

# Linear Mixed-Effects Model for Analyzing Experimental Data\*

Wooyeol Lee†

Department of Psychology, Chungbuk National University

This paper introduces a linear mixed-effects model with crossed random effects as an alternative to repeated measures analysis of variance (RM-ANOVA). With lmer function in the lme4 package, one of the R packages, two kinds of the linear mixed-effects model are described regarding the model specification, an estimation method, model comparison criteria, and an inference method. The use of the linear mixed-effects model is illustrated through an empirical example data. The performance of hypothesis testing is compared via a Monte Carlo simulation study between the mixed-effects model and variance of analysis framework.

**Keywords:** linear mixed-effects model,  $F1/F2$  analysis, R package, repeated-measures ANOVA

1 차원고접수 20.02.07; 수정본접수: 20.04.09; 최종게재결정 20.04.17

심리학 실험 연구에서 측정하는 대상은 대부분 직접 관찰하기 어렵다. 예를 들어, 인지심리학 연구자들은 흔히 주의, 기억과 같은 인간의 인지 능력이 실험조건에 따라서 어떻게 달라지는지에 관심이 있다. 이때 측정하려는 대상은 직접 관찰할 수 없으므로, 인간의 행동이나 생리적 반응과 같이 간접 측정값을 사용하여 측정된다. 간접측정값은 측정하려는 대상과 관련 없는 변인들의 영향을 줄이기 위해 흔히 여러 시행에 걸쳐 측정한다.

실험조건 안에서의 시행들은 같은 문항을 반복적으로 사용하는 대신에 복수의 문항들로 이루어진다. 예를 들어, 참가자가 사람의 얼굴을 기억하고 재인하는 과제를 수행한다면, 한 조건 안에서 여러 사람의 얼굴 자극이 제시된다. 이는 학습효과 또는 단순 노출 효과와 같이 연구자가 원하지 않은 효과를 없애기 위함이다(Gordon & Holyoak, 1983). 또한, 참가자 내 설계는 인지심리학에서 빈번히 사용되는 실험설계 방법이다. 참가자 내 설계에서 참가자는 한 조건의 모든 수준에 노출된다. 참가자 내 설계를 사용하면 적은 수의 참가자로 높은 검정력(power)을 얻을 수 있는 장점이 있다. 이렇

게 얻어진 데이터에서 각 반응은 실험조건, 참가자, 그리고 문항에 의해 설명될 수 있다.

문항효과(item effect)는 종속변인을 측정하기 위해 사용된 문항이 반응에 영향을 미치는 것을 의미한다. 예를 들어, 반응의 정확률이 종속변인으로 측정될 때 문항에 따라서 정확률이 다르다면 문항의 난이도가 문항효과로 간주할 수 있다. 참가자의 개인차를 연구하는 측정 분야에서 문항 반응 이론(IRT: item response theory)을 통해 문항효과가 널리 연구되어 온 것에 비해, 집단 간 차이에 관심을 두는 실험심리학 분야에서 문항효과는 흔히 무시되어 왔다(문항 반응 이론의 개괄은 Embretson & Reise, 2000을 참고).

참가자 내 설계를 사용한 연구에서 행동 반응을 종속변인으로 사용할 때 전통적으로 반복측정 변량분석(RM-ANOVA: repeated measures analysis of variance)이 통계분석방법으로 가장 널리 사용되어왔다. 반복측정 변량분석을 사용하여 실험데이터를 분석할 때, 매 시행에서의 반응을 실험조건의 각 수준 안에서 평균한 반응이 종속변인으로 사용된다. 예를 들어, 참가자 내 설계에서 두 개의 조건이

\* 본 연구는 2019학년도 충북대학교 학술연구지원사업의 연구비 지원에 의해 연구되었습니다(과제번호 201910046401).

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

각각 두 개의 수준을 갖고 있다면, (총참가자 수) × 2 × 2 개의 크기를 가진 종속변인에 대해 반복측정 변량분석을 통해 조건의 상호작용과 주효과의 유의성을 검증한다.

위와 같은 분석에서는 참가자의 개인차만이 연구자가 조작한 실험조건 이외에 반응을 설명할 수 있는 유일한 요소로 가정된다. 그에 비해서, 문항의 변산성이 반응에 영향을 미칠 가능성은 무시한다. 이러한 가정의 근거는 문항들에 걸쳐 얻어진 평균값을 종속변인으로 사용하면 문항이 설명하는 분산은 한 조건 내에서 시행 수가 증가할수록 작아지리라는 것이다. 이는 생리적 반응을 측정하는 연구자의 관점에서 신호대 잡음비(signal-to-noise ratio)의 감소로 이해될 수 있다 (Luck, 2005). 하지만 유의미한 문항효과를 무시할 경우 1종 오류와 2종 오류가 모두 높아질 가능성이 있다는 사실이 알려져 왔다(Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013; Raaijmakers, Schrinemakers, & Gremmen, 1999). 즉, 실제 효과의 유무와 상관없이 통계적 판정의 오류를 초래할 확률이 증가할 수 있다.

언어심리학 분야에서는 문항효과를 고려하지 않은 분석의 문제점은 “language-as-fixed-effect fallacy”로 알려졌고 오래전부터 논의되어왔다(Baayen et al., 2008; Barr et al., 2013; Clark, 1973; Raaijmakers, 2003; Raaijmakers et al., 1999). Clark(1973)은 이 문제점에 대처하는 방안으로 Quasi-*F*를 사용할 것을 제안하였다. Quasi-*F*는 참가자와 문항처럼 두 개의 무선 요소(random component)가 있을 때 그것을 모두 고려한 통계량이다. 하지만 Quasi-*F*는 그 계산의 어려움 때문에 실제로 거의 사용되지 않았다. 대신, 참가자분석과 문항분석을 모두 사용하여 보고하는 방식이 언어실험의 결과를 보고할 때 관행적으로 사용되어왔다. 앞으로 이 방법을 *F1/F2* 분석이라 부른다. 이는 조건 간 평균 차이가 참가자분석(*F1*)과 문항분석(*F2*)에서 모두 유의할 때만 그 차이가 유의미하다고 간주하는 것이다.

한편, 선형 혼합효과 모형(LMM: linear mixed-effects model)은 최근 언어심리학을 비롯한 실험심리학 분야에서 실험데이터를 분석하는 방법으로 점차 사용빈도가 높아지고 있다. 2008년 Journal of Memory and Language의 특별호에서는 실험데이터를 다루기 위한 기존 방법들의 한계점을 논의하고 그것을 극복하기 위한 새로운 방법들을 제안하였다. Baayen 등(2008)은 변량분석의 대안으로서 선형 혼합효과 모형을 소개하였다. 특히, 참가자효과와 문항효과를 하나의 모형에 포함하여, 문항효과를 고려하기 위해 독립된 분석을 할 수밖에 없었던 *F1/F2*분석의 한계점을 극복하였다. 같은 저널에 게재된 Barr 등(2013)은 선형 혼합효과 모형 중에

서도 무선효과(random effect)를 실험설계에 따라 최대고려한 무선 기울기(random slope) 모형을 기본으로 사용할 것을 제안하였다.

Baayen 등(2008)과 Barr 등(2013)은 시뮬레이션 연구를 통해 특정 상황에서 *F1/F2*분석과 선형 혼합효과 모형의 수행을 1종 오류와 2종 오류의 관점에서 보여주었다. 하지만 이들이 사용한 실험설계 상황은 너무 단순하다는 한계점이 있다. 이들 논문에서는 두 개의 수준을 가진 하나의 실험조건 상황에서 선형 혼합효과 모형을 적용하였다. 이는 대응표본 *t*-검정이 적합한 실험설계방법으로, 심리학 실험에서 주요인 이상을 포함한 실험설계가 일반적임을 고려하면 너무 단순하다. 선형 혼합효과 모형에서 실험조건과 같이 범주형 변인이 독립변인으로 사용되면 범주(또는 실험조건의 수준)의 수가 증가함에 따라 무선효과의 구조가 복잡해진다. 따라서, 기존 문헌의 시뮬레이션 연구 결과는 경험적 연구자들에게 어떤 경우에 선형 혼합효과 모형이 변량분석보다 유용한지에 대한 정보를 충분히 제공해주지 못한다.

인지심리학 맥락에서의 선형 혼합효과 모형에 대한 튜토리얼로서 Singmann과 Kellen(2020)과 Winter(2013)를 참고할 수 있다. 이들은 인지심리학 연구자들에게 친숙한 예제를 사용하여 선형 혼합효과 모형의 사용을 보였으며, 모형의 구조와 종류, 모형 비교방법, 추론방법에 대해 자세한 설명을 포함하였다. 하지만, 이들은 변량분석과 비교하는 과정이 없이 선형 혼합효과 모형을 새로운 맥락에서 소개함에 그침으로써 변량분석에 익숙한 연구자들을 위해 새로운 모형이 언제, 어떤 경우에 유용한지에 대한 정보가 부족하다.

외국과 비교하면, 한국에서는 아직 변량분석의 대안으로서의 선형 혼합효과 모형은 사용되는 빈도가 낮다. ‘한국심리학회지:일반’과 ‘한국심리학회지:인지 및 생물’에 2008년 1월부터 2019년 12월 현재까지 출판된 논문을 대상으로 선형 혼합효과 모형의 사용빈도를 조사하였다. 조사 기간 동안 ‘한국심리학회지:일반’에는 실험 연구를 위해 선형 혼합효과 모형 결과를 보고한 논문이 단 한 건도 없었고, 하나의 논문에서 *F1/F2* 분석 결과를 보고하였다(Lee, 2009). ‘한국심리학회지:인지 및 생물’에는 네 편의 논문에서 선형 혼합효과 모형으로 분석한 결과를 보고하였고(Ahn, Kim, & Choi, 2019; Bae & Yi, 2019(b); Lee & Choi, 2019; Yoon, Kang, An, & Koh, 2010), 23편의 논문에서 *F1/F2* 분석 결과를 보고하였다.<sup>1)</sup> 조사 결과는 최근까지도 한국에서 실험

1) 23개 논문은 다음과 같다. Bae와 Lee(2017), Bae와 Yi(2010), Bae와 Yi(2016), Bae와 Yi(2019a), Bae와 Yi(2019b), Choi와 Koh(2012), Choi, Lee, 및 Nam(2008), Choi, Lee, Kang, 및 Nam(2015), Kang,

데이터를 분석하기 위해 선형 혼합효과 모형을 사용하는 빈도가 매우 낮음을 보여준다.

본 논문의 목적은 인지심리학 연구자들에게 실험 연구에서 얻어진 데이터를 분석하는 방법으로서 반복측정 변량분석의 대안으로서의 선형 혼합효과 모형을 소개하는 것이다. 이를 위해 기존 문헌들에서처럼 선형 혼합효과 모형의 구조와 종류, 모형 비교, 모형의 해석, 추론방법에 관해 기술한다. 추가로, 변량분석에 익숙한 연구자들이 선형 혼합효과 모형의 사용에 앞서서 흔히 갖는 궁금점에 대해 논의한다.  $2 \times 2$  참가자 내 설계에서의 상황을 설정하고, 수식과 함께 R(R Core Team, 2019)의 lme4 패키지의 lmer 함수를 가지고 시연을 보인다.

앞으로 본 논문은 다음과 같이 구성된다. 첫째,  $F1/F2$  분석방법에 관해 설명하고 그 분석의 가정과 한계점에 관해 기술한다. 둘째, 선형 혼합효과 모형의 설정, 추정, 모형 비교방법, 그리고 추론 및 결과 해석에 관해 기술한다. 셋째, 예제데이터를 가지고  $F1/F2$  분석과 선형 혼합효과 모형을 사용한 분석 결과를 비교한다. 넷째, 임의의 상황을 설정하고, 몬테카를로 시뮬레이션 실험을 통해 반복측정 변량분석,  $F1/F2$  분석과 선형 혼합효과 모형 사이에서 가설검정의 수행을 비교한다. 마지막으로, 연구의 함의와 한계점을 논의한다.

### F1/F2 분석

본 논문에 걸쳐, 데이터는 다음의 실험설계로부터 얻었음을 가정한다. 종속변인은 연속형 변인으로  $2 \times 2$ 의 참가자 내 설계를 통해 얻어진다. 즉, 한 참가자는 첫 번째 조건의 모든 수준( $a=1, 2$ )과 두 번째 조건의 모든 수준( $b=1, 2$ )에 노출된다.

종속변인은  $J$ 명의 참가자( $j=1, \dots, J$ )로부터 각 조건의 각 수준에서  $I$ 개의 문항( $i=1, \dots, I$ )을 통해 측정되며 동일한 문항 세트가 모든 수준에서 반복된다. 따라서 얻어지는 총반응수는  $N=2 \times 2 \times J \times I$ 이다.

### 분석방법

반응  $y_{jiab}$  이  $j$ 번째 참가자와  $i$ 번째 문항으로부터 첫 번째 조

Nam, Lim, 및 Nam(2016), Kim과 Park(2016), Koh, Hong, Yoon, 및 Cho(2008), Kwon과 Nam(2011), Lee, Woo, 및 Lee(2019), Lee와 Kwon(2012), Min과 Yi(2010), Nam, Baik, Lim, 및 Nam(2014), Noh, So, 및 Kim(2017), Park, Yi, Abe, 및 Liu(2008), Ryu, Nam, Kim, 및 Baik(2016), Yi와 Bae(2009), Yi, Park, Abe, Liu, 및 Zhang(2010), Yoon과 Koh(2010), Yoon, Kang, An, 및 Koh(2010).

건의  $a$ 번째 수준과 두 번째 조건의  $b$ 번째 수준에서 얻어졌다고 하자. 반복측정 변량분석을 사용하여 참가자분석을 시행한다고 하면 먼저 종속변인을 (1)과 같이 새롭게 정의한다.

$$y_{jab}^* = \bar{y}_{j.ab} = \frac{\sum_{i=1}^I y_{jiab}}{I} \quad (1)$$

위에 정의된 새로운 종속변인은 한 참가자가 실험조건의 각 수준에서 보인 반응을 문항에 걸쳐 평균한 값을 의미한다.

변량분석은 일반적으로 종속변인의 제곱합을 실험조건이 설명하는 제곱합과 조건이 설명하지 못하는 제곱합으로 분해한 후 그 평균제곱(분산)의 크기를 비교한다. 반복측정 변량분석은 참가자효과를 통제하기 위해 전체 제곱합을 실험조건, 참가자, 그리고 나머지에 의해 설명되는 제곱합으로 분해한다. 두 조건의 주효과와 상호작용은 각 효과에 대한 분산과 참가자효과를 제외한 후 설명되는 분산을  $F$  통계량을 사용하여 비교한다.

반면, 문항 분석( $F2$ )은 새로운 종속변인을 (2)와 같이 정의한다.

$$y_{iab}^* = \bar{y}_{.iab} = \frac{\sum_{j=1}^J y_{jiab}}{J} \quad (2)$$

분석방법은  $F1$  분석과 같다.

### 가정과 한계점

반복측정 변량분석의 가정 가운데, 문항효과가 무시되었을 때 위협받을 수 있는 가정은 독립성 가정이다. 독립성 가정을 만족하려면, 실험조건 이외에 참가자의 개인차가 유일하게 반응을 설명해야 한다. 하지만, 실험에 사용된 문항이 반응에 영향을 미친다면 개별 반응을 문항에 걸쳐 평균한 참가자의 반응을 종속 변인으로 사용하더라도 그 변인의 변산성에 영향을 미칠 수 있다. 이는 검정통계량이 영가설 아래의  $F$ 분포를 정확히 따르지 않는 결과를 가져온다. Clark(1973)은 문항효과를 무시할 경우 검정통계량  $F$ 의 크기가 과대 추정되어 1종 오류의 가능성이 증가함을 보였다.

만약 문항효과가 존재하더라도 그것이 참가자의 효과와 독립적이라면, 문항의 수를 증가시킨 후 문항에 걸쳐 평균한 값을 종속변인으로 사용함으로써 문항의 효과를 줄일 수 있다. 이는 평균값을 사용하면 문항의 효과가 그 수의 제공근

에 비례하여 작아지기 때문이다. 하지만 만약 문항의 수가 작거나, 문항의 효과와 참가자의 효과가 상호작용한다면, 평균값을 종속변인으로 사용하는 것이 문제를 해결해주지 못한다.

F1/F2 분석의 더욱 큰 문제점은 이 방법이 통계적 이론에 근거하기보다는 임시방편의 방법(ad-hoc)에 지나지 않는다는 점이다. F1/F2 분석은 참가자분석과 문항분석을 하나의 데이터에 중복하여 시행함으로써 참가자분석만을 했을 때보다 더 보수적으로 효과를 검정하는 목적이 있다. 하지만 참가자분석과 문항분석을 각각 유의수준  $\alpha = .05$  수준에서 시행했다고 하더라도 F1/F2 분석이 항상  $\alpha = .05$ 에서 올바른 결론임을 보장하지 않는다. 그리고 각각의 검정 역시 위반된 가정을 토대로 하기때문에 개별 검정의 결과를 신뢰하기 어렵다.

### 선형 혼합효과 모형

선형 혼합효과 모형은 연속 변인인 종속변인을 고정효과(fixed effect)와 무선평과(random effect)의 합으로 표현한 통계모형이다(Gelman & Hill, 2007; Pinheiro & Bates, 2000; Raudenbush & Bryk, 2002; Snijders & Bosker, 2011). 선형 혼합효과 모형은 사용되는 분야에 따라서 다층모형(multilevel model) 또는 위계적 선형 모형(hierarchical linear model)으로도 알려져 있다(Raudenbush & Bryk, 2002; Snijders & Bosker, 2011)<sup>2)</sup>. 이 모형은 종단 연구, 메타분석, 그리고 자료가 국가/지역/학교/병원 등의 단위로 층화 추출(stratified sampling)되었을 때와 같이 자료 특성이나 자료 수집 과정 때문에 반응이 독립성 가정을 만족하지 못하는 여러 맥락에서 사용되어왔다.

한편, 실험데이터에서 종속변인은 실험조건 외에도 참가자와 문항에 의해 동시에 설명되므로 복수의 무선평과가 필요하다. 따라서 본 논문에서 선형 혼합효과 모형은 구체적으로 교차 무선평과를 가진 선형 혼합효과 모형(linear mixed-effects model with crossed random effects)을 가리킨다.

이 장에서는 두 가지 선형 혼합효과모형의 설정과 추정방법, 모형 비교방법, 그리고 추론방법에 대해 개괄한다.

### 모형 설정

무선 절편 모형(M1).  $2 \times 2$  참가자 내 설계에서 모든 문항이 모든 조합의 수준에서 반복되어 사용될 때 무선 절편만을 가진(random intercept only) 선형 혼합효과 모형을 다음과 같이 설정하면 종속변인  $y_{jiab}$ 은 (3)과 같은 식에 의하여 설명될 수 있다. 앞으로 이 모형을 편의상 M1으로 지칭한다.

$$y_{jiab} = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2b} + \beta_3 x_{1a} x_{2b} + s_j + w_i + e_{jiab} \quad (3)$$

선형 혼합효과 모형과 같은 선형 모형(linear model)에 범주형 독립변인이 포함될 경우 코딩 스킴(coding scheme)을 필요로 한다. 더미 코딩(dummy coding)을 사용할 경우  $x_{1a}$ 와  $x_{2b}$ 은 각각 첫 번째 실험조건( $x_{11} = 0, x_{12} = 1$ )과 두 번째 실험조건( $x_{21} = 0, x_{22} = 1$ )을 나타낸다. 두 조건 사이의 상호작용 효과를 나타내기 위해  $x_{1a} x_{2b}$  항이 모형에 포함되어 있다.

고정효과  $\beta$ 는 각각 다음을 의미한다.  $\beta_0$ 은 a=1, b=1 수준에서의 평균 반응,  $\beta_1$ 은 a=1 수준과 a=2 수준에서의 평균 반응 차이,  $\beta_2$ 은 b=1 수준과 b=2 수준에서의 평균 반응 차이, 그리고  $\beta_3$ 은 두 실험조건 사이의 상호작용을 나타낸다.

독립변인의 각 수준 사이의 평균 차이가 없다는 영가설만을 검정하는 반복측정 변량분석과 달리, 선형 혼합효과 모형은 각 수준의 평균값이 모수로서 추정되거나 고정효과의 합으로 계산된다.  $2 \times 2$  설계의 각 실험조건의 평균값은 표 1과 같이 모수화(parameterization)된다.

두 무선평과  $s_j$ 와  $w_i$ 는 j번째 참가자의 반응 경향성과 i번째 문항의 반응 경향성을 각각 나타낸다. 고정효과와 마찬가지로 개별 참가자 또는 반응의 평균은 효과들의 합으로 계산된다. 예를 들어, 1번 참가자가 a=1, b=1 수준에서 보인 평균 반응값은  $\beta_0 + s_1$ 로 계산된다. 마찬가지로, a=2, b=2 수준

**Table 1.** Parametrization for mean according to experimental conditions

Experimental condition	Parametrization for mean
a=1, b=1	$\beta_0$
a=2, b=2	$\beta_0 + \beta_1$
a=1, b=2	$\beta_0 + \beta_2$
a=2, b=2	$\beta_0 + \beta_1 + \beta_2 + \beta_3$

2) 본 논문에서 소개되는 모형은 엄밀히 말하면 참가자와 문항이 반응과 중복적으로 내재(nested) 관계가 있으므로 위계적 구조에서 벗어나 있다. 하지만 관련 문헌들에서 특수한 형태의 다층모형으로 다루어지고 있다(Snijders & Bosker, 2011).

에서 사용된 2번 문항이 참가자에 걸쳐 얻어진 반응값의 평균은  $\beta_0 + \beta_1 + \beta_2 + \beta_3 + w_2$ 로 계산된다. 무선효과는 개별 구성원이 가지고 있는 크기가 모수로서 추정되는 대신, 구성원에 걸쳐 평균이 0으로 고정된 정규분포를 따른다고 가정한다. 개별 참가자와 개별 문항의 무선효과는 각각  $s_j \sim N(0, \tau^2)$ 와  $w_i \sim N(0, \omega^2)$ 를 따르며, 분산인  $\tau^2, \omega^2$ 가 추정해야 할 모수가 된다. 개별 구성원의 효과 크기는 분포의 분산이 추정된 후에 독립적인 과정을 통해 계산된다.

마지막으로 각 반응의 잔차 역시  $e_{jiab} \sim N(0, \sigma^2)$ 의 분포를 따른다고 가정한다. 종합해서, 위 모형이 가지고 있는 모수의 수는 총 7개( $\beta_0, \beta_1, \beta_2, \beta_3, \tau^2, \omega^2, \sigma^2$ )가 된다.

lmer 함수에서 위 모형은 (4)와 같이 설정된다.

$$y \sim a*b + (1|subj)+(1|item) \quad (4)$$

위 표현은  $\sim$ 를 기준으로 좌변의 y는 우변의 효과들의 합으로 나타냄을 의미한다. 고정효과  $a*b$ 는 두 독립변인 a, b의 주효과와 상호작용을 포함한다. 무선효과는 1를 기준으로 왼쪽에는 무선효과에 대응하는 고정효과, 오른쪽에는 무선효과를 갖는 단위를 표시한다. 왼쪽의 1은 절편(intercept), 혹은 전체 평균(overall mean)을 가리킨다. 예를 들어, (1|subj)은 개별 참가자가 가진 하나의 무선효과 또는 조건에 상관없이 보이는 반응 경향성을 의미한다.

무선 기울기 모형(M2). 무선 기울기(random slope)를 가진 선형혼합효과 모형은 (5)와 같이 나타낼 수 있다. 편의상 이 모형을 앞으로 M2라고 부른다.

$$y_{jiab} = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2b} + \beta_3 x_{1a} x_{2b} + s_{0j} + s_{1j} x_{1a} + s_{2j} x_{2b} + s_{3j} x_{1a} x_{2b} + w_i + e_{jiab} \quad (5)$$

$$\begin{bmatrix} s_{0j} \\ s_{1j} \\ s_{2j} \\ s_{3j} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & & & \\ \tau_{10} & \tau_1^2 & & \\ \tau_{20} & \tau_{21} & \tau_2^2 & \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_3^2 \end{bmatrix} \right) \quad (6)$$

M2는 M1의 모수를 모두 포함하면서 추가적인 복잡성을 가지고 있는 구조이다. 위 모형이 M1과 다른 점은 한 참가자가 4개의 무선효과를 갖는다는 점이다.  $s_{0j}, s_{1j}, s_{2j}, s_{3j}$ 는  $\beta_0, \beta_1, \beta_2, \beta_3$ 에 대한 j번째 참가자의 개인차를 의미한다. 즉, M2는 a=1, b=1 수준에서의 평균 반응, 두 실험조건

의 주효과, 그리고 상호작용 효과에 대해 개인차를 표현할 수 있다. 4개의 참가자 무선효과는 (6)과 같이 각 효과의 평균이 0으로 고정된 다변량 정규분포를 따른다고 가정한다. 다변량 정규분포는 각 효과의 분산과 효과들 사이의 공분산을 모수로 갖는다. 따라서 M2가 가지고 있는 총 모수의 수는 16개(고정효과 4개 + 참가자 무선효과의 분산/공분산 10개 + 문항 무선효과의 분산 1개 + 잔차의 분산 1개)이다.

lmer 함수에서 위 모형은 (7)과 같이 설정된다.

$$y \sim a*b + (a*b|subj)+(1|item) \quad (7)$$

M1의 설정 (4)과 비교하여 다른 점은 참가자 무선효과의 설정을 (a\*b|subj)으로 표현한 부분이다. 이는 모든 고정효과에 대한 참가자 무선효과가 존재함을 의미한다.

### 추정

선형 혼합효과 모형의 추정을 위해서는 최대우도추정(MLE: maximum likelihood estimation)과 베이지언 추정이 사용될 수 있다. lmer 함수는 최대우도추정을 사용한다. 선형 혼합효과 모형을 위한 베이지언 접근은 Gelman과 Hill (2007)을 참고할 수 있다. 모형에 무선효과가 포함되어 있을 경우, 최대우도추정을 사용할 때 우도를 계산하기 위해 반복적인(iterative) 절차를 필요로 한다. 따라서 선형 혼합효과 모형의 추정값은 통계 프로그램이 반복적인 절차를 위해 사용하는 알고리즘에 따라 그 값이 정확히 일치하지 않을 수도 있다. lmer 함수는 PLS(penalized least squares) 알고리즘을 사용한다고 알려져 있다(Bates et al., 2015).

### 모형 비교

연구자들은 연구자가 설정한 조건의 효과, 즉 고정효과의 유의성에 관심이 있다. 하지만 고정효과의 검정결과는 모형이 올바르게 설정되었을 때만 그 결과를 믿을 수 있다. 선형 혼합효과 모형을 사용할 때는, 무선효과의 구조를 결정하기 위해 M1, M2를 비교한 후 선택된 모형에 기반하여 그 결과를 해석한다.

M1와 M2의 관계와 같이 하나의 모형이 다른 모형의 특수한 사례일 때 널리 사용되는 모형 비교의 방법으로는 우도비검정(LRT: likelihood ratio test)이 있다. 우도비검정은 두 모형의 적합도를 나타내는 값인 편차(deviance) 차이가 카이제곱분포를 따른다는 가정을 기반으로 한다. 보통의 경우, 카이제곱분포의 자유도로서 두 모형이 가진 모수의 개수 차이를 사용한다. 하지만 그 차이에 해당하는 모수가 분산일

때는 이 편차의 차이가 어떤 혼합분포(mixture distribution)을 따른다고 알려져 있다(Molenbergh & Verbeke, 2007; Zhang & Lin, 2008). 본 논문에서 사용된 예에서는 M1에 비해 M2에 추가된 9개의 모수 가운데 분산을 나타내는 모수가 3개( $\tau_1^2, \tau_2^2, \tau_3^2$ )이므로  $\chi_{6,.95}^2$ 과  $\chi_{9,.95}^2$ 의 50:50 혼합인 14.76을 임계값(critical value)으로 사용한다.

**추론**

고정효과는 모형 비교과정에서 선택된 최종 모형을 기반으로 하여 추론이 이루어진다. 두 개의 주효과와 상호작용에 대한 추론은 각각  $\beta_1, \beta_2, \beta_3$ 을 검정함으로써 이루어진다. Wald 검정은 고정효과에 대한 가설 검정을 위해 흔히 사용하는 방법이다. 실험조건에 대한 고정효과  $\beta$ 에 대한 영가설과 대안 가설은 (8)과 같다.

$$H_0 : \beta = 0 \text{ vs. } H_1 : \beta \neq 0 \tag{8}$$

검정통계량  $T_{Wald}$ 는 추정값  $\hat{\beta}$ 를 그 표준오차  $SE(\hat{\beta})$ 로 나눈 값을 사용한다. 표본의 크기가 충분히 클 때,  $T_{Wald}$ 는 영가설 아래에서  $Z$  분포를 따른다는 가정을 기반으로 한다. 따라서, 검정통계량이 1.96보다 클 때, 검정하려는 고정효과가 통계적으로 유의하다는 결론을 내린다.

**적용**

이 장에서는 예제데이터를 사용하여 선형 혼합효과 모형을 적합하고, 그 결과를 F1/F2 분석 결과와 비교한다. 결과의 반복을 위해 예제데이터와 선형 혼합효과 모형을 위한 R코드를 다음의 저장소에서 내려받을 수 있다: <https://osf.io/vx39k/>.

**데이터 설명**

Song과 Lee(2009)는 한국어 화자들이 문장의 의미를 이해하는데 있어서 주어 연속성과 주어 언급 여부의 효과를 알아보았다. 연구자들은 만약 이전 문장들과 검사문장의 주어가 다르면 검사문장에 주어가 언급되는 것이 생략될 때보다 문장의 의미를 더 이해하기 쉽게 만들지만, 주어가 이전 문장들과 같을 때는 검사문장의 주어가 생략되는 것이 언급되는 것보다 문장의 이해를 더 쉽게 만들 것으로 예측하였다. 실험과제를 통해서 연구자들은 성인 참가자들을 대상으로 자기 조절 읽기 과제(self-paced reading task)를 실시하고 검사문

장의 반응시간을 측정하였다. 주어의 연속성(연속 대 비연속) × 주어의 언급 여부(언급 대 생략)를 참가자 내 변인으로 하는 2 × 2 반복측정 변량분석 결과 주어 연속성의 주효과만 유의미하게 나타났다. 즉, 참가자들은 검사문장에서 주어의 언급 여부와는 상관없이 이전 문장들과 검사문장의 주어가 같을 때 문장의 의미를 더 빨리 이해하였다.

본 연구에서 사용된 데이터는 Song과 Lee(2009)의 실험 자극을 수정하여 실험을 진행하여 얻어졌고 주 저자의 허락을 받아 본 논문에 사용되었다.3) 새로 실시한 실험에는 총 25명이 참여하였다. 총 32개의 검사문장이 주어 연속성, 주어 언급 여부의 각 수준에 걸쳐서 사용되었다. 표 2에서 보이는 것처럼, 실험은 라틴 방진 설계(Latin-square design)를 사용하였다. 실험 자극은 네 가지의 리스트로 나누어졌고, 각 참가자에게 네 가지 중 하나의 리스트가 주어졌다. 하나의 리스트에는 두 조건의 모든 수준에서 8개의 검사문장이 겹치지 않게 사용되었다. 리스트에 걸쳐 검사문장은 역균형화되었다.

종합하여, 얻어진 총반응 수는 800(=25×32)개였다. 각 참가자의 평균값에서 3 표준편차 범위 밖으로 정의된 이상값 기준에 의해 11개가 분석에서 제외되어 총 789개(98.6%)의 반응이 사용되었다.

**기술통계와 F1/F2 분석**

표 3에 조건에 따른 검사문장의 평균 반응시간과 그 표준편차가 제시되었다. 참가자가 검사문장에 반응한 시간의 전체 평균은 1815.16(sd= 1007.75)였다. 반응시간은 두 실험조건의 수준에 따라 변화가 있었다. 검사문장의 주어가 맥락문장과 같을 때, 주어 언급 조건과 주어 생략 조건에서의 반응시간은 각각 1872.74(sd=939.85)와 1456.27(sd=740.46)였다. 검사문장의 주어가 맥락문장과 달랐을 때, 주어 언급 조건과 주어 생략 조건에서 검사문장의 평균 반응시간은 2068.16(sd= 1107.51)과 1869.39(sd=1107.45)였다.

언어실험 연구에서의 관행대로 참가자분석과 문항분석을 시행하였다(표 4). 참가자분석 결과 주어 연속성과 주어 언급 여부 사이의 상호작용( $F(1,24)=16.754, p<.001$ )과 두 주효과(주어연속성:  $F(1,24)=13.589, p<.005$ ; 주어언급여부:  $F(1,24)=6.333, p<.05$ )가 모두 유의하였다. 반면, 문항분석에서는 상호작용은 유의하지 않았고( $F(1,31)=1.842, p=.184$ ), 두 주효과만 유의하였다(주어연속성:  $F(1,31)=28.318, p<.001$ ; 주어언급여부:  $F(1,31)=16.851, p<.001$ ).

3) 두 가지가 수정되었다. 첫째로, 맥락문장의 수를 두 개에서 세 개로 늘리고, 둘째로 맥락문장의 형식을 “~에게”의 형식으로 통일하였다.

**Table 2.** Illustration for the Latin square design

	Item 1-8	Item 9-16	Item 17-24	Item 25-32
List 1	a=1, b=1	a=1, b=2	a=2, b=1	a=2, b=2
List 2	a=1, b=2	a=1, b=1	a=2, b=2	a=2, b=1
List 3	a=2, b=1	a=2, b=2	a=1, b=1	a=1, b=2
List 4	a=2, b=2	a=2, b=1	a=1, b=2	a=1, b=1

**Table 3.** Descriptive statistics (msec, standard deviation in parenthesis)

	Subject continued	Subject shifted	Total
Subject mentioned	1872.74(939.85)	2068.16(1107.51)	1969.71(1029.81)
Subject omitted	1456.27(740.46)	1869.39(1107.45)	1661.79(962.38)
Total	1663.98(870.08)	1968.27(1110.53)	1815.16(1007.75)

**Table 4.** Results for *F1/F2* analyses

Effect	<i>F1</i> analysis		<i>F2</i> analysis	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Subject continuity (A)	16.754	0.000	28.318	0.000
Subject form (B)	13.589	0.001	16.385	0.000
A*B	6.333	0.019	1.842	0.184

**선형 혼합효과 모형**

두 종류의 선형 혼합효과 모형(즉, M1과 M2)을 예제데이터에 적합하였다. 표 5는 lme4 패키지의 summary 함수를 통해 얻은 각 모형의 모수의 수와 편차를 보여준다.

우도비검정은 두 모형 사이의 편차 차이를 적절한 카이제곱분포와 비교한다. 표 4의 M1과 M2 편차의 차이값인 37은 이전 장에서 구한 임계값 14.76보다 큰데, 이것은 M2에 추가된 모수들이 유의미하게 모형 적합성을 증가시켰음을 의미한다. 따라서 우도비검정에 의해 M2가 선택되었다고 말할 수 있다.

다음으로, 모형 선택과정에서 선택된 M2를 기반으로 하여 결과를 해석한다. 하지만 본 논문에서는 시연을 위한 목적으로 summary 함수를 통해 얻은 M1과 M2의 모수에 대한 추정값을 제시하였다(표 6). M1과 M2에서 추정값과 그 추론 결과는 거의 일치하였다.

**Table 5.** Model comparison results

	M1	M2
Number of parameters	7	16
Deviance	12,863	12,826

각 고정효과는 다음과 같이 해석된다.  $\hat{\beta}_0=1873.14$ 는 주어 연속/주어 언급 조건에서의 평균 반응시간에 대한 추정값이다.  $\hat{\beta}_1=194.94$ 는 앞의 조건과 주어 비연속/주어 언급 조건 사이의 평균 반응 차이에 대한 추정값으로서 검사문장의 주어가 맥락문장과 바뀌었을 때, 주어 언급 조건에서의 평균 반응시간의 추정값은  $\hat{\beta}_0+\hat{\beta}_1=2068.08$ 이 된다.  $\hat{\beta}_2=-418.31$ 은 주어 연속/주어 언급 조건과 주어 연속/주어 비언급 조건 사이의 차이값에 대한 추정값을 나타낸다. 즉, 검사문장의 주어가 맥락문장과 같을 때, 주어 비언급 조건에서의 평균 반응시간에 대한 추정값은  $\hat{\beta}_0+\hat{\beta}_2=1454.83$ 이 된다. 마지막으로,  $\hat{\beta}_3=218.38$ 은 주어 연속성의 효과가 주어 언급 여부에 따라서 달라지는 정도를 가리킨다. 따라서 검사문장의 주어가 맥락문장과 달라질 때, 주어 비언급 조건에서의 평균 반응시간의 추정값은 모든 고정효과를 합한 값인  $\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3=1868.15$ 가 된다.

Wald 검정을 통한 유의성 검정결과, 주어 연속성과 주어 언급의 주효과, 그리고 상호작용효과가 모두 통계적으로 유의하였다. 상호작용 효과를 예를 들어 검정절차를 살펴보면, 검정통계량 2.001은 상호작용 효과의 추정값( $\hat{\beta}_3=218.38$ )을

**Table 6.** Results from linear mixed-effects models

Parameter	M1				M2			
	Estimate	SE	Z	p	Estimate	SE	Z	p
$\beta_0$	1872.75	130.30	14.372	.000	1873.14	113.89	16.447	.000
$\beta_1$	196.70	78.76	2.497	.006	194.94	94.09	2.072	.019
$\beta_2$	-418.53	78.33	-5.343	.000	-418.31	104.49	-4.004	.000
$\beta_3$	221.07	114.14	1.989	.023	218.38	109.16	2.001	.023
$sd(s_j)$ ( $sd(s_{0j})$ )	509.7	.	.	.	403.8	.	.	.
$sd(s_{1j})$	.	.	.	.	280.2	.	.	.
$sd(s_{2j})$	.	.	.	.	362.9	.	.	.
$sd(s_{3j})$	.	.	.	.	116.4	.	.	.
$corr(s_0, s_1)$	.	.	.	.	0.33	.	.	.
$corr(s_0, s_2)$	.	.	.	.	-0.31	.	.	.
$corr(s_0, s_3)$	.	.	.	.	0.90	.	.	.
$corr(s_1, s_2)$	.	.	.	.	0.79	.	.	.
$corr(s_1, s_3)$	.	.	.	.	-0.11	.	.	.
$corr(s_2, s_3)$	.	.	.	.	-0.69	.	.	.
$sd(w_i)$	335.2	.	.	.	340.2	.	.	.
$\sigma$	779.6	.	.	.	747.9	.	.	.

그 표준오차( $SE(\hat{\beta}_3)=109.16$ )로 나눔으로써 계산되었다. 추정값의 분포가 정규분포이고 그 표준편차가 추정값의 표준오차라는 가정 아래에서 검정통계량은 Z분포를 따르게 되며, 검정통계량은 유의수준  $\alpha=.05$ 에서의 임계값 1.96과 비교하여 통계적 유의성을 판단한다.

summary 함수를 통해 보이는 결과에서 무선효과는 분산과 공분산 대신 표준편차와 상관계수의 단위로 제시된다. M1에서 참가자 무선효과의 표준편차는 509.7였다. 이는 참가자들 간의 반응 변산성으로 해석될 수 있고, 이 크기는 모든 조건에 걸쳐 같다고 가정한다. 문항 무선효과의 표준편차는 335.2였고, 마찬가지로 문항효과와 변산성의 크기로 해석할 수 있다. 모형이 종속변인을 설명하고 남은 잔차의 표준편차는 779.6이었다. M2에서 참가자 무선효과의 표준편차는 403.8, 280.2, 362.9, 116.4였다. 이는 고정효과  $\beta_0, \beta_1, \beta_2, \beta_3$ 에 대응하는 참가자의 개인차에 대한 변산성의 크기로 해석할 수 있다. 예를 들어, 주어 연속/주어 언급 조건에서 검사문장에 대한 평균 반응시간은  $\hat{\beta}_0=1873.14$ 이고 참가자에

따른 개인차는  $\hat{\tau}_0 = sd(\hat{s}_{0j}) = 403.8$ 의 표준편차를 가진 정규분포를 따른다. 만약 한 참가자가 평균보다 1 표준편차 큰 개인차를 보였다면 그 참가자의 평균 반응시간의 추정값은  $1873.14+403.8=2276.94$ 이다. 4개의 참가자 무선효과는 다변량 정규분포를 따른다고 가정하여 무선효과 사이의 상관계수가 모수로서 추정된다. M2에서 참가자 무선효과들은  $-.69$  ( $corr(s_2, s_3)$ )에서  $.90$  ( $corr(s_0, s_3)$ ) 사이의 상관관계를 보였다.

무선효과는 분산(또는 표준편차)의 형태로 모수화되어 있지만, 개별 참가자와 문항의 효과를 추정하는 것도 가능하다. ranef 함수는 무선효과가 설정된 분포를 따른다는 가정 아래에서의 효과를 추정하여 보여준다. 예를 들어, 1번 참가자의 무선효과로  $[\hat{s}_{01}, \hat{s}_{11}, \hat{s}_{21}, \hat{s}_{31}]=[-278.41, -188.77, -84.81, -48.68]$ 을 얻었다. 이 참가자는 각각  $\beta_0, \beta_1, \beta_2, \beta_3$ 에 해당하는 효과에서 평균적인 참가자보다 낮은 효과를 보였음을 알 수 있다. 마찬가지로, 1번 문항의 무선효과로  $\hat{w}_1 = 128.90$ 을 얻었는데, 이를 통해 참가자들은 이 문항을 응답하는데 평균적인 문항보다 반응시간이 더 길었음을 알 수

있다.

개별 반응의 추정값은 추정된 고정효과와 무선평과와의 합으로 나타낼 수 있다. lme4 패키지의 predict 함수는 개별 반응의 추정값을 보여준다. 예를 들어, 첫 번째 참가자가 주어 비연속/주어 언급 조건에서 첫 번째 문항에 대해  $\hat{y}_{2111}=1729.79$ 의 반응시간을 보였음을 예측했다. 이는 해당하는 효과들의 합  $\hat{\beta}_0 + \hat{\beta}_1 + \hat{s}_{01} + \hat{s}_{11} + \hat{w}_1 = 1729.79$ 로 계산될 수 있다.

### 비교

F1/F2 분석에서는 주어 연속성과 주어 언급여부 사이의 상호작용이 유의하지 않았으며, 상호작용은 참가자분석에서만 유의하였고 문항 분석에서는 그 크기가 통계적으로 유의하지 않았다. 이처럼 F1/F2 검정은 두 번의 검정이 필요하다. 문항분석에서 실험조건의 효과가 통계적으로 유의하지 않은 것은 문항효과가 존재함을 의미한다. 만약 예제데이터에서 문항효과가 없다면, 개별 반응을 참가자에 걸쳐 평균한 값은 실험조건에 의한 효과만으로 설명될 것이다. 따라서 각 조건 안에서 반응들은 변산성을 보이지 않았을 것이고 모든 실험설계에 포함된 모든 효과가 유의미할 것이다. 하지만 반복측정 변량분석의 가정을 만족하지 못하기 때문에 개별 검정은 그 결과를 신뢰하기 어렵다.

반면, 선형 혼합효과 모형은 F1/F2 분석 결과와 달리 유의한 상호작용을 보였다. 모형 선택과정을 거친 이후에 적용된 최종 모형은 하나의 모형에 반응의 영향을 주는 실험조건, 참가자, 문항을 모두 고려하였다. 또한, 선형 혼합효과 모형은 F1/F2 분석이 제공하지 못하는 정보를 제공한다. 선형 혼합효과 모형은 고정효과, 참가자 무선평과와 문항 무선평과의 크기를 모수로 포함하고 있고 모형 비교를 통해 그 크기뿐 아니라 무선평과의 구조를 파악할 수 있다. 무선평과는 연구자에게 유용한 정보가 될 수 있다. 참가자 또는 문항의 특성이 연구자의 관심사가 되거나, 또는 진단의 목적으로 사용될 수도 있다.

### 시뮬레이션 연구

실험연구자가 새로운 분석방법을 접하게 될 때 흔히 갖는 질문 가운데 하나는 다음과 같다. 위에 소개된 예제와 같이 만약 두 가지 분석도구를 사용한 결과가 일치하지 않을 때 어떤 결과를 신뢰할 수 있을 것인가? 본 단락에서는 몬테카를로 시뮬레이션 방법을 사용하여 고정효과를 탐지하는 데 있

어서 분석도구에 따른 결과를 비교하였다. 임의의 상황을 설정하고, 문항효과가 존재할 경우와 그렇지 않은 경우, 두 가지 조건에서 각각 1000개의 데이터셋을 생성하였다. 그 후, 변량분석과 선형 혼합효과 모형을 사용하여 목표로 하는 고정효과에 대해 검증하고 그 결과를 비교하였다.

## 방 법

### 시뮬레이션 조건과 생성모형

시뮬레이션 조건은 예제데이터의 실험설계방법과 같은 상황에서 그 추정값을 참고, 수정하여 모수를 설정하였다. 실험설계방법으로서  $2 \times 2$  참가자 내 설계를 사용하고 라틴 방진설계를 통해 문항을 조건에 걸쳐 역균형화하였다. 그리고 두 가지 조건을 변화시켰다. 첫째, 문항효과 유무에 따라서 생성모형의 문항무선평과의 분산을  $500^2$  또는  $0$ 으로 설정하였다. 둘째, 상호작용 효과의 유무에 따라서 고정효과 크기를  $200$  또는  $0$ 으로 설정하였다. 반면 참가자 수(40명)와 문항 수(40개)는 고정되었다. 각 조건에서 사용한 생성모형은 표 7과 같다. 또한, 무선평과의 분산, 공분산에 대한 모수를 (9)-(11)에 제시하였다.

$$\begin{bmatrix} s_{0j} \\ s_{1j} \\ s_{2j} \\ s_{3j} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 250000 & & & \\ 75000 & 250000 & & \\ 175000 & 175000 & 250000 & \\ -75000 & -75000 & -175000 & 250000 \end{bmatrix} \right), \quad (9)$$

$$w_i \sim N(0, 500^2), \quad (10)$$

$$e_{jiab} \sim N(0, 500^2) \quad (11)$$

### 검정방법

생성된 각각의 데이터셋에 대해, 다음 세 가지 방법을 통해 고정효과 유의성이 검증되었다. 첫째, 반복측정 변량분석을 이용하여 상호작용의 유의성을 검증하였다(이하 F1분석). 둘째, F1/F2분석을 이용하여 상호작용의 유의성을 검증하였다. 참가자분석과 문항분석이 모두 유의한 경우에만 효과가 유의하다고 판단하였다. 셋째, 선형 혼합효과 모형을 사용한 유의성 검정으로서 다음 두 단계 과정을 사용하였다. 먼저, M1과 M2를 모두 데이터셋에 적합한 후, 우도비검정을 통해 모형을 선택하였다. 다음으로, 선택된 모형에서 상호작용의 추정값에 대해 Wald검정을 통해 유의성을 판단하였다.

**Table 7.** Simulation conditions and data-generating models

	Item effect	Fixed effect	Data-generating model
Condition 1	Yes	Yes	$y_{jab} = 2000 + 200x_{1a} - 500x_{2b} + 200x_{1a}x_{2b} + s_{0j} + s_{1j}x_{1a} + s_{2j}x_{2b} + s_{3j}x_{1a}x_{2b} + w_i + e_{jab}$
Condition 2	No	Yes	$y_{jab} = 2000 + 200x_{1a} - 500x_{2b} + 200x_{1a}x_{2b} + s_{0j} + s_{1j}x_{1a} + s_{2j}x_{2b} + s_{3j}x_{1a}x_{2b} + e_{jab}$
Condition 3	Yes	No	$y_{jab} = 2000 + 200x_{1a} - 500x_{2b} + s_{0j} + s_{1j}x_{1a} + s_{2j}x_{2b} + s_{3j}x_{1a}x_{2b} + w_i + e_{jab}$
Condition 4	No	No	$y_{jab} = 2000 + 200x_{1a} - 500x_{2b} + s_{0j} + s_{1j}x_{1a} + s_{2j}x_{2b} + s_{3j}x_{1a}x_{2b} + e_{jab}$

**검정방법의 평가**

검정방법의 수행은 1종 오류의 비율과 검정력을 통해 평가 되었다. 1종 오류는 상호작용의 고정효과가 존재하지 않을 때, 검정도구가 그 효과를 유의하다고 판단한 것을 가리킨다. 고정효과가 존재하지 않는 조건(조건3, 4)의 1000개의 데이터셋에서 유의한 결과를 보인 데이터셋의 비율을 1종 오류의 비율로 정의하였다. 검정력은 상호작용의 고정효과가 유의할 때, 검정도구가 그 효과를 유의하다고 판단할 확률을 가리킨다. 1종 오류의 비율과 마찬가지로, 고정효과가 존재하는 조건(조건 1, 2)의 1000개의 데이터셋에서 유의한 결과를 보인 데이터셋의 비율을 검정력으로 정의하였다.

**결 과**

각 조건별 검정방법에 따른 검정력과 1종 오류 비율을 표 8에 제시하였다. 결과는 다음과 같이 요약할 수 있다. 검정력에 관하여 기술하면, 변량분석을 사용한 F1분석과 F1/F2분석은 문항효과가 존재할 경우, 그렇지 않은 경우에 비하여 검정력이 크게 감소하였다. 반면, 선형 혼합효과모형은 문항효과 유무에 따른 검정력의 감소가 관찰되지 않았다. 뿐만 아니라 두 조건 모두에서 변량분석을 사용했을 때보다 검정력이 높았다. 1종 오류 비율에 관하여 기술하면, 영가설 검

정절차에서 1종 오류 비율은 유의수준을 .05로 통제하므로 이상적인 경우에 .05에 가까울 것이 예측된다. 문항효과가 존재하지 않을 때, F1분석을 사용하였을 때만 .05에 근접한 (.054) 1종 오류 비율이 관찰되었다. F1분석과 F1/F2분석 모두, 문항효과가 존재할 경우 존재하지 않을 때보다 더 보수적인 1종 오류 비율을 나타냈다. 선형 혼합효과 모형은 .036-.037로 유의수준에 비하여 1종 오류 비율이 다소 낮았지만, 문항효과의 유무에 따라 그 수치가 크게 변하지 않았다.

**종합논의**

본 논문은 심리학 연구자들에게 실험데이터를 분석하는 도구로서의 선형 혼합효과 모형을 소개하였다. 한국에서 출판된 심리학 실험 연구에서 아직 선형 혼합효과 모형의 사용빈도가 높지 않음을 보였고, 변량분석 관점에서 문항분석을 위해 관행적으로 쓰이고 있는 F1/F2 분석의 한계점을 논의하고, 선형 혼합효과 모형이 그 대안으로 사용할 수 있음을 제안하였다. 앞으로, 변량분석을 주요 분석 도구로서 사용하는 연구자가 새로운 분석 도구를 사용하기 전에 흔히 갖게 되는 질문을 통해 논의를 전개하고자 한다.

**Table 8.** Results for simulation study

Method	Power		Type I error rate	
	Condition 1 (with item effect)	Condition 2 (without item effect)	Condition 3 (with item effect)	Condition 4 (without item effect)
F1 analysis	.484	.578	.034	.054
F1/F2 analyses	.407	.472	.022	.037
Linear mixed-effects model	.602	.605	.037	.036

### 반복측정 변량분석을 사용한 추론이 신뢰로운 경우와 그렇지 못한 경우는 각각 언제인가?

모형 비교 관점에서 데이터 분석 도구를 이해하는 것은 연구자들이 새로운 모형을 더 쉽게 받아들이는 데 도움이 될 수 있다. 다시 말해서, 변량분석을 일반 선형 모형(GLM: general linear model) 관점에서 이해한다면, 반복측정 변량분석을 복수의 무선효과를 고려할 수 없는 선형 혼합효과 모형의 특수한 형태로 볼 수 있다. 따라서, 반복측정 변량분석의 사용이 신뢰로운 결과를 낳는 경우는 선형모형으로서의 반복측정 변량분석의 가정이 참일 경우이다. 즉, 1) 연구자의 가설에 따른 독립변인의 효과가 참이고, 2) 독립변인의 효과 이외에 참가자의 개인차만이 오로지 종속변인의 변산성을 설명하고, 3) 2)의 개인차는 하나의 정규분포로서 설명되는 경우이다. 반면, 이 중 하나의 가정이라도 만족하지 못한다면 반복측정 변량분석의 추론 결과는 정확하지 않을 수 있다. 본 논문의 시뮬레이션 연구 결과도 생성모형에 문항효과를 포함하고 있지 않을 때는 반복측정 변량분석의 수행이 선형 혼합효과 모형보다 크게 떨어지지 않았지만, 문항의 변산성이 생성모형에 포함될 때 반복측정 변량분석은 검정력에 문제를 보였다.

### 참가자 간 설계의 연구에는 선형 혼합효과 모형을 사용할 수 없는가?

선형 혼합효과 모형은 데이터에 포함된 구성단위 때문에 종속변인의 독립성이 위반되는 것을 고려하기 위해 주로 사용되어왔다. 실험데이터 맥락에서는 참가자가 그 단위에 해당하기 때문에 이를 고려하기 위해 참가자 내 설계에서 주로 논의가 되어왔다. 하지만 선형 혼합효과 모형은 종속변인을 설명하는 효과들이 상수이거나(고정효과) 또는 분포를 따른다(무선효과) 가정 아래에서 비교적 자유롭게 구성할 수 있기 때문에 참가자 간 설계에서도 사용할 수 있다. 보통의 경우 참가자 간 설계의 연구를 위해서는 조건에 상관없이 모든 참가자의 효과를 위해 하나의 무선효과를 설정할 수 있을 것이다. 만약, 조건의 수준마다 참가자효과 분산의 크기가 다를 것이라고 가정한다면(즉, 등분산성 가정이 위협된다면), 조건의 각 수준마다 무선효과를 설정하고 분산의 크기를 추정할 수 있다. 또한, 본문에서 언급된 것처럼 라틴 방진법이나 혼합설계(mixed design, 일명 분할구획설계(split-plot design))에도 적용 가능하다. Baayen 등(2008)에서 위에 언급된 설계방법을 포함한 몇 가지 실험설계상황에서 선형 혼합효과 모형의 용례를 제시하였다.

### 선형 혼합효과 모형은 반복측정 변량분석에 비해 어떤 추가적인 정보를 제공하는가?

첫째, 통계적 유의성만을 검정하는 반복측정 변량분석에 비해 선형 혼합효과 모형은 연구자가 관심있는 효과의 크기가 모수로서 추정된다. 둘째, 모형 비교 과정에서 최종적으로 선택된 모형을 통해 무선효과의 구조를 이해할 수 있다. 예를 들어, 무선 절편 모형에 비해 무선 기울기 모형이 선택되었다면, 참가자의 개인차는 전체적인 반응 경향성뿐 아니라 실험조건의 효과에서도 나타남을 의미한다. 셋째, 사용된 모형이 참이라는 가정 아래에서 개별 참가자 또는 문항효과들의 크기도 사후에 파악할 수 있다. 개별 단위의 효과는 모수로서 추정되지는 않지만, 분포적 가정을 이용하여 계산된다.

### 선형 혼합효과 모형은 반복측정 변량분석에 비해 언제 더 정확한 결론을 보장하는가?

데이터에 반복측정 변량분석의 가정을 위협하는 특성이 있지만, 선형 혼합효과 모형에서 그것을 고려할 수 있을 때 선형 혼합효과 모형의 사용이 권장될 수 있을 것이다. 특히, 선형 혼합효과 모형은 앞서 설명한 반복측정 변량분석의 2)번 가정이 위반되었을 때 유용한 대안으로 고려될 수 있다. 한 가지의 원인으로 독립변인의 효과를 제외한 모든 변산성을 설명할 수 없을 때, 적절한 구조의 무선효과를 설정함으로써 하나의 모형 안에 종속변인이 갖는 변산성의 원인을 모두 포함한다. 그 구조가 실제와 가까울수록 연구자가 탐지하기를 원하는 효과에 대한 추론 역시 정확할 것이다. 하지만, 선형 혼합효과 모형의 수행을 구체적인 실험 맥락들에 걸쳐 평가한 연구는 아직 없다.

### 선형 혼합효과 모형을 사용하기 전에 어떤 점을 고려해야 하는가?

선형 혼합효과 모형 또한 몇 가지 가정이 있고, 그 가정이 위협될 경우 올바른 결론을 보장하지 않는다. 따라서, 연구자는 선형 혼합효과 모형을 사용하기 전에 자신의 데이터가 모형의 가정을 만족하는지 확인해야 한다. 선형 혼합효과 모형의 가정 또한 다른 통계모형들과 마찬가지로 설정된 모형이 참인 모형임을 가정한다. 즉, 독립변인이 빠짐없이 모형에 포함되어 있어야 하고(=연구자의 연구가설이 참), 무선효과의 구조 또한 실제의 구조와 일치하여야 한다. 또한, 잔차를 포함한 모든 무선효과의 분포가 설정된 것과 일치해야 한다. 선형 혼합효과 모형의 가정에 대한 논의를 위해서 McNeish 등(2016)을 참고할 수 있다.

**M1, M2 이외의 모형을 사용할 수 있는가?**

사용할 수 있다. Baayen 등(2008)과 Barr 등(2013)에서 실험데이터를 분석을 위한 선형 혼합효과 모형(본 논문의 무선 기울기(M2) 모형)을 사용할 것을 권장하였다. 특히, Barr 등(2013)은 고정효과뿐 아니라 무선효과에서도 모수의 수를 최대화한 모형(본 논문의 무선 기울기(M2) 모형)을 사용할 것을 권장하였다. 하지만 Barr 등(2013)이 시연을 위해 사용한 설계방법이 너무 단순하여서 경험적 연구자들이 이 권장 사항을 따르기 어려운 점이 있다. 무선효과의 구조는 실험조건 수, 실험조건 수준의 수가 늘어날수록 빠른 속도로 복잡해진다. 따라서 M2는 데이터의 크기에 비해 쉽게 너무 복잡한 모형이 되고, 그 결과는 수렴실패(convergence failure)로 이어진다. 만약 연구자가 무선효과의 구조에 대한 선형적 가정이 있다면 그에 따라 무선효과의 구조를 설정하여 사용할 수 있다. 모형 비교 접근의 여러 맥락에서 모형 선택의 기준으로 데이터 주도(data-driven) 방법과 이론 주도(theory-driven) 방법을 모두 고려하는 것은 일반적이다(예: Snijder & Bosker, 2011, Ch.6.4).

관련된 논점으로, M2는 참가자효과의 무선 기울기를 포함하였지만, 문항효과에는 절편만을 포함하였다. 그렇다면 참가자효과뿐만 아니라 문항효과에도 무선 기울기를 모형에 포함할 수 있을까? M2와 같이 모수를 최대로 설정한 모형에 문항 무선 기울기가 포함되지 않은 이유는 그것이 단일 해를 산출할 수 없는(unidentifiable) 모형이기 때문이다(Barr et al., 2013). 하지만 이론 주도적 관점에서, 만약 연구자가 자신들의 문항이 조건의 주효과나 상호작용에 걸쳐 서로 다른 변산성을 가진 효과가 있다는 가정이 있다면, 모형에 참가자 무선 기울기를 제외하고 문항 무선 기울기만을 포함하는 것은 가능할 것이다.

**제한점**

다음의 질문들은 본 논문의 제한점으로서 추후 연구들을 통해 탐구되어야 할 것이다. 본 논문에서 사용된 시뮬레이션 연구설계는 제한된 상황에서 반복측정 변량분석과 선형 혼합효과 모형의 수행을 비교하였기 때문에 이 결과를 현행 연구들이 가진 실험설계방법에서 변량분석을 사용하는 것이 어느 경우에 어떤 문제를 일으키는지 일반화하기 어렵다. 독립변인의 수와 수준의 수, 참가자 수, 문항 수, 효과 크기를 체계적으로 조작한 시뮬레이션 연구가 이 문제에 대해 연구자들에게 유용한 정보를 줄 수 있을 것이다. 또한, 실험데이터 맥락에서 선형 혼합효과 모형의 추정방법, 모형 비교방법, 추론을 위한 가설검정 방법을 평가하는 연구들도 필요할 것이다.

**References**

Ahn, J., Kim, T. H., & Choi, W. (2019). The effects of visual complexity and character structure on Hangul perception. *The Korean Journal of Cognitive and Biological Psychology, 31*, 135-146.

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*, 390-412.

Bae, S., & Lee, D. (2017). Individual differences in the morphological decomposition of Hanja words. *The Korean Journal of Cognitive and Biological Psychology, 29*, 455-462.

Bae, S., & Yi, K. (2010). Processing of orthography and phonology in Korean word recognition. *The Korean Journal of Cognitive and Biological Psychology, 22*, 369-385.

Bae, S., & Yi, K. (2016). The morphological processing of Korean compound words with Saisios. *The Korean Journal of Cognitive and Biological Psychology, 28*, 691-698.

Bae, S., & Yi, K. (2019a). The influence of word type and compositionality on the word length effect in Korean. *The Korean Journal of Cognitive and Biological Psychology, 31*, 39-52.

Bae, S., & Yi, K. (2019b). Individual differences in reading spaced and unspaced compound noun phrases. *The Korean Journal of Cognitive and Biological Psychology, 31*, 253-264.

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*, 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Choi, S., & Koh, S. (2012). The effects of real world knowledge and case-markers on semantic relation processing during Korean sentence reading: An eye-tracking study. *The Korean Journal of Cognitive and Biological Psychology, 24*, 89-105.

Choi, W., Lee, C., & Nam, K. (2008). Cross-linguistic semantic priming effect for Korean-English unbalanced bilinguals. *The Korean Journal of Cognitive and Biological Psychology, 20*, 357-372.

- Choi, W., Lee, C., Kang, J., & Nam, K. (2015). The lexical inhibition of the phonological information in Korean visual word recognition. *The Korean Journal of Cognitive and Biological Psychology, 27*, 561-581.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and generalization of the “mere exposure” effect. *Journal of Personality and Social Psychology, 45*, 492-500.
- Kang, J., Nam, S., Lim, H., & Nam, K. (2016). ERP indices of Korean derivational prefix morphemes separated from the semantic and orthographic information. *The Korean Journal of Cognitive and Biological Psychology, 28*, 409-430.
- Kim, J., & Park, K. (2016). The implicit causality effect observed in Korean interpersonal verbs does not require causal connective. *The Korean Journal of Cognitive and Biological Psychology, 28*, 221-239.
- Koh, S., Hong, H., Yoon, S., & Cho, P. (2008). The frequency effect in Korean noun eojeols: An eye-tracking study. *The Korean Journal of Experimental Psychology, 20*, 21-37.
- Kwon, Y., & Nam, K. (2011). The relationship between morphological family size and syllabic neighborhoods density in Korean visual word recognition. *The Korean Journal of Cognitive and Biological Psychology, 23*, 301-319.
- Lee, H., & Choi, W. (2019). Predictability effects modulated by age during sentence reading: An eye-tracking study. *The Korean Journal of Cognitive and Biological Psychology, 31*, 17-38.
- Lee, J.-H. (2009). Anaphoric reference resolution in expository text: the effects of demonstratives type. *The Korean Journal of Psychology: General, 28*, 547-569.
- Lee, K. E., Woo, Y.-H., & Lee, H.-W. (2019). Translation priming effects in unbalanced Korean-English bilinguals. *The Korean Journal of Cognitive and Biological Psychology, 31*, 211-221.
- Lee, Y., & Kwon, N. (2012). The effect of information status of noun phrase on Korean sentence reading: An eye-tracking study. *The Korean Journal of Cognitive and Biological Psychology, 24*, 149-166.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2016). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods, 22*, 114-140.
- Molenberghs, G., & Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician, 61*, 22-27.
- Nam, S., Baik, Y., Lim, H., & Nam, K. (2014). Different time courses of orthographic, morphological, and semantic activation during Korean prefixed derivational word recognition. *The Korean Journal of Cognitive and Biological Psychology, 26*, 1-20.
- Noh, S. R., So, Y.-S., & Kim, M. (2017). Aging and situation models in Korean sentence comprehension. *The Korean Journal of Cognitive and Biological Psychology, 29*, 189-196.
- Park, K., Yi, K., Abe, J., & Liu, Y. (2008). A cross-linguistic study on representation and processing of Hanja words: Reading aloud. *The Korean Journal of Experimental Psychology, 20*, 179-202.
- Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. New York, NY: Springer.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raaijmakers, J. G. W., Schrinemakers, 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*, 416-426.
- Raaijmakers, J. G. W. (2003). A further look at the “language-as-fixed-effect fallacy” *Canadian Journal of Experimental Psychology, 57*, 141-151.
- Raudenbush, S. & Bryk, A. (2002). *Hierarchical linear models: Applications and data Analysis methods*. New York, NY: Sage.
- Ryu, J., Nam, K., Kim, D., & Baik, Y. (2016). Do Korean learners of English use spelling-to-sound regularity information during English word recognition?. *The Korean Journal of Cognitive and Biological Psychology, 28*, 1-24.
- 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* (pp.4-31). New York, NY: Routledge.
- Snijders, T., & Bosker, R. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed). Thousand Oaks, CA: Sage.
- Song, H., & Lee, W. (2009). The influence of discourse information on Korean adults' sentence processing. *Language Facts and Perspectives*, 24, 141-154.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [<http://arxiv.org/pdf/1308.5499.pdf>]
- Yi, K., & Bae, S. (2009). Morphological processing of native Korean words. *The Korean Journal of Cognitive and Biological Psychology*, 21, 233-247.
- Yi, K., Park, K., Abe, J., Liu, Y., & Zhang, Y. (2010). A cross-linguistic study on representation and processing of Hanja words: Naming and lexical decision. *The Korean Journal of Cognitive and Biological Psychology*, 22, 277-291.
- Yoon, S., & Koh, S. (2010). The effect of age of acquisition on fixation durations in Korean reading: An eye tracking study. *The Korean Journal of Cognitive and Biological Psychology*, 22, 129-142.
- Yoon, S. O., Kang, W.-S., An, J., & Koh, S. (2010). The frequency and length effect on eye fixation in Korean reading. *The Korean Journal of Cognitive and Biological Psychology*, 22, 215-232.
- Zhang, D., & Lin, X. (2008). Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In D. B. Dunson (Ed.) *Random effect and latent variable model selection*. (pp.19-36). New York, NY: Springer.

# 실험데이터 분석을 위한 선형 혼합효과 모형

## 이우열

충북대학교 심리학과

본 논문은 반복측정 변량분석의 대안으로서 교차 무선효과를 가진 선형 혼합효과 모형을 소개한다. 두 종류의 선형 혼합효과 모형의 설정, 추정방법, 모형 비교방법, 추론방법에 관해 R패키지 중 하나인 lme4 패키지의 lmer 함수를 가지고 설명한다. 예제데이터를 통해 선형 혼합효과 모형의 용례를 보인다. 또한, 몬테카를로 시뮬레이션 실험을 통해 특정 상황에서 선형 혼합효과 모형과 변량분석 사이에서 가설검정 수행을 비교한다.

**주제어:** 선형 혼합효과 모형,  $F1/F2$ 분석, R패키지, 반복측정 변량분석