

Ventajas metodológicas del uso del análisis de clases latentes (LCA) para investigar en comunicación. Aplicación empírica al estudio de la confianza en la ciencia

---

*Methodological Advantages of Using Latent Class Analysis (LCA) for Communication Research: An Empirical Application to the Study of Trust in Science*

**Felipe Núñez.** Universidad Carlos III de Madrid (España)

Doctorando en Investigación de Medios en la Universidad Carlos III de Madrid (UC3M), con una beca nacional FPU. Su investigación se centra en el cambio climático y la percepción social de la ciencia, investigando cómo los medios influyen en la comprensión y las actitudes del público hacia los problemas ambientales.

ORCID: <https://orcid.org/0000-0002-5898-700X>

**Carlos Elías.** Universidad Carlos III de Madrid (España)

Catedrático de Periodismo de la Universidad Carlos III de Madrid (UC3M) y catedrático europeo Jean Monnet “EU, desinformación y fake news”. Está especializado en la intersección entre medios, ciencia, tecnología y desinformación. Su último libro es *Science on the Ropes. Decline of Scientific Culture in the era of Fake News* (Springer, 2019). ORCID: <https://orcid.org/0000-0002-1330-4324>

Artículo recibido: 09/04/2026 – Aceptado: 25/05/2026

**Resumen:**

Este artículo fundamenta y defiende el uso del Análisis de Clases Latentes (LCA) como herramienta metodológica para la investigación en comunicación, con especial atención a sus ventajas frente a los métodos de clasificación tradicionales. A partir de una revisión conceptual de los enfoques supervisados y no supervisados, se argumenta que el LCA ofrece ventajas sustantivas al incorporar incertidumbre en la asignación de los casos, modelar la heterogeneidad social desde una perspectiva probabilística y trabajar con variables categóricas u ordinales, frecuentes en estudios de comunicación. Se discuten además los criterios de selección del modelo, los indicadores de ajuste y los diagnósticos de clasificación. Como ilustración empírica, se aplica el paquete poLCA en R a datos de la Encuesta de Percepción Social de la Ciencia y la Tecnología en España (FECYT,

2024), identificando cuatro perfiles latentes de confianza en los científicos. El trabajo concluye destacando el potencial del LCA como puente entre teoría y medición en la investigación cuantitativa en comunicación.

**Palabras clave:**

Análisis de Clases Latentes; Segmentación de audiencias; Métodos cuantitativos

**Abstract:**

*This article grounds and advocates for the use of Latent Class Analysis (LCA) as a methodological tool for communication research, with particular attention to its advantages over traditional classification methods. Drawing on a conceptual review of supervised and unsupervised approaches, it argues that LCA offers substantive advantages by incorporating uncertainty in case assignment, modeling social heterogeneity from a probabilistic perspective, and handling categorical and ordinal variables commonly found in communication studies. Model selection criteria, goodness-of-fit indicators, and classification diagnostics are also discussed. As an empirical illustration, the poLCA package in R is applied to data from the Social Perception of Science and Technology Survey in Spain (FECYT, 2024), identifying four latent profiles of trust in scientists. The article concludes by highlighting the potential of LCA as a bridge between theory and measurement in quantitative communication research.*

**Keywords:**

*Latent class analysis; audience segmentation; Quantitative methods*

## 1. Introducción

Clasificar constituye una de las operaciones más básicas y, al mismo tiempo, más trascendentes en la producción de conocimiento científico. Implica organizar y estructurar la realidad agrupando los elementos de estudio en categorías definidas por características compartidas, ya sean observables o inferidas (Bailey, 2003; Sokal, 1974). Este proceso requiere establecer criterios explícitos de inclusión y exclusión, que determinen qué objetos, individuos o fenómenos son considerados equivalentes dentro de una categoría. Asimismo, dichos criterios operan como definiciones taxonómicas, en la medida en que delimitan el conjunto mínimo de propiedades necesarias para identificar una entidad o fenómeno como miembro de una clase específica (Hodgson, 2019).

En las ciencias sociales, la clasificación cumple una función epistemológica central, ya que permite identificar uniformidades y patrones predecibles en la diversidad de los comportamientos humanos y sociales. A través de la categorización, el investigador logra simplificar y ordenar la complejidad del mundo social, transformando la experiencia empírica en estructuras conceptuales que facilitan la formulación de teorías y la comprensión de las dinámicas colectivas (Turner, 1987).

En este contexto, las categorías sociales funcionan como herramientas analíticas que agrupan individuos o grupos que comparten atributos comunes, como la nacionalidad, la edad, el género o la ideología política, y permiten describir, comparar e interpretar

fenómenos sociales. En este sentido, una categoría social puede definirse como un conjunto de elementos que presentan cierta similitud relevante entre sí (Forsyth, 2019). Por ejemplo, los ciudadanos de Irlanda constituyen la categoría “irlandeses”, que delimita una identidad nacional.

Sin embargo, cuando las categorías se aplican a individuos, su alcance trasciende lo meramente descriptivo. Estas categorías pueden adquirir un carácter performativo y relacional, ya que aluden a procesos psicológicos e interpersonales asociados a la identidad social (Tajfel, 1974). Así, las personas no solo son clasificadas dentro de un grupo, sino que también se reconocen y actúan en función de dicha pertenencia. Un irlandés, por ejemplo, puede celebrar el Día de San Patricio como expresión de su identidad colectiva y de su conexión simbólica con el grupo al que pertenece (Forsyth, 2019).

Dentro de la investigación en comunicación, clasificar es una tarea esencial para comprender cómo se estructuran, producen y consumen los mensajes mediáticos. Los investigadores de medios clasifican audiencias, contenidos y prácticas comunicativas con el fin de detectar patrones de comportamiento, segmentar públicos y analizar dinámicas discursivas. Por ejemplo, la clasificación de audiencias permite identificar grupos de usuarios con intereses o hábitos mediáticos similares (Metag y Schäfer, 2018; Napoli, 2011), lo que facilita comprender fenómenos como la polarización informativa (Bail, 2021) o la creación de cámaras de eco (Flaxman et al., 2016). En el análisis del contenido, la clasificación de mensajes en géneros, marcos o temas contribuye a reconocer las estrategias narrativas y los sesgos en la producción informativa (Budak et al., 2016; Entman, 1993). Asimismo, en el entorno digital contemporáneo, los algoritmos de las plataformas de comunicación operan de forma continua como mecanismos automáticos de clasificación, agrupando usuarios y contenidos en función de sus interacciones y preferencias (Beer, 2017). Estas formas de clasificación —humanas o automatizadas— reflejan, en última instancia, modelos de interpretación social que definen qué información se considera relevante, cómo se organiza el espacio mediático y qué identidades se construyen en él.

Desde un punto de vista cuantitativo, los procedimientos de clasificación se abordan a partir de las técnicas de análisis multivariante, una rama de la estadística que permite considerar las interrelaciones entre muchas variables de forma simultánea (Bartholomew et al., 2008). En estadística, inicialmente se utilizó el término *clustering* para englobar esta serie de técnicas, ya que *classification* se utilizaba para referirse al análisis discriminante, un procedimiento mediante el cual se desarrolla una regla de clasificación para colocar una observación multivariada en una clase previamente establecida (Hartigan, 2015).

Las técnicas que funcionan de esta forma, es decir, conociendo de antemano las categorías o grupos en los que queremos clasificar nuestras observaciones, se conocen como métodos supervisados (James et al., 2013, p. 26). En su vertiente predictiva, estos métodos siguen un proceso general que comienza con el entrenamiento de un modelo a partir de un conjunto de datos previamente etiquetados. El modelo identifica patrones en dichos datos y aprende a reconocer las características que distinguen cada categoría.

Posteriormente, aplica ese conocimiento para clasificar nuevas observaciones no etiquetadas dentro de las categorías predefinidas.

Dentro de esta modalidad encontramos desde los modelos clásicos de regresión lineal y logística hasta técnicas más modernas como *support vector machines*. En el campo de la comunicación, un ejemplo ilustrativo sería el siguiente: un medio digital dispone de un corpus de artículos etiquetados en secciones temáticas (deportes, política, tecnología) y utiliza esos textos para entrenar un modelo de clasificación automática. El modelo analiza los patrones lingüísticos característicos de cada sección y, una vez entrenado, puede asignar a nuevas noticias no etiquetadas la categoría correspondiente. Así, un artículo que contenga referencias a elecciones, partidos políticos o gobierno sería clasificado automáticamente como perteneciente a la sección de política.

En contraposición a los métodos supervisados se encuentran los métodos no supervisados, aquellos que no parten de categorías o etiquetas predefinidas, sino que buscan descubrir estructuras o patrones implícitos en los datos y generar agrupamientos a partir de ellos (James et al., 2013, p. 26). El objetivo de estos procedimientos es establecer una clasificación emergente, basada exclusivamente en la forma en que las observaciones se relacionan entre sí, sin intervención directa por parte del investigador en la definición de las clases. Un ejemplo de aplicación en el ámbito de la comunicación sería el siguiente: un medio digital recopila un amplio corpus de artículos sin etiquetas ni secciones temáticas asignadas. El modelo analiza los textos y detecta similitudes en el uso del lenguaje, agrupándolos automáticamente en distintos conjuntos. Sin supervisión humana, identifica que ciertos artículos contienen con frecuencia términos como elecciones o gobierno y los agrupa dentro de un tema latente relacionado con política.

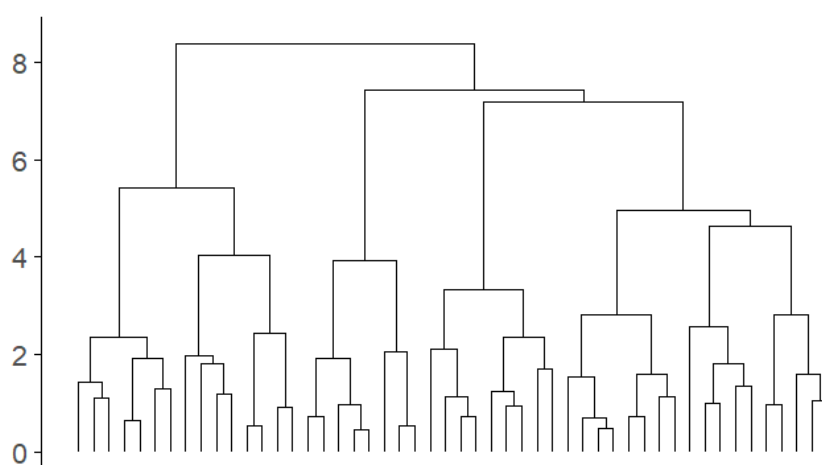
Dentro de los modelos no supervisados encontramos las técnicas de *clustering*, cuyo objetivo es determinar si las observaciones se dividen en grupos relativamente distintos. Por ejemplo, en un estudio de segmentación de audiencias podríamos observar múltiples características de los públicos, como el código postal, los ingresos familiares y los hábitos de consumo mediático. Podríamos pensar que las audiencias se dividen en diferentes grupos, como, por ejemplo, los que están dispuestos a pagar por una plataforma audiovisual y los que no. Si tuviéramos a nuestra disposición datos sobre suscripciones, podríamos llevar a cabo un análisis supervisado. Sin embargo, si esta información no está disponible, podemos intentar agrupar a las audiencias en función de cómo se comportan las variables que sí tenemos a nuestra disposición.

A su vez, las técnicas de *clustering* pueden dividirse en jerárquicas y no jerárquicas. En el primer caso, el proceso de agrupación produce una jerarquía secuencial en la que los subconjuntos de un nivel se agregan para formar los subconjuntos del siguiente nivel. Este procedimiento no requiere conocer de antemano el número de conglomerados deseado y permite observar cómo las observaciones se van fusionando en etapas sucesivas. La representación gráfica más común de este proceso es el dendrograma (figura 1), una gráfica esquemática que muestra la secuencia completa de fusiones y la distancia a la que cada elemento se une a otro grupo.

Por otra parte, los modelos no jerárquicos engloban un conjunto más diverso de técnicas que, a diferencia de los anteriores, no imponen una estructura de anidamiento entre los grupos. Entre ellas, una de las más utilizadas es el método *k-means*, que divide un conjunto de datos en *k* grupos distintos, previamente especificados y mutuamente excluyentes, a partir de un criterio de optimización (Figura 2). La idea subyacente a esta técnica es minimizar la variabilidad interna de cada grupo y, al mismo tiempo, maximizar la diferencia entre grupos (Steinley, 2006). En otras palabras, la partición será más adecuada cuanto menor sea la variación dentro de cada conglomerado respecto a la media de su grupo.

**Figura 1**

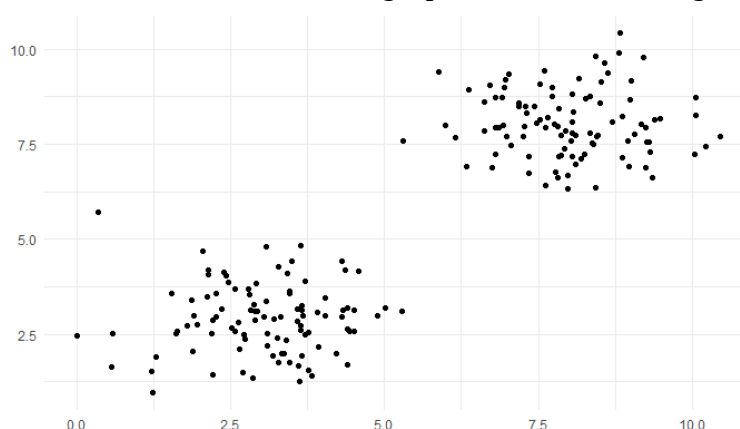
*Dendrograma del análisis de conglomerados jerárquico*



*Nota.* Fuente: Elaboración propia a partir de datos simulados.

**Figura 2**

*Distribución de observaciones agrupadas mediante el algoritmo K-means*



*Nota.* Fuente: Elaboración propia a partir de datos simulados.

Si bien estas técnicas han resultado sumamente útiles para la segmentación de públicos y el análisis de patrones mediáticos, presentan limitaciones importantes cuando se pretende capturar la variabilidad latente que subyace a las respuestas observadas. En particular, los métodos de *clustering* tradicionales suponen que cada observación pertenece a un único grupo de forma determinista, sin estimar la probabilidad de pertenencia a otras categorías. Frente a esta perspectiva, los modelos probabilísticos de clasificación, como el LCA, asignan a cada observación una probabilidad de pertenencia a las distintas clases que identifique el modelo. Esta aproximación resulta especialmente valiosa cuando los límites entre categorías no son nítidos, como ocurre en la identificación ideológica, las percepciones mediáticas o las orientaciones de consumo.

En este contexto, el presente artículo tiene un doble objetivo. Por un lado, ofrecer una fundamentación metodológica rigurosa del LCA como herramienta de clasificación para la investigación en comunicación, situándolo en relación con otros métodos cuantitativos y discutiendo sus ventajas, limitaciones y criterios de aplicación. Por otro lado, ilustrar su uso práctico mediante una aplicación empírica a datos reales, con el fin de que el lector pueda valorar tanto el procedimiento técnico como su potencial interpretativo en estudios de audiencias, actitudes y comportamiento mediático.

## 2. Propuesta metodológica: Análisis de clases latentes (LCA)

### 2.1. Fundamentación y objetivos metodológicos

El LCA fue introducido por Paul Lazarsfeld (1950; 1968) como un modelo matemático destinado a identificar tipologías no observadas en los datos empíricos. Su propósito original era comprender mejor las diferencias individuales en actitudes, opiniones y comportamientos a partir de respuestas observadas en encuestas, en el contexto de los estudios sociales desarrollados durante la Segunda Guerra Mundial. Este origen histórico es especialmente significativo para los estudios de comunicación, ya que el LCA surge dentro del mismo marco intelectual que dio forma a la *Mass Communication Research*, corriente en la que Lazarsfeld fue pionero al vincular la teoría social con métodos estadísticos de análisis de datos.

Posteriormente, el método fue ampliado y formalizado matemáticamente por Leo Goodman (1974), quien estableció un marco estadístico para la identificación y estimación de los parámetros del modelo. Y, desde entonces, el LCA se ha consolidado como una técnica estadística diseñada para detectar grupos latentes dentro de un conjunto de observaciones categóricas, modelando la heterogeneidad no observable que subyace a los datos (Nylund-Gibson y Choi, 2018; Weller et al., 2020).

Su fundamento reside en el análisis de los patrones de covariación entre múltiples variables manifiestas —u observadas—, a partir de los cuales se infieren las clases o tipologías subyacentes (Collins y Lanza, 2010; McCutcheon, 1987). En términos simples, si pensamos en una encuesta, el LCA parte de la idea de que las personas que responden de forma parecida en varias preguntas probablemente comparten una característica común, no medida directamente, que explica las similitudes. De esta forma, el LCA extiende los principios del *clustering* tradicional, pero incorporando un marco

probabilístico que permite estimar la pertenencia de cada individuo a una clase latente, junto con la incertidumbre asociada a esa asignación; y además, al igual que los modelos factoriales, asume que ese elemento subyacente es la razón que impulsa las puntuaciones de los ítems observables (Kline, 2016).

Desde un punto de vista técnico, el LCA busca encontrar el modelo que mejor representa los patrones de relación entre las variables observadas (Nylund-Gibson y Choi, 2018). Y, en la práctica, para el caso de una encuesta, esto significa que el método intenta descubrir cuántos grupos ocultos existen en los datos, cuántas personas pertenecen aproximadamente a cada grupo y qué probabilidades tiene cada persona de dar determinadas respuestas dentro de cada clase (Nylund-Gibson y Choi, 2018). De este modo, una vez que sabemos a qué clase pertenece una persona, sus respuestas a las preguntas ya no están relacionadas entre sí, porque cualquier relación entre ellas se debe únicamente a esa característica común que define su clase.

Según esta técnica, si dos preguntas de una encuesta —como “confía en la televisión” y “confía en la prensa”— suelen tener respuestas parecidas, no es porque una respuesta cause la otra, sino porque ambas reflejan el mismo rasgo latente —como el nivel general de confianza en los medios—. Esto se llama independencia local, y es un principio por el cual las variables observadas son estadísticamente independientes entre sí una vez controlada la clase latente, lo que implica que la covariación entre los ítems se explica enteramente por la variable latente subyacente (Visser y Depaoli, 2022).

El LCA ofrece, por tanto, una vía metodológica para representar empíricamente diferentes categorías de constructos teóricos complejos, traduciendo dimensiones abstractas en configuraciones estadísticas interpretables. De esta manera, se convierte en un puente entre los modelos de clasificación conceptuales y la evidencia empírica que los sustenta.

## 2.2. LCA frente a otros modelos de clasificación

El LCA se diferencia de los métodos tradicionales de agrupamiento —como el *k-means* o los métodos jerárquicos— tanto en su fundamentación teórica como en su estructura estadística. Mientras que los procedimientos clásicos de *clustering* se basan en la proximidad geométrica entre observaciones en un espacio de variables, el LCA se apoya en un modelo probabilístico que asume la existencia de una variable latente categórica que explica las asociaciones observadas entre los ítems (Magidson y Vermunt, 2002).

En los métodos tradicionales de agrupamiento, como *k-means* o *k-median*, la asignación de cada individuo a un grupo es determinista: cada caso pertenece a un único clúster según su distancia a un centroide. En cambio, el LCA introduce un enfoque probabilístico y flexible, permitiendo que cada observación tenga una probabilidad de pertenecer a varias clases. De este modo, la clasificación no se concibe como una etiqueta fija, sino como una distribución probabilística que refleja la posición relativa del individuo en un espacio latente de significados o comportamientos (Magidson y Vermunt, 2002).

Por otra parte, los métodos jerárquicos ofrecen una representación útil de las relaciones entre los casos mediante dendrogramas, pero carecen de una base probabilística formal

que permita evaluar la calidad del modelo o realizar inferencias sobre la estructura subyacente. El LCA, en cambio, permite comparar modelos con distinto número de clases mediante criterios de información (AIC, BIC) y ofrece estimaciones de incertidumbre asociadas a cada asignación (Nylund et al., 2007).

Otra diferencia sustantiva reside en la naturaleza de las variables admitidas. Mientras que *k-means* y sus variantes operan con datos métricos o de intervalo, el LCA está especialmente diseñado para variables categóricas u ordinales, lo que lo hace particularmente adecuado para analizar datos de encuestas o cuestionarios con escalas tipo Likert, muy comunes en investigación en comunicación (Oberski, 2016).

### 2.3. Interpretación del modelo y discusión metodológica

Antes de realizar el LCA, los investigadores deben seleccionar el estimador, es decir, el procedimiento estadístico mediante el cual se obtendrán las estimaciones de los parámetros del modelo. En la mayoría de los paquetes estadísticos, este estimador viene predefinido; en el caso del paquete *poLCA* de R (Linzer y Lewis, 2011), el método utilizado es el de Máxima Verosimilitud (*Maximum Likelihood*, ML). Este procedimiento estima los parámetros del modelo —tales como las probabilidades de pertenencia a cada clase latente y las probabilidades condicionales de respuesta para cada ítem dentro de cada clase— maximizando la función de verosimilitud, es decir, buscando el conjunto de valores que hace que la probabilidad de observar los datos empíricos sea la más alta posible dado el modelo especificado. Concretamente, *poLCA* implementa este proceso mediante algoritmos iterativos de *Expectation–Maximization* (EM) y *Newton–Raphson*, que garantizan la convergencia hacia un óptimo local de la verosimilitud. Este estimador es ampliamente aceptado en la literatura especializada por su robustez y eficiencia en muestras moderadas a grandes (Linzer y Lewis, 2011). En otras palabras, el programa ajusta repetidamente los valores del modelo hasta encontrar aquellos que hacen que las respuestas observadas sean las más coherentes con la estructura de clases propuesta.

El paquete *poLCA* se ha convertido en una de las implementaciones más utilizadas para aplicar LCA a datos de encuesta con variables politómicas. Sus aplicaciones típicas incluyen el análisis de encuestas de opinión, el estudio de estilos de vida y elección del consumidor, y otros fenómenos sociales y comportamentales. En el campo de la comunicación y áreas afines, el LCA se ha empleado para identificar tipologías de usuarios en redes sociales en función de su comportamiento político (Philipps et al., 2025), para segmentar públicos según sus repertorios de búsqueda de información durante la pandemia de COVID-19 (Johansson et al., 2024), o para caracterizar distintos perfiles de audiencia en comunicación de riesgo y salud (Piltch-Loeb et al., 2024). Estos antecedentes ilustran el potencial del método para capturar heterogeneidad latente en poblaciones con comportamientos mediáticos y actitudinales complejos.

Una vez seleccionado el estimador, el procedimiento estándar del LCA consiste en ejecutar una serie de modelos, empezando por uno de una sola clase y añadiendo progresivamente más clases —dos, tres, cuatro, etc.—, comparando cada modelo con el anterior. Se continúa hasta identificar el modelo que ofrece el mejor equilibrio entre ajuste estadístico y sentido teórico, ya que agregar más clases no siempre mejora la calidad del

modelo. En la práctica, la elección del número de clases se basa en varios criterios estadísticos y de interpretabilidad. Entre los índices de ajuste más usados están el Criterio de Información Bayesiano (BIC), considerado el más confiable (Nylund-Gibson y Choi, 2018; Weller et al., 2020), y otros como el Akaike Information Criterion (AIC), donde valores más bajos indican mejor ajuste. No obstante, aunque los indicadores estadísticos son útiles, el modelo seleccionado debe tener coherencia teórica (Nylund et al., 2007).

Además de evaluar el ajuste global del modelo, es necesario examinar los diagnósticos de clasificación, ya que aportan información sobre la precisión con la que se asigna a los individuos a las clases latentes. Uno de los indicadores más utilizados es la probabilidad posterior promedio, que refleja el grado en que el modelo predice correctamente la pertenencia a las clases, que idealmente debería estar por encima de 0,80. Otro indicador es la entropía, que evalúa cuán claramente el modelo distingue las clases, y que no debería ser inferior a 0,60. Asimismo, se recomienda observar el tamaño de las clases, evitando que sean demasiado pequeñas, aunque esta regla es flexible siempre que el modelo tenga sentido conceptual (Weller et al., 2020).

La interpretación substantiva del modelo, por tanto, requiere combinar los resultados estadísticos con el marco teórico de la investigación. Los nombres o etiquetas asignadas a las clases no emergen del modelo de manera automática, sino que son construcciones interpretativas que reflejan el significado conceptual de los patrones hallados (Magidson y Vermunt, 2002). Esta etapa es crucial para garantizar que la solución estadística tenga coherencia teórica y relevancia analítica dentro del campo de estudio.

#### 2.4. Limitaciones de LCA

El LCA, pese a sus ventajas, presenta un conjunto de limitaciones que el investigador debe considerar antes de su aplicación. En primer lugar, requiere un tamaño muestral mayor que los métodos de *clustering* tradicionales, y su complejidad computacional aumenta de forma significativa conforme se incrementa el número de clases o de ítems incluidos en el modelo. En segundo lugar, la solución obtenida puede corresponder a un máximo local de la función de verosimilitud en lugar del máximo global, por lo que se recomienda ejecutar el modelo múltiples veces con distintos valores iniciales aleatorios para garantizar la estabilidad de los resultados (Linzer y Lewis, 2011). En tercer lugar, la interpretación y denominación de las clases no emerge automáticamente del modelo, sino que depende del juicio teórico del investigador, lo que puede introducir un grado de subjetividad en la fase de rotulación (Brusco et al., 2017). Finalmente, el principio de independencia local —que asume que las variables observadas son estadísticamente independientes entre sí una vez controlada la clase latente— puede ser violado en la práctica, lo que requiere diagnóstico y, en su caso, ajustes en la especificación del modelo (Visser y Depaoli, 2022).

### 3. Aplicación práctica de un análisis de clases latentes

Con el fin de ilustrar la aplicabilidad del LCA, presentamos un ejemplo empírico utilizando el lenguaje estadístico R y el paquete *poLCA* (Linzer y Lewis, 2011). Este entorno ofrece una implementación flexible y de libre acceso que permite ajustar modelos

de clases latentes sobre datos categóricos, visualizar resultados y comparar diferentes soluciones de clasificación.

Para este caso práctico hemos utilizado datos reales de la Encuesta de Percepción Social de la Ciencia y la Tecnología en España (EPSCYT), elaborada por la Fundación Española para la Ciencia y la Tecnología (FECYT). Esta encuesta, de carácter bienal, constituye uno de los instrumentos más consolidados para el estudio de las actitudes ciudadanas hacia la ciencia en el contexto español, y en su edición de 2024 contó con un total de 4.521 personas entrevistadas (FECYT, 2024). En nuestro ejercicio hemos trabajado con la última edición disponible, correspondiente al año 2024, con el propósito de explorar cómo se estructura la confianza en los científicos. Para ello, nos centramos en el bloque temático dedicado a dicha confianza, que incluye una batería de ítems evaluados en una escala de cinco puntos, desde "totalmente en desacuerdo" a "totalmente de acuerdo". Tras seleccionar los casos con respuesta completa en todos los ítems del bloque, la muestra analítica empleada en el LCA fue de 1168 participantes. El conjunto de ítems seleccionados aborda distintas dimensiones de la confianza en los científicos, que desglosamos en la Tabla 1.

**Tabla 1**

*Ítems que miden la confianza en los científicos*

Ítems	Descripción
Competencia	Los/as científicos/as son expertos/as en su campo.
Rigor	Los/as científicos/as trabajan de acuerdo a reglas y procedimientos estándar.
Altruismo	Los/as científicos/as investigan por el interés común.
Transparencia	Los/as científicos/as informan a la ciudadanía sobre los resultados relevantes de su investigación.
Participación	Los/as científicos/as tienen suficientemente en cuenta a la ciudadanía en su investigación.

*Nota.* Elaboración propia a partir de datos de FECYT (2024).

Con el fin de evaluar la adecuación de este conjunto de indicadores para su utilización como un constructo coherente, examinamos la consistencia interna mediante el coeficiente alfa de Cronbach, que alcanzó un valor de 0,74. Este nivel se considera indicativo de una fiabilidad aceptable para fines de investigación social aplicada, lo que respalda la interpretación del conjunto como una escala unidimensional de confianza en los científicos y, por tanto, su idoneidad para el análisis posterior a través de modelos de clases latentes.

Mientras que el coeficiente alfa de Cronbach permite comprobar que los ítems seleccionados funcionan como indicadores de un mismo constructo subyacente, el LCA

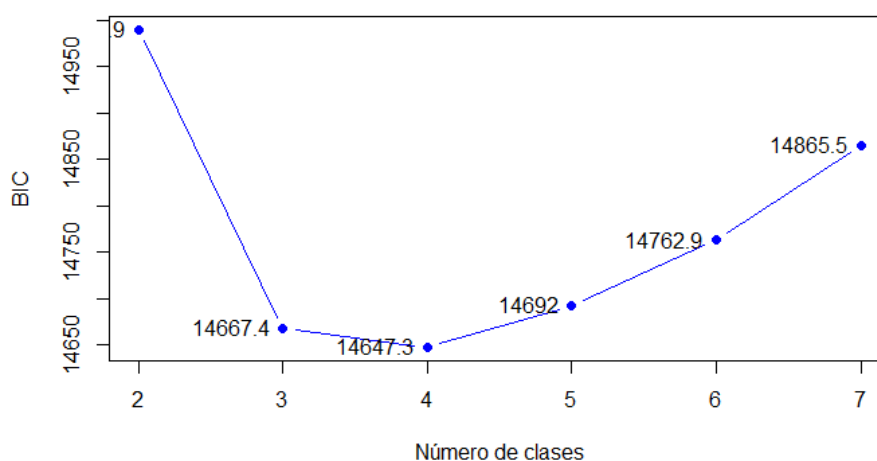
posibilita identificar cómo se estructura dicho constructo dentro de la población, revelando los patrones heterogéneos de confianza en los científicos y los perfiles latentes asociados. Con este propósito, estimamos modelos con un número de clases que oscila entre dos y siete, con el fin de determinar la solución óptima en términos estadísticos y sustantivos.

Comparando los valores del BIC de cada modelo, encontramos que el más bajo aparece con cuatro clases. Por esta razón, concluimos que este modelo logra el mejor equilibrio entre ajuste y parsimonia. A partir de cinco clases, el BIC vuelve a aumentar, lo que sugiere sobreajuste. En términos simples, el BIC evalúa qué tan bien se ajusta el modelo a los datos, pero penaliza la complejidad; es decir, castiga los modelos que introducen demasiados parámetros o clases innecesarias. Cuanto más bajo es el valor del BIC, mejor es el equilibrio entre ajuste y simplicidad (Figura 3).

Además del BIC, los indicadores de clasificación muestran un modelo de 4 clases bien definido y distribuido: la clase 1 concentra el 52.1% de los casos, la clase 2 el 18.1%, la clase 3 el 18.5% y la clase 4 el 11.4%. La probabilidad posterior máxima de pertenencia alcanza un valor global de 0,864, lo que indica que los individuos están asignados a su clase con un alto grado de certeza. A nivel de clase, esta medida también resulta consistente: clase 1 = 0,879; clase 2 = 0,806; clase 3 = 0,884; clase 4 = 0,856, lo que refleja una separación suficiente entre clases. Por otra parte, la entropía es de 0,749, lo que confirma una calidad de clasificación global alta.

**Figura 3**

*Valores de BIC según el número de clases*



*Nota.* Fuente: Elaboración propia

Una vez definidas las cuatro clases, procedemos a interpretar su significado a partir de los patrones de respuesta característicos de cada una (Figura 3). La interpretación de las clases no se basa únicamente en los porcentajes de pertenencia, sino en la configuración relativa de las probabilidades de respuesta en los distintos ítems, lo que permite

comprender cómo se combinan las dimensiones de la confianza en los científicos en cada perfil ciudadano:

Clase 1 - confiados (50.83 % de la muestra): poseen una confianza alta y generalizada. Las personas de esta clase muestran niveles altos de acuerdo (categorías 4 y 5) en todos los ítems. No obstante, confían más en los elementos técnicos (competentes, rigurosos, altruistas) que en los de apertura a la ciudadanía (transparentes, participación ciudadana).

Clase 2 – pragmáticos (18.07 % de la muestra): poseen una confianza moderada y selectiva, con bajo reconocimiento en las dimensiones más cercanas a la labor participativa con la ciudadanía. Este grupo muestra acuerdo alto en la competencia y, en menor medida, en el rigor de los científicos, pero ofrece respuestas más críticas en transparencia, altruismo y apertura al cambio. Se puede interpretar como un perfil de confianza pragmática, que valora la capacidad técnica, pero desconfía de las motivaciones o la apertura del sistema científico.

Clase 3 - entusiastas (18.49 % de la muestra): muestran una confianza muy alta, con acuerdos casi unánimes en todos los ítems, con proporciones superiores al 70–80 % de respuestas “totalmente de acuerdo”. Es el grupo de confianza entusiasta y plena, que no solo valora la competencia científica, sino también la transparencia y el altruismo de la comunidad científica.

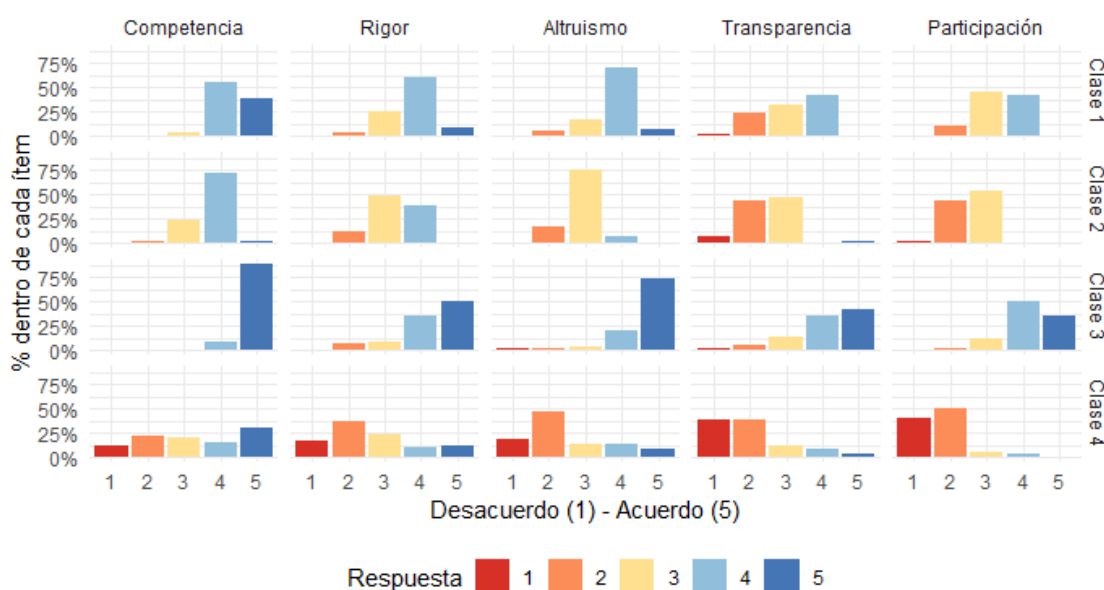
Clase 4 - escépticos (11.4 % de la muestra): muestran una desconfianza generalizada. Las mayores probabilidades de respuesta están en las categorías 1, 2 y 3, especialmente para los ítems relacionados con “interés común” y “tener en cuenta a la ciudadanía”. Este grupo percibe a los expertos como poco transparentes y receptivos, y cuestiona tanto su legitimidad como su cercanía con la sociedad. Se trata de una minoría crítica que expresa baja confianza institucional y epistemológica, situándose en el extremo escéptico del continuo.

En este punto, ya tenemos clasificadas a las observaciones de nuestra muestra en cuatro clases latentes, donde cada una de las cuales refleja distintos perfiles de confianza en los científicos. Una opción para profundizar el análisis consiste en examinar qué factores sociodemográficos o actitudinales se asocian con la pertenencia a cada clase. Por ejemplo, podemos analizar si variables como la ideología política o el nivel educativo influyen en la probabilidad de formar parte de un perfil u otro.

Para ilustrar esta posibilidad, ampliamos el modelo de clases latentes incorporando la variable de ideología política como covariable explicativa. En este tipo de modelos, la ideología no modifica directamente las respuestas a los *ítems*, sino que afecta la probabilidad de pertenecer a cada clase latente. De este modo, el modelo estima cómo cambia la composición de las clases a lo largo del eje ideológico; es decir, si las personas de izquierda, centro o derecha presentan distintas probabilidades de ubicarse en cada perfil de confianza.

**Figura 4**

*Proporción de cada respuesta dentro de cada ítem y clase.*



*Nota.* Fuente: Elaboración propia

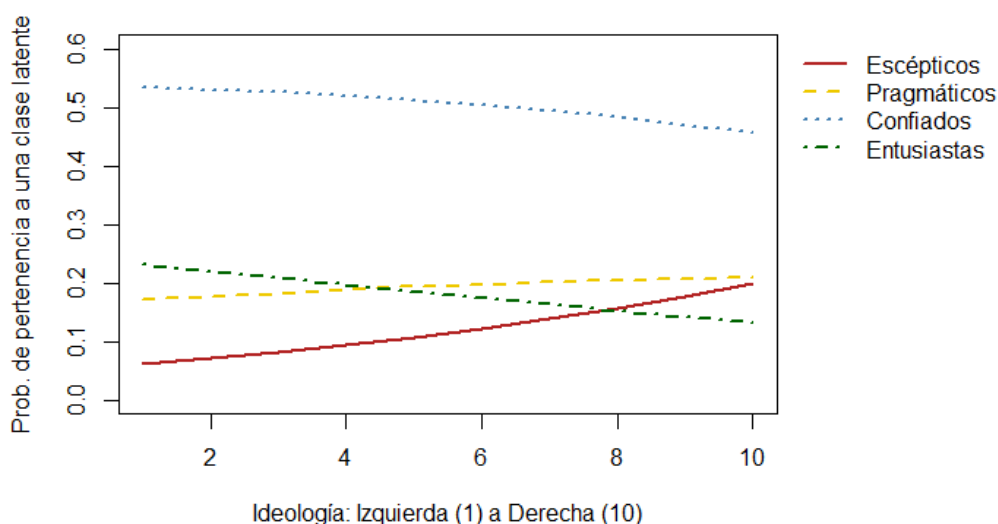
Con el objetivo de analizar de manera más clara cómo varía la pertenencia a cada clase latente en función de la ideología política, estimamos las probabilidades predichas de pertenencia a cada perfil a lo largo del eje ideológico, evaluado a partir de una escala entre 1 y 10. A partir de los coeficientes del modelo de clases latentes con la ideología política como covariable, calculamos para cada valor de ideología la probabilidad estimada de que un individuo pertenezca a una de las cuatro clases identificadas —que hemos identificado como “escépticos”, “pragmáticos”, “confiados” y “entusiastas”—. Lo que hace el modelo en este caso es transformar los coeficientes estadísticos en probabilidades comprensibles —que, en cada valor de ideología, suman siempre 1—, lo que nos permite ver cómo cambiaría la composición de los perfiles en personas hipotéticas situadas en distintos puntos del espectro ideológico.

Una forma clara de ver cómo cambia la composición de los perfiles de confianza en la ciencia conforme se avanza desde las posiciones más de izquierda (1) hasta las más de derecha (10) puede verse en la Figura 5. Se observa que los “escépticos” (línea roja) incrementan su probabilidad de pertenencia conforme aumenta la autoubicación ideológica hacia la derecha, lo que sugiere que las personas más conservadoras tienden a concentrarse en esta clase. En contraste, los “confiados” (línea azul) presentan la tendencia opuesta: su probabilidad disminuye progresivamente entre los individuos más de derecha, siendo la clase predominante entre los sectores ideológicamente más de izquierda. Por su parte, los “pragmáticos” (línea amarilla) y los “entusiastas” (línea verde) mantienen perfiles relativamente estables, aunque con leves variaciones: los primeros muestran una ligera caída hacia los extremos de derecha, mientras que los segundos tienden a incrementarse levemente conforme la ideología se desplaza hacia posiciones más conservadoras. En conjunto, estos resultados indican que la orientación ideológica

incide de forma diferenciada en la pertenencia a los distintos tipos de actitudes latentes identificados por el modelo.

**Figura 5**

*Ideología como predictor de clase latente*

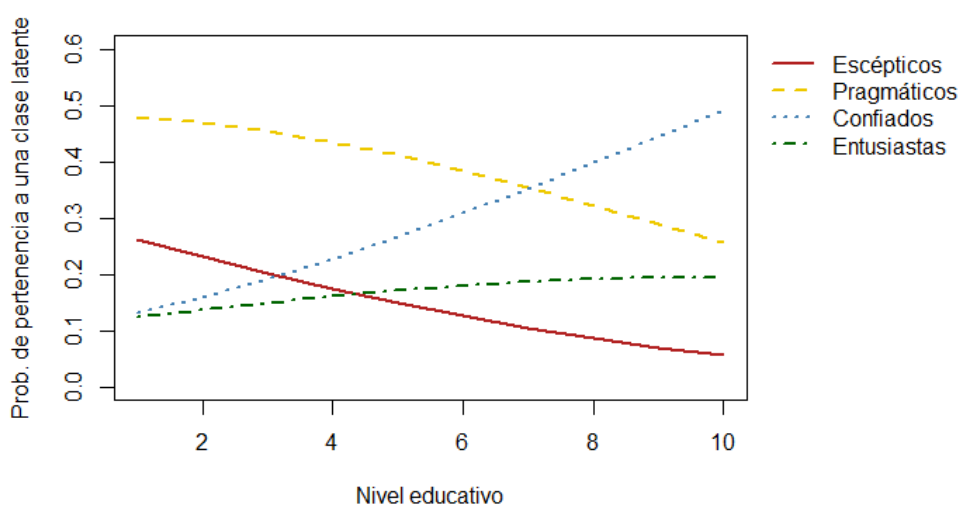


Nota. Fuente: Elaboración propia

Otra posible vía de análisis consiste en analizar cómo se comportan los grupos latentes cuando la covariable introducida en el modelo es el nivel educativo. Este enfoque permite observar si las probabilidades de pertenencia a cada clase varían según el grado de educación formal alcanzado, lo que ofrecería una perspectiva complementaria sobre la estructura latente identificada.

**Figura 6**

*Nivel de estudios como predictor de la clase latente*



Nota. Fuente: Elaboración propia

En la Figura 6 observamos cómo el nivel educativo actúa como covariable en la predicción de la pertenencia a las clases latentes. A medida que aumenta el nivel educativo, la probabilidad de pertenecer al grupo de los escépticos (línea roja) disminuye de forma constante, lo que sugiere que las personas con menor formación tienden a concentrarse en esta clase. En cambio, la probabilidad de integrar el grupo de los Confiados (línea azul) crece de manera sostenida conforme se incrementa el nivel educativo, indicando una asociación positiva entre educación y confianza en las instituciones o en los objetos de análisis considerados. Por su parte, los pragmáticos (línea amarilla) muestran una tendencia ligeramente descendente, mientras que los Entusiastas (línea verde) presentan un leve aumento a lo largo del eje educativo. En conjunto, estos resultados sugieren que el nivel educativo introduce diferencias significativas en la composición de los grupos latentes, reforzando la idea de que la educación constituye un factor relevante en la formación de actitudes y percepciones diferenciadas dentro de la población analizada.

#### 4. Conclusiones

El ejercicio metodológico desarrollado a lo largo de este trabajo permite extraer varias conclusiones de orden tanto técnico como epistemológico. En primer lugar, el LCA demuestra ser una herramienta especialmente valiosa para la investigación en comunicación cuando los fenómenos de interés se expresan mediante constructos abstractos y dimensiones no observables de forma directa. Frente a los métodos de *clustering* tradicionales, el LCA incorpora una lógica probabilística que reconoce la ambigüedad y la variabilidad inherentes al comportamiento humano. En lugar de imponer categorías rígidas, el modelo permite que las clases emerjan de los propios datos, estimando para cada individuo una probabilidad de pertenencia y reflejando así la incertidumbre natural que acompaña las actitudes sociales.

Desde una perspectiva más amplia, el empleo de modelos de clases latentes invita a reconsiderar el modo en que la investigación en comunicación aborda la diversidad social. Al construir las clases sobre la base de variables latentes previamente definidas y validadas, el modelo no solo segmenta poblaciones, sino que también traduce dimensiones conceptuales en configuraciones empíricas interpretables. En este sentido, el LCA no reemplaza la reflexión teórica, sino que obliga al investigador a pensar con precisión qué constructos subyacen a los indicadores observados y viceversa, cómo elegir indicadores capaces de capturar elementos no observables.

Más allá del caso empírico presentado, el LCA tiene un amplio potencial de aplicación en distintos ámbitos de la investigación en comunicación. En el estudio de audiencias, puede emplearse para identificar perfiles de consumo mediático, patrones de exposición informativa selectiva o tipologías de usuarios digitales (Metag y Schäfer, 2018; Napoli, 2011). En el análisis de encuestas sobre actitudes hacia los medios, la desinformación o la confianza institucional, el método permite capturar heterogeneidad latente que los análisis de regresión convencionales no detectan (Weller et al., 2020). Asimismo, en investigaciones sobre comunicación política, salud o ciencia, el LCA ha demostrado ser útil para construir tipologías de públicos con distintos niveles de implicación, escepticismo o receptividad (Nylund-Gibson y Choi, 2018). En todos estos contextos, la

técnica ofrece una alternativa rigurosa y teóricamente fundamentada a los métodos de *clustering* tradicionales, especialmente cuando los datos provienen de escalas Likert o variables categóricas, y cuando el investigador sospecha que la población estudiada no es homogénea.

## 5. Referencias bibliográficas

- Bail, C. (2021). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press.
- Bailey, K. D. (2003). *Typologies and taxonomies: An introduction to classification techniques* (Nachdr.). Sage Publ.
- Bartholomew, D. J., Steele, F., Moustaki, I., y Galbraith, J. I. (Eds.). (2008). *Analysis of multivariate social science data* (2. ed). CRC Press.
- Beer, D. (2017). The social power of algorithms. *Information, Communication y Society*, 20(1), 1-13. <https://doi.org/10.1080/1369118X.2016.1216147>
- Brusco, M. J., Shireman, E., y Steinley, D. (2017). A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. *Psychological Methods*, 22(3), 563-580. <https://doi.org/10.1037/met0000095>
- Budak, C., Goel, S., y Rao, J. M. (2016). Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly*, 80(S1), 250-271. <https://doi.org/10.1093/poq/nfw007>
- Collins, L. M., y Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social behavioral, and health sciences*. Wiley.
- Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4), 51-58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Flaxman, S., Goel, S., y Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80(S1), 298-320. <https://doi.org/10.1093/poq/nfw006>
- Forsyth, D. R. (2019). *Group dynamics* (Seventh edition). Cengage.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215-231. <https://doi.org/10.1093/biomet/61.2.215>
- Hartigan, J. A. (2015). Statistical Clustering. En *International Encyclopedia of the Social y Behavioral Sciences* (pp. 392-396). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.42176-8>
- Hodgson, G. M. (2019). Taxonomic definitions in social science, with firms, markets and institutions as case studies. *Journal of Institutional Economics*, 15(2), 207-233. <https://doi.org/10.1017/S1744137418000334>
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: With applications in R*. Springer.
- Johansson, S., Johansson, B., y Johansson, J. (2024). The Dynamics of Information-Seeking Repertoires: A Cross-Sectional Latent Class Analysis of Information-Seeking During the COVID-19 Pandemic. *Mass Communication and Society*, 27(4), 599-626. <https://doi.org/10.1080/15205436.2023.2258863>

- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (Fourth edition). The Guilford Press.
- Lazarsfeld, P. (1950). The logical and mathematical foundation of latent structure analysis. En P. Lazarsfeld, S. A. Stouffer, E. Guttman, P. Suchman, S. A. Star, y J. A. Clausen (Eds.), *Studies in social psychology World War II: Measurement and prediction* (Vol. 4, pp. 361-412). Princeton University Press.
- Lazarsfeld, P., y Henry, N. (1968). *Latent Structure Analysis*. Houghton Mifflin.
- Linzer, D., y Lewis, J. (2011). poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42(10), 1-29.
- Magidson, J., y Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20, 36-43.
- McCutcheon, A. (1987). *Latent Class Analysis*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412984713>
- Metag, J., y Schäfer, M. S. (2018). Audience Segments in Environmental and Science Communication: Recent Findings and Future Perspectives. *Environmental Communication*, 12(8), 995-1004. <https://doi.org/10.1080/17524032.2018.1521542>
- Napoli, P. M. (2011). *Audience evolution: New technologies and the transformation of media audiences*. Columbia University Press.
- Nylund, K. L., Asparouhov, T., y Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- Nylund-Gibson, K., y Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440-461. <https://doi.org/10.1037/tps0000176>
- Oberski, D. (2016). Mixture Models: Latent Profile and Latent Class Analysis. En J. Robertson y M. Kaptein (Eds.), *Modern Statistical Methods for HCI* (pp. 275-287). Springer International Publishing. [https://doi.org/10.1007/978-3-319-26633-6\\_12](https://doi.org/10.1007/978-3-319-26633-6_12)
- Philipps, G., Kelm, O., Friess, D., y Ziegele, M. (2025). From Unpolitical Sporadics to Political Activists: Identifying political social media user types. *European Journal of Communication*, 40(2), 188-205. <https://doi.org/10.1177/02673231251327860>
- Piltch-Loeb, R., Silver, D., Kim, Y., y Abramson, D. (2024). COVID-19 Vaccine Information Seeking Patterns and Vaccine Hesitancy: A Latent Class Analysis to Inform Practice. *Journal of Public Health Management y Practice*, 30(2), 183-194. <https://doi.org/10.1097/PHH.0000000000001834>
- Sokal, R. R. (1974). Classification: Purposes, Principles, Progress, Prospects: Clustering and other new techniques have changed classificatory principles and practice in many sciences. *Science*, 185(4157), 1115-1123. <https://doi.org/10.1126/science.185.4157.1115>
- Steinley, Douglas. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1-34. <https://doi.org/10.1348/000711005X48266>

- Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, 13(2), 65-93. <https://doi.org/10.1177/053901847401300204>
- Turner, J. C. (1987). *Rediscovering the social group: A self-categorization theory* (1. publ. in paperback). Blackwell.
- Visser, M., y Depaoli, S. (2022). A Guide to Detecting and Modeling Local Dependence in Latent Class Analysis Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(6), 971-982. <https://doi.org/10.1080/10705511.2022.2033622>
- Weller, B. E., Bowen, N. K., y Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 46(4), 287-311. <https://doi.org/10.1177/0095798420930932>

**Financiación:** Ministerio de Universidades [FPU/2020].

**Traducción al inglés:** realizada por los propios autores.

**Conflicto de intereses:** los autores declaran que no existen.

#### HOW TO CITE (APA 7<sup>a</sup>)

Núñez Sánchez, F., & Elías, C. (2026). Ventajas metodológicas del uso del análisis de clases latentes (LCA) para investigar en comunicación. Aplicación empírica al estudio de la confianza en la ciencia. *Comunicación & Métodos*, 8(1), 6-23. <https://doi.org/10.35951/v8i1.266>