



Aidesoc
Formación y
Desarrollo

Introducción al análisis de la Teoría de Respuesta al Ítem

Antonio Matas Terrón

Ediciones Aidesoc 2010

Introducción al análisis de la Teoría de Respuesta al Ítem
Antonio Matas Terrón

Edita: Aidesoc.net



Licencia:

Esta obra está bajo una licencia Attribution-NonCommercial de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/> o envíe una carta a Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Introducción al análisis de la Teoría de Respuesta al Ítem

Introducción

La Teoría de Respuesta al Ítem es un área de desarrollo de la [Psicometría](#) que trata de medir [rasgos](#) latentes a través de una serie de modelos matemáticos. Como modelo de variables latentes forma parte de los [análisis de datos multivariante](#). En general, los modelos de variables latentes son modelos de [regresión multivariantes](#) que enlazan una [variable respuesta](#) con otra no observada (Rizopoulos, 2006).

Los principios de esta teoría se encuentran en el [Danés Geog Rasch](#). No obstante, los conceptos básicos de esta teoría se pueden encontrar en investigadores como [Alfred Binet](#) o [Louis L. Thourston](#).

La TRI se ha desarrollado en los últimos años, registrándose numerosas aplicaciones en el ámbito de la evaluación del rendimiento académico, la medición de las actitudes, la medición de las competencias (deportivas, lingüística, etc.) y por supuesto, la construcción de test. Muestra de este desarrollo son los grupos de investigación que están trabajando en este terreno: [EDIMⁱ](#) (Grupo Universidad Autónoma de Madrid), [Grupo HUM-369ⁱⁱ](#) (Universidad de Málaga).

La Teoría de Respuesta al Ítem (TRI) o Teoría de Rasgo Latente (TRL) asume que entre la respuesta que una persona ofrece a una pregunta o prueba, y el rasgo o habilidad (puede entenderse también la competencia, actitud, aptitud, etc.) de la persona, existe una relación no lineal, que se puede expresar en términos probabilísticos. El conjunto de cálculos que se aplican para ello determina los distintos modelos y resultados de la TRI.

Teoría de Respuesta al ítem frente a la Teoría Clásica de los Test

La TRI es una alternativa efectiva a la [Teoría Clásica de los Test](#) (TCT) para el análisis de los ítems y de los test, así como para la estimación de las puntuaciones “[verdadera](#)” de las personas que responden a dichos ítems y tests. La TCT, aunque es una estrategia muy útil y muy difundida, tiene ciertos inconvenientes que la TRI trata de resolver en gran parte.

En la TCT las características de los ítems o pruebas (tales como la [dificultad](#) o la [discriminación](#)) dependen de las características de la muestra que contesta al test, y por tanto, también de las características de la población de referencia. Igualmente, las características del test (tales como la [fiabilidad](#) y la [validez](#)) dependen además de la muestra, del número de ítems que compone el test. De esta forma, la estimación de la puntuación verdadera, depende de las características de la muestra y del propio test. Otros problemas de la TCT están en el cumplimiento del supuesto independencia entre puntuación verdadera y error, así como del principio de [homocedasticidad](#).

En la TRI, esa puntuación verdadera perseguida (el nivel de aptitud, rasgo, habilidad, competencia, etc.), es considerado independiente del ítem, y por lo tanto, la respuesta del sujeto ante el ítem, depende sólo de este nivel. En este sentido, se dice que el rasgo es la [Variable Independiente](#), y la respuesta la [Variable Dependiente](#).

La unidad básica de estudio en la TRI es el ítem. La relación entre la respuesta de la persona al ítem y el rasgo que este tiene, se estima mediante una [función probabilística](#), representada mediante la [Curva Característica del Ítem \(CCI\)](#). La importancia de la TRI está en el nivel de rasgo, de forma similar a la importancia de la puntuación verdadera en la TCT. Sin embargo, al contrario que ésta, en la TRI, el valor del rasgo es independiente del test y del ítem utilizado para estimarlo. Otra gran

ventaja de la TRI es que no se necesita especificar su distribución de forma que no es necesario, por ejemplo, garantizar la [normalidad](#) de los datos.

De forma resumida, tal como escribíamos en otro documento (Matas, 2001), la TRI tiene tres ventajas:

- Invarianza de los [parámetros](#) de los ítems respecto a la muestra que se calcula. Es decir, que los parámetros del ítem no cambian aunque las personas que contesten sean distintas.
- Invarianza del parámetro del rasgo del sujeto respecto al instrumento utilizado para estimarlo. Es decir, que el nivel de habilidad de la persona no depende del test.
- Proporciona medidas locales de precisión a través de la función de precisión. Es decir, que ofrece un indicador de la precisión de las estimaciones para niveles concretos.

Supuestos de la TRI

Los supuestos básicos que exige el modelo de rasgo latente son básicamente tres:

- El principal supuesto es que se asume la existencia de una variable no observada, latente, que explicaría las respuestas de las personas a una prueba o ítem.
- Unidimensionalidad, es decir, el ítem o prueba mide un solo rasgo.
- Por otra lado debe cumplirse la independencia local, es decir, que la respuesta a un ítem no influye en la respuesta dada a ningún otro. Esto permite afirmar que la probabilidad de responder correctamente a un conjunto de ítems es el producto de las probabilidades de contestar correctamente a cada ítem por separado.

En ciertas ocasiones es difícil comprobar la independencia local de los ítems, por tal motivo, es habitual comprobar solamente la unidimensionalidad a través de un [análisis factorial](#) o alguna prueba similar.

Modelos de la TRI

En la TRI se suelen distinguir tres modelos distintos en función del número de [parámetros](#) que se estiman:

- Modelo de un parámetro, también llamado modelo de Rasch. Se estima la dificultad del ítem (b).
- Modelo de dos parámetros. Se estima la dificultad (b) y la discriminación del ítem (a).
- Modelo de tres parámetros, o modelo de Birnbaum. Este modelo se incluye en la TRI aunque en esencia puede considerarse como un enfoque alternativo a la TRI. El modelo de Birnbaum estima la dificultad (b), la discriminación (a) y el pseudoazar o adivinación (c).

También se puede hablar de modelos en función del número de respuesta de los ítems, así tendremos:

- Modelo dicotómico, donde las respuestas a los ítems tienen dos opciones, normalmente

valoradas como 1 y 0.

- Modelos politómicos. En este caso los ítems tiene más de dos opciones, siendo generalmente escalas de apreciación graduadas. Por ejemplo del tipo “Indique en qué grado está de acuerdo con las afirmaciones, siendo 1 nada de acuerdo, 2 algo de acuerdo y 3 muy de acuerdo”.

Por último, hay que tener en cuenta también el algoritmo de estimación del nivel de habilidad o rasgo. En este sentido, las funciones habituales para relacionar las respuestas al test y el nivel de habilidad suelen ser la [función logística](#) y la [función normal](#).

Resultados de la TRI

La TRI genera una serie de resultados de especial interés:

- [Parámetros](#). Como se ha comentado anteriormente, tanto la dificultad, como la discriminación son dos elementos esenciales en la valoración de un ítem. Lo mismo puede decirse sobre el pseudoazar. Estos elementos tienen implicaciones directas en la construcción de test y pruebas académicas.
- [Curva característica del ítem](#). Es un elemento fundamental en la TRI porque, por un lado, ofrece una representación gráfica de las características del ítem, y por otra parte muestra la relación entre el nivel de habilidad y la respuesta al ítem, en términos probabilísticos.
- [Puntuación verdadera](#) en el test. Se identifica con la suma de las probabilidades estimadas con relación al nivel de habilidad, para el total de ítems del test.
- [Curva características del test](#). Permite ver de manera gráfica, la relación entre puntuación verdadera y habilidades. La Curva Característica del Test (CCT) muestra la relación existente entre la puntuación total en un test (no solamente en un ítem como sucede con la CCI) y el nivel de habilidad de una persona.
- [Nivel de información](#). Es un indicador de la precisión de la estimación.

Curva Característica del Ítem (CCI)

La TRI afirma que si una persona tiene un nivel n de conocimiento sobre geografía, por ejemplo, tendrá asociado a dicho valor una probabilidad de que conteste correctamente a una pregunta de geografía. Expresado a la inversa, si una persona que responde correctamente a un ítem, es posible suponer que esta persona posee un determinado nivel de conocimiento.

Para cada nivel de conocimiento, aptitud, habilidad o rasgo existe una probabilidad asociada de contestar correctamente al ítem. Esta probabilidad es pequeña para sujetos con bajo nivel de rasgo, y alta para sujetos con altos niveles de rasgo. La relación entre el nivel de aptitud y la probabilidad de respuesta correcta al ítem se representa con la curva características del ítem (CCI). En la práctica la CCI se suele representar en una escala situada entre -3 y $+3$ puntos de habilidad. En el eje x se sitúan los valores de la habilidad y en el eje y las probabilidades asociadas.

La CCI depende de tres parámetros básicos, o parámetros del ítem. Estos son el índice de dificultad, el índice de discriminación y el índice de azar (o pseudo-azar).

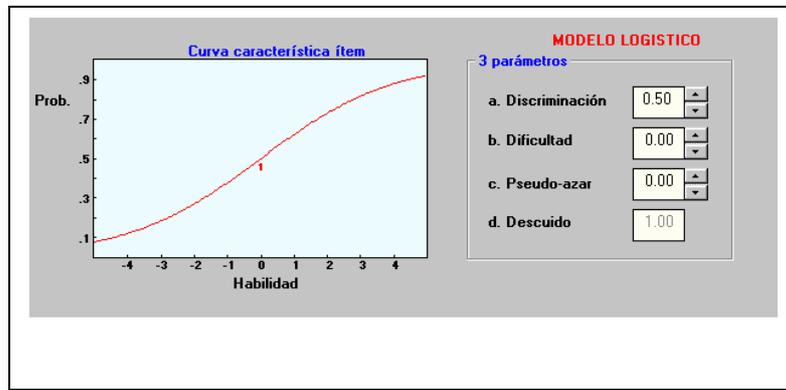


Gráfico 1. Curva característica de un ítem

Parámetros del ítem

Como se ha indicado, los parámetros son tres: dificultad, discriminación y pseudo-azar. El índice de dificultad se suele indicar con la letra “b” y se define como el punto de la escala de habilidad donde la probabilidad de responder correctamente el ítem es igual a 0,5. Es decir, que coincide con el grado de habilidad donde la curva “señala” una probabilidad de 0,5. Dicho de otra forma, con ese nivel de habilidad, la persona tienen una probabilidad del 50% de contestar correctamente al ítem.

El índice de discriminación indica hasta que punto la pregunta puede discriminar entre sujetos que tienen un nivel alto de habilidad de aquellos que tienen un nivel bajo. El índice de discriminación se suele representar con la letra “a”. Esta propiedad se refleja en la pendiente de la curva, así, una mayor pendiente indicará una mayor discriminación y viceversa.

El coeficiente de azar o pseudoazar, se representa con la letra “c” e indica la probabilidad de contestar correctamente un ítem sin tener el nivel de habilidad exigido. Este parámetro trata de representar el efecto del azar.

Función de información

La función de información es un algoritmo que indica la precisión con que se realiza la estimación de los parámetros. Así, para cada nivel de habilidad se atribuye un valor de error en la estimación. Partiendo de la idea de que la varianza de un estimador es un indicador de su grado de precisión o imprecisión (de forma que una mayor varianza sugiere una menor precisión) entonces, la cantidad de información se puede calcular como la inversa de dicha varianza. De esta forma a mayor precisión, mayor información, y al contrario.

Por un lado se puede calcular la función de información de un ítem y la función de información del test completo:

- Función de información del ítem (FII). Aunque se puede calcular para todos los niveles de habilidad, en los modelos de uno y dos parámetros, los ítems miden con mayor precisión alrededor del valor de dificultad.
- Función de información del test (FIT). Calcula la cantidad de información proporcionada en el test total para los distintos niveles de habilidad. Se calcula sumando las funciones de información de todos los ítems para cada nivel de habilidad.

A partir de la función de información es posible comparar la eficacia relativa (ER) de un test con relación a otro test sobre cada nivel de habilidad. Si la ER de un test con relación a otro tiende a 1 para un nivel de habilidad, indicaría que ambos son similares en cuanto a precisión en dicho nivel.

Procedimiento para el análisis de un test por medio de la TRI

De forma esquemática los pasos a seguir para analizar las características de un test según la TRI, y estimar los niveles de rasgo de las personas que contesten al test son los siguientes:

- a) Preparar los datos para su análisis.
- b) Analizar si los datos cumplen los supuestos básicos de la TRI.
- c) Estimar los parámetros del modelo elegido (uno, dos o tres parámetros). Elaborar los resúmenes y gráficas correspondientes.
- d) Analizar el ajuste del modelo a los datos. Si el ajuste no es adecuado volver a la fase anterior recurriendo a otro modelo.
- e) Estimar los niveles de habilidad de los participantes.

Para realizar los cálculos se utiliza el paquete “ltm” (Rizopoulos, 2006) que es una librería del programa R (R Development core Team, 2011).

Conclusiones

La TRI aporta un enfoque alternativo a la Teoría Clásica de los Test para analizar las características de los ítems. En otras características, permite analizar la dificultad y discriminación de los ítems, así como la posibilidad de contestar correctamente un ítem de forma azarosa. Junto con estos elementos, la TRI permite conocer la estructura de un ítem a través de su Curva Característica. Además, determina la función de información, que permite valorar la precisión de la estimación que realiza el test.

Estos elementos son especialmente útiles en el ámbito de la educación y la personalidad, cuando se trabajan con pruebas objetivas de evaluación y diagnóstico. En estos campos es útil tener información tanto de la eficacia del instrumento de evaluación, como del resultado de la propia evaluación en las personas.

Referencias

Rizopoulos, D. (2006). ltm: An R package for latent variables modeling and item response theory analysis. *Journal of statistical software*, 17(5), 1-25. ([http://www.google.es/url?sa=t&source=web&cd=1&ved=0CBwQFjAA&url=http%3A%2F%2Frsirt.googlecode.com%2Ffiles%2FltmPackage.pdf&rct=j&q=Rizopoulos%2C%20D.%20\(2006\).%20ltm%3AAn%20R%20package%20for%20latent%20variables%20modelin%20and%20item%20reponse%20theory%20analysis.%20Journal%20of%20statistical%20software&ei=yLwQToTQBcnHsgbD_NzgDg&usq=AFQjCNHrwYYit6aDrNw-1_dWIKq6WinPeg&sig2=u0T7V1JMPjGes_SgKOq79Q&cad=rja](http://www.google.es/url?sa=t&source=web&cd=1&ved=0CBwQFjAA&url=http%3A%2F%2Frsirt.googlecode.com%2Ffiles%2FltmPackage.pdf&rct=j&q=Rizopoulos%2C%20D.%20(2006).%20ltm%3AAn%20R%20package%20for%20latent%20variables%20modelin%20and%20item%20reponse%20theory%20analysis.%20Journal%20of%20statistical%20software&ei=yLwQToTQBcnHsgbD_NzgDg&usq=AFQjCNHrwYYit6aDrNw-1_dWIKq6WinPeg&sig2=u0T7V1JMPjGes_SgKOq79Q&cad=rja))

R Development Core Team (2011). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Matas, A. (2001). *Medición y Diagnóstico. Introducción a la medición educativa aplicada al Diagnóstico en Educación*. Sevilla: Kronos.

Glosario

Puntuación verdadera: dentro de la teoría clásica de los test, es el valor del rasgo que se pretende medir realmente. Es la puntuación que se obtendría si no existiesen errores en la medida.

Dificultad de un ítem: es un indicador del nivel de exigencia de un ítem para que la persona consiga un determinado valor. Por ejemplo, un ítem de un examen es difícil si exige un elevado nivel de conocimiento de la materia para ser contestados correctamente.

Discriminación de un ítem: capacidad que tiene el ítem de diferenciar entre personas con distinto nivel de habilidad.

Fiabilidad: estabilidad o constancia de las medidas obtenidas con un mismo test. También indica la precisión o exactitud de las medidas del test.

Validez: indicador de que el test mide realmente aquello que pretende medir.

Homocedasticidad: igualdad de varianza (es una palabra que no se recoge en el diccionario de la Real Academia de la Lengua Española).

Variable independiente: variable (fenómeno o suceso) que controla el investigador con la intención de provocar una respuesta o un cambio en el sistema.

Variable dependiente: fenómeno que se pretende medir para identificar posibles cambios.

Curva característica del test (CCT): representación que muestra la relación entre la ejecución de una persona en un test y su nivel de habilidad.

- i http://www.uam.es/personal_pdi/psicologia/fjabad/cv/tai.html
- ii <http://hum369.uma.es>