## set seed and path, load libraries

set.seed(1234567)

path<- ("c:/users/USER/jeewon")

setwd(path)

library(MASS)

library(lme4)

## number of replications

**nrep = 1000**

## conditions (adapted from Matuschek et al.(2017))

# number of persons: 30명, 50명

# number of items/level: 10문항, 20문항

**⇒** fixed effect 실험조건 수 (b0, b1)

# **fixed effect** (b0): **2000**

# **fixed effect** (b1): **0 (Type 1 error rate ---> 두 조건 간 차이가 *없음*),**

**25 (hit rate = power rate ---> 두 조건 간 차이가 있음)**

# person **random effect ⇒ ∴** fixed effect 수만큼 2개(tau0, tau1) **+ corr**

# **tau**0(sd): 100

# **tau**1(sd): 20

**# corr(tau0,tau1): 0.6**

# item **random effect**(sd)

#omega = **20**, 60, 100

# sigma(=sd(**e**)): 300

## set parameters (모수 추정 - 총 7)

nperson=30; ⇒ 30명이 두 실험조건 모두 측정하였으므로 **총 600개의 반응**

nitem=10; ⇒ 하나의 **실험조건**에 10문항

**#fixed effect**

b0 = 2000;

b1 = 25;

#person random effect (=각 조건(2개)별 개인차)

---→ **각 고정효과(tau0, tau1) 별로 참가자 무선효과가 있기 때문에 random slope model.**

① tau0 = 100;

② tau1 = 20;

**r=0.6;**

omega=20; # **item random effect** (20, 60, 100 중)

sigma=300 # 잔차(=300)

S = **matrix**(c(tau0^2, r\*tau0\*tau1, r\*tau0\*tau1, tau1^2), nrow=2)

**## design matrix**

#id

**j** = rep(c(1**:nperson**), each=2\*nitem) # 1st ~ **30th**까지(=참가자 수) 각 수준(총 2개)마다 10문항씩 repeat (총 600 반응)

**i** = rep(c(1:**nitem**), 2\*nperson) # 1부터 **10th**까지(=문항수) 60번씩 (총 600개 숫자)

# fixed (x0과 x1의 의미??)

**x0** **= 1**

**x1** = rep(c(rep(0, **nitem**), rep(1,**nitem**)), nperson(=30)) **#0(10번), 1(10번) 각각 30명씩**

**###### spaces to save (loop를 돌림 - 이전 루프들 모두 ll**(numeric(0)**, fixed, sig 저장)**

**ll <- numeric(0)**

**fixed <- numeric(0)**

**sig<-numeric(0)**

**#### FOR LOOP <---- for loop로 parameter estimation 1000번**

for (n in 1:nrep)

{

***# generate random effects* ⇒ (distribution 만들기)**

**s** = **mvrnorm**(nperson, mu=c(0,0), Sigma=S) # **참가자 무선효과** 30명

(30,2 (**두** 조건에 대한 참가자 X의 distribution))

**w1** = **rnorm**(nitem(=10), mean=0, sd=**omega**) # **문항효과** (하나의 level당 **10개** 문항)

**e** = **rnorm**(nperson\*nitem**\*2**, mean=0, sd=sigma) # **잔차 600**

***# person random***

**ss** = **rep**(**s**[,1]**,** each=20) \* **x0**

**+ rep**(**s**[,2]**,** each=20) \* **x1 (=0(10개) &1(10개)\*30명) -->2번째 실험조건**

**⇒** each가 20인 이유는 10문항씩 2개의 조건이기 때문 즉 1사람이 10문항을 2조건에서 풂

***# item random***

w= rep(**w1**, 2\*nperson) ⇒ 10문항을 2\*30(=60)번 반복

**(10문항 x 60개의 set 다 다른 숫자)**

***# y***

y = (**b0\*x0** + b1\***x1**) + **ss** + w + e

***## data frame***

dat<- data.frame (**j**(id), **i**(id), **x0(=1)**, **x1(0 & 1)**, ss, w, e, y) ⇒ 8개

**###### estimation**

**m0 = lmer (y ~ 1 + (1+x1| j)** **+ (1|i),** data=dat, REML=**F**) ⟸ FALSE is used in case of **comparing models with different “Fixed effects”** (during the simplification of model)

REML (=restricted maximum likelihood) **⇒ REML이 False인 이유**

**m1** = **lmer (y ~ 1 + x1 + (1+x1| j) + (1|i),** data=dat, REML=F)

**## several things to save**

# b1

fixed1 <- summary(**m1**)$coefficients[c(2,4)] --- estimate & standard error of **x1**

**fixed** <- rbind(**fixed**, fixed1)

z\_value <- summary(m1)$coefficients[c(2)] / summary(m1)$coefficients[c(4)]

p\_value <- pnorm(z\_value, mean=0, sd=1, lower.tail=FALSE, log.p=FALSE)

if (p\_value<0.05) {

sig1 <- 1

} else {

sig1 <- 0

}

sig <- rbind(sig, sig1)

# ***log-likelihood***

ll0<- summary(m0)$**logLik**

ll1<- summary(m1)$**logLik**

ll\_n<- c(ll0, ll1)

ll <- rbind(ll, ll\_n)

**}**

**## END of FOR LOOP**

## save files

# fixed effect

colnames(fixed) <- c("m1est","m1se")

write.table(fixed, file="fixed.txt", **sep="\t"**, **row.names = F, col.names = T)** --> **tab 구분**

**# significance of each repetition**

**colnames(sig) <-”sig”**

**write.table(sig, file=”sig.txt”, sep=”\t”, row.names=F, col.names = T)**

**-------------calculate the proportion of significance**

**------------fixed effect x1 sig if higher than 95% ??**

**# log-likelihood**

colnames(ll) <- c("m0ll","m1ll")

write.table(ll, file="ll.txt", sep="\t", row.names = F, col.names = T)

3조건 이상의 복잡한 실험조건 상황 설정 - simulation 돌려보기

power가 언제 가장 높은지?

데이터를 만들고 모형을 통해서 답을 찾음.

문제를 잘 맞춘 건지...

MLE(최대 우도) 추정법 -

**GEE**(일반화 추정 방정식) 추정법 -

PQL(벌점 준우도 추정법) -

**randomized blocks design (RBD)**

\*logistic regression = Generalized linear model

\*mixed effect? fixed & random

\*분포로 표헌됨 -- random

반복측정을 하였기 때문에 random slope가 가능

RBD - 영향을 미치는 요소들을 random effect로

모수 추정 - parameter estimation

BLUE - Best linear unbiased estimation - 베타값을 얻게 되면 추정~ (추정값이 얼마나 모수를 대표하느냐?)

BLUP - Best linear unbiased **prediction** - 패턴을 찾음 - estimation이 이루어진 이후에 prediction이 이루어짐

random effect(계산되어짐)는 fixed effect(추정되어짐)가 정해진 다음 가능,

추정 // 예측

추정을 위해서 반복이 필요 ~ !

\*ML(maximum likelihood) - 모수를 모두 찾아야... (분산은 0부터 무한대) - 미분 - 편미분

근사(approximation) - 근사하는 방법을 찾아내는 대신 정확도가 떨어진다.

PQL // Bayesian -

비교방법 / 추정방법

\*\*공변량 & random effect 간 차이