

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

El parámetro de desviación discriminante en la Ley del Juicio Comparativo

Antonio Jesús Herencia-Leva y Cristino Pérez Meléndez
Universidad de Granada

En diferentes publicaciones (López e Hidalgo, 2001, p. 331; Luce et al., 1990; Cliff, 1993; Lee, 2001) se ha ofrecido evidencia acerca de las limitaciones que las pruebas de ajuste presentan a la hora de comprobar los supuestos un modelo que utiliza una representación funcional. Un ejemplo de tales tipos de modelos son los basados en la Ley del Juicio Comparativo de Thurstone (1927a). Los resultados de este estudio no recomiendan interpretar el parámetro de desviación discriminante de la Ley del Juicio Comparativo como un índice del grado de confusión que provocan los estímulos en los jueces y a su vez abre la posibilidad de que realmente sean los valores escalares obtenidos con dicho modelo los que se deban interpretar como un índice del grado confusión que provocan en los sujetos.

Discriminal desviation parameter in the Law of Comparative Judgment. In different publications (López & Hidalgo, 2001, p. 331; Luce et al., 1990; Cliff, 1993; Lee, 2001) it has offered itself evidence brings over of the limitations that the tests of adjustment present at the moment of verify the suppositions of a model that uses a functional representation. An example of such types of models they are the based ones on the Law of the Thurstone's comparative Judgment. The results of this study do not recommend to interpret the parameter of disciminal desviation of the Law of the Comparative Judgment, as an index of the degree of confusion that provoke the stimuli in the judges and in turn will be the possibility that really they are the scale values obtained with the above mentioned model, which should be interpreted as an index of the degree confusion that they provoke in the subjects.

La adecuación de los modelos de escalamiento se realiza normalmente utilizando una prueba de ajuste en la cual se realiza una estimación del grado de desviación de las predicciones realizadas por el modelo (Rojas, González, Padilla y Pérez, 2000, p. 296; Hair, Anderson, Tatham y Black, 1999). En diferentes publicaciones (López e Hidalgo, 2001, p. 331; Luce et al., 1990; Cliff, 1993; Lee, 2001) se ha ofrecido evidencia acerca de las limitaciones que tales pruebas presentan a la hora de comprobar los supuestos del modelo u orientar en las posibles razones que pueden explicar que no se ajuste a los datos.

La comprobación de los supuestos de un modelo es importante en los contextos de escalamiento de estímulos y respuestas, ya que existen ejemplos en la literatura en los que se ha conseguido ajustar los datos a un modelo que no es coherente con el tipo de tarea

e instrucciones dadas a los sujetos (Torgenson, 1958; Coombs, 1964). En otras ocasiones ocurre que las ecuaciones funcionales que caracterizan a dos modelos que proceden de enfoques diferentes no se pueden diferenciar empíricamente (Maydeu-Olivares, 1999; Böckenholt y Tsai, 2001; Tsai y Böckenholt, 2001).

Estos dos aspectos ponen de relieve las limitaciones que presentan las representaciones funcionales y las técnicas de ajuste. Ejemplos de modelos que hacen uso de representaciones funcionales se pueden encontrar en la clásica Ley del Juicio Comparativo de Thurstone (1927a). Dicha ley parte de una teoría *encubierta* que trata de explicar cómo tiene lugar un proceso de juicio. Se trata de una teoría especialmente atractiva, debido a que los diferentes casos que se contemplan en la Ley permiten hablar de la presencia en el proceso de juicio de diferentes tipos de componentes. Uno de ellos es el grado en que los estímulos son confundidos por los jueces de la muestra. Este componente se define operativamente como el parámetro de desviación discriminante asociada a los estímulos.

En este trabajo se pretende responder a dos preguntas fundamentales: por un lado, ¿se puede interpretar el parámetro de desviación discriminante asociada a los estímulos como un índice del grado confusión que éstos provocan en los jueces? Por otro lado,

de ser afirmativa la respuesta a dicha pregunta, ¿qué significado tienen los valores escalares asignados a los estímulos?

Para responder a la primera pregunta se hace uso de pruebas de ajuste diseñadas para los casos V y IV de la Ley del Juicio Comparativo. En caso de obtenerse sólo ajuste para el caso IV, este aspecto se interpreta (Thurstone, 1927a; Guilford, 1954; Torgenson, 1954) como evidencia de que existen estímulos que tienden a ser más confundidos que otros. Para comprobar si esta respuesta es correcta se hace uso de diferentes tipos de índices que se basan en el análisis de los juicios individuales dados por los sujetos, en concreto, se hace uso de pruebas de transitividad (Hendel, 1977; Bezembinder, 1981; Hull y Buhyoff, 1981; Rawlin y Meglino, 1989; Monsuur y Storcken, 1997) e índices que reflejan cómo se distribuyen a lo largo de los estímulos diferentes tipos de errores cometidos por los jueces (Fishburn, 1970; 1973a, 1973b). Llegados a este punto cabe preguntarse: ¿existirá alguna relación entre la consistencia de los juicios de los sujetos y los resultados obtenidos con las técnicas de ajuste?; ¿existirá alguna relación entre la distribución de los errores cometidos por los jueces y los valores que adopta el parámetro de desviación típica asociada a los estímulos?

Para analizar la interpretación que cabe realizar de los valores escalares asignados a los estímulos se construyen matrices de confusión a partir de la distribución de errores cometidos por los jueces. En este caso, ¿se podrán interpretar los valores escalares como índices del grado en que los estímulos son confundidos?

Método

Sujetos

Se utilizó un total de 405 sujetos procedentes de diferentes facultades de la Universidad de Granada. Tal como se presenta en el apartado siguiente del material, se construyeron tres cuestionarios diferentes para evaluar los constructos de individualismo, colectivismo y autoritarismo y para cada uno se construyeron cuatro formas paralelas. La tabla 2 recoge la distribución de los sujetos a lo largo de los diferentes cuestionarios. No se registraron variables sociodemográficas por no estar éstas incluidas en ninguno de los objetivos de este estudio.

Material

Se hizo uso del marco teórico de Schwartz (1990) para concretar las sentencias a utilizar como estímulos en el escalamiento de los constructos individualismo, colectivismo y autoritarismo. En

| | Forma A | Forma B | Forma C | Forma D | Total |
|----------------|---------------|--------------|---------------|---------------|----------------|
| Individualismo | 44 26,35% | 22 13,17% | 54 32,33% | 47 28,14% | 167 41,20% |
| Colectivismo | 38 25,85% | 40 27,20% | 33 22,45% | 36 24,50% | 147 36,30% |
| Autoritarismo | 24 26,37% | 22 24,18% | 21 23,08% | 24 26,37% | 91 22,50% |
| Total | 106 26,17% | 84 20,74% | 108 26,67% | 107 26,42% | 405 100,00% |

la Tabla III se recogen las diez sentencias empleadas, de modo que las cuatro primeras hacen referencia al constructo Colectivismo, las cuatro últimas al de Individualismo y las dos restantes son representativas de ambos constructos.

Con estas sentencias se confeccionaron cuatro cuestionarios para cada uno de los tres constructos –dando lugar a 12 cuestionarios en total–. Estos cuestionarios se diseñaron para controlar el efecto del orden de presentación de los estímulos en la serie de pares; los cuatro cumplían los criterios que en referencia a este aspecto se suelen citar (Guilford, 1954, p. 196, 305 y ss.; Torgenson, 1958, p. 168; Wells, 1991). En cada uno de estos cuestionarios se presentaba una tabla en la que aparecían los estímulos enfrentados entre sí por pares.

Procedimiento

Los cuestionarios incluían una hoja con instrucciones y un escenario en el cual se describían las características más relevantes del constructo que debían tener en cuenta a la hora de juzgar la adecuación de los estímulos. Ya que se pretendía escalar tres constructos, se diseñaron tres escenarios diferentes. La hoja de instrucciones era la misma para los 12 cuestionarios, pero los escenarios variaban dependiendo del tipo de constructo que se pretendía que tuvieran en cuenta los jueces. En los escenarios se describía a una persona que se comportaba de acuerdo con las características que se atribuyen al constructo en consideración. En ningún momento se incluyó alguna referencia directa que pudiera conectarse luego con alguna de las sentencias que debían juzgar. Este material se incluyó para intentar que todos los jueces tuvieran el mismo conocimiento del constructo.

Cada uno de los cuestionarios se aplicó a una muestra diferente y tanto las instrucciones como el escenario se leyeron en voz alta.

Análisis

Como se menciona al principio de este estudio, se va a trabajar con tres tipos de pruebas: pruebas de ajuste, de transitividad e índices de distribución de errores. Finalmente, se describe el tratamiento que recibieron los datos para obtener las matrices de confusión.

| Valores | Sentencias |
|---------------|--|
| Universalismo | Tener en cuenta el bienestar de las demás personas. |
| Benevolencia | Prestar atención a la gente importante para uno. |
| Tradicición | Mantener las costumbres y tradiciones de la cultura a la que se pertenece. |
| Conformidad | Actuar de acuerdo con las normas establecidas en la sociedad. |
| Seguridad | Considerar que la sociedad permite realizar lo que interesa a todo el mundo. |
| Poder | Preocuparse por obtener un alto estatus social. |
| Logro | Poner énfasis en la consecución de las metas personales. |
| Hedonismo | Hacer las cosas con las que uno disfruta. |
| Estimulación | Acumular tantas experiencias como se pueda. |
| Autodirección | Ser libre para decidir lo que uno cree y hace. |

Pruebas de ajuste

Dado que se van a utilizar los modelos de escalamiento de Thurstone y Comrey, para el primer modelo haremos uso de la prueba de ajuste diseñadas por Mosteller (Guilford, 1954, p. 163 y ss.) para los casos IV y V de la Ley del Juicio Comparativo.

Pruebas de transitividad

En lo que se refiere a las pruebas de transitividad se hará uso de las condiciones de Transitividad Estocástica Débil (TED). Una descripción de tales pruebas se puede encontrar en Hendel (1977) y Bezembinder (1981), entre otros. Se aplicó la chi-cuadrado con la finalidad de ver si el número de inconsistencias cometidas por los jueces era inferior al que cabría esperar por azar.

Índices de distribución de errores

Se calcula el porcentaje de errores de sustitución y omisión contemplados en el modelo de Fishburn (1970; 1973b). Con dicho modelo se asigna un intervalo a cada estímulo, permitiendo detectar estos dos tipos de errores: cuando estímulos con intervalos no solapados son confundidos y cuando estímulos con intervalos solapados son juzgados como diferentes, respectivamente (Half, Ortony y Anderson, 1976).

Los resultados obtenidos con este cálculo se comparan con la distribución de valores de desviación discriminante asociada a los estímulos. Se hace uso de la prueba U de Mann-Whitney para someter a comprobación la hipótesis de que los estímulos extremos presentan desviaciones discriminantes más altas que los estímulos centrales, tratando con ello de replicar los resultados obtenidos por Guilford (1954).

Matrices de confusión

Se utiliza la distribución de errores de sustitución de un estímulo a lo largo de los restantes, como un índice de la similitud que éste guarda con cada uno de ellos, obteniendo una matriz de confusión (Contasti y Castro, 1985). Posteriormente, se procede a escalar los estímulos haciendo uso de las técnicas de escalamiento multidimensional. Se utiliza como índice de ajuste el valor de «Stress» arrojado por la configuración introducida.

Resultados

En lo que se refiere a los índices de ajuste para el modelo de Thurstone, el mejor ajuste se obtiene para el constructo Colectivismo en el caso IV de la Ley del Juicio Comparativo (χ^2 -27, 0.8471- = 19.6036). Para el caso V, en cambio, el mejor ajuste se obtiene para el constructo Autoritarismo (χ^2 -36, 0.3237- = 39.3161). Para el constructo Individualismo no se obtiene ajuste en el caso V (χ^2 -36, 0.0006- = 70.0715) y para el caso IV éste es muy pobre (χ^2 -27, 0.0942- = 37.0528).

En los resultados del análisis de la propiedad transitiva se obtiene que, para la prueba TED, en los tres constructos los sujetos se muestran consistentes (el valor de la Chi-Cuadrado, por constructos, igual a: -146, 0.0000- = 6451.5330 para Colectivismo, -90, 0.0000- = 2412.4170 para Autoritarismo y -157, 0.0000- = 8084.7500 para individualismo).

En cuanto a la hipótesis de la distribución de las desviaciones discriminantes contempladas en el caso IV del modelo de Thurstone, utilizando la prueba U de Mann-Whitney, los valores obtenidos por constructos fueron $z = 2,34520$ ($p < 0,01902$) para individualismo, $z = 0,00000$ ($p < 1,00000$) para autoritarismo y $z = 1,91880$ ($p < 0,05502$) para colectivismo. Por tanto, en dos de las tres comparaciones se obtiene apoyo a que las desviaciones discriminantes son mayores para los estímulos extremos.

A continuación se analizó la distribución de los dos tipos de errores contemplados en el modelo. Bajo la hipótesis de que la distribución de los errores de sustitución tiende a acumularse en la parte central del continuo, se aplicó la prueba no paramétrica U de Mann-Whitney, obteniéndose los siguientes valores por constructo: $z = 0,85280$ ($p = 0,39377$) para Autoritarismo, $z = 2,39317$ ($p = 0,01671$) para Colectivismo y $z = 2,05129$ ($p = 0,04025$) para Individualismo. La misma prueba se aplicó en este caso para los errores de omisión, obteniéndose que los resultados eran no significativos para los tres constructos. Por tanto, en los tres casos se puede afirmar que las inconsistencias vinculadas a los juicios de dominancia se tienden a acumular en la parte central.

Finalmente, se construyó la matriz de confusión y se encontró que sólo en la configuración con dos dimensiones se obtenía un valor del estadístico Stress aceptable para uno de los constructos (consultar Contasti y de Castro, 1985). En el resto de los constructos el valor de este estadístico era superior a 0,20. El constructo para el que se obtiene ajuste es el de colectivismo, siendo el valor del Stress de 0,08171. La correlación entre los valores escalares obtenidos en la primera dimensión y los obtenidos con el procedimiento de Thurstone es negativa y superior a 0,90 ($p < 0,01$). Por tanto, se puede afirmar que los juicios de indiferencia tienden a darse en menor grado para aquellos estímulos con diferencias en valor escalar más grande.

Discusión

En la introducción se planteaban dos grandes preguntas: una hacía referencia a la interpretación que se podía realizar del parámetro de desviación discriminante asociado a los estímulos en la Ley de Thurstone, como un índice del grado en que éstos eran confundidos por los sujetos. En referencia a este aspecto se ha obtenido evidencia de que los parámetros de desviación discriminante son más elevados para los estímulos con puntuaciones extremas. Posteriormente, se ha analizado la distribución de errores cometidos por los jueces y se ha encontrado que éstos tienden a darse en mayor proporción entre los estímulos centrales. Tampoco se encuentra una relación entre los resultados obtenidos en las pruebas de ajuste y la consistencia con que los jueces realizan sus juicios.

Respecto a la segunda pregunta, en la que se hacía referencia a la interpretación que cabría realizar de los valores escalares asignados a los estímulos, los resultados obtenidos escalando los estímulos haciendo uso de una matriz de confusión indican que existe una correlación negativa entre los valores escalares obtenidos mediante ambos procedimientos.

En los modelos de última generación existentes actualmente (Böckenholt y Tsai, 2001; Tsai y Böckenholt, 2001) se contempla la influencia de las diferencias individuales en los juicios realizados por los sujetos, haciendo que sea imposible diferenciar la aproximación centrada en el escalamiento de estímulos, de la centrada en el escalamiento de respuestas. Se hace necesario, pues, construir nuevos índices que permitan estudiar el comportamiento de los parámetros introducidos en dichos modelos.

Finalmente, se quiere resaltar la problemática existente en la confusión entre el lenguaje matemático con que se expresan relaciones entre variables y el lenguaje que se utiliza posteriormente para hablar sobre los parámetros que un determinado modelo incluye. Los resultados de este estudio no recomiendan interpretar el parámetro de desviación discriminante de la Ley del Juicio Comparativo, como un índice del grado de confusión que provocan los estímulos en los jueces y a su vez abre la posibilidad de que real-

mente sean los valores escalares obtenidos con dicho modelo los que se deban interpretar como un índice del grado de confusión que provocan en los sujetos.

Agradecimientos

Este trabajo ha sido realizado parcialmente bajo los proyectos DGICYI, PB98-1305 y TIC2002-03411.

Referencias

- Bezembinder, T.G.G. (1981). Circularity and consistency in paired comparisons. *British Journal of Mathematical and Statistical Psychology*, 34, 16-37.
- Böckenholt, U. y Tsai, R.C. (2001). Individual differences in paired comparison data. *British Journal of Mathematical and Statistical Psychology*, 54, 265-77.
- Candel, M.J.J.M. (1997). A probabilistic feature model for unfolding tested for perfect and imperfect nestings. *Journal of Mathematical Psychology*, 41, 414-30.
- Cliff, N. (1993). What is and isn't measurement. En G. Keren y C. Lewis (Eds.), *A handbook for the data analysis in the behavioral sciences: methodological issues* (pp. 59-93). New Jersey: Lawrence Erlbaum Associates, Publishers.
- Contasti, M. y de Castro, C. (1985). *Aproximación métrica a variables ordinales*. Colección Estudio. Caracas: Editorial de Humanidades y Educación, Universidad Central de Venezuela.
- Coombs, C.H. (1964). *A theory of data*. New York: John Wiley & Sons.
- Fishburn, P.C. (1970). Intransitive indifference with unequal indifference intervals. *Journal of Mathematical Psychology*, 7, 144-9.
- Fishburn, P.C. (1973a). *The Theory of Social Choice*. New Jersey: Princeton University Press.
- Fishburn, P.C. (1973b). Interval representations for interval orders and semiorders. *Journal of Mathematical Psychology*, 10, 91-105.
- Guilford, J.P. (1954). *Psychometric Methods*. New York: McGraw-Hill Book Company. Second Edition.
- Halford, H.M., Ortony, A. y Anderson, R.C. (1976). A context-sensitive representation of word meanings. *Memory and Cognition*, 4(4), 378-83.
- Hendel, D.D. (1977). Intransitivity on paired-comparisons instruments: the relationship of the total circular triad scores to stimulus circular triads. *Applied Psychological Measurement*, 1(3), 403-11.
- Hull, R.C. y Buhoff, G.J. (1981). On the law of comparative judgement: scaling with intransitive observers and multidimensional stimuli. *Educational and Psychological Measurement*, 41, 1.083-9.
- Lee, M.D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, 45, 149-66.
- López, J.A. e Hidalgo, M.D. (2001). Un estudio de potencia y tasa de error tipo I del estadístico de razón de verosimilitud en la TRI. *Psicothema*, 13(2), 330-4.
- Luce, R.D., Krantz, D.H., Suppes, P. y Amos, A. (1990). *Foundations of Measurement. Vol. 3: Representation, axiomatization and invariance*. San Diego, California: Academic Press.
- Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*, 64 (3), 325-40.
- Monsuur, H. y Storcken, T. (1997). Measuring intransitivity. *Mathematical Social Sciences*, 34, 125-52.
- Rawlin, E.C. y Meglino, B.M. (1989). The transitivity on work values: hierarchical preference ordering of social desirable stimuli. *Organizational Behavior and Human Decision Processes*, 44, 494-508.
- Rojas, A.J., González, A., Padilla, J.L. y Pérez, C. (2000). Comparación de estrategias de ajuste de los datos al modelo de crédito parcial. *Psicothema*, 12(2), 296-302.
- Schwartz, S.H. (1990). Individualism-collectivism: critique and proposed refinements. *Journal of Cross-Cultural Psychology*, 21(2), 139-57.
- Thurstone, L.L. (1927a). A law of comparative judgment. *Psychological Review*, 34, 273-86.
- Thurstone, L.L. (1927b). Three Psychophysical laws. *Psychological Review*, 34, 424-32.
- Torgenson, W.S. (1958). *Theory and Methods of Scaling*. New York: John Wiley and Sons, Inc. Cuarta Impresión, 1963.
- Tsai, R.C. y Böckenholt, U. (2001). Maximum likelihood estimation of factor and ideal point models for paired comparison data. *Journal of Mathematical Psychology*, 45, 795-811.
- Wells, A.J. (1991). Optimal presentation orders for the method of paired comparisons. *British Journal of Mathematical and Statistical Psychology*, 44, 181-93.
- Hair, J.F., Anderson, R.E., Tatham, R.L. y Black, W.C. (1999). *Análisis Multivariante*. Madrid: Prentice-Hall.