

# *Equating*: Una Breve Introducción

**Jorge González**

Facultad de Matemáticas, Pontificia Universidad Católica de Chile  
LIES: Laboratorio Interdisciplinario de Estadística Social

Seminario de Estadística Educacional  
Septiembre 1, 2017, PUC, Chile



## LIES

Laboratorio Interdisciplinario de Estadística Social, Facultad de Matemática, Pontificia Universidad Católica de Chile

### LIES

Quiénes somos

Talleres de Investigación y Seminarios ▾

A quiénes servimos ▾

Formación continua y apoyo técnico

Publicaciones

En los medios

LIES es un espacio en el que diferentes académicos y estudiantes desarrollan un *ethos interdisciplinario*. Uno de los lenguajes comunes que intentamos desarrollar es el *modelamiento estructural de fenómenos sociales*. En este enfoque, utilizamos diferentes herramientas con el fin de modelar algunos aspectos esenciales que subyacen a las problemáticas sociales que nos convocan. Dicho modelamiento está constantemente monitoreado por consideraciones sustantivas, y por lo tanto queda abierto a mejores comprensiones de los comportamientos sociales.

Orientación de la investigación:

- Construcción de instrumentos para medir aprendizaje
- Comparabilidad de puntajes generados por instrumentos de evaluación aplicados en diferentes contextos
- Confiabilidad y validez: asegurar calidad técnica de instrumentos para el levantamiento de datos
- Evaluación de políticas públicas: relacionar mejoras esperadas con reformas implementadas
- Evaluaciones de impacto: determinar el potencial efecto de intervenciones implementadas



## Noticias

Inicio del proceso de Capacitación en Evaluación de Políticas USACH

30 Agosto, 2017

Seminario de Estadística Educacional: Equating

30 Agosto, 2017

Inicio Seminario de Estadística Educacional II Semestre

24 Agosto, 2017

Ernesto San Martín, Director de LIES, fue promovido a la categoría de Profesor Titular

30 Junio, 2017

Jorge González, Co-Director de LIES, publica su libro "Applying Test Equating Methods"

30 Marzo, 2017

# Outline

- 1 Introducción
- 2 El problema estadístico en equating
- 3 Inferencia estadística en equating
- 4 Problemas abiertos en equating

EDUCACIÓN NACIONAL

## Resultados del Simce revelan preocupante baja en comprensión de lectura

Oriana Miranda y Tania González | Martes 7 de junio 2016 12:56 hrs.



Las mujeres subieron sus resultados en matemáticas y ciencias, mientras los hombres los bajaron. En comprensión de lectura, en tanto, las mujeres mantuvieron los resultados del año anterior, mientras los hombres disminuyeron doce puntos.

# Las razones por las cuales el Simce de Lectura cayó 15 puntos en octavo básico

Autor: [Paulina Salazar](#)

La Agencia de Calidad dio a conocer resultados, en esta área y matemática, entre 2013 y 2014. Expertos califican la baja en el rendimiento de esta etapa escolar como "atípica y sugieren análisis.

RESULTADOS SIMCE 2014



## Uso de puntaje PSU en dos procesos de admisión consecutivos

Desde el Proceso de Admisión 2011 está vigente el uso de Puntaje PSU por dos procesos consecutivos. Acá te entregamos los requisitos y modo para hacer uso de éste.

Corre la voz! 🗣️

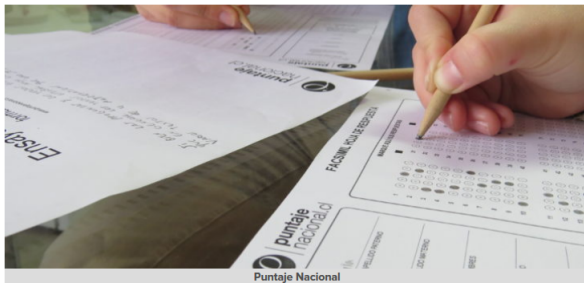


En caso que hayas rendido la Prueba de Selección Universitaria (PSU) el Proceso de Admisión 2016, rendida en 2015, puedes hacer uso del puntaje obtenido para postular el actual Proceso de


**Sociedad**

Lunes 19 Junio de 2017 | Publicado a las 16:59 · Actualizado a las 17:27

## 4 consejos para definir bien cuál prueba electiva rendir al momento de dar la PSU



1420 Visitas

URL Corta: <http://rbb.cl/h57c> 

Publicado por [Sebastián Asencio](#)

## Seguridad y ensayos

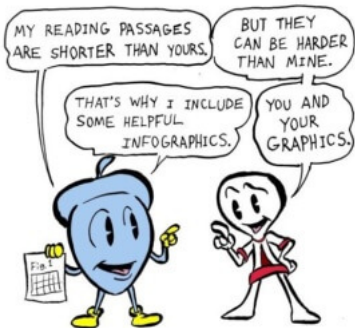
"Siempre será mejor rendir la prueba específica en la cual te sientas más seguro y sistemáticamente has obtenido los mejores resultados en los ensayos y notas del colegio, escuela o liceo. Esto, porque la postulación a las universidades no diferencia entre rendir una especialidad de Ciencias u otra", comenta Fabián Martínez, creador del sitio web Puntaje Nacional.



SIAD

Un ejemplo de ello es que para estudiar Ingeniería se puede rendir Física, Química o Biología. "Lo único que importa es obtener el puntaje que requiere la universidad", agrega el experto.





## SAT Reading vs. ACT Reading

David Recine for Magoosh.com

# Preguntas

- ¿Cómo podemos asegurar que se avanzó (retrocedió) en los puntajes de una prueba SIMCE?
- ¿Cómo podemos asegurar que fui admitido (rechazado) de manera justa considerando que mi puntaje PSU resultó de una prueba con menor (mayor) dificultad?
- ¿Cómo podemos reportar un puntaje único en la prueba PSU de ciencias?
- ¿Cómo podemos utilizar de manera justa los puntajes que provienen de dos pruebas distintas que miden un mismo atributo?

# Equating

## Equating

Los métodos de Equating han sido desarrollados para contestar a estas y otras preguntas.

## Objetivos de esta charla

- Introducir los modelos de equating para la comparabilidad de puntajes.
- Discutir algunos problemas abiertos en equating.

## Elementos involucrados

- Mediciones
- Escalas de medición
- Instrumentos de medición

## Transformación de mediciones: temperatura

- ¿A cuantos grados en la escala Fahrenheit equivalen 0 grados celsius?
- Función de transformación de escalas

$$F = \left( \frac{9}{5} \times C \right) + 32$$

- Resultado:  $32^{\circ}F$  significan lo mismo que  $0^{\circ}C$

## Transformación de mediciones: temperatura

- ¿A cuantos grados en la escala Fahrenheit equivalen 0 grados celsius?
- Función de transformación de escalas

$$F = \left( \frac{9}{5} \times C \right) + 32$$

- Resultado:  $32^{\circ}F$  significan lo mismo que  $0^{\circ}C$

## Transformación de mediciones: temperatura

- ¿A cuantos grados en la escala Fahrenheit equivalen 0 grados celsius?
- Función de transformación de escalas

$$F = \left( \frac{9}{5} \times C \right) + 32$$

- Resultado:  $32^{\circ}F$  significan lo mismo que  $0^{\circ}C$



## Propósitos del equating

- Realizar una transformación estadística de los puntajes de distintas formas de una prueba que ajusta por diferencias en dificultad, de manera que los puntajes puedan ser usadas de manera intercambiable.
  - Las distintas formas deben estar construidas de acuerdo al mismo contenido y con iguales especificaciones estadísticas
  - El equating es necesario pues es imposible desarrollar formas exactamente iguales en dificultad a lo largo de todo el rango de puntajes
- Debido a que las diferencias entre formas (dificultad) y diferencias entre grupos (habilidades) están confundidas, es necesario controlar por las últimas antes de realizar equating.

## Propósitos del equating

- Realizar una transformación estadística de los puntajes de distintas formas de una prueba que ajusta por diferencias en dificultad, de manera que los puntajes puedan ser usadas de manera intercambiable.
  - Las distintas formas deben estar construidas de acuerdo al mismo contenido y con iguales especificaciones estadísticas
  - El equating es necesario pues es imposible desarrollar formas exactamente iguales en dificultad a lo largo de todo el rango de puntajes
- Debido a que las diferencias entre formas (dificultad) y diferencias entre grupos (habilidades) están confundidas, es necesario controlar por las últimas antes de realizar equating.

## Propósitos del equating

- Realizar una transformación estadística de los puntajes de distintas formas de una prueba que ajusta por diferencias en dificultad, de manera que los puntajes puedan ser usadas de manera intercambiable.
  - Las distintas formas deben estar construidas de acuerdo al mismo contenido y con iguales especificaciones estadísticas
  - El equating es necesario pues es imposible desarrollar formas exactamente iguales en dificultad a lo largo de todo el rango de puntajes
- Debido a que las diferencias entre formas (dificultad) y diferencias entre grupos (habilidades) están confundidas, es necesario controlar por las últimas antes de realizar equating.

## Modelos estadísticos para equating

- Al igual que cualquier otro modelo estadístico, variables aleatorias, distribuciones de probabilidad y parámetros también juegan un rol en equating
- Establecer una teoría estadística para equating no sólo permite comprender mejor los métodos que existen hasta ahora, sino que también abre nuevas oportunidades de investigación en el área.

## Modelos estadísticos para equating: notación

- $X$  y  $Y$  son dos formas que serán equiparadas
- $X_i$ , ( $i = 1, \dots, n_x$ ) y  $Y_j$ , ( $j = 1, \dots, n_y$ ) son variables aleatorias que denotan los puntajes de  $n_x$  y  $n_y$  individuos a quienes se han administrados las formas  $X$  e  $Y$ , respectivamente.
- Espacios muestrales:  $X \in \mathcal{X}$ ,  $Y \in \mathcal{Y}$
- $F_X(x)$  y  $F_Y(y)$  son las funciones de distribución de  $X$  e  $Y$ , respectivamente.

## Modelos estadísticos para equating

- El problema estadístico de interés en equating será modelar la relación entre el puntaje en una forma de la prueba y su equivalente en la otra forma que se quiere equiparar.
- Matemáticamente, se debe definir una función que tome valores en  $\mathcal{X}$  y entregue como resultados un valor en  $\mathcal{Y}$ .

### Definition (Función de equating)

Sean  $\mathcal{X}$  y  $\mathcal{Y}$  dos espacios muestrales. Una función  $\varphi : \mathcal{X} \mapsto \mathcal{Y}$  se denomina *función de equating*.

## Modelos estadísticos para equating

- El problema estadístico de interés en equating será modelar la relación entre el puntaje en una forma de la prueba y su equivalente en la otra forma que se quiere equiparar.
- Matemáticamente, se debe definir una función que tome valores en  $\mathcal{X}$  y entregue como resultados un valor en  $\mathcal{Y}$ .

### Definition (Función de equating)

Sean  $\mathcal{X}$  y  $\mathcal{Y}$  dos espacios muestrales. Una función  $\varphi : \mathcal{X} \mapsto \mathcal{Y}$  se denomina *función de equating*.

## Definición: Braun and Holland (1982)

### Definition (Braun and Holland's (1982) definition of equating)

Sean  $X$  y  $Y$  dos formas de una prueba, cada una generando los puntajes  $X$  y  $Y$ , respectivamente. Las formas  $X$  y  $Y$  se dicen *equiparadas* en la población  $T$  por  $\varphi(x)$  si  $F_Y(y) = F_{\varphi(x)}(y)$

- Con esta definición, se puede obtener una forma explícita para la función de equating que equipara  $X$  a  $Y$  en  $T$ . En efecto:

### Theorem

El requerimiento  $F_Y(y) = F_{\varphi(x)}(y)$  implica que

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



## Definición: Braun and Holland (1982)

### Definition (Braun and Holland's (1982) definition of equating)

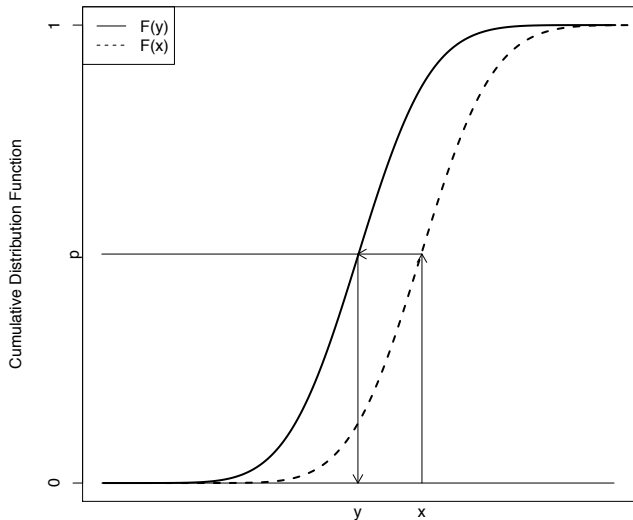
Sean  $X$  y  $Y$  dos formas de una prueba, cada una generando los puntajes  $X$  y  $Y$ , respectivamente. Las formas  $X$  y  $Y$  se dicen *equiparadas* en la población  $T$  por  $\varphi(x)$  si  $F_Y(y) = F_{\varphi(x)}(y)$

- Con esta definición, se puede obtener una forma explícita para la función de equating que equipara  $X$  a  $Y$  en  $T$ . En efecto:

### Theorem

El requerimiento  $F_Y(y) = F_{\varphi(x)}(y)$  implica que

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



## Equating lineal

- Si se asume una familia paramétrica de localización y escala para la distribución de puntajes de  $X$  y  $Y$  tal que

$$F_X(x) = H\left(\frac{x - \mu_X}{\sigma_X}\right)$$
$$F_Y(y) = H\left(\frac{y - \mu_Y}{\sigma_Y}\right),$$

entonces

$$\varphi(x; \boldsymbol{\theta}) = F_Y^{-1}(F_X(x; \mu_X, \sigma_X); \mu_Y, \sigma_Y) = \mu_Y + \frac{\sigma_Y}{\sigma_X}(x - \mu_X),$$

## Equating equipercantil

- Determinar los percentiles para las distribuciones de puntajes de ambas pruebas a ser equiparadas
- Parear percentiles equivalentes para un cierto puntaje

Puntaje	Percentiles	
	Forma X	Forma Y
10	99	99
9	95	93
8	90	80
7	80	62
6	65	42
5	45	25
4	25	15
3	13	10
2	5	5
1	1	1

## Equating equipercantil

- Determinar los percentiles para las distribuciones de puntajes de ambas pruebas a ser equiparadas
- Parear percentiles equivalentes para un cierto puntaje

Puntaje	Percentiles	
	Forma X	Forma Y
10	99	99
9	95	93
8	90	<b>80</b>
7	<b>80</b>	62
6	65	42
5	45	<b>25</b>
4	<b>25</b>	15
3	13	10
2	5	5
1	1	1

## Equating equipercentil

- ¿cuál es el puntaje equivalente en  $Y$  para 6 puntos en la forma  $X$ ?
- Problema: puntajes discretos!

## Equating equipercentil

- ¿cuál es el puntaje equivalente en  $Y$  para 6 puntos en la forma  $X$ ?
- Problema: puntajes discretos!

## Función de equating equipercantil: observaciones

- Todos los modelos que serán descritos están basados en la función  $\varphi(x) = F_Y^{-1}(F_X(x))$ .
- Esto implica que todos los métodos de equating se ven afectados por la naturaleza discreta de los puntajes (y distribuciones), lo que impide el cálculo de la inversa.
- En la práctica, se utilizan aproximaciones *continuas* de las distribuciones de puntajes originalmente discretas (*continuización*)
- Métodos de continuización típicamente utilizados en equating incluyen interpolación lineal, técnicas de kernel smoothing y modelos log-lineales polinomiales.



## Modelos paramétricos, semiparamétricos y no-paramétricos

- Los modelos estadísticos en equating admiten estimadores paramétricos, semiparamétricos y no-paramétricos de  $\varphi$  (González and von Davier, 2013).
  - **Paramétrico:**  $\varphi = \varphi(\cdot, \theta_x, \theta_y)$
  - **Semiparamétrico:**  $\varphi = \varphi(\cdot, F_X, F_Y, \theta_x, \theta_y)$
  - **No-paramétrico:**  $\varphi = \varphi(\cdot, F_X, F_Y)$

## Ejemplos de estimadores paramétricos, semiparamétricos y no-paramétricos

- **Paramétrico:** Equating lineal, Tucker, Levine, IRT True-score equating (Kolen and Brennan, 2004)
- Ejemplo: equating lineal

$$\varphi(x; \boldsymbol{\theta}) = \mu_Y + \frac{\sigma_Y}{\sigma_X}(x - \mu_X),$$

## Ejemplos de estimadores paramétricos, semiparamétricos y no-paramétricos

- **Semiparamétrico**: Kernel equating (von Davier et al., 2004; González and von Davier, 2016); Local equating (van der Linden, 2011), y combinaciones (Wiberg et al., 2014)
- Ejemplo: Kernel equating

$$\varphi(x; \theta) = F_Y^{-1}(F_X(x; \mathbf{r}), \mathbf{s}),$$

$$F_X(x) = \sum_j K(R_j(x))r_j, \quad r_j = P(X = x_j),$$

$$R_j(x) = \frac{x - a_X x_j - (1 - a_x)\mu_X}{a_X h_X}, \quad a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_V^2 h_X^2}$$

## Ejemplos de estimadores paramétricos, semiparamétricos y no-paramétricos

- **Semiparamétrico**: Kernel equating (von Davier et al., 2004; González and von Davier, 2016); Local equating (van der Linden, 2011), y combinaciones (Wiberg et al., 2014)
- Ejemplo: Kernel equating

$$\varphi(x; \boldsymbol{\theta}) = F_Y^{-1}(F_X(x; \mathbf{r}), \mathbf{s}),$$

$$F_X(x) = \sum_j K(R_j(x))r_j, \quad r_j = P(X = x_j),$$

$$R_j(x) = \frac{x - a_X x_j - (1 - a_x)\mu_X}{a_X h_X}, \quad a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_V^2 h_X^2}$$

# Ejemplos de estimadores paramétricos, semiparamétricos y no-paramétricos

- **No-paramétrico**: equipercentile equating

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

## Inferencia estadística: métodos de estimación

Dependiendo del tipo de modelo;

- Método de momentos entrega medias y varianzas muestrales  $\theta = (\mu_X, \mu_Y, \sigma_X, \sigma_Y)$  en equating lineal.
- Máxima verosimilitud para estimar  $\theta_1 = (r, s)$  en kernel equating utilizando ya sea modelos log-lineales o modelos IRT.
- Kernel density estimation para estimar  $\theta_2 = (F_X, F_Y)$  en kernel equating.
- Maximum likelihood, Empirical Bayes, Maximum a posteriori, expected a posteriori para estimar  $\theta$  en local equating

## Inferencia estadística: métodos de estimación

Dependiendo del tipo de modelo;

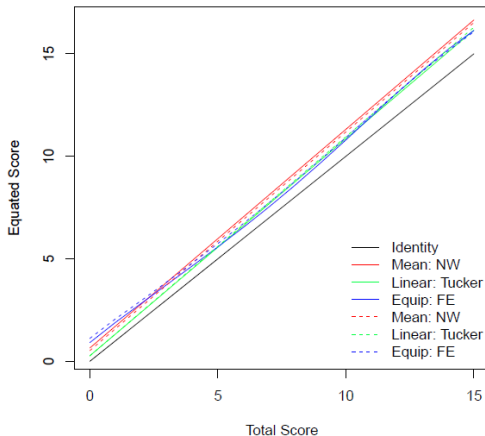
- **Método de momentos** entrega medias y varianzas muestrales  $\theta = (\mu_X, \mu_Y, \sigma_X, \sigma_Y)$  en equating lineal.
- **Máxima verosimilitud** para estimar  $\theta_1 = (r, s)$  en kernel equating utilizando ya sea modelos log-lineales o modelos IRT.
- **Kernel density estimation** para estimar  $\theta_2 = (F_X, F_Y)$  en kernel equating.
- **Maximum likelihood, Empirical Bayes, Maximum a posteriori, expected a posteriori** para estimar  $\theta$  en local equating

## Inferencia estadística: métodos de estimación

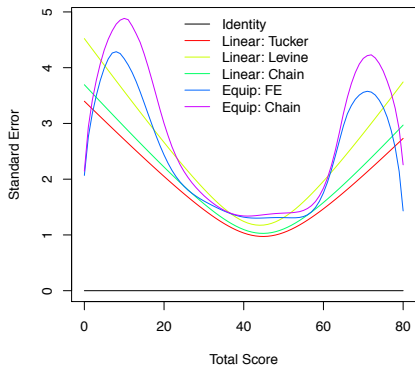
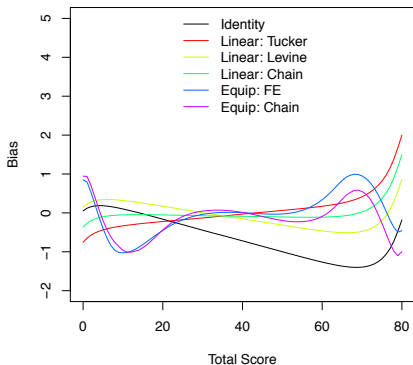
- **Fórmula recursiva** para estimar  $F_{X|\theta}$  y  $F_{Y|\theta}$  en local equating (Lord and Wingersky, 1984; González et al., 2016)
- **Bayesian nonparametrics** para estimar  $F_X$  and  $G_Y$  (Karabatsos and Walker, 2009).
- **Dependent Bayesian nonparametrics** para estimar  $F_X$  y  $F_Y$  cuando se incorporan covariables en el análisis (González et al., 2015)



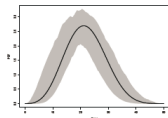
## Ejemplos



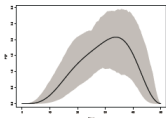
## Examples



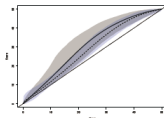
## Ejemplos



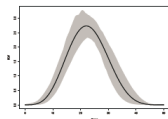
(a) Municipal in Form X



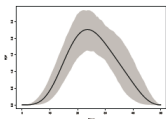
(b) Municipal in Form Y



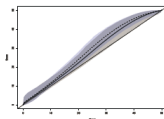
(c) Municipal in Form X to Y



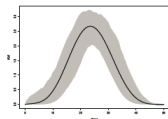
(d) Subsidized in Form X



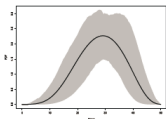
(e) Subsidized in Form Y



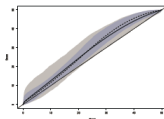
(f) Subsidized in Form X to Y



(g) Private in Form X



(h) Private in Form Y

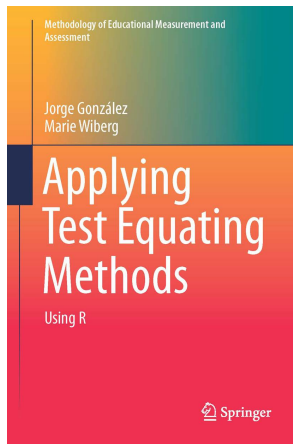


(i) Private in Form X to Y

## Software para equating

- R packages
  - Métodos tradicionales de equating: `equate` (Albano, 2014)
  - Kernel equating: `kequate` (Andersson et al., 2013)
  - Métodos tradicionales y otros para equating: `SNSequate` (González, 2014)

# Libro



## Otros puntajes y escalas

- Modelos IRT, puntajes IRT, escala IRT.
- No sólo datos binarios (correcto/incorrecto)
- No sólo instrumentos que miden atributos cognitivos
- ¿Métodos de equating para puntuaciones politómicas?
- Proyecto Doctoral: Francisca Calderón.

## Otros puntajes y escalas

- Modelos IRT, puntajes IRT, escala IRT.
- No sólo datos binarios (correcto/incorrecto)
- No sólo instrumentos que miden atributos cognitivos
- ¿Métodos de equating para puntuaciones politómicas?
- Proyecto Doctoral: Francisca Calderón.

# Coherencia de métodos existentes y fundamentación estadística

- ¿Es correcto hablar de *kernel* equating?
- Equating basado en convoluciones
- Equating que intenta preservar al máximo la naturaleza discreta de los puntajes
- Proyecto Doctoral: Inés Varas.



## Coherencia de métodos existentes y fundamentación estadística

- ¿Es correcto hablar de *kernel* equating?
- Equating basado en convoluciones
- Equating que intenta preservar al máximo la naturaleza discreta de los puntajes
- Proyecto Doctoral: Inés Varas.

## Problemas de indentificabilidad

- Identificación parcial de funciones de distribución de puntajes (González, San Martín, in progress).
- Linking de parámetros en modelos IRT (Proyecto Doctoral: Gabriel Muñoz)

## Uso de covariables

- Equating con covariables

# Gracias por su atención!

**Jorge González B.**

Department of Statistics, Faculty of Mathematics,  
Pontificia Universidad Católica de Chile  
Av. Vicuña Mackenna 4860, Macul, Santiago, Chile  
E-mail: [jorge.gonzalez@mat.uc.cl](mailto:jorge.gonzalez@mat.uc.cl)  
web: [www.mat.uc.cl/~jorge.gonzalez](http://www.mat.uc.cl/~jorge.gonzalez)

This research is supported by grant Fondecyt 1150233

- Andersson, B., K. Bränberg, and M. Wiberg (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software* (6), 1–25.
- González, J. (2014). SNSequate: Standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software* (7), 1–30.
- González, J., A. F. Barrientos, and F. A. Quintana (2015). A Dependent Bayesian Nonparametric Model for Test Equating. In R. Millsap, D. Bolt, L. van der Ark, and W.-C. Wang (Eds.), *Quantitative Psychology Research*, pp. 213–226. Springer International Publishing.
- González, J. and A. A. von Davier (2016). Comparisons of the Epanechnikov, adaptive, and Gaussian kernels in test equating. *In preparation*.
- González, J. and M. von Davier (2013). Statistical models and inference for the true equating transformation in the context of local equating. *Journal of Educational Measurement* (3), 315–320.
- González, J., M. Wiberg, and A. A. von Davier (2016). A note on the Poisson's binomial distribution in item response theory. *Applied Psychological Measurement* (4), 302–310.
- Karabatsos, G. and S. Walker (2009). A Bayesian nonparametric approach to test equating. *Psychometrika* (2), 211–232.
- Kolen, M. and R. Brennan (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.
- Lord, F. and M. Wingersky (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement* (4), 453–461.
- van der Linden, W. J. (2011). Local observed-score equating. In A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking*, pp. 201–223. New York: Springer.
- von Davier, A. A., P. Holland, and D. Thayer (2004). *The kernel method of test equating*. New York: Springer.
- Wiberg, M., W. J. van der Linden, and A. A. von Davier (2014). Local observed-score kernel equating. *Journal of Educational Measurement* 51, 57–74.