

구조방정식모형:

기본 개념과 활용 사례

강사: 노법래(세명대학교)

# 특강 내용

- 1부: 구조방정식 개관
  - 왜 구조방정식인가
  - 분석 관련 주요 개념
  
- 2부: 활용 사례
  - 유형화, 잠재집단분석(LCA)
  - 궤적 분석, 잠재성장모형
  - 종단 자료, 종단적 구조방정식
  - 집단 비교, 성향점수 매칭과의 활용

# 1부: 구조방정식 개관

# 구조방정식 개관: 연구 사례

Published in final edited form as:

*Child Dev.* 2007 ; 78(1): 70–95. doi:10.1111/j.1467-8624.2007.00986.x.

## Income Is Not Enough: Incorporating Material Hardship Into Models of Income Associations With Parenting and Child Development

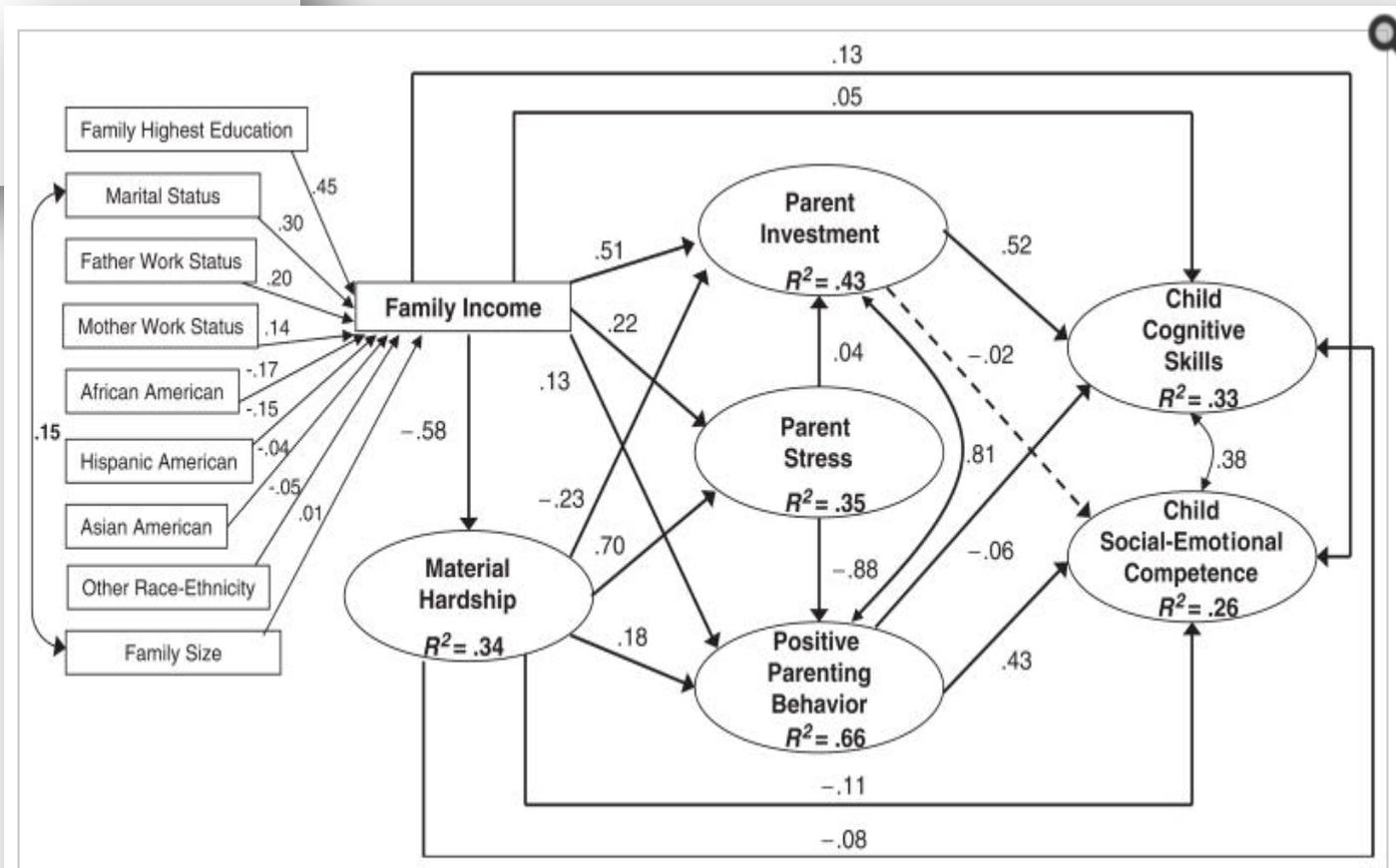
Elizabeth T. Gershoff,  
University of Michigan

J. Lawrence Aber,  
New York University

C. Cybele Raver, and  
University of Chicago

Mary Clare Lennon  
Columbia University

- 통상적인 회귀분석 모형과 무엇이 다르나요?



# 구조방정식 개관

## 왜 구조방정식인가?

통상적인 회귀분석으로 하기 어려운 것...

- 매커니즘 혹은 causal process의 규명
- 다수의 outcome 다루기(매개, 최종)
- 잠재변수(latent variable) 구조의 동시 고려
- 역인과관계의 모형화

...즉, 구조적 접근(structural approach)

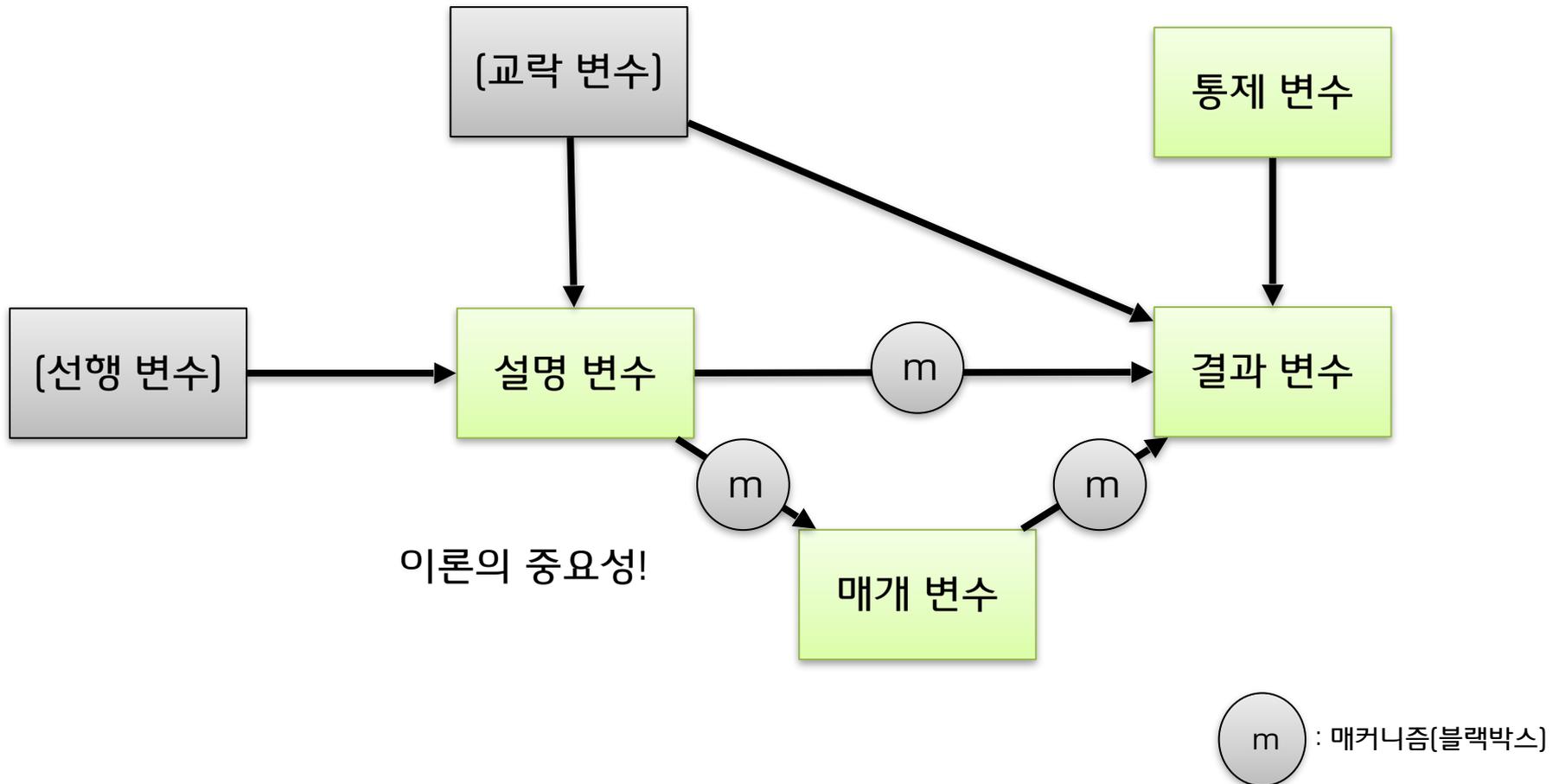
# 구조방정식 개관: 인과 과정 (causal process)

- 설명 변수(causal variable)는 결과 변수에 대해 인과적 선행성 혹은 외생적(exogeneity) 성격을 가질 필요가 있음.
- 실세계에서의 양방향의 인과성(reciprocal causal relationships)에 대한 고려가 필요. 그러나 인과적 방향을 특정하였다면, 그에 대한 충분한 이론적 설명이 뒤따를 필요가 있음.

# 구조방정식 개관: 인과 과정 (causal process)

- 설명 변수와 결과 변수는 서로 공변(covariation)하는 관계에 있어야 하며, 이때 교락 변수(confounding)의 존재로 인한 허위적인 관련성이 있지 않은지 살펴 볼 필요가 있음.
- 교락 변수를 다루는 방법론적 전략
  - 대리변수의 구성
  - 성향점수의 추정
  - 구조방정식 활용 모형화

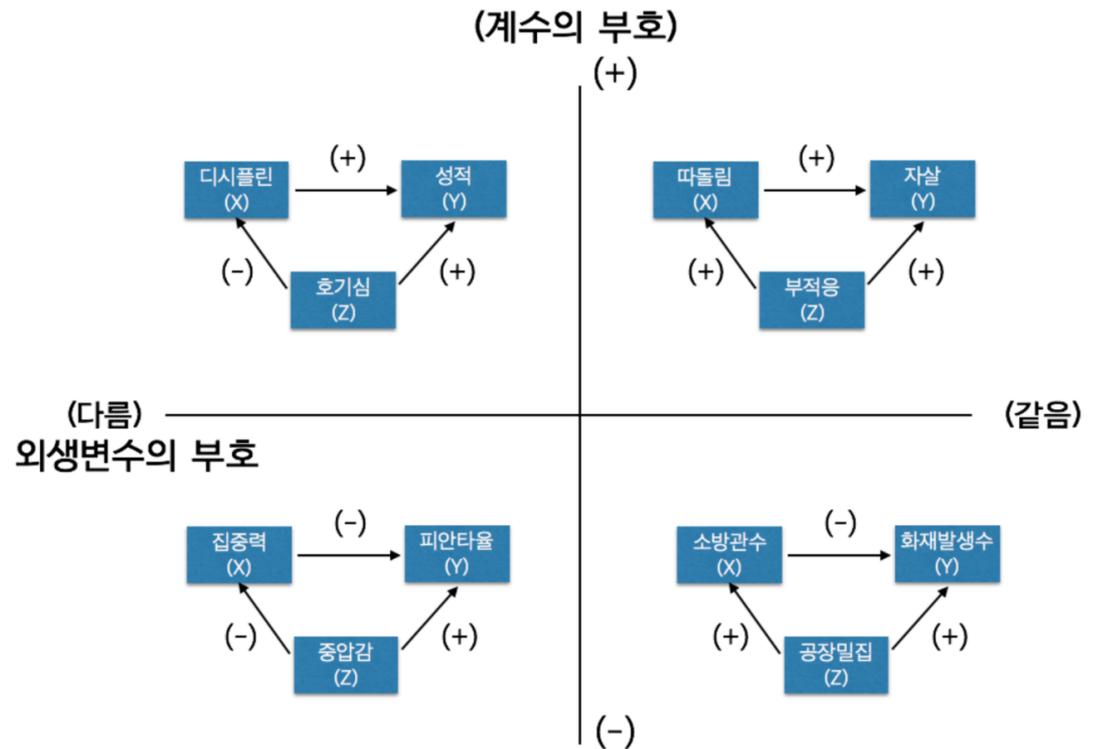
# 구조방정식 개관: 인과 과정(causal process)



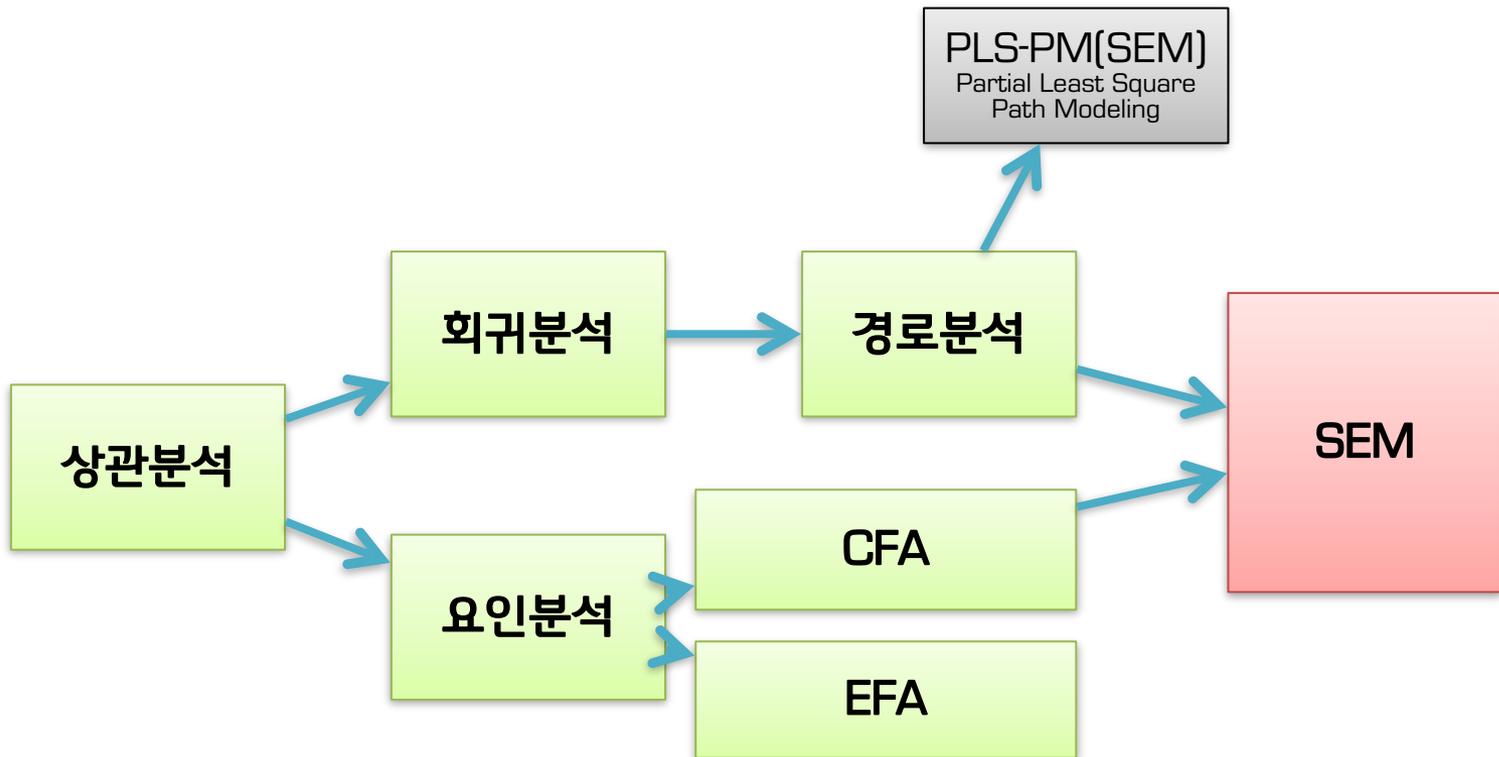
# 구조방정식 개관: 인과 과정 (causal process)

- 내적 타당성

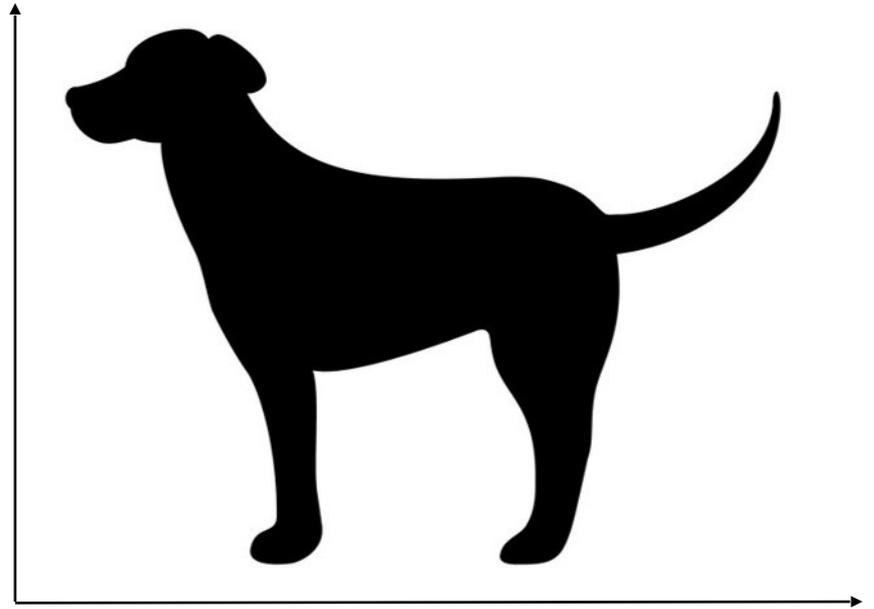
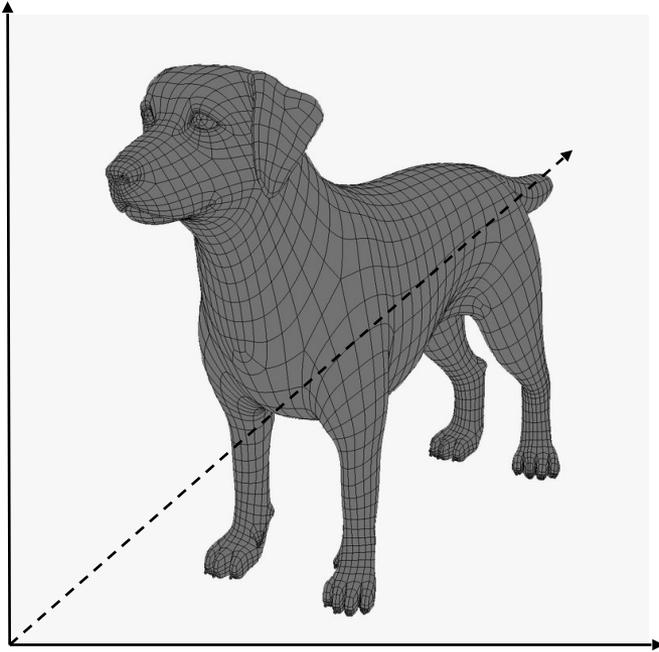
- 탐색적으로 확인된 인과 관계의 타당성을 저해 상황



# 구조방정식 개관: 기법 계통도



# 요인분석: 차원축약 & 고유값



# 단차원성(unidimensionality)

- 구조방정식 분석을 하기에 앞서 척도(측정 모형)에 대한 충분한 검토가 필요
- 복수의 측정 변인을 하나의 공통 차원(요인)으로 묶는(그룹핑) 하는 과정에서 단차원성의 고려는 중요함.
- 단차원성은, 한 척도에 속한 여러개의 문항이 정말 “하나의 변수”를 포착하는 것인가의 문제.[Mcdonard, 1998]
- 단차원성의 검토에 있어 Scree plot을 통한 elbow method는 간명하고 직관적인 방법[Cattle, 1966].

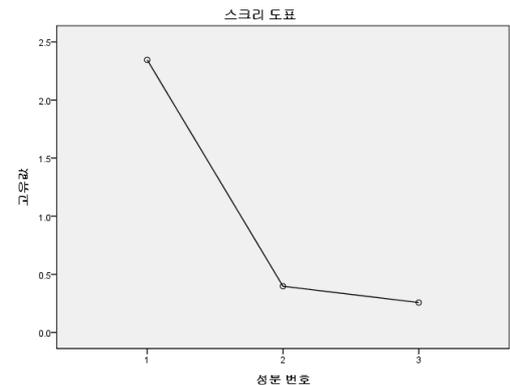
# 단차원성(unidimensionality)

- 단차원성의 측정

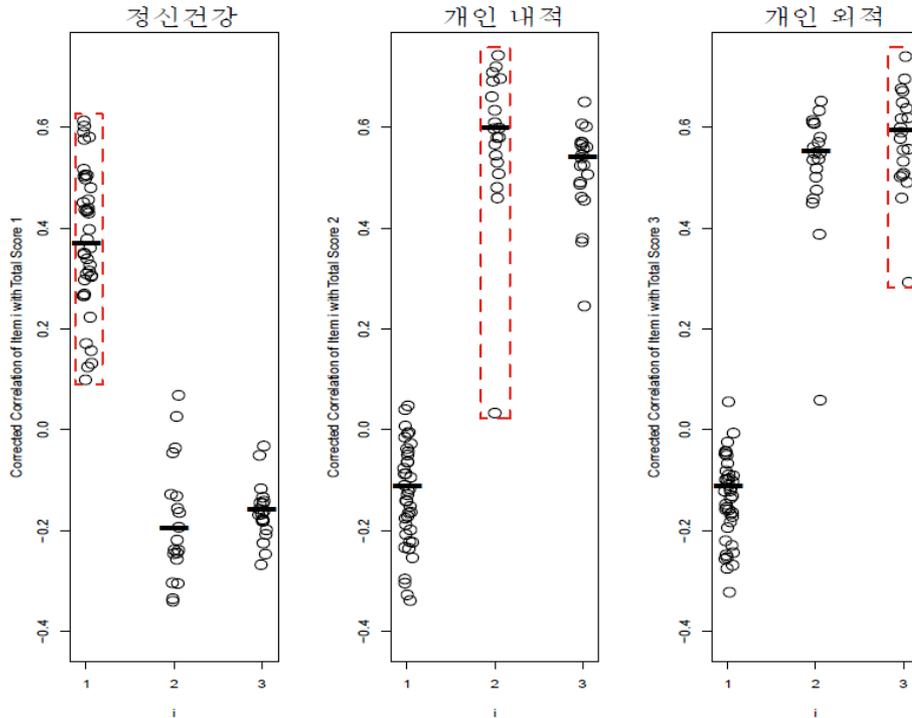
- 신뢰도 평가지표(e.g. Cronbach alpha, gamma six, Dillon-Goldstein's rho 등). 통상적으로  $>.9$  기준을 활용함.

- 상관계수 행렬의 first eigenvalue와 second eigenvalue의 검토. 전자가 1이상 이면서 후자가 1미만인 경우 단차원성을 확보한 것으로 탐색적 결론.

- “Elbow Method”



# 단차원성(unidimensionality)



정신건강		개인 내적		개인 외적	
문항	요인총점-항목 절대 상관계수	문항	요인총점-항목 절대 상관계수	문항	요인총점-항목 절대 상관계수
s18	0.27	s53	0.58	s54	0.58
s32	0.27	s56	0.58	s47	0.56
s4	0.27	s83	0.57	s77	0.56
s25	0.22	s66	0.54	s69	0.53
s44	0.19	s85	0.53	s62	0.51
s37	0.17	s75	0.51	s63	0.5
s12	0.16	s60	0.48	s78	0.5
s34	0.13	s64	0.46	s68	0.49
s35	0.12	s81	0.31	s48	0.46
s28	0.10	s57	0.03	s59	0.29

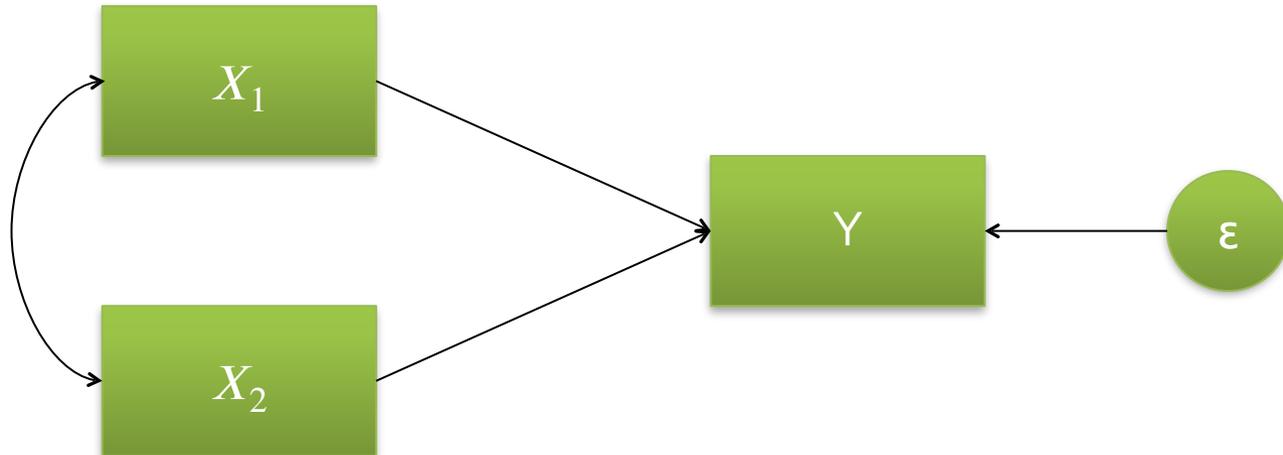
- 척도의 신뢰도를 높이고 단차원성을 확보하기 위해서 문항 간 상관구조에 대한 검토가 필요하며, 이를 통해 문항 삭제, 수정 등에 대한 분석적 고려가 있어야 함.

# 분석 자료의 일반적 특성

- 변수 수준: 연속변수(등간/비율척도)
- 표본 크기: 표본 크기의 최소 수준에 대한 구체적 가이드라인이 있는 것은 아니나 다음을 권장.
  - $n > 100$  (최소  $n > 50$ )
  - $n >$  측정 변수의 수 10~5배 수준
- 변인 상관 수준: 실제 연산에 활용되는 자료는 상관관계 행렬이며 이는 통계 패키지 등에서 자동으로 진행됨. 따라서 관측 변수 간 평균 상관계수의 절대값이 0.3 이상일 것을 권장함. 측정 변인 간 상관관계 수준이 낮다면 신뢰도 측정 등에 있어서 기대한 효과를 거두기 어려움.

# 구조방정식의 특성: 회귀분석과의 관련성

- 회귀식 :  $Y = a_0 + a_1X_1 + a_2X_2 + \varepsilon$

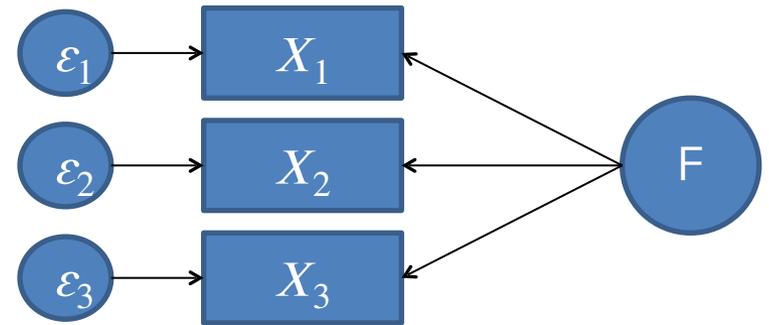


☞ 회귀식을 SEM의 한 “특정한 예”로 볼수도 있음.

# 구조방정식의 특성: 요인분석과의 관련성

- 예제 연구상황 : “생활만족”(F)을 측정하기 위해 “주거만족”(X<sub>1</sub>), “가족관계만족” (X<sub>2</sub>), “여가생활만족” (X<sub>3</sub>)의 세 측정 변수를 사용하는 경우.

- 식: 
$$X_1 = a_1 F + \varepsilon_1$$
$$X_2 = a_2 F + \varepsilon_2$$
$$X_3 = a_3 F + \varepsilon_3$$



☞ [확증적]요인분석 또한, SEM의 한 특수한 예로서 이해할 수 있음.

# 변수의 성격

## [측정 수준에 따른 분류]

- 잠재변수 (Latent Variable)
- 관측(측정)변수 (Observed Variable)

## [분석 모형 내 위치에 대한 분류]

- 외생변수 (Exogenous Variable)
- 내생변수 (Endogenous Variable)
- 오차변수 (Error Variable)

# 효과 추정

- 직접효과(direct effects): 독립변수가 종속변수에 직접 미치는 효과
- 간접효과(indirect effects): 독립변수가 하나 이상의 매개변수를 통하여 종속 변수에 영향을 미치는 효과.
- 총효과(total effect) = 직접효과 + 간접효과
  - 이때, 직접효과 = 0, |간접효과| > 0 → 완전매개구조
  - |직접효과| > 0 & |간접효과| > 0 → 불완전매개구조

# 구조방정식 모형에서의 오차

- 오차 변수는 모형이 설명하지 못한 변이에 대한 정보를 포함하며 두 유형으로 대별할 수 있음.
  - 잠재구조의 경우 -> 측정 오차 (Measurement error)
    - 잠재변수-측정변수 구조를 완전히 모형화하지 못함에 따른 오차.
    - 측정변수 측에 오차변수가 연결됨.
  - 회귀구조의 경우 -> 구조적 오차 (Structural error)
    - 내생변수가 회귀식에 포함된 외생변수에 의해서 설명이 되지 않은 부분.
    - 내생변수측에 오차변수가 연결됨

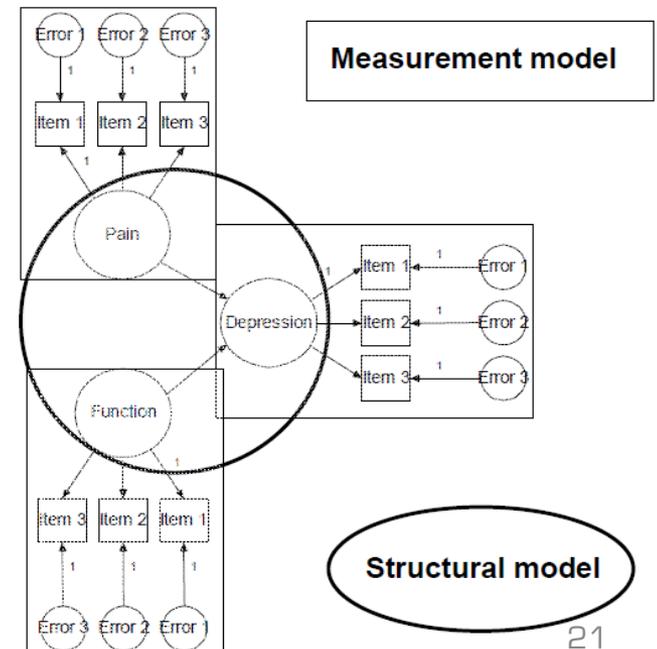
# 정리: 경로 + [잠재변수] = SEM

- 측정 모델(measurement model)

-> 잠재변수와 측정변수 간 구조

- 구조 모델(structural model)

-> 요인 & 측정 변인 간 회귀구조



# 모형 적합도

**절대적합지수 (Absolute fit index)**

**: CMIN( $\chi^2$ ), GFI, RMR, RMSEA, NCP, ECVI**

**증분적합지수 (Incremental fit index)**

**: TLI, NFI, RNI, BFI, CFI**

**간명적합지수 (Parsimonious fit index)**

**: AGFI, PNFI, PGFI, Normed  $\chi^2$ , AIC**

# 모형 적합도

모형적합도	값의 범위	일반적 해석 기준
*CMIN(chi-square값)	-	유의도(p)가 0.05이상
GFI (Goodness of fit index)	0 - 1	0.9 이상
AGFI (Adjusted GFI)	0 - 1	0.9 이상
RMR (Root-mean-square residual	0 - 1	0.05 이하 양호
<u>RMSEA (Root mean square error of approximation)</u>	0 - 1	0.10~0.08 : 보통 0.08~0.05 : 양호 0.05 이하 : 좋음
<u>TLI (Tucker-Lewis index)</u>	0 - 1	0.9 이상
<u>NFI (Normed fit index)</u>	0 - 1	0.9 이상
PGFI (Parsimony goodness of fit index)	0 - 1	클수록 양호
AIC (Akaike information criterion) BIC (Bayesian information criterion)	0 이상의 값	모형 간 비교에 활용 (수치가 작을수록 적합도가 높음)

\*CMIN : (샘플의 크기가 커질수록 기각역( $p < .05$ )에 속할 가능성이 크므로 신뢰도가 높은 지수라고 하기에 한계가 있음)

# 분석의 일반 절차

- Step 1 : 관심있는 변수들을 정하기(관측, 잠재)
- Step 2 : 변수간의 관계에 대한 구조 결정(잠재구조와 회귀 구조)
- Step 3
  - 통계적 평가 : 모형적합도, 경로계수의 부호, 유의미성, 크기
  - 이론적 평가 : 간명성, 결과의 이론적 설명 가능성
- Step 4
  - 경로구조에 대한 해석(직접, 간접, 순환 구조 등)
  - 다중집단 비교를 하는 경우: 집단 간 추정치 차이 검토
  - 궤적을 다루는 경우(e.g. LGM): 집단별 궤적 형태에 대한 비교
  - 유형을 다루는 경우(e.g. LCA): 유형수의 타당성 및 유형 간 비교

## 2부: 활용 사례들

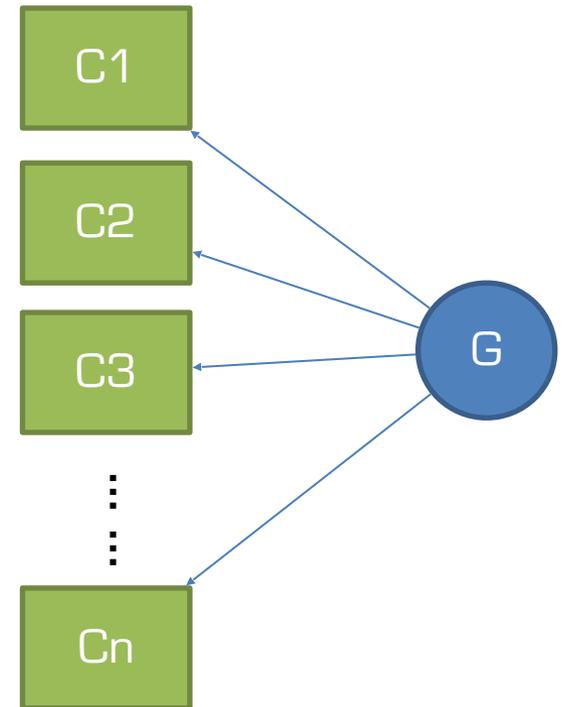
# 유형화 분석

- 잠재집단모형(Latent Class Model)
- 잠재프로파일모형(Latent Profile Model)
- 잠재평균모형(Latent Mean Model)
- 문항반응이론 (Item Response Theory)

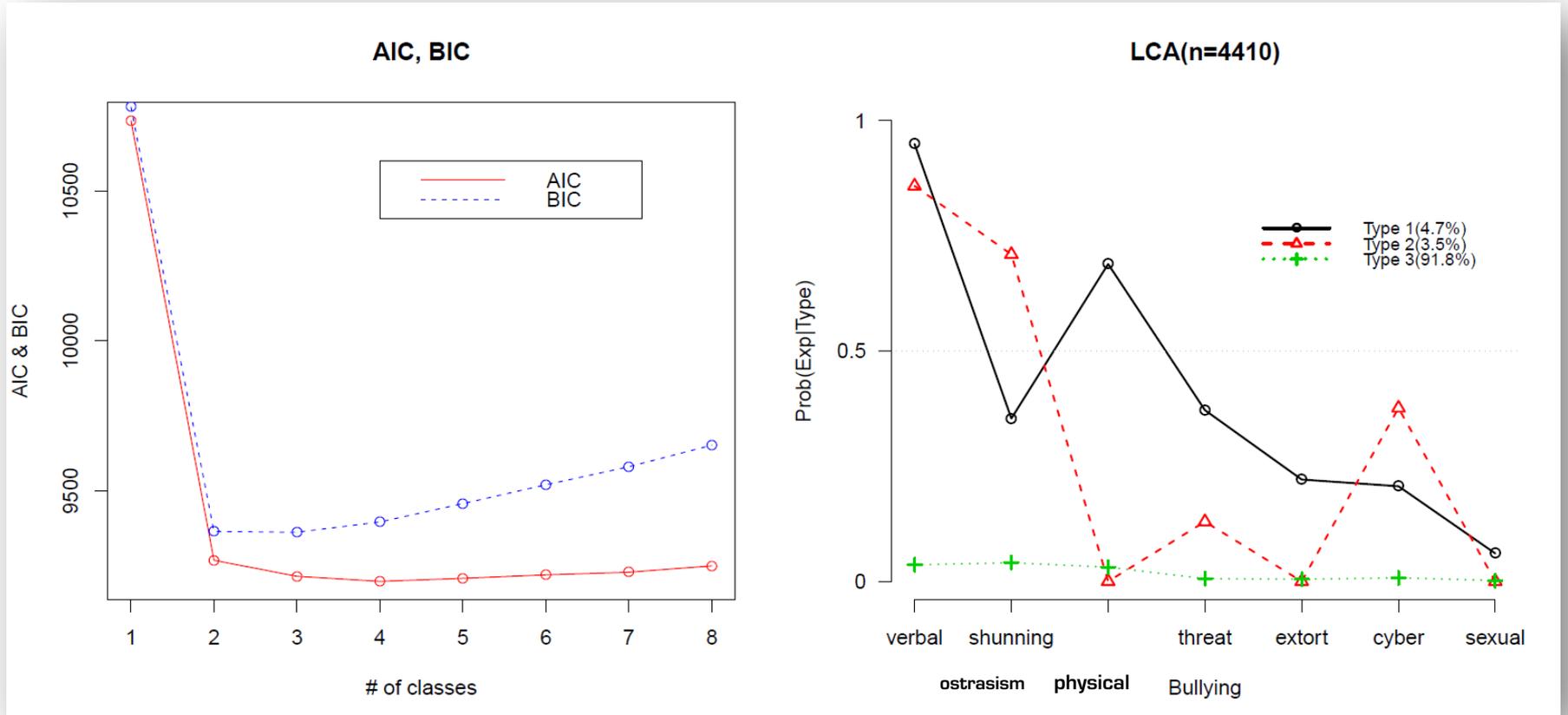
		측정변인	
		연속	명목
잠재변인	연속	(      )	IRT
	명목	LPA	LCA

# 유형화 분석: LCA

- 유형화 기법, unsupervised.
- 명목변수(2분변수)의 활용.
- 명징한(distinct) (잠재)집단이 존재할 것이라는 가정.



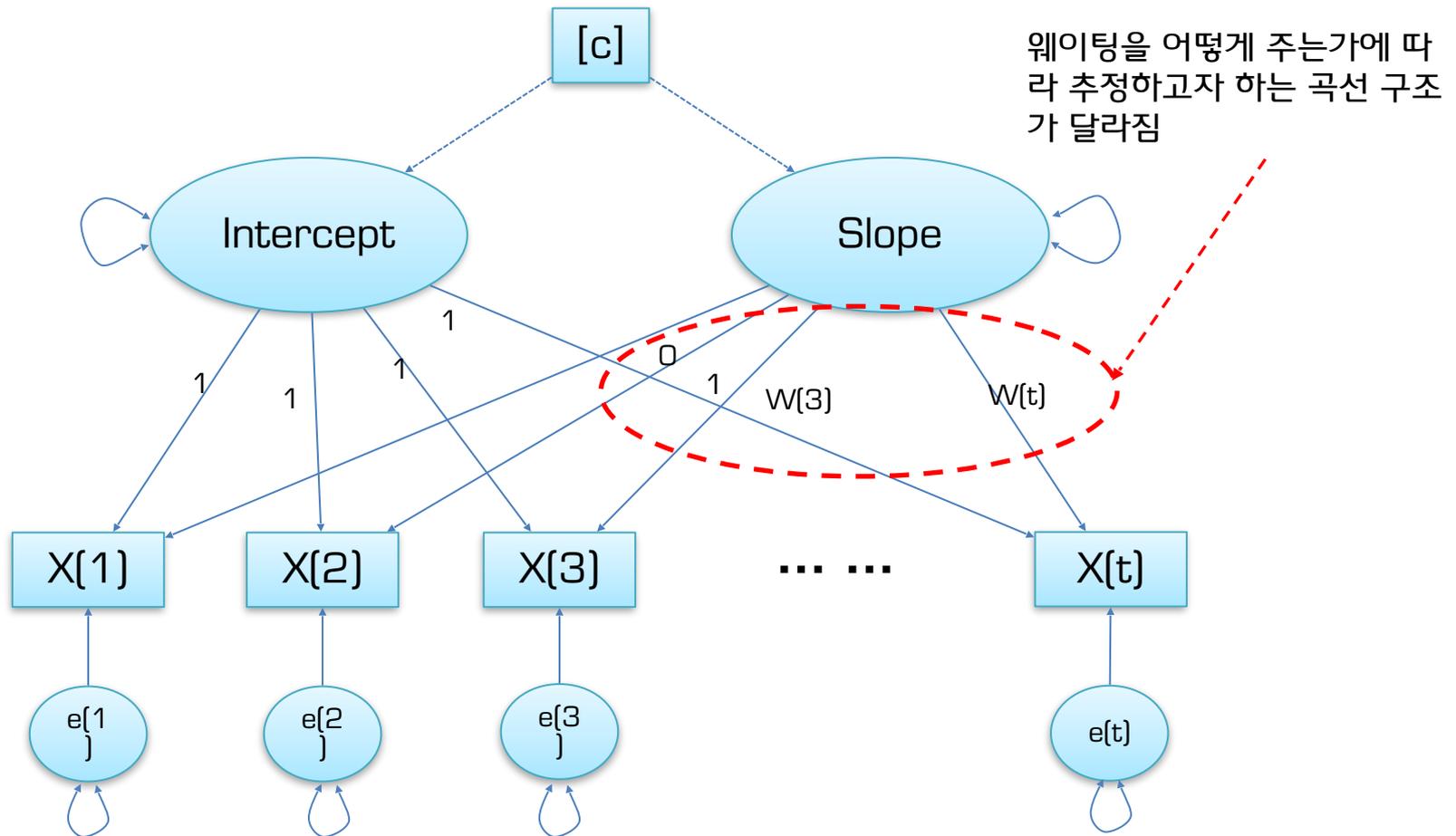
# 유형화 분석: LCA



Roh et al.(2015)

# 궤적 분석, LGM

- Mean structure의 종단적 변화 모형화



# 궤적 분석, LGM

반복측정(종단) 데이터, 일반적으로 wide type

절편(I)과 기울기(S)의 평균 추정을 통해 종단적 변화를 모형화

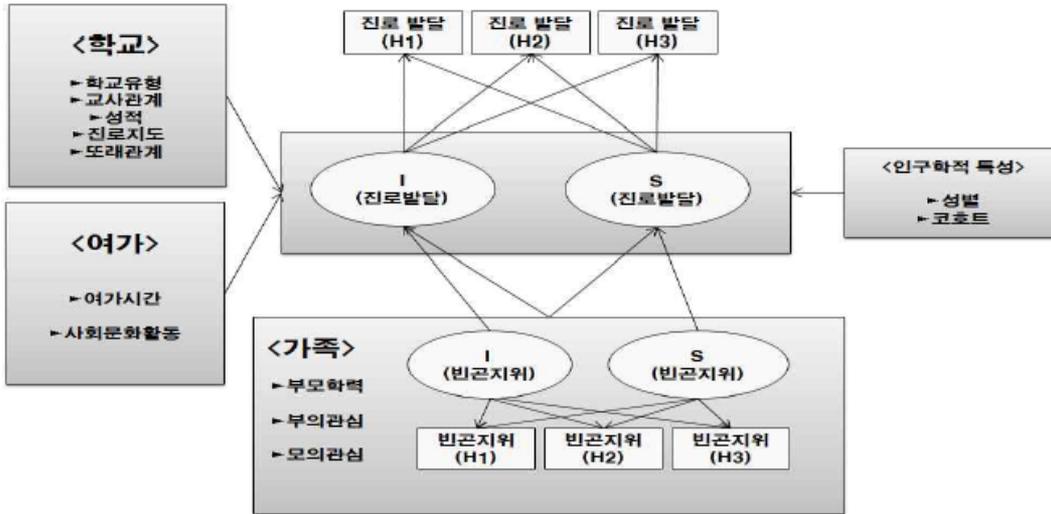
- 추정된 I평균은 추정된 초기 평균을, 추정된 S평균은 시기별 변화량의 총평균을 의미함.

한 집단의 평균적 변화 궤적을 포착, 다집단 비교 가능.

I와 S의 상관관계에 대한 검토.

궤적에 영향을 미치는 영향요인 포함 가능.

# 궤적 분석, LGM



노법래(2013)

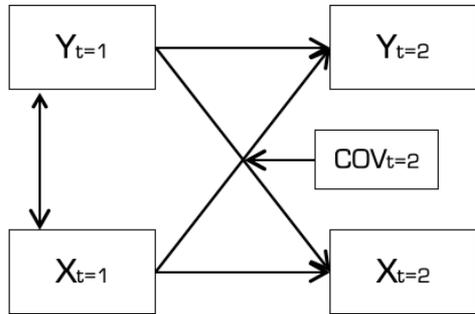
표 2  
잠재성장모형 분석 결과 요약

결과 변수	변수유형	변수	Estimate	S.E.	P> z
진로 발달 (I)	가족변수	부모학력	.168	.137	.222
		부의관심	.033	.135	.807
		모의관심	.333	.145	.022
	학교변수	빈곤지위(I)	-1.990	4.823	.680
		학교유형	-.933	.952	.327
		교사관계	.185	.615	.764
여가변수	성적	2.653	.477	.000	
	진로지도	.983	.414	.018	
	또래관계	1.800	.792	.023	
인구학적 변수	여가시간	.144	.223	.517	
	사회문화활동	.839	.423	.047	
진로 발달 (S)	가족변수	성별	-1.373	.693	.048
		코호트	-1.394	.692	.044
		부모학력	-.031	.088	.728
	학교변수	부의관심	-.020	.087	.816
		모의관심	-.028	.093	.766
		빈곤지위(S)	-17.802	8.064	.027
여가변수	학교유형	.912	.611	.136	
	교사관계	-.405	.395	.306	
	성적	-.072	.306	.814	
인구학적 변수	진로지도	.186	.266	.484	
	또래관계	.498	.509	.327	
	여가시간	.015	.143	.917	
인구학적 변수	사회문화활동	-.086	.272	.753	
	성별	.053	.445	.906	
		코호트	.523	.444	.239

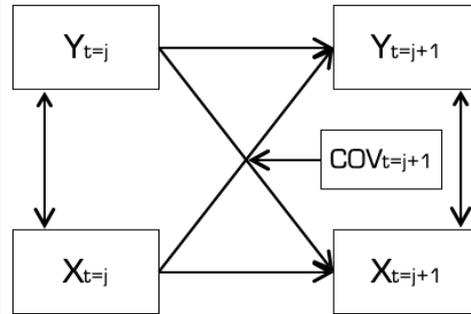
CFI = .964 RMSEA = .019 SRMR = .023 TLI = .945

# 종단 자료, 종단적 구조방정식

Cross-lagged panel model

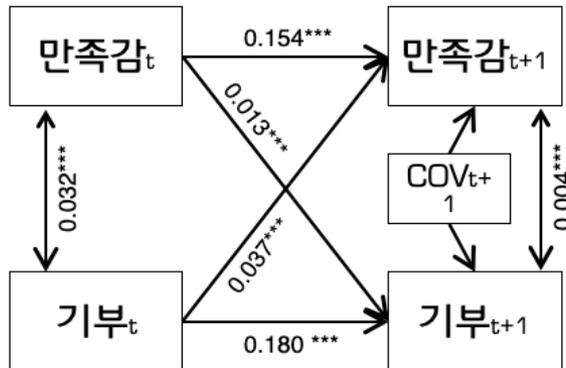


$t = \{1, 2\}$

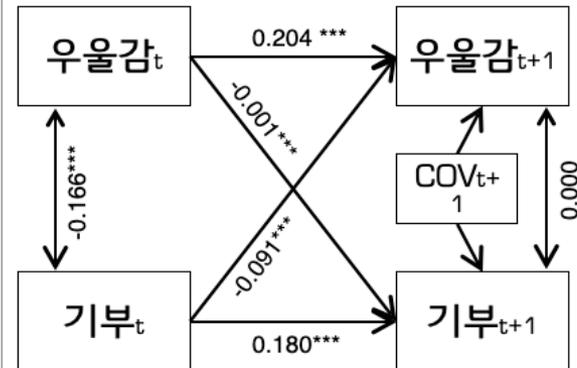


$t = \{1, 2, \dots, j, \dots, k, k+1\}$

모형 1

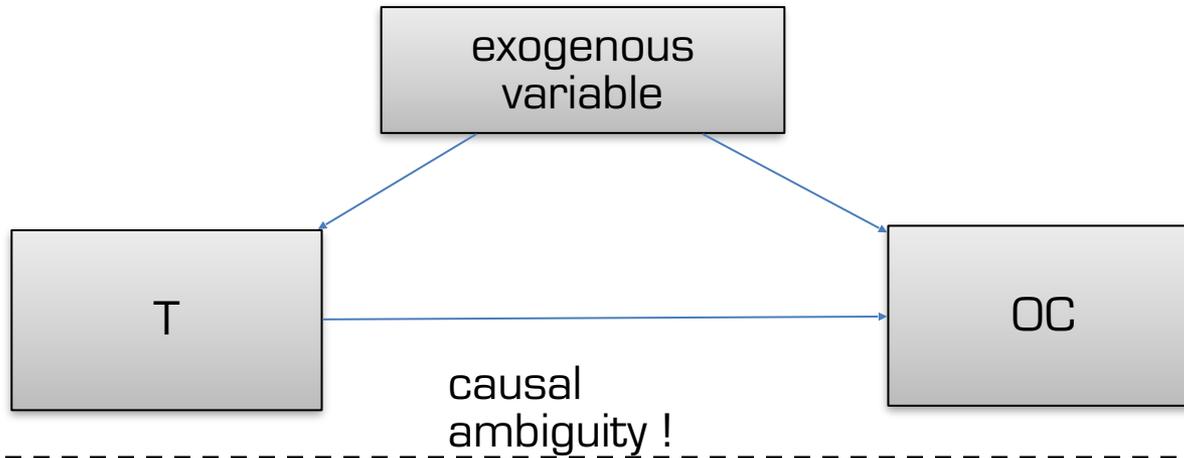


모형 2

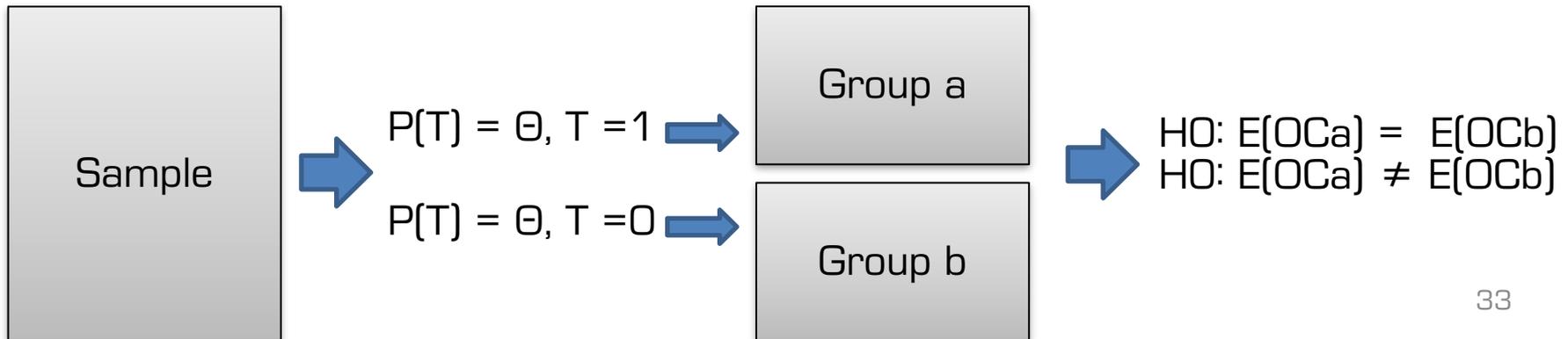


# 집단비교, 성향점수 매칭과의 결합

Non-experimental design



Mimic balanced experiment



**감사합니다!**