

Análisis de Clases Latentes con JAMOVÍ

(Latent Class Analysis, LCA)

F. Javier Herrero Díez¹, Iván Diego Rodríguez² y Ana B. Bernardo Gutiérrez³

herrero@uniovi.es; ivan.diego@valnalon.com; bernardoana@uniovi.es

Este archivo se publica bajo Licencia Creative Commons CC BY-NC 4.0:
<https://creativecommons.org/licenses/by-nc/4.0/>



¹ Área de Metodología de las Ciencias del Comportamiento de la Universidad de Oviedo, Asturias, España

² Departamento de Formación. Ciudad Industrial del Valle del Nalón, SAU

³ Área de Psicología evolutiva de la Universidad de Oviedo, Asturias, España

Índice

1. INTRODUCCIÓN.....	3
2. ANÁLISIS DE CLASES LATENTES EN JAMOVÍ.....	5
2.1 Iniciar el Análisis de Clases Latentes en módulo snowRMM en JAMOVÍ.....	6
2.2 Seleccionar variables.....	6
2.3. Determinar el número de clases del modelo y otros parámetros.....	7
2.4. Entender las medidas de ajuste del modelo.....	8
2.5. Interpretar el modelo.....	9
2.6. Comparar modelos.....	10
2.7. Comparar variable a variable los patrones de respuesta de cada clase.....	11
2.8. Comprobar el tamaño de cada clase latente.....	16
2.9. Interpretar gráficos.....	17
2.10 Describir y etiquetar las clases.....	20
2.11 Asignar cada caso de la muestra a una clase.....	21
3. REDACCIÓN DE LOS RESULTADOS DEL ANÁLISIS.....	22
4. REFERENCIAS.....	23

Índice de tablas

Tabla 1: Comparación de modelos.....	10
Tabla 2: Tamaño de clases latentes.....	16

Índice de figuras

Fig. 1: Esquema general de selección Lu (2024).....	3
Fig. 2: Inicio de análisis de clases latentes en módulo snowRMM.....	6
Fig. 3: Seleccionar variables.....	6
Fig. 4: Determinar las condiciones del modelo.....	7
Fig. 5: Parámetros de ajuste del modelo.....	8
Fig. 6: Gráfico de codo [elbow plot].....	17
Fig. 7: LCA Plot.....	18
Fig. 8: Item by Class.....	19
Fig. 9: Profile plot.....	19
Fig. 10: Activación de la casilla Person class.....	21

1. INTRODUCCIÓN

El Análisis de Clases Latentes (ACL) es una metodología ampliamente utilizada en las ciencias del comportamiento para identificar estructuras latentes subyacentes en datos observacionales. Perteneciente a la familia de los modelos de variables latentes, el ACL permite clasificar individuos o casos en grupos no observables directamente, denominados clases latentes. Estas clases representan subpoblaciones homogéneas en términos de un conjunto de respuestas observadas (Weller et al., 2020).

A la hora de elegir este modelo, entre todos los posibles modelos que se pueden elegir para grandes bases de datos, y teniendo en cuenta que este enfoque asume que las variables observadas son categorías discretas observadas, podemos tener como referencia el diagrama de flujo siguiente:

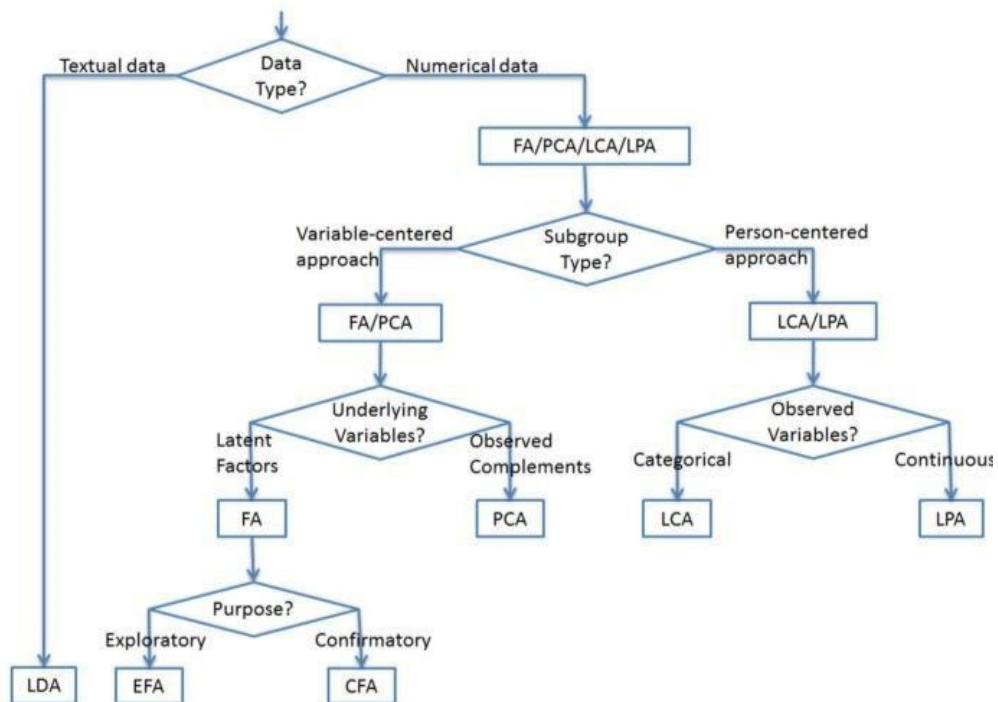


Fig. 1: Esquema general de selección Lu (2024)

El Análisis de Clases Latentes (ACL) es una técnica estadística diseñada para identificar subgrupos latentes mutuamente excluyentes dentro de una población, basándose en los patrones de respuesta observados en un conjunto de variables categóricas. El objetivo principal del ACL es desentrañar la estructura latente de los datos y describir la heterogeneidad existente en la población estudiada, permitiendo clasificar a los individuos en función de su probabilidad de pertenecer a cada clase latente.

El ACL se ha aplicado ampliamente en estudios psicológicos, educativos y de salud, entre otros, para explorar heterogeneidad en poblaciones. Ejemplos específicos que podemos citar incluyen:

- Clasificación de trastornos psicológicos: Identificación de subtipos de trastornos de ansiedad o depresión basados en patrones sintomáticos.
- Segmentación educativa: Determinación de perfiles de estudiantes según estilos de aprendizaje o niveles de compromiso.
- Estudios de salud pública: Clasificación de individuos según patrones de comportamiento de riesgo, como consumo de sustancias.

Es ideal para usar con escalas Likert, siempre que asumamos que las variables son de naturaleza ordinal y no continua.

Como hemos dicho el modelo, de naturaleza probabilística, postula la existencia de una variable subyacente no observada que divide la población en clases latentes mutuamente exclusivas y exhaustivas.

A nivel más formal supongamos, $\mathbf{y}_i=(y_{i1}, y_{i2}, \dots, y_{ij})$ como un vector de respuestas a un conjunto de J elementos categóricos. Bajo el supuesto de independencia local, la probabilidad de observar el patrón de respuesta para el i -ésimo individuo puede expresarse (Vermunt & Magidson, 2002) como:

$$p(\mathbf{y}_i | \boldsymbol{\pi}, z_i) = \sum_{k=1}^K \pi_k \prod_{j=1}^J p(y_{ij} | z_i = k)$$

...donde:

- $\boldsymbol{\pi}$ es un vector de *probabilidades de clase* o *pesos de mezcla* con la suma de 1, y π_k es la probabilidad de que el i -ésimo individuo pertenezca a la k -ésima clase ($k = 1, \dots, K$)

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_K)$$

- K es el número de clases latentes
- z_i es la pertenencia a la clase latente del individuo i , que se supone que sigue una distribución multinomial, y $p(y_{ij} | z_i = k)$ es la función de densidad de probabilidad condicional (PDF) de la respuesta y_{ij} en la clase k .

Los estadísticos de ajuste clásicos del modelo, aplicados a determinar el número de clases latentes, son:

- BIC (Criterio de información bayesiano)
- AIC (Criterio de información de Akaike)

Los valores más bajos de ambos estadísticos indican un mejor ajuste del modelo, teniendo en cuenta que BIC funciona mejor en muestras grandes, y que en caso de tener menos de 300 casos se deberían usar los dos de forma conjunta.

El algoritmo general de trabajo consiste en comenzar con el modelo de dos clases y se van añadiendo clases hasta que el modelo no satisfaga nuestros criterios de convergencia, definidos *a priori*.

Por último, señalar las similitudes metodológicas (aunque no son intercambiables) entre el LCA y el análisis factorial, ya que son técnicas útiles para la reducción de la complejidad de los datos y se refieren a constructos latentes (no observados) que se infieren a partir de los datos. Proporcionan enfoques complementarios para el análisis de datos complejos, compartiendo una base metodológica común en la modelización de constructos latentes y la reducción de dimensionalidad, aunque su aplicación depende del tipo de constructo latente que se desea modelar y del propósito analítico subyacente (explorar patrones estructurales frente a clasificar individuos).

2. ANÁLISIS DE CLASES LATENTES EN JAMOMI

En este ejemplo vamos a utilizar el software **JAMOMI** (versión 2.6.17), y la librería *snowRMM* (también es posible usar la librería *snowLatent* con el mismo propósito) para realizar un análisis de clases latentes a partir de cinco variables tipo Likert sobre destinos potenciales en caso de abandono

- Dejar de estudiar
- Incorporarse al mundo laboral
- Cambiar de grado
- Cambiar de universidad
- Cursar estudios de Formación Profesional

El instrumento se administró a una muestra de 2120 estudiantes de la Universidad de Oviedo a principios del curso 2024-2025 que debía especificar el grado de acuerdo o desacuerdo con cada uno de estos destinos de manera independiente en función de una escala tipo Likert de 4 niveles:

1. Muy en desacuerdo
2. En desacuerdo
3. De acuerdo
4. Muy de acuerdo

Se decide aplicar el Análisis de Clases Latentes con el objetivo de identificar subgrupos latentes mutuamente excluyentes dentro de una población, basándose en los patrones de respuesta observados para este conjunto de cinco variables categóricas.

2.1 Iniciar el Análisis de Clases Latentes en módulo snowRMM en JAMOVI

Para realizar el Análisis de Clases Latentes, nos situamos en el apartado Análisis y hacemos click en el módulo snowRMM. Si no aparece, deberemos instalarlo desde la biblioteca JAMOVI. Al hacer click sobre el módulo (círculo rojo) se abre un menú desplegable y seleccionamos la opción *Latent Class Analysis* [Análisis de Clases Latentes] (rectángulo rojo)

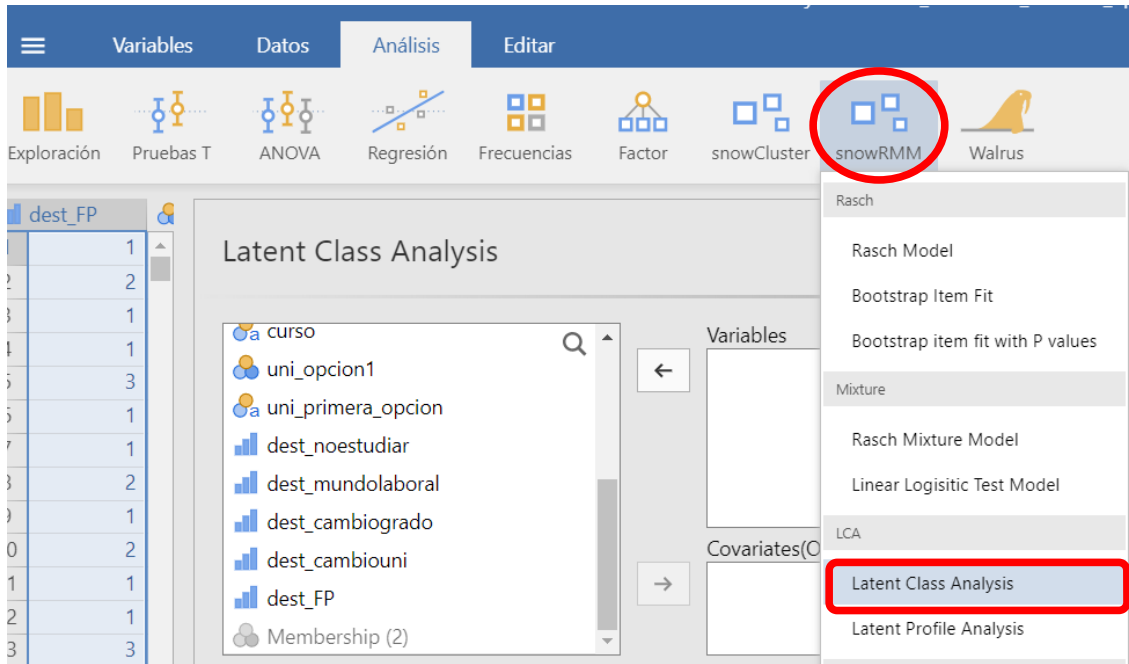


Fig. 2: Inicio de análisis de clases latentes en módulo snowRMM

2.2 Seleccionar variables

El siguiente paso será seleccionar las variables en la caja de la izquierda y arrastrarlas al apartado Variables situado a la derecha (rectángulo rojo).

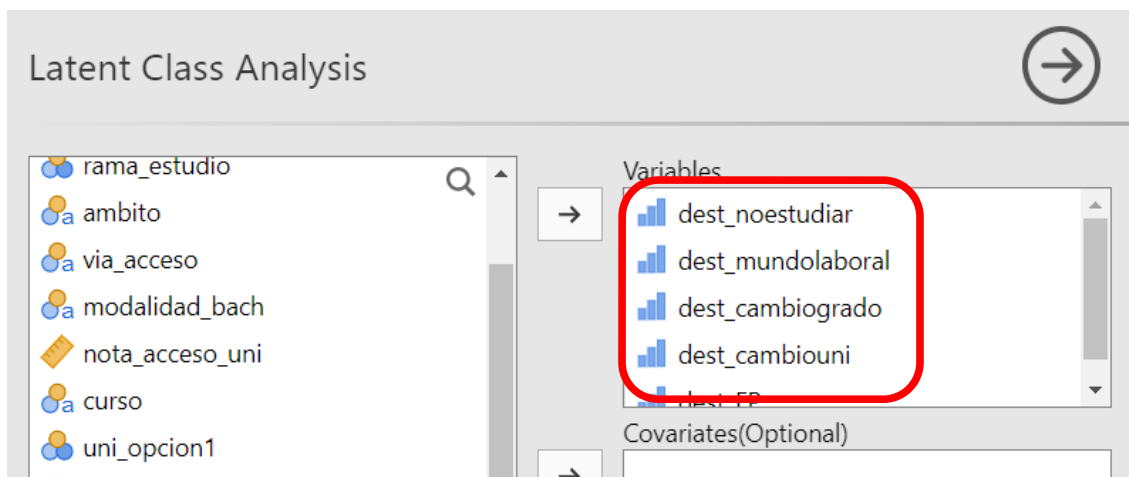


Fig. 3: Seleccionar variables

2.3. Determinar el número de clases del modelo y otros parámetros

Ahora abrimos el desplegable Analysis (rectángulo rojo) donde indicaremos las condiciones para el modelo. En nuestro ejemplo hemos previsto que el número de clases latentes será 4 y lo indicamos en la caja *Class* (rectángulo rojo) y activamos las opciones de menú que se indican en la figura.

The screenshot displays the 'Latent Class Analysis' software interface. At the top, the title 'Latent Class Analysis' is visible. Below the title, there are two columns of variables. The left column contains variables: ambito, via_acceso, modalidad_bach, nota_acceso_uni, curso, uni_opcion1, uni_primera_opcion, and Membership (2). The right column contains variables: dest_noestudiar, dest_mundolaboral, dest_cambiogrado, dest_cambiouni, and dest_EP. Below these columns is a 'Covariates(Optional)' section. A red box highlights the 'Analysis' dropdown menu. Below the 'Analysis' menu, there are three sections: 'Fit', 'Probability', and 'Membership'. The 'Fit' section includes a 'Class' dropdown menu set to '4', and checkboxes for 'Fit measures', 'Model comparison', and 'Elbow plot'. The 'Probability' section includes checkboxes for 'Class size', 'Item response', and 'Predicted cell counts'. The 'Membership' section includes a checkbox for 'Person class'. Below these sections is a 'Plots' section with checkboxes for 'LCA plot', 'Item by class', and 'Profile plot'.

Fig. 4: Determinar las condiciones del modelo

2.4. Entender las medidas de ajuste del modelo

Ajuste del modelo

El primer resultado del análisis es la tabla *Model fit* que muestra una serie de parámetros relacionados con el ajuste del modelo

Class	Log-likelihood	Resid.df	AIC	AIC3	BIC	SABIC	CAIC	Entropy	G ²	G ² p	χ ²	χ ² p
Model fit												
4	-9797.595	960.000	19721.190	19784.190	20077.718	19877.560	20140.718	0.736	978.215	0.334	1718.785	<.001

Nota. G²=Likelihood ratio statistic; χ²=Pearson Chi-square goodness of fit statistic; Entropy=entropy R² statistic (Vermunt & Magidson, 2013, p. 71)

Fig. 5: Parámetros de ajuste del modelo

A continuación, vamos a desglosar el significado de las columnas de la tabla anterior:

Class

Indica el número de clases latentes especificadas en el modelo. En este caso, se está analizando un modelo con 4 clases.

Log-likelihood

Logaritmo de la verosimilitud del modelo ajustado. Un valor más alto (menos negativo) indica mejor ajuste, siempre que el modelo sea comparable con otros. En este caso, el valor es -9798, lo que sirve para calcular criterios de información como el AIC, BIC, etc.

Resid.df

Grados de libertad residuales. Refleja el número de datos observados menos los parámetros estimados. Aquí, hay 960 grados de libertad residuales.

AIC

Akaike Information Criterion. Mide el ajuste del modelo penalizando la complejidad (número de parámetros). Modelos con un AIC más bajo son preferidos. El AIC en este caso es 19721.

AIC3

Modificación del AIC. Similar al AIC, pero añade una penalización mayor por complejidad. Aquí, el valor es 19784.

BIC

Bayesian Information Criterion. Similar al AIC, pero penaliza más agresivamente la complejidad. Modelos con BIC más bajo son preferidos, especialmente cuando el tamaño muestral es grande. Aquí, el BIC es 20078.

SABIC

Sample-size Adjusted BIC. Variante del BIC que ajusta la penalización según el tamaño muestral. Es útil para estudios con muestras pequeñas o medianas. Aquí, el SABIC es 19878.

CAIC

Consistent Akaike Information Criterion. Otra variante del AIC que ajusta los grados de penalización para evitar sobreajuste. Aquí, el CAIC es 20141.

Entropy

Entropía. Mide el grado de separación entre las clases latentes. Cuanto más cercana a 1, mejor capacidad de clasificación del modelo, es decir, una buena separación y una alta precisión en la asignación de los individuos a sus respectivas clases. En este caso, la entropía es 0.736, lo cual indica un buen grado de separación entre clases. Aunque existen autores que citan el umbral $>.80$ (Celeux & Soromenho, 1996), también se puede usar criterios más liberales ($>.60$; Muthén, 2008).

G^2

Likelihood Ratio Statistic. Evalúa la discrepancia entre los datos observados y los predichos por el modelo. Un valor bajo y una significación mayor que 0.05 ($p > 0.05$) sugiere que el modelo ajusta bien los datos. Con un valor G^2 de 978 y una significación ($p = 0.334$), el modelo no tiene una discrepancia significativa respecto a los datos observados.

χ^2

Chi-square goodness-of-fit statistic. Similar al G^2 , pero menos robusto ante tamaños de muestra grandes. Aquí el valor χ^2 es 1719, y la significación es $p < .001$, lo que sugiere una discrepancia significativa entre el modelo y los datos.

2.5. Interpretar el modelo

- **Entropía:** Un valor de **0.736** indica que las clases latentes están relativamente bien diferenciadas. Se considera aceptable cuando supera 0.80.
- **G^2 y χ^2 :** Aunque el G^2 sugiere un buen ajuste ($p > 0.05$), el χ^2 muestra una discrepancia significativa ($p < 0.001$). Esto podría deberse a un tamaño muestral grande ($n=2120$), ya que el χ^2 es sensible a ello.
- **Criterios de ajuste (AIC, BIC, etc.):** Puesto que los valores absolutos no son interpretables por sí mismos, estos criterios se deben comparar con otros modelos con distinto número de clases. El modelo con los valores más bajos será el mejor. En este caso, se debe comparar con modelos con 3 o 5 clases para determinar si el de 4 clases es la mejor solución.

2.6. Comparar modelos

Para confirmar si el modelo con 4 clases es óptimo, es preciso comparar estos resultados con los modelos con un menor o mayor número de clases.

JAMOVI genera una segunda tabla más detallada para la comparación de modelos con distinto número de clases posibles:

Model comparison								
Class	AIC	AIC3	BIC	SABIC	CAIC	Log-likelihood	χ^2	G ²
1	21780	21795	21865	21817	21880	-10875	72038	3133
2	20182	20213	20357	20259	20388	-10060	9825	1503
3	20021	20068	20287	20137	20334	-9963	9994	1310
4	19721	19784	20078	19878	20141	-9798	1719	978

Tabla 1: Comparación de modelos

Según los valores, el modelo con 4 clases es el que presenta mejor ajuste, ya que tiene los valores más bajos en AIC, BIC, SABIC, CAIC, χ^2 y G², y el log-likelihood más alto. No obstante, a la hora de seleccionar el número de clases del modelo es igualmente importante considerar su interpretabilidad en función de las teorías sobre el fenómeno de estudio.

En el caso que nos ocupa, la teoría identifica 3 o 4 trayectorias posibles tras el abandono de estudios de grado. Tinto (1994) distingue “salida del programa de estudios”, “salida institucional” y “salida del sistema” que supondría el abandono del sistema universitario. Por su parte, Kalogrides y Grodsky (2011) plantean cuatro posibles trayectorias: permanencia en el grado y universidad elegida hasta la graduación, cambio de universidad, paso a Formación Profesional, o abandono completo de los estudios. Tieben (2024) por su parte distingue tres trayectorias en el caso de abandono del sistema universitario: el paso a la Formación Profesional, la incorporación al mercado de trabajo y el retorno a estudios de Educación Superior más adelante.

2.7. Comparar variable a variable los patrones de respuesta de cada clase

Destino en caso de abandono: Dejar de estudiar

La distribución de los integrantes de cada clase en función de su grado de acuerdo con la opción de dejar estudios completamente en caso de abandono del grado es la siguiente:

	Pr(1)	Pr(2)	Pr(3)	Pr(4)
class 1:	0.945	0.044	0.000	0.011
class 2:	0.950	0.043	0.002	0.005
class 3:	0.231	0.328	0.336	0.104
class 4:	0.329	0.657	0.015	0.000

- Clase 1 y 2: La mayor parte de los integrantes de estas dos clases muestran un alto grado de desacuerdo (Pr(1)) con la posibilidad de dejar de estudiar en caso de abandono del grado (94.5% y 95% respectivamente).
- Clase 3: Presenta el mayor porcentaje de personas que valoran la opción de dejar de estudiar en caso de abandono. Un 33,6 % se muestra de acuerdo (Pr(3)) y un 10,4% muy de acuerdo (Pr(4)) con esta afirmación.
- Clase 4. Al igual que las clases 1 y 2, dejar de estudiar en caso de abandono tampoco es una opción pero el grado de desacuerdo es menor. Sólo un 32,9% se muestra muy en desacuerdo (Pr(1)) frente al 65,7% que está en desacuerdo (Pr(2)).

Destino en caso de abandono: Incorporarse al mundo laboral

La distribución de los integrantes de cada clase en función de su grado de acuerdo con incorporación al mercado laboral en caso de abandono del grado es la siguiente:

	Pr(1)	Pr(2)	Pr(3)	Pr(4)
class 1:	0.760	0.240	0.000	0.000
class 2:	0.878	0.110	0.012	
class 3:	0.032	0.159	0.591	0.217
class 4:	0.078	0.790	0.132	0.000

- Clases 1 y 2: No valoran en absoluto la posibilidad de incorporarse al mercado laboral en caso de abandono. Un 87,8% de los estudiantes asignados a la clase 2 se muestran muy en desacuerdo con esta afirmación mientras que en la clase 1, este porcentaje se sitúa en el 76%.
- Clase 3: La mayoría de sus componentes muestran un alto grado de acuerdo con esta opción como destino en caso de abandono. Un 59,1% se muestra de acuerdo y otro 21,7% muy de acuerdo.
- Clase 4: Tampoco contemplan esta opción pero el grado de desacuerdo no es tan alto como en las clases 1 y 2. Un 79% expresa su desacuerdo pero sólo un 8% está muy en desacuerdo. Otra diferencia con respecto a las clases 1 y 2 es que hay un 13,2% valoraría esta opción.

Destino en caso de abandono: Cambio de grado

La distribución de los integrantes de cada clase en función de su grado de acuerdo con cambiar de grado en caso de abandono de los estudios que cursan actualmente es la siguiente:

Probability of dest_cambiogrado				
	Pr(1)	Pr(2)	Pr(3)	Pr(4)
class 1:	0.010	0.192	0.490	0.308
class 2:	0.735	0.208	0.057	0.000
class 3:	0.269	0.190	0.330	0.211
class 4:	0.208	0.513	0.242	0.037

- Clase 1: Es la clase que se muestra más propensa a cambiar de grado. Un elevado porcentaje de integrantes se muestra de acuerdo (49%) o muy de acuerdo (30,8%) con la posibilidad de cambiar de grado. Es decir, prácticamente 8 de cada 10 valorarían cambiar de grado en caso de abandono.
- Clase 2: En esta clase se observa un posicionamiento claramente en contra de la opción de cambiar de grado. Más del 94% de sus integrantes se muestra muy en desacuerdo (73,5%) o en desacuerdo (20,8%).
- Clase 3: Junto con la clase 2, aunque en menor medida, la mayoría de sus integrantes valora el cambio de grado en caso de abandono. Un 33 % se muestra de acuerdo y otro 21,1% muy de acuerdo.
- Clase 4: Son mayoría los que adoptan un posicionamiento claramente negativo con respecto a esta posibilidad (20,8% muy en desacuerdo, 51,3% en desacuerdo) pero en menor medida que la clase 2.

Destino en caso de abandono: Cambio de universidad

La distribución de los integrantes de cada clase en función de su grado de acuerdo con cambiar de universidad en caso de abandono de los estudios que cursan actualmente es la siguiente:

Probability of dest_cambiouni				
	Pr(1)	Pr(2)	Pr(3)	Pr(4)
class 1:	0.225	0.274	0.347	0.154
class 2:	0.643	0.176	0.143	0.038
class 3:	0.365	0.156	0.295	0.184
class 4:	0.210	0.491	0.281	0.018

- Clase 1: Al igual que con la variable cambio de estudios, es la clase que cuenta con el mayor porcentaje de integrantes que se muestran de acuerdo con esta posibilidad (34,7% de acuerdo y 15,4% muy de acuerdo) pero en menor medida que el cambio de grado. Es decir, esta clase tiende a anteponer el cambio de grado al cambio de universidad en caso de abandono.
- Clase 2: De nuevo, es la clase que se muestra más claramente en contra. Más del 81% de sus integrantes se manifiesta muy en desacuerdo (64,3%) o en desacuerdo (17,6%) con esta posibilidad.
- Clase 3: Son mayoría los que se muestran en desacuerdo (52%) pero a la vez esta clase presenta los mayores porcentajes de personas que se muestran muy de acuerdo (18,4%) o de acuerdo (29,5%) con esta afirmación.
- Clase 4: Son mayoría los que adoptan un posicionamiento claramente negativo con respecto a esta posibilidad (21% muy en desacuerdo, 49,1% en desacuerdo) pero en menor medida que la clase 1.

Destino en caso de abandono: Cursar estudios de Formación Profesional

La distribución de los integrantes de cada clase en función de su grado de acuerdo con matricularse en un ciclo formativo de FP en caso de abandono de los estudios que cursan actualmente es la siguiente:

Probability of dest_FP				
	Pr(1)	Pr(2)	Pr(3)	Pr(4)
class 1:	0.637	0.205	0.124	0.033
class 2:	0.936	0.045	0.015	0.005
class 3:	0.329	0.237	0.297	0.137
class 4:	0.250	0.637	0.113	0.000

- Clase 1: Son mayoría los que no valoran esta opción (63,7% muy en desacuerdo y 20,5% en desacuerdo).
- Clase 2: De nuevo es la clase que muestra un mayor grado de desacuerdo con esta vía de salida en caso de abandono (93,6%)
- Clase 3: . El 43,4 % se muestran muy de acuerdo (29,7%) o de acuerdo (13,7%) con cursar estudios de FP en caso de abandono. Aunque sigue siendo una opción minoritaria, es la clase más propensa a explorar esta opción.
- Clase 4: Al igual que con el resto de destinos, también se muestran abiertamente en desacuerdo con esta posibilidad (25% muy en desacuerdo, 63.7% en desacuerdo) pero en menor medida que la clase 1.

2.8. Comprobar el tamaño de cada clase latente

La siguiente tabla que nos muestra JAMOVİ refleja el porcentaje de casos de la muestra que el modelo asigna a cada clase.

Size of each latent class	
	Probability
1	0.161
2	0.517
3	0.129
4	0.193

Tabla 2: Tamaño de clases latentes

Cada número de la primera columna representa a una de las clases latentes, una categoría o grupo no observable directamente a partir de los datos, pero que se infiere a partir de las respuestas a las variables observadas. La columna Probabilidad indica la proporción de casos en la muestra asignados a cada clase latente por el modelo. Por ejemplo, el 16.1% de los casos pertenecen a la clase latente 1.

La mayoría de los casos quedan agrupados en la clase 2 (51,7%), seguida de la 4, la 1 y la 3. Si las clases tuviesen tamaños muy similares indicarían una mayor heterogeneidad en la población, pero si una clase es mucho más grande que las demás, como es este ejemplo, indica una mayor homogeneidad de los sujetos analizados.

2.9. Interpretar gráficos

Elbow plot (gráfico de codo)

Como complemento a los datos numéricos para comparar el ajuste del modelo (apartado 2.6) , podemos usar el gráfico de codo, "Elbow Plot", para ayudar a determinar el número óptimo de clases. El gráfico representa los valores de varios índices de ajuste del modelo (AIC, AIC3, BIC, SABIC, CAIC) para diferentes números de clases latentes. El objetivo es encontrar un punto en el gráfico donde la mejora del ajuste disminuya significativamente al aumentar el número de clases (similar a un "codo" que se dobla).

En este caso, parece haber un codo relativamente claro en 2 clases. Los índices AIC, BIC, SABIC y CAIC disminuyen considerablemente al pasar de 1 a 2 clases. Al pasar de 2 a 3 y de 3 a 4 clases, la disminución es mucho menor.

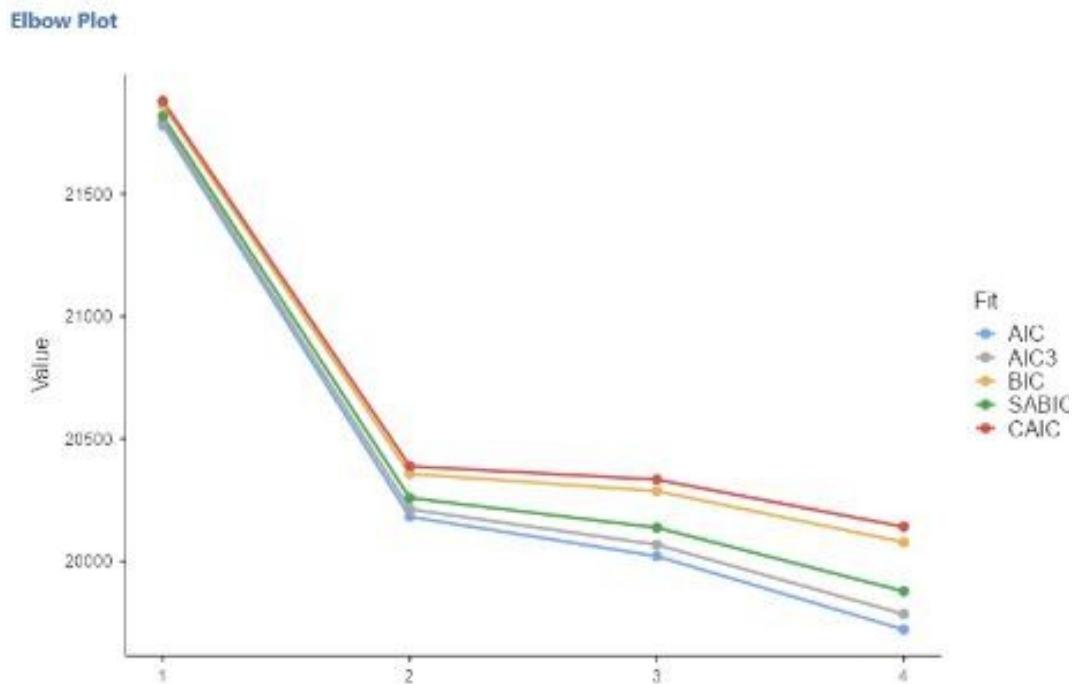


Fig. 6: Gráfico de codo [elbow plot]

Basándose en el gráfico anterior, un modelo con 2 clases latentes parece ser el más adecuado para los datos. Ofrece un buen equilibrio entre el ajuste del modelo y la complejidad. Aunque el ajuste mejora ligeramente con más clases, la mejora es marginal y probablemente no justifica la complejidad adicional. No obstante, debemos tener en cuenta, que a veces, un modelo con un número ligeramente mayor de clases puede ser más interpretable, incluso si el ajuste no mejora drásticamente. Por tanto, en este caso se han usado 4, por cuestiones teóricas, y como apoyo visual a la toma de decisiones se han comprobado los casos de 5 y 6 clases, donde el codo para 4 clases ofrece una inflexión clara:

LCA Plot

Este gráfico representa de manera separada para cada clase la probabilidad de cada respuesta para las cinco variables incorporadas en el modelo:

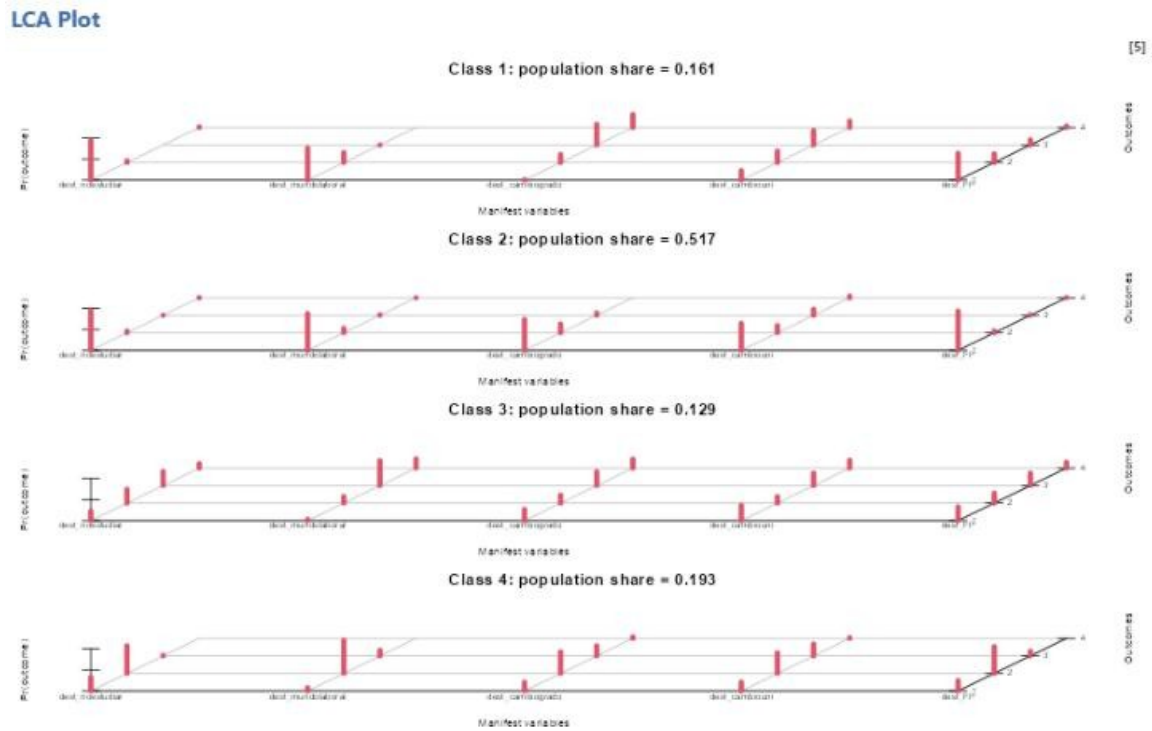


Fig.

7: LCA Plot

Item by Class

Este gráfico muestra intuitivamente la prevalencia de cada respuesta dentro de cada clase:

Item by class

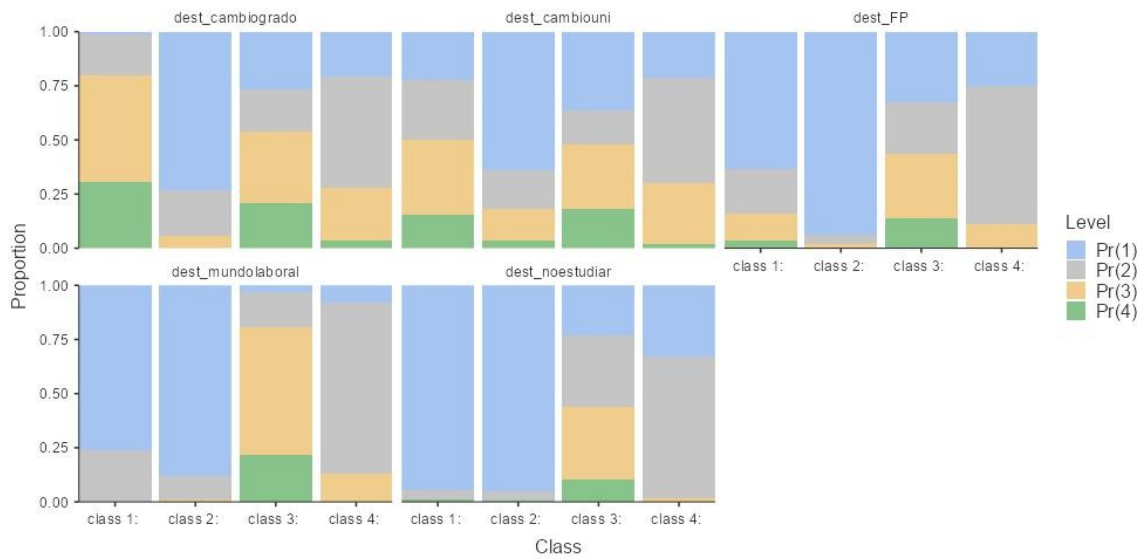


Fig. 8: Item by Class

Profile Plot

El gráfico muestra las probabilidades de respuesta para cada variable dentro de cada clase latente del LCA. A diferencia de los gráficos anteriores, este visualiza las diferencias entre las clases de una manera más directa, permitiendo comparar fácilmente el perfil de respuesta de cada clase.

Profile Plot

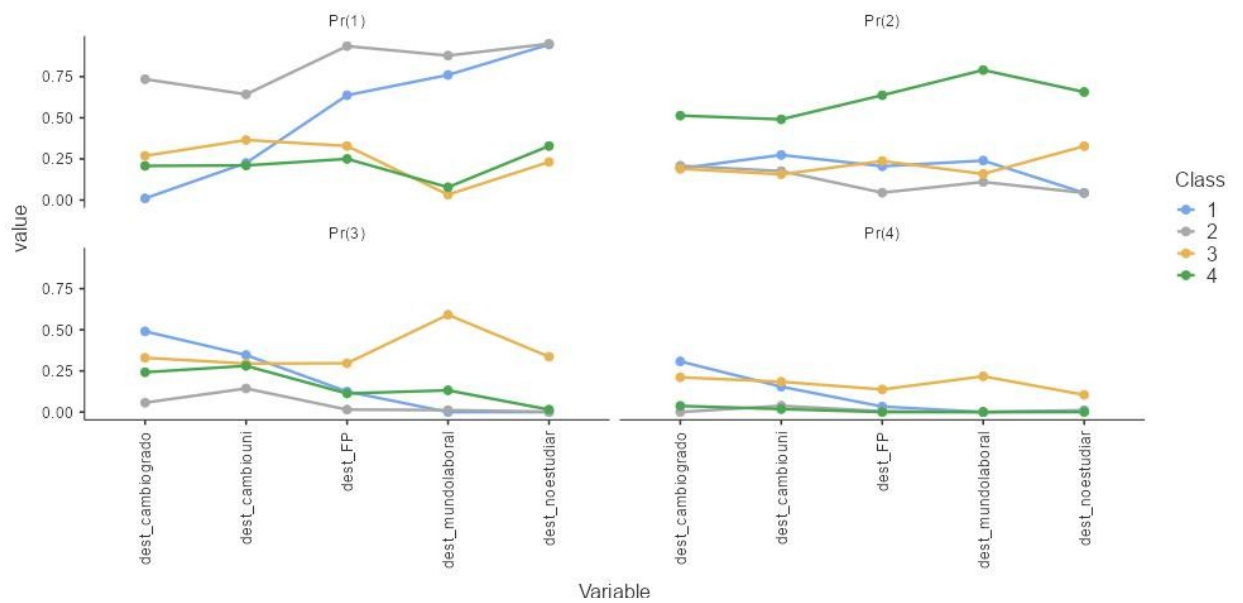


Fig. 9: Profile plot

Este gráfico permite identificar rápidamente qué variables contribuyen más a la diferenciación entre las clases, para ello se busca observar las mayores diferencias en la altura de las líneas entre las distintas clases para cada variable.

2.10 Describir y etiquetar las clases

Una vez analizados los patrones de respuesta para las cinco variables incorporadas al modelo elegido es necesario describir y etiquetar las clases del modelo elegido.

Clase 2 (51,7%) – permanencia en mismo grado : Estudiantes que en el momento de la encuesta no valoran ninguna otra posibilidad que no sea continuar estudiando el grado en el que estaban matriculados en el momento de la encuesta. Representa la mayoría de los casos (51,7%)

Clase 4 (19,3%) - permanencia en grado pero en menor medida: Tampoco valoran por el momento ninguna opción de cambio o de salida pero su grado de desacuerdo es algo menor que el de los integrantes de la clase 2.

Clase 1 (16,1%) – cambio de grado: Clase conformada por un conjunto de estudiantes más propensos a cambiar de grado o, en menor medida, a cambiar de universidad en caso de abandono. Descartan mayoritariamente cualquiera de las otras tres opciones que implican una salida del sistema universitario (dejar estudios, incorporarse al mercado laboral o cursar estudios de Formación Profesional)

Clase 3 (19,3%) - abandono del Sistema Universitario: Estudiantes con una mayor predisposición a salirse del sistema universitario en caso de abandono priorizando la incorporación al mercado de trabajo (80,8% de acuerdo o muy de acuerdo). En menor medida, contemplan la posibilidad de dejar de estudiar (44%) o cursar estudios de FP (43,4%).

2.11 Asignar cada caso de la muestra a una clase

Al activar la casilla *Person class*, JAMOVİ crea la variable *Membership* que determina a que clase pertenece cada caso de la muestra. Esto nos permite realizar análisis adicionales para ver diferencias entre clases y relación con otras variables (ej. comparación de medias, análisis de contingencia, análisis de regresión).

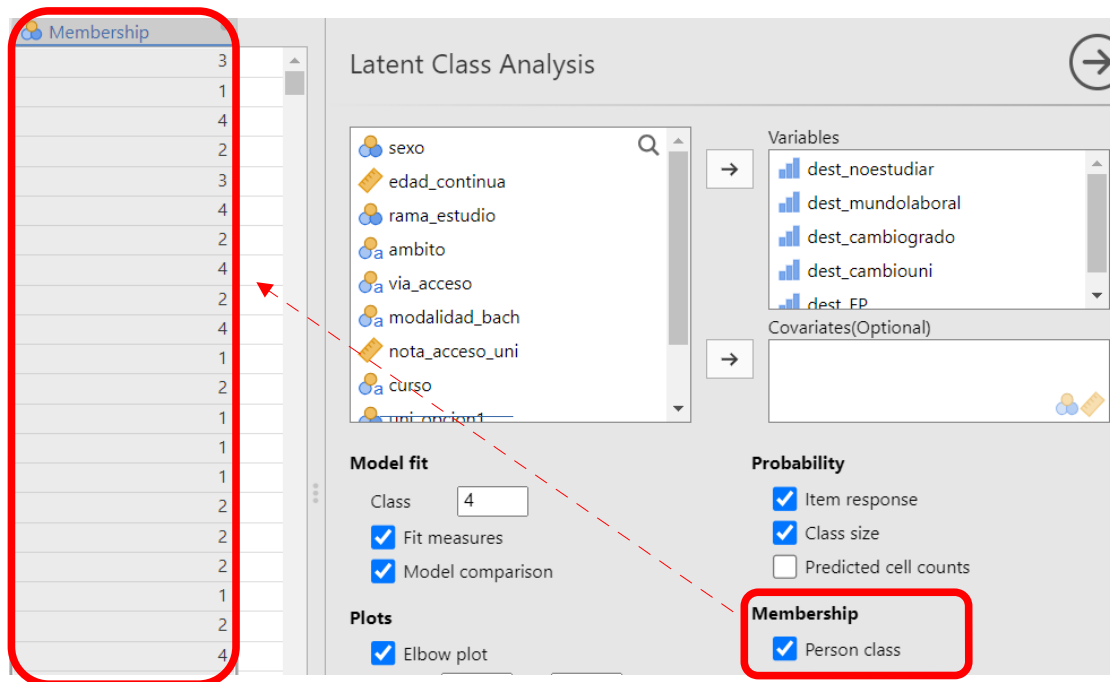


Fig. 10: Activación de la casilla *Person class*

3. REDACCIÓN DE LOS RESULTADOS DEL ANÁLISIS

Versión 1

“Se exploró una gama de modelos LCA, de dos a seis clases, para determinar el número de clases que mejor representaban los patrones de comportamiento en caso de abandono de estudios de grado. Se consideraron el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC) para determinar el número óptimo de clases conceptualmente significativas (Collins y Lanza, 2010). Además, se calcularon los valores G^2 del modelo estimado para seleccionar el modelo de mejor ajuste. Tomando como criterio que tanto los valores G^2 , AIC y BIC más bajos como indicadores del modelo que mejor ajusta (Chan y Shek, 2018).

La entropía se utilizó como criterio complementario para determinar el modelo final, tomándose como un criterio de calidad para la clasificación, con valores más altos entre 0 y 1 indicando un mejor ajuste (Chan y Shek, 2018). En este estudio el umbral que hemos usado es el liberal de $>.60$, propuesto por Muthén (2008).

Todos los análisis de LCA se realizaron utilizando el paquete snowRMM en el software JAMOVl, versión 2.6.17.”

Versión 2

“Para comparar los modelos se utilizaron varios índices de ajuste, incluido el criterio de información de Akaike (AIC), el criterio de información bayesiano criterio (BIC) y su alternativa ajustada al tamaño de la muestra (ABIC), donde los valores más bajos de AIC, BIC y ABIC muestran modelos que se ajustan mejor. Para determinar la calidad de la clasificación de pertenencia a una clase, se consideró el índice de entropía (Nylund et al., 2007) tomando como referencia un criterio estricto ($>.80$) (Clark, 2010).

Los análisis de LCA se realizaron manejando el paquete snowRMM del software JAMOVl, versión 2.6.17.”

Versión 3

“Para determinar el modelo de mejor ajuste, se evaluaron varios indicadores de ajuste, incluidos los criterios de información de Akaike (AIC) (Akaike, 1987), los criterios de información bayesianos (BIC) (Schwarz, 1978) y el BIC ajustado (SABIC). Los valores más pequeños en estos indicadores reflejan un mejor ajuste del modelo (Dan et al., 2018). La precisión de la clasificación se evaluó utilizando la entropía, que varía de 0 a 1. Un valor de entropía más alto significa una clasificación más precisa, con una criterio de entropía mayor a $.8$ (Zhang et al., 2017).

En el presente estudio, se empleó el software JAMOVl, versión 2.6.17, junto con la librería snowRMM, para llevar a cabo un análisis de clases latentes (Latent Class Analysis, LCA). El objetivo principal de este análisis fue categorizar de manera empírica a los estudiantes universitarios de la Universidad de Oviedo (UNIOVI) en función de sus patrones de respuesta al cuestionario administrado.”

4. REFERENCIAS

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332.
- Celeux, G. & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2), 195-212. <https://doi.org/10.1007/BF01246098>
- Chan, P. & Shek, C. (2018). Latent class analysis of student artefacts. *Proceedings of the 10th International Conference on Education Technology and Computers*, 302-307. <https://doi.org/10.1145/3290511.3290521>
- Clark, S. L. (2010). *Mixture modeling with behavioral data*. University of California.
- Dan, S., Zheng, L., & Jian, Y. (2018). Symptom characteristics of stroke patients latent class analysis. *Chinese Nursing Research*, 32(20).
- Kalogrides, D., & Grodsky, E. (2011). Something to Fall Back On: Community Colleges as a Safety Net. *Social Forces*, 89(3), 853-877. <https://doi.org/10.1353/sof.2011.0019>
- Lu, Z. (2024). Choosing Among PCA, FA, LCA, LPA, and LDA. *Fudan J. Hum. Soc. Sci.* <https://doi.org/10.1007/s40647-024-00431-2>
- McCutcheon, A.L. (1987). *Latent class analysis*. Sage.
- Muthén, B. O. (2008). What is a good value of entropy? <http://www.statmodel.com/discussion/messages/13/2562.html?1487458497>
- Nylund, K. L., Asparouhov, T., & 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>
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Tieben, N. (2024). Destinations after higher education non-completion: The role of social background and pre-tertiary vocational qualifications. *Empirical Research in Vocational Education and Training*, 16(1), 9. <https://doi.org/10.1186/s40461-024-00161-5>
- Tinto, V. (1994). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226922461.001.0001>

Vermunt, J.K. & Magidson J. (2002). Latent class cluster analysis. En: Hagnaars, J.A. & McCutcheon, A.L. (editors). *Applied Latent Class Analysis* , 89-106. Cambridge University Press)

Weller, B. E., Bowen, N. K., & 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>

Zhang, H., Wang, J., Li, X., & Wang, Y. (2017). Influence of parental marital status and family atmosphere on the adolescent depressive symptoms and the moderating function of character strengths. *Guangdong Med J*, 38(4), 598-603.