

Modelo multidimensional de Teoría de Respuesta al Ítem Jerárquico*

Diana Stella Antonio
Álvaro Montenegro

Universidad Nacional de Colombia

*Esta investigación recibió financiamiento del Instituto Colombiano Para la Evaluación de la Educación – ICFES a través de la Convocatoria de Estudiantes de la Oficina Asesora de Gestión de proyectos de Investigación

Justificación

- Análisis de resultados de pruebas internacionales.
- Evaluación de aspectos del sistema de educación de cada país.
- Comparación de los resultados entre países.
- Es difícil lograr la invarianza de los ítems.
- Se pueden presentar diferencias entre poblaciones.
- Modelos TRI para analizar los resultados.
- Un ítem puede ser sensible a diferencias en múltiples dimensiones.
- Los análisis con modelos TRIM pueden ayudar a aclarar tanto las diferencias entre las poblaciones, como las razones del funcionamiento diferencial del ítem (Reckase 2009).

Introducción

- Los resultados de los análisis deben ser comparables entre clases de respondientes.
- Los trazos latentes de los respondientes son multidimensionales y provienen de una única distribución.
- En el modelo propuesto se incluyen dos conjuntos de parámetros de los ítems para modelar las diferencias entre clases.
- El modelo propuesto puede ser usado con ítems invariantes y no invariantes.

Modelo propuesto

$$P(Y_{igk} = 1 | \theta_{ig}, \xi_{kg}) = \frac{e^{a'_{kg} \theta_{ig} + d_{kg}}}{1 + e^{a'_{kg} \theta_{ig} + d_{kg}}}$$

Y_{igk} variable aleatoria que representa la respuesta Y del individuo i de la subpoblación g al ítem k

a_{kg} vector de parámetros de discriminación del ítem k en la subpoblación g

d_{kg} parámetro de intercepto del ítem k en la subpoblación g

θ_{ig} vector de trazos latentes de la persona i en la subpoblación g

Datos PERCE

- Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado (PERCE) (LLCE, 1997).
- Subpruebas: Matemática (32 preguntas) y Lenguaje (19 preguntas).
Cuestionarios A y B.
- Muestra: respuestas del mismo estudiante en las dos subpruebas, seleccionados aleatoriamente, Bolivia (252 datos), Colombia (194 datos) y Cuba (156 datos). 6 % de los estudiantes de cada país del cuestionario B.
- Dimensiones de los trazos latentes: 3 dimensiones resultado de un análisis paralelo conjunto.

Trazo latente Unidimensional G

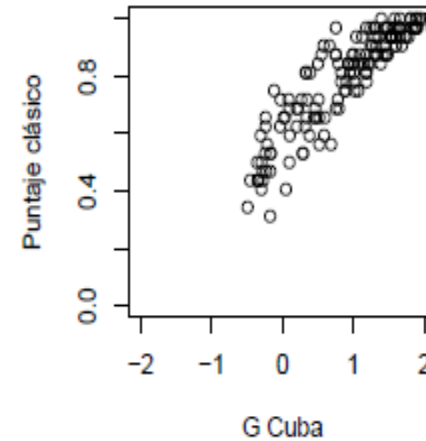
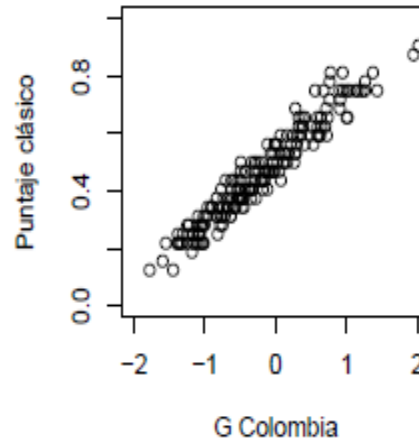
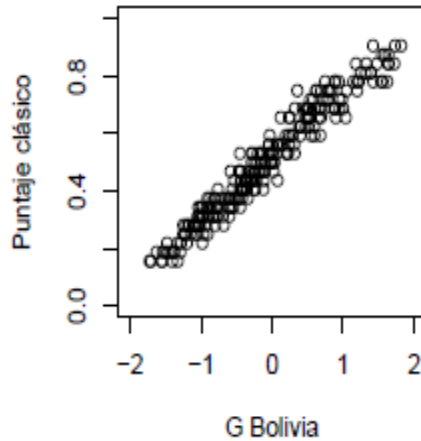
$$G = w * \theta$$

(Wang, 1986)

- w es el primer vector propio de la matriz $A * A^T$, dirección principal de la subprueba
- θ trazo latente estimado con el modelo TRIMJ

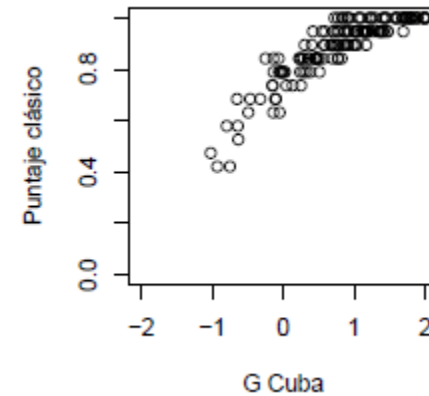
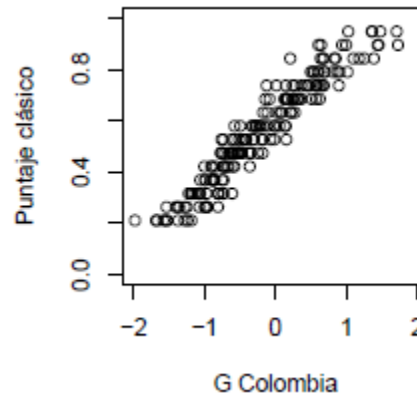
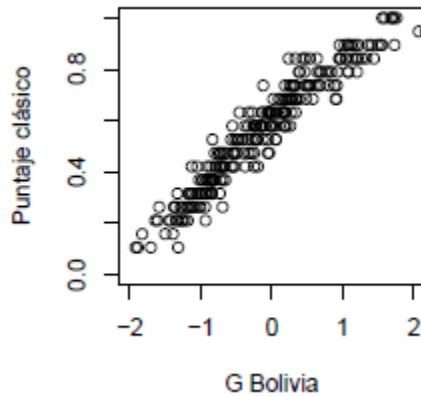
G es el trazo latente unidimensional aproximado

Resultados en Matemáticas



Puntaje clásico vs. trazos estimados con los parámetros nacionales de los ítems de Matemática por país.

Resultados en Lenguaje



Puntaje clásico vs. trazos estimados con los parámetros nacionales de los ítems de Lenguaje por país.

Discriminación multidimensional (MDISC) internacional y por país con TRIMJ

MDISC Matemática	Bolivia	Colombia	Cuba	Internacional
Item 2	1.48	1.58	2.25	1.54
Item 3	1.29	1.27	2.38	1.28
Item 4	2.13	2.19	1.72	2.14
Item 5	0.95	0.6	2.84	0.79
Item 7	0.78	0.68	7.87	0.74
Item 8	1.21	1.23	3.86	1.22
Item 10	1.84	1.81	2.22	1.82
Item 16	2.01	1.61	2.43	1.84
Item 18	1.64	2.19	2.03	1.89
Item 28	0.45	0.81	1.93	1.03
MDISC Lenguaje	Bolivia	Colombia	Cuba	Internacional
Item 36	2.15	1.82	1.9	1.91

Dificultad multidimensional por país (MDIFF) con TRIMJ

MDIFF Matemática	Bolivia	Colombia	Cuba	Internacional
Item 14	0.8	0.9	0.21	0.74
Item 16	0.53	0.86	0.25	0.55
Item 19	1.28	1.39	0.26	0.94
Item 21	0.74	0.48	0.16	0.46
Item 28	2.51	1.2	0.58	1.05
Item 29	0.4	0.29	-0.63	0
Item 30	-0.3	-0.49	-0.72	-0.51
Item 31	0.91	0.79	-0.72	0.3
MDIFF Lenguaje	Bolivia	Colombia	Cuba	Internacional
Item 43	0.32	-0.75	-1.4	-0.62
Item 44	0.15	1	-0.66	0.16
Item 48	-1.58	-2.13	-1.88	-1.87

Evaluación del modelo TRIMJ

Comparación DIC modelos TRIM y TRIMJ

Parámetro	TRIM	TRIMJ
a1	66.542	66.542
a2	66.542	30.115
a3	66.542	53.012
Response	31356.7	30875.5
Total	31556.4	31099.5

Porcentaje de valores atípicos y extremos

	Outliers	Valores extremos
Global	0.09	0.04
Bolivia	0.03	0.01
Colombia	0.03	0
Cuba	0.2	0.12

HABILIDADES
EFICIENCIA

Conclusiones

- El modelo TRIMJ se ajusto mejor a los datos, que el modelo TRIM de dos parámetros de Reckase (2009).
- En la muestra de datos de la prueba PERCE, los parámetros de los ítems mostraron diferencias entre países.
- Los resultados de Cuba fueron diferentes en la mayoría de los análisis desarrollados: parámetros de los ítems, distribuciones de los trazos latentes.
- Los individuos de la muestra de datos reales que respondieron igual a los ítems de una subprueba, tuvieron diferentes trazos latentes.

Bibliografía

- Bolt, D. (2003), Estimation of compensatory and noncompensatory multidimensional item response models using markov chain monte carlo, *Applied Psychological Measurement* 26(6).
- Casassus, J., F. J. P. J. and Cusato, S. (1998), *Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados en Tercero y Cuarto Grado*, Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.
- Fox, J. (2010), *Bayesian item response modeling*, Springer Science+Business Media
- Reckase, M. (2009), *Multidimensional item response theory*, Michigan State University