

LPCM-Win 1.0: programa para analizar modelos logísticos lineales de la familia de Rasch

Pedro M. Hontangas

Universitat de València

Descripción del programa

LPCM-Win 1.0 recoge las aportaciones derivadas de la investigación de Fisher y sus colaboradores durante los últimos 30 años sobre diversas extensiones del modelo de Rasch y sus aplicaciones a la medida del cambio. La descripción detallada del fundamento psicométrico de este conjunto de modelos se puede encontrar en las obras de Fischer y Molenaar (1995) y van der Linden y Hambleton (1997).

Se trata de una aplicación Windows escrita en lenguaje C++, que dispone de un interfaz gráfico de menús y cuadros de diálogo que hacen su manejo relativamente fácil, una vez se conocen las características y el fundamento psicométrico de los modelos. Esta versión del programa ha sido elaborada por Fischer y Ponocny-Seliger (1998) y es distribuida por la empresa holandesa ProGAMMA (<http://www.gamma.rug.nl>). Existe una versión previa para el sistema operativo DOS, denominada LPCM-Dos, elaborada por Seliger y Fisher (1994).

Finalidad

El objetivo del programa es fundamentalmente estimar los parámetros de un amplio conjunto de modelos de la Teoría de la Respuesta al ítem (TRI) pertenecientes a la familia de Rasch y analizar el cambio en los parámetros, producido por el efecto de diversos tratamientos en diseños de medidas repetidas. Una descripción más detallada de sus aplicaciones o de sus objetivos específicos requiere la presentación de los grupos de modelos y los tipos de diseños utilizados, que a continuación señalaremos.

Modelos

LPCM-Win 1.0 contiene varios tipos de modelos para ítems dicotómicos y politómicos, tanto unidimensionales como multidimensionales,

con o sin parámetros de efectos de tratamientos o de cambio. El más general es el modelo de crédito parcial lineal, que da nombre al programa (LPCM, Linear Partial Credit Model). El resto de los modelos se pueden considerar casos especiales de éste, sometido a diferentes restricciones. Los modelos básicos que se pueden estimar son: el *modelo de Rasch* (RM, Rasch model), el *modelo de escala de estimación* o clasificación (RSM, rating scale model), el *modelo de crédito parcial* (PCM, partial credit model) y el *modelo logístico lineal* (LLTM, linear logistic test model). Estos modelos tienen varias modalidades y extensiones.

En el *modelo de Rasch*, el *modelo de escala de estimación* y el *modelo de crédito parcial* se trata principalmente de estimar los parámetros de dificultad de los ítems (θ_i, w_i, w_{ik}) y la habilidad de los sujetos (θ_j):

A1. En el caso de ítems dicotómicos, el *modelo de Rasch* describe la probabilidad de responder correctamente ($x_{ij}=1$) un ítem I_i de dificultad θ_i por un sujeto S_j de habilidad θ_j mediante la expresión (1). Por tanto, un solo parámetro caracteriza a ítems y a sujetos respectivamente.

$$P(x_{ij} = 1 | I_i, S_j) = \frac{\exp(\theta_j - \theta_i)}{1 + \exp(\theta_j - \theta_i)} \quad (1)$$

B1. En el caso de ítems politómicos con m categorías de respuesta, el *modelo de escala de estimación* incorpora parámetros para describir el atractivo de las categorías (w_k), siendo los mismos para todos los ítems. La probabilidad de elegir la categoría k es definida en la expresión (2). Por su parte, el *modelo de crédito parcial* permite que los parámetros de las categorías de diferentes ítems sean distintos. La expresión (3) define la probabilidad de elegir la categoría k en este modelo.

$$P(x_{ij} = k | I_i, S_j) = \frac{\exp[k(\theta_j - \theta_i) + w_k]}{\sum_{h=0}^m \exp[h(\theta_j - \theta_i) + w_h]} \quad (2)$$

$$P(x_{ij} = k | I_i, S_j) = \frac{\exp[k(\theta_j - \theta_i) + w_{ik}]}{\sum_{h=0}^m \exp[h(\theta_j - \theta_i) + w_{ih}]} \quad (3)$$

El *modelo logístico lineal* parte de la estimación de los parámetros del modelo de Rasch y pretende: a) en su versión componencial, analizar los procesos o elementos responsables de las diferencias encontradas en la dificultad de los ítems; b) en su versión factorial, analizar la influencia de diversas condiciones y tratamientos en la dificultad que tienen los ítems para los grupos de sujetos que los han recibido; y c) en su versión del cambio, por un lado, analizar lo mismo que en b) cuando se dispone de medidas tomadas en varias ocasiones (o análisis del efecto en los parámetros de los ítems) y, por otro lado, analizar el cambio individual producido por estas condiciones y tratamientos (o análisis de la modificabilidad de la habilidad de los sujetos).

A2. El *modelo logístico lineal componencial* descompone el parámetro de dificultad θ_i de cada ítem, como se indica en la expresión (4), en una serie de 'parámetros básicos', β_q , que corresponden a los p procesos cognitivos, operaciones mentales o condiciones experimentales que intervienen en su resolución. Los parámetros básicos tienen pesos, w_{iq} , normalmente una variable dummy (0,1), que indican si el componente interviene ($w_{iq}=1$) o no ($w_{iq}=0$) y en qué grado en cada ítem. Además, se añade la constante c como factor de normalización de la escala. En consecuencia, la probabilidad de acertar el ítem se obtiene sustituyendo el valor θ_i de la expresión (1) por el contenido de la expresión (4).

$$\theta_i = \sum_{q=1}^p w_{iq} \beta_q + c \quad (4)$$

B2. Las demás las versiones del modelo se operacionalizan de una manera similar, pero se centran en estimar el efecto que tienen en los parámetros de ítems y sujetos distintos factores o tratamientos en uno o varios momentos temporales. En unos casos se aplican los mismos ítems en una sola ocasión a sujetos que han recibido distintas condiciones (v.gr., ítems en formato de lápiz y papel vs. ítems presentados por ordenador) o que pertenecen a diferentes grupos definidos por características socio-demográficas (v.gr., hombres vs. mujeres) o de otra índole. Esta versión también se conoce como *modelo de Rasch multifactorial o de múltiples facetes* (MRM, multifaceted Rasch model) y su planteamiento es semejante al ofrecido en el programa FACETS (Linacre, 1994); no obstante, LPCM-Win sólo permite controlar el análisis simultáneo de tres factores o facetes.

C2. En los análisis longitudinales se tienen varias medidas en diferentes ocasiones (v.gr., antes y después de una terapia; antes, durante y

después de un curso de formación). En este caso se le denomina *modelo logístico lineal para la medida del cambio*. El análisis se puede efectuar desde dos puntos de vista complementarios. El primero es el análisis del efecto en los parámetros de los ítems y se centra en las diferencias de dificultad que tienen los ítems para sujetos sometidos a distintas condiciones. Es posible analizar la existencia de efectos principales entre condiciones o tratamientos, su interacción y la tendencia temporal. El énfasis se pone en los ítems y en las semejanzas o diferencias entre grupos de sujetos. El segundo es el análisis de modificabilidad de los sujetos y se centra en el aumento o disminución de la habilidad estimada debida a los diferentes tratamientos. El énfasis se sitúa, pues, en el cambio individual experimentado por los sujetos. El programa permite formular y contrastar hipótesis jerárquicas sobre diferentes efectos en los modelos de las versiones B2 y C2.

Por otra parte, el *modelo logístico lineal* es aplicable tanto a ítems dicotómicos unidimensionales como a ítems politómicos multidimensionales.

A3. Desde el punto de vista de la dimensionalidad, el programa implementa el denominado *modelo logístico lineal con supuestos relajados* (LLRA, linear logistic model with relaxed assumptions), que no requiere el supuesto de unidimensionalidad del modelo de Rasch tradicional. Se trata de un modelo multidimensional en el que cada ítem puede medir un proceso, factor o rasgo diferente, pudiendo estar todos ellos relacionados o ser independientes. También es posible emplear una combinación de ítems unidimensionales y multidimensionales en el denominado *modelo logístico lineal híbrido* (Hybrid LLRA). Ambos modelos admiten cualquiera de las variantes descritas en el apartado C2.

B3. Desde el punto de vista del formato de respuesta, los modelos politómicos de escala de estimación y de crédito parcial han sido adaptados para analizar los componentes de la dificultad, dando lugar a los denominados *modelo de escala de estimación lineal* (LRSM, linear rating scale model) y *modelo de crédito parcial lineal* (LPCM, linear partial credit model). En este caso se procede de una forma similar a la descrita en A2, pero con los parámetros de las expresiones (2) y (3). A su vez, estos modelos han sido extendidos para analizar el cambio en las mismas condiciones que las variantes de los modelos dicotómicos unidimensionales, multidimensionales e híbridos recogidas en los apartados C2 y A3.

Finalmente, se encuentran los *modelos definidos por el usuario*. Esta opción, denominada 'customized models', es muy interesante ya que permite definir modelos no especificados previamente, estableciendo las correspondientes restricciones de tipo lineal en una matriz de diseño. Queda,

pues, abierta la posibilidad de implementar modelos ajustados a las necesidades de cada investigador.

Estimación de parámetros

El programa emplea el método de máxima verosimilitud condicional (CML) para estimar los parámetros de los ítems. Las funciones simétricas elementales son obtenidas combinando los denominados algoritmo suma y algoritmo diferencia. En la resolución de las ecuaciones no lineales usa una técnica Quasi-Newton basada en el algoritmo Broyden-Fletcher-Goldfarb-Shano y la regla longitud-paso Armijo. Los parámetros de los sujetos son estimados por el método de máxima verosimilitud (ML) y la aproximación numérica de Newton. Estos parámetros sólo pueden ser estimados en el modelo de Rasch, el modelo de escala de estimación, el modelo de crédito parcial y el modelo logístico lineal. El uso exclusivo de métodos de máxima verosimilitud implica que no se puede obtener estimaciones finitas para ítems acertados o fallados por todos los sujetos, ni para sujetos que aciertan o fallan todos los ítems.

Ajuste

Respecto al ajuste individual de ítems y sujetos, el programa ofrece los errores típicos de estimación de los parámetros, pero no proporciona tests de significación estadística del ajuste de cada ítem, ni de los patrones de respuesta de los sujetos (v.gr., chi-cuadrado, residuales). Respecto al ajuste global de los modelos, los índices se basan en el principio de invarianza de la TRI. El procedimiento consiste en dividir la muestra en dos submuestras y efectuar una calibración separada para cada una y otra para la muestra completa. El ajuste se obtiene mediante un test de razón de verosimilitud entre las estimaciones de las submuestras y la muestra completa (partes vs todo). En este sentido, una característica a destacar del programa es que ofrece un amplio número de opciones para dividir la muestra en base a criterios internos (v.gr., la media, la mediana, una puntuación total, las respuestas a un ítem) o externos (variables clasificatorias incluidas en el fichero de datos, como el sexo de los sujetos).

Otro aspecto interesante del programa es que tiene una prueba de significación estadística de la diferencia de parámetros entre dos submuestras de sujetos. La prueba es similar a la técnica propuesta por Lord (1980) para analizar el funcionamiento diferencial de los ítems (DIF). Esta información es útil para detectar ítems problemáticos y, si se utiliza un criterio externo al dividir la muestra, el análisis puede ser equivalente a un estudio de DIF. Por otra parte, el programa indica el número de iteraciones, permitiendo detectar problemas de convergencia, y el logaritmo de la verosimilitud del modelo, permitiendo comparar modelos alternativos. En muchos casos, es el propio

programa el que hace la comparación directamente. Por ejemplo, en el análisis de los componentes de la dificultad de los ítems, compara la verosimilitud de las estimaciones del modelo logístico lineal y la del modelo de Rasch para determinar el grado de ajuste del primero (la adecuación de los componentes propuestos para explicar la dificultad de los ítems). También se ofrece una representación gráfica en la que se puede visualizar la correspondencia entre los parámetros de dificultad estimados en diferentes calibraciones (v.gr., muestra 1 vs. muestra 2, modelo 1 vs. modelo 2).

Adicionalmente, el programa proporciona una serie de pruebas estadísticas para detectar el incumplimiento del supuesto de unidimensionalidad (test de Martin-Löf), la existencia de matrices de datos mal condicionadas y la utilización de matrices de diseño con problemas de rango.

En el contexto de los modelos de análisis del cambio, las hipótesis sobre los distintos efectos de factores y tratamientos se contrastan mediante tests de razón de verosimilitud condicional. El programa tiene prefijadas las cuatro hipótesis que suelen ser de mayor interés o las que son de uso más común, pero permite que el usuario pueda especificar otras hipótesis alternativas.

Documentación

Los materiales facilitados con el programa son muy claros y están muy bien estructurados. El manual del programa es excelente. El capítulo 1 describe las diferentes opciones disponibles y la forma de operar. En el resto (capítulos 2 a 12) se presenta la teoría psicométrica que subyace a cada grupo de modelos, la secuencia detallada de operaciones necesarias para estimarlos, incluyendo un ejemplo, y un breve comentario de los resultados. Además, el programa se distribuye con los ficheros necesarios para ejecutar ejemplos de prácticamente todo tipo de modelos.

Limitaciones

El número de sujetos es ilimitado, sin embargo, el número de ítems depende de la cantidad de parámetros a estimar, fijada en un máximo de 180. Esta circunstancia da lugar a que se puedan utilizar distinto número de ítems en cada tipo de modelo; por ejemplo, se pueden utilizar hasta 181 ítems en el modelo de Rasch, 85 ítems de 6 categorías en el modelo de escala de estimación o 36 ítems de 6 categorías en el modelo de crédito parcial.

Otras de las limitaciones detectadas son: a) la imposibilidad de escribir ficheros de instrucciones, encadenar análisis automáticamente o trabajar mediante procesamiento diferido o en 'batch' (es cierto que dispone de una utilidad para guardar las especificaciones de los análisis y usarlas posteriormente, pero no es suficiente, ya que al utilizarlas hay que confirmar

manualmente cada uno de los datos); b) la ausencia de estadísticos descriptivos sobre los ítems, los sujetos, los grupos y los momentos temporales, teniendo que utilizarse otros programas para obtener esta información; y c) el tratamiento de los datos faltantes, pues equipara los datos de diseños incompletos y la ausencia de respuestas, siendo su control responsabilidad del usuario.

Valoración general

La utilidad de LPCM-Win 1.0 para estimar el *modelo de Rasch*, el *modelo de escala de estimación* y el *modelo de crédito parcial* es sustancialmente inferior a la ofrecida por otros programas como, por ejemplo, BIGSTEP (Linacre y Wright, 1997), QUEST (Adams y Khoo, 1995), PARSCALE (Muraki y Bock, 1996) o BILOG-MG (Zimowski, Muraki, Mislevy, y Bock, 1996). Estos programas proporcionan al usuario mucha más información de las propiedades psicométricas de los ítems (v.gr., índices de ajuste, función de información, etc.) y disponen de un amplio número de opciones para controlar el proceso de estimación (v.gr, métodos de estimación, criterios utilizados en los algoritmos, etc.). Una afirmación similar puede hacerse respecto al *modelo de Rasch multifactorial o de múltiples facetas* en comparación con otros programas como, por ejemplo, FACETS (Linacre, 1997).

Ahora bien, el programa LPCM-Win 1.0 es especialmente útil, por un lado, en el análisis de los componentes de la dificultad de los ítems y, por otro lado, en el análisis del cambio. Aquí reside su contribución sustancial. En estos aspectos no existe software comercial alternativo que se le pueda comparar, por el momento.

REFERENCIAS

- Adams, R.J. y Khoo, S.T. (1995). *QUEST 2.1. The Interactive Rasch Test Analysis System*. Camberwell, Australian Council for Educational Research.
- Muraki, E. y Bock, R.D. (1996). *PARSCALE. IRT based test scoring and items analysis for graded open-ended exercises and performance tasks* (version 3). Chicago, IL: Scientific Software International.
- Fischer, G.H. y Molenaar, I.W. (1995). *Rasch models. Foundations, recent developments and applications*. New York: Springer-Verlag.
- Fisher, G.H. y Ponocny-Seliger, E. (1998). *Structural Rasch Modeling. Handbook of the usage of PLCM-WIN 1.0*. Groningen, the Netherlands: ProGAMMA.
- Linacre, J. (1997). *A user's guide to FACETS. Rasch model compute program*. Chicago, IL: MESA Press.
- Linacre, J. y Wright, B. (1997). *A user's guide to BIGSTEPS. Rasch model compute program*. Chicago, IL: MESA Press.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Seliger, E. y Fisher, G.H. (1994). *Program description - LRSMG, LRSM, LLTM, and LPCM, with applications to scale analysis and measuring change*. Vienna: Department of Psychology of the University of Vienna.
- van der Linden, W.J. y Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- Zimowski, M.F., Muraki, E., Mislevy, R.J. y Bock, R.D. (1996). *BILOG-MG. Multiple-group IRT analysis and test maintenance for binary items*. Chicago, IL: Scientific Software International.