

# 패널 다층모형 & 성장곡선분석

(Panel Multilevel Models & Growth Curve Analysis)

황 선 재

충남대학교 사회학과  
sunjaeh@gmail.com

2019/09/26

# 2019 한국복지패널 학술대회 특강

# 강의개요

2

- 다층모형 소개
  - ▣ 정의 및 의의
    - 다층자료를 분석대상으로 하는 회귀분석 기법
  - ▣ 다층모형의 기초
  - ▣ 예시 w/ Stata
  
- 다층모형 활용: 성장곡선분석
  - ▣ 정의 및 의의
    - 다층모형을 활용한 패널자료 분석 기법 중 하나
  - ▣ 성장곡선분석 ‘기본’ 모형
  - ▣ 예시 w/ Stata

3

## 다층모형 소개

# 정의

4

- 다층모형(Multi-Level Models: MLM)이란?
  - ▣ 다층자료를 분석대상으로 하는 회귀분석 기법
    - 다층자료(multilevel data)는 (복수의) 관측값이 특정 개체/집단에 속하거나 군집되어 있는 자료를 지칭
  - ▣ 다층자료의 예
    - 1) 횡단 다층자료
      - 가족 별 가족 구성원을 관측한 자료
        - Level 1: 가족구성원, Level 2: 가족
      - 학군 별, 학교 별, 학생들을 관측한 자료
        - Level 1: 학생, Level 2: 학교, Level 3: 학군
    - 2) 패널 다층자료
      - 동일한 개체를 대상으로 복수의 시간에 걸쳐 측정한 자료 (즉, 패널자료)
        - Level 1: 개체내 시계열 관측치, Level 2: 개체

# 정의

5

- 다층모형의 (또 다른) 명칭
  - ▣ 위계선형모형 (Hierarchical Linear Model: HLM)
  - ▣ 임의계수모형 (Random-Coefficient Model)
  - ▣ 공분산 구성 모형 (Covariance Component Model)
  - ▣ 혼합효과모형 (Mixed-Effect Model)
    - ‘고정효과’ 부분과 ‘임의효과’ 부분이 혼합된 모형
      - 패널 자료 분석에서 사용하는 고정효과(fixed-effect model) 및 임의효과(random-effect model)와는 의미가 다름에 유의!!
      - 혼란을 피하기 위해 이 모형의 고정효과 부분은 ‘고정요소’로, 임의효과 부분은 ‘임의요소’로 지칭
    - Stata에서 실제로 다층모형을 추정하는 방식
  - ▣ 혼합모형 (Mixed Model)

# 의의

6

- 다층모형이 왜 필요한가?
  - ▣ 통계적 이유
    - 다층자료에 본질적으로 존재하게 마련인 동시적 상관(주로 횡단자료의 경우)과 자기상관(주로 종단자료의 경우)의 문제를 명시적으로 다룸
    - 같은 층위에 속한 변수들 간의 관계 뿐 아니라, 서로 다른 층위에 속한 변수들간의 관계도 (보다 직관적으로) 모형화 가능
  - ▣ 실질적 이유
    - 개체의 속성, 개체가 속한 집단/구조의 속성, 그리고 이들의 상호관계를 한꺼번에 명시적으로 다룬다는 측면에서 가장 사회과학적인 통계적 방법론 중 하나

# 다층모형의 기초

7

## □ 회귀분석의 (단순) 확장으로서의 MLM

- 1) 선형회귀 모형 (linear regression model)

$$y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij}$$

- 2) 임의절편 모형 (random-intercept model)

$$y_{ij} = \alpha_j + \beta x_{ij} + \varepsilon_{ij}$$

- 3) 임의계수 모형 (random-coefficient model)

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \varepsilon_{ij}$$

- 4) 다층/혼합효과 모형

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \varepsilon_{ij}$$

# 다층모형의 기초

8

## □ 회귀분석의 확장으로서의 MLM (계속)

### □ 4-1) 다층/혼합효과 ‘기본’ 모형: 분리형

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \varepsilon_{ij} \quad (\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)) \quad \text{(Level 1)}$$

$$\alpha_j = \gamma_0 + \gamma_1 z_j + u_{1j} \quad (u_{1j} \sim N(0, \sigma_{u1}^2)) \quad \text{(Level 2)}$$

$$\beta_j = \delta_0 + \delta_1 k_j + u_{2j} \quad (u_{2j} \sim N(0, \sigma_{u2}^2))$$

### □ 4-2) 다층/혼합효과 ‘기본’ 모형: 결합형

$$\begin{aligned} y_{ij} &= (\gamma_0 + \gamma_1 z_j + u_{1j}) + (\delta_0 + \delta_1 k_j + u_{2j}) x_{ij} + \varepsilon_{ij} \\ &= (\gamma_0 + \gamma_1 z_j + \delta_0 x_{ij} + \delta_1 k_j x_{ij}) + (u_{1j} + u_{2j} x_{ij} + \varepsilon_{ij}) \\ &\quad \text{‘고정요소’} \qquad \qquad \qquad \text{‘임의요소’} \end{aligned}$$



# 예시: Stata & HSB Data

9

## □ Stata w/ *High School and Beyond* Data

```
. use http://www.stata-press.com/data/mlmus2/hsb, clear
. xtsum
```

Variable		Mean	Std. Dev.	Min	Max	Observations
schoolid	overall	5277.898	2499.578	1224	9586	N = 7185
	between		2547.683	1224	9586	n = 160
	within		0	5277.898	5277.898	T-bar = 44.9063
mathach	overall	12.74785	6.878246	-2.832	24.993	N = 7185
	between		3.117651	4.239781	19.71914	n = 160
	within		6.186706	-6.926784	30.71674	T-bar = 44.9063
ses	overall	.0001434	.7793552	-3.758	2.692	N = 7185
	between		.4139706	-1.193946	.8249825	n = 160
	within		.660588	-3.650597	2.856222	T-bar = 44.9063
sector	overall	.4931106	.4999873	0	1	N = 7185
	between		.4976359	0	1	n = 160
	within		0	.4931106	.4931106	T-bar = 44.9063

# 예시: 선형회귀 모형

10

```
. regress mathach ses
```

Source	SS	df	MS	Number of obs	=	7,185
Model	44233.1556	1	44233.1556	F(1, 7183)	=	1074.69
Residual	295643.779	7,183	41.1588165	Prob > F	=	0.0000
Total	339876.934	7,184	47.3102637	R-squared	=	0.1301
				Adj R-squared	=	0.1300
				Root MSE	=	6.4155

mathach	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ses	3.18387	.0971209	32.78	0.000	2.993485	3.374256
_cons	12.7474	.0756864	168.42	0.000	12.59903	12.89576

# 예시: 임의절편 모형

11

```
. xtmixed mathach ses || schoolid:, mle nolog
```

Mixed-effects ML regression  
Group variable: schoolid

Number of obs = 7,185  
Number of groups = 160

Obs per group:

min = 14  
avg = 44.9  
max = 67

Log likelihood = -23320.502

Wald chi2(1) = 511.98  
Prob > chi2 = 0.0000

‘고정요소’ {

mathach	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ses	2.391499	.1056926	22.63	0.000	2.184346	2.598653
_cons	12.65762	.1873212	67.57	0.000	12.29048	13.02477

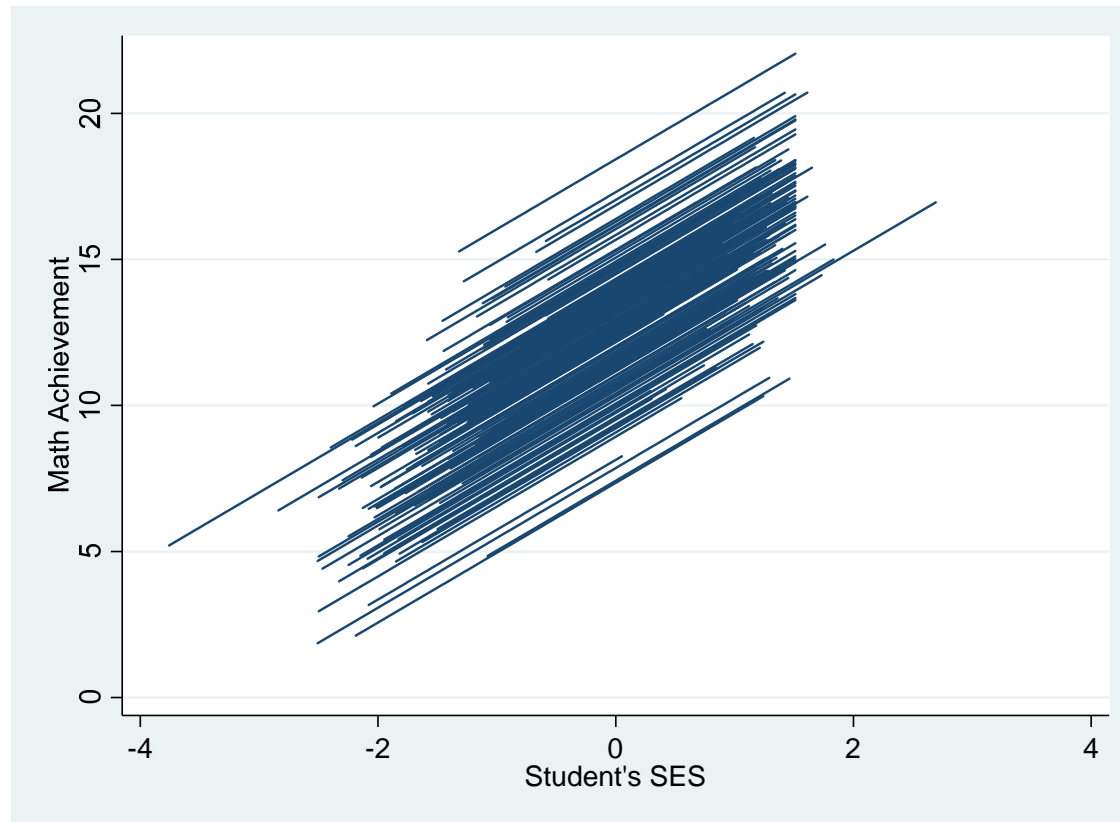
‘임의요소’ {

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
schoolid: Identity				
sd(_cons)	2.174516	.1491543	1.900978	2.487414
sd(Residual)	6.085211	.0513769	5.985342	6.186745

LR test vs. linear model: chibar2(01) = 456.94 Prob >= chibar2 = 0.0000

# 예시: 임의절편 모형

12



# 예시: 임의계수 모형

13

```
. xtmixed mathach ses || schoolid: ses, mle nolog
```

```
Mixed-effects ML regression      Number of obs   =    7,185
Group variable: schoolid         Number of groups =     160

Obs per group:
    min =          14
    avg =         44.9
    max =           67

Wald chi2(1)      =    412.75
Prob > chi2      =     0.0000

Log likelihood = -23318.373
```

mathach	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ses	2.396654	.1179671	20.32	0.000	2.165443	2.627865
_cons	12.65302	.1895413	66.76	0.000	12.28153	13.02452

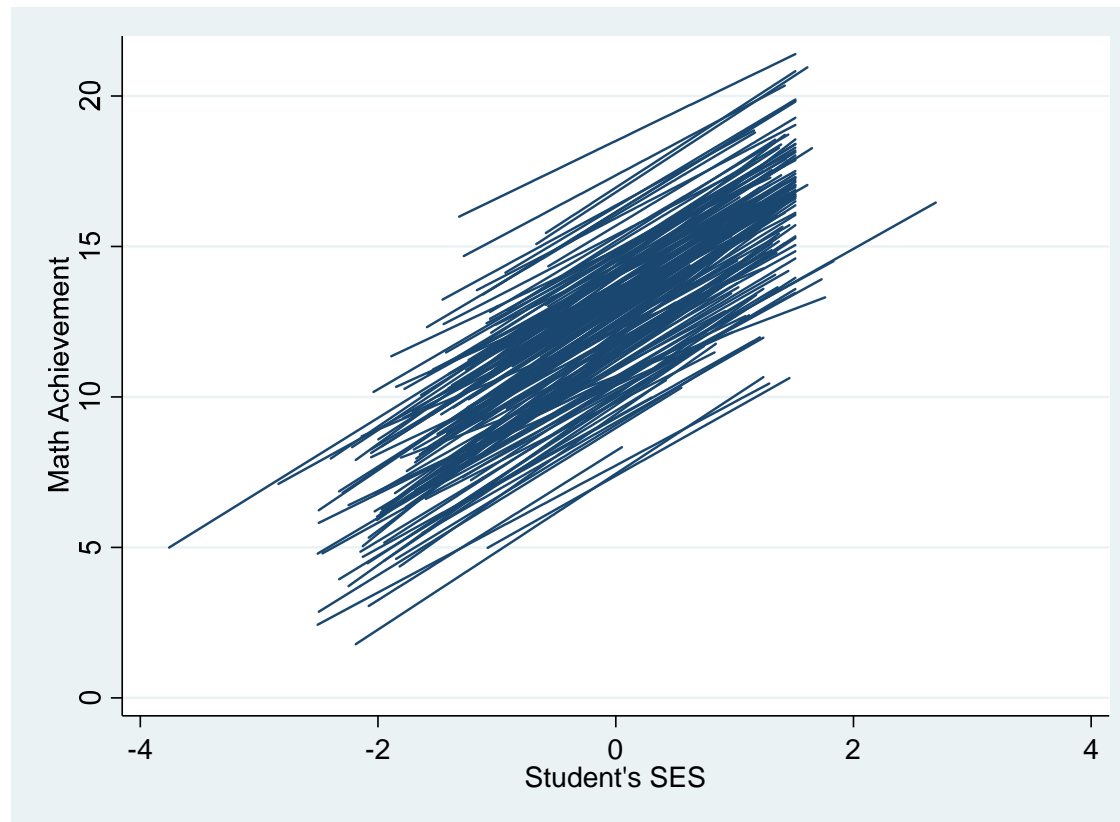
Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
schoolid: Independent				
sd(ses)	.6400501	.1815615	.3670757	1.116021
sd(_cons)	2.19312	.1519763	1.914595	2.512164
sd(Residual)	6.06823	.0518184	5.967513	6.170647

```
LR test vs. linear model: chi2(2) = 461.19      Prob > chi2 = 0.0000
```

Note: LR test is conservative and provided only for reference.

# 예시: 임의계수 모형

14



## 예시: 다층 모형 (모형)

15

### □ 회귀분석의 확장으로서의 MLM (8쪽과 비교)

#### □ 1) 다층/혼합효과 ‘기본’ 모형: 분리형

$$math_{ij} = \alpha_j + \beta_j ses_{ij} + \varepsilon_{ij} \quad (\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)) \quad \text{(Level 1)}$$

$$\alpha_j = \gamma_0 + \gamma_1 sector_j + u_{1j} \quad (u_{1j} \sim N(0, \sigma_{u1}^2)) \quad \text{(Level 2)}$$

$$\beta_j = \delta_0 + \delta_1 sector_j + u_{2j} \quad (u_{2j} \sim N(0, \sigma_{u2}^2))$$

#### □ 2) 다층/혼합효과 ‘기본’ 모형: 결합형

$$\begin{aligned} math_{ij} &= (\gamma_0 + \gamma_1 sector_j + u_{1j}) + (\delta_0 + \delta_1 sector_j + u_{2j}) ses_{ij} + \varepsilon_{ij} \\ &= (\gamma_0 + \gamma_1 sector_j + \delta_0 ses_{ij} + \delta_1 sector_j ses_{ij}) + (u_{1j} + u_{2j} ses_{ij} + \varepsilon_{ij}) \end{aligned}$$

‘고정요소’

‘임의요소’

# 예시: 다층 모형 (추정)

16

```
. xtmixed mathach ses i.sector i.sector#c.ses || schoolid: ses, mle nolog
```

Mixed-effects ML regression  
Group variable: schoolid

Number of obs = 7,185  
Number of groups = 160

Obs per group:

min = 14  
avg = 44.9  
max = 67

Log likelihood = -23284.091

Wald chi2(3) = 617.69  
Prob > chi2 = 0.0000

mathach	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
ses	2.953381	.1405383	21.01	0.000	2.677931 3.228831
sector Catholic	2.137317	.3389178	6.31	0.000	1.47305 2.801584
sector#c.ses Catholic	-1.312292	.2118813	-6.19	0.000	-1.727572 -.8970125
_cons	11.79835	.2269272	51.99	0.000	11.35358 12.24312

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
schoolid: Independent			
sd(ses)	.0001712	.0002742	7.42e-06 .0039513
sd(_cons)	1.905169	.1367453	1.655151 2.192953
sd(Residual)	6.068864	.0512393	5.969263 6.170127

LR test vs. linear model: chi2(2) = 338.63

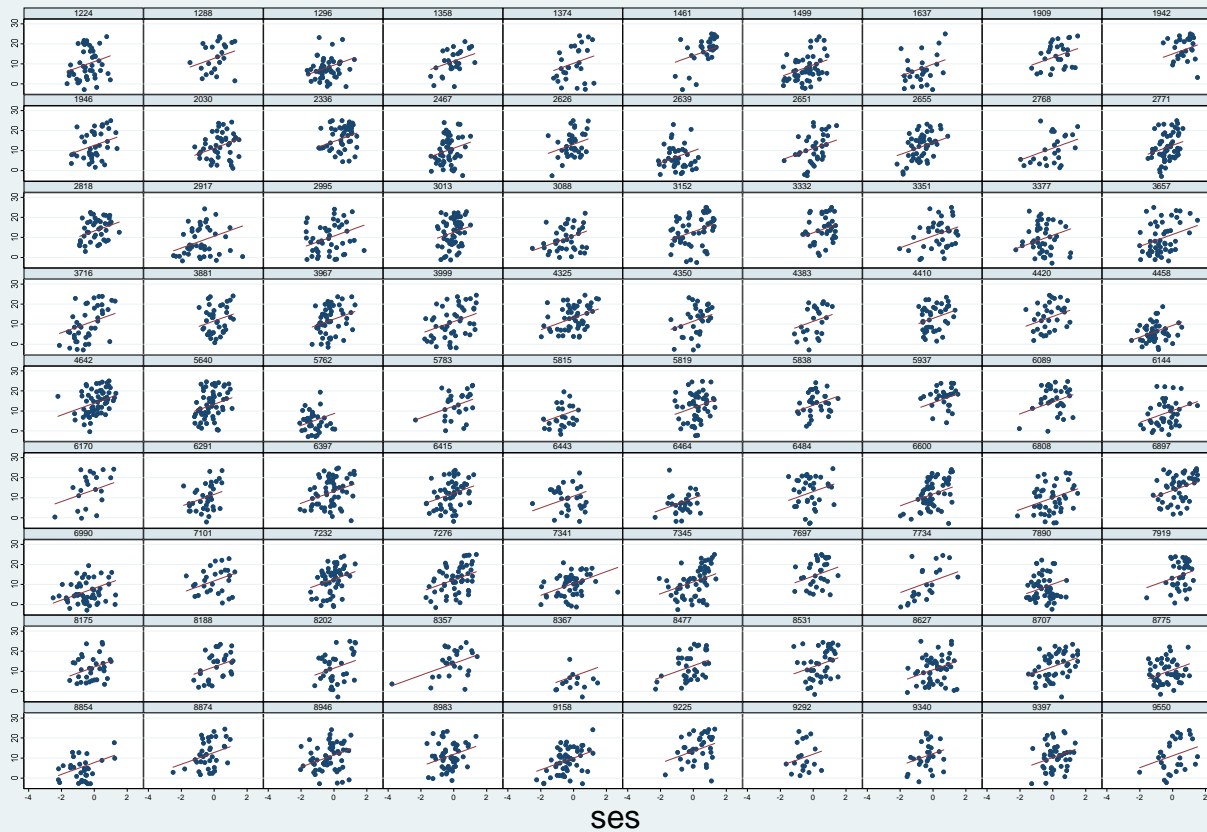
Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.



# 예시: 다층 모형 (그래프 for 'Public')

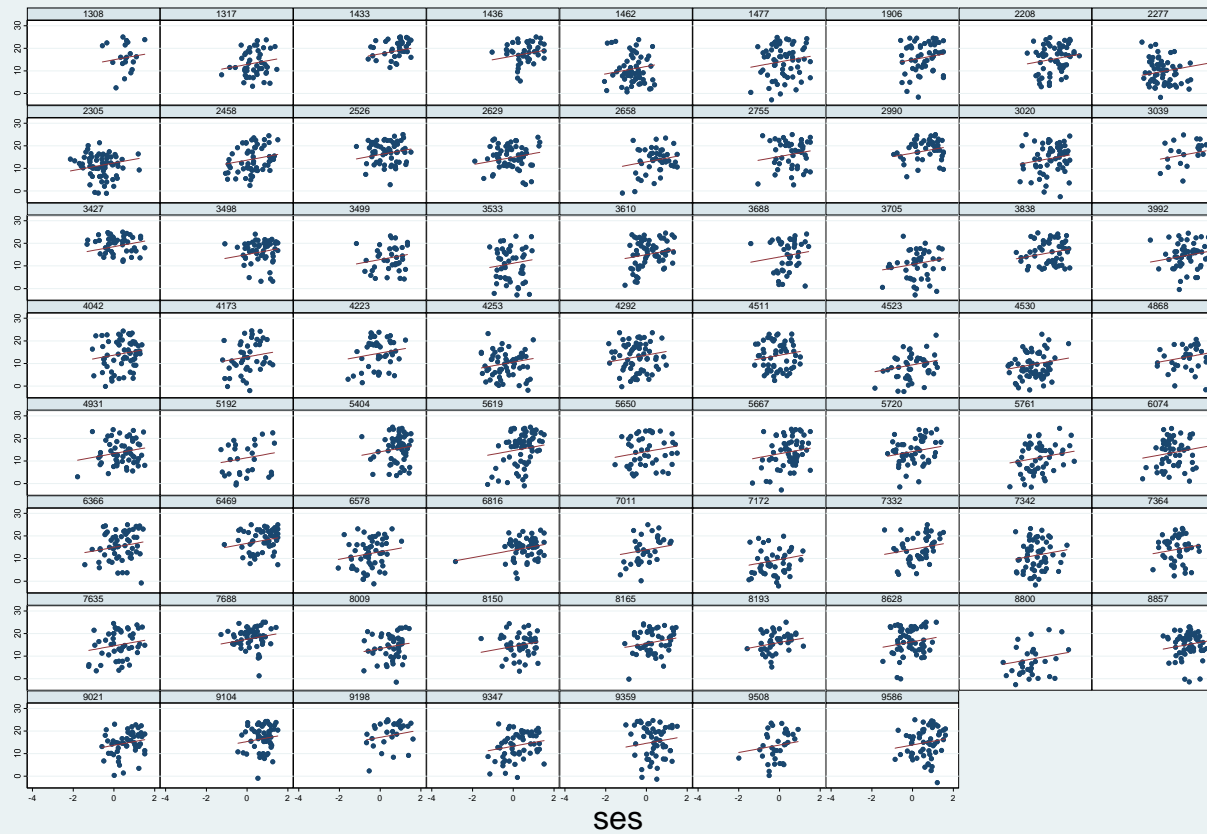
17



Graphs by schoolid

# 예시: 다층 모형 (그래프 for 'Catholic')

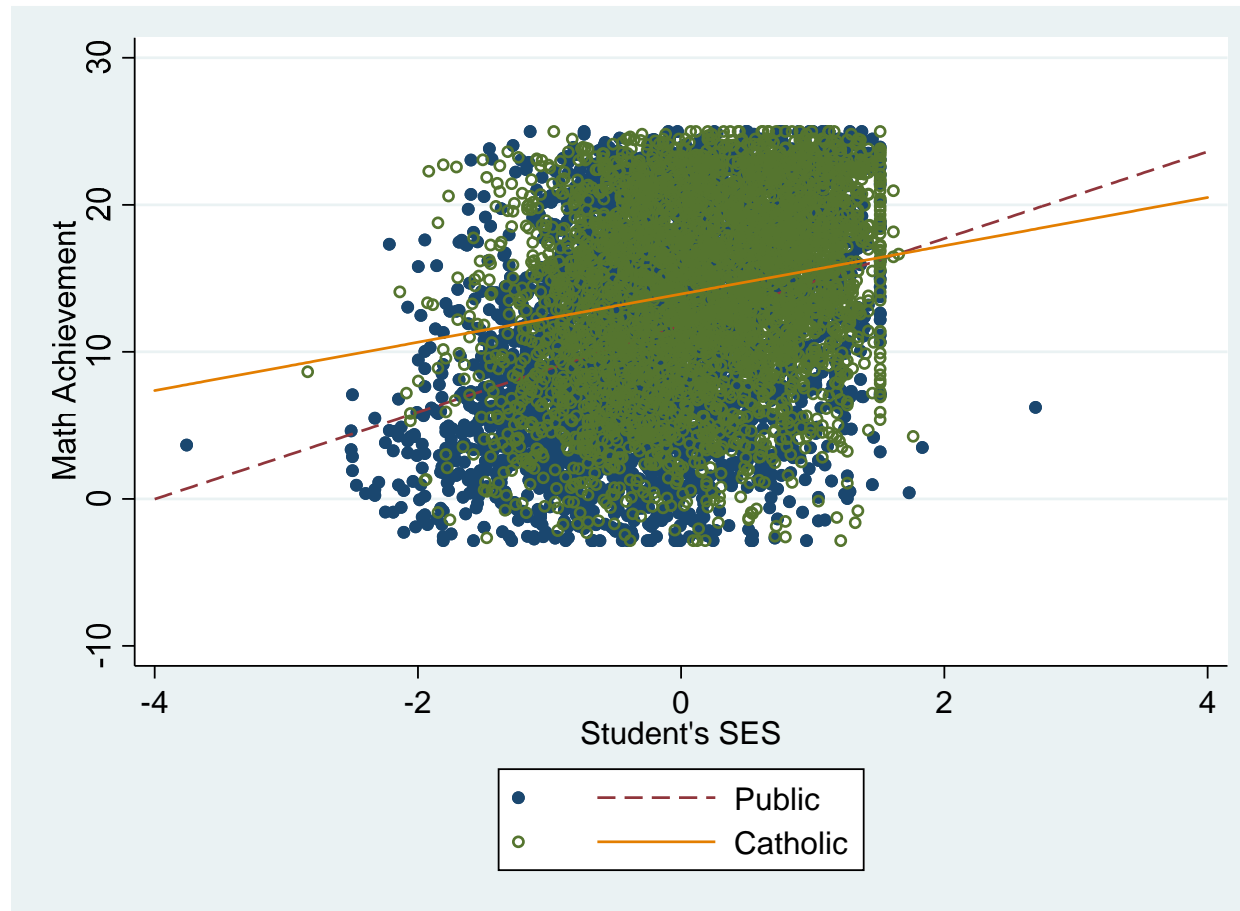
18



Graphs by schoolid

## 예시: 다층 모형 (그래프 for Both)

19



20

## 다층모형 활용: 성장곡선분석

# 정의 및 의의

21

- 패널자료는 특정 개체를 시간에 따라 반복적으로 관찰하여 기록한 자료이며, 앞서 논의한 바와 같이 그 구조상 다층자료 중 하나로 간주할 수 있음.
  - ▣ 즉, 패널자료란 개체별로 복수의 시계열 관측치가 존재하는 다층자료
    - Level 1: 개체별 시계열 관측치, Level 2: 개체
- 성장곡선분석(Growth Curve Analysis)은 이러한 패널자료 개체의 특정 변수가 시간에 따라 어떻게 변해가는지, 그 변화의 양상이 개체별로 다른지를 분석하는, 패널자료용 다층모형 분석기법 중 하나
  - ▣ 1) 시간의 흐름에 따른 종속변수의 변화양상을 개체별로 추정
  - ▣ 2) 개체별 변화양상의 차이를 독립변수로 설명

# 성장곡선분석 ‘기본’ 모형

22

- 성장곡선분석
  - 다층/혼합효과 ‘기본’ 모형: 분리형

$$y_{it} = \alpha_i + \beta_i \text{TIME}_{it} + \varepsilon_{it} \quad (\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2))$$

$$\alpha_i = \gamma_0 + \gamma_1 z_i + u_{1i} \quad (u_{1i} \sim N(0, \sigma_{u1}^2))$$

$$\beta_i = \delta_0 + \delta_1 k_i + u_{2i} \quad (u_{2i} \sim N(0, \sigma_{u2}^2))$$

- 다층/혼합효과 ‘기본’ 모형: 결합형

$$\begin{aligned} y_{it} &= (\gamma_0 + \gamma_1 z_i + u_{1i}) + (\delta_0 + \delta_1 k_i + u_{2i}) \text{TIME}_{it} + \varepsilon_{it} \\ &= (\gamma_0 + \gamma_1 z_i + \delta_0 \text{TIME}_{it} + \delta_1 k_i \text{TIME}_{it}) + (u_{1i} + u_{2i} \text{TIME}_{it} + \varepsilon_{it}) \end{aligned}$$

‘고정요소’,

‘임의요소’,

# 예시: Stata w/ Weight x Age Data

23

```
. use http://www.stata-press.com/data/mlmus2/asian, clear
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	198	2491.227	1379.389	45	4975
occ	198	2.111111	1.031375	1	5
age	198	1.080552	.787069	.1149897	2.546201
weight	198	8.837091	3.02723	3.1	17.2
brthwt	198	3106.081	537.5595	1575	4270
gender	198	1.494949	.5012419	1	2

	id	occ	age	weight	brthwt	gender
1	45	1	.136893	5.171	4140	1
2	45	2	.657084	10.86	4140	1
3	45	3	1.21834	13.15	4140	1
4	45	4	1.42916	13.2	4140	1
5	45	5	2.27242	15.88	4140	1
6	258	1	.19165	5.3	3155	2
7	258	2	.687201	9.74	3155	2
8	258	3	1.12799	9.98	3155	2
9	258	4	2.30527	11.34	3155	2
10	287	1	.134155	4.82	3850	1
11	287	2	.70089	9.09	3850	1
12	287	3	1.16906	11.1	3850	1
13	287	4	2.2423	16.8	3850	1
14	483	1	.747433	5.76	2875	2
15	483	2	1.01848	6.92	2875	2
16	483	3	2.24504	9.53	2875	2
17	725	1	.120465	4.4	3280	2
18	725	2	2.30527	12.25	3280	2
19	800	1	1.12252	10.89	3900	1
20	800	2	2.26146	12.7	3900	1
21	801	1	.11499	3.9	2520	1
22	801	2	2.28063	12	2520	1
23	832	1	.117728	3.52	2460	2
24	832	2	.635181	7.22	2460	2
25	832	3	2.30801	12.8	2460	2
26	833	1	.134155	5.16	3605	1
27	833	2	2.2642	13	3605	1
28	944	1	.125941	4.39	3200	1
29	944	2	.996578	8.14	3200	1
30	944	3	2.35729	10.89	3200	1

# 예시: 선형회귀 모형

24

```
. regress weight age age2
```

Source	SS	df	MS	Number of obs	=	198
Model	1523.59303	2	761.796513	F(2, 195)	=	527.26
Residual	281.738331	195	1.44481195	Prob > F	=	0.0000
Total	1805.33136	197	9.16411856	R-squared	=	0.8439
				Adj R-squared	=	0.8423
				Root MSE	=	1.202

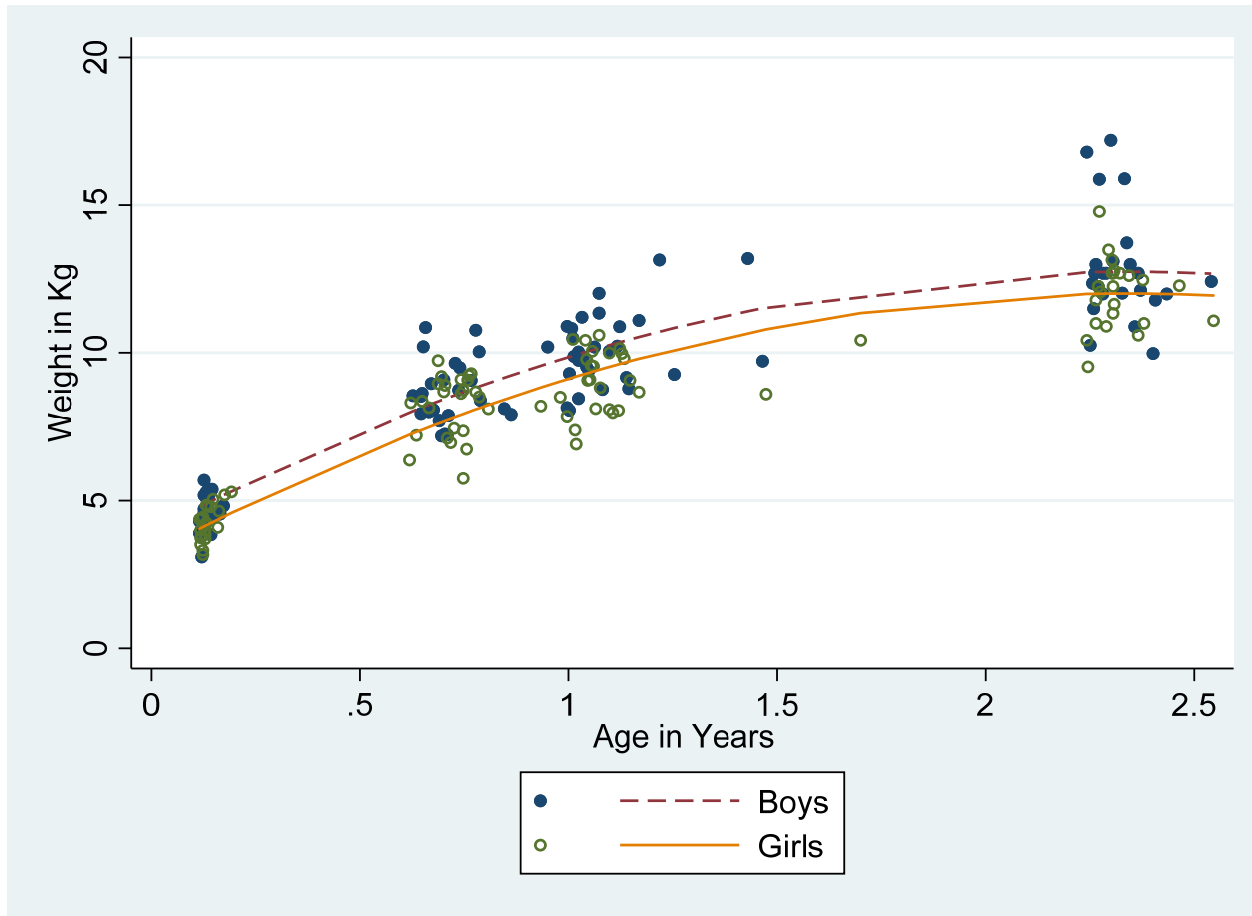
  

weight	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	7.503384	.430679	17.42	0.000	6.653997	8.352771
age2	-1.600168	.161087	-9.93	0.000	-1.917864	-1.282471
_cons	3.583899	.2183076	16.42	0.000	3.153352	4.014447



# 예시: 선형회귀 모형

25



# 예시: 임의절편 모형

26

```
. xtmixed weight age age2 || id:, mle nolog
```

Mixed-effects ML regression  
Group variable: id

Number of obs = 198  
Number of groups = 68

Obs per group:  
min = 1  
avg = 2.9  
max = 5

Log likelihood = -276.83266

Wald chi2(2) = 2623.63  
Prob > chi2 = 0.0000

weight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	7.817918	.2896529	26.99	0.000	7.250209	8.385627
age2	-1.705599	.1085984	-15.71	0.000	-1.918448	-1.49275
_cons	3.432859	.1810702	18.96	0.000	3.077968	3.78775

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Identity				
sd(_cons)	.9182257	.0973788	.7458966	1.130369
sd(Residual)	.7347063	.0452564	.6511507	.8289837

LR test vs. linear model: chibar2(01) = 78.07

Prob >= chibar2 = 0.0000

# 예시: 임의계수 모형

27

```
. xtmixed weight age age2 || id: age age2, mle nolog
```

Mixed-effects ML regression  
Group variable: id

Number of obs = 198  
Number of groups = 68

Obs per group:

min = 1  
avg = 2.9  
max = 5

Log likelihood = -258.28823

Wald chi2(2) = 1908.80  
Prob > chi2 = 0.0000

weight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	7.703184	.2328483	33.08	0.000	7.24681	8.159559
age2	-1.657892	.087368	-18.98	0.000	-1.82913	-1.486654
_cons	3.492399	.1410613	24.76	0.000	3.215924	3.768874

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(age)	.5049391	.1035995	.3377505	.7548871
sd(age2)	.1024006	.0777392	.0231259	.4534265
sd(_cons)	.7256802	.101173	.5521693	.9537141
sd(Residual)	.5420471	.0456327	.4595978	.6392875

LR test vs. linear model: chi2(3) = 115.16

Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

## 예시: 성장곡선분석 (모형)

28

- 성장곡선분석 (22쪽과 비교)
  - 다층/혼합효과 ‘기본’ 모형: 분리형

$$weight_{it} = \alpha_i + \beta_{1i}age_{it} + \beta_{2i}age_{it}^2 + \varepsilon_{it} \quad (\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2))$$

$$\alpha_i = \gamma_0 + \gamma_1girl_i + u_{1i} \quad (u_{1i} \sim N(0, \sigma_{u1}^2))$$

$$\beta_{1i} = \delta_0 + u_{2i} \quad (u_{2i} \sim N(0, \sigma_{u2}^2))$$

$$\beta_{2i} = \tau_0 + u_{3i} \quad (u_{3i} \sim N(0, \sigma_{u3}^2))$$

- 다층/혼합효과 ‘기본’ 모형: 결합형

$$y_{it} = (\gamma_0 + \gamma_1girl_i + u_{1i}) + (\delta_0 + u_{2i})age_{it} + (\tau_0 + u_{3i})age_{it}^2 + \varepsilon_{it}$$

$$= (\gamma_0 + \gamma_1girl_i + \delta_0age_{it} + \tau_0age_{it}^2) + (u_{1i} + u_{2i}age_{it} + u_{3i}age_{it}^2 + \varepsilon_{it})$$

‘고정요소’,

‘임의요소’,

# 예시: 성장곡선분석 (추정)

29

```
. xtmixed weight age age2 i.girl || id: age age2, mle nolog
```

Mixed-effects ML regression  
Group variable: id

Number of obs = 198  
Number of groups = 68

Obs per group:  
min = 1  
avg = 2.9  
max = 5

Log likelihood = -253.71942

Wald chi2(3) = 1958.16  
Prob > chi2 = 0.0000

weight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	7.701928	.2316034	33.25	0.000	7.247993	8.155862
age2	-1.65784	.0872627	-19.00	0.000	-1.828871	-1.486808
girl						
Girls	-.6230483	.1990072	-3.13	0.002	-1.013095	-.2330014
_cons	3.804559	.1676813	22.69	0.000	3.475909	4.133208

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
id: Independent				
sd(age)	.4883836	.1028249	.323257	.7378604
sd(age2)	.1071308	.0735085	.0279163	.4111229
sd(_cons)	.6573844	.0988615	.4895651	.8827309
sd(Residual)	.5450756	.0457708	.4623598	.6425891

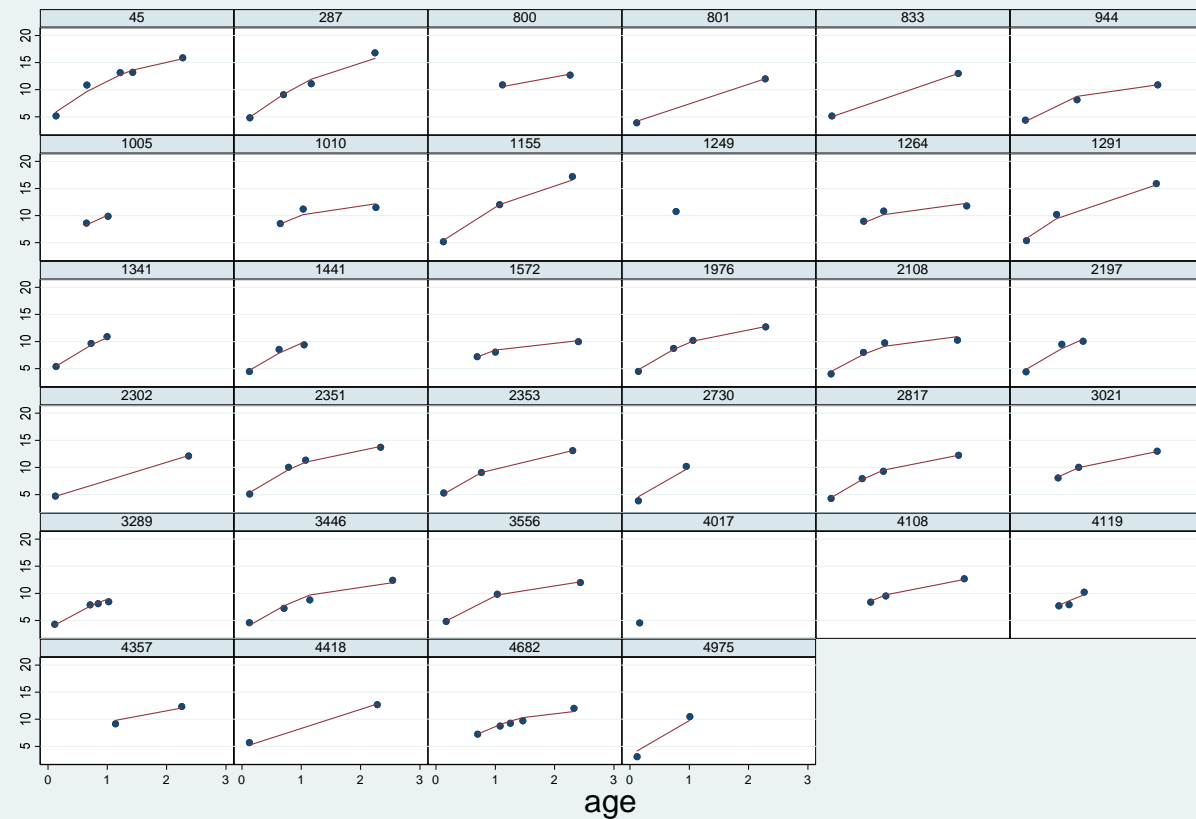
LR test vs. linear model: chi2(3) = 104.46

Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

# 예시: 성장곡선분석 (그래프 for 'Boys')

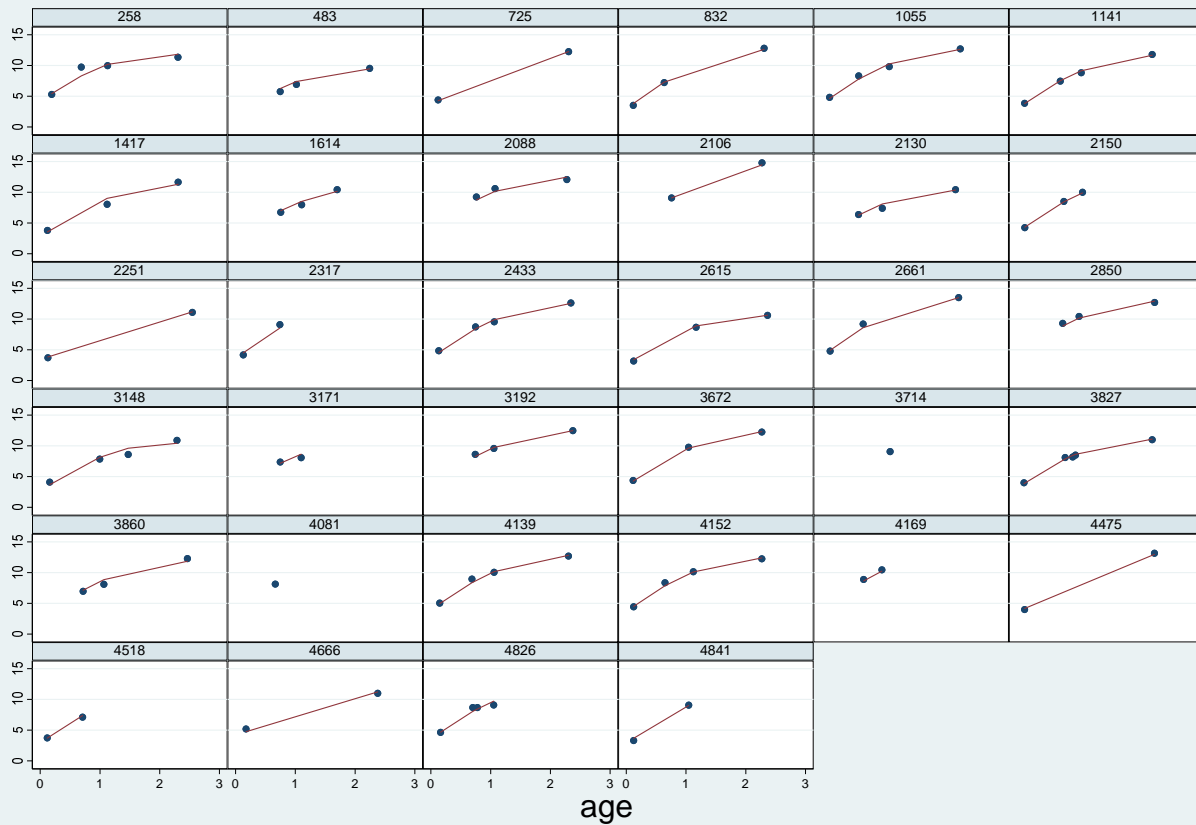
30



Graphs by id

# 예시: 성장곡선분석 (그래프 for 'Girls')

31



Graphs by id

## 참고문헌: 다층모형/성장곡선분석

32

### □ Books

- ▣ Rabe-Hesketh and Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata* (3<sup>rd</sup>).
- ▣ Raudenbush and Bryk. 2002. *Hierarchical Linear Models* (2<sup>nd</sup>).
- ▣ Gelman and Hill. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*.
- ▣ Kreft and Leeuw. (1998). *Introducing Multilevel Modeling*.
- ▣ Singer and Willett. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*.