

2012년 한국아동·청소년패널2010 데이터분석방법론세미나 자료집

일시 | 2012년 7월 5일(목) 10:00~18:00

장소 | 한국교원단체총연합회관 2층 다산홀

주최 | 한국청소년정책연구원



세 부 일 정

「2012년 한국아동·청소년패널2010 데이터분석방법론 세미나」

사회 : 이경상(한국청소년정책연구원 선임연구위원)

시 간	내 용	비 고
09:30~10:00	등록	
10:00~10:05	개회 / 오전 강의 선생님 소개	이경상(사회자)
10:05~10:10	환영인사	이재연(한국청소년정책연구원장)
10:10~13:00	오전 강의	강상진 교수
13:00~15:00	점심식사·휴식	
15:00~15:05	오후 강의 선생님 소개	이경상(사회자)
15:05~18:00	오후 강의	이기종 교수
18:00	폐회	이경상(사회자)

목 차

■ 종단자료 분석을 위한 다층모형	1
▶ 강상진(연세대학교 교육학과 교수)	
■ 잠재성장모형	43
▶ 이기종(국민대학교 교육학과 교수)	

종단자료 분석을 위한 다층모형

강 상 진

(연세대학교 교육학과 교수)

종단자료 분석을 위한 다층모형 Multilevel Model for Longitudinal Data Analysis

강상진
연세대학교

한국 청소년 정책연구원
패널데이터 분석방법 세미나
2012. 7. 5

목 차

Part 1. Cross-Sectional vs. Longitudinal Multilevel Data

Part 2. 패널 자료 분석의 방법론적 특성

Part 3. Centering in Growth Curve Modeling

Part 4. Multilevel Models for Panel Data Analysis

Part 5. Computer Lab

Part 6. Growth Modeling에서 반복측정 분산분석, SEM,
HLM 비교

Part .1

Cross-Sectional vs. Longitudinal Multilevel Data

A. Cross-Sectional Multilevel Data

1. students within schools

School 1	School 2	.	.	.	School J
student 1	student 1	.	.	.	student 1
student 2	student 2	.	.	.	student 2
.
.
.
student n_1	student n_2	.	.	.	student n_j

A. Cross-Sectional Multilevel Data

2. employees within branches

Branch 1	Branch 2	.	.	.	Branch J
employee 1	employee 1	.	.	.	employee 1
employee 2	employee 2	.	.	.	employee 2
.
.
.
employee n_1	employee n_2	.	.	.	employee n_J

5

A. Cross-Sectional Multilevel Data

3. residents within Neighbors (residents within districts of Seoul)

Neighbor 1	Neighbor 2	.	.	.	Neighbor J
resident 1	resident 1	.	.	.	resident 1
resident 2	resident 2	.	.	.	resident 2
.
.
.
resident n_1	resident n_2	.	.	.	resident n_J

6

B. Longitudinal Multilevel Data

1. Panel Data

repeated measures within persons (2-level data)

Person 1	Person 2	.	.	.	Person n
obs 1	obs 1	.	.	.	obs 1
obs 2	obs 2	.	.	.	obs 2
.
.
.
obs T ₁	obs T ₂	.	.	.	obs T _n

7

B. Longitudinal Multilevel Data

time-person-schools (3-level data)

sch 1			sch 2			.	sch J				
s ₁	s ₂	.	s _{n1}	s ₁	s ₂	.	s _{n2}	.	s ₁	.	s _{nJ}
obs1	obs1	.	obs1	obs1	obs1	.	obs1	.	obs1	.	obs1
obs2	obs2	.	obs2	obs2	obs2	.	obs2	.	obs2	.	obs2
.
.
obs T ₁₁	obs T ₂₁	.	obs T _{n1}	obs T ₁₂	obs T ₂₂	.	obs T _{n2}	.	obs T _{1J}	.	obs T _{nJ}

8

B. Longitudinal Multilevel Data

2. Cohort Data

repeated measures of cohorts

1970			1980			1990		
sch 1	-	sch J	sch 1	-	sch J	sch 1	-	sch J
s_1	-	s_1	s_1	-	s_1	s_1	-	s_1
.
.
.
s_{n1}	-	s_{nJ}	s_{n1}	-	s_{nJ}	s_{n1}	-	s_{nJ}

9

B. Longitudinal Multilevel Data

3. Trends Data

students with a quantity variable within schools

	中 1	中 2	中 3
sch 1	s_1	s_1	s_1
	.	.	.
	s_{n1}	s_{n1}	s_{n1}
sch 2	s_1	s_1	s_1
	.	.	.
	s_{n2}	s_{n2}	s_{n2}
:	:	:	:
sch J	.	.	.
	.	.	.
	s_{nJ}	s_{nJ}	s_{nJ}

10

C. Cross-Classified Multilevel Data

1. school x neighbor data (Cross-Sectional Data)

	neighbor 1	neighbor 2	...	neighbor k
sch 1	s_1	s_1		s_1
	:	:	...	:
	s_{n11}	s_{n12}		s_{n1k}
sch 2	s_1	s_1		s_1
	:	:	...	:
	s_{n21}	s_{n22}		s_{n2k}
:	:	:	:	:
sch J	s_1	s_1		s_1
	:	:	...	:
	s_{nJ1}	s_{nJ2}		s_{nJk}

11

C. Cross-Classified Multilevel Data

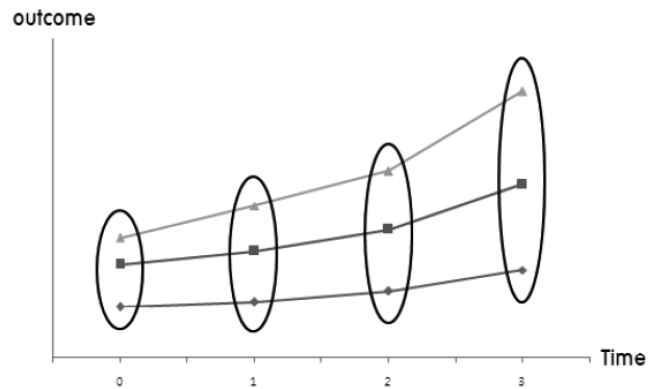
2. school x year data (Trends or Cohort Data)

	year 1	year 2	year 3	...	year k
sch 1	s_1	s_1	s_1		s_1
	:	:	:	...	:
	s_{n11}	s_{n12}	s_{n13}		s_{n1k}
sch 2	s_1	s_1	s_1		s_1
	:	:	:	...	:
	s_{n21}	s_{n22}	s_{n23}		s_{n2k}
:	:	:	:	:	:

12

D. Cross-sectional vs Longitudinal Analysis

1. underlying Latent changes are the same.



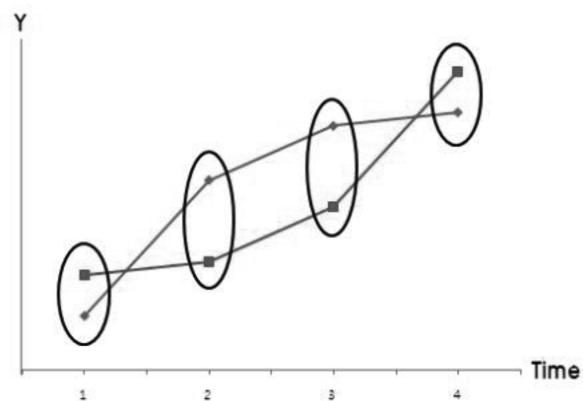
(모든 현상은 지속되고 있음

⇒ cross-sectional analysis is inherently limited.)

13

D. Cross-sectional vs Longitudinal Analysis

2. underlying Latent changes are not the same.



14

아동·청소년 관련 국내 패널 조사

명칭	한국노동패널 (KLIPS)	청년패널 (YP2001/YP2007)	한국 청소년패널 (KYPS)	한국교육·고용패널 (KEEP)	한국교육 중단연구 (KELS)	한국아동패널 (PSKC)	한국 아동·청소년패널 (NYPI)
조사 기간	1998년~	2001년~	중2:2003~ 초4:2004~	2004년~	2005년~	2008년~	2010년~
표집 단위 (표본 크기)	가구 (5,000 가구)	가구 (11,000 ~12,000 청년 가구원)	초4학생, 학부모 (2,500명) 중2학생, 학부모 (3000명)	중3학생 일반고3 실업고3 (각 2,000 명)	중1학생, 학부모, 교사, 학교장 (6,999명)	가구 (2,561가구 - 신생아, 부모)	초1학생 초4학생 중1학생 및 학부모 (각2200명)
관련 조사 변수	<ul style="list-style-type: none"> • 가구원 • 근로자 • 미취업자 	<ul style="list-style-type: none"> • 중·고등 학생 및 대학·대학원생 • 취업자 • 미취업자 	<ul style="list-style-type: none"> • 청소년 • 가구 	<ul style="list-style-type: none"> • 학생 • 학교, 행정가 및 담임교사 • 가구 	<ul style="list-style-type: none"> • 학생 • 가정 • 학교 • 교육활동 관련변수 	<ul style="list-style-type: none"> • 아동 • 부모 • 지역사회 • 육아활동 관련변수 	<ul style="list-style-type: none"> • 아동·청소년 발달 및 환경 관련 변수

15

Part .2

패널 자료 분석의 방법론적 특성

A. 패널 자료 분석의 방법론적 특성 예시

Data : 청소년 패널 자료 (중2 패널)

N=50명

Y=학생의 한달 평균 용돈

X=학년(중2~고2 ⇒ 0,1,2,3)

[raw data : obs within persons]

	N	평균	표준편차	최소값	최대값
학생의 한달 평균 용돈(Y)	200	44,990	29,023.468	2,000	150,000
학년(X)	200	1.5	1.121	0	3
학년 제공 값(X2)	200	3.5	3.509	0	9

17

A. 패널 자료 분석의 방법론적 특성 예시

□ 분석1 : $Y_t = \beta_0 + \beta_1 a_t + \beta_2 a_t^2 + e_t$

분산분석^b

모형	제곱합	자유도	평균제곱	F	유의확률
1 선형회귀분석	8.520E+10	2	4.260E+10	101.809	.000 ^a
잔차	8.243E+10	197	418429442		
합계	1.676E+11	199			

a. 예측값: (상수), X2, X

b. 종속변수: 학생의 한달 평균 용돈

계수^a

모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	22040,000	2819,603		7.817	.000
	X	5570,000	4528,029	.215	1,230	.220
	X2	4170,000	1446,426	.504	2,883	.004

a. 종속변수: 학생의 한달 평균 용돈

- 문제점 ?

18

A. 패널 자료 분석의 방법론적 특성 예시

■ 분석2 : times within persons (2 stage analysis)

$$Y_{ti} = \beta_{0i} + \beta_{1i}a_{ti} + \beta_{2i}a_{ti}^2 + e_{ti}, \quad e_{ti} \sim N(0, \sigma^2)$$

$$\hat{\beta}_{0i} = \gamma_{00} + r_{0i}$$

$$\hat{\beta}_{1i} = \gamma_{10} + r_{1i}$$

$$\hat{\beta}_{2i} = \gamma_{20} + r_{2i} \quad \cdot \quad [r_{0i}, r_{1i}, r_{2i}]^T \sim N(O, T)$$

19

A. 패널 자료 분석의 방법론적 특성 예시

■ Level-1 OLS regressions

Level-2 Unit	INTRCPT1	X slope	X2 slope
1	16750.00000	1750.00000	6250.00000
2	30500.00000	-14500.00000	12500.00000
3	30000.00000	10000.00000	0.00000
4	16000.00000	21000.00000	5000.00000
5	30750.00000	750.00000	1250.00000
6	50250.00000	10250.00000	-1250.00000
		.	
		.	
45	30500.00000	5500.00000	2500.00000
46	30500.00000	5500.00000	2500.00000
47	40500.00000	5500.00000	2500.00000
48	20250.00000	2750.00000	1250.00000
49	31500.00000	-8500.00000	12500.00000
50	22500.00000	17500.00000	2500.00000

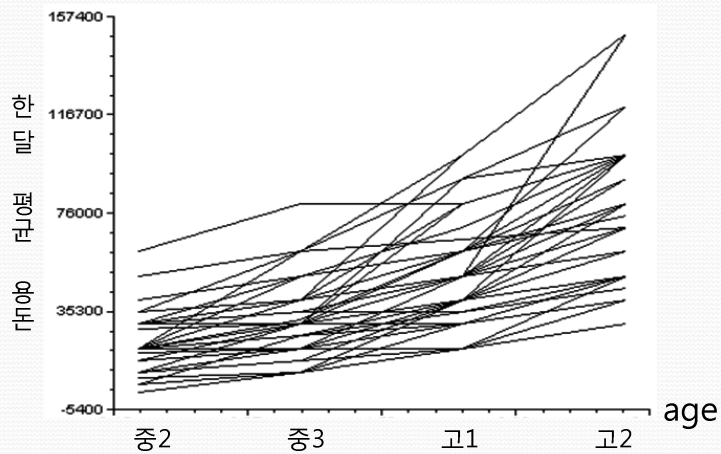
The average OLS level-1 coefficient for INTRCPT1 = 22040.00000

The average OLS level-1 coefficient for X = 5570.00000

The average OLS level-1 coefficient for X2 = 4170.00000

20

A. 패널 자료 분석의 방법론적 특성 예시



- 문제점 ?

21

B. Some Methodological Problems Addressed

Growth Curve Analysis

- Assess reliability of instruments for studying change
- Estimation of correlation of initial status and growth rate
- Study correlates of change
- Incorporate time-invariant and time-varying covariates
- improved estimation of individual growth

22

C. Problems of Growth Modeling

research topics

- 1) 학업성취도의 변화, 어휘 습득량의 변화
- 2) 정의적 특성의 변화 (동기; 효능감, 존중감, ..)
- 3) 상담횟수에 따른 내담자의 변화
- 4) 문제행동, 일탈행동의 변화
- 5) 개인의 성장발달과정
- 6) 처치, 프로그램, 정책의 효과
- 7) 계층, 집단간 격차 이해
- 8) 정신적 질병의 영향 요인(risk factor) 탐색

23

C. Problems of Growth Modeling

research questions

Q. 개인(집단)의 행동은 어떤 모습으로 변하는가? 변화의 곡선은 무엇인가?
어떤 패턴을 갖는가?

Q. 왜 성장곡선은 개인(집단) 간에 차이가 있는가?
(학습량의 차이, 학습속도, 프로그램의 효과...)

→ 이 질문은 개인(집단)의 성장곡선을 정확하게 확인하지 않으면
응답 불가능함

"For studying individual growth, change itself should be modeled." Collins & Sayer (2000).

24

D. 개인의 성장곡선을 분석하기 위한 통계모형의 조건

1. 개인 내적 변화 양상을 통계모형에 모수로 나타낼 수 있어야 함.
→ 개인별 성장모형의 설정.
2. 개인간에 어떻게 & 왜 성장모형에 차이가 나는가를 설명 가능해야 함. 개인별 성장곡선간의 비교를 허용해야 함.
→ 개인별 변화모수의 개인간 변산을 설명하는 통계모형의 설정

25

D. 개인의 성장곡선을 분석하기 위한 통계모형의 조건

3. 자료의 구조(다층구조)를 반영하여야 함.
 - ▣ 단층구조 VS 다층구조(2층, 3층 구조)
 - ▣ 일원 다층구조 (One-way nested structure) :
반복측정치 \subset 개인 \subset 집단
 - ▣ 이원 다층구조 (Two-way nested structure) :
반복측정치 \subset 개인 \subset 집단A X 집단B
(예: 집단A=중학교, 집단B=고등학교)

26

D. 개인의 성장곡선을 분석하기 위한 통계모형의 조건

4. 준거변수의 분포를 반영하여야 함.

- 연속변수: 정규분포
- 이분변수: 베르누이 분포, 이항분포
- 다분변수: 다항분포
- 서열변수: 누적 확률분포
- 빈도변수: 포아송 분포

27

E. General Model for Growth Curve Analysis

1. 수준-1 모형: 개인별 성장곡선의 정의

■ 일반성장모형

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + \pi_{2i}a_{it}^2 + \pi_{3i}a_{it}^3 + \dots + \pi_{pi}a_{it}^p + e_{it} \quad e_{it} \sim N(0, \sigma^2)$$

- 선형성장모형: $Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + e_{it} \quad e_{it} \sim N(0, \sigma^2)$

- 비선형모형: $Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + \pi_{2i}a_{it}^2 + e_{it} \quad e_{it} \sim N(0, \sigma^2)$

28

E. General Model for Growth Curve Analysis

▣ Time variant covariate 모형

개인 내적 변화에 영향을 주는 요인 (time variant variable) 추가 가능

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + \pi_{2i}a_{it}^2 + \pi_{3i}Z_{it} + e_{it} \quad e_{it} \sim N(0, \sigma^2)$$

예) : 개인의 사건체험(부모 이혼, 비행친구 사귀)

29

E. General Model for Growth Curve Analysis

2. 수준-2 모형: 변화 모수(change parameter)의 개인간 분포 설명

$$\pi_{0i} = \beta_{00} + \sum \beta_{0q} X_{qi} + r_{0i}$$

$$\pi_{pi} = \beta_{p0} + \sum \beta_{pq} X_{qi} + r_{pi} \quad \text{for } P=1, 2, \dots, P; q=1, 2, \dots, q$$

$$[\pi_{0i}, \pi_{1i}, \dots, \pi_{Pi}]^T \sim N\left(0, \begin{matrix} \rightarrow & \begin{bmatrix} \tau_{00}, \tau_{01}, \dots, \tau_{0P} \\ \tau_{10}, \tau_{11}, \dots, \tau_{1P} \\ \dots \\ \tau_{P0}, \tau_{P1}, \dots, \tau_{PP} \end{bmatrix} \end{matrix}\right)$$

주) 변화모수(예, π_{0i}, π_{pi})의 변산을 설명하는 개인차 변수 (X_{qi})가 모형에 포함 가능

30

E. General Model for Growth Curve Analysis

3. 수준-3 모형: 개인 특성 변수 효과의 집단간 변산 설명

$$\beta_{00j} = \gamma_{000} + \sum \gamma_{00s} W_{sj} + u_{00j}$$

$$\beta_{0qj} = \gamma_{0q0} + \sum \gamma_{0qs} W_{sj} + u_{0qj}$$

$$\beta_{p0j} = \gamma_{p00} + \sum \gamma_{p0s} W_{sj} + u_{p0j}$$

$$\beta_{pqj} = \gamma_{pq0} + \sum \gamma_{pqs} W_{sj} + u_{pqj}$$

$$\vec{U} \sim N(\vec{0}, T_{\beta}),$$

$$\vec{U} = [u_{00j}, u_{01j}, \dots, u_{0Qj}, u_{10j}, \dots, u_{p0j}, u_{11j}, \dots, u_{1Qj}, \dots, u_{pQj}]^T$$

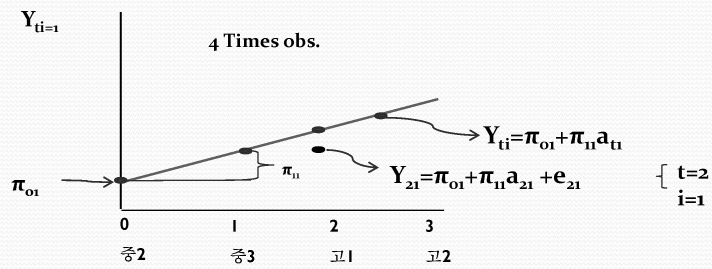
31

Part .3

Centering in Growth Curve Modeling

A. 종단모형의 명세화 방법과 Centering

- 1) Time effect in one person
 - No centering model



π_{01} = initial status of person 1.

status of person 1 at $a_{t1}=0$

π_{11} =Growth rate of person 1 over obs period.

Growth of person 1 for one unit change in time metric.

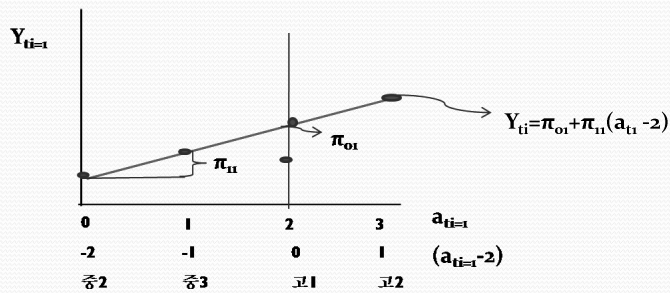
33

A. 종단모형의 명세화 방법과 Centering

- Centering at time = L

Time effect in one person

Model : $Y_{ti} = \pi_{0i} + \pi_{1i}(a_{ti} - 2) + e_{ti}$

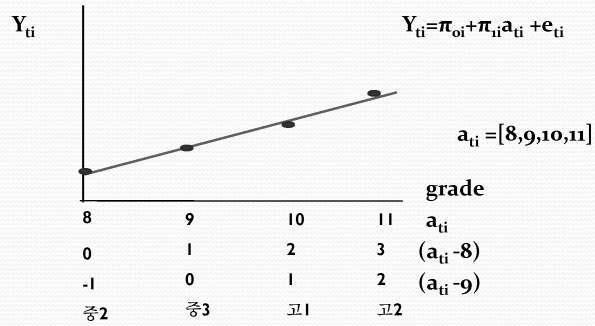


π_{01} 의 의미?

π_{11} 의 의미?

34

A. 종단모형의 명세화 방법과 Centering

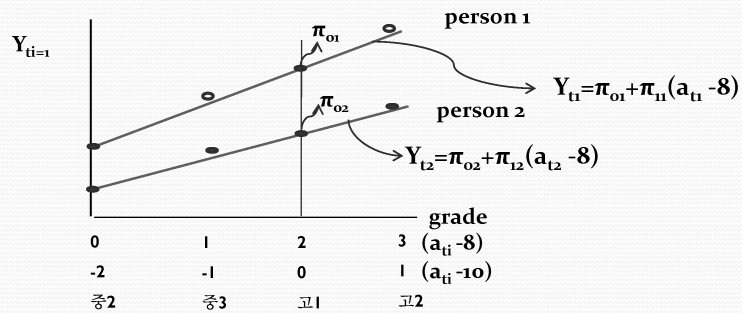


35

A. 종단모형의 명세화 방법과 Centering

1. 2) Time Effects for two persons

Model : $Y_{ti} = \pi_{0i} + \pi_{1i}(a_{ti} - 8) + e_{ti}$



Centering at Time Level (within person) Model

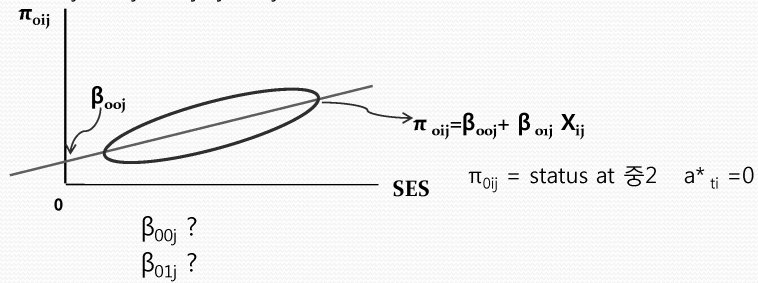
Options	Predictor (coding)	회귀계수의 의미	
		π_{0i}	π_{1i}
no centering	$a_{ti} = [8, 9, 10, 11]$	무의미 수준	성장률
centering at some meaningful time L	$(a_{ti}-L) = [8-L, 9-L, 10-L, 11-L]$	시점 L에서의 수준	성장률

36

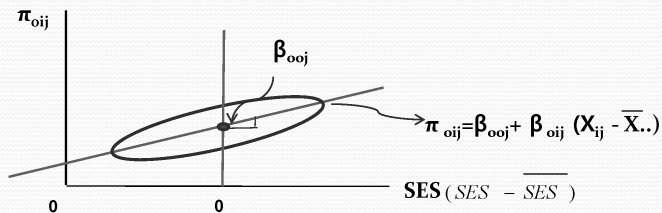
A. 종단모형의 명세화 방법과 Centering

2. 1) Person variable effects in one school

$$\text{Model : } \pi_{0ij} = \beta_{00j} + \beta_{01j} X_{ij} + r_{ij}$$



$$\text{Model : } \pi_{0ij} = \beta_{00j} + \beta_{01j} (X_{ij} - \bar{X}_{..}) + r_{ij}$$

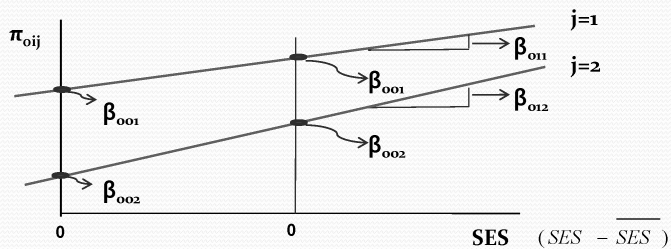


37

A. 종단모형의 명세화 방법과 Centering

2. 2) Two schools (excellence vs. equity)

$$\text{Model : } \pi_{01j} = \beta_{00j} + \beta_{01j} SES^*_{ij} + r_{01j}$$



* choice of comparative status

* choice of meaningful status

※ What variability do you want to explain?

$\beta_{001} > \beta_{002}$: Excellence

$\beta_{011} < \beta_{012}$: Equity

38

A. 종단모형의 명세화 방법과 Centering

Centering Options	Predictor	회귀계수의 의미	
		β_{00j}	β_{01j}
Group mean centering	$X_{ij} - \bar{X}_{.j}$	학교별 평균	변수 X의 선형효과
Grand mean centering	$X_{ij} - \bar{X}_{..}$	학교별 교정평균	The same
Centering at a constant	$X_{ij} - C$	학교별 교정평균 의 절편	The same
No Centering	X_{ij}	절편	The same

39

Part .4

Multilevel Models for Panel Data Analysis

A. Basic Design for Growth Curve Analysis

집단	학교	학생	측정시점			
			T1	T2	T3	T4
실험 집단	Sch 1	S11	-	-	-	-
		.	-	-	-	-
		Sn1	-	-	-	-

	Sch J ₁	S1J	-	-	-	-
		S2J	-	-	-	-
SnJ		-	-	-	-	
비교 집단	Sch 1	S11	-	-	-	-
		.	-	-	-	-
		Sn1	-	-	-	-

	Sch J ₂	S1J	-	-	-	-
		.	-	-	-	-
SnJ		-	-	-	-	

※ 학생들이 학교에 내재 됨(국내 모든 교육종단자료)
 ※ 각 수준(시간, 개인, 집단)에서 연구자의 가설검증 가능.
 시간 : 여름방학의 효과, Critical Event의 효과
 개인 : 남학생 vs. 여학생, 문제학생 vs. 일반학생, SES차이
 집단 : 도시 vs. 지방, 중학생 vs. 고등학생, 인문계 vs. 실업계

41

A. Basic Design for Growth Curve Analysis

<분석자료의 특징: 실험연구자료 vs. 비실험연구자료>

■ Experimental Study:

- Two Groups are Equivalent. (무선배치 적용)
- 통계적 통제 필수사항 아님.
- Complete data (모든 개체의 반복측정 횟수 동일)
- 결측치는 사례수에서 제외
- Fixed time interval (측정시점간 거리 동일)
- 연구자가 측정시점 통제.

분석방법 → Anova Model(Split-Plot Design),
 Multilevel Growth Modeling

42

A. Basic Design for Growth Curve Analysis

■ Non-Experimental Study(Quasi-Experimental Study):

- Two groups are not equivalent.(자연집단 조사)
- 통계적 통제 필수사항 임.
- Incomplete data
(개체별 반복측정 횟수는 다를 수 있음)
- 불균형 자료 분석 가능해야 함.
- Variable time interval (측정시점간 거리 다양)
- 연구자의 연구상황에 따른 측정시점 선택 가능.
- 조사대상자의 상황적 조건 반영 가능

분석방법 → Multilevel Growth Modeling

43

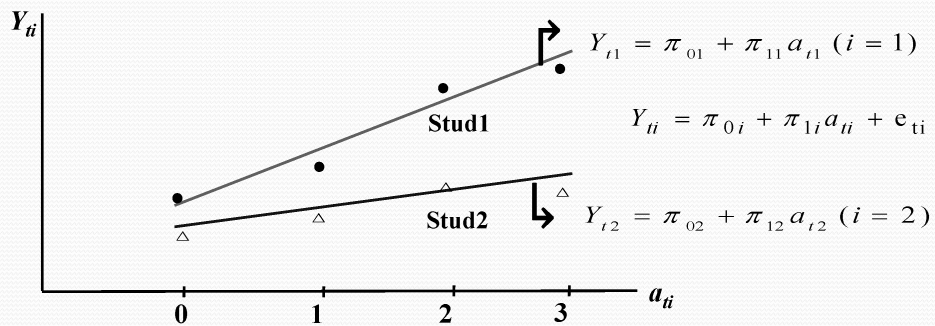
A. Basic Design for Growth Curve Analysis

■ 반복측정 분산분석 모형(R-ANOVA)의 문제점

- 1) 개인별 성장곡선을 추정하지 않음.
- 2) 실험연구(무선배치 적용) 자료 분석에 제한 됨.
유사실험연구자료, 조사연구자료의 분석에 부적절.
- 3) 관찰시점과 관찰횟수가 모든 개인에게 동일하여야 함.
→ 관찰시점이 개인간 다르거나, 관찰횟수가 다르면 결측치로
처리 또는 자료를 변조해야 함.
- 4) time variant 공변수의 도입이 불가능함.
- 5) 연속변수를 독립변수로 도입 불가 (K-1개의 가변수로 처리함)
- 6) 위계적 자료의 분석모형에서 오차항(Error term)의 결정 어려움.

44

B. Linear Growth Model



- ▣ Underlying principle is the linear function of time
- ▣ The latent change (linear) is the same for all subjects

45

B. Linear Growth Model

▣ 연구질문:

1. 모집단에서의 평균성장선은 무엇인가?
2. 개인별 성장선은 전체 평균성장선에서 어느 정도 변산을 갖는가?
개인차는 어느 정도인가?
3. 개인별 성장선은 어느 정도 정밀하게 추정되는가?
4. 개인별 출발점 위치와 성장률은 상관을 갖는가?
5. 왜 개인들의 출발점 위치와 성장률은 다른가?

46

B. Linear Growth Model

1. Two-level Linear Model

1) 기초모형(Unconditional Model): 개인별 성장곡선의 추정

$$\text{Level-1(time)} \quad Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + e_{it}, \quad e_{it} \sim N(0, \sigma^2)$$

$$\text{Level-2(person)} \quad \begin{aligned} \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} + r_{1i} \end{aligned}$$

$$\begin{bmatrix} r_{0i} \\ r_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{00} \end{bmatrix} \right)$$

◆ 통합모형(Combined Model)

$$\begin{aligned} Y_{it} &= (\beta_{00} + r_{0i}) + (\beta_{10} + r_{1i})a_{it} + e_{it} \\ &= \beta_{00} + \beta_{10}a_{it} + r_{0i} + r_{1i}a_{it} + e_{it} \end{aligned}$$

47

B. Linear Growth Model

1. Two-level Linear Model

■ Fixed Effects

β_{00} : Average Status at $a_{it} = 0$
If $a_{it} = (0, 1, 2, \dots)$, Average initial status.

β_{10} : Average of Linear growth rate over the observation period.

■ Random Effects

π_{1i} : $a_{it} = 0$ 에서 각 개인의 진점수(latent value)

π_{1i} : 관찰기간 동안 개인 i의 성장률

τ_{00} : 출발점 개인차의 분산

τ_{11} : 성장률의 개인차 분산

$\tau_{01} = \tau_{10}$: 출발점 위치와 성장률의 공분산

48

B. Linear Growth Model

1. Two-level Linear Model

▣ 평균성장 곡선

$$Y_{it} = \beta_{00} + \beta_{10} a_{it}$$

▣ 변화모수 추정치의 신뢰도:

$$rel(\pi_{0i}) = \frac{\tau_{00}}{\tau_{00} + \sigma^2 / T_i} \quad rel(\pi_{1i}) = \frac{\tau_{11}}{\tau_{11} + \sigma^2 / T_i}$$

- 신뢰도가 높으면, 개인별 해석 가능
- 신뢰도가 낮으면, 전체 성장곡선 중심으로 해석

49

B. Linear Growth Model

1. Two-level Linear Model

2) 연구모형(Conditional Model): 성장곡선의 개인차 설명

$$\text{Level-1(time)} \quad Y_{it} = \pi_{0i} + \pi_{1i} a_{it} + e_{it}, \quad e_{it} \sim N(0, \sigma^2)$$

$$\text{Level-2(person)} \quad \begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01}(\text{차차}) + \sum \beta_{0p} X_{pi} + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11}(\text{차차}) + \sum \beta_{1p} X_{pi} + r_{1i} \end{aligned}$$

$$\begin{bmatrix} r_{0i} \\ r_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

◆ 통합모형(Combined Model)

$$\begin{aligned} Y_{it} &= (\beta_{00} + \beta_{01}(\text{차차}) + \sum \beta_{0p} X_{pi}) \\ &+ (\beta_{10} + \beta_{11}(\text{차차}) + \sum \beta_{1p} X_{pi}) a_{it} \\ &+ [r_{0i} + r_{1i} a_{it}] + e_{it} \end{aligned}$$

50

B. Linear Growth Model

1. Two-level Linear Model

■ Fixed Effects

For π_{0i} (개인별 출발점 위치)

β_{00} : 비교집단의 출발점 교정평균

β_{01} : 처치집단과 비교집단의 출발점 공변수의 효과

β_{0p} : 개인별 출발점 위치에 대한 개인수준 공변수의 효과

For π_{1i} (개인별 선형성장률)

β_{10} : 비교집단의 교정된 선형성장률(관찰기간 동안)

β_{11} : 처치집단과 비교집단의 교정된 선형성장률의 차이

β_{1p} : 개인별 선형성장률에 대한 개인수준 공변수의 효과

51

B. Linear Growth Model

1. Two-level Linear Model

■ Random Effects

τ_{00} : 출발점 개인차의 잔차 분산

τ_{11} : 성장률 개인차의 잔차 분산

$\tau_{01} = \tau_{10}$: 출발점 위치와 성장률의 잔차 공분산

■ 평균성장곡선

비교집단 :

$$Y_{t.}^c = (\beta_{00} + \sum \beta_{0p} \bar{X}_{p.}) + (\beta_{10} + \sum \beta_{1p} \bar{X}_{p.})a_{t.}$$

실험집단 :

$$Y_{t.}^E = (\beta_{00} + \beta_{01} + \sum \beta_{0p} \bar{X}_{p.}) + (\beta_{10} + \beta_{11} + \sum \beta_{1p} \bar{X}_{p.})a_{t.}$$

52

B. Linear Growth Model

1. Two-level Linear Model

Level 1

$$ATTIT = \pi_{0i} + \pi_{1i}(\text{age16}) + e_i$$

Level 2

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{female}) + e_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{female}) + e_{1i}$$

53

B. Linear Growth Model

HLM 분석 결과 예시

Problem Title: NYS

Summary of the model specified (in equation format)

Level-1 Model

$$Y = P0 + P1*(AGE16) + E$$

Level-2 Model

$$P0 = B00 + B01*(FEMALE) + R0$$

$$P1 = B10 + B11*(FEMALE) + R1$$

54

B. Linear Growth Model

HLM 분석 결과 예시

Tau

INTRCPT1,P0	0.04470	-0.00097
AGE16,P1	-0.00097	0.00318

Tau (as correlations)

INTRCPT1, P0	1.000	-0.082
AGE16, P1	-0.082	1.000

Random level-1 coefficient Reliability estimate

INTRCPT1, P0	0.835
AGE16, P1	0.456

55

B. Linear Growth Model

The outcome variable is ATTTT

Final estimation of fixed effects

(with robust standard errors)

Fixed Effect	Standard		Approx.		
	Coefficient	Error	T-ratio	d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	0.523577	0.020252	25.853	239	0.000
FEMALE, B01	-0.069763	0.029206	-2.389	239	0.018
For AGE16 slope, P1					
INTRCPT2, B10	0.033804	0.007659	4.414	239	0.000
FEMALE, B11	-0.002887	0.010578	-0.273	239	0.785

56

B. Linear Growth Model

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	R0	0.21143	0.04470	234	1703.16415	0.000
AGE16 slope, R1		0.05636	0.00318	234	446.29282	0.000
level-1,	E	0.16943	0.02871			

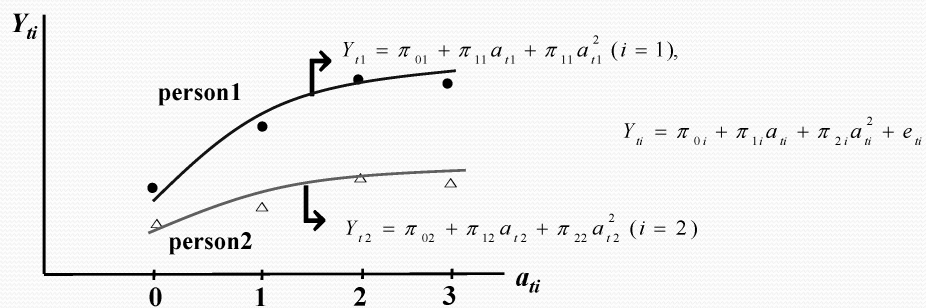
Statistics for current covariance components model

Deviance = -92.836635

Number of estimated parameters = 4

57

C. Non-Linear Growth Model



- Underlying principle is a quadratic function of time
- The latent change (quadratic change) is the same for all subjects

58

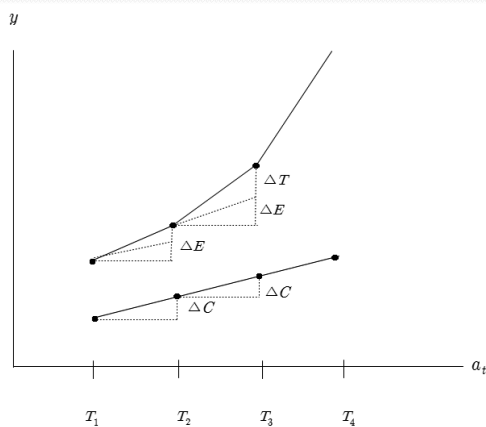
C. Non-Linear Growth Model

■ 연구질문:

1. 모집단에서의 평균성장곡선은 무엇인가?
2. 개인별 성장곡선은 전체 평균성장곡선에 어느 정도 변산을 갖는가?
개인차는 어느 정도인가?
3. 개인별 성장곡선은 어느 정도 정밀하게 추정되는가?
4. 각 시점에서의 성장률은 어느 정도 인가?
5. 개인별 출발점 위치와 성장률은 상관을 갖는가?
6. 개인별 출발점 위치와 성장률의 차이를 어떻게 설명하는가?
개인의 배경변수 및 특성, 집단의 특성 및 정책과 상관이 있는가?
(성장률: 선형 성장률, 가속 또는 감속 변화율)

59

C. Non-Linear Growth Model



실험집단(E): 0—0—X—0—0

비교집단(C): 0—0—O—0—0

ΔC : 비교집단의
자연성장(Development growth)부분

ΔE : 실험집단의 자연성장

ΔT : 처치의 고유효과
(Value Added growth) at time 3.

※ The growth curve is non-linear in nature!!

☆ Non-Linear model is more appropriate for growth modeling

☆ 처치의 고유효과 추정을 위해서는 적어도 3회 반복 측정이 필요함.

60

C. Non-Linear Growth Model

1) 기초모형(Unconditional Model): 개인별 성장곡선의 추정

Level-1

$$Y_{it} = \pi_{0i} + \pi_{1i}(a_{it} - L) + \pi_{2i}(a_{it} - L)^2 + e_{it}, \quad e_{it} \sim N(0, \sigma^2)$$

Level-2

$$\begin{aligned} \pi_{0i} &= \beta_{00} + r_{0i} & \begin{bmatrix} r_{0i} \\ r_{1i} \\ r_{2i} \end{bmatrix} & \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right) \\ \pi_{1i} &= \beta_{10} + r_{1i} \\ \pi_{2i} &= \beta_{20} + r_{2i} \end{aligned}$$

■ Fixed Effects: $\beta_{20} + r_{2i}$

측정시점 L ($a_{it}=L$)에서의 전체 평균

β_{00} : 측정기간 선형 성장률

β_{10} : 측정기간 동안 가속(감속)성장률

β_{20} : (acceleration rate)

61

C. Non-Linear Growth Model

■ Random Effects

τ_{00} : 측정시점 L 에서의 개인차 분산

τ_{11} : 측정기간 선형성장률의 개인차 분산

τ_{22} : 측정기간 가속성장률의 개인차 분산

$\tau_{01}, \tau_{02}, \tau_{12}$: 개인별 성장곡선 모수간의 공분산

■ 전체 평균 성장곡선:

$$Y_{it} = \beta_{00} + \beta_{01}(a_{it} - L) + \beta_{20}(a_{it} - L)^2$$

62

C. Non-Linear Growth Model

2) 연구모형(Conditional Model) : 성장곡선의 개인차 설명

$$L-1 \quad Y_{it} = \pi_{0i} + \pi_{1i}(a_{it} - L) + \pi_{2i}(a_{it} - L)^2 + e_{it}, \quad e_{it} \sim N(0, \sigma^2)$$

$$L-2 \quad \begin{matrix} \pi_{0i} = \beta_{00} + \beta_{01}(\text{처치}) + r_{0i} \\ \pi_{1i} = \beta_{10} + \beta_{11}(\text{처치}) + r_{1i} \\ \pi_{2i} = \beta_{20} + \beta_{21}(\text{처치}) + r_{2i} \end{matrix} \quad \begin{bmatrix} r_{0i} \\ r_{1i} \\ r_{2i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} & \tau_{02} \\ \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right)$$

▣ 통합모형

$$\begin{aligned} Y_{it} = & \beta_{00} + \beta_{01}(\text{처치}) + (\beta_{10} + \beta_{11}(\text{처치}))(a_{it} - L) \\ & + (\beta_{20} + \beta_{21}(\text{처치}))(a_{it} - L)^2 \\ & + r_{0i} + r_{1i}(a_{it} - L) + r_{2i}(a_{it} - L)^2 + e_{it} \end{aligned}$$

63

C. Non-Linear Growth Model

(1) 평균성장곡선

비교집단의 평균 성장곡선(처치=0)

$$Y_{it}^C = \beta_{00} + \beta_{10}(a_{it} - L) + \beta_{20}(a_{it} - L)^2$$

처치집단의 평균 성장곡선(처치=1)

$$Y_{it}^E = (\beta_{00} + \beta_{01}) + (\beta_{10} + \beta_{11})(a_{it} - L) + (\beta_{20} + \beta_{21})(a_{it} - L)^2$$

(2) 각 집단별 자연성장 부분

비교집단의 자연 성장분

$$\Delta^C = Y_{it=2}^C - Y_{it=1}^C = \beta_{10} + \beta_{20}(3 - 2L)$$

처치집단의 자연 성장분

$$\Delta^E = Y_{it=2}^E - Y_{it=1}^E = (\beta_{10} + \beta_{11}) + (\beta_{20} + \beta_{21})(3 - 2L)$$

64

C. Non-Linear Growth Model

(3) 처치의 Net Effect(Value Added Progress)

$$\text{at time=3} \quad Y_{t=3}^E - (Y_{t=2}^E + \Delta^E) = 2(\beta_{20} + \beta_{21})$$

$$\text{at time=4} \quad Y_{t=4}^E - (Y_{t=3}^E + \Delta^E) = 4(\beta_{20} + \beta_{21})$$

(4) 각 개인별 time=t 에서의 성장률

$$\frac{d(Y_{ii})}{d(a_{ii})} = \pi_{1i} + 2\pi_{2i}(a_{ii} - L)$$

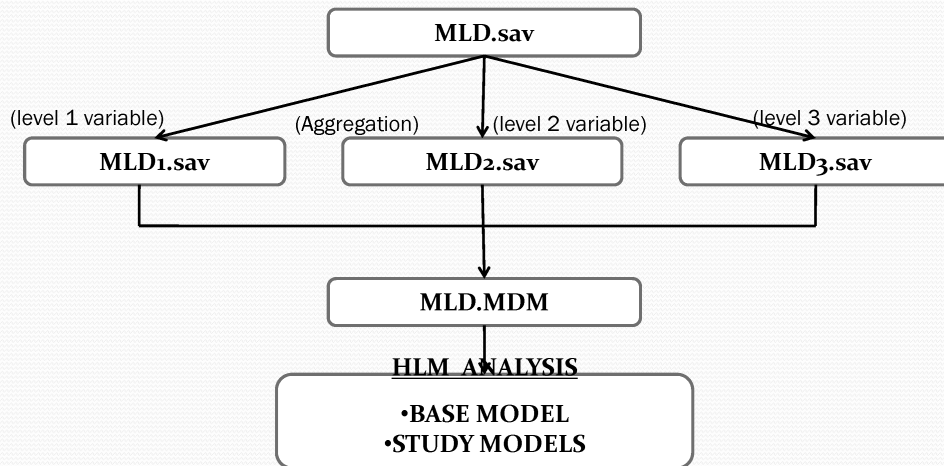
※ 모든 개인의 성장률은 시간 (a_{ii})에 따라 다름.

65

Part .5

Computer Lab

A. HLM data 생성절차



67

B. Computer Lab

분석자료

[FILE NAME: Nys21.sav]

List of variables on the working file

Name		Position
ID	student id	1
	Print Format: F4	
	Write Format: F4	
ATTIT	a 9-item scale assessing attitudes favorable to deviant behavior	2
	Print Format: F3.2	
	Write Format: F3.2	
AGE16	age of students	3

68

B. Computer Lab

분석자료

[FILE NAME: Nys22.sav]

List of variables on the working file

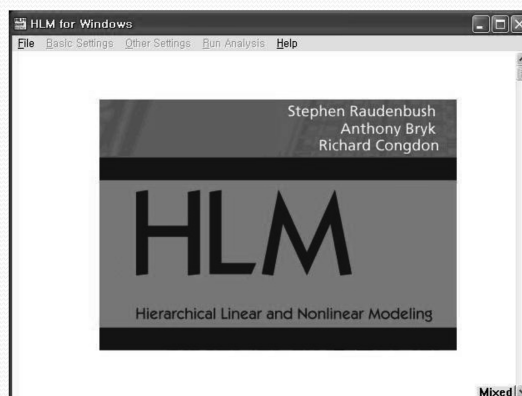
Name	Position
ID student id	1
Print Format: F4	
Write Format: F4	
FEMALE an indicator for gender (1=female, 0=male)	2

69

B. Computer Lab

➤ Window 환경에서의 HLM 분석

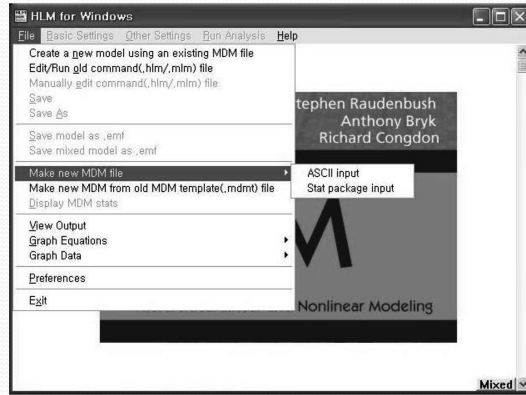
- 바탕 화면에서 [hlm6] 또는 [whlm] 단축 icon을 누르면, 다음의 화면이 나타난다.



- 위의 화면에서 마우스로 [File] 단추를 누르면 다음과 같은 화면의 변화를 볼 수 있다.

70

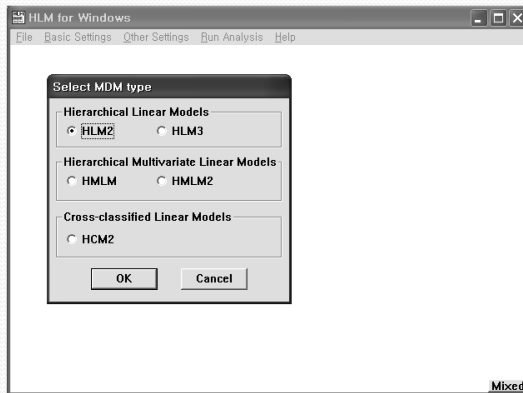
B. Computer Lab



- 위의 화면은 HLM의 투입자료로 Ascii code 자료와 기존의 통계 Package (e.g., SPSS, SAS, SYSTAT)의 자료를 모두 활용할 수 있음을 보이고 있다.

71

B. Computer Lab

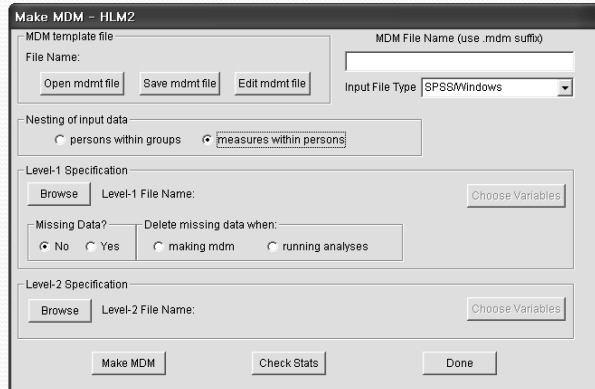


- HLM3: 3층위 자료에 적용됨 (예, 학생-학급-학교)
- (임시혁, 1997, 교육평가연구; 강상진, 1998, 교육평가연구)
- HMLM: 다변량 2-level Model.
- HMLM2: 다변량 3-level Model.

72

B. Computer Lab

- 위의 화면에서 HLM2를 선택하면, 다음의 화면이 나타나고 여기서 MDM file의 모든 요건을 선택한 후에 [Make MDM] 단추를 누르면 MDM file이 생성된다.



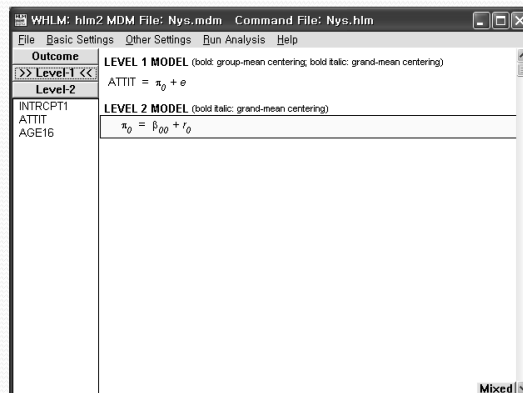
- 위의 화면에서 MDM file의 생성을 위한 조건을 선택하는 순서는 다음과 같다.
1.MDM file name; 2.Input File type; 3.Nesting of input data; 4.Model Specification 5. Save Responsive file(MDMT file); 6. Make MDM

73

B. Computer Lab

2. 다층모형에 의한 분석방법

- [Make MDM] 단추를 누르면, MDM file에 있는 변수들의 기술통계량이 화면에 나타나고, 이를 확대해서 보려면 [Check Stats] 단추를 누르면 메모 화면이 나타나, 이를 프린트할 수 있다. 그 후에 다음의 화면이 나타나서 HLM 분석을 시도할 수 있다.

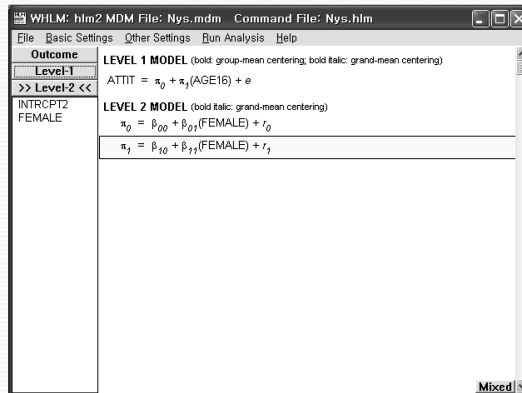


- 위의 화면에서 2-level HLM 분석을 위한 모든 모형을 명세화 할 수 있다. 위의 모형 명세화는 level-1과 level-2의 모형에서 예측변수가 없는 모형이다.

74

B. Computer Lab

- 다음의 화면은 level-1과 level-2에서 모두 예측변수가 있는 모형을 명세화한 것이다.



The screenshot shows the WHLM software interface with the following content:

```
WHLM: hlm2 MDM File: Nys.mdm Command File: Nys.hlm
File Basic Settings Other Settings Run Analysis Help
Outcome
Level-1
>> Level-2 <<
INTRCPT2
FEMALE

LEVEL 1 MODEL (bold group-mean centering, bold italic: grand-mean centering)
ATTIT =  $\pi_0 + \pi_1(\text{AGE16}) + e$ 

LEVEL 2 MODEL (bold italic: grand-mean centering)
 $\pi_0 = \beta_{00} + \beta_{01}(\text{FEMALE}) + r_0$ 
 $\pi_1 = \beta_{10} + \beta_{11}(\text{FEMALE}) + r_1$ 
```

- 위의 화면에서 [File]를 선택하여 [Save As]로 *.hlm file를 저장한 후에 [Run Analysis] 단추를 누르면 자료 분석이 이루어진다.

75

Part .6

Growth Modeling에서 반복측정
분산분석, SEM, HLM 비교

1. 반복측정 분산분석 모형(R-ANOVA)의 문제점

- 1) 개인별 성장곡선을 추정하지 않음.
- 2) 실험연구(무선배치 적용) 자료 분석에 제한 됨.
유사실험연구자료, 조사연구자료의 분석에 부적절.
- 3) 관찰시점과 관찰횟수가 모든 개인에게 동일하여야 함.
==> 관찰시점이 개인간 다르거나, 관찰횟수가 다르면
결측치로 처리 또는 자료를 변조해야 함.
- 4) time variant 공변수의 도입이 불가능함.
- 5) 연속변수를 독립변수로 도입 불가
(K-1개의 가변수로 처리함)
- 6) 위계적 자료의 분석모형에서 오차항(Error term)의 결정 어려움.

77

2. Advantage of HLM

- 1) R-ANOVA에서 지적된 모든 제한점 극복.
- 2) 관찰횟수와 관찰시점의 개인간 차이 허용.
- 3) 각 수준에서 모형의 설명력 제공.
- 4) 각 수준별 모수 추정의 측정오차 극복
L-1 모수 : $\hat{\pi}$ (추정치) = π (모수) + ϵ_{π} (측정오차)
 $VAR(\hat{\pi}) = \tau_{\pi} + \sigma_{\pi}^2 = \text{모수분산} + \text{오차분산}$
L-2 모수 : $\hat{\beta}$ (추정치) = β (모수) + ϵ_{β} (측정오차)
 $VAR(\hat{\beta}) = \tau_{\beta} + \sigma_{\beta}^2 = \text{모수분산} + \text{오차분산}$
- 5) 모형 명세화의 논리적 일관성.

78

3. HLM vs. SEM

	HLM	SEM
자료의 구조	다층자료	단층자료 (modeling difficulty)
다층자료의 다양성	multi-way, multi-layer nested	only one-way nested
종단자료의 관찰시점	flexible	fixed
관찰횟수	flexible	fixed
Missing values at Micro level	delete, impute	robust (use covariance matrix)
불균형자료 (case missing)	account for MAR	account for MAR
Model 적합성	$R_1^2, R_2^2,$ $-2 \text{ LogL} \sim \chi^2$	$\chi^2 (H_0: \Sigma_\theta = \Sigma_S),$ GFI, AGFI, RMSEA, NFI, NNFI, CFI
편의성	Easy to use? choice of the software.	

79

● 수고하셨습니다!

80



잠재성장모형

이 기 종

(국민대학교 교육학과 교수)

잠재성장모형(LGM)

이기종
국민대학교 교육학과
rhee0408@kookmin.ac.kr
02)910-4736, 010-9900-2277

Model fit

$$H_0: \Sigma = \Sigma(\theta)$$

Exact fit

진술된 영가설이 정확하게 맞다는 가정

test statistic: chi-square, $(N-1)F$

$$F_{ML} = \log |\Sigma(\theta)| + \text{tr}[\mathbf{S}\Sigma(\theta)^{-1}] - \log |\mathbf{S}| - (p+q)$$

Close fit

영가설이 근사(비슷)하게 맞다는 가정

test statistic: RMSEA

$$\begin{aligned} \text{RMSEA} &= \sqrt{\frac{F}{df}} \\ &= \sqrt{\max\left[F - \frac{df}{N-1}, 0\right]} \end{aligned}$$

Parameter fit

추정된 자유모수 평가

test statistic: student t

If alpha = .05, GT 1.96 in absolute value

Check improper solutions

Heywood case, negative variance

잠재성장모형(LGM)

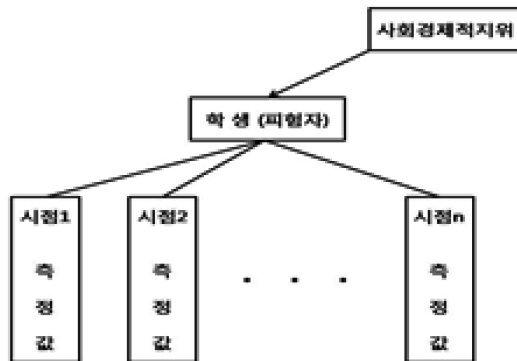
- 구조방정식모형(SEM)으로 성장궤적을 탐구하는 것
- 피험자 개개인으로부터 여러 번 반복해서 관찰된 측정값이 초기상태 및 변화율의 두 잠재변수로 잠재화되어 측정되고, 이 두 잠재변수간의 관계가 잠재변수모형을 통해 밝혀짐
- 초기상태= 절편, 변화율(궤적)= 기울기

Longitudinal data(종단자료)

- cohort data
- panel data
- 둘 다 여러 번 반복관찰되는 것은 동일하나 연구대상 표본에서 차이
- 엄밀하게는 종단자료는 패널에만 국한

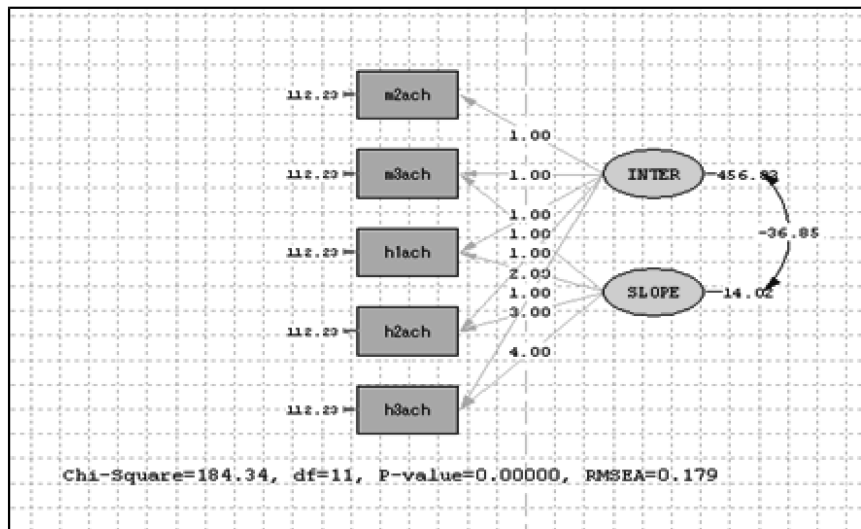
종단자료의 구성

- 피험자에 내재된 이중구조

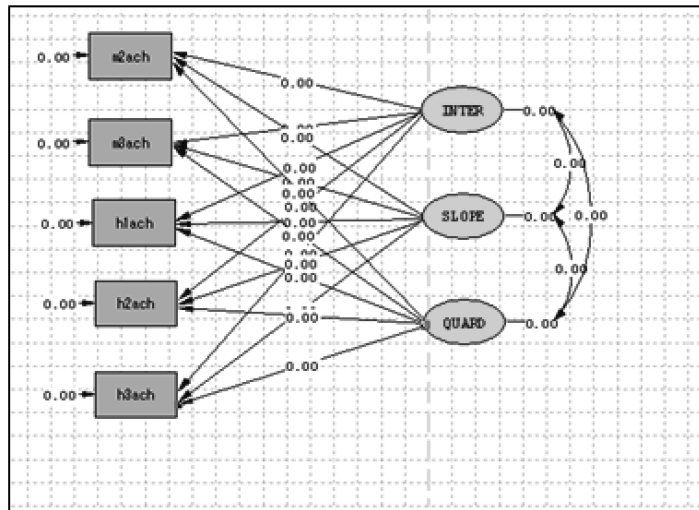


[그림 8-1] 2중구조 예

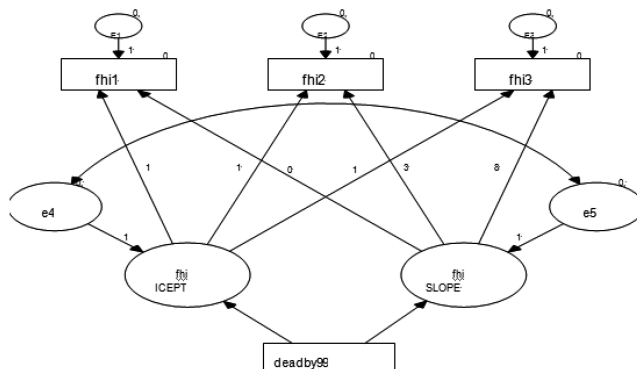
Unconditional model 예시



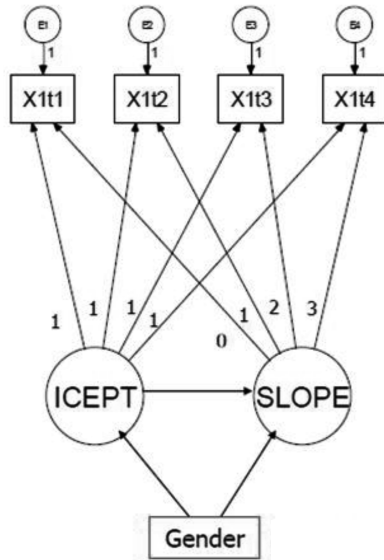
Quadratic term 예시



Conditional model 예시



사망으로 인한
손실(deadby99)이
초기상태,
기울기에
유의미한 관계?



남성이 여성에 비해
초기상태 및 변화율에서
차이가 있는가?

LGM 분석 필요요건

- 종속변수가 연속적이어야
- 측정단위가 같아야, equating 가능
- 적어도 3회의 측정, three waves
- 시간구조가 동일간격이어야

세미나자료집 12-S30

**2012년 한국아동·청소년패널2010
데이터분석방법론세미나 자료집**

인 쇄 2012년 7월 4일

발 행 2012년 7월 5일

발행처 한국청소년정책연구원

서울특별시 서초구 태봉로 114

발행인 이 재 연

등 록 1993. 10. 23 제 21-500호

인쇄처 (주)문영사 전화 02)2263-5087 대표 김희자
