

Sobre el uso de variables latentes en métodos de equating

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

8 de enero de 2026

Contenido

- 1 Equiparación de pruebas (equating)
- 2 Modelo latente para variables ordinales
- 3 LE-C: ilustración para el diseño de Grupo Único
- 4 Simulación y aplicación
- 5 Discusión

Contenido

- 1 Equiparación de pruebas (equating)
- 2 Modelo latente para variables ordinales
- 3 LE-C: ilustración para el diseño de Grupo Único
- 4 Simulación y aplicación
- 5 Discusión

Métodos de equiparación

Formulación general

Familia de modelos y métodos estadísticos que se utilizan para hacer comparables los puntajes entre dos o más versiones de un instrumento, de modo que los puntajes en distintas formas puedan usarse de manera intercambiable (González and Wiberg, 2017; Wiberg et al., 2024).

Algunas notaciones

- Consideremos dos formas de prueba, X e Y .
- Los puntajes obtenidos en X e Y son variables aleatorias denotadas por X e Y , definidas en las escalas \mathcal{X} y \mathcal{Y} , respectivamente.
- Las formas X e Y se administran a n_x y n_y examinados, de modo que $\{X_i\}_{i=1}^{n_x}$ y $\{Y_j\}_{j=1}^{n_y}$ se asumen como realizaciones de X e Y , respectivamente, es decir:

$$X_i \sim F_X(x) \quad i = 1, \dots, n_x,$$

$$Y_j \sim F_Y(y) \quad j = 1, \dots, n_y,$$

donde $X \sim F_X(x)$ y $Y \sim F_Y(y)$.

Diseños de *equating*

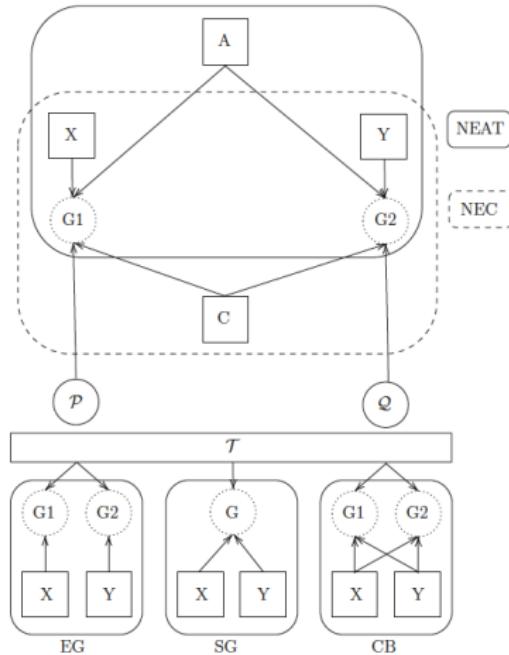


Figura 1: Diseños de *equating*: SG: diseño de grupo único; EG: diseño de grupos equivalentes; CB: diseño contrabalanceado; NEAT: diseño de grupos no equivalentes con test ancla y NEC: grupos no equivalentes con covariables (Wiberg et al., 2024).

Desde un punto de vista estadístico

Transformación de equiparación

Una función $\varphi : \mathcal{X} \rightarrow \mathcal{Y}$ que satisface el requisito de simetría se denomina transformación de equiparación. (Wiberg et al., 2024)

Transformación equipercentil

Braun and Holland (1982) establecieron que $\varphi(\cdot)$ equipara \mathcal{X} con \mathcal{Y} si, para $x \in \mathcal{X}$ e $y \in \mathcal{Y}$,

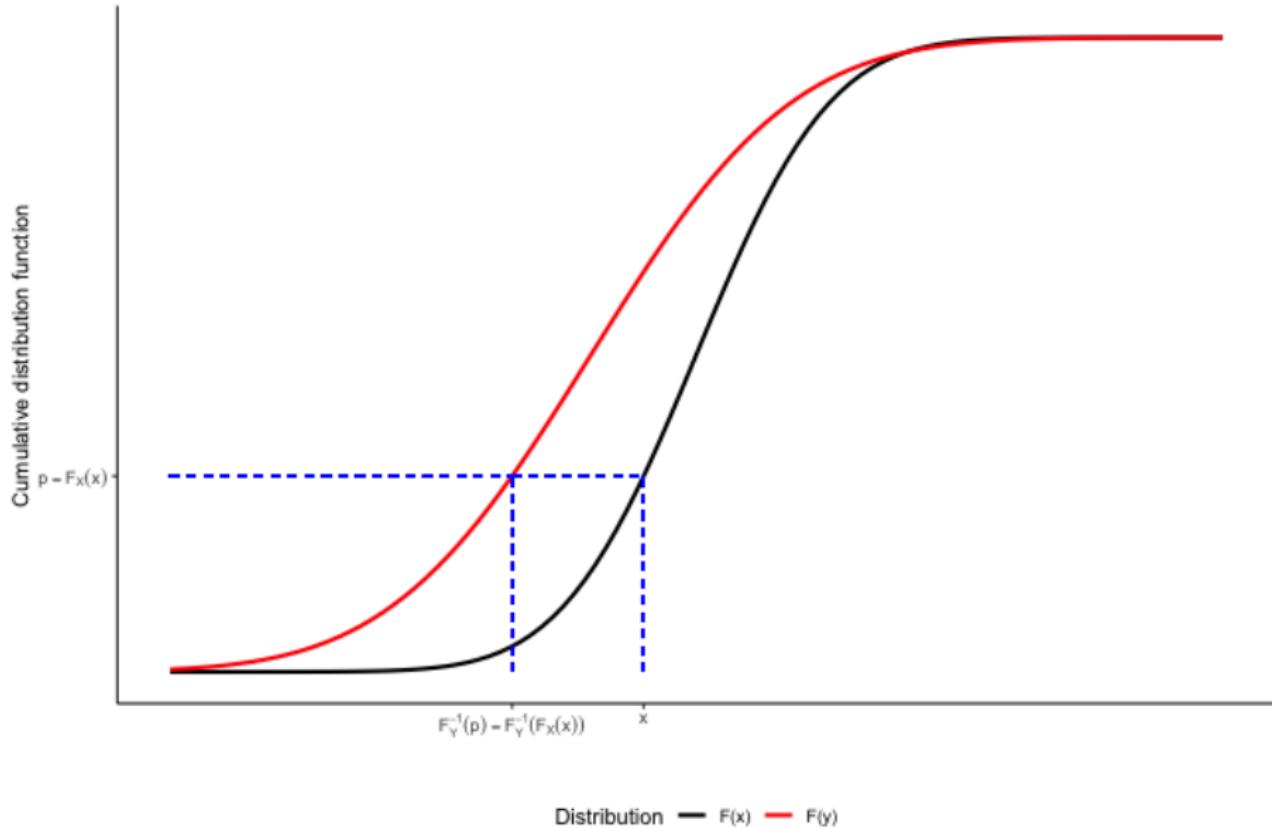
$$F_Y(y) = F_{\varphi(x)}(y) .$$

Esta función se conoce como la **función equipercentil** (Angoff, 1971).

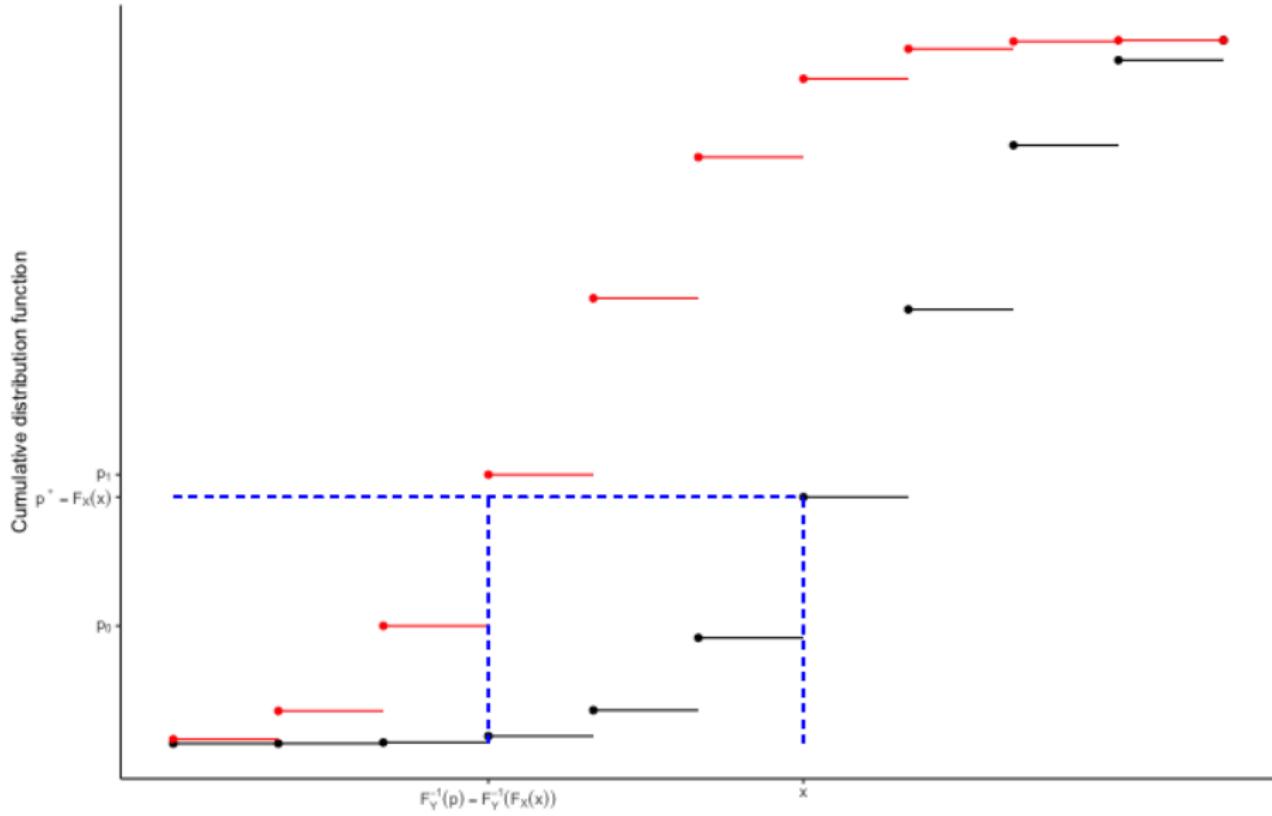
$$\varphi(x) = F_Y^{-1}(F_X(x)) ,$$

donde $F_W^{-1}(\cdot)$ denota la inversa de la CDF de la variable aleatoria W .

Transformación equipercentil



Transformación equipercentil

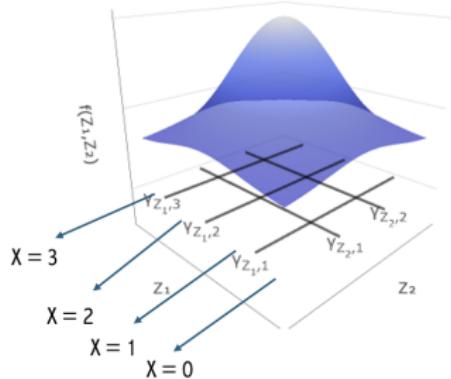


Contenido

- 1 Equiparación de pruebas (equating)
- 2 Modelo latente para variables ordinales
- 3 LE-C: ilustración para el diseño de Grupo Único
- 4 Simulación y aplicación
- 5 Discusión

Puntajes como variables ordinales

- Consideramos los puntajes de pruebas como variables aleatorias ordinales. Utilizaremos la representación latente de estas variables:



donde

$$-\infty = \gamma_{Z_1,0} < \gamma_{Z_1,1} < \gamma_{Z_1,2} < \gamma_{Z_1,3} < \gamma_{Z_1,4} = \infty$$

$$-\infty = \gamma_{Z_2,0} < \gamma_{Z_2,1} < \gamma_{Z_2,2} < \gamma_{Z_2,3} = \infty$$

Un modelo Bayesiano no paramétrico

- Kottas et al. (2005) propusieron un modelo de mezcla de proceso de Dirichlet para la distribución latente \mathbf{Z} :

$$\mathbf{z}_i \mid F \quad F(\mathbf{z}) = \int N(\mathbf{z}; \boldsymbol{\theta}) dG(\boldsymbol{\theta}) \quad G \sim p(G),$$

donde $G \sim DP(M, G_0)$.

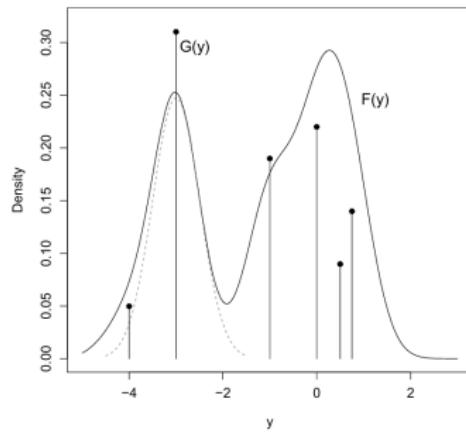


Figura 2: Representación gráfica de DP y DPM (Müller and Rodriguez, 2013).

Un modelo Bayesiano no paramétrico

- La construcción stick-breaking (Sethuraman, 1994) puede escribirse como sigue:

$$G(\cdot) = \sum_{h=1}^{\infty} p_h \delta_{\theta_h}(\cdot), \quad (1)$$

donde δ_{θ_h} denota una masa puntual en θ_h , con $\theta_h \stackrel{iid}{\sim} G_0$, y los pesos $\{p_h\}_{h=1}^{\infty}$ se obtienen a partir de la representación recursiva:

$$p_1 = \nu_1, \quad p_h = \nu_h \prod_{j < h} (1 - \nu_j), \quad \nu_h \stackrel{iid}{\sim} \text{Beta}(1, M), \quad h > 1,$$

- Ishwaran (2000) e Ishwaran and James (2001) propusieron una versión finita de (1) considerando H componentes. Así, el DPM puede aproximarse por:

$$p(z_i | \boldsymbol{p}, \boldsymbol{\theta}) = \sum_{h=1}^H p_h N(z_i | \theta_h).$$

Contenido

- 1 Equiparación de pruebas (equating)
- 2 Modelo latente para variables ordinales
- 3 LE-C: ilustración para el diseño de Grupo Único
- 4 Simulación y aplicación
- 5 Discusión

Representación latente para el diseño de Grupo Único

- Bajo el diseño SG, los puntajes son datos bivariados denotados por:

$$V_i = (X_i, Y_i) \quad i = 1, \dots, n = n_x = n_y.$$

- Suponemos que $\mathcal{X} = \{0, \dots, C_X\}$ y $\mathcal{Y} = \{0, \dots, C_Y\}$.
- Sea $Z_i = (Z_{i,X}, Z_{i,Y})$ el vector latente asociado a V_i , entonces:

$$V_i = (\delta_l, \omega_k) \iff \left\{ \gamma_{X,l} < Z_{i,X} \leq \gamma_{X,l+1} \right\} \cap \left\{ \gamma_{Y,k} < Z_{i,Y} \leq \gamma_{Y,k+1} \right\}$$

- $l = 0, \dots, C_X$, $k = 1, \dots, C_Y$,
- $\gamma_X = (\gamma_{X,0}, \dots, \gamma_{X,C_X+1})$,
- $\gamma_Y = (\gamma_{Y,0}, \dots, \gamma_{Y,C_Y+1})$.

Propuesta: el modelo

- El modelo latente BNP para el vector latente \mathbf{Z}_i puede escribirse como:

$$\mathbf{Z}_1, \dots, \mathbf{Z}_n | \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_H, K_1, \dots, K_n \stackrel{ind}{\sim} N_2(\mathbf{z}_i | \boldsymbol{\mu}_{K_i}, \boldsymbol{\Sigma}_{K_i})$$

$$K_1, \dots, K_n | \mathbf{p} \stackrel{iid}{\sim} \sum_{h=1}^H p_h \delta_h(K_i)$$

$$\boldsymbol{\theta}_h | \tau, \boldsymbol{\Psi} \sim N_2(\boldsymbol{\mu}_h | \tau, \boldsymbol{\Sigma}_h) \text{IWish}_2(\boldsymbol{\Sigma}_h | \nu_0, \boldsymbol{\Psi}),$$

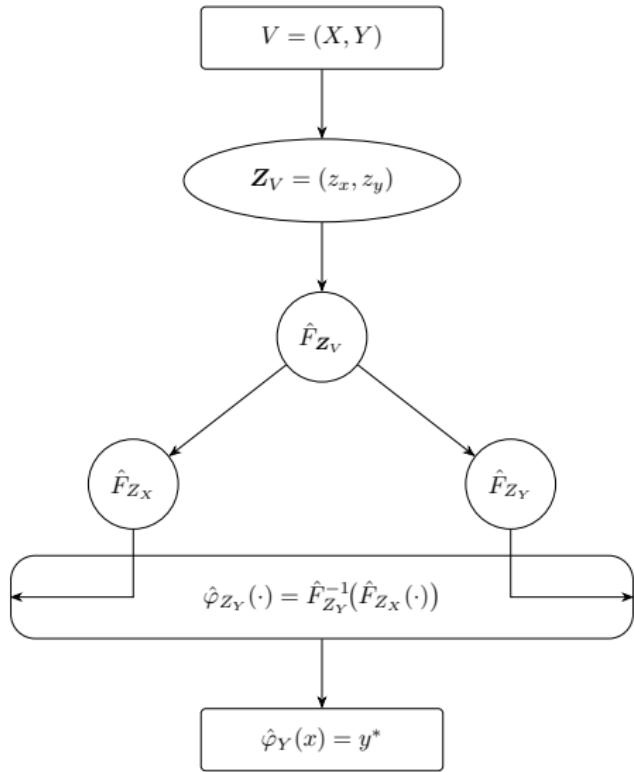
$$\tau \sim N_2(\tau | \tau_0, T_0),$$

$$\boldsymbol{\Psi} \sim \text{Wish}(c_0, C_0),$$

$$M \sim \text{Gamma}(a, b),$$

- El modelo se estima usando el algoritmo blocked Gibbs sampler propuesto por Ishwaran and James (2001).

Método Equating latente



Propuesta: el método LE-C

- Motivado por el método *Equating latente* (LE; Varas et al., 2019, 2020), una versión *continua* considera los siguientes pasos:
 1. Versión continua de los puntajes: $\boldsymbol{v} = (\delta_l, \omega_k)$ corresponde a $\boldsymbol{z} = (z_X, z_Y)$,
 2. Estimación de funciones de distribución acumulada: \hat{F}_{Z_X} y \hat{F}_{Z_Y} ,
 3. Estimación de la función equipercentil latente:
$$\hat{\varphi}_{Z_Y}(\cdot) = \hat{F}_{Z_Y}^{-1}(\hat{F}_{Z_X}(\cdot)),$$
 4. Puntajes equiparados latentes: para cada $x \in \mathcal{X}$ (equivalentemente z_x) se obtiene el conjunto $\{z_Y^* : z_Y^* = \hat{\varphi}_{Z_Y}(z_x)\}$, reescalado al rango $[0, C_Y]$,
 5. Estimación del error estándar de *equating* (SEE)

Contenido

- 1 Equiparación de pruebas (equating)
- 2 Modelo latente para variables ordinales
- 3 LE-C: ilustración para el diseño de Grupo Único
- 4 Simulación y aplicación
- 5 Discusión

Simulación

- Para el vector latente \mathbf{Z} consideramos un modelo de mezcla:

$$f(\mathbf{z}) = \sum_{k=1}^K N_2(\mathbf{z} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- Los umbrales fueron puntos equidistantes entre $\gamma_{X,1} = \gamma_{Y,1} = -20$ y $\gamma_{X,C_X} = \gamma_{Y,C_Y} = 20$.
- La versión finita del DPM se evaluó con $H = 25$.
- Los hiperparámetros se fijaron en $a = 10$, $b = 2$, $c_0 = 4$, $\nu_0 = 4$, $\tau_0 = (0, 0)^\top$, $C_0 = T_0 = \text{diag}((40/4)^2, (40/4)^2)$.
- Se realizó un estudio Monte Carlo con 100 conjuntos de datos, considerando tamaños muestrales: $n_1 = 250$, $n_2 = 500$ y $n_3 = 1000$.

Escenario 1

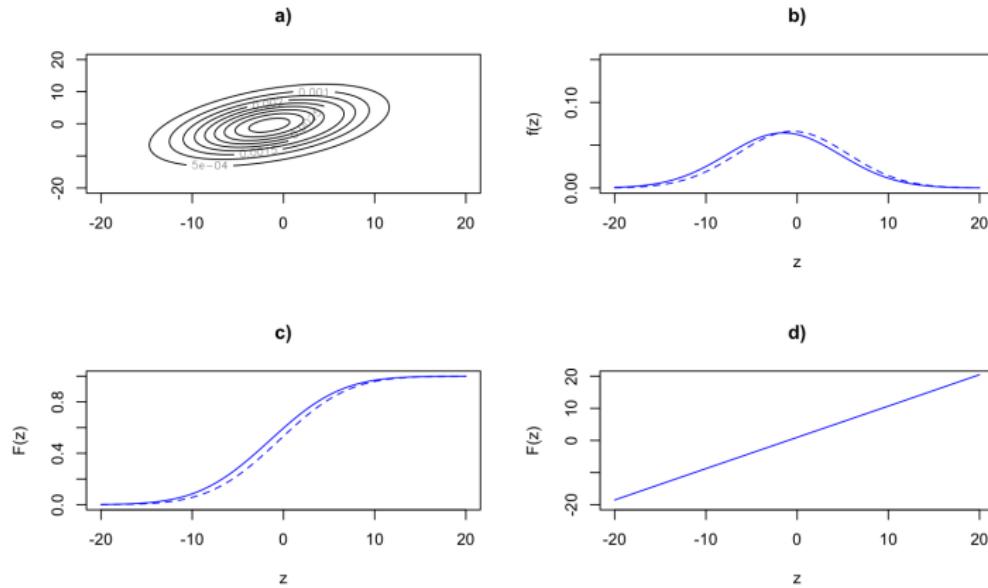


Figura 3: Escenario 1: *a)* distribución latente bivariada (curvas de nivel), *b)* densidad marginal, *c)* CDF marginal, *d)* función equipercenital real.

Escenario 1

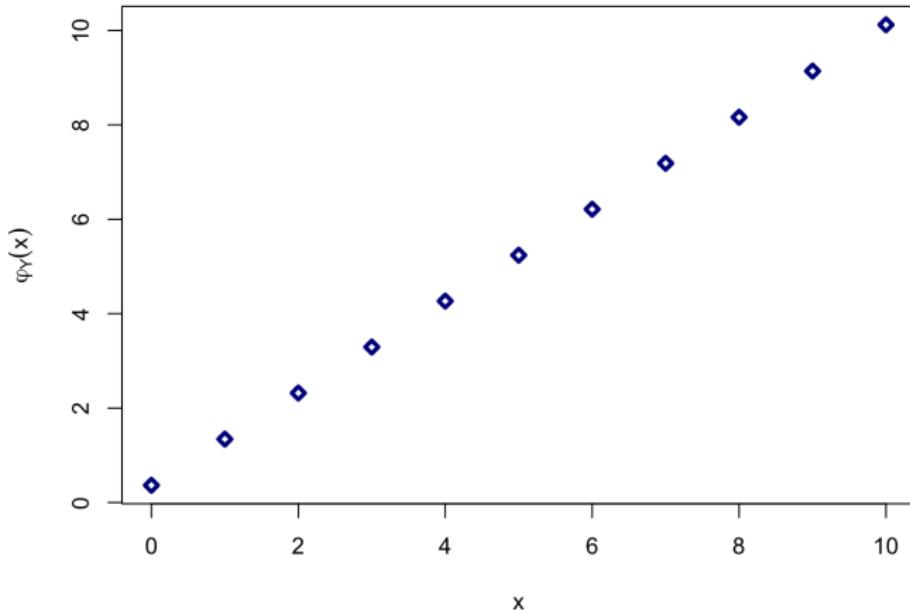


Figura 4: Escenario 1: puntajes equiparados reales.

Escenario 1

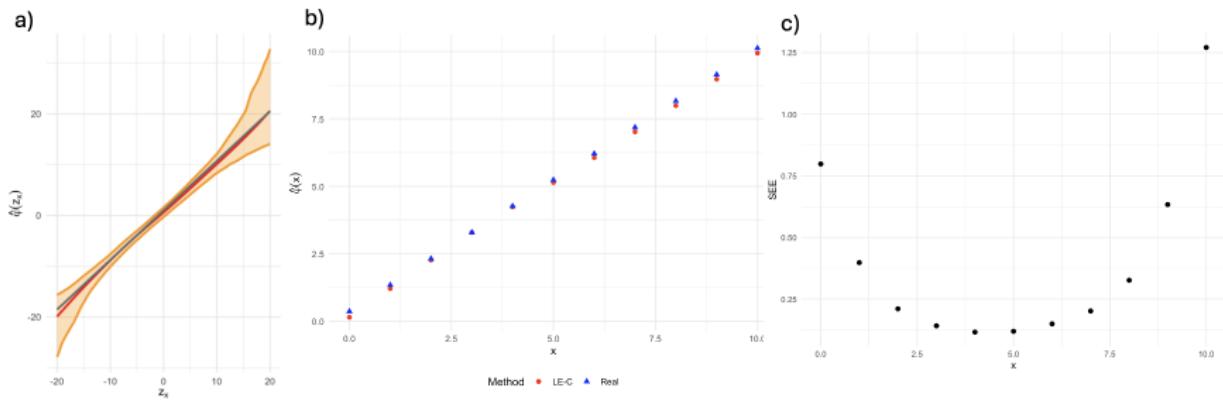


Figura 5: Escenario 1 - resultados ($n = 250$): a) función equipercentil latente estimada; b) puntajes equiparados discretos; c) SEE.

Escenario 2

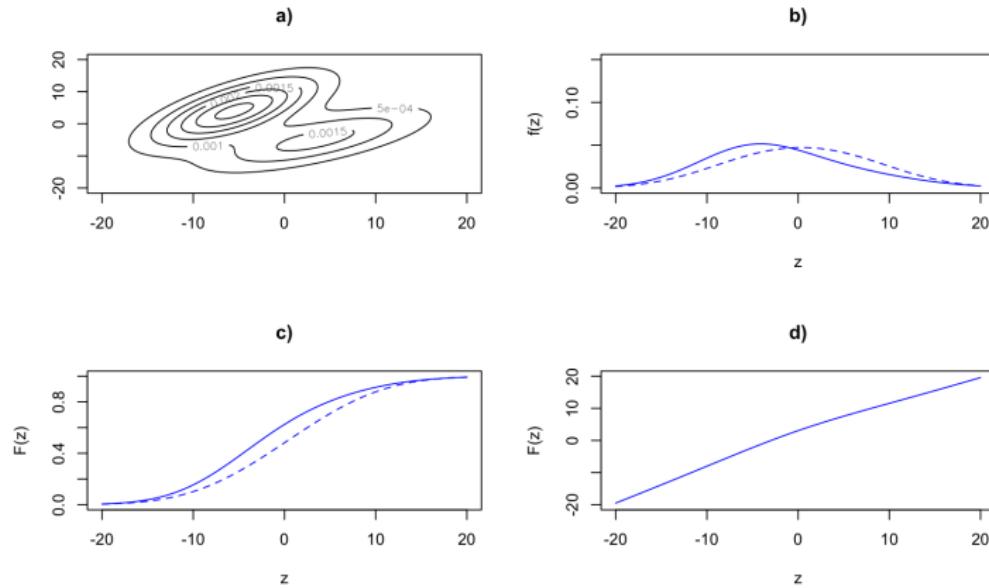


Figura 6: Escenario 2: *a)* distribución latente bivariada (curvas de nivel), *b)* densidad marginal, *c)* CDF marginal, *d)* función equipercenital real.

Escenario 2

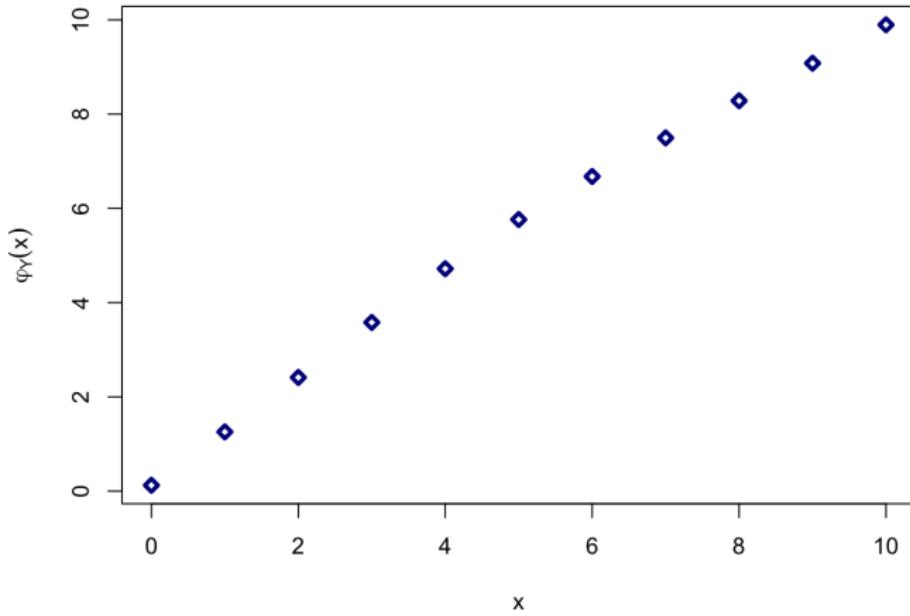


Figura 7: Escenario 2: puntajes equiparados reales.

Escenario 2

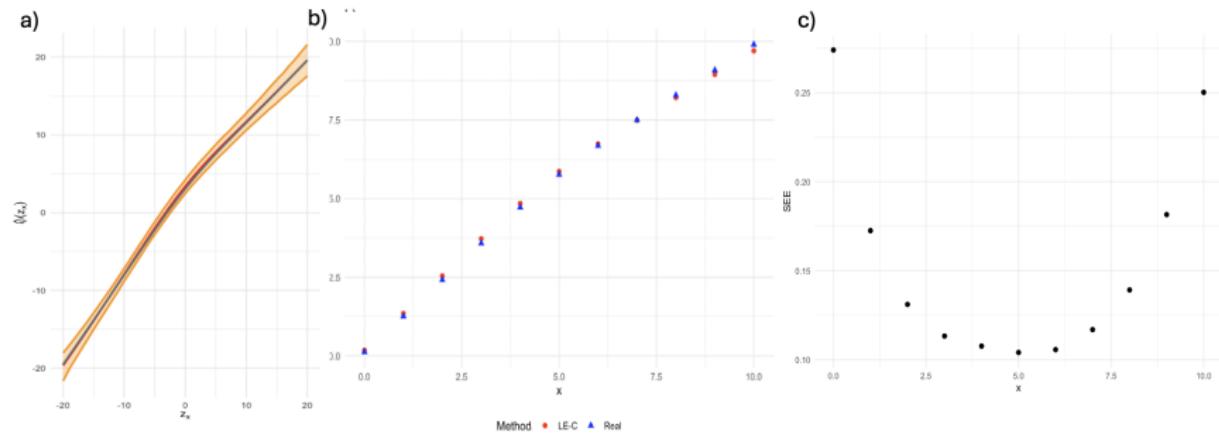


Figura 8: Escenario 2 - resultados ($n = 1000$): a) función equipercentil latente estimada; b) puntajes equiparados; c) SEE.

Cuadro 1: Datos simulados: distancia L_2 estimada entre la función equipercenital latente real y su estimación bajo el método LE-C.

Escenarios	n_1	n_2	n_3
Escenario 1	9.953	7.368	5.903
Escenario 2	10.338	7.278	5.126

Cuadro 2: Datos simulados: distancia L_2 estimada entre el vector de puntajes equiparados continuos verdaderos (en toda la escala) y su estimación a partir de LE-C, el método kernel (KE) y el método equipercenital (EQ).

Método	Escenario 1			Escenario 2		
	n_1	n_2	n_3	n_1	n_2	n_3
LE-C	0.907	0.481	0.445	0.879	0.546	0.453
KE	0.534	0.484	0.447	0.626	0.550	0.485
EQ	1.557	1.219	0.931	0.923	0.727	0.572

Aplicación: puntajes de prueba de Matemática

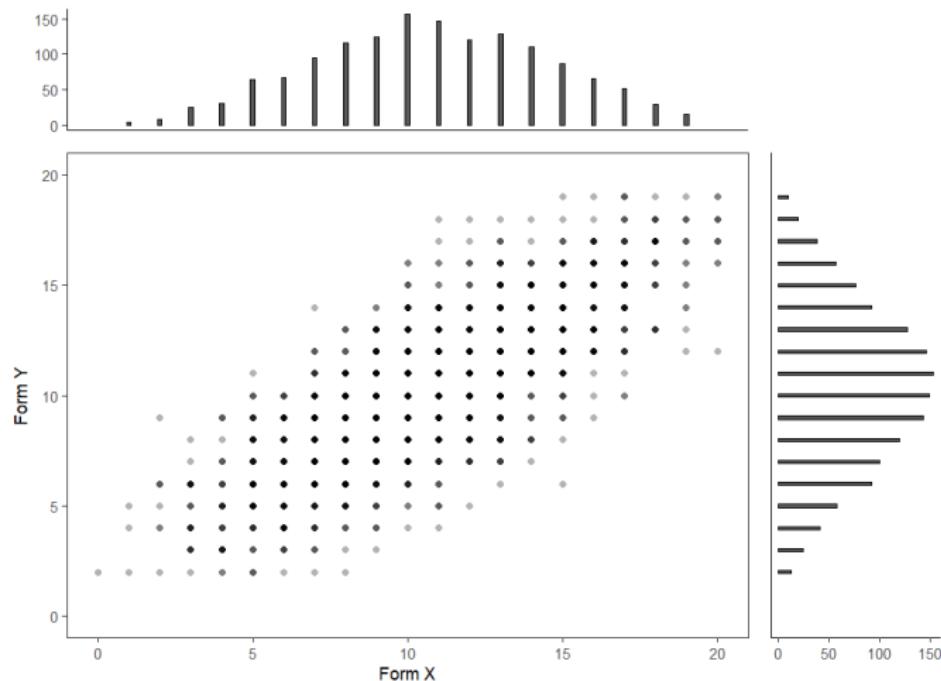


Figura 9: MathSG: distribuciones marginal y conjunta de las formas de puntaje de la prueba de Matemática.

Aplicación: puntajes de prueba de Matemática

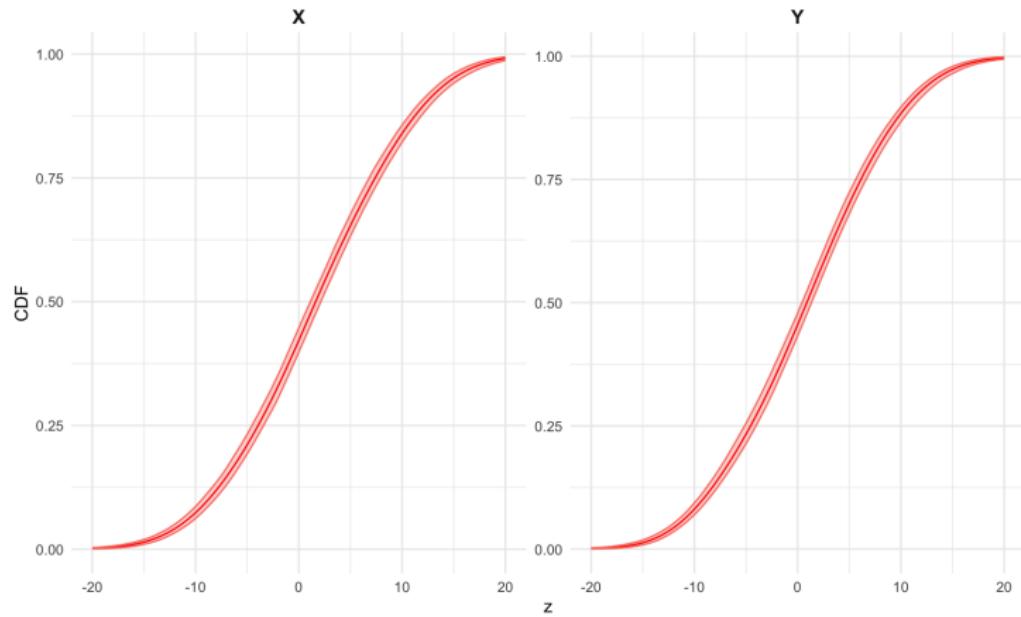


Figura 10: Distribución posteriori predictiva marginal latente para las formas X e Y (línea roja). El intervalo de credibilidad del 95 % se muestra como el área sombreada.

Aplicación: puntajes de prueba de Matemática

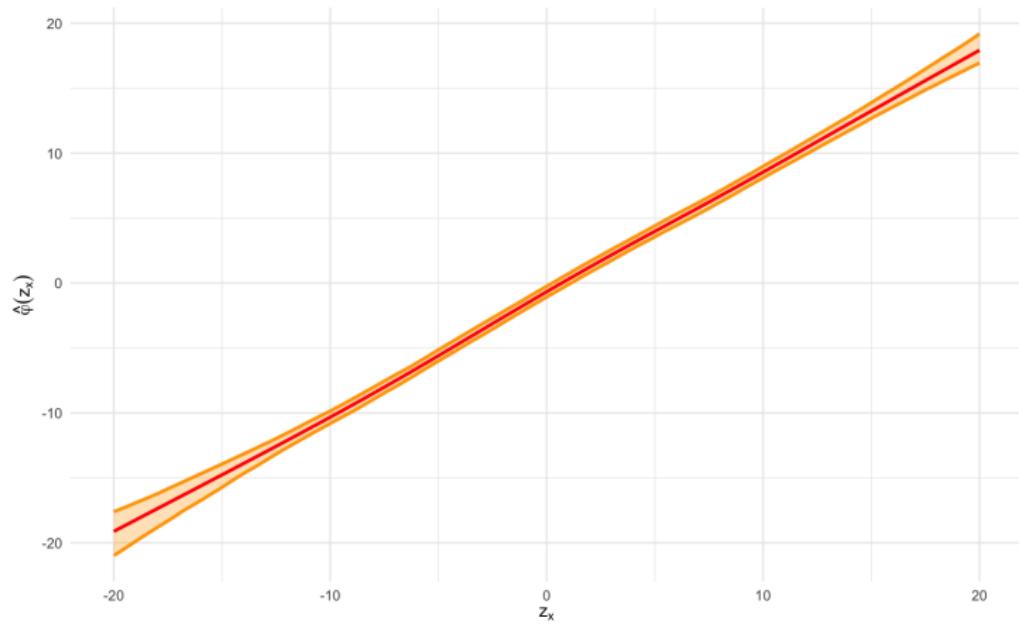


Figura 11: Función equipercentil latente estimada y su intervalo creíble del 95 %.

Aplicación: puntajes de prueba de Matemática

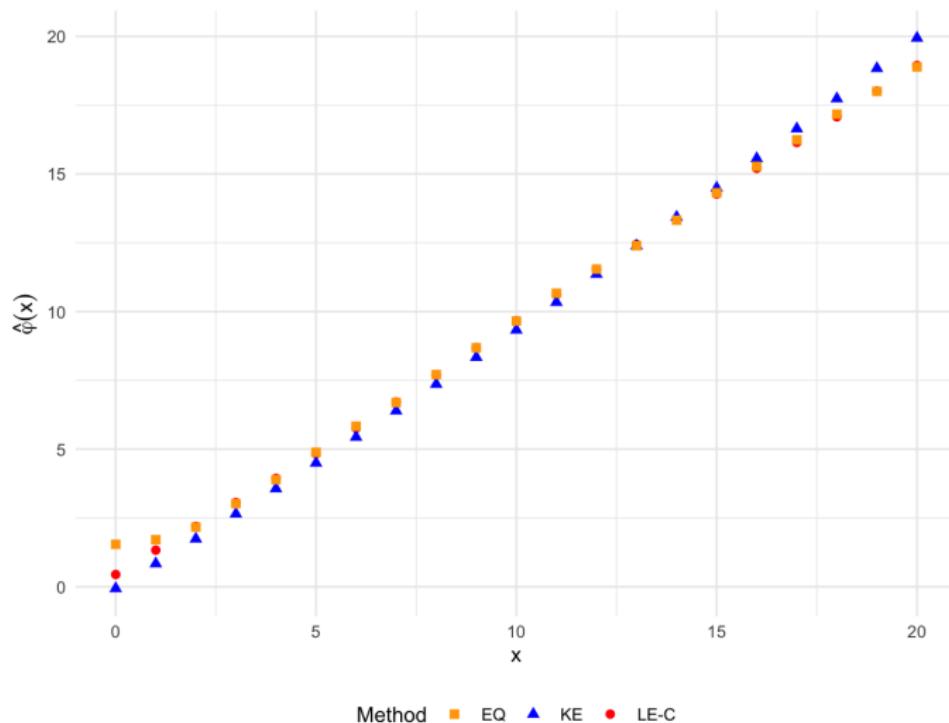


Figura 12: MathSG: puntajes equiparados bajo distintos métodos.

Aplicación: puntajes de prueba de Matemática

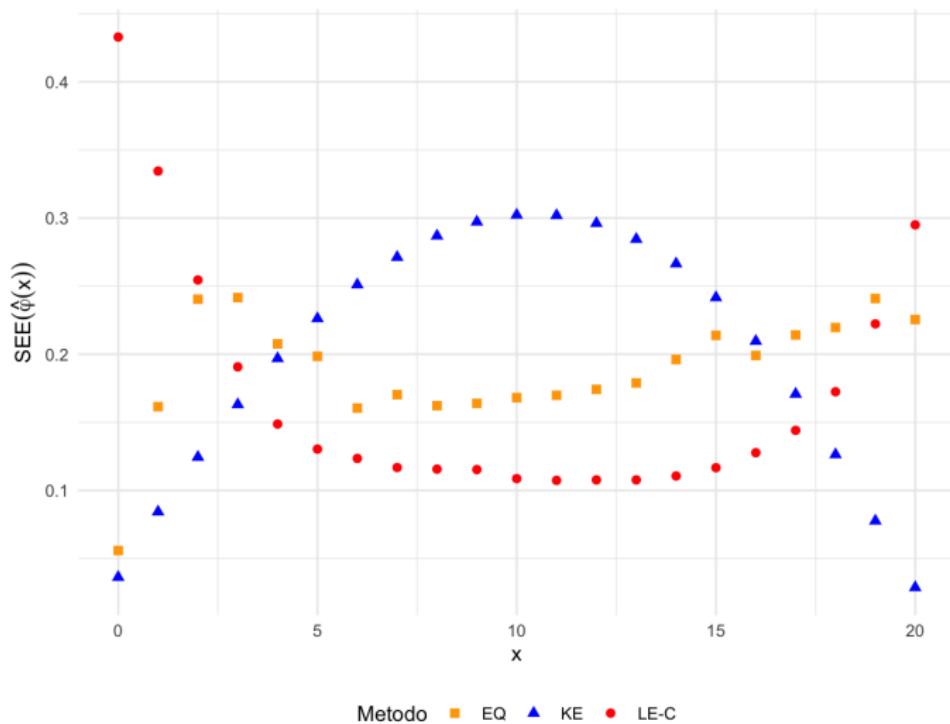


Figura 13: MathSG: error estándar de equiparación estimado.

Contenido

- 1 Equiparación de pruebas (equating)
- 2 Modelo latente para variables ordinales
- 3 LE-C: ilustración para el diseño de Grupo Único
- 4 Simulación y aplicación
- 5 Discusión

Discusión

- Considerando los puntajes de pruebas como variables ordinales, hemos propuesto un procedimiento de equiparación basado en ideas del método LE.
- Aprovechando el modelo Bayesiano, el procedimiento LE-C satisface la propiedad de simetría de la transformación de equiparación.
- Para tamaños muestrales moderados y grandes, la propuesta es tan competitiva como los métodos kernel normales y, en la mayor parte de la escala, los SEE son menores que en los métodos tradicionales de equiparación.

Referencias

- González, J., Wiberg. M. (2017). Applying Test Equating Methods. Springer International Publishing.
- von Davier, A. A., Holland, P., Thayer, D. (2004). The Kernel Method of Test Equating. Springer.
- Wiberg, M., González, J., von Davier, A. A. (2024). Generalized Kernel Equating with Applications in R. Boca Raton, FL: Chapman and Hall.
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.), pp. 508–600.
- Braun, H., Holland, P. (1982). Observed-score test equating: a mathematical analysis of some ETS equating procedures. In P. Holland and D. Rubin (Eds.), Test equating, Volume 1, pp. 9–49. New York: Academic Press.

Referencias

- Kottas, A., P. Muller, and F. Quintana (2005). Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphics Statistics*, 14.3, pp. 610–625.
- Ishwaran, H. and L. F. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 96(453), 161–173.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.

¡Gracias por su atención!
imvaras@uc.cl

Esta investigación fue financiada parcialmente por la Agencia Nacional de Investigación y Desarrollo (ANID), a través del proyecto FONDECYT 11251581.