콜로키움자료집 19-S32

제22차 NYPI 역량강화 콜로키움

잠재프로파일 분석의 이론과 실제

일 시 2019년 9월 27일(금) 15:00~16:30

장 소 한국청소년정책연구원 7층 대회의실

주 최 한국청소년정책연구원

한국청소년정책연구원
National Youth Policy Institute

잠재프로파일 분석의 이론과 실제

노언경
(서울특별시교육청교육연구정보원 교육정책연구소)

잠재프로파일 분석의 이론과 실제

서울특별시교육청교육연구정보원 교육정책연구소

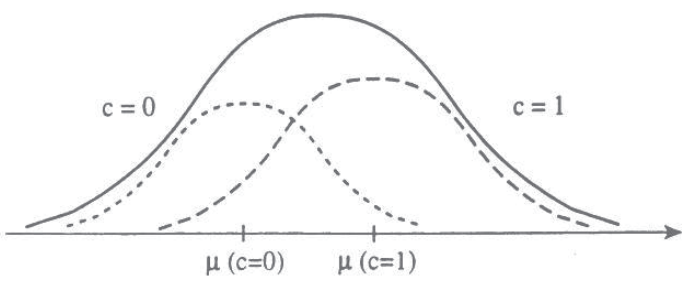
노언경

차 례

- 혼합모형의 기초
- 잠재계층 수 결정 기준
- 잠재계층 분류와 외부변수와의 관계 분석
- 혼합모형 추정의 문제
- 잠재프로파일분석 적용 논문 예시
- Mplus 명령어

혼합모형의 기초

■ 혼합모형 분포의 예



출처: Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marchoulides & R. E. Schmacker (eds.), New Developments and Techniques in Structural Equation Modeling (pp. 1-33). Lawrence Erlbaum Associates.

■ 사람 중심적 접근 vs 변수 중심적 접근 (Bergman & Magnusson, 1997)

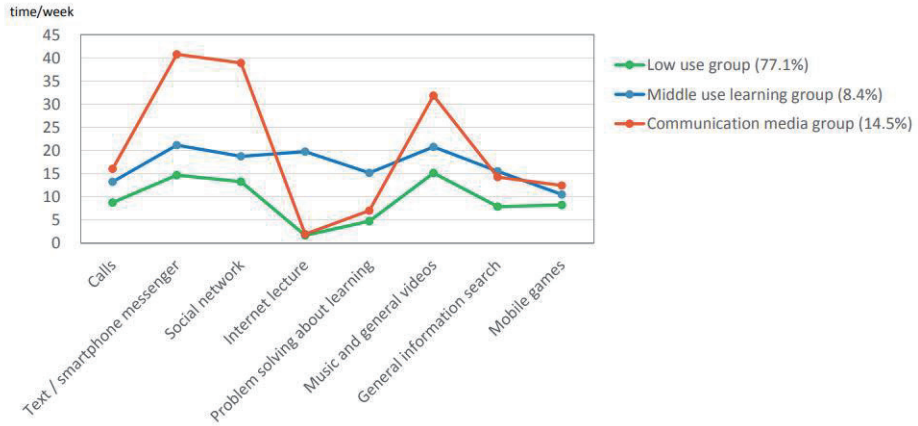
○ 사람 중심적 접근 (Person-centered approach)

개인들 간의 관계 분석
응답에 따라 개인들을 집단으로 유형화
예) 군집분석, 잠재계층분석, 혼합모형

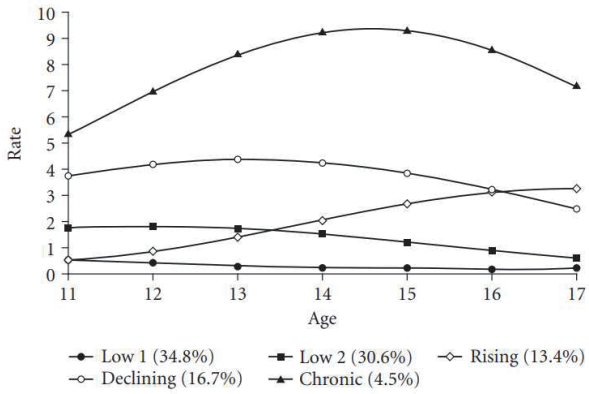
○ 변수 중심적 접근 (Variable-centered approach)

변수들 간의 관계 분석
종속변수와 독립변수 간 관계 설명
예) 회귀분석, 요인분석, 구조방정식 모형

■ 잠재계층 분류의 예 : 휴대전화 사용 목적 (횡단)



■ 잠재계층 분류의 예 : 폭력 비행 변화 (종단)



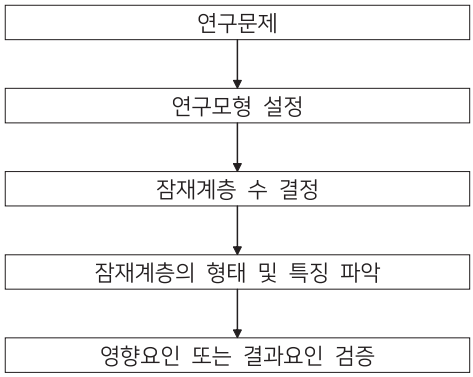
■ 혼합모형의 종류 (Muthén, 2001)

Summary of Techniques Using Latent Classes

Class	*Outcome/ Indicator Scale*	*Number of Time Points*	*Number of Outcomes/ Time Points*	*Within-Class Variation*
LCA	Categorical (u)	Single	Multiple	No
LPA	Continuous (y)	Single	Multiple	No
LCGA	Categorical (u) Continuous (y)	Multiple	Single	No
LTA	Categorical (u)	Multiple	Multiple	No
GMM	Continuous (y)	Multiple	Single Multiple	Yes
GGMM	Categorical (u) Continuous (y)	Multiple	Single Multiple	Yes

LCA – latent class analysis, LPA – latent profile analysis, LCGA – latent class growth analysis, LTA – latent transition analysis, GMM – growth mixture modeling, GGMM – general growth mixture modeling.

■ 혼합모형 연구의 흐름

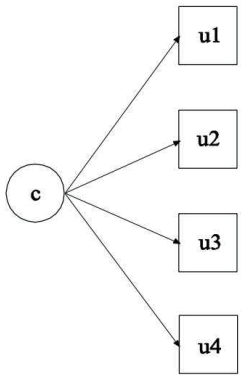


◎ 연구문제

- 잠재계층이 몇 개로 분류되는지
- 분류된 잠재계층은 어떤 형태이며 어떤 특징을 갖는지
- 잠재계층 분류에 영향을 주는 요인은 무엇인지
- 잠재계층별 다른 결과를 나타내는 요인은 무엇인지

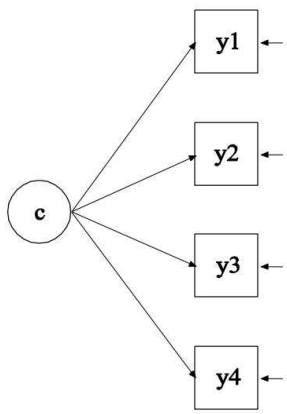
◎ 연구모형

- 잠재계층 분석 (Latent Class Analysis)




```
DATA:      FILE IS [      ].dat;
VARIABLE:  NAMES ARE u1-u4;
           CLASSES = c (2);
           CATEGORICAL = u1-u4;
ANALYSIS:  TYPE = MIXTURE;
```


- 잠재프로파일 분석 (Latent Profile Analysis)




```
DATA:      FILE IS [      ].dat;
VARIABLE:  NAMES ARE y1-y4;
           CLASSES = c (2);
ANALYSIS:  TYPE = MIXTURE;
```

잠재계층 수 결정 기준

■ 통계적 기준 : ① 정보지수 (Information Criteria)

AIC (Akaike IC)

BIC (Bayesian IC)

Sample-Size Adjusted BIC

→ 작을수록 더 좋은 모형

■ 통계적 기준 : ② 모형비교 검증

LMRLRT (Lo, Mendell, Rubin Likelihood Ratio Test)(Lo, Mendell, & Rubin, 2001)

BLRT (Bootstrapped Likelihood Ratio Test)(McLachlan & Peel, 2000)

→ k개 모형과 k-1개 모형의 비교
- 유의하면 k개 모형 선택
- 유의하지 않으면 k-1개 모형 선택

■ 통계적 기준 : ③ 분류의 질

Entropy

→ 0~1 범위, 1에 가까울수록 더 좋은 모형

low 0.4
medium 0.6
high 0.8 (Clark, 2010)

■ 통계적 기준 : ④ 잠재계층 비율

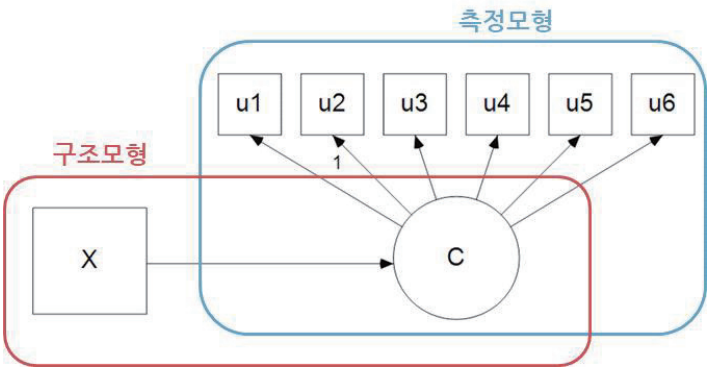
비율이 너무 작은 경우(예. 1% 또는 5% 미만) 집단으로 구분하기 부적절할 수 있음
현상에 따라 적절하다고 판단되는 경우도 있으므로 검토 필요(예. 비행집단)

■ 해석적 기준

→ 이론적 근거, 선행연구 고려

잠재계층 분류와 외부변수와의 관계 분석

■ 외부변수가 포함된 혼합모형
(외부변수: 잠재계층 분류의 영향요인 또는 결과요인)



■ 외부변수가 포함된 혼합모형의 분석방법

◎ 1단계 접근법 (One-step approach) : 최종 연구모형에 포함된 모든 변수들을 동시에 분석

<문제점>
- 변수의 조합이 달라질 때마다 잠재계층 분류가 달라짐
- 응용 연구자의 연구 논리에 적합하지 않음

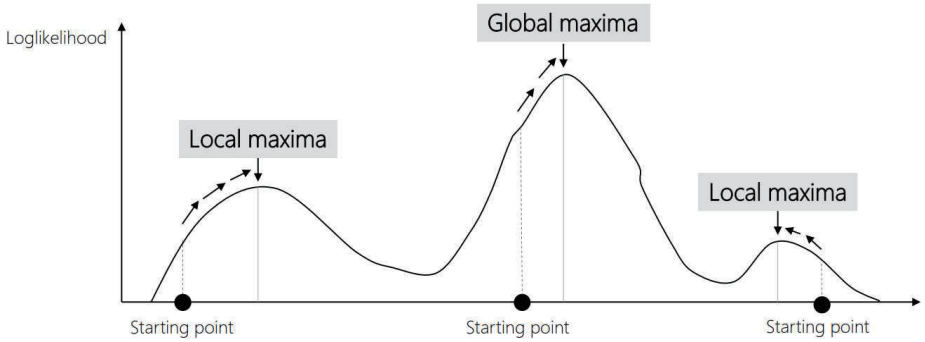
◎ 3단계 접근법 (Three-step approach) : 잠재계층이 분류된 이후에 외부변수와의 관계 분석

- 1단계 : 잠재계층 분류
- 2단계 : 각 개인이 속할 가능성이 가장 높은 잠재계층 추정, 분류오류율 계산하여 할당
- 3단계 : 잠재계층 분류와 외부변수와의 관계 분석

→ 최근 시뮬레이션 연구들에 의하면 3단계 접근법의 추정이 더 우수함

혼합모형 추정의 문제

■ Global maxima와 Local maxima



◎ Global maxima의 예

Loglikelihood	Seed	Initial Stage Starts
-836.899	902278	21
-836.899	366706	29
-836.899	903420	5
-836.899	unperturbed	0
-836.899	27071	15
-836.899	967237	48
-836.899	462953	7
-836.899	749453	33
-836.899	637345	19
-836.899	392418	28

THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED. RERUN WITH AT LEAST TWICE THE RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.

◎ Local maxima의 예

Loglikelihood	Seed	Initial Stage Starts
-835.247	902278	21
-837.132	366706	29
-840.786	903420	5
-840.786	unperturbed	0
-840.786	27071	15
-853.684	967237	48
-867.123	462953	7
-890.442	749453	33
-905.512	637345	19
-956.774	392418	28

WARNING: THE BEST LOGLIKELIHOOD VALUE WAS NOT REPLICATED. THE SOLUTION MAY NOT BE TRUSTWORTHY DUE TO LOCAL MAXIMA. INCREASE THE NUMBER OF RANDOM STARTS.

◎ Local maxima일 때 해결방법 (Muthén & Muthén, 1998-2017)

- Default

```
STARTS = 20 4;
STITERATIONS = 10;
```

- 1번째 제안

```
STARTS = 100 10;
```

- 2번째 제안

```
STARTS = 100 10;
STITERATIONS = 20;
```

- 3번째 제안

```
STARTS = 500 100;
```


잠재프로파일 분석 적용논문 예시

■ 예시논문

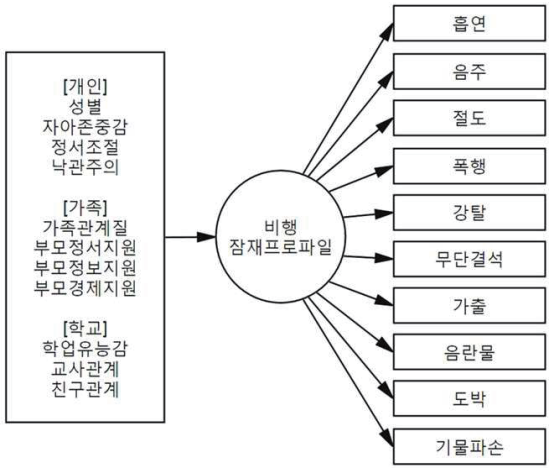
노언경, 정송, 홍세희 (2014). 잠재프로파일 분석을 통한 아동·청소년 비행 유형 분류 및 영향요인 검증. 한국청소년연구, 25(4), 211-240.

◎ 연구문제

1. 아동과 청소년의 비행 유형에 따라 유사한 집단을 유형화할 때 몇 가지 형태로 분류되며, 초등학생과 중학생 집단에서 차이가 있는가?

2. 집단을 유형화하는 데 있어서 개인, 가족, 학교 관련 변수들 중에 영향을 미치는 변수는 무엇인가?

◎ 연구모형

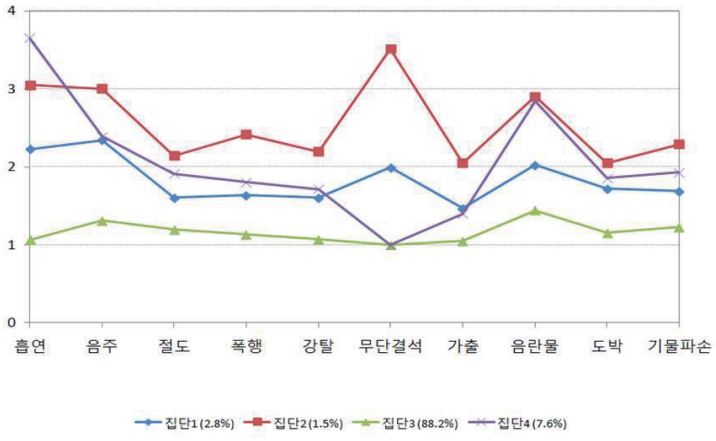


◎ 연구결과 : 집단 수 결정

분류 기준		잠재프로파일의 수			
		2개	3개	4개	5개
분류의 질	Entropy	.998	.998	.995	.986
정보지수	AIC	43629.3	40722.5	36409.2	37279.0
	BIC	43878.6	41102.3	36919.6	37920.0
	SABIC	43745.2	40899.0	36646.4	37576.9
모형비교 검증*	LMRLRT	.000	.038	.608	.256
	BLRT	.000	.000	.000	1.000

* LMRLRT와 BLRT는 p값을 제시하였음.

◎ 연구결과 : 분류된 잠재프로파일의 형태



◎ 연구결과 : 분류된 잠재프로파일의 특성

영역		중간수준	고수준/무단결석	저비행	흡연/음란물
개인	성별	0.615	0.732	0.517	0.806
	자아존중감	2.926	2.762	3.085	2.924
	정서조절	2.482	2.309	2.681	2.366
	낙관주의	2.718	2.601	2.869	2.700
가족	가족관계 질	2.915	2.746	3.228	2.943
	부모정서지원	2.782	2.870	3.141	2.842
	부모정보지원	2.821	2.854	3.141	2.935
	부모경제지원	3.224	3.146	3.394	3.276
학교	학업유능감	2.393	2.292	2.840	2.431
	교사관계	2.483	2.382	2.615	2.389
	친구관계	3.214	3.065	3.342	3.335

◎ 연구결과 : 잠재프로파일 분류의 영향요인

영역		집단1 vs 집단2		집단1 vs 집단3		집단1 vs 집단4		집단2 vs 집단3		집단2 vs 집단4		집단3 vs 집단4	
		계수	표준오차	계수	표준오차	계수	표준오차	계수	표준오차	계수	표준오차	계수	표준오차
개인	성별	0.510	0.439	-0.459	0.250	1.095 ***	0.308	-0.968 **	0.369	0.585	0.406	1.553 ***	0.192
	자아존중감	-0.160	0.358	-0.133	0.223	-0.035	0.277	0.027	0.296	0.125	0.336	0.098	0.193
	정서조절	-0.381	0.462	0.050	0.222	-0.552 *	0.270	0.431	0.413	-0.172	0.438	-0.602 **	0.175
	낙관주의	0.118	0.437	-0.132	0.255	0.120	0.309	-0.249	0.366	0.002	0.402	0.251	0.197
가족	가족관계 질	-0.815 *	0.371	0.184	0.262	-0.066	0.308	1.000 ***	0.277	0.749 *	0.318	-0.251	0.183
	부모정서지원	0.873 *	0.426	0.391	0.257	-0.020	0.303	-0.482	0.351	-0.893 *	0.385	-0.411 *	0.183
	부모정보지원	0.208	0.354	0.095	0.220	0.257	0.256	-0.113	0.284	0.049	0.311	0.162	0.147
	부모경제지원	-0.210	0.362	-0.024	0.235	-0.006	0.274	0.186	0.291	0.204	0.317	0.018	0.161
학교	학업유능감	-0.073	0.408	1.155 ***	0.241	0.308	0.271	1.227 ***	0.344	0.381	0.361	-0.846 ***	0.146
	교사관계	-0.070	0.363	-0.371	0.191	-0.300	0.210	-0.301	0.317	-0.230	0.328	0.071	0.107
	친구관계	-0.064	0.294	-0.067	0.196	0.420	0.264	-0.003	0.233	0.484	0.284	0.487 *	0.198

주1. 집단1=중간수준 집단, 집단2=고수준/무단결석 집단, 집단3=저비행 집단, 집단4=흡연/음란물 집단

주2. 각 집단별 비교 시 앞에 있는 집단이 참조집단(reference group)임.

*p<.05, **p<.01, ***p<.001

Mplus 명령어

■ 기본 명령어

TITLE: // 제목

DATA: FILE IS _____.dat; // 불러올 파일명

VARIABLE: NAMES ARE __ __ __ __ __ __ __; // 자료의 변수명
USEVARIABLES = __ __; // 사용할 변수명

ANAYSIS: // 분석방법

MODEL: // 연구모형

OUTPUT: // 결과

MODEL:

BY	short for measured by -- defines latent variables example: f1 BY y1-y5;
ON	short for regressed on -- defines regression relationships example: f1 ON x1-x9;
@	fixes a parameter at a default value or a specific value example: y1@ y2@0;

	Growth Language	Alternative
Intercept only	MODEL: i | y1-y4@1;	MODEL: i BY y1-y4@1; [y1-y4@0 i];
Linear	MODEL: i s | y1@0 y2@1 y3@2 y4@3;	MODEL: i BY y1-y4@1; s BY y1@0 y2@1 y3@2 y4@3; [y1-y4@0 i s];
Linear with free time scores	MODEL: i s | y1@0 y2@1 y3 y4;	MODEL: i BY y1-y4@1; s BY y1@0 y2@1 y3 y4; [y1-y4@0 i s];
Quadratic	MODEL: i s q | y1@0 y2@1 y3@2 y4@3;	MODEL: i BY y1-y4@1; s BY y1@0 y2@1 y3@2 y4@3; q BY y1@0 y2@1 y3@4 y4@9; [y1-y4@0 i s q];

■ 혼합모형 분석을 위한 명령어

VARIABLE:
CLASSES = c(4); // 잠재계층 수
AUXILIARY = _____ (R3STEP); // Three-step approach

- predictor : R3STEP (Vermunt, 2010; Asparouhov & Muthén, 2012b)
- distal outcome : BCH (Bakk & Vermunt, 2015), DCATEGORICAL (Lanza et al., 2013)

ANALYSIS:
TYPE=MIXTURE; // 혼합모형 분석

MODEL:
%OVERALL% // 전체 연구모형

%c#1% // 잠재계층별 연구모형

```
OUTPUT:  TECH11 TECH14;  // LMRLRT, BLRT 분석결과

SAVEDATA: FILE IS _____.dat;  // 저장할 파일명
SAVE=CPROB;  // 개인별 잠재계층 사후확률 저장

PLOT: TYPE IS PLOT3;  // 그래프
series is g1-g5 (*);
```

MEMO