

14 강

데이터분석방법론2

# 경시적자료분석 LMM 변량계수모형

대전대학교 빅데이터인공지능학과 강우창 교수

## 1 제 14강. 경시적자료 LMM 변량계수모형

- 1 경시적 자료
- 2 LMM : 경시적 자료의 특성 탐색
- 3 LMM : 경시적 자료의 분석모형 구축
- 4 LMM : 경시적 자료 모형적합 결과 해석
- 5 LMM : 모형진단



# 학습개요 및 목표

이번 강의에서는 경시적자료를 분석하는 LMM 변량계수모형을  
예제자료 중심으로 소개합니다.

- 1 경시적 자료의 통계적 특성을 이해하고 해당 자료에 적합한 LMM 분석모형을 제시할 수 있다.
- 2 경시적 자료에 LMM을 올바르게 적합시키고 분석결과를 해석할 수 있다.



01

제 14강. 경시적 자료분석 LMM 변량계수모형

# 경시적 자료

# 1. 예제자료

## ▶ 자폐아 자료(Anderson et al., 2009)

### ➤ 연구 목적

- 자폐아에서 초기 소통능력이 성장하면서 사회화정도에 어떻게 영향을 주는가?

### ➤ 자료 수집

- 자폐(ASD: Autism Spectrum Disorder) 또는 전반적발달장애(PDD: Pervasive Developmental Disorder)가 있는 158명의 어린 아이를 관찰한 연구
- 각 어린이는 두 살 때 소통발달정도 (1=low, 2=medium, 3=high)을 측정
- 2, 3, 5, 9, 13세 때에 사회화의 정도를 관측

# 1. 예제자료

## ▶ 자폐아 자료(Anderson et al., 2009)

### ➤ 자료 구조(<https://websites.umich.edu/~bwest/chapter6.html>)

age	vsae	sicdegp	childid
2	6	3	1
3	7	3	1
5	18	3	1
9	25	3	1
13	27	3	1
2	17	3	3
3	18	3	3
5	12	3	3
9	18	3	3
13	24	3	3
2	12	3	4
3	14	3	4
5	38	3	4
9	114	3	4
2	17	3	19

- **개체 간 특성 변수[Subject(Level 2) variables]**
  - ✓ childid: **아이의 고유 번호**
  - ✓ sicdegp: **2세 때의 소통발달정도**(1=low, 2=medium, 3=high)
- **개체 내 특성 변수[Time-varying(Level 1) variables]**
  - ✓ Age: **관측 나이**(2, 3, 5, 9, 13)
  - ✓ vsae: **사회화정도**(Vineland Socialization Age Equivalent: patient reported socialization)

## 2. 자료 준비

- <https://websites.umich.edu/~bwest/chapter6.html> 에서 자료를 다운받아 **autism.csv** 파일로 저장.
- **autism.csv** 파일이 아래 폴더  
[C:\강위창\방통대\데이터분석방법론2\강의노트\_2024\예제자료]에 있다고 가정

### 자료 불러오기

```
autism <- read.csv("~/autism.csv", h = T)
attach(autism)
head(autism, 10)
```

02

제 14강. 경시적 자료분석 LMM 변량계수모형

# LMM : 경시적 자료의 특성 탐색



# 1. 요인(변수)의 구분 : 고정요인 vs. 변량요인

age	vsae	sicdegp	childid
2	6	3	1
3	7	3	1
5	18	3	1
9	25	3	1
13	27	3	1
2	17	3	3
3	18	3	3
5	12	3	3
9	18	3	3
13	24	3	3
2	12	3	4
3	14	3	4
5	38	3	4
9	114	3	4
2	17	3	19

- 고정요인 (fixed factors)
  - Level 1 : 나이(age)
  - Level 2 : 2세 때의 소통발달 정도(sicdegp)
- 변량요인 (random factors)
  - Level 2 : 아이(childid)

## 2. 탐색적 자료분석 : 기술 통계량

### ▶ 소통발달정도(sicdegp) x 나이(age) 분할표

#### 분할표

```
> # 소통발달정도(sicdegp) * 나이(age) 분할표
> addmargins(table(sicdegp, age))
```

	age					
sicdegp	2	3	5	9	13	Sum
1	50	48	29	37	28	192
2	66	64	36	48	41	255
3	40	38	26	35	26	165
Sum	156	150	91	120	95	612



- Balanced data 아니다.
- Drop out & intermittent missing 발생

## 2. 탐색적 자료분석 : 기술 통계량

### ▶ sicdegp x age 별 사회화정도 점수(vsae) 분포

#### R 결과

	sicdegp	age	vsae	MEAN	MEDIAN	MIN	Q.25.	Q.75.	MAX	SD
2	1	2	50	7.00	6.5	1	5.0	9.0	13	2.73
3	1	3	48	12.02	11.0	5	9.0	13.0	45	6.26
4	1	5	29	15.03	13.0	6	11.0	16.0	42	7.92
5	1	9	37	25.56	15.0	3	11.8	22.2	130	28.42
1	1	13	28	37.11	18.5	7	12.0	46.0	126	35.54
7	2	2	66	8.67	8.5	1	7.0	11.0	18	3.54
8	2	3	64	14.08	13.0	4	10.0	18.0	39	6.20
9	2	5	36	17.69	17.0	4	11.0	24.0	35	8.00
10	2	9	48	32.12	21.5	6	15.0	43.2	110	23.40
6	2	13	41	58.83	42.0	9	18.0	73.0	198	50.27
12	3	2	40	12.40	12.0	6	10.0	15.0	20	3.43
13	3	3	38	21.24	20.0	7	16.2	24.0	63	9.38
14	3	5	26	33.92	33.5	12	24.2	37.8	77	15.78
15	3	9	35	64.14	65.0	15	34.0	76.0	171	34.59
11	3	13	26	88.69	74.5	24	56.8	119.8	192	46.34

#### • vsae 평균은

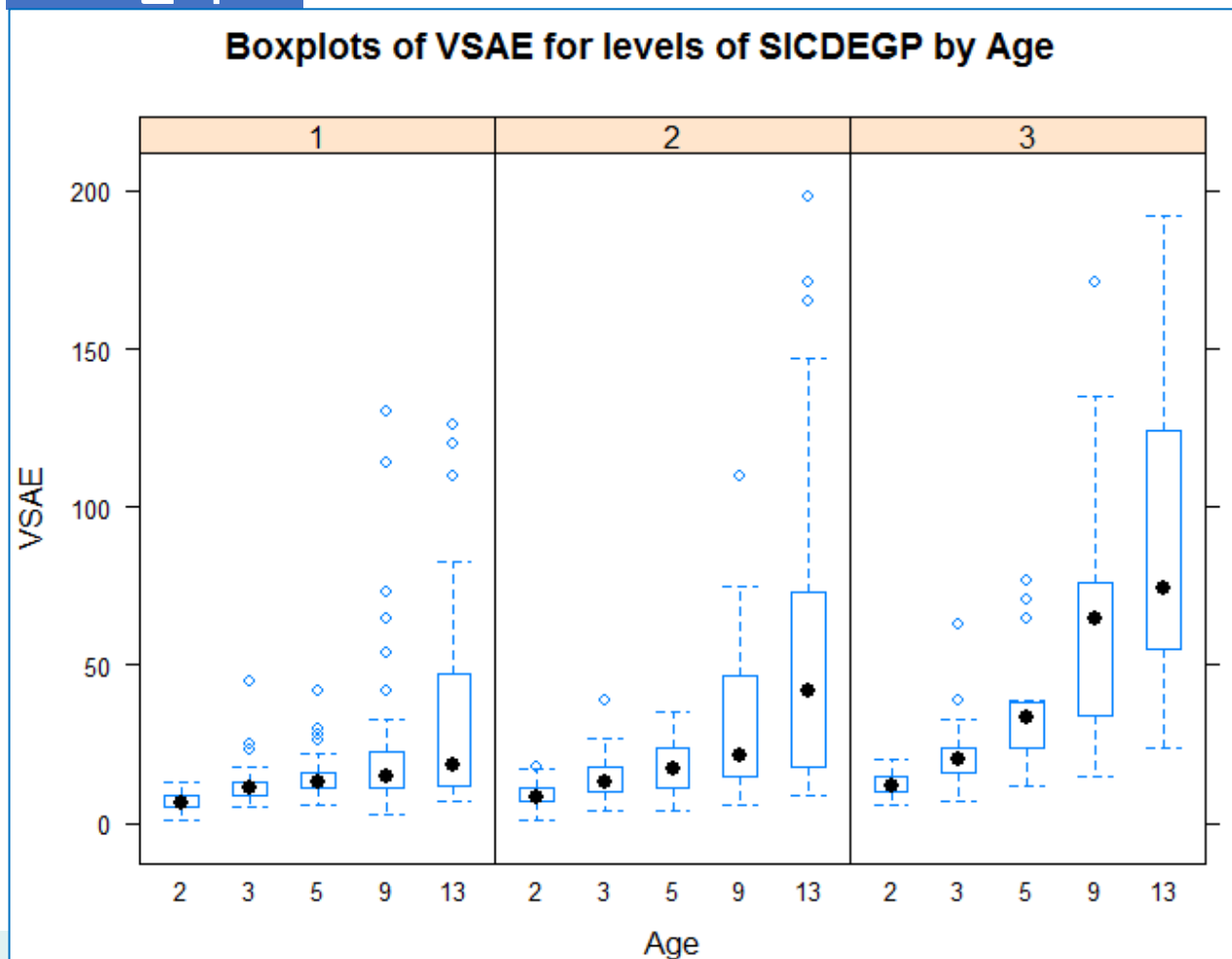
- age가 증가함에 따라 증가한다.
- sicdegp 수준이 증가함에 따라 증가한다.

#### • 각 수준조합에서 중앙값<평균, 즉 right-skewed 분포

## 2. 탐색적 자료분석 : 박스그림

### ▶ sicdegp x age 별 사회화정도점수 (vsae) 분포

R 결과

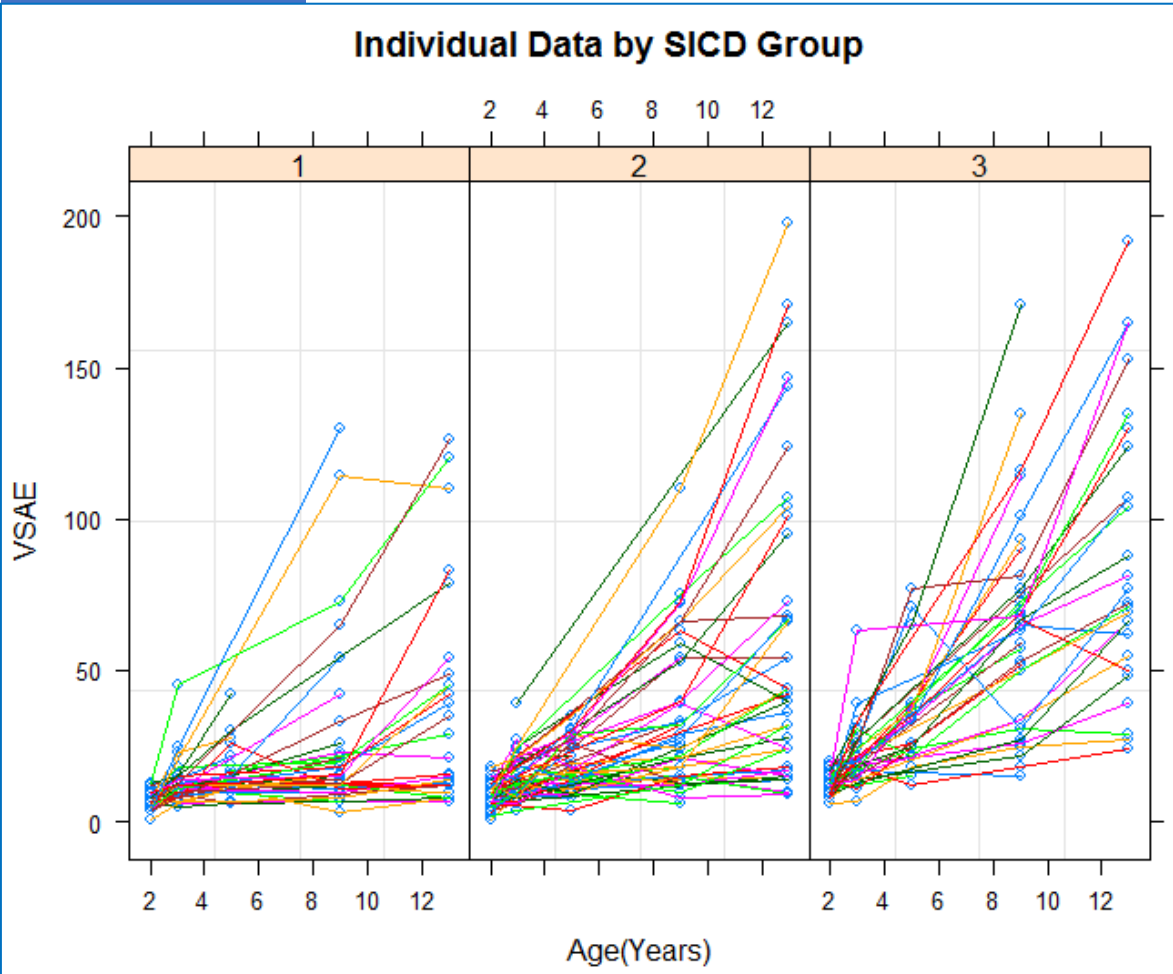


- vsae 평균은
  - age가 증가함에 따라 증가한다.
  - sicdegp 수준이 증가함에 따라 증가한다.
- 각 수준조합에서 중앙값 < 평균, 즉 right-skewed 분포
- vsae의 분산은 age가 증가함에 따라 증가한다.

## 2. 탐색적 자료분석

### ▶ 아이 별 age에 따른 vsae 변화

R 결과



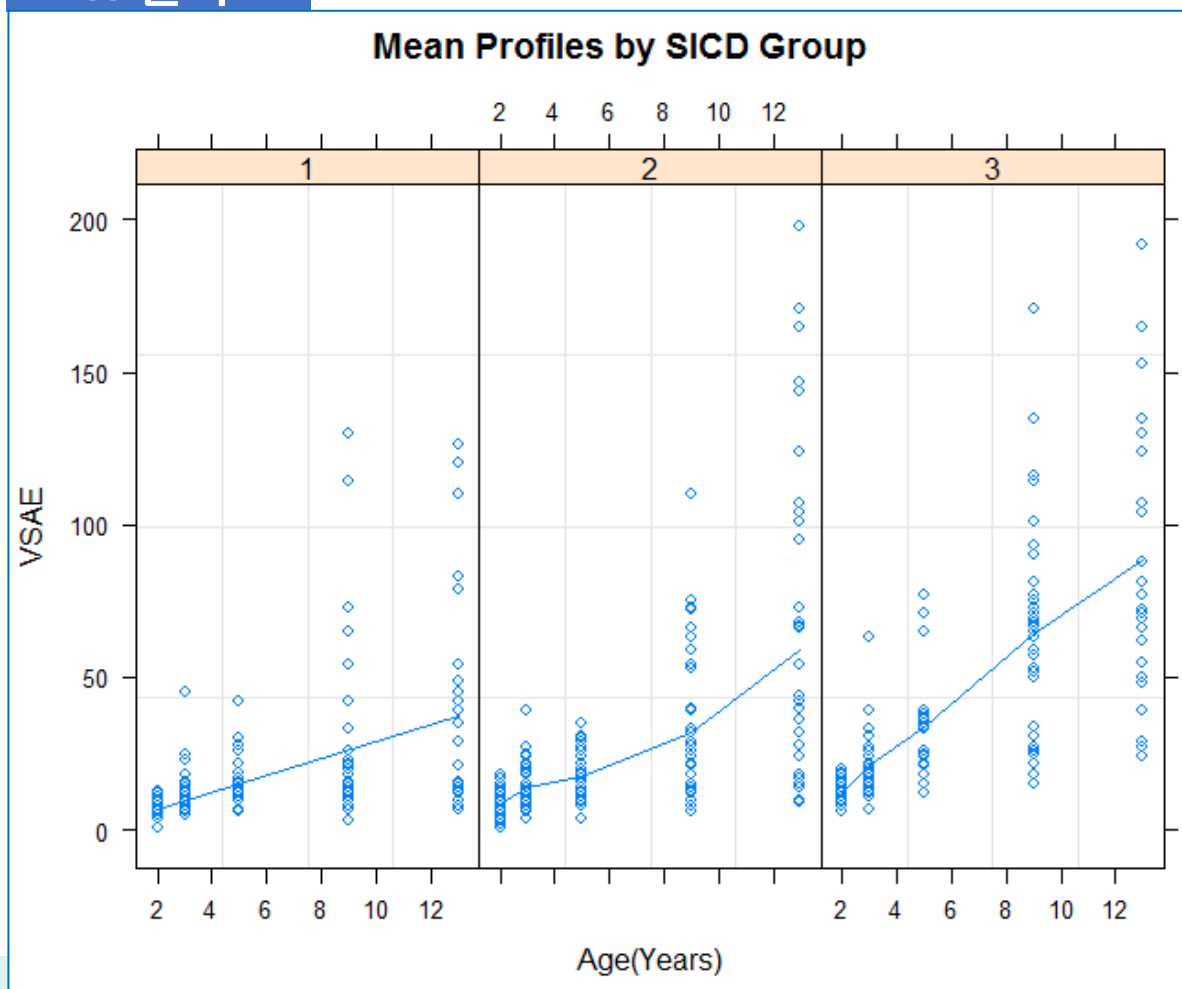
- vsae 는 age가 증가함에 따라 증가하는 경향이 뚜렷하지만
  - 증가하는 형태는 아이에 따라 뚜렷한 차이가 있다.
  - sicdegp=1에서는 나이가 증가하여도 vsae 가 거의 증가하지 않는 아이들도 있다.
- age=2 에서 vsae 의 분산은 다른 나이의 분산과 비교하여 작고 sicdegp 수준에 따라서도 변화가 거의 없다.



## 2. 탐색적 자료분석

### ▶ sicdegp 각 수준에서 age에 따른 vsae 평균변화

R 결과



- vsae 평균은
  - sicdegp 수준이 1인 군과 3인 군에서는 직선의 경향이 나타난다.
  - 반면에 sicdegp 수준이 2인 군에서는 곡선(2차 곡선)의 경향성이 보인다.

## 2. 탐색적 자료분석 : R code

### ▶ Code 1

```
# 소통발달정도 (sicdegp) *나이 (age) 별 사회화정도점수 (vsae)의 분포
library(Hmisc)
g <- function(x){c(N=length(x),MEAN=mean(x,na.rm=TRUE),MEDIAN=median(x,na.rm=TRUE),
  MIN=min(x,na.rm=TRUE),Q=quantile(x,probs=c(0.25,0.75),na.rm=TRUE),
  MAX=max(x,na.rm=TRUE),SD=sd(x,na.rm=TRUE),options(digits=3))}

summarize(vsae, by=llist(sicdegp, age), g)

## 소통발달정도 (sicdegp) *나이 (age) 별 사회화정도점수 (vsae)의 boxplots 그리기
# 소통발달정도 (sicdegp)와 나이 (age)에 대한 범주형변수 생성
sicdegp.f <- factor(sicdegp)
age.f <- factor(age)

# 생성된 범주형변수 기존 자료와 합치기
autism.updated <- data.frame(autism, sicdegp.f, age.f)

library(lattice)
trellis.device(color=T)

bwplot(vsae ~ age.f|sicdegp.f, aspect = 2, data=autism.updated, na.rm=TRUE,
  ylab="VSAE", xlab="Age",
  main = "Boxplots of VSAE for levels of SICDEGP by Age")
```

## 2. 탐색적 자료분석 : R code

### ▶ Code 2

```
## 소통발달정도 (sicdegp) 수준 별 나이 (age)에 따른 사회화정도점수 (vsae) 변화
# 아이 별 나이 (age)에 따른 사회화정도점수 (vsae) 변화
library(nlme)
autism.g1 <- groupedData(vsae~age | childid, outer= ~sicdegp.f,
                        data=autism.updated)

plot(autism.g1, display="childid", outer=TRUE, aspect=2, key=F,
     xlab="Age(Years)", ylab="VSAE", main="Individual Data by SICD Group")

# sicdegp 각 수준에서 나이 (age)에 따른 사회화정도점수 (vsae) 평균 변화
autism.g2 <- groupedData(vsae~age|sicdegp, order.groups=F,
                        data=autism.updated)

plot(autism.g2, display = "sicdegp", aspect=2, key=F,
     xlab = "Age(Years)", ylab="VSAE", main = "Mean Profiles by SICD Group")
```

- 아이 별 그래프 형식은 library 'nlme' 에서 제공
- 'nlme' 의 groupedData() 함수를 사용하여 그래프의 성격에 맞게 자료를 그룹화

03

제 14강. 경시적 자료분석 LMM 변량계수모형

# LMM : 경시적 자료의 분석모형 구축

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 변량계수모형(Random coefficient model) 이란?

### ➤ 회귀계수의 변동성에 대한 모형

- 회귀계수가 개체에 따라 변화하는 경우, 이를 모형화하는 한 방법은 회귀계수에 개체효과를 더하여 모형화 하는 것.

### ➤ 모형 : 예

- $Y_{ij}$ :  $j$ 번째 개체의  $i$ 번째 관측값
- $X_{ij}$ : 공변량

$$Y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{ij} + \varepsilon_{ij}$$

여기서  $U_j = (u_{0j}, u_{1j})' \sim^{iid} N(0, D)$ ,  $\varepsilon_{ij} \sim^{iid} N(0, \sigma^2)$  이고 서로 독립.

- $E(Y_{ij}) = \beta_0 + \beta_1 x_{ij}$
- $E(Y_{ij}|U_j) = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{ij}$
- $V(Y_{ij}) = V(u_{0j}) + x_{ij}^2 V(u_{1j}) + 2x_{ij} Cov(u_{0j}, u_{1j}) + V(\varepsilon_{ij})$



# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 하향식(Top-Down) 모형구축 전략

### ❖ 자료탐색 결과

#### ➤ 평균모형의 설정

- sicdegp의 수준 2에서 age의 2차 곡선 경향성이 있는 듯이 보임. 수준1과 2에서는 직선 경향성 보임.
  - $(age)^2$ 항과 sicdegp 항과 교호작용 항을 포함시킴.
  - 즉, age,  $(age)^2$ , sicdegp, sicdegp\*age, sicdegp\*  $(age)^2$  항 들을 평균모형에 포함시킴.

#### ➤ 변량모형의 설정

- age가 증가함에 따라 vsae가 증가하는 형태는 각 아이에서 다르게 나타남.
- age가 증가함에 따라 vsae의 분산이 증가하고 있음.
  - 절편, age,  $(age)^2$ 의 회귀계수에 변량효과를 추가.

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 하향식(Top-Down) 모형구축 전략 : Step1

### ➤ 최초 설정 모형(Model1)

$$\begin{aligned}
 Y_{ij} &= (\beta_0 + u_{0j}) + (\beta_1 + u_{1j}) * age_{ij} + (\beta_2 + u_{2j}) * (age_{ij})^2 \\
 &\quad + \beta_3 * sic2_j + \beta_4 * sic3_j + \beta_5 * age_{ij} * sic2_j + \beta_6 * age_{ij} * sic3_j \\
 &\quad + \beta_7 * (age_{ij})^2 * sic2_j + \beta_8 * (age_{ij})^2 * sic3_j + \epsilon_{ij} \\
 &= \beta_0 + \beta_1 * age_{ij} + \beta_2 * (age_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\
 &\quad + \beta_5 * age_{ij} * sic2_j + \beta_6 * age_{ij} * sic3_j \\
 &\quad + \beta_7 * (age_{ij})^2 * sic2_j + \beta_8 * (age_{ij})^2 * sic3_j \\
 &\quad + u_{0j} + u_{1j} * age_{ij} + u_{2j} * (age_{ij})^2 + \epsilon_{ij}
 \end{aligned}$$

} 고정효과 모형

} 변량효과 모형

여기서  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$U_j = (u_{0j}, u_{1j}, u_{2j})' \sim iid N(0, D)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립,  
 $D$ 는  $U_j$ 의  $3 \times 3$  분산-공분산 행렬.

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 하향식(Top-Down) 모형구축 전략 : Step2

### ➤ 변량 요인 선택

#### Model1 적합

```
### (나이-2) 변수 생성
age.2 <- age - 2
autism.updated <- subset(data.frame(autism.updated, age.2), !is.na(vsae))

### 변량계수모형 분석 자료준비
autism.grouped <- groupedData(vsae ~ age.2 | childid, data=autism.updated,
                              order.groups = F)

### 변량효과 구조선택
model.1.fit <- lme(vsae ~ age.2+ I(age.2^2)+ sicdegp.f+
                  age.2:sicdegp.f+ I(age.2^2):sicdegp.f,
                  random = ~ 1+ age.2+ I(age.2^2), method="REML",
                  data = autism.grouped)

summary(model.1.fit)
```

### ➤ 결과

#### 출력 결과

```
> ### 변량효과 구조선택
> model.1.fit <- lme(vsae ~ age.2+ I(age.2^2)+ sicdegp.f+
+                   age.2:sicdegp.f+ I(age.2^2):sicdegp.f,
+                   random = ~ 1+ age.2+ I(age.2^2), method="REML",
+                   data = autism.grouped)
Error in lme.formula(vsae ~ age.2 + I(age.2^2) + sicdegp.f + age.2:sicdegp.f + :
  nlminb problem, convergence error code = 1
  message = iteration limit reached without convergence (10)
> summary(model.1.fit)
Error in summary(model.1.fit) : 객체 'model.1.fit'를 찾을 수 없습니다
```

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 변량 요인 선택

### ➤ 변량절편( $u_{0j}$ )의 의미

- $age_{ij} = 2$ 일 때, 즉  $age.2_{ij} = 0$  일 때,

$$\begin{aligned}
 Y_{ij} &= (\beta_0 + u_{0j}) + (\beta_1 + u_{1j}) * age.2_{ij} + (\beta_2 + u_{2j}) * (age.2_{ij})^2 \\
 &\quad + \beta_3 * sic2_j + \beta_4 * sic3_j + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\
 &\quad + \beta_7 * (age.2_{ij})^2 * sic2_j + \beta_8 * (age.2_{ij})^2 * sic3_j + \epsilon_{ij} \\
 &= \beta_0 + \beta_3 * sic2_j + \beta_4 * sic3_j + u_{0j} + \epsilon_{ij}
 \end{aligned}$$

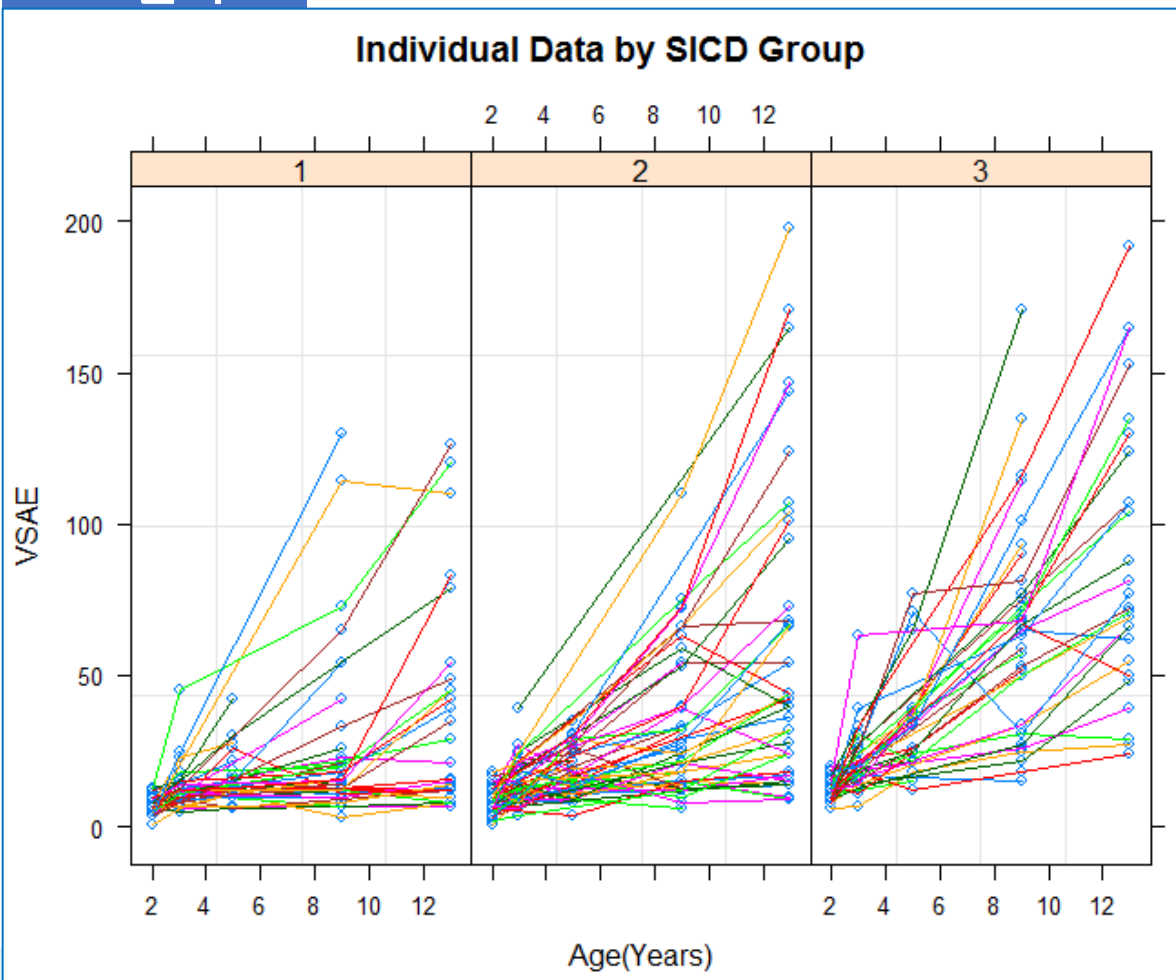


- 변량절편  $u_{0j}$ 는 2세 때의 각 아이의 효과를 모형화 하는 것.

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 아이 별 age에 따른 vsae 변화: 그래프(revisit)

R 결과



- age=2 에서 vsae 의 분산은 다른 나이의 분산과 비교하여 작고 scicdegp 수준에 따라서도 변화가 거의 없다.

➤ 변량절편  $u_{0j}$  모형에서 제거



# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 변량절편( $u_{0j}$ )의 제거

### ➤ 수정 모형(Model2)

$$\begin{aligned}
 Y_{ij} = & \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\
 & + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\
 & + \beta_7 * (age.2_{ij})^2 * sic2_j + \beta_8 * (age.2_{ij})^2 * sic3_j \\
 & + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}
 \end{aligned}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$U_j = (u_{1j}, u_{2j})' \sim^{iid} N(0, D)$ ,  $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$  이고 서로 독립,

$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ 는  $U_j$ 의  $2 \times 2$  분산-공분산 행렬.

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 변량절편( $u_{0j}$ )의 제거 모형(Model2) 적합

### Model2 적합

```
model.2.fit <- lme(vsae ~ age.2+ I(age.2^2)+ sicdegp.f+
                  age.2:sicdegp.f+ I(age.2^2):sicdegp.f,
                  random = ~ age.2+ I(age.2^2) - 1, method="REML",
                  data = autism.grouped)
summary(model.2.fit)
```

### 적합 결과(일부)

```
Linear mixed-effects model fit by REML
Data: autism.grouped
      AIC      BIC    logLik
4641.276 4698.457 -2307.638

Random effects:
Formula: ~age.2 + I(age.2^2) - 1 | childid
Structure: General positive-definite, Log-Cholesky parametrization
           StdDev   Corr
age.2      3.8298024 age.2
I(age.2^2) 0.3625824 -0.317
Residual   6.2047262

Fixed effects: vsae ~ age.2 + I(age.2^2) + sicdegp.f + age.2:sicdegp.f + I(age.2^2):sicdegp.f
              Value Std.Error   DF   t-value p-value
(Intercept)   8.353853  0.7352792  446  11.361471  0.0000
age.2          2.307043  0.7492601  446   3.079096  0.0022
I(age.2^2)     0.069383  0.0787034  446   0.881581  0.3785
sicdegp.f2     1.378406  0.9721759  155   1.417856  0.1582
sicdegp.f3     5.415784  1.0934935  155   4.952735  0.0000
age.2:sicdegp.f2 0.549878  0.9929795  446   0.553766  0.5800
age.2:sicdegp.f3 3.296437  1.0915657  446   3.019916  0.0027
I(age.2^2):sicdegp.f2 0.004917  0.1032861  446   0.047603  0.9621
I(age.2^2):sicdegp.f3 0.134615  0.1133821  446   1.187271  0.2358
```

# 1. 자폐아 자료 모형구축 : 변량계수모형

▶  $(age.2_{ij})^2$  변량계수의 유의성 검정

➤ 귀무가설 모형:  $(age.2_{ij})^2$  변량계수는 유의하지 않다  
(Model2.a).

$$Y_{ij} = \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\ + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\ + \beta_7 * (age.2_{ij})^2 * sic2_j + \beta_8 * (age.2_{ij})^2 * sic3_j \\ + u_{1j} * age.2_{ij} + \epsilon_{ij}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$u_{1j} \sim iid N(0, \sigma_1^2)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립.

# 1. 자폐아 자료 모형구축 : 변량계수모형

▶  $(age.2_{ij})^2$  변량계수의 유의성 검정

➤ 대립가설 모형:  $(age.2_{ij})^2$  변량계수는 유의하다 (Model2).

$$Y_{ij} = \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\ + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\ + \beta_7 * (age.2_{ij})^2 * sic2_j + \beta_8 * (age.2_{ij})^2 * sic3_j \\ + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}$$

$$\text{여기서 } age.2_{ij} = (age_{ij} - 2), \quad sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}, \quad sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$$

$$U_j = (u_{1j}, u_{2j})' \sim iid N(0, D), \quad \epsilon_{ij} \sim iid N(0, \sigma^2) \text{ 이고 서로 독립,}$$

$$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \text{는 } U_j \text{의 } 2 \times 2 \text{ 분산-공분산 행렬.}$$

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ $(age.2_{ij})^2$ 변량계수의 유의성 검정

### ➤ 가능도비 검정 통계량과 분포

- 가능도비검정통계량  $LR$ 의 분포

$$LR = -2\log\left(\frac{L_{H_0}}{L_{H_1}}\right) \approx^{H_0} 0.5 * \chi^2(1) + 0.5 * \chi^2(2)$$

여기서  $L_{H_0}$ 과  $L_{H_1}$ 은

각각 귀무가설과 대립가설에서 구한 REML값이고  $\chi^2(k)$ 는 자유도  $k$ 인 카이제곱 분포.



# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ $(age.2_{ij})^2$ 변량계수의 유의성 검정

### ➤ 두 가설 모형 적합

#### R cdoe

```
# 귀무가설 모형 적합과 결과
model.2a.fit <- update(model.2.fit, random = ~ age.2 - 1)
summary(model.2a.fit)
# 대립가설 모형 적합결과
summary(model.2.fit)
```

#### 적합 결과

```
> # 귀무가설 모형 적합결과
> summary(model.2a.fit)
Linear mixed-effects model fit by REML
Data: autism.grouped
      AIC      BIC    logLik
4721.203 4769.587 -2349.601
> # 대립가설 모형 적합결과
> summary(model.2.fit)
Linear mixed-effects model fit by REML
Data: autism.grouped
      AIC      BIC    logLik
4641.276 4698.457 -2307.638
```



- REML 추정 사용
- 두 모형에서 로그가능도 값 차이=  
2349.601 - 2307.638 = 41.963

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ $(age.2_{ij})^2$ 변량계수의 유의성 검정

### ➤ 검정결과

#### 검정통계량 값과 유의확률

1) 검정통계량의 관측값:

$$LR = -2 \log \left( \frac{L_{H_0}}{L_{H_1}} \right) = -2 \text{ Res Log Likelihood}_{H_0} + 2 \text{ Res Log Likelihood}_{H_1} \\ = 2 * (41.963) \approx 83.9$$

2) 유의확률(p-value):

$$\Pr(LR \geq 83.9) = 0.5 * \Pr(X_1^2 \geq 83.9) + 0.5 * \Pr(X_2^2 \geq 83.9) < 0.0001$$

여기서  $X_1^2 \sim \chi^2(1)$ ,  $X_2^2 \sim \chi^2(2)$ .

#### 유의확률 계산

```
> # (age.2_ij)^2 변량계수에 대한 유의성검정
> 0.5*(1-pchisq(83.9,1)) + 0.5*(1-pchisq(83.9,2))
[1] 0
```

•  $(age.2_{ij})^2$ 의 변량계수는 통계적으로 유의함(**Model2**).

➤  $age.2_{ij}$ 의 변량계수는 모형에 포함 시킴.

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 하향식(Top-Down) 모형구축 전략 : Step3

### 변량 요인 선택 결과 모형

$$\begin{aligned}
 Y_{ij} = & \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\
 & + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\
 & + \beta_7 * (age.2_{ij})^2 * sic2_j + \beta_8 * (age.2_{ij})^2 * sic3_j \\
 & + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}
 \end{aligned}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$ ,

$U_j = (u_{1j}, u_{2j})' \sim iid N(0, D)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립,

$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ 는  $U_j$ 의  $2 \times 2$  분산-공분산 행렬.

➤ 고정효과 모형선택1: 교호작용  $[(age.2)^2 * sicdegp]$  의 유의성 검정

# 1. 자폐아 자료 모형구축 : 변량계수모형

▶ 교호작용  $(age.2)^2 * sicdegp$  의 유의성 검정

➤ 귀무가설 (Model3)

$$H_0: \beta_7 = \beta_8 = 0$$

➤ 대립가설 (Model2)

$$H_1: \beta_7 \neq 0 \text{ or } \beta_8 \neq 0$$

# 1. 자폐아 자료 모형구축 : 변량계수모형

▶ 교호작용  $(age.2)^2 * sicdegp$  의 유의성 검정

➤ 귀무가설 모형(**Model3**)

$$Y_{ij} = \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\ + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\ + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$U_j = (u_{1j}, u_{2j})' \sim iid N(0, D)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립,

$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$  는  $U_j$  의  $2 \times 2$  분산-공분산 행렬.

# 1. 자폐아 자료 모형구축 : 변량계수모형

▶  $(age.2)^2 * sicdegp$  의 유의성 검정

## ➤ 가능비 검정통계량과 분포

- 가능도비검정통계량  $LR$ 의 분포

$$LR = -2\log\left(\frac{L_{H_0}}{L_{H_1}}\right) \approx^{H_0} \chi^2(2)$$

여기서  $L_{H_0}$ 과  $L_{H_1}$ 은  
각각 귀무가설과 대립가설에서 구한 최대가능도값

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ $(age.2)^2 * sicdegp$ 의 유의성 검정

### ➤ 모형적합과 가능도비 검정결과

#### R cdoe

```
# 대립가설 모형
model.2.ml.fit <- update(model.2.fit, method="ML")
# 귀무가설 모형
model.3.ml.fit <- update(model.2.ml.fit, fixed= ~age.2+I(age.2^2)+sicdegp.f
                                         +age.2:sicdegp.f)
anova(model.2.ml.fit, model.3.ml.fit)
```

#### 검정결과

```
> anova(model.2.ml.fit, model.3.ml.fit)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model.2.ml.fit	1	13	4636.444	4693.819	-2305.222			
model.3.ml.fit	2	11	4634.314	4682.862	-2306.157	1 vs 2	1.869704	0.3926

• 교호작용  $(age.2)^2 * sicdegp$ 의 효과는 통계적으로 유의하지 않았다(p-value=0.3926)

➤ 간결성의 원칙에 따라 **귀무가설모형(Model3)** 선택



# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ 교정효과 모형선택 결과1: Model3

$$Y_{ij} = \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\ + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\ + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$U_j = (u_{1j}, u_{2j})' \sim iid N(0, D)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립,

$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ 는  $U_j$ 의  $2 \times 2$  분산-공분산 행렬.

➤ 고정효과 모형선택2: 교호작용 ( $age.2 * sicdegp$ ) 의 유의성 검정

# 1. 자폐아 자료 모형구축 : 변량계수모형

▶  $age.2 * sicdegp$ 의 유의성 검정

➤ 귀무가설 (Model3.a)

$$H_0: \beta_5 = \beta_6 = 0$$

$$\begin{aligned} \text{즉, } Y_{ij} = & \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 \\ & + \beta_3 * sic2_j + \beta_4 * sic3_j \\ & + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij} \end{aligned}$$

➤ 대립가설(Model3)

$$H_1: \beta_5 \neq 0 \text{ or } \beta_6 \neq 0$$

# 1. 자폐아 자료 모형구축 : 변량계수모형

## ▶ *age.2 \* sicdegp* 의 유의성 검정

### ➤ 모형적합과 가능도비 검정결과

#### 귀무가설모형 적합과 가설검정

```
# age.2*sicdegp 유의성 검증
model.3a.ml.fit <- update(model.3.ml.fit, fixed= ~age.2+I(age.2^2)+sicdegp.f)
anova(model.3a.ml.fit, model.3.ml.fit)
```

#### 검정결과

```
> anova(model.3a.ml.fit, model.3.ml.fit)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model.3a.ml.fit	1	9	4653.696	4693.417	-2317.848			
model.3.ml.fit	2	11	4634.314	4682.862	-2306.157	1 vs 2	23.38232	<.0001

- 교호작용 *age.2 \* sicdegp* 의 효과는 통계적으로 유의하였다(p-value<0.0001)

➤ **대립가설모형(Model3) 선택(최종 적합모형!!)**

04

제 14강. 경시적 자료분석 LMM 변량계수모형

# LMM : 경시적 자료 모형적합 결과 해석

# 1. 최종 분석모형과 적합결과 해석

## ▶ 자폐아 자료에 대한 최종분석 모형

### ➤ 분석모형(Model3) 기술

$$Y_{ij} = \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\ + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\ + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$U_j = (u_{1j}, u_{2j})' \sim iid N(0, D)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립,

$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ 는  $U_j$ 의  $2 \times 2$  분산-공분산 행렬.

# 1. 최종 분석모형과 적합결과 해석

## ▶ 최종 모형 적합 결과

### 모형 적합

```
## 최종 모형 적합
model.3.fit <- lme(vsae ~ age.2+ I(age.2^2)+ sicdegp.f+
                  age.2:sicdegp.f,
                  random = ~ age.2+ I(age.2^2) - 1, method="REML",
                  data = autism.grouped)
summary(model.3.fit)
```

### 결과

```
> summary(model.3.fit)
Linear mixed-effects model fit by REML
Data: autism.grouped
      AIC      BIC    logLik
4633.57 4681.991 -2305.785

Random effects:
Formula: ~age.2 + I(age.2^2) - 1 | childid
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev      Corr
age.2         3.8110274 age.2
I(age.2^2)    0.3556805 -0.306
Residual      6.2281389

Fixed effects: vsae ~ age.2 + I(age.2^2) + sicdegp.f + age.2:sicdegp.f
              Value Std.Error   DF   t-value p-value
(Intercept)   8.475894 0.7094431 448  11.947249  0.0000
age.2         2.080709 0.6482319 448   3.209822  0.0014
I(age.2^2)    0.109008 0.0427795 448   2.548125  0.0112
sicdegp.f2    1.364819 0.9215857 155   1.480946  0.1407
sicdegp.f3    4.987639 1.0379064 155   4.805480  0.0000
age.2:sicdegp.f2 0.572512 0.7960151 448   0.719222  0.4724
age.2:sicdegp.f3 4.068041 0.8797676 448   4.623995  0.0000
```

# 1. 최종 분석모형과 적합결과 해석

## ▶ 최종 모형 적합 결과

### ➤ 모평균의 추정식: $\hat{E}(Y_{ij})$

$$\begin{aligned}\hat{E}(Y_{ij}) &= \hat{\beta}_0 + \hat{\beta}_1 * age.2_{ij} + \hat{\beta}_2 * (age.2_{ij})^2 + \hat{\beta}_3 * sic2_j + \hat{\beta}_4 * sic3_j \\ &\quad + \hat{\beta}_5 * age.2_{ij} * sic2_j + \hat{\beta}_6 * age.2_{ij} * sic3_j \\ &= 8.48 + 2.08 * age.2_{ij} + 0.11 * (age.2_{ij})^2 + 1.36 * sic2_j + 4.99 * sic3_j \\ &\quad + 0.57 * age.2_{ij} * sic2_j + 4.07 * age.2_{ij} * sic3_j\end{aligned}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$



## 2. 고정효과와 적합결과 해석

### ▶ 최종 모형 적합 결과

#### ➤ $\hat{E}(Y_{ij})$ 의 해석

구분	age=2	age=x
sicdegp=1	$\hat{\beta}_0 = 8.48$	$\hat{\beta}_0 + \hat{\beta}_1x + \hat{\beta}_2x^2 = 8.48 + 2.08x + 0.11x^2$
sicdegp=2	$\hat{\beta}_0 + \hat{\beta}_3 = 8.48 + 1.36$	$\hat{\beta}_0 + \hat{\beta}_3 + (\hat{\beta}_1 + \hat{\beta}_5)x + \hat{\beta}_2x^2$ $= 8.48 + 1.36 + (2.08 + 0.57)x + 0.11x^2$
sicdegp=3	$\hat{\beta}_0 + \hat{\beta}_4 = 8.48 + 4.99$	$\hat{\beta}_0 + \hat{\beta}_4 + (\hat{\beta}_1 + \hat{\beta}_6)x + \hat{\beta}_2x^2$ $= 8.48 + 4.99 + (2.08 + 4.07)x + 0.11x^2$

- $\hat{\beta}_0 = 8.48(se = 0.71)$  은 sicdegp=1, age=2 인 아이들의 vsae 평균에 대한 추정치
- $\hat{\beta}_3 = 1.36(0.92)$ ,  $\hat{\beta}_4 = 4.99(1.04)$  는 각각 age=2 인 아이들의 sicdegp=1 vs. sicdegp=2, sicdegp=1 vs. sicdegp=3 비교에서 vsae 평균차이에 대한 추정치. sicdegp 각 수준에서 추정된 age의 영향 곡선의 절편의 차이.
- $\hat{\beta}_1 = 2.08(0.65)$  는 sicdegp=1인 아이들에서 추정된 age 영향 곡선의 일차항 계수.
- $\hat{\beta}_5 = 0.57(0.80)$ ,  $\hat{\beta}_6 = 4.07(0.88)$  는 추정된 age 영향 곡선의 sicdegp=1 vs. sicdegp=2, sicdegp=1 vs. sicdegp=3 비교에서 일차항 계수 값의 차이.
- $\hat{\beta}_2 = 0.11(0.04)$  는 sicdegp 각 수준에 상관없이 추정된 age 영향 곡선의 이차항 계수.

## 2. 고정효과와 적합결과 해석

### ▶ 최종 모형 적합 결과

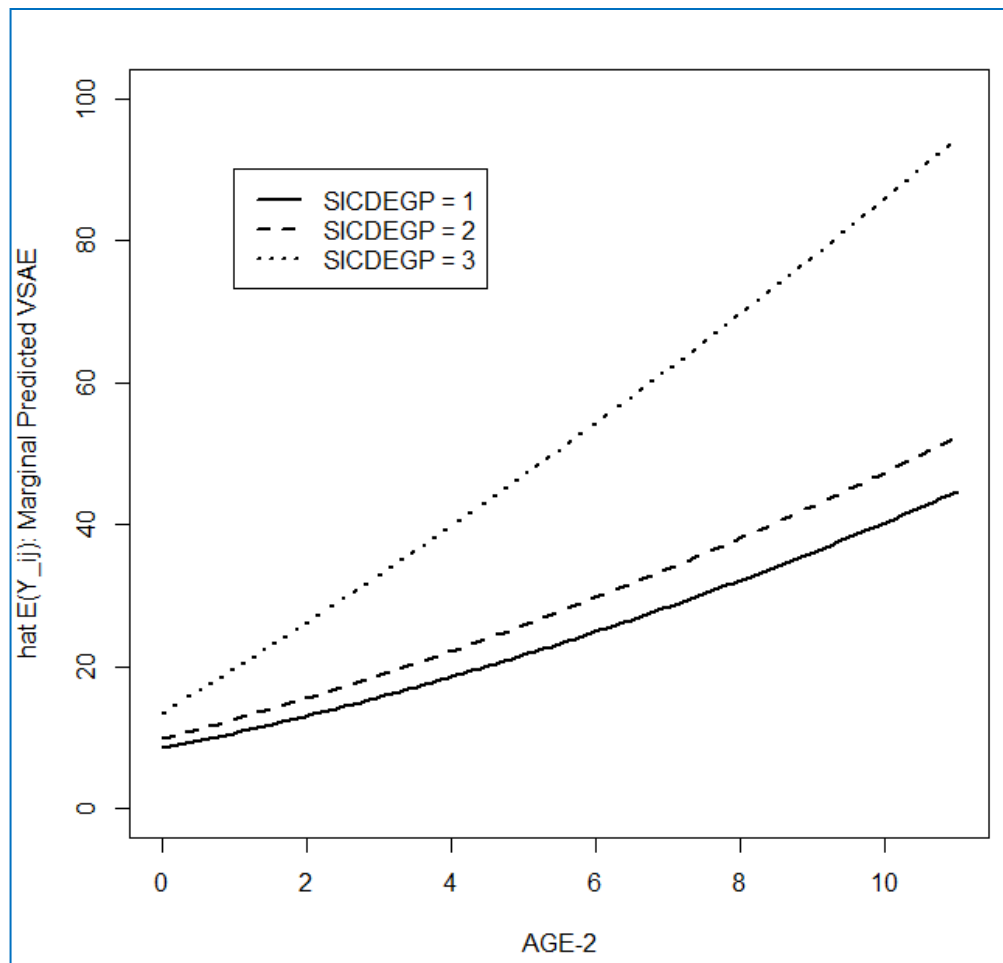
#### ➤ 고정효과 모수의 신뢰구간 추정

```
> intervals(model.3.fit)
Approximate 95% confidence intervals

Fixed effects:
              lower      est.      upper
(Intercept)  7.08164412  8.4758938  9.8701435
age.2        0.80675643  2.0807093  3.3546622
I(age.2^2)    0.02493412  0.1090076  0.1930811
sicdegp.f2   -0.45567009  1.3648185  3.1853071
sicdegp.f3    2.93737183  4.9876387  7.0379056
age.2:sicdegp.f2 -0.99187569  0.5725116  2.1368988
age.2:sicdegp.f3  2.33905737  4.0680411  5.7970249
```

## 2. 고정효과와 적합결과 해석

### ➤ $\hat{E}(Y_{ij})$ 의 그래프



➤ 2세 시기의 소통발달정도(sicdegp)가 주어졌을 때 나이(age)의 영향

- ✓ 나이(age)가 증가함에 따라 사회화정도(vsae)는 통계적으로 유의한 2차 곡선의 형태를 띠며 증가하였다.

➤ 2세 시기의 소통발달정도(sicdegp)의 영향

- sicdegp 1수준과 2수준에 있는 아이들의 나이에 따른 사회화정도의 발달곡선 형태는 통계적인 측면에서 유사하였다.
- 반면에 sicdegp 3수준 아이들의 나이에 따른 사회화정도의 발달곡선은 1수준 및 2수준 아이들의 그것과 통계적으로 유의한 차이가 있었으며 발달이 빠른 것으로 나타났다.

## 2. 고정효과 적합결과 해석

### ▶ $\hat{E}(Y_{ij})$ 의 그래프와 신뢰구간 추정 R code

```
# sicdegp 각 수준에서 추정된 age 영향곡선
curve(0.11*x^2 + 6.15*x + 13.47, 0, 11, xlab= "AGE-2",
      ylab= "hat E(Y_ij): Marginal Predicted VSAE",
      lty = 3, ylim=c(0,100), lwd = 2)
curve(0.11*x^2 + 2.65*x + 9.84, 0, 11, add=T, lty = 2, lwd = 2)
curve(0.11*x^2 + 2.08*x + 8.48, 0, 11, add=T, lty = 1, lwd = 2)
legend(1,90,c("SICDEGP = 1", "SICDEGP = 2","SICDEGP = 3"),
      lty = c(1,2,3), lwd = c(2,2,2))

# 모수의 신뢰구간 추정
intervals(model.3.fit)
```

### 3. 분산-공분산 모수 적합결과 해석

#### ▶ 자폐아 자료에 대한 최종분석 모형

#### ➤ 분석모형(Model3) 기술

$$Y_{ij} = \beta_0 + \beta_1 * age.2_{ij} + \beta_2 * (age.2_{ij})^2 + \beta_3 * sic2_j + \beta_4 * sic3_j \\ + \beta_5 * age.2_{ij} * sic2_j + \beta_6 * age.2_{ij} * sic3_j \\ + u_{1j} * age.2_{ij} + u_{2j} * (age.2_{ij})^2 + \epsilon_{ij}$$

여기서  $age.2_{ij} = (age_{ij} - 2)$ ,  $sic2_j = \begin{cases} 1 & sicdegp = 2 \\ 0 & otherwise \end{cases}$ ,  $sic3_j = \begin{cases} 1 & sicdegp = 3 \\ 0 & otherwise \end{cases}$

$U_j = (u_{1j}, u_{2j})' \sim iid N(0, D)$ ,  $\epsilon_{ij} \sim iid N(0, \sigma^2)$  이고 서로 독립,

$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ 는  $U_j$ 의  $2 \times 2$  분산-공분산 행렬.

### 3. 분산-공분산 모수 적합결과 해석

#### ▶ 최종 모형 적합 결과

##### 결과

```
> summary(model.3.fit)
Linear mixed-effects model fit by REML
Data: autism.grouped
      AIC      BIC    logLik
4633.57 4681.991 -2305.785

Random effects:
Formula: ~age.2 + I(age.2^2) - 1 | childid
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev      Corr
age.2         3.8110274 age.2
I(age.2^2)    0.3556805 -0.306
Residual      6.2281389
```

- $\hat{\sigma}^2 = (6.23)^2 = 38.79$

- $\hat{D} = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{pmatrix} = \begin{pmatrix} 3.81^2 & -0.31 * 3.81 * 0.36 \\ -0.31 * 3.81 * 0.36 & 0.36^2 \end{pmatrix}$   
 $= \begin{pmatrix} 14.52 & -0.43 \\ -0.43 & 0.13 \end{pmatrix}$

05

제 14강. 경시적 자료분석 LMM 변량계수모형

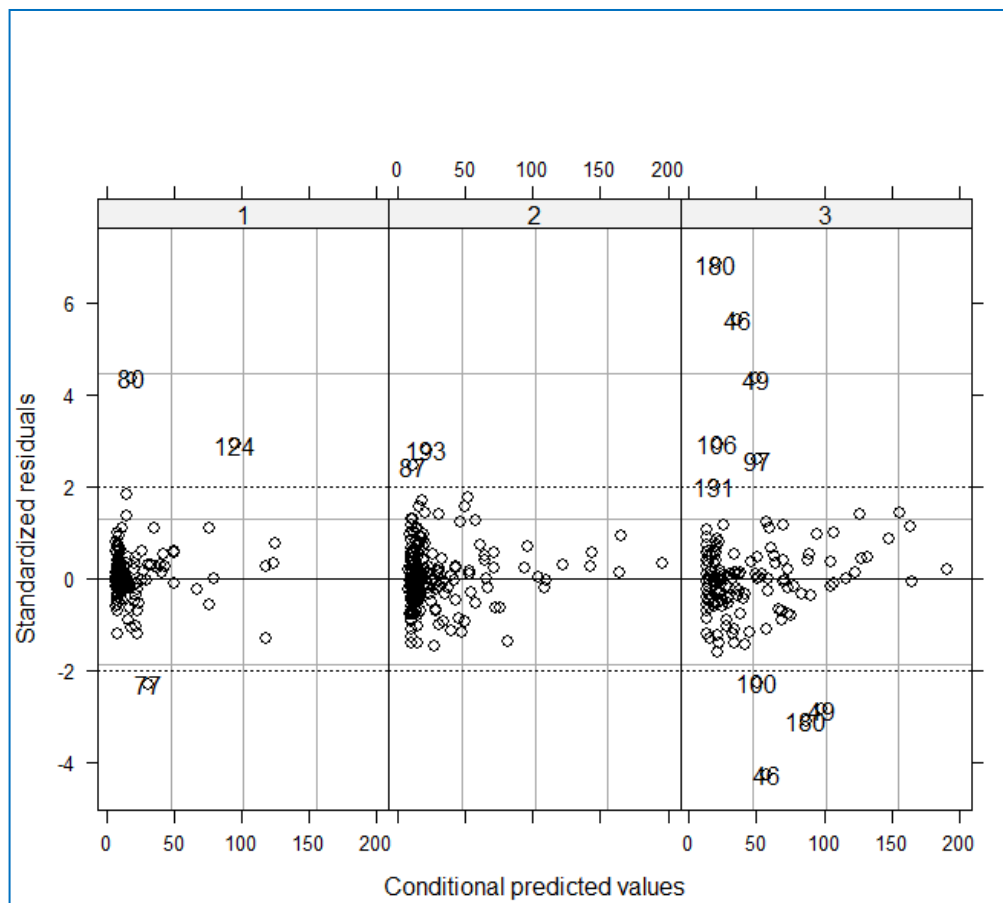
# LMM : 모형진단



# 1. 조건부 잔차 분석

▶ 조건부 잔차:  $r_{ij}^c = Y_{ij} - X_{ij}\hat{\beta} - Z_{ij}\hat{U}_j$

## ▶ 조건부 예측값 vs. 표준화 조건부 잔차 그림



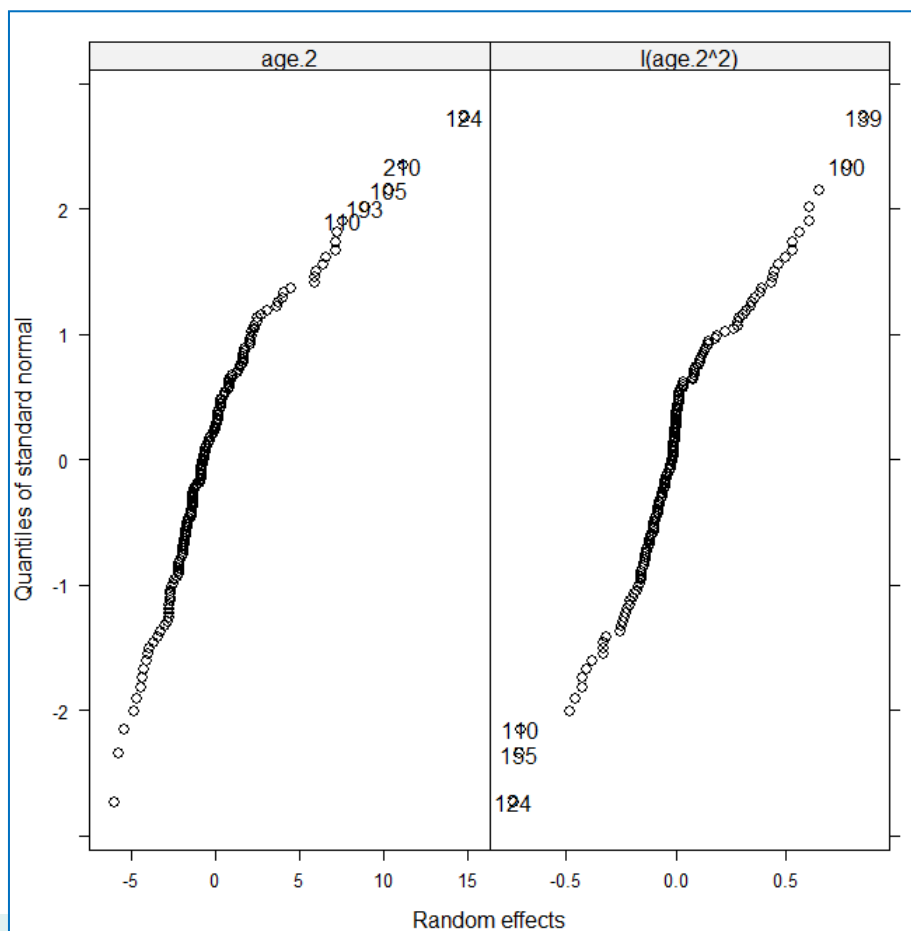
- 조건부 예측값이 증가할수록 잔차는 감소하는 경향이 있으며 이상치로 의심되는 값들이 존재한다.



## 2. EBLUPs 그림 : 변량효과와 분포 가정 진단

▶ 변량효과 예측값:  $\widehat{U}_j = (\widehat{u}_{1j}, \widehat{u}_{2j})'$

➤ 변량효과 예측값의 Normal Q-Q plots



- 직선으로터 어느정도의 이탈이 나타나고 있으며, 124번 관측치는 두 변량효과 모두에서 이상치로 제시되고 있다.

# 3. 이상치 영향력 진단

## ▶ 46번, 124번, 180번 관측값의 영향력 진단: 모수추정값의 변화

### 해당 아이 제외

Approximate 95% confidence intervals

Fixed effects:

	lower	est.	upper
(Intercept)	7.35728574	8.5248757	9.6924656
age.2	0.52948756	1.7242760	2.9190645
I(age.2^2)	0.03387995	0.1152578	0.1966356
sicdegp.f2	-0.21050317	1.3153294	2.8411620
sicdegp.f3	2.46228761	4.2033426	5.9443975
age.2:sicdegp.f2	-0.54872561	0.9447972	2.4383200
age.2:sicdegp.f3	2.73384769	4.4074270	6.0810063

attr(,"label")  
[1] "Fixed effects:"

Random Effects:

Level: childid

	lower	est.	upper
sd(age.2)	3.0211393	3.5678380	4.21346604
sd(I(age.2^2))	0.3027171	0.3660752	0.44269412
cor(age.2,I(age.2^2))	-0.4538808	-0.2526656	-0.02693256

### 모두 포함

Approximate 95% confidence intervals

Fixed effects:

	lower	est.	upper
(Intercept)	7.08164412	8.4758938	9.8701435
age.2	0.80675643	2.0807093	3.3546622
I(age.2^2)	0.02493412	0.1090076	0.1930811
sicdegp.f2	-0.45567009	1.3648185	3.1853071
sicdegp.f3	2.93737183	4.9876387	7.0379056
age.2:sicdegp.f2	-0.99187569	0.5725116	2.1368988
age.2:sicdegp.f3	2.33905737	4.0680411	5.7970249

attr(,"label")  
[1] "Fixed effects:"

Random Effects:

Level: childid

	lower	est.	upper
sd(age.2)	3.1954537	3.8110274	4.54518548
sd(I(age.2^2))	0.2889373	0.3556805	0.43784105
cor(age.2,I(age.2^2))	-0.5137880	-0.3062872	-0.06493179



- 46번, 124번, 180번 관측값을 제외하고 분석한 결과, 모수추정값은 조금 변화하였지만 추정값의 방향과 유의성에는 변화가 없었다. 즉, 이들 관측값이 모수추정값에 미치는 영향을 크지 않았다.

## 4. 모형 진단 결과

### ▶ 모형진단 결과와 새로운 분석 방안

#### ➤ 표준화 조건부 잔차 진단 결과

- 오차에 대한 등분산성과 정규성 가정이 타당한지에 대한 의심이 생김
- 오른쪽으로 긴 꼬리를 가지는 분포형태가 나타남

#### ➤ 새로운 분석 방안

- 오차분산에 대한 이분산성 (heteroscedasticity) 가정
- $Y_{ij}$ 에 대한 새로운 분포 가정(예: Gamma 분포)
- $Y_{ij}$ 에 대한 변수 변환(예: log 변환)

## 5. R code

### ▶ R code: 모형 진단과 영향력 진단

#### R code

```
### 잔차진단
## 조건부 예측값 vs. 표준화 조건부 잔차 그림
plot(model.3.fit, resid(., type="p") ~ fitted(.) | factor(sicdegp),
      xlab="Conditional predicted values", layout=c(3,1),
      aspect=2, abline=c(0,2,-2), lty=c(1,3,3), id = 0.05)

## age.2 vs. 표준화 조건부 잔차 그림
plot(model.3.fit, resid(., type="p") ~ age.2,
      xlab="Age - 2", aspect=2,
      abline=c(0,2,-2), lty=c(1,3,3), id = 0.05)

## 표준화 조건부 잔차에 대한 Normal Q-Q plots
qqnorm(model.3.fit, ~resid(., type="p") | factor(sicdegp),
        layout=c(3,1), aspect = 2, id = 0.05)

### 변량효과에 대한 진단
## 변량효과에 대한 Normal Q-Q plots
qqnorm(model.3.fit, ~ranef(.), id = 0.05)

## 변량효과 예측값들 간의 산점도
pairs(model.3.fit, ~ranef(.) | factor(sicdegp),
      layout=c(3,1), aspect= 2, id = 0.05)

### 조건부 예측값 vs. 관측값
plot(model.3.fit, vsae ~ fitted(.) | factor(sicdegp), id = 0.05,
      layout=c(3,1), xlab="Conditional predicted values", aspect= 2)

### 영향력 진단
## 46번, 124번, 180번 아이의 영향력진단
autism.grouped2 <- autism.grouped[(autism.grouped$schildid != 46
  & autism.grouped$schildid != 124 & autism.grouped$schildid != 180),]
model.3.fit.out <- update(model.3.fit, data = autism.grouped2)
summary(model.3.fit.out)
intervals(model.3.fit.out)
```



# 수고하셨습니다.

---