

NCDIF y Regresión logística como procedimientos para detectar funcionamiento diferencial del ítem (DIF) en Saber 5° y 9°



Víctor H. Cervantes* - Martha Ligia Cuevas**
Subdirección de Estadística – Instituto Colombiano para la Evaluación de la Educación (ICFES)

Introducción

- Validez.
- Interpretaciones iguales de las puntuaciones por género, zona, sector y nivel socioeconómico.
- Equidad.
- Funcionamiento Diferencial del Ítem (DIF o FDI).
- ¿Impacto? ¿Sesgo?
- Procedimientos específicos.

Índice no compensatorio de DIF (NCDIF)

Propuesto como una alternativa basada en la teoría de respuesta al ítem (TRI) para la detección de DIF (Raju, van der Linden & Fleer, 1995).

$$NCDIF_i = \int_{-\infty}^{\infty} [P_{iF}(\theta) - P_{iR}(\theta)]^2 f_F(\theta) d\theta$$

Puede verse afectado por variables como razón de tamaños entre los grupos que se comparan, tamaño de muestra, y parece estar relacionado con parámetros específicos de los ítems como dificultad y discriminación (Bolt, 2002; Oshima, Raju & Nanda, 2006).

La prueba estadística propuesta para probar la hipótesis de no DIF no se aproxima a la distribución χ^2 inflando las tasas de error tipo I para muestras grandes (Raju et al, 2006).

Regresión Logística - RL

- Usado en psicología y educación.
- Coeficientes para el grupo, puntuación total y su interacción.
- Prueba de significancia.
- Variable dependiente: respuesta de los examinados.
- Suele verse afectado por razón de tamaños de muestra, tamaño de muestra, dificultad y discriminación (Herrera, 2005; Santana, 2009).

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta + \beta_2g + \beta_3\theta g$$

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta + \beta_2g$$

$$\text{logit}(P(U = 1)) = \beta_0 + \beta_1\theta$$

Objetivos y preguntas de investigación

- ¿Cómo funcionan el NCDIF y la RL bajo diferentes condiciones teniendo en cuenta su potencia y tasas de error tipo I?
- Identificar algunas variables que pueden afectar el NCDIF y la RL.
- Encontrar los mejores puntos de corte para el NCDIF y el ΔR^2 en la detección de DIF, tomando en cuenta algunas variables que los afectan

Metodología

Estudio de simulación.

Variables independientes

- Tamaño de muestra: n = 7500, 8500, 26000 y 33000.
- Razón de tamaños entre los grupos: 1, 3, 4, 6, 10 y 20 para NCDIF, y 1, 3, 4, 6, 10, 20, 150 y 250 para RL.
- Longitud de la prueba: K = 18 y 26.
- Magnitud de DIF: No DIF, dos ítems con DIF (pequeño y grande), y cuatro ítems con DIF (pequeño y grande).

220 condiciones experimentales para NCDIF y 270 para RL.

Réplicas por condición: R = 200.

Dos ítems con DIF uniforme y dos ítems con DIF no uniforme cuando la condición tenía cuatro ítems.

Variables fijas: dificultad y discriminación.

Variables dependientes y modelos

- NCDIF (medida de tamaño del efecto).
- ΔR^2 de McFadden (medida de tamaño del efecto, se introduce el efecto de grupo para DIF uniforme y de la interacción para DIF no uniforme).
- Modelos lineales para definir cuáles variables afectaban la medida de tamaño del efecto.

Puntos de Corte

- Curvas operador receptor (COR o ROC).
- Para cada condición.
- Para cada combinación de variables con efecto sobre el NCDIF y el ΔR^2 .
- Selección de los mejores puntos de corte en forma eficiente, es decir, manteniendo buenas tasas de error tipo I y potencia.

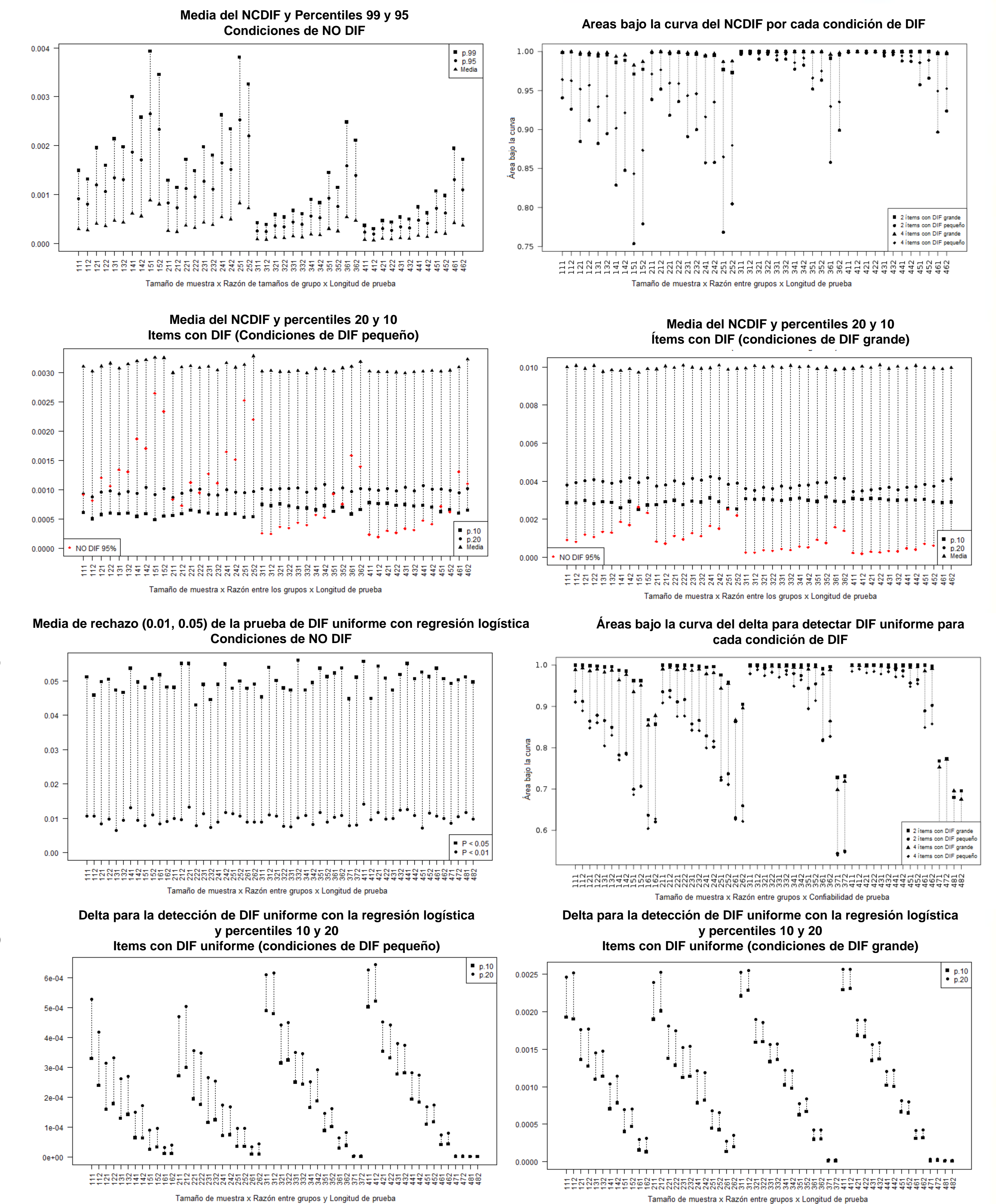
Resultados

Tabla 1. Efectos en el NCDIF, el ΔR^2 para DIF uniforme y el ΔR^2 para DIF No uniforme

Variable	NCDIF		ΔR^2 para DIF uniforme		ΔR^2 para DIF No uniforme	
	F	p	F	p	F	p
Tamaño	364,008	1,52E-195	70,61	9,69E-44	27,601	2,95E-17
Razón	185,955	6,48E-165	473,287	0	133,511	6,E-145
Magnitud de DIF	2,442	0,0447	892,215	0	638,029	5,E-113
Longitud de prueba	4,733	0,03	3,983	0,046		
Tamaño x Magnitud de DIF			3,778	9,64E-05	4,081	0,007
Razón x Magnitud de DIF			17,012	3,40E-58	30,403	7,42E-39
Magnitud de DIF x Longitud de prueba					4,122	0,043

Referencias

Bolt, D. (2002) A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied measurement in education*, 15,113-141.
 Herrera, A. N. (2005) *Efecto del tamaño de muestra y razón de tamaños de muestra en la detección de funcionamiento diferencial de los ítems*. Tesis de Doctorado en Evaluación y Tecnología Informática en Ciencias del Comportamiento. Universidad de Barcelona, España.
 Oshima, T., Raju, N. & Nanda, A. A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of educational measurement*, 43, 1-17.
 Raju, N., van der Linden, W. & Fleer, P. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied psychological measurement*, 19, 353-368.
 Santana, A. C. (2009) *Efecto de la razón de tamaños de muestra en la detección de funcionamiento diferencial de los ítems a través del procedimiento de regresión logística*. Tesis de Maestría en Psicología. Universidad Nacional de Colombia, Colombia.



Discusión

- Aunque algunos puntos de corte fueron obtenidos para cada condición y algunas variables tienen efecto sobre las medidas de tamaño del efecto, estos puntos de corte pudieron ser simplificados sin perder potencia ni control sobre las tasas de error Tipo I para algunas condiciones.
- Es necesario contar con puntos de corte específicos de las medidas de tamaño del efecto para condiciones particulares.
- El NCDIF parece poco eficiente para detectar DIF cuando uno de los grupos tiene un tamaño de muestra pequeño.
- Los resultados concuerdan con otros estudios como el de Bolt (2002) y Santana (2009).



* vcervantes@icfes.gov.co
 ** mcuevas@icfes.gov.co