

제13회 한국복지패널 학술대회  
특강자료

# 패널 다층모형 분석

2020. 9. 24.(목)

김진석  
(서울여자대학교 사회복지학과)



# 패널 다층모형 분석

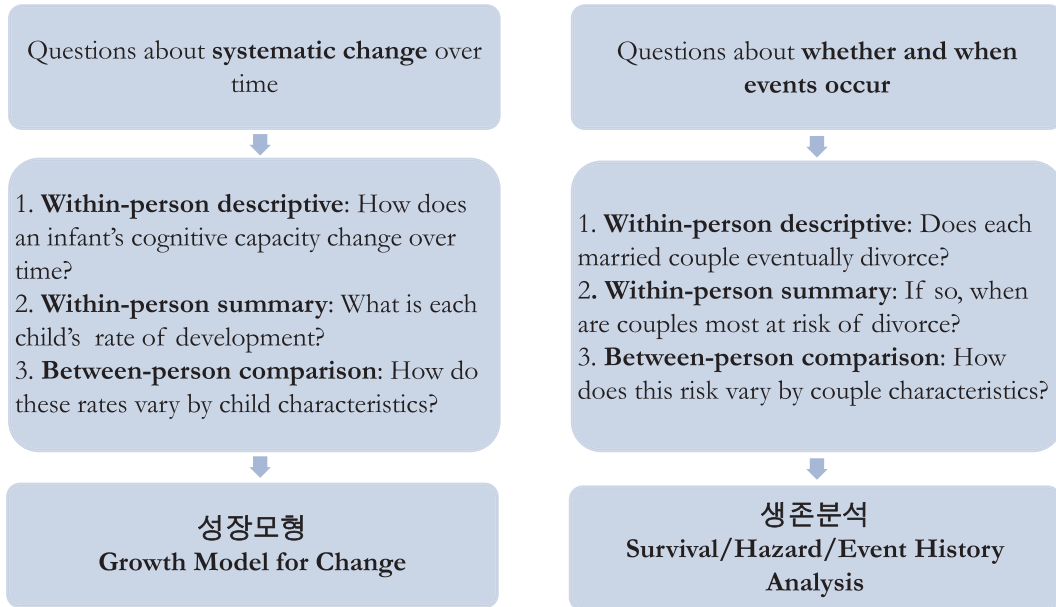
제13회 한국복지패널 학술대회

2020년 9월 24일

김진석 (서울여자대학교)

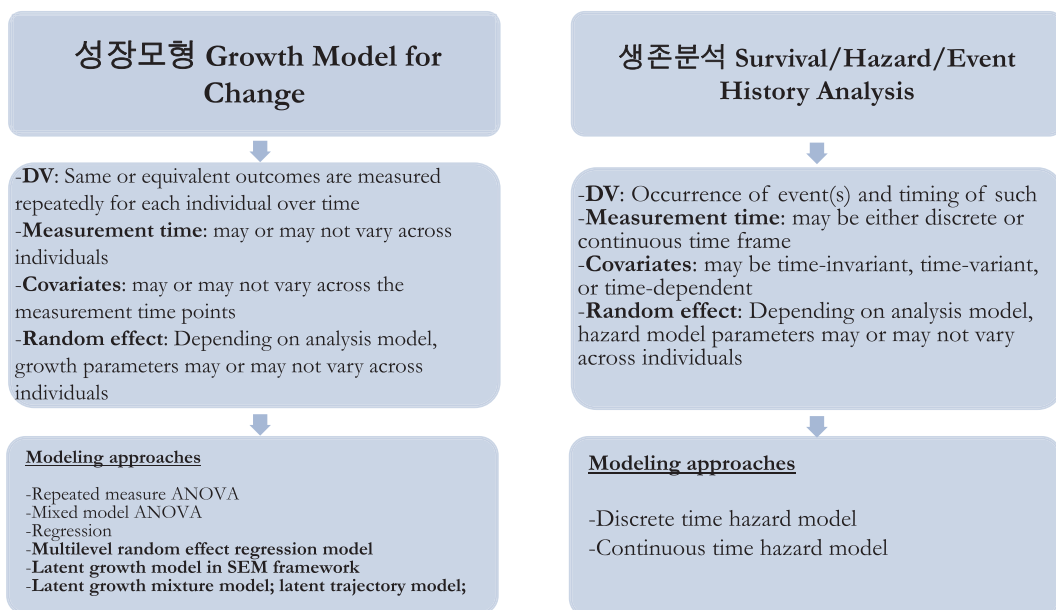
종단자료 및 패널자료에 대한 이해 - review

# Longitudinal data analysis(LDA) – overview



Adapted from Singer & Willet, (2006).

# Longitudinal data analysis(LDA) – overview (cont'd)

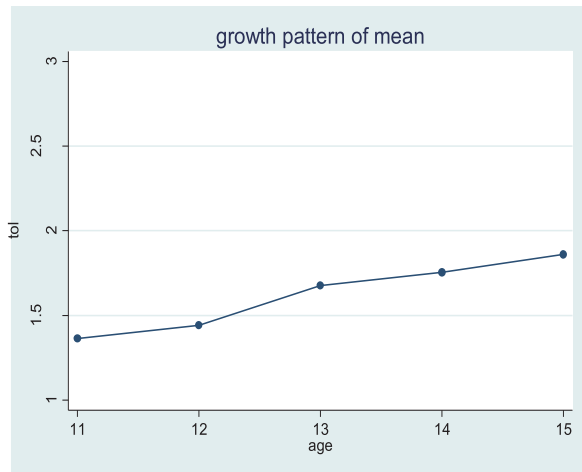
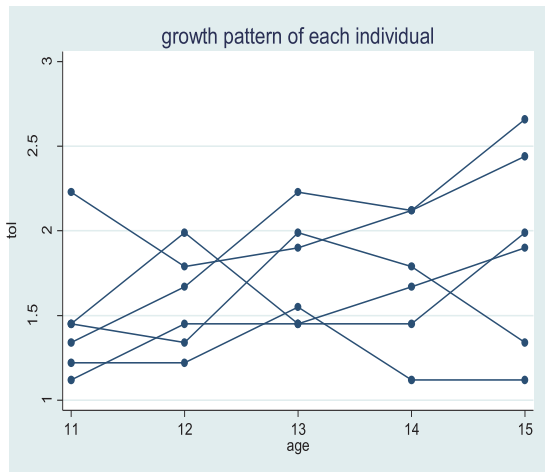


## 패널 자료 분석의 모형

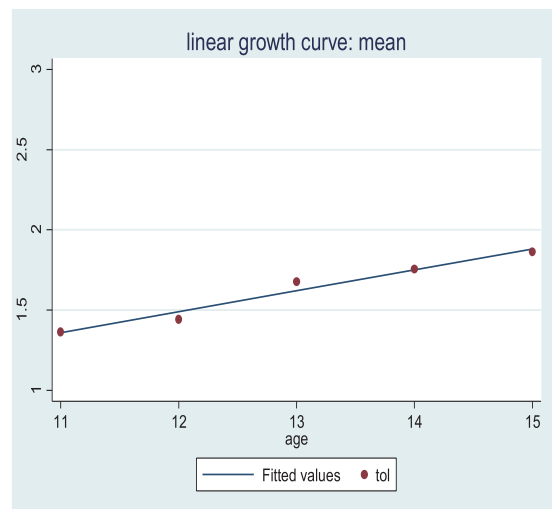
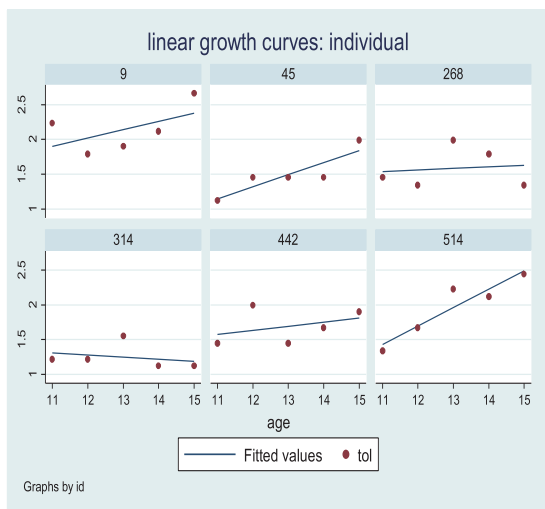
## 시간에 따른 변화에 대한 성장모형 – Example: NYS data

- NYS: National Youth Survey (Singer & Willett, Chap 2)
- DV: 이상행동에 대한 용인도 [tol]
  - [tol] score가 높을 수록, 이상행동에 대한 용인도가 높음
  - 11, 12, 13, 14, 15살에 각각 측정
- IVs:
  - 성별 [male]
  - 연구시작시 (age of 11) 이상행동에 대한 노출정도 [exposure]

# Descriptive growth pattern



# Growth pattern – linearly fitted



# 성장모형 - 다양한 접근법

(표본수 = n; 반복측정횟수 = T)

- 시간변수를 이용하여 (repeated) DV 에 대한 단일 선형회귀모형
  - 단일 회귀모형;  $n \times T$  관측치 observations
  - 모두에게 동일하게 적용되는 intercept와 slope
  - 모두가 같은 횟수의 측정을 할 필요 없음; 같은 시간에 측정할 필요 없음
  - Within-subject association과 individual difference에 대한 고려가 전혀 없음

```
. reg tol age if id!=0
```

Source	SS	df	MS			
Model	2.73790524	1	2.73790524	Number of obs =	80	
Residual	16.6693625	78	.213709776	F( 1, 78) =	12.81	
Total	19.4072677	79	.245661617	Prob > F =	0.0006	
				R-squared =	0.1411	
				Adj R-squared =	0.1301	
				Root MSE =	.46229	

tol	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1308125	.036547	3.58	0.001	.0580529	.2035721
_cons	-.0811874	.4779146	-0.17	0.866	-1.032642	.8702674



# 성장모형 - 다양한 접근법

(표본수=n; 반복측정횟수=T)

- 각 개인당 하나씩의 선형회귀 성장모형
  - (n regression)X(T observations)
  - 각 개인별로 서로 다른 절편과 기울기
  - 모두가 같은 횟수의 측정을 할 필요 없음; 같은 시간에 측정할 필요 없음
  - 개인별 차이에 대한 고려 가능; 개인간 차이에 대한 모형은 가능하지 않음
  - 선형곡선모형에 대한 기술적인 아이디어를 얻기에 좋은 방법

```
. bysort id: reg to1 age if id!=0
```

-> id = 9

Source	SS	df	MS
Model	.141610014	1	.141610014
Residual	.317390122	3	.105796707
Total	.459000135	4	.114750034

Number of obs =	5
F( 1, 3) =	1.34
Prob > F =	0.3311
R-squared =	0.3085
Adj R-squared =	0.0780
Root MSE =	.32526

to1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.119	.1028575	1.16	0.331	-.2083385 .4463386
_cons	.5929999	1.345037	0.44	0.689	-3.687507 4.873507

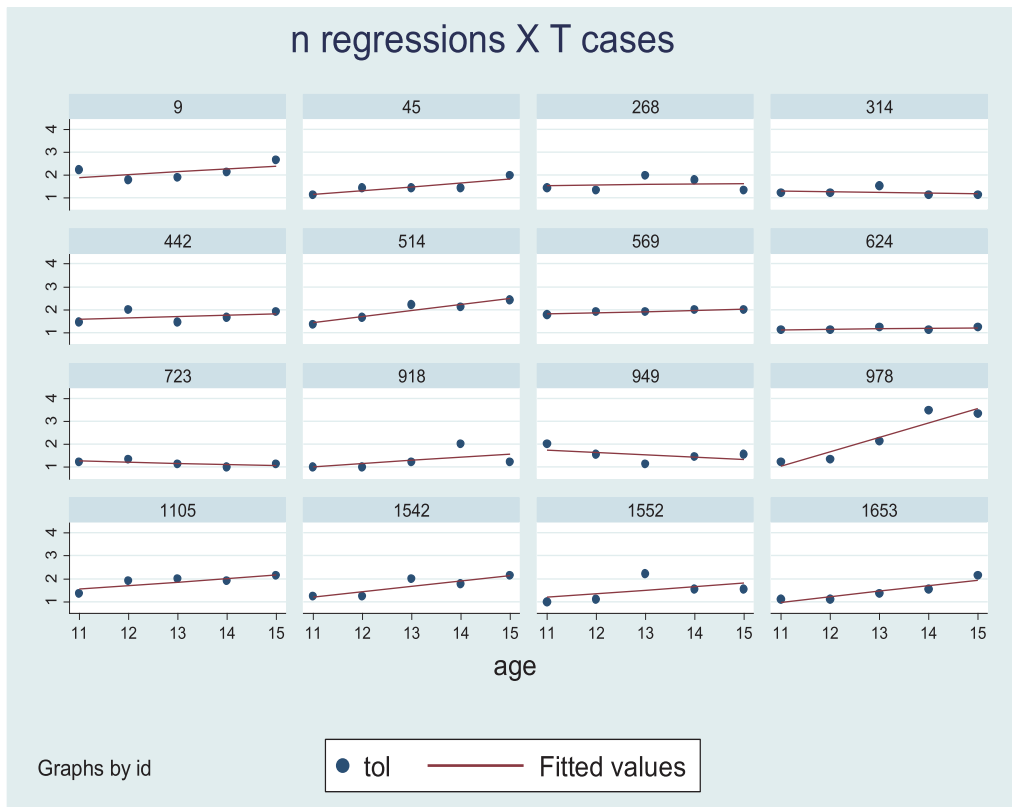
-> id = 45

Source	SS	df	MS
Model	.302760003	1	.302760003
Residual	.088919991	3	.029639997
Total	.391679994	4	.097919998

Number of obs =	5
F( 1, 3) =	10.21
Prob > F =	0.0495
R-squared =	0.7730
Adj R-squared =	0.6973
Root MSE =	.17216

to1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	.174	.0544426	3.20	0.049	.0007393 .3472607
_cons	-.77	.7119297	-1.08	0.359	-3.035678 1.495678

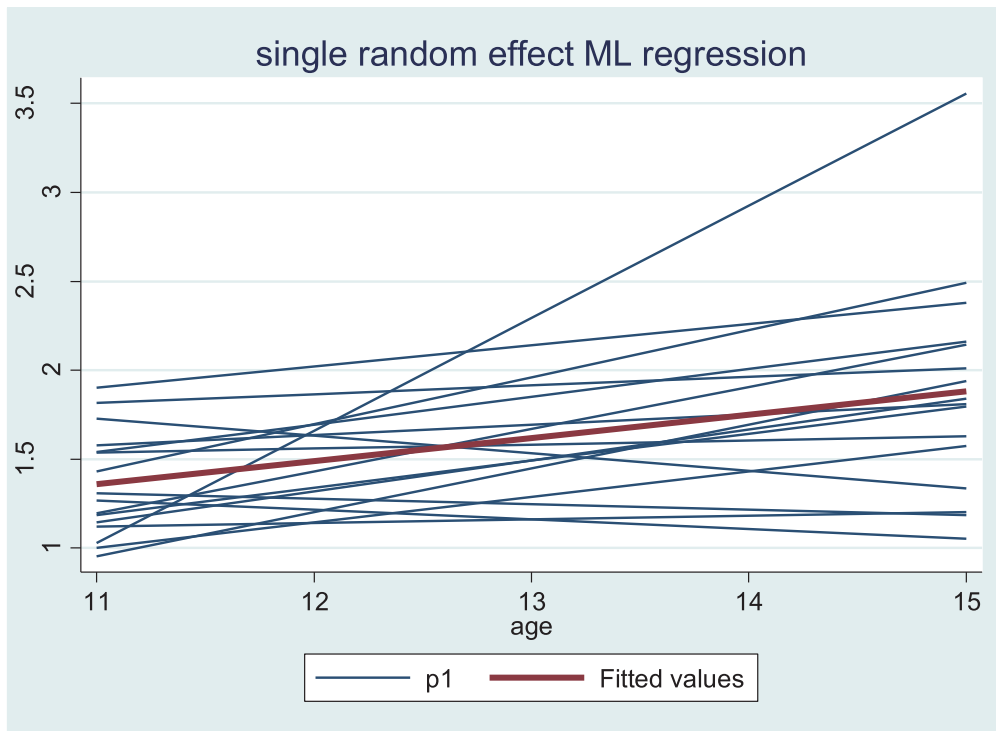




## 성장모형 - 다양한 접근법

(표본수 = n; 반복측정횟수 = T)

- 모두에게 적용되는 한 개의 무작위효과 MLM
  - 한 개의 MLM (or HLM);  $n \times T$  observations
  - 각 개인별로 서로 다른 절편과 기울기 적용가능
  - 모두가 같은 횟수의 측정을 할 필요 없음; 같은 시간에 측정할 필요 없음
  - 동일인에 대한 다른 시점의 관측치 사이의 상관관계에 대한 고려가 가능 within-subject association
  - 개인간 차이에 대한 고려와 모형화가 가능 between-subject modeling
  - MLM w/ 무작위 효과 접근법 혹은 latent variable (SEM) 접근법이 모두 가능



## 패널자료분석 모형 review

- Pooled ols

$$Y_{it} = \beta_0 + \beta X_{it} + \varepsilon_{it}$$

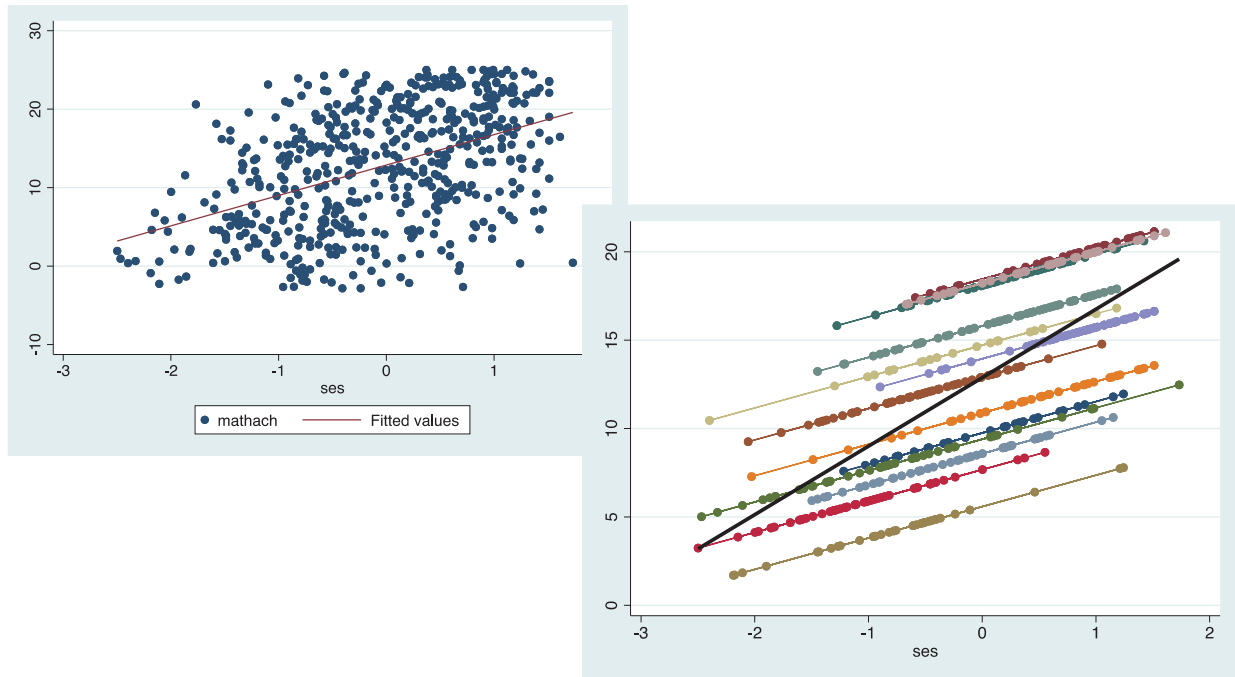
- 패널데이터에 대한 고려가 없음
- Underestimated SE, inflated Type-1 error

- Within individual model (fixed effect model)

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - X_{i.}) + \varepsilon_{ij} - \varepsilon_i.$$

- $\varepsilon_{it}$ : time-varying within individual error
- Time-invariant variable의 효과를 추정할 수 없음

# Pooled ols vs. FE



제 13회 한국복지패널 학술대회

패널 다층모형 분석

17

## 패널자료분석 모형 (계속)

- Between individual model

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_i + \zeta_i + \bar{\varepsilon}_i$$

– Time effect를 무시한 모형

- Random effect model

$$Y_{it} = \beta_0 + \zeta_i + \beta_1 X_{it} + \varepsilon_{ij}$$

$$\zeta_i | X_{it} \sim N(0, \psi), \quad \varepsilon_{ij} | X_{it}, \zeta_i \sim N(0, \theta)$$

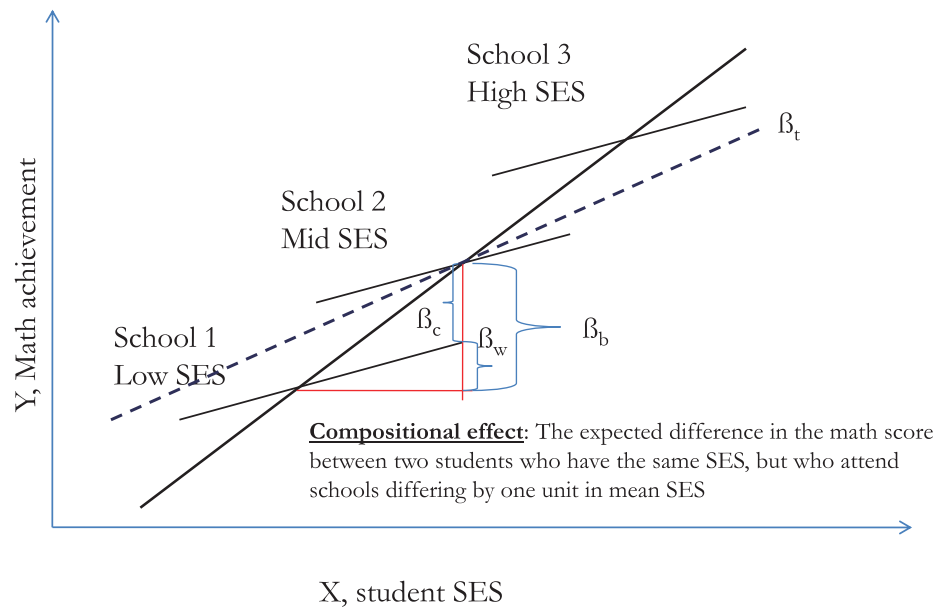
- Multilevel model (MLM); Hierarchical linear model (HLM)
- Random effects를 통해 개인간 차이에 대한 고려
- Within individual association를 통계적으로 통제; no type-I error inflation

제 13회 한국복지패널 학술대회

패널 다층모형 분석

18

# Compositional, between, within effects – an example



## FE vs. RE

$$Y_{it} = \beta_1 X_{it} + (u_i + e_{it}) = \beta_1 X_{it} + v_{it}$$

### Fixed-effects model

- Within estimator model
- Useful when unobserved individual effects may be related to predictor ( $\text{corr}(v_{it}, x_{it}) \neq 0$ )
- No estimation for TIC is available

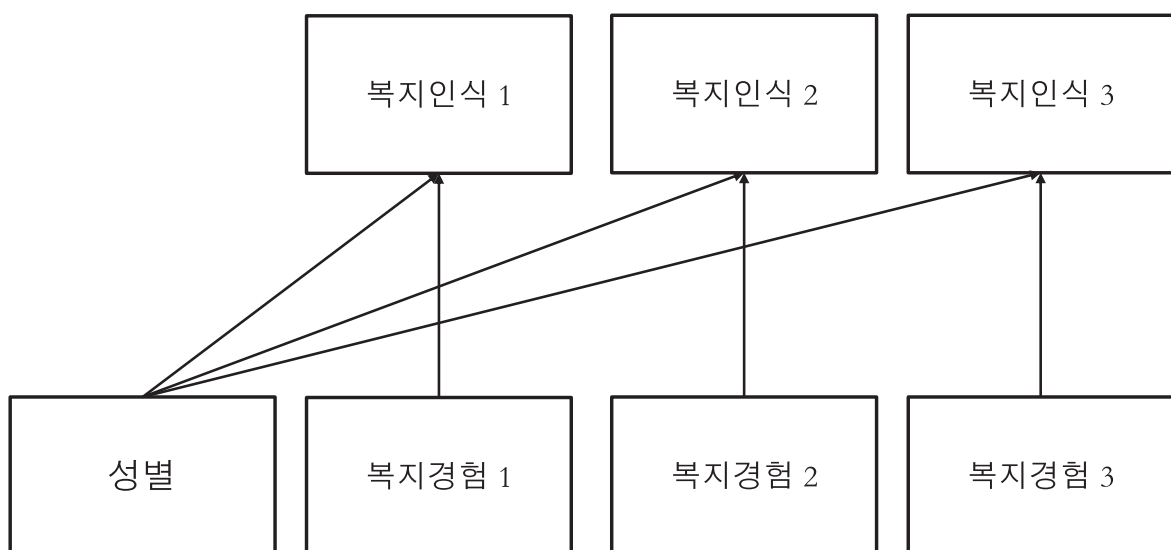
### Random-effects model

- Within + Between estimator model
- Useful when unobserved individual effects is assumed to be random and uncorrelated with the predictors
- The effect of TIC may be estimated

- **Hausman test ( $H_0: \text{corr}(u_i, x_{it}) = 0$ ): if significant, FE is the answer**
- **Likelihood ratio test (LRT) between FE and RE models in ML estimation**

## Panel data analysis within MLM framework

## 패널자료분석의 예: 한국복지패널



# Random intercept 성장곡선 모형- w/o covariate

Level-1 model —  $wperc_{ij} = \pi_{0i} + \pi_{1i} year_{ij} + \varepsilon_{ij}$   
 $Var(\pi_{0i}) = \sigma_0^2, Var(\pi_{1i}) = \sigma_1^2$

Level-2 model —  $\pi_{0i} = \gamma_{00} + \zeta_{0i}$   
 $\pi_{1i} = \gamma_{10} + \zeta_{1i}$

j=time point  
i=individual  
wperc=outcome  
year=time score  
 $\pi_0$ =intercept  $\pi_1$ =slope

$$wperc_{ij} = (\gamma_{00} + \zeta_{0i}) + (\gamma_{10} + \zeta_{1i})year_{ij} + \varepsilon_{ij}$$

$$= (\gamma_{00} + \gamma_{10} year_{ij}) + (\zeta_{0i} + \zeta_{1i})year_{ij} + \varepsilon_{ij}$$

Fixed effects: same for all workplace

Random effects: different for each workplace

Random intercept model: if  $Var(\pi_{0i}) = \sigma_0^2 \neq 0, Var(\pi_{1i}) = \sigma_1^2 = 0$

Random slope model: if  $Var(\pi_{0i}) = \sigma_0^2 = 0, Var(\pi_{1i}) = \sigma_1^2 \neq 0$

# Random intercept 성장곡선 모형- w/o covariate

```
. mixed welfcon02 year||h_pid:, nolog
```

```
Mixed-effects ML regression      Number of obs   =   13,433
Group variable: h_pid            Number of groups =    6,314

Obs per group:
    min =         1
    avg =         2.1
    max =         5

Wald chi2(1) =       36.66
Prob > chi2  =       0.0000

Log likelihood = -6466.748
```

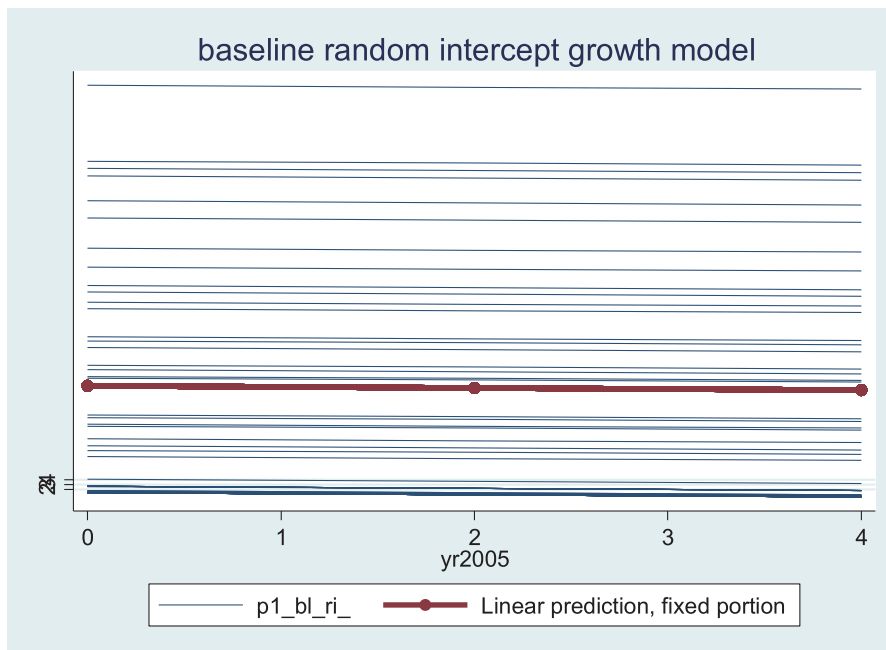
welfcon02	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
year	.0055325	.0009137	6.06	0.000	.0037417 .0073233
_cons	-8.125525	1.838684	-4.42	0.000	-11.72928 -4.521771

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
<b>h_pid: Identity</b>			
var(_cons)	.0432715	.0019292	.0396508 .0472228
var(Residual)	.1178329	.0018994	.1141682 .1216151

LR test vs. linear model:  $\text{chibar2}(01) = 710.12$  Prob >= chibar2 = 0.0000

- welfcon02=복지인식
- 무작위 절편모형의 무작위 효과에 대한 LRT 결과가 유의함
- 이는 시간에 따른 welfcon02의 시작점에서의 값이 h\_pid별로 달라짐을 의미

# Random intercept 성장곡선



## Random intercept and slope 성장모형 in stata

```
. mixed welfcon02 year||h_pid: year, nolog
```

Mixed-effects ML regression  
Group variable: h\_pid

Number of obs = 13,433  
Number of groups = 6,314

Obs per group:  
min = 1  
avg = 2.1  
max = 5

Log likelihood = -6466.1867

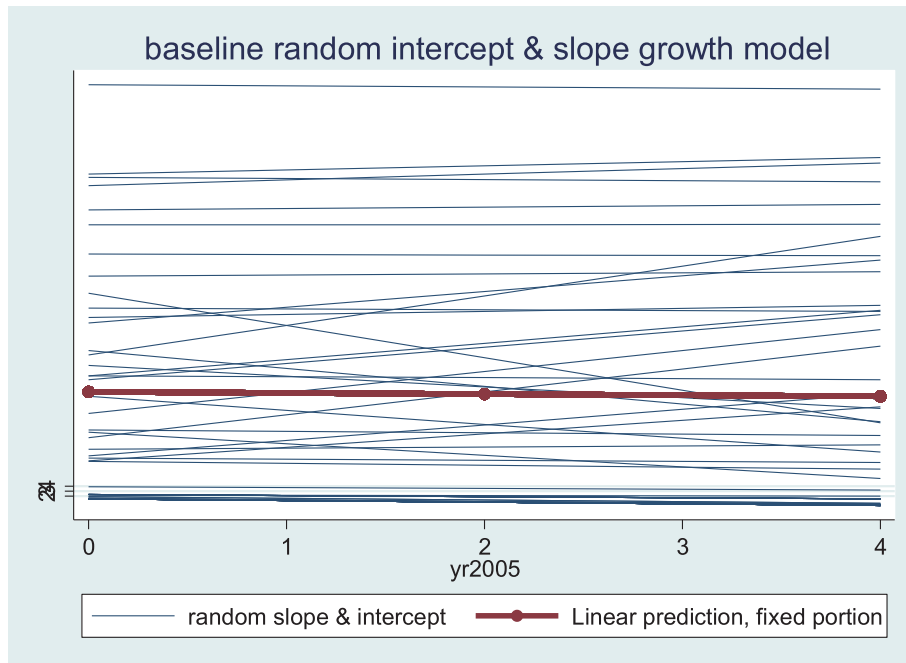
Wald chi2(1) = 36.55  
Prob > chi2 = 0.0000

welfcon02	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
year	.0055226	.0009135	6.05	0.000	.0037322	.0073131
_cons	-8.105771	1.838361	-4.41	0.000	-11.70889	-4.502649

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
<b>h_pid: Independent</b>				
var(year)	1.07e-08	4.77e-10	9.80e-09	1.17e-08
var(_cons)	2.83e-08	1.75e-08	8.45e-09	9.50e-08
var(Residual)	.1178047	.0019025	.1141343	.1215932

LR test vs. linear model: chi2(2) = 720.24 [Screenshot](#) Prob > chi2 = 0.0000

# Random intercept and slope 성장모형



제13회 한국복지패널 학술대회

패널 다층모형 분석

27

## Random intercept 성장모형 w/ TIC in stata

```
. mixed welfcon02 year female||h_pid:, nolog
```

```
Mixed-effects ML regression      Number of obs   =   13,433
Group variable: h_pid            Number of groups =    6,314

Obs per group:
    min =         1
    avg =         2.1
    max =         5

Wald chi2(2)   =   44.86
Prob > chi2    =   0.0000

Log likelihood = -6462.6802
```

welfcon02	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
year	.0055707	.0009136	6.10	0.000	.0037802	.0073613
female	-.0232408	.0081455	-2.85	0.004	-.0392057	-.0072758
_cons	-8.189454	1.838345	-4.45	0.000	-11.79254	-4.586365

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
<b>h_pid: Identity</b>				
var(_cons)	.0431502	.0019268	.0395343	.0470968
var(Residual)	.1178252	.0018991	.1141612	.1216068

```
LR test vs. linear model: chibar2(01) = 716.35      Prob >= chibar2 = 0.0000
```

제13회 한국복지패널 학술대회

패널 다층모형 분석

28



Random intercept  
성장모형 w/ TVC in  
stata

```
. mixed welfcon02 year welfever female||h_pid:, nolog
Mixed-effects ML regression           Number of obs   =   13,433
Group variable: h_pid                 Number of groups =    6,314

Obs per group:
      min =          1
      avg =         2.1
      max =          5

Wald chi2(3)   =    51.61
Prob > chi2    =    0.0000

Log likelihood = -6459.3135
```

welfcon02	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
year	.0056118	.0009135	6.14	0.000	.0038215	.0074021
welfever	.0407326	.0156952	2.60	0.009	.0099705	.0714947
female	-.0243574	.0081551	-2.99	0.003	-.040341	-.0083738
_cons	-8.274207	1.838169	-4.50	0.000	-11.87695	-4.671462

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
h_pid: Identity				
var(_cons)	.0431361	.0019256	.0395225	.0470802
var(Residual)	.1177611	.0018979	.1140993	.1215403

LR test vs. linear model: **chibar2(01) = 716.97**      Prob >= chibar2 = **0.0000**

패널자료에 대한 잠재변수모형적 접근 -  
잠재성장모형 Latent Growth Model(LGM)

# 다시 ML 성장모형 – 잠재변수 모형에서 재조명

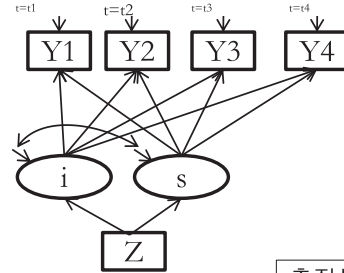
## • ML틀에서의 성장모형

$$Y_{it} = \pi_{0i} + \pi_{1i}X_{it} + \varepsilon_{it}$$

$$\pi_{0i} = \gamma_{00} + \gamma_{01}Z_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}Z_i + \zeta_{1i}$$

## • 잠재변수 틀에서의 성장모형



$$Y_{ti} = W_t i_i + X_t s_i + \varepsilon_{ti}$$

$$i_i = \gamma_{00} + \gamma_{01}Z_i + \zeta_{0i}$$

$$s_i = \gamma_{10} + \gamma_{11}Z_i + \zeta_{1i}$$

측정부분:  
- 제약: 잠재변수  $i$ 의 계수는 1로 고정  
- ML성장모형의 1레벨 모형과 비교

구조부분:  
ML성장모형에서 2레벨 모형과 비교

# 성장모형의 비교: ML vs. LGM

## • ML framework

- $X_{ti}$  가 변수의 기능
  - 유연성이 있으며 개인에 따라 변할 수 있음
- TVC의 경사는 개인에 따라 변할 수 있음

## • SEM framework

- $x_t$  가 매개변수/모수로 기능
  - 성장함수가 유연하게 정의됨
- TVC의 경사가 시간지점에 따라 변할 수 있음 (we will see this later)

이러한 점을 염두에 두고 보면, 위의 두 접근법은 서로 동등함!!

$$Y_{ti} = \pi_{0i} + \pi_{1i}X_{ti} + \pi_{2i}TVC_{ti} + \varepsilon_{ti}$$

$$\pi_{0i} = \gamma_{00} + \gamma_{01}Z_i + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}Z_i + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}Z_i + \zeta_{2i}$$

$$Y_{ti} = W_t i_i + X_t s_i + \varepsilon_{ti}$$

$$i_i = \gamma_{00} + \gamma_{01}Z_i + \zeta_{0i}$$

$$s_i = \gamma_{10} + \gamma_{11}Z_i + \zeta_{1i}$$

## 성장모형의 비교: ML vs. LGM

- ML framework

- Time은 변수

- 예: i-th 응답자의 chores

- 1<sup>st</sup> wave:

$$\text{chores}_{1i} = \pi_{0i} + \pi_{1i}X_{1i} + \varepsilon_{1i}, (X_{1i} = 0)$$

- 2<sup>nd</sup> wave:

$$\text{chores}_{2i} = \pi_{0i} + \pi_{1i}X_{2i} + \varepsilon_{2i}, (X_{2i} = 1)$$

- 3<sup>rd</sup> wave:

$$\text{chores}_{3i} = \pi_{0i} + \pi_{1i}X_{3i} + \varepsilon_{3i}, (X_{3i} = 2)$$

- SEM framework

- Time은 parameter

- 예: i-th 응답자의 chores

- 1<sup>st</sup> wave :

$$\text{chores}_{1i} = w_1i_i + x_1s_i + \varepsilon_{1i}, (\text{set } x_1 \equiv 0, w_1 \equiv 1)$$

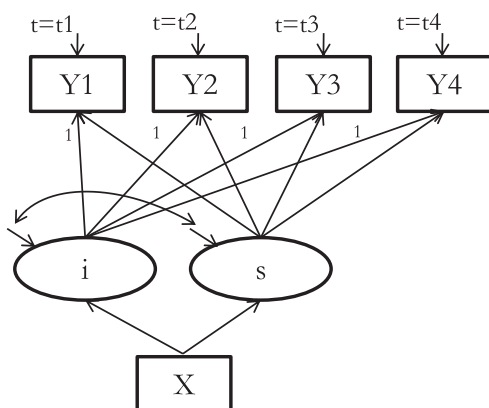
- 2<sup>nd</sup> wave :

$$\text{chores}_{2i} = w_2i_i + x_2s_i + \varepsilon_{2i}, (\text{set } x_2 \equiv 1, w_2 \equiv 1)$$

- 3<sup>rd</sup> wave :

$$\text{chores}_{3i} = w_3i_i + x_3s_i + \varepsilon_{3i}, (\text{set } x_3 \equiv 2, w_3 \equiv 1)$$

## 구조방정식 틀에서 선형성장모형



- 선형 성장모형을 기술하기 위해서는 두 개의 잠재변수 (i & s) 가 필요

- 2차 성장곡선 quadratic growth model 을 위해서는 세 개 (i, s, & q) 필요함

- Model estimation 의 초점은 [i] 와 [s]의 기대값/평균값 과 분산

- Mean structure model estimate

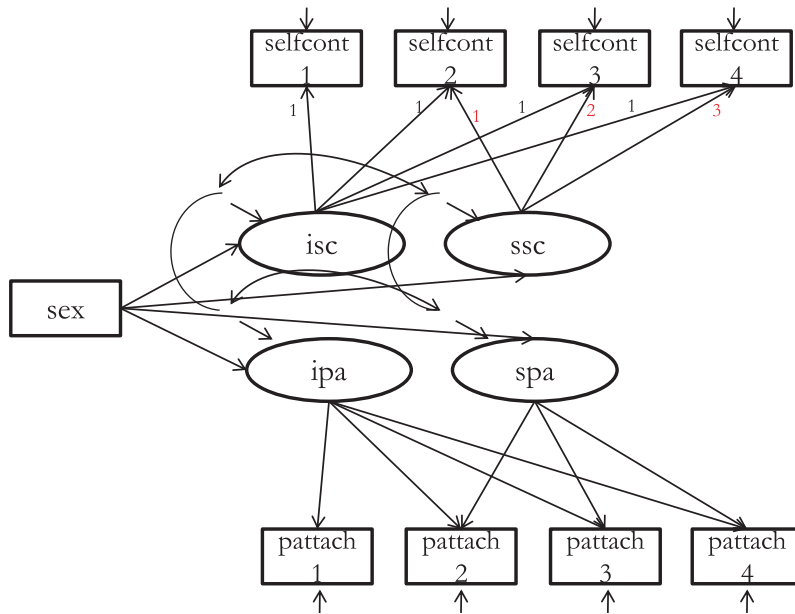
## MLM vs. LGM

## 성장/변화에 대한 LGM 접근의 강/약점

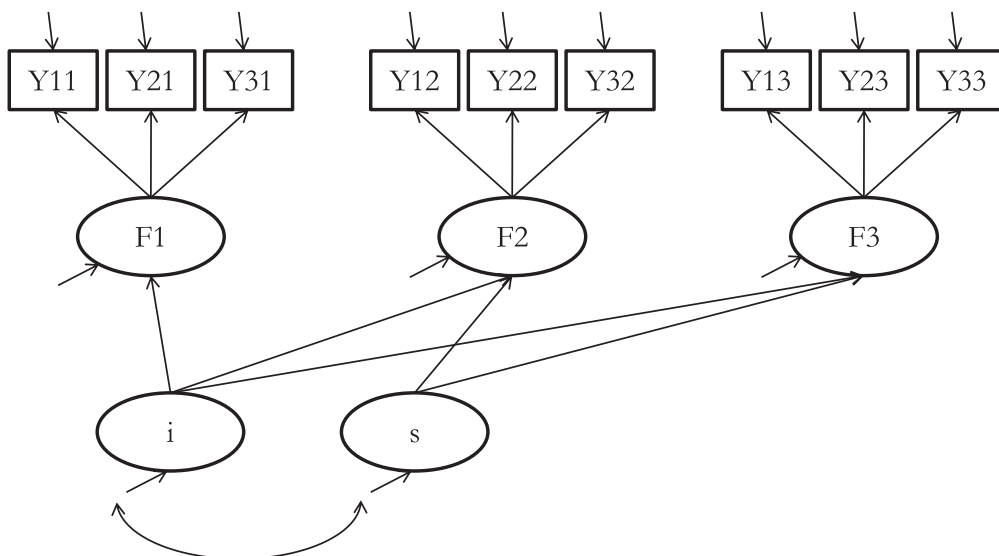
- 유연한 성장곡선 모형
- 개인별로 다른 측정시간의 처리가 용이
- 다중 집단에 대한 모형이 가능
- 랜덤효과 사이의 회귀분석 모형의 설정이 용이
- 다중 변화과정의 모형이 가능
- 복수 지표사용이 가능
- 성장모형이 다른 더 큰 모형의 일부로 작용할 수 있음
- 연속적이지 않은 잠재변수의 활용이 가능: growth mixtures

From Muthen & Muthen, (2008), Mplus short courses Topic 3

# 다중변화과정 multiple process growth curve 모형

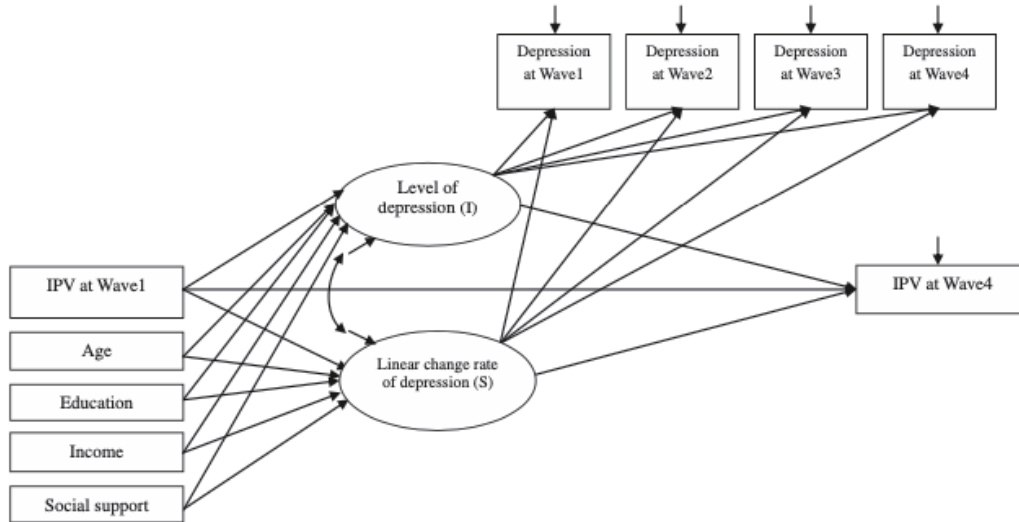


# 복수지표성장곡선 모형 multiple indicator growth curve model



# 모형의 일부로서의 성장모형 LGM as a part of the overall model

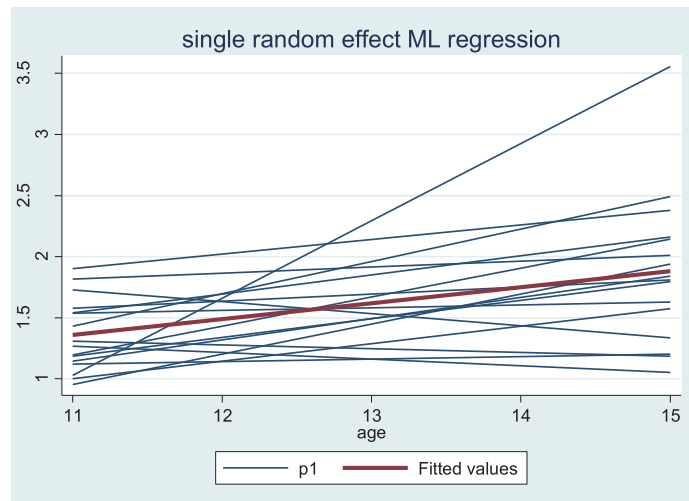
*J. Kim, J. Lee / Social Science & Medicine 99 (2013) 42–48*



Note: Social support, income, age, education level were measured at Wave 1

Beyond “regular” approach

# Latent growth mixture model



## Latent growth mixture model (cont'd)

- 사람중심의 접근법 Person-oriented approach
- 잠재궤적모형 latent trajectory analysis (LTA)/잠재계층성장모형 latent class growth analysis (LCGA)/ 잠재성장혼합모형 latent growth mixture model (GMM) approach
  - 이전 slide에서 언급된 접근법과 완전히 다른 가정에서 출발: group based, finite mixture model
    - 동일하지 않은 모집단이 혼재되어 있다고 가정
  - 각각 다른 양상을 보이는 subgroup or subpopulation (i.e., class)을 구분하는 것이 주된 연구과제
  - 각각의 class는 서로 다른 growth parameter를 가질 수 있음

