XIII Jornadas de Investigación y Segundo Encuentro de Investigadores en Psicología del Mercosur. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires, 2006.

Análisis del DIF en una Escala de Voluntad de Trabajo mediante los procedimientos de Mantel-Haenszel y Breslow-Day, regresión logística y el criterio ETS.

Galibert, María Silvia, Aguerri, María Ester, Lozzia, Gabriela y Abal, Facundo Juan Pablo.

Cita:

Galibert, María Silvia, Aguerri, María Ester, Lozzia, Gabriela y Abal, Facundo Juan Pablo (2006). *Análisis del DIF en una Escala de Voluntad de Trabajo mediante los procedimientos de Mantel-Haenszel y Breslow-Day, regresión logística y el criterio ETS. XIII Jornadas de Investigación y Segundo Encuentro de Investigadores en Psicología del Mercosur. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-039/551>

ARK: <https://n2t.net/ark:/13683/e4go/Nqx>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

ANÁLISIS DEL DIF EN UNA ESCALA DE VOLUNTAD DE TRABAJO MEDIANTE LOS PROCEDIMIENTOS DE MANTEL-HAENSZEL Y BRESLOW-DAY, REGRESIÓN LOGÍSTICA Y EL CRITERIO ETS

Galibert, María Silvia; Aguerri, María Ester; Lozzia, Gabriela; Abal, Facundo Juan Pablo
UBACyT, Universidad de Buenos Aires - CONICET - ANPCyT PICT

RESUMEN

Se estudia el funcionamiento diferencial de los ítems (DIF), según el sexo, en una escala de voluntad de trabajo que se halla en elaboración. El análisis del DIF se lleva a cabo mediante los procedimientos de Mantel-Haenszel y de Breslow-Day, la regresión logística y el criterio de clasificación del Educational Testing Service (ETS). Los datos fueron recogidos de dos muestras de estudiantes de la Facultad de Psicología, Universidad de Buenos Aires, cuyos resultados se comparan. En una muestra, un ítem fue detectado con DIF por el procedimiento de Mantel-Haenszel además de ser clasificado como sospechoso de DIF por el criterio ETS y otro ítem exhibió DIF no uniforme por el análisis de regresión logística pero este resultado no fue sostenido por el test de Breslow-Day. En la otra muestra, tres ítems resultaron con una magnitud de DIF intermedia pero no estadísticamente significativa. Del estudio comparativo entre métodos y muestras se concluye que en la escala no hay suficiente evidencia de ítems con DIF según el sexo. Sin embargo, convendría seguir analizando aquéllos que en algún momento fueron señalados como tales para descartar cualquier posible sesgo y administrarlos junto con un mayor número de ítems para obtener resultados más confiables.

Palabras clave

DIF Voluntad de trabajo Mantel-Haenszel Breslow-Day

ABSTRACT

DIF ANALYSIS IN A WILL-TO-WORK SCALE BY MEANS OF THE MANTEL-HAENSZEL AND BRESLOW-DAY PROCEDURES, LOGISTIC REGRESSION AND THE ETS CRITERION

Differential Item Functioning (DIF) according to genders is analyzed in a will-to-work scale that is being built. DIF analysis is carried out by means of the Mantel-Haenszel procedure, the Breslow-Day test, logistic regression and the Educational Testing Service (ETS) criterion. The data were gathered from two samples of students of the Psychology School, University of Buenos Aires. The results obtained from both samples are compared. In one of them, one item was detected as having DIF by the Mantel-Haenszel procedure and it was also classified as suspected of DIF by the ETS criterion; another one displayed nonuniform DIF when applying logistic regression but this result was not supported by the Breslow-Day test. In the other sample, three items had a medium magnitude of DIF but it was not statistically significant. From the comparative study among methods and samples it is followed that there is no strong evidence of DIF between genders. However, in order to discard any potential bias, it is advisable to keep under study the items pointed out with possible DIF. They should be administrated with a larger number of items so as to obtain more reliable results.

Key words

DIF Will to work Mantel-Haenszel Breslow-Day

INTRODUCCIÓN

El análisis del Funcionamiento Diferencial de los Ítems (DIF) forma parte de los procesos de construcción y validación de una escala. Existe DIF cuando sujetos de igual nivel en el rasgo que se desea medir, por ejemplo una habilidad, tienen distinta probabilidad de responder correctamente al ítem según el grupo al que pertenecen. Si un ítem tiene DIF puede haber alguna característica asociada al grupo de pertenencia que afecta la respuesta al ítem y es diferente de la variable que se desea medir, con la cual se confunde. En otras palabras, puede manifestar una falta de unidimensionalidad que afecta a la validez.

Se detecta DIF a través de diversos métodos de contraste de hipótesis estadísticas. Básicamente, consiste en el rechazo de la hipótesis nula que postula su ausencia. También existen medidas que expresan su magnitud. Fidalgo y Ferreres (2002, p. 493) aconsejan la utilización de las medidas del DIF como guía de aceptación de la significación estadística.

Cuando se detecta DIF es de interés hallar una razón que lo explique; es decir, interpretarlo en términos de sesgo. Esto es importante principalmente para identificar posibles fuentes de sesgo. En ese caso el ítem es excluido de la escala. Cuando no se halla una explicación clara, puede sospecharse que se trata de una falsa detección; es decir, un resultado debido al azar que lleva a cometer error de tipo I. Estudios de simulación han mostrado que algunos contrastes de hipótesis tienen una tasa de error de tipo I superior a la nominal (5%). En particular esto ocurre con la prueba χ^2 de Mantel-Haenszel, presentada por Mantel y Haenszel (1959) y aplicada al estudio del DIF por Holland y Thayer (1988). Por ello, antes de descartar el ítem con DIF es preferible, de ser posible, estudiar su comportamiento en sucesivas administraciones.

El DIF se clasifica en uniforme y no uniforme según que un grupo se vea favorecido por el ítem para todo nivel del rasgo en la misma proporción o bien que dicha proporción varíe según los niveles. En el primer caso no hay interacción entre el grupo de pertenencia y la habilidad y en el segundo sí la hay. Camilli y Shepard (1994) y Holland y Wainer (1993) presentan diversos métodos para detectar la presencia de DIF uniforme y no uniforme.

Para la medición de actitudes y preferencias se suelen utilizar escalas tipo Likert. Si bien en estos casos no hay "respuestas correctas", sí se puede hablar de respuestas que están en la dirección de un mayor o menor nivel del rasgo que se desea medir. Si éstas se dicotomizan asignando 1 y 0 respectivamente, se pueden utilizar los modelos antedichos para el análisis del DIF. El 1 resume entonces el conjunto de respuestas asociadas con un mayor nivel en el rasgo medido y el 0 a las asociadas a un menor nivel.

En este trabajo se analiza el funcionamiento diferencial entre sexos de los ítems de una escala de Voluntad de Trabajo durante el proceso de elaboración de la misma. Este constructo se ha caracterizado como la actitud por la cual una persona se aboca a sus obligaciones llevándolas a cabo con responsabilidad, precisión, sin dilaciones y motivada por su propio deseo

de realizar un buen trabajo. Se comparan los resultados obtenidos en dos administraciones de la escala que tuvieron lugar durante su proceso de construcción, todavía no finalizado.

MÉTODO

Participantes

Estudiantes de segundo año de Psicología de la Universidad de Buenos Aires.

Administración Año 2003

Tamaño de muestra: 1141

Composición por sexo: 83% mujeres y 17% varones.

Composición por edad: 68% entre 18 y 21 años, 23% entre 22 y 28 y 9% a partir de 29.

Administración Año 2005

Tamaño de muestra: 834

Composición por sexo: 79% mujeres y 21% varones.

Composición por edad: 71% entre 18 y 21 años, 27% entre 22 y 28 y 2% a partir de 29.

Instrumentos

El protocolo administrado en 2003 consiste de un cuestionario con un número inicial de 21 enunciados de cuatro opciones al modo de las escalas Likert. Las opciones hacen referencia a las frecuencias con que se presentan ciertas conductas donde se manifiestan las características del constructo Voluntad de Trabajo. Para la administración 2003 las opciones fueron *casi nunca, a veces, con frecuencia y casi siempre*. En la administración del 2005 se agregó una categoría central: *casi nunca, pocas veces, a veces, con frecuencia y casi siempre*. El cuestionario también contiene datos personales como edad, sexo y condición laboral. No se registró la identidad de los sujetos con el fin de favorecer una mayor sinceridad. Después del análisis clásico de ítems quedaron seleccionados 9 enunciados que volvieron a ser administrados en 2005 junto con otros ítems para ampliar la escala. Es sobre estos nueve que se lleva a cabo el análisis comparativo del DIF.

METODOLOGÍA

Se aplicó el procedimiento de Mantel-Haenszel, la regresión logística y la prueba de Breslow-Day, estos métodos proporcionan un test de significación estadística sobre el cual basar la decisión. El primero permite decidir acerca de la presencia o ausencia de DIF, el segundo permite además decir si el DIF es uniforme o no uniforme mientras que el último se aplica para la detección del DIF no uniforme. También se consideró la clasificación del ítem según su DIF utilizada por el Educational Testing Service (ETS) sobre la base de la magnitud y significancia del estadístico MH D-DIF, definido por Holland y Thayer (1988), y la clasificación basada en la sola magnitud del estadístico MH D-DIF. El ETS considera los siguientes tipos de ítems, propuestos por Zieky (1993), A: con DIF muy pequeño, B: con DIF intermedio (o sospechoso de DIF), o C: con DIF grande. Un ítem es clasificado 'tipo C' si el valor absoluto del estadístico MH D-DIF es mayor o igual que 1.5 y significativamente mayor que 1, y clasificado como 'tipo A' si MH D-DIF es menor que 1 o no es significativamente diferente de cero. El resto de los ítems son clasificados 'tipo B'.

Estos procedimientos requieren que los ítems tomen sólo los valores 0 o 1; por lo que hubo que dicotomizarlos. A las dos categorías que expresaban un mayor grado de voluntad de trabajo del cuestionario de 2003 se les asignó 1 y a las otras el 0. En la administración de 2005 se presentó la dificultad de una clase central que hubo que determinar si se le asignaba 1 o 0. A fin de que los resultados de ambas administraciones fueran comparables se adoptó como criterio que las distribuciones de frecuencias de las variables dicotomizadas fueran lo más similares posibles entre las dos administraciones. Examinando las mismas se observó que esto se lograba asignando 1 a la clase central.

Se utilizó el programa EZDIF (Waller, 1998) que proporciona el valor del estadístico χ^2 de Mantel-Haenszel y su probabilidad

asociada, el valor del estadístico MH D-DIF, la clasificación del ítem según el ETS y el análisis de regresión logística y el SAS (Statistical Analysis System, 1989) para obtener el valor del estadístico de la prueba de Breslow-Day y su probabilidad asociada.

RESULTADOS

En la administración del año 2003 el ítem 26: "*Me dejo estar en el cumplimiento de mis deberes*" fue señalado como sospechoso de DIF a favor de las mujeres, según el criterio del ETS y el procedimiento Mantel-Haenszel. En efecto, habiéndose rechazado la hipótesis nula de la prueba χ^2 de Mantel-Haenszel ($p=0,036$) fue clasificado como B, con una medida MH D-DIF = -1,56. No fue igualmente detectado por el método de la regresión logística ni la prueba de Breslow-Day.

Si bien el ítem 32: "*Aunque me sienta cansado termino con la tarea que me impuse*" no fue señalado como con DIF al 5% por la prueba χ^2 de Mantel-Haenszel, estuvo cerca del límite de rechazo, pues $p = 0.064$, esto también ocurrió con la prueba de Breslow-Day ($p=0.0505$) pero fue detectado con DIF no uniforme por la regresión logística pues la interacción entre Voluntad de Trabajo y Sexo resultó significativa ($p=0.003$).

En cuanto al ítem 24: "*Cuando iba a la escuela estudiaba como máximo lo necesario para estudiar*" fue señalado con DIF no uniforme por el análisis de regresión logística por cuanto la interacción entre Voluntad de Trabajo y Sexo resultó significativa ($p=0.000$), este resultado no fue sostenido por la prueba de Breslow-Day, además fue considerado sin DIF a partir del procedimiento de Mantel-Haenszel y el criterio del ETS.

En la administración del 2003, los métodos empleados y la clasificación del ETS coincidieron en considerar libres de DIF a los seis ítems restantes.

Para los datos recogidos en la administración de 2005 no se detectó DIF al 5% en ninguno de los ítems con la prueba χ^2 de Mantel-Haenszel ni con la clasificación del ETS ni DIF no uniforme con la prueba de Breslow-Day; pero el mencionado ítem 32 tiene una magnitud de DIF a favor de los varones (MH-D-DIF = 1.227) que lo posicionaría entre los sospechosos aunque no fue significativa al 5%. Lo mismo ocurre con el ítem 34: "*Doy muchas vueltas antes de ponerme a trabajar*" (MH-D-DIF = 1.114) y con el ítem 20: "*Prefiero un menor esfuerzo a un mejor resultado*" (MH-D-DIF = -1.171), en este caso a favor de las mujeres. Según el análisis de regresión logística el ítem 15: "*Prefiero aquellas actividades que no demandan mayor esfuerzo*" presenta DIF no uniforme pues la interacción entre Voluntad de Trabajo y Sexo resultó significativa ($p=0.006$) y el resto de los ítems estarían libres de DIF.

DISCUSIÓN

Del análisis comparativo entre métodos y muestras no ha surgido ningún ítem que exhibiera fehacientemente DIF. De hecho, en la segunda administración todos los ítems fueron clasificados tipo A, según el criterio del ETS; es decir, sin DIF. El único ítem con potencial DIF que muestra cierta estabilidad en ambas muestras, es el ítem 32 cuyo valor p de significación lo posiciona próximo a ser detectado con DIF, del tipo no uniforme según lo obtenido en el análisis de regresión logística y la prueba de Breslow-Day, y en la segunda administración tiene una magnitud MH-D-DIF intermedia pero no de significación estadística.

Posibles causas de las diferencias en la detecciones de DIF entre ambas administraciones pueden ser, por un lado, el efecto de haber tenido que dicotomizar a partir de distinto número de categorías. Los sujetos de 2005 correspondientes a la clase central, que fueron asignados al 1 podrían haber estado repartidos entre el 0 y el 1 de no haber existido esa clase central. Por otra parte, los métodos de detección pueden no resultar estables cuando el conjunto de ítems es pequeño.

Convendría, pues, antes de aventurarse a una explicación del DIF o a proceder a descartar los ítems donde alguna vez se

detectó, volver a administrarlos junto con un conjunto mayor de ítems y utilizando el mismo número de categorías de la escala para verificar si se reiteran o no algunos de estos resultados. Entre tanto, puede sostenerse la unidimensionalidad de la escala de Voluntad de Trabajo respecto del sexo.

BIBLIOGRAFÍA

- Breslow, N.E. y Day, N.E. (1980). *Statistical methods in cancer research*. Vol. I: The analysis of case-control studies. Scientific Publication N° 32. Lyon: International Agency for Research on Cancer.
- Camilli, G., y Shepard, L. (1994). *Methods for identifying biased test items*. Thousand Oaks: Sage.
- 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.
- 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.
- Holland, P.W., y Wainer, H. (1993). *Differential Item Functioning* Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- SAS Institute Inc., SAS/STAT® *User's Guide*, Version 6, Fourth Edition, Volume 1, Cary, N.C.: SAS Institute Inc., 1989. 943 pp.
- 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.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in item development. In P.W. Holland & H. Wainer (Eds.) *Differential Item Functioning* (pp. 337-347). Hillsdale, NJ: Lawrence Erlbaum.