

한국복지패널조사를 활용한 삶의 경로 추적

발표 순서

- 패널자료와 생애사 연구
- 분석 갈래 개괄
- 분석 모형과 분석 사례
 - T : Trajectory Classification
 - Tm : Longitudinal SEM
 - S : Sequence Analysis
 - Sm : Multi-state Model
 - Sm : Mixed Latent Markov Model

패널자료와 생애사 연구

패널자료와 생애사 연구

실증적 생애사 연구

- 삶이란 시간의 흐름을 따라 다층적(multilevel), 다차원적(multidimensional), 연결된(linked) 구조로 이해되어야 함(Josef Brüde et al., 2019).
- 생애사적 접근(life-course perspective)은 사회과학 현상 규명에서 핵심적인 패러다임으로 자리잡음(Elder, 1994).
- “(S)ingle events should not be isolated from each other,” but rather need “to be understood in their continuity” (Aisenbrey & Fasang, 2010)
- “Holistic approach”(Raffaella Piccarre, 2019)
- 삶의 다양한 측면을 다루고 있는 한국복지패널과 같은 데이터의 성숙은 생애사적 접근에 입각한 실증 연구의 가능성을 더욱 확대하고 있으며, 관측 시점의 누적은 평균적 효과에 기반한 연구를 넘어선 삶의 경로를 본격적으로 다루는 연구를 촉발함.

패널자료와 생애사 연구

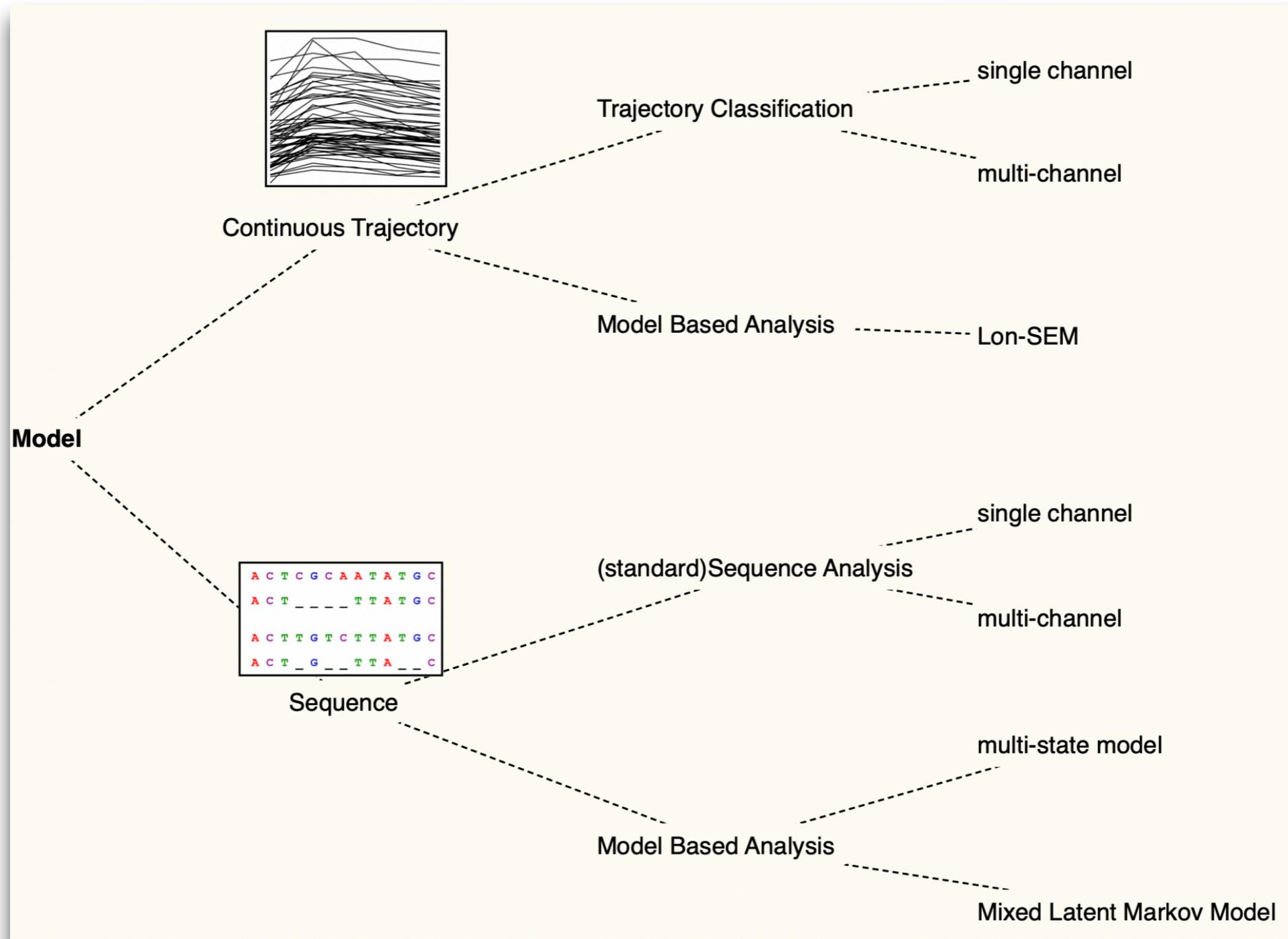
연구 문제와 관련한 주요 질문

- 왜 holistic한 접근이 타당한가?
- 분석에서 시간의 의미는 무엇인가? (연령, 누적 경험, 혹은 관찰 시점)
- 다루고자 하는 현상(속성)은 질적인가 양적인가?
- 얻고자 하는 결과는 전반적인 유형화인가, 특정 상태로의 변환 “확률”인가?
- 연구 성격은 탐색적(혹은, data-driven)인가, 모형 검증적인가?
-

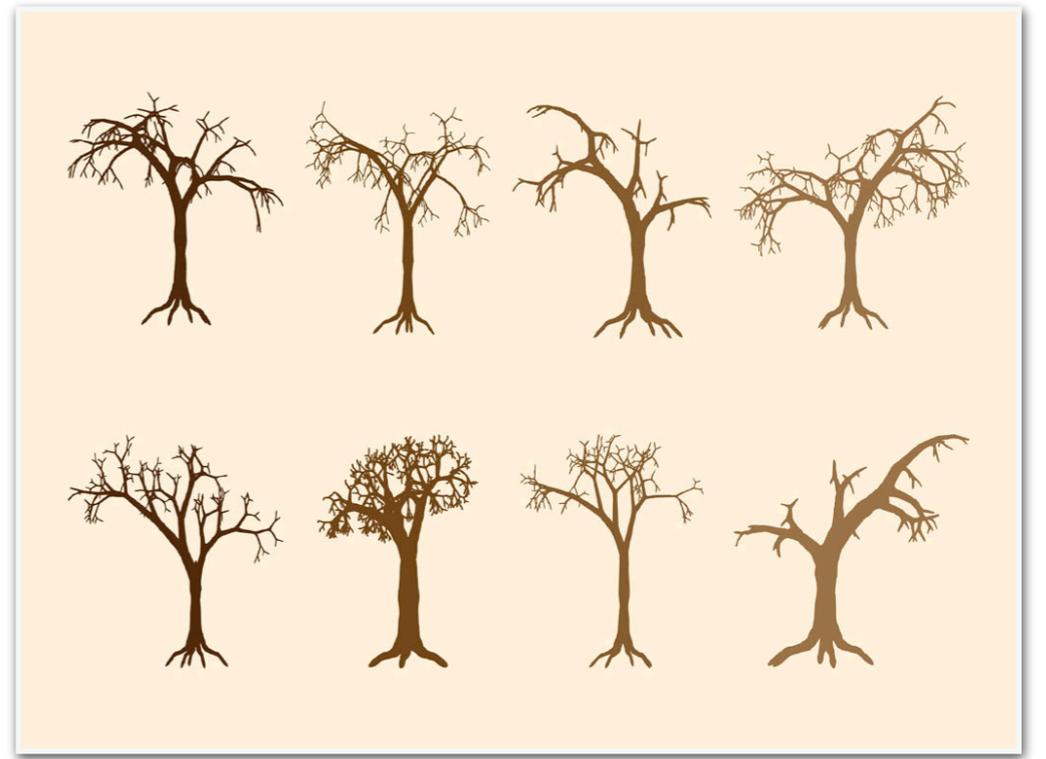
분석 갈래 개괄

분석 갈래 개괄

탐색적인 분석 계통도와 주요 개념



- Data mining & algorithm culture vs. Statistical culture

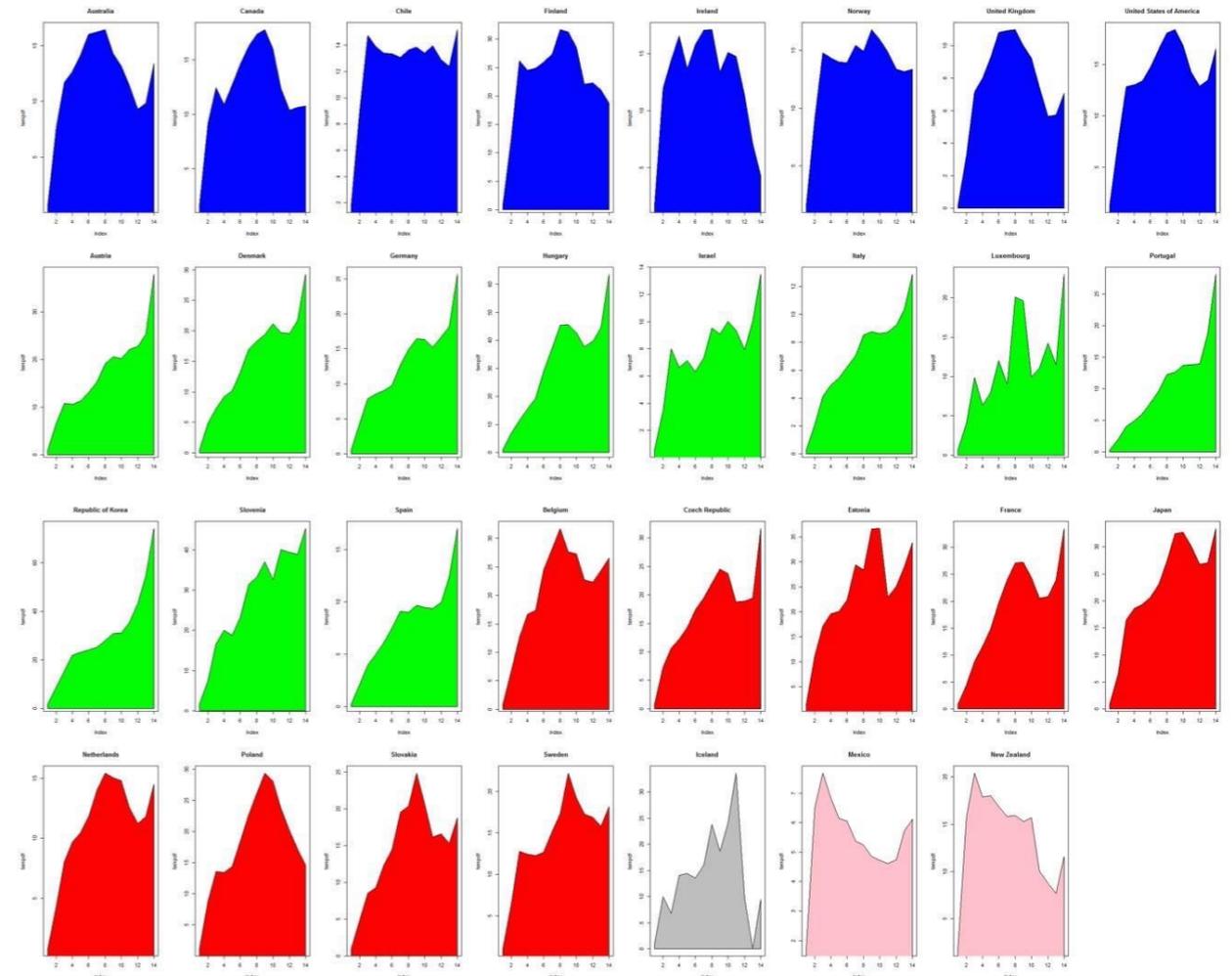


분석 모형과 사례

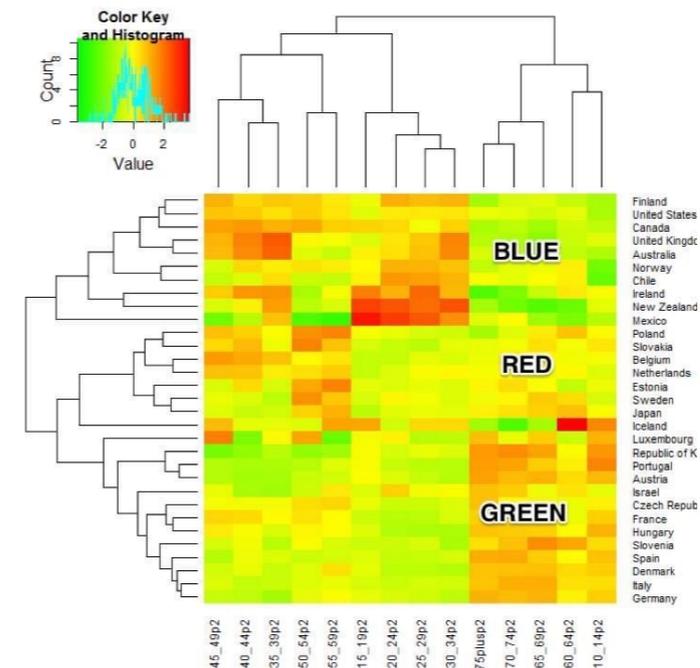
(Continuous) Trajectory

Trajectory Classification

- Partitioning(e.g. k-means clustering)기법의 종단 자료에 대한 응용.
- 반복 측정 자료 전체의 정보를 활용한 변인별 변화 유형의 탐색적 검토.
- 종단 자료의 장점을 활용한 변인 간 관련성에 대한 기술적 검토.
- 자료의 안정성 및 이상성에 대한 검토(오류 혹은 새로운 발견)



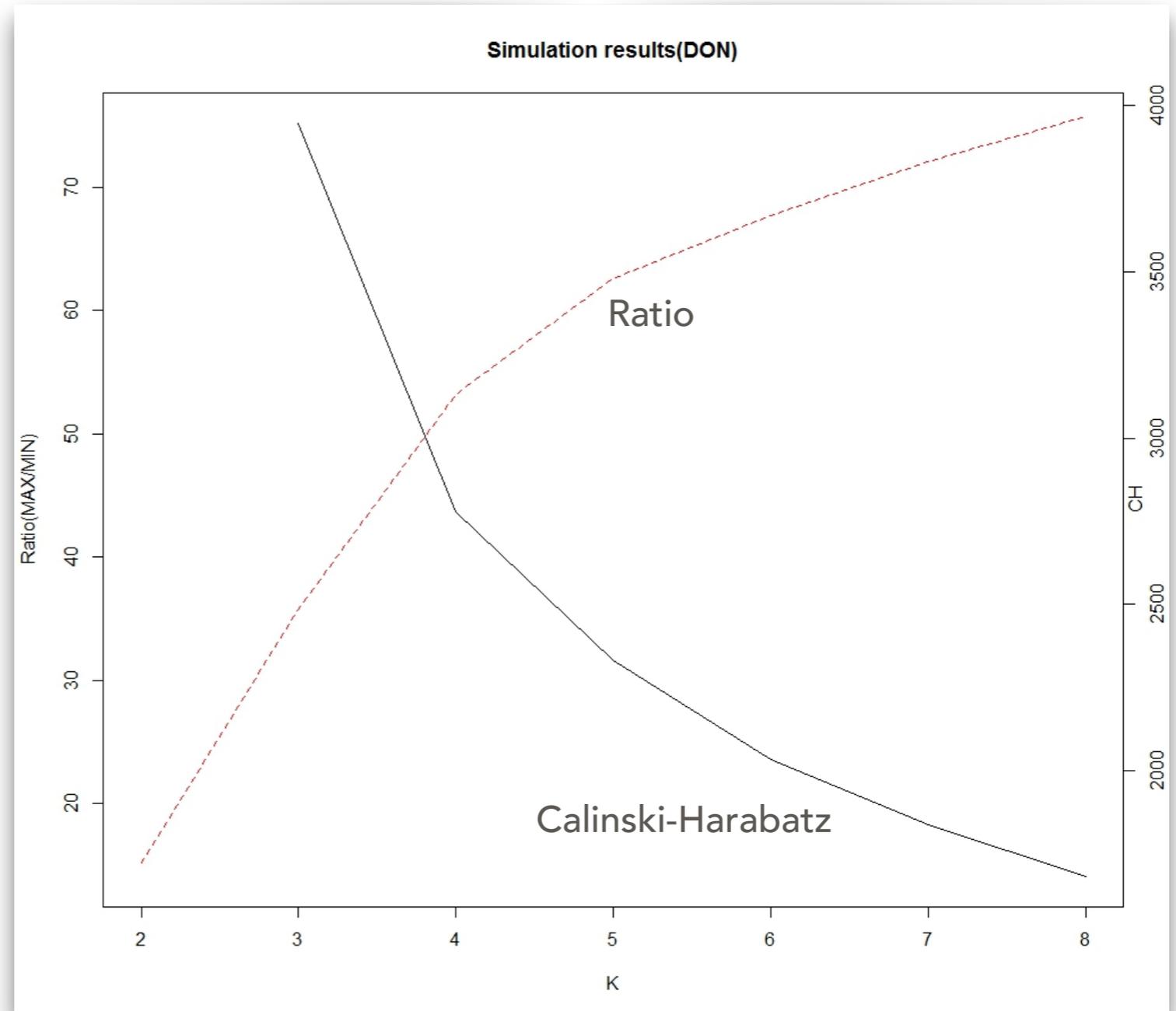
노법래(2015)



TC 분석 사례

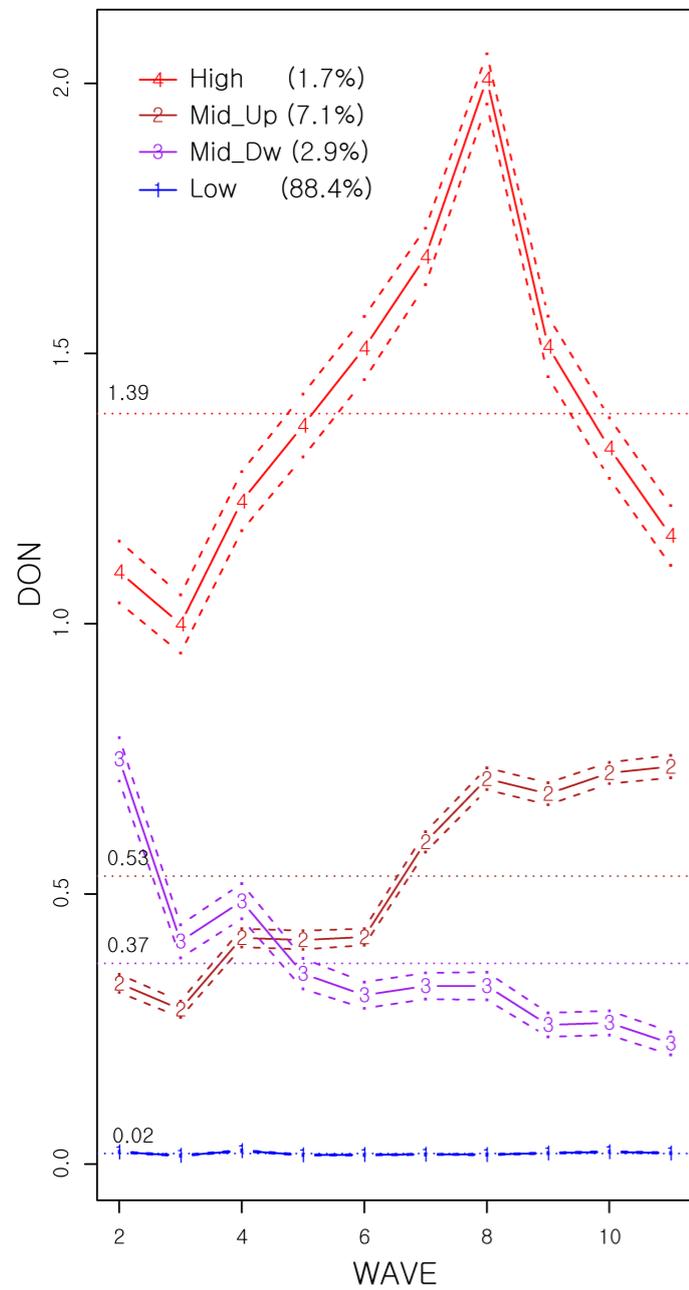
모형(유형수)의 결정

- 기부 수준(소득대비 기부액)의 종단 궤적에 대한 유형화 결과 시뮬레이션
- 조건당 10회 반복 후, 평균화
- Elbow-method 활용시, 최적 모형은 $k=4$ 로 판단.

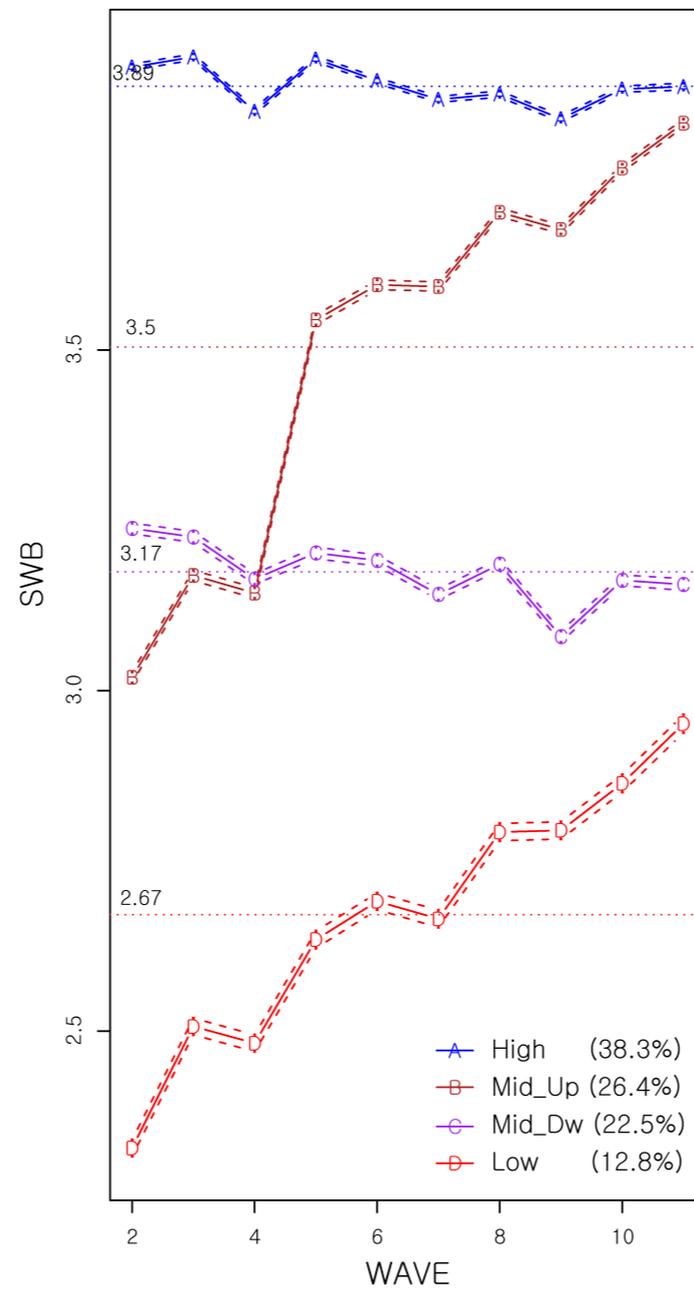


TC 분석 사례

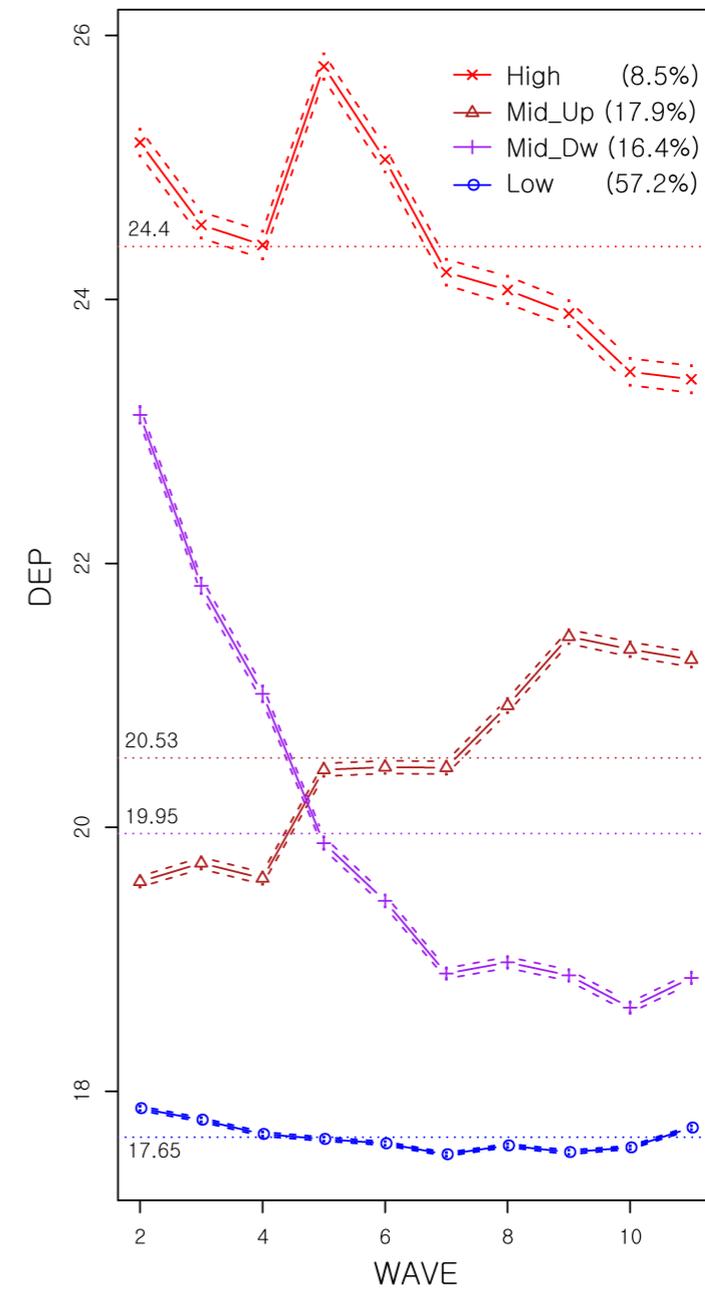
Don.Rates Level



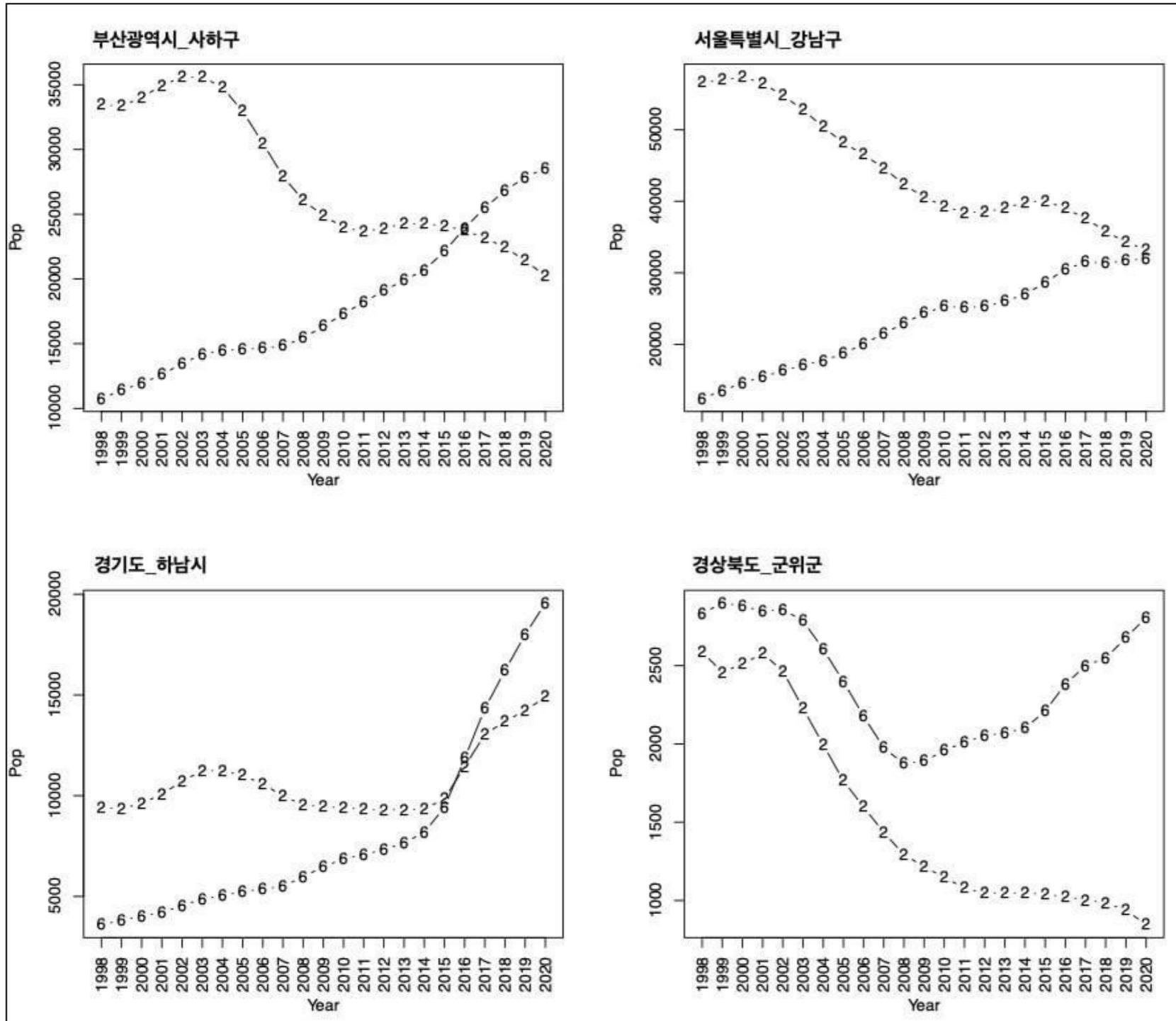
SWB



Depression



- 이런 경우는?



Joint Trajectory Analysis

결합궤적분석

- Measuring Distance : 관측시점을 j 라고 하고 측정변인을 X 라 할 때, 복수의 측정변인으로 구성된 두 관측 케이스 y_1, y_2 간 거리는,

$$Dist(y_{1..}, y_{2..}) = \sum_{j,X} \sqrt[p]{|y_{1jX} - y_{2jX}|^p}$$

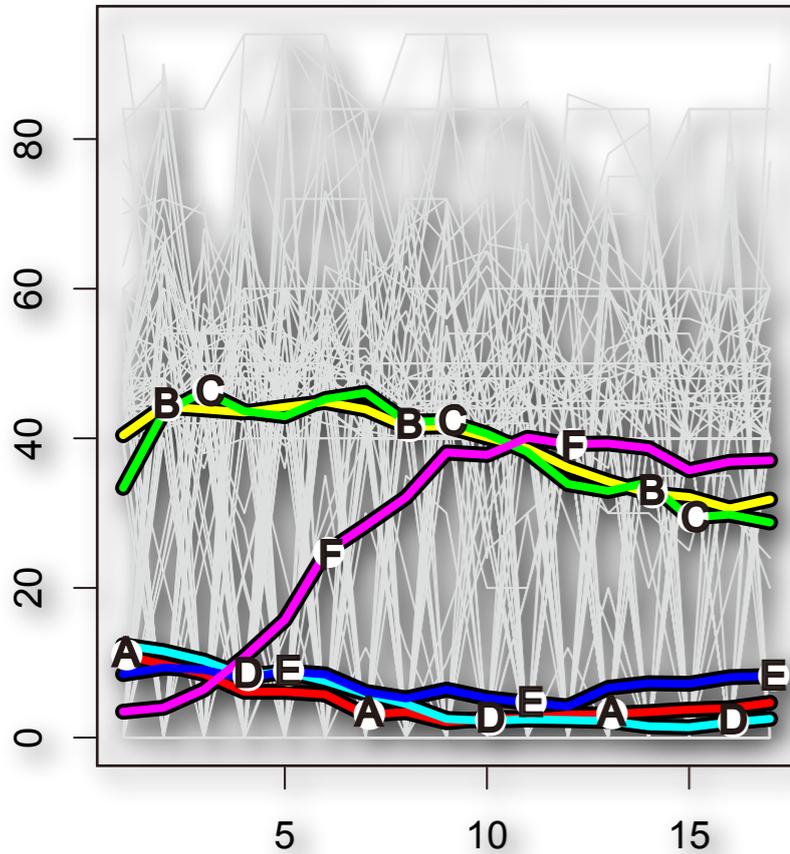
- [Normalization] : 스케일이 다른 복수의 궤적 정보를 활용할 때, 각 궤적을 표준화하는 것이 유형간 거리 계산에서 필요함. 측정값 y_{ijX} 의 표준화 값 y'_{ijX} 은, $y'_{ijX} = (y_{ijX} - \bar{y}_{..X})/SD_{..X}$
- k-means Classification and Calinsky & Harabasz creterion(1974):

$$CH(k) = M_{between}/M_{within} * (n - k)/(k - 1)$$

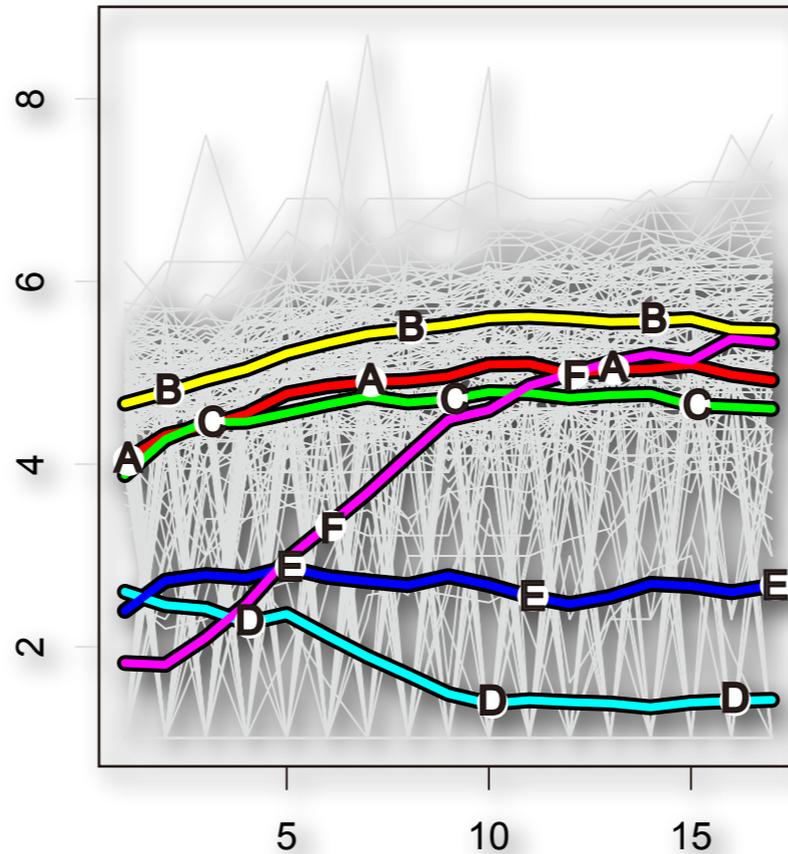
JTA 분석 사례

노동시간-임금-삶의만족도 결합궤적(남성)

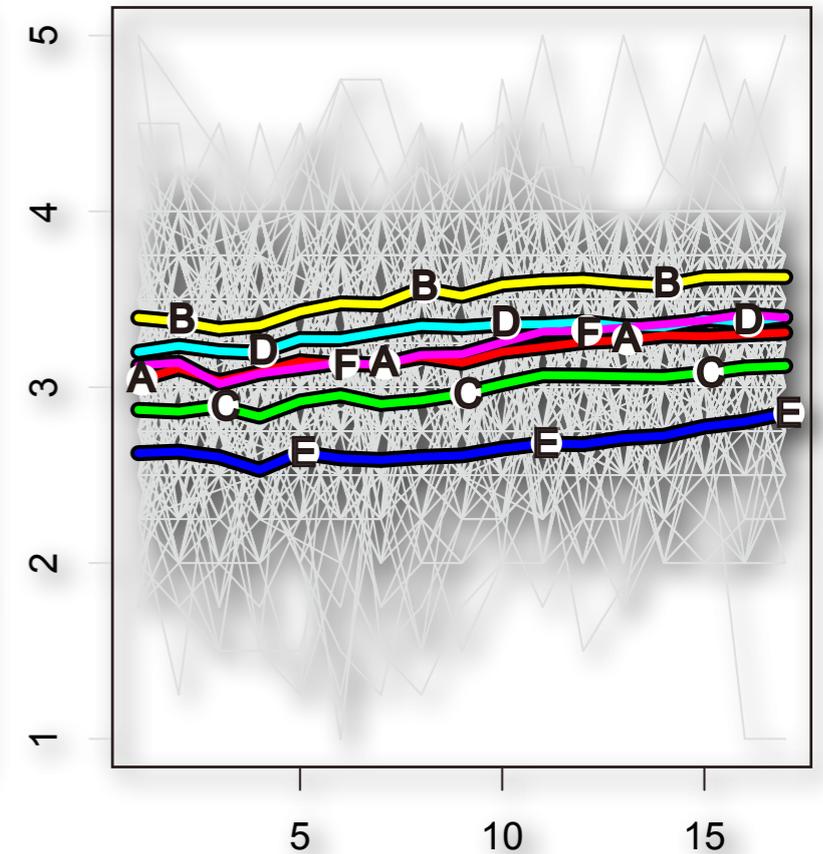
Avg. working time



Log wage



Life satisfaction



노법래(2017)

- JT A : 적은 노동시간 - 평균 이상 소득 - 중간 행복 (자영자형)?
- JT B : 많은 노동시간 - 가장 높은 소득 - 가장 높은 행복 (참여형1)?
- JT C : 많은 노동시간 - 평균 수준 소득 - 낮은 행복(참여형2)?
- JT D : 적은 노동시간 - 가장 낮은 소득 - 중간 행복 (은퇴형)?
- JT E : 적은 노동시간 - 낮은 소득 - 가장 낮은 행복(배제형)?
- JT F : 노동시간 상승 - 소득 상승 - 중간 행복(신규진입형)?

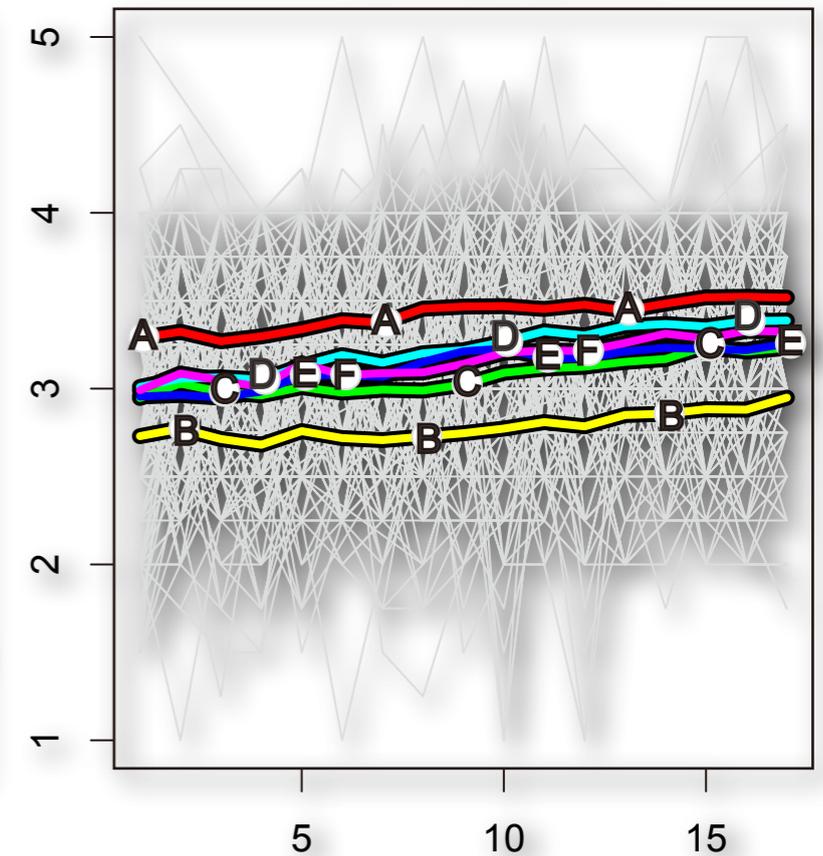
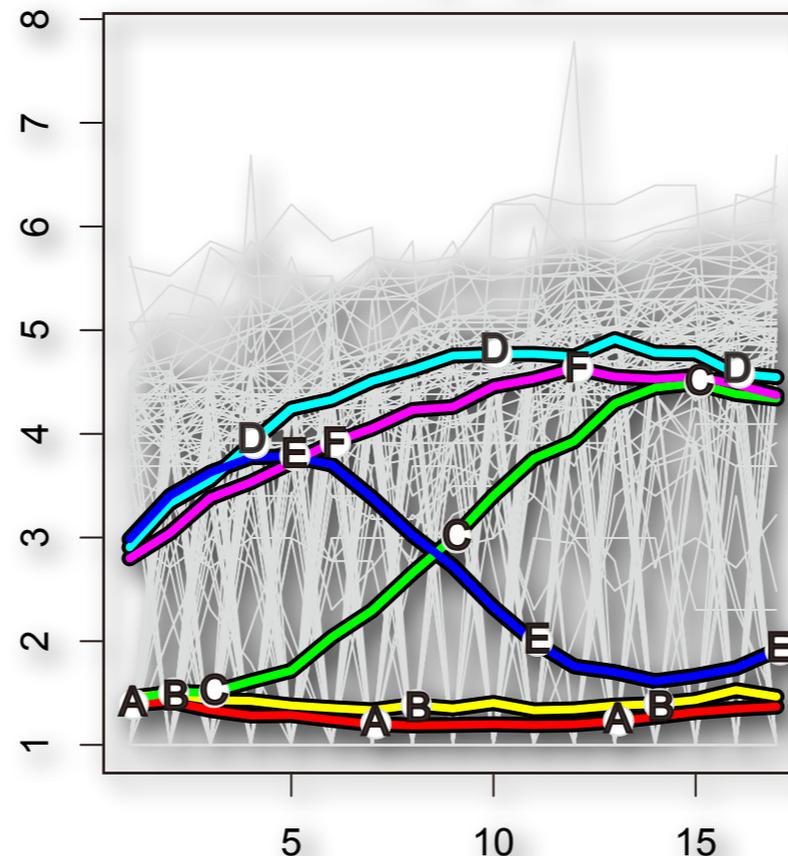
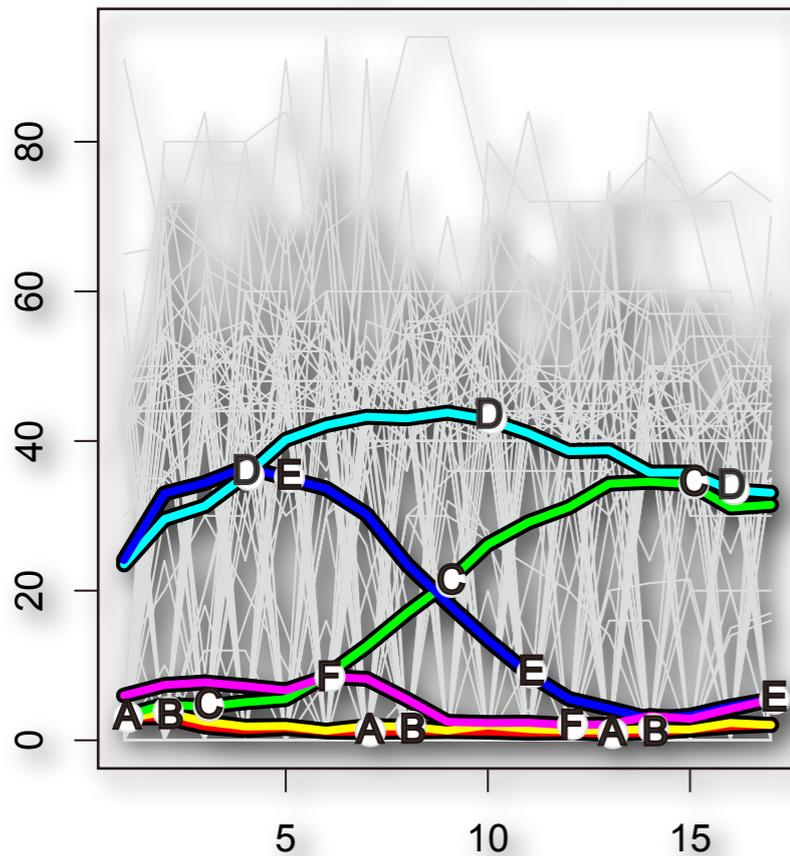
JTA 결과

노동시간-임금-삶의만족도 결합궤적(여성)

Avg. working time

Log wage

Life satisfaction



Wave

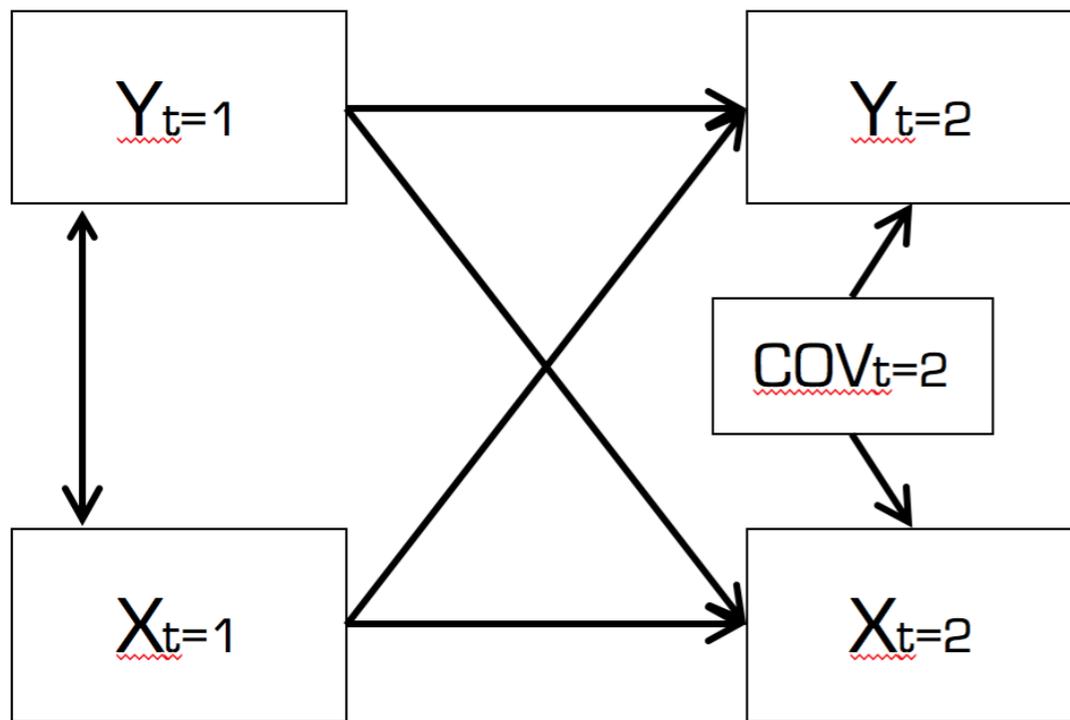
노법래(2017)

- JT A : 적은 노동시간 - 낮은 소득 - 가장 높은 행복(미참여1)?
- JT B : 적은 노동시간 - 낮은 소득 - 가장 낮은 행복(미참여2)?
- JT C : 노동시간 증가- 소득 증가 - 평균적 행복참(신규진입형)?
- JT D : 가장 많은 노동시간 - 가장 높은 소득 - 중간 행복 (참여형)?
- JT E : 노동시간 감소 - 소득 감소 - 평균적 행복(이탈형)?
- JT F : 적은 노동시장 - 높은 소득 - 중간 행복(자영자형)?

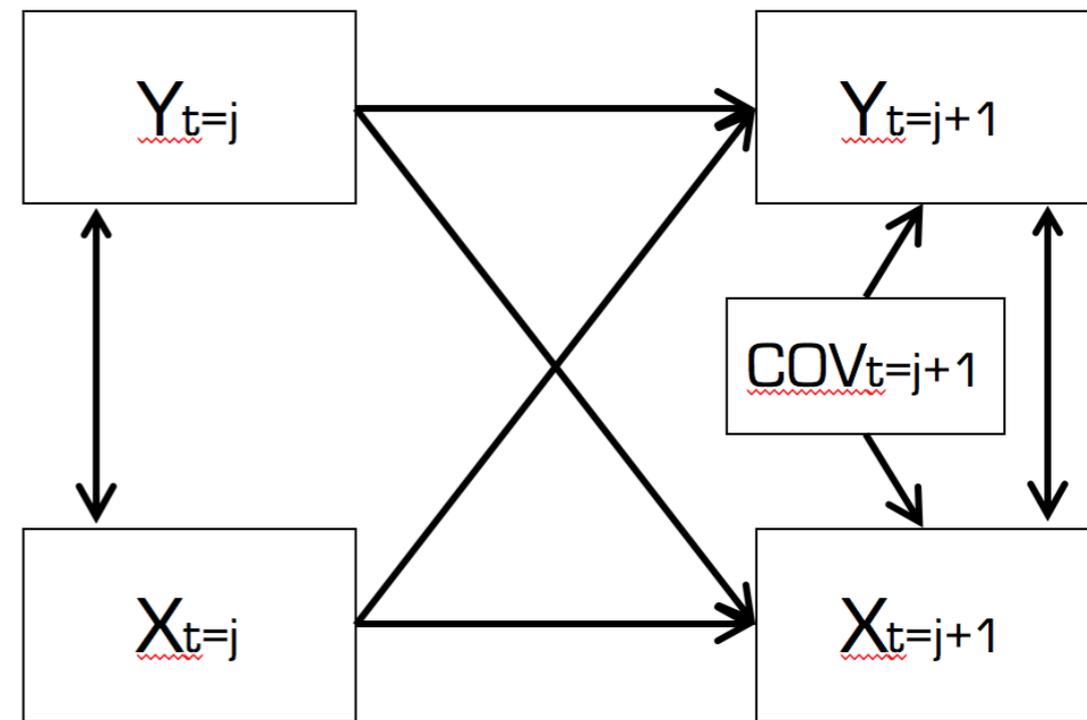
Longitudinal SEM

분석 개요(CLPM 예시)

Cross-lagged panel model



$t = \{1, 2\}$

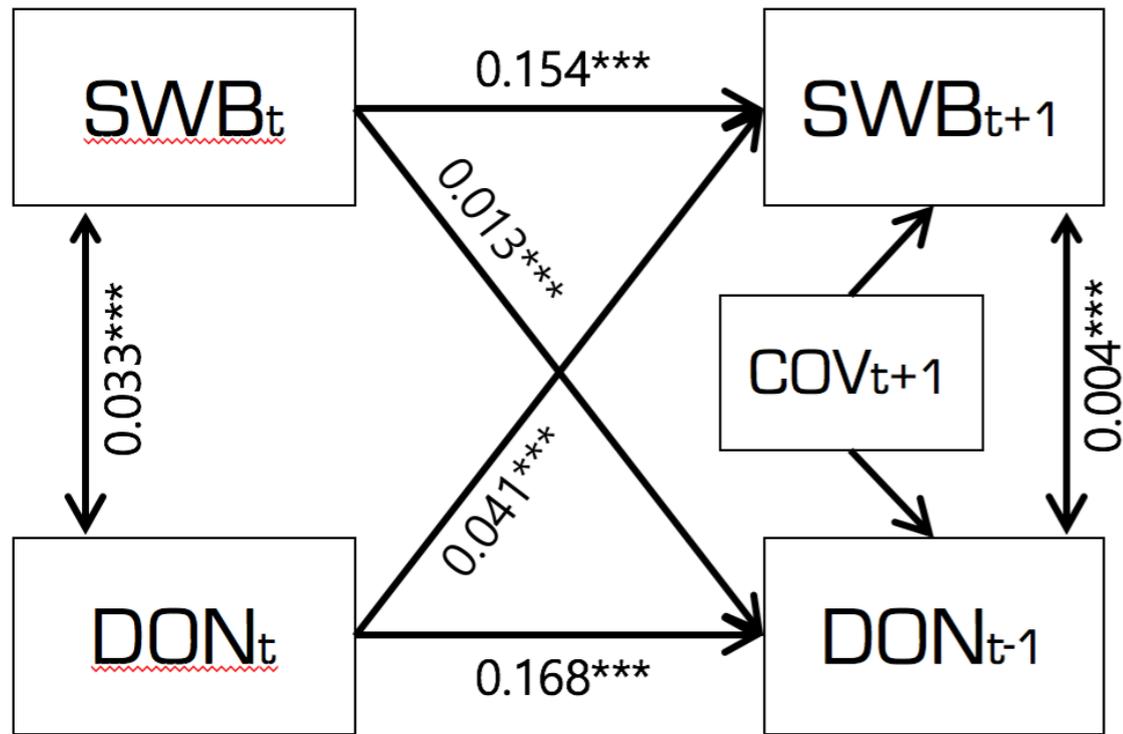


$t = \{1, 2, \dots, j, \dots, k, k+1\}$

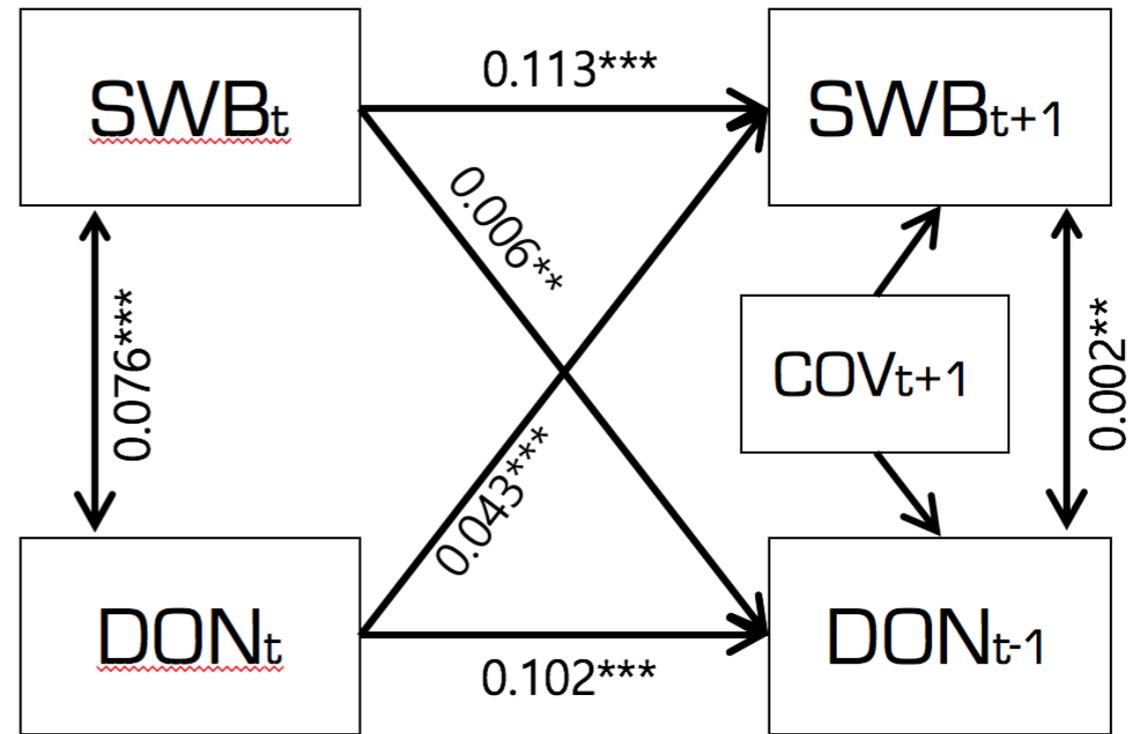
L-SEM 분석 사례

기부와 삶의 만족도

Results(SWB ~ DON)



Non-weighted

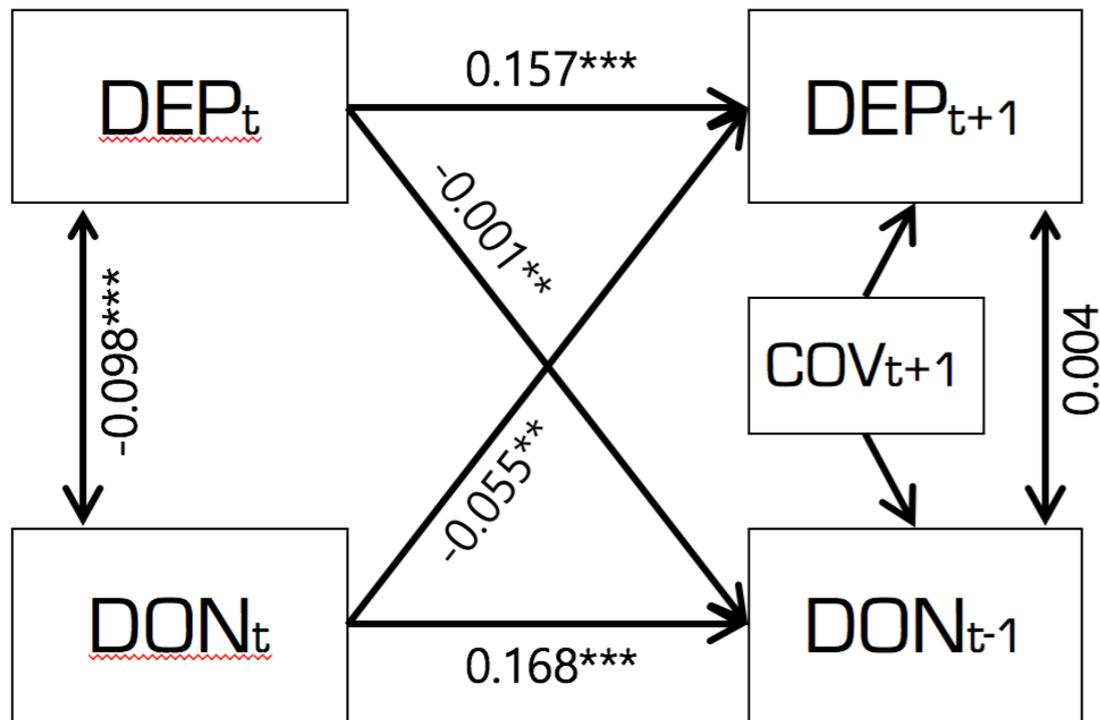


Weighted

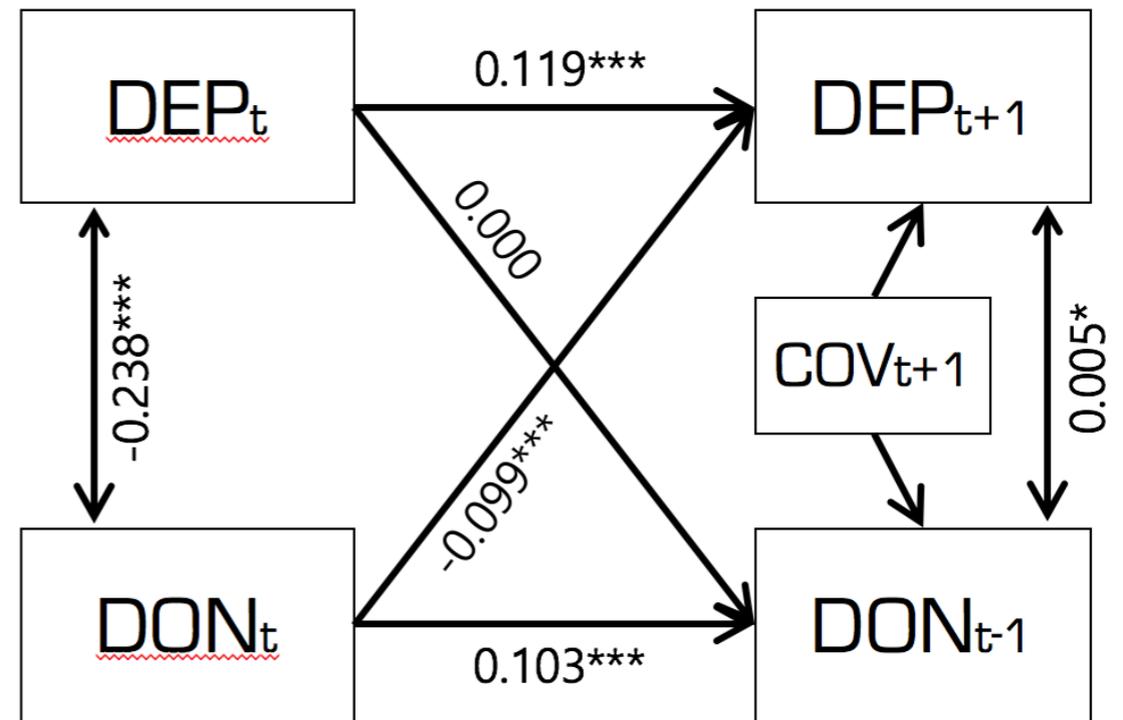
LSEM 분석 사례

기부와 우울

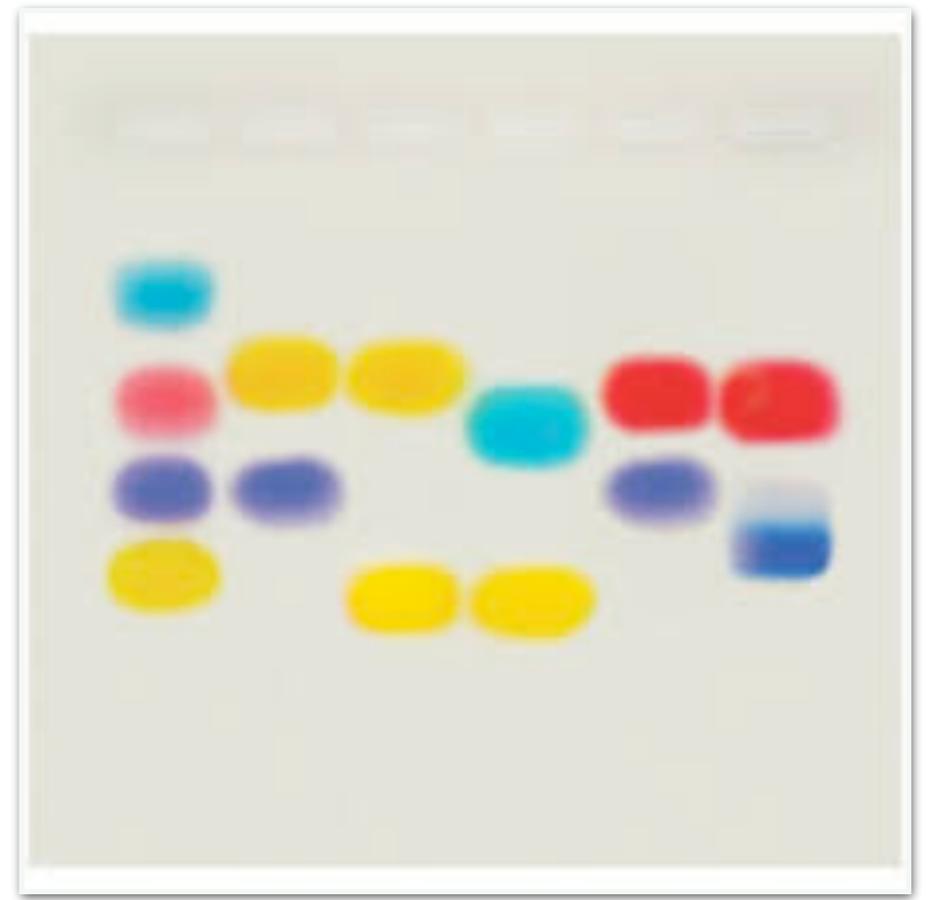
Results(DEP ~ DON)



Non-weighted



Weighted

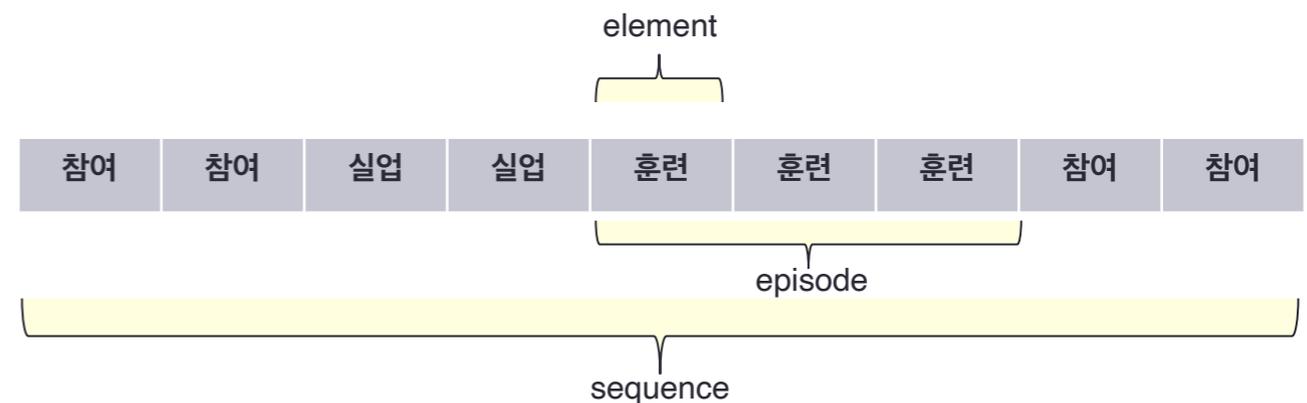


분석 모형과 사례

Sequence

Sequence Analysis

- **Sequencing** : define “event” under one-dimensional vector(or single-channel)



- **Optimal matching(OM)**

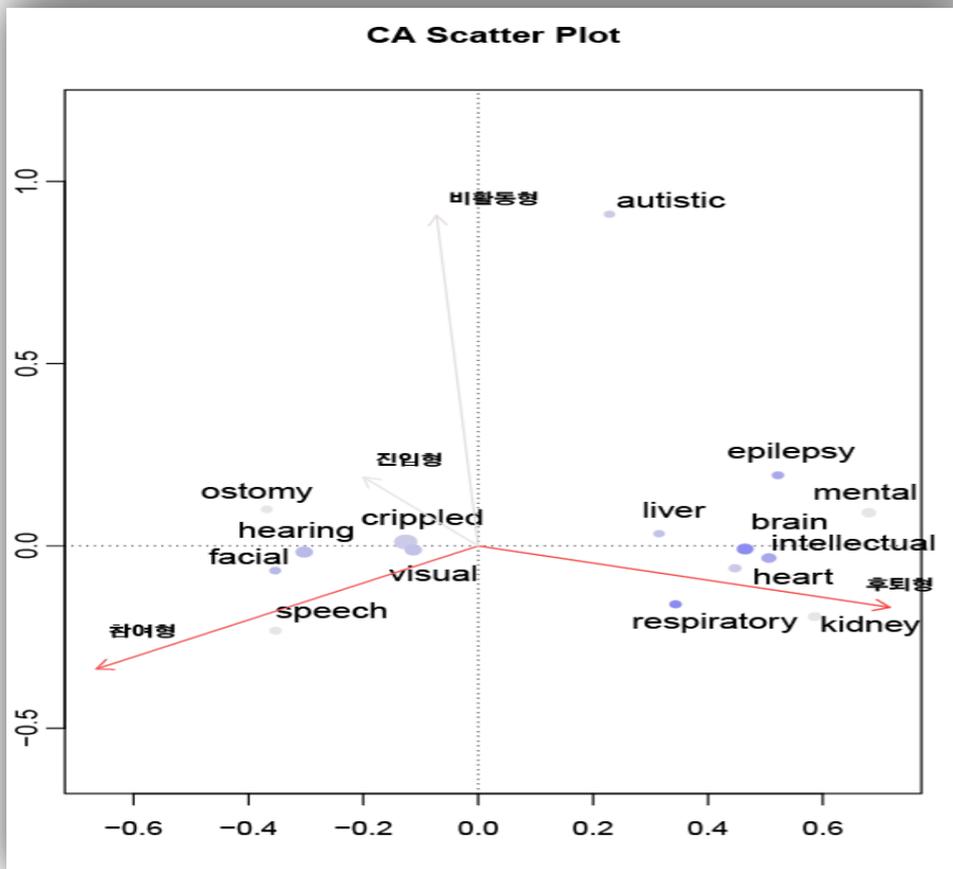
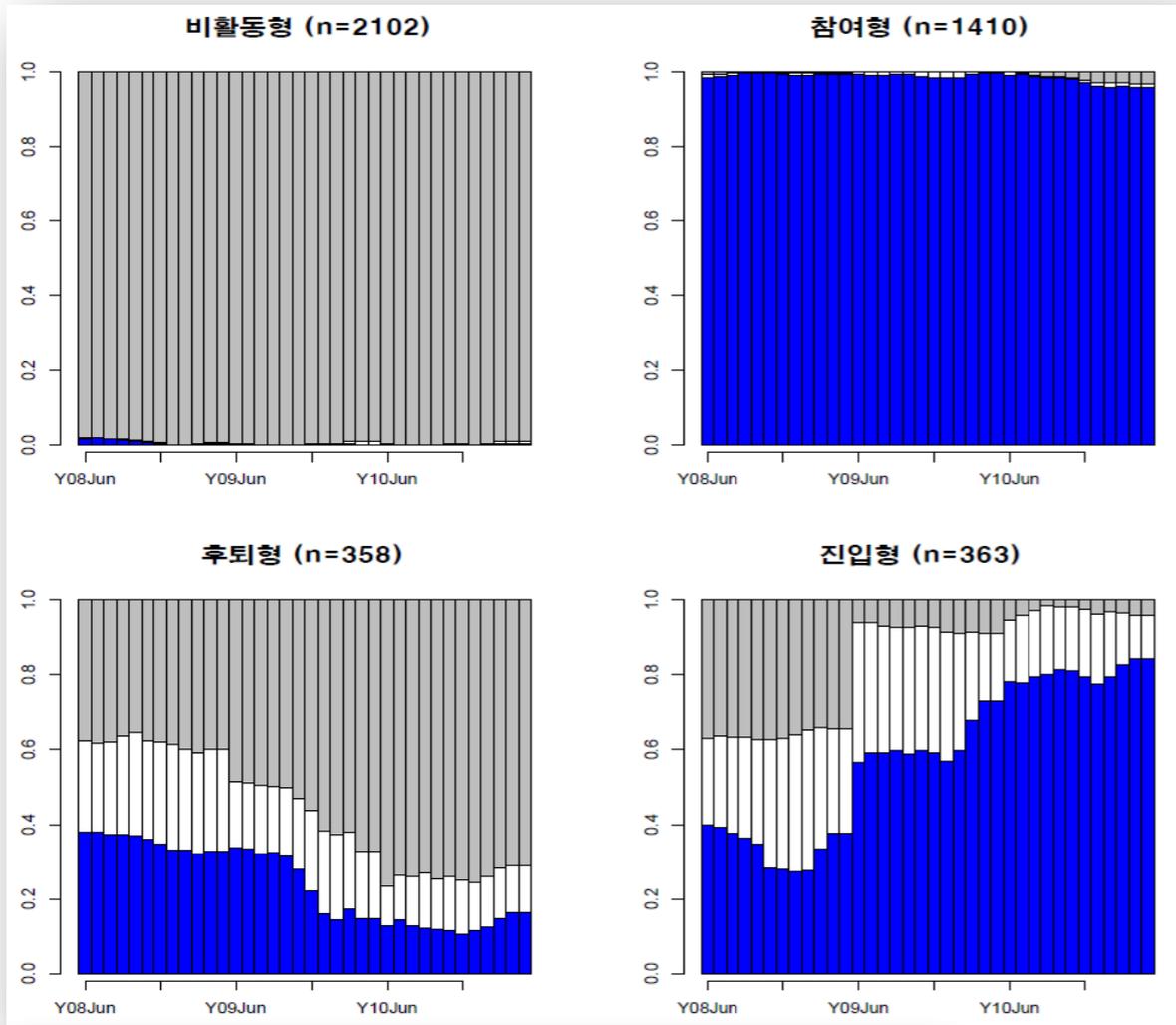
$$\min(Diff_{S_1, S_2} = cost(insertion, deletion, substitution) \mid Diff_{S_1, S_2} = 0)$$

- **Clustering Analysis(hierarchical, k-means)**

Sequence Analysis

- 시퀀스 분석의 일반적 과정
 - 묘사(description)
 - 시각화(visualization)
 - 비교와 군집화(comparison & grouping)
 - 확증적 분석(confirmatory analysis)

SA 분석 사례



	비활동형 / 참여형	진입형 / 참여형	후퇴형 / 참여형
	Exp(b)	Exp(b)	Exp(b)
(ref.=감각장애)			
지체장애	1.377 **	0.987	1.354
기타내부장애	2.486 ***	1.277	1.658
정신장애	4.066 ***	1.519	2.322 ***
(ref.=20s)			
30s	0.622	0.498 *	0.602
40s	0.548 *	0.477 *	0.525
50s	0.883	0.427 **	0.662
60s	2.054 *	0.447 *	1.307
여성 (ref.=남성)	5.522 ***	2.327 ***	2.807 ***
중증 장애 (ref.=경증)	2.161 ***	0.925	0.955
로그 균등화소득	1.17 ***	1.072 **	1.141 ***
수급경험	1.685 ***	1.074	1.385 ***
교육연수	1.051 ***	1.014	1.034
주관건강 좋음(ref.=건강나쁨)	0.38 ***	0.943	0.818
직업훈련경험	0.842	2.591 ***	2.353 ***
차별경험	1.508 ***	1.151 *	1.444 ***

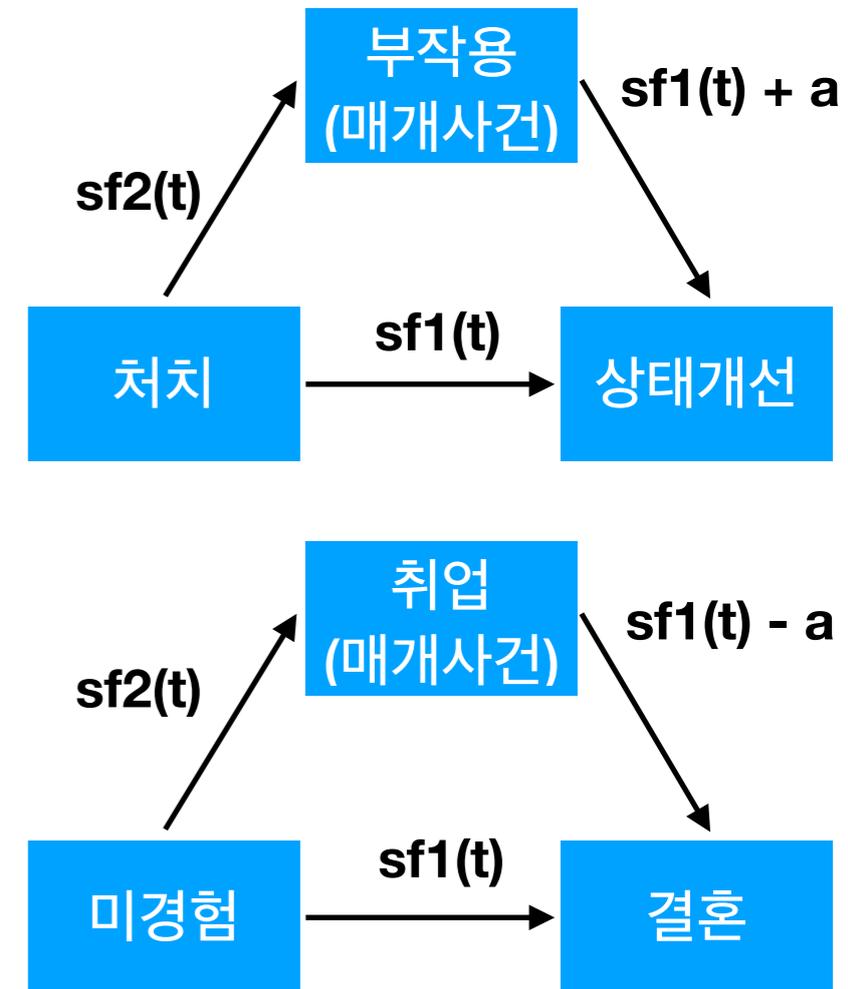
Multi-state Model

- Multi-state Model(Putter et al. 2007)

Transition intensity: $\lambda_{ij}(t) = \lim_{\Delta t \downarrow 0} \frac{\text{Prob}(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$

Cumulative hazard: $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s) ds$

- 두 가지 이상의 상태 변화에 적용
- 사건 발생(상태 변화)의 시간에 따른 모형화
- “매개사건” 다루기(이후 사건 발생 확률에 영향을 미치는 사전 변화)
- missing(혹은 절단) 데이터에 대한 추정 기법



Multi-state Model

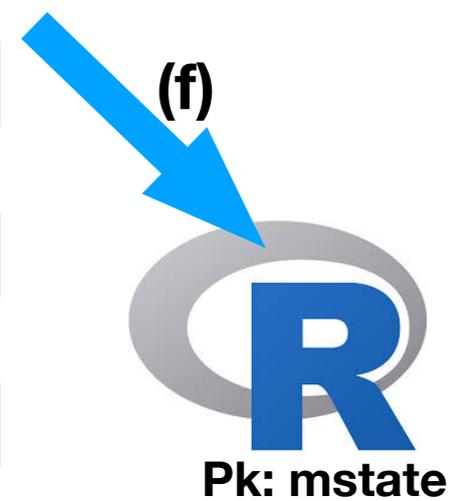
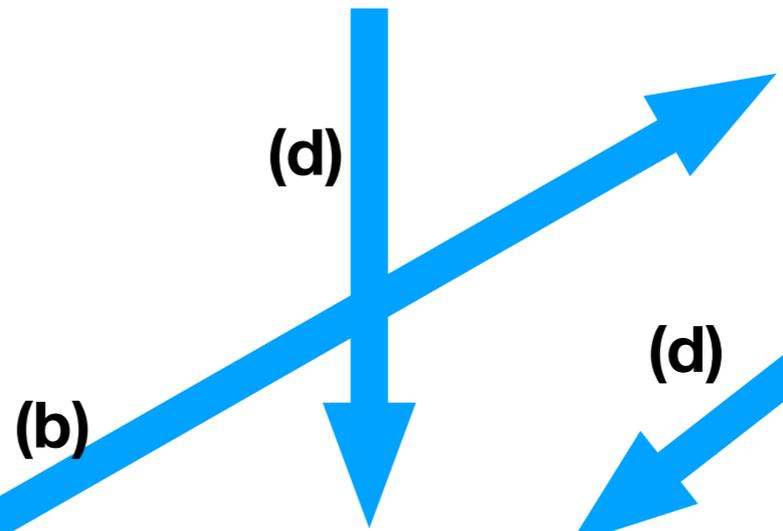
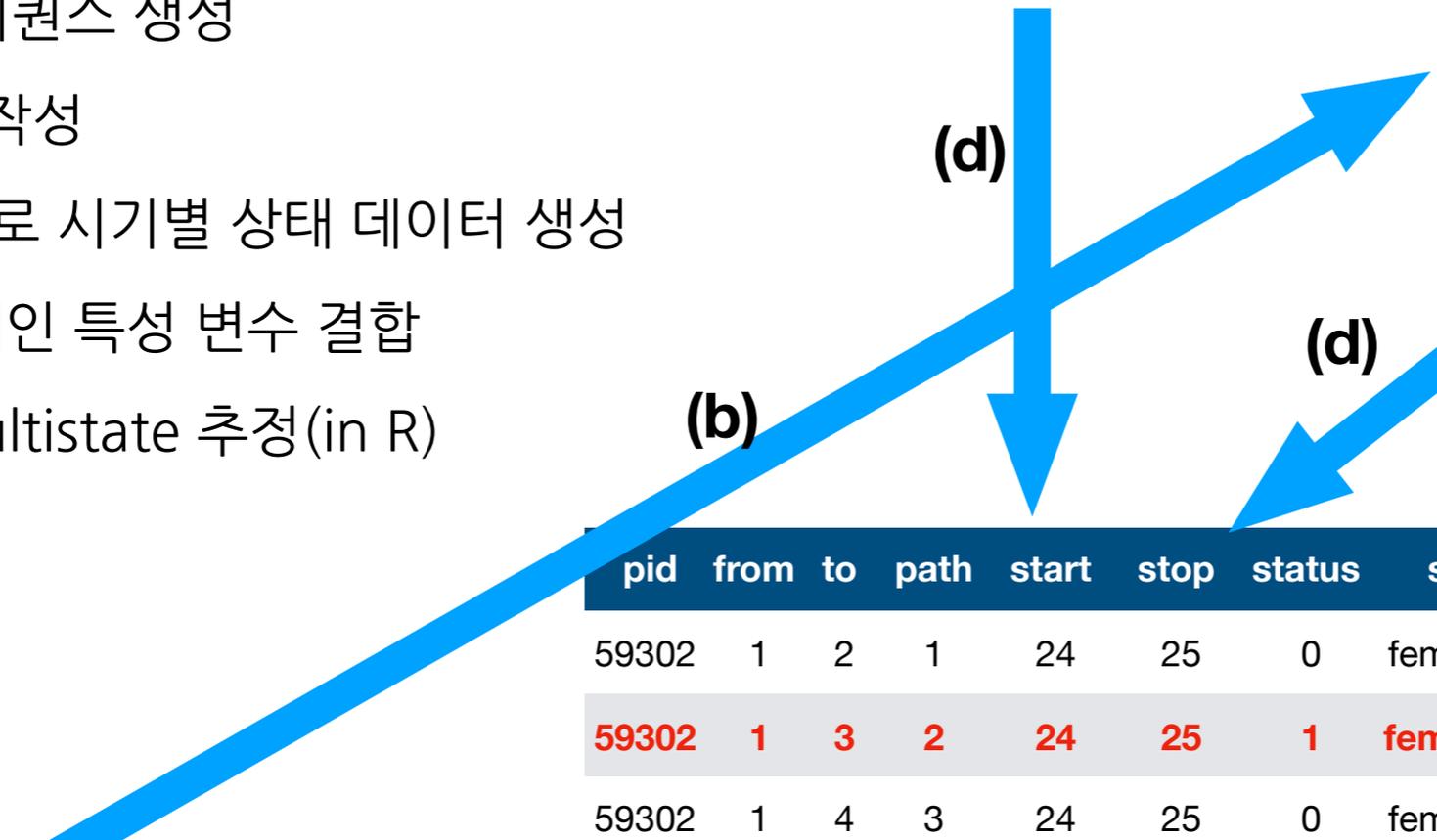
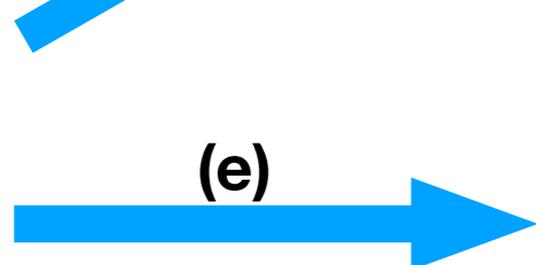
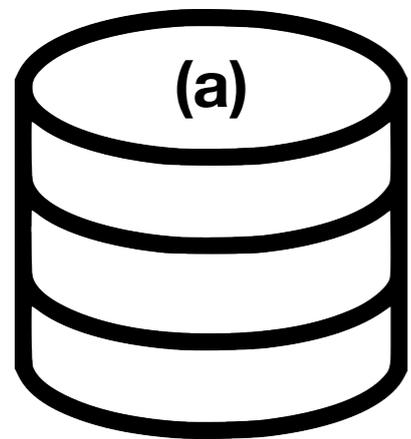
- (a) Long-type 자료상에서 상태 판정 변수 만들기
- (b) 개인별 (상태)시퀀스 생성
- (c) 전환 매트릭스 작성
- (d) (b), (c) 바탕으로 시기별 상태 데이터 생성
- (e) (d) 데이터에 개인 특성 변수 결합
- (f) (e) 활용하여 multistate 추정(in R)

(c)

		to					
		emc	eMc	Emc	WMc	wmC	WMC
from	emc	-	1	2	3	4	5
	eMc	NA	-	NA	6	7	8
	Emc	NA	NA	-	9	NA	10
	WMc	NA	NA	NA	-	NA	11
	wmC	NA	NA	NA	NA	-	12
	WMC	NA	NA	NA	NA	NA	-

pid	seq
254201	66666666666666..
456001	66666666666666..
310901	66666666666666..
456501	66666666666666..
456201	666666666666....
309003	333255556.....
4802	333225555555...
268101	3332224422255..
662002	333133333333...
365404	.33333466666666.
690203	.33333466666666.

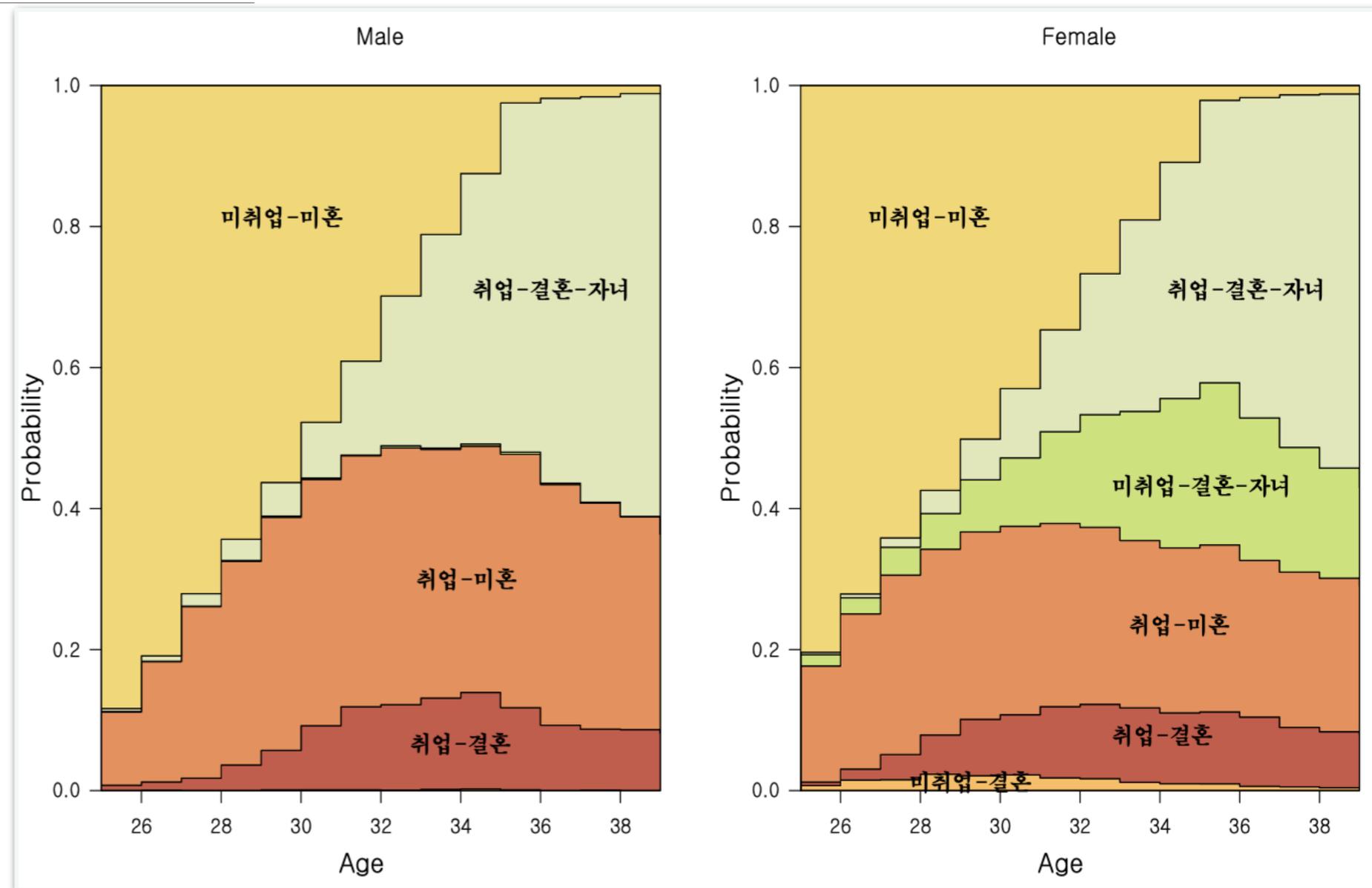
pid	from	to	path	start	stop	status	sex	birth
59302	1	2	1	24	25	0	female	1982
59302	1	3	2	24	25	1	female	1982
59302	1	4	3	24	25	0	female	1982
59302	1	5	4	24	25	0	female	1982
59302	1	6	5	24	25	0	female	1982
59302	3	4	9	25	32	1	female	1982
59302	3	6	10	25	32	0	female	1982
59302	4	6	11	32	34	1	female	1982



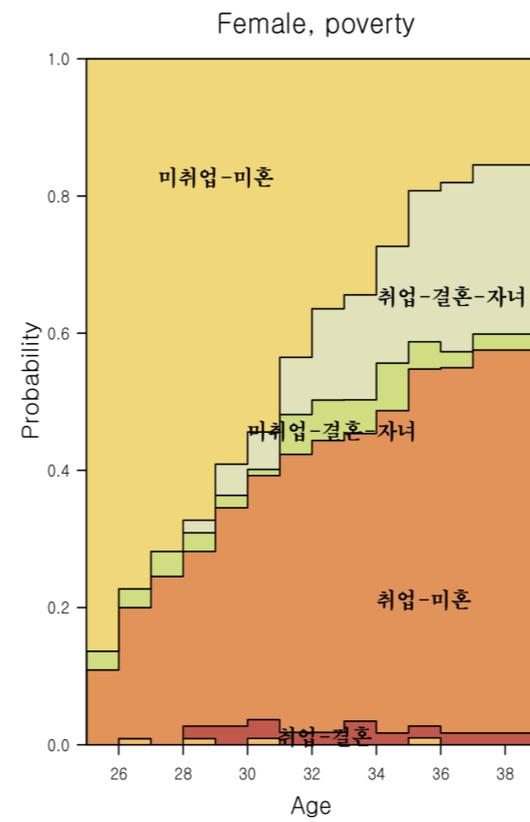
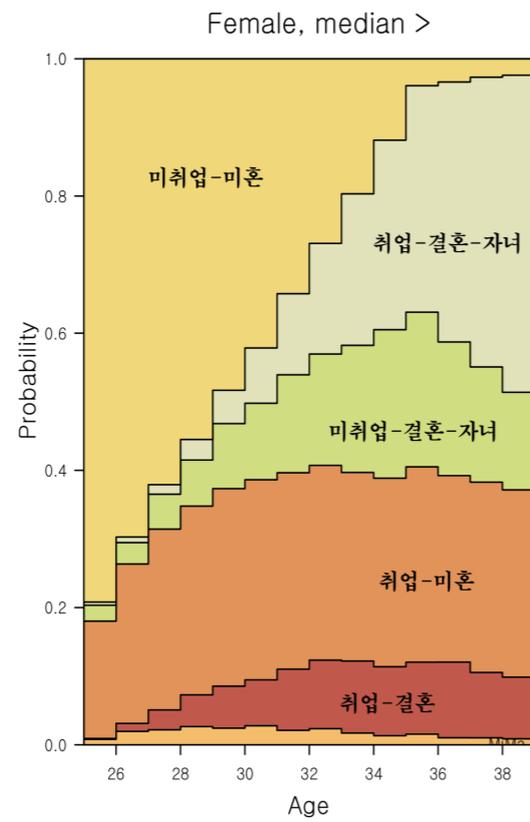
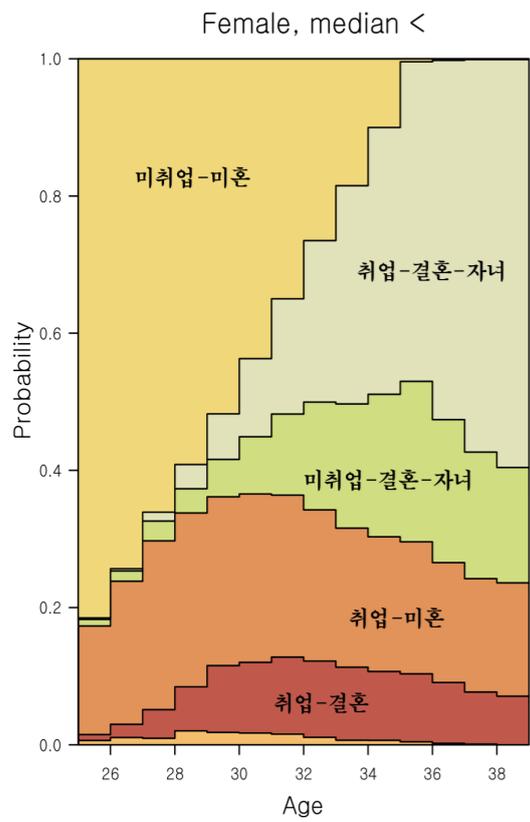
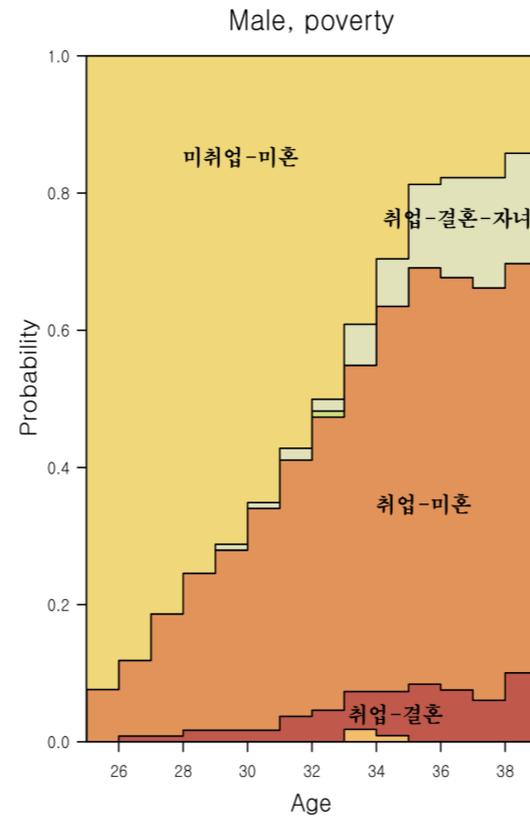
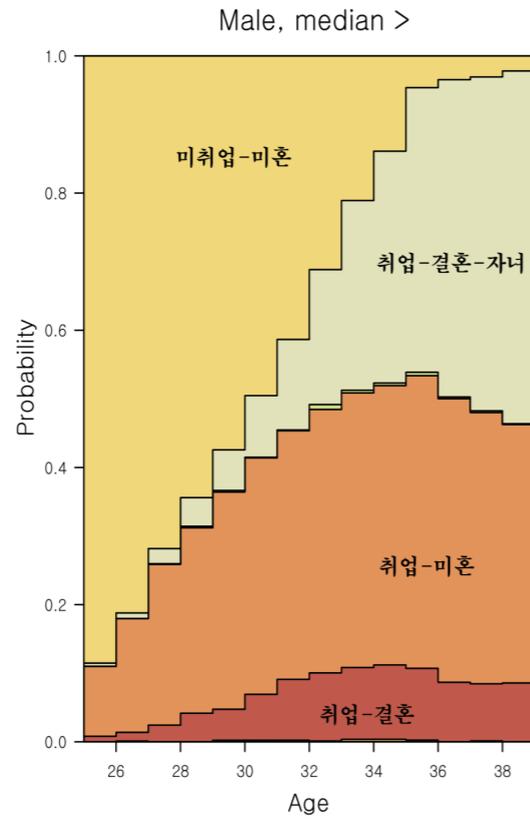
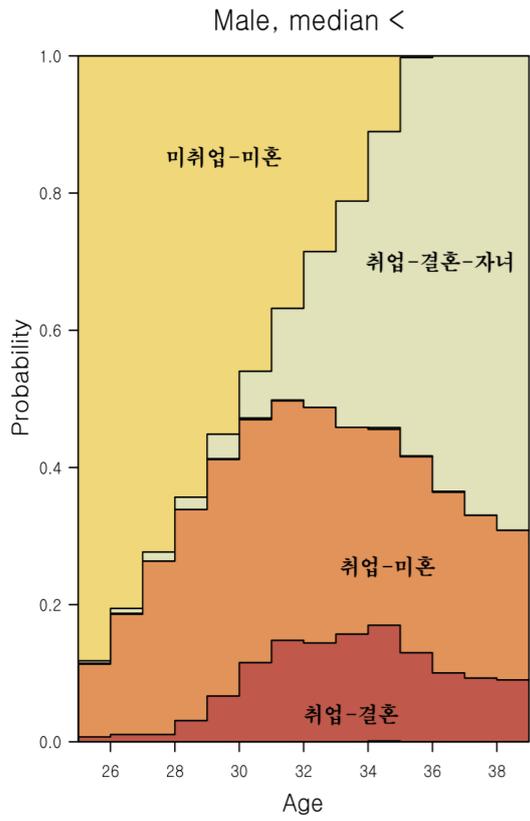
Pk: mstate

MSM 분석 사례

성별	연령	상태(%)					
		미취업* 미혼	미취업* 결혼	취업* 미혼	취업* 결혼	미취업* 결혼* 자녀	취업* 결혼* 자녀
남성	25	88.35	0.00	10.41	0.76	0.06	0.41
	30	47.77	0.12	34.93	9.09	0.18	7.91
	35	2.47	0.13	35.93	11.64	0.31	49.53
	39	1.14	0.00	28.22	8.17	0.06	62.41
여성	25	80.38	0.72	16.45	0.5	1.61	0.33
	30	42.98	2.23	26.73	8.53	9.70	9.82
	35	2.10	0.97	23.69	10.19	23.00	40.06
	39	1.12	0.20	21.19	7.63	13.94	55.92

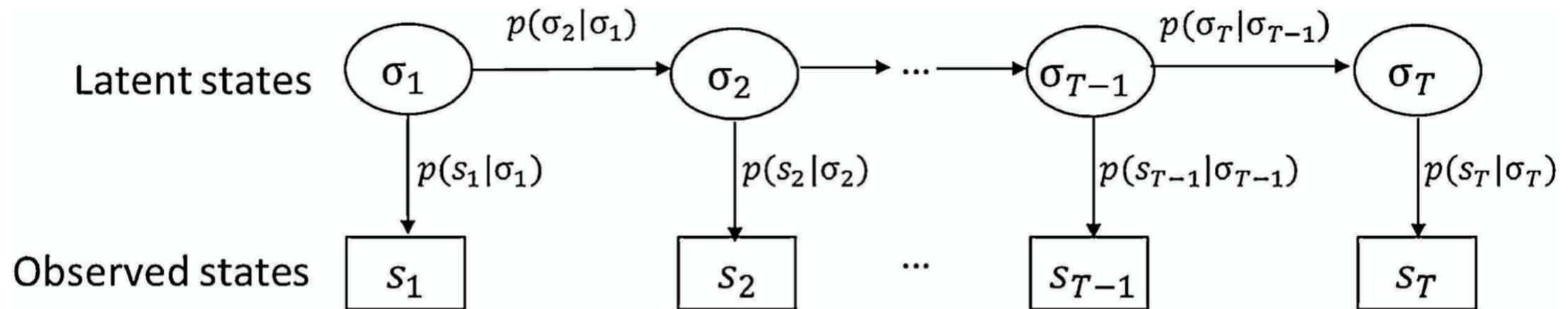


MSM 분석 사례



Hidden Markov Model

- Hidden Markov Model(HMM)

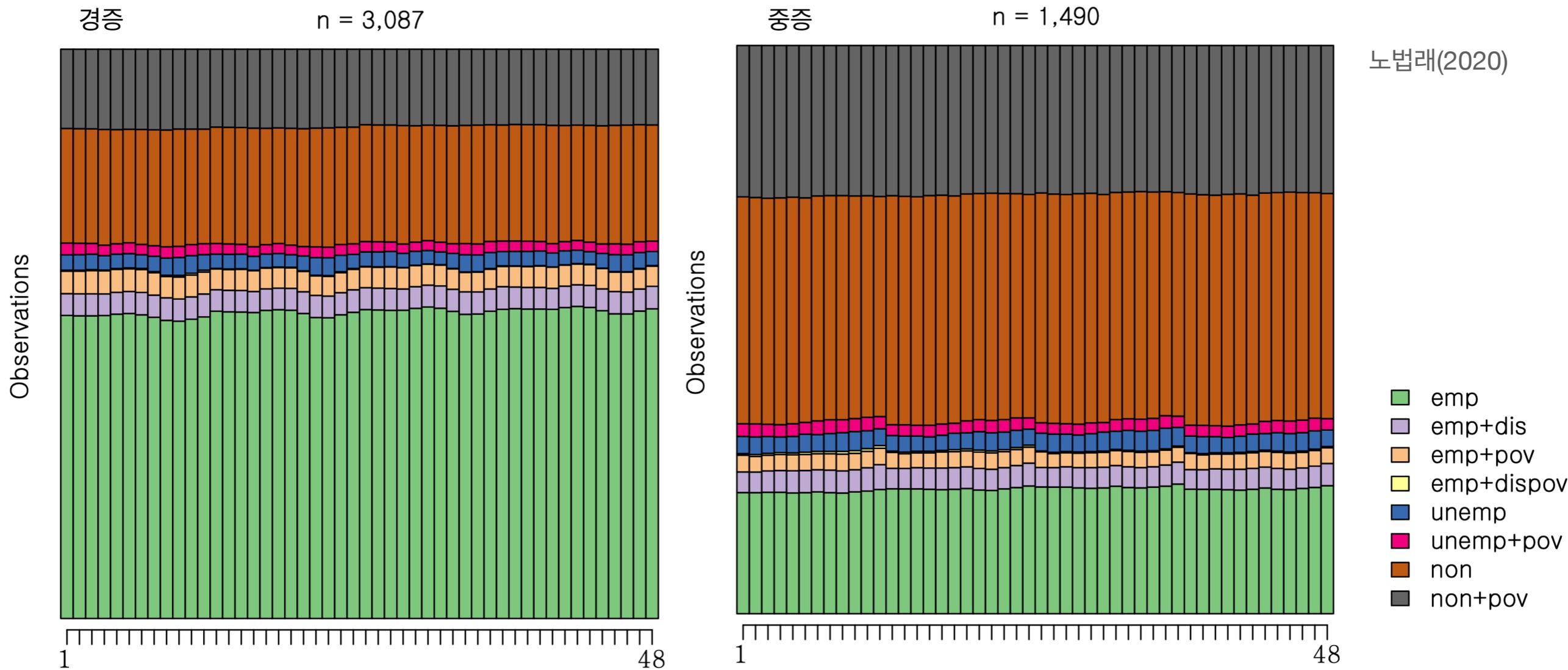


[그림 1] HMM의 도식화(Piccarreta & Stuber, 2019:10)

- Mixture HMM
 - Status의 이행 패턴(sequence)를 유형화하고, 유형에 속할 확률에 영향을 미치는 요인(covariates)의 영향에 대한 검토를 포함.
 - 단일한 상태(status)가 아닌 시퀀스가 결과 변수가 됨.

HMM 분석 사례

장애심도에 따른 고용 이행 경로



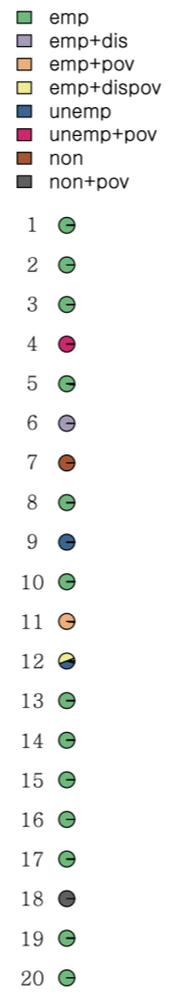
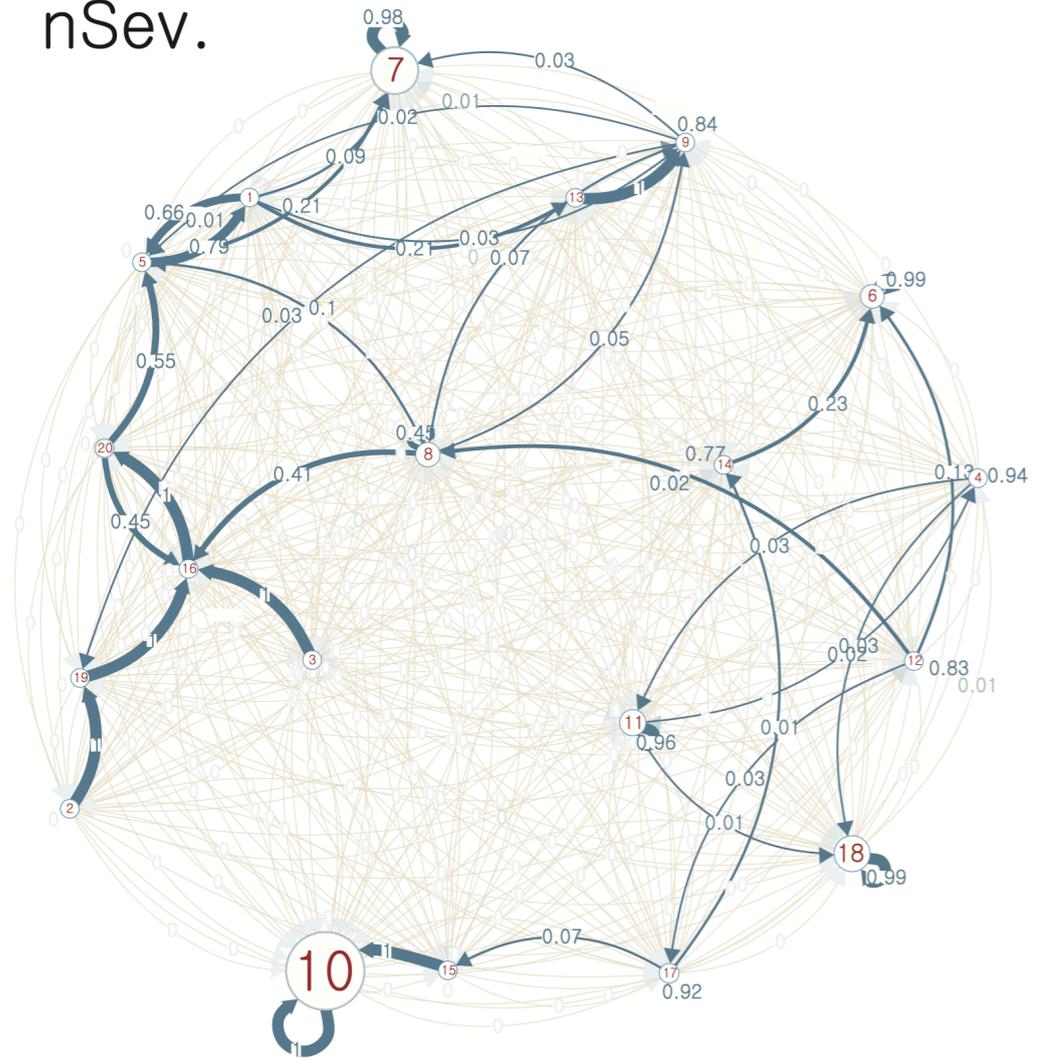
노법래(2020)

- 장애심도(경증/중증)에 따라 고용 상태에서 전반적인 차이가 있는 것으로 나타남.
- 중증 장애 집단의 경우 고용 상태인 집단보다 비경활 집단의 비중이 컸으며, 고용 상태인 경우도 빈곤이나 차별을 경험하는 집단이 상대적으로 큰 비중을 차지하는 것으로 관찰됨.

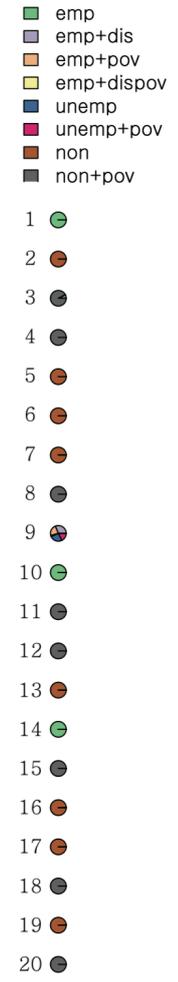
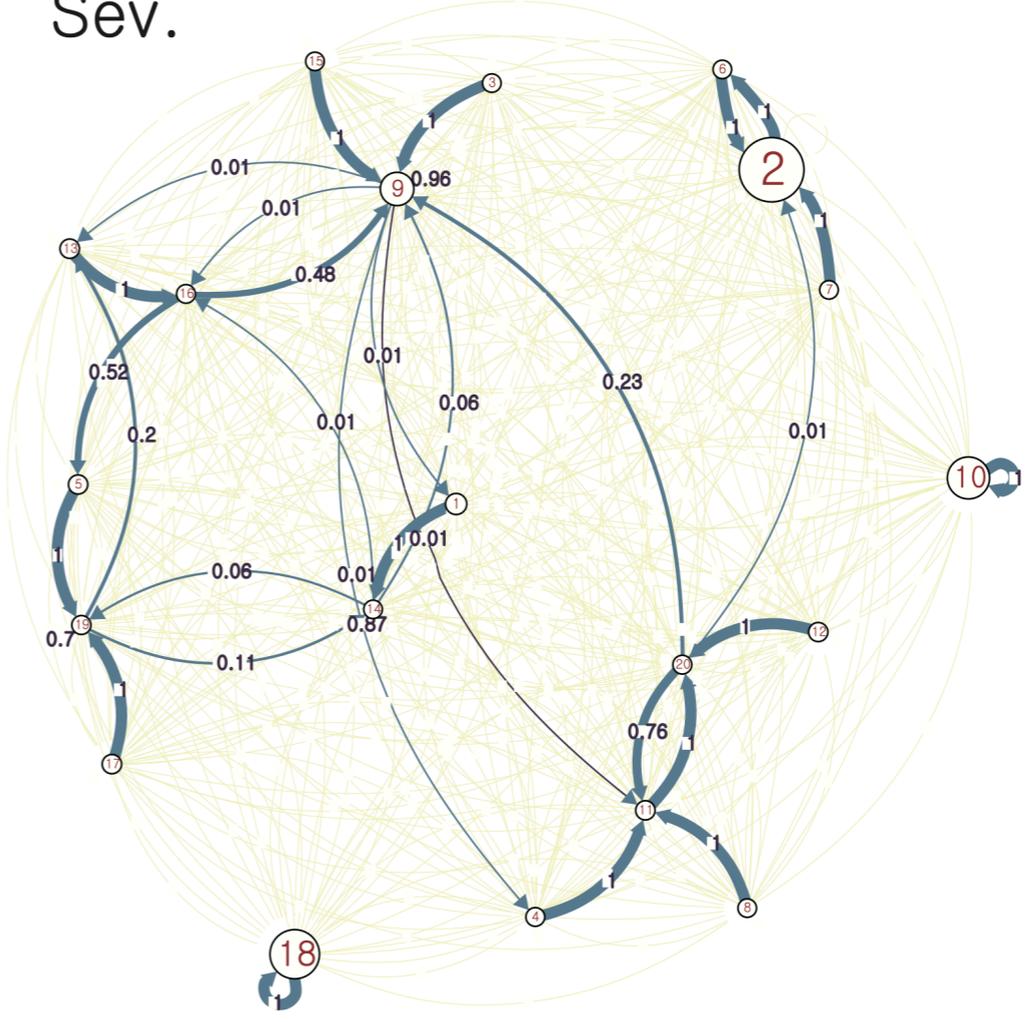
HMM 분석 사례

장애심도에 따른 고용 이행 경로

nSev.



Sev.



노법래(2020)

- 경증 집단(nSEV)은 중증 집단(Sev)에 비해 고용의 질이 높고 안정적인 고용을 유지하는 경로가 더욱 명확하게 관찰됨.
- 중증집단의 비경활 상태에서 머물거나(->2), 비경활과 빈곤을 동시에 경험하는 경로에 머물러 있는 집단이 다수였음. 상태 이행을 보이는 집단은 주로 불안정한 노동과 실업을 동시에 경험하는 상태로의 이행(->9)이 주요함.
- 예외적으로 안정적인 고용 지위를 유지하거나(10), 여기로 이행하는 경로(->14)도 소수 관찰됨.

HMM 분석 사례

Mixture Hidden Markov Model

• 본 연구에서는 고용 상태의 시퀀스 유형을 크게 5개로 대별함.

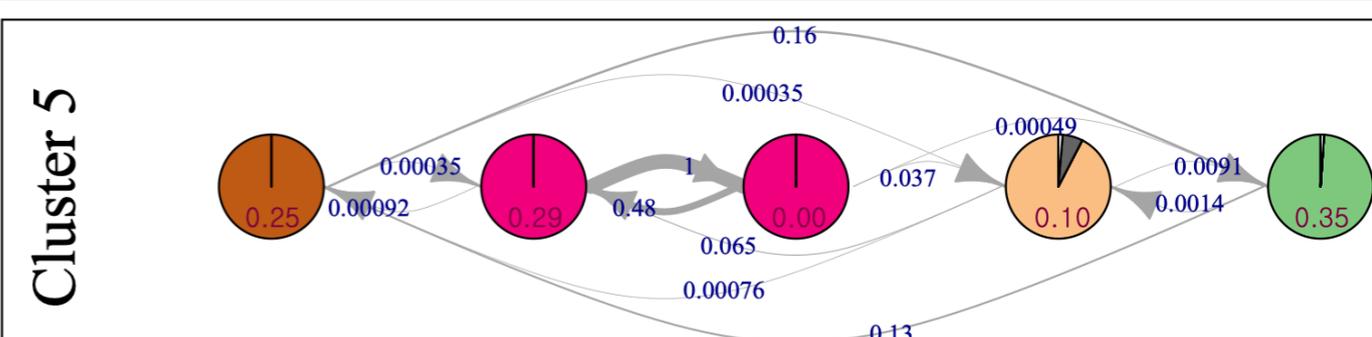
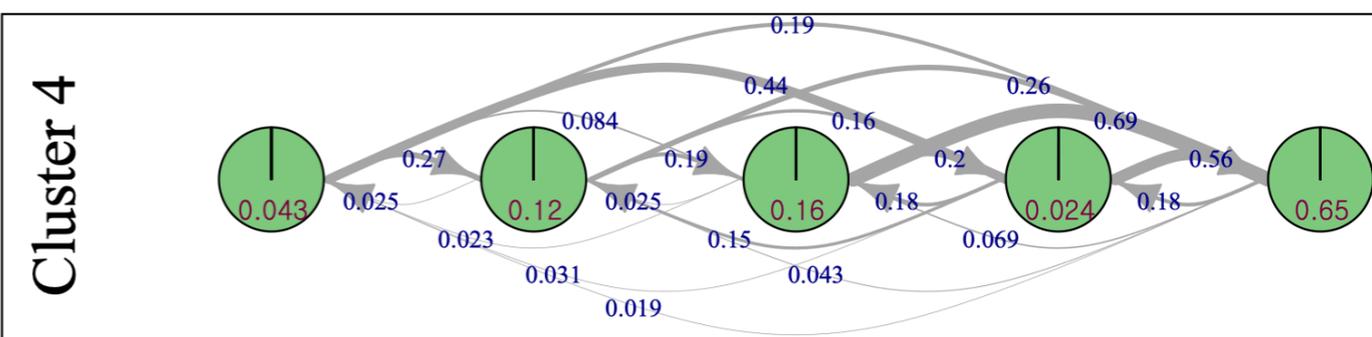
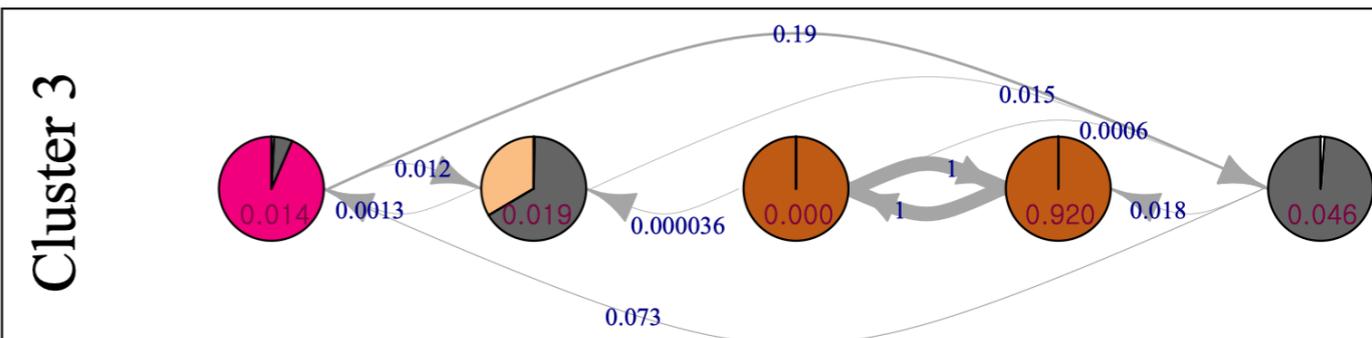
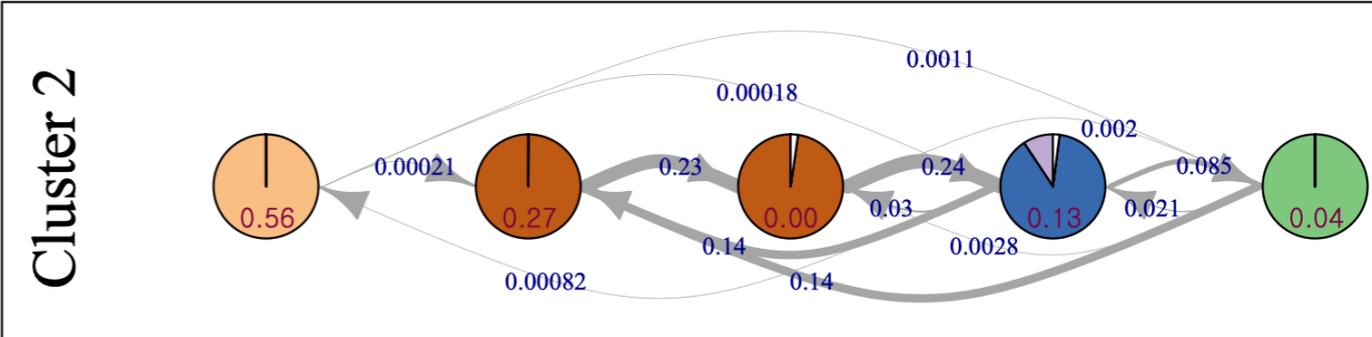
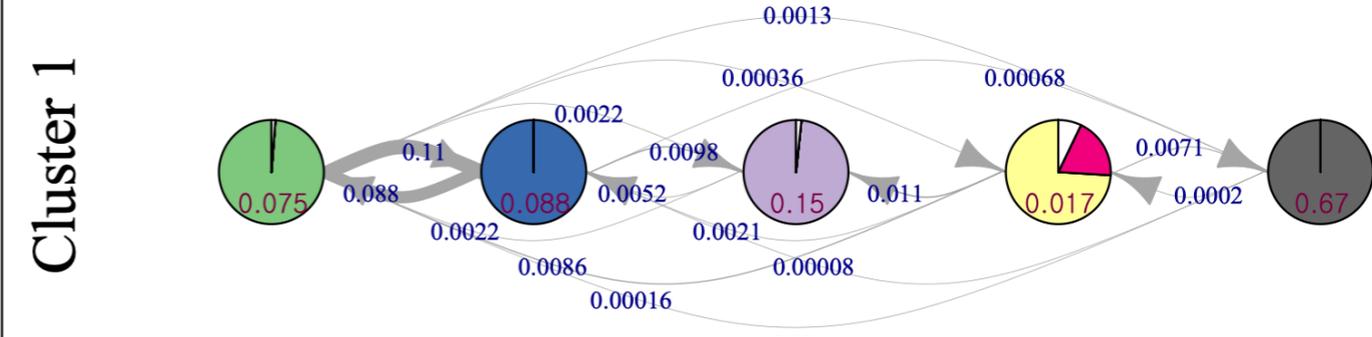
• **C1:** 빈곤 비경활 상태와 차별, 빈곤 노동, 실업을 주로 오가는 전형적인 불안정한 고용이행 집단[**불안정집단**]

• **C2:** C1과 유사한 불안정 집단이지만 노동빈곤 경험이 좀 더 주요한 집단[**노동빈곤집단**]

• **C3:** 비경활이 주로 관찰되는 집단[**비경활집단**]

• **C4:** 안정적인 고용 이력을 보이는 집단[**안정집단**]

• **C5:** 실업과 비경활 고용을 비교적 다양하게 오가는 노동시장 참여-미참여 사이의 이동을 활발하게 보이는 집단[**유동집단**]



노법래(2020)



HMM 분석 사례

Mixture Hidden Markov Model	노동빈곤		비경활		안정		유동적	
	Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	exp(b)	z	exp(b)	z	exp(b)	z	exp(b)	z
(Intercept)	0.15	-4.12	0.45	-3.19	1.09	0.33	0.24	-3.31
성별_남성	1.00		1.00		1.00		1.00	
성별_여성	1.18	1.11	2.10	8.49	0.60	-5.73	1.29	1.73
연령_2030	1.00		1.00		1.00		1.00	
연령_4050	0.48	-4.31	0.53	-6.37	0.80	-2.38	0.58	-3.28
연령_60+	0.60	-2.00	0.68	-2.61	0.64	-3.11	0.73	-1.33
심도_경증	1.00		1.00		1.00		1.00	
심도_중증	0.61	-2.51	1.54	4.08	0.36	-9.25	0.39	-4.30
유형_내부	1.00		1.00		1.00		1.00	
유형_정신	0.81	-0.59	0.50	-4.09	0.68	-1.89	1.01	0.01
유형_신체	1.46	1.29	0.79	-1.54	1.72	3.36	1.07	0.26
유형_감각	1.19	0.56	0.72	-2.02	1.35	1.71	1.07	0.23
교육연수	1.05	2.30	1.03	2.48	1.13	9.94	1.01	0.71
Log-likelihood: -21133.92 BIC: 46315.99								

- **심도_중증**: 비경활 상태로의 일관된 이행에 영향을 미침.
- **유형_정신장애**: 불안정 고용으로의 이행 가능성이 큼.
- **유형_내부장애**: 불안정 고용으로 이행과 비경활 상태로의 이행에 영향을 지님.
- **성별_여성**: 불안정 고용과 비경활 상태 이행으로 영향력을 지님.
- **연령_장년층 이상**: 40대 이상일수록 불안정 노동 이력을 경험할 가능성이 커짐.
- **연령_청년층**: 비경활 상태에 머물 확률이 상대적으로 높음.
- **교육**: 교육수준은 안정적 고용지위와 비경활 상태에 이행에 각각 관여함.

노법래(2020)

감사합니다!