

Jordan Muiños, Federico M.

 **Federico M. Jordan Muiños**
federico.muinos@uflo.edu.ar
Universidad de Flores, Argentina

PSOCIAL
Universidad de Buenos Aires, Argentina
ISSN-e: 2422-619X
Periodicidad: Semestral
vol. 7, núm. 1, 2021
psocial@sociales.uba.ar

URL: <http://portal.amelica.org/ameli/jatsRepo/123/1232225009/index.html>

Resumen: En las investigaciones psicológicas resulta de amplia relevancia conocer si los cuestionarios utilizados miden los constructos latentes de interés. A tal efecto, el Análisis Factorial Confirmatorio se emplea frecuentemente en la literatura especializada para aportar evidencias de validez a las escalas empleadas. Sin embargo, en la actualidad existe cierta disparidad de criterios respecto de los puntos de corte que deben tenerse en consideración. Por este motivo, el presente trabajo se propone revisar la literatura en torno a los diferentes puntos de corte tradicionales de los índices de ajuste más utilizados. Se concluye que para seleccionar los índices e interpretar los resultados se debe tener en cuenta que estos puntos de corte pueden cambiar por diferentes motivos, tales como el tamaño de la muestra.

Abstract: In psychological research, it is of great relevance to know whether the questionnaires used measure the latent constructs of interest. For this purpose, Confirmatory Factor Analysis is frequently used in the specialized literature to provide evidence of validity to the scales used. However, there is currently a certain disparity in the criteria regarding the cut-off points to be taken into consideration. For this reason, this paper aims to review the literature on the different cut-off points for the most commonly used fit indexes. It is concluded that to select the indexes and interpret the results, it should be should take into account that these cut-off points may change for different reasons, such as sample size.

Keywords: punto de corte, índices de ajuste, análisis factorial confirmatorio, validez, cut-off value, fit indexes, confirmatory factor analysis, validity .

INTRODUCCIÓN

Uno de los métodos que emplea la investigación científica para recolectar y analizar datos es el método cuantitativo (Cadena-Iñiguez et al., 2017). Entre las herramientas que utiliza este método para recopilar datos se encuentra el uso de los cuestionarios, cuyo diseño tiene en cuenta la comprensión del participante y la claridad de los ítems o reactivos (Kamp, 2018). Los evaluados responden a estos instrumentos a través de una escala tipo Likert (ordinal), que tradicionalmente suele incluir 3 o 4 puntos de anclaje (Wu & Leung, 2017). De esta manera, es posible medir una variable latente por medio de indicadores observables y acceder a una medida indirecta del constructo. Sin embargo, es importante que la relación entre estos indicadores observables y las variables latentes sea sólida para que las inferencias que se obtengan resulten precisas y correctas (Batista-Foguet, et. al., 2004).

Para evaluar la calidad de un instrumento se realizan análisis de evidencias de confiabilidad y validez. El primero refiere a que, si se aplican las mismas condiciones, las mediciones realizadas con el instrumento obtienen el mismo resultado en diferentes momentos, escenarios y poblaciones (Manterola et al., 2018). El segundo indica en qué grado el instrumento mide la variable que dice medir (Hernández Sampieri et. al, 2014).

De acuerdo con la literatura, es posible distinguir diferentes análisis de evidencias de validez: validez aparente o de aspecto, de contenido, externa e interna. La validez aparente es una evaluación subjetiva realizada por el autor del instrumento y por los usuarios del mismo. El objetivo es evaluar si el instrumento parece una encuesta formal o un instrumento de medida; se tiene en cuenta la organización del instrumento, si los ítems son entendibles, razonables y claros, el léxico empleado, la claridad de las instrucciones y organización del instrumento (Lagunes Córdoba, 2017).

La validez de contenido analiza en qué grado los ítems de un instrumento son representativos y relevantes de un dominio específico del constructo a evaluar (Hernández Sampieri, et. al., 2014). Se tienen en cuenta principalmente cuatro elementos: definición, relevancia, representación del dominio y qué tan apropiado es el procedimiento de construcción del instrumento (Almanasreh et al., 2019).

La validez externa es la capacidad de generalizar las inferencias obtenidas con el instrumento a otras poblaciones o muestras (Findley, 2021). Esta validez se suele yuxtaponer con la validez interna (Reiss, 2019), la cual es el grado en que los resultados obtenidos representan la población estudiada y que no hay un error metodológico (Patino & Carvalho Ferreira, 2018).

El realizar los diferentes análisis para encontrar evidencias de confiabilidad y validez permite evitar los errores aleatorios o no sistemáticos y sistemáticos. El primero de estos errores compromete la confiabilidad del instrumento, pues el mismo está asociado a las variaciones azarosas que ocurren en las mediciones repetidas y no puede evitarse. Los factores que influyen son: el tamaño muestral, la variabilidad individual e interindividual y la magnitud de las diferencias (a mayor diferencia en la comparación, la probabilidad de que se deba al azar será menor) (Barraza et al., 2019). El segundo tipo de error, los sistemáticos, ocurre cuando las variaciones ocurren de forma predecible y se sobreestima o subestiman los resultados obtenidos en medidas repetidas y conlleva a conclusiones erróneas. Estos se encuentran relacionados con tres tipos de sesgos: el de selección (inadecuada selección de los participantes), de información o medición (son defectos que ocurren durante la medición) y de confusión (cuando se llega a conclusiones erróneas en la relación entre las variables) (Hernández-Ávila et al, 2000; Manterola & Otzen, 2015).

La validez interna, por su parte, es la más utilizada en las investigaciones científicas. Para llevar a cabo dicho análisis se suele utilizar el Análisis Factorial Confirmatorio.

ANÁLISIS FACTORIAL CONFIRMATORIO

Este análisis permite contrastar la hipótesis con la relación entre indicadores y dimensiones latentes (Batista-Foguet et al, 2004). De esta forma, analizando la estructura interna del mismo, se puede verificar que el instrumento realmente evalúe el constructo que dice medir y testar si los datos empíricos recogidos por el instrumento se ajustan al modelo teórico que subyace al mismo (Smith & McMillan, 2001). Este ajuste es evaluado mediante dos clases de índices que son incrementales y absolutos (Domínguez-Lara, 2019).

Los índices de ajuste incrementales son aquellos índices que evalúan la mejora del modelo propuesto en relación a un modelo base (McNeish et al., 2017). Ejemplos de ellos son el CFI (*Comparative fit index*), el GFI (*Goodness of fit index*) y el TLI (*Tucker-Lewis inndex*), entre otros.

Cuando el CFI otorga un valor mayor o igual .95 se considera que el modelo se ajusta a la muestra (Lai, 2020). Para el GFI se recomienda un punto de corte igual o superior a .89 en una muestra de 100 casos, mientras que en muestras mayores se recomienda un corte mayor o igual a .93 (Cho et al, 2020). El punto de corte recomendado para el TLI es superior a .90 (Xia & Yang, 2019). En los índices de ajuste incrementales

generalmente se considera que un valor de superior o igual a .90 es adecuado, y si es mayor o igual a .95 se lo considera óptimo, siendo el puntaje ideal es de 1 (Hu & Bentler, 1999; Kline, 2011; Brown, 2015; Escobedo Portillo et. al., 2016).

Los índices de ajuste absolutos indican en qué grado el modelo observado en la matriz de covarianza iguala a la matriz de covarianza del modelo implícito; mientras menor sea el resultado, mejor ajusta el modelo (Chen, 2007). Entre ellos, se encuentra el RMSEA (*Root mean squared error of approximation*), para el cual se recomienda como mínimo una muestra de 200 casos, o emplearlo con precaución en muestras menores (Curran et. al., 2003).

Otro índice de ajuste absoluto que se suele utilizar es el SRMR (*Standardized root mean square residual*). Este índice se interpreta como la covarianza residual estandarizada de la muestra (Maydeu-Olivares et al., 2017). En general, el SRMR es más eficaz que el RMSEA en rechazar modelos que no se ajustan estrechamente, en especial en muestras chicas, iguales o menores a 200 casos (Shi et al., 2019).

Cuando el RMSEA otorga un valor igual o inferior a .05 se considera que el modelo se ajusta de forma adecuada a la muestra (Lai, 2020). Para el SRMR se recomienda un punto de corte menor o igual a .09 en una muestra de 100 casos o menos, mientras que para muestra mayor a 100 casos, se recomienda un punto de corte de .08 o menos (Cho et. al., 2020).

Otro indicador que se tiene en cuenta para evaluar el ajuste de la muestra al modelo propuesto es el chi cuadrado (χ^2). Si su valor es estadísticamente significativo (ej. $p < .05$), el ajuste del modelo es pobre en comparación a la muestra. Caso contrario ($p \geq .05$), se considera que el modelo se ajusta de forma adecuada a la muestra (Walker & Smith, 2017). Cabe señalar que un inconveniente que presenta el χ^2 es su sensibilidad al tamaño de la muestra. Esto se debe a que en muestras relativamente grandes suele aumentar las diferencias entre la matriz de varianza-covarianza de la muestra y la matriz del modelo propuesto, a comparación de lo que debería ser (Lewis, 2017).

Según explica Rojas-Torres (2020), en los modelos de una única estructura factorial, los índices de ajuste TLI, CFI, RMSEA y SRMR poseen muy poca variabilidad ante valores extremos. A su vez, menciona que al aumentar la muestra, la variabilidad disminuye. Por lo cual, los puntos de corte clásicos de estos índices son robustos ante la contaminación de los datos. Sin embargo, se recomienda que en muestras de 100 o menos casos, se consideren nada más los puntos de corte clásico del SRMR y el CFI (Rojas-Torres, 2020). Además, el SRMR es un indicador robusto en diferentes métodos de estimación (Shi & Maydeu-Olivares, 2019).

Sin embargo, según indican McNeish et al. (2017), no existe un solo criterio de corte que pueda aplicarse a cualquier índice en los modelos de variable latente. Por el contrario, se deben tener en cuenta diferentes variables, tales como la complejidad del modelo, el tamaño de la muestra, el número de indicadores por factor, la calidad de la medición o el número de alternativas de respuesta de los indicadores observables utilizados, entre otras.

DISCUSIÓN

Al indagar un constructo teórico es necesario contar con un instrumento que lo evalúe de forma correcta. A tal efecto, se realizan diferentes análisis que puedan aportar evidencias de confiabilidad y validez. Entre estas últimas el análisis más utilizado es el AFC, permitiendo confirmar de forma estadística si el instrumento mide la variable latente del estudio.

Por este motivo, es necesario comprender qué miden los índices utilizados. De esta manera se puede saber si el cuestionario presenta suficientes evidencias de validez como para ser empleado en el campo de aplicación, o si aún requiere de mejoras que reduzcan el error de medición al momento de relevar la información sobre el constructo latente. Pues, el saber que el instrumento mide lo que dice medir, reduce la posibilidad de errores sistemáticos en la investigación efectuada.

Además, conocer cuáles son los puntos de corte tradicionales permite validar instrumentos en base a los parámetros aceptados por la comunidad científica en general. Aunque, como explica Lai (2020), si bien existen puntos de cortes consensuados por la comunidad científica, es importante destacar que no son universalmente aplicables porque estos ignoran las características del modelo y de los datos. Una alternativa que ofrece el autor es ajustar los puntos de corte de acuerdo a las características del modelo y de los datos.

Al seguir las sugerencias de Lai (2020) es importante justificar por qué y en base a qué se realiza el cambio en los puntos de corte tradicionales.

REFERENCIAS

- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214 – 221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Barraza, F., Arancibia, M., Madrir, E., & Papuzinski, C. (2019). Conceptos generales en bioestadística y epidemiología clínica: error aleatorio y error sistemático. *Medwave*, 19(7). <https://doi.org/10.5867/medwave.2019.07.7687>
- Batista-Foguet, J. M., Coenders, G., & Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina Clínica*, 122 (1), 21-7. <https://doi.org/10.1157/13057542>
- Brown, T. (2015). *Confirmatory factor analysis for applied research*. Guilford Press.
- Cadena-Iñiguez, P., Rendón-Medel, R., Aguilar-Ávila, J., Salinas-Cruz, E., del Rosario de la Cruz-Morales, F., & Sangerman-Jarquín, D. M. (2017). Métodos cuantitativos, métodos cualitativos o su combinación en la investigación: un acercamiento en las ciencias sociales. *Revista Mexicana de Ciencias Agrícolas*, 8 (7), 1603 - 1617.
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling*, 14 (3), 464 - 504. <https://doi.org/10.1080/10705510701301834>
- Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA. *Sociological Methods & Research*, 32 (2), 208 - 252. <https://doi.org/10.1177/0049124103256130>
- Cho, G., Hwang, H., Sarstedt, M., & Ringle, Ch. M. (2020). Cutoff criteria for overall model fit indexes in generalized structured component analysis. *Journal of Marketing Analytics*. <https://doi.org/10.1057/s41270-020-00089-1>
- Domínguez-Lara, S. (2019). Correlación entre residuales en análisis factorial confirmatorio: una breve guía para su uso e interpretación. *Interacciones*, 5 (3), 1-7. <https://doi.org/10.24016/2019.v5n3.207>
- Escobedo Portillo, M. T., Hernández Gómez, J. A., Estebané Ortega, V. E., & Martínez Moreno, G. (2016). Modelos de ecuaciones estructurales: características, fases, construcción, aplicación y resultados. *Ciencia y Trabajo*, 18(55), 16-22. <https://doi.org/10.4067/S0718-24492016000100004>
- Findley, M. G., Kikuta, K., & Denly, M. (2021). External Validity. *Annual Review of Political Science*, 24, 365 – 393. <https://doi.org/10.1146/annurev-polisci-041719-102556>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. P. (2014). *Metodología de la Investigación*. McGraw Hill Education.
- Hernández-Ávila, M., Garrido, F., & Salazar-Martínez, E. (2000). Sesgos en estudios epidemiológicos. *Salud Pública de México*, 42(5), 438 – 446. <https://doi.org/10.1590/S0036-3634200000500010>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1 – 55. <https://doi.org/10.1080/10705519909540118>
- Kamp, K., Wyatt, G., Dudley-Brown, S., Brittain, K., & Given, B. (2018). Using cognitive interviewing to improve questionnaires: An exemplar study focusing on individual and condition-specific factors. *Applied Nursing Research*. <https://doi.org/10.1016/j.apnr.2018.06.007>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. The Guilford Press.

- Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality*, 80, 17 - 22. <https://doi.org/10.1016/j.jrp.2019.03.007>
- Lagunes Córdoba, R. (2017). Recomendaciones sobre los procedimientos de construcción y validación de instrumentos y escalas de medición en la psicología de la salud. *Psicología y Salud*, 27(1), 5 - 18.
- Lai, K. (2020). Fit Difference Between Nonnested Models Given Categorical Data: Measures and Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, <https://doi.org/10.1080/10705511.2020.1763802>
- Lewis, T. F. (2017). Evidence Regarding the Internal Structure: Confirmatory Factor Analysis. *Measurement and Evaluation in Counseling and Development*. <https://doi.org/10.1080/07481756.2017.1336929>
- McNish, D., An, J., & Hancock, G. R. (2017). The Thorny Relation Between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. *Journal of personality assessment*. <https://doi.org/10.1080/00223891.2017.1281286>
- Manterola, C., & Otzen, T. (2015). Los sesgos en investigación clínica. *International Journal of Morphology*, 33(3), 1156 - 1164. <https://doi.org/10.4067/S0717-95022015000300056>
- Reiss, J. (2019). Against external validity. *Synthese*, 196, 3103 - 3121. <https://doi.org/10.1007/s11229-018-1796-6>
- Rojas-Torres, L. (2020). Robustez de los índices de ajuste del análisis factorial confirmatorio a los valores extremos. *Revista de Matemática: Teoría y Aplicaciones*, 27(2), 403-424. <https://doi.org/10.15517/rmta.v27i2.33677>
- Shi, D., & Maydeu-Olivares (2019). The Effect of Estimation Methods on SEM Fit Indices. *Educational and Psychological Measurement*, 1 - 25. <https://doi.org/10.1177/0013164419885164>
- Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2019). Assessing Fit in Ordinal Factor Analysis Models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*. <https://doi.org/10.1080/10705511.2019.1611434>
- Smith, T. D., & McMillan, B. F. (2001). *Primer of Model Fit Indices in Structural Equation Modeling*. Southwest Educational Research Association.
- Patino, C. M., & Carvalho Ferreira, J. (2018). Internal and external validity: can you apply research study results to your patients? *Jornal Brasileiro de Pneumologia*, 44(03), 183. <https://doi.org/10.1590/S1806-37562018000000164>
- Walker, D. A., & Smith, T. J. (2017). Computing Robust, Bootstrap-Adjusted Fit Indices for Use With Nonnormal Data. *Measurement and Evaluation in Counseling and Development*, 50, 131 - 137. <https://doi.org/10.1080/07481756.2017.1326748>
- Wu, H., & Leung, S. O. (2017). Can Likert Scales be Treated as Interval Scales?—A Simulation Study. *Journal of Social Service Research*, 43 (4), 527 - 532. <https://doi.org/10.1080/01488376.2017.1329775>
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Method*, 51, 409 - 428. <https://doi.org/10.3758/s13428-018-1055-2>