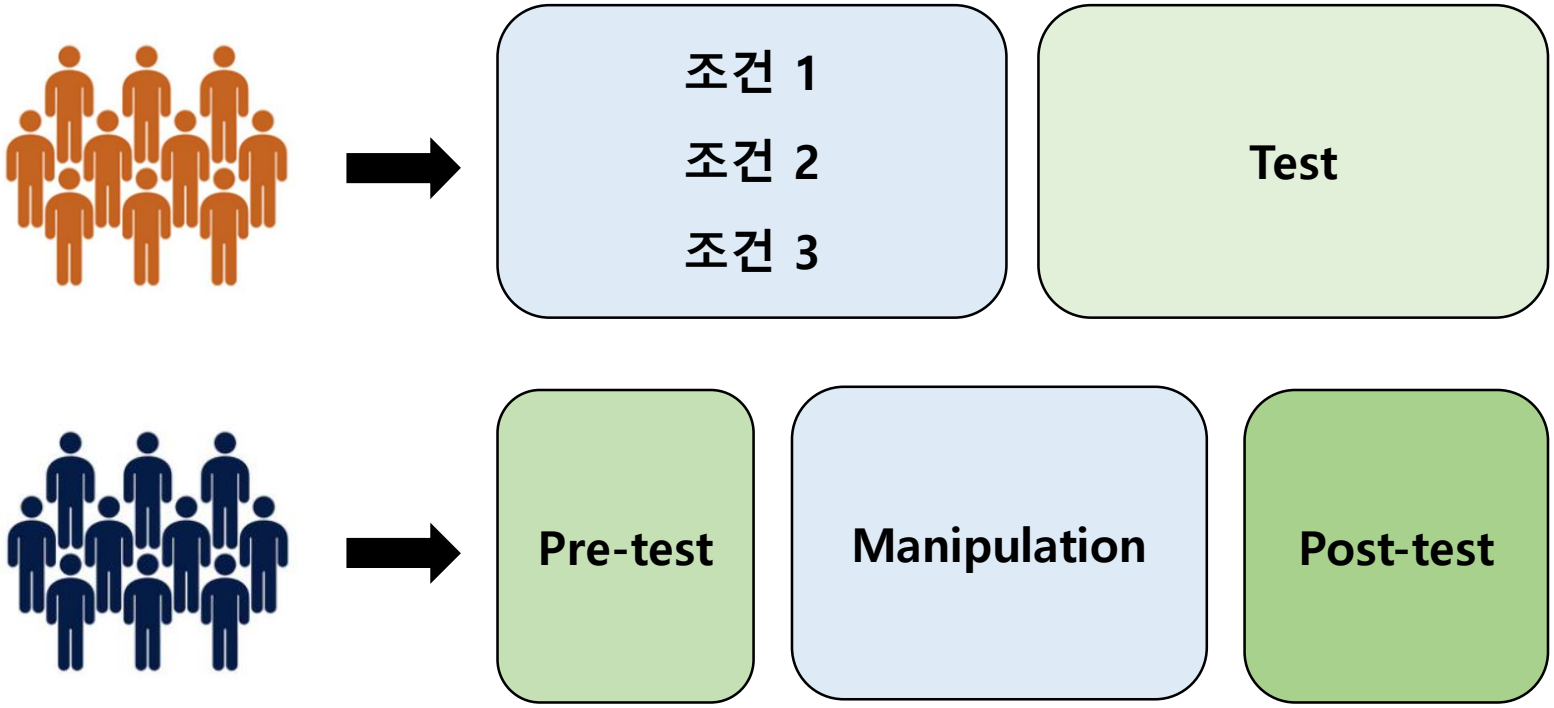
Introduction to Linear Mixed Effect Models (LMMs)

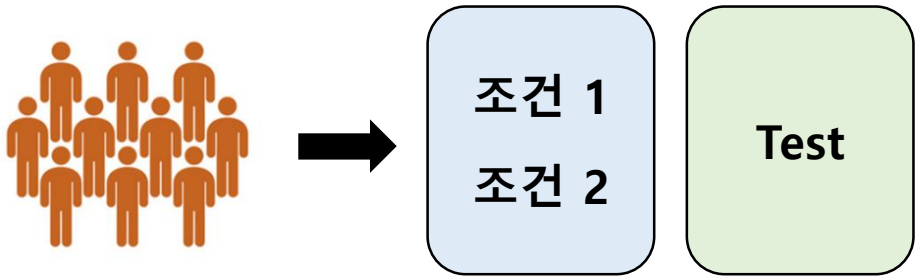
Taehoon Kim

20. 10. 07.

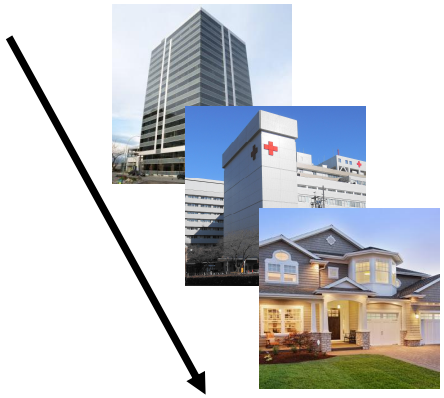
Within-Subject / Repeated Measures Design



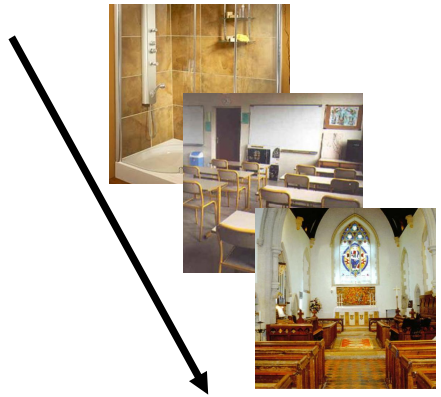
Within-Subject / Repeated Measures Design



조건 1



조건 2



SN	Trial	Cond	Item	···	Response
1	1	1	1		1520
1	2	1	2		2530
···	···	···	···		
1	30	2	15		1870
2	1	1	1		2250
2	2	1	2		2460
···	···	···	···		
2	30	2	15		1900
3	1	1	1		1820
3	2	2	2		2000
···	···	···	···		···
20	1	1	1		3200
20	2	1	2		2890
···	···	···	···		···
20	30	2	15		3450

RM ANOVA

SN	Trial	Cond	Item	···	Resp
1	1	1	1		1520
1	2	1	2		2530
···	···	···	···		
1	30	2	15		1870
2	1	1	1		2250
2	2	1	2		2460
···	···	···	···		
2	30	2	15		1900
3	1	1	1		1820
3	2	2	2		2000
···	···	···	···		···
20	1	1	1		3200
20	2	1	2		2890
···	···	···	···		···
20	30	2	15		3450

SN	Cond1	Cond2
1	1800	1920
2	2240	2360
3	1600	2320
···	···	···
20	2900	3210

Repeated Measures ANOVA

Within Subjects Effects

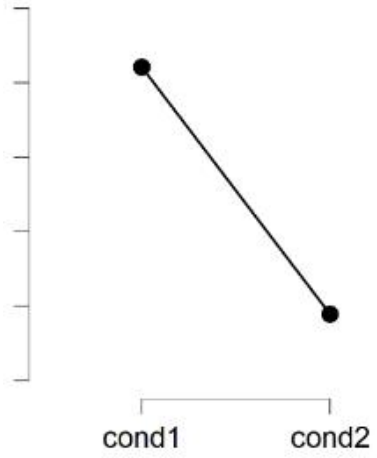
Cases	Sum of Squares	df	Mean Square	F	p
cond	2755.600	1	2755.600	26.265	< .001
Residuals	1993.400	19	104.916		

Note. Type III Sum of Squares

Post Hoc Tests

Post Hoc Comparisons - cond

		Mean Difference	SE	t	p_{bonf}
cond1	cond2	16.600	3.239	5.125	< .001



RM ANOVA

SN	Trial	Cond	Item	···	Resp
1	1	1	1		1520
1	2	1	2		2530
···	···	···	···		
1	30	2	15		1870
2	1	1	1		2250
2	2	1	2		2460
···	···	···	···		
2	30	2	15		1900
3	1	1	1		1820
3	2	2	2		2000
···	···	···	···		···
20	1	1	1		3200
20	2	1	2		2890
···	···	···	···		···
20	30	2	15		3450

SN	Cond1	Cond2
1	1800	1920
2	2240	2360
3	1600	2320
···	···	···
20	2900	3210

Repeated Measures ANOVA

Within Subjects Effects

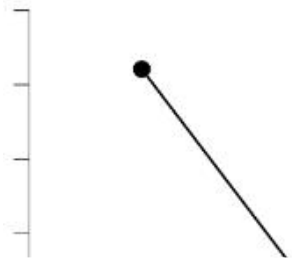
Cases	Sum of Squares	df	Mean Square	F	p
cond	2755.600	1	2755.600	26.265	< .001
Residuals	1993.400	19	104.916		

Note. Type III Sum of Squares

Post Hoc Tests

Post Hoc Comparisons - cond

		Mean Difference	SE	t	p_{bonf}
cond1	cond2	16.600	3.239	5.125	< .001



참가자에 의한 변산성은 일부 고려되지만,

Item에 의한 변산성은 무시된다

Linear Mixed Effect Model

- 조건 별 평균이 아닌 모든 개별 시행의 반응을 활용한다.
- 관심 변인의 효과 (고정 효과fixed effect)와 참가자, 아이템에 의한 변산 (무선효과random effect)를 단일 모형에서 다루어 관심 변인의 효과를 더욱 정확히 추정할 수 있다.
- 참가자와 아이템의 무선 효과 구조를 추정하여 더 많은 정보를 얻을 수 있다.

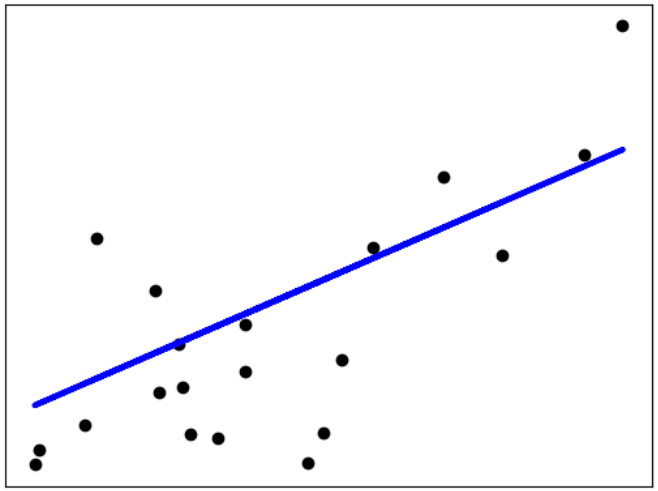
- 무선 효과 (random effect)
  - 무선 절편 (random intercept): 개별 참가자, 항목의 기저선
  - 무선 기울기 (random slope): 개별 참가자, 항목에서 관심 변인의 효과

Baayen et al., 2008; Singmann & Kellen, 2017

Linear Model

Fixed Effect

$$y_{i,j,k} = \beta_0 + \beta_\delta X_{i,j,k} + \epsilon_{i,j,k},$$



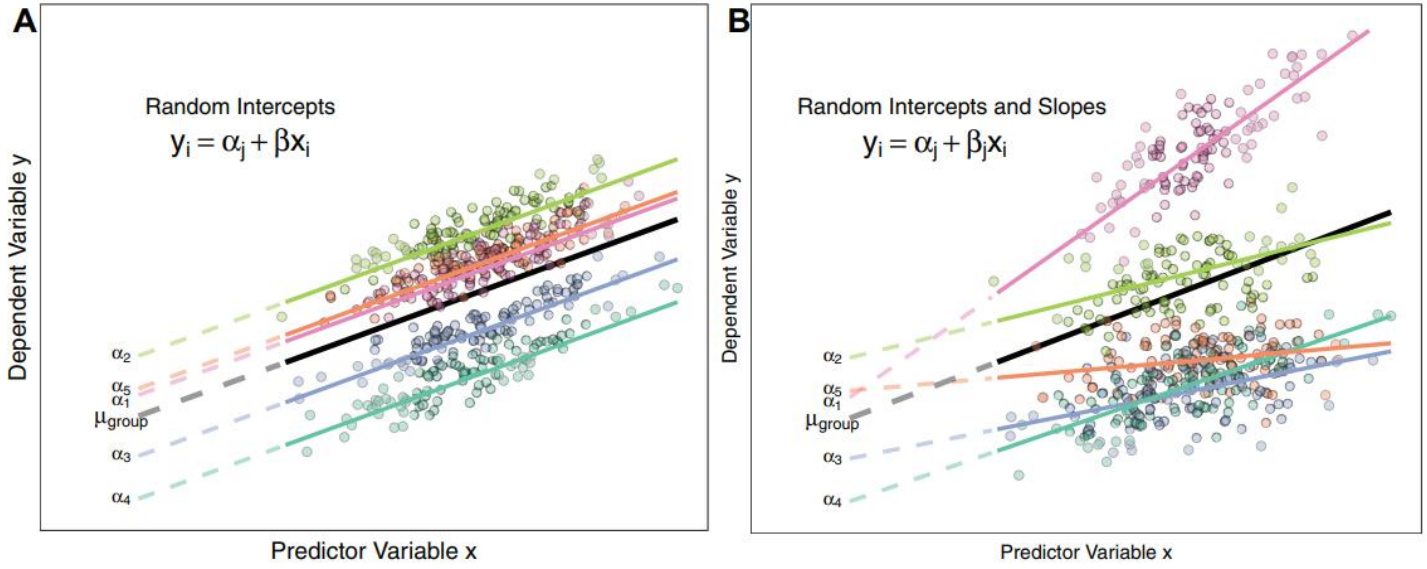
Baayen et al., 2008; Singmann & Kellen, 2017; Harrsion et al., 2018. PeerJ.

Linear Mixed Effect Model

Random Intercept Random Slope

$$y_{i,j,k} = \beta_0 + S_{0,i} + I_{0,j} + (\beta_\delta + S_{\delta,i} + I_{\delta,j})X_{i,j,k} + \epsilon_{i,j,k},$$

$$i = 1, 2, \ldots, I, \qquad j = 1, 2, \ldots, J, \qquad k = 1, 2,$$



Baayen et al., 2008; Singmann & Kellen, 2017; Harrsion et al., 2018. PeerJ.

Practice

- R version 4.0.2 / Rtools40 / R studio
- 또는 R Cloud (https://rstudio.cloud/plans/free)
- Python에선 pymer4

- LME_tutorial.R

- Packages
  - lme4
  - lmerTest
  - afex
  - emmeans
  - …

- dat_tutorial.csv

SN	Trial	Cond	Item	RT
1	1	1	1	2045
1	2	1	2	2597
…	…	…	…	…
1	14	2	7	2768
2	1	1	1	2524
2	2	1	2	2307
…	…	…	…	…
2	14	2	7	2133
3	1	1	1	2851
3	2	2	2	2039
…	…	…	…	…
4	1	1	1	2319
4	2	1	2	1812
…	…	…	…	…
6	14	2	7	2313

Practice

- R version 4.0.2 / Rtools40 / R studio
- 또는 R Cloud (https://rstudio.cloud/plans/free)
- Python에선 pymer4

- LME_tutorial.R

- Packages
  - lme4
  - lmerTest
  - afex
  - emmeans
  - …

- RT
- Preference Score
- Accuracy
- Correct, incorrect (GLMM)
- Beta …

- dat_tutorial.csv

SN	Trial	Cond	Item	RT
1	1	1	1	2045
1	2	1	2	2597
…	…	…	…	…
1	14	2	7	2768
2	1	1	1	2524
2	2	1	2	2307
…	…	…	…	…
2	14	2	7	2133
3	1	1	1	2851
3	2	2	2	2039
…	…	…	…	…
4	1	1	1	2319
4	2	1	2	1812
…	…	…	…	…
6	14	2	7	2313

Practice

- **Repeated Measures ANOVA**

SN	Cond	RT
1	c1	2467.86
1	c2	2172.86
2	c1	2700.14
2	c2	2463.57
3	c1	2652.57
3	c2	2362.14
4	c1	1463.00
4	c2	1413.43
5	c1	1773.00
5	c2	1606.57
6	c1	1098.71
6	c2	944.86

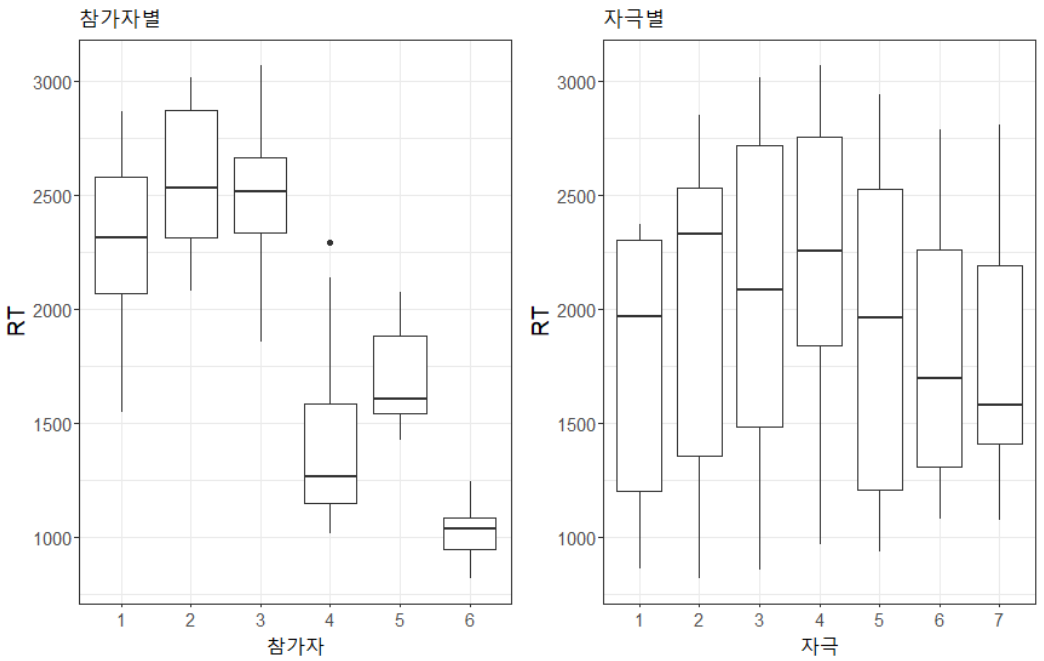
```
## RM ANOVA
rt.aov <- aov_ez(id="SN", dv = "RT", data = p1rtL, within = c("Cond"))
nice(rt.aov, es="pes")
# pes = partial eta-squared
```

```
Anova Table (Type 3 tests)

Response: RT
  Effect   df     MSE        F  pes p.value
1   Cond 1, 5 4441.37 26.65 ** .842    .004
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1
```

Practice

- **Repeated Measures ANOVA**

SN	Cond	RT
1	c1	2467.86
1	c2	2172.86
2	c1	2700.14
2	c2	2463.57
3	c1	2652.57
3	c2	2362.14
4	c1	1463.00
4	c2	1413.43
5	c1	1773.00
5	c2	1606.57
6	c1	1098.71
6	c2	944.86



Practice

- **Linear Mixed Effect Model**

By-subject random intercept model	y ~ x + (1 | Subject) + e
By-subject random intercept & slope model	y ~ x + (1 + x | Subject) + e
Cross-random effect Model (full model)	y ~ x + (1 + x | Subject) + (1 + x | Item) + e

```
# Cross-Random Effect Model (Full Model)
rt.lmer.full <- lmer(RT ~ Cond + (1 + Cond|SN) + (1 + Cond|Item), tdat)
summary(rt.lmer.m1)
anova(rt.lmer.m1)
```

```
# 무선 효과 구조 확인
coef(rt.lmer.full) # 추정된 무선 효과
ranef(rt.lmer.full) # 추정된 참가자, 항목별 변화량 (grandMean - coefficient)
coef(rt.lmer.full)$SN$`(Intercept)` - ranef(rt.lmer.full)$SN$`(Intercept)` # F1 Intercept, b0
coef(rt.lmer.full)$Item$`(Intercept)` - ranef(rt.lmer.full)$Item$`(Intercept)` # F2 Intercept, b0
```

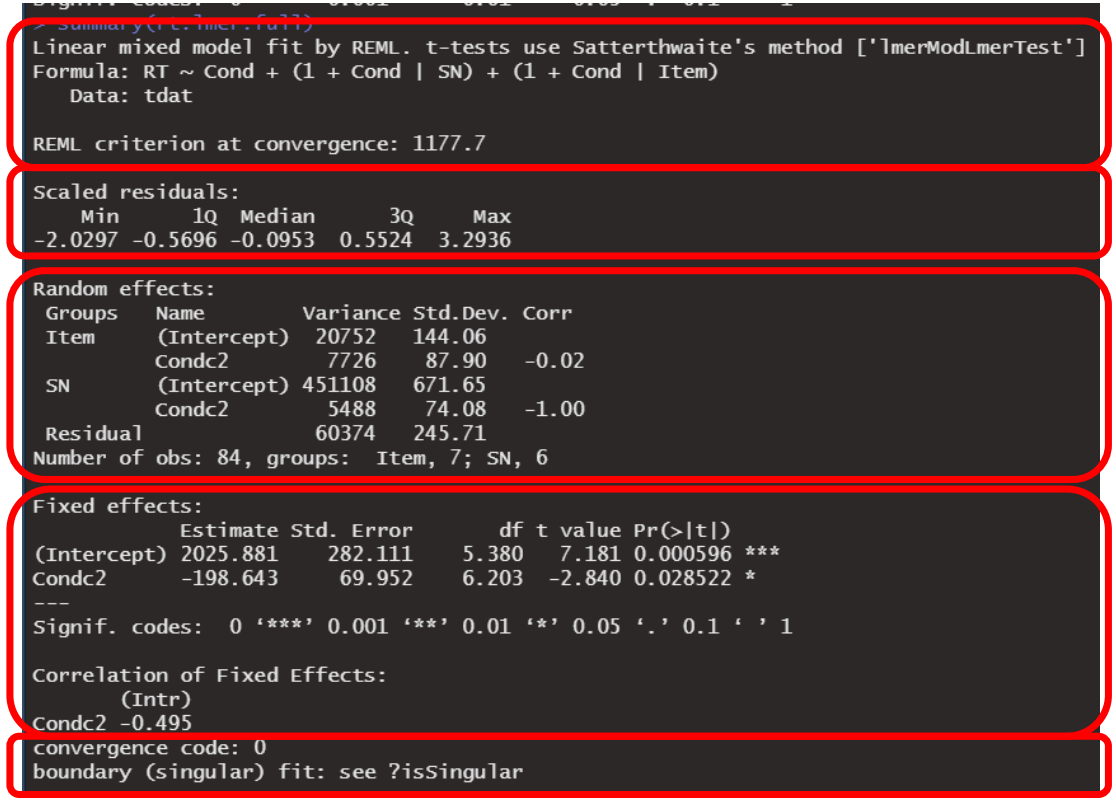
Practice

- **Linear Mixed Effect Model**

```
> summary(rt.lmer.full)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ Cond + (1 + Cond | SN) + (1 + Cond | Item)
   Data: tdat

REML criterion at convergence: 1177.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0297 -0.5696 -0.0953  0.5524  3.2936

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Item     (Intercept)  20752   144.06
          Condc2        7726    87.90   -0.02
 SN       (Intercept) 451108   671.65
          Condc2        5488    74.08   -1.00
 Residual              60374   245.71
Number of obs: 84, groups:  Item, 7; SN, 6

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) 2025.881    282.111    5.380   7.181 0.000596 ***
Condc2      -198.643     69.952    6.203  -2.840 0.028522 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr)
Condc2 -0.495
convergence code: 0
boundary (singular) fit: see ?isSingular
```

Practice

- **Linear Mixed Effect Model**



```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: RT ~ Cond + (1 + Cond | SN) + (1 + Cond | Item)
   Data: tdat

REML criterion at convergence: 1177.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.0297 -0.5696 -0.0953  0.5524  3.2936

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Item     (Intercept)  20752   144.06
          Condc2        7726    87.90   -0.02
 SN       (Intercept) 451108   671.65
          Condc2        5488    74.08   -1.00
 Residual              60374   245.71
Number of obs: 84, groups:  Item, 7; SN, 6

Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) 2025.881    282.111    5.380   7.181 0.000596 ***
Condc2      -198.643     69.952    6.203  -2.840 0.028522 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr)
Condc2 -0.495
convergence code: 0
boundary (singular) fit: see ?isSingular
```

- **Model 구조, Fitting 방법, 통계적 검정 방법**
- **Residual의 분포 정보**
- **Random Effect**
  - Residual에서 Subject, Item이 차지하는 비중
  - Intercept와 Slope 간의 상관
  - …
- **Fixed Effect**
  - Intercept와 Slope의 계수
  - Satterthwaite’s t-statistics과 p값
  - Fixed Effect 간의 상관 관계
- **오류 코드**
  - Convergence Error
  - Singular Fit

Practice

- **Linear Mixed Effect Model**

```
> anova(rt.lmer.full)
Type III Analysis of Variance Table with Satterthwaite's method
     Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
Cond 486850  486850     1 6.2031  8.0638 0.02852 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> coef(rt.lmer.full) # 추정된 무선 효과
$Item
  (Intercept)    Condc2
1    1886.825 -188.1743
2    2061.876 -123.6755
3    2157.888 -240.5962
4    2217.014 -158.9080
5    2002.202 -186.7391
6    1912.310 -227.8266
7    1943.051 -264.5803

$SN
  (Intercept)     Condc2
1    2439.798 -244.29770
2    2710.929 -274.20340
3    2634.819 -265.80843
4    1511.685 -141.92728
5    1777.794 -171.27895
6    1080.261  -94.34138
```

- 개별 항목, 참가자의 절편과 기울기

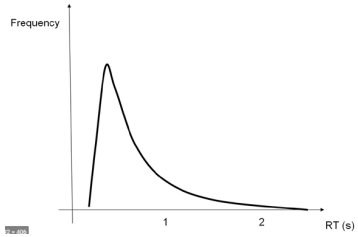
```
> ranef(rt.lmer.full) # 추정된 참가자, 항목별 변화량 (grandMean - coefficient)
$Item
  (Intercept)    Condc2
1  -139.05561  10.46859
2    35.99518  74.96734
3   132.00749 -41.95339
4   191.13260  39.73487
5   -23.67899  11.90378
6  -113.57113 -29.18377
7   -82.82955 -65.93742

$SN
  (Intercept)    Condc2
1    413.9169 -45.65484
2    685.0485 -75.56054
3    608.9379 -67.16557
4   -514.1959  56.71558
5   -248.0872  27.36391
6   -945.6201 104.30147
```

- 개별 항목, 참가자의 무선 효과 -> BLUPs

Generalized Linear Mixed Model (GLMM)

- 데이터의 분포가 정규성을 가지지 않을 때, 데이터에 적합한 분포를 선택하여 모델링할 수 있다.

- **GLMM for RT Data**
  - 반응 시간 분포는 편포가 되어 있는 경우가 많다.



    따라서 반응 시간 분포에 더 적합한 분포를 가정하여 LMM을 사용한다. (예. Inverse-gaussian)

- **GLMM for Correct/Incorrect Data**
  - 정확반응은 1(맞음), 0(틀림)으로 구성되어, 정규분포가 아닌 이항분포를 따른다.

    따라서 binomial distribution을 가정하여 LMM을 사용한다.

참고 자료

이해하기 쉬운 LMM 튜토리얼

- Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. arXiv preprint arXiv:1308.5499.

LMM에 대한 소개 & 분석 팁

- 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.
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. New methods in cognitive psychology, 4-31.

Correctness 분석에 대한 GLMM 소개

- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of Memory and Language, 59(4), 434-446.

RT 분석에 대한 GLMM 소개

- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. Frontiers in psychology, 6, 1171.