

# **Una medida del funcionamiento diferencial del ítem obtenida a partir de los parámetros del modelo logístico ajustado. Su aplicación a un estudio con datos reales.**

Aguerri, María Ester, Galibert, María Silvia, Lozzia, Gabriela S. y Attorresi, Horacio.

Cita:

Aguerri, María Ester, Galibert, María Silvia, Lozzia, Gabriela S. y Attorresi, Horacio (2005). *Una medida del funcionamiento diferencial del ítem obtenida a partir de los parámetros del modelo logístico ajustado. Su aplicación a un estudio con datos reales. XII Jornadas de Investigación y Primer Encuentro de Investigadores en Psicología del Mercosur. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-051/406>

# UNA MEDIDA DEL FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM OBTENIDA A PARTIR DE LOS PARÁMETROS DEL MODELO LOGÍSTICO AJUSTADO. SU APLICACIÓN A UN ESTUDIO CON DATOS REALES.

Aguerri, María Ester; Galibert, María Silvia; Lozzia, Gabriela S.; Attorresi, Horacio  
Facultad de Psicología, Universidad de Buenos Aires.

## Resumen

En el estudio del funcionamiento diferencial del ítem (DIF) se aplican medidas que dan cuenta de su magnitud y sentido. Una medida de sencilla interpretación es el logaritmo de la razón común de las chances de Mantel-Haenszel (MH-LOR), que vale cero cuando el ítem no presenta funcionamiento diferencial, es positivo cuando el ítem favorece a uno de los grupos y negativo cuando favorece al otro. En este trabajo se analiza el DIF mediante MH-LOR y otra medida obtenida en el marco de la Teoría de Respuesta al Ítem según se ajuste el modelo logístico de uno, dos o tres parámetros sobre datos reales y se comparan los resultados tanto en cuanto a su magnitud como a su error estándar. El ajuste del modelo de un parámetro condujo a resultados similares a MH-LOR, pues le corresponde la menor suma de los cuadrados de las diferencias y la mayor correlación. Pero el resultado más destacable es en cuanto a la similitud de los respectivos errores estándar; ya que la suma de los cuadrados de las diferencias es casi cero y la correlación es marcadamente más alta que cuando se considera el modelo de dos o tres parámetros.

## Palabras Clave

DIF -MH-LOR -MODELO LOGÍSTICO

## Abstract

A MEASURE OF DIFFERENTIAL ITEM FUNCTIONING OBTAINED FROM THE PARAMETERS OF THE FITTED LOGISTIC MODEL. ITS APPLICATION TO A STUDY WITH REAL DATA.

In differential item functioning analysis, some measures are applied in order to express the magnitude and sense of DIF. One of them, which has a simple interpretation, is the Mantel-Haenszel Log Odds Ratio, MH-LOR, whose value is zero when the item does not have a DIF, positive when it benefits one of two groups or negative when it benefits the other one. In this work DIF is analyzed by means of MH-LOR and another measure obtained within the frame of the Item Response Theory when one, two or three-parameters logistic models are fitted to real data. These results are compared with respect to their magnitudes as well as their standard errors. When the one-parameter logistic model was fitted, the results were similar to MH-LOR; in fact the sum of the squared differences was the smallest and the correlation was the highest. But the most meaningful result was referred to the similitude of their respective standard errors because the sum of squared differences is almost zero and the correlation is higher than those that are obtained when the two or three-parameters logistic model are considered.

## Key words

DIF- MH-LOR -LOGISTIC

## INTRODUCCIÓN

En la evaluación psicológica y educativa cuando ante un ítem, que dice medir una capacidad, se detecta que sujetos igualmente capaces pero pertenecientes a diferentes grupos no tienen la misma posibilidad de responderlo correctamente, se está en presencia de un ítem que no mide lo que afirma medir. La cuestión planteada se encuentra en el campo de los estudios de validez dentro de la Teoría Clásica de Tests y puede analizarse con procedimientos adecuados desde la Teoría de Respuesta al Ítem (TRI). En este marco se dice que tal ítem presenta funcionamiento diferencial (Differential Item Functioning, DIF).

Para la detección del DIF se utilizan medidas y pruebas de hipótesis. Camilli y Shepard (1994) distinguen dos enfoques, los llamados métodos de Tablas de Contingencia, haciendo alusión a la forma en que se pueden presentar los datos, y los métodos que suponen la aplicación de algún modelo de la Teoría de Respuesta al Ítem.

El logaritmo de la razón común de las chances de Mantel-Haenszel (Mantel-Haenszel Log Odds Ratio, MH-LOR) es una medida del DIF considerada entre los métodos de Tablas de Contingencia. La razón común de las chances es una medida de asociación para tablas del tipo  $2 \times 2 \times q$  presentada por Mantel y Haenszel (1959) y aplicada al estudio del DIF por Holland y Thayer (1988). En la literatura psicométrica acerca del DIF un grupo suele denominarse grupo de referencia y el otro grupo focal. Un ítem no tiene DIF cuando, para todo nivel  $j$  del puntaje total, la chance de contestar correctamente el ítem (cociente entre la cantidad de sujetos que acierta el ítem y la que lo falla) en el grupo de referencia es igual a la que se verifica en el grupo focal, luego la razón de las chances es 1 para todo nivel  $j$  del puntaje total y también lo es la razón común por lo que en ese caso MH-LOR es cero; si el grupo de referencia tiene ventaja respecto del grupo focal MH-LOR resulta positivo y si el aventajado es el grupo focal MH-LOR es negativo.

A partir de MH-LOR se obtiene una medida del DIF muy difundida, el delta-DIF de Mantel Haenszel (MH D-DIF) que se obtiene como:

$$\text{MH D-DIF} = -2.35 * \text{MH-LOR}$$

Una de las razones de su amplia difusión posiblemente sea que el Educational Testing Service utiliza una clasificación de los ítems según su DIF sobre la base de la magnitud y significancia del estadístico MH D-DIF. Tal clasificación es aconsejada por Fidalgo y Ferreres (2002, p.494) como guía de aceptación de la significación estadística en el estudio del DIF. Roussos y Stout (1996) y Aguerri, Galibert, Zanelli, y Attorresi (2005), entre otras investigaciones, analizan sobre datos simulados, las ventajas y limitaciones del estudio del DIF mediante el estadístico MH D-DIF; Ferreres, González-Romá, y Gómez (2002) y Liu y Feigenbaum (2004), entre otros, lo aplican sobre datos reales. Fidalgo (1996) haciendo referencia al procedimiento Mantel-Haenszel, dentro del cual se encuentra MH D-DIF, afirma que se comporta de manera óptima cuando los ítems se ajustan al modelo de un parámetro pero sigue dando buenos resultados aún cuando eso no ocurra. Donoghue, Holland y Thayer (1993) afirman que MH D-DIF estima a una cantidad que es proporcional a la diferencia de los parámetros de dificultad del ítem y analizan el efecto de la violación de los

supuestos requeridos. Sus estudios se basan sobre conjuntos de datos, tal que el parámetro de discriminación es constante dentro de cada conjunto de datos aunque entre los conjuntos de datos sí varía.

En este trabajo, desarrollado en el marco de una investigación basada en estudios de simulación, se analiza el funcionamiento diferencial de ítemes de razonamiento verbal mediante la medida MH-LOR y otra obtenida en el marco de la Teoría de Respuesta al Ítem según se ajuste el modelo de uno, dos o tres parámetros sobre datos reales y se comparan los resultados tanto en cuanto a su magnitud como a su error estándar.

## METODOLOGÍA

### Muestra

La muestra se compone de 349 egresantes del ciclo medio de enseñanza de la Ciudad de Buenos Aires y 865 ingresantes a la carrera de Psicología de la Universidad de Buenos Aires (Galibert, 2000).

### Instrumento

La prueba de razonamiento verbal, integrada por 20 ítemes, mide la habilidad para identificar y discriminar relaciones (Attorresi; Pano; Fernández Liporace y Cayssials, 1994). Los ítemes presentan un par de palabras base entre las cuales existe algún tipo de relación y cuatro opciones de pares de palabras donde se debe elegir aquél cuya relación es la más cercana a la del par base.

### Medidas del DIF

Camilli y Shepard (1994, p.116) presentan la expresión para el cálculo de MH-LOR. En ella intervienen, para cada nivel del puntaje total, la cantidad de sujetos del grupo de referencia que respondió correctamente el ítem, la cantidad de sujetos del mismo grupo que respondió incorrectamente el ítem, las respectivas cantidades en el grupo focal y el total de sujetos con dicho nivel del puntaje total.

En el marco de la Teoría de Respuesta al Ítem se presenta una medida del DIF que resulta de la diferencia de los parámetros de dificultad. En base a tal diferencia puede calcularse  $1.7*a*(bF - bR)$ , donde  $a$  es el parámetro de discriminación del ítem en ambos grupos, y  $bF$  y  $bR$  son los parámetros de dificultad del ítem en el grupo focal y en el grupo de referencia respectivamente. A esta medida del DIF la llamamos LOR-TRI. Camilli y Shepard (1994, p.118) muestran que  $1.7*a*(bF - bR)$  coincide con MH-LOR cuando se ajusta el modelo de dos parámetros con el mismo parámetro de discriminación en ambos grupos. En ese caso la expresión es equivalente, salvo la constante, a la que presentan Donoghue, Holland y Thayer (1993, p.141) para MH D-DIF, pero no lo es cuando se ajusta el modelo de dos parámetros y el parámetro de discriminación difiere entre grupos.

El programa BILOG-MGTM (Zimowski, Muraki, Mislav, y Bock, 1996) permite ajustar el modelo logístico de tres parámetros de modo que para cada ítem el parámetro  $c$ , de aciertos por azar, es el mismo para los dos grupos así como también es igual la potencia discriminatoria del ítem en los dos grupos, es decir:  $cR = cF$  y  $aR = aF$ . El programa proporciona estimaciones de  $bF$  y  $bR$ , y la diferencia de los mismos con su respectivo error estándar. En el caso de ajustar el modelo de dos parámetros  $cR = cF = 0$  y en el un parámetro se agrega la condición de que todos los ítemes tienen igual potencia discriminatoria, es decir  $aR = aF = a$ .

### Procedimiento

El estudio se realizó sobre 18 ítemes pues los ítemes 14 y 19, que presentaban correlación biserial negativa, fueron excluidos del análisis para favorecer el supuesto de unidimensionalidad de los modelos de la TRI por ajustar. Mediante el programa EZDIF de Waller (1998) se registró para cada ítem la clasificación del ETS según su DIF, MH D-DIF y su error estándar. Dividiendo estos últimos por  $-2.35$  y  $2.35$  respectivamente se obtuvieron MH-LOR y su error estándar. Con el programa BILOG-MG TM (Zimowski, Muraki, Mislav, y

Bock, 1996) se registraron, para cada ítem, LOR-TRI1, LOR-TRI2 y LOR-TRI3 y sus respectivos errores estándar según se consideró el modelo de uno, dos y tres parámetros.

## RESULTADOS

A partir del valor de MH-LOR quince ítemes fueron clasificados libres de DIF y los ítemes 4, 13 y 18, siguiendo la clasificación del ETS, fueron señalados sospechosos de DIF. Los dos primeros a favor del grupo de los ingresantes a la Facultad de Psicología y el último a favor de los egresantes del ciclo medio. A partir de los valores de LOR-TRI1 dieciséis ítemes fueron señalados sin DIF, sólo el ítem 4 y el ítem 18 presentaron valores como para ser categorizados como sospechosos de DIF en el mismo sentido que MH-LOR. Según el valor de LOR-TRI2 diecisiete ítemes fueron señalados sin DIF, sólo el ítem 18 fue señalado como sospechoso en el mismo sentido que lo hizo MH-LOR. Al considerar el valor LOR-TRI3 quince ítemes fueron señalados sin DIF, el ítem 4 resultó sospechoso de DIF a favor de los ingresantes a la Facultad de Psicología y los ítemes 12 y 18 sospechosos de DIF a favor de los egresantes del ciclo medio. Por tanto el porcentaje de coincidencias en la clasificación de los ítemes según su DIF entre MH-LOR y los diferentes LOR-TRI es 94.44% para el ajuste del modelo de un parámetro y 88.88% para el de dos y el de tres parámetros.

Para estudiar la similitud de las medidas se calcularon las sumas de los cuadrados de las diferencias entre MH D-DIF y LOR-TRI para cada uno de los modelos, y se obtuvo 0.0181, 0.1282 y 0.1417 para el de uno, dos y tres parámetros respectivamente. Tales sumas para los respectivos errores estándar fueron 0.0006, 0.5710 y 1.0508.

La correlación obtenida entre MH-LOR vs. LOR-TRI1, MH-LOR vs. LOR-TRI2, y MH-LOR vs. LOR-TRI3 fue 0.9965, 0.9901 y 0.9917, y entre los respectivos errores estándar fue 0.9813, 0.7627 y 0.7515.

## CONCLUSIONES

La medida del DIF obtenida en el marco de la Teoría de Respuesta al Ítem al ajustar el modelo de un parámetro condujo a resultados similares a MH-LOR (coinciden en el 94.44% de las decisiones), le corresponde la menor suma de los cuadrados de las diferencias y la mayor correlación. Pero el resultado más destacable es en cuanto a la similitud de los respectivos errores estándar ya que la suma de los cuadrados de las diferencias es casi cero y la correlación es marcadamente más alta que cuando se considera el modelo de dos o tres parámetros. Los ítemes de esta prueba, por presentar cuatro alternativas de las cuales sólo una es la correcta, son susceptibles de ser modelizados con el modelo logístico de tres parámetros (con parámetro de aciertos por azar no nulo) y sin embargo los resultados en magnitud y error estándar de MH-LOR se asemejan a los de LOR-TRI1. Si estos resultados se sostienen en los estudios de simulación que se están llevando a cabo, significaría que basta con ajustar el modelo de un parámetro (aunque el modelo adecuado sea el de tres) para obtener una medida del DIF similar a MH-LOR.

## BIBLIOGRAFÍA

- Aguerri, M. E.; Galibert, M. S.; Zanelli, M. L. y Attorresi, H. F. (2005) Detección errónea del Funcionamiento Diferencial del Ítem. Una comparación de métodos. *Psicothema*, 17, 2, 335-340. En prensa
- Attorresi, H.; Pano, C.; Fernández Liporace, M. y Cayssials, A. (1994) Evaluación de la habilidad para identificar y discriminar relaciones. *Anuario de Investigaciones N° 3*. Facultad de Psicología, UBA. pp. 27-34.
- Camilli, G. y Shepard, L. (1994) *Methods for identifying biased test items*. Thousand Oaks: Sage.
- Donoghue, J. R.; Holland, W. P. y Thayer, D. T. (1993) A Monte Carlo study of factors that affect the Mantel-Haenszel and standardization measures of differential item functioning. En P. W. Holland y H. Wainer (Eds.) *Differential Item Functioning*. (pp. 137-166) Hillsdale, NJ: Erlbaum

- Ferreres, D., González-Romá, V. y Gómez, J. (2002) Funcionamiento diferencial de los ítems en una situación de contacto de lenguas. *Psicothema*, 14, 2, 483-490.
- Fidalgo, A. (1996) Funcionamiento diferencial de los ítems, en J. Muñiz (coord.): *Psicometría*, Madrid, Universitas
- Fidalgo, A. y Ferreres, D. (2002) Supuestos y consideraciones en los estudios empíricos sobre el funcionamiento diferencial de los ítems. *Psicothema*, 14, 2, 491-496.
- Galibert, M. S. (2000) *Modelización psicométrica de un test de razonamiento verbal en los marcos de la Teoría Clásica de Tests y de la Teoría de Respuesta al Ítem*. Tesis de Magister Scientiae en Biometría. Universidad de Buenos Aires. Inédito.
- Holland, P. W. y Thayer, D. T. (1988) Differential item functioning and the Mantel-Haenszel procedure. En H. Wainer & H.I. Braun (Eds.), *Test Validity* (pp. 129 -145) Hillsdale, NJ: Lawrence Erlbaum.
- Liu, J. y Feigenbaum, M. D. (2004) Prototype analysis of Spring 2003 New SAT® Field Trial. Ponencia presentada en el encuentro anual del National Council on Measurement in Education, San Diego. Memorandum inédito. Princeton, NJ: Educational Testing Service
- Mantel N. y Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Roussos, L. & Stout, W. (1996) Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33, 2, 215-230.
- Waller, N. G. (1998). EZDIF: Detection of Uniform and Nonuniform Differential Item Functioning with Mantel-Haenszel and Logistic Regression Procedures. *Applied Psychological Measurement*, 22, 2, 391.
- Zimowski, M., Muraki, E., Mislevy, R. y Bock, R. (1996). *BILOG-MGTM: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. [Computer program] Scientific Software International, Inc.