

VENTAJAS DE LA UTILIZACIÓN DE LA TEORÍA DE RESPUESTA AL ÍTEM NO PARAMÉTRICA PARA LA EVALUACIÓN DE APRENDIZAJES

Mario Javier Luzardo Verde¹

Pilar Cecilia Rodríguez Morales²

Simposio Evaluación de la calidad de los sistemas educativos. Retos y perspectivas

Resumen

La Teoría de Respuesta al Ítem (TRI) surge en el ámbito de la evaluación de aprendizajes como forma de solucionar los problemas clásicos en medición educativa. El enfoque más popular de la TRI es mediante el uso de modelos paramétricos tales como los modelos logísticos de uno, dos y tres parámetros. Los principales cuestionamientos al enfoque paramétrico son que la dependencia entre la probabilidad y el rasgo es forzada a ser de determinada forma funcional y no se toma en cuenta hechos tales como la no monotonicidad de los ítems o desviaciones del formato impuesto. Por estos motivos se desarrollaron varios métodos basados en enfoques no paramétricos. Con el objetivo de estimar la curva característica del ítem (CCI) y estudiar el ajuste se utilizó el procedimiento de TRI no paramétrico desarrollado por Luzardo y Rodríguez (2015) y se aplicó a una prueba de evaluación de competencias en Matemática para estudiantes de ingreso a la Universidad (Rodríguez, 2016).

El método aplicado es una alternativa atractiva para estimar las CCI en casos en que estas deban ser crecientes, como por ejemplo en los contextos educativos. Es sencillo de programar y no iterativo. Esta es una ventaja importante respecto a los métodos paramétricos.

¹ Magister en Matemática por la Universidad de la República y Magister en Metodología de las Ciencias del Comportamiento y de la Salud por la Universidad Autónoma de Madrid. Profesor Agregado del Instituto de Fundamentos y Métodos de la Facultad de Psicología y del Departamento de Matemática y Aplicaciones del Centro Universitario Regional del Este, Universidad de la República, Uruguay. mluzardov@gmail.com.

² Doctora en Educación por la Universidad Nacional de Educación a Distancia, España. Profesora Adjunta, Coordinadora de la Unidad de Apoyo a la Enseñanza del Centro Universitario Regional del Este, Universidad de la República, Uruguay. prodriguez@cure.edu.uy. Expositora en el Simposio.

Introducción

Con el fin de optimizar la medición de constructos psicológicos o cognitivos y por lo tanto, mejorar la toma de decisiones y resolver ciertos problemas de medición, se ha pasado de la utilización de la Teoría Clásica de los Tests (TCT) a la Teoría de Respuesta al Ítem (TRI). La TRI surge para dar respuesta a los problemas centrales en evaluación de aprendizajes. La TRI toma los ítems como unidad de análisis, permite describir algunas propiedades psicométricas del instrumento mediante indicadores invariantes, es decir, que no dependen de la muestra en que se aplique. Esa se puede considerar como su mayor contribución *“la posibilidad de obtener mediciones invariantes respecto de los instrumentos utilizados y de los sujetos implicados”* (Muñiz, 1997).

La TRI propone soluciones para las limitaciones de la TCT como: la invarianza de los parámetros que permite que el valor de los parámetros de los ítems no depende de la muestra de donde se obtiene; la precisión con la que cada persona es medida según su nivel de rasgo y en función de los ítems concretos que se le hayan aplicado, y además, los indicadores de bondad de ajuste permiten estudiar el grado en que los datos ajustan al modelo (Abad et al., 2011).

La TRI se diferencia de la teoría clásica de los test por utilizar modelos basados en las características de los ítems en vez de las del test, donde las características de los ítems son independientes del grupo en que el ítem se ha calibrado y las puntuaciones del rasgo no dependen de las puntuaciones obtenidas en cada test particular. En la TRI se puede obtener una medida de la precisión para cada puntuación del rasgo, lo que la distingue claramente de la teoría clásica y para evaluar la fiabilidad no se requieren de tests estrictamente paralelos.

La teoría de respuesta al ítem establece una relación funcional entre la respuesta del examinado a cada ítem y el rasgo latente responsable de tal realización, al que se nota θ . En la mayoría de los modelos se asume que esta función depende solo de un rasgo, es decir, son unidimensionales. La función que da la probabilidad de obtener determinada puntuación en el ítem condicionado al rasgo, se denomina curva característica del ítem (CCI).

Las tres hipótesis que sustentan la TRI son: la dimensionalidad del espacio latente, la independencia local y la ausencia de factores de velocidad. La dimensionalidad del espacio latente supone que hay un conjunto de rasgos o aptitudes que subyacen en la respuesta de los individuos a un conjunto de ítems. Cada rasgo es un número real,

por lo tanto, si existen k rasgos, estos determinan puntos del espacio R^k . A los modelos que suponen que un único rasgo es el que determina las respuestas se les llama unidimensionales. Es claro que este supuesto nunca se cumplirá estrictamente, por lo tanto, lo que se pide es que para un conjunto de ítems del test exista un factor dominante que explique la realización del test, a este factor se le llama aptitud medida por el test. Los modelos que consideran más de un factor son llamados multidimensionales.

Si se cumple el supuesto de unidimensionalidad, al aplicar a r subpoblaciones de examinados un test, las distribuciones condicionadas de las puntuaciones del test para cada nivel de rasgo son iguales para todas las subpoblaciones. En caso de no cumplirse lo anterior estamos en presencia de funcionamiento diferencial del ítem.

La independencia local refiere a que los n ítems que constituyen el test, condicionados a un valor constante de θ son independientes. Es importante notar que la independencia local no implica que los ítems son no correlacionados, sino que son independientes condicionalmente (al nivel de rasgo). Hambleton y Swaminathan (1991) muestran que el principio de unidimensionalidad es equivalente a la independencia local (condicionado a solo un rasgo), pues si suponemos que el test es unidimensional y mide una variable latente θ , si no se da la independencia local resultaría que algunos sujetos tendrían mayor probabilidad de responder correctamente para el mismo nivel de rasgo, lo cual niega la unidimensionalidad. Recíprocamente, si se da la independencia local tenemos que las probabilidades de respuesta solo dependen de un único rasgo. La independencia local también puede expresarse con datos multidimensionales (condicionando al espacio completo).

En los modelos de TRI se supone que los tests se aplican en ausencia de factores de velocidad, es decir, que la puntuación que obtiene un sujeto en un ítem, se debe solo a su nivel de rasgo, pero no a falta de tiempo. Esta suposición está implícita en el supuesto de unidimensionalidad.

Desarrollo

Modelos no paramétricos basados en regresión con núcleos

El enfoque más popular para la TRI es mediante el uso de modelos paramétricos tales como los modelos logísticos de uno, dos y tres parámetros. La bibliografía sobre estos modelos es extensa, por ejemplo, Lord (1980); Hambleton, Swaminathan y Rogers (1991); Van der Linden y Hambleton (1997) y Boomsma, Van Duijn y Snijders (2001). Estos modelos determinan la forma de la CCI dependiendo de un número pequeño de

parámetros, pero no toman en cuenta la no monotonía ni apartamientos sistemáticos de la forma y son poco flexibles (Douglas, 1997; Douglas y Cohen, 2001; Ramsay, 1991).

Si en la estimación de modelos paramétricos la suposición de unidimensionalidad e independencia local son violadas, las estimaciones de los parámetros de los ítems y de la habilidad son pobres.

Uno de los principales cuestionamientos que se pueden hacer al enfoque paramétrico, es que la dependencia entre la probabilidad y el rasgo es forzada a ser de determinada forma funcional y no se toma en cuenta hechos tales como la no monoticidad de los ítems o desviaciones del formato impuesto. También está de por medio la complejidad de los cálculos, tanto si se emplea el paradigma de Birnbaum (máxima verosimilitud marginal con el algoritmo EM) como procedimientos bayesianos.

Existen varios métodos alternativos basados en enfoques no paramétricos por ejemplo el modelo monótono homogéneo y el modelo de doble monoticidad.

Una comparación entre los modelos paramétricos y no paramétricos se hace en Meijer, Sijtsma y Smid (1990), y en De Koning, Sijtsma y Hamers (2001). Ramsay (1991) presenta un método que estima las CCI mediante regresión no paramétrica con núcleos e implementada en el software TestGraf (Ramsay, 2000). Douglas (1997) provó la consistencia conjunta para el caso unidimensional; en tanto, Luzardo y Forteza (2014) prueban condiciones para la consistencia conjunta en el caso multidimensional. Este método tiene varias ventajas: la primera de ellas es la no imposición de ninguna restricción sobre la forma funcional de las CCI, y además, el método de estimación es no iterativo y muy fácil de programar.

Modelo Isótono

Se presenta un enfoque no paramétrico para estimar las CCI cuando estas deben ser monótonas. A los efectos de estimar funciones monótonas hay una gran variedad de métodos: por ejemplo, el propuesto por Wright (1981); Friedman y Tibshirani (1984) y Mammen (1991). Finalmente, el modelo que se utiliza para la estimación de la CCI es el propuesto por Luzardo y Rodríguez (2015) que está basado en el método desarrollado por Dette et al. (2006) que proponen estimar funciones monótonas a partir de estimadores no monótonos.

El proceso tiene dos etapas: la primera, utiliza el estimador de la CCI por medio de de regresión no paramétrica con núcleos, y la segunda, usa este resultado para estimar la función de densidad de la inversa de la CCI.

Integrando la función de densidad se obtiene un estimador isótono de la inversa de la CCI, luego se simetriza respecto a la bisectriz de cuadrado unitario para obtener el estimador de la CCI.

Aplicación del método no paramétrico basado en núcleos

Con la finalidad de estimar la curva característica del ítem y estudiar el ajuste al modelo logístico se utilizó el procedimiento de TRI no paramétrico desarrollado por Luzardo y Rodríguez (2015) y se aplicó a una prueba de evaluación de competencias en Matemática para estudiantes de ingreso a la Universidad (Rodríguez, 2016).

Para medir el ajuste de las personas usamos el estadístico L_0 de Levine y Rubin (1979) y su versión estandarizada L_z propuesta por Drasgow et al. (1985). A través de este análisis se encontró que el 96.45 % de los sujetos ajustaron bien.

Para el análisis de ajuste de los ítems se utilizó el Q_1 de Yen. Como este test tiene una distribución chi cuadrado que crece rápidamente y puede llevar a rechazar la hipótesis nula de ajuste, se decidió aplicar un método no paramétrico. Se utilizó el método propuesto por Luzardo y Rodríguez (2015).

En este gráfico se puede observar la curva estimada mediante el modelo 2P (en negro) y la curva isótona no paramétrica (en verde).

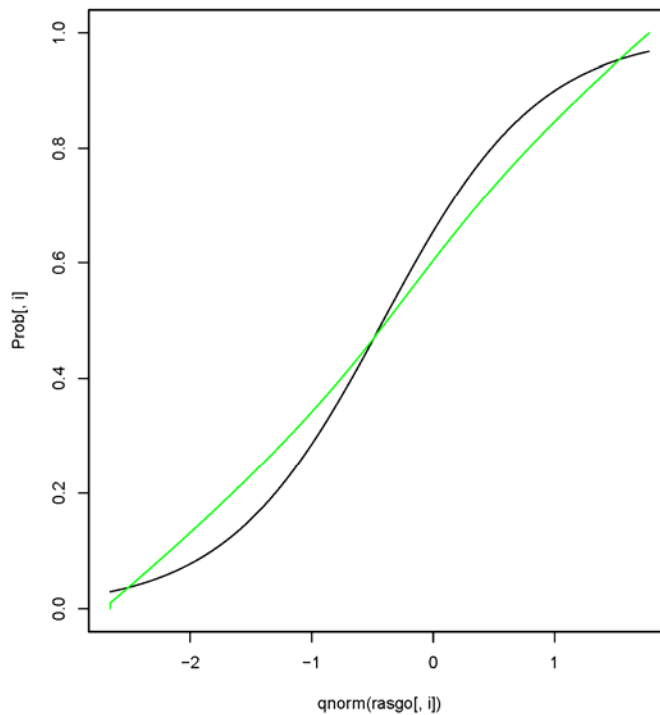


Figura 1: Gráfica de las curvas estimadas mediante el modelo 2P y el modelo isótono

Las distancias entre las dos curvas para cada ítem se muestran en la tabla siguiente donde se ve que los valores son pequeños con respecto a los aceptados mediante simulación por lo que mantenemos el modelo 2P de los ítems.

Conclusiones

El método no paramétrico de Luzardo y Rodríguez (2015) es una alternativa atractiva para estimar las CCI en casos en que estas deban ser crecientes, como por ejemplo en los contextos educativos. Es un método de mucha flexibilidad, sencillo de programar y no iterativo. Esta es una ventaja importante respecto a los métodos paramétricos, tanto cuando se trata de modelos unidimensionales, como en los modelos multidimensionales. Es justamente en el caso multidimensional donde aparece toda su potencia respecto a los métodos paramétricos, ya que la cantidad de iteraciones en estos últimos es elevada. También se obtienen funciones suaves a partir del método original, cosa que no es así en otros procedimientos no paramétricos. Se puede mencionar como limitante que al trabajar en el intervalo $[0;1]$ no obtenemos

la distribución original del rasgo y la estimación de las CCI en el dominio de la habilidad. Sin embargo, esta dificultad es aparente dado que las escalas son equivalentes ante transformaciones monótonas. Si conocemos la distribución del rasgo en el caso unidimensional o las marginales en el caso multidimensional se obtendrán las escalas originales. Otra característica importante es que la dimensión del rasgo latente debe ser reducida, ya que la cantidad de observaciones necesarias para una correcta estimación crece potencialmente con el número de dimensiones. Si bien la prueba de evaluación diagnóstica utilizada en este estudio presenta muy buenas propiedades psicométricas (Rodríguez, 2016) conviene seguir explorando la utilización de procedimientos no paramétricos, especialmente en constructos multidimensionales o en aquellos donde los ajustes con el modelo paramétrico no han sido tan buenos, como la prueba de evaluación diagnóstica implementada por Rodríguez (2016).

Bibliografía

- Abad, F., Olea, J., Ponsoda, V., García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.
- Boomsma, A., van Duijn, M., & Snijders, T. (Eds.). (2001). *Essays on item response theory* (Vol. 157). New York: Springer Science & Business Media.
- De Koning, E., Sijtsma, K., & Hamers, J. H. (2003). Construction and validation of test for inductive reasoning. *European Journal of Psychological Assessment*, 19(1), 24.
- Dette, H, Neumeyer, N. y Pilz, K. (2006) A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3), 469-490.
- Douglas, J. (1997): Joint Consistency of Nonparametric Item Characteristic Curve and Ability Estimation. *Psychometrika*, 62, 7-28.
- Douglas, J., Kim, H.R., Habing, B. y Gao, F. (1998): Investigating local dependence with conditional covariance functions. *Journal of Educational and Behavioral Statistics*, 23, 129-151.
- Douglas, J. y Cohen, A (2001): Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25, 234-243.
- Drasgow, F., Levine, M.V., Williams, B., McLaughlin, M.E., y Candell, G.L. (1989): Modeling incorrect responses to multiple-choice item with multilinear formula score theory. *Applied Psychological Measurement*, 13, 285-299.

- Friedman, J., & Tibshirani, R. (1984). The monotone smoothing of scatterplots. *Technometrics*, 26(3), 243-250.
- Hambleton, R.; Swaminathan, H. y Rogers, H. (1991). *Fundamentals of Item Response Theory*. Thousand Oaks: Sage Publications.
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational and Behavioral Statistics*, 4(4), 269-290.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York: Routledge.
- Luzardo, M. y Forteza, D. (2014). *Modelo no paramétrico multidimensional para la estimación de los rasgos y de las curvas características del ítem mediante regresión no paramétrica con núcleos*. Montevideo: CSIC.
- Luzardo, M., y Rodríguez, P. (2015). A Nonparametric Estimator of a Monotone Item Characteristic Curve. En L. A. van der Ark, D. Bolt, W.Ch. Wang, J. A. Douglas y S. M. Chow (Eds), *Quantitative Psychology Research* (pp. 99-108). Suiza: Springer International Publishing.
- Mammen, E. (1991). Estimating a smooth monotone regression function. *The Annals of Statistics*, 724-740.
- Meijer, R. R., Sijtsma, K., & Smid, N. G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 14(3), 283-298.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Ramsay, J. O. (1991): Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 611-630.
- Ramsay, J. O. (2000). Differential equation models for statistical functions. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 225-240.
- Rodríguez, P. (2016). *Creación y establecimiento de estándares para la evaluación de la calidad de la educación superior: un modelo adaptado a los centros universitarios regionales de la Udelar*. Tesis doctoral. Disponible en: <http://espacio.uned.es/fez/view/tesisuned:Educacion-Prrodriguez>
- Van der Linden, W. y Hambleton, R. (1997): *Handbook of Modern Item Response Theory*, New York, Springer-Verlag.