

REVISTA SIGMA

Departamento de Matemáticas y Estadística

Volumen XX N.º 2 (2024), páginas 25–34

Universidad de Nariño

Revisión Literaria en Coeficientes Policóricos y Tetracóricos

Juan Camilo López Caicedo ¹

Carlos Andres Mesías Ocaña ²

Abstract: The growing significance of polychoric and tetrachoric correlation coefficients in statistical research is examined in this work, with a focus on psychometrics and related fields. These coefficients provide a precise tool for studying correlations among variables in social sciences, since many of the variables are categorical or ordinal in nature. Though polychoric and tetrachoric coefficients are more useful in managing ordinal data, older techniques like Pearson correlations are still applicable.

A review of the literature indicates that the frequency of publications has been on the rise and has peaked in 2009, 2014, and 2016. This indicates times of increased research activity as well as a persistent interest in the use and advancement of these coefficients.

In summary, this study offers a thorough grasp of the applicability and promise of polychoric and tetrachoric correlation coefficients in statistical research, emphasizing their function in the comprehension and examination of student's mental health through an example.

Keywords. Polychoric correlations, tetrachoric correlations, psychometrics, ordinal data analysis, statistical methods.

Resumen: En este trabajo se examina la creciente importancia de los coeficientes de correlación policóricos y tetracóricos en la investigación estadística, con un enfoque en la psicometría y campos relacionados. Estos coeficientes proporcionan una herramienta precisa para estudiar correlaciones entre variables en ciencias sociales, dado que muchas de las variables son de naturaleza categórica u ordinal. Aunque los coeficientes policóricos y tetracóricos son más útiles en el manejo de datos ordinales, técnicas más antiguas como las correlaciones de Pearson siguen siendo aplicables.

Una revisión de la literatura indica que la frecuencia de publicaciones ha estado en aumento y alcanzó su punto máximo en 2009, 2014 y 2016. Esto indica períodos de mayor actividad investigativa así como un interés persistente en el uso y avance de estos coeficientes.

¹Maestría en Estadística Aplicada, Facultad Ciencias Naturales y Exactas, Universidad de Nariño. Correo: jclopez23a@udenar.edu.co

²Maestría en Estadística Aplicada, Facultad Ciencias Exactas y Naturales, Universidad de Nariño. Correo: comesias23a@udenar.edu.co

En resumen, este estudio ofrece una comprensión exhaustiva de la aplicabilidad y promesa de los coeficientes de correlación policóricos y tetracóricos en la investigación estadística, enfatizando su función en la comprensión y examen de la salud mental de los estudiantes a través de un ejemplo.

Palabras Clave. Correlaciones policóricas, correlaciones tetracóricas, psicometría, análisis de datos ordinales, métodos estadísticos.

1. Introducción

Debido a la gran cantidad de variables que se recolectan para el estudio de diversos fenómenos principalmente los de tipo social, en los que la mayoría de características son de origen categórico tanto nominal y ordinal, las teorías estadísticas han propuesto análisis que buscan reducir la dimensionalidad de las matrices resultantes, tales como el análisis de componentes principales, análisis factorial exploratorio, análisis factorial confirmatorio, entre otros.

Los coeficientes de correlación tetracóricos y policóricos son una respuesta a la necesidad de precisión que se busca en los estudios de tipo social, específicamente los relacionados a la psicometría cuya finalidad radica en la predicción de cierto tipo de comportamientos basados en una serie de variables, siendo las más destacables las categóricas de escala dicotómica y politómica (Richaud, 2005).

Es importante aclarar que en psicometría clásica los estudios exploratorios, constituyen métodos que tienen por objeto establecer evidencias de validez de constructo. Para que estos procedimientos tengan lugar, se debe partir de matrices de correlaciones que permitan estimar las covariaciones entre los elementos que componen la herramienta, siendo las más utilizadas por los profesionales las correlaciones de Pearson (Freiberg et al. 2013). Sin embargo, en psicometría, actualmente se proponen reemplazar el modelo de correlación de Pearson.

En este orden de ideas, diferentes autores proponen aplicar a los diferentes tipos de correlación dependiendo de su nivel de medición, si las variables son continuas, deben responder al modelo Pearson. Empero, si todas son dicotómicas se utiliza una matriz de correlación tetracórica (TCC) y si la totalidad es ordinal o algunas son ordinales y otras dicotómicas, corresponde el uso de una matriz de correlación policórica (PCC) (Freiberg et al. 2013).

Una de las herramientas utilizadas para realizar este tipo de análisis, es el lenguaje R, el cual ofrece muchas herramientas para desarrollar el ACP, AFE y el AFC, estos elementos se encuentran fundamentados en el coeficiente de correlación de Pearson, pero que por su naturaleza, puede presentar limitaciones al uso de algunos estadísticos. Por ello, surgen alternativas que buscan solucionar las limitaciones estadísticas y mejorar la calidad de los resultados obtenidos. El usar este lenguaje, favorece el aprendizaje básico de estadística, con la ventaja de ser ejecutado en diferentes sistemas operativos (Ruiz y Puga, 2016).

Considerando lo previo, el presente se basa en una revisión sistemática de literatura que por medio de herramientas de inteligencia artificial, identificó 38 estudios relacionados con la aplicación de los coeficientes en análisis multivariados, principalmente en estudios psicométricos. Así pues, el objetivo de este estudio es identificar las publicaciones que a través del transcurso de los años, se han tratado, con respecto a los coeficientes de correlación tetracóricos y policóricos.

Así pues, en primer lugar, el presente expone una introducción a los coeficientes policóricos y tetracóricos, seguido de una revisión literaria con 38 estudios en las diferentes áreas del

conocimiento, luego, apoyado del software R, se desarrolla un ejercicio práctico a partir de una base de datos propia creada con fines académicos en la que se recolectó información de 50 sujetos acerca de posibles trastornos de ansiedad y depresión. Por último, las conclusiones del estudio son presentadas.

2. Desarrollo del Tema

En la presente sección, se presenta los descubrimientos relacionados con el análisis conjunto de las características y el contenido de los 38 manuscritos (expuestos en red dentro de la ilustración I) que componen el corpus de este estudio. Según se observa en la Tabla I, existe un interés académico moderado en analizar los coeficientes policóricos y tetracóricos. Igualmente, los autores más presentes en la literatura son Shaobo Jin, Fritz Drasgow, Guangjian Zhang, Jaehwa Choi, Scott L. Hershberger. Por otra parte, de acuerdo con la ilustración II, los años más prolíferos para publicación son: 2009, 2014 y 2016.

El análisis de la tendencia en la frecuencia de publicaciones a lo largo de los años (2002 – 2023) revela una dinámica fluctuante en la actividad investigativa. A partir del año 2006, se observa un incremento progresivo en la cantidad de artículos, alcanzando un punto álgido en 2009 con un total de cinco publicaciones. Posteriormente, la frecuencia se reduce de forma relativamente estable, con algunos años destacando por un mayor número de contribuciones (como 2014 y 2016), aunque también se registran períodos de menor actividad. Los últimos años presentan indicios de una ligera recuperación en la frecuencia de publicaciones, sugiriendo un posible renacimiento del interés o una nueva fase de investigación en el área abordada.

La revista *Structural Equation Modeling: A Multidisciplinary Journal* es la fuente más prolífica, su objetivo es publicar trabajos académicos evaluados por pares de todas las disciplinas académicas interesadas en el modelado de ecuaciones estructurales. Además, los 38 artículos de esta revista recibieron un total de 698 citas (18.36 en promedio). Por último, es imperioso resaltar que *Journal descripto* cuenta con 5 citas.

Tabla 1. Compilación de Artículos

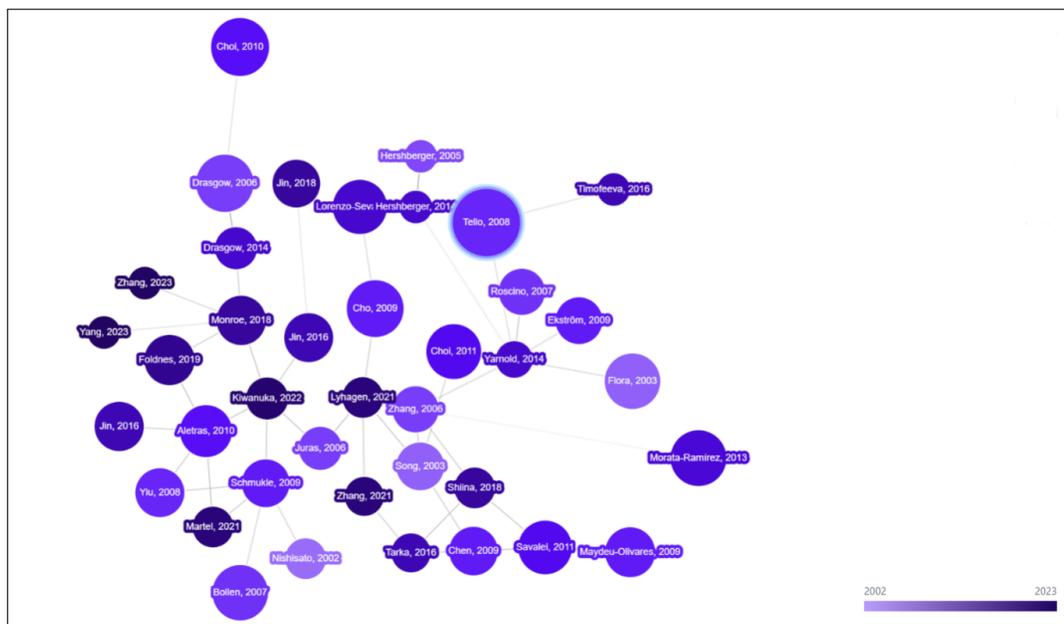
Número	Título	Autor	Año	Citas
1	Quantification of Ordinal Variables: A Critical Inquiry into Polychoric and Canonical Correlation	Shizuhiko Nishisato	2002	3
2	Higher Order Factor Structure Of A Self-Control Test: Evidence From Confirmatory Factor Analysis With Polychoric Correlations	David B. Flora	2003	48
3	Full Maximum Likelihood Estimation of Polychoric and Polyserial Correlations With Missing Data	Xinyuan Song	2003	12
4	Polychoric Correlation	Scott L. Hershberger	2005	0
5	Bootstrap Fit Testing, Confidence Intervals, and Standard Error Estimation in the Factor Analysis of Polychoric Correlation Matrices	Guangjian Zhang	2006	9

6	Polychoric and Polyserial Correlations	Fritz Drasgow	2006	66
7	Application of tetrachoric and polychoric correlation coefficients to forecast verification	Josip Juras	2006	5
8	A Polychoric Instrumental Variable (PIV) Estimator for Structural Equation Models with Categorical Variables	Kenneth A. Bollen	2007	48
9	A Generalization of the Polychoric Correlation Coefficient	Annarita Roscino	2007	9
10	Estimating the polychoric correlation from misclassified data	Choi-Fan Yiu	2008	15
11	A Comparison of Maximum Likelihood and Expected A Posteriori Estimation for Polychoric Correlation Using Monte Carlo Simulation	Jinsong Chen	2009	14
12	Testing Categorized Bivariate Normality With Two-Stage Polychoric Correlation Estimates	Alberto Maydeu-Olivares	2009	18
13	A generalized definition of the polychoric correlation coefficient	Joakim Ekström	2009	10
14	Accuracy of the Parallel Analysis Procedure With Polychoric Correlations	Sun-Joo Cho	2009	61
15	Exploring bipolarity of affect ratings by using polychoric correlations	Stefan C. Schmukle	2009	12
16	Development and preliminary validation of a questionnaire to measure satisfaction with home care in Greece: an exploratory factor analysis of polychoric correlations	Vassilis Aletras	2010	26
17	Correlational analysis of ordinal data: from Pearson's r to Bayesian polychoric correlation	Jaehwa Choi	2010	75
18	A Comparison of Maximum Likelihood and Bayesian Estimation for Polychoric Correlation Using Monte Carlo Simulation	Jaehwa Choi	2011	43
19	What to Do About Zero Frequency Cells When Estimating Polychoric Correlations	Victoria Savalei	2011	35
20	Construct Validity of Likert Scales through Confirmatory Factor Analysis: A Simulation Study Comparing Different Methods of Estimation Based on Pearson and Polychoric Correlations	María de los Ángeles Morata-Ramírez	2013	51
21	POLYMAT-C: a comprehensive SPSS program for computing the polychoric correlation matrix	Urbano Lorenzo-Seva	2014	38
22	Polychoric Correlation	Scott L. Hershberger	2014	0
23	Polychoric and Polyserial Correlations	Fritz Drasgow	2014	4

24	UniODA vs. Polychoric Correlation: Number of Lambs Born Over Two Years	Paul R. Yarnold	2014	1
25	A Simulation Study of Polychoric Instrumental Variable Estimation in Structural Equation Models	Shaobo Jin	2016	16
26	Asymptotic Robustness Study of the Polychoric Correlation Estimation	Shaobo Jin	2016	15
27	Generalizations of the polychoric correlation approach for analyzing survey data	Anastasiia Timofeeva	2016	0
28	Modele równań strukturalnych a zmienne mierzone na skali porzadkowej polichorycznej w analizie danych marketingowych / Structural equation models (SEM) and variables measured on the polychoric ordinal scale in the analysis of marketing data	Piotr Tarka	2016	2
29	Selecting polychoric instrumental variables in confirmatory factor analysis: An alternative specification test and effects of instrumental variables	Shaobo Jin	2018	12
30	Polychoric Correlations for Ordered Categories Using the EM Algorithm	Kenpei Shiina	2018	3
31	Contributions to Estimation of Polychoric Correlations	Scott Monroe	2018	16
32	Pernicious Polychorics: The Impact and Detection of Underlying Non-normality	Njål Foldnes	2019	19
33	PolychoricRM: A Computationally Efficient R Function for Estimating Polychoric Correlations and their Asymptotic Covariance Matrix	Guangjian Zhang	2021	2
34	The polychoric dual-component wealth index as an alternative to the DHS index: Addressing the urban bias	P. Martel	2021	3
35	Robust polychoric correlation	Johan Lyhagen	2021	3
36	Polychoric Correlation With Ordinal Data in Nursing Research	Frank Kiwanuka	2022	4
37	Revisiting Savalei's (2011) Research on Remediating Zero-Frequency Cells in Estimating Polychoric Correlations: A Data Distribution Perspective	Tong-Rong Yang	2023	0
38	Iteratively Reweighted Least Squares Method for Estimating Polyserial and Polychoric Correlation Coefficients	Peng Zhang	2023	0

Fuente: Elaboración propia

Ilustración 1. Red de Artículos



Fuente: Elaboración propia

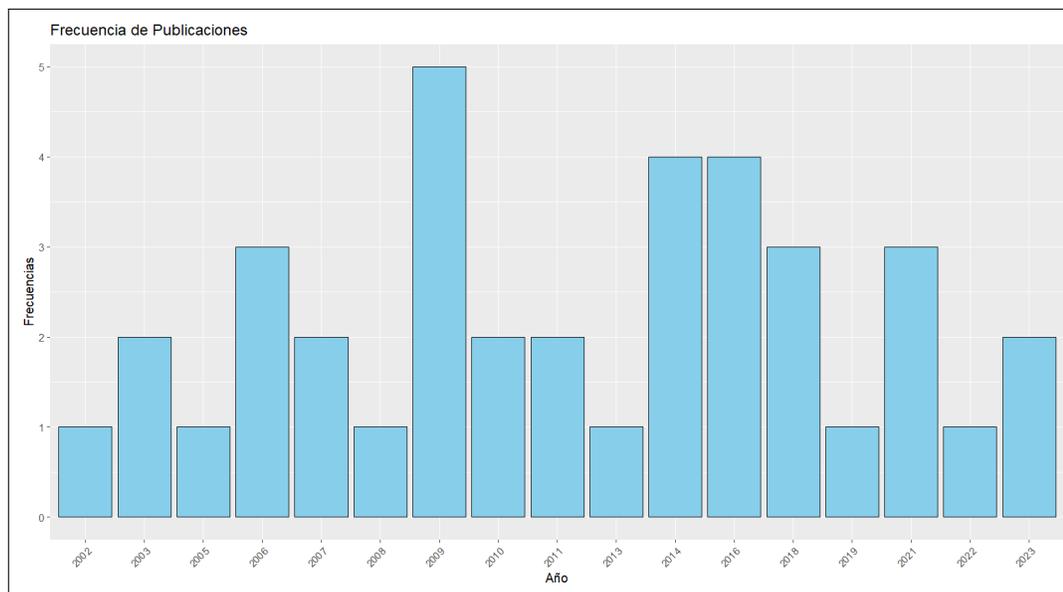
Tabla 2. Compilación de Artículos

Autor	Frecuencia
Shaobo Jin	3
Fritz Drasgow	2
Guangjian Zhang	2
Jaehwa Choi	2
Scott L. Hershberger	2
Alberto Maydeu-Olivares	1
Anastasiia Timofeeva	1
Annarita Roscino	1
Choi-Fan Yiu	1
David B. Flora	1
Frank Kiwanuka	1
Jinsong Chen	1
Joakim Ekström	1
Johan Lyhagen	1
Josip Juras	1
Kenneth A. Bollen	1
Kenpei Shiina	1
María de los Ángeles Morata-Ramírez	1
Njål Foldnes	1
P. Martel	1
Paul R. Yarnold	1
Peng Zhang	1
Piotr Tarka	1
Scott Monroe	1
Shizuhiko Nishisato	1

Stefan C. Schmukle	1
Sun-Joo Cho	1
Tong-Rong Yang	1
Urbano Lorenzo-Seva	1

Fuente: Elaboración propia

Ilustración 2. Tendencia de Fechas de Publicaciones



Fuente: Elaboración propia

En general, los hallazgos muestran un panorama dinámico en el campo de la aplicación de coeficientes policóricos y tetracóricos, se identifican temáticas dentro de áreas como: Estadística y Análisis de Datos, Modelado de Ecuaciones Estructurales (SEM), Psicometría, Métodos de Estimación de Correlaciones, Análisis de Datos Ordinales, Investigación de Mercados, Validación de Cuestionarios, Estimación de Causalidad, Métodos de Estimación de Parámetros, Instrumentos de Investigación.

Considerando lo anterior, se identifican posibles áreas de investigación futura dentro de la exploración de nuevas aplicaciones y adaptaciones de coeficientes policóricos en campos como la psicología, la sociología, la economía, la epidemiología y otras disciplinas donde los datos ordinales son comunes, igualmente, la evaluación sistemática y comparativa de diferentes métodos de estimación de coeficientes, así como de sus ventajas, limitaciones y condiciones de aplicación óptima.

3. Aplicación en R

La salud mental es un tema de creciente importancia en la sociedad actual, principalmente debido a su alta incidencia, especialmente en universitarios (Martínez-Libano, Gonzalez , & Pereira , 2022). La base de datos usada, tuvo como objetivo establecer la incidencia entre factores sociodemográficos y las problemáticas en Salud Mental. La población objeto de estudio fueron estudiantes del programa de Psicología de quinto y noveno semestre. Fue

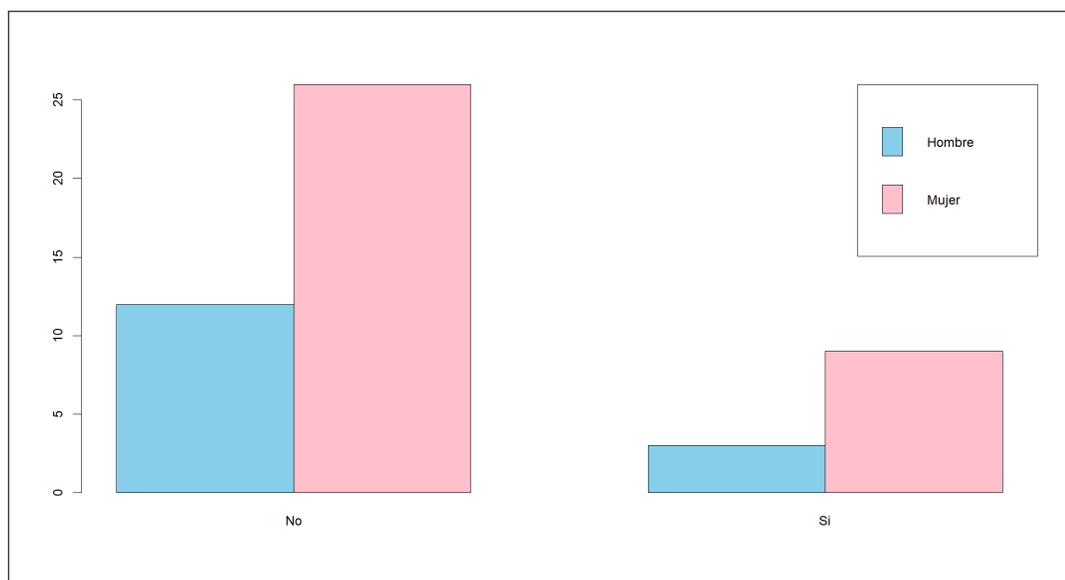
desarrollado en la última semana del periodo lectivo en el año 2022-02. Contó con un método empírico-analítico y un diseño no experimental, transversal, los datos fueron recolectados con instrumentos validados en la población objetivo.

Para el cálculo del coeficiente de correlación tetracórico se escogieron las variables sexo y diagnósticos psiquiátricos, estas dos variables cumplen con la condición de ser categóricas y dicotómicas. Se procede a realizar la tabla de contingencia y un diagrama de barras los que se observan a continuación:

	No	Si
Hombre	12	3
Mujer	26	9

Fuente: Elaboración propia

Ilustración 3. Diagrama de Barras Sexo



Fuente: Elaboración propia

La forma de asociación indicada para este tipo de variables es la aplicación de la Prueba Chi_Cuadrada que arroja un resultado de 0.00522. Este resultado, permite concluir que existe evidencia suficiente para afirmar que las variables sexo y diagnósticos psiquiátricos tienen una asociación, en este caso débil, esto sustentado por el resultado del coeficiente V de Cramer (0.06131393).

Coeficientes de correlación como el de Pearson, Spearman y en menor medida el de Kendall están diseñados para determinar la existencia o no de relaciones entre variables cuantitativas, pero cuando se van a desarrollar ACP, AFE y/o AFC se necesitan generar matrices de correlaciones que están determinadas por los coeficientes antes mencionados, una opción similar a estas mediciones la brinda el coeficiente de correlación tetracórico, que en la misma escala de interpretación arroja un resultado a partir de la tabla de contingencia 0.11. Para este ejemplo, el resultado obtenido permite identificar que hay una correlación débil entre las dos variables analizadas.

En cuanto al cálculo del coeficiente de correlación policórico, se escogieron las variables estrato socioeconómico y estado civil, las dos cumplen con ser variables categóricas y politómicas. El coeficiente de correlación policórico, es la respuesta a la mala utilización del coeficiente

de correlación de Pearson ya que, está diseñado para otro tipo de variables y en este caso el cálculo de los dos arrojan los resultados expuestos en la tabla IV.

Tabla 4. Coeficientes de Correlación

Pearson	Polychoric
-0.08940361	-0.1761

Fuente: Elaboración propia

Los dos coeficientes evidencian la misma relación inversa débil entre las dos variables, no se observan diferencias muy marcadas en sus resultados. Es importante resaltar que el cálculo del coeficiente de correlación policórico arroja una salida mas amplia que el coeficiente de Pearson.

4. Conclusiones

En resumen, el estudio demuestra que los coeficientes de correlación policóricos y tetracóricos son un tema de importancia en la investigación estadística, particularmente en psicometría y disciplinas relacionadas. Estos coeficientes ofrecen una herramienta precisa para analizar las relaciones entre muchas variables en el estudio social debido a su naturaleza categórica. Aunque las técnicas convencionales, como las correlaciones de Pearson, siguen siendo útiles, se está observando un cambio hacia el uso de coeficientes policóricos y tetracóricos debido a su capacidad para manejar datos ordinales de manera más efectiva.

El análisis de la literatura muestra una tendencia a la baja en la frecuencia de publicaciones a lo largo de los años. Hubo tres picos notables en la frecuencia de publicaciones en 2009, 2014 y 2016. Esto indica que hay un interés constante en la aplicación y el desarrollo de estos coeficientes, así como períodos de mayor actividad investigativa.

La exploración de nuevas aplicaciones y adaptaciones de los coeficientes policóricos en campos diversos como la psicología, la sociología, la economía y la epidemiología, donde los datos ordinales son comunes, es una de las áreas de investigación futuras. Además, se determina que se requiere una evaluación sistemática y comparativa de varios métodos de estimación de coeficientes, así como una evaluación de sus ventajas, desventajas y condiciones de aplicación ideales.

Este estudio proporciona un análisis completo de la relevancia y el potencial de los coeficientes de correlación policóricos y tetracóricos en la investigación estadística, enfatizando su papel en la comprensión y el análisis de datos ordinales en una variedad de disciplinas académicas.

Referencias

- [1] Freiberg , A., Beatriz, J., De la Iglesia, G., & Fernández, M. (2013). Correlaciones Policóricas Y Tetracóricas En Estudios Factoriales Exploratorios Y Confirmatorios. *Ciencias Psicológicas*, 7(2), 151-164.
- [2] Martinez-Libano, J., Gonzalez , N., & Pereira , J. (2022). Las Redes Sociales y su Influencia en la Salud Mental de los Estudiantes Universitarios: Una Revisión Sistemática. *Reidocrea*, 11(4), 44-57.

- [3] Richaud, M. (2005). Desarrollos del analisis factorial para el estudio de item dicotomicos y ordinales. *Interdisciplinaria*, 22(2), 237-251.
- [4] Ruiz-Ruano, A., Puga, J. (2016). R Como Entorno Para El Análisis Estadístico En Evaluación Psicológica. *Papeles del Psicólogo*, 37(1), 74-79.