

전남대학교 교육학과 BK21 플러싱 교육 연구단 2024년 상반기 연구분석방법 온라인 워크숍

교육학 연구를 위한 잠재계층모형(latent class model)의 이해와 분석 실습

장창성, Ph.D.

한양대학교 한국교육문제연구소 박사후연구원

순서



잠재계층모형 이해



잠재계층모형 활용



잠재계층분석 실습



질의응답

강연자 소개

연구영역 & 주제 | RESEARCH INTEREST & TOPIC



고등·평생교육 정책 및 평가



성인교육훈련(adult education and training) 및 일 관련 학습(work-related learning)



학교에서 노동시장(school-to-work) 및 노동시장에서 학교(work-to-school)로의 전환



고등교육기관(higher education institutions)의 조직개발 및 학습

연구방법 | METHODOLOGY



고등·평생교육 분야 대용량 데이터 분석



교육연구 머신러닝(machine learning) 알고리즘 기법 활용



교육연구 신경과학/뇌기반 접근 적용

잠재계층모형 이해

잠재계층모형 이해

개념 및 종류

■ 잠재계층모형의 개념

- 분석대상의 응답패턴(response pattern)을 기반으로 개인 및 집단의 특성을 실증적으로 분류하는 혼합모형기법(mixture modeling technique)의 일종
- 모집단 내 분포와 크기가 다른 하위집단의 존재를 가정하며, (1)데이터 내 숨겨진(hidden) 패턴을 식별하고, (2)식별된 잠재집단과 관측 변수 간의 연관성 이해

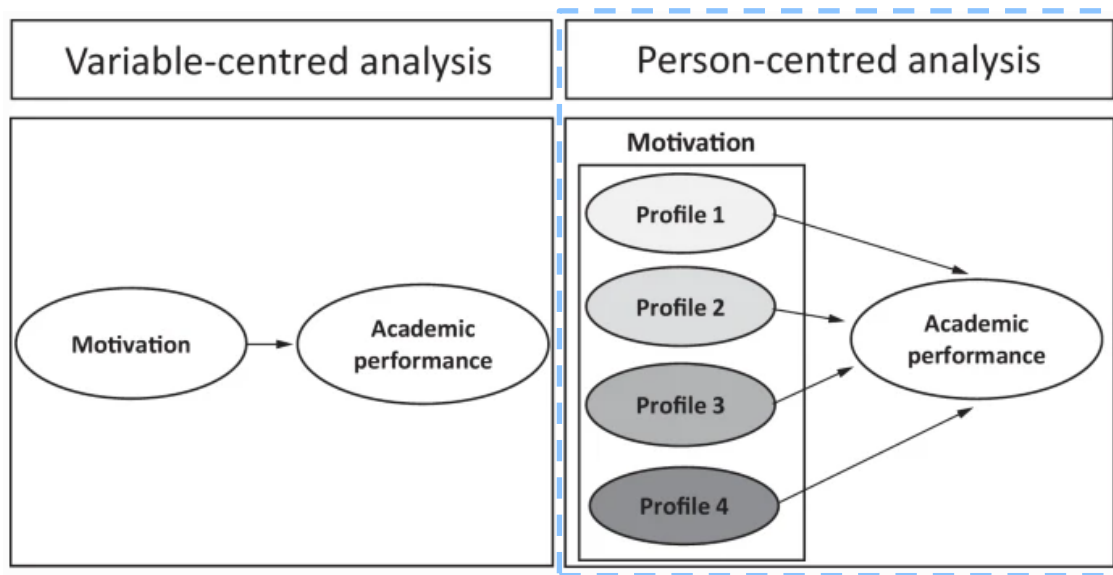
■ 잠재계층모형의 종류

- | | | |
|--------------------------------------|---|------|
| ○ 잠재계층분석(latent class analysis) | } | 횡단분석 |
| ○ 잠재프로파일분석(latent profile analysis) | | |
| ○ 잠재전이분석(latent transition analysis) | } | 종단분석 |
| ○ 잠재성장분석(latent growth analysis) | | |

잠재계층모형 이해

주요용어 및 분석원리

- 변수중심적 접근 *vs.* 인간중심적 접근(Kusurkar et al., 2021)
 - 변수중심적(variable-centered) 접근: 상관분석 및 회귀분석 등은 변수와 변수 간의 관계를 하나의 요약치(예: mean value)로 설명
 - 인간중심적(person-centered) 접근: 개인의 이질성에 근거하여 집단을 분류하고 관련 변수들의 영향력을 파악



잠재계층모형 이해

주요용어 및 분석원리

- **이질성(heterogeneity)**
 - 특정 변수 및 응답 패턴에 있어 모집단 내 개인 및 집단 간 이질적인 연관성을 의미하며, 관측 변수에 대한 이질적 분포를 문항반응확률(item response probability)에 따라 추정
 - 이질적 응답 패턴을 통해 개인의 경험, 가치 및 인식 등에 따라 모집단 내 잠재집단을 통계적으로 분류(오영교, 차성현, 2018)
- **집단 소속 확률(membership probability)**
 - 변수와 변수 간의 관계는 특정 개인 및 집단의 응답/행동 패턴에 따라 다르게 나타날 수 있으며, 각 집단에 속할 확률을 추정함으로써 집단 간 차이 분석(신태수, 2010)

잠재계층모형 이해

잠재계층모형 활용 장점

- 통계적 활용의 용이성

- 사전에 결정되지 않은 적합한 집단의 개수나 크기를 변수의 유형 및 척도와 관계 없이 세분화 가능하며, 문항반응확률에 따라 유사한 응답/행동 패턴을 공유하는 집단의 개수와 크기 결정
- 인구통계학적 변수와 같은 다양한 외생변수(exogenous variable)를 포함하는 확장된 형태의 분석 가능

- 집단 분류의 통계적 엄밀성

- 잠재집단의 개수를 결정하는데 있어 모형 적합도(model fit) 지수를 활용하며, 최적의 잠재집단 식별 및 분류의 객관성과 정확성 확보
- 모형 적합도 지수: AIC(Akaike information criterion), CAIC(consistent Akaike Information Criterion), BIC(Bayesian Information Criterion), BLRT(parametric bootstrapped likelihood ratio test), entropy 등

잠재계층모형 이해

분석절차(Sinha et al., 2022; Weller et al., 2020)

1. 연구설계(study design)

- 관측지표(indicator) 선정: 잠재계층모형에서 관측지표는 집단을 분류하는 주요 변인을 의미하며, 연구목적 및 질문에 부합하는 관측지표 선정

2. 데이터 설정(data set-up)

- 데이터 구조 파악: 전체 표본 크기, 관측 변수의 분포 및 결측치 등을 검토하며, 일반적으로 $N = 200$ 이상의 샘플 수 권장
- 결측치 처리: 잠재집단 식별 및 정확도 추정은 극단치(outlier), 데이터 누락 등에 의해 더 크게 영향을 받을 수 있음

잠재계층모형 이해

분석절차 (Sinha et al., 2022; Weller et al., 2020)

3. 모형 적합도 분석(fitting models to the data)

- k 집단 수를 순차적으로 늘리면서 다수의 모형을 구성하고, 데이터 구조에 적합한 최적의 집단 수 결정을 위한 모형 적합도 지수 산출
- 모형의 복잡성이 증가함에 따라(즉, 집단 수가 늘어남에 따라) 각 집단의 관측치 수는 적어지며, 일반화 가능성이 감소할 수 있음 → 관측치 수가 가장 적은 집단의 $n = 50$ 또는 전체 샘플의 5% 미만일 경우 권장되지 않음

4. 모형 평가(evaluating the models)

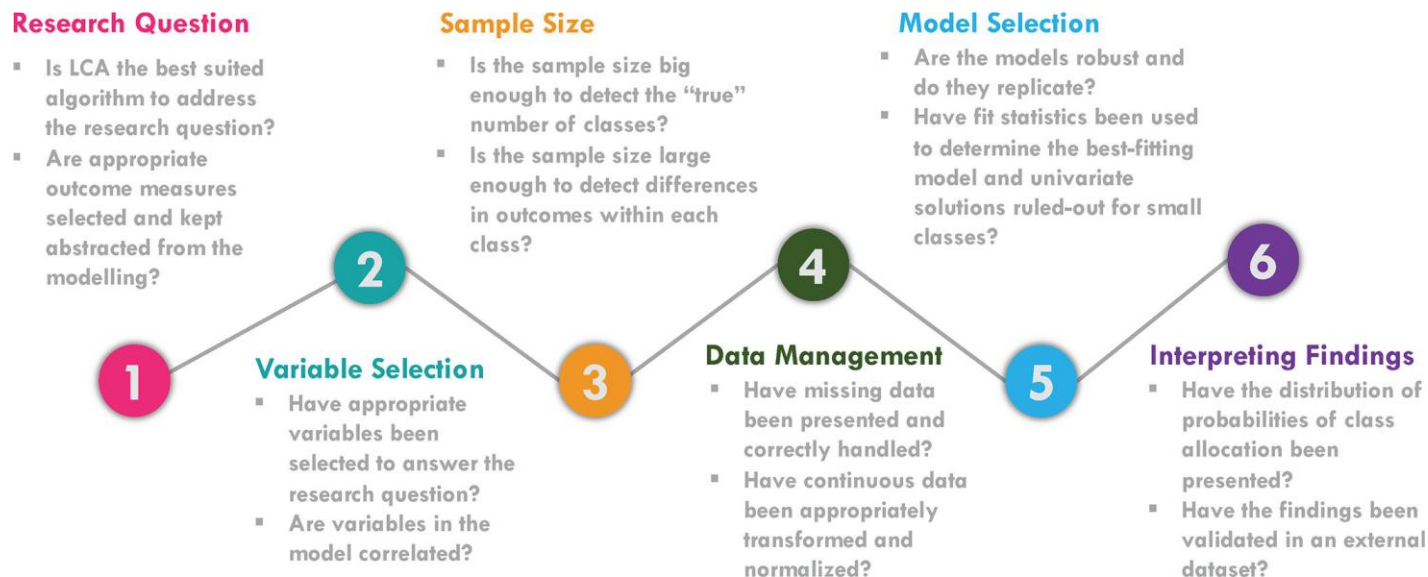
- 산출된 모형 적합도 지수(AIC, CAIC, BIC, BLRT, entropy 등)에 근거하여 최적의 잠재집단 수 결정
- 무엇이 최적의 모형인가? *Note that best fit does not always mean a good fit* – 집단 분류에 따른 해석가능성(interpretability) 고려 필요

잠재계층모형 이해

분석절차(Sinha et al., 2022; Weller et al., 2020)

5. 최종 모형 해석(interpreting the final model)

- 집단 비교: 각 집단의 관측지표별 사후 확률(posterior probability)값에 근거하여 집단별 특성 파악
- 외부 검증: 잠재계층모형에 외생변수를 투입하여 사후분석을 실시하며, 각 집단에 소속될 확률에 영향을 미치는 예측요인과 성과변수 등과의 연관성을 바탕으로 연구결과 해석



잠재계층모형 이해

분석 프로그램

- **Mplus**(Muthén & Muthén, 2000): <https://www.statmodel.com/>
- **RStudio**(R Core Team, 2021)
 - poLCA 패키지: <https://cran.r-project.org/web/packages/poLCA/poLCA.pdf>
 - mclust 패키지: <https://mclust-org.github.io/mclust/>
- **jamovi**(The jamovi Project, 2022)
 - snowLatent 모듈: <https://github.com/hyunsooseol/snowLatent>

잠재계층모형 활용

잠재계층모형 활용

연구배경: 연구목적 및 내용

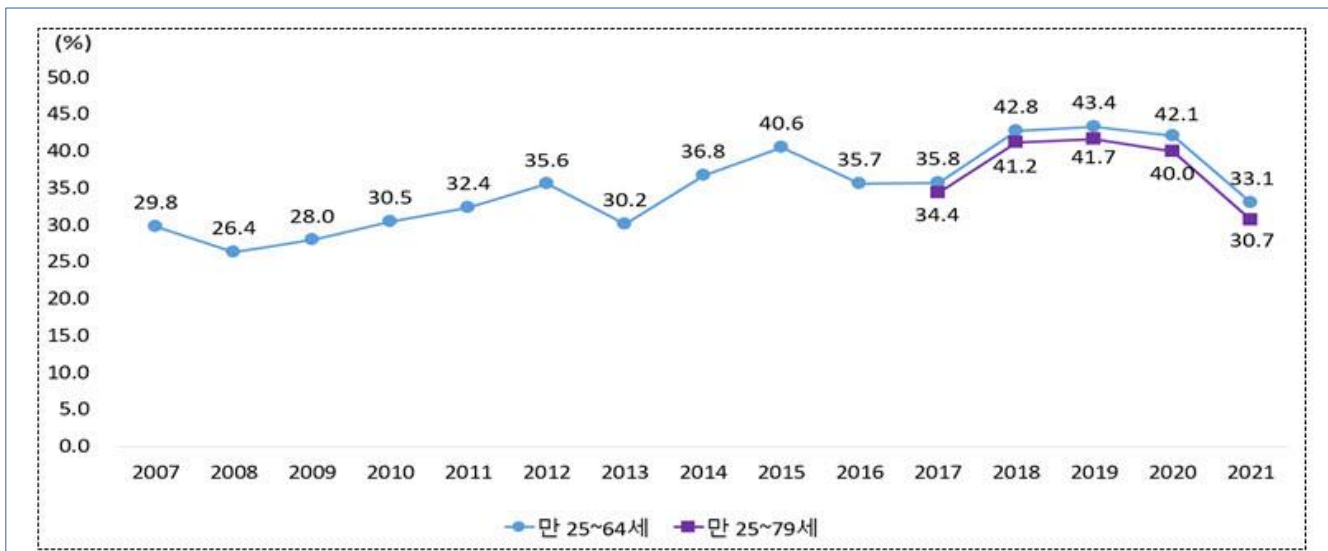
- **연구주제:** 장창성, 전하람, 김재형(2023). 교육수준에 따른 재직자의 평생학습 참여 유형 및 성과분석: 잠재계층분석. *평생교육학연구*, 29(4), 149–180. <http://doi.org/10.52758/kjle.2023.29.4.149>
- **연구목적 및 기대효과**
 - 우리나라 재직자의 평생학습 참여 유형을 파악하고 각 유형별 평생학습 참여 예측요인 및 성과의 차이를 분석함으로써 재직자 평생학습 참여 활성화를 위한 실증적인 근거 제시
 - 재직자 집단별 맞춤형 평생학습 프로그램 운영 및 관련 정책 개발 등 평생학습 활성화를 위한 구체적인 실천방안 논의
- **연구문제**
 1. 교육수준에 따른 재직자의 평생학습 참여 유형은 어떻게 분류될 수 있는가?
 2. 교육수준에 따른 재직자의 평생학습 참여 유형 분류에 영향을 미치는 요인은 무엇인가?
 3. 교육수준에 따른 재직자의 평생학습 참여 유형별 평생학습 참여 성과에 어떠한 차이가 있는가?

잠재계층모형 활용

연구배경: 연구 필요성

▪ 고용노동시장의 변화와 평생학습

- 생애초기 단선적인 성장경로만으로는 안정적인 경제적 지위를 보장할 수 없으며, 이에 따라 지속적인 경력개발 및 전문성 향상을 위한 기술 숙련 및 재교육 등 직업관련 평생학습의 중요성 강조(최동선, 2021)
- 2000년대 중반 이후 평생학습 참여율의 지속적 증가

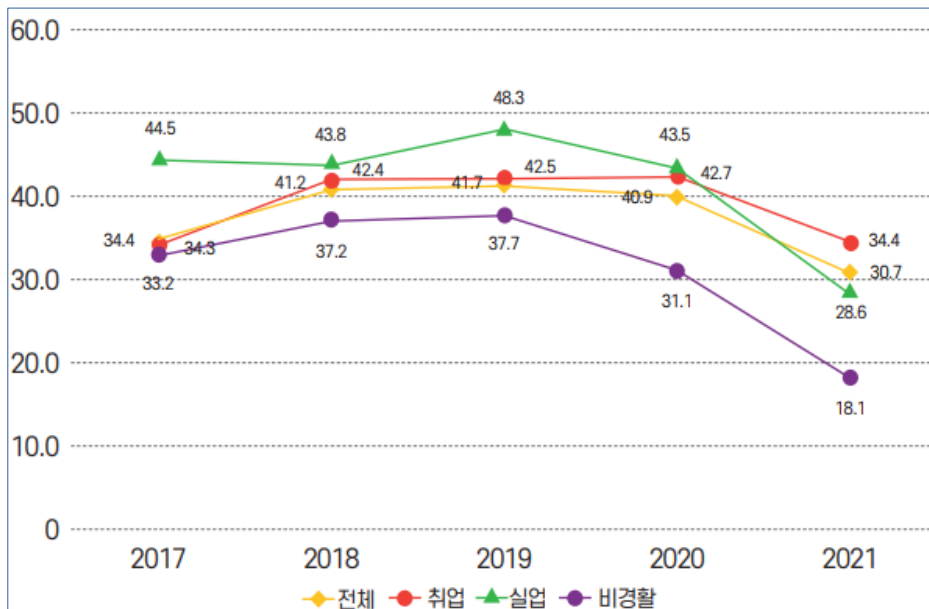


[그림 1] 연도별 평생학습참여율(2007-2021)

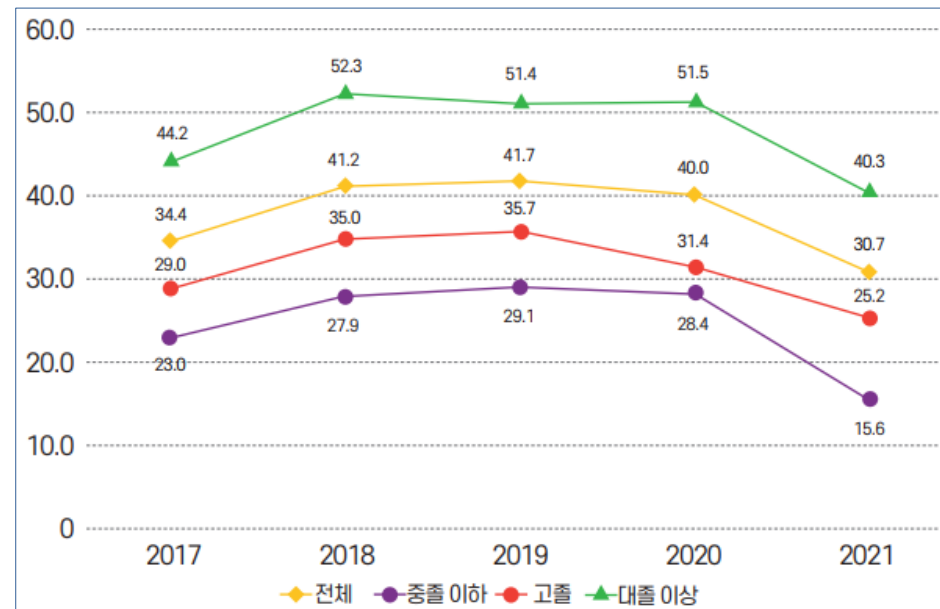
잠재계층모형 활용

연구배경: 연구 필요성

- 재직자 교육수준에 따른 평생학습 참여 격차
 - 재직자의 참여율은 34.4%로, 전체 인구 평균(30.4%)에 비해 높은 수준이나, 교육수준 등 집단간 차이에 의해 평생학습 참여 격차 심화
 - 계층간 불균형한 교육기회 및 자원 배분에 따른 평생학습 참여 및 성과의 불평등 시사(채창균, 2021)



[그림 2] 경제활동상태별 평생학습 참여율(2017-2021)



[그림 3] 학력별 평생학습 참여율(2017-2021)

잠재계층모형 활용

연구방법 및 연구변인

- **분석자료:** 2022년 한국교육개발원 평생학습 개인실태조사 데이터
- **연구대상:** 9,968명의 전체 표본 중 (1)만 65세 이상이고, (2)경제활동에 종사하지 않으며, (3)비형식교육 또는 무형식학습에 참여한 경험이 없는 응답자 3,305명을 제외한 총 6,663명
- **관측변인(indicator)**
 - 학력: 이분형 더미 변수(0 = 대졸 미만, 1 = 대졸 이상)
 - 비형식교육: 직업능력향상교육, 인문교양교육, 문화예술스포츠교육, 시민참여교육 등 참여 여부를 측정하는 이분형 더미 변수(0 = 비참여, 1 = 참여)
 - 무형식학습: 전체 11개의 유형 중 응답률이 높은 6개 이분형 더미 변수(0 = 비참여, 1 = 참여)
- **성과변인(distal outcomes)**
 - 삶의 질 향상 관련 2개 하위변수: (1)사회참여 만족도, (2)경제적 안정감 등(1 = 전혀 도움이 안됨, ... 5 = 매우 도움이 됨)

잠재계층모형 활용

연구방법 및 연구변인

- 공변인(covariates): 인구배경학적 요인, 개인수준 요인 및 직업관련 요인



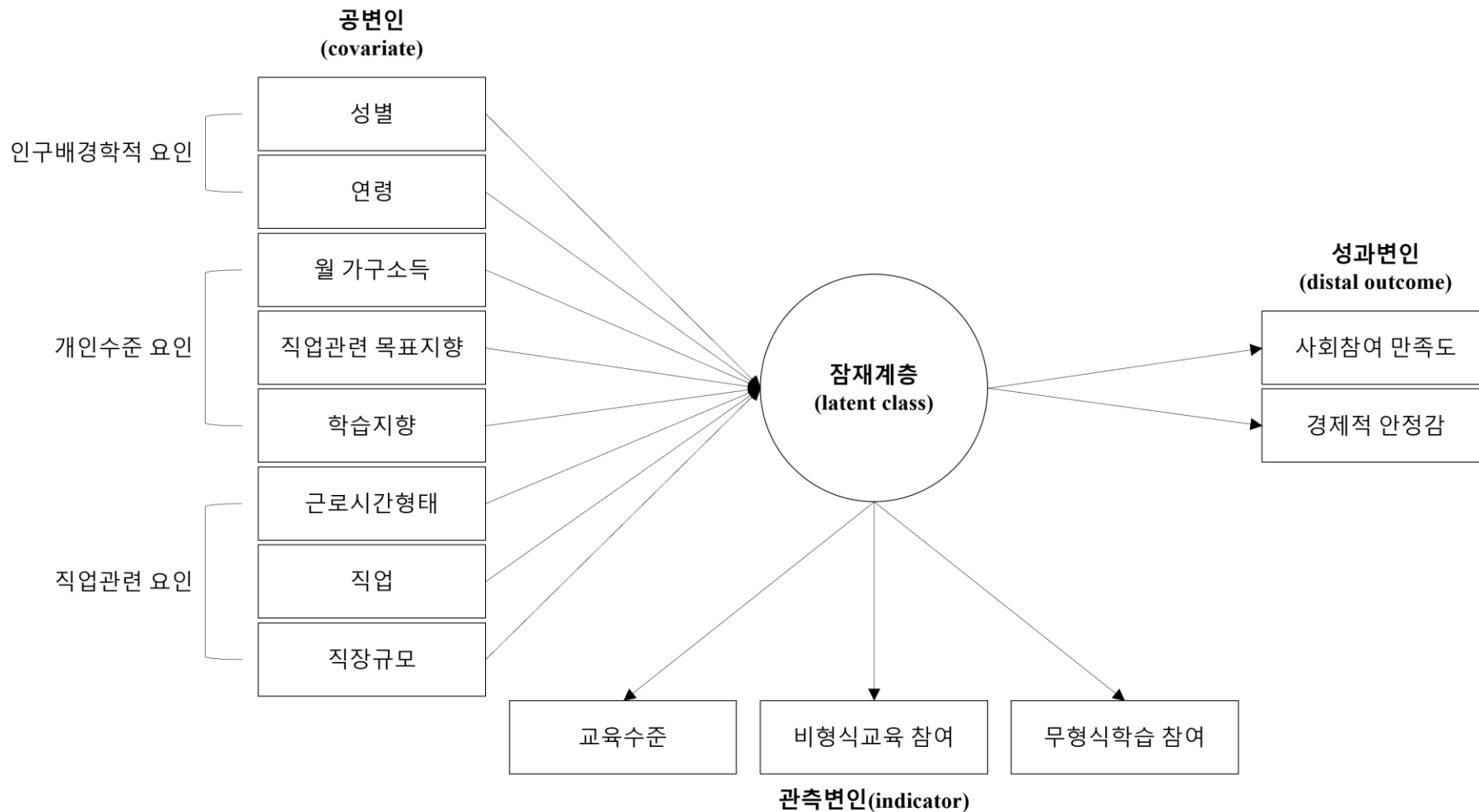
변수명	측정
성별	0 = 여성, 1 = 남성
연령	1 = 25-34세, ... 4 = 55-64세
월 가구소득	1 = 150만원 미만, ... 4 = 500만원 이상
직업관련 목표지향	5점 리커트 척도
학습지향	5점 리커트 척도



변수명	측정
근로시간형태	0 = 시간제, 1 = 전일제
직업	0 = 비사무직, 1 = 사무직
직장규모	1 = 10명 이하, ... 5 = 300명 이상

잠재계층모형 활용

연구모형



잠재계층모형 활용

분석방법: 잠재계층분석

▪ 분석원리

- 혼합모형(mixture modeling) 기법의 하나로, 모집단 내에 분포와 크기가 다른 하위집단이 존재한다는 가정 하에 수행(Muthén & Muthén, 2000)
- 특정 문항에 대해 유사한 응답패턴을 보이는 표본을 동일한 집단으로 분류하고, 각 집단에 속할 확률(membership probability)을 추정함으로써 모집단 내 집단 간 차이를 분석

▪ 분석절차: 3단계 접근법(3-step approach)

- 1단계: 잠재집단 분류 및 식별(무조건모형) – 정보지표(AIC, CAIC, BIC 등), χ^2 차이 검증(BLRT p -value), 분류의 질(entropy) 등 고려하여 최적의 집단 수 결정
- 2단계: 무조건모형에 공변인을 투입하여 잠재집단 분류에 영향을 미치는 예측요인 탐색
- 3단계: 성과변인을 활용하여 식별된 잠재집단 유형별 차이 검증

잠재계층모형 활용

연구결과: 교육수준에 따른 평생학습 참여 잠재집단 분류

- 분석결과: 4개의 잠재집단 식별

<표 1> 교육수준에 따른 평생학습 참여 잠재계층 분류 기준

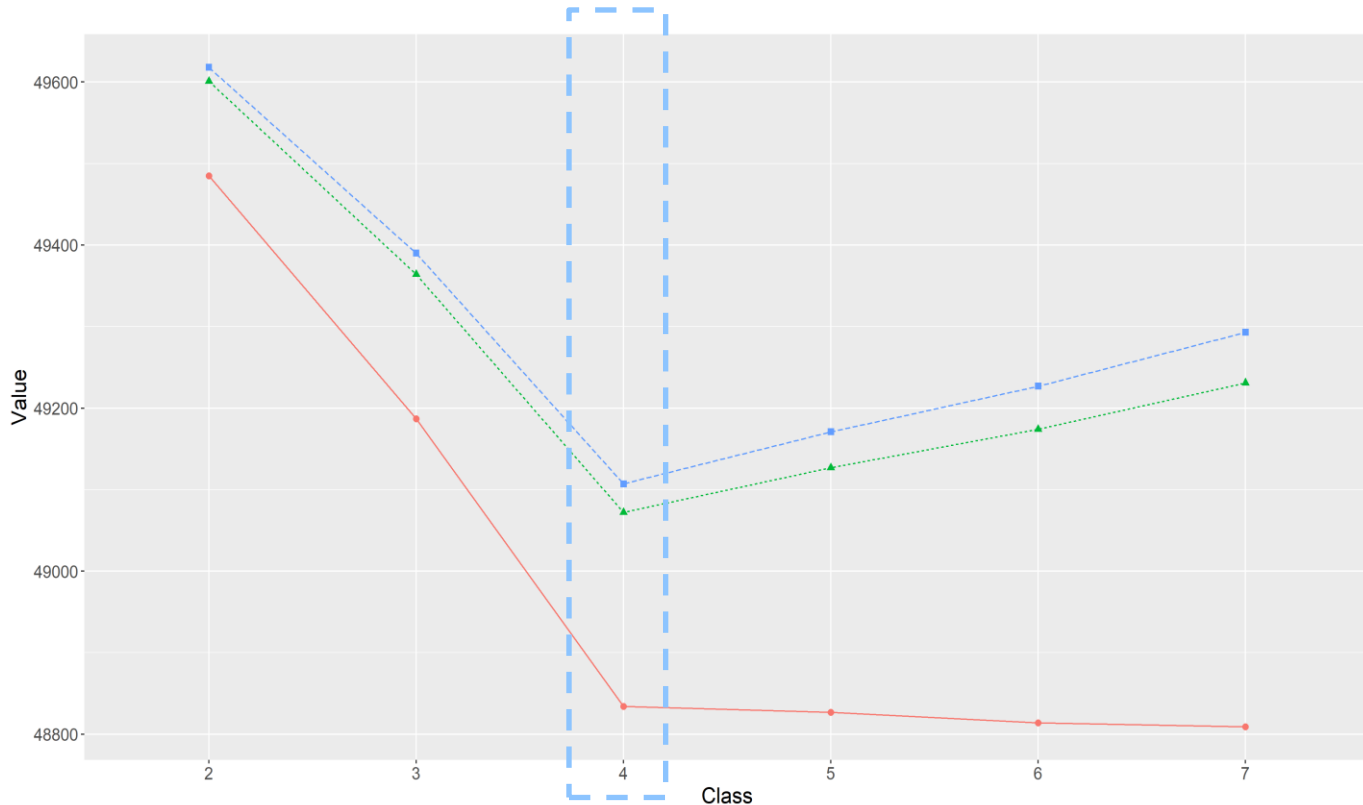
잠재계층 수	정보지표				χ^2 차이 검증	분류의 질
	log-likelihood	AIC	CAIC	BIC	BLRT(p)	entropy
2	-24726	49485	49618	49601	.000	.661
3	-24567	49187	49390	49364	.000	.733
4	-24382	48834	49107	49072	.000	.689
5	-24370	48827	49171	49127	.000	.664
6	-24354	48814	49227	49174	.020	.686
7	-24342	48809	49293	49231	.160	.607

- BLRT p 값 기준: 잠재계층의 수가 6개인 모형과 5개인 모형을 비교할 때까지는 χ^2 차이가 유의미한 것으로 나타났으나, 그 이후로는 유의미하지 않음
- entropy 기준: 잠재계층 수가 3개일 경우 .733, 4개일 경우 .689

잠재계층모형 활용

연구결과: 교육수준에 따른 평생학습 참여 잠재집단 분류

- 분석결과: 4개의 잠재집단 식별



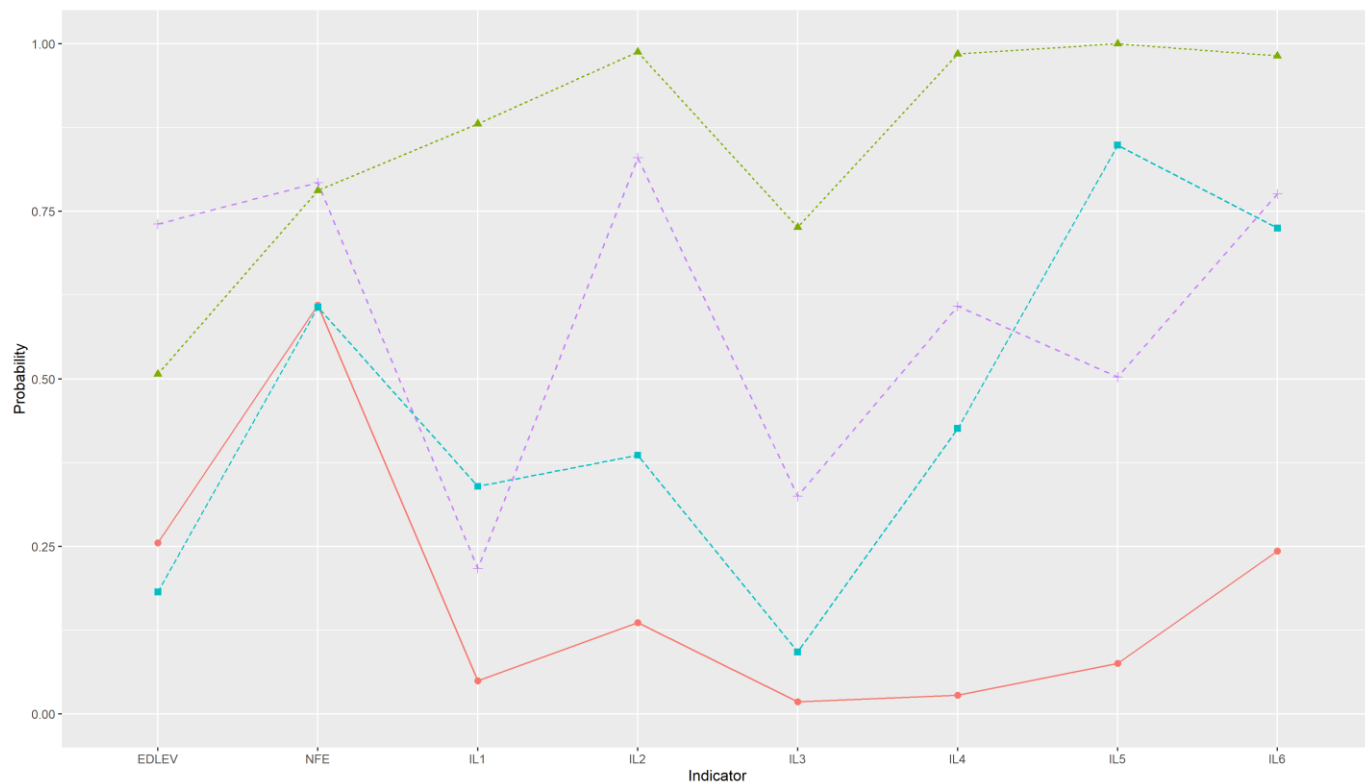
[그림 3] 교육수준에 따른 평생학습 참여 잠재계층 수 증가에 따른 정보지표 변화

- 스크리 도표(scree plot)의 기울기가 완만해지거나 급격히 변화하는 구간에서 요인의 수를 결정
- AIC: 잠재계층의 수가 4개로 증가할 때까지는 큰 폭으로 감소하지만 이후로는 큰 변화를 보이지 않음
- CAIC 및 BIC: 잠재계층의 수가 4개로 증가할 때까지는 지속적으로 감소하다가 이후 증가

잠재계층모형 활용

연구결과: 교육수준에 따른 평생학습 참여 잠재집단 분류

- 분석결과: 4개의 잠재집단 식별



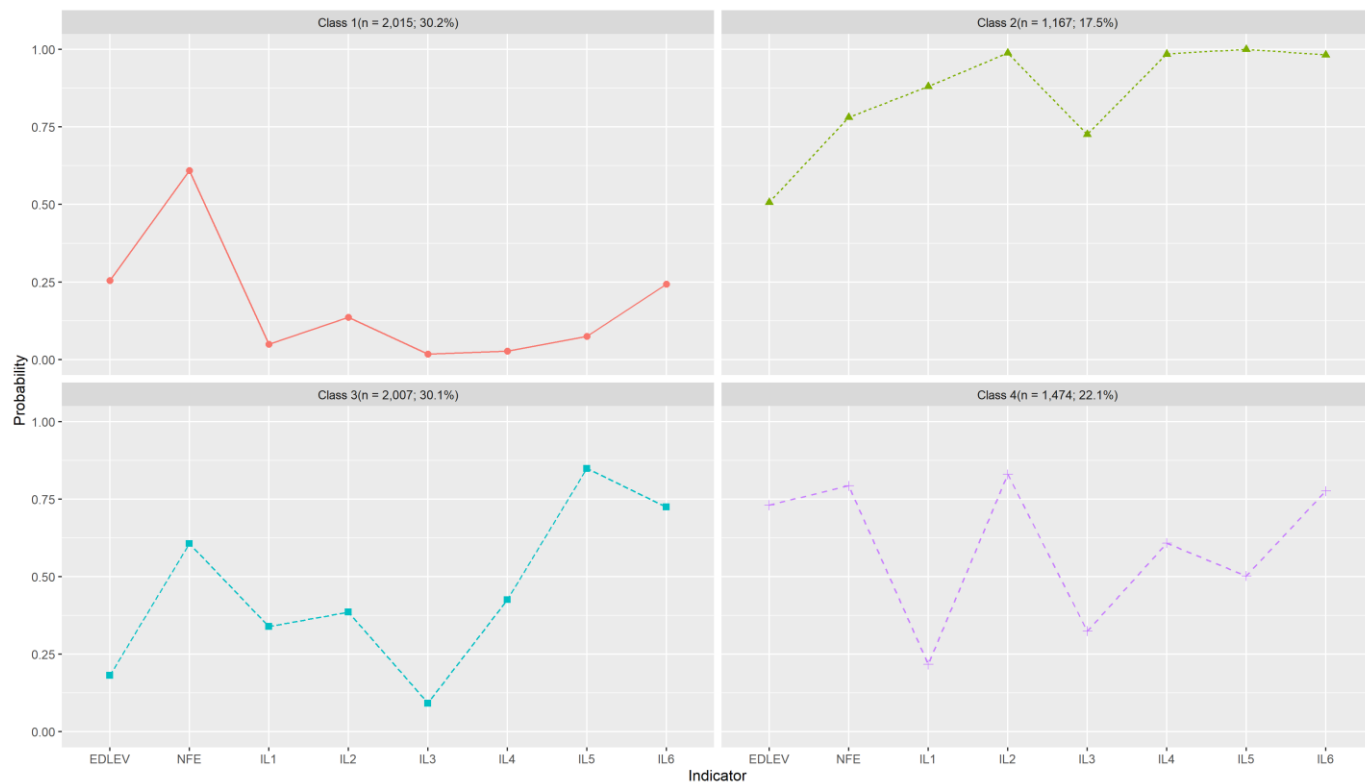
- **Class 1:** 저학력-저참여 ($n = 2,015$; 30.2%)
- **Class 2:** 고학력-고참여 ($n = 1,167$; 17.5%)
- **Class 3:** 저학력-전통적 학습방법 의존형 ($n = 2,007$; 30.1%)
- **Class 4:** 고학력-매체기반형 ($n = 1,474$; 22.1%)

[그림 4] 교육수준에 따른 평생학습 참여 유형

잠재계층모형 활용

연구결과: 교육수준에 따른 평생학습 참여 잠재집단 분류

■ 분석결과: 4개의 잠재집단 식별



- **Class 1:** 저학력-저참여 ($n = 2,015$; 30.2%)
- **Class 2:** 고학력-고참여 ($n = 1,167$; 17.5%)
- **Class 3:** 저학력-전통적 학습방법 의존형 ($n = 2,007$; 30.1%)
- **Class 4:** 고학력-매체기반형 ($n = 1,474$; 22.1%)

[그림 5] 교육수준에 따른 평생학습 참여 유형

잠재계층모형 활용

연구결과: 교육수준에 따른 평생학습 참여 유형과 영향요인

- 평생학습 참여 유형과 예측변수들 간의 연관성 분석: 다항 로지스틱 회귀분석

<표 2> 교육수준에 따른 평생학습 참여 유형 영향요인

요인	변수	준거집단: 집단1				준거집단: 집단3			
		집단2		집단4		집단2		집단4	
		b	OR	b	OR	b	OR	b	OR
인구배경학적 요인	성별	.071	1.068	.305**	1.263	.058	1.056	.292**	1.254
	연령	-.327***	.614	-.732***	.921	-.394***	.517	-.799***	.776
개인수준 요인	월 가구소득	-.033	.967	.305***	1.263	-.073	.924	.265***	1.233
	직업관련 목표지향	1.084***	1.662	.160	1.148	.949***	1.613	.026	1.025
	학습지향	.020	1.020	.413***	1.339	-.027	.972	.366**	1.307
직업관련 요인	근로시간형태	-.313	.633	.282	1.246	-.017	.982	.578**	1.439
	직업	-.054	.945	.740***	1.523	.166	1.153	.959***	1.617
	직장규모	.093*	1.089	.137**	1.128	.156***	1.145	.200***	1.182

* $p < .05$, ** $p < .01$, *** $p < .001$; OR = odds ratio

주. 집단 1: 저학력-저참여, 집단 2: 고학력-고참여, 집단 3: 저학력-전통적학습방법의존형, 집단 4: 고학력-매체기반형

- **고학력-고참여**: 연령($b = -.327, p < .001$)이 낮고, 직업관련 목표지향 참여동기($b = 1.084, p < .001$)가 높으며, 더 큰 규모의 직장($b = .093, p < .05$)에 재직하고 있을 확률이 높음
- **고학력-매체기반 학습자**: 남성($b = .292, p < .01$)일 확률이 높고, 연령($b = -.799, p < .001$)이 낮으며, 높은 월 평균 가구소득($b = .265, p < .001$)과 학습지향 참여동기($b = .366, p < .01$)를 가지고 있고, 전일제($b = .578, p < .001$), 사무직($b = .959, p < .01$)일 확률이 높으며, 더 큰 규모의 직장($b = .200, p < .001$)에 근무하고 있을 확률이 높음

잠재계층모형 활용

연구결과: 교육수준에 따른 평생학습 참여 유형별 성과분석

평생학습 참여 유형별 성과 차이 분석: 일원배치 분산분석

<표 3> 교육수준에 따른 평생학습 참여 유형별 평생학습 참여 성과의 차이

성과변인	집단1 (n=2,015)		집단2 (n=1,167)		집단3 (n=2,007)		집단4 (n=1,474)		F(p)	사후검증 (Scheffe)
	M	SE	M	SE	M	SE	M	SE		
사회참여만족도	3.82	.015	3.62	.022	3.73	.014	3.72	.018	22.441***	1>3,4>2
경제적안정감	3.38	.018	3.03	.025	3.31	.016	3.18	.021	57.105***	1>3>4>2

***p<.001

주. 집단 1: 저학력-저참여, 집단 2: 고학력-고참여, 집단 3: 저학력-전통적 학습방법의존형, 집단 4: 고학력-매체기반형

- 모든 영역에 걸쳐 '저학력-저참여' 집단의 평생학습 참여 성과에 대한 인식이 가장 높았고, '고학력-고참여' 집단의 인식이 가장 낮았음
- 전반적으로 학력이 낮은 재직자 집단(집단 1, 집단 3)이 고학력 집단(집단 2, 집단 4)에 비해 평생학습 참여 성과에 대해 긍정적으로 인식
- 비록 저학력 재직자 집단의 평생학습 참여는 고학력 집단에 비해 낮은 경향을 보이지만, 다양한 측면에 있어 평생학습의 효과성 또는 만족도에 대해 학력이 낮은 집단이 더 높은 기대와 가치를 부여하고 있음을 시사

잠재계층모형 활용

결론 및 논의: 연구결과 요약

▪ 교육수준에 따른 평생학습 참여 잠재집단 분류

- (1)저학력-저참여, (2)고학력-고참여, (3)저학력-전통적 학습방법 의존형, (4)고학력-매체기반형
- 전반적으로 저학력 집단에 비해 고학력 집단이 평생학습 프로그램에 참여할 확률이 높았으며, 재직자 집단간 교육수준의 차이는 비형식교육 참여 확률과 밀접하게 연관
- 무형식학습 참여 맥락에서 집단 3과 집단 4를 비교할 때, 저학력 집단은 가족, 직장동료 등 주변인으로부터의 조언이나 텔레비전, 라디오를 통한 지식 습득 등 전통적 학습방식에 의존하는 경향이 있었으며, 반면 고학력 집단은 미디어, 온라인 매체 등을 활용한 새로운 학습방식을 선호할 확률이 높았음

잠재계층모형 활용

결론 및 논의: 연구결과 요약

- **교육수준에 따른 평생학습 참여 유형과 예측요인 간 관계**
 - 평생학습 참여에 있어 고학력 집단의 경우 저학력 집단에 비해 개인수준에서는 연령이 낮고 학습동기가 높으며, 직업관련 수준에서는 큰 규모의 직장 재직할 가능성이 높았음
 - 학습 참여 맥락에서 직장내 사회자본으로써 ‘유의미한 타인(significant others)’ 또는 다양한 학습 프로그램 운영 및 지원과 정적인 연관이 있는 것으로 알려진 직장규모 또는 업무환경의 영향을 보고한 연구결과(Brown & Bimrose, 2018; Kwon, 2019)를 지지
- **교육수준에 따른 평생학습 참여 유형별 성과 차이**
 - 전반적으로 저학력 집단이 고학력 집단에 비해 평생학습 참여 성과에 대해 긍정적으로 인식
 - 학습자 스스로 본인의 부족한 부분을 인식하고, 극복하고자 노력하는 행위가 삶의 기회로 연결될 때 평생학습에 대한 가치 형성(권재현, 윤선영, 2009)
 - 고학력자 집단에게 있어 평생학습의 참여의 효과성 및 만족도 제고 필요성 시사

잠재계층모형 활용

결론 및 논의: 시사점

▪ 교육격차와 평생학습 참여

- 학력은 평생학습 참여를 예측하는 유의미한 변수이며(한송희 외, 2007; 황정원, 길혜지, 2017), 평생학습 참여는 성인의 경제적·사회적 성과에 영향
- 교육수준에 따른 평생학습 참여 격차는 재직자의 경제적 지위 및 노동시장 성과의 격차를 야기할 수 있으며, 따라서 저학력 집단의 평생학습 참여 활성화를 위한 국가 및 조직수준의 논의 필요

▪ 포용적(inclusive) 성인학습 체제 구축

- 평생교육 기회 양적 확대의 이면에 존재할 수 있는 평생학습 참여 격차 및 이에 따른 사회·경제적 불평등 재생산에 대한 면밀한 고려 필요
- 대학평생교육 재정지원사업의 확대 운영(교육부, 2019), 정부-산업-대학간 파트너십을 통한 저학력·저숙련 재직자 맞춤형 평생학습 프로그램의 도입(OECD, 2008), 평생학습 참여 활성화를 위한 재정지원 및 자원확보(Desjardins & Lee, 2016) 등 다양한 차원의 방안 검토 필요

잠재계충분성 실습

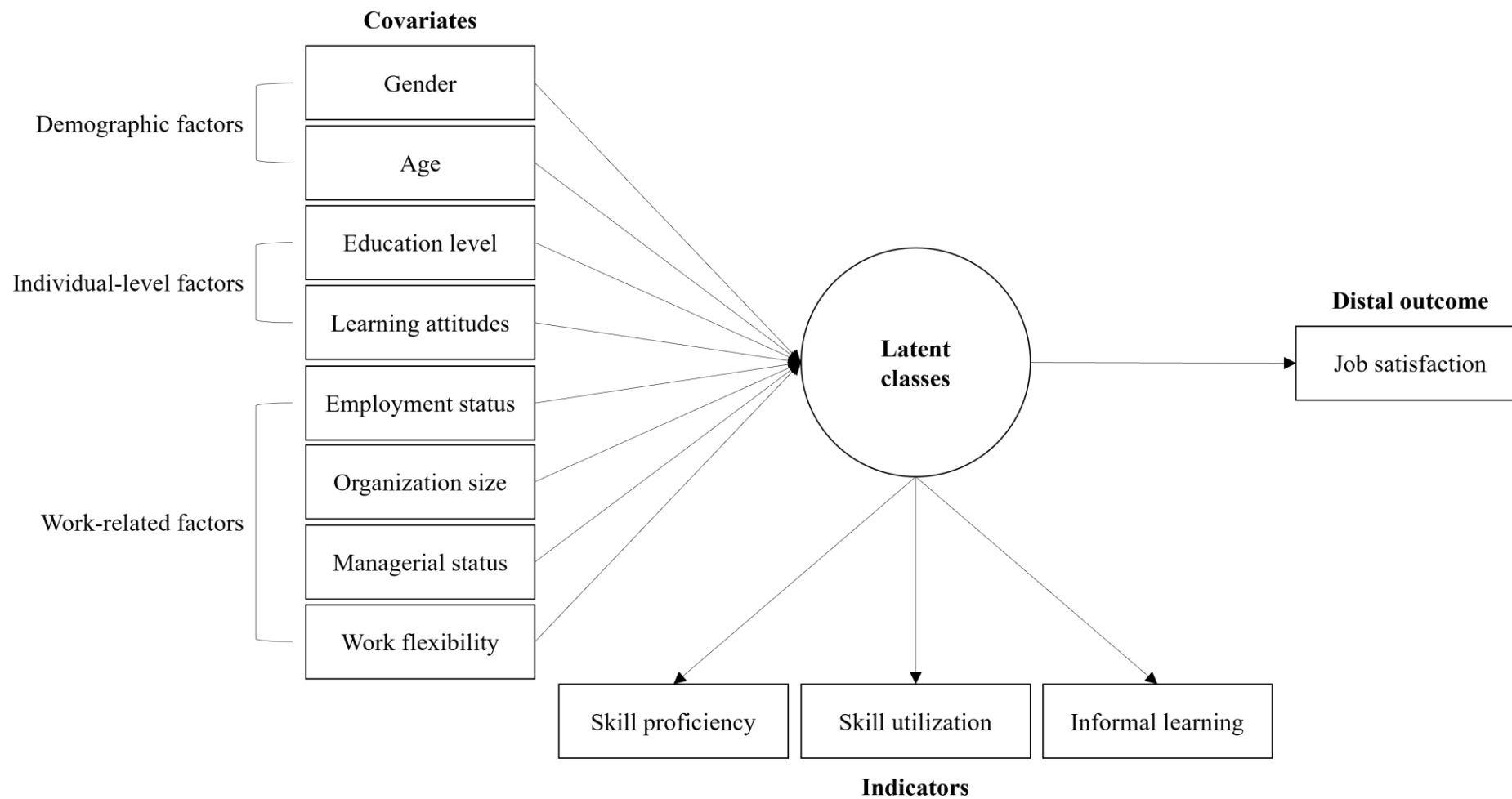
잠재계층분석 실습

실습 개요

- 연구주제: Skill Gap, Informal Workplace Learning, and Job Satisfaction Among Workers: A Latent Class Analysis (manuscript in progress)
- 분석자료 및 연구대상
 - The 2017 U.S. data from the Program for the International Assessment of Adult Competencies (PIAAC) survey (access link: <https://github.com/csjiang1227/LCA-SkillGap>)
 - Total sample size: 1,446 adult workers aged 25 to 65 years old who were currently in the workforce
- 분석 프로그램
 - LCA: The 'snowLatent' module (Seol, 2022) in *jamovi* software version 2.3.28 (download link: <https://www.jamovi.org/download.html>).
 - Visualization: The 'ggplot2' package (Wickham, 2016) in *RStudio* version 2023.06.1.

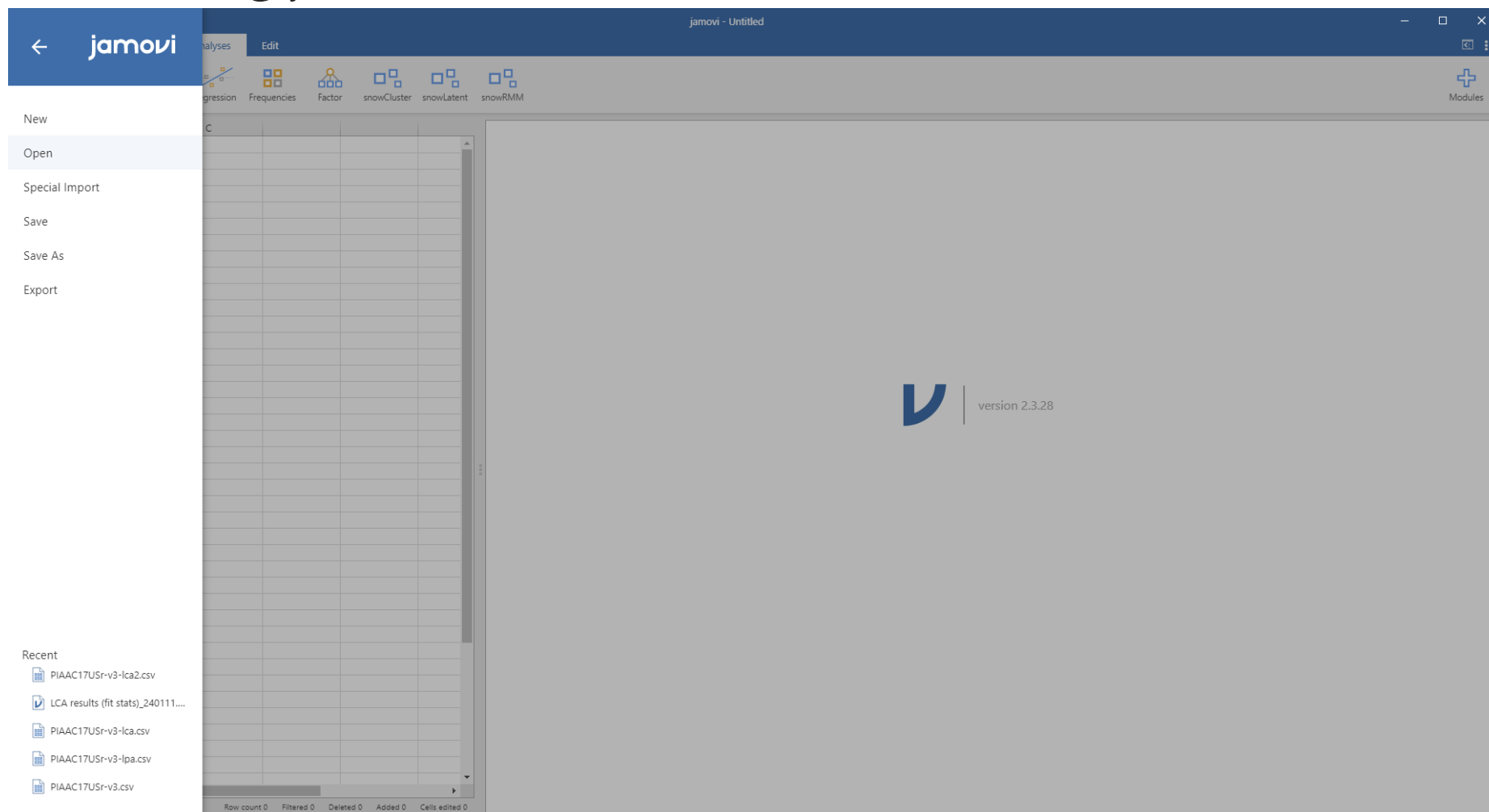
잠재계층분석 실습

연구모형



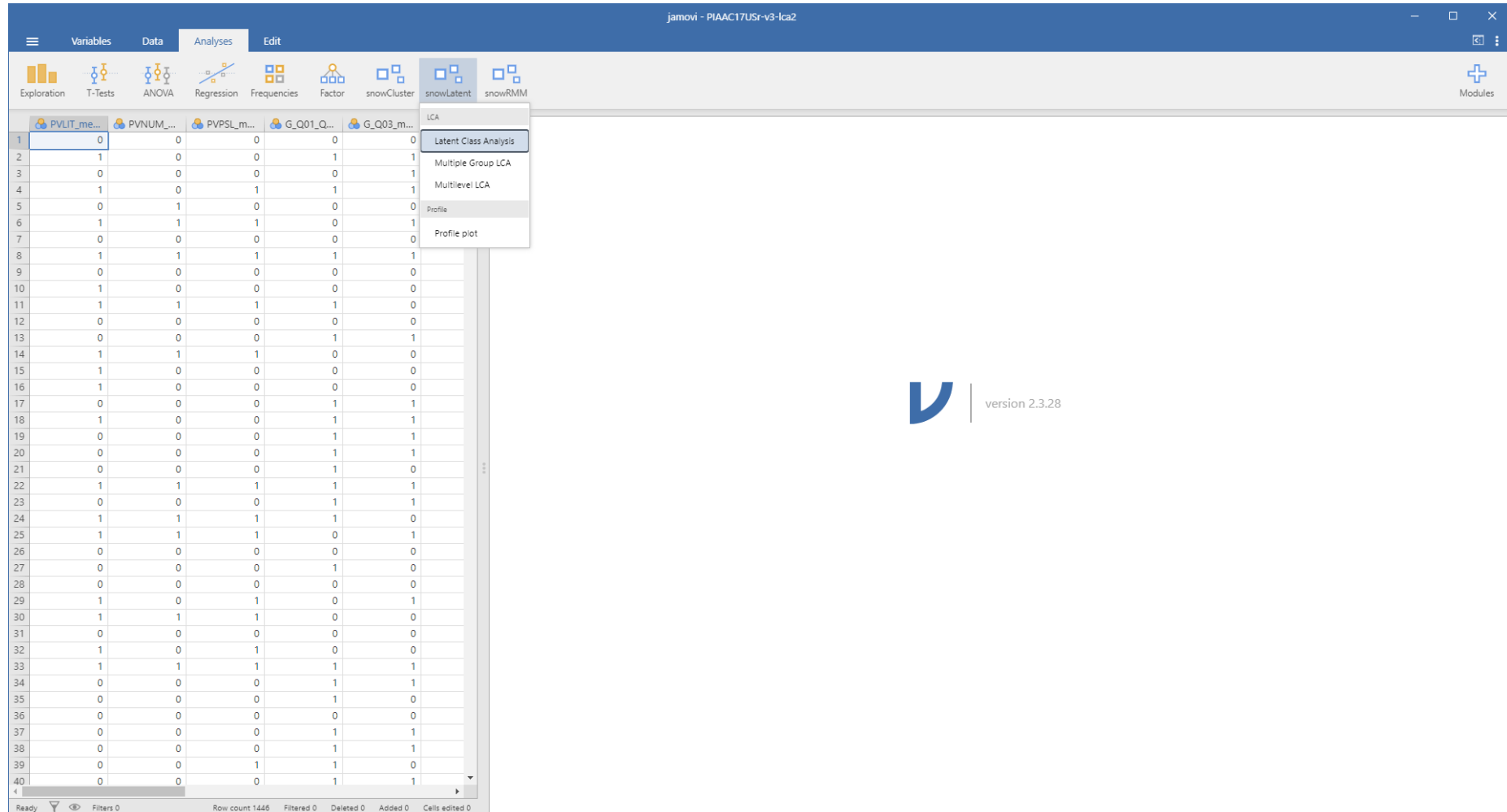
잠재계층분석 실습

STEP 1(using *jamovi*): 데이터 불러오기



잠재계층분석 실습

STEP 1(using *jamovi*): Latent Class Analysis 모듈 선택



The screenshot shows the Jamovi software interface. The 'Analyses' menu is open, and 'Latent Class Analysis' is selected. The main window displays a dataset with 40 rows and 6 columns. The status bar at the bottom indicates 'Ready', 'Filters 0', 'Row count 1446', 'Filtered 0', 'Deleted 0', 'Added 0', and 'Cells edited 0'.

	PVLIT_me...	PVNUM_...	PVPSL_m...	G_Q01_Q...	G_Q03_m...
1	0	0	0	0	0
2	1	0	0	1	1
3	0	0	0	0	1
4	1	0	1	1	1
5	0	1	0	0	0
6	1	1	1	0	1
7	0	0	0	0	0
8	1	1	1	1	1
9	0	0	0	0	0
10	1	0	0	0	0
11	1	1	1	1	0
12	0	0	0	0	0
13	0	0	0	1	1
14	1	1	1	0	0
15	1	0	0	0	0
16	1	0	0	0	0
17	0	0	0	1	1
18	1	0	0	1	1
19	0	0	0	1	1
20	0	0	0	1	1
21	0	0	0	1	0
22	1	1	1	1	1
23	0	0	0	1	1
24	1	1	1	1	0
25	1	1	1	0	1
26	0	0	0	0	0
27	0	0	0	1	0
28	0	0	0	0	0
29	1	0	1	0	1
30	1	1	1	0	0
31	0	0	0	0	0
32	1	0	1	0	0
33	1	1	1	1	1
34	0	0	0	1	1
35	0	0	0	1	0
36	0	0	0	0	0
37	0	0	0	1	1
38	0	0	0	1	1
39	0	0	1	1	0
40	0	0	0	1	1

잠재계층분석 실습

STEP 1(using jamovi): 모형 적합도 분석 및 모형 평가

Variables

Data

Analyses

Edit

Exploration

T-Tests

ANOVA

Regression

Frequencies

Factor

snowCluster

snowLatent

snowRMM

Modules

Latent Class Analysis

PVPSL_mean2

G_Q05_mean

L_Q04_mean

→

PVLIT_mean2

PVNUM_mean2

G_Q01_Q02_mean

G_Q03_mean

D_Q13A

→

Covariates(Optional)

Analysis options

Class

7

Bootstrap

50

Analysis

Model fit

☒ Fit measures

☒ Absolute model fit

☐ Relative model fit(Class>2)

☐ Prevalences for latent class

☐ Item response(Rho)

☐ Posterior

☐ Class membership

Additional output

Plot

☐ Profile plot

☐ Item by class

Angle of x-axis labels

0

☒ Elbow plot

Logistic regression

☐ Coefficients with covariates

Results

Latent Class Analysis

To get started:

1. Latent Class Analysis(LCA) based on [glca](#) R package.

2. The rationale of snowLatent module is described in the [documentation](#).

3. The result table does not printed if the results from glca R package are not available.

4. Feature requests and bug reports can be made on my [GitHub](#).

Model information

Call:

glca::glca(formula = formula, data = data, nclass = nc, seed = 1)

Manifest items : PVLIT_mean2 PVNUM_mean2 G_Q01_Q02_mean G_Q03_mean D_Q13A D_Q13B D_Q13C

Model : Latent class analysis

Number of latent classes : 7

Number of observations : 1446

Number of parameters : 55

log-likelihood : -6028

G-squared : 80.46

AIC : 12167

BIC : 12457

Model fit

Class	Log-likelihood	AIC	CAIC	BIC	Entropy	df	G ²
7	-6028	12167	12512	12457	0.749	72	80.5

[3]

Absolute model fit

Class	Log-likelihood	AIC	CAIC	BIC	Entropy	df	G ²	p
2	-6338	12705	12800	12785	0.804	112	698.9	< .001
3	-6179	12405	12549	12526	0.687	104	382.5	< .001
4	-6105	12271	12466	12435	0.729	96	232.6	< .001

잠재계층분석 실습

STEP 1(using jamovi): 모형 적합도 분석 및 모형 평가

Variables

Data

Analyses

Edit

Exploration

T-Tests

ANOVA

Regression

Frequencies

Factor

snowCluster

snowLatent

snowRMM

Modules

Latent Class Analysis

PVPSL_mean2

G_Q05_mean

L_Q04_mean

Membership

→

Variables

PVLIT_mean2

PVNUM_mean2

G_Q01_Q02_mean

G_Q03_mean

D_Q13A

→

Covariates(Optional)

Analysis options

Class

4

Bootstrap

50

Analysis

Model fit

☒ Fit measures

☒ Absolute model fit

☐ Relative model fit(Class>2)

Probability

☒ Prevalences for latent class

☒ Item response(Rho)

☐ Posterior

☒ Class membership

Plot

☒ Profile plot

☐ Item by class

Angle of x-axis labels 0

☐ Elbow plot

Logistic regression

☐ Coefficients with covariates

> Additional output

Results

Latent Class Analysis

To get started:

1. Latent Class Analysis(LCA) based on [glca](#) R package.

2. The rationale of snowLatent module is described in the [documentation](#).

3. The result table does not printed if the results from glca R package are not available.

4. Feature requests and bug reports can be made on my [GitHub](#).

Model information

Call:

glca::glca(formula = formula, data = data, nclass = nc, seed = 1)

Manifest items : PVLIT_mean2 PVNUM_mean2 G_Q01_Q02_mean G_Q03_mean D_Q13A D_Q13B D_Q13C

Model : Latent class analysis

Number of latent classes : 4

Number of observations : 1446

Number of parameters : 31

log-likelihood : -6105

G-squared : 232.6

AIC : 12271

BIC : 12435

Model fit

Class	Log-likelihood	AIC	CAIC	BIC	Entropy	df	G ²
4	-6105	12271	12466	12435	0.729	96	233

[3]

Absolute model fit

Class	Log-likelihood	AIC	CAIC	BIC	Entropy	df	G ²	p
2	-6338	12705	12800	12785	0.804	112	699	< .001
3	-6179	12405	12549	12526	0.687	104	382	< .001
4	-6105	12271	12466	12435	0.729	96	233	< .001

잠재계층분석 실습

STEP 1(using jamovi): 모형 적합도 분석 및 모형 평가

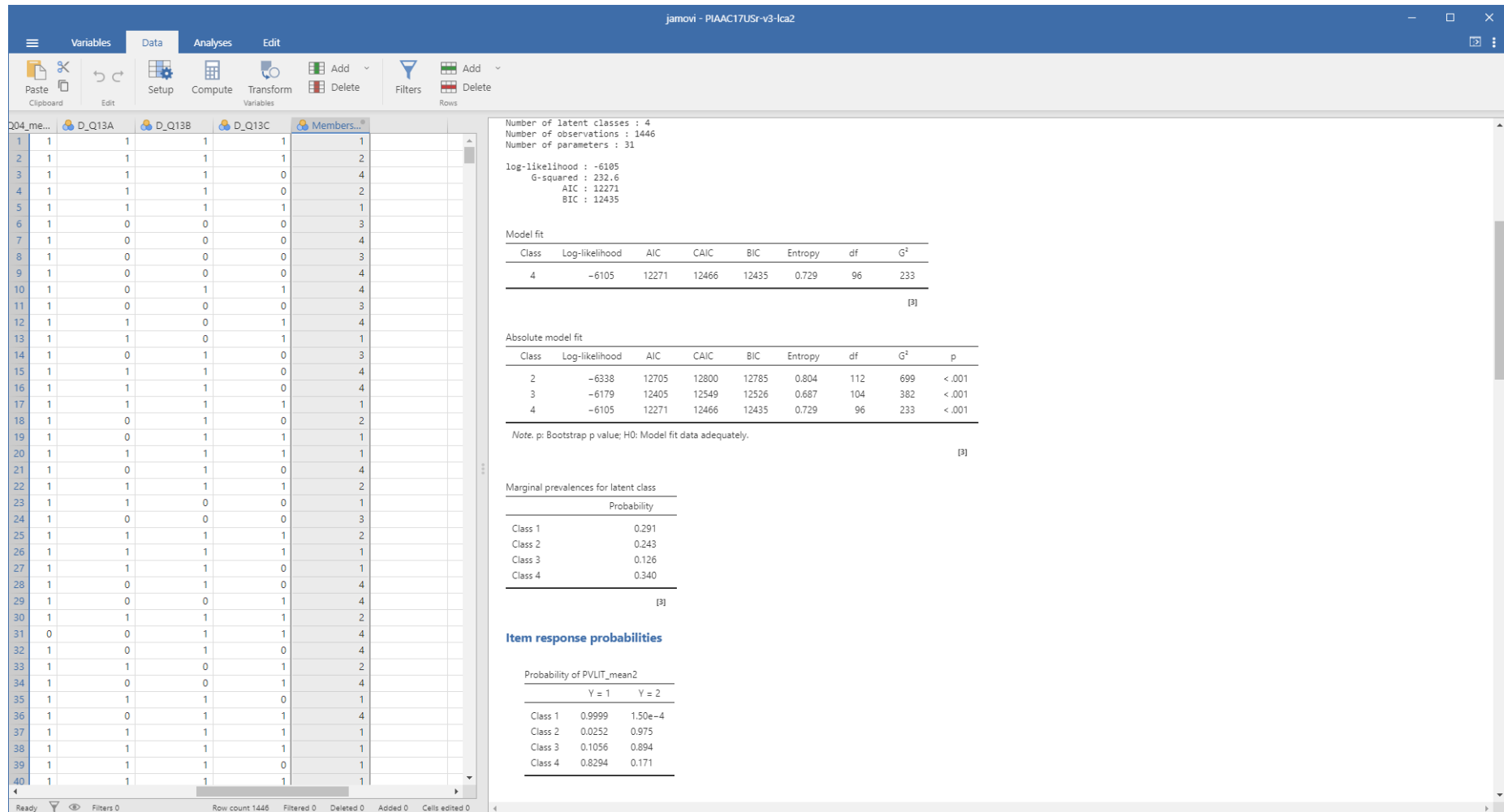
Table 1
Comparison of Model Fit Indices for Latent Class Solutions

Number of classes	Model fit indices					
	<i>LL</i>	<i>AIC</i>	<i>BIC</i>	<i>CAIC</i>	<i>BLRT p</i>	<i>Entropy</i>
2-class	-6338	12705	12785	12800	.000	.804
3-class	-6179	12405	12526	12549	.000	.687
4-class	-6105	12271	12412	12451	.000	.741
5-class	-6064	12206	12425	12466	.100	.729

Note. *N* = 1,446; LL = log-likelihood, BIC = Bayesian information criteria, AIC = Akaike information criteria, CAIC = consistent Akaike information criteria, BLRT = bootstrapped log-likelihood ratio test; Bold-faced values indicate the best-fit solution.

잠재계층분석 실습

STEP 1(using *jamovi*): Membership 칼럼 원 데이터로 가져오기



jamovi - PIAAC17USr-v3-lca2

Number of latent classes : 4
Number of observations : 1446
Number of parameters : 31

log-likelihood : -6105
G-squared : 232.6
AIC : 12271
BIC : 12435

Model fit

Class	Log-likelihood	AIC	CAIC	BIC	Entropy	df	G ²
4	-6105	12271	12466	12435	0.729	96	233

[3]

Absolute model fit

Class	Log-likelihood	AIC	CAIC	BIC	Entropy	df	G ²	p
2	-6338	12705	12800	12785	0.804	112	699	< .001
3	-6179	12405	12549	12526	0.687	104	382	< .001
4	-6105	12271	12466	12435	0.729	96	233	< .001

Note, p: Bootstrap p value; H0: Model fit data adequately.

[3]

Marginal prevalences for latent class

	Probability
Class 1	0.291
Class 2	0.243
Class 3	0.126
Class 4	0.340

[3]

Item response probabilities

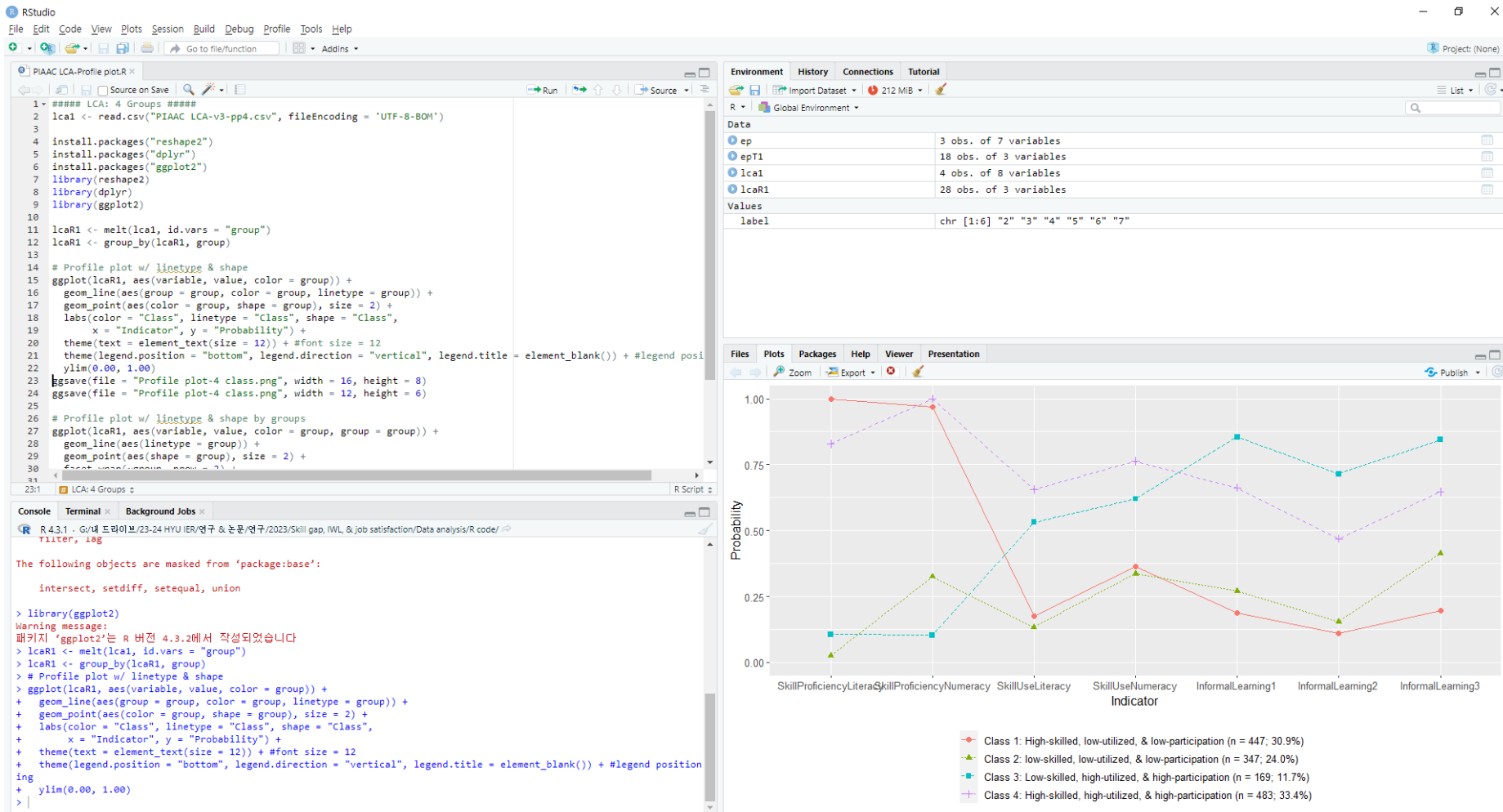
Probability of PVLIT_mean2

	Y = 1	Y = 2
Class 1	0.9999	1.50e-4
Class 2	0.0252	0.975
Class 3	0.1056	0.894
Class 4	0.8294	0.171

Ready Filters 0 Row count 1446 Filtered 0 Deleted 0 Added 0 Cells edited 0

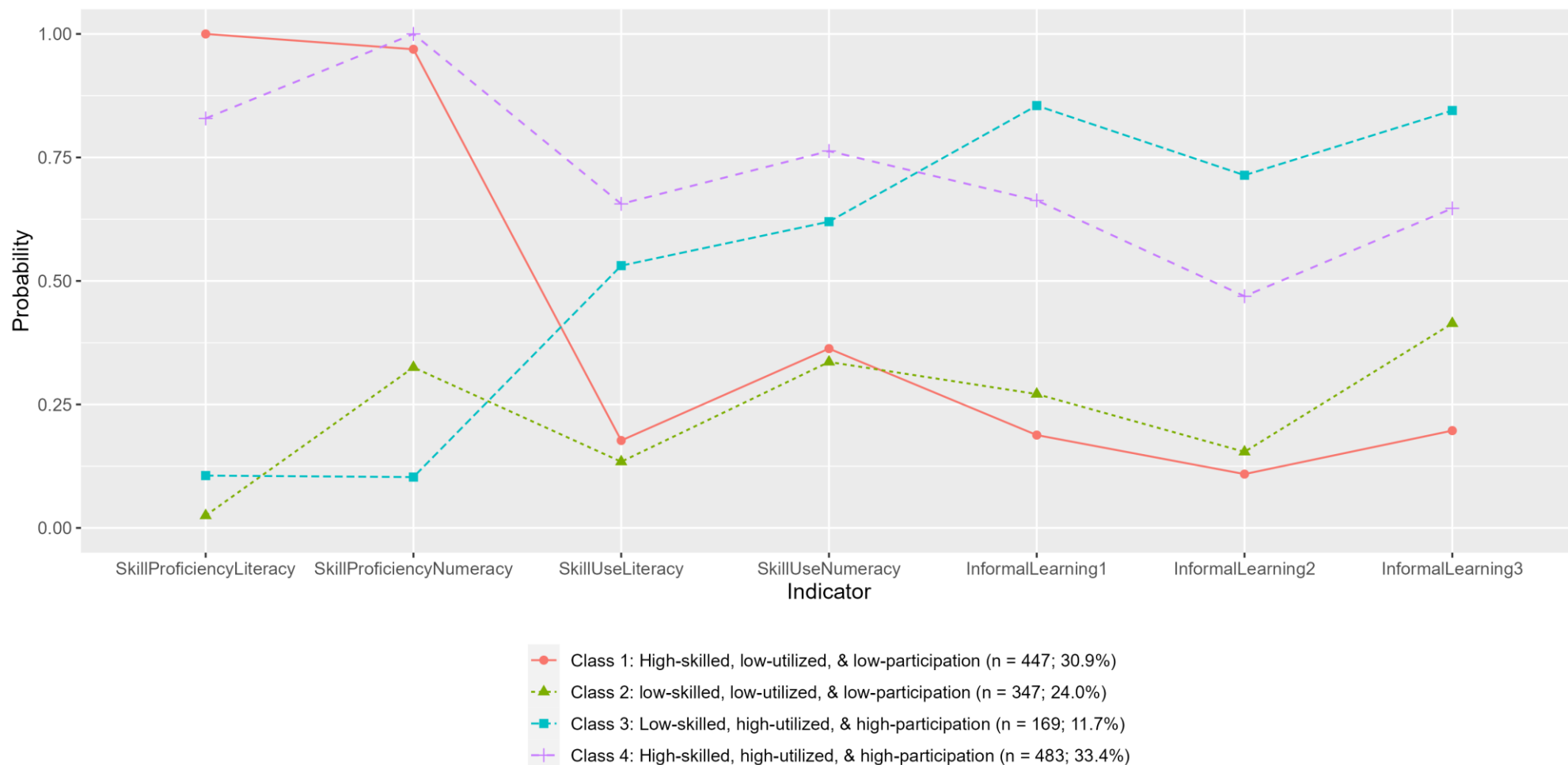
잠재계층분석 실습

STEP 1(using RStudio): 잠재계층분석 결과 도식화



잠재계층분석 실습

STEP 1(using *RStudio*): 잠재계층분석 결과 도식화



잠재계층분석 실습

STEP 2(using SPSS): 잠재집단 분류 영향요인

PIAAC17USr-v3.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래픽(G) 유틸리티(U) 확장(X) 창(W) 도움말(H) Meta Analysis KoreaPlus(P)

가계통계분석(E) > 보고서(R) > 기술통계량(E) > 표(B) > 평균 비교(M) > 일반선형모형(G) > 일반화 선형 모형(Z) > 혼합 모형(X) > 상관분석(C) > 회귀분석(R) > 로그선형분석(Q) > 선형(W) > 분류분석(F) > 차원 축소(D) > 척도분석(A) > 비모수검정(N) > PS Matching > 시계열 분석(T) > 생존분석(S) > 다중반응(U) > 결속값 분석(Y) > 다중대체(I) > 복합 표본(L) > 시뮬레이션(I) > 품질관리(Q) > 공간과 시간 모형화(S) > 다이어트 마케팅(K)

1 SEQID 숫자 5
2 GENDER_R 숫자 1
3 AGE10LF... 숫자 3
4 EDCAT6r1 숫자 2
5 EDCAT6r2 숫자 2
6 MONTHLY... 숫자 1
7 PVLIT_mean 숫자 8
8 PVNUM_m... 숫자 8
9 PVPSL_mean 숫자 8
10 PVLIT_mean2 숫자 8
11 PVNUM_m... 숫자 8
12 PVPSL_mean... 숫자 8
13 I_Q04_mean 숫자 8
14 I_Q05F 숫자 1
15 I_Q07_mean 숫자 8
16 I_Q08 숫자 1
17 C_Q07 숫자 2
18 C_Q09_C 숫자 2
19 D_Q03 숫자 1
20 D_Q06A 숫자 1
21 D_Q08A 숫자 1
22 D_Q10_T1 숫자 2
23 D_Q11_mean 숫자 8
24 D_Q14 숫자 1
25 G_Q01_Q0... 숫자 8
26 G_Q03_mean 숫자 8
27 G_Q05_mean 숫자 8
28 FAET12 숫자 1
29 FAET12JR 숫자 1
30 NFE12 숫자 1
31 B_Q12_comb 숫자 8
32 B_Q12_co... 숫자 8
33 NFE12JR 숫자 1
34 D_Q13_mean 숫자 8
35 D_Q13_co... 숫자 8
36 D_Q13_co... 숫자 8
37 D_Q13 co... 숫자 8

가정제공 분석(E) > 보고서(R) > 기술통계량(E) > 표(B) > 평균 비교(M) > 일반선형모형(G) > 일반화 선형 모형(Z) > 혼합 모형(X) > 상관분석(C) > 회귀분석(R) > 로그선형분석(Q) > 선형(W) > 분류분석(F) > 차원 축소(D) > 척도분석(A) > 비모수검정(N) > PS Matching > 시계열 분석(T) > 생존분석(S) > 다중반응(U) > 결속값 분석(Y) > 다중대체(I) > 복합 표본(L) > 시뮬레이션(I) > 품질관리(Q) > 공간과 시간 모형화(S) > 다이어트 마케팅(K)

다항 로지스틱 회귀

종속변수(D): membership_lca2(마지막)
참조범주(N):
요인(F): GENDER_R, AGE10LFSEXT, EDCAT6r2
공변량(C):

확인 불여방기(R) 재설정(R) 취소 도움말

데이터 보기 변수 보기

다항 로지스틱(M)...

IBM SPSS Statistics 프로세서 준비 완료 유니코드 설정

잠재계층분석 실습

STEP 2(using SPSS): 잠재집단 분류 영향요인

Table 2

Multinomial Logistic Regression Analysis Results Across Class Membership by Covariates

	ref. = Class 1 ^a		ref. = Class 2				ref. = Class 3				ref. = Class 4	
	Class 2 ^b		Class 3 ^c		Class 4 ^d		Class 3		Class 4		Class 4	
	B	OR	B	OR	B	OR	B	OR	B	OR	B	OR
<i>Demographic factor</i>												
Gender (ref. = female)	.280	1.323	.568**	1.765	-.273	.761	.289	1.334	.553**	1.425	.841***	1.569
Age	-.150	.861	.148	1.160	.023	1.023	-.298**	.653	-.172*	.812	-.126	.882
<i>Individual-level factor</i>												
Education level (ref. = less than a bachelor's degree)	-1.437***	.791	1.191***	3.289	.092	1.096	-.247	.781	1.345***	1.740	1.099***	1.667
Learning attitudes	-.383**	.533	.418*	1.342	.605***	1.454	.801***	1.551	.988***	1.628	-.187	.830
<i>Work-related factor</i>												
Employment status (ref. = full-time)	-.105	.900	-.082	.921	-.185	.831	.023	1.023	-.080	.923	-.103	.902
Organization size	-.003	.997	-.117	.889	-.132*	.877	-.114	.892	-.129*	.879	-.014	.986
Managerial status (ref. = non-managerial)	.000	1.000	-.227	.797	-.402**	.669	-.228	.796	.403*	1.332	-.175	.839
Work flexibility	.218	1.244	.042	1.043	-.063	.939	-.176	.839	.281**	1.245	-.106	.900
Intercept	-3.386***		-.395		3.329***		2.990**		6.714***		3.724***	
<i>Goodness of fit</i>												
-2 LL	33342.734											
χ^2 (df)	263.379 (24)***											
Cox & Snell R ²	.174											
Nagelkerke R ²	.188											

Note. N = 1,446; LL = log-likelihood, OR = odds ratio.

^a Class 1: High-skilled, low-utilized, & low-participation (n = 447; 30.9%), ^b Class 2: low-skilled, low-utilized, & low-participation (n = 347; 24.0%), ^c Class 3: Low-skilled, high-utilized, & high-participation (n = 169; 11.7%), ^d Class 4: High-skilled, high-utilized, & high-participation (n = 483; 33.4%).

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

잠재계층분석 실습

STEP 2(using SPSS): 잠재집단별 성과 차이

PIAAC17USr-v3.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래픽(G) 유틸리티(U) 확장(X) 창(W) 도움말(H) Meta Analysis KoreaPlus(P)

거름제곱 분석(M) > 보고서(R) > 기술통계량(E) > 베이저안 통계량(B) > 평균 비교(M) > 일반선형모형(G) > 일반화 선형 모형(Z) > 혼합 모형(X) > 상관분석(C) > 회귀분석(R) > 로그선형분석(Q) > 신경망(W) > 분류분석(E) > 차원 축소(D) > 척도분석(A) > 비모수검정(N) > PS Matching > 시계열 분석(T) > 생존분석(S) > 다중반응(U) > 결속값 분석(Y) > 다중대체(I) > 복합 표본(L) > 시뮬레이션(O) > 품질관리(Q) > 공간과 시간 모형화(S) > 다차원트 마케팅(K)

이름	유형	너비	분석	결과값	알	맞춤	속도	역할
1	SEQID	숫자	5					
2	GENDER_R	숫자	1					
3	AGEG10LF...	숫자	3					
4	EDCAT6r1	숫자	2					
5	EDCAT6r2	숫자	2					
6	MONTHLYI...	숫자	1					
7	PVLIT_mean	숫자	8					
8	PVNUM_m...	숫자	8					
9	PVPSL_mean	숫자	8					
10	PVLIT_mean2	숫자	8					
11	PVNUM_m...	숫자	8					
12	PVPSL_me...	숫자	8					
13	I_Q04_mean	숫자	8					
14	I_Q05F	숫자	1					
15	I_Q07_mean	숫자	8					
16	I_Q08	숫자	1					
17	C_Q07	숫자	2					
18	C_Q09_C	숫자	2					
19	D_Q03	숫자	1					
20	D_Q06A	숫자	1					
21	D_Q08A	숫자	1					
22	D_Q10_T1	숫자	2					
23	D_Q11_mean	숫자	8					
24	D_Q14	숫자	1					
25	G_Q01_Q0...	숫자	8					
26	G_Q03_mean	숫자	8					
27	G_Q05_mean	숫자	8					
28	FAET12	숫자	1					
29	FAET12JR	숫자	1	0	Participated in f... {0, No}...	4 - 9	8	명목형
30	NFE12	숫자	1	0	Participated in ... {0, No}...	5 - 9	8	명목형
31	B_Q12_comb	숫자	8	0	Participated in ... {0, No}...	지정양음	12	명목형
32	B_Q12_co...	숫자	8	0	Participated in ... {0, No}...	지정양음	12	명목형
33	NFE12JR	숫자	1	0	Participated in ... {0, No}...	5 - 9	8	명목형
34	D_Q13_mean	숫자	8	0	Informal learnin... {1, Never}...	지정양음	12	명목형
35	D_Q13_co...	숫자	8	0	Informal learnin... {0, Low}...	지정양음	13	명목형
36	D_Q13_co...	숫자	8	0	Informal learnin... {0, Low}...	지정양음	13	명목형
37	D_Q13 co...	숫자	8	0	Informal learnin... {0, Low}...	지정양음	13	명목형

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 유니코드: 설정

잠재계층분석 실습

STEP 2(using SPSS): 잠재집단별 성과 차이

Table 3
Descriptive Statistics and F Test for Distal Outcome by Class Membership

Distal outcome	Descriptive statistics								<i>F</i> test	Post hoc <i>Scheffé</i> test
	Class 1 ^a		Class 2 ^b		Class 3 ^c		Class 4 ^d			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Job satisfaction	3.92	.908	3.92	.880	4.16	.800	4.13	.801	8.39***	3, 4 > 1, 2

Note. N = 1,446; SD = standard deviation.
^a Class 1: High-skilled, low-utilized, & low-participation (n = 447; 30.9%), ^b Class 2: low-skilled, low-utilized, & low-participation (n = 347; 24.0%), ^c Class 3: Low-skilled, high-utilized, & high-participation (n = 169; 11.7%), ^d Class 4: High-skilled, high-utilized, & high-participation (n = 483; 33.4%).
 *** p < .001

참고문헌

교육부(2022). 제5차 평생교육진흥 기본계획('23~'27년): 평생학습 진흥방안. 교육부.

권재현, 윤선영(2009). 성인학습자들의 학습태도가 평생학습 참여 및 학습의 사회적 성과에 미치는 영향 분석. *평생교육학연구*, 15(3), 85-110.

신태수(2010). 종단프로파일 분석과 군집분석을 이용한 잠재집단연구: 성장혼합모형과 비교를 통하여. *교육평가연구*, 23(3), 641-664.

오영교, 차성현(2018). 대학생의 진로도구준비 유형 분석: 잠재계층분석방법(Latent Class Analysis)을 활용하여. *통계연구*, 23(3), 39-64. <https://doi.org/10.22886/jkos.2018.23.3.39>

장창성, 전하람, 김재형(2023). 교육수준에 따른 재직자의 평생학습 참여 유형 및 성과분석: 잠재계층분석. *평생교육학연구*, 29(4), 149-180. <http://doi.org/10.52758/kjle.2023.29.4.149>

채창균(2021). 직업능력향상을 위한 평생학습 참여 실태. 통계개발원(편), *한국의 사회동향 2021*(pp. 149-157). 통계청.

최동선(2021). 미래 환경 변화에의 대응을 위한 고등교육 기관에서의 평생직업교육 내실화 방안(기본연구2021-05). 한국직업능력연구원.

한송희, 신태수, 양은아(2007). 평생학습참여 결정요인에 관한 연구: 학습자배경, 학습패턴 및 자기보고 핵심역량이 평생학습참여에 미치는 영향. *평생교육학연구*, 13(2), 93-118.

황정원, 길혜지(2017). 비형식 평생학습 참여에 영향을 미치는 학습자 요인 분석: 직업능력향상 및 인문교양 프로그램을 중심으로. *평생학습사회*, 13(4), 105-127.

<https://doi.org/10.26857/JLLS.2017.11.13.4.105>

Brown, A., & Bimrose, J. (2018). Drivers of learning for the low-skilled. *International Journal of Lifelong Education*, 37(2), 151-167. <https://doi.org/10.1080/02601370.2017.1378934>

Desjardins, R., & Lee, J. (2016). *Earnings and employment benefits of adult higher education in comparative perspective: Evidence based on the OECD Survey of Adults Skills (PIAAC)*. Retrieved from PIAAC Gateway website:

https://static1.squarespace.com/static/51bb74b8e4b0139570dddf020/t/578d1dcc414fb589c6502f6b/1468865997731/Desjardins_Lee_PIAAC.pdf

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical & Experimental Research*, 24(6), 882-891. <https://doi.org/10.1111/j.1530-0277.2000.tb02070.x>

Kusurkar, R. A., Mak-van der Vossen, M., Kors, J., Grijpma, J. W., van der Burgt, S. M., Koster, A. S., & de la Croix, A. (2021). 'One size does not fit all': The value of person-centered analysis in health professions education research. *Perspectives on medical education*, 10(4), 245-251. <https://doi.org/10.1007/s40037-020-00633-w>

Kwon, K. (2019). The long-term effect of training and development investment on financial performance in Korean companies. *International Journal of Manpower*, 40(6), 1092-1109. <https://doi.org/10.1108/IJM-10-2017-0286>

OECD. (2008). *Tertiary education for the knowledge society: Volume 1 and Volume 2*. OECD publishing. <https://doi.org/10.1787/9789264046535-en>

Seol, H. (2022). *snowLatent: Latent class analysis for jamovi*. <https://github.com/hyunsooseol/snowLatent>

Sinha, P., Calfee, C. S., & Delucchi, K. L. (2022). Practitioner's guide to latent class analysis: Methodological considerations and common pitfalls. *Critical Care Medicine*, 49(1), e63-e79. <https://doi.org/10.1097/CCM.0000000000004710>

The jamovi Project. (2022). *jamovi* (Version 2.3). <https://www.jamovi.org>

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: A guide to best practice. *Journal of Black Psychology*, 46(4), 287-311. <https://doi.org/10.1177/0095798420930932>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>

질의응답