

VALIDACIÓN DE LA ESCALA MEXICANA DE SEGURIDAD ALIMENTARIA A TRAVÉS DE CORRELACIÓN TETRACÓRICA CON BASE EN LA DISTRIBUCIÓN NORMAL ASIMÉTRICA BIVARIADA

VALIDATION OF MEXICAN FOOD SECURITY SCALE THROUGH TETRACHORIC CORRELATION BASED ON BIVARIATE SKEW NORMAL DISTRIBUTION

Ana Laura **Ramírez Jiménez**^{1*}, Martha Elva **Ramírez Guzmán**², Elizabeth **González Estrada**²

¹Oyametla Mz. 73 Lt. 8, Colonia San Bernabé Ocotepec, Magdalena Contreras, 10300, CDMX. (laurar_j@hotmail.com). ²Estadística. Campus Montecillo. Colegio de Postgraduados. 56230. Montecillo, Estado de México.

RESUMEN

La evaluación de la seguridad alimentaria en México se realiza por medio de la Escala Mexicana de Seguridad Alimentaria (EMSA) compuesta por reactivos de respuesta dicotómica que miden el acceso a los alimentos en los hogares con y sin menores de 18 años. Una manera de validar la EMSA es a través del Análisis de Factores Confirmatorio (AFC) al contrastar la estructura de las variables latentes con los datos observados. La correlación entre los reactivos de la EMSA se obtiene con la correlación tetracórica de una tabla de contingencia de doble entrada, la cual asume variables latentes con distribución Normal. Sin embargo, el supuesto de simetría restringe la admisión de variables que no cumplen con este supuesto, como las económicas que se caracterizan por su distribución asimétrica. El objetivo de esta investigación fue proponer un estimador de la correlación tetracórica para diferentes niveles de asimetría. Este coeficiente se denominó $\tilde{\rho}_{t_{SN}}$ el cual minimiza la diferencia entre la proporción de coincidencias de las variables dicotómicas en estudio y la probabilidad de la distribución Normal Asimétrica bivariada. El $\tilde{\rho}_{t_{SN}}$ se caracterizó por su consistencia y robustez estadística en los diferentes niveles de asimetría considerados. Los modelos AFC con los datos de la EMSA 2014 confirmaron la estructura teórica de la escala para adultos. En hogares con menores, el AFC sugirió una confusión en el entendimiento de los reactivos asociados con experiencias de hambre. El modelo sugiere que la falta de dinero o recursos limita la variedad de los alimentos y elimina algún tiempo de comida tanto en adultos como en menores. Estos hallazgos solo fueron visibles en el AFC con la inclusión del coeficiente $\rho_{t_{SN}}$ propuesto. El método de estimación propuesto permite analizar

ABSTRACT

Food security in Mexico is evaluated using the Mexican Food Security Scale (EMSA), composed of dichotomous-response items that measure access to food in households with and without children (members under 18 years-old). One way to validate EMSA is through Confirmatory Factor Analysis (CFA), by comparing the structures of latent variables with the observed data. The correlation between EMSA items is obtained using tetrachoric correlation on a double-entry contingency table, which assumes latent variables with a normal distribution. However, the symmetry assumption does restrict the admission of variables that do not satisfy this assumption, such as economic variables, which are characterized by a skewed distribution. The aim of this research was to propose an estimator of the tetrachoric correlation for different levels of skewness. This coefficient was denoted $\tilde{\rho}_{t_{SN}}$, which minimizes the difference between the proportion of coincidences of the dichotomous variables under study and the probability of the Bivariate Skew Normal distribution. $\tilde{\rho}_{t_{SN}}$ was characterized by its statistical consistency and robustness at the different levels of skewness considered. The CFA models with EMSA 2014 data confirmed the theoretical structure of the scale for adults. In households with children, the CFA suggested a confusion in the comprehension of items related to hunger experiences. The model suggests that a lack of money or resources limits the variety of food and forces either adults or children to skip one meal. These findings were only visible by the CFA with the inclusion of the proposed coefficient $\rho_{t_{SN}}$. The proposed estimation method helps analyze the correlation between dichotomous variables with a certain level of skewness, which can be defined with knowledge of the topic under study.

* Autor para correspondencia ♦ Author for correspondence.

Recibido: abril, 2020. Aprobado: agosto, 2020.

Publicado como ARTÍCULO en *Agrociencia* 55: 55-71. 2021.

la correlación entre variables dicotómicas con cierto nivel de asimetría, el cual puede definirse a partir del conocimiento del tema en estudio.

Palabras clave: correlación tetracórica, normal asimétrica, simulación Monte Carlo, análisis de factores confirmatorio, validez, Escala Mexicana de Seguridad Alimentaria.

INTRODUCCIÓN

La medición de la seguridad alimentaria en México se realiza por medio de la Escala Mexicana de Seguridad Alimentaria (EMSA). Esta escala es un conjunto de 12 reactivos de respuesta dicotómica que permite la identificación de cuatro posibles grados de inseguridad alimentaria: inseguridad alimentaria severa; inseguridad alimentaria moderada; inseguridad alimentaria leve, y seguridad alimentaria. A partir de los grados de inseguridad alimentaria se construye el indicador de carencia por acceso a la alimentación, de modo que la población considerada como carente es aquella que vive en hogares con inseguridad alimentaria moderada o severa (CONEVAL, 2010).

Ya que las escalas se utilizan para cuantificar conceptos teóricos que no se pueden medir directamente, la validez de dichos instrumentos constituye un elemento básico en la comprensión del tema en cuestión. Es decir, que las escalas, como todos los instrumentos de medición, deben tener la capacidad de medir el constructo que pretenden cuantificar, de modo que, los resultados que se generan a partir de ellas adquieren sentido únicamente si provienen de una herramienta válida desde el punto de vista científico (Pérez-Gil, 2000). La validez se refiere a la capacidad de reflejar el rasgo que se pretende medir por medio de las variables observables, es decir, el grado en el cual la teoría y la evidencia empírica respaldan las conclusiones obtenidas por medio de los resultados de la escala. Asimismo, en el caso de las escalas de seguridad alimentaria, la validez posibilita la identificación de los grupos de riesgo y vulnerables en el tema de la alimentación y enfrentar de manera efectiva la problemática de la seguridad alimentaria.

Con base en la metodología propuesta por FAO (2012), Villagómez *et al.* (2014) realizaron pruebas de validación a la EMSA por medio del ordenamiento de reactivos, coeficiente de Cronbach, paralelismo de curvas de prevalencia y el modelo de Rasch;

Key words: tetrachoric correlation, skew normal, Monte Carlo simulation, confirmatory factor analysis, validity, Mexican Food Security Scale (EMSA).

INTRODUCTION

Food security in Mexico is measured using the Mexican Food Security Scale (EMSA). This scale is a set of 12 dichotomous-response items that allow the identification of four possible degrees of food insecurity: severe food insecurity, moderate food insecurity, mild food insecurity and food security. From the degrees of food insecurity, the indicator of lack of access to food is constructed; thus, the population that is considered deprived is formed by those who live in households with moderate or severe food insecurity (CONEVAL, 2010).

Since the scales are used to quantify theoretical concepts that cannot be measured directly, the validity of such instruments constitutes a basic element in the comprehension of the subject at hand. That is, scales, like all measurement instruments, must be able to measure the construct which they intend to quantify, so that results produced from them are only meaningful if they come from a scientifically validated tool (Pérez-Gil, 2000). Validity refers to the ability to reflect the trait to be measured using observable variables, *i.e.* the degree to which theory and empirical evidence support the conclusions obtained by the results of the scale. Likewise, in the case of food security scales, validity helps identifying groups at risk and vulnerable groups in terms of food security, in order to tackle the problem of food security in an effective way.

Based on the methodology proposed by FAO (2012), Villagómez *et al.* (2014) performed validation tests on EMSA by sorting of items, calculating the Cronbach coefficient, the parallelism of prevalence curves and the estimation of Rasch model. They concluded that the EMSA has the conceptual and technical solidity of its results. However, there are statistical alternatives to analyze the validity of the scale, and therefore reinforce its scientific basis. The Confirmatory Factor Analysis (CFA) is one of them, and uses correlation matrix of the observed variables as an input. In addition, the CFA is considered an adequate analytical tool to empirically test the theoretical structure of an instrument. From the relationship between the latent characteristics it

concluyeron que la EMSA cuenta con la solidez conceptual y técnica de sus resultados. Sin embargo, existen otras alternativas estadísticas para analizar la validez de la escala y con ello, reforzar su fundamento científico. El Análisis de Factores Confirmatorio (AFC) es una de ellas y utiliza como insumo la matriz de correlación de las variables observadas. Además, el AFC se considera como una herramienta analítica apropiada para probar de manera empírica la estructura teórica de un instrumento. A partir de la relación entre los rasgos latentes que representa y sus variables manifiestas, permite la incorporación de una teoría o hipótesis establecida previamente para formalizar el modelo que será puesto a prueba.

La correlación entre los reactivos de la EMSA se obtiene por medio de la correlación tetracórica, la cual se basa en variables latentes con distribución Normal. Sin embargo, el supuesto de simetría restringe la admisión de variables que no cumplen con este supuesto, como las económicas que se caracterizan por su distribución asimétrica.

Por lo tanto, en esta investigación se propone un método para estimar la correlación tetracórica que incluye el nivel de asimetría y realizar la validación de la EMSA por medio del AFC con matrices de correlación que consideran distintos niveles de asimetría.

MATERIALES Y MÉTODOS

Fuente de información

La Escala Mexicana de Seguridad Alimentaria (EMSA) es un conjunto de 12 reactivos que miden el concepto teórico de la seguridad alimentaria a través de preguntas con las cuales es posible conocer, si por falta de dinero o recursos, los hogares enfrentan cambios tanto en la cantidad como en la calidad de los alimentos, e incluso, si han tenido experiencias de hambre. Los reactivos son de respuesta dicotómica (Sí=1, No=0), se ordenan en un continuo de severidad y permiten distinguir el acceso a los alimentos de los adultos y el de los menores de 18 años (Villagómez *et al.*, 2014). Los reactivos de la EMSA son los siguientes:

Reactivos para adultos

ALIM_01: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez usted o algún adulto en su hogar tuvo una alimentación basada en muy poca variedad de alimentos?

ALIM_02: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez usted o algún adulto en su hogar dejó de desayunar, comer o cenar?

represents, and its manifest variables, CFA helps incorporate a previously established theory or hypothesis to formalize the model to be tested.

The correlation between EMSA items is obtained by tetrachoric correlation, which is based on latent variables with a normal distribution. However, the symmetry assumption does restrict the admission of variables that do not satisfy this assumption, such as economic variables, which are characterized by their skewed distribution.

Therefore, this research proposes a method to estimate the tetrachoric correlation that includes the level of skewness; and to validate EMSA using CFA with correlation matrices considering different levels of skewness.

MATERIALS AND METHODS

Data source

The Mexican Food Security Scale (EMSA) is a set of 12 items that measure the theoretical concept of food security using questions to find out if households face changes in food quantity and quality, or even if they have experienced hunger, due to lack of money or resources. The items have dichotomous responses (Yes=1, No=0), they are ordered in a continuum of severity and are used to identify degrees of access to food in adults and members under 18 years-old (Villagómez *et al.*, 2014). The EMSA items are as follows:

Items for adults

ALIM_01: In the past three months, due to lack of money or resources, have you or has any adult in your household had a diet based on a very little variety of food?

ALIM_02: In the past three months, due to lack of money or resources, have you or has any adult in your household had to skip breakfast, lunch or dinner?

ALIM_03: In the past three months, due to lack of money or resources, have you or has any adult in your household eaten less than you think you should have?

ALIM_04: In the past three months, due to a lack of money or resources, have you run out of food?

ALIM_05: In the past three months, due to lack of money or resources, were you or was any adult in your household hungry, but did not eat?

ALIM_06: In the past three months, due to lack of money or resources, have you or has any adult in your household eaten only once a day or not at all in an entire day?

ALIM_03: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez usted o algún adulto en su hogar comió menos de lo que usted piensa debía comer?

ALIM_04: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez se quedaron sin comida?

ALIM_05: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez usted o algún adulto en su hogar sintió hambre pero no comió?

ALIM_06: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez usted o algún adulto en su hogar sólo comió una vez al día o dejó de comer todo un día?

Reactivos para menores de 18 años

ALIM_07: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez algún menor de 18 años en su hogar tuvo una alimentación basada en muy poca variedad de alimentos?

ALIM_08: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez algún menor de 18 años en su hogar comió menos de lo que debía?

ALIM_09: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez tuvieron que disminuir la cantidad servida en las comidas a algún menor de 18 años del hogar?

ALIM_10: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez algún menor de 18 años en su hogar sintió hambre pero no comió?

ALIM_11: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez algún menor de 18 años en su hogar se acostó con hambre?

ALIM_12: En los últimos tres meses, por falta de dinero o recursos, ¿alguna vez algún menor de 18 años en su hogar comió una vez al día o dejó de comer todo un día?

Los resultados de la EMSA son de acceso público y están disponibles en la página oficial del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL).

El Análisis de Factores Confirmatorio es una técnica estadística que se utiliza en la validación de escalas. Con base en la correlación existente entre las variables medibles se contrasta la correspondencia teórica con el constructo que se quiere cuantificar a partir de la especificación de un modelo hipotético (Suárez, 2015).

Análisis de Factores Confirmatorio

El Análisis de Factores Confirmatorio (AFC) es una técnica estadística que permite especificar la relación entre la variable latente o constructo que se desea medir y las variables observables, también llamadas variables manifiestas.

Items for individuals under 18 years-old

ALIM_07: In the past three months, due to lack of money or resources, has any underage person in your household had a diet based on a very little variety of food?

ALIM_08: In the past three months, due to lack of money or resources, has any underage person in your household eaten less than you think they should have?

ALIM_09: In the past three months, due to lack of money or resources, did you ever have to serve less food to any underage person in your household?

ALIM_10: In the past three months, due to lack of money or resources, was any underage person in your household hungry but did not eat?

ALIM_11: In the past three months, due to lack of money or resources, did any underage person in your household go to bed hungry?

ALIM_12: In the past three months, due to lack of money or resources, has any underage person in your household eaten only once a day or not at all in an entire day?

The results of the EMSA are of public access and available on the official website of the Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL).

The Confirmatory Factor Analysis is a statistical technique used to validate scales. Based on the existing correlation between measurable variables, the theoretical correspondence is contrasted with the construct to be quantified from the specification of a hypothetical model (Suárez, 2015).

Confirmatory Factor Analysis

The Confirmatory Factor Analysis (CFA) is a statistical technique to specify the relation between a latent variable or construct to be measured and observed variables, also called manifest variables.

The matrix representation of the existing relation between observed and latent variables corresponds to the following model:

$$\underline{X} = \underline{\Lambda}\underline{F} + \underline{\varepsilon} \quad (1)$$

where $\underline{X}$ is the p -dimensional vector of observable endogenous indicators, $\underline{\Lambda}$ is the $p \times m$, dimensional factorial loadings matrix, $\underline{F}$ is the m -dimensional vector of factors or latent variables and $\underline{\varepsilon}$ is the p -dimensional vector of measurement errors. The CFA model considers that the factors and the measurement errors are independent, *i.e.* $\text{corr}(\underline{F}, \underline{\varepsilon}) = 0$.

La representación matricial de la relación existente entre las variables observables y las variables latentes corresponde al siguiente modelo:

$$\underline{X} = \Lambda \underline{F} + \underline{\varepsilon} \quad (1)$$

donde $\underline{X}$ es el vector de p indicadores endógenos observables, Λ es la matriz de cargas factoriales de dimensión $p \times m$, $\underline{F}$ es el vector de m factores o variables latentes y $\underline{\varepsilon}$ es el vector de los p errores de medición. El modelo AFC considera que los factores y los errores de medición son independientes, es decir, $\text{corr}(\underline{F}, \underline{\varepsilon}) = 0$.

Si $\underline{F} \sim N_m(\underline{0}, I)$ y $\underline{\varepsilon} \sim N_p(\underline{0}, \Theta_\varepsilon)$, entonces $\underline{X} \sim N_p(\underline{0}, \Sigma)$. Dado que $\text{Cov}(\underline{F}, \underline{X}) = \Lambda$, la matriz de varianzas y covarianzas Σ puede expresarse como:

$$\Sigma = \Lambda \Lambda' + \Theta_\varepsilon \quad (2)$$

En la diagonal principal de Θ_ε se encuentran las varianzas de los errores de medición de cada indicador endógeno observable y fuera de la diagonal están las covarianzas entre los errores de medición. Por lo tanto, la matriz Θ_ε forma parte del modelo de AFC. Lo anterior representa una de las principales ventajas del AFC, dado que permite incluir explícitamente al error de medición (ε) en el modelo, a diferencia de los modelos clásicos de regresión.

La hipótesis principal a contrastar en el AFC es que si el modelo es correcto, la matriz de varianzas y covarianzas poblacional (Σ) puede reproducirse por la matriz de varianzas y covarianzas asociada con el modelo teórico ($\Sigma(\theta)$). Donde θ puede sustituirse por $\hat{\theta}$, un estimador de máxima verosimilitud, mínimos cuadrados ordinarios, mínimos cuadrados generalizados, entre otros, a fin de que la diferencia entre $\Sigma(\hat{\theta})$ y Σ sea mínima. Dado que Σ no es conocida, se utiliza la matriz de covarianzas de la muestra (S) como el estimador de Σ en (2).

Medidas de bondad de ajuste

Kline (2011) propuso que la evaluación del ajuste del modelo se realice en función de índices de ajuste globales e individuales. Los índices globales evalúan el grado de ajuste del modelo en su conjunto, mientras que los índices individuales permiten la evaluación de los parámetros del modelo. Los índices de ajuste global que se usan con mayor frecuencia se enuncian a continuación (Acock, 2013).

Índice Comparativo de Ajuste (CFI, por sus siglas en inglés). Compara el ajuste del modelo propuesto y el ajuste de un modelo

If $\underline{F} \sim N_m(\underline{0}, I)$ and $\underline{\varepsilon} \sim N_p(\underline{0}, \Theta_\varepsilon)$, then $\underline{X} \sim N_p(\underline{0}, \Sigma)$. Given $\text{Cov}(\underline{F}, \underline{X}) = \Lambda$, variances and covariances matrix Σ can be expressed as:

$$\Sigma = \Lambda \Lambda' + \Theta_\varepsilon \quad (2)$$

In the main diagonal Θ_ε there are the variances of the measurement errors of each observed endogenous indicator; and outside the diagonal, there are the covariances between measurement errors. Hence, matrix Θ_ε is a part of the CFA model. This is one of the main advantages of the CFA, since it helps explicitly include measurement error (ε) in the model, unlike classic regression models.

The main hypothesis to contrast in the AFC is that, if the model is correct, variance and covariance population matrix (Σ) can be reproduced with the variance and covariance matrix associated with the theoretical model ($\Sigma(\theta)$), where θ can be substituted for $\hat{\theta}$, a maximum likelihood estimator, ordinary least squares, generalized least squares, and others, so that the difference between $\Sigma(\hat{\theta})$ and Σ be minimal. Given that Σ is unknown, sample covariance matrix (S) is used as the estimator of Σ in (2).

Goodness-of-fit Indices

Kline (2011) proposed that evaluation of the goodness of fit of the model be carried out based on global and individual fit indices. Global indices evaluate the degree of fit of the model in its entirety, whereas individual indices help evaluate parameters of the model. The most frequently used global fit indices are presented below (Acock, 2013).

Comparative Fit Index (CFI). It compares the adjustment of the proposed model and the adjustment of a model of independent variables, *i.e.* the one in which all covariances are equal to zero. The CFI is defined as follows:

$$CFI = 1 - \frac{\text{máximo}[(\chi_m^2 - g_m^l), 0]}{\text{máximo}[(\chi_0^2 - g_0^l), 0]}$$

where χ_m^2 and g_m^l are the values of the chi-square statistic and the degrees of freedom for the proposed model, respectively; χ_0^2 and g_0^l are the values of the chi-square statistic and the degrees of freedom for the independent variables model, respectively.

Root Mean Square Error of Approximation (RMSEA). It measures the difference between $\Sigma(\hat{\theta})$ and S. This is a

de variables independientes; es decir, en el cual todas las covarianzas son cero. El CFI se calcula como:

$$CFI = 1 - \frac{\text{máximo}[(\chi_m^2 - g_m), 0]}{\text{máximo}[(\chi_0^2 - g_0), 0]}$$

donde χ_m^2 y g_m son los valores del estadístico ji-cuadrada y los grados de libertad del modelo propuesto, respectivamente; χ_0^2 y g_0 son los valores del estadístico ji-cuadrada y los grados de libertad del modelo de variables independientes, respectivamente.

Error de aproximación cuadrático medio (RMSEA, por sus siglas en inglés). Mide la diferencia entre las matrices $\Sigma(\hat{\theta})$ y S . Este es un índice de parsimonia, ya que su valor disminuye a medida que el modelo tiene más grados de libertad o el tamaño de muestra (n) crece.

$$RMSEA = \sqrt{\text{máximo}\left\{\left[\frac{(\chi_m^2 - g_m), 0}{g_m(n-1)}\right], 0\right\}}$$

Residuo cuadrático medio estandarizado (SRMR, por sus siglas en inglés). Es la raíz cuadrada del promedio de los residuales al cuadrado. Si el cálculo de los residuales se realiza a partir de las matrices de correlación, el SRMR estará en términos de la correlación residual media y facilita su interpretación porque las correlaciones están en la misma escala, lo cual no sucede con las covarianzas que varían respecto a la variable y a la unidad de medida.

Residuales estandarizados (STD RES, por sus siglas en inglés). Diferencia estandarizada entre las covarianzas observadas (S) y las covarianzas estimadas por el modelo $\Sigma(\hat{\theta})$.

Los índices de ajuste individual, o índices de ajuste del modelo de medida, se refieren a la evaluación de la significancia estadística del vector de parámetros estimados $\hat{\theta}$, los porcentajes de varianza explicada por el factor y el porcentaje explicado por los errores de medición, así como la magnitud y el sentido de las cargas factoriales.

Cuando las variables observables consideradas en el AFC son de tipo cuantitativo, la estimación de los parámetros del modelo se realiza a partir de la matriz de varianzas y covarianzas, mientras que para variables de tipo cualitativo se realiza a partir de la matriz de correlación (Byrne, 2016). En el caso de variables dicotómicas, la prueba χ^2 de independencia de dos variables no indica la magnitud ni la dirección de la asociación. Una alternativa para ello es la correlación tetracórica (Pearson, 1900), la cual permite

parsimony index, since its value decreases as the model has more degrees of freedom or the sample size (n) grows.

$$RMSEA = \sqrt{\text{máximo}\left\{\left[\frac{(\chi_m^2 - g_m), 0}{g_m(n-1)}\right], 0\right\}}$$

Standardized Root Mean Square Residual (SRMR). It is the square root of the average square of residuals. If the calculation of residuals is obtained from correlation matrices, the SRMR will result in terms of the mean residual correlation, thus making its interpretation easier, because correlations are in the same scale. This is not the case with covariances, due to they vary in relation to the variable and the measurement unit.

Standardized Residuals (STD RES). Standardized difference between observed covariances (S) and the covariances estimated using the model $\Sigma(\hat{\theta})$.

Individual fit indices, or measurement model fit indices, refer to the evaluation of statistical significance of the vector of estimated parameters $\hat{\theta}$, the proportion of variance explained by the factor and the proportion explained by measurement errors, as well as the magnitude and the direction of the factor loadings.

When observed variables considered in the CFA are quantitative, estimation of parameters of the model is carried out from the variance and covariance matrix, whereas for qualitative variables, is carried out from the correlation matrix (Byrne, 2016). In the case of dichotomous variables, the χ^2 test of independence of two variables does not indicate the magnitude nor the direction of the association. An alternative for this is the tetrachoric correlation (Pearson, 1900), which helps finding the magnitude and the direction of the association. Kline (2011) proposed the use of the tetrachoric correlation in the CFA model (1) to analyze dichotomous variables.

Tetrachoric correlation

If we consider that the measurement of two variables, X and Y , with dichotomous response (0 and 1) of n individuals is summarized in a contingency table, where a , b , c and d represent the observed frequencies of each combination of levels of X and Y ; then, the proportions related to these frequencies are displayed in Figure 1.

Let us suppose that observed variables X and Y are related to continuous latent variables X^* and Y^* , respectively. Let z_1 and z_2 be the thresholds to identify each one of the categories of dichotomous variables X and Y as follows:

conocer la magnitud y la dirección de la asociación. Kline (2011) propuso utilizar la correlación tetracórica en el modelo de AFC (1) para analizar variables dicotómicas.

Correlación tetracórica

Si se considera que la medición de dos variables, X y Y , con respuesta dicotómica (0 y 1) de n individuos se resume en una tabla de contingencia, donde a , b , c y d representan las frecuencias observadas de cada combinación de los niveles de X y Y ; entonces, las proporciones asociadas con estas frecuencias se muestran en la Figura 1.

Al suponerse que las variables observables, X y Y , se relacionan con variables latentes continuas X^* y Y^* , respectivamente; y cuando z_1 y z_2 son umbrales que permiten identificar cada una de las categorías de las variables dicotómicas X y Y de la siguiente manera:

$$X = \begin{cases} 0 & \text{si } X^* \leq z_1 \\ 1 & \text{si } X^* > z_1 \end{cases} \quad Y = \begin{cases} 0 & \text{si } Y^* \leq z_2 \\ 1 & \text{si } Y^* > z_2 \end{cases}; \text{ se asume que el vector aleatorio } (X^*, Y^*)' \text{ tiene una distribución Normal estándar}$$

bivariada $N_2(\mathbf{0}, \mathbf{\Omega})$, donde $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}$ y ρ_t es la correlación entre las variables latentes X^* y Y^* .

Los cuantiles que definen las proporciones marginales de X y Y a partir de la función de distribución Normal Estándar Φ son:

$$z_1 = \Phi^{-1}\left(\frac{a+b}{n}\right) \\ z_2 = \Phi^{-1}\left(\frac{a+c}{n}\right)$$

Los umbrales z_1 y z_2 definen el punto de corte que divide a la distribución Normal estándar bivariada en cuatro cuadrantes, cuyas probabilidades corresponden a las proporciones observadas (Figura 1).

El coeficiente de correlación tetracórica ρ_t , definido por Pearson (1900), satisface que:

$$\frac{a}{n} = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} \varphi(x^*, y^*, \rho_t) dx^* dy^*$$

donde $\varphi(\cdot)$ es la función de densidad de la Normal estándar bivariada.

Este coeficiente de correlación no considera el nivel de asimetría en las variables latentes, el cual puede incorporarse por medio de la distribución normal asimétrica.

		Y		
		0	1	
X	0	$\frac{a}{n}$	$\frac{b}{n}$	$\frac{a+b}{n}$
	1	$\frac{c}{n}$	$\frac{d}{n}$	$\frac{c+d}{n}$
		$\frac{a+c}{n}$	$\frac{b+d}{n}$	1

Figura 1. Tabla de proporciones de las combinaciones de las variables dicotómicas X y Y .

Figure 1. Table of proportions for the combinations of dichotomous variables X and Y .

$$X = \begin{cases} 0 & \text{si } X^* \leq z_1 \\ 1 & \text{si } X^* > z_1 \end{cases} \quad Y = \begin{cases} 0 & \text{si } Y^* \leq z_2 \\ 1 & \text{si } Y^* > z_2 \end{cases};$$

The random vector $(X^*, Y^*)'$ is assumed to have a bivariate standard normal distribution $N_2(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}$ and ρ_t is the correlation between latent variables X^* and Y^* .

The quantiles that define marginal proportions of X and Y from a standard normal distribution function Φ are:

$$z_1 = \Phi^{-1}\left(\frac{a+b}{n}\right) \\ z_2 = \Phi^{-1}\left(\frac{a+c}{n}\right)$$

Thresholds z_1 and z_2 define the cut-off point dividing bivariate standard normal distribution into four quadrants, the probabilities of which correspond to the observed proportions (Figure 1).

The tetrachoric correlation coefficient ρ_t , defined by Pearson (1900), satisfies:

$$\frac{a}{n} = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} \varphi(x^*, y^*, \rho_t) dx^* dy^*$$

where, $\varphi(\cdot)$ is the normal bivariate standard density function.

This correlation coefficient does not consider the level of skewness in the latent variables, which can be incorporated through a skew normal distribution.

Distribución Normal Asimétrica

Azzallini (1985) dio a conocer de manera formal la definición de la familia de distribuciones Normales Asimétricas (SN, por sus siglas en inglés), como una generalización de la familia de distribuciones Normales. La familia SN incluye, además de los parámetros de localidad y escala, un parámetro de forma (α) que permite regular la asimetría de la distribución. La distribución SN ha sido de gran interés porque comparte muchas de las propiedades de la distribución Normal (Azzallini y Capitanio, 2014).

Distribución Normal Asimétrica univariada

Una variable aleatoria continua X tiene distribución Normal Asimétrica con parámetro de forma α , denotada como $X \sim SN(\alpha)$, si su función de densidad es:

$$f_X = 2 \varphi(x) \Phi(\alpha x) \quad \forall x \in \mathbb{R}; \alpha \in \mathbb{R}$$

donde $\varphi(\cdot)$ y $\Phi(\cdot)$ corresponden a las funciones de densidad y distribución de la Normal estándar, respectivamente.

Distribución Normal Asimétrica bivariada

Sean $\underline{X} = (X_1, X_2)'$ una variable aleatoria continua en $\mathbb{R}^2$, $\varphi_2(\cdot)$ la función de densidad de la distribución Normal bivariada con media $\underline{0}$ y matriz de correlación $\underline{\Omega} = \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}$, $\Phi(\cdot)$ la función de distribución de la Normal estándar y $\underline{\alpha} = (\alpha_1, \alpha_2)' \in \mathbb{R}^2$. La función de densidad de $\underline{X}$ es:

$$f_{\underline{X}} = 2\varphi_2(\underline{x}, \underline{\Omega})\Phi(\underline{\alpha}'\underline{x}) \quad (3)$$

donde $\underline{X}$ tiene distribución Normal Asimétrica bivariada con parámetro de forma $\underline{\alpha}$ y se denota como $\underline{X} \sim SN_2(\underline{\Omega}, \underline{\alpha})$.

Distribuciones marginales de la Normal Asimétrica bivariada

Si $\underline{X} = (X_1, X_2)' \sim SN_2(\underline{\Omega}, \underline{\alpha})$, las distribuciones marginales de X_1 y X_2 son $SN(\bar{\alpha}_1)$ y $SN(\bar{\alpha}_2)$ (Azzallini y Capitanio, 1999), respectivamente, donde los parámetros de forma son:

$$\bar{\alpha}_1 = \frac{\alpha_1 + \omega\alpha_2}{\sqrt{1 + (1 - \omega^2)\alpha_2^2}} \quad (4)$$

$$\bar{\alpha}_2 = \frac{\alpha_2 + \omega\alpha_1}{\sqrt{1 + (1 - \omega^2)\alpha_1^2}} \quad (5)$$

Skew Normal Distribution

Azzallini (1985) formally published the definition of Skew Normal distribution family (SN), as a generalization of the Normal distribution family. The SN family includes location and scale parameters, and also a parameter of shape (α), which helps regulate the skewness of the distribution. SN distribution has been of great interest, because it shares many properties of the normal distribution (Azzallini and Capitanio, 2014).

Univariate Skew Normal Distribution

A continuous random variable X has a skew normal distribution with a parameter of shape α , denoted as $X \sim SN(\alpha)$, if its density function is:

$$f_X = 2 \varphi(x) \Phi(\alpha x) \quad \forall x \in \mathbb{R}; \alpha \in \mathbb{R}$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ correspond to the density and distribution functions of the standard Normal, respectively.

Bivariate Skew Normal Distribution

Let $\underline{X} = (X_1, X_2)'$ be a continuous random variable in $\mathbb{R}^2$, $\varphi_2(\cdot)$, the density function of bivariate normal distribution with mean $\underline{0}$ and correlation matrix $\underline{\Omega} = \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}$, $\Phi(\cdot)$ the standard normal distribution function and $\underline{\alpha} = (\alpha_1, \alpha_2)' \in \mathbb{R}^2$. Then, the density function of $\underline{X}$ is:

$$f_{\underline{X}} = 2\varphi_2(\underline{x}, \underline{\Omega})\Phi(\underline{\alpha}'\underline{x}) \quad (3)$$

where $\underline{X}$ has a bivariate skew normal distribution with a parameter of shape $\underline{\alpha}$ and is denoted as $\underline{X} \sim SN_2(\underline{\Omega}, \underline{\alpha})$.

Marginal distributions of Bivariate Skew Normal Distribution

Let be $\underline{X} = (X_1, X_2)' \sim SN_2(\underline{\Omega}, \underline{\alpha})$; then the marginal distributions of X_1 and X_2 are $SN(\bar{\alpha}_1)$ and $SN(\bar{\alpha}_2)$ (Azzallini and Capitanio, 1999), respectively, where the parameters of shape are:

$$\bar{\alpha}_1 = \frac{\alpha_1 + \omega\alpha_2}{\sqrt{1 + (1 - \omega^2)\alpha_2^2}} \quad (4)$$

$$\bar{\alpha}_2 = \frac{\alpha_2 + \omega\alpha_1}{\sqrt{1 + (1 - \omega^2)\alpha_1^2}} \quad (5)$$

Coefficiente de correlación en la distribución Normal Asimétrica bivariada

Sea $\underline{X} = (X_1, X_2)'$ con distribución $SN_2(\underline{\Omega}, \underline{\alpha})$. El coeficiente de correlación entre X_1 y X_2 (Azzallini y Dalla-Valle, 1996) es:

$$\rho_{X_1, X_2} = \frac{\omega - \frac{2}{\pi} \delta_1 \delta_2}{\sqrt{\left(1 - \frac{2}{\pi} \delta_1^2\right) \left(1 - \frac{2}{\pi} \delta_2^2\right)}}$$

$$\text{donde } \underline{\delta} = (\delta_1, \delta_2)' = \frac{\underline{\Omega} \underline{\alpha}}{\sqrt{1 + \underline{\alpha}' \underline{\Omega} \underline{\alpha}}}$$

Propuesta para calcular la correlación tetracórica con base en la distribución Normal Asimétrica

Si se considera que el vector aleatorio $\underline{X} = (X_1, X_2)'$ tiene distribución Normal Asimétrica bivariada $SN_2(\underline{\Omega}_{SN}, \underline{\alpha})$, donde $\underline{\Omega}_{SN} = \begin{pmatrix} 1 & \omega_{SN} \\ \omega_{SN} & 1 \end{pmatrix}$; entonces, el coeficiente ω_{SN} satisface que:

$$\frac{a}{n} = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} f_{\underline{X}}(x_1, x_2, \omega_{SN}) dx_1 dx_2 \quad (6)$$

donde $f_{\underline{X}}(\cdot)$ es la función de densidad de la distribución $SN_2(\underline{\Omega}_{SN}, \underline{\alpha})$ definida en (3).

De acuerdo con (4) y (5), los cuantiles z_1 y z_2 se obtienen como:

$$z_1 = \Phi_{\alpha_1}^{-1} \left(\frac{a+b}{n} \right) \quad (7)$$

$$z_2 = \Phi_{\alpha_2}^{-1} \left(\frac{a+c}{n} \right) \quad (8)$$

La correlación tetracórica con base en la Normal Asimétrica, denotada como $\rho_{\tau_{SN}}$, está dada por:

$$\rho_{\tau_{SN}} = \frac{\omega_{SN} - \frac{2}{\pi} \delta_1 \delta_2}{\sqrt{\left(1 - \frac{2}{\pi} \delta_1^2\right) \left(1 - \frac{2}{\pi} \delta_2^2\right)}} \quad (9)$$

Correlation coefficient in Bivariate Skew Normal Distribution

Let $\underline{X} = (X_1, X_2)'$ be with distribution $SN_2(\underline{\Omega}, \underline{\alpha})$; then, the correlation coefficient between X_1 and X_2 (Azzallini and Dalla-Valle, 1996) is:

$$\rho_{X_1, X_2} = \frac{\omega - \frac{2}{\pi} \delta_1 \delta_2}{\sqrt{\left(1 - \frac{2}{\pi} \delta_1^2\right) \left(1 - \frac{2}{\pi} \delta_2^2\right)}}$$

$$\text{donde } \underline{\delta} = (\delta_1, \delta_2)' = \frac{\underline{\Omega} \underline{\alpha}}{\sqrt{1 + \underline{\alpha}' \underline{\Omega} \underline{\alpha}}}$$

Proposal to calculate tetrachoric correlation based on Skew Normal Distribution

If the random vector $\underline{X} = (X_1, X_2)'$ is considered to have a bivariate skew normal distribution $SN_2(\underline{\Omega}_{SN}, \underline{\alpha})$ where $\underline{\Omega}_{SN} = \begin{pmatrix} 1 & \omega_{SN} \\ \omega_{SN} & 1 \end{pmatrix}$, then the coefficient ω_{SN} satisfies:

$$\frac{a}{n} = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} f_{\underline{X}}(x_1, x_2, \omega_{SN}) dx_1 dx_2 \quad (6)$$

where $f_{\underline{X}}(\cdot)$ is the density function of distribution $SN_2(\underline{\Omega}_{SN}, \underline{\alpha})$ defined in (3).

According to (4) and (5), quantiles z_1 and z_2 are obtained as:

$$z_1 = \Phi_{\alpha_1}^{-1} \left(\frac{a+b}{n} \right) \quad (7)$$

$$z_2 = \Phi_{\alpha_2}^{-1} \left(\frac{a+c}{n} \right) \quad (8)$$

The tetrachoric correlation based on skewed normal, denoted as $\rho_{\tau_{SN}}$, is given by:

$$\rho_{\tau_{SN}} = \frac{\omega_{SN} - \frac{2}{\pi} \delta_1 \delta_2}{\sqrt{\left(1 - \frac{2}{\pi} \delta_1^2\right) \left(1 - \frac{2}{\pi} \delta_2^2\right)}} \quad (9)$$

Procedimiento para la estimación Monte Carlo del coeficiente de correlación tetracórica con base en la distribución Normal Asimétrica

Debido a que no es posible evaluar analíticamente la doble integral en (6) para despejar a ω_{SN} y luego obtener a $\rho_{t_{SN}}$ en (9), el problema de estimación de dichos parámetros se aborda a través de simulación Monte Carlo.

Ya que la proporción $\frac{a}{n}$ representa el área de la distribución Normal Asimétrica bivariada determinada por los cuantiles z_1 y z_2 , la ecuación (6) puede expresarse como la probabilidad conjunta de dos eventos; es decir:

$$P(X_1 \leq z_1, X_2 \leq z_2) = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} f_{\omega_{SN}}(x_1, x_2, \omega_{SN}) dx_1 dx_2 = \frac{a}{n}$$

El procedimiento de estimación consiste en los siguientes pasos:

1. Definir una partición de tamaño k del intervalo $[-1, 1]$, al cual pertenece $\omega_{SN, i}$.
2. Para cada valor de $\omega_{SN, i}$ elemento de la partición, se realiza el siguiente procedimiento:
 - 2.1. Simular una muestra pseudo aleatoria de tamaño n de la distribución $SN_2(\Omega_{SN, i}, \underline{\alpha}_0)$, donde $\underline{\alpha}_0 = (\alpha_{0,1}, \alpha_{0,2})'$.
 - 2.2. Contar la proporción de parejas (x_1, x_2) de la muestra simulada que cumplan simultáneamente que $x_1 \leq z_1$ y $x_2 \leq z_2$. Los cuantiles z_1 y z_2 se determinan a partir de las expresiones (7) y (8) que son función de las proporciones marginales observadas.
 - 2.3. Para no depender de los resultados de una muestra, los pasos 2.1 y 2.2 se repiten l veces. En cada una de las l muestras asimétricas bivariadas se obtiene la proporción $\frac{m_{ij}}{n}$, $j=1, 2, \dots, l$, de parejas en la muestra que cumplen con la condición establecida en el paso 2.2.
 - 2.4. La proporción asociada a $\omega_{SN, i}$ es $\left(\frac{m_i}{n}\right) = \frac{\sum_{j=1}^l \frac{m_{ij}}{n}}{l}$, $i=1, 2, \dots, k$.
3. La proporción observada $\frac{a}{n}$ se compara con cada una de las proporciones $\frac{m_{ij}}{n}$ obtenidas en 2.4, de modo que el valor $\omega_{SN, i}$ asociado con la diferencia mínima corresponde a la estimación de ω_{SN} denotado como $\tilde{\omega}_{SN}$, el cual permite obtener la estimación de $\rho_{t_{SN}}$, es decir, $\tilde{\rho}_{t_{SN}}$, de acuerdo con (9).

RESULTADOS Y DISCUSIÓN

El análisis de la consistencia de $\tilde{\rho}_{t_{SN}}$ se realizó por medio de un estudio de simulación Monte Carlo con 1000 repeticiones con el paquete *sn* (Azzallini, 2018) del programa R Core Team (2019). A partir de muestras

Monte Carlo estimation procedure for tetrachoric correlation coefficient based on Skew Normal Distribution

Due to the impossibility of analytically analyzing the double integral in (6) to find ω_{SN} and then obtain $\rho_{t_{SN}}$ in (9), the problem of estimating these parameters is tackled using a Monte Carlo simulation.

Since the proportion $\frac{a}{n}$ represents the area of the bivariate skew normal distribution determined by quantiles z_1 and z_2 , equation (6) can be expressed as the joint probability of two events, *i.e.* as follows:

$$P(X_1 \leq z_1, X_2 \leq z_2) = \int_{-\infty}^{z_2} \int_{-\infty}^{z_1} f_{\omega_{SN}}(x_1, x_2, \omega_{SN}) dx_1 dx_2 = \frac{a}{n}$$

The estimation procedure consists of the following steps:

1. Defining a k -sized partition of interval $[-1, 1]$, to which $\omega_{SN, i}$ belongs.
2. For each value of $\omega_{SN, i}$ element of the partition, the following procedure is applied:
 - 2.1. Simulating a pseudo-random n -sized sample of the $SN_2(\Omega_{SN, i}, \underline{\alpha}_0)$ distribution, where $\underline{\alpha}_0 = (\alpha_{0,1}, \alpha_{0,2})'$.
 - 2.2. Counting the proportion of vectors (x_1, x_2) in the simulated sample that simultaneously satisfy that $x_1 \leq z_1$ and $x_2 \leq z_2$. Quantiles z_1 and z_2 are determined from the expressions (7) and (8), which are functions of the observed marginal proportions.
 - 2.3. In order to avoid dependency on the results of a sample, steps 2.1 and 2.2 are repeated l times. Each of the l bivariate skew samples gives the proportion $\frac{m_{ij}}{n}$, $j=1, 2, \dots, l$, of vector pairs in the sample that comply with the condition established in step 2.2.
 - 2.4. The proportion associated to $\omega_{SN, i}$ is $\left(\frac{m_i}{n}\right) = \frac{\sum_{j=1}^l \frac{m_{ij}}{n}}{l}$, $i=1, 2, \dots, k$.
3. The $\frac{a}{n}$ observed proportion is compared with each proportion $\frac{m_{ij}}{n}$ obtained in 2.4, so that value $\omega_{SN, i}$ related to the minimum difference, corresponds to the estimation of ω_{SN} denoted as $\tilde{\omega}_{SN}$, which helps obtain the estimation of $\rho_{t_{SN}}$ that is, $\tilde{\rho}_{t_{SN}}$, according to (9).

RESULTS AND DISCUSSION

A consistency analysis of $\tilde{\rho}_{t_{SN}}$ was performed using a Monte Carlo simulation study with 1000 iterations using the *sn* package (Azzallini, 2018) of the R Core Team (2019) program. From pseudo-random n -sized samples with $SN_2(\Omega_{SN}, \underline{\alpha})$ distribution,

pseudo aleatorias de tamaño n de la distribución $SN_2(\Omega_{SN}, \underline{\alpha})$, donde $\omega_{SN} = 0.9$, $\alpha_1 = \alpha_2 = 0.8$ y $\rho_{tSN} = 0.825$, se construyeron variables dicotómicas con base en los cuantiles asociados con proporciones marginales fijas $\left(\frac{a+b}{n}\right)_0 = 0.7064$ y $\left(\frac{a+c}{n}\right)_0 = 0.8550$. La tabla de proporciones de las variables dicotómicas construidas proporcionó el insumo necesario para la aplicación del método de estimación de ρ_{tSN} , donde $k=150$, $l=30$, y $n \in \{30, 100, 500, 1000, 5000, 10\ 000\}$.

El estudio de simulación Monte Carlo mostró que $\tilde{\rho}_{tSN}$ es un estimador robusto al nivel de asimetría y consistente en Error Cuadrático Medio (ECM). Los resultados señalaron una reducción importante de la varianza, y en consecuencia del ECM, a partir de las muestras de tamaño $n=100$. Además, la estimación fue más cercana al valor de referencia $\rho_{tSN} = 0.825$ conforme aumentó el tamaño de muestra, lo cual fue evidencia de la consistencia del estimador propuesto (Cuadro 1).

La distribución de las simulaciones mostró que las estimaciones se concentran alrededor del valor de referencia ρ_{tSN} (representado por la línea vertical) a medida que el tamaño de muestra aumenta (Figura 2).

El método propuesto permitió la identificación de la relación inversa entre nivel de asimetría y la magnitud de la correlación $\tilde{\rho}_{tSN}$ (Cuadro 2).

Los resultados mostraron que, ante la presencia de asimetría, la correlación tiende a ser menor a la obtenida en el caso $\alpha_0 = 0$. Sin embargo, el grado de asociación no se vio afectado, pues en todos los escenarios definidos se observó una correlación

where $\omega_{SN} = 0.9$, $\alpha_1 = \alpha_2 = 0.8$ and $\rho_{tSN} = 0.825$, dichotomous variables were constructed, based on the quantiles associated with fixed marginal proportions $\left(\frac{a+b}{n}\right)_0 = 0.7064$ and $\left(\frac{a+c}{n}\right)_0 = 0.8550$. The proportions table of the constructed dichotomous variables provided the required input for the application of the estimation method of ρ_{tSN} , where $k=150$, $l=30$, and $n \in \{30, 100, 500, 1000, 5000, 10\ 000\}$.

Monte Carlo simulation study showed that $\tilde{\rho}_{tSN}$ is a robust estimator at the skewness level and it is a mean-squared-error consistent estimator (ECM). The results indicated a significant reduction in variance, and consequently, on its ECM, from samples of size $n=100$. In addition, the estimation came closer to the reference value $\rho_{tSN} = 0.825$ as the sample size increased, which was evidence of the consistency of the proposed estimator (Table 1).

The distribution of simulations showed that estimations are concentrated around the reference value ρ_{tSN} (represented by the vertical line) as the sample size increases (Figure 2).

The proposed method helped identifying the inverse relationship between level of skewness and magnitude of correlation $\tilde{\rho}_{tSN}$ (Table 2).

Results showed that, given the presence of skewness, the correlation tends to be lower than the one obtained when $\alpha_0 = 0$. However, the degree of association was not affected, since in all the defined scenarios, a high positive correlation was observed, along with a marginal decrease in correlation from $\alpha_0 = 2.7$.

Cuadro 1. Estimación del coeficiente ρ_{tSN} , según tamaño de muestra. Simulación Monte Carlo con 1000 repeticiones.

Table 1. Estimation of coefficient ρ_{tSN} , according to sample size. Monte Carlo simulation with 1000 iterations.

Tamaño de muestra (n)	$\tilde{\rho}_{tSN}$				
	Media	Mediana	Sesgo	ECM	Varianza
30	0.6791	0.7690	-0.14586	0.13790	0.11675
100	0.8178	0.8349	-0.00721	0.01481	0.01478
500	0.8271	0.8349	0.00211	0.00302	0.00302
1 000	0.8240	0.8349	-0.00103	0.00149	0.00149
5 000	0.8250	0.8349	0.00003	0.00030	0.00030
10 000	0.8245	0.8349	-0.00050	0.00017	0.00017

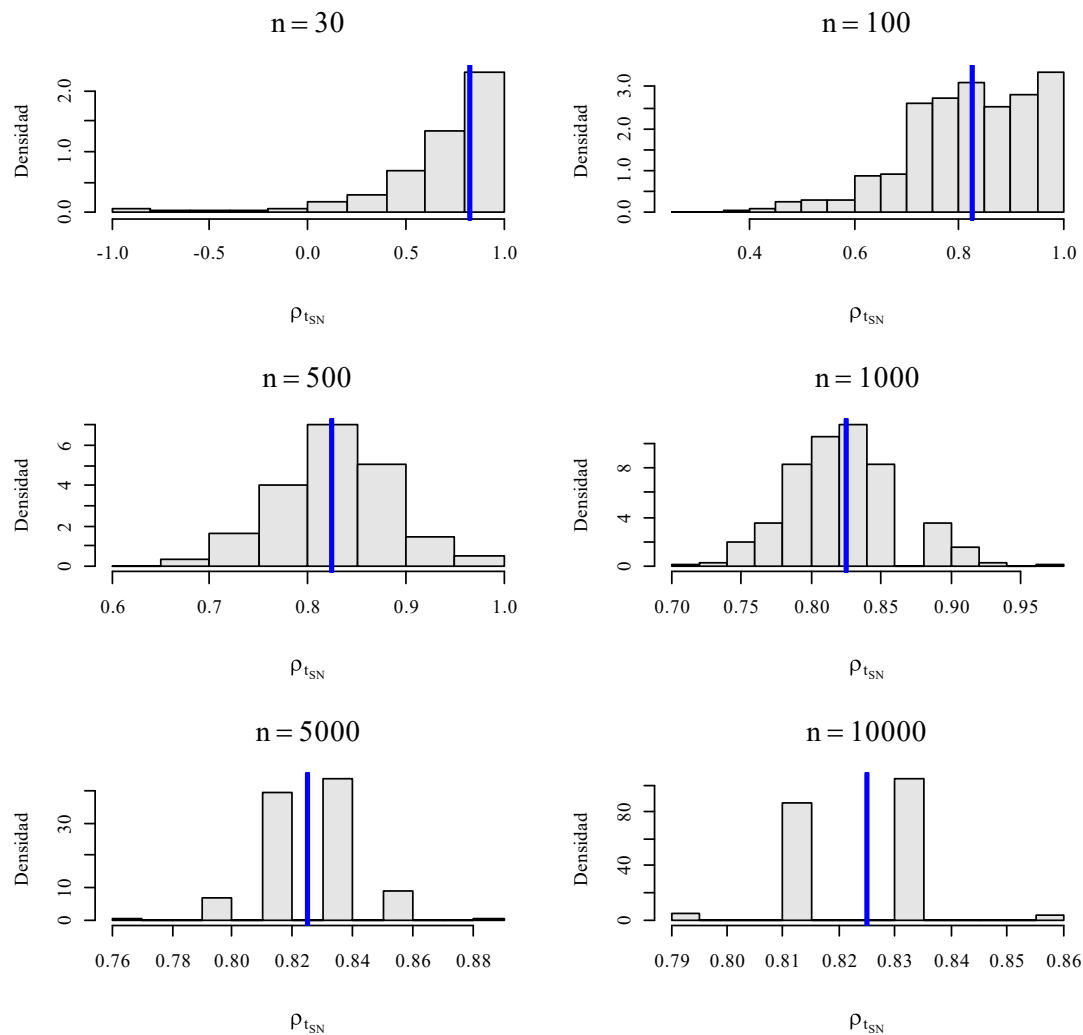


Figura 2. Distribución de $\hat{\rho}_{t_{SN}}$, según tamaño de muestra.

Figure 2. Distribution of $\hat{\rho}_{t_{SN}}$, according to sample size.

Cuadro 2. Estimación del coeficiente $\rho_{t_{SN}}$, según parámetro de forma. Primeros dos reactivos HSM, EMSA 2014.

Table 2. Estimation of coefficient $\rho_{t_{SN}}$, according to parameter of shape. First two items HSM, EMSA 2014.

α_0	$\frac{a}{n}$	$\frac{m}{n}$	$\left \frac{m}{n} - \frac{a}{n} \right $	$\tilde{\omega}_{SN}$	$\tilde{\rho}_{t_{SN}}$
0.0	0.6947	0.6946	0.0001	0.8657	0.8657
0.8	0.6947	0.6942	0.0005	0.9194	0.8575
1.5	0.6947	0.6936	0.0011	0.9194	0.8218
2.7	0.6947	0.6938	0.0008	0.9194	0.8034
3.4	0.6947	0.6939	0.0007	0.9194	0.7997
4.2	0.6947	0.6939	0.0007	0.9194	0.7974
5.0	0.6947	0.694	0.0007	0.9194	0.7961
6.0	0.6947	0.694	0.0007	0.9194	0.7951
7.0	0.6947	0.6941	0.0005	0.9194	0.7945

positiva alta y una disminución marginal en la correlación a partir de $\alpha_0 = 2.7$.

Validación de la EMSA por medio del Análisis de Factores Confirmatorio

El contraste de las estructuras por cada tipo de hogar se realizó con la información de la EMSA 2014. La muestra consta de un total de 58 125 hogares a nivel nacional, de los cuales 22 486 son hogares sin menores (HSM) y 35 639 son hogares con menores (HCM). Las Figuras 3 y 4 muestran las estructuras hipotéticas utilizadas en el AFC correspondientes a cada tipo de hogar.

Para estimar los modelos se contemplaron las matrices de correlación obtenidas bajo tres niveles de asimetría: $\alpha_0 = 0$, $\alpha_0 = 0.8$ y $\alpha_0 = 2.7$. Dichas matrices se construyeron a partir del método propuesto para la estimación de ρ_{iSN} . Los resultados señalaron que existe una correlación positiva alta entre los reactivos de la escala, lo cual refleja el grado de coherencia entre ellos como variables manifiestas del concepto teórico de la seguridad alimentaria en los hogares.

Las medidas de bondad de ajuste para los HSM (Cuadro 3) mostraron un buen ajuste, con excepción del RMSEA. En todos los casos, los residuales estandarizados indicaron discrepancia mínima entre la matriz de correlaciones observada y la matriz de correlaciones reproducida por el modelo. Por lo tanto, los reactivos de la EMSA cuantificaron en forma adecuada el concepto teórico de la seguridad alimentaria

Validation of EMSA using Confirmatory Factor Analysis

The contrast of the structures for each type of household was carried out with the information from EMSA 2014. The sample consists of a total of 58 125 households nationwide, 22 486 of which are households without children (HSM) and 35 639 are households with children (HCM). Figures 3 and 4 show hypothetical structures used in CFA corresponding to each type of household.

To estimate the models, the correlation matrices obtained under three levels of skewness were considered: $\alpha_0 = 0$, $\alpha_0 = 0.8$ and $\alpha_0 = 2.7$. These matrices were created using the proposed method for the estimation of ρ_{iSN} . Results indicated that there is a high positive correlation between items of the scale, which reflects the degree of coherence between them as manifest variables of the theoretical concept of food security in households.

The goodness-of-fit indices for HSM (Table 3) showed a good fit, with the exception of RMSEA. In all cases, the standardized residuals indicated a minimum discrepancy between the observed correlation matrix and the correlation matrix reproduced by the model. Therefore, EMSA items adequately quantified the theoretical concept of food security in HSM, and they support the relationship between theoretical structure and empirical evidence.

The CFA model for HCM with the original structure proposed (Figure 4) did not show an

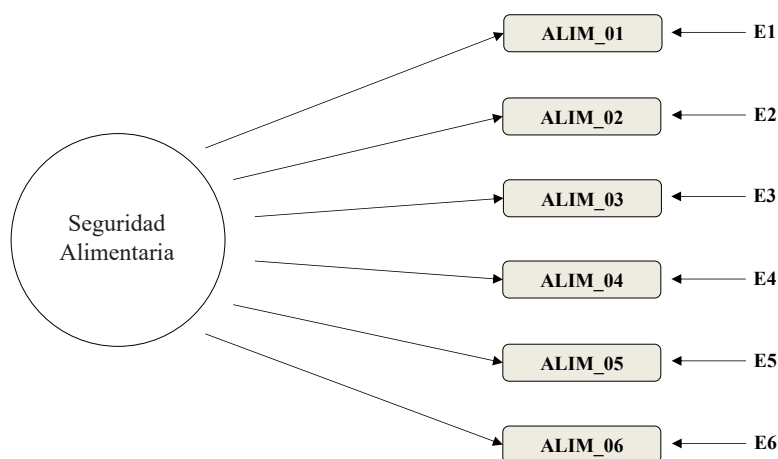


Figura 3. Modelo factorial para la seguridad alimentaria en HSM.
Figure 3. Factorial model for food security in HSM.

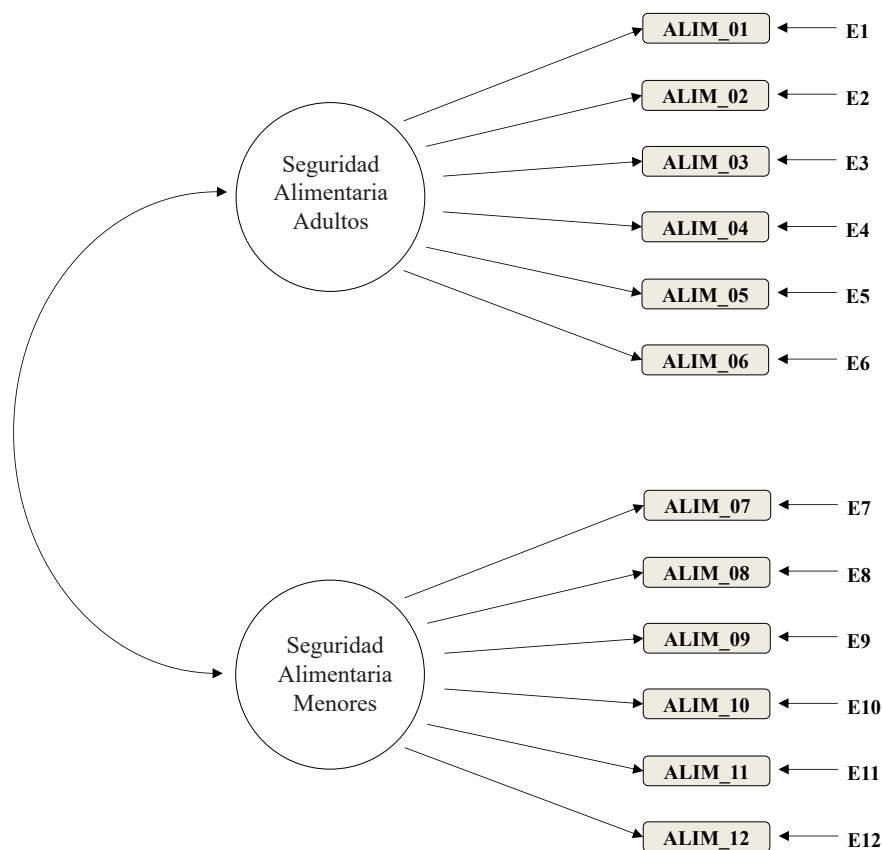


Figura 4. Modelo factorial para la seguridad alimentaria en HCM
Figure 4. Factorial model for food security in HCM

en los HSM y respaldan la correspondencia entre la estructura teórica y la evidencia empírica.

El modelo de AFC para HCM con la estructura original planteada (Figura 4) no mostró un ajuste adecuado para ninguno de los niveles de asimetría considerados. Sin embargo, al incorporar en el modelo tres relaciones adicionales entre los errores de medición (E) de las variables *ALIM_01*, *ALIM_06*,

adequate fit for any of the levels of skewness considered. However, by incorporating into the model three additional relations between the measurement errors (E) of the variables *ALIM_01*, *ALIM_06*, *ALIM_07*, *ALIM_10*, *ALIM_11* and *ALIM_12* (Figure 5), the global fit indices improved substantially. These additional relationships were obtained from the modification indices, which

Cuadro 3. Medidas de bondad de ajuste para los modelos de HSM, 2014.
Table 3. Goodness-of-fit indices for HSM models, 2014.

Medida	Nivel de asimetría			Criterio de aceptación
	$\alpha_0 = 0$	$\alpha_0 = 0.8$	$\alpha_0 = 2.7$	
CFI	0.964	0.958	0.953	Mayor a 0.90
RMSEA	0.190	0.196	0.190	Menor a 0.08
SRMR	0.015	0.019	0.024	Menor a 0.08
STD_RES (valor máximo)	0.038	0.061	0.072	Menor a 0.10

ALIM_07, *ALIM_10*, *ALIM_11* y *ALIM_12* (Figura 5), las medidas de ajuste global mejoraron sustancialmente. Dichas relaciones adicionales se obtuvieron a partir de los índices de modificación, los cuales estiman la variación que puede producirse en el modelo al introducir una nueva relación entre variables que no se haya considerado y contribuye con la mejora del ajuste del modelo.

La relación entre los errores de medición *E1* y *E7* indicó que la presencia de una situación adversa que provoca cambios en la variedad de los alimentos tiene repercusiones a nivel hogar, de modo que las afectaciones son tanto para los adultos como para los menores de 18 años. Una situación similar ocurre cuando la severidad de la inseguridad alimentaria se incrementa al nivel de eliminar algún tiempo de comida para todos los integrantes del hogar, lo cual se señala por la relación entre *E6* y *E12*. Estas relaciones son reflejo del marco conceptual en el que se basa la EMSA, de modo que cada una de las dimensiones

estimate the variation that can occur in the model by introducing a new relationship between variables that had not been considered, and contributes to the improvement of the fit of the model.

The relationship between measurement errors *E1* and *E7* indicated that the presence of an adverse situation that causes changes in the variety of food has repercussion at the household level, so the impacts are both for adults and underage individuals. A similar situation occurs when the severity of food insecurity increases until it deprives everyone in the household from some meal, which is pointed out by the relationship between *E6* and *E12*. These relationships reflect the conceptual framework upon which EMSA is based, so each one of the underlying dimensions in the measurement of household food security affects children after adults (FAO, 2012). Finally, the relationship between *E10* and *E11* suggests a confusion in the comprehension of items related to hunger experiences.

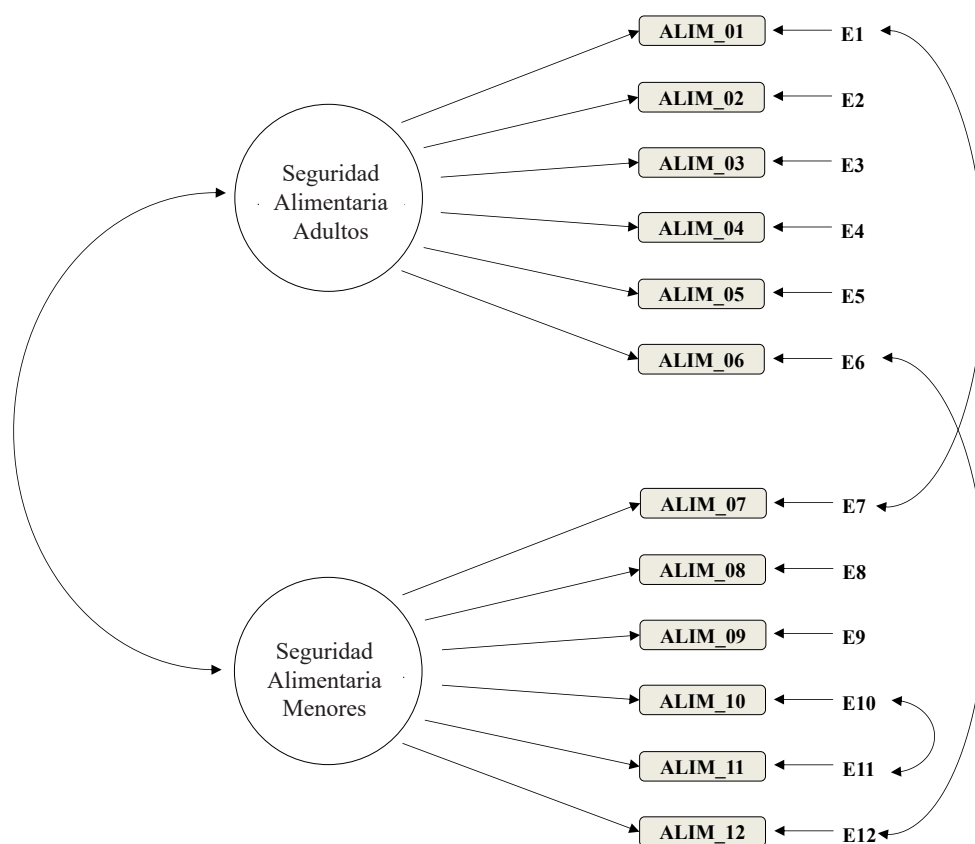


Figura 5. Modelo factorial para la seguridad alimentaria en HCM, de acuerdo con los índices de modificación.
Figure 5. Factorial model for food security in HCM, according to modification indices.

subyacentes a la medición de la seguridad alimentaria en los hogares afecta a los menores después de que ha afectado a los adultos (FAO, 2012). Por último, la relación entre $E10$ y $E11$ sugiere confusión en el entendimiento de los reactivos asociados con experiencias de hambre.

La incorporación de las relaciones entre los errores de medición (Figura 5) contribuyó con un mejor ajuste (Cuadro 4).

CONCLUSIONES

El método de estimación de $\rho_{t_{SN}}$ permitió analizar el comportamiento de la correlación entre variables dicotómicas por medio de variables latentes en presencia de diferentes niveles de asimetría. El estimador Monte Carlo $\tilde{\rho}_{t_{SN}}$ presentó un comportamiento consistente y de poca varianza.

De acuerdo con los resultados de los modelos del Análisis de Factores Confirmatorio, los reactivos de la EMSA explicaron en forma adecuada el concepto teórico de la seguridad alimentaria para los hogares sin menores. En los hogares con menores, los resultados indicaron colinealidad entre los errores de medición de las variables que corresponden a la variedad de los alimentos, la eliminación de algún tiempo de comida y las experiencias de hambre.

En tal caso se sugiere revisar los reactivos correspondientes a $ALIM_{10}$ y $ALIM_{11}$ pues existe confusión en su redacción. La nueva estructura teórica se confirmó y fue robusta en presencia de niveles de asimetría diferentes. Estos hallazgos solo fueron visibles al considerar las matrices de correlación obtenidas por medio del estimador $\tilde{\rho}_{t_{SN}}$ propuesto.

The incorporation of relationships between measurement errors (Figure 5) contributed to a better fit (Table 4).

CONCLUSIONS

The estimation method of $\rho_{t_{SN}}$ helped analyzing the behavior of correlation between dichotomous variables by latent variables in the presence of different levels of skewness. The Monte Carlo estimator $\tilde{\rho}_{t_{SN}}$ showed a consistent and low variance behavior.

According to the results of Confirmatory Factor Analysis models, EMSA items adequately explained the theoretical concept of food security for households without children. In households with children, the results indicated collinearity between measurement errors of the variables regarding variety of food, the elimination of some meal and hunger experiences.

In such a case, a revision is suggested of items $ALIM_{10}$ and $ALIM_{11}$, because there was confusion in their wording. The new theoretical structure was confirmed and it was robust in presence of different levels of skewness. These findings were only visible when considering correlation matrices obtained using the proposed estimator $\tilde{\rho}_{t_{SN}}$.

—End of the English version—

---*---

Cuadro 4. Medidas de bondad de ajuste para los modelos de HCM, 2014.
Table 4. Goodness-of-fit indices for HCM models, 2014.

Medida	Nivel de asimetría			Criterio de aceptación
	$\alpha_0 = 0$	$\alpha_0 = 0.8$	$\alpha_0 = 2.7$	
CFI	0.811	0.914	0.895	Mayor a 0.90
RMSEA	0.305	0.183	0.188	Menor a 0.08
SRMR	0.032	0.031	0.043	Menor a 0.08
STD_RES (valor máximo)	0.091	0.083	0.155	Menor a 0.10

LITERATURA CITADA

- Acock, A. 2013. *Discovering Structural Equation Modeling using Stata*. Stata Press. Texas. 306 p.
- Azzallini, A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*. 12: 171 - 178.
- Azzallini, A., and A. Dalla-Valle. 1996. The multivariate skew-normal distribution. *Biometrika*. 83: 715 - 726.
- Azzallini, A., and A. Capitanio. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 61: 579 - 602.
- Azzallini, A., and A. Capitanio. 2014. *The skew-normal and related families*. Institute of Mathematical Statistics Monographs. Cambridge University Press. 262 p.
- Azzallini, A. 2018. The R package sn: the skew-normal and related distributions such as the skew-t (version 1.5-3). *Università di Padova*. Italia. 79 p.
- Byrne, B. 2016. *Structural equation modeling with AMOS: basic concepts, applications and programming*. 3a. ed. Taylor and Francis Group. New York. 438 p.
- CONEVAL (Consejo Nacional de Evaluación de la Política de Desarrollo Social). 2010. *Metodología para la medición multidimensional de la pobreza en México*. México. 128 p.
- FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). 2012. *Escala Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA): manual de uso y aplicaciones*. Roma. 85 p.
- Kline, R. 2011. *Principles and practice of structural equation modeling*. 3a. ed. The Guilford Press. New York. 427 p.
- Pearson, K. 1900. *Mathematical contributions to the theory of evolution*. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London*. 195: 1 - 47.
- Pérez-Gil, J., S. Chacón, y R. Moreno. 2000. Validez de constructo: el uso de análisis factorial exploratorio - confirmatorio para obtener evidencias de validez. *Psicothema*. 12: 442 - 446.
- Suárez, A. 2015. La estadística en la validación de escalas, una visión práctica para su construcción o su adaptación. *Revista I3+*. 2: 46 - 61.
- Villagómez, P., P. Hernández, B. Carrasco, K. Barrios, R. Pérez, y H. Melgar. 2014. Validez estadística de la Escala Mexicana de Seguridad Alimentaria y la Escala Latinoamericana y Caribeña de Seguridad Alimentaria. *Salud Pública de México*. 56: S5 - S11.

