



Programa
Universitario
de Estudios
del **Desarrollo**
UNAM

Reliability / Confiabilidad

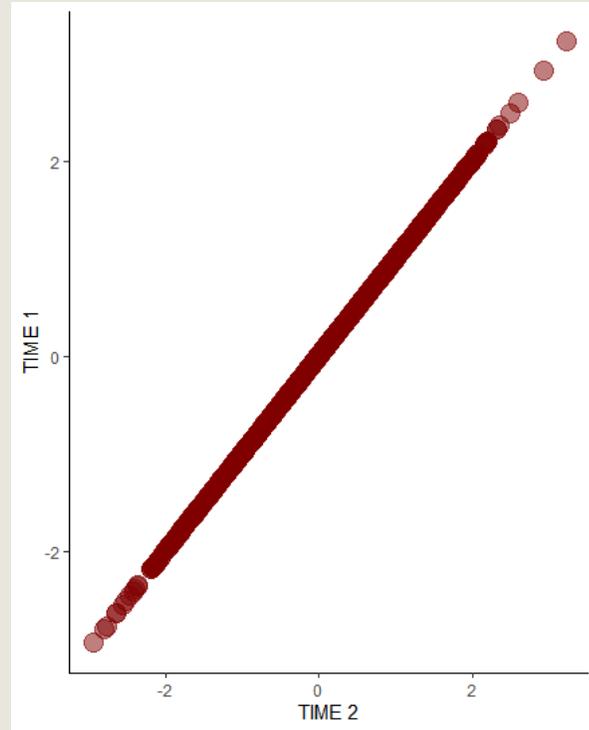
Héctor Nájera

Curso internacional de teorías y métodos contemporáneos para la
medición de la pobreza multidimensional

29 Nov 2021

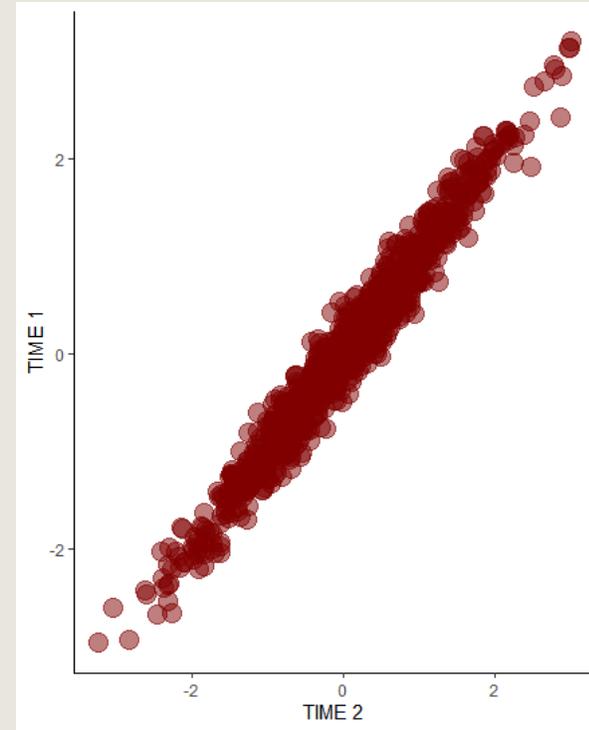
Confiabilidad como consistencia

Sin cambios en T1 y T2



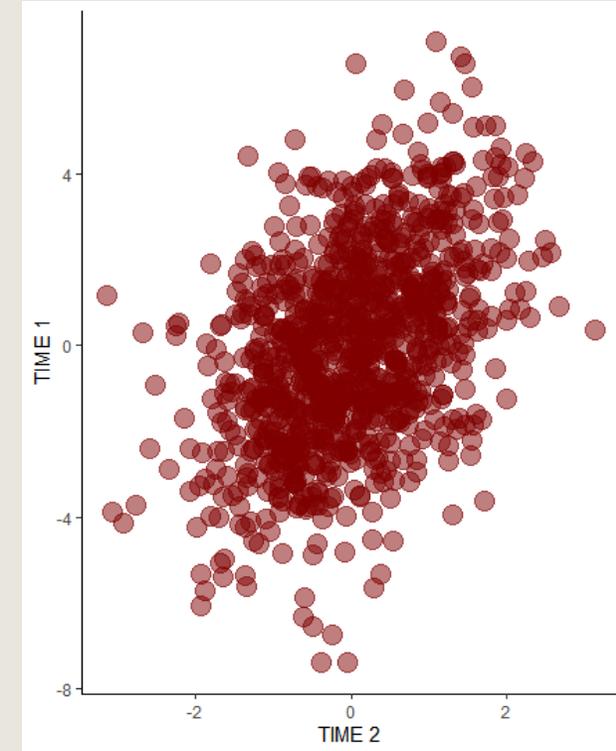
Confiamos en los **scores**, instrumento y su modelo subyacente

Pequeño error T1 y T2

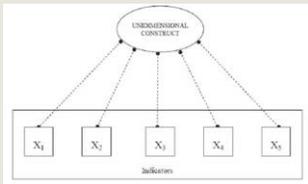


Confiamos en los **scores**, instrumento y su modelo subyacente –toleramos pequeñas variaciones–

Mucho error T1 y T2

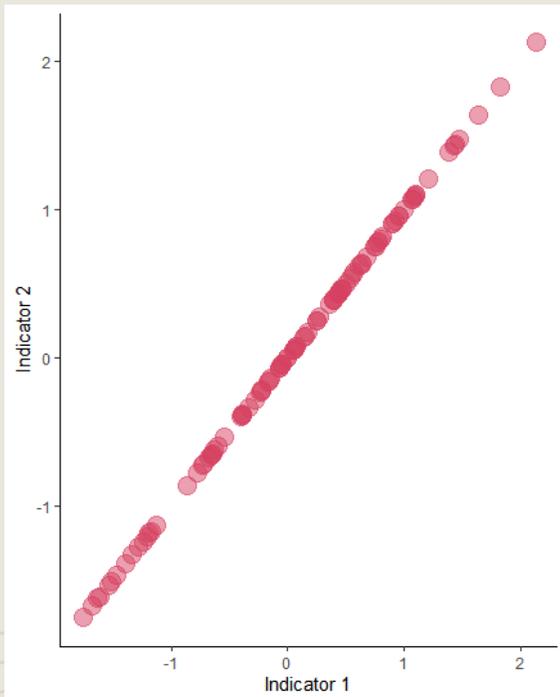


No confiamos en los **scores**

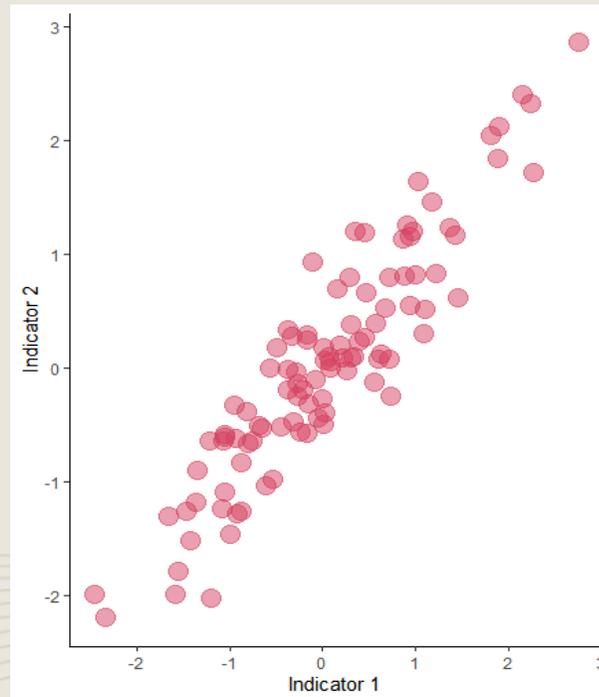


Fundamentos del principio de confiabilidad

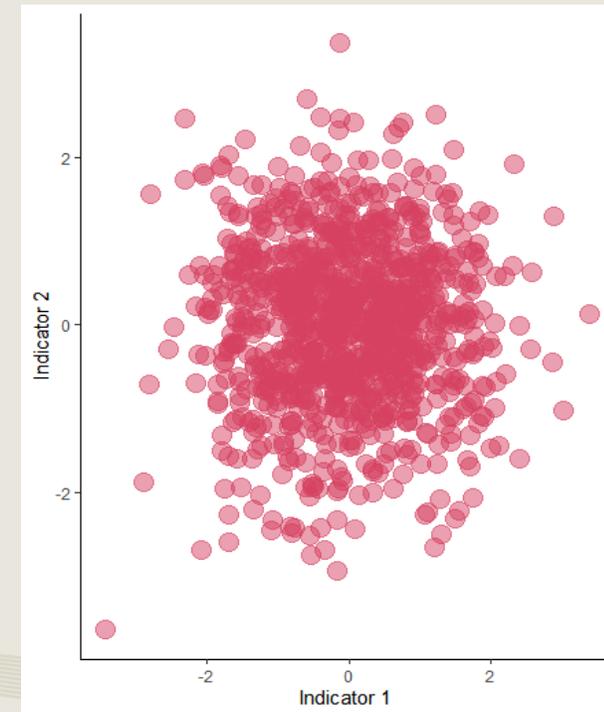
Si la variabilidad total de dos indicadores se explica **por la misma** fuente:



Si la variabilidad de dos indicadores se explica **mayormente** por la misma fuente:



Si la variabilidad de dos indicadores **no tiene fuente común**:



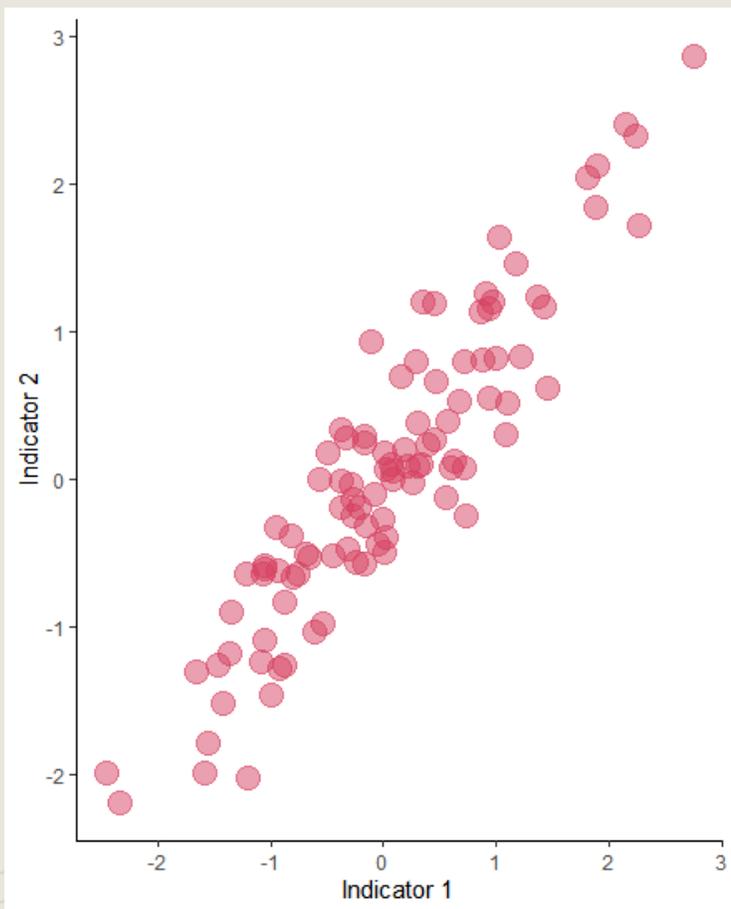
Las relaciones entre manifestaciones de **un mismo fenómeno** se **atenúan** por **error de medición**



Confiabilidad y variables latentes



Lo que teorizamos es que los que es observable son las manifestaciones de pobreza: A cambios en pobreza -latente-, cambios en las lecturas que hacemos



Las privaciones son manifestaciones - consecuencias- de la pobreza

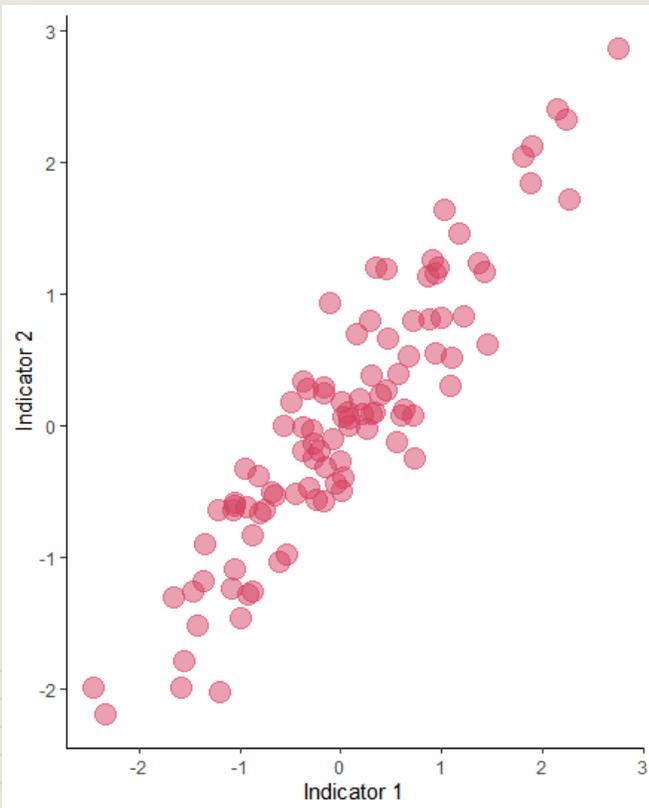


Monotonicidad: Caídas de ingreso y aumentos de pobreza

Señal y ruido: Error de medición

Confiabilidad = consistencia

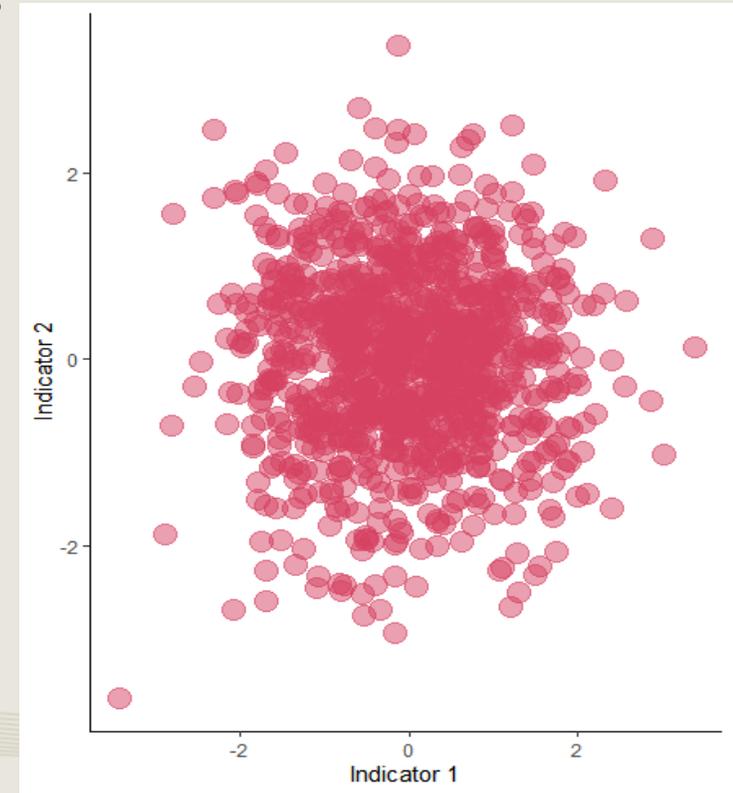
(Sin) Cambios en pobreza,
(Sin) Cambios en los indicadores



Ruido = Error

Toda aquella variabilidad que no me interesa

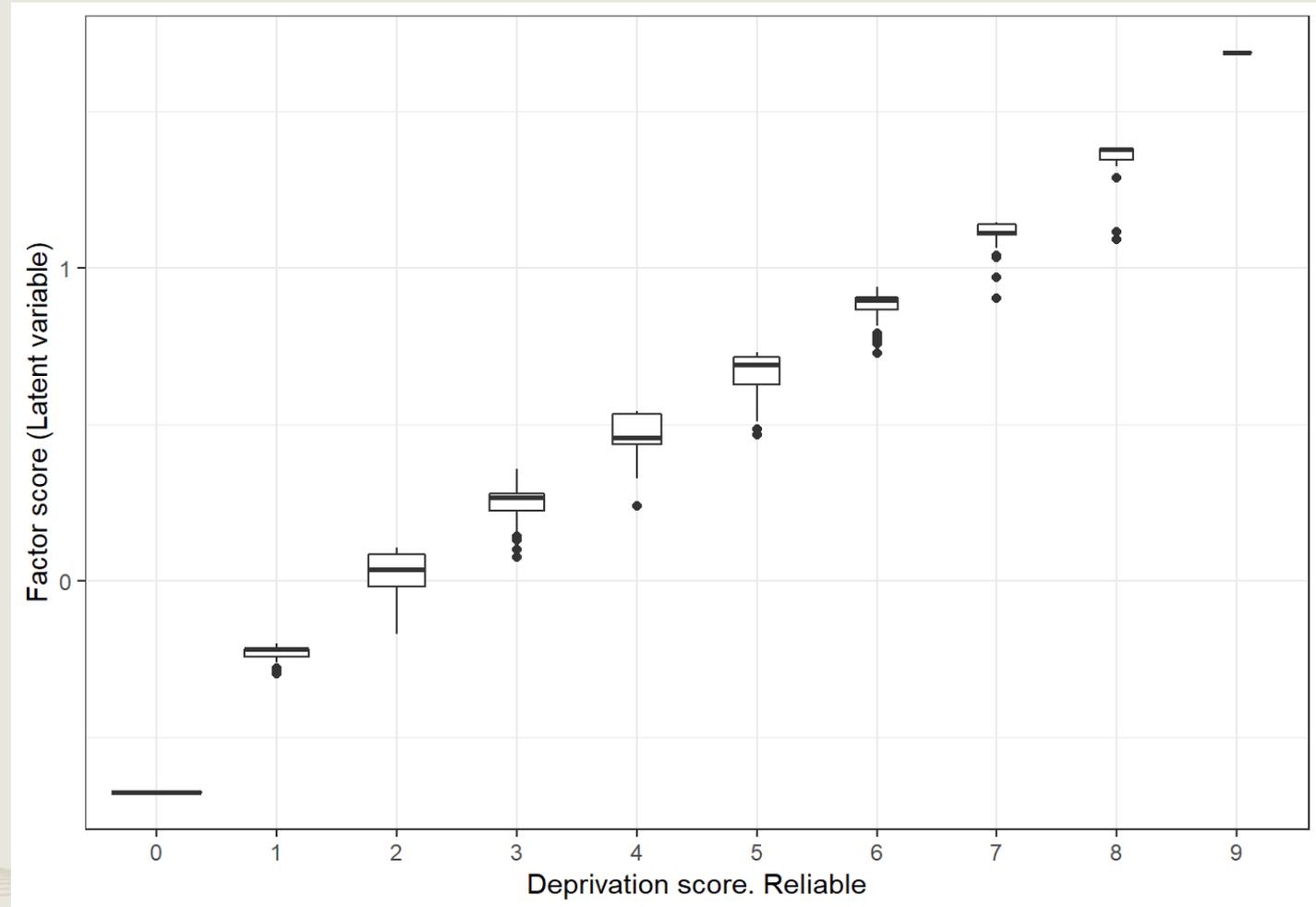
Cambios en el fenómeno de interés no se registran en mis observaciones



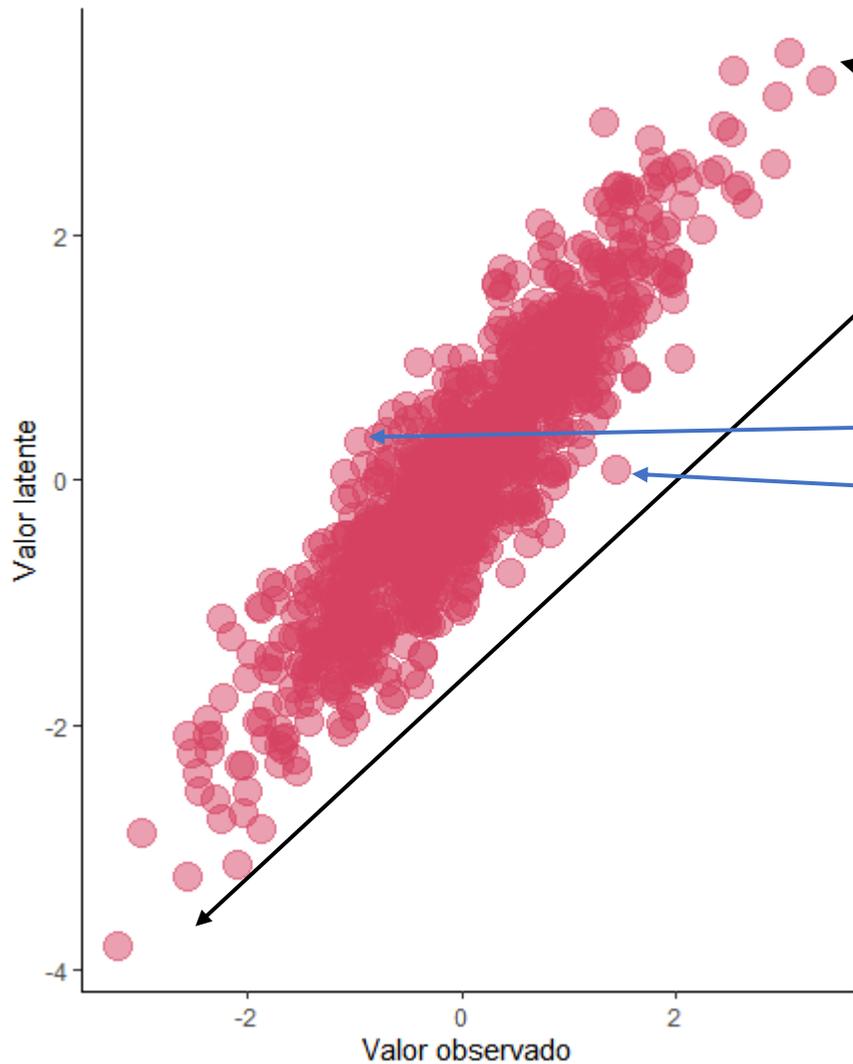


Confiabilidad y scores/puntajes

- Estadísticamente se entiende como la **razón de señal y ruido** de una medición
- Conceptualmente significa que los scores son **homogéneos**, i.e. su variabilidad representa una misma señal
- La consecuencia de alta confiabilidad son **ordenamientos** poblacionales "limpios"



Confiabilidad: Es general y puntual



Puedo distinguir con claridad los extremos -No necesito **muy alta** confiabilidad-

Pero qué tal y si quiero hacer distinciones en este punto dada una línea de pobreza: Scores observados distintos, niveles similares de pobreza

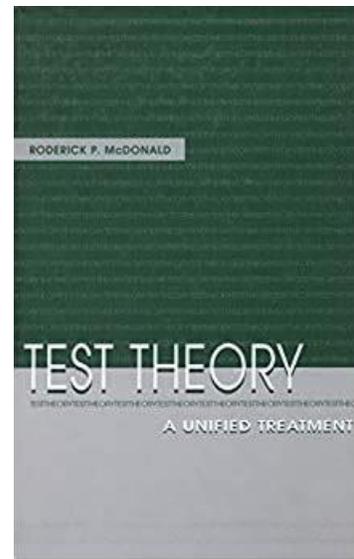
Medidas de pobreza con baja confiabilidad hacen sentido en los extremos (Ruralidad, Indígena, Alto y bajo nivel educativo)

Estimadores de confiabilidad

- En los 1920s: Spearman-Brown -Split-half reliability-
- En los 1950s: Gutman y Cronbach - Lambdas y Alfa-
- En los 1970s: William Revelle -Beta-
- En el 2000: McDonald -Omega y Omega h-. *(Se ha empezado a usar en los últimos 5 años).*

Estimadores clásicos - exploratorios-

Estimador a partir de variables latentes y ecuaciones estructurales



¿Cuál usar? ¿Qué diferencias tienen?

El mejor estimador es aquel que mejor se ajusta al **modelo de medición** en cuestión:

- El mejor estimador es **omega** por su potencia y flexibilidad
- **Omega h** es útil para medidas multidimensionales
- **Alfa** es útil bajo modelos de medición muy estrictos
- **Beta** es una versión pesimista de la confiabilidad -heterogeneidad de la escala-

Theory and Methods | [Published: 11 December 2008](#)

Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma

[William Revelle](#) & [Richard E. Zinbarg](#)

Psychometrika **74**, Article number: 145 (2009) | [Cite this article](#)

4757 Accesses | 708 Citations | 4 Altmetric | [Metrics](#)

Estimating Generalizability to a Latent Variable Common to All of a Scale's Indicators: A Comparison of Estimators for ω_h

[Richard E. Zinbarg](#), [Iftah Yovel](#), [William Revelle](#), , , more...

[Show all authors](#)

First Published March 1, 2006 | Research Article

<https://doi.org/10.1177/0146621605278814>

[Article information](#)



Abstract

The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality

AUTHORS
[Gjalt-Jorn Peters](#)

Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates

[David B. Flora](#)

First Published November 6, 2020 | Research Article

Check for updates

<https://doi.org/10.1177/2515245920951747>

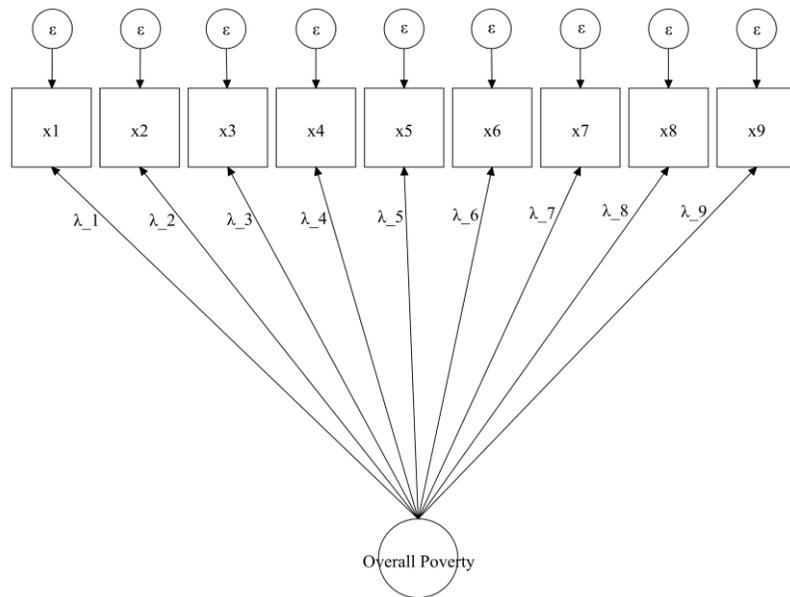
[Article information](#)



Modelo de medición y confiabilidad

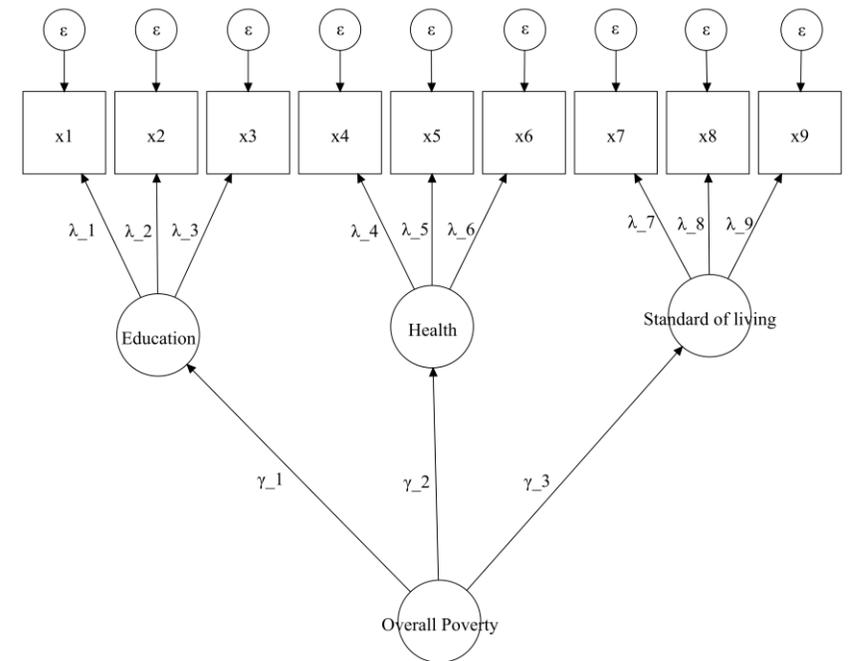
Alfa, beta y omega darán el mismo resultado si λ 's, σ , μ son iguales:

Tests-paralelos y equivalencia Tau



A medida que nos desviamos de ese modelo ideal, omega es el mejor estimador

En ciencias sociales requerimos modelos más flexibles



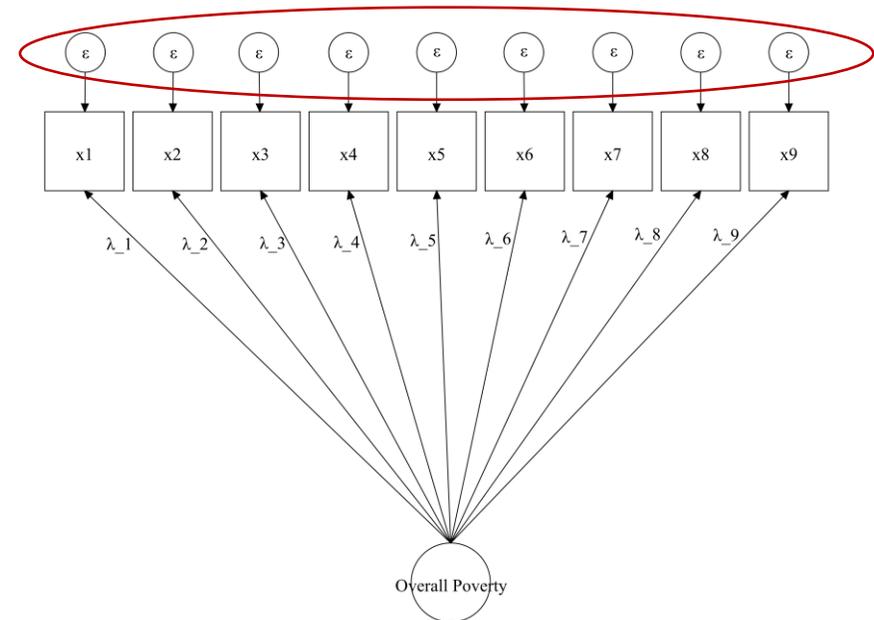
Omega y omega_h

Estimación de confiabilidad

Condiciona en el modelo de medición, podemos estimar el cociente de varianza de interés respecto a la varianza total

% de ruido

% de la varianza explicada por el fenómeno de interés



Por tanto, todos los estimadores de confiabilidad van de cero a uno. Donde ≈ 1 implica mayor confiabilidad de los scores

¿Qué tan alto debe ser el valor de confiabilidad?

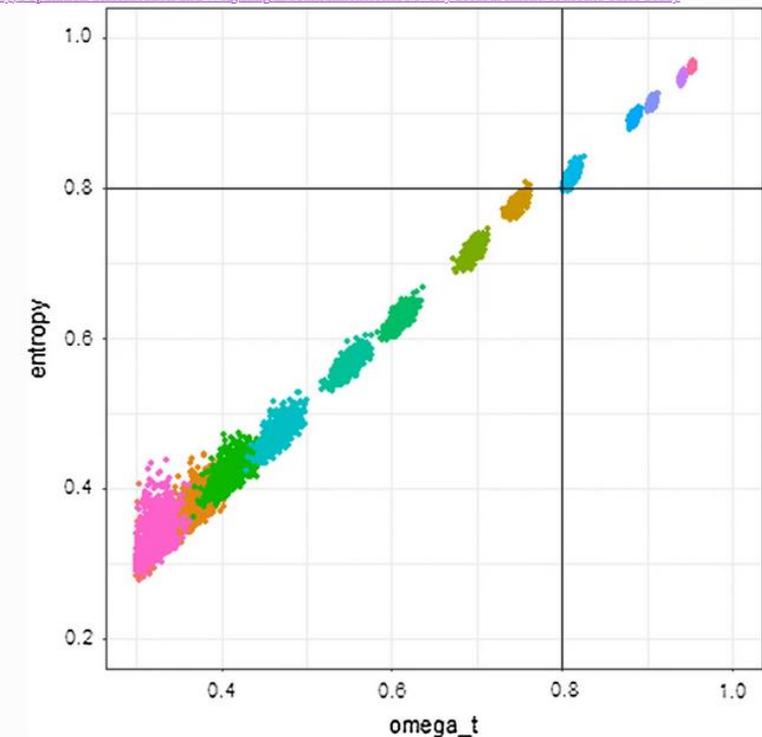
- Depende para qué. Clasificaciones muy crudas –alto y bajo–, no se requieren valores altos
- Clasificaciones precisas requieren valores altos: $\omega > .8$
- Conforme decrece la confiabilidad –aumenta el error de medición– se empeora nuestra capacidad de distinguir personas:
 - Mismo score puede significar diferente pobreza latente
 - Distintos scores pueden significar niveles similares de pobreza latente
- La clasificación pobre y no pobre tendrá mayores tasas de falsos positivos y negativos –**Error de clasificación**–

Reliability, Population Classification and Weighting in Multidimensional Poverty Measurement: A Monte Carlo Study

[Héctor E. Nájera Catalán](#) 

Social Indicators Research **142**, 887–910 (2019) | [Cite this article](#)

From: [Reliability, Population Classification and Weighting in Multidimensional Poverty Measurement: A Monte Carlo Study](#)

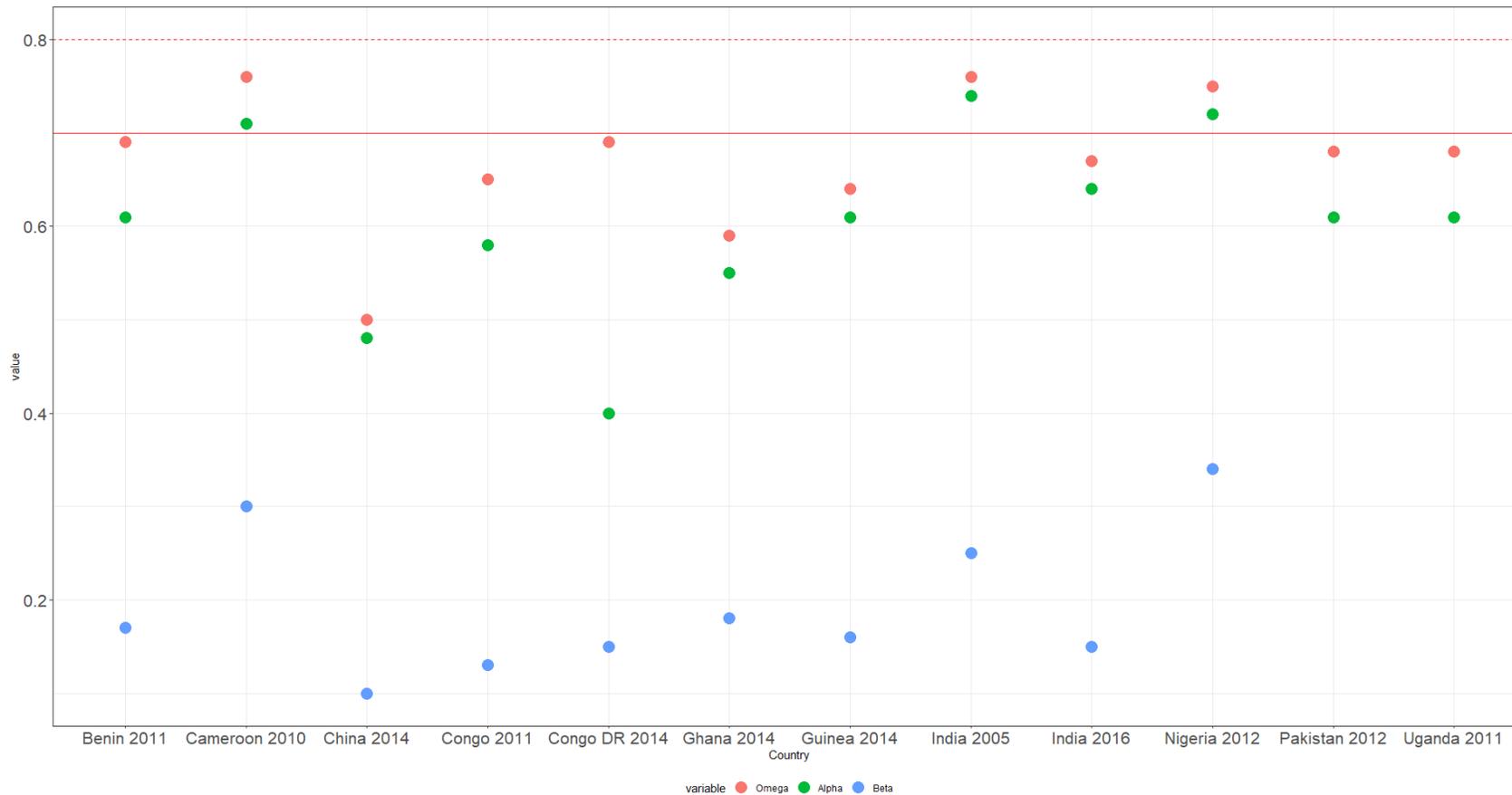


Unidimensional models ω and E

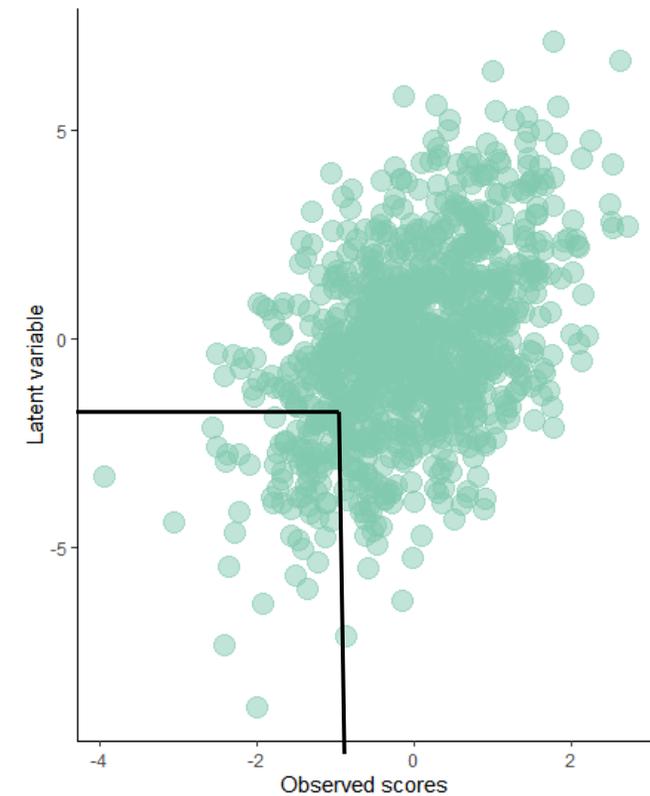
¿Qué sabemos de la
confiabilidad global de
algunas medidas?

OPHI-MPI: Modelo flexible. Confiabilidad máxima.

$\omega < .8 \approx > 10\%$
error de clasificación



Esto pasa en el fondo



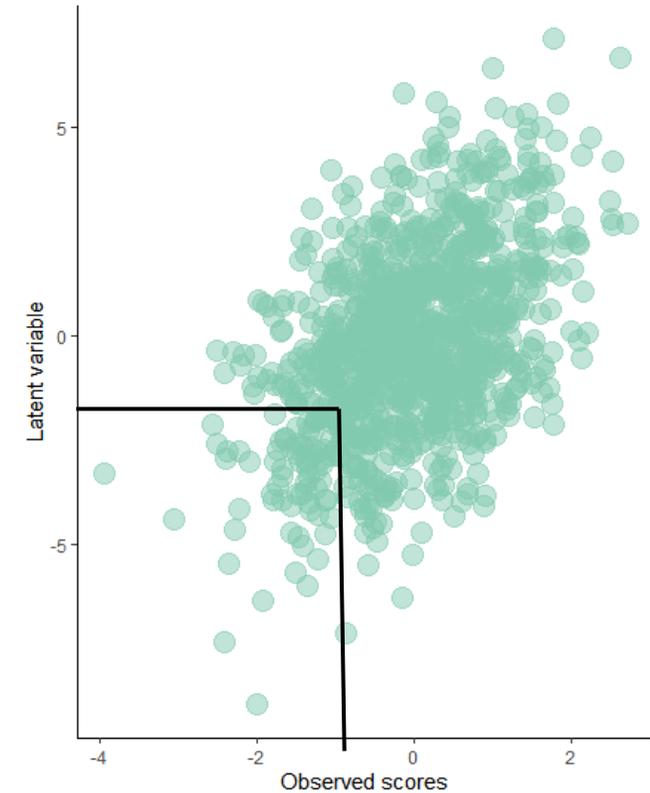
Indice de activos MPI

$\alpha < .6 \approx > 20\%$
error de clasificación

Volmer y Alkire (2019)

Esto pasa en el fondo

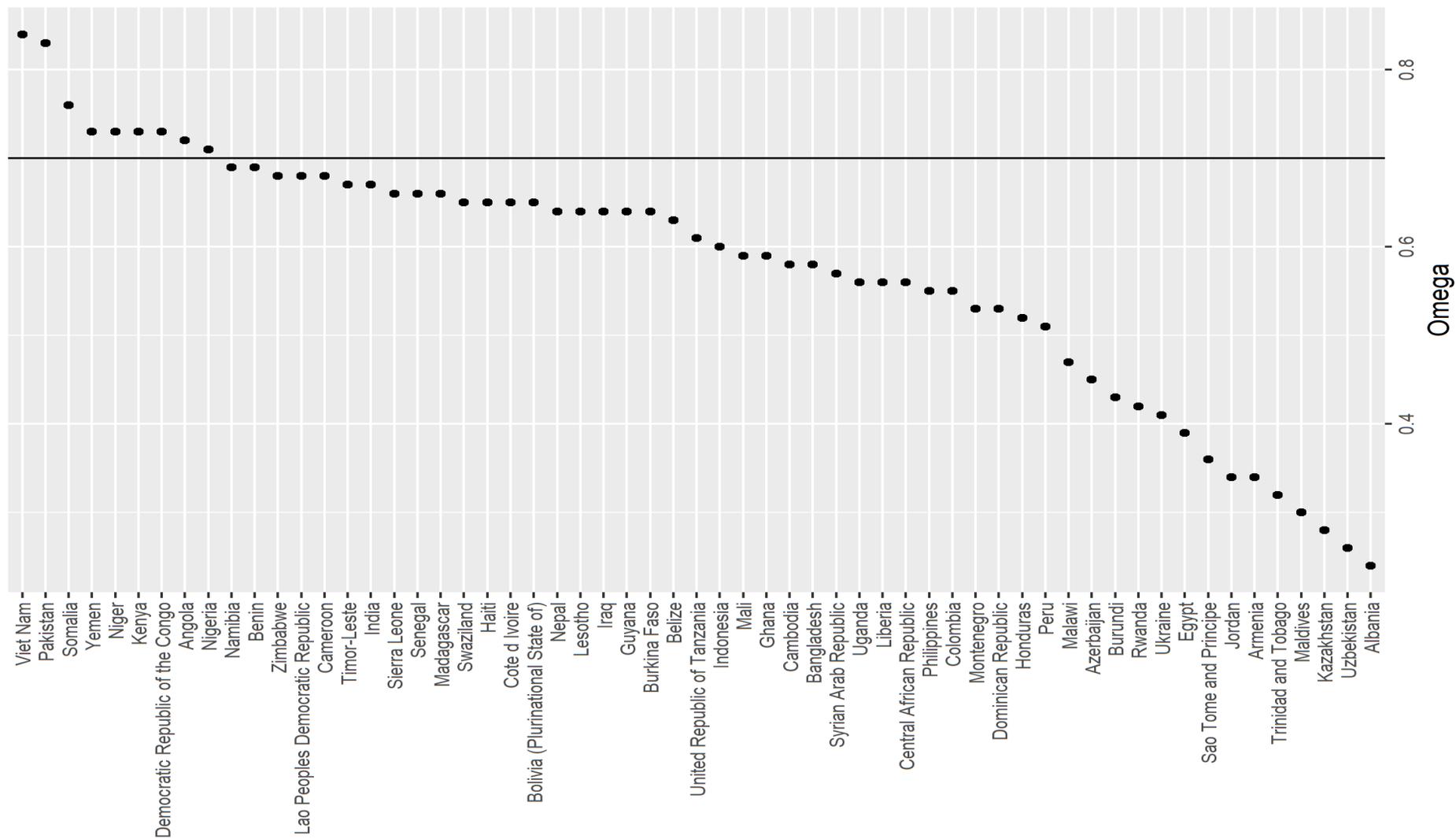
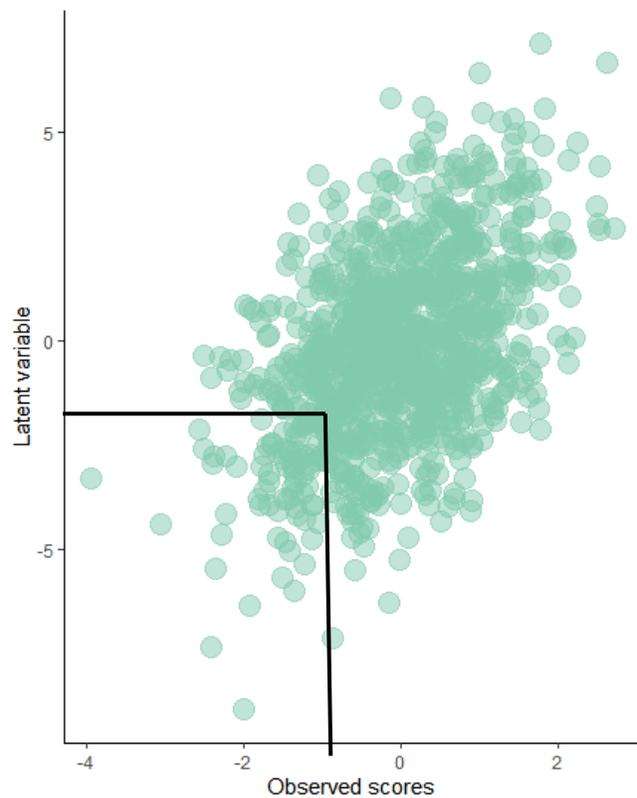
	MPI-O	MPI-I	MPI-I, alt 1 (3ha)	MPI-I, alt 1 (0.3ha)	MPI-I, alt 2	MPI-I, alt 3 (3ha)	MPI-I, alt 3 (0.3ha)	MPI-I, alt 4 (3ha)	MPI-I, alt 4 (0.3ha)	MPI-N, version 1	MPI-N, version 2	MPI-N, version 3	MPI-N, version 3 minus bicycle and animal cart
Pooled	0.583	0.4776	0.5360	0.52	0.4866	0.4970	0.4897	0.5146	0.4969	0.742	0.7034	0.6129	0.6779
Armenia	0.2356	0.2973	0.2074	0.2469	0.286	0.2233	0.2492	0.2071	0.2463	0.513	0.4982	0.3087	0.3172
Angola	0.6896	0.4964	0.5651	0.5364	0.5066	0.5319	0.5084	0.5605	0.5354	0.7627	0.7365	0.6972	0.7531
Bangladesh	0.4523	0.4667	0.4103	0.4407	0.4785	0.4103	0.4407	0.4103	0.4407	0.5727	0.5333	0.5155	0.54
Brazil	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.3685	0.5753	0.5753	0.4577	0.4577
DR Congo	0.6256	0.4759	0.518	0.4796	0.4671	0.5078	0.471	0.5152	0.48	0.6982	0.6372	0.638	0.7105
Côte d'Ivoire	0.511	0.4444	0.4932	0.4658	0.4643	0.4919	0.4644	0.4906	0.4627	0.6346	0.5586	0.5273	0.6195
Colombia	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.5625	0.6703	0.6781	0.6238	0.6073
Egypt	0.2954	0.3795	0.305	0.305	0.383	0.383	0.383	0.305	0.305	0.382	0.3601	0.3471	0.3982
Ethiopia	0.6398	0.4028	0.4883	0.4303	0.4175	0.487	0.4292	0.4876	0.43	0.6636	0.6753	0.6651	0.6814
Guatemala	0.6611	0.5167	0.547	0.5366	0.5369	0.5473	0.5367	0.5467	0.5363	0.7434	0.7291	0.6659	0.7126
Haiti	0.6338	0.4333	0.4898	0.4995	0.4432	0.4901	0.4997	0.4898	0.4995	0.691	0.6767	0.6302	0.6829
India	0.5534	0.4905	0.5255	0.515	0.4905	0.4848	0.4944	0.5195	0.5104	0.7251	0.6757	0.5795	0.6567
Indonesia	0.6811	0.4935	0.5282	0.4896	0.4868	0.5244	0.4889	0.527	0.488	0.6829	0.6821	0.6306	0.6702
Kenya	0.5833	0.5035	0.5046	0.5106	0.5019	0.5028	0.5106	0.5038	0.5089	0.5158	0.5207	0.5538	0.5755
Cambodia	0.5675	0.4429	0.4612	0.4573	0.4496	0.4611	0.4573	0.4611	0.4573	0.5739	0.5464	0.5039	0.5802



UNICEF 2004

$\omega < .7 \approx > 15\%$
error de clasificación

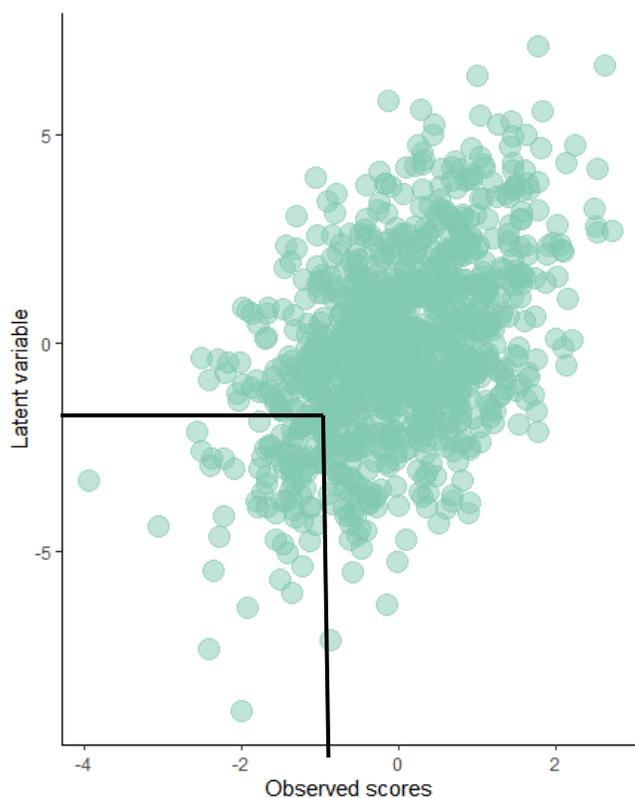
Esto pasa en el fondo



MPI-LA

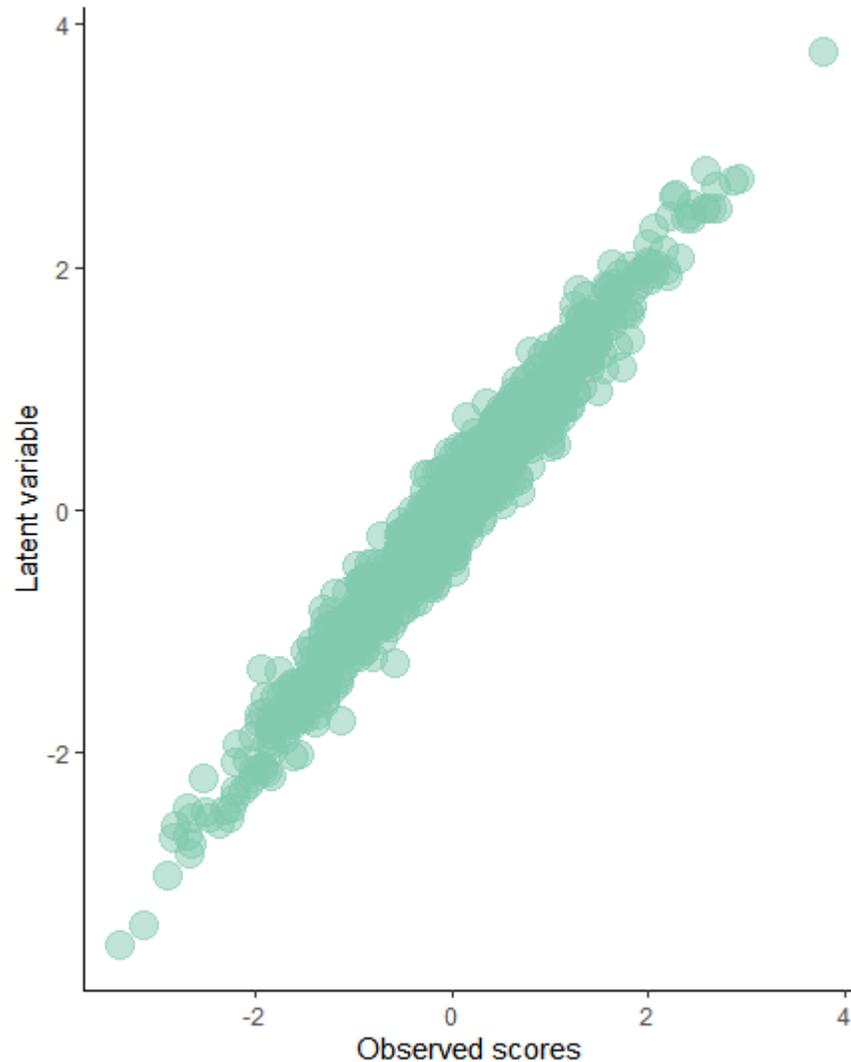
$\omega < .6 \approx > 20\%$
error de clasificación

Esto pasa en el fondo



Country	Year	α	β	ω
Argentina	2005	0.63	0.38	0.71
	2012	0.51	0.32	0.60
Bolivia	2003	0.64	0.56	0.68
	2012	0.65	0.34	0.76
Brazil	2005	0.52	0.04	0.62
	2012	0.45	0.00	0.57
Chile	2003	0.46	0.35	0.58
	2011	0.27	0.11	0.33
Mexico	2004	0.75	0.07	0.81
	2012	0.64	0.24	0.69
Uruguay	2005	0.54	0.02	0.67
	2012	0.43	0.38	0.54

Pobreza relativa: Europa. Método consensual



$\omega > .9 \approx < 5\%$
error de clasificación

Table 15: Omega Analysis, national level, 2014

	Unidimensional		Townsend		
	Omega	BIC	Omega	Omega_h	BIC
Belgium	0.97	83593	0.95	0.65	86226
Bulgaria	0.94	155878	0.93	0.63	160109
Czech Republic	0.95	125497	0.95	0.64	127748
Denmark	0.94	73381	0.94	0.64	74320
Germany	0.94	188040	0.93	0.64	192065
Estonia	0.93	111990	0.93	0.65	113873
Ireland	0.94	124365	0.94	0.65	126486
Greece	0.92	226947	0.93	0.63	230458
Spain	0.96	245426	0.94	0.65	250931
France	0.95	182346	0.94	0.64	186617
Croatia	0.94	121422	0.94	0.64	123073
Italy	0.96	385840	0.94	0.65	397204
Cyprus	0.93	107605	0.93	0.64	109133
Latvia	0.92	157885	0.92	0.63	158775
Lithuania	0.92	123433	0.92	0.63	125872
Luxembourg	0.97	39261	0.95	0.65	40550
Hungary	0.95	242356	0.93	0.64	247061
Malta	0.94	109760	0.92	0.64	111200
Netherlands	0.95	127708	0.95	0.64	130432
Austria	0.96	70340	0.94	0.64	71973
Poland	0.94	310111	0.93	0.64	315812
Portugal	0.94	159328	0.94	0.65	162471
Slovenia	0.94	207387	0.94	0.64	211409
Slovakia	0.93	139511	0.93	0.64	141971
Finland	0.95	126804	0.94	0.64	128633
Sweden	0.93	53870	0.93	0.64	54715

Source: EU-SILC 2014 cross-sectional data, authors' computation.

México

La calidad de la medición multidimensional en México ha venido a menos

Esto se contempló en los lineamientos de la medición oficial

Ya se hizo un primer ajuste en 2018 pero no sabemos si fue suficiente

Cuadro 1. Estimación del error de clasificación del índice de derechos sociales. México 2008-2018

Año	Omega	
2008	0.75	
2010	0.72	
2012	0.71	
2014	0.72	
2016	0.68	
2018	0.68	

México: Error de clasificación

Gracias a desarrollos teóricos y adelantos computacionales hoy podemos estimar el error de clasificación directamente con distintos métodos

Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left

Jerry Hausman

JOURNAL OF ECONOMIC PERSPECTIVES
VOL. 15, NO. 4, FALL 2001
(pp. 57-67)



Journal of Econometrics
Volume 87, Issue 2, December 1998, Pages 239-269



Misclassification of the dependent variable in a discrete-response setting

J.A. Hausman a , Jason Abrevaya b, F.M. Scott-Morton b

Cuadro 1. Estimación del error de clasificación del índice de derechos sociales. México 2008-2018

Año	Omega	Error FN [ICr 95%]	Error FP [ICr 95%]
2008	0.75	6 [2-10]	1 [0-2]
2010	0.72	8 [5-11]	1 [0-2]
2012	0.71	11 [8-14]	1 [0-3]
2014	0.72	9 [6-12]	1 [0-2]
2016	0.68	13 [11-16]	1 [0-2]
2018	0.68	14 [11-17]	0 [0-1]

Misclassification error, binary regression bias and reliability in multidimensional poverty measurement: An estimation approach based on Bayesian modelling.

Héctor Nájera (Manuscript accepted for publication. Measurement Journal)

Confiabilidad de ítems: Teoría de respuesta al ítem

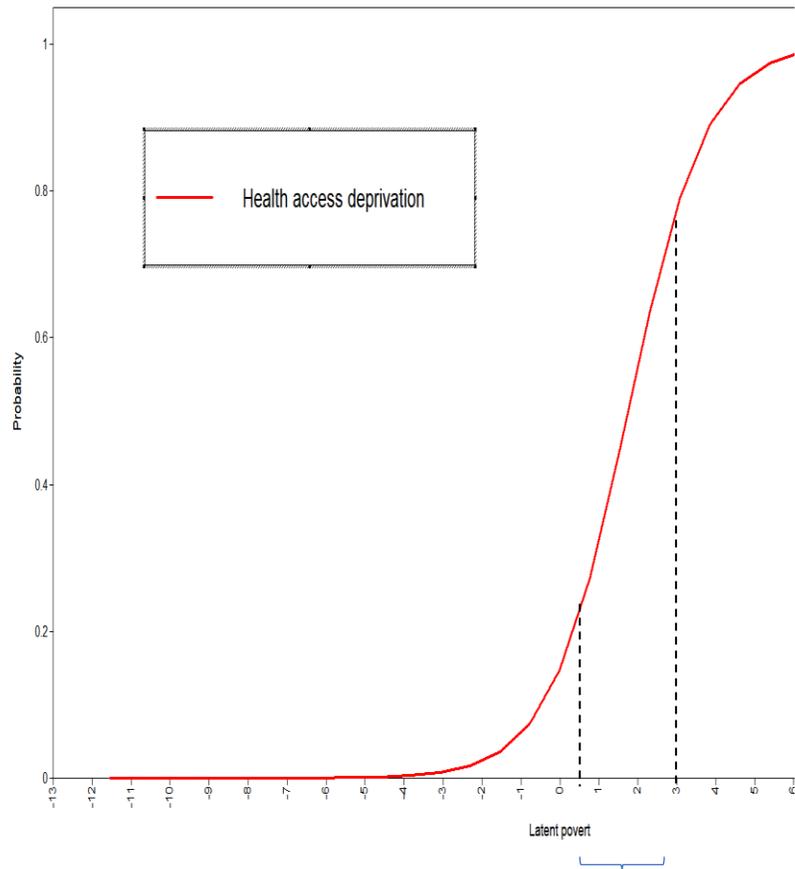
El análisis de confiabilidad puede ayudarnos a identificar ítems problemáticos:

- Aquellos cuya variabilidad no proviene del fenómeno de interés (**Discriminación**)
- aquellos cuya variabilidad es tan baja, que no aportan información (**Severidad**)

Modela la probabilidad de tener privación condicional en: El nivel latente de pobreza, la discriminación del ítem y su severidad.

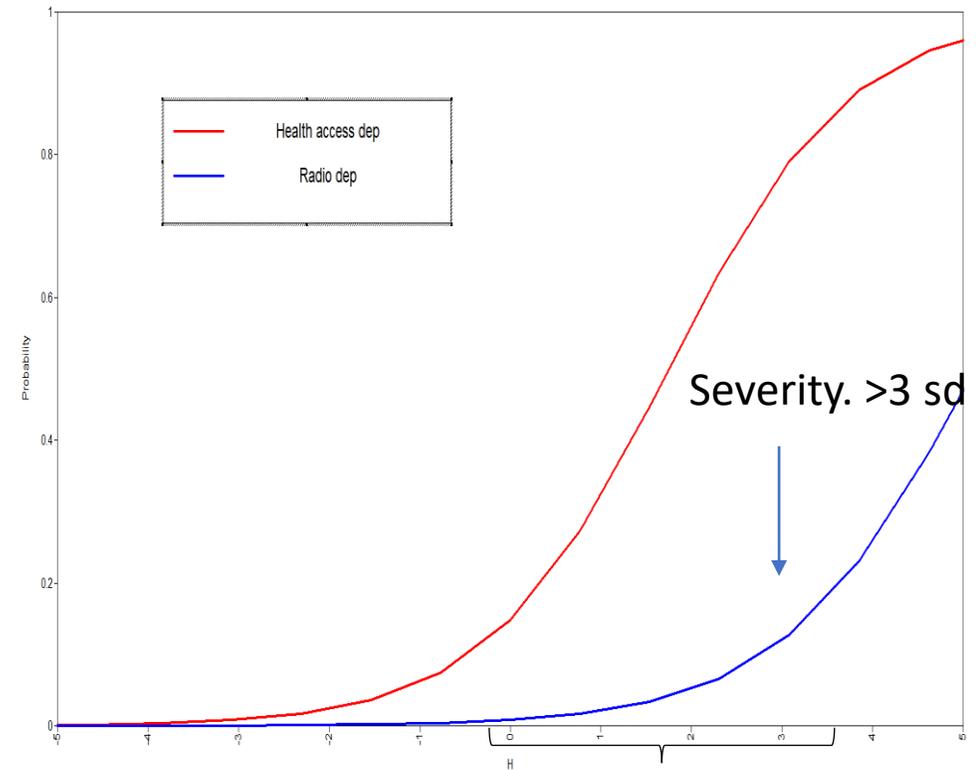
Teoría de respuesta al ítem

Carencia por acceso efectivo a servicios de salud



Changes in latent poverty lead to changes in probability

Carencia por falta de radio



Changes in latent poverty don't lead to important changes in probability

Examples

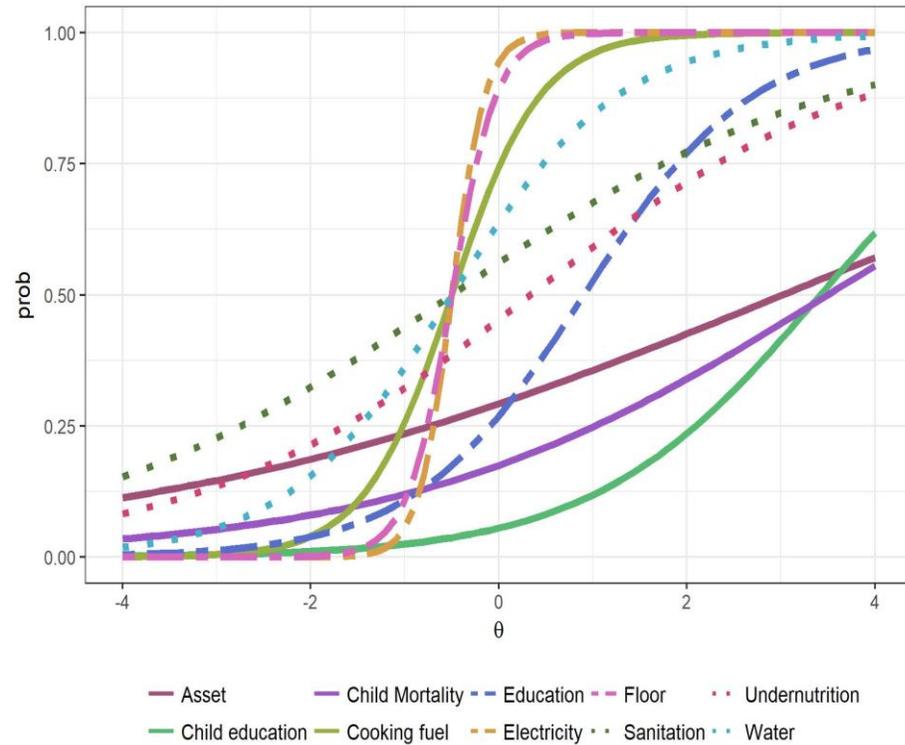
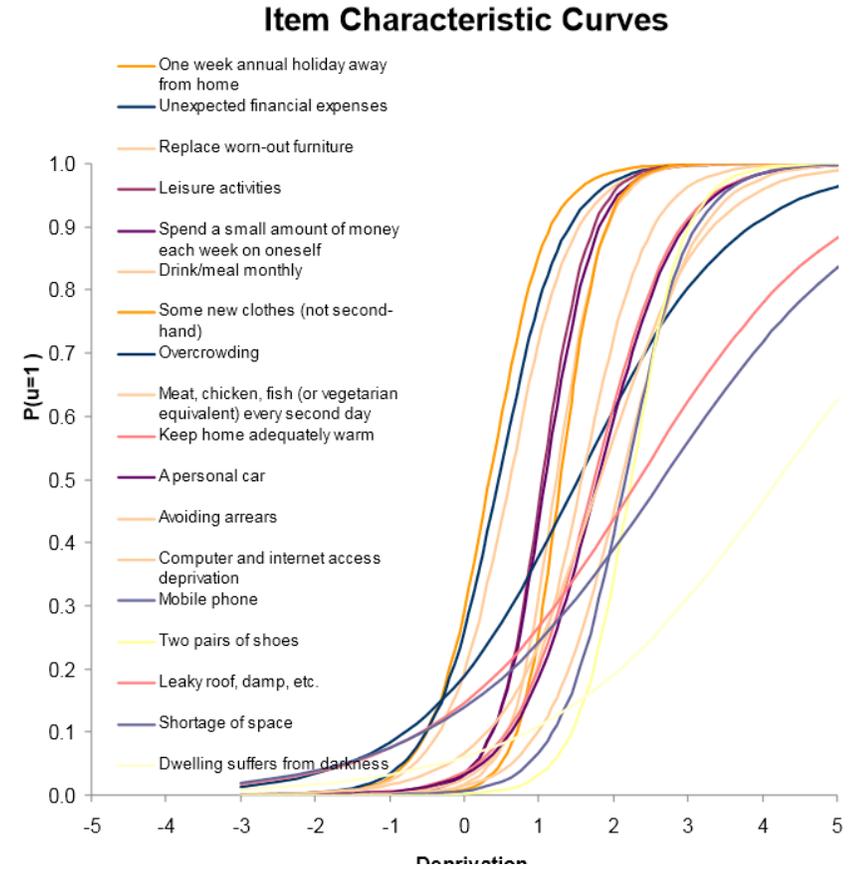
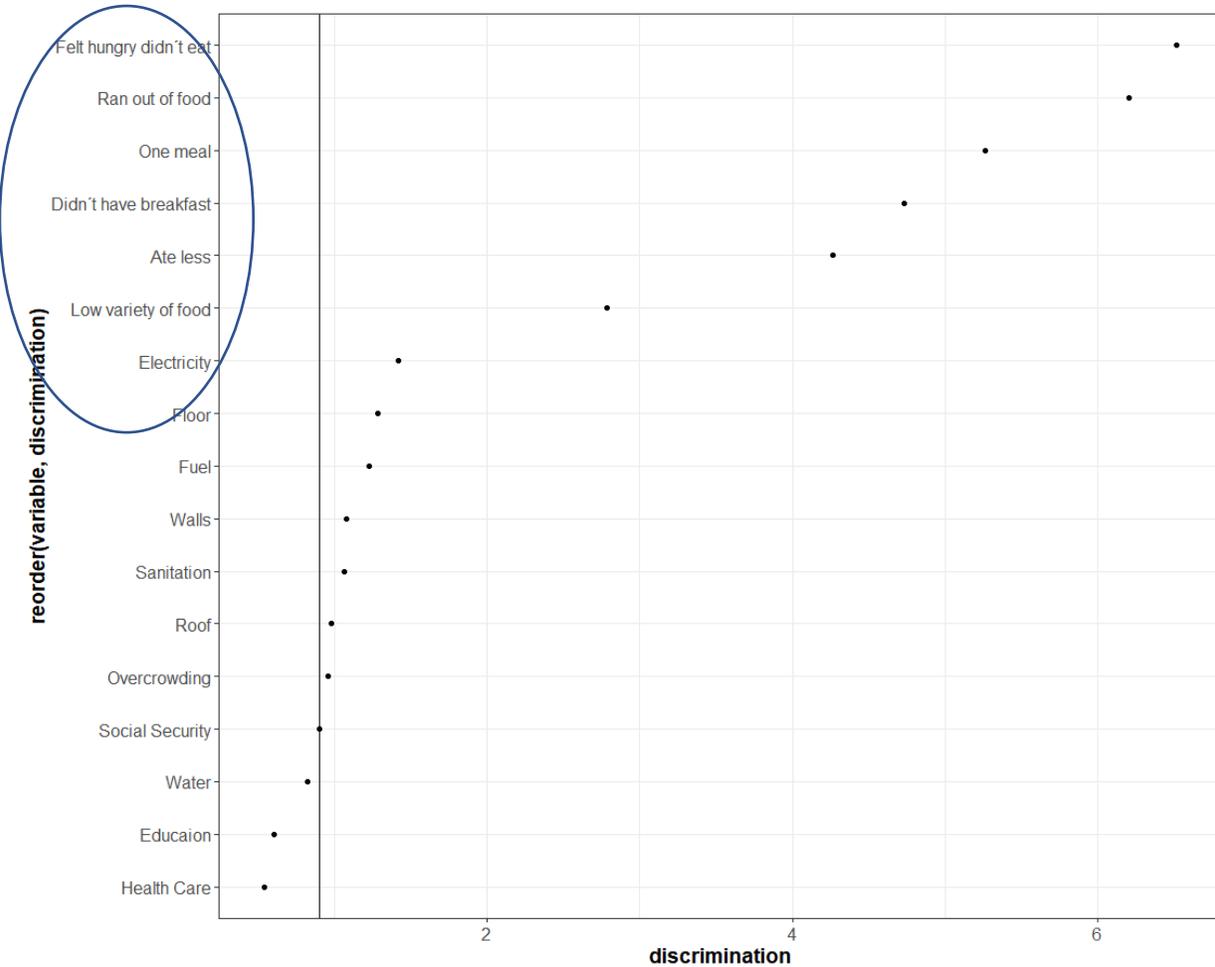


Figure 8: IRT results for the 18 items retained for the whole population MD indicator after the validity and reliability (Cronbach's alpha) tests, 2009

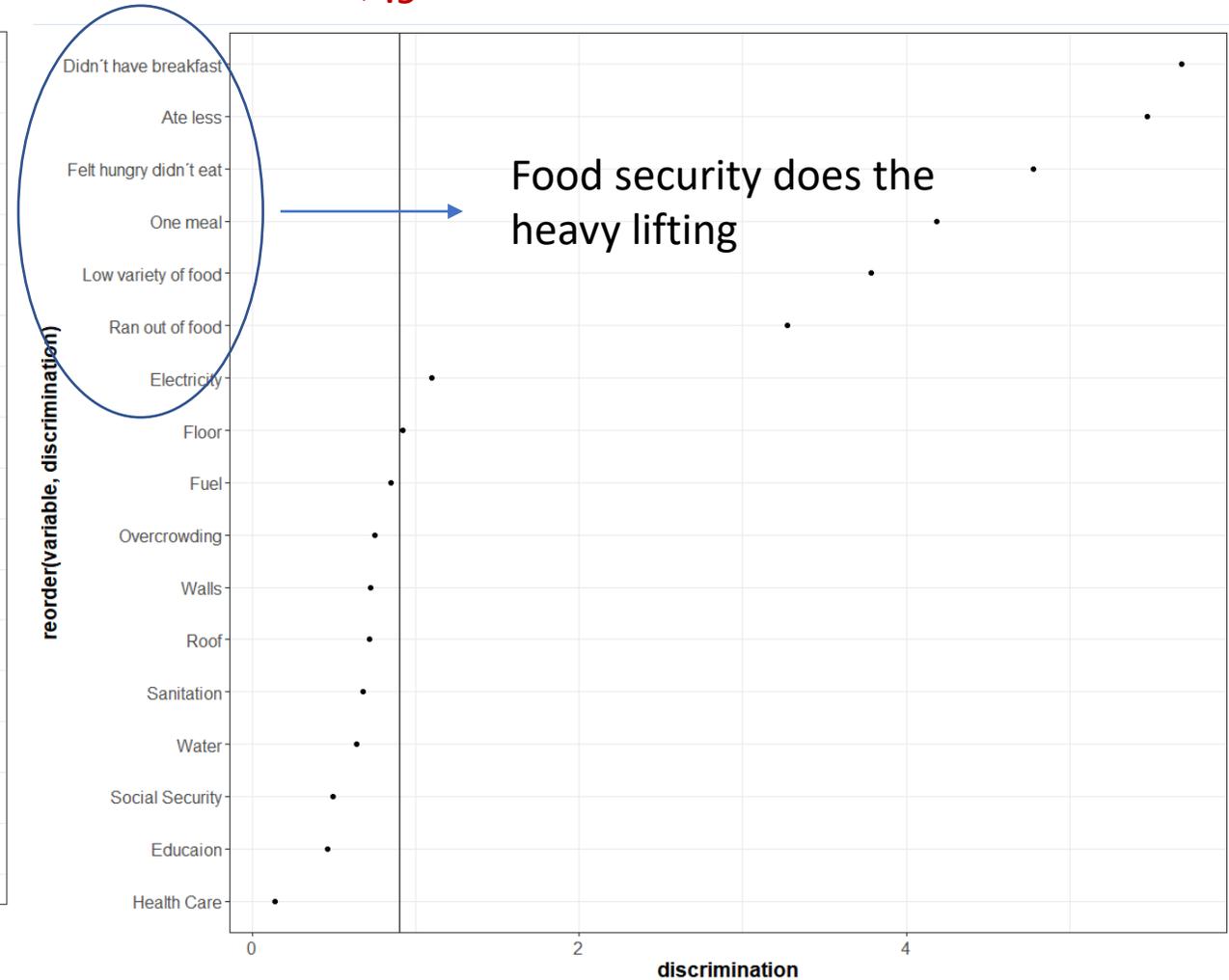


Results Model 1: 2P-IRT. Discrimination

>.9
2008



>.9
2018



What are you weighting for?



En contexto de alta confiabilidad los pesos hacen muy poco:
Los scores observados correlacionan altamente con los latentes

Siempre los pesos diferenciales hacen más daño

Los pesos tal vez ayudan en casos de baja confiabilidad -pero cómo saberlo-

Los pesos pueden tener otros objetivos y no sólo la clasificación

What are you weighting for? Misconceptions

Endogenous weights and multidimensional poverty: A cautionary tale ☆

Indranil Dutta ^a, Ricardo Nogales ^b, Gaston Yalonetzky ^c

Show more ▾

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.jdeveco.2021.102649>

[Get rights and content](#)

Highlights

- A large and growing body of work uses endogenous (data driven) weights to compute multidimensional poverty.



Articles

The Importance of Reliability and Construct Validity in Multidimensional Poverty Measurement: An Illustration Using the Multidimensional Poverty Index for Latin America (MPI-LA)

Héctor E. Nájera Catalán & David Gordon

Pages 1763-1783 | Received 07 Feb 2018, Accepted 01 Mar 2019, Published online: 26 Sep 2019

EL ejemplo que usan es particular y con baja confiabilidad:

Los pesos endógenos -como cualquier otro tipo de peso- son dañinos en contexto de baja confiabilidad

En contextos de alta confiabilidad los pesos "endógenos" son óptimos y los pesos iguales son mejores que los "ad hoc".

Es falso que una parte importante y creciente use pesos diferenciales, mucho menos en contextos de alta confiabilidad

Los scores de privación del método consensual son simples.



Conclusiones

- Confiabilidad es una propiedad de los scores y es **condición necesaria** para cualquier ejercicio de medición
- Confiabilidad es una **condición necesaria** para la *clasificación* de poblaciones
- Confiabilidad **no es suficiente** ¿Por qué?
 - No hemos dicho nada sobre la variable latente
 - Un conjunto de variables puede tener alta confiabilidad y decirnos nada sobre la pobreza
- Es la validez lo que permite hablar de la representación del fenómenos de interés.