



# NUEVOS ENFOQUES EN EL MODELADO ESTRUCTURAL LONGITUDINAL

Delfino Vargas

Seminario de Altos Estudios del Desarrollo  
5 de Febrero 2020



# La medición de lo inobservable

- Paul Felix Lazarsfeld, matemático y físico austriaco, fundó el Instituto de Investigación Aplicada en Psicología Social en Austria.
- Emigró a los Estados Unidos en 1933 y trabajó en la Universidad de Princeton como director de la Office of Radio Research.
- Propone la construcción de conceptos a partir de los índices empíricos. Una nueva forma de ejemplificar el concepto de medición.
- Hoy día esto se ha formalizado como las variables latentes a partir de indicadores observables. Que constituyen la base de la Teoría de la Medida.



Paul Lazarsfeld (1901-1976)

# Modelos de Segunda Generación

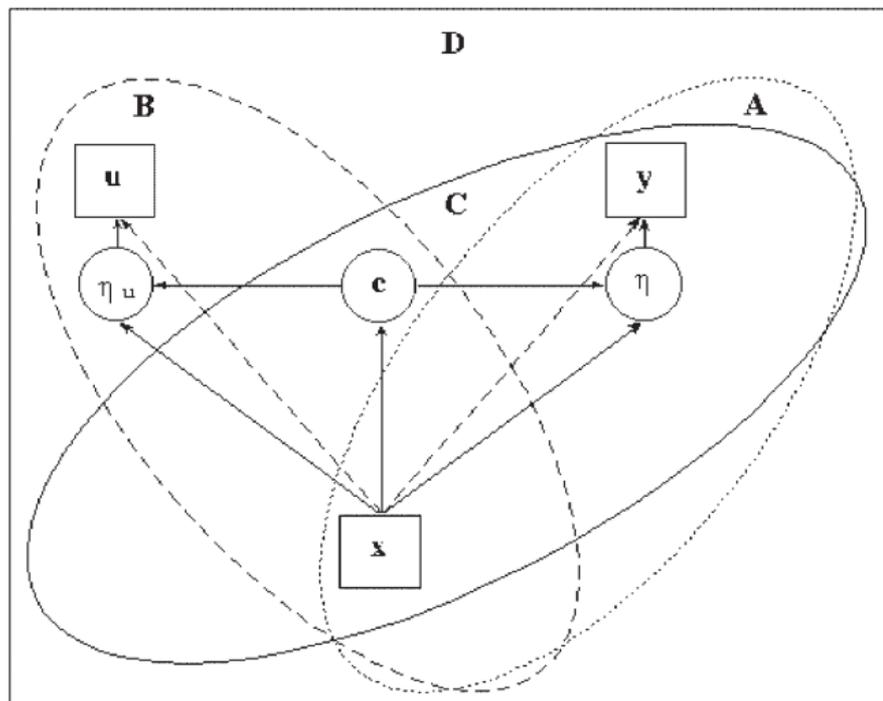


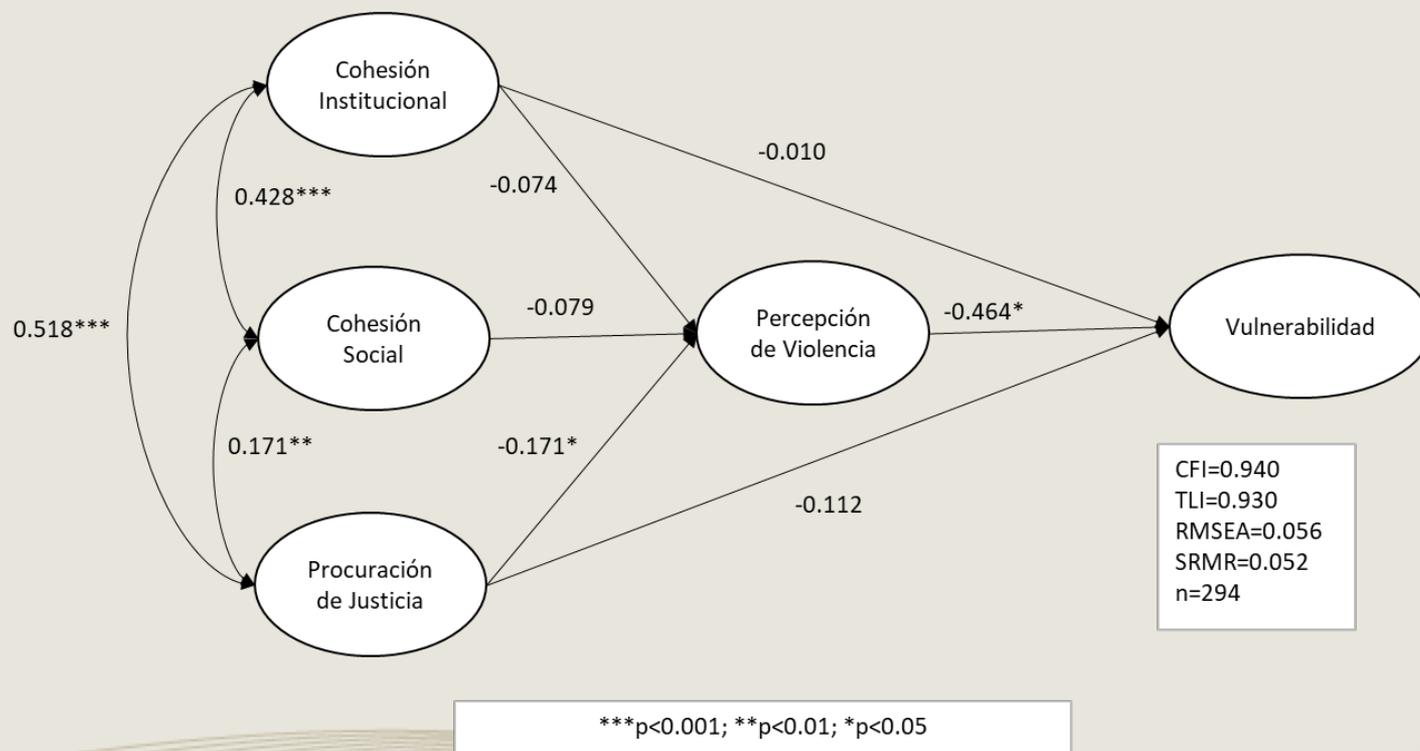
Figure 1: A general latent variable modeling framework

Fuente: Muthen (2002:83) "Beyond SEM: General Latent Variable Modeling". Behaviormetrika (29)1: 81-117

- Una forma diferente de formalizar la nueva propuesta de SEM es ubicarlo como la combinación de modelos que configuran **grupos de variables** (variable-oriented approach), con modelos que configuren **grupos de observaciones** (person-centered approach).
- **Modelo A:** Análisis de Clases Latentes con variables continuas (condicional/no-condicional). Incluye modelos de curvas de crecimiento, análisis factorial confirmatorio.
- **Modelo B:** Análisis de Clases Latentes con variables categóricas (condicional/no-condicional). Incluye análisis factorial confirmatorio, Modelos longitudinales.

# Sistemas de Hipótesis Complejas

- La naturaleza de las hipótesis es compleja, la causalidad en sí no es fácil de establecer. En su lugar se busca la “explicación causal”.





- La medición social debe pasar por varios filtros, que incluyen: la confiabilidad, la validez, la invariancia factorial, entre otros.
- La **confiabilidad** se relaciona más con analizar la estructura de correlación de los constructos (e.g.  $\alpha$ -Chronbach,  $\Omega$ -McDonald, etc.).
- La **validez**. Hay varias maneras de formalizarla, pero la más común es atribuirle al soporte conceptual que enmarca al estudio.
- La **invarianza factorial** tiene que ver con el grado de comparabilidad de la medida entre grupos.



# El índice de rezago social

- Las instituciones mexicanas encargadas de la evaluación de la política de desarrollo social requieren de mejores técnicas en la medición del rezago social de los municipios, y es de vital importancia identificar los municipios con mayor rezago que permitan la focalización de programas sociales para combatir la pobreza.
- El presente trabajo constituye una aportación metodológica para medir el rezago social en México usando la Encuesta Nacional de Ingreso y Gasto de los Hogares, de los años 2000, 2005 2010 y 2015. la información está disponible a nivel municipal.
- Actualmente el CONEVAL utiliza 11 variables para producir un índice de rezago social, con que posteriormente reclasifica los municipios en cinco categorías usando técnicas de Dalenius-Hodges.



# Variables del Rezago Social

<b>Características integrantes del hogar</b>	15 + Analfabeta
	6-14 años no asiste a la escuela
	15 + con educación básica incompleta
	Sin acceso a servicios de salud
<b>Características del vivienda</b>	Piso de tierra
	Sin escusado
	Sin agua entubada
	Sin drenaje
	Sin energía eléctrica
<b>Activos del hogar</b>	Sin lavadora
	Sin refrigerador

Fuente: Vargas D., & Valdés, S. (2019). A longitudinal study of social lag: regional inequalities of growth in Mexico 2000 to 2015. *The Journal of Chinese Sociology*, 6(1), 12. doi: 10.1186/s40711-019-0100-6



# El índice de Rezago Social

Componente	2000		2005		2010		2015	
	Valor propio	Proporción Var						
1	6.544	<b>59.5%</b>	5.911	<b>53.7%</b>	5.379	<b>48.9%</b>	5.191	<b>47.2%</b>
2	0.962	8.7%	1.170	<b>10.6%</b>	1.236	<b>11.2%</b>	1.232	<b>11.2%</b>
3	0.862	7.8%	1.036	9.4%	1.043	9.5%	1.104	10.0%
4	0.739	6.7%	0.760	6.9%	0.764	6.9%	0.886	8.1%
5	0.494	4.5%	0.531	4.8%	0.676	6.1%	0.699	6.4%
6	0.437	4.0%	0.491	4.5%	0.598	5.4%	0.566	5.2%
7	0.357	3.3%	0.379	3.4%	0.396	3.6%	0.406	3.7%
8	0.243	2.2%	0.261	2.4%	0.370	3.4%	0.330	3.0%
9	0.164	1.5%	0.214	1.9%	0.279	2.5%	0.299	2.7%
10	0.138	1.3%	0.185	1.7%	0.189	1.7%	0.195	1.8%
11	0.060	0.6%	0.063	0.6%	0.070	0.6%	0.092	0.8%

- En el año 2000 una sola CP explica cerca del 60% de la varianza total, en tanto que para el 2005, 2010 y 2015, se requieren de dos componentes para explicar el 60%.



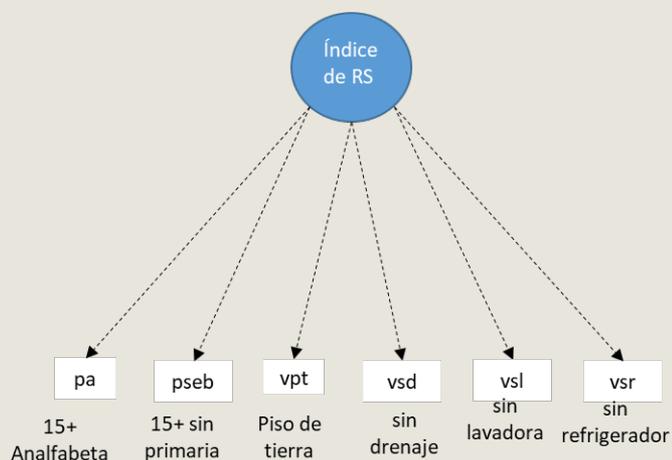
# Análisis factorial confirmatorio de segundo orden

- Una forma de encontrar un índice es tomar en cuenta las dimensiones que lo componen.
- Las cargas factoriales mayores que 0.7 indican un buen ajuste. Sin embargo esto no ocurre para las variables *pnae*, *pss*, *vss*, *vsa*, *vse*.
- Por otro lado la confiabilidad omega reporta valores inferiores a 0.7 para el año 2015
- Se sugiere conservar las variables *pa*, *pseb*, *vpt*, *vsd*, *vsl* y *vsr* que muestran el principio de invarianza factorial (i.e, las cargas factoriales son similares)

F1 (educación)		2000	2005	2010	2015
pa	15+ analfabeta	<b>0.857</b>	<b>0.891</b>	<b>0.926</b>	<b>0.940</b>
pnae	6-14 años no asiste a la escuela	0.541	0.513	0.495	0.373
pseb	15+ con educación básica incompleta	<b>0.870</b>	<b>0.845</b>	<b>0.836</b>	<b>0.824</b>
pss	Sin derechohabencia a servicios de salud	0.718	0.629	0.327	-0.098
	Omega	0.840	0.819	0.761	0.645
F2 (vivienda)					
vpt	Piso de tierra	<b>0.881</b>	<b>0.910</b>	<b>0.735</b>	<b>0.849</b>
vss	Sin escusado	0.506	0.252	0.325	0.273
vsa	Sin agua entubada	0.630	0.621	0.560	0.519
vsd	Sin drenaje	<b>0.841</b>	<b>0.787</b>	<b>0.797</b>	<b>0.764</b>
vse	Sin energía eléctrica	0.654	0.576	0.605	0.617
	Omega	0.835	0.781	0.750	0.755
F3 (activos del hogar)					
vsl	Sin lavadora	<b>0.942</b>	<b>0.953</b>	<b>0.961</b>	<b>0.952</b>
vsr	Sin refrigerador	<b>0.986</b>	<b>0.966</b>	<b>0.951</b>	<b>0.943</b>
	Omega	0.964	0.959	0.955	0.946
G (2o Orden)					
F1		0.932	0.892	0.875	0.866
F2		0.992	0.958	0.923	0.923
F3		0.909	0.928	0.945	0.946
	CFI	0.874	0.885	0.887	0.913
	RMSEA	0.168	0.149	0.136	0.115



# Análisis factorial confirmatorio de primer orden



Var.	Descripción	2000	2005	2010	2015
pa	15+ analfabeta	0.787	0.795	0.800	0.807
pseb	15+ con educación básica incompleta	0.747	0.707	0.708	0.705
vpt	Piso de tierra	0.837	0.846	0.701	0.783
vsd	Sin drenaje	0.804	0.739	0.714	0.700
vsl	Sin lavadora	0.941	0.945	0.954	0.941
vsr	Sin refrigerador	0.970	0.957	0.948	0.936
	CFI	0.925	0.923	0.921	0.921
	RMSEA	0.224	0.218	0.209	0.209
	SRMR	0.044	0.044	0.044	0.044
	alfa-Cronbach	0.927	0.918	0.896	0.889
	omega	0.940	0.933	0.919	0.923



- Se reduce el número de indicadores de RS, de 11 a 6 usando técnicas de análisis factorial exploratorio y confirmatorio (AFE y AFC, respectivamente) y se encuentra que las cargas factoriales estandarizadas son muy similares para los años de estudio con un ajuste satisfactorio (CFI=0.921-0.925, SRMR=0.04,  $\alpha$ -Cronbach de 0.89-0.93; Omega de 0.92-0.94).
- Usando el principio de parsimonia se tienen 6 variables para el índice de rezago social que explican lo mismo que el índice original con 11 variables.



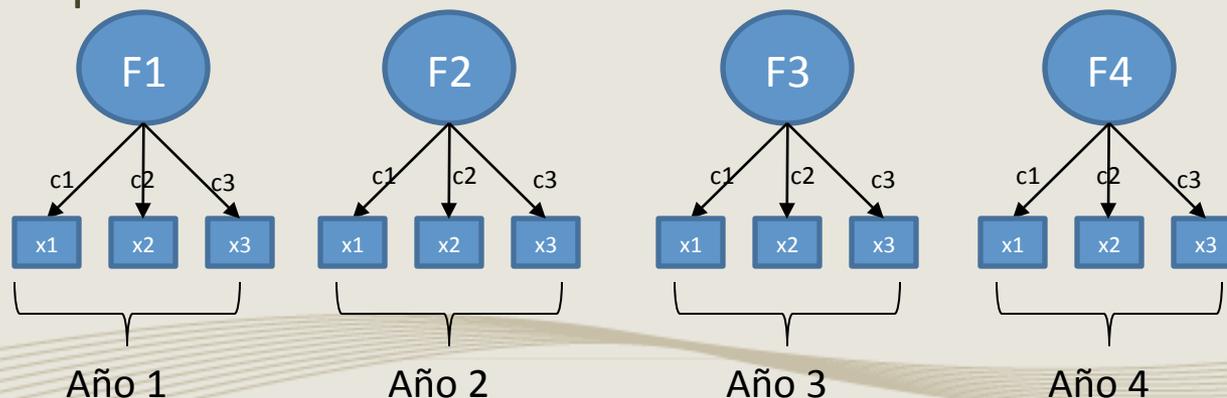
# Invarianza factorial

- La invarianza factorial se clasifica de acuerdo a diferentes enfoques. Una es la invarianza factorial **débil** (Widaman & Reise, 1997), este tipo de invarianza se modela por las restricciones impuestas al modelo. Digamos que la invarianza factorial débil solo impone restricciones en las cargas
- La invarianza factorial **fuerte** impone mayores restricciones tanto en las cargas factoriales como en la estructura de la matriz de covarianzas (Millsap & Meredith, 2007).
- Otro tipo de restricciones se podrían imponer, creando un tercer tipo de invarianza factorial en la cual se restringen las medias,
- La invarianza factorial longitudinal tiene más sentido la propiedad de comparabilidad de las mediciones en el tiempo



# Invarianza Factorial Longitudinal

- Se usa el método de máxima verosimilitud para estimar las cargas factoriales.
- La invarianza factorial se calcula usando la propiedad de que las cargas se parezcan a lo largo del tiempo y de esta manera no pierden la comparabilidad de las mediciones.
- La invarianza factorial se modela haciendo que las cargas sean “similares” en cada medición. Esto permite modelar la media y la varianza en cada medición permitiendo estimar el efecto del tiempo.





# El índice de marginación

Cargas factoriales estandarizadas para las variables de marginación. México 1990, 1995, 2000, 2005, 2010 y 2015

Variables	1990	1995	2000	2005	2010	2015
% Población de 15 años o más analfabeta	0.88	0.87	0.91	0.92	0.92	0.94
% Población de 15 años o más sin primaria completa	0.90	0.91	0.92	0.91	0.94	0.93
% Ocupantes en viviendas sin drenaje ni excusado	0.78	0.80	0.52	0.38	0.37	0.34
% Ocupantes en viviendas sin energía eléctrica	0.67	0.59	0.59	0.50	0.43	0.44
% Ocupantes en viviendas sin agua entubada	0.66	0.60	0.52	0.50	0.45	0.43
% Ocupantes en viviendas con algún nivel de	0.73	0.71	0.73	0.71	0.63	0.67
% Ocupantes en viviendas con piso de tierra	0.86	0.87	0.83	0.78	0.57	0.66
% Población ocupada con ingreso de hasta 2 s.m.	0.60	0.62	0.59	0.57	0.59	0.55
% Población en localidades con menos de 5,000	0.71	0.79	0.81	0.78	0.83	0.68
$\alpha$ -Cronbach	0.90	0.89	0.87	0.84	0.81	0.77
$\Omega$	0.92	0.92	0.91	0.89	0.87	0.86
CFI	0.89	0.87	0.91	0.81	0.83	0.83
RMSEA	0.16	0.18	0.19	0.20	0.18	0.17
SRMR	0.05	0.06	0.06	0.07	0.07	0.07

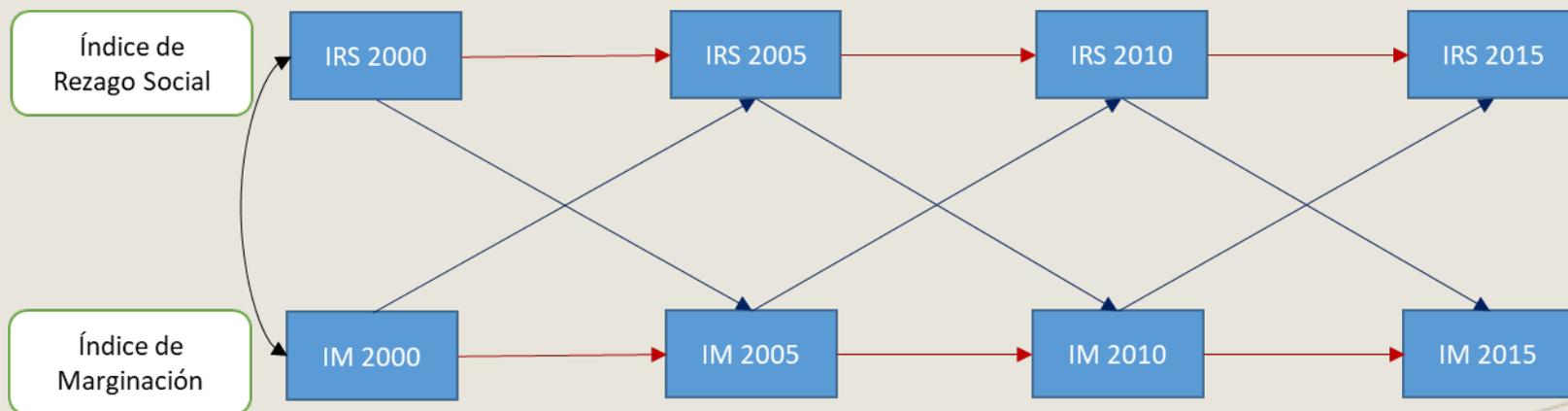


- Para modelar el tiempo
  - Modelos con rezago cruzado
  - Modelos con cambios temporales (¿probar efectos de crisis económicas?)
  - Modelos con tiempos semicontinuos (observaciones que varían con el tiempo)
  - Medición de la invarianza longitudinal
  - Análisis multigrupos de modelos longitudinales (diseños acelerados)
  - Modelos longitudinales condicionales y no condicionales
- Modelos MIXTOS con curvas de crecimiento latente
  - Análisis de clases latentes (LCA)
  - Análisis de crecimiento de curvas latentes (LCGA)
  - Análisis de transición latente (LTA)
  - Procesos Longitudinales en Paralelo
  - Vinculación con R



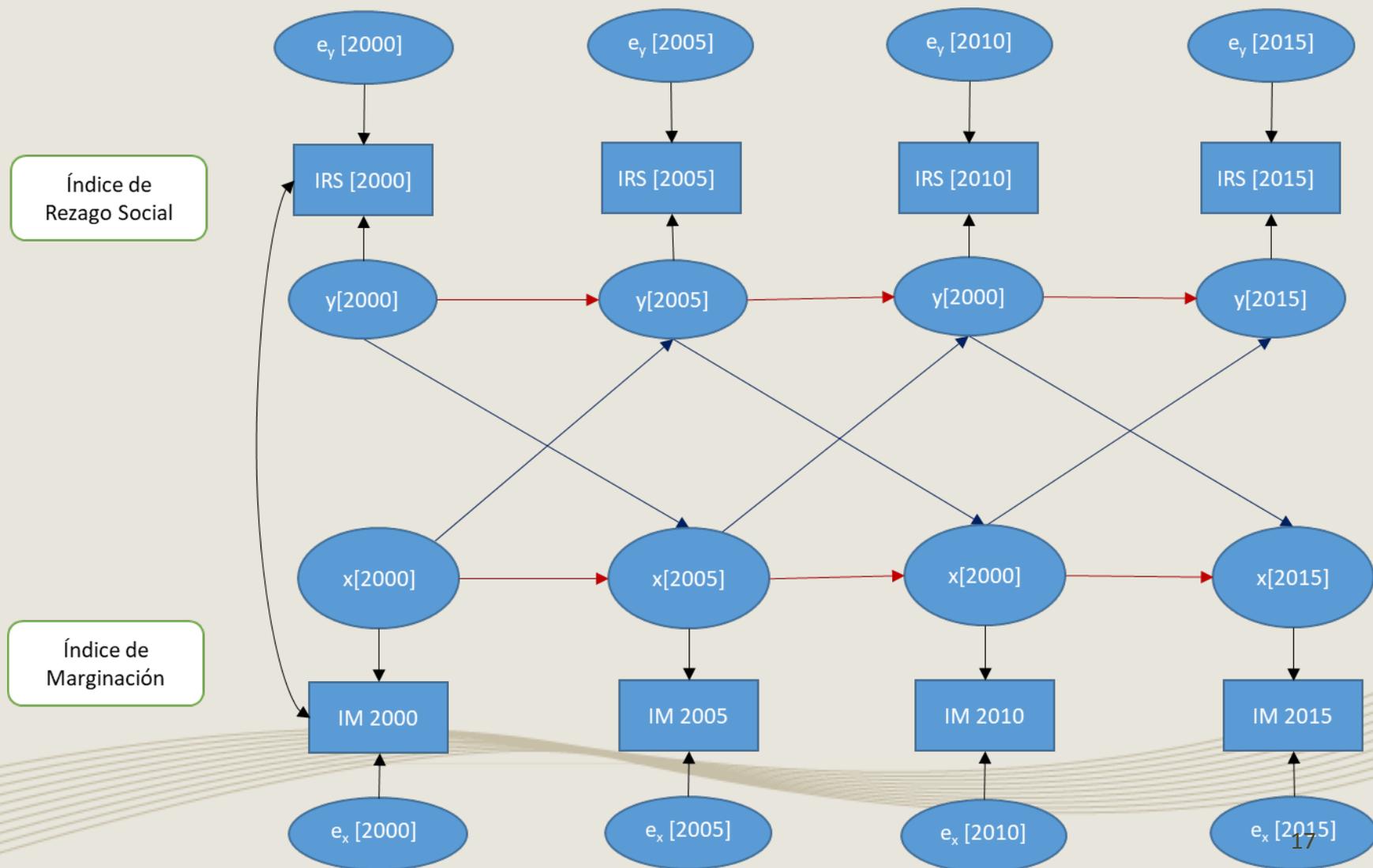
# Dos procesos en paralelo. Modelo con Rezago Cruzado

- Se pueden proponer dos procesos medidos ya sea con indicadores dicotómicos o continuos. Por ejemplo, la marginación se mide con 9 indicadores (Cortes y Vargas, 2017) y el rezago con 6 indicadores (Vargas y Valdés, 2019). Entonces, podemos postular un modelo de rezago cruzado entre la marginación y el rezago a lo largo del tiempo.





# Modelo Bivariado de RC





# Posibilidades de modelado longitudinal

- Modelo de curvas latentes de crecimiento (CLC)
- Modelos con efectos lineales y cuadráticos
- Modelos con el tiempo medido de manera métrica
- Modelos longitudinales por partes (piecewise)
- Modelos con variable de respuesta distal
- Medición de invarianza longitudinal

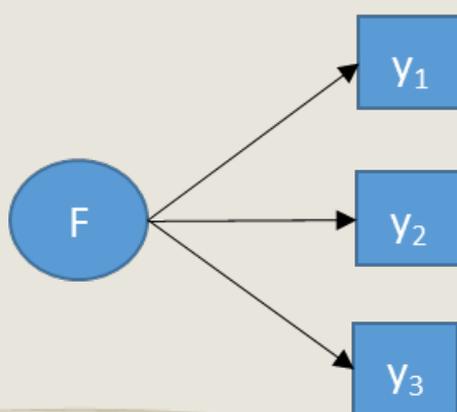


# Aplicaciones de análisis de clases latentes

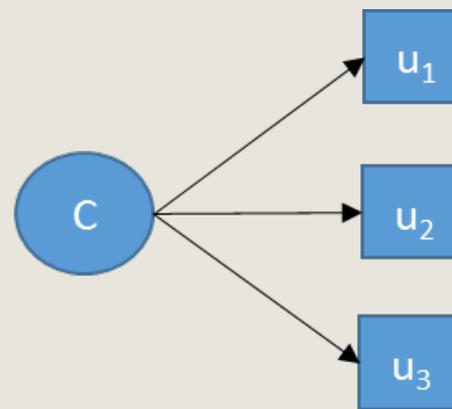
- Identificar clases latentes
- Clasificar individuos
- Identificar covariables que predicen la pertenencia a las clases
- Identificar variables distales que predicen un efecto final a partir de las pertenencia a las clases latentes.
- Identificar clases de procesos en paralelo

# Análisis de Clases Latentes

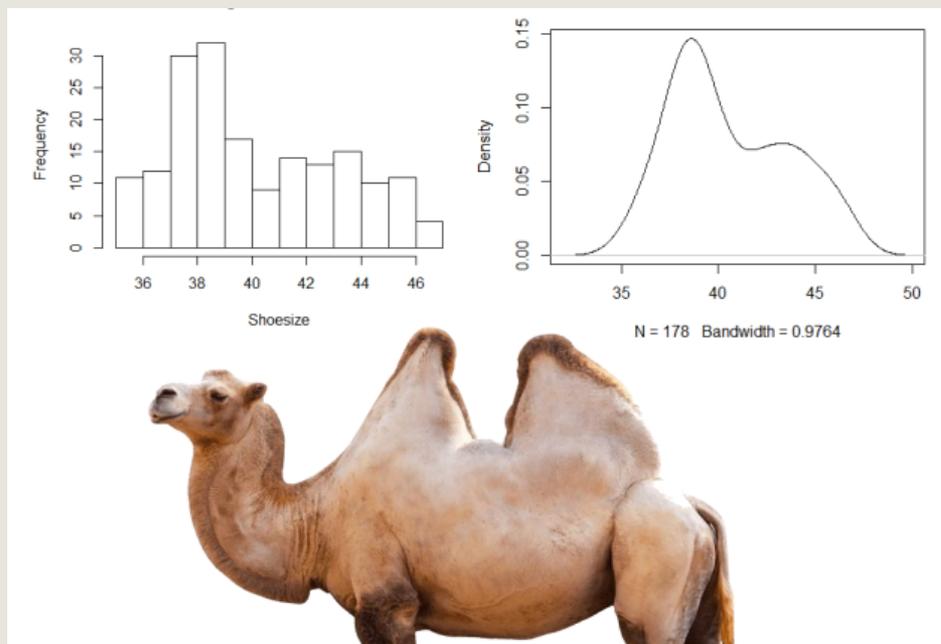
		Variables observadas	
		continuas	categoricas
Variable latente	continuas	<b>Análisis Factorial</b>	Teoría de respuesta al ítem (IRT)
	categoricas	Análisis de Perfiles Latentes	<b>Análisis de Clases Latentes</b>



Análisis Factorial



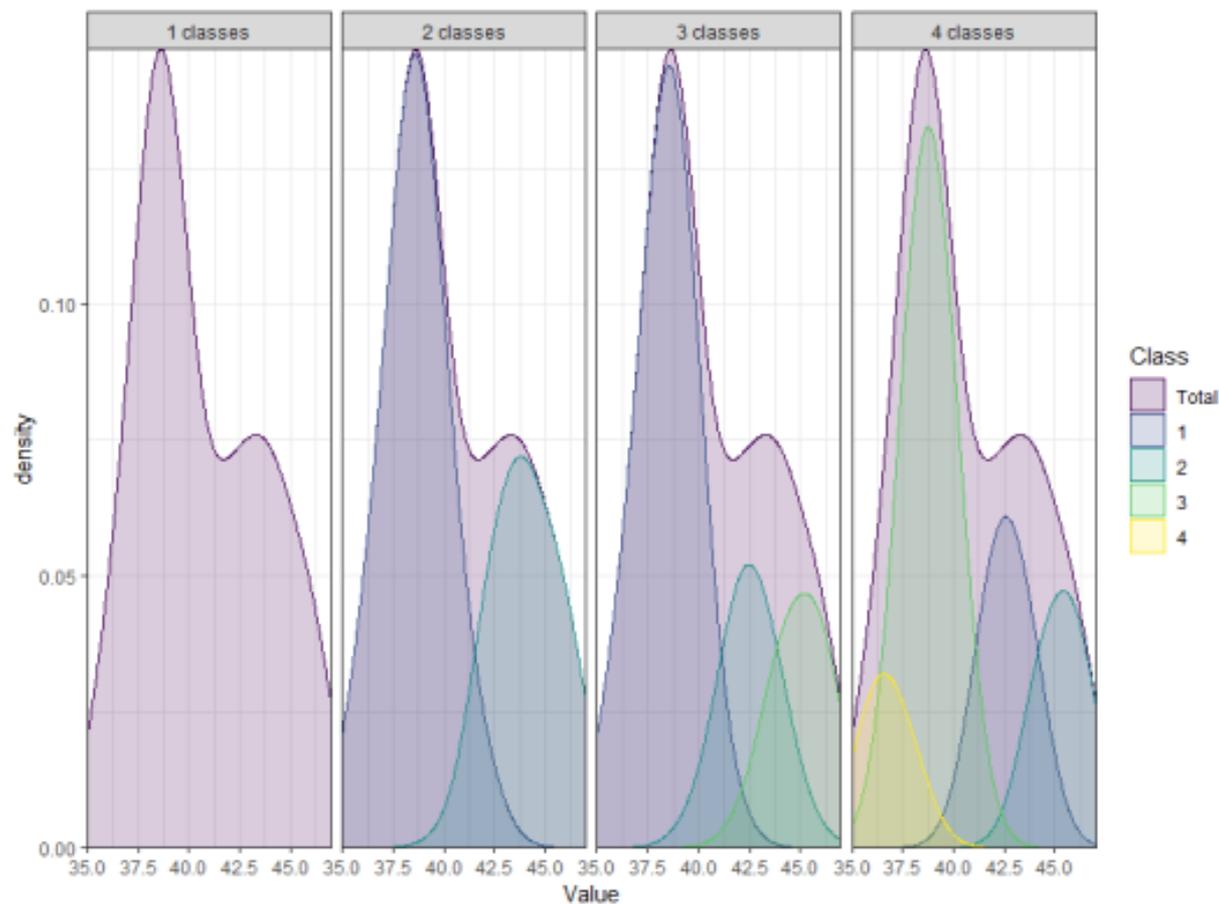
Análisis de Clases Latentes



- Los modelos de clases latentes en realidad provienen de una mezcla de distribuciones.
- El objetivo consiste en separar dichas distribuciones en grupo, llamadas clases latentes.



# Análisis de clases latentes





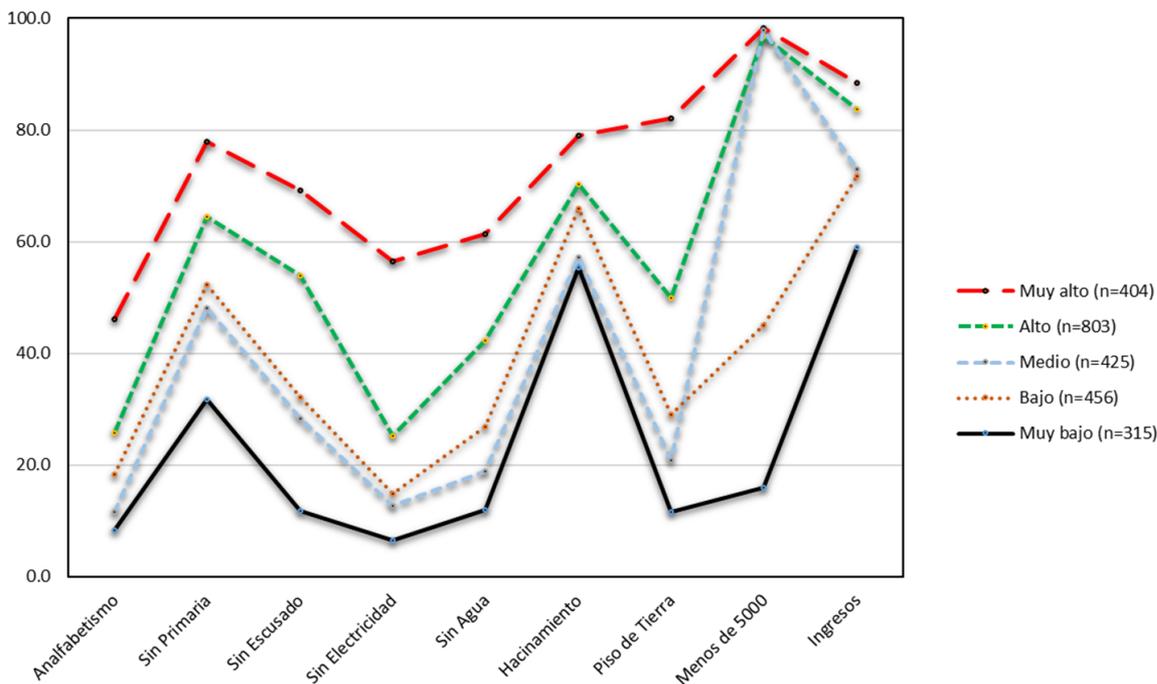
# Tipos de análisis transversal y longitudinal

- Análisis de clases latentes (LCA)
- Análisis de clases de crecimiento latente (LCGA)
- Modelos de mezclas de crecimiento (GMM)
- Análisis de transición latente (LTA)



# Gráficas de las Variables de Marginación 1990

Solución de 5 Clases Latentes de Marginación 1990



Notar que las líneas no se cruzan, lo que es un indicador de que las categorías están bien construidas.

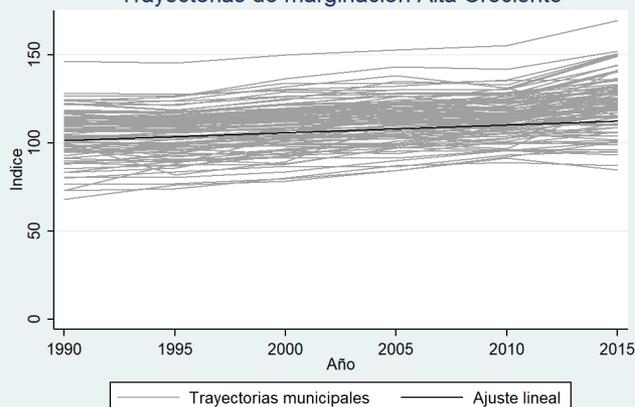
- En el grupo correspondiente a la categoría “Muy alta” marginación (n=404) están incluidos los municipios que presentan los puntajes más elevados en todas las variables de estudio
- Si comparamos dichos puntajes contra sus correspondientes de la clase “Muy baja” marginación (n=315) se observa que hay un gradiente marcado (del puntaje mayor al menor).

Puntajes promedio de las variables de marginación usando clases latentes para 1990.

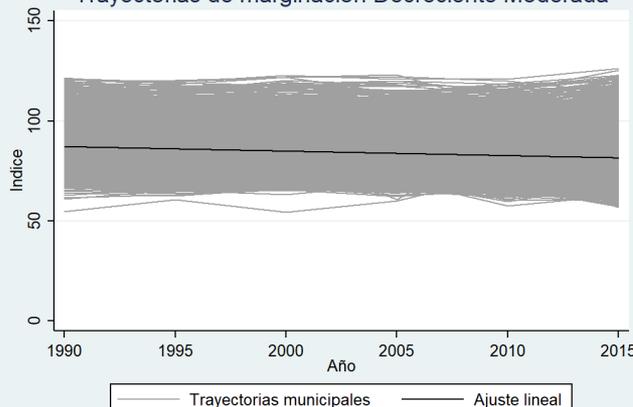


# Trayectorias municipales de marginación 1990-2015

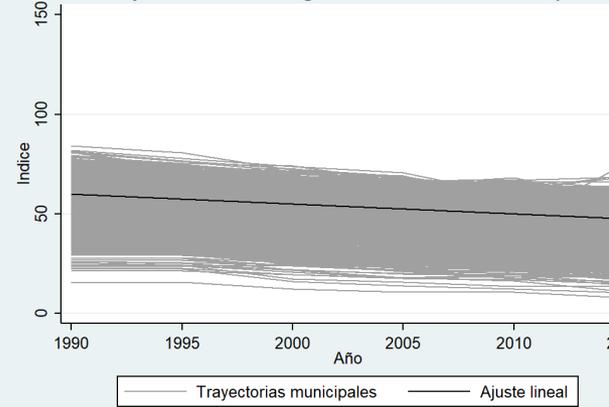
### Trayectorias de marginación Alta Creciente



### Trayectorias de marginación Decreciente Moderada



### Trayectorias de marginación Decreciente Rápida



Marginación Alta y Creciente (n=98)

$$Y_{ij} = 101.23 + 2.22 t$$

(1.57) (0.49)

Marginación Decreciente Moderada (n=1427)

$$Y_{ij} = 87.06 - 1.13 t$$

(0.95) (0.13)

Marginación Decreciente Rápida (n=932)

$$Y_{ij} = 59.78 - 2.48 t$$

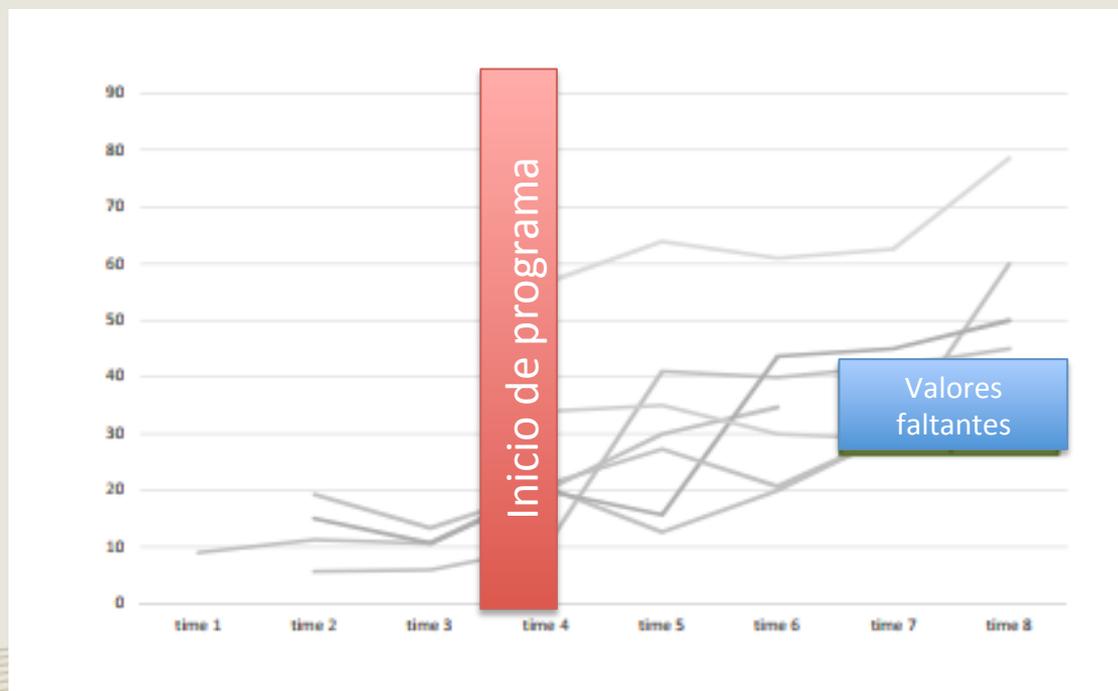
(1.43) (0.07)

Cortes y Vargas (2017) "Origen es Destino: Análisis longitudinal de la marginación. México 1990-2015"



# Modelos CLC por partes

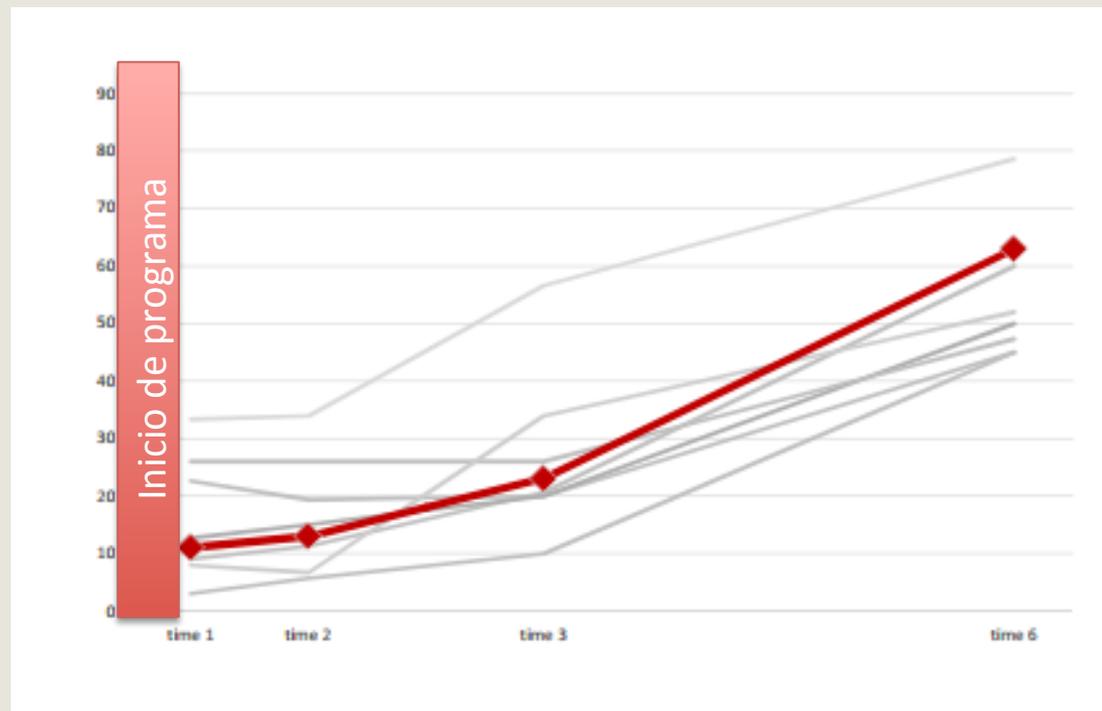
- Podemos agregar una variable de respuesta distal. Por ejemplo, probar si las trayectorias de violencia son diferentes antes y después de un programa





# Las ocasiones de medición no son equidistantes

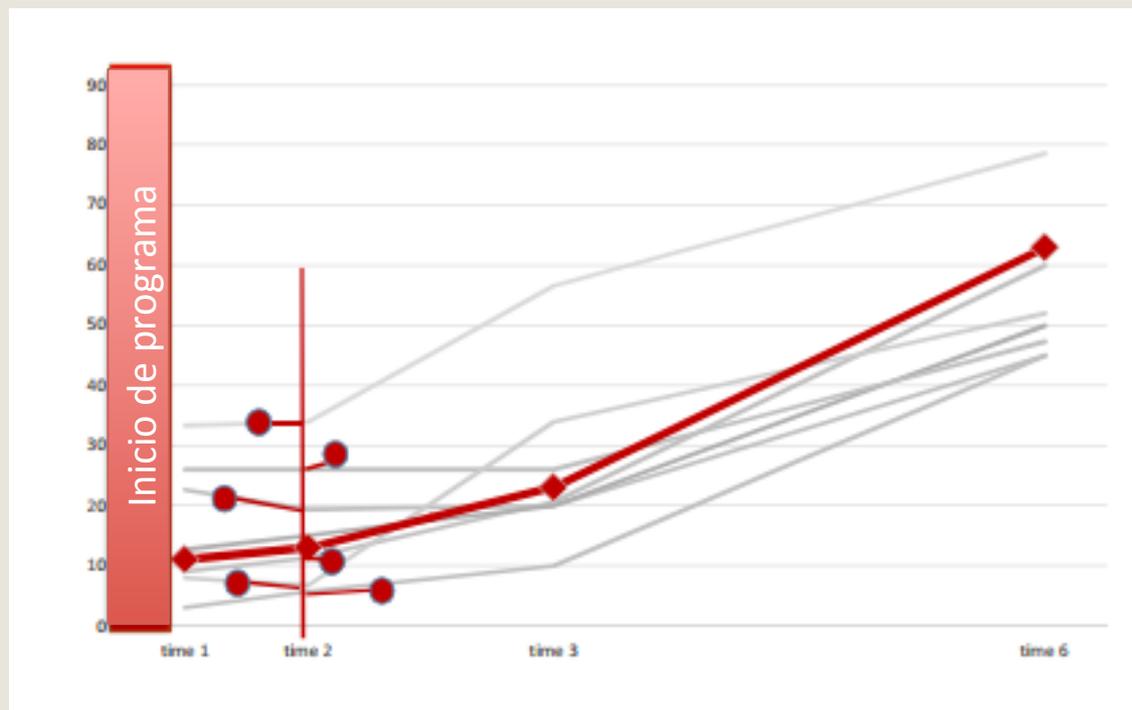
- Las mediciones no son equidistantes





# El tiempo es cronométrico e irregular

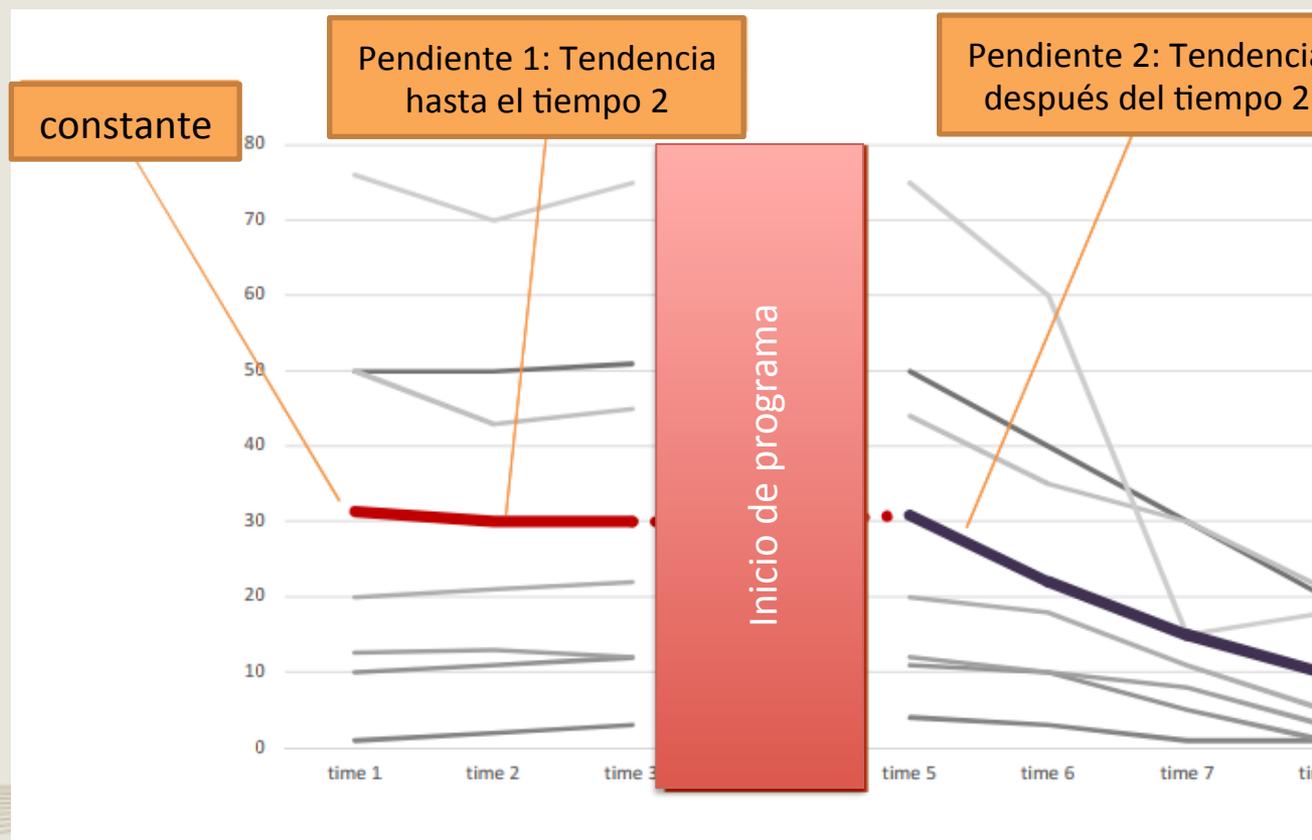
- Se pueden tener diferentes medidas dentro de las ocasiones de medición
- Ver por ejemplo el caso de la ENASEM, que tiene mediciones no equidistantes en el tiempo.
- Medir el deterioro de la calidad de vida de los adultos mayores.





# Modelos para analizar períodos

- Si se tienen diferentes trayectorias de violencia a nivel municipal
- Se puede analizar si hay una tendencia diferente antes y/o después de un tiempo fijo

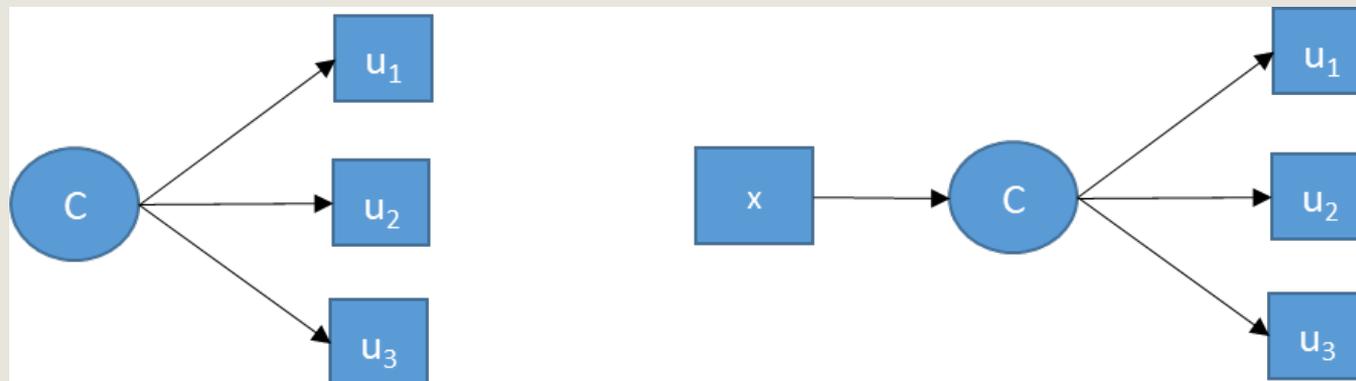




# Análisis de transición latente

- Análisis de transición latente
  - Las mismas clases latentes medidas al tiempo  $t$  y  $t+1$
  - Las clases latentes tienen los mismo indicadores
- Puntos reflexionar
  - ¿la medición es invariante en el tiempo?
  - SI: Usar el modelo “mover/stayer”
  - NO: usar el modelo de transición latente.

		tiempo $t+1$	
		LCA 1	LCA 2
Tiempo $t$	LCA 1	$p_{11}$	$p_{12}$
	LCA 2	$p_{21}$	$p_{22}$



LCA no condicional

LCA condicional

- Hasta ahora solo hemos hecho LCA no condicional
- MPLUS y R interactúan, muchos de los procesos están automatizados.
- MPLUS y STATA también interactúan.



- La medición ha sido la parte esencial del modelado estructural. Los indicadores formativos (observables) construyen a las variables latentes (no observables).
- La llamada “subjetividad” se mide y se evalúa el error de medición, así como la confiabilidad.
- Las medidas longitudinales nos permiten evaluar la tendencias en el tiempo.
- El modelado longitudinal abre nuevas perspectivas de análisis donde el interés radica en medir las tendencias.



---

# GRACIAS