분산구조가 비구조적이고, 참가자 무선효과의 분산이 0.49, 고정효과의 크기가 0.2인 조건이었으며, 이 조건에서의 복구비율은 0.835였다. 무선기울기 모형의 시뮬레이션 조건에 걸친 복구비율의 범위는 0.937에서 0.970이었다. 자료생성모형의 참가자 무선효과 구조가 무선절편

표 4. 실험조건 효과 추정값( $\hat{\beta}_3$ )의 편향

$\beta_3$	$r^2$	참가자 무선효과의 구조		
		무선절편	분산성분	비구조
0	0.09	-0.001	-0.002	0.003
	0.25	0.002	-0.006	-0.005
	0.49	-0.007	0.004	-0.008
0.2	0.09	0.000	-0.004	-0.005
	0.25	-0.003	0.008	-0.005
	0.49	-0.006	0.009	-0.005
0.5	0.09	0.003	0.005	-0.001
	0.25	-0.006	-0.001	0.006
	0.49	0.003	0.001	-0.005

주. 무선절편 모형과 무선기울기 모형의 추정값은 모든 조건에서 동일함

표 5. 추정된 표준오차( $SE(\hat{\beta}_3)$ )의 편향

$\beta_3$	$r^2$	참가자 무선효과의 구조					
		무선절편		분산성분		비구조	
		M1	M2	M1	M2	M1	M2
0	0.09	0.000	0.008	-0.002	0.009	-0.011	0.004
	0.25	0.000	0.007	-0.013	0.002	-0.027	-0.002
	0.49	0.005	0.012	-0.021	-0.002	-0.041	0.001
0.2	0.09	0.002	0.009	-0.006	0.005	-0.014	0.001
	0.25	-0.005	0.002	-0.012	0.002	-0.026	0.000
	0.49	-0.001	0.006	-0.020	-0.002	-0.046	-0.005
0.5	0.09	-0.006	0.003	-0.007	0.004	-0.005	0.009
	0.25	0.001	0.008	-0.012	0.003	-0.020	0.005
	0.49	-0.002	0.006	-0.015	0.003	-0.039	0.003

주. M1: 무선절편 모형; M2: 무선기울기 모형

표 6. 95% 신뢰구간을 이용한 모수복구비율

$\beta_3$	$r^2$	참가자 무선효과의 구조					
		무선절편		분산성분		비구조	
		M1	M2	M1	M2	M1	M2
0	0.09	0.949	0.955	0.944	0.961	0.926	0.950
	0.25	0.958	0.966	0.930	0.954	0.889	0.937
	0.49	0.957	0.970	0.911	0.954	0.868	0.947
0.2	0.09	0.951	0.966	0.949	0.962	0.924	0.950
	0.25	0.941	0.948	0.928	0.949	0.886	0.939
	0.49	0.950	0.960	0.914	0.944	0.835	0.941
0.5	0.09	0.947	0.959	0.941	0.957	0.939	0.962
	0.25	0.947	0.956	0.926	0.954	0.903	0.953
	0.49	0.951	0.966	0.919	0.949	0.874	0.940

주. M1: 무선절편 모형; M2: 무선기울기 모형

모형일 때, 또는 참가자 무선효과의 분산이 하는 복구비율을 나타냈다. 0.09인 시뮬레이션 조건에서 주로 95%를 초과

탐지율

1종 오류비율

표 7에 최대모형 접근과 모형비교 접근의 시뮬레이션 조건에 따른 1종 오류비율이 제시되었다. 다음 시뮬레이션 조건만 제외하고 적절한 수준의 1종 오류비율을 보였다. 자료생성모형이 무선절편 모형이고 참가자 무선효과와 분산이 0.25와 0.49 수준일 때 최대모형 접근은 각각 0.023과 0.030의 1종 오류비율을 보였다. 비구조 조건에서는 최대모형 접근과 모형비교 접근에서의 1종 오류비율은 일치하였다. 간단한 공분산구조 조건에서는 최대모형 접근이 모형비교 접근에서보다 더 작은 1종

오류비율을 보였으나, 그 차이는 상대적으로 크지 않았다. 가장 큰 차이를 보인 시뮬레이션 조건은 자료생성모형이 무선절편 모형, 참가자 무선효과와 분산이 0.25인 조건이었고, 이 조건에서의 1종 오류비율의 차이는 0.019였다.

검정력

최대모형 접근과 모형비교 접근의 시뮬레이션 조건에 따른 통계적 검정력 결과는 다음과 같이 요약될 수 있다(표 8). 첫째, 고정효과와 크기가 0.5일 때, 두 접근 모두에서 참가자 무선효과와 구조와 분산의 크기와 상관없이 0.8 이상의 검정력을 보였고, 최대모형 접근과 모

표 7. 실험조건 효과( $\beta_3$ ) 검정에서의 1종 오류 비율

$\beta_3$	$\tau^2$	참가자 무선효과와 구조					
		무선절편		분산성분		비구조	
		최대 모형	모형비교	최대 모형	모형비교	최대 모형	모형비교
0	0.09	0.045	0.051	0.039	0.041	0.050	0.055
	0.25	0.023	0.042	0.046	0.046	0.063	0.063
	0.49	0.030	0.043	0.046	0.046	0.053	0.053

표 8. 실험조건 효과( $\beta_3$ ) 검정에서의 검정력

$\beta_3$	$\tau^2$	참가자 무선효과와 구조					
		무선절편		분산성분		비구조	
		최대 모형	모형비교	최대 모형	모형비교	최대 모형	모형비교
0.2	0.09	0.339	0.408	0.310	0.340	0.301	0.316
	0.25	0.369	0.408	0.332	0.332	0.253	0.253
	0.49	0.336	0.373	0.305	0.305	0.231	0.231
0.5	0.09	0.992	0.993	0.974	0.976	0.973	0.975
	0.25	0.987	0.989	0.951	0.951	0.940	0.940
	0.49	0.989	0.992	0.924	0.924	0.849	0.849



형비교 접근 사이의 검정력 차이도 모든 실험 조건에 걸쳐 0.003 이하로 작았다. 둘째, 고정 효과의 크기가 0.2일 때, 두 접근이 산출한 검정력은 0.231에서 0.408의 범위를 보였다. 공분산구조가 무선절편이 아니거나 참가자 무선 효과의 분산이 0.25 또는 0.49인 조건에서는 최대모형 접근과 모형비교 접근에서의 검정력이 일치하였고 참가자 무선효과의 분산이 0.09인 조건에서는 모형비교 접근과 최대모형 접근의 검정력 차이가 공분산구조에 따라서 0.015-0.030이었다. 공분산구조가 무선절편인 조건에서는 모형비교 접근을 사용하는 것이 최대모형 접근에서보다 검정력이 더 높았다. 검정력의 차이는 0.039(참가자 무선효과 분산: 0.025)에서 0.069(참가자 무선효과 분산: 0.09)였다.

표 9에 모형비교 접근에서 각 시뮬레이션 조건에서 생성된 1000개의 자료 중 무선기울기 모형이 분석모형으로 선택된 비율이 제시되었다. 가장 두드러진 결과는 자료생성모형이 무선절편 모형일 때 무선기울기 모형이 선

택되는 비율은 모든 시뮬레이션 조건에 걸쳐 최대 1.2%로 매우 낮았다는 점이다. 또한, 참가자 무선효과의 분산이 작은 크기인 시뮬레이션 조건에서 무선기울기 모형선택비율이 87%보다 낮았다. 이 결과는 자료생성모형에 무선기울기 구조가 없거나 참가자 무선효과의 분산이 0.09인 조건에서 모형비교 접근이 최대모형 접근에 비해서 1중 오류비율이 유의수준에 근접한 반면 검정력이 더 높은 결과에 대한 근거를 제공한다. 반면, 자료생성모형이 무선기울기 모형이고 참가자 무선효과의 분산이 0.25 이상인 조건에서는 모형선택 과정에서 무선기울기 모형이 99% 이상 선택되었다.

### 논 의

본 연구에서는 선형 혼합효과 모형으로 심리학 실험자료를 분석하기 위한 적절한 절차, 특히 분석에 사용할 모형을 선택하는 전략이 실험조건에 상호작용효과 검정에 미치는 영향

표 9. 모형비교 접근의 모형비교 과정에서 무선기울기 모형이 선택된 비율

$\beta_3$	$\tau^2$	참가자 무선효과의 구조		
		무선절편	분산성분	비구조
0	0.09	0.009	0.620	0.837
	0.25	0.011	0.999	1.000
	0.49	0.008	1.000	1.000
0.2	0.09	0.003	0.646	0.862
	0.25	0.012	0.999	1.000
	0.49	0.009	1.000	1.000
0.5	0.09	0.006	0.599	0.847
	0.25	0.011	0.996	1.000
	0.49	0.007	1.000	1.000

을 조사하였다. 구체적으로, 다음 두 가지 질문을 제기하였다: 첫째, 자료의 무선효과의 구조에 따라서 무선절편 모형과 무선기울기 모형이 실험조건 효과를 추정하는 능력은 어떠한가? 둘째, 자료의 무선효과의 구조에 따라서 최대모형 접근과 모형비교 접근이 실험조건 효과를 탐지하는 능력은 어떠한가? 앞으로, 각 연구질문에 대한 연구결과의 해석과 실험연구자들을 위한 함의, 그리고 연구의 제한점이 논의된다.

무선절편 모형과 무선기울기 모형의 모수복구 결과는 이론적으로 예측되는 결과와 일치하였다(Gurka, Edwards, & Muller, 2011). 즉, 참가자 무선효과의 공분산 구조가 자료생성모형과 분석모형 사이에서 일치할 때는 추정값과 그 표준오차가 편향되지 않았고, 그 결과 모수복구비율 역시 이론적으로 예측되는 값과 대체로 일치하였다. 반면, 자료생성모형과 분석모형이 일치하지 않을 때는 추정값의 표준오차 편향으로 인해서 모수복구비율이 이론적으로 예측되는 값과 일치하지 않았다. 만약 분석모형인 무선절편 모형이 자료생성모형보다 더 간단한 모형이라면 추정값의 표준오차가 과소추정되고 모수복구비율 역시 이론적으로 예측되는 크기보다 작아지는 반면, 분석모형인 무선기울기 모형이 자료생성모형보다 더 복잡한 모형이라면 추정값의 표준오차가 실제보다 크게 추정되고 모수복구비율이 이론적으로 예측되는 크기보다 컸다.

한편, 본 연구에서 편향의 방향보다 더 중요한 질문은 시뮬레이션 조건이 “얼마나” 평가지표에 영향을 주는가일 것이다. 이 질문은 다시 두 개의 질문으로 나누어진다. 첫째는 실제 자료의 참가자 무선효과의 구조가 복잡할 때 무선절편 모형의 사용이 실험조건 효과

추정에 미치는 영향은 어떠한가? 무선절편 모형의 표준오차의 편향 크기는 자료생성모형의 참가자 무선효과가 비구조인 공분산구조에서 분산성분 구조에서보다 더 컸다. 이는 실험조건 효과의 개인차 분산의 크기가 크고, 각 실험조건 효과 사이의 상관이 작거나 비일관적일 때 무선절편 모형을 사용하는 것은 실험조건 효과의 유의성을 과도하게 잘못 탐지해 낼 가능성이 있음을 시사한다. 그리고 본 시뮬레이션 조건에서 큰 크기의 참가자 무선효과 분산( $\tau^2=0.49$ )은 잔차 분산( $\sigma^2=1$ )의 약 50%이었는데, 이는 실제 실험자료를 반영한 것이다(Lee, 2020). 이 수준에서 모수복구비율은 최소 83%까지 작아져서 사전에 설정한 95% 신뢰구간과 큰 차이를 보였다. 종합하면, 실제 자료의 공분산구조가 무선절편 모형보다 더 복잡한 경우에 분석모형으로 무선절편 모형을 사용할 때는 1종 오류의 가능성이 상당히 커지는 위험에 노출되고, 이는 Barr 등(2013)에서의 결론과도 일치한다. 둘째, 실제 자료의 참가자 무선효과가 단순할 때 무선기울기 모형을 사용하는 것은 실험조건 효과 추정에 얼마나 큰 영향을 미치는가? 본 연구결과는 그 반대의 경우만큼은 아니더라도 표준오차의 편향에 의한 유의성 검정결과에 영향을 미칠 수 있음을 보여준다. 편향의 크기가 상대적으로 크지 않았더라도 일관적으로 표준오차를 과대 추정하였고, 그에 따라서 모수복구비율이 최대 97%까지 상승하였다. 이 결과는 영가설을 그만큼 덜 빈번하게 기각한다는 의미이므로 검정력의 감소로 이어진다. 무선기울기 모형은 모형선택 전략에서 최대모형에 해당하므로 검정력은 다음 단락에서 논의된다.

실험자료 분석을 위한 선형 혼합효과 모형 설정에서 모형선택전략은 연구자들 사이에서

제안하는 바가 일치되어 있지 않은 주제이다 (Barr et al, 2013; Matuschek et al., 2017). 본 연구결과는 연구자가 자료의 무선효과 구조에 대한 사전지식이 없다면 모형비교를 먼저 시행한 후 선택된 모형으로 실험효과 검정을 하는 것이 비교절차 없이 무선키움기 모형을 사용하는 것에 비하여 실험연구자에게 더 나은 전략임을 시사한다. 그 첫 번째 근거로서, 적어도 본 시뮬레이션 조건에 따라 조작된 참가자 무선효과의 구조와 분산 크기의 모든 수준에서 모형비교 접근은 1종 오류의 비율을 유의수준에 가깝게 산출하였다. 반면, 최대모형 접근은 자료생성모형이 무선킨편 모형일 때 지나치게 보수적인 영가설 기각 비율을 보였다. 두 번째 근거로서, 모형비교 접근은 자료생성모형이 무선키움기 모형일 때는 최대모형 접근과 동등한 수준의 검정력을, 그리고 자료생성모형이 무선킨편 모형일 때는 최대모형 접근보다 더 높은 검정력을 산출하였다. 특히, 실험조건 효과크기가 작을 때 이 차이가 두드러졌다. 시뮬레이션 결과를 추가적으로 분석했을 때 서론에서 예측했던 대로, 참가자 무선효과의 공분산구조가 무선킨편이 아니고 분산 크기가 0.25 이상인 조건에서는 모형비교 과정에서 최소 99% 이상 무선키움기 모형이 분석모형으로 선택되었다. 반면, 참가자 무선효과의 공분산구조가 무선킨편 모형인 조건에서는 무선키움기 모형이 분석모형으로 선택된 비율이 최대 1.2% 이하였다. 그렇다면 남아있는 질문 중 한 가지는 최대모형 접근의 사용을 제안한 Barr 등(2013)과 본 연구에서 다른 결론을 얻은 이유일 것이다. Barr 등(2013)의 시뮬레이션 조건 조작은 모든 무선효과의 분산과 공분산에 대한 모수를 균등분포로부터 매 시행에서 무선킨적으로 추출하는 방

식으로 이루어졌다. 이와 같은 조작은 많은 시행에서 현실적인 가정과 맞지 않는 모수의 조합으로부터 실험이 이루어질 가능성을 내포하고 있다. 또한, 서론에서 언급했듯, Barr 등(2013)에서 반영한 실험설계는 현실에서 시행되고 있는 실험설계에 비해 너무 간단해서 최대모형과 무선킨편 모형의 복잡성이 큰 차이가 나지 않는다. 반면, 본 연구에서는 실제 실험자료를 반영한 선행연구를 참고하여 시뮬레이션 조건과 수준을 설정하였고, Barr 등(2013)에 비하여 현실적인 실험설계에서 시뮬레이션 연구를 수행하였다.

표본크기의 맥락에서 본 연구결과는 기존 문헌에서의 결과와 어떻게 비교될 수 있는가? 본 연구와 기존문헌들(Barr et al., 2013; Matuschek et al., 2017)은 사용한 참가자 수와 문항 수의 크기뿐 아니라 시뮬레이션 조건에 사용한 모수가 모두 달라서 그 결과를 직접 비교하기는 어렵다. 다만, 24-50의 참가자 크기를 사용한 세 연구 모두에서 모수추정에서의 문제를 보고하거나 발견하지 못했고, 참가자 수를 시뮬레이션 조건으로 조작한 Luke (2017)에서도 30명 내외의 참가자 수는 1종 오류비율만을 고려했을 때 간단한 실험설계를 위해 적절한 크기임을 보여주었다. 본 연구에서는 추가적으로, 실험조건 효과의 크기를 조작하여 효과크기에 따라 연구자가 기대할 수 있는 검정력에 대한 증거를 제공하였다. 본 연구결과는 표준화한 단위에서 중간크기의 효과를 성공적으로 탐지하기 위해서 30명의 참가자 수는 충분하지만, 작은크기의 효과를 성공적으로 탐지하기 위해서는 그 이상의 참가자 수가 필요함을 보여준다.

본 연구의 제한점은 다음과 같다. 첫째, 본 연구에서 사용한 모형비교 접근에서의 방법이

일반적인 의미에서의 모형비교보다 단순하다. 본 연구에서는 무선절편 모형과 무선기울기 모형만을 비교한 후 선택된 모형을 분석모형으로 사용했다. 반면, 일반적으로 모형비교는 모형의 복잡성에 미치는 요소들을 다양하게 조합하여 후보모형을 만들고 각 모형의 적합성을 순차적으로, 또는 동시에 비교하여 최종 모형을 선정한다(Henderson & Denison, 1989). 후보모형의 선정방법과 모형선택 전략은 실험설계가 복잡해질수록 본 연구에서 제안된 것보다 더 복잡해질 필요가 있다. 그 이유는 모형의 복잡성 면에서 실험설계가 복잡해질수록 무선절편 모형과 무선기울기 모형 사이의 모수의 수 차이가 급격히 벌어질 뿐 아니라 두 모형 모두 실제 자료의 무선효과 구조를 잘 반영하기 어려울 수 있기 때문이다. Barr 등(2013)과 Matuscheck 등(2017)에서는 무선효과의 구조를 설명하기 위해 서로 포함관계가 아닌(non-nested) 여러 후보모형을 제시하고 모형비교 과정에 포함하였다. 하지만, 그 모형 가운데 일부는 특별한 경우가 아니면 현실에서 만족하기 어려운 가정으로 보인다. 다만, 두 연구에서 모두 제안한 분산성분 모형은 복잡한 실험설계에서 중간단계의 대안모형으로 사용 가능할 것이다. 이 모형은 lmer 함수에서도 간단하게 설정할 수 있다(Bates et al., 2015). 추후 연구를 통해 실험조건의 수 또는 수준의 수가 큰 실험설계를 위한 모형설정 방법을 조사해볼 수 있을 것이다.

둘째, 본 연구에서는 모형비교 방법으로 LRT를 사용하였지만 실험자료 맥락에서 이 검정방법의 수행에 관한 연구가 더 필요하다. LRT는 비록 본 연구상황에서처럼 비교되는 두 모형이 서로 포함관계인 경우에 사용할 수 있지만 복잡한 모형에 추가되는 모수가 공분

산구조의 분산과 공분산일 때는 자유도가 모수의 수 차이인 카이제곱분포가 LRT를 위한 참조분포(reference distribution)로서 적절하지 않을 수 있다(Molenbergh & Verbeke, 2007; Zhang & Lin, 2008). 따라서, 실험자료 분석을 위한 선형 혼합효과모형에서 모형비교 방법 또한 추후 연구문제로 제안될 수 있다.

종합하여, 본 연구는 선형 혼합효과 모형을 실험자료 분석을 위한 도구로 활용할 때 종속변인의 변산성을 설명하는 요인들의 구조와 크기를 고려하여 분석모형을 선택하는 것이 중요하다는 것을 시사한다. 분석모형이 자료의 특성을 온전히 반영하지 못하거나 반대로 자료의 특성에 비해 지나치게 복잡하다면 실험조건 효과에 대한 추론이 부정확해질 수 있다. 분석모형을 선택하기 위한 전략으로서, 만약 연구자가 자료의 무선효과 구조와 크기에 대한 사전가설이 없다면 모형비교를 통해 선택된 모형에 기반하여 연구가설을 검증하는 것이 모형비교 절차 없이 분석모형을 선택하는 것에 비하여 선호된다. 제한된 시뮬레이션 조건과 수준만을 조작한 한계점에도 불구하고 본 연구결과는 선형 혼합효과 모형을 분석모형으로 고려하는 심리학 실험연구자들에게 실제적인 정보를 제공할 수 있을 것이다.

## 참고문헌

- Baayen, H., Davidson, D., & Bates, D. (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>
- 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., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Available from arXiv:1506.04967 (stat.ME).  
<https://doi.org/10.48550/arXiv.1506.04967>
- Bates D, Mächler, M, Bolker, B, Walker, S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.  
<https://doi.org/10.18637/jss.v067.i01>
- Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1), Article 251524590260351.  
<https://doi.org/10.1177/2515245920960351>
- Brown, M. B. & Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*, 69, 364-367.  
<http://dx.doi.org/10.1080/01621459.1974.10482955>
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.  
<https://doi.org/10.4324/9780203771587>
- Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, 23(2), 298-317.  
<https://doi.org/10.1037/met0000148>
- Fisher, R. A. (1921). Studies in crop variation: 1. An examination of the yield of dressed grain from broadbalk. *The Journal of Agricultural Science*, 11(2), 107-135.  
<https://doi.org/10.1017/S0021859600003750>
- Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511790942>
- Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician*, 60(1), 19-26.  
<https://doi.org/10.1198/000313006X90396>
- Gurka, M. J., Edwards, L. J., & Muller, K. E. (2011). Avoiding bias in mixed model inference for fixed effects. *Statistics in Medicine*, 30, 2696-2707.  
<https://doi.org/10.1002/sim.4293>
- Henderson, D. A. & Denison, D. R. (1989). Stepwise regression in social and psychological research. *Psychological Reports*, 64(1), 251-257.  
<https://doi.org/10.2466/pr0.1989.64.1.251>
- 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, 54-69.  
<https://doi.org/10.1037/a0028347>
- Kwok, O.-M., 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, 557-592.  
<https://doi.org/10.1080/00273170701540537>

- LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. of California Publications in Statistics*, 1, 277-330.  
<https://worldcat.org/ko/title/1841672>
- Lee, W. (2020). Linear mixed-effects model for analyzing experimental data. *The Korean Journal of Cognitive and Biological Psychology*, 32(2), 197-211.  
<https://doi.org/10.22172/cogbio.2020.32.2.006>
- Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violation revisited. A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66(4), 579-619.  
<https://doi.org/10.3102/00346543066004579>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavioral Research*, 49, 1494-1502.  
<https://doi.org/10.3758/s13428-016-0809-y>
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3), 86-92.  
<https://doi.org/10.1027/1614-2241.1.3.86>
- Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, 6, Article 2.  
<https://doi.org/10.3389/fpsyg.2015.00002>
- 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.  
<https://doi.org/10.1016/j.jml.2017.01.001>
- Maxwell, S. E., & Delaney, H. D. (2017). *Designing experiments and analyzing data* (3rd Ed.). Lawrence Erlbaum Associates.  
<https://doi.org/10.4324/9781315642956>
- Meteyard, L., & Davies, R. A. (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>
- Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science*, 15, 125-130. <https://doi.org/10.2172/1054744>
- 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>
- 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>
- Pinheiro, J. & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer.  
<https://doi.org/10.1007/b98882>
- Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41(3), 416-426.  
<https://doi.org/10.1006/jmla.1999.2650>
- Rao, P. (1971). Some notes on misspecification in

- multiple regressions. *The American Statistician*, 25(5), 37-39.  
<https://doi.org/10.1080/00031305.1971.10477302>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.  
<https://uk.sagepub.com/en-gb/asi/hierarchical-linear-models/book9230>
- Scariano, S. M. & Davenport, J. M. (1987). The effects of violations of independence assumptions in the One-way ANOVA. *The American Statistician*, 41(2), 123-129.  
<https://doi.org/10.1080/00031305.1987.10475459>
- Singmann, H., & Kellen, D. (2020). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology*. Psychology Press.  
<https://doi.org/10.4324/9780429318405-2>
- 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>
- Tomarken, A. J. & Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*, 99(1), 90-99.  
<https://doi.org/10.1037/0033-2909.99.1.90>
- Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91, 217-221.  
<https://doi.org/10.2307/2291398>
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4), 595-601.  
<https://doi.org/10.1214/aoms/1177729952>
- Zhang, D., & Lin, X. (2008). Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In *Random effect and latent variable model selection* (pp. 19-36). Springer, New York, NY.  
[https://doi.org/10.1007/978-0-387-76721-5\\_2](https://doi.org/10.1007/978-0-387-76721-5_2)
- 1차원고접수 : 2022. 07. 18.  
2차원고접수 : 2022. 09. 01.  
2차원고접수 : 2022. 10. 11.  
최종게재결정 : 2022. 10. 21.

## Comparison of procedures for detecting the experimental condition effect in a linear mixed-effects model

Yu Cheol Shin

Wooyeol Lee

Department of Psychology, Chungbuk National University

In a linear mixed-effects model for psychological experimental data analysis, the effect of the model selection procedure on detecting the experimental condition effect was investigated through a Monte Carlo simulation study. Specifically, while changing the complexity of the random effect structure of the data-generating model, the type I error rate and power were compared between the model selection strategies. As a result, the maximal model approach (or the random slope model) showed relatively low statistical power under the condition that the structure of the random component of the data was simple. On the other hand, when the model comparison approach was used, the Type I error rate approached the significance level, and the power was superior to or equivalent to that of the maximal model approach in all simulation conditions of this study. Finally, we discussed the points that experimental researchers should consider when using the linear mixed-effects model as an analysis tool.

*Key words* : linear mixed-effects models, experimental data, Monte Carlo simulation, parameter recovery, Type I error, power