



Aplicaciones del análisis de correspondencia, simple y múltiple. Clases latentes

Applications of correspondence analysis, simple and multiple. Latent classes

Aplicações de análise de correspondência, simples e múltiplas. Classes latentes

Yoel Hernández-Navarro ^I
yoel.hernandez@uta.edu.ec
<https://orcid.org/0000-0002-0178-4949>

Jesús E. Sánchez-García ^{II}
jesanch64@yahoo.com
<https://orcid.org/0000-0001-8137-2882>

Correspondencia: yoel.hernandez@uta.edu.ec

Ciencias técnicas y aplicadas
Artículo de investigación

***Recibido:** 05 de julio de 2020 ***Aceptado:** 20 de agosto 2020 * **Publicado:** 07 de septiembre de 2020

- I. Máster en Ciencias Matemáticas Mención Probabilidades y Estadísticas, Licenciado en Educación en la Especialidad de Matemática Computación, Investigador Independiente, Ecuador.
- II. Investigador Independiente, Ecuador.

Resumen

Se exponen ejemplos clásicos y de aplicaciones reales sobre la vinculación de varios métodos de análisis de datos categóricos.

Palabras Claves: Análisis de las correspondencias; Análisis de las frecuencias de configuraciones; análisis de clases latentes.

Abstract

Classical and real-life application examples of linking various methods of categorical data analysis are presented.

Keywords: Analysis of the correspondences; Analysis of the frequencies of configurations; latent class analysis.

Resumo

São apresentados exemplos de aplicativos clássicos e reais de vinculação de vários métodos de análise de dados categóricos.

Palavras-chave: Análise das correspondências; Análise das frequências das configurações; análise de classe latente.

Introducción

El trabajo contiene aplicaciones a problemas, uno real y otro de libro de texto, para poner de manifiesto no sólo las características del análisis con esos métodos, sino también para discutir, sobre la base de los ejemplos, algunas ideas que pueden ser de interés para los especialistas, especialmente con el fin de establecer una forma de análisis más cabal y abarcadora.

Consideraciones acerca la vinculación de los métodos

La idea de utilizar varios métodos para llegar a una mejor interpretación de un conjunto de datos no es nueva. En el caso de datos continuos es muy frecuente combinar, por ejemplo, el análisis de componentes principales con el cluster analysis y, en ciertas ocasiones, con el análisis discriminante.

Estas combinaciones se hacen intuitivamente, sin que, en la mayoría de las ocasiones, medien consideraciones teóricas que las justifiquen, sino la práctica y la capacidad de interpretación del analista, así como el conocimiento profundo del material de estudio.

De igual forma, podría procederse en el caso del análisis de datos categóricos y trabajar con los diversos métodos de manera concatenada para lograr una interpretación cabal.

No obstante, lo antes expuesto permite, hasta cierto punto, un trabajo menos dependiente de la intuición y constituye, de hecho, un interesante ensayo de justificación del uso conjunto de diversas técnicas.

En este sentido, es importante destacar que todavía existe un campo de investigación que podría llegar a ser muy productivo y que podría desembocar en una justificación más completa que permitiera la formulación de una metodología de trabajo para el análisis de datos categóricos.

Aplicaciones prácticas

En la interpretación sólo se hará hincapié en los aspectos de interés para el presente trabajo. Las implicaciones propias del análisis dentro de las disciplinas correspondientes se salen de este marco y queda abiertas para el trabajo conjunto con los especialistas. Los datos fueron procesados con el programa LEM 1.0 (VERMUNT, 1997).

Formas de suicidio en Alemania según sexo y año (Krauth & Lienert, 1973; von Eye et al., 1995).

Aquí interesa conocer si los suicidas de ambos sexos se valen de los mismos métodos para llevar a cabo su intención y, asimismo, si esto depende del tiempo que ha transcurrido desde el fin de la II Guerra Mundial.

Las variables en estudio son: Año (1944, 1952), Sexo (Masculino, Femenino), Forma de suicidio (gas, cuerda, soporífero, ahogado, venas, disparo, salto). Por tanto, se trata de una tabla de 3 entradas de dimensión 2 x 2 x 7. Los datos se presentan en forma tabular a continuación:

Tabla 1: Suicidios en Alemania

Año y forma de suicidio	Sexo	
	Masculino	Femenino
1952, gas	52	47
1952, cuerda	31	14
1952, soporífero	44	97
1952, ahogado	20	10
1952, venas	22	5
1952, disparo	3	0
1952, salto	2	2
1944, gas	16	61
1944, cuerda	76	35
1944, soporífero	7	9
1944, ahogado	19	54
1944, venas	15	4
1944, disparo	35	11
1944, salto	9	2

Resultados

Aquí sólo se presentará el análisis de los resultados. El gráfico ternario no se presenta porque los resultados no son en absoluto explicativos.

1. Análisis de las frecuencias de configuraciones (AFC)

Se determinaron 2 tipos: (1952, soporíferos, mujeres) y (1944, disparo, hombres)

Figura 1: Configuraciones en el plano determinado por el ACM

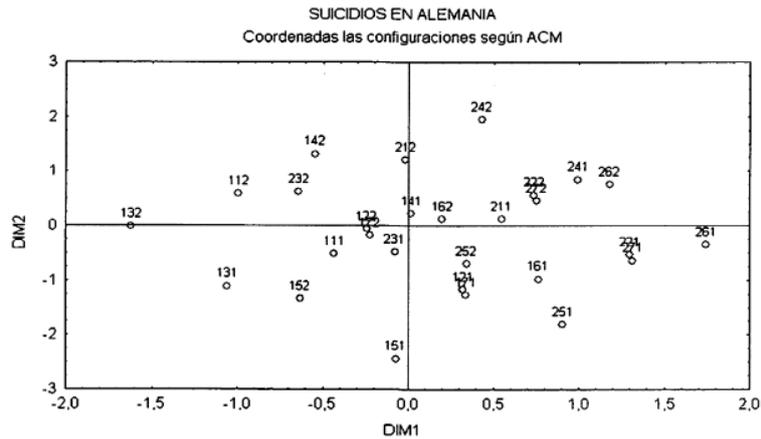
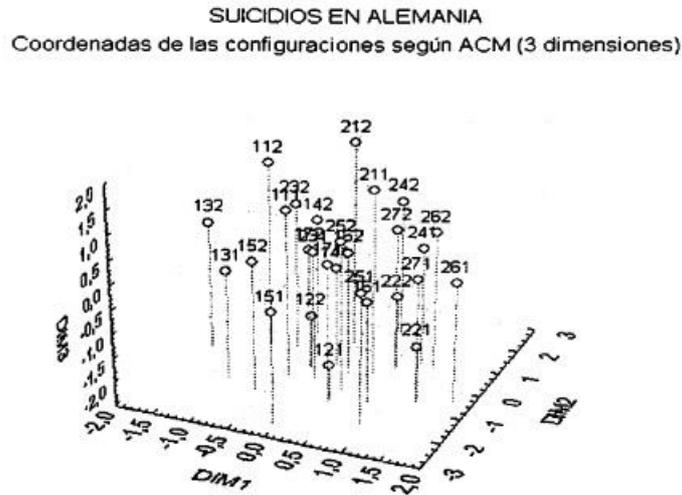


Figura 2: Configuraciones en el espacio determinado por el ACM (3 dimensiones)



Análisis de clases latentes

El ajuste del modelo de 3 clases no fue bueno. No obstante, no se buscó qué modelo ajustaba mejor para poder apreciar qué relaciones se tienen entre ambos métodos en este caso.

Se hizo la correspondiente transformación de rescalamiento de los datos. Los resultados aparecen en la tabla siguiente:

Tabla 2: Suicidios en Alemania (Rescalados)

Variables manifiestas	Clase latente		
	1	2	3
	0.4467	0.3124	0.2409
Año 1952	0.8644	0.1357	0
Año 1944	0.0338	0.4871	0.4791
Gas	0.5545	0.1110	0.3344
Cuerda	0.162	0.6181	0.2208
Soporífero	0.9292	0.0341	0.0368
Ahogado	0.2268	0.2538	0.5193
Venas	0.4097	0.5125	0.0779
Disparo	0	0.7752	0.2247
Salto	0.1861	0.6857	0.1297
Masculino	0.3751	0.6248	0
Femenino	0.5183	0	0.4818

Figura 3: Gráfico del análisis de las correspondencias múltiples

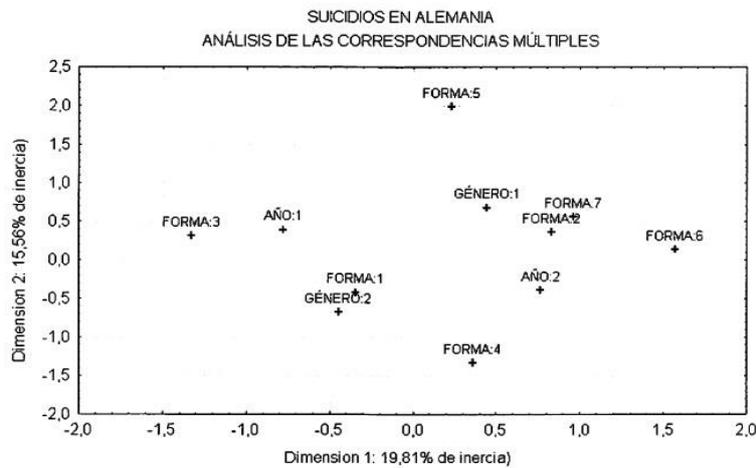


Figura 4: Gráfico de los parámetros rescalados (2 dimensiones)

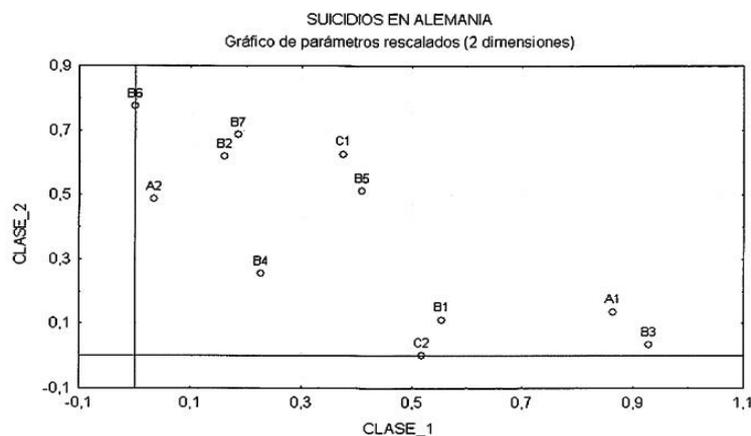
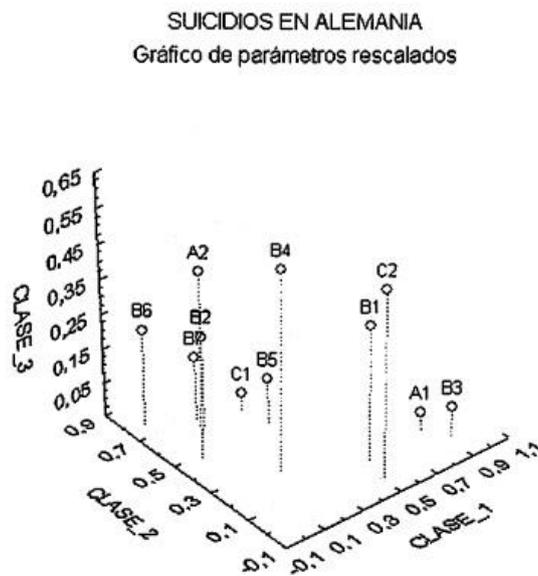


Figura 5: Gráfico de los parámetros rescalados (3 dimensiones)

Discusión

Se encontraron dos tipos, uno que habla de que el año 1952 las mujeres se suicidaban con soporíferos (1,3,2), mientras que el otro tipo da que en 1944 los hombres se suicidaban con disparos (2,6,1). Cuando se analizan las coordenadas del ACM (Figs. 2 y 3) se aprecia que efectivamente estas dos configuraciones conforman los extremos del gráfico. Un aspecto interesante que se ve también es la división de los años, con lo cual podría concluirse que hay un cambio en la forma de suicidios en esos dos momentos.

El análisis de clases latentes coincide completamente con los resultados del ACM. En ellos se aprecia una división de los años (Figs. 4, 5 y 6). Al aplicar el análisis de los ángulos en el gráfico del ACM se ve la asociación clara de las mujeres con la forma de suicidio por soporífero en el año 1952, mientras que la forma asociada al hombre aparece en el otro extremo. Es bueno señalar que la representación bidimensional no tiene un gran porcentaje de inercia, por lo que las conclusiones no tienen la fuerza que tendrían en el caso de un alto porcentaje.

Como se aprecia, aunque el modelo de clases latentes no se ajustó, se puede hacer una interpretación con el ACM que coincide con los resultados del ACL.

Desarrollo del lenguaje del niño cubano (González, Sánchez & López, 2006)

Se encuestaron 213 niños con edades entre 3 y 4 años, en el período de septiembre de 1993 a enero de 1994, en 8 provincias del país (López et al., 2000). Aquí sólo se presenta un estudio parcial referido a la expresión de los componentes del grupo léxico sustantivos (A), verbos (B), adjetivos (C) y adverbios (D). A continuación, se dan los datos:

Tabla 3: Evolución del lenguaje en niños cubanos

SUSTANTIVOS	VERBOS	ADJETIVOS	ADVERBIOS	
			Incorrecto	Correcto
Incorrecto	Incorrecto	Incorrecto	52	3
Incorrecto	Incorrecto	Correcto	5	4
Incorrecto	Correcto	Incorrecto	8	5
Incorrecto	Correcto	Correcto	3	2
Correcto	Incorrecto	Incorrecto	14	10
Correcto	Incorrecto	Correcto	10	6
Correcto	Correcto	Incorrecto	18	19
Correcto	Correcto	Correcto	17	46

Resultados

Aquí sólo se presentará el análisis de los resultados. Al igual que en el caso anterior, en este también se prescinde del gráfico ternario.

1. Análisis de las frecuencias de configuraciones (AFC)

Se determinaron 2 tipos correspondientes a la configuración (incorrecto, incorrecto, incorrecto, incorrecto) y (correcto, correcto, correcto, correcto), A continuación, se presentan los gráficos, según el ACM, de las configuraciones en 2 y 3 dimensiones:

Figura 6: Configuraciones en el plano determinado por el ACM

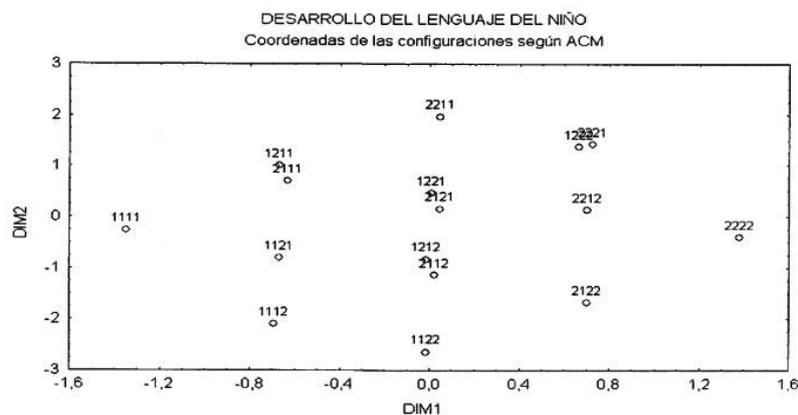
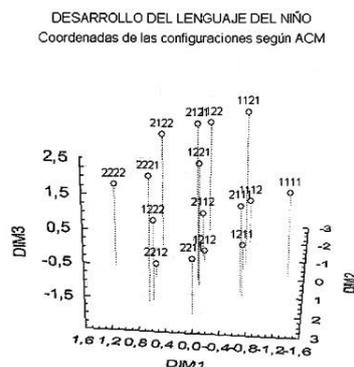


Figura 7: Configuraciones en el espacio determinado por el ACM.



2. Análisis de clases latentes

Se ajustó el modelo de 3 clases latentes con un valor de $L2 = 1.9983$ ($p = 0.1575$). En el trabajo original, también se ajustó el modelo para clases ordinales, los interesados se remiten a la publicación González et al. (2006), pero aquí sólo se trabajará con el de 3 clases por simplicidad. Se hizo la correspondiente transformación para obtener los datos rescalados que aparecen en la tabla que sigue:

Tabla 4: Desarrollo del lenguaje (parámetros rescalados)

Variables manifiestas	Clase latente		
	1	2	3
	0.2309	0.5555	0.2143
Sustantivo Incorrecto	0.600	0.401	0
Sustantivo Correcto	0	0.652	0.348
Verbo Incorrecto	0.457	0.544	0
Verbo Correcto	0.015	0.567	0.419
Adjetivo Incorrecto	0.399	0.602	0
Adjetivo Correcto	0.014	0.496	0.491
Adverbio Incorrecto	0.387	0.562	0.051
Adverbio Correcto	0	0.546	0.455

Figura 8: Gráfico del análisis de las correspondencias múltiples

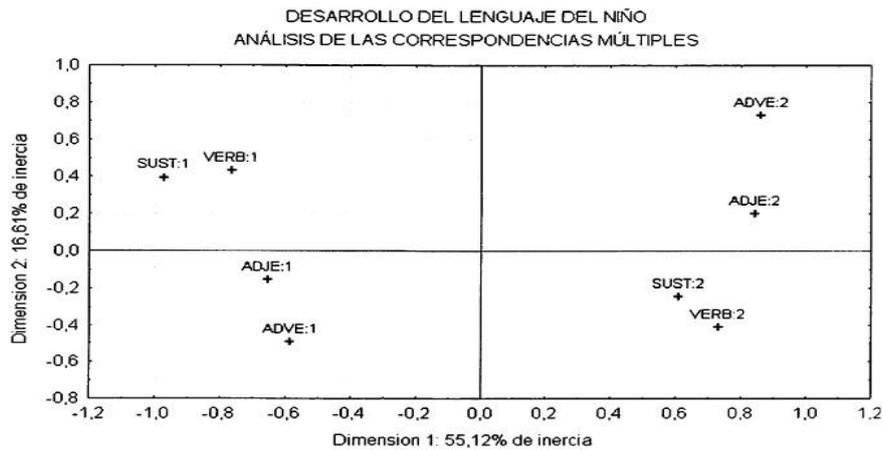


Figura 9: Gráfico de los parámetros rescalados (2 dimensiones)

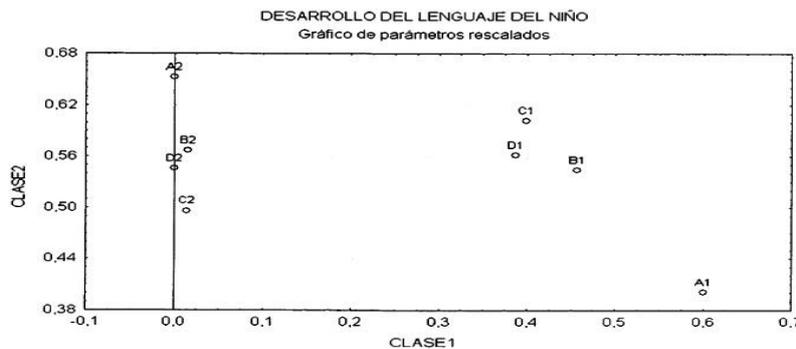
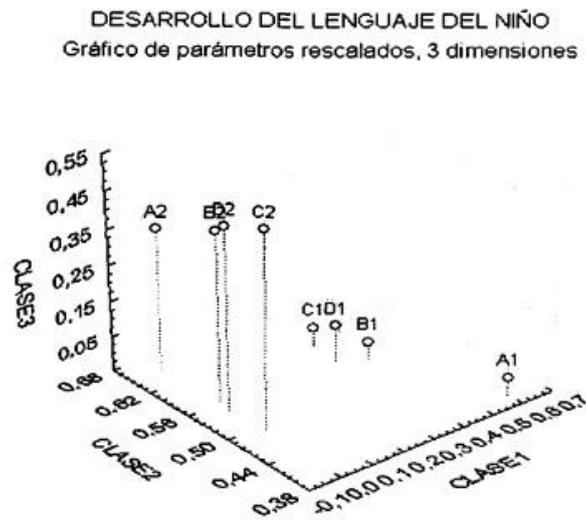


Figura 10: Gráfico de los parámetros rescalados (3 dimensiones)

Discusión

En el AFC se encontraron dos tipos, uno correspondiente a los niños que contestaron incorrectamente todos los ítems (1,1,1,1) y el otro formado por los que respondieron correctamente (2,2,2,2). Como se aprecia en el gráfico del ACM asociado a las configuraciones (Fig. 6), estas dos aparecen en los extremos, mientras que entre ellas se ordenan las configuraciones en el sentido de ir respondiendo correctamente mayor cantidad de ítems.

El ACL de 3 clases repite la estructura del AFC, ya que presenta una clase para los que responden correctamente a todos los ítems, otra para los que fallan y una tercera clase que reúne a todos aquellos que fallan en alguna. Como se aprecia en los gráficos correspondientes (Figs. 8, 9 y 10), el ACM da una estructura semejante con 2 dimensiones.

Evidentemente, en este conjunto de datos hay una componente ordinal que, si bien no se consideró explícitamente, sí sale de todos los análisis.

Al considerar el estudio del desarrollo del lenguaje con los métodos combinados se obtiene una visión más completa del fenómeno en estudio.

Conclusiones y recomendaciones

Si bien se considera que este es un estudio preliminar, no obstante, pueden formularse algunas conclusiones que resultan de interés, porque pudieran considerarse, a su vez, puntos de arranque para nuevos trabajos.

- El uso de las posibilidades gráficas del AC, en sus dos variantes, permite una visualización de los resultados de los otros dos métodos planteados, al tiempo que permite un análisis más completo, cuando se consideran sus propios resultados.
- En los ejemplos y aplicaciones realizados en este trabajo, se aprecian nuevos aspectos del problema que permiten una visión más profunda de la naturaleza de los datos.

Referencias

1. AGRESTI, A. (2002): *Categorical Data Analysis (2nd Edition)*, Wiley, Nueva York
2. BENZÉCRI, J.-P et collaborateurs (1973) : *L'Analyse des Donnés. L'Analyse des Correspondences*, Dunoá, París
3. BENZÉCRI, J.-P. (1992): *Correspondence Analysis Handbook*, Marcel Dekker, Inc., Nueva York, 665 + xi pp.
4. FERNÁNDEZ, R. S. M. (2011). *Análisis de correspondencias simples y múltiples*. Universidad Autónoma de Madrid: Facultad de Ciencias Económicas y Empresariales.
5. GABRIEL, K. R. (1971): The biplot graphic display of matrices with application to principal component analysis, *Biometrika* 58(3), 453-467 pp.
6. GONZALEZ, D. A. (2006): Dos enfoques para el análisis de clases latentes ordinales. En: *Revista de la Facultad de Matemática y Computación de la Universidad de La Habana. Cuba. Vol. 27, Núm. 1.*
7. GONZALEZ, D. A. (2006): Algunas consideraciones prácticas acerca de la estimación de parámetros en el modelo clásico de clases latentes. En: *Revista de la Facultad de Matemática y Computación de la Universidad de La Habana. Cuba. Vol. 27, Núm. 1.*
8. GONZÁLEZ, A., SÁNCHEZ, J.E. y LÓPEZ, Marcia (2006): Dos enfoques para el análisis de clases latentes ordinales. *Investigación Operacional*, 27(1), pp. 4-11
9. GOODMAN, L.A. (1997): *Statistical Methods, Graphical Displays, and Tukey's Ladder of Re-expression in the Analysis of Nonindependence in Contingency Tables: Correspondence. Analysis, Association Analysis, and the Midway View of Nonindependence*, en: BRILLINGER, D., FERNHOLZ, L.T. & MORGENTHALER, S.: *The Practice of Data Analysis: Essays in Honor of John W. Tukey*, Princeton, Nueva Jersey, Princeton University Press, pp0. 101-132

10. GOODMAN, L. A. (2000) : The Analysis of Cross-Classified Data: Notes on a Century of Progress in Contingency Table Analysis, and Some Comments on Its Prehistory and Its Future, Marcel Dekker, Inc., New York, 231 + i pp.
11. GREENACRE, M. (1984): Theory and Applications of Correspondence Analysis, Academic Press, Londres
12. GREENACRE, M. (2006) : Tying up the loose ends in simple correspondence analysis, Economic Working Paper 940.
13. GREENACRE, M. (2005): From correspondence analysis to múltiple and joint correspondence analysis.
14. JAMBU, M. (1991) : Exploratory and Multivariate Data Analysis, Academic Press, Inc., Boston, 474 + xv pp.
15. KRAUTH, J. y LIENERT, G.A. (1973) : Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin, Freiburg, Albert
16. LAUTSCH, E. y PLICHTA, M.M. (2003): Configural Frequency Analysis (CFA), Múltiple Correspondence Analysis (MCA) and Latent Class Analysis (LCA): An empirical comparison, Psychology Science 45(2), 298-323 pp.
17. LAZARFELD, P.F. (1950): The logical and mathematical foundation of latent structure analysis. En: STOUFFER, S. A. et al. (Eds.): The American Soldier, Vol. IV, Measurement and Prediction, Princeton
18. LAZARFELD, P.F. y HENRY, N. W. (1968): Latent Structure Analysis, Houghton Mifflin, Boston
19. LEEUW, J. & Van der HEIJDEN, P.G.M. (1991): Reduced rank models for contingency tablas, Biometrika, 78, pp. 229-232
20. LÓPEZ, M., REGAL, N. , PASCUAL, M., GONZÁLEZ, A. y SÁNCHEZ, J.E. (2000): Desarrollo del lenguaje del niño cubano menor de 18 meses. Revista Cubana de Pediatría, 72(1), 32-39 (pdf)
21. MOOD, A.M., GRAYBILL, R.A. y BOES, D.C. (1974): Introduction to the Theory of Statistics (3rd Edition), MacGraw Hill, Londres
22. MONTGOMERY, D, C. (1991) : Design and Analysis of Experiments, Third Edition, John Wiley and Sons, Nueva York, 649 + xvii pp.
23. RAO, C.R. y SZÉKELY, G.J. (Eds.) (2000) : Statistics for the 21st Century. Methodologies for Applications of the Future, Marcel Dekker, Inc., Nueva York
24. Van der HEIJDEN, GILULA, P. G. y van der ARK, L. A. (1999) : An extended study into the relationships between correspondence analysis and latent class analysis, 40 + vii pp.

25. VERMUNT, J. K. (1997): LEM 1.0: A general programa for the analysis of categorical data. Tilburg: Tilburg university Von EYE, A. y NIEDERMEIER, K.E. (1999) : Statistical Analysis of Longitudinal Categorical Data in the Social and Behavioral Sciences, Lawrence Erlbaum Associates, Mahwah, Nueva Jersey.
26. Von EYE, A., SPIEL, C. y ROVINE, M.J. (1995): Concepts of nonindependence in configural frequency analysis, Journal of Mathematical Sociology, 20, pp. 41-54

©2020 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).