

Modelamiento Estadístico de Fenómenos Educativos

Ernesto San Martín

Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Chile

The Economics School of Louvain, Université catholique de Louvain, Belgium

LIES Laboratorio Interdisciplinario de Estadística Social, Pontificia Universidad Católica de Chile, Chile

Seminario de Educación

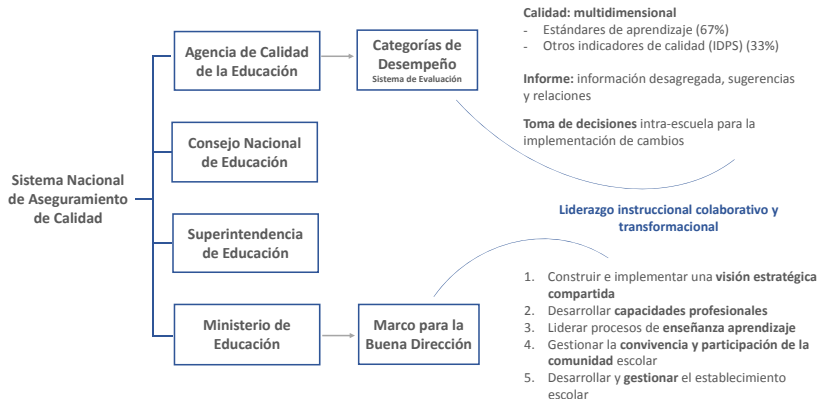
LIES, 24 de agosto de 2017

- En Chile se han ido implementado leyes con la intención de cambiar y mejorar el sistema chileno de educación.
- Leyes relevantes:
 - Ley 20.529, aprobada el 2011, sobre el Sistema Nacional de Aseguramiento de la Calidad.
 - Ley 20.903, aprobada el 2016, sobre el Sistema de Desarrollo Profesional Docente.
 - Ley 20.845, aprobada el 2015, sobre Inclusión escolar que regula la admisión de los y las estudiantes, elimina el financiamiento compartido y prohíbe el lucro en establecimientos educacionales que reciben aportes del estado.

- Proyectos de ley relevantes:
 - Educación superior, en particular lo relacionado con sistemas de inserción universitaria y procesos de selección universitaria.
 - Fortalecimiento de la educación pública.

Sistema Nacional de Aseguramiento de la Calidad

Mejora de los procesos de aprendizaje de los estudiantes,
por medio del desarrollo de capacidades docentes y directivas



(Tomado de T. González y E. San Martín, 2017)

- Todos estas leyes y proyectos de ley conllevan un sinnúmero de fenómenos educacionales que requieren, para su comprensión y evaluación, de **modelamiento estructural**.
- **Modelamiento estructural:**
 - No medir sin teoría (Koopmans, 1947).
 - Población observada (Fisher, 1922) versus estructura (Koopmans & Reiersol, 1950).
 - Descomponer un sistema simultáneo en uno recursivo (Wold, 1949; 1954).
 - Modelos estructurales constituyen explicaciones de fenómenos sociales (educacionales) pues dichos modelos especifican **mecanismos** (Hedström & Swedberg, 1998; Demeulanaere, 2011; Little, 2011): en efecto, un mecanismo puede explicarse cuando **se lo descompone en sub-mecanismos caracterizados por distribuciones condicionales de una variable endógena condicionalmente a un conjunto de variables exógenas** (Russo, Wunsch & Mouchart, 2011; Wunsch, Mouchart & Russo, 2014).

- Mejoramiento continuo de los aprendizajes:
 - ¿Cómo medir aprendizajes de forma coherente con nociones sustantivas de aprendizaje?
 - Aprendizaje como cambios de desempeño a lo largo del tiempo (estándares de aprendizaje?).
 - Aprendizaje axiomatizado desde la perspectiva de Learning Spaces: Aprender progresivamente; conocer más no impide aprender cosas nuevas.
 - La medición del aprendizaje depende de la perspectiva sustantiva asumida.

Problema 1

- ACE está trabajando en producir pruebas de progreso en matemáticas para aplicarlas eventualmente en 5o año de enseñanza básica.
 - Decimales y operaciones.
 - Fracciones y operaciones.
 - Algebra (polinomios, productos notables, etc.)
- ¿Cómo incluir las llamadas habilidades o competencias? Teoría de competencias de Korossy.

Teoría Clásica de Tests

- El modelo Rasch anterior no tiene relación alguna con los modelos psicométricos que incluyen variables latentes.
- La gran mayoría de los modelos con variables latentes pueden ser sumergidos en el contexto de la Teoría Clásica de Tests.
- Sean $Y_{i1}, Y_{i2}, \dots, \theta_i$ v.a. definidas sobre $L^2(M, \mathcal{M}, P)$. La v.a. θ se conoce como variable latente y está caracterizada por el Axioma de Independencia Local (Débil): para todo i ,

$$Y_{ij} \perp Y_{il} \mid \theta_i \quad \forall j \neq l,$$

que es equivalente a

$$Y_{ij} - E(Y_{ij} \mid \theta_i) \perp Y_{il} - E(Y_{il} \mid \theta_i).$$

- El Axioma de Independencia Local (Débil) implica que

$$\begin{aligned} \text{cov}(Y_{ij}, Y_{il}) &= E[\text{cov}(Y_{ij}, Y_{il} | \theta_i)] + \text{cov}[E(Y_{ij} | \theta), E(Y_{il} | \theta_i)] \\ &= 0 + \text{cov}[E(Y_{ij} | \theta), E(Y_{il} | \theta_i)] \\ &\neq 0. \end{aligned}$$

- θ_i explica la correlación entre los observables por medio de $E(Y_{ij} | \theta_i)$.
- De ahí entonces que la descomposición fundamental de la TCT es la siguiente:

$$Y_{ij} = E(Y_{ij} | \theta) + [Y_{ij} - E(Y_{ij} | \theta)],$$

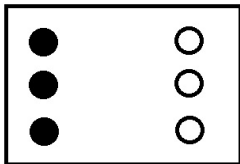
donde por construcción $E(Y_{ij} | \theta_i) \perp [Y_{ij} - E(Y_{ij} | \theta_i)]$.

Problema 2

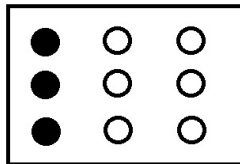
- Practicamente todos los modelos psicométricos son casos particulares de la TCT: especificar $E(Y_{ij} | \theta_i)$.
- ¿Qué son las variables latentes? Discusión epistemológica y técnica.
- No observable \neq No observado.
- La formalización que hacemos de la TCT enfatiza el aspecto no-observado. Pero las investigaciones probabilísticas en torno al Axioma de la Causa Común (Hofer-Szabó, Rédei, Szabó, 2013) ofrecen formalismos que parecen enfatizar el aspecto no-observable.

Problema de la teoría clásica de probabilidades

- Una urna se escoge de forma aleatoria. Dada la urna, fichas son extraídas con reemplazo.



Urn A
Probability of choosing it is $1/3$



Urn B
Probability of choosing it is $2/3$

- El observador **no conoce la urna que fue escogida**. Solo observa el patrón seleccionado, como por ejemplo $(1, 0, 0)$.
- ¿Cuál es la probabilidad que haya sido escogida la urna *A* o la *B*?
- Buscamos la urna **más probable**.

Problema de la teoría clásica de probabilidades

- Usando el teorema de Bayes-Laplace, el problema se resuelve:

$$P(A | 1, 0, 0) = \frac{P(1, 0, 0 | A) P(A)}{P(1, 0, 0)} = \frac{\frac{1}{2} \times \left(\frac{1}{2}\right)^2 \times \frac{1}{3}}{\frac{91}{3 \times 8 \times 27}} = \frac{27}{91};$$

$$P(B | 1, 0, 0) = \frac{64}{91}.$$

- Dada la urna, las extracciones con reemplazo aseguran **independencia condicional**.
- El problema que estamos resolviendo es un **problema de clasificación**.

Problema 3: ¿Variables latentes o invariantes?

- Para responder, hay que releer a Spearman (1904).
- Aquí Spearman desarrolló la **psicología correlacional** como herramienta para analizar resultados de experimentos, los cuales constituyen los elementos esenciales de la **psicología experimental**.
- Spearman explícitamente se ancla en la tradición desarrollada por Wundt (1901): una experiencia única
 - puede ser considerada en cuanto al contenido que se presenta, como **experiencia mediata**. Para ello hay que hacer abstracción del factor subjetivo presente en todas las experiencias. Estas experiencias son analizadas por las **ciencias naturales**;
 - puede ser considerada en relación con la aprehensión de ese contenido, lo que hace referencia al sujeto que experimenta, lo que constituye la **experiencia inmediata**. Estas experiencias son analizadas por la **psicología**.
- Para Wundt, existe una concordancia de método entre ambas formas de analizar la experiencia única.

- Resumen de las reflexiones de Spearman (1904):

Identidades de la ciencia	Experiencia única	
	Experiencia externa (Ciencias naturales)	Experiencia interna (Psicología)
Primer orden	Mismas consecuencias bajo mismas circunstancias	Uniformidades funcionales (ej. Coincidencia entre inteligencia y discriminaciones sensoriales)
Segundo orden	Persistencia de la sustancia: invarianza que la permite identidad de primer orden	Sentido interno al cual se accede por introspección (rompe el paralelo)

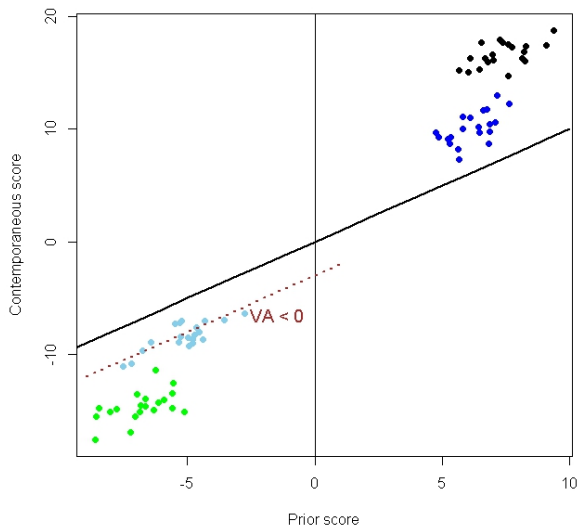
- Por lo tanto, lo que se buscan son **invariantes**: éstos son representados por las variables latentes.

Problema 4: Uso de los Modelos con efectos aleatorios

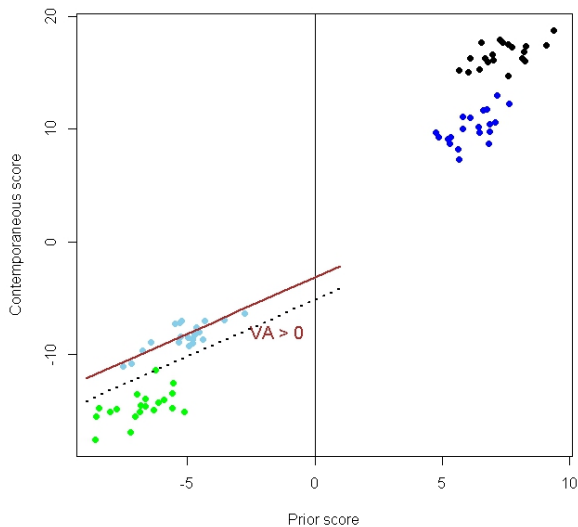
- No interesa estimar o predecir la verdadera “habilidad” matemática de un estudiante como si se tratase de una realidad ontológica en sí.
- Se trata más bien de un problema de clasificación.
- Cuando clasificamos, inducimos relaciones de equivalencia y por tanto estamos definiendo conceptos.
- El desafío es ir a observar los objetos clasificados y mejorar inductivamente la caracterización de los (nuevos) conceptos.

- ¿Cómo definir la efectividad de una escuela?
- ¿Cómo caracterizar un aspecto de la escuela en relación a los estudiantes?
Axioma de Independencia Condicional.

El rol de las covariables en la definición de valor agregado



El rol de las covariables en la definición de valor agregado



Problema 5

- Efectividad escolar y causalidad.
- Definición de causalidad en ciencias sociales.

- La teoría sustantiva permite identificar:
 - una **variable de respuesta** de interés, que denotaremos por Y ;
 - una **variable de tratamiento** que causa (en un sentido probabilístico) Y ; la denotaremos por Z ;
 - variables que definen diversos **contextos en los cuales se espera observar la relación causal entre Z e Y** ; estas variables las denotaremos por X .
- **Consecuencia:** las covariables X ya no constituyen variables explicativas de Y , mucho menos variables de “control”, sino que caracterizan contextos en los cuales se (espera) observa(r) la relación causal entre Z e Y .
- La **relación causal entre Z e Y para diferentes contextos caracterizados por X se resume en la distribución condicional $P(Y | Z, X)$: la **relación causal entre Z e Y es probabilística.****

- Y_1 e Y_0 variables de respuesta en dos escenarios ficticios.
- Interesa compara $P(Y_1 | X)$ y $P(Y_0 | X)$.
- Por el Teorema de Probabilidades Totales,

$$P(Y_1 = 1 | X) = P(Y_1 = 1 | X, Z = 1)P(Z = 1 | X) + P(Y_1 = 1 | X, Z = 0)P(Z = 0 | X).$$

Pero $P(Y_1 = 1 | X, Z = 0)$ no es identificado.

- Solución estándar: Ignorabilidad Fuerte

$$P(Y_1 = 1 | X, Z = 0) = P(Y_1 = 1 | X, Z = 1).$$

- Identificación parcial: $0 \leq P(Y_1 = 1 | X, Z = 0) \leq 1$, luego

$$P(Y_1 = 1 | X, Z = 1)P(Z = 1 | X) \leq P(Y_1 = 1 | X) \leq$$

$$P(Y_1 = 1 | X, Z = 1)P(Z = 1 | X) + P(Z = 0 | X).$$

Problema 6

- $P(Z = 0) \leq q$ dado, luego $P(Z = 1) \geq 1 - q$.
- Entonces

$$P(Y_1 = 1 | X, Z = 1)\nu \leq P(Y_1 = 1 | X) \leq$$

$$P(Y_1 = 1 | X, Z = 1)P(Z = 1 | X) + \gamma,$$

donde

$$\nu = \max\{P(Z = 1), 1 - q\}, \quad \gamma = \min\{q, P(Z = 0)\}.$$

Problema 6

- En general queremos acotar $P(Z = 1 | X = x)$. Sabemos que

$$P(Z = 0, X = x) \leq P(Z = 0),$$

por lo que tenemos

- $P(Z = 0)P(X = x) \leq P(Z = 0, X = x) \leq P(Z = 0)$;
- $P(Z = 0, X = x) \leq P(Z = 0)P(X = x) \leq P(Z = 0)$;

- Otro problema de sesgo de auto-selección que es relevante en educación es el Consentimiento Informado (CI) (Ley 20.120; ley 19.628):
 - Cambiar el reporte científico?
 - Impacto de las restricciones éticas sobre la metodología.
 - Evaluación de políticas y comportamiento inductivo (Neyman, 1957); ver González y San Martín (2017).
 - Relación del estadístico con policy y politics (Pielke, 2007).

- Panorama de las investigaciones aplicadas que estamos desarrollando
- Organizar una agenda de seminarios.
- Enfatizar el hecho de que intentamos motivar nuestra investigación en la agenda pública educacional.