

Modelos Rasch: ¿cuán (in-)coherentemente son presentados y utilizados?*

Rasch Models: How (In-)Coherently are they Showed and Used?

Ernesto San Martín¹

Pontificia Universidad Católica de Chile, Chile

Resumen. El modelo Rasch es ampliamente usado para el análisis de datos educacionales. En la práctica, se reportan los estimadores de la dificultad de los ítems y los estimadores de las habilidades de los individuos. Sin embargo, nunca se explicita qué significado tiene el término “dificultad” y cuál el término “habilidad”. Los significados de estos términos no dependen de las estimaciones; al contrario, estas últimas han de interpretarse en función del significado de los primeros. En este trabajo se muestra que el significado de las nociones de “dificultad” y “habilidad” dependen de la forma en que el modelo Rasch se especifica. En la literatura psicométrica hay dos maneras de especificar dicho modelo: una que sólo se limita a modelar los observables; la otra, que además de los observables, modela no-observables. La primera forma de especificación se debe al mismo Rasch, mientras que la segunda fue desarrollada por Lord.

Palabras clave. Axioma de independencia local, rasgo latente, dificultad de un ítem, proporción de chances, probabilidad inversa.

Abstract. Rasch models are widely used for the analysis of educational data. In practice, estimates of difficulties of items and abilities of examinees are reported. However, the meaning of the terms “difficulty” and “abilities” are never made explicit. The meaning of these terms does not depend on the estimations; they should be interpreted with respect to the eventual meaning of both item difficulties and individual abilities. This paper shows that the meaning of the terms “difficulty” and “ability” depends on the way in which the Rasch model is specified. In the psychometric literature, Rasch models are specified in two different ways: one specifies the observable only, whereas the other one specifies both observable and unobservable. The first specification is due to Rasch himself, the second one is due to Lord.

Keywords. Local independence axiom, latent trait, item difficulty, odds ratio, inverse probability.

¹ Ernesto San Martín. Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Chile; Center for Operations Research and Econometrics, Belgium. Dirección Postal: Vicuña Mackenna 4860, Macul, Santiago, Chile. Email: esanmart@mat.puc.cl

*Agradecimientos: Este trabajo ha sido parcialmente financiado por el Proyecto ANILLO SOC 1107 Statistics for Public Policy in Education, de la Corporación Nacional de Ciencia y Tecnología del Gobierno de Chile.



Problema: a modo de introducción

Supongamos que 10 estudiantes respondieron un test compuesto de seis ítems de selección múltiple. Una vez revisado el test, los resultados se resumen en la tabla 1.

Datos como estos se obtienen a partir de aplicaciones masivas, como el SIMCE en Chile, las pruebas SABER en Colombia, o PISA por la OECD. Los análisis psicométricos que se hacen se reducen a dos tipos: por un lado se estima un puntaje individual, el cual típicamente se interpreta como el grado de dominio que un determinado estudiante tiene en relación a lo que mide el test; de hecho, a veces se relaciona dicho puntaje con una supuesta habilidad medida por el test, en cuyo caso dicha habilidad corresponde a un rasgo latente. Por otro lado, se estiman propiedades de los ítems, concretamente la dificultad, el grado de discriminación y el grado de respuesta al azar. Las propiedades de los ítems están basadas en modelos de Teoría de Respuesta al Ítem (TRI), aunque también suelen reportarse indicadores de dificultad y de discriminación (i.e. correlación ítem-test) basados en la Teoría Clásica de Test (TCT).

La práctica resumida en el párrafo anterior plantea un problema que utiliza una terminología estándar (a saber, habilidad de un individuo y dificultad de un ítem) independientemente del tipo de modelo psicométrico que se utilice para analizar dichos datos. Puede argüirse

Tabla 1

Matriz de respuestas individuales a un conjunto de ítems

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Estudiante 1	0	0	1	1	0	1
Estudiante 2	1	0	0	1	1	1
Estudiante 3	1	1	0	0	0	0
Estudiante 4	1	1	0	1	0	0
Estudiante 5	0	0	1	1	0	0
Estudiante 6	1	0	1	0	0	0
Estudiante 7	1	1	1	1	0	1
Estudiante 8	1	1	0	0	0	1
Estudiante 9	1	0	0	0	1	1
Estudiante 10	0	0	0	1	0	1

que, por ejemplo, las estimaciones de las dificultades de los ítems bajo el esquema de la TCT está casi perfectamente correlacionado con las estimaciones de las dificultades que se obtienen al aplicar un modelo Rasch o un modelo 2PL. Pero, como ya es ampliamente reconocido en la investigación cuantitativa, correlación no implica explicación, por lo que no resulta coherente suponer que a una misma terminología subyace una misma significación.

El objetivo de este artículo es mostrar cómo el significado de dificultad de un ítem y habilidad de un individuo dependen de la especificación del modelo psicométrico utilizado. Concretamente, nos focalizaremos en dos maneras de especificar el modelo Rasch, las cuales se aclararán en los párrafos siguientes.

Primera Parte

Heterogeneidad de la información observada

Para definir rigurosamente habilidad individual y dificultad de un ítem, es necesario considerar dos hechos:

- Cuando se observan las respuestas a un ítem dadas por una determinada población de individuos, se constata que dichas respuestas son heterogéneas.
- Cuando se observan las respuestas que un estudiante da a un conjunto de ítems, también se observa un comportamiento heterogéneo.

La heterogeneidad puede ser representada formalmente por medio de probabilidades. Para ser más específicos, definamos una variable aleatoria Y_{pi} que toma el valor 1 cuando la persona p responde correctamente el ítem i ; y toma el valor 0 en caso contrario. La heterogeneidad de la variable aleatoria Y_{pi} se representa por medio de la probabilidad que dicha variable tiene de ser igual a 1 ó a 0: más alta (respectivamente, más baja) es esta probabilidad, más chances (respectivamente, menos chances) hay que la persona p responda correctamente el ítem i . Si denotamos por π dicha probabilidad, entonces escribimos $P[Y_{pi} = 1] = \pi$, y consecuentemente $P[Y_{pi} = 0] = 1 - \pi$. De forma compacta, decimos que Y_{pi} es una variable aleatoria Bernoulli de parámetro π , y lo denotamos por $Y_{pi} \sim \text{Bern}(\pi)$.

Bajo estas consideraciones, la tabla de datos inicial corresponde a una posible realización de 60 distribuciones Bernoulli de parámetro π , todas mutuamente independientes entre sí –es decir, que la realización de cualquiera de ellas no proporciona información alguna acerca de la realización de cualquiera de las restantes.

Pero de inmediato surgen dos preguntas: para un determinado ítem, ¿por qué asumir que la probabilidad de responder correctamente es la misma para cada persona? Para una persona determinada, ¿por qué asumir que la probabilidad de responder correctamente cada ítem es la misma? Una forma de representar la heterogeneidad que se observa tanto a nivel de ítems, como de personas, es precisamente asumir que la probabilidad de responder correctamente un ítem por una determinada persona depende tanto del ítem como de la persona. Más precisamente, se supone que el proceso generador de datos que subyace a la tabla inicial de observaciones está caracterizado por dos condiciones:

- a. Las variables aleatorias Y_{pi} 's son mutuamente independientes.
- b. Para cada persona p e ítem i , $Y_{pi} \sim \text{Bern}(\pi_{pi})$, donde el parámetro $\pi_{pi} \in (0,1)$. Enfatizamos que π_{pi} es distinto de 0 y de 1: si π_{pi} fuera igual a 0 ó a 1, entonces no habría heterogeneidad alguna.

Modelo Rasch

Separación entre la característica individual y la característica de un ítem. Como ya ha sido mencionado anteriormente, la especificación anterior implica que $P(Y_{pi} = 1) = \pi_{pi}$ y $P(Y_{pi} = 0) = 1 - \pi_{pi}$. Ambas igualdades pueden resumirse por medio del siguiente cociente de chances

$$\frac{P[Y_{pi} = 1]}{P[Y_{pi} = 0]} = \lambda_{pi}$$

donde

$$\lambda_{pi} = \frac{\pi_{pi}}{1 - \pi_{pi}}$$

El parámetro λ_{pi} (que corresponde a un número real) representa la razón entre la probabilidad que la persona p responda correctamente el ítem i , y la probabilidad de que lo responda incorrectamente. Así, por ejemplo, si $\lambda_{pi} = 3$, entonces la probabilidad que la persona p responda correctamente el ítem i triplica la probabilidad de que lo responda incorrectamente. De manera similar, si $\lambda_{pi} < 1$, entonces la probabilidad de que la persona p responda correctamente el ítem i es menor que la probabilidad de que lo responda incorrectamente.

El modelo Rasch introduce una hipótesis fundamental, que consiste en “separar” la característica individual de la característica del ítem; ver Rasch (1960). Específicamente, supone que el parámetro λ_{pi} se descompone de la siguiente manera:

$$\lambda_{pi} = \frac{\varepsilon_p}{\eta_i} \quad (1)$$

donde ε_p representa una característica de la persona p , y η_i representa una característica del ítem i . Para entrever a qué tipo de características nos referimos, notemos lo siguiente:

- a. Supongamos que η_i está fijo. Si ε_p crece (respectivamente, decrece) indefinidamente, entonces en algún momento $\lambda_{pi} > 1$ (respectivamente, $\lambda_{pi} < 1$) y

por lo tanto la probabilidad que la persona p responda correctamente el ítem i será mayor (respectivamente, menor) que la probabilidad de responderlo incorrectamente. En términos meramente intuitivos, podemos afirmar que ε_p representa la “habilidad de la persona p ”. No hay que olvidar que esta afirmación es relativa al valor fijo de η_i .

b. Supongamos que ε_p está fijo. Si η_i crece (respectivamente, decrece) indefinidamente, entonces en algún momento $\lambda_{pi} < 1$ (respectivamente, $\lambda_{pi} > 1$) y por lo tanto la probabilidad que la persona p responda correctamente el ítem i será menor (respectivamente, mayor) que la probabilidad de que lo responda incorrectamente. En términos meramente intuitivos, podemos afirmar que η_i representa la “dificultad del ítem i ”. No hay que olvidar que esta afirmación es relativa al valor fijo de ε_p .

Las afirmaciones anteriores se basan en una comparación entre ε_p y η_i . Esto requiere que los valores que se asignan a los parámetros ε_p y η_i estén únicamente determinados. Sin embargo, la descomposición (1) sufre de una indeterminación. En efecto, supongamos que $\lambda_{pi} = 3$. Como ya hemos indicado, esto significa que la probabilidad de que la persona p responda correctamente el ítem i triplica la probabilidad de responderlo incorrectamente. La pregunta ahora es a qué característica individual y a qué característica del ítem corresponde este valor del parámetro λ_{pi} . Pues bien, observemos las siguientes igualdades:

$$\lambda_{pi} = 3 = \frac{3}{1} = \frac{30}{10} = \frac{1}{1/3} = \dots$$

Es decir, el parámetro $\lambda_{pi} = 3$, que tiene un significado bien preciso, está asociado a diferentes combinaciones de características de un individuo y de un ítem: una “habilidad” igual a 3 y una “dificultad” igual a 1; o una “habilidad” igual a 30 y una “dificultad” igual a 10; o una “habilidad” igual a 1 y una “dificultad” igual a 1/3. Es decir, dado un valor de λ_{pi} , no existe un único valor de ε_p y un único valor de η_i , lo cual implica que los términos de dificultad y de habilidad no están bien definidos y, por tanto, son meros nombres a los que se les puede atribuir cualquier significado que no esté sustentado por el proceso que genera los datos

bajo análisis. Técnicamente, esto corresponde a un problema de identificación (para detalles, ver San Martín, González & Tuerlinckx, 2009; San Martín, 2015), el cual requiere ser resuelto a fin de asegurar un significado a los parámetros ε_p y η_i .

Significados de “habilidad” y “dificultad”

Para resolver dicha indeterminación, es necesario introducir una restricción de identificación. Una posible restricción es la siguiente:

$$\eta_1 = 1 \quad (2)$$

Bajo esta restricción, se deduce que

$$\frac{P(Y_{pi} = 1)}{P(Y_{pi} = 0)} = \lambda_{pi} = \varepsilon_p \quad (3)$$

para toda persona p . Es decir, bajo la restricción $\eta_1 = 1$, lo que típicamente se llama “habilidad de una persona p ” corresponde al cociente de la probabilidad que dicha persona responda el ítem estándar (así llamaremos al ítem con la etiqueta 1) correctamente sobre la probabilidad de que lo responda incorrectamente. Más aún, si $\varepsilon_p > 1$ tiene un significado bien preciso: la persona p tiene una probabilidad de responder correctamente el ítem estándar mayor que la de responderlo incorrectamente.

En relación al significado de η_i , notemos que

$$\frac{P(Y_{pi} = 1)/P(Y_{pi} = 0)}{P(Y_{pi} = 1)/P(Y_{pi} = 0)} = \frac{\lambda_{pi}}{\lambda_{pi}} = \frac{\varepsilon_p}{\varepsilon_p/\eta_i} = \eta_i \quad (4)$$

Es decir, lo que llamamos “dificultad del ítem i ” corresponde al *odd ratio* entre el ítem 1 y el ítem i para cada persona p . Así, si $\eta_i > 1$, entonces la razón de chances de una respuesta correcta de la persona p al ítem estándar es más grande que la razón de chances de una respuesta correcta al ítem i . En otras palabras, es más probable para la persona p responder correctamente el ítem estándar que el ítem i (esto es, $\eta_i > \eta_1 = 1$).

Representación simultánea habilidades-dificultades. Las consideraciones anteriores sugieren cómo las habilidades pueden compararse con las dificultades. En efecto,

$$\epsilon_p > \eta_i \Leftrightarrow P(Y_{pi} = 1) > P(Y_{pi} = 0) \quad (5)$$

para cada persona p y cada ítem i . Es decir, un individuo p tiene una habilidad superior a la dificultad del ítem i si la probabilidad que la persona p tiene de responder correctamente el ítem i es mayor que la probabilidad que tiene de responderlo incorrectamente.

Discusión. El modelo Rasch, así como otras modelos de Teoría de Respuesta al Ítem, son usados intensivamente en los sistemas educacionales nacionales e internacionales para analizar los resultados que una población de estudiantes obtiene después de la aplicación de un test. Estos análisis se reducen a estimar las habilidades de los individuos, en base a las cuales se calcula un puntaje individual; y a estimar las dificultades de los ítems (como otras características de los ítems). A pesar de esto, no es del todo claro lo que se entiende por “dificultad de un ítem” y por “habilidad de un estudiante” ya que, como lo hemos dicho en la introducción, se habla de la dificultad de un ítem tanto si se la estima usando Teoría Clásica de Test, o un modelo Rasch u otro modelo de Teoría de Respuesta al Ítem, como el 2PL o 3PL.

Tanto la habilidad de una persona, como la dificultad de un ítem, deben ser interpretadas con respecto a un proceso generador de datos, el cual se formaliza en términos probabilísticos. Este proceso no hace sino representar la heterogeneidad que se observa tanto en las respuestas que una persona da a un conjunto de ítems, así como la heterogeneidad que se observa cuando un grupo de estudiantes responde un determinado ítem. Asumido esto, hemos especificado el modelo Rasch de la siguiente manera:

- a. $\{Y_{pi} : p = 1, \dots, N; i = 1, \dots, I\}$ son mutuamente independientes.
- b. Para cada p e i , $Y_{pi} \sim \text{Bern}(\pi_{pi})$ para cada persona p y cada ítem i , con $\pi_{pi} \in (0,1)$.
- c. El parámetro π_{pi} se descompone como

$$\pi_{pi} = \frac{\epsilon_p / \eta_i}{1 + \epsilon_p / \eta_i} \quad (6)$$

donde ϵ_p representa la habilidad de una persona p , y η_i representa la dificultad del ítem i . Esta parametrización puede ser poco familiar al lector. Sin embargo, si utiliza la siguiente reparametrización

$$\theta_p = e^{\epsilon_p}, \quad \eta_i = e^{\beta_i}$$

entonces

$$P(Y_{pi} = 1) = \frac{e^{\theta_p \cdot \beta_i}}{1 + e^{\theta_p \cdot \beta_i}}$$

que es la forma estándar de presentar el modelo Rasch.

Lo que hemos constatado en las secciones anteriores es que el significado de ϵ_p y de η_i dependen exclusivamente del proceso generador de datos, que en este caso se reduce a distribuciones de probabilidad Bernoulli con parámetros dados por (6). Este significado se hace explícito sólo si se elimina una indeterminación en la igualdad (1), o equivalentemente (6). Una forma de eliminar dicha indeterminación es por medio de la restricción (2). Bajo esta restricción, es posible determinar el significado de la “habilidad de una persona p ” y de la “dificultad de un ítem i ”. Estos significados están explicitados en las ecuaciones (3) y (4). Es más, tanto las habilidades como las dificultades pueden compararse entre sí, y dicha comparación tiene un significado explícito dado por la desigualdad (5).

Tres observaciones se imponen:

- a. El desarrollo que hemos hecho acá para encontrar la interpretación de los parámetros del modelo Rasch es posible hacerla para los modelos 2PL y 3PL. Dejemos constatado aquí que para el modelo 3PL no es posible interpretar las características de los ítems y que, en particular, lo que llamamos “parámetro de adivinanza” no tiene significado alguno. Para detalles, ver San Martín, González y Tuerlinckx (2014).
- b. El significado de las habilidades, así como de las dificultades, se basa exclusivamente en probabilidades de respuestas correctas/incorrectas. Es por ello que este modelo debe considerarse como un modelo descriptivo en el sentido de que no se provee ninguna explicación psicológica o educacional de las habilidades y dificultades. Es posible, sin embargo, extender este

modelo de manera de incorporar un aspecto explicativo tanto de las habilidades como de las dificultades. Para detalles y referencias, ver San Martín y Rolin (2013).

c. En la construcción del modelo Rasch no se ha supuesto que la habilidad individual sea un rasgo latente, el cual se representa por medio de una variable aleatoria. Sin embargo, las modernas especificaciones del modelo Rasch lo asumen, aunque tienden a presentar el significado de “habilidad” y “dificultad” en la línea discutida hasta aquí; ver por ejemplo Embretson y Reise (2000). La pregunta que es necesario hacer, entonces, es cómo se modifican los significados de “dificultad” y “habilidad” bajo este nuevo esquema. Esto es lo que discutiremos en la segunda parte de este trabajo.

Segunda Parte

Descomposición Marginal-Condicional Fundamental

Los modelos de Teoría de Respuesta al Ítem se especifican desde una perspectiva completamente diferente a la discutida en la primera parte de este trabajo. El contexto general sobre el cual dichos modelos se especifican asume dos procesos aleatorios secuenciales: el primero de ellos describe la heterogeneidad de una característica individual llamada constructo o rasgo latente; el segundo dice relación con la heterogeneidad de las respuestas que se dan a un ítem o a un test por un individuo con un determinado nivel de habilidad. Si denotamos por Y la respuesta a un test o a un ítem, y por θ el rasgo o constructo latente, entonces los dos procesos aleatorios anteriores corresponden a una descomposición marginal-condicional de la distribución conjunta de (Y, θ) , la que denotamos por $p(Y, \theta)$; esto es,

$$p(Y, \theta) = p(\theta)p(Y|\theta) \quad (7)$$

donde $p(\theta)$ corresponde a la distribución de probabilidad que representa la heterogeneidad de los rasgos o constructos latentes de una determinada población de individuos, y $p(Y|\theta)$ corresponde a la distribución condicional que representa la heterogeneidad de las respuestas a un test o ítem cuando el individuo que responde tiene un rasgo latente dado por θ . El aspecto fundamental que

es necesario enfatizar es que la descomposición (7) asume que existe una relación de dependencia entre el rasgo o constructo latente θ y la respuesta observable Y , o, en otras palabras, que la respuesta observable Y está determinada por el rasgo latente θ .

En este contexto, el problema psicométrico consiste en “estimar” el nivel de habilidad θ de una persona después que se ha observado el puntaje Y que ha obtenido en una prueba o en un ítem. Más precisamente, dado que el rasgo latente es una variable aleatoria cuya distribución está dada por $p(\theta)$, el objetivo consiste en obtener la distribución $p(\theta|Y)$. Es importante notar que esta distribución, junto a la distribución de los puntajes observados $p(Y)$, constituyen una descomposición alternativa de la distribución conjunta $p(Y|\theta)$, esto es,

$$p(Y|\theta) = p(Y)p(\theta|Y). \quad (8)$$

Combinando las igualdades (7) y (8) es posible obtener una forma explícita de calcular la distribución de probabilidad $p(\theta|Y)$:

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)}$$

En esta igualdad, conocida como el Teorema de Bayes-Laplace, la única componente que puede estimarse a partir de las observaciones es $p(Y)$; las otras dos son supuestas y, en algún grado, justificadas por las aplicaciones. En la práctica, se reporta como “estimador” de θ la media de la distribución condicional $p(\theta|Y)$, estimador que se conoce como “Empirical Bayes estimator”.

Especificación del Modelo Rasch

Consideremos nuevamente la matriz de datos de la introducción. Como antes, denotaremos por $Y_{pi} = 1$ si la persona p responde correctamente el ítem i , en caso contrario diremos que $Y_{pi} = 0$. Las respuesta de la persona p a todos los ítems será llamado patrón de respuesta y será denotado por Y_p . Teniendo en cuenta el contexto descrito en la sección anterior, debemos afirmar que los observables están dados por los patrones de respuesta de cada individuo, a saber

Y_p, Y_{p2}, \dots, Y_n . El problema consiste en especificar el modelo estadístico que genera dichas observaciones. En lugar de limitarse a especificar una distribución de los observables, lo que se hace es suponer que existe una estructura subyacente responsable de generar los observables. Esta estructura subyacente corresponde a la descomposición marginal-condicional de la distribución conjunta de (Y_p, θ_p) . Los componentes de dicha descomposición serán llamado modelo estructural, para distinguirlo del modelo estadístico. Digamos de paso que este paradigma fue formalizado por Koopmans y Reiersol (1950), y aplicado a los modelos de Teoría de Respuesta al Ítem por Lord (1952).

Modelo estructural. El modelo estructural está definido por las siguientes hipótesis:

a. Los rasgos latentes θ_p 's son mutuamente independientes e idénticamente distribuidos, con distribución común G^σ —es decir, una distribución conocida salvo el parámetro de escala. Típicamente, se asume que dicha distribución corresponde a una normal de media 0 y de varianza σ^2 . Esta hipótesis indica relación con la selección de los individuos: lo que se selecciona es un rasgo latente, del cual dependen las respuestas de dicho individuo a un conjunto de ítems.

b. Los patrones de respuesta Y_p 's son mutuamente independientes condicionalmente a $(\theta_p, \theta_{p2}, \dots, \theta_n)$. Esto es, si se conociesen los rasgos latentes de cada individuo, el patrón de respuesta de un individuo no sería informativo para inferir el patrón de respuesta de cualquiera de los restantes individuos.

c. El patrón de respuesta Y_p solo depende de θ_p . Esto es, el patrón Y_p de la persona p solo depende del rasgo latente θ_p de dicha persona, y no del rasgo latente de las restantes personas.

d. Para cada individuo p , sus respuestas $\{Y_{p1}, Y_{p2}, \dots, Y_{pi}\}$ son mutuamente independientes condicionalmente al rasgo latente θ_p . Esto es conocido como el Axioma de Independencia Local. Discutiremos su significado en la siguiente subsección.

e. La distribución condicional de Y_{pi} dado θ_p corresponde a una Bernoulli de parámetro

$$\frac{e^{\theta_p - \beta_i}}{1 + e^{\theta_p - \beta_i}} \quad (9)$$

Es importante enfatizar que la expresión (9) corresponde a la probabilidad condicional $P(Y_{pi} = 1 | \theta_p)$, mientras que en el paradigma discutido en la primera parte de este trabajo, esto corresponde a la probabilidad marginal $P(Y_{pi} = 1 | \theta_p)$. Por otro lado, la probabilidad condicional $P(Y_{pi} = 1 | \theta_p)$ puede especificarse de diversas formas, como por ejemplo el modelo 2PI o el modelo 3PL.

¿Cuál es la diferencia entre el modelo estructural y el modelo estadístico? El modelo estructural induce un modelo estadístico, que corresponde al proceso generador de las observaciones — en nuestro caso, los patrones de respuesta. Para entender la diferencia y la relación que existe entre ambos modelos, parece instructivo introducir un símil, que está basado en la Teoría Clásica de Probabilidades, tal y como fue desarrollada por Laplace (1774). De hecho, el paradigma introducido por Laplace (1774) está completamente asumido en el paradigma explicitado en la sección sobre la descomposición marginal-condicional fundamental.

Supongamos que se tiene dos urnas: la urna A, que representa a las personas cuyo rasgo latente es “bajo”; la urna B, que representa a las personas cuyo rasgo latente es “alto”. Cada urna contiene bolitas blancas y negras: las bolitas blancas corresponden a “respuestas correctas”, mientras que las bolitas negras corresponden a “respuestas incorrectas”. Supongamos que conocemos la composición de cada urna: la urna A está compuesta por 3 bolitas blancas y 17 negras; la urna B está compuesta por 14 bolitas blancas y 4 bolitas negras. Ahora bien, describamos la extracción de las bolitas:

a. En primer lugar, definimos un procedimiento aleatorio que nos permita seleccionar cada urna. Supongamos, por ejemplo, que la urna A se selecciona con probabilidad $\frac{2}{3}$, mientras que la urna B se selecciona con probabilidad $\frac{1}{3}$.

b. Una vez que se selecciona una urna, se extraen con reemplazo, por ejemplo, 4 bolitas.

El esquema anterior tiene su correspondencia con el problema psicométrico que estamos modelando. En efecto,

a. La distribución de probabilidad bajo la cual se seleccionan las urnas corresponde a la distribución del rasgo latente θ_p .

b. Una vez que se escoge una urna, por ejemplo la urna B, la extracción con reemplazo de las bolitas asegura que dichas extracciones son mutuamente independientes. Esto es, condicionalmente a la urna, las extracciones son mutuamente independientes. Lo que corresponde al Axioma de Independencia Local.

Ahora bien, teniendo en cuenta el esquema de extracción de bolitas ya especificado, Laplace (1774) enuncia dos problemas. El primer problema consiste en calcular la probabilidad de extraer, por ejemplo, la secuencia (b, n, b, b). Evidentemente, la probabilidad de extraer dicha secuencia depende de la urna que se escoja: si se escoge la urna A, entonces aplicando el Axioma de Independencia Local se tiene que

$$P(b, n, b, b | A) = P(b | A)P(n | A)P(b | A)P(b | A) = \left(\frac{3}{20}\right)^3 \times \frac{17}{20}.$$

Y si se escoge la urna B, entonces aplicando el Axioma de Independencia Local,

$$P(b, n, b, b | B) = P(b | B)P(n | B)P(b | B)P(b | B) = \left(\frac{14}{18}\right)^3 \times \frac{4}{18}.$$

Pero la elección de la urna es aleatoria, por lo que es necesario combinar estos resultados teniendo en cuenta la distribución de elección de las urnas; esto es,

$$P(b, n, b, b) = P(b, n, b, b | A)P(A) + P(b, n, b, b | B)P(B) \quad (10),$$

donde $P(A) = \frac{2}{3}$ y $P(B) = \frac{1}{3}$. Lo que llamamos modelo estructural corresponde a la probabilidad de escoger cada urna y a la probabilidad condicional de extraer bolitas de cada urna. Lo que llamamos modelo estructural corresponde a la probabilidad de escoger cada urna y a la probabilidad condicional de extraer bolitas de cada urna. Lo que llamamos modelo

estadístico corresponde a la probabilidad de la secuencia (b, n, b, b), que se obtiene después de promediar las probabilidades condicionales de extraer la secuencia (b, n, b, b) de cada urna, donde dicho promedio se determina teniendo en cuenta el “peso” de cada urna.

En el contexto psicométrico, el modelo estadístico corresponde a la distribución de probabilidad de los patrones de respuesta Y_p , el cual se obtiene promediando las probabilidades condicionales con respecto a la distribución del rasgo latente. Más precisamente, para $y_p \in \{0, 1\}^I$ (es decir, un vector de largo I cuyas coordenadas son 0 ó 1),

$$P(Y_p = y_p) = e^{-\sum_{1 \leq i \leq I} y_{pi} \beta_i} \times E \left\{ \frac{e^{y_p + \theta}}{\prod_{1 \leq i \leq I} (1 + e^{\theta - \beta_i})} \right\} \quad (11)$$

donde y_{p+} corresponde al total de 1's que tiene el vector y_p , y la esperanza de la igualdad anterior se calcula con respecto a la distribución del rasgo latente. La igualdad (11) nos permite además concluir que el modelo estadístico está parametrizado por $(\beta_p, \beta_2, \dots, \beta_p, \sigma)$, por lo que el significado de los “parámetros de dificultad” β_i 's, así como del “parámetro de escala” σ deben precisarse con respecto al modelo estadístico. Este aspecto será discutido en la siguiente sección.

Significado de Axioma de Independencia Local. Usando el mismo esquema anterior, podemos apreciar el impacto que tiene el Axioma de Independencia Local sobre la extracción de una secuencia, como por ejemplo (b, n, b, b). Este impacto consiste en que se induce una dependencia entre las extracciones consideradas por separado. Más precisamente, se puede verificar que

$$P(b, n, b, b) \neq P(b)P(n)P(b)P(b).$$

En el contexto psicométrico, el Axioma de Independencia Local implica que las respuestas de un individuo a diferentes ítems están correlacionadas entre sí. Más precisamente,

$$\text{cov}(Y_{pi}, Y_{pj}) = \text{cov} \left(\frac{e^{y_{pi}(\theta - \beta_i)}}{1 + e^{(\theta - \beta_i)}}, \frac{e^{y_{pj}(\theta - \beta_j)}}{1 + e^{(\theta - \beta_j)}} \right) \neq 0$$

En otras palabras, las respuestas Y_{pi} 's están relacionadas entre sí porque cada respuesta está relacionada con el rasgo latente θ_p (Lazarsfeld, 1954).

El problema inverso. El segundo problema mencionado por Laplace (1774) puede enunciarse de la siguiente manera: supongamos que se observa la secuencia (b, n, b) . ¿Cuál es la urna más probable de la que proviene esta secuencia? Para responder a esta pregunta, es necesario calcular las probabilidades condicionales $P(A|b, n, b)$ y $P(B|b, n, b)$. Estas se calculan usando el Teorema de Probabilidades Inversa o Fórmula de Bayes-Laplace:

$$P(A|b, n, b) = \frac{P(b, n, b|A)P(A)}{P(b, n, b)}$$

Similarmente para $P(B|b, n, b)$. Se puede verificar que $P(B|b, n, b) > P(A|b, n, b)$, por lo que la urna más probable de la que la secuencia (b, n, b) proviene es la urna B.

En términos psicométricos, el problema inverso consiste en calcular la probabilidad condicional del rasgo latente una vez que se ha observado un patrón de respuesta.

Significados de “dificultad de un ítem” y de “habilidad individual”

Significado de la dificultad de un ítem. Para poder interpretar el significado de los parámetros de dificultad, así como del parámetro de escala, es necesario realizar un análisis de identificación. Dicho análisis ha sido recientemente realizado por San Martín y De Boeck (2015), concluyéndose lo siguiente:

a. El parámetro σ representa la dependencia entre dos ítems distintos. Más aún, esta dependencia es la misma para todos los pares de ítems.

b. El parámetro β_i es una función tanto de la proporción de respuestas correctas del ítem i , como de la dependencia común entre ítems, dependencia capturada por el parámetro σ .

Para apreciar de forma explícita estas afirmaciones, consideremos el llamado modelo Normal-Ojiva, el cual asume que $P(Y_{pi} = 1 | \theta_p) = \Phi(\theta_p - \beta_i)$, donde $\Phi(\cdot)$ es la función de distribución de una normal estándar, y que la distribución del rasgo latente θ_p es una

distribución normal de media 0 y varianza σ^2 . Sea

$$\begin{pmatrix} U_i \\ U_j \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad (12)$$

donde

$$\rho = \frac{\sigma^2}{1 + \sigma^2}$$

Entonces, considerando la distribución conjunta de (Y_{pi}, Y_{pj}) para cualquier individuo p y para cualquier par de ítems i y j , se tiene que

$$P(Y_{pi} = 1) = P(U_i \leq -\gamma_i) = \Phi(-\gamma_i)$$

$$P(Y_{pj} = 1) = P(U_j \leq -\gamma_j) = \Phi(-\gamma_j)$$

$$P(Y_{pi} = 1, Y_{pj} = 1) = P(U_i \leq -\gamma_i, U_j \leq -\gamma_j)$$

con

$$\gamma_i = \frac{\beta_i}{\sqrt{1 + \sigma^2}}, \quad \gamma_j = \frac{\beta_j}{\sqrt{1 + \sigma^2}} \quad (13)$$

Para detalles, ver San Martín & Rolin (2007). Notemos que $P(Y_{pi} = 1)$ corresponde a la probabilidad marginal de que cualquier individuo p responda correctamente el ítem i , y que puede ser estimado por medio de la proporción de individuos que efectivamente respondieron dicho ítem correctamente —es decir, la dificultad empírica del ítem según la TCT. Similarmente, $P(Y_{pi} = 1, Y_{pj} = 1)$ corresponde a la probabilidad conjunta marginal de responder correctamente ambos ítems i y j . El Axioma de Independencia Local asegura que hay una relación de dependencia entre Y_{pi} e Y_{pj} , por lo que dicha probabilidad conjunta es distinta al producto $P(Y_{pi} = 1) \times P(Y_{pj} = 1)$. Dicha probabilidad conjunta puede ser estimada por medio de la proporción de individuos que efectivamente respondieron correctamente ambos ítems.

De las igualdades anteriores se deduce que, para cada ítem i ,

$$\gamma_i = -\Phi^{-1} \left(P(Y_{pi} = 1) \right), \quad (14)$$

que se puede interpretar diciendo que γ_i es una función decreciente de la dificultad empírica del ítem i . Por otro lado, se puede verificar que $P(Y_{pi} = 1, Y_{pj} = 1)$ se puede escribir como

$$P(Y_{pi}=1, Y_{pj}=1) = E \left\{ \Phi \left(\frac{-\gamma_j - \rho U_i}{\sqrt{1 + \rho^2}} \right) 1_{\{U_i \leq \gamma_i\}} \right\} \quad (15)$$

donde $1_{\{U_i \leq \gamma_i\}}$ es igual a 1 cuando efectivamente $U_i \leq \gamma_i$, y 0 en caso contrario. Teniendo en cuenta que γ_i y γ_j son conocidos (ver ecuación (14)) el término de la derecha de la igualdad (15) es una función de ρ , que denotamos por $H(\rho)$. Se puede verificar que la función $H(\cdot)$ es estrictamente creciente y, por tanto, invertible, por lo que ρ es una función tanto de $P(Y_{pi}=1)$ y $P(Y_{pj}=1)$ (a través de γ_i y γ_j), así como de la probabilidad marginal conjunta $P(Y_{pi}=1, Y_{pj}=1)$. En otras palabras, ρ , y por tanto σ (gracias a la igualdad (12)), representa la dependencia que hay entre las respuestas a cualquier par de ítems.

Teniendo en cuenta las ecuaciones (13) y (14), se deduce que, para todo ítem i ,

$$\beta_i = -\Phi^{-1}(P(Y_{pi}=1)) \times \sqrt{1 + \sigma^2}. \quad (16)$$

Es decir, lo que llamamos dificultad del ítem es una función tanto de la dificultad empírica del ítem, así como de la dependencia entre cualquier par de ítems. La igualdad (16) puede interpretarse de la siguiente manera:

a. Si la dificultad empírica del ítem i es al menos 0.5 (es decir, si al menos el 50% de los individuos responde el ítem correctamente), entonces β_i tiene signo negativo. Más aún, β_i es una función decreciente de la dependencia entre cualquier par de ítems: más pequeña (respectivamente, más grande) es esta dependencia, menos negativo (más negativo) es β_i .

b. Si la dificultad empírica del ítem i es a lo más 0.5 (es decir, si a lo más el 50% de los individuos responde el ítem correctamente), entonces β_i tiene signo positivo. Más aún, β_i es una función creciente de la dependencia entre cualquier par de ítems: más pequeña (respectivamente, más grande) es esta dependencia, más positivo (respectivamente, menos positivo) es β_i .

En otras palabras, para los ítems cuya dificultad empírica es a lo más 0.5, la escala de dificultades del modelo Rasch es una contracción de la escala de

dificultades empíricas. Para los ítems cuya dificultad empírica es a lo menos 0.5, la escala de dificultades del modelo Rasch es una extensión de la escala de dificultades empíricas².

Significado de habilidad individual. En relación con la “habilidad individual” (rasgo latente), como hemos dicho anteriormente, la práctica común consiste en reportar como “estimador del rasgo latente” la media condicional de la distribución condicional $p(\theta | Y_p = y_p)$. El esquema laplaciano desarrollado en la sección sobre el Modelo de Rasch, nos permite entender el significado de esta “estimación”: se trata de un problema de clasificación. En efecto, una vez que se observa un patrón de respuesta Y_p , interesa saber cuál es el grupo de habilidad más probable al que pertenece el individuo cuyo patrón de respuesta es Y_p .

Discusión final

El análisis estadístico de un conjunto de observaciones requiere especificar el proceso generador de dichos datos. Este proceso corresponde a lo que hemos llamado modelo estadístico. Formalmente hablando, el modelo estadístico corresponde a una familia de distribuciones de probabilidad etiquetadas por parámetros. Los parámetros tienen un significado que debe ser derivado a partir del modelo estadístico. El significado de los parámetros nunca es algo nominal o impuesto.

En este trabajo hemos mostrado que el significado de “dificultad de un ítem” y “habilidad de una persona” es completamente diferente dependiendo de cómo se especifica el modelo estadístico. La primera especificación, debida al propio Rasch, y que en la jerga estadística y econométrica se denomina modelo Rasch efecto fijo, se caracteriza por el hecho de que todas las respuestas de todos los individuos, además de las respuestas de un mismo individuo, son mutuamente independientes entre sí. En este contexto, tanto la dificultad de un ítem, como la habilidad de una persona

²Estas afirmaciones corrigen lo expuesto en San Martín y De Boeck (2015, p. 7), donde no se advirtió que el signo de $\Phi^{-1}[P(Y_{pi}=1)]$ puede ser negativo o positivo, dependiendo si $P(Y_{pi}=1) < 0.5$ o $P(Y_{pi}=1) > 0.5$.

se pueden escribir en función de probabilidades de respuesta correcta e incorrecta, tal y como lo muestran las ecuaciones (3) y (4). Este resultado depende de una restricción de identificación, a saber que la dificultad del ítem estándar es igual a 0. Una consecuencia importante es que, en este contexto, se pueden comparar habilidades individuales y dificultades de ítems. Dicha comparación se hace en términos de probabilidades de respuesta correcta e incorrecta; ver ecuación (5).

Resulta interesante constatar que el procedimiento de standard setting llamado Bookmark utiliza la representación simultánea habilidades-dificultades para clasificar a las personas luego que un conjunto de jueces ha establecido un puntaje de corte teniendo en cuenta las dificultades de los ítems (Cizek, 2012). Estrictamente hablando, este procedimiento tiene sentido solo en la especificación de un modelo Rasch en la cual las respuestas a los ítems no son sólo mutuamente independientes, sino que además no están relacionadas entre sí por medio de un rasgo latente no observable.

La otra forma de especificar el modelo Rasch, debida a Lord, y que en la jerga estadística y econométrica se denomina modelo Rasch efecto aleatorio, está dentro de una tradición completamente distinta y que, en términos históricos, tiene su origen en Laplace y en su concepción de probabilidades inversas. El supuesto fundamental es que las respuestas que un individuo da a un grupo de ítems están correlacionadas entre sí porque cada una está relacionada con el rasgo latente (o habilidad individual no observable). Desde esta perspectiva, el problema psicométrico consiste en calcular la probabilidad condicional del constructo latente condicionalmente a las respuestas dadas por un individuo. Cuando este problema se examina a la luz del paradigma de Laplace, se puede afirmar que se trata de un problema de clasificación: aquí no interesa estimar la verdadera habilidad de una persona, sino que interesa clasificarla en un grupo de habilidades después que dichos grupos han sido definidos por medio de la distribución del rasgo latente.

Esta conclusión se refrenda aún más si nos hacemos

la siguiente pregunta: si el rasgo latente es no-observable, ¿cómo es posible tener una idea precisa de su distribución de probabilidad? Una manera de responder a esta pregunta es especificar un modelo Rasch, donde la distribución del rasgo latente es completamente desconocida y por tanto se convierte en un parámetro de interés. Se ha demostrado que dicha distribución no es identificada, por lo que no tiene significado alguno; para detalles, ver San Martín, Jara, Rolin y Mouchart (2011) y San Martín, Rolin y Castro (2013). Por lo tanto, una manera de justificar el procedimiento de estimación de la probabilidad condicional de las habilidades dado los patrones de respuesta es atendiendo al problema de clasificación.

En este contexto, las dificultades de los ítems no tienen la misma significación que en el caso del modelo Rasch efecto fijo. Las dificultades de los ítems son una función tanto de la dificultad empírica de cada ítem, así como del grado de dependencia entre todo par de ítems. Así, la escala de las dificultades empíricas es contraída o extendida (dependiendo de si los ítems tienen dificultades empíricas mayores a 0.5, o menores a 0.5) en un factor que depende precisamente de dicha dependencia.

La interpretación de un fenómeno observable u observado depende completamente del modelo estadístico propuesto o inducido. El caso de datos educacionales no es la excepción, aunque se pone escasa atención a la significación de los términos “dificultad” y “habilidad”, términos que son los que precisamente ayudan a construir el discurso sustantivo de los datos. Este trabajo quiere enfatizar este asunto, mostrando cómo un modelo estadístico tiene una lectura sustantiva que puede y debe dialogar con la teoría psicológica y educacional que hace usos de estos modelos.

Referencias

- Cizek, G. (2012). *Setting Performance Standards: Foundations, Methods, and Innovations*. Routledge.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates, Publishers.

- Koopmans, T. J. & Reiersol, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics*, 21, 165-181.
- Laplace, P.S. (1774). *Mémoire sur la probabilité des cause par les événements*. Paris: Académie Royal des Sciences de Paris.
- Lazarsfeld, P. F. (1954). A conceptual introduction to latent structure analysis. In: P. F. Lazarsfeld (Ed.), *Mathematical Thinking in the Social Sciences*, pp. 349-387. New York: Russell & Russell.
- Lord, F. (1952). *A Theory of Test Scores*. Psychometric Monograph No. 7.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research (Expanded Edition, 1980). Chicago: The University Chicago Press.
- San Martín, E. (2015). Identification of Item Response Theory Models. In: R.K. Hambleton & W. van der Linden (Eds.), *Handbook of Item Response Theory: Models, Statistical Tools, and Applications*, Volumen 2, Chapter 8. En Imprenta.
- San Martín, E. & De Boeck, P. (2015). What do you mean by a difficult item? On the interpretation of the difficulty parameter in Rasch models. In: Roger E. Millsap, Daniel M. Bolt, L. Andries van der Ark, Wen-Chung Wang (Eds.), *Quantitative Psychology Research*, Springer Proceedings in Mathematics & Statistics, 89, Chapter 1.
- San Martín, E., González, J. & Tuerlinckx, F. (2009). Identified Parameters, Parameters of Interest and Their Relationships. *Measurement: Interdisciplinary Research & Perspective*, 7, 97-105.
- San Martín, E., González, J. & Tuerlinckx, F. (2014). On the unidentifiability of the fixed-effects 3PL model. *Psychometrika*, 80, 450-467. DOI: 10.1007/S11336-014-9404-2.
- San Martín, E., Jara, A., Rolin, J.-M. & Mouchart, M. (2011). On the Bayesian nonparametric generalization of IRT-type models. *Psychometrika*, 76, 385-409.
- San Martín, E. & Rolin, J. -M. (2007). Identifiability and Estimability of Parametric Rasch-Type Models. Discussion Paper 0702, Institut de Statistique, Université catholique de Louvain, Belgium. Disponible en <http://sites.uclouvain.be/IAP-Stat-Phase-V-VI/ISBApub/dp2007/DP0702.pdf>.
- San Martín, E. & Rolin, J. -M. (2013). Identification of parametric Rasch-type models. *Journal of Statistical Planning and Inference*, 143, 116-130.
- San Martín, E., Rolin, J.-M. & Castro, M. (2013). Identification of the 1PL model with guessing parameter: parametric and semi-parametric results. *Psychometrika*, 78, 341-379.

Recibido: 23 de abril de 2015

Aceptado: 18 de setiembre de 2015