Revista de Investigación
Hatun Yachay Wasi 1(1), 158 - 174 (2022)
ISSN: 2955 - 8255

ARTÍCULO ORIGINAL
Recibido: mayo 2022
Aceptado: julio 2022

EFECTO DEL TAMAÑO DE MUESTRA EN LOS ÍNDICES DE AJUSTE DEL ANÁLISIS FACTORIAL CONFIRMATORIO SOBRE EL CUESTIONARIO PHQ - 9 APLICADO A LA POBLACIÓN PERUANA

EFFECT OF SAMPLE SIZE ON THE ADJUSTMENT INDICES OF THE CONFIRMATORY FACTORIAL ANALYSIS ON THE PHQ - 9 QUESTIONNAIRE APPLIED TO THE PERUVIAN POPULATION

Moisés Tarazona Cochachin[1] Zoraida Huamán Gutiérrez[1] Emerson Norabuena Figueroa[1]
Roger Norabuena Figueroa[1]

[1]Universidad Nacional Mayor de San Marcos, Perú.

Correspondencia:
Dr. Roger Norabuena Figueroa
rnorabuenaf@unmsm.edu.pe

Para citar este artículo: Tarazona, M., Huamán, Z., Norabuena, E., & Norabuena, R. (2022). Efecto del tamaño de muestra en los índices de ajuste del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana. *Revista de Investigación Hatun Yachay Wasi 1*(1), 158 - 174. https://doi.org/10.57107/hyw.v1i1.18

RESUMEN

En la presente investigación, el objetivo fue determinar el efecto del tamaño de muestra en los índices de ajuste del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana. La población estuvo representada por 34476 registros de la encuesta ENDES – INEI - 2018, y el tamaño de muestra estuvo conformado por muestras elegidas mediante el muestreo Bootstrap con 100 réplicas en cantidad de 50, 100, 250, 500, 750, 1000, 2500, 5000 y 10000 casos. El tamaño de muestra presentó un efecto directo en los índices, de ajuste del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana; asimismo, un efecto directo con concavidad hacia arriba en el X^2, X^2/gl, AIC y BIC, un efecto directo con concavidad hacia abajo en el CFI, TLI, NNFI, NFI, RNI, GFI, PNFI y AGFI, y en algunas medidas de bondad de ajuste un efecto indirecto en el RMSEA, RMR y SRMR. Se contrasta un efecto directo e inverso del tamaño de muestra sobre las medidas de bondad de ajuste.

Palabras clave: medidas de bondad de ajuste, análisis factorial confirmatorio, cuestionario PHQ - 9.

Esta obra está publicada bajo la licencia CC BY 4.0

ABSTRACT

This investigation aimed to determine the effect of the size of the sample in the adjustment indices of the confirmatory factor analysis on the PHQ - 9 questionnaire applied to the Peruvian population. The population was represented by 34,476 records from the ENDES - INEI - 2018 survey, and the sample size was made up of samples chosen by Bootstrap sampling with 100 replicates in the amount of 50, 100, 250, 500, 750, 1000, 2500, 5000 and 10000 cases. The sample size had a direct effect on the indices of adjustment of the confirmatory factor analysis on the PHQ - 9 questionnaire applied to the Peruvian population; likewise, a direct effect with upward concavity in X^2, X^2/gl, AIC and BIC, a direct effect with downward concavity in CFI, TLI, NNFI, NFI, RNI, GFI, PNFI and AGFI, and in some measures of goodness of fit an indirect effect on the RMSEA, RMR and SRMR. A direct and inverse effect of the sample size on the measures of goodness of fit is contrasted.

Keywords: goodness of fit measures, confirmatory factor analysis, PHQ - 9 questionnaire.

INTRODUCCIÓN

En la actualidad, muchos métodos de modelamiento son empleados para estudiar los fenómenos que requieren la utilización de un conjunto complejo de variables. El análisis factorial confirmatorio (AFC), como caso particular del modelo de ecuaciones estructurales (SEM), estudia las construcciones teóricas de un constructo y sus indicadores, para analizarlos y conocer su viabilidad. Debido a que, no existe un criterio único, para la evaluación de ajuste del modelo teórico obtenido como resultado del AFC, se desarrolló una amplia gama de índices de ajuste, las cuales son afectados directamente por el tamaño de muestra y método de estimación (Escobedo et al., 2016).

Los índices de ajuste proporcionan información sobre la idoneidad del modelo para explicar los datos (Keith, 2019); sin embargo, éstos índices de ajuste se ven influenciados principalmente por el tamaño de muestra; así como, por la complejidad del modelo propuesto y el método de estimación.

Por otro lado, el tamaño de muestra se refiere al número de participantes que se incluye en un estudio de investigación. Si se realiza adecuadamente el cálculo, no solamente se establecerán conclusiones con sustento estadístico y se limitarán los riesgos potenciales a los sujetos incluidos en la investigación, sino, que además el estudio podrá planearse adecuadamente para optimizar costos económicos y tiempo (Rendón - Macías & Villasís - Keever, 2017). Así, el tamaño de muestra proporciona garantías de que el AFC refleje o se ajuste al comportamiento teórico planteado, por medio de los índices de ajuste.

Por ello, es importante evaluar los índices de ajuste del AFC aplicado a datos reales y sin aceptación de supuestos *a priori*, ni métodos de estimación seleccionados de antemano. Es así que las estadísticas de prueba del modelo son las estadísticas de ajuste SEM originales.

Generalmente son estadísticas de chi-cuadrado que prueban la hipótesis de ajuste exacto de que no hay diferencia entre las covarianzas predichas por el modelo, dadas las estimaciones de los parámetros, y la matriz de covarianza de la población. Rechazar esta hipótesis dice que: (1) los datos contienen información de covarianza referida en contra del modelo, y (2) el investigador debe explicar las discrepancias modelo-datos que exceden las esperadas por error de muestreo (Kline, 2016).

La prueba de chi-cuadrado que se acaba de describir es una prueba de aceptación-apoyo donde la hipótesis nula representa la creencia del investigador de que el modelo es correcto; por lo tanto, es el fracaso para rechazar la hipótesis nula, o la ausencia de significancia estadística (p. Ej., $p \geq 0.05$), lo que respalda el modelo. Esta lógica es "al revés" de la prueba más típica de rechazo - apoyo donde el rechazo de la hipótesis nula (p. Ej., $p < 0.05$), apoya la teoría del investigador.

De los dos, las pruebas de aceptación-apoyo son lógicamente más débiles porque el hecho de no refutar una afirmación (la hipótesis del ajuste exacto), no significa que la afirmación sea verdadera. La baja potencia en las pruebas de aceptación-soporte significa que hay pocas posibilidades de detectar un modelo falso.

Esto significa que analizar el modelo en una muestra que es demasiado pequeña (es decir, de baja potencia), hace que sea más probable que se retenga el modelo; sin embargo, en las pruebas de rechazo-apoyo, la penalización por baja potencia debido a un tamaño de muestra insuficiente es que las hipótesis del investigador tienen menos probabilidades de ser respaldadas (Sheldon, 2021). De esta manera, una estadística de prueba del modelo es como un detector de humo: si suena la alarma, puede haber o no un incendio (discrepancia grave entre modelo y datos), pero es prudente tratar la alarma con seriedad (se debe realizar una evaluación más detallada de ajuste).

Los índices de ajuste aproximados no son pruebas de significación, por lo que no hay una decisión binaria sobre si rechazar o aceptar una hipótesis nula al igual que no hay una demarcación de los límites del error de muestreo. En cambio, estos índices están destinados a ser medidas continuas de correspondencia modelo-datos. Algunos se escalan como estadísticas de "maldad de ajuste" donde los valores más altos indican peor ajuste, pero otros son medidas de "bondad de ajuste" donde los valores más altos indican un mejor ajuste.

Los valores de algunos índices de bondad de ajuste están más o menos estandarizados, de modo que, su rango es 0 - 1.0, donde un valor de 1.0 indica el mejor resultado. A continuación, se describen cuatro categorías de índices de ajuste aproximados (Kline, 2016). Estas categorías no son mutuamente excluyentes

porque algunos índices se pueden clasificar en más de uno. Los índices de ajuste absoluto miden qué tan bien un modelo *a priori*, explica los datos. Ese modelo es el modelo del investigador porque no hay otro punto de referencia para un índice de ajuste absoluto. Explicar los datos no dice por sí solo que el modelo sea adecuado. Esto se debe a que se puede hacer un modelo mal especificado para explicar los datos agregando parámetros libres al punto donde no quedan grados de libertad; es decir, la mayoría de los modelos recién identificados explicarán perfectamente las covarianzas observadas.

Los índices de ajuste incrementales (relativos, comparativos) miden la mejora relativa en el ajuste del modelo del investigador sobre un modelo de referencia. El modelo de referencia suele ser el modelo de independencia (nulo), que supone covarianzas de cero entre las variables endógenas. El modelo nulo en el programa de modelado estadístico Mplus, usa las covarianzas de muestra entre las variables exógenas, pero otros modelos de ecucaciones estructurales, como EQS y LISREL, también fijan las covarianzas entre pares de variables exógenas medidas a cero (Morata, 2017). Se debe verificar del modelo nulo en la documentación de su herramienta informática SEM. La suposición de cero covarianzas es a menudo inverosímil. Esta es la razón por la que los índices de ajuste incremental basados en el modelo nulo efectivamente dicen, ¿Cómo está mi modelo, en comparación con el peor modelo que existe?, esto significa que el modelo nulo es un argumento "hombre de paja" que probablemente sea falso.

Los índices ajustados por parsimonia incluyen en sus fórmulas una corrección o "penalización" por la complejidad del modelo. La misma pena puede ser vista como una "recompensa" por parsimonia. El modelo de parsimonia se define como la escasez de parámetros libremente estimados en relación con el número de observaciones. Parsimonia está relacionado con, pero los dos no son sinónimos. Esto se debe a que, no es una medida proporcional de la relación entre observaciones y parámetros. Por ejemplo, el valor “p”, puede ser relativamente alto cuando hay muchas variables observadas, pero aún puede ser proporcionalmente pequeño en comparación con el número de observaciones cuando hay muchos parámetros libres en un modelo muy complejo. Se define la relación de parsimonia (PR) como la relación del modelo del investigador sobre los grados de libertad del modelo nulo. El modelo ideal (el más parsimonioso) tendría un PR de 1.0, que dice que el modelo tiene tantos grados de libertad como hay disponibles en los datos (según el modelo nulo). Dichos modelos son potencialmente más no confirmables con los datos que los modelos en los que el valor del RP es inferior a 1.0 (Santos, 2017).

Los índices de ajuste predictivo estiman el ajuste del modelo en muestras de réplica hipotéticas del mismo tamaño y extraídas aleatoriamente de la misma población que la muestra original. Por lo tanto, estos índices pueden verse como basados en la población en lugar en lugar de la muestra.

También pueden corregir los grados de libertad del modelo o el tamaño de la muestra. Existe un contexto específico, para los índices de ajuste predictivo que se describe más adelante, pero la mayoría de las aplicaciones de SEM no se incluyen en él.

Las fórmulas de algunos índices de ajuste aproximados incluyen estadísticas de prueba del modelo. Existe una relación similar en los análisis de datos más estándar entre las estadísticas de prueba y las medidas del tamaño del efecto: muchos tamaños de efectos se pueden expresar como funciones de las estadísticas de prueba y viceversa (Kline, 2016). La relación entre las estadísticas de prueba del modelo y los índices de ajuste aproximados significa que ambos se basan en los mismos supuestos de distribución. Si estos supuestos son insostenibles, entonces los valores tanto del índice de ajuste aproximado, como del estadístico de prueba correspondiente (y su valor p) pueden ser inexactos.

Muchos estudios fueron realizados controlando el método de estimación y el tamaño de muestra, en la cual incluyen el cumplimiento de la normalidad multivariante y la complejidad del modelo estimado, para evaluar los índices de ajuste del AFC; sin embargo, estas evaluaciones de los índices de ajuste no fueron estudiados ampliamente en aplicación a casos reales, particularmente al cuestionario PHQ - 9 (Patient Health Questionnaire-9). De ello, radica el interés de fijar tamaños de muestra que evidencian la la invariabilidad de los índices de ajuste del AFC sobre el cuestionario PHQ - 9, aplicado a la población peruana.

MATERIALES Y MÉTODOS

Para llevar a cabo la presente investigación se utilizaron registros de la encuesta ENDES - INEI 2018, el tamaño de muestra estuvo conformado por muestra elegidas mediante el muestreo Bootstrap con 100 réplicas en cantidad de 50, 100, 250, 500, 750, 1000, 2500, 5000 y 10000 tamaño de muestra. Se aplicó el análisis factorial confirmatorio y se obtuvieron las medidas de bondad de ajuste para cada uno de los tamaños de muestra. El cálculo se realizó en base al cuestionario de salud del paciente (PHQ - 9), el cual es una herramienta de diagnóstico para trastornos de salud mental utilizada por profesionales de la salud que es rápida y fácil de completar para los pacientes (Borghero et al., 2018).

Los modelos analizados en SEM generalmente asumen causalidad probabilística, no causalidad determinista. Esto último significa que, dado un cambio en una variable causal, se observa la misma consecuencia en todos los casos en la variable de resultado. En contraste, la causalidad probabilística permite que ocurran cambios en los resultados con alguna probabilidad < 1.0.

La estimación de estas probabilidades (efectos) con datos de muestra, generalmente se basa en supuestos de distribución específicos, como la normalidad. La causalidad como una relación funcional entre dos variables cuantitativas se conserva en este punto de vista, pero se supone que los efectos causales cambian una distribución de probabilidad (Kline, 2016).

Los análisis factoriales confirmatorias llevadas a cabo se realizaron utilizando el método de máxima verosimilitud para la extracción de factores ortogonales. Las decisiones para aceptación o rechazo de la H_0 se hicieron a un nivel de confianza del 95 %.

RESULTADOS

La Tabla 1, muestra los valores poblacionales para cada una de las condiciones establecidas. Como puede observarse el valor de X^2 presenta dependencia del tamaño de la muestra, pues tiende a comportarse como se espera, en función del nivel de significación, a partir de tamaños muestrales por encima de 500 sujetos.

El análisis del ajuste del modelo factorial hipotetizado se realizó mediante el análisis factorial confirmatorio (AFC) con el programa estadístico Lavaan de R, usándose el estimador WLSMV (mínimos cuadrados ponderados con media y varianza ajustadas).

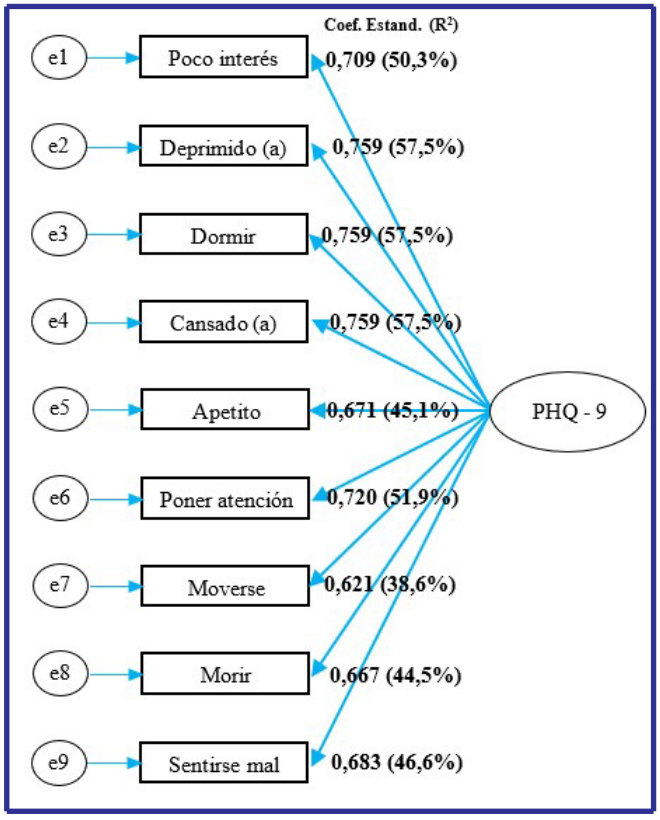
TABLA 1

Medidas de bondad de ajuste del modelo factorial confirmatorio paramétrico sobre el cuestionario PHQ - 9 aplicado a la población peruana, ENDES 2018

Índice de bondad de Ajuste		Valor
Absoluto	Chi-Cuadrado	9020,030
	Chi-Cuadrado / GL	334,075
	GFI	0,943
	RMR	0,015
	SRMR	0,047
	RMSEA	0,098
Parsimónico	AGFI	0,906
	PNFI	0,694
Comparativo	AIC	478202,180
	BIC	478354,244
Incremental	CFI	0,925
	TLI	0,901
	RNI	0,925
	NFI	0,925
	NNFI	0,901

FIGURA 1

Parámetros del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana, ENDES 2018



Este estimador tiene su importancia en la robustez de sus resultados en situaciones de no normalidad y escala ordinal de las variables analizadas (Kline, 2015; Lei & Wu, 2012), en la prueba piloto, para determinar la validez interna y luego se procesó en el programa SPSS v 21.

El tamaño de muestra presenta un efecto directo en los índices de ajuste del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana, tal como se muestran en las Figuras 2 al 16, donde se evidencia que el tamaño de muestra presenta un efecto directo con concavidad hacia arriba en el X^2, X^2/gl, AIC y BIC, efecto directo con concavidad hacia abajo en el CFI, TLI, NNFI, NFI, RNI, GFI, PNFI y AGFI, y en algunas medidas de bondad de ajuste un efecto indirecto en el RMSEA, RMR y SRMR. Por lo que se contrasta un efecto directo e inverso del tamaño de muestra sobre las medidas de bondad de ajuste.

FIGURA 2

Comportamiento de la medida X² según tamaño de muestra

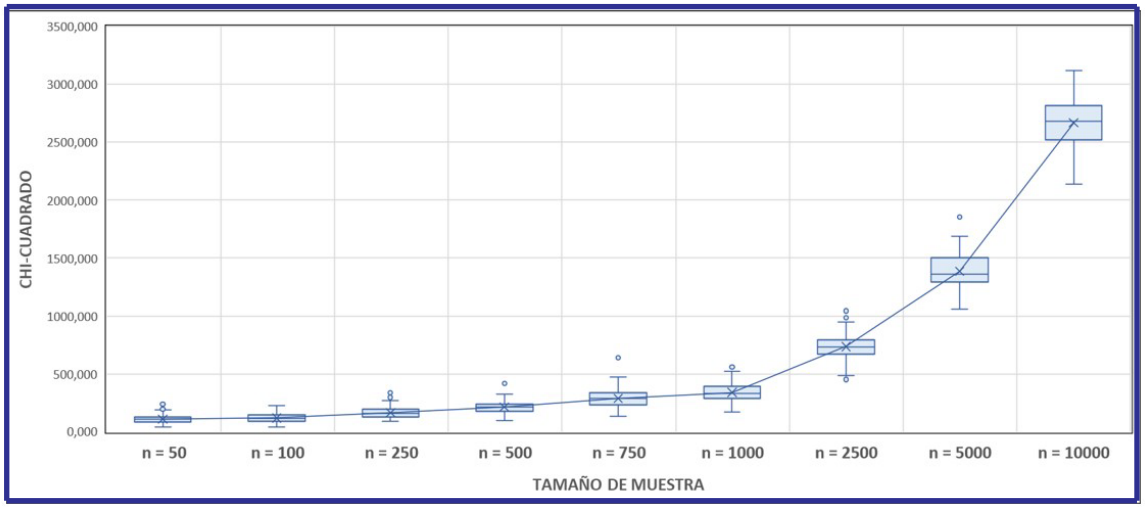


FIGURA 3

Comportamiento de la medida X²/gl según tamaño de muestra

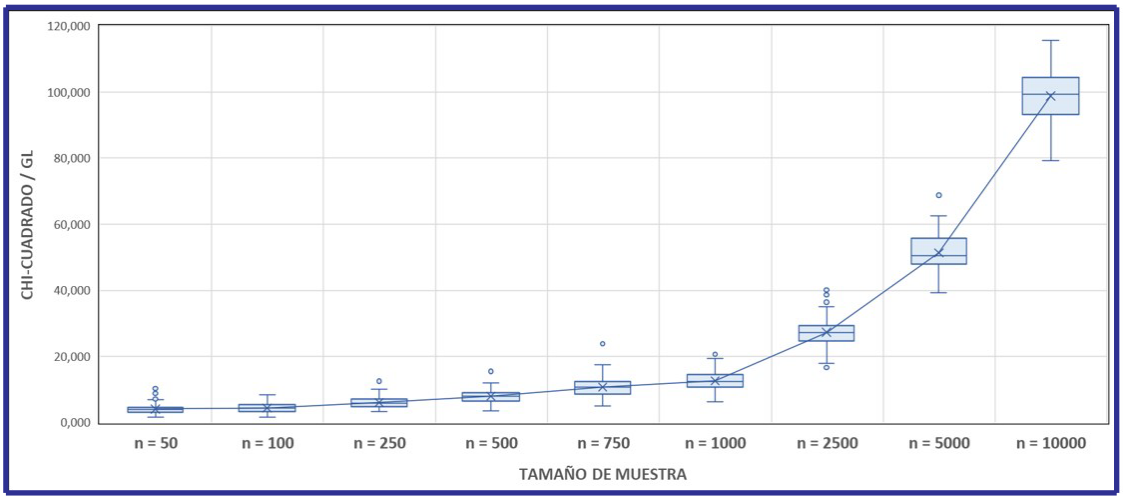


FIGURA 4

Comportamiento de la medida GFI según tamaño de muestra

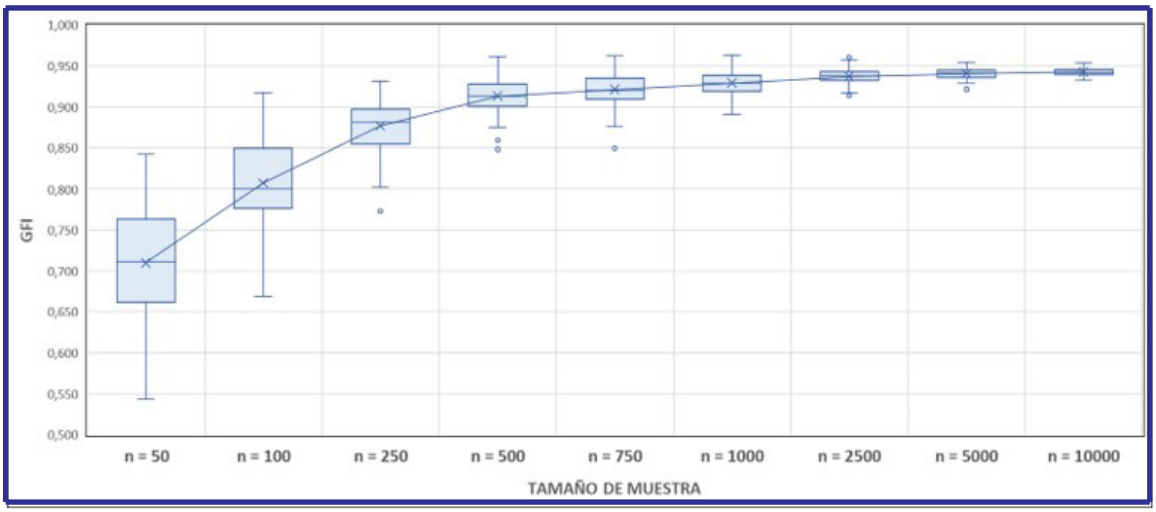


FIGURA 5

Comportamiento de la medida RMR según tamaño de muestra

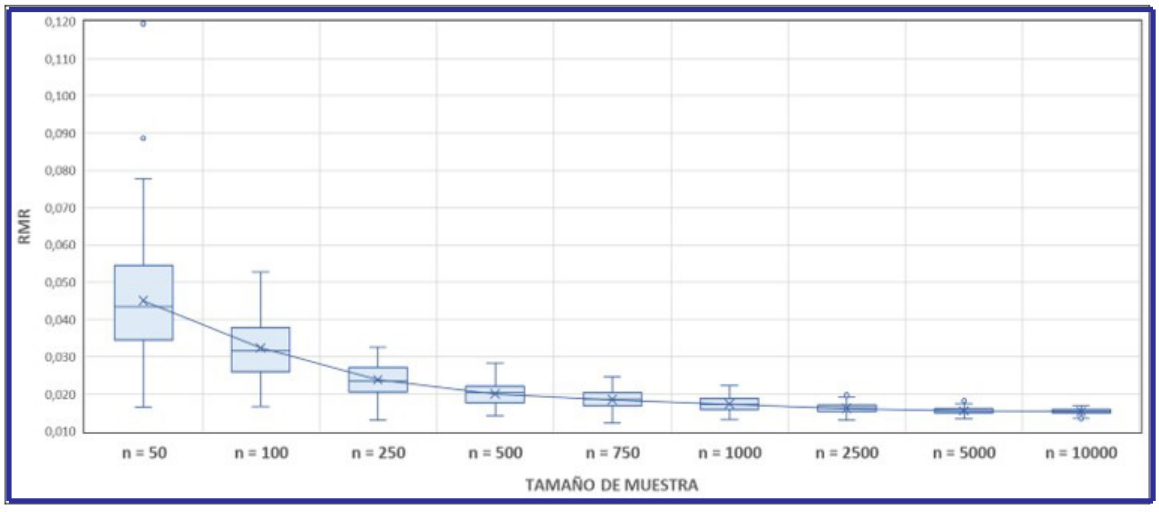


FIGURA 6

Comportamiento de la medida SRMR según tamaño de muestra

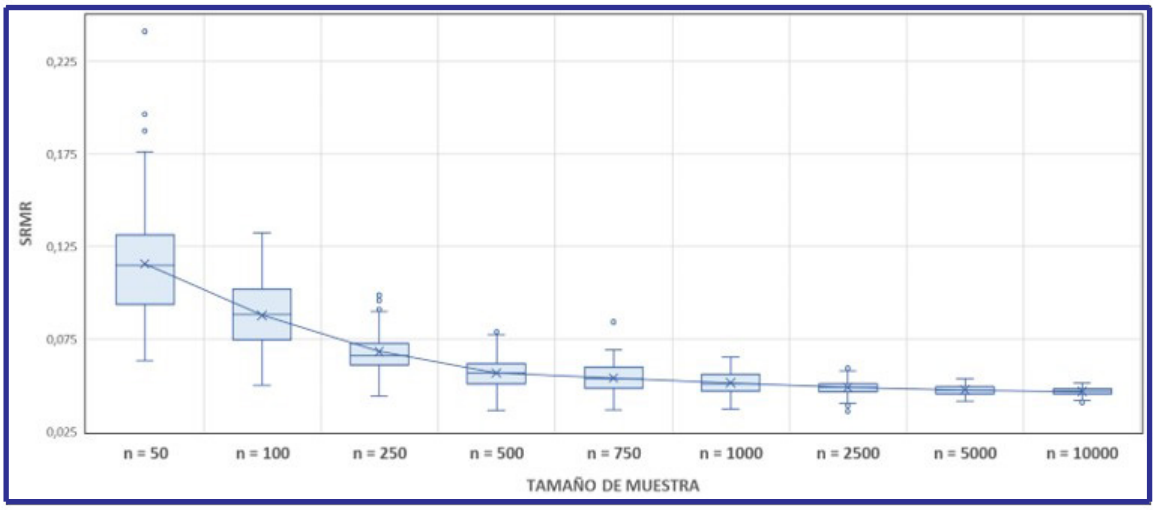


FIGURA 7

Comportamiento de la medida RMSEA según tamaño de muestra

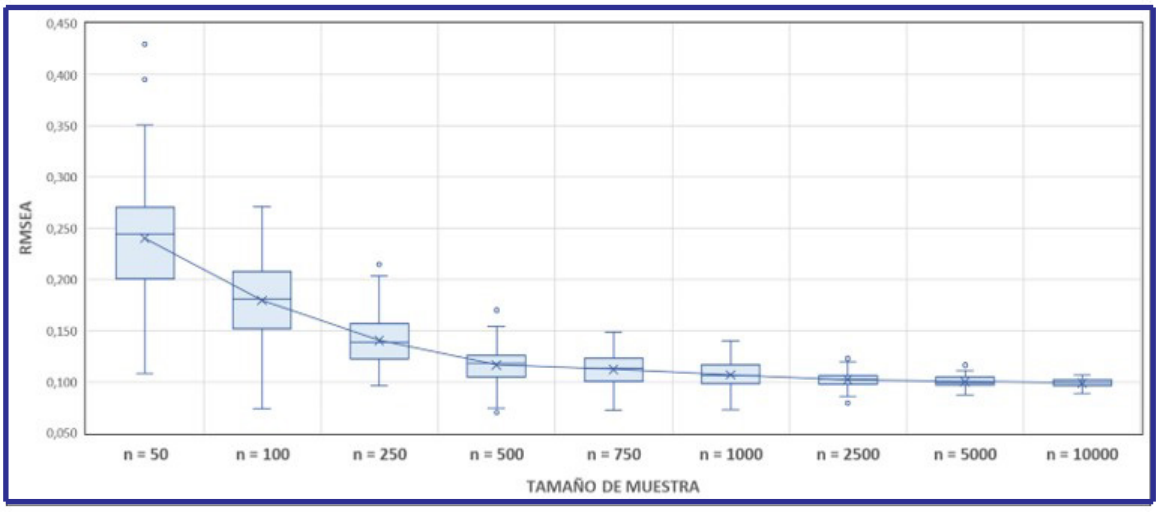


FIGURA 8

Comportamiento de la medida AGFI según tamaño de muestra

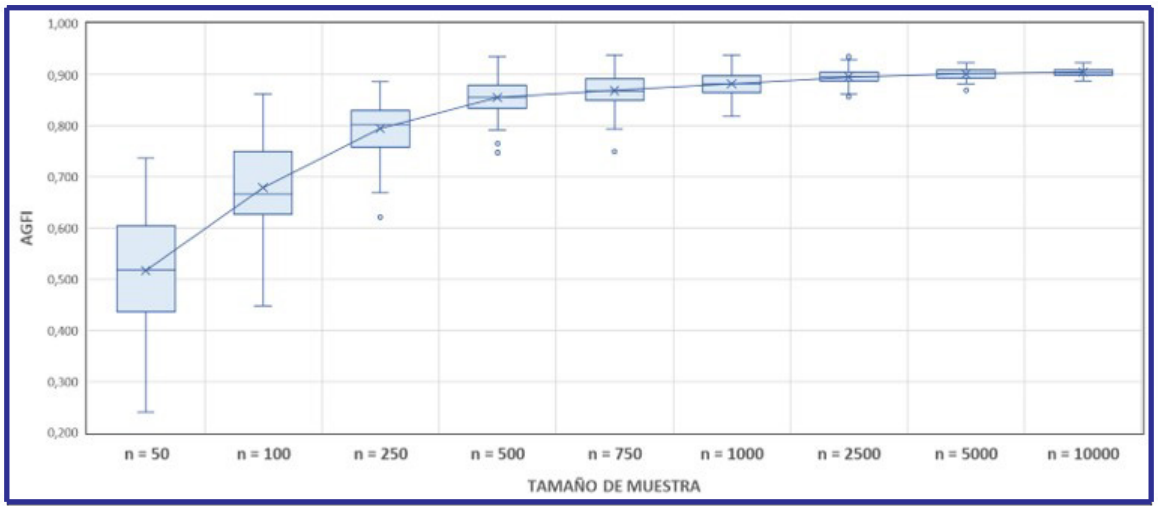


FIGURA 9

Comportamiento de la medida PNFI según tamaño de muestra

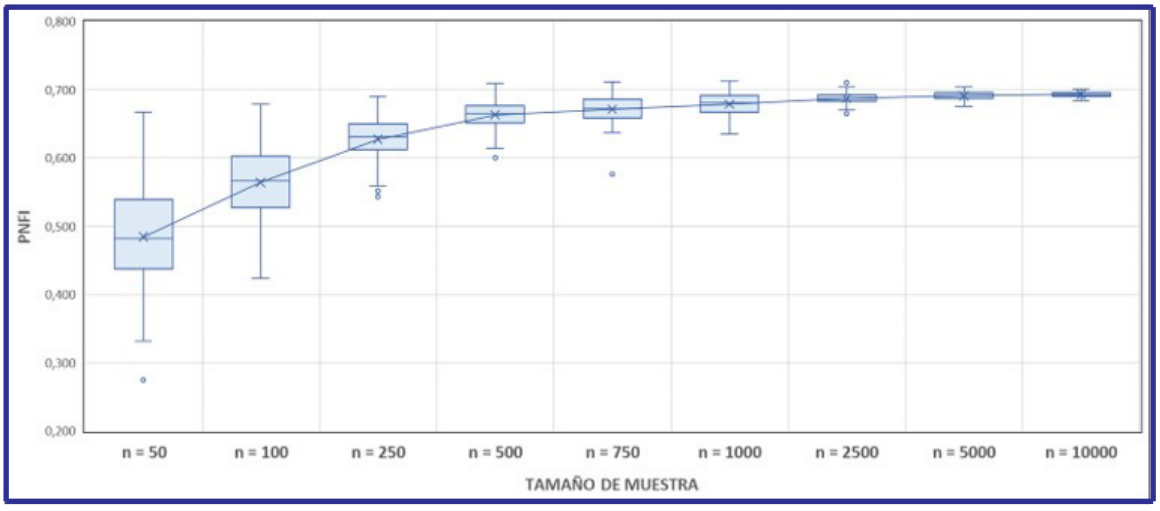


FIGURA 10

Comportamiento de la medida AIC según tamaño de muestra

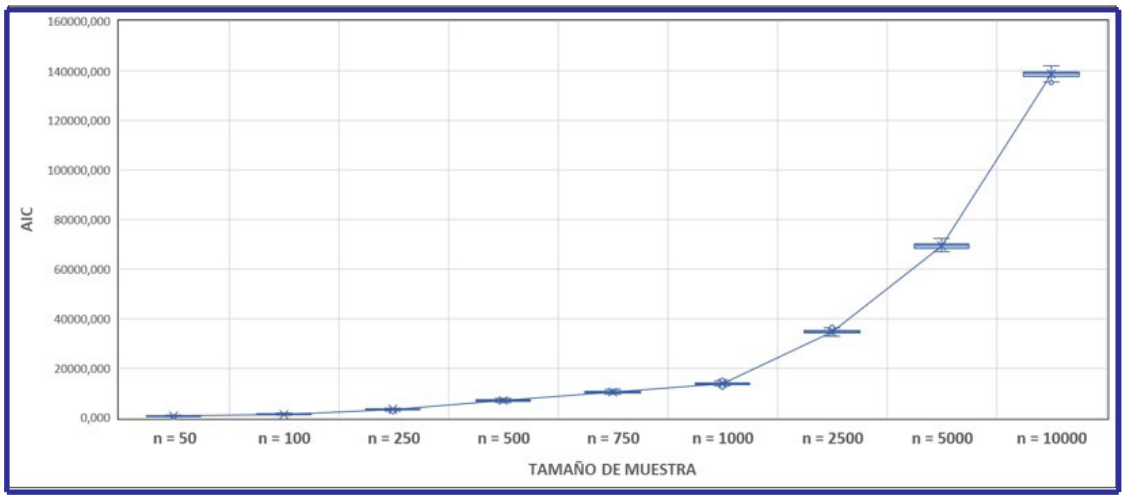


FIGURA 11

Comportamiento de la medida BIC según tamaño de muestra

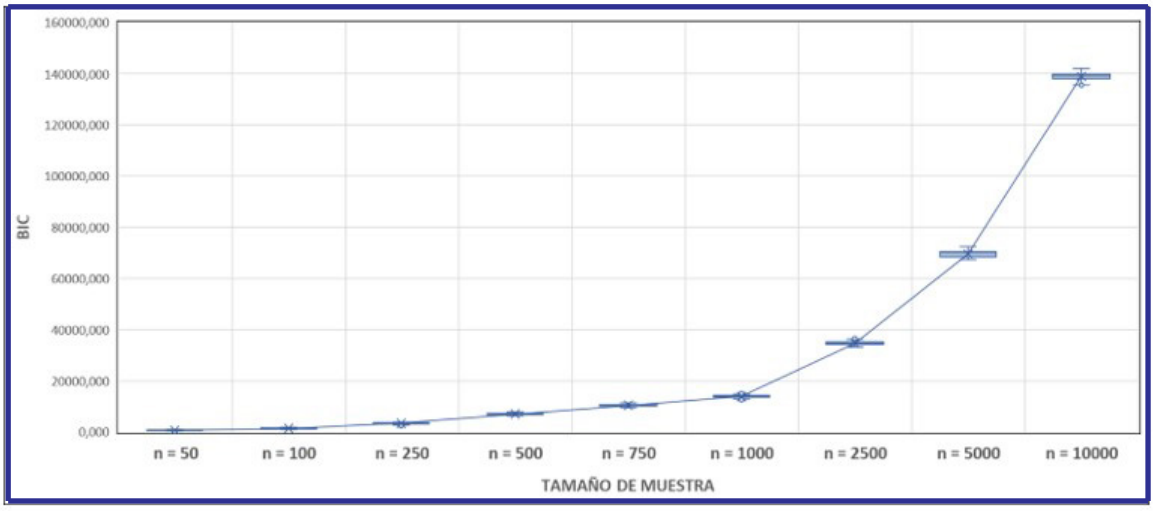


FIGURA 12

Comportamiento de la medida CFI según tamaño de muestra

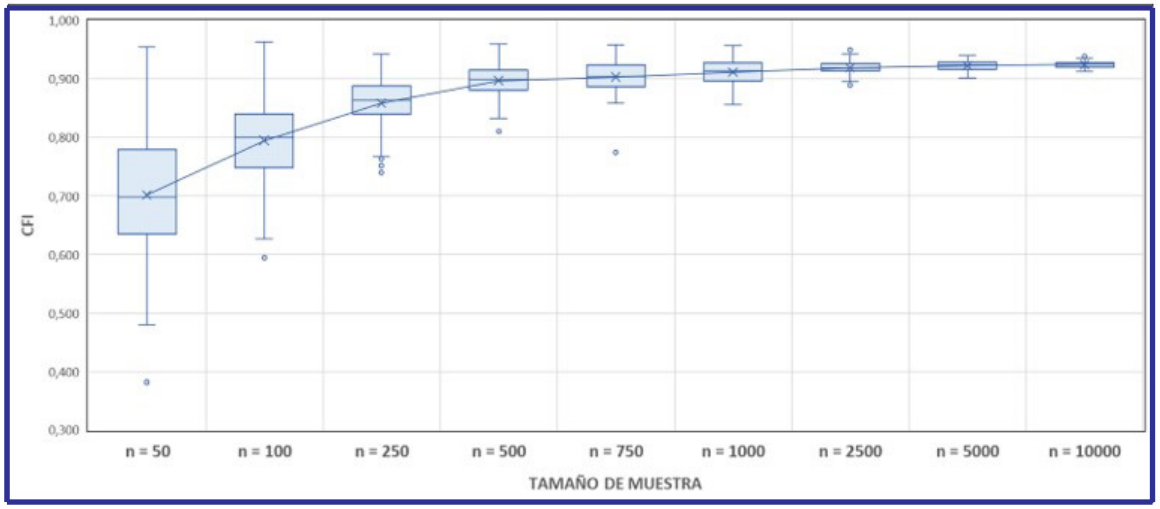


FIGURA 13

Comportamiento de la medida TLI según tamaño de muestra

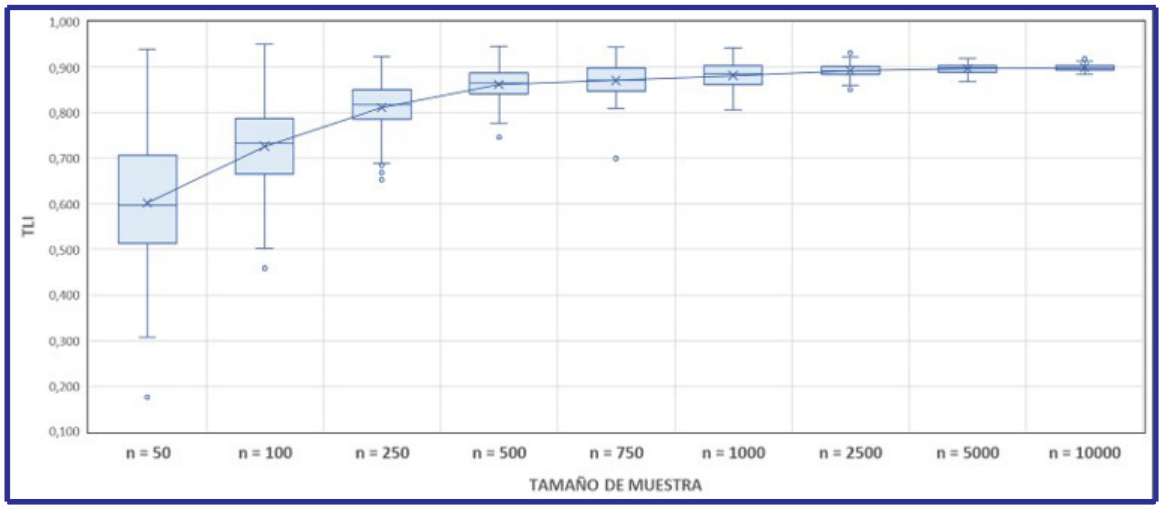


FIGURA 14

Comportamiento de la medida RNI según tamaño de muestra

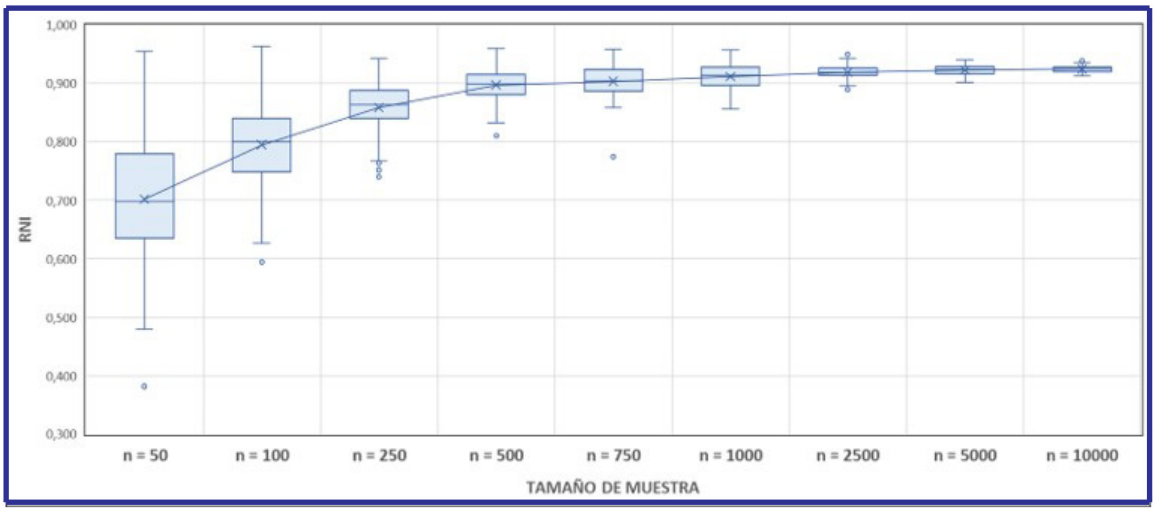


FIGURA 15

Comportamiento de la medida NFI según tamaño de muestra

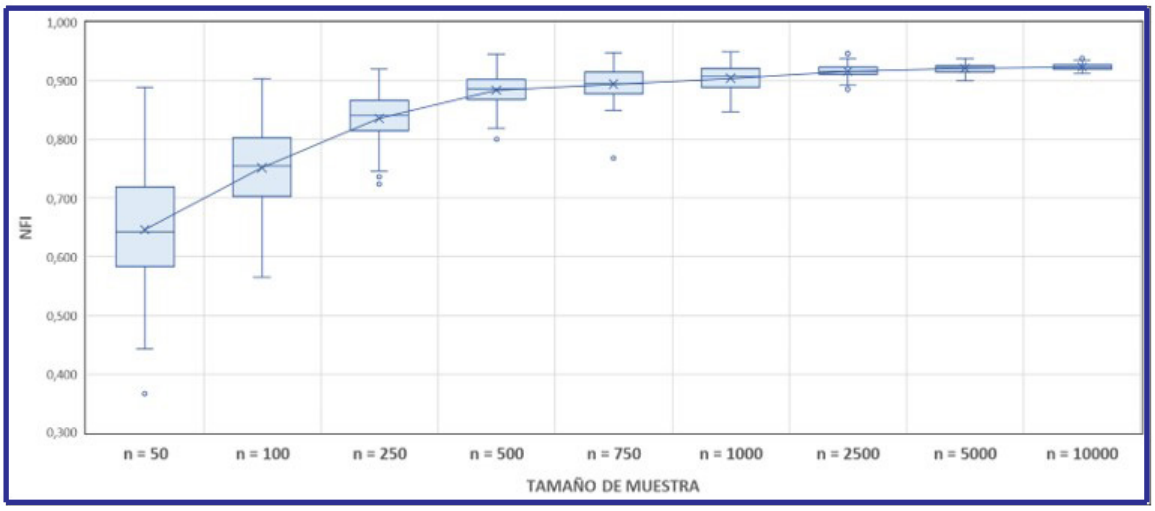
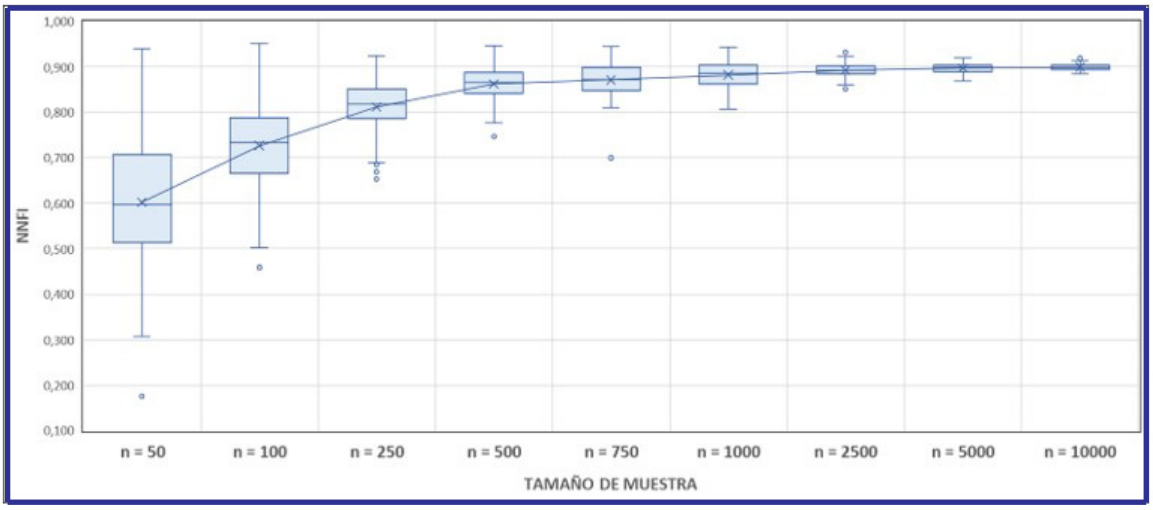


FIGURA 16

Comportamiento de la medida NNFI según tamaño de muestra



DISCUSIÓN

Las medidas de bondad de ajuste en su totalidad presentan efecto respecto al tamaño de muestra, tal como lo indicaron Cangur & Ercan (2015) y Ramírez et al., (1969); asimismo, en cuanto al GFI se observa un crecimiento cóncavo hacia abajo a medida que el tamaño de muestra crece lo cual coincide con los resultados obtenidos por Miles & Shevlin, (1998), donde concluyeron que el rendimiento del GFI mejora al aumentar el tamaño de la muestra. Los valores del RMSEA son favorecidos con los tamaños de muestras grandes, tal como lo indicaron MacCallum et al. (1996), en la que recomendaron el uso del RMSEA en casos con muestras muy grandes. Del mismo modo, los resultados del RMR y SRMR son favorecidos a medida que se cuenta con tamaños de muestras grandes. Por otro lado, los resultados muestran que las medidas de bondad de ajuste ya sea en tamaños de muestras pequeños o grandes no significan plenamente un mal o buen ajuste del modelo estimado, tal como lo indica labobucci (2010), en no tomar las reglas generales demasiado en serio, más por el contrario buscarle interpretabilidad y ser parsimonioso en el modelamiento de las variables de estudio.

CONCLUSIONES

- El tamaño de muestra presenta efecto directo o indirecto según corresponda en los índices de ajuste del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana.

- El tamaño de muestra presenta efecto directo en los índices de ajuste absoluto del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana.

- El tamaño de muestra presenta efecto directo en los índices de ajuste parsimónico del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana.

- El tamaño de muestra presenta efecto directo en los índices de ajuste comparativo del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana.

- El tamaño de muestra presenta efecto directo en los índices de ajuste predictivo del análisis factorial confirmatorio sobre el cuestionario PHQ - 9 aplicado a la población peruana.

AGRADECIMIENTOS

A los docentes de la Facultad de Ciencias Matemáticas por su participación y colaboración en el desarrollo de la investigación.

REFERENCIAS BIBLIOGRAFICAS

Borghero, F., Martínez, V., Zitko, P., Vöhringer, P., Cavada, G., & Rojas, G. (2018). Tamizaje de episodios depresivos en adolescentes. Validación PHQ - 9. *Revista Médica de Chile. 146*(4), 479 – 486. https://dx.doi.org/10.4067/s0034-98872018000400479

Cangur, S., & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods, 14*(1), 152 – 167. https://doi.org/10.22237/jmasm/1430453580

Escobedo, M., Hernández, J., Estebané, V., & Martínez, G. (2016). Modelos de ecuaciones estructurales: Características, fases, construcción, aplicación y resultados. *Ciencia & Trabajo, 18*(55), 16 - 22. https://dx.doi.org/10.4067/S0718-24492016000100004

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90 – 98. https://doi.org/10.1016/J.JCPS.2009.09.003

Instituto Nacional de Estadística e Informática [INEI] (2018). Encuesta Demográfica y de *Salud Familiar [ENDES].* https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1656/index1.html

Keith, T. (2019). *Multiple Regression and Beyond - An Introduction to Multiple Regression and Structural Equation Modeling (2nd ed).* https://doi.org/10.4324/9781315749099

Kline, R. (2016). *Principles and Practice of Structural Equation Modeling (4th ed).* https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462523344

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130 – 149. https://doi.org/10.1037/1082-989X.1.2.130

Miles, J., & Shevlin, M. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences, 25*(1), 85 – 90. https://doi.org/10.1016/S0191-8869(98)00055-5

Morata, M. (2017). *Métodos de estimación y sus implicaciones para la validación de constructo mediante análisis factorial confirmatorio de escalas tipo likert. Un estudio de simulación.* [Tesis Doctoral. Universidad Nacional de Educación a Distancia]. http://e-spacio.uned.es/fez/eserv/tesisuned:ED-Pg-MCySMamorata/MORATA_RAMIREZ_MAngeles_Tesis.pdf