

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



TITULO

**Modelo de Regresión de Clases Latentes: factores asociados a
la valoración de una universidad privada**

**TESIS PARA OPTAR POR EL GRADO DE MAGISTER EN
ESTADÍSTICA**

Presentado por:

Lucía Wiener Ramos

Asesor: Luis Valdivieso Serrano

Miembros del jurado:

Dr. Luis Valdivieso Serrano

Dr. Cristian Bayes Rodríguez

Dr. Carlos Veliz Capuñay

Lima, Diciembre 2015

Agradecimientos

Agradezco a mis padres por su cariño y ánimos durante todo este proceso. Y a mi asesor por su apoyo y paciencia.



Índice general

1. Introducción	1
1.1. Consideraciones Preliminares	1
1.1.1. Análisis de clases latentes	1
1.1.2. El modelo de regresión de clases latentes	2
1.1.3. Estudio sobre los factores asociados a la valoración de una universidad privada	3
1.2. Objetivos	4
1.3. Organización del Trabajo	4
2. El modelo lineal general de variables latentes	6
2.1. Modelo lineal general de variable latente	6
2.1.1. Conceptos generales	6
2.1.2. Principio de Independencia condicional	6
2.1.3. El modelo lineal general para variables latentes	7
2.1.4. Estimación de parámetros	8
2.2. Análisis de clases latentes	8
2.2.1. Variables manifiestas politómicas	9
2.2.2. Variables manifiestas politómicas ordenadas	10
2.2.3. Estimación por máxima verosimilitud y el algoritmo EM	10
2.2.4. Estimación por máxima verosimilitud para el caso binario	11
2.2.5. Error estándar para el caso binario	12
2.2.6. Estimación por máxima verosimilitud para el caso politómico	14
3. Modelos de regresión de clases latentes	17
3.1. Extensión para variables manifiestas politómicas	18
3.2. Interpretación de los coeficientes de regresión	19
3.3. Marginalización del modelo de regresión de clases latentes	20
3.4. Problema de Identificabilidad	20
3.5. Estimación de parámetros	21
3.5.1. Estimación de parámetros para el caso dicotómico	21
3.5.2. Estimación de parámetros para el caso politómico	25
3.6. Error estándar	29
3.7. Selección del modelo	29
3.8. Implementación computacional	30

4. Aplicación	31
4.1. Metodología del recojo de información	31
4.1.1. Método	31
4.1.2. Población de estudio	32
4.2. Metodología de análisis	32
4.3. Constructos de valoración	34
4.4. Factores asociados a los constructos de valoración	38
5. Conclusiones	46
5.1. Conclusiones	46
5.2. Sugerencias para investigaciones futuras	47
A. Sintaxis del poLCA	49
A.1. Sintaxis del poLCA	49
Bibliografía	50



Capítulo 1

Introducción

1.1. Consideraciones Preliminares

En diversos campos de análisis, especialmente en las ciencias sociales y humanas, se identifican constructos teóricos a los cuales queremos aproximarnos pero que no son directamente observables ni medibles, como por ejemplo, la calidad o satisfacción con un servicio, el nivel de estrés, el nivel de conocimiento en matemáticas, entre otras. Este tipo de constructos son llamados variables latentes y su valor solo puede ser aproximado a través de variables observadas o manifiestas que si pueden ser medidas ([Bartholomew et al., 2011](#)).

Dependiendo del tipo de variables manifiestas y latentes que estemos tratando (categórica o continua), se asumen cuatro tipos de modelos de variables latentes:

Variable Manifiesta	Variable Latente	
	Continua	Categórica
Continua	Análisis Factorial	Modelos de Perfiles Latentes
Categórica	Modelos de Trazo Latente	Modelos de clases Latentes

Cuadro 1.1: Modelos de Variables Latentes

Como se observa en el Cuadro 1.1, cuando las variables manifiestas y las variables latentes que se quiere conocer son categóricas, el modelo de variable latente a aplicar es el análisis de clases latentes (ACL).

1.1.1. Análisis de clases latentes

El modelo de clases latentes es utilizado para el análisis de datos categóricos multivariados y permite establecer relaciones entre las variables observadas, suponiendo que hay una estructura subyacente explicada por una o más variables latentes que generan esta relación. Es a través de este modelo que se puede definir un número de clases latentes que represente patrones de respuesta de un conjunto de variables manifiestas, y calcular la probabilidad de que una observación, a la que en adelante llamaremos individuo, pertenezca a una clase ([Collins and Lanza, 2010](#)).

Para una mejor comprensión, supongamos que tenemos p variables manifiestas dicotómicas; es decir, variables que tomarán tan solo dos valores, los cuales identificaremos sin pérdida de generalidad como 1 y 0 (más adelante extenderemos esto a un modelo politómico en el cual las variables podrían tomar dos o más valores), y sea Y la variable latente categórica subyacente con $j = 0, 1, 2, \dots, K - 1$ clases latentes, las cuales corresponden a las categorías

de esta variable.

Si denotamos por $\eta_j = P(Y = j)$ a la probabilidad de que un individuo pertenezca a la clase latente j , se debe de cumplir naturalmente que

$$\sum_{j=0}^{K-1} \eta_j = 1.$$

La probabilidad de responder positivamente a una variable manifiesta i , condicionada por la pertenencia a la clase latente j , se representará en adelante por π_{ij} ; es decir, si X_i denota a la posible respuesta dada por un individuo a la variable manifiesta i ,

$$\pi_{ij} = P(X_i = 1 | Y = j).$$

Sea $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ el vector de respuestas de un individuo a las p variables manifiestas y sea $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ el valor observado o patrón de respuesta correspondiente a este individuo. La probabilidad de observar este patrón particular de respuestas condicionada a la pertenencia a una clase latente j , se obtendrá como:

$$P(\mathbf{X}=\mathbf{x}|Y = j) = \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i},$$

donde se asume aquí un supuesto de independencia condicional.

Por otro lado, si buscamos la probabilidad de pertenencia a una clase latente j , dado un vector de respuestas \mathbf{x} , podríamos invocar al teorema de Bayes y obtener la siguiente probabilidad de clasificación a posteriori:

$$P(Y = j|\mathbf{X}=\mathbf{x}) = \frac{P(\mathbf{X}=\mathbf{x}|Y = j)\eta_j}{P(\mathbf{X}=\mathbf{x})}$$

donde

$$P(\mathbf{X}=\mathbf{x}) = \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i},$$

o más compactamente

$$P(Y = j|\mathbf{X}=\mathbf{x}) = \frac{\eta_j \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}}{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}}.$$

En el análisis de clases latentes será entonces de interés estimar estas probabilidades, para lo cual requeriremos estimar las probabilidades π_{ij} y las probabilidades de pertenencia a la clase η_j .

1.1.2. El modelo de regresión de clases latentes

Una limitante del análisis de clases latentes es que la probabilidad de pertenencia a una clase latente es igual en todos los individuos que tienen los mismos patrones de respuesta a las variables manifiestas que sirven de proxy para la variable subyacente latente, sin tomar en cuenta las características propias de los individuos, las cuales pudieran influir en su patrón

de respuesta, y por lo tanto, en la probabilidad de pertenecer a una clase latente, tales como el sexo, la edad, el nivel socioeconómico, etc.

Una extensión al análisis de clases latentes es el análisis de regresión de clases latentes, el cual considera covariables como las antes mencionadas y permite modelar con mayor información la probabilidad de pertenencia a una clase dependiendo de estas. (Clogg (1981), Formann (1992)).

Este tipo de análisis es similar al de la regresión logística, con la diferencia que la variable respuesta es latente en vez de ser directamente observable. Las covariables son ingresadas para analizar su posible efecto en la prevalencia de un individuo a una clase latente, verificando si esta inclusión es significativa o no, y si lo es, que tipo de efecto genera. Como en cualquier tipo de regresión, el grupo de covariables puede incluir variables categóricas o numéricas (Agresti, 1990).

Sea $\mathbf{Z} = (Z_1, Z_2, \dots, Z_m)'$ el vector de covariables de un individuo y sea $\mathbf{z} = (z_1, z_2, \dots, z_m)'$ su valor observado. Estas covariables como dijimos se introducen para intentar explicar la pertenencia del individuo a una clase j mediante el siguiente modelo de regresión logística multinomial:

$$P(Y = j | \mathbf{Z} = \mathbf{z}) = \frac{e^{\beta_{0j} + \sum_{k=1}^m \beta_{kj} z_k}}{1 + \sum_{l=1}^{K-1} e^{\beta_{0l} + \sum_{k=1}^m \beta_{kl} z_k}}$$

donde $j = 0, 1, 2, \dots, K - 1$. Ello produce estimaciones del efecto de las covariables para cada clase latente en comparación con una clase latente de referencia, por lo que se obtienen $K - 1$ coeficientes de regresión β_{mj} correspondientes a cada covariable, además del intercepto β_{0j} para cada clase latente con excepción de la designada como referencia, que asumiremos aquí, sin pérdida de generalidad, es la clase 0.

Los coeficientes de regresión pueden ser exponenciados y transformados en odds ratio para su interpretación. Estos se calculan en relación a la clase referencial elegida y refleja el cambio que la covariable genera.

Si bien en el ACL estimábamos dos conjuntos de parámetros; las probabilidades de respuesta a los ítems π_{ij} y las probabilidades de pertenencia a una clase latente η_j , en el modelo de regresión de clases latentes estimamos las primeras probabilidades pero no las segundas. En su lugar, se estiman los coeficientes de regresión β_{mj} y la pertenencia a una clase latente se expresa como función de los coeficientes de regresión y los valores individuales de las covariables correspondientes. En nuestro modelo las covariables se relacionan con la pertenencia a una clase pero no con la probabilidad de respuesta a los ítems (Collins and Lanza, 2010).

1.1.3. Estudio sobre los factores asociados a la valoración de una universidad privada

Los modelos de regresión de clases latentes han sido aplicados en diversos campos de investigación, en psicología para estudiar actitudes hacia el consumo de drogas (Chung et al., 2006) y el aborto (Chung, 2003); en ciencias sociales, buscando identificar roles de género en países musulmanes (Blaydes and D.A., 2007) y en el Japón (Yamaguchi, 2000); en ciencias políticas (Breen, 2000), relacionando la orientación política y aspectos sociodemográficos (Quinn, 2004); en el marketing, mediante la segmentación de mercados para conocer tipos

CAPÍTULO 1. INTRODUCCIÓN

de consumidores (Kamakura et al., 1994), entre otros.

Una de sus aplicaciones más interesantes esta orientada a identificar niveles de valoración y definir constructos de calidad. Por citar Morrison et al. (2011) aplican un modelo de regresión de clase latente para identificar constructos de calidad en la percepción de excelencia en programas de PhD. A partir del diseño de un modelo que considera 26 variables orientadas a conocer niveles de satisfacción con distintos aspectos de los estudios doctorales, ellos optaron por un modelo de dos clases, una clase donde predominaba la evaluación de la calidad académica para evaluar la calidad de programa, y otro donde la evaluación se daba desde un enfoque multi dimensional. Uno de los principales factores o variables en el modelo de regresión que ellos encontraron como significativas en su estudio fué la reputación académica de la facultad.

En el presente trabajo estaremos interesados en realizar un estudio similar al citado, basado en el Sistema de Seguimiento a Egresados de una universidad privada de pregrado del país, el cual busca contar con información sobre la situación y valoración de los egresados, específicamente en lo que se refiere a los estudios luego de egresar, situación laboral al momento del recojo de información, experiencia laboral previa y valoración con distintos aspectos de la universidad en general, de su especialidad, de la imagen institucional y de los estudios generales llevados.

El recojo de información se hace de manera censal a todas las promociones que egresan; en este caso nos centraremos en los alumnos que egresaron los años 2009, 2010, 2011 y 2012. Cada grupo fue encuestado a 2 años de su egreso.

De manera similar al estudio de Morrison et al. (2011), será de gran interés aquí conocer que variables intervienen primordialmente cuando se evalúa una universidad privada de pregrado, identificar constructos de valoración, e indagar si esta evaluación se ve afectada o no por características personales de los egresados con data obtenida de la encuesta o de la misma universidad.

1.2. Objetivos

Hacer una revisión teórica del modelo de regresión de clases latentes y aplicarla a un conjunto de datos reales, buscando identificar constructos y covariables orientados a explicar niveles de valoración general con una universidad privada de pregrado. De manera específica:

- Revisar la literatura acerca del análisis de clases latentes.
- Estudiar los modelos de regresión de clases latentes.
- Aplicar el modelo de regresión de clases latentes a un conjunto de datos reales.

1.3. Organización del Trabajo

En el Capítulo 2 se presenta consideraciones generales acerca del modelo lineal general de variables latentes y el modelo de clases latentes. En el Capítulo 3 se estudian los modelos de regresión de clases latentes, la estimación de sus parámetros y su implementación computacional. En el Capítulo 4 se presenta los resultados de la aplicación del modelo a un conjunto de datos reales orientados a conocer la valoración de una universidad privada. En

el Capítulo 5 se presenta algunas conclusiones, recomendaciones y futuras extensiones que se podrían derivar de este trabajo.



Capítulo 2

El modelo lineal general de variables latentes

2.1. Modelo lineal general de variable latente

2.1.1. Conceptos generales

En muchas áreas de estudio, particularmente en las ciencias sociales y la psicología, es de gran interés aproximarnos a conceptos o constructos teóricos, los cuales, si bien no pueden ser observados directamente, asumimos que ejercen un efecto sobre variables que si pueden ser medidas, permitiéndonos una aproximación a estos conceptos y un mejor entendimiento de las variables medidas.

Estos conceptos teóricos son comúnmente llamados variables latentes, y su modelo estadístico específica la distribución conjunta de las variables manifiestas (u observadas) y latentes.

Siguiendo a [Bartholomew et al. \(2011\)](#), denotemos por $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ al vector columna p -dimensional de variables manifiestas u observadas. Por otro lado, sea $\mathbf{Y} = (Y_1, Y_2, \dots, Y_q)'$ el vector de variables latentes. Dado que solo \mathbf{X} puede observarse, toda inferencia se basará en la distribución conjunta de este vector, cuya función de densidad se expresará de la siguiente manera:

$$f(\mathbf{x}) = \int h(\mathbf{y})g(\mathbf{x}|\mathbf{y})d\mathbf{y}$$

siendo $h(\mathbf{y})$ la distribución a priori de \mathbf{Y} y $g(\mathbf{x}|\mathbf{y})$ la distribución condicional de \mathbf{X} dado que $\mathbf{Y} = \mathbf{y}$. Nuestro propósito es aproximarnos a la distribución de \mathbf{Y} a partir de la observación que hemos hecho de \mathbf{X} . Esta información se resume en la densidad condicional a posteriori $h(\mathbf{y}|\mathbf{x})$, deducida del Teorema de Bayes:

$$h(\mathbf{y}|\mathbf{x}) = \frac{h(\mathbf{y})g(\mathbf{x}|\mathbf{y})}{f(\mathbf{x})}$$

Así, para conocer esta distribución condicional, es preciso conocer h y g .

2.1.2. Principio de Independencia condicional

El modelo de variable latente asume que las variables observadas son independientes, condicionadas a la presencia de la variable latente. Así, la dependencia entre las variables observadas, componentes de \mathbf{X} , es inducida por las variables latentes en \mathbf{Y} . El vector de variables latentes debe ser de una dimensión q tal que de conocerse estas variables, se pueda garantizar la independencia en los componentes de \mathbf{X} ; vale decir:

$$g(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^p g_i(x_i|\mathbf{y}), \quad (2.1)$$

donde g_i denota a la densidad condicional de la variable manifiesta X_i dado que $\mathbf{Y} = \mathbf{y}$. Si esto ocurre, lo cual asumiremos en adelante, se debe cumplir que

$$f(\mathbf{x}) = \int h(\mathbf{y}) \prod_{i=1}^p g_i(x_i|\mathbf{y}) d\mathbf{y}.$$

2.1.3. El modelo lineal general para variables latentes

Una distribución conveniente para las densidades g_i en (2.1) viene dada por la familia exponencial de un parámetro (Bartholomew et al., 2011). Suponiendo que las variables latentes se resumen en un parámetro θ_i , el modelo está definido por:

$$g_i(x_i|\mathbf{y}) = g_i(x_i|\theta_i) = F_i(x_i)G_i(\theta_i)\exp(\theta_i u_i(x_i)), \quad i = 1, 2, \dots, p \quad (2.2)$$

donde F_i y G_i son funciones que dependen de la variable manifiesta y el parámetro latente respectivamente, y θ_i es alguna función de \mathbf{y} . Esta función la asumiremos lineal; es decir, que explícitamente vendrá dada por:

$$\theta_i = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \dots + \alpha_{iq}y_q, \quad (2.3)$$

donde los α_{ij} serán parámetros desconocidos.

Algunos miembros particulares de la familia exponencial son las distribuciones Bernoulli, Gamma y Normal.

Por ejemplo, si X_i tiene una distribución de Bernoulli, la distribución condicional de X_i dado que $\mathbf{Y} = \mathbf{y}$ se puede definir de la siguiente manera:

$$g_i(x_i | \mathbf{y}) = \pi_i^{x_i} (1 - \pi_i)^{1-x_i} = (1 - \pi_i) \exp(x_i \text{logit}(\pi_i))$$

donde $x_i \in \{0, 1\}$ y π_i es la probabilidad de que X_i sea 1. En este caso, lo que antes habíamos definido por θ_i es igual a $\text{logit}(\pi_i) = \log(\pi_i/(1 - \pi_i))$. Así, $G_i(\theta_i)$ es igual a $1 - \pi_i$, y $u_i(x)$ igual a x . De esta manera, el modelo lineal general se completa especificando que

$$\text{logit}(\pi_i) = \alpha_{i0} + \sum_{j=1}^q \alpha_{ij} y_j.$$

Esta función logit es conocida como una función de enlace. En el caso $q = 1$, el modelo de variable latente binario se reduce a:

$$\pi_i(y) = \frac{1}{1 + e^{-\alpha_{i0} - \alpha_{i1}y}}$$

el cual se le conoce también, en la teoría de respuesta al ítem, como un modelo de dos parámetros.

Los parámetros α_{ij} en (2.2) son desconocidos y deben ser estimados en base de la data particular que se obtenga de la distribución conjunta de \mathbf{X} , la cual, recordemos, depende

también de la distribución a priori.

2.1.4. Estimación de parámetros

Para poder estimar los parámetros es necesario revisar algunos supuestos que se deben hacer acerca de la distribución a priori $h(\mathbf{y})$. Sin pérdida de generalidad asumiremos que \mathbf{Y} tiene una distribución normal q -variada de media $\mathbf{0}$ y matriz de varianza igual a la identidad, lo cual denotaremos por $\mathbf{N}_q(\mathbf{0}, \mathbf{I})$. Esto desde que, si las variables componentes latentes fueran independientes, cualquier cambio en la localización podría ser absorbida por la constante α_{i0} , y todo cambio de escala por los parámetros α_{ij} para $j > 0$.

Aquí la independencia solo es asumida para facilitar la interpretación; el modelo general podría soportar también variables latentes que no son independientes.

La estimación de los parámetros del modelo suele hacerse, desde el punto de vista clásico, por el método de máxima verosimilitud. La función de máxima verosimilitud se forma a partir de la distribución conjunta de los patrones de respuesta a las variables manifiestas en los n individuos de una m.a.:

$$\begin{aligned}
 L(\boldsymbol{\alpha}) &= \prod_{h=1}^n f(\mathbf{x}_h) = \prod_{h=1}^n \int h(\mathbf{y}) \prod_{i=1}^p g_i(x_{ih}|\mathbf{y}) d\mathbf{y} \\
 &= \prod_{h=1}^n \int h(\mathbf{y}) \left(\prod_{i=1}^p F_i(x_{ih}) G_i(\theta_i) \right) \exp \left(\sum_{i=1}^p u_i(x_{ih}) \theta_i \right) d\mathbf{y}
 \end{aligned} \tag{2.4}$$

donde x_{ih} denota la respuesta del individuo h a la variable manifiesta X_i , $\boldsymbol{\alpha} = [\alpha_{i0}, \alpha_{i1}, \dots, \alpha_{iq}]$ y $\theta_i = \alpha_{i0} + \sum_{j=1}^q \alpha_{ij} y_j$.

La optimización directa de (2.4) resulta en general complicada, sobre todo si q es grande, dada la integral presente. En este sentido, dos alternativas muy populares para este tipo de modelos son el uso del algoritmo de Esperanza-Maximización (EM) y el uso de métodos Bayesianos que utilizan cadenas de Markov de Montecarlo para generar muestras de la distribución a posteriori.

2.2. Análisis de clases latentes

El modelo de clases latentes asume que tanto las variables manifiestas como latentes son categóricas. El hecho que consideremos a la variable latente como categórica puede ser producto de una evidencia a priori o teórica, o simplemente tomada así por una cuestión práctica. Para mayor detalle ver [Lazarsfeld and Henry \(1968\)](#)

Las bondades de este tipo de análisis son varios: reducen la complejidad de un conjunto de datos agrupándolos en clases, permite estimar varias probabilidades, y permite analizar datos categóricos sin la necesidad de aplicar transformaciones.

Para empezar, partimos de dos supuestos para la versión estándar del modelo.

1. Dentro de cada clase, todos los individuos tienen las mismas probabilidades de respuesta a las variables manifiestas.
2. Se cumple la independencia condicional; es decir, las respuestas entre los individuos son independientes dado que estos pertenecen a una misma clase.

Para una mejor comprensión analicemos primero un modelo de clases latentes con variables manifiestas binarias, en el que la variable latente Y es unidimensional y presenta K categorías o clases que los denotaremos con el índice $j = 0, 1, 2, \dots, K - 1$. Sea π_{ij} la probabilidad de obtenerse una respuesta positiva a la variable manifiesta X_i , dado que la variable latente esté en la clase j ; es decir, $\pi_{ij} = P(X_i = 1 | Y = j)$ y sea η_j la probabilidad a priori de que un individuo pertenezca a la clase j . Entonces la función de probabilidad conjunta del vector de respuestas manifiestas o patrones de respuesta \mathbf{X} viene dado por:

$$f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = \sum_{j=0}^{K-1} P(\mathbf{X} = \mathbf{x} | Y = j)P(Y = j) = \sum_{j=0}^{K-1} g(\mathbf{x} | j)\eta_j, \quad (2.5)$$

donde

$$g(\mathbf{x} | j) = \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}.$$

De esta manera, cada individuo es descrito por dos dimensiones: el vector de variables manifiestas \mathbf{X} y el indicador de pertenencia a clase.

La probabilidad a posteriori de que un individuo con un patrón de respuestas $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ pertenezca a la clase j es por tanto:

$$h(j | \mathbf{x}) = P(Y = j | \mathbf{X} = \mathbf{x}) = \frac{\eta_j \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}}{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}} \quad (2.6)$$

Esta función es utilizada luego como regla de clasificación para asignar a un individuo a la clase que tiene mayor probabilidad de pertenencia, dado el patrón de respuestas \mathbf{x} que tenga.

El principal problema estadístico en el modelo consiste en la estimación de los parámetros η_j y π_{ij} , y el cálculo del ajuste del modelo. Otro problema será también identificar las clases latentes subyacentes e interpretarlas de manera que tengan un sentido con la data.

2.2.1. Variables manifiestas politómicas

La extensión al caso de variables manifiestas politómicas es directa. Si tenemos $c_i \geq 2$ categorías en una variable manifiesta i , podríamos definir para cada una de las posibles categorías $s = 1, 2, \dots, c_i - 1$ las variables binarias:

$$X_i(s) = \begin{cases} 1, & \text{si la respuesta de un individuo a la variable manifiesta } i \text{ está en la categoría } s, \\ 0, & \text{en caso contrario} \end{cases}$$

con lo que $\sum_{s=1}^{c_i-1} X_i(s) = 1$. Así, si identificamos la respuesta de un individuo a la variable manifiesta i con el vector columna $\mathbf{X}_i = [X_i(0), \dots, X_i(c_i - 1)]'$, el vector de respuestas de un individuo se podrá representar en el vector columna $\mathbf{X}' = [\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_p]$.

Si definimos por $\pi_{ij}(s)$ a la probabilidad de que un individuo en una clase j de una respuesta positiva en la categoría s de la variable i ; es decir, $\pi_{ij}(s) = P(X_i(s) = 1 | Y = j)$, entonces, la función de probabilidad conjunta del vector de respuestas manifiestas o patrones de respuesta \mathbf{x} viene dada por:

$$f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)}. \quad (2.7)$$

y la distribución a posteriori es

$$h(j | \mathbf{x}) = P(Y = j | \mathbf{X} = \mathbf{x}) = \frac{\eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)}}{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)}}. \quad (2.8)$$

2.2.2. Variables manifiestas politómicas ordenadas

En el caso de que las variables manifiestas politómicas no sean nominales, sino ordinales, se requieren consideraciones extras.

En este caso, para preservar la característica ordinal de la variable manifiesta componente X_i , debemos modelar la probabilidad acumulada de responder positivamente en una categoría igual o mayor a la categoría s de esta variable, dado que el individuo pertenece a la clase j :

$$\Pi_{ij}(s) = \pi_{ij}(s) + \pi_{ij}(s+1) + \dots + \pi_{ij}(c_i-1).$$

La idea es estimar los $\Pi_{ij}(s)$ y obtener de aquí los $\pi_{ij}(s) = \Pi_{ij}(s) - \Pi_{ij}(s+1)$. De esta manera, la probabilidad conjunta de \mathbf{X} viene dada por:

$$f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)} = \sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} (\Pi_{ij}(s) - \Pi_{ij}(s+1))^{x_i(s)}, \quad (2.9)$$

donde $x_i(s) = 1$, si un individuo responde positivamente a la categoría s de la i -ésima variable y $x_i(s) = 0$ en caso contrario.

La distribución a posteriori viene dada en este contexto por:

$$P(Y = j | \mathbf{X} = \mathbf{x}) = \frac{\eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} (\Pi_{ij}(s) - \Pi_{ij}(s+1))^{x_i(s)}}{\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} (\Pi_{ij}(s) - \Pi_{ij}(s+1))^{x_i(s)}}. \quad (2.10)$$

2.2.3. Estimación por máxima verosimilitud y el algoritmo EM

Como mencionamos en la Sección 2.1.4 la optimización directa de la función de verosimilitud es en general complicada para los modelos de variables latentes dada la gran cantidad de parámetros involucrados. Los algoritmos de Esperanza Maximización (EM) permiten precisamente la maximización de la función de verosimilitud cuando los procedimientos estándares son numéricamente difíciles o inviables.

[Dempster et al. \(1977\)](#) introdujo este algoritmo para el cálculo de parámetros estimados por máxima verosimilitud, cuando se contaban con datos no observados o casos perdidos.

En nuestro contexto de variables latentes tales datos no observados los podemos entender como los provenientes de las variables latentes en estudio.

El algoritmo EM consiste en alternar pasos de esperanza (E), en el que uno obtiene la esperanza de la función de verosimilitud mediante la inclusión de las variables latentes como si fueran observables, y el paso de maximización (M), en el cual se obtienen estimadores

de máxima verosimilitud de los parámetros mediante la maximización de la verosimilitud esperada hallada en el paso E. Los parámetros hallados en el paso M son luego usados para hallar la esperanza en el paso E y el proceso se repite de manera iterativa.

Este algoritmo en su forma más general especifica la log-verosimilitud completa del modelo, el cual incluye tanto a las variables manifiestas como latentes de la siguiente manera:

$$\mathcal{L}_c = \sum_{h=1}^n \log f(\mathbf{X}_h, \mathbf{Y}_h) = \sum_{h=1}^n (\log g(\mathbf{X}_h | \mathbf{Y}_h) + \log p(\mathbf{Y}_h)) = \sum_{h=1}^n \left(\sum_{i=1}^p \log g_i(X_{ih} | \mathbf{Y}_h) + \log p(\mathbf{Y}_h) \right)$$

De esta manera, se alternan las etapas E, que evalúan $E[\mathcal{L}_c | \mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_n = \mathbf{x}_n]$, y las etapas M que maximizan esta función de manera iterativa. Dado que la distribución a priori no contiene a los parámetros de interés, \mathcal{L}_c podría reemplazarse para simplificar por:

$$\mathcal{L}^* = \sum_{h=1}^n \log g(\mathbf{X}_h | \mathbf{Y}_h) = \sum_{h=1}^n \left(\sum_{i=1}^p \log g_i(X_{ih} | \mathbf{Y}_h) \right).$$

2.2.4. Estimación por máxima verosimilitud para el caso binario

Siguiendo las ecuaciones (2.4) y (2.5), la función de log-verosimilitud para una muestra aleatoria de tamaño n del vector de variables manifiestas viene dada por:

$$\mathcal{L} = \sum_{h=1}^n \log \left(\sum_{j=0}^{K-1} g(\mathbf{x}_h | j) \eta_j \right) = \sum_{h=1}^n \log \left(\sum_{j=0}^{K-1} \eta_j \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \right). \quad (2.11)$$

La maximización de \mathcal{L} está naturalmente sujeta a la restricción

$$\sum_{j=0}^{K-1} \eta_j = 1,$$

lo cual equivale a la maximización del Lagrangiano:

$$Lg = \mathcal{L} + \lambda \left(\sum_{j=0}^{K-1} \eta_j - 1 \right),$$

siendo λ el multiplicador de Lagrange.

Las derivadas parciales de esta función con respecto a las probabilidades η_j y π_{ij} vienen dadas respectivamente por:

$$\frac{\partial Lg}{\partial \eta_j} = \sum_{h=1}^n \frac{g(\mathbf{x}_h | j)}{f(\mathbf{x}_h)} + \lambda = \sum_{h=1}^n \left(\prod_{i=1}^p \frac{1}{f(\mathbf{x}_h)} \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \right) + \lambda$$

y

$$\frac{\partial Lg}{\partial \pi_{ij}} = \sum_{h=1}^n \eta_j \frac{\partial g(\mathbf{x}_h | j)}{\partial \pi_{ij}} \frac{1}{f(\mathbf{x}_h)}.$$

Dado que

$$\begin{aligned} \frac{\partial g(\mathbf{x}_h | j)}{\partial \pi_{ij}} &= \frac{\partial}{\partial \pi_{ij}} \left(\exp \left(\sum_{i=1}^p (x_{ih} \log(\pi_{ij}) + (1 - x_{ih}) \log(1 - \pi_{ij})) \right) \right) = g(\mathbf{x}_h | j) \left(\frac{x_{ih}}{\pi_{ij}} - \frac{1 - x_{ih}}{1 - \pi_{ij}} \right) \\ &= \frac{1}{\pi_{ij}(1 - \pi_{ij})} (x_{ih} - \pi_{ij}) g(\mathbf{x}_h | j); \end{aligned}$$

se tiene

$$\frac{\partial Lg}{\partial \pi_{ij}} = \frac{\eta_j}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^n (x_{ih} - \pi_{ij}) \frac{g(\mathbf{x}_h | j)}{f(\mathbf{x}_h)}.$$

Expresando estas ecuaciones en términos de las probabilidades a posteriori $h(j | \mathbf{x}_h) = \eta_j \frac{g(\mathbf{x}_h | j)}{f(\mathbf{x}_h)}$ en (2.6), e igualando a 0, se obtiene de la restricción para los η_j 's que $\lambda = -n$. Esto consecuentemente, nos remite a las siguientes ecuaciones de estimación:

$$\eta_j = \frac{1}{n} \sum_{h=1}^n h(j | \mathbf{x}_h) \quad (2.12)$$

y

$$\pi_{ij} = \frac{1}{n\eta_j} \sum_{h=1}^n x_{ih} h(j | \mathbf{x}_h). \quad (2.13)$$

Claramente de conocerse $h(j | \mathbf{x}_h)$ estas ecuaciones serían sencillas de resolver, lo cual motiva el siguiente algoritmo recursivo de estimación:

1. Elegir una distribución inicial para las probabilidades a posteriori $h(j | \mathbf{x}_h)$.
2. Usar (2.12) y (2.13) para obtener una primera aproximación de los estimadores de máxima verosimilitud para los η_j y π_{ij} .
3. Substituir las estimaciones anteriores en (2.6) para mejorar la distribución de las probabilidades a posteriori $h(j | \mathbf{x}_h)$.
4. Retomar el paso 2 para obtener una segunda aproximación de los parámetros y continuar el ciclo hasta alcanzar la convergencia.

Se sugiere utilizar en este algoritmo distintas distribuciones iniciales para que la solución obtenida no corresponda a un máximo local.

2.2.5. Error estándar para el caso binario

Una de las limitaciones del algoritmo anterior y del algoritmo EM es que estos no nos proveen de los errores estándar de estimación. Para obtener estos podríamos optar por utilizar la matriz de información de Fisher observada, la cual involucra el cálculo de la matriz Hessiana de \mathcal{L} (matriz de segundas derivadas parciales de \mathcal{L}).

Recordando la restricción

$$\sum_{j=0}^{K-1} \eta_j = 1,$$

la ecuación (2.5) para el patrón de respuestas de un individuo h podría expresarse como:

$$f(\mathbf{x}_h) = g(\mathbf{x}_h|0) + \sum_{j=0}^{K-1} \eta_j (g(\mathbf{x}_h|j) - g(\mathbf{x}_h|0)), \quad (2.14)$$

siendo $g(\mathbf{x}_h|j) = \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}$.

Las derivadas parciales de la función de log-verosimilitud, $\mathcal{L} = \sum_{h=1}^n \log(f(\mathbf{x}_h))$, pueden obtenerse, en términos de la distribución a posteriori, por:

$$\frac{\partial \mathcal{L}}{\partial \eta_j} = \sum_{h=1}^n \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \eta_j} = \sum_{h=1}^n \frac{g(\mathbf{x}_h|j)}{f(\mathbf{x}_h)}$$

y

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}} = \sum_{h=1}^n \frac{1}{f(\mathbf{x}_h)} \frac{\partial f(\mathbf{x}_h)}{\partial \pi_{ij}} = \frac{\eta_j}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^n (x_{ih} - \pi_{ij}) \frac{g(\mathbf{x}_h|j)}{f(\mathbf{x}_h)}, \quad (2.15)$$

desde que

$$\begin{aligned} \frac{\partial f(\mathbf{x}_h)}{\partial \pi_{ij}} &= \eta_j \frac{\partial g(\mathbf{x}_h|j)}{\partial \pi_{ij}} = \eta_j \frac{\partial}{\partial \pi_{ij}} (\pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \prod_{l \neq i} \pi_{lj}^{x_{lh}} (1 - \pi_{lj})^{1-x_{lh}}) \\ &= \eta_j \left(\prod_{l \neq i} \pi_{lj}^{x_{lh}} (1 - \pi_{lj})^{1-x_{lh}} \right) \frac{x_{ih} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})} (\pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}) \\ &= \eta_j g(\mathbf{x}_h|j) \frac{x_{ih} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}. \end{aligned}$$

De otro lado, para las derivadas parciales de segundo orden de \mathcal{L} , tenemos en primer lugar:

$$\frac{\partial^2 \mathcal{L}}{\partial \eta_k \partial \eta_j} = - \sum_{h=1}^n \frac{g(\mathbf{x}_h|j)}{f(\mathbf{x}_h)^2} \frac{\partial}{\partial \eta_k} f(\mathbf{x}_h) = - \sum_{h=1}^n \frac{g(\mathbf{x}_h|j)}{f(\mathbf{x}_h)^2} g(\mathbf{x}_h|k).$$

se cumple que

$$\frac{g(\mathbf{x}_h|j)}{f(\mathbf{x}_h)} = \frac{h(j|\mathbf{x}_h)}{\eta_j}$$

y por tanto

$$\frac{\partial^2 \mathcal{L}}{\partial \eta_k \partial \eta_j} = - \sum_{h=1}^n \frac{h(j|\mathbf{x}_h)h(k|\mathbf{x}_h)}{\eta_j \eta_k}.$$

Por otro lado, recordando lo visto en (2.15), tenemos que:

$$\frac{\partial^2 \mathcal{L}}{\partial \eta_k \partial \pi_{ij}} = \frac{\eta_j}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^n (x_{ih} - \pi_{ij}) g(\mathbf{x}_h|j) \frac{\partial}{\partial \eta_k} \left(\frac{1}{f(\mathbf{x}_h)} \right),$$

donde el último término viene según (2.14) dado por

$$\frac{\partial}{\partial \eta_k} \left(\frac{1}{f(\mathbf{x}_h)} \right) = \frac{1}{f(\mathbf{x}_h)^2} \frac{\partial}{\partial \eta_k} f(\mathbf{x}_h) = \frac{g(\mathbf{x}_h|0) - g(\mathbf{x}_h|k)}{f(\mathbf{x}_h)f(\mathbf{x}_h)}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \eta_k \partial \pi_{ij}} = \frac{\eta_j}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^n (x_{ih} - \pi_{ij}) h(j|\mathbf{x}_h) \left(\frac{h(0|\mathbf{x}_h)}{\eta_0} - \frac{h(k|\mathbf{x}_h)}{\eta_k} \right). \quad (2.16)$$

Así,

$$\frac{\partial^2 \mathcal{L}}{\partial \eta_j \partial \pi_{ij}} = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^n (x_{ih} - \pi_{ij}) \frac{h(j|x_h)}{\eta_j} (1 - h(j|x_h)).$$

Finalmente

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{lk} \partial \pi_{ij}} = \frac{-\eta_j}{\pi_{ij}(1 - \pi_{ij})} \sum_{h=1}^n \frac{x_{ih} - \pi_{ij} g(\mathbf{x}_h|j)}{f(\mathbf{x}_h) f(\mathbf{x}_h)} \eta_k g(\mathbf{x}_h|k) \frac{x_{lh} - \pi_{lk}}{\pi_{lk}(1 - \pi_{lk})}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{lk} \partial \pi_{ij}} = - \sum_{h=1}^n \frac{(x_{ih} - \pi_{ij})(x_{lh} - \pi_{lk}) h(j|\mathbf{x}_h) h(k|\mathbf{x}_h)}{\pi_{ij}(1 - \pi_{ij}) \pi_{lk}(1 - \pi_{lk})}$$

en el caso que $l \neq i$ y $k \neq j$. Esto último no es difícil generalizar para los demás casos resultando en una expresión de la forma:

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{lk} \partial \pi_{ij}} = \sum_{h=1}^n (x_{ih} - \pi_{ij})(x_{lh} - \pi_{lk}) h(j|\mathbf{x}_h) \times \frac{\delta_{jk}(1 - \delta_{il}) - h(k|\mathbf{x}_h)}{\pi_{ij}(1 - \pi_{ij}) \pi_{lk}(1 - \pi_{lk})},$$

donde

$$\delta_{rs} = \begin{cases} 1, & \text{si } r = s, \\ 0, & \text{en caso contrario.} \end{cases}$$

Una estimación de la matriz de varianza-covarianza asintótica de los estimadores de máxima verosimilitud viene entonces dada por la inversa de la matriz hessiana de orden $(K(p+1) - 1) \times (K(p+1) - 1)$ cuyas entradas hemos aquí obtenido.

2.2.6. Estimación por máxima verosimilitud para el caso politómico

Al igual que en el caso binario, la función de verosimilitud se puede escribir por (2.7) de la siguiente manera:

$$\mathcal{L} = \sum_{h=1}^n \log f(\mathbf{x}_h) = \sum_{h=1}^n \log \left(\sum_{j=0}^{K-1} \eta_j g(\mathbf{x}_h|j) \right)$$

donde

$$g(\mathbf{x}_h|j) = \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_{ih}(s)}.$$

Si bien en el caso binario solo se tenía la restricción $\sum_{j=0}^{K-1} \eta_j = 1$, en el caso politómico se mantiene esta restricción y se requiere adicionalmente la siguiente:

$$\sum_{s=1}^{c_i-1} \pi_{ij}(s) = 1,$$

para $i = 1, 2, \dots, p$ y $j = 0, 1, 2, \dots, c_i$.

De esta manera, la función a maximizar en este caso es la siguiente:

$$Lg = \mathcal{L} + \lambda \left(\sum_{j=0}^{K-1} \eta_j - 1 \right) + \sum_{j=0}^{K-1} \sum_{i=1}^p \phi_{ij} \left(\sum_{s=1}^{c_i-1} \pi_{ij}(s) - 1 \right), \quad (2.17)$$

siendo λ y ϕ_{ij} multiplicadores de Lagrange indeterminados.

Las derivadas parciales con respecto a η_j son:

$$\frac{\partial Lg}{\partial \eta_j} = \sum_{h=1}^n \frac{g(\mathbf{x}_h | j)}{f(\mathbf{x}_h)} + \lambda = \sum_{h=1}^n \frac{h(j | \mathbf{x}_h)}{\eta_j} + \lambda.$$

Igualando a 0 esta expresión y utilizando la restricción para los η_j , obtendremos como en el caso binario que:

$$\eta_j = \frac{1}{n} \sum_{h=1}^n h(j | \mathbf{x}_h). \quad (2.18)$$

De otro lado, como

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}(s)} = \sum_{h=1}^n \frac{1}{f(\mathbf{x}_h)} \eta_j \frac{\partial g(\mathbf{x}_h | j)}{\partial \pi_{ij}(s)}$$

y

$$\frac{\partial g(\mathbf{x}_h | j)}{\partial \pi_{ij}(s)} = \frac{\partial}{\partial \pi_{ij}(s)} \exp \left(\sum_{i=1}^p \sum_{s=0}^{c_i-1} x_{ih}(s) \log(\pi_{ij}(s)) \right) = \frac{g(\mathbf{x}_h | j) x_{ih}(s)}{\pi_{ij}(s)}$$

se tiene que

$$\frac{\partial Lg}{\partial \pi_{ij}(s)} = \eta_j \sum_{h=1}^n \frac{g(\mathbf{x}_h | j) x_{ih}(s)}{f(\mathbf{x}_h) \pi_{ij}(s)} + \phi_{ij} = \sum_{h=1}^n \frac{h(j | \mathbf{x}_h) x_{ih}(s)}{\pi_{ij}(s)} + \phi_{ij}.$$

Igualando a 0

$$0 = \sum_{h=1}^n h(j | \mathbf{x}_h) x_{ih}(s) + \pi_{ij}(s) \phi_{ij} \quad (2.19)$$

y sumando ambos lados sobre s obtendremos que

$$0 = \sum_{h=1}^n h(j | \mathbf{x}_h) + \phi_{ij}$$

y

$$\phi_{ij} = - \sum_{h=1}^n h(j | \mathbf{x}_h).$$

Así, substituyendo en (2.19) se obtienen las ecuaciones de estimación.

$$\pi_{ij}(s) = \frac{\sum_{h=1}^n h(j | \mathbf{x}_h) x_{ih}(s)}{\sum_{h=1}^n h(j | \mathbf{x}_h)} = \frac{\sum_{h=1}^n h(j | \mathbf{x}_h) x_{ih}(s)}{n \hat{\eta}_j}. \quad (2.20)$$

En el caso que las variables manifiestas sean politómicas ordenadas, (2.20) puede escribirse como:

$$\Pi_{ij}(s) = \hat{\Pi}_{ij}(s+1) + \frac{\sum_{h=1}^n x_{ih}(s)h(j|\mathbf{x}_h)}{n\hat{\eta}_j} \quad (2.21)$$

Así (2.18), (2.20) ó (2.21), pueden resolverse con el algoritmo EM antes expuesto para el caso binario:

1. Elegir una distribución inicial para las probabilidades a posteriori $h(j|\mathbf{x}_h)$.
2. Usar (2.18) y (2.20) ó (2.21) para obtener una primera aproximación de los estimadores de máxima verosimilitud para los $\hat{\eta}_j$ y $\hat{\pi}_{ij}$.
3. Substituir las estimaciones anteriores en (2.8) ó (2.10) para mejorar la distribución de las probabilidades a posteriori $h(j|\mathbf{x}_h)$.
4. Retomar el paso 2 para obtener una segunda aproximación de los parámetros y continuar el ciclo hasta alcanzar la convergencia.



Capítulo 3

Modelos de regresión de clases latentes

El modelo de regresión de clases latentes es una extensión del análisis de clases latentes, el cual permite la inclusión de covariables en la predicción de la pertenencia a estas clases. Mientras que en el análisis de clases latentes la probabilidad de pertenencia a una clase estaba solo definida por el patrón de respuestas de las variables manifiestas, \mathbf{X} ; en los modelos de regresión de clases latentes, esta prevalencia varía por individuo, en función de sus covariables observadas.

Sea $\mathbf{Z} = (Z_1, Z_2, \dots, Z_m)'$ un vector columna m -dimensional de covariables. Por el teorema de probabilidad total y la asunción de independencia condicional entre las variables manifiestas y que extenderemos a las covariables se tiene que:

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}) = \sum_{j=0}^{K-1} P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}, Y = j) P(Y = j | \mathbf{Z} = \mathbf{z}),$$

donde

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}, Y = j) &= \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, Y = j)}{P(\mathbf{Z} = \mathbf{z}, Y = j)} \\ &= \frac{P(\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} | Y = j) P(Y = j)}{P(\mathbf{Z} = \mathbf{z} | Y = j) P(Y = j)} = \frac{P(\mathbf{X} = \mathbf{x} | Y = j) P(\mathbf{Z} = \mathbf{z} | Y = j) P(Y = j)}{P(\mathbf{Z} = \mathbf{z} | Y = j) P(Y = j)} \\ &= P(\mathbf{X} = \mathbf{x} | Y = j) \end{aligned} \quad (3.1)$$

De esta manera, la probabilidad de tener un vector de respuestas en las variables manifiestas dadas las covariables es:

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}) = \sum_{j=0}^{K-1} P(\mathbf{X} = \mathbf{x} | Y = j) \eta_j(\mathbf{z}),$$

donde

$$\eta_j(\mathbf{z}) = P(Y = j | \mathbf{Z} = \mathbf{z}).$$

Se observan dos cosas: primero, que las covariables no se relacionan con las probabilidades de respuesta a las variables manifiestas al interior de cada clase latente, y segundo, que esta ecuación es bastante similar a la presentada en (2.5), con la diferencia que el parámetro

que indica la probabilidad de pertenencia a la clase se define en función de las covariables. De esta manera, una extensión de regresión del modelo de clase latente para una variable concomitante, propuesta por [Dayton and Macready \(1988\)](#) y citado por [Bandeem-Roche et al. \(1997\)](#) para el caso de variables manifiestas binarias, plantea que:

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}) = \sum_{j=0}^{K-1} \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i} \eta_j(\mathbf{z}) \quad (3.2)$$

donde $\pi_{ij} = P(X_i = 1 | Y = j)$ y $\eta_j(\mathbf{z})$ viene dado por un modelo logístico multinomial con categoría de referencia 0, la cual se expresa como:

$$\eta_j(\mathbf{z}) = \frac{e^{\beta_{0j} + \sum_{k=1}^m \beta_{kj} z_k}}{1 + \sum_{l=1}^{K-1} e^{\beta_{0l} + \sum_{k=1}^m \beta_{kl} z_k}} = \frac{e^{\mathbf{z}' \boldsymbol{\beta}_j}}{1 + \sum_{l=1}^{K-1} e^{\mathbf{z}' \boldsymbol{\beta}_l}}, \quad j = 0, 1, 2, \dots, K-1 \quad (3.3)$$

donde $\mathbf{z}' = [1, \mathbf{z}']$, $\boldsymbol{\beta}_0 = 0$ y $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \beta_{2j}, \dots, \beta_{mj})'$ es el vector de coeficientes correspondientes a la clase latente j . Si tenemos m covariables, el vector $\boldsymbol{\beta}_j$ tendrá tamaño $m + 1$; un coeficiente por cada covariable más un intercepto.

La ecuación (3.3) representa un submodelo donde la relación condicional entre covariables y la pertenencia a una clase latente, se basa en el parámetro $\boldsymbol{\beta}_j$. De esta manera, en el modelo de regresión de clases latentes los parámetros a estimar son los $K - 1$ vectores de coeficientes $\boldsymbol{\beta}_j$ y las probabilidades de una respuesta positiva condicional de cada variable manifiesta i a cada clase latente j , π_{ij} .

Las distribuciones a posteriori se definen según (3.1) mediante:

$$\begin{aligned} P(Y = j | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) &= \frac{P(Y = j, \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})}{P(\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})} \\ &= \frac{P(\mathbf{X} = \mathbf{x} | Y = j, \mathbf{Z} = \mathbf{z}) P(Y = j, \mathbf{Z} = \mathbf{z})}{P(\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})} \\ &= \frac{P(\mathbf{X} = \mathbf{x} | Y = j) P(Y = j | \mathbf{Z} = \mathbf{z}) P(\mathbf{Z} = \mathbf{z})}{P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}) P(\mathbf{Z} = \mathbf{z})} \\ &= \frac{P(\mathbf{X} = \mathbf{x} | Y = j) P(Y = j | \mathbf{Z} = \mathbf{z})}{\sum_{l=0}^{K-1} P(\mathbf{X} = \mathbf{x} | Y = l) P(Y = l | \mathbf{Z} = \mathbf{z})}. \end{aligned}$$

De esta manera, la distribución a posteriori de clasificación en el caso binario viene dada por:

$$h(j|x, z) = P(Y = j | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \frac{\eta_j(\mathbf{z}) \prod_{i=1}^p \pi_{ij}^{x_i} (1 - \pi_{ij})^{1-x_i}}{\sum_{l=0}^{K-1} \eta_l(\mathbf{z}) \prod_{i=1}^p \pi_{il}^{x_i} (1 - \pi_{il})^{1-x_i}}. \quad (3.4)$$

3.1. Extensión para variables manifiestas politómicas

La extensión para el caso en que las variables manifiestas sean politómicas se da de manera similar a lo visto en las Sección 2.2.1.

En el caso de variables politómicas nominales con $c_i \geq 2$ categorías para la variable

manifiesta i , podríamos definir en cada posible categoría $s = 1, 2, \dots, c_i - 1$ las variables binarias:

$$X_i(s) = \begin{cases} 1, & \text{si la respuesta de un individuo a la variable manifiesta } i \text{ está en la categoría } s, \\ 0, & \text{en caso contrario} \end{cases}$$

de tal manera que $\sum_{s=1}^{c_i-1} X_i(s) = 1$. Así, si identificamos la respuesta de un individuo a la variable manifiesta i con el vector columna $\mathbf{X}_i = [X_i(0), \dots, X_i(c_i-1)]'$, el vector de respuestas de un individuo se podrá representar mediante el vector columna, $\mathbf{X}' = [\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_p]$.

Si definimos por $\pi_{ij}(s)$ a la probabilidad de que un individuo en la clase j de una respuesta positiva en la categoría s de la variable i ; es decir, $\pi_{ij}(s) = P(X_i(s) = 1 | Y = j)$, entonces, la función de probabilidad conjunta del vector de respuestas manifiestas o patrones de respuesta \mathbf{x} , dadas las covariables, viene dada por:

$$P(\mathbf{X} = \mathbf{x} | \mathbf{Z} = \mathbf{z}) = \sum_{j=0}^{K-1} \eta_j(\mathbf{z}) \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)} \quad (3.5)$$

y su distribución a posteriori es

$$P(Y = j | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \frac{\eta_j(\mathbf{z}) \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)}}{\sum_{l=0}^{K-1} \eta_l(\mathbf{z}) \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{il}(s)^{x_i(s)}}. \quad (3.6)$$

En los modelos de regresión de clases latentes el reto es decidir el número de clases a considerar, estimar los parámetros del modelo e identificar el efecto que genere el ingreso de las covariables a las probabilidades de pertenencia a una clase. A continuación veremos cada aspecto por separado.

3.2. Interpretación de los coeficientes de regresión

El modelo de regresión no lineal planteado en la ecuación (3.3) fue introducido con el fin de poder explicar las probabilidades de pertenencia a una clase latente en términos de un conjunto de covariables. Para ello asumimos un esquema multinomial con enlace logístico, en virtud de que este enlace nos permite precisamente una mejor interpretación de los efectos que ejercen las covariables en las probabilidades de pertenencia. Mas concretamente, la componente β_{kj} asociada al efecto que ejerce la k -ésima covariable z_k del vector de covariables $\tilde{\mathbf{z}} = (1, z_1, \dots, z_m)'$ en la pertenencia a la clase j se interpreta como el cambio en proporción de los odds (o cocientes entre la probabilidad de pertenencia a la clase j con respecto a la clase de referencia 0) cuando la k -ésima covariable se incrementa en una unidad y las demas covariables se mantienen constantes. Más formalmente se tiene que

$$\frac{\eta_j(\mathbf{z}_k^+) / \eta_0(\mathbf{z}_k^+)}{\eta_j(\mathbf{z}) / \eta_0(\mathbf{z})} = \frac{P(Y=j | \mathbf{z}_k^+) / P(Y=0 | \mathbf{z}_k^+)}{P(Y=j | \mathbf{z}) / P(Y=0 | \mathbf{z})} = e^{\beta_{kj}}, \quad (3.7)$$

donde $\mathbf{z}_k^+ = (1, z_1, \dots, z_{k-1}, z_k + 1, z_{k+1}, \dots, z_m)'$.

Note en particular, que si z_k es una variable dicotómica, entonces $(e^{\beta_{kj}} - 1) * 100\%$

representa el cambio (aumento si es positivo o disminución si es negativo) en el odds de pertenecer a la clase j con relación a la clase de referencia cuando la covariable z_k pasa del valor 0 al 1 y las demás covariables se mantienen constantes.

3.3. Marginalización del modelo de regresión de clases latentes

Es interesante observar que de considerarse a las covariables en la ecuación (3.5) como aleatorias, la distribución conjunta marginal de las variables manifiestas posee la misma estructura que se había expuesto en el análisis de clases latentes (2.7). En efecto, ello se sigue desde que:

$$\begin{aligned}
 P(\mathbf{X} = \mathbf{x}) &= \int P(\mathbf{X} = \mathbf{x} \mid \mathbf{Z} = \mathbf{z}) dG(\mathbf{z}) = \sum_{j=0}^{K-1} \left[\int \eta_j(\mathbf{z}) dG(\mathbf{z}) \right] \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)} \\
 &= \sum_{j=0}^{K-1} \eta_j^* \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_i(s)} \tag{3.8}
 \end{aligned}$$

donde los $\eta_j^* = \int \eta_j(\mathbf{z}) dG(\mathbf{z})$ representan la probabilidad media de pertenencia a la clase latente j de acuerdo a la función de distribución G del vector de covariables \mathbf{Z} y cumplen que $\sum_{j=0}^{K-1} \eta_j^* = 1$.

La propiedad de marginalización (3.8) permite calcular entonces el número de clases en un modelo de regresión de clases latentes utilizando los mismos métodos del análisis de clases latentes, evitando de esta manera las complejidades que implican la inserción de las covariables.

3.4. Problema de Identificabilidad

Un posible problema que se puede dar en los modelos de clases latentes es la no identificabilidad local de los parámetros, donde diremos que el modelo es localmente identificable en un valor particular θ^* del vector de parámetros si la función de log-verosimilitud es únicamente determinada por los parámetros en una vecindad de θ^* (Vermunt and Magidson (2013), Chung (2003)). Una condición necesaria y suficiente para la identificabilidad local es que la matriz hessiana de la función de log-verosimilitud en θ^* tenga un rango de columna completa. Bandeen-Roche et al. (1997) extienden esta condición al modelo de clase latente con covariables usando la propiedad de marginalización y demuestran que si

1. El modelo marginal (3.8) es localmente identificable.
2. La matriz de covariables para los n individuos $(\mathbf{z}_1, \dots, \mathbf{z}_n)'$ tiene un rango columna completo.
3. Existe al menos un individuo cuya distribución de probabilidades de pertenencia a las clases no tenga entradas nulas.

Entonces el modelo de regresión de clase latente es localmente identificable.

3.5. Estimación de parámetros

3.5.1. Estimación de parámetros para el caso dicotómico

El modelo de regresión de clases latentes puede estimarse usándose el principio de máxima verosimilitud. La función de verosimilitud para el caso dicotómico viene dada por:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{h=1}^n P(\mathbf{X} = \mathbf{x}_h | \mathbf{Z} = \mathbf{z}_h) = \prod_{h=1}^n \sum_{j=0}^{K-1} P(X = x_h | Y = j) \eta_j(\mathbf{z}_h) \\ &= \prod_{h=1}^n \sum_{j=0}^{K-1} \eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \end{aligned} \quad (3.9)$$

donde el vector de parámetros a estimar es:

$$\boldsymbol{\theta} = (\pi_{10}, \pi_{11}, \dots, \pi_{p(K-1)}, \boldsymbol{\beta}) = (\pi_{10}, \pi_{11}, \dots, \pi_{p(K-1)}, \boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_{K-1})$$

El logaritmo de la función de máxima verosimilitud viene entonces dado por:

$$\mathcal{L} = \sum_{h=1}^n \log \left(\sum_{j=0}^{K-1} \eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \right) \quad (3.10)$$

Las derivadas parciales de la función de log-verosimilitud con respecto a los π_{ij} y los $\boldsymbol{\beta}_j$ se pueden derivar fácilmente, pero la ecuación resultante es no lineal; por ello, la estimación debe resolverse numéricamente. Dentro de los procedimientos iterativos que usan derivadas parciales, el algoritmo de Newton-Rhapson (basado en las primeras y segundas derivadas parciales del logaritmo de la función de máxima verosimilitud) es especialmente atractivo, ya que genera muestras estimadas para las varianzas y covarianzas como subproducto de las estimaciones (Dayton and Macready, 1988).

Para estimar la probabilidad de pertenencia a una clase en función de las covariables de un individuo h , usaremos como en (3.3) el modelo logístico multinomial:

$$\eta_j(\mathbf{z}_h) = \frac{e^{\mathbf{z}'_h \boldsymbol{\beta}_j}}{1 + \sum_{l=1}^{K-1} e^{\mathbf{z}'_h \boldsymbol{\beta}_l}} = \frac{e^{\beta_{0j} + \sum_{k=1}^m \beta_{kj} z_{kh}}}{1 + \sum_{l=1}^{K-1} e^{\beta_{0l} + \sum_{k=1}^m \beta_{kl} z_{lh}}} \quad (3.11)$$

En base a las ecuaciones (3.2) y (3.11), se obtiene la siguiente función de score (vector de derivadas parciales de la log-verosimilitud), expresada en términos de las probabilidades a posteriori $h(j | \mathbf{x}, \mathbf{z}) = P(Y = j | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z})$ definidas en (3.4).

Proposición 3.1

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}} = \sum_{h=1}^n \frac{h(j | \mathbf{x}_h, \mathbf{z}_h) (x_{ih} - \pi_{ij})}{\pi_{ij} (1 - \pi_{ij})} \quad (3.12)$$

y

$$\frac{\partial \mathcal{L}}{\partial \beta_{kj}} = \sum_{h=1}^n (h(j | \mathbf{x}_h, \mathbf{z}_h) - \eta_j(\mathbf{z}_h)) z_{kh}, \quad (3.13)$$

para $i = 1, 2, \dots, p$; $j = 1, 2, \dots, K - 1$ y $k = 0, 1, \dots, m$.

Demostración: Puesto que en la función de log-verosimilitud

$$\mathcal{L} = \sum_{h=1}^n \log \left(\sum_{j=0}^{K-1} \eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \right) = \sum_{h=1}^n \log(A_h)$$

el término

$$A_h = \sum_{j=0}^{K-1} \eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}},$$

puede descomponerse como:

$$A_h = \sum_{\substack{l=0 \\ l \neq j}}^{K-1} \eta_l(\mathbf{z}_h) \prod_{i=1}^p \pi_{il}^{x_{ih}} (1 - \pi_{il})^{1-x_{ih}} + \eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \prod_{\substack{m=1 \\ m \neq i}}^p \pi_{mj}^{x_{mh}} (1 - \pi_{mj})^{1-x_{mh}},$$

se tiene que

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \pi_{ij}} &= \sum_{h=1}^n \frac{1}{A_h} \eta_j(\mathbf{z}_h) \frac{x_{ih} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})} \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \prod_{\substack{m=1 \\ m \neq i}}^p \pi_{mj}^{x_{mh}} (1 - \pi_{mj})^{1-x_{mh}} \\ &= \sum_{h=1}^n \frac{\eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}}{\sum_{j=0}^{K-1} \eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}} \times \frac{x_{ih} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}, \end{aligned}$$

lo cual nos conduce a (3.12).

De otro lado,

$$\frac{\partial \mathcal{L}}{\partial \beta_{kj}} = \sum_{h=1}^n \frac{1}{A_h} \sum_{\substack{l=0 \\ l \neq j}}^{K-1} \prod_{i=1}^p \pi_{il}^{x_{ih}} (1 - \pi_{il})^{1-x_{ih}} \frac{\partial}{\partial \beta_{kj}} (\eta_l(\mathbf{z}_h)) + \sum_{h=1}^n \frac{1}{A_h} \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}} \frac{\partial}{\partial \beta_{kj}} (\eta_j(\mathbf{z}_h)),$$

donde

$$\begin{aligned} \frac{\partial}{\partial \beta_{kj}} (\eta_j(\mathbf{z}_h)) &= \frac{\partial}{\partial \beta_{kj}} \left(\frac{e^{\mathbf{z}'_h \boldsymbol{\beta}_j}}{1 + \sum_{l=1}^{K-1} e^{\mathbf{z}'_h \boldsymbol{\beta}_l}} \right) = \frac{e^{\mathbf{z}'_h \boldsymbol{\beta}_j} z_{kh} (1 + \sum_{l=1}^{K-1} e^{\mathbf{z}'_h \boldsymbol{\beta}_l}) - (e^{\mathbf{z}'_h \boldsymbol{\beta}_j}) (e^{\mathbf{z}'_h \boldsymbol{\beta}_j}) z_{kh}}{(1 + \sum_{m=1}^{K-1} e^{\mathbf{z}'_h \boldsymbol{\beta}_m})^2} \\ &= (1 - \eta_j(\mathbf{z}_h)) \eta_j(\mathbf{z}_h) z_{kh} = -\eta_j(\mathbf{z}_h)^2 z_{kh} + \eta_j(\mathbf{z}_h) z_{kh} \end{aligned}$$

y

$$\frac{\partial}{\partial \beta_{kj}} (\eta_l(\mathbf{z}_h)) = \frac{\partial}{\partial \beta_{kj}} \left(\frac{e^{\mathbf{z}'_h \boldsymbol{\beta}_l}}{1 + \sum_{m=1}^{K-1} e^{\mathbf{z}'_h \boldsymbol{\beta}_m}} \right) = \frac{e^{\mathbf{z}'_h \boldsymbol{\beta}_l} e^{\mathbf{z}'_h \boldsymbol{\beta}_j} z_{kh}}{(1 + \sum_{l=1}^{K-1} e^{\mathbf{z}'_h \boldsymbol{\beta}_j})^2} = -\eta_l(\mathbf{z}_h) \eta_j(\mathbf{z}_h) z_{kh},$$

si $j \neq l$.

Así,

$$\frac{\partial \mathcal{L}}{\partial \beta_{kj}} = - \sum_{h=1}^n \frac{1}{A_h} \sum_{l=0}^{K-1} \prod_{i=1}^p \pi_{il}^{x_{ih}} (1 - \pi_{il})^{1-x_{ih}} \eta_l(\mathbf{z}_h) \eta_j(\mathbf{z}_h) z_{kh}$$

$$\begin{aligned}
 & + \sum_{h=1}^n \frac{1}{A_h} \prod_{i=1}^p \pi_{il}^{x_{ih}} (1 - \pi_{il})^{1-x_{ih}} \eta_j(\mathbf{z}_h) z_{kh} \\
 & = \sum_{h=1}^n h(j | \mathbf{x}_h, \mathbf{z}_h) z_{kh} - \sum_{h=1}^n \eta_j(\mathbf{z}_h) z_{kh} \left(\sum_{l=0}^{K-1} h(l | \mathbf{x}_h, \mathbf{z}_h) \right) = \sum_{h=1}^n (h(j | \mathbf{x}_h, \mathbf{z}_h) - \eta_j(\mathbf{z}_h)) z_{kh},
 \end{aligned}$$

que es precisamente la fórmula (3.13). \square

Note que igualando (3.12) a 0, que corresponde a una de las condiciones de primer orden para encontrar los estimadores de máxima verosimilitud, uno puede despejar los π_{ij} en términos de las probabilidades de clasificación a posteriori (3.4) como:

$$\pi_{ij} = \frac{\sum_{h=1}^n x_{ih} h(j | \mathbf{x}_h, \mathbf{z}_h)}{\sum_{h=1}^n h(j | \mathbf{x}_h, \mathbf{z}_h)} \quad (3.14)$$

Al igual que en la Sección 2.2.3, estas probabilidades serían simples de estimar si se conociera la función $h(j | \mathbf{x}_h, \mathbf{z}_h)$.

Como se menciono anteriormente, para hallar los valores estimados de θ , se puede usar un algoritmo modificado de Esperanza-Maximización (EM) con un paso de Newton-Rhapson. Dado que este algoritmo requiere del cálculo de la matriz Hessiana de \mathcal{L} , el cual será también utilizado para la determinación de los errores estándares, presentaremos seguidamente la estructura de tal matriz.

Proposición 3.2 *Las derivadas parciales de segundo orden de la función log-verosimilitud (3.10) vienen dadas por*

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \pi_{ij}} = \sum_{h=1}^n \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} (1 - \delta_{ai}) - h(j | \mathbf{x}_h, \mathbf{z}_h)) (x_{ah} - \pi_{aq}) (x_{ih} - \pi_{ij})}{\pi_{aq} (1 - \pi_{aq}) \pi_{ij} (1 - \pi_{ij})} \quad (3.15)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \beta_{kj}} = \sum_{h=1}^n \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)) z_{kh} (x_{ah} - \pi_{aq})}{\pi_{aq} (1 - \pi_{aq})}. \quad (3.16)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_{rq} \beta_{kj}} = \sum_{h=1}^n (h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)) - \eta_q(\mathbf{z}_h) (\delta_{qj} - \eta_j(\mathbf{z}_h))) z_{rh} z_{kh}, \quad (3.17)$$

donde $a, i = 1, 2, \dots, p$; $q, j = 0, 1, \dots, K - 1$ y $r, k = 0, 1, \dots, m$, siendo los índices q y j distintos de 0 para el caso de los coeficientes de regresión.

Demostración: Al igual que en la demostración de la Proposición 3.1, consideremos el término A_h que aparece en el argumento de los logaritmos de la función de log-verosimilitud. Como se calculó allí, se cumple que

$$\frac{\partial}{\partial \pi_{ij}} (\eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}) = \frac{\partial A_h}{\partial \pi_{ij}} = \eta_j(\mathbf{z}_h) \frac{x_{ih} - \pi_{ij}}{\pi_{ij} (1 - \pi_{ij})} \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}},$$

lo cual implica que

$$\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{ij}} = \frac{\partial}{\partial \pi_{ij}} \left(\frac{\eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}}{A_h} \right) = \frac{h(j | \mathbf{x}_h, \mathbf{z}_h) (1 - h(j | \mathbf{x}_h, \mathbf{z}_h)) (x_{ih} - \pi_{ij})}{\pi_{ij} (1 - \pi_{ij})}$$

y

$$\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{aq}} = \frac{\partial}{\partial \pi_{aq}} \left(\frac{\eta_j(\mathbf{z}_h) \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}}{A_h} \right) = - \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) h(j | \mathbf{x}_h, \mathbf{z}_h) (x_{ah} - \pi_{aq})}{\pi_{aq} (1 - \pi_{aq})},$$

cuando $q \neq j$. En resumen:

$$\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{aq}} = \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)) (x_{ah} - \pi_{aq})}{\pi_{aq} (1 - \pi_{aq})}. \quad (3.18)$$

Así, si $a \neq i$ se cumple que

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \partial \pi_{ij}} = \sum_{h=1}^n \frac{\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{aq}} (x_{ih} - \pi_{ij})}{\pi_{ij} (1 - \pi_{ij})} = \sum_{h=1}^n \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)) (x_{ah} - \pi_{aq}) (x_{ih} - \pi_{ij})}{\pi_{aq} (1 - \pi_{aq}) \pi_{ij} (1 - \pi_{ij})}.$$

En caso contrario, si $a = i$ y $q \neq j$,

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{iq} \partial \pi_{ij}} = \sum_{h=1}^n \frac{\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{iq}} (x_{ih} - \pi_{ij})}{\pi_{ij} (1 - \pi_{ij})} = - \sum_{h=1}^n \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) h(j | \mathbf{x}_h, \mathbf{z}_h) (x_{ih} - \pi_{iq}) (x_{ih} - \pi_{ij})}{\pi_{iq} (1 - \pi_{iq}) \pi_{ij} (1 - \pi_{ij})},$$

mientras que para $a = i$ y $q = j$ se tiene que el siguiente desarrollo:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \pi_{ij}^2} &= \sum_{h=1}^n \frac{(x_{ih} - \pi_{ij})^2 h(j) (1 - h(j)) - h(j) \pi_{ij} (1 - \pi_{ij}) - h(j) (x_{ih} - \pi_{ij}) (1 - 2\pi_{ij})}{\pi_{ij}^2 (1 - \pi_{ij})^2} \\ &= - \sum_{h=1}^n \left(\frac{h(j)^2 (x_{ih} - \pi_{ij})^2}{\pi_{ij}^2 (1 - \pi_{ij})^2} - \frac{h(j) (x_{ih}^2 - x_{ih})}{\pi_{ij}^2 (1 - \pi_{ij})^2} \right), \end{aligned}$$

donde para abreviar hemos escrito $h(j) = h(j | \mathbf{x}_h, \mathbf{z}_h)$. Dado que nuestras variables manifiestas son dicotómicas, esta expresión finalmente se reduce a

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{ij}^2} = - \sum_{h=1}^n \frac{h(j | \mathbf{x}_h, \mathbf{z}_h)^2 (x_{ih} - \pi_{ij})^2}{\pi_{ij}^2 (1 - \pi_{ij})^2},$$

con lo que la fórmula (3.15) resume los casos aquí considerados.

De otro lado, por (3.18) se tiene que

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \partial \beta_{kj}} = \sum_{h=1}^n \frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{aq}} z_{kh} = \sum_{h=1}^n \frac{h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)) z_{kh} (x_{ah} - \pi_{aq})}{\pi_{aq} (1 - \pi_{aq})}.$$

Finalmente,

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_{rq} \partial \beta_{kj}} = \sum_{h=1}^n \left(\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \beta_{rq}} - \frac{\partial \eta_j(\mathbf{z}_h)}{\partial \beta_{rq}} \right) z_{kh},$$

donde

$$\frac{\partial \eta_j(\mathbf{z}_h)}{\partial \beta_{rq}} = (\delta_{qj} - \eta_j(\mathbf{z}_h)) \eta_q(\mathbf{z}_h) z_{rh}$$

y

$$\frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \beta_{rq}} = \frac{\partial}{\partial \beta_{rq}} \left(\frac{\eta_j(\mathbf{z}_h) \varrho_{jh}}{A_h} \right) = \frac{\frac{\partial \eta_j(\mathbf{z}_h)}{\partial \beta_{rq}} \varrho_{jh} A_h - \eta_j(\mathbf{z}_h) \varrho_{jh} \frac{\partial A_h}{\partial \beta_{rq}}}{A_h^2},$$

siendo $\varrho_{jh} = \prod_{i=1}^p \pi_{ij}^{x_{ih}} (1 - \pi_{ij})^{1-x_{ih}}$. Más aún, por lo visto en la demostración de la proposición 3.1

$$\begin{aligned} \frac{\partial A_h}{\partial \beta_{rq}} &= \sum_{\substack{l=0 \\ l \neq q}}^{K-1} \frac{\partial \eta_l(\mathbf{z}_h)}{\partial \beta_{rq}} \varrho_{lh} + \frac{\partial \eta_q(\mathbf{z}_h)}{\partial \beta_{rq}} \varrho_{qh} = - \sum_{\substack{l=0 \\ l \neq q}}^{K-1} \eta_l(\mathbf{z}_h) \eta_q(\mathbf{z}_h) \varrho_{lh} z_{rh} - (\eta_q(\mathbf{z}_h)^2 - \eta_q(\mathbf{z}_h)) \varrho_{qh} z_{rh} \\ &= \left(-\eta_q(\mathbf{z}_h) \sum_{l=0}^{K-1} \eta_l(\mathbf{z}_h) \varrho_{lh} + \eta_q(\mathbf{z}_h) \varrho_{qh} \right) z_{rh} = \eta_q(\mathbf{z}_h) (\varrho_{qh} - A_h) z_{rh}. \end{aligned}$$

Así,

$$\begin{aligned} \frac{\partial h(j | \mathbf{x}_h, \mathbf{z}_h)}{\partial \beta_{rq}} &= \left(\frac{(\delta_{qj} - \eta_j(\mathbf{z}_h)) \eta_q(\mathbf{z}_h) \varrho_{jh} A_h - \eta_j(\mathbf{z}_h) \varrho_{jh} \eta_q(\mathbf{z}_h) (\varrho_{qh} - A_h)}{A_h^2} \right) z_{rh} \\ &= (\delta_{qj} - \eta_j(\mathbf{z}_h)) h(q | \mathbf{x}_h, \mathbf{z}_h) - h(q | \mathbf{x}_h, \mathbf{z}_h) (h(j | \mathbf{x}_h, \mathbf{z}_h) - \eta_j(\mathbf{z}_h)) = h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)). \end{aligned}$$

y

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_{rq} \partial \beta_{kj}} = \sum_{h=1}^n (h(q | \mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j | \mathbf{x}_h, \mathbf{z}_h)) - \eta_q(\mathbf{z}_h) (\delta_{qj} - \eta_j(\mathbf{z}_h)) z_{rh} z_{kh}).$$

□

Conocidas entonces las derivadas parciales de primer y segundo orden de la función de log-verosimilitud, el algoritmo antes comentado para la estimar $\boldsymbol{\theta}$ se resume como:

1. Elegir una estimación inicial $\hat{\boldsymbol{\theta}}_0$ para $\boldsymbol{\theta}$.
2. Usar (3.4) para calcular $h(j | \mathbf{x}, \mathbf{z})$ con la última estimación disponible de $\boldsymbol{\theta}$.
3. Maximizar actualizando los coeficientes de regresión de acuerdo a la fórmula:

$$\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_{t-1} + (-D_{\boldsymbol{\beta}}^2 \mathcal{L}(\hat{\boldsymbol{\beta}}_{t-1}))^{-1} \nabla_{\boldsymbol{\beta}} \mathcal{L}(\hat{\boldsymbol{\beta}}_{t-1}), \quad (3.19)$$

donde $\nabla_{\boldsymbol{\beta}} \mathcal{L}(\hat{\boldsymbol{\beta}}_{t-1})$ es el gradiente y $D_{\boldsymbol{\beta}}^2 \mathcal{L}(\hat{\boldsymbol{\beta}}_{t-1})$ la matriz Hessiana de la función de log-verosimilitud \mathcal{L} con respecto al vector de parámetros de regresión $\boldsymbol{\beta}$ estimado en el paso $t - 1$ anterior.

4. Actualizar los $\hat{\pi}_{ij}$ mediante (3.14) con el $h(j | \mathbf{x}, \mathbf{z})$ estimado bajo el último $\boldsymbol{\beta}$ estimado y los $\hat{\pi}_{ij}$ previos.
5. Repetir los pasos 2, 3 y 4 hasta alcanzar la convergencia.

3.5.2. Estimación de parámetros para el caso politómico

En el caso de que las variables manifiestas sean politómicas, la función de log-verosimilitud viene dada por:

$$\mathcal{L} = \sum_{h=1}^n \log \left(\sum_{j=0}^{K-1} \eta_j(\mathbf{z}_h) \prod_{i=1}^p \prod_{s=1}^{c_i-1} \pi_{ij}(s)^{x_{ih}(s)} \pi_{ij}(c_i)^{x_{ih}(c_i)} \right), \quad (3.20)$$

donde el término $\pi_{ij}(c_i)$ se calcula mediante

$$\pi_{ij}(c_i) = 1 - \sum_{s=1}^{c_i-1} \pi_{ij}(s) \quad (3.21)$$

dada la restricción de que las probabilidades de que la variable manifiesta i condicionada a la clase latente j debe de tomar siempre alguno de sus c_i posibles valores.

Una manera de tratar la estimación de los parámetros en el caso de las variables manifiestas politómicas podría ser dicotomizando las distintas categorías de las variables manifiestas. Esta estrategia, sin embargo no toma en cuenta la restricción (3.21).

Con el fin de maximizar la función de log-verosimilitud (3.20) será conveniente obtener tanto el vector gradiente como la matriz hessiana de esta función, de manera tal que la restricción (3.21) se incluya ya en la definición de esta función y ella se vea solo como función de los parámetros libres; en otras palabras buscaremos obtener las derivadas de la función de log-verosimilitud, que denotaremos por \mathcal{L}^- solo como función del vector de parámetros de probabilidades $\boldsymbol{\pi}_{ij} = [\pi_{ij}(1), \dots, \pi_{ij}(c_i - 1)]$ y no como función del vector de probabilidades completo $\boldsymbol{\pi}_{ij}^+ = [\pi_{ij}, \pi_{ij}(c_i)]$. La siguiente proposición nos brinda precisamente estas derivadas.

Proposición 3.3 *Para el caso politómico se cumple que:*

a) *El gradiente de la función de log-verosimilitud con respecto a las probabilidades de respuesta $\boldsymbol{\pi}_{ij}$ viene dado por:*

$$\nabla \mathcal{L}^- = G_i \nabla \mathcal{L},$$

donde G_i es una matriz $c_i - 1 \times c_i$ que contiene en las primeras $c_i - 1$ columnas a la matriz identidad de orden $c_i - 1$ y en la última al vector de -1 's. Más aún, la componente $s \in \{1, 2, \dots, c_i\}$ del vector gradiente de la log-verosimilitud completa no restringida $\nabla \mathcal{L}$ toma la forma:

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}(s)} = \sum_{h=1}^n \frac{h(j|\mathbf{x}_h, \mathbf{z}_h) x_{ih}(s)}{\pi_{ij}(s)},$$

para cualesquieras $i = 1, 2, \dots, p; j = 0, 1, \dots, K - 1$ y $s = 1, 2, \dots, c_i$.

b) *Las derivadas parciales con respecto al vector de probabilidades y coeficientes de regresión de la función de log-verosimilitud se caracterizan por*

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_{kj} \partial \pi_{aq}(s)} = \sum_{h=1}^n h(j|\mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(q|\mathbf{x}_h, \mathbf{z}_h)) \left(\frac{x_{ah}(s)}{\pi_{aq}(s)} - \frac{x_{ah}(c_a)}{\pi_{aq}(c_a)} \right) z_{kh},$$

para cualesquieras $i = 1, 2, \dots, p; j = 0, 1, \dots, K - 1$ y $s = 1, 2, \dots, c_i$.

c) *Las entradas de la matriz hessiana para la función de log-verosimilitud con respecto a los*

vectores de probabilidad π_{aq} y π_{ij} vienen dadas por:

$$\frac{\partial^2 \mathcal{L}^-}{\partial \pi_{aq} \pi_{ij}} = G_a \frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \pi_{ij}} G_i^\top,$$

donde la componentes (t, s) de la matriz hessiana $\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \pi_{ij}}$ de log-verosimilitud completa no restringida viene dada por:

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq}(t) \partial \pi_{ij}(s)} = \sum_{h=1}^n h(q|\mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j|\mathbf{x}_h, \mathbf{z}_h)) \frac{x_{ah}(t)}{\pi_{aq}(t)} \frac{x_{ih}(s)}{\pi_{ij}(s)} - h(j|\mathbf{x}_h, \mathbf{z}_h) \frac{x_{ih}(s)}{\pi_{ij}^2(s)} \delta_{st} \delta_{ia} \delta_{qj},$$

para cualesquieras $i, a = 1, 2, \dots, p; j, q = 0, 1, \dots, K-1; t = 1, 2, \dots, c_a$ y $s = 1, 2, \dots, c_i$.

Demostración: Comentaremos en primer lugar que la función de log-verosimilitud para el caso politómico tiene, en términos de los $\eta_j(\mathbf{z}_h)$, la misma forma funcional que en el caso dicotómico y por tanto las derivadas parciales de ella con respecto al vector de los parámetros de regresión β son exactamente las mismas a este caso. De otro lado, sea

$$\mathcal{L}^-(\pi_{ij}, \beta) = \mathcal{L}\left(\left(\pi_{ij}, 1 - \sum_{s=1}^{c_i-1} \pi_{ij}(s)\right), \beta\right)$$

la función de log-verosimilitud con respecto a los $c_i - 1$ parámetros libres de π_{ij} . Note que \mathcal{L}^- es la composición de la función de log-verosimilitud completa \mathcal{L} con la función $g(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n, 1 - \sum_{i=1}^n x_i)$ y por tanto podríamos aplicar la regla de la cadena a fin de obtener

$$\nabla \mathcal{L}^-(\pi_{ij}, \beta) = G_i \nabla \mathcal{L}(\pi_{ij}^+, \beta),$$

donde G_i es una matriz $c_i - 1 \times c_i$ que contiene en las primeras $c_i - 1$ columnas a la matriz identidad de orden $c_i - 1$ y en la última al vector columna de -1 's. De otro lado, las componente $s \in \{1, 2, \dots, c_i\}$ del último vector gradiente viene dado por:

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}(s)} = \sum_{h=1}^n \frac{1}{A_h} \frac{\partial A_h}{\partial \pi_{ij}(s)}$$

donde

$$A_h = \sum_{l=0}^{K-1} \eta_l(\mathbf{z}_h) \prod_{m=1}^p \prod_{u=1}^{c_m} \pi_{ml}(u)^{x_{mh}(u)} = \sum_{l=0, l \neq j}^{K-1} \eta_l(\mathbf{z}_h) \prod_{m=1}^p \prod_{u=1}^{c_m} \pi_{ml}(u)^{x_{mh}(u)} + \eta_j(\mathbf{z}_h) \alpha_{hj}$$

y

$$\alpha_{hj} = \prod_{m=1}^p \prod_{u=1}^{c_m} \pi_{mj}(u)^{x_{mh}(u)} = \prod_{m=1, m \neq i}^p \prod_{u=1}^{c_m} \pi_{mj}(u)^{x_{mh}(u)} \prod_{u=1, u \neq s}^{c_m} \pi_{ij}(u)^{x_{ih}(u)} \pi_{ij}(s)^{x_{ih}(s)}$$

Luego,

$$\frac{\partial A_h}{\partial \pi_{ij}(s)} = \frac{\eta_j(\mathbf{z}_h) \alpha_{hj} x_{ih}(s)}{\pi_{ij}(s)}$$

Más aún, dado que la distribución a posteriori viene dada por:

$$h(j|\mathbf{x}_h, \mathbf{z}_h) = \frac{\eta_j(\mathbf{z}_h) \prod_{i=1}^p \prod_{s=1}^{c_i} \pi_{ij}(s)^{x_{ih}(s)}}{A_h} = \frac{\eta_j(\mathbf{z}_h) \alpha_h}{A_h}$$

tenemos que

$$\frac{\partial \mathcal{L}}{\partial \pi_{ij}(s)} = \sum_{h=1}^n \frac{h(j|\mathbf{x}_h, \mathbf{z}_h) x_{ih}(s)}{\pi_{ij}(s)}.$$

Veamos ahora las derivadas cruzadas entre los coeficientes de regresión en β y las probabilidades de respuesta π_{ij} . Utilizando lo anteriormente desarrollado, se tiene que

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_{kj} \partial \pi_{aq}(s)} = \sum_{h=1}^n \frac{\partial h(q|\mathbf{x}_h, \mathbf{z}_h)}{\partial \beta_{kj}} \left(\frac{x_{ah}(s)}{\pi_{aq}(s)} - \frac{x_{ah}(c_a)}{\pi_{aq}(c_a)} \right)$$

Más aún similarmente a lo visto en la proposición 3.2 se tiene que

$$\frac{\partial h(q|\mathbf{x}_h, \mathbf{z}_h)}{\partial \beta_{kj}} = h(j|\mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(q|\mathbf{x}_h, \mathbf{z}_h)) z_{kh},$$

de donde se obtiene que

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_{kj} \partial \pi_{aq}(s)} = \sum_{h=1}^n h(j|\mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(q|\mathbf{x}_h, \mathbf{z}_h)) \left(\frac{x_{ah}(s)}{\pi_{aq}(s)} - \frac{x_{ah}(c_a)}{\pi_{aq}(c_a)} \right) z_{kh}.$$

Finalmente, una nueva aplicación de la regla de la cadena, nos provee de la siguiente matriz hessiana para la función de log-verosimilitud con respecto a los vectores de probabilidad π_{aq} y π_{ij}

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \pi_{ij}} = G_a \frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \pi_{ij}} G_a^\top,$$

donde la matriz hessiana $\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq} \pi_{ij}}$ de log-verosimilitud completa no restringida tiene componente $(t, s) \in \{1, 2, \dots, c_a\} \times \{1, 2, \dots, c_i\}$ a

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq}(t) \pi_{ij}(s)} = \sum_{h=1}^n \frac{\partial h(j|\mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{aq}(t)} \left(\frac{x_{ih}(s)}{\pi_{ij}(s)} \right) - h(j|\mathbf{x}_h, \mathbf{z}_h) \frac{x_{ih}(s)}{\pi_{ij}^2(s)} \delta_{st} \delta_{ia} \delta_{qj}.$$

No es difícil obtener por un argumento similar a lo desarrollado en la primera parte de esta demostración que:

$$\frac{\partial h(j|\mathbf{x}_h, \mathbf{z}_h)}{\partial \pi_{aq}(t)} = h(q|\mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j|\mathbf{x}_h, \mathbf{z}_h)) \frac{x_{ah}(t)}{\pi_{aq}(t)}.$$

Así, reemplazando en la expresión anterior:

$$\frac{\partial^2 \mathcal{L}}{\partial \pi_{aq}(t) \partial \pi_{ij}(s)} = \sum_{h=1}^n h(q|\mathbf{x}_h, \mathbf{z}_h) (\delta_{qj} - h(j|\mathbf{x}_h, \mathbf{z}_h)) \frac{x_{ah}(t)}{\pi_{aq}(t)} \frac{x_{ih}(s)}{\pi_{ij}(s)} - h(j|\mathbf{x}_h, \mathbf{z}_h) \frac{x_{ih}(s)}{\pi_{ij}^2(s)} \delta_{st} \delta_{ia} \delta_{qj}$$

□

3.6. Error estándar

Una de las limitaciones del algoritmo EM es que este no nos provee de los errores estándar de estimación. Como alternativa para ello utilizaremos la matriz de información de Fisher observada, la cual involucra el cálculo de la matriz Hessiana $D^2\mathcal{L}$ cuyas entradas fueron descritas en las proposiciones 3.1 y 3.2. Recordemos, con base en la teoría asintótica de los estimadores de máxima verosimilitud, que $(-D_{\theta}^2\mathcal{L}(\hat{\theta}))^{-1}$ es un estimador de la matriz de varianza-covarianza del estimador de máxima verosimilitud $\hat{\theta}$. Así, los errores estándar de las componentes de este estimador se pueden hallar simplemente como las raíces cuadradas de las entradas en la diagonal de esta matriz.

3.7. Selección del modelo

Los métodos más usados para encontrar el mejor balance entre una data de tamaño n , el ajuste del modelo y el número de parámetros a ser estimados, son el Bayesian Information Criterion (BIC) y el Akaike Information Criterion (AIC). Si Υ representa el número total de parámetros estimados y Λ el valor máximo de la función de log-verosimilitud, entonces:

$$AIC = -2\Lambda + 2\Upsilon$$

$$BIC = -2\Lambda + \Upsilon \log(n)$$

Tanto el AIC como el BIC son medidas de calidad del modelo estadístico, buscando un equilibrio entre la bondad de ajuste del modelo y su complejidad, penalizando el número de parámetros calculados y el número de datos. Es importante señalar que estos métodos no son pruebas estadísticas tan solo indicadores de cual de todos los modelos analizados es el mejor. De un conjunto de modelos analizados, el modelo deseado será aquel que tenga un menor valor de AIC y BIC. En el caso del BIC, el término de penalización por el aumento de parámetros es incluso mayor; de esta manera se busca controlar el sobreajuste.

Por otro lado, es posible construir otros indicadores de ajuste del modelo como el estadístico Chi-Cuadrado χ^2 de Pearson y el de la razón de verosimilitud (G^2).

Si q_x denota la frecuencia observada de casos para el patrón x tomado de cualesquiera de los $2^{\sum_{i=1}^p c_i}$ posibles patrones de respuesta a las p variables manifiestas, y \hat{Q}_x denota a la frecuencia esperada para este patrón; es decir,

$$\hat{Q}_x = n \sum_{j=0}^{K-1} \hat{\eta}_j \prod_{i=1}^p \prod_{s=1}^{c_i-1} \hat{\pi}_{ij}(s)^{x_{i(s)}} \quad (3.22)$$

estos indicadores estadísticos se definen como:

$$\chi^2 = \sum_x (q_x - \hat{Q}_x)^2 / \hat{Q}_x \quad (3.23)$$

$$G^2 = 2 \sum_x q_x \log(q_x / \hat{Q}_x) \quad (3.24)$$

Dado que el χ^2 de Pearson mide la diferencia entre la distribución observada y la teórica

dada por el modelo, el objetivo es seleccionar un modelo donde tanto χ^2 y G^2 se minimicen, lo cual implica un mayor ajuste.

3.8. Implementación computacional

Para la estimación del modelo de regresión de clases latentes usaremos el paquete poLCA, implementado en el programa estadístico R (Linzer and Lewis, 2011).

El paquete poLCA estima el modelo de clase latente utilizando la función presentada en (2.17) y la metodología descrita en el capítulo 2. poLCA aprovecha la naturaleza iterativa de los algoritmos para hacer posible la estimación del modelo, incluso cuando hay casos perdidos en las variables manifiestas.

De igual manera, poLCA incluye como salidas de su análisis a los indicadores de ajuste Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), y los estadísticos Chi-Cuadrado χ^2 de Pearson y de la razón de verosimilitud (G^2).

En el caso de la extensión e inclusión de covariables, como se ha visto en la parte teórica, el modelo selecciona arbitrariamente una clase latente que sirva como la clase referencial. El paquete poLCA selecciona la primera clase, y asume que los log-odds de la pertenencia a priori a una clase latente con respecto a la referencial son funciones lineales de las covariables.

La función de log-verosimilitud con la cual trabaja el paquete es la presentada en (3.20). Para hallar los valores de los $\hat{\beta}_j$ y $\hat{\pi}_{ij}$ que maximicen esta función, poLCA usa un algoritmo de Esperanza-Maximización (EM) modificado, con un paso de Newton-Rhaphson, el cual presentamos en la Sección 3.4.

Capítulo 4

Aplicación

En este Capítulo se desarrollará una aplicación del modelo de regresión de clases latentes a la data del Sistema de Seguimiento a Egresados de pregrado de una Universidad privada del país. Este sistema tiene como objetivo producir indicadores que permita a la universidad conocer la situación del egresado a dos y cinco años de su egreso, lo cual contribuya a mejorar la calidad de la formación que ofrece la universidad. Más específicamente la información del Sistema de Seguimiento a Egresados permite:

- Que las autoridades conozcan la proyección, las dificultades y los éxitos que encuentran sus estudiantes cuando terminan su etapa de formación y empiezan su vida profesional.
- Conocer los niveles de satisfacción que tienen los egresados respecto a la formación y servicios brindados por la Universidad y qué valoran cuando empiezan su ejercicio profesional.
- Reflexionar sobre los planes de estudios al interior de cada especialidad, alimentando el proceso de mejora continua.

Para la presente aplicación, se utilizarán los resultados de la encuesta aplicada a los egresados de las cohortes 2009, 2010, 2011 y 2012 que hayan pasado por los Estudios Generales de la universidad. A continuación se presenta las características de la metodología aplicada para el recojo de la información, así como el período de medición y tasa de respuesta institucional alcanzada en cada año.

4.1. Metodología del recojo de información

4.1.1. Método

- Alcance: censal, asumiendo como población objetivo a una cohorte de egresados

Cohorte	Ciclo de egreso	Año de medición	Encuestados	Tasa de respuesta institucional
2009	2009-1, 2009-2, 2010-0	2011	1173	75.9 %
2010	2010-1, 2010-2, 2011-0	2012	1239	80.7 %
2011	2011-1, 2011-2, 2012-0	2014	1348	80.6 %
2012	2012-1, 2012-2, 2013-0	2014	1378	81.1 %

Cuadro 4.1: Cohortes de egresados a analizar

- Instrumento: Cuestionario
- Método de entrevista: Encuestas telefónicas

4.1.2. Población de estudio

A continuación se presentan algunas definiciones.

- Egresado: alumno que ha cumplido con aprobar todos los cursos de su plan de estudios así como certificar su conocimiento del idioma inglés.
- Cohorte de egresados: grupo de alumnos que hayan egresado en los ciclos 1, 2 del año de referencia y el ciclo de verano (enero y febrero) del año siguiente.

Si bien esta es la definición oficial de egresado que hace la universidad, para seleccionar a la población se considera un parámetro más, el cual solo considera a los egresados que hayan certificado el conocimiento del idioma inglés máximo un año después de haber concluido su plan de estudios; de esta manera, se evitan los sesgos que pueden generar la presencia de egresados que en la práctica tienen más tiempo fuera de la universidad.

Es importante señalar que estos egresados permanecieron en la universidad un promedio de 6.8 años; por otro lado, si se analiza al grupo de egresados que no contestaron la encuesta, se identifica que existe un ligero sesgo por escala, esto quiere decir, que la tasa de no respuesta a la encuesta es ligeramente mayor en los egresados de escala de pago más alta.

4.2. Metodología de análisis

Para esta aplicación se utilizará la data de la encuesta aplicada a los egresados de las cohortes del 2009, 2010, 2011 y 2012 que hayan pasado por los Estudios Generales de la universidad (5138 casos en total). En el presente análisis, se excluye a los egresados de facultades que no pasan por Estudios Generales, debido a que parte de las preguntas manifiestas en evaluación tienen que ver con el paso por esta etapa en la universidad.

Si bien la encuesta tiene varios ejes de análisis, para la construcción de las clases latentes nos vamos a concentrar en cuatro ejes de preguntas que nos permitan medir la valoración general que se tiene de la universidad:

1. Importancia y utilidad que se le da a Estudios Generales (2 preguntas)
2. Valoración que se tiene de la Imagen Institucional de la Universidad (8 preguntas)
3. Desarrollo de capacidades en la formación (13 preguntas)
4. Valoración con aspectos de la formación y servicios que brinda la universidad (9 preguntas)

Todas las variables son construidas en escala likert, con un puntaje del 1 al 5, siendo, en todos los casos, el puntaje 1 lo más bajo (menos satisfecho, menos desarrollado, muy malo, etc.) y 5 lo más alto (totalmente satisfecho, totalmente desarrollado, muy bueno, etc.).

Un primer paso en el análisis fue la aplicación del coeficiente de alfa de cronbach para confirmar la fiabilidad de las variables manifiestas entre ejes. Dado los buenos resultados del

coeficiente en cada caso, se decide considerar todas las variables manifiestas de cada eje para el cálculo de las clases latentes:

1. Importancia y Utilidad que se le da a Estudios Generales (2 preguntas)

- a) EEGG1: Qué tan importante considera que ha sido pasar por estudios generales
- b) EEGG2: Qué tan útil considera que ha sido pasar por estudios generales

2. Valoración que se tiene de la Imagen Institucional de la Universidad (8 preguntas)

- a) VII1: Qué tan cercana considera que es la relación de la universidad con sus egresados
- b) VII2: Considera que la universidad es: Plural
- c) VII3: Considera que la universidad es: Líder en el mercado educativo
- d) VII4: Considera que la universidad es: Moderna
- e) VII5: Considera que la universidad es: Investigadora
- f) VII6: Considera que la universidad es: Innovadora
- g) VII7: Considera que la universidad es: Contribuye con el desarrollo del país
- h) VII8: Considera que la universidad es: Reconocida internacionalmente

3. Desarrollo de capacidades en la formación (13 preguntas)

- a) CF1: Considera que la formación le permitió desarrollar: Autonomía para el aprendizaje
- b) CF2: Considera que la formación le permitió desarrollar: Trabajo en Equipo
- c) CF3: Considera que la formación le permitió desarrollar: Liderazgo
- d) CF4: Considera que la formación le permitió desarrollar: Creatividad
- e) CF5: Considera que la formación le permitió desarrollar: Capacidad reflexiva y crítica
- f) CF6: Considera que la formación le permitió desarrollar: Iniciativa y capacidad de hacer propuestas
- g) CF7: Considera que la formación le permitió desarrollar: Capacidad de comunicación oral
- h) CF8: Considera que la formación le permitió desarrollar: Dominio de Idiomas
- i) CF9: Considera que la formación le permitió desarrollar: Dominio de Tecnologías
- j) CF10: Considera que la formación le permitió desarrollar: Resolución de Problemas
- k) CF11: Considera que la formación le permitió desarrollar: Capacidad de análisis y síntesis
- l) CF12: Considera que la formación le permitió desarrollar: Capacidad de Investigación

Clases	Resultados						
	Obs.	Parámetros	G^2	χ^2	Grados de libertad	AIC	BIC
1	4439	128	283897.3	9.042174e+45	4311	358712.4	359531.4
2	4439	257	264422.3	1.568473e+40	4182	339495.4	341139.8
3	4439	386	253382.6	2.240797e+41	4053	328713.6	331183.3
4	4439	515	246592.5	1.455179e+33	3924	322181.6	325476.7
5	4439	644	243860.2	1.874381e+32	3795	319707.3	323827.7
6	4439	773	241807.2	4.643516e+30	3666	317912.8	322858.1

Cuadro 4.2: Estimación de número de clases latentes

m) CF13: Considera que la formación le permitió desarrollar: Capacidad de comunicación escrita

4. Valoración con aspectos de la formación y servicios que brinda la universidad (9 preguntas)

- a) VFS1: Qué tan satisfecho se encuentra con: La actualidad del plan de estudios de su especialidad
- b) VFS2: Qué tan satisfecho se encuentra con: La organización de los cursos de su especialidad
- c) VFS3: Cómo calificaría a la universidad: En la formación académica
- d) VFS4: Cómo calificaría a la universidad: En la calidad de los docentes
- e) VFS5: Cómo calificaría a la universidad: En la difusión de investigaciones actualizadas sobre su especialidad
- f) VFS6: Cómo calificaría a la universidad: En el Sistema de Bibliotecas
- g) VFS7: Cómo calificaría a la universidad: En la Infraestructura
- h) VFS8: Cómo calificaría a la universidad: En la inserción de los egresados al mercado laboral
- i) VFS9: Cómo calificaría a la universidad: En el vínculo con el sector privado

4.3. Constructos de valoración

Al momento de elegir el número de clases latentes, es importante considerar dos aspectos: parsimonia (más simple es mejor) y facilidad de interpretación (Collins and Lanza, 2010).

Si bien los resultados observados en el Cuadro 4.2 nos sugieren que un modelo con 5 y 6 clases se ajustan ligeramente mejor que los demás modelos, por un tema de interpretación y simpleza, preferimos quedarnos con un modelo de 4 clases latentes. La disminución del AIC y del BIC entre 4 y 5 clases es bastante pequeño, por lo que concluimos que elegir un modelo de 4 clases no afecta el ajuste. Es importante observar que si bien se trabajó con una base de 5138 individuos, debido a que se remueven los casos perdidos, obtenemos un total de 4439 observaciones.

Más aún, si observamos los estadísticos de prueba de manera visual (Figura 4.1), se confirma que los indicadores de ajuste disminuyen rápidamente y se estabilizan a partir de cuatro clases, por lo que optamos por un modelo de cuatro clases latentes.

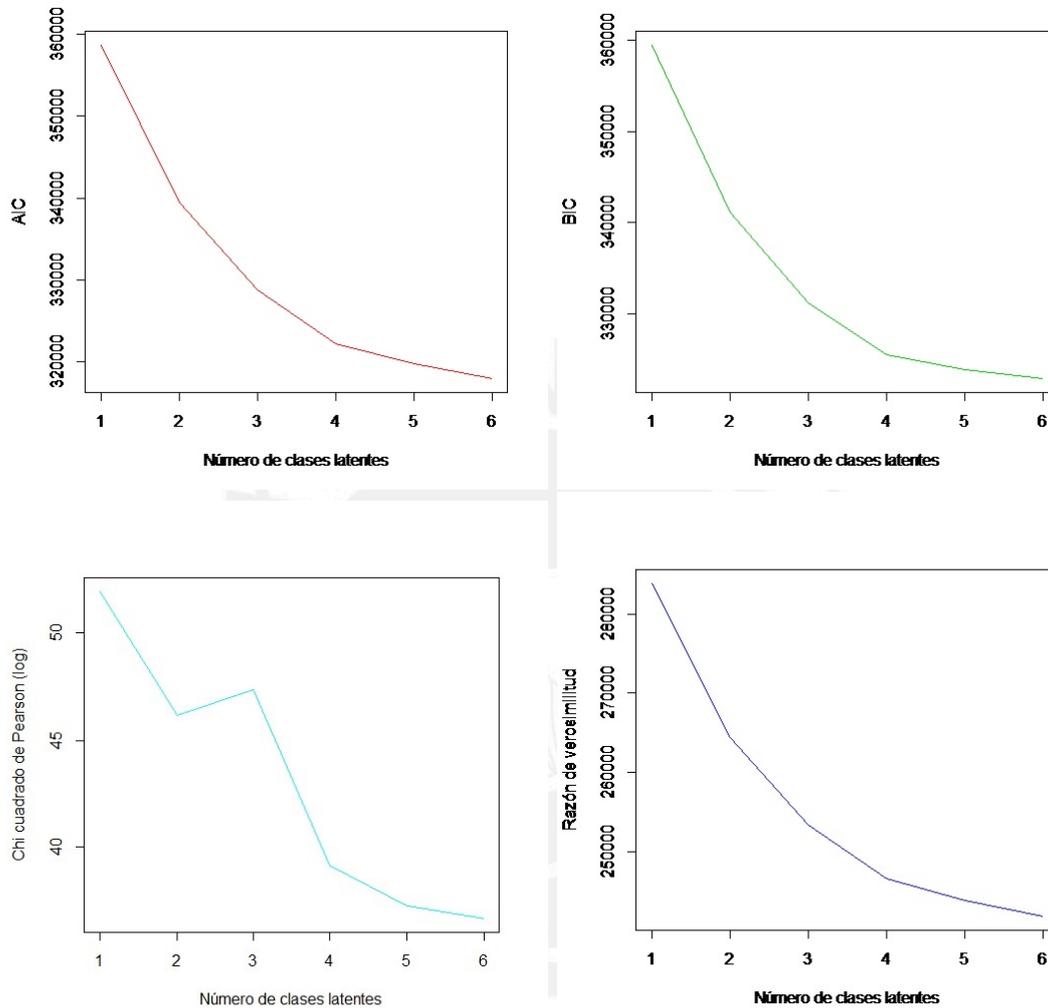


Figura 4.1: Estadísticos de prueba para selección de número de clases latentes

A continuación comentaremos los resultados presentados en los Cuadros 4.3, 4.4, 4.5, 4.6, 4.7 y 4.8.

Con respecto a la prevalencia a cada clase, se observa que la primera clase agrupa al 33 % de la población, la clase 2 al 20 %, la clase 3 al 36 %, y la clase 4 al 11 %.

La primera clase, la cual agrupa al 33 % de la población, tiene mayores probabilidades de considerar muy importante y muy útil su paso por Estudios Generales, y de considerar entre desarrollada y totalmente desarrollada por la formación las capacidades recibidas, principalmente la "Capacidad reflexiva y crítica", la "Capacidad de análisis y síntesis" y la "Capacidad de investigación".

Con respecto a la percepción que se tiene de la imagen de la universidad a nivel institucional, la clase 1 tiene mayores probabilidades de considerar que la universidad cumple con los atributos presentados, principalmente valorándola como "Plural", "Líder en el mercado

educativo educativo”, ”Investigadora”, que ”Contribuye con el desarrollo del País”, y que es ”Reconocida internacionalmente”.

De igual manera, esta clase tiene mayores probabilidades de sentirse satisfecho con la actualidad del plan de estudios y con la organización de los cursos de su especialidad, de considerar muy buena la formación académica, buena la difusión de investigaciones de su especialidad, y entre buena y muy buena la calidad de los docentes, la Infraestructura, el Sistema de Bibliotecas, la inserción al mercado laboral, y el vínculo con el sector privado.

De esta manera, concluimos que esta clase agrupa a los egresados que tienen una valoración muy alta de todos los aspectos evaluados.

La clase 2, la cual agrupa al 20% de egresados, tiene un patrón de respuestas difuminado con respecto a la importancia y utilidad de haber pasado por Estudios Generales, de manera que las probabilidades de respuesta se distribuyen de manera homogénea entre las opciones que van entre medianamente importante o útil a totalmente importante o útil.

Algo similar sucede con respecto al desarrollo de capacidades por parte de la formación; las capacidades de ”Trabajo en equipo”, ”Liderazgo”, ”Creatividad”, ”Iniciativa y capacidad de hacer propuestas” y ”Capacidad de comunicación oral”, tienen una patrón de respuestas que se distribuye entre las opciones 2, 3 y 4, es decir entre poco desarrollado, medianamente desarrollado y desarrollado, mientras que ”Autonomía para el aprendizaje”, ”Capacidad reflexiva y crítica”, y ”Capacidad de análisis y síntesis”, tienen mayor probabilidad de ser consideradas como desarrolladas; por el contrario, las capacidades ”Dominio de Idiomas” y ”Dominio de Tecnologías” tienen una mayor probabilidad de ser consideradas como nada o poco desarrolladas por la formación.

Con respecto a la percepción que se tiene de la imagen de la universidad, la clase 2 tiene una mayor probabilidad de considerar que la universidad es medianamente ”Plural”, ”Líder en el mercado educativo”, ”Moderna”, ”Investigadora”, ”Innovadora”, que ”Contribuye con el desarrollo del País”, y es ”Reconocida internacionalmente” (opciones de respuesta 3 y 4).

De igual manera, esta clase tiene menores probabilidades de sentirse satisfecho con la actualidad del plan de estudios y con la organización de los cursos de su especialidad, de considerar entre mala y regular la difusión de investigaciones de su especialidad, y entre regular y bueno el resto de aspectos evaluados: la formación académica, la calidad de los docentes, el Sistema de Bibliotecas, la Infraestructura, y el vínculo con el sector privado.

De esta manera concluimos que esta clase agrupa a los egresados que tienen una valoración relativamente baja de los aspectos evaluados.

La clase 3, que agrupa al 36% de los egresados, tiene mayores probabilidades de considerar medianamente importante y útil haber pasado por Estudios Generales. Con respecto a las capacidades desarrolladas por la formación, la clase 3 se caracteriza por considerar medianamente desarrolladas o desarrolladas por la universidad todas las capacidades analizadas.

Con respecto a la percepción que se tiene de la imagen de la universidad, la clase 3 tiene una mayor probabilidad de valorar todos estos aspectos con el puntaje 4; es decir, que la universidad es cercana a los atributos evaluados.

Por otro lado, esta clase tiene mayores probabilidades de sentirse entre medianamente satisfecho y satisfecho con la actualidad del plan de estudios y con la organización de los

cursos de su especialidad, de considerar buena la formación académica y la calidad de los docentes, entre regular y buena la difusión de investigaciones, y buena la Infraestructura, el Sistema de Bibliotecas, la inserción al mercado laboral, y el vínculo con el sector privado.

De esta manera concluimos que esta clase agrupa a los egresados que se encuentran satisfechos con todos los aspectos evaluados.

Por último, observamos que la clase cuatro, la cual agrupa al 11 % de la población, se caracteriza por considerar su paso por Estudios Generales entre útil y muy útil, e importante y muy importante. Con respecto a las capacidades desarrolladas por la formación de la universidad, la clase cuatro tiene mayores probabilidades de considerar que todas las capacidades presentadas han sido desarrolladas; en algunos casos, como "Capacidad reflexiva y crítica", "Resolución de Problemas", y "Capacidad de Investigación", es más probable que consideren que la formación les ha permitido desarrollar totalmente estas capacidades.

En lo que respecta al eje de valoración de la imagen institucional de la universidad, la clase 4 se diferencia del resto de clases al valorar muy bajo estos aspectos. De esta manera, esta clase tiene mayores probabilidad de considerar a la universidad como nada "Plural", nada "Líder en el mercado educativo", entre nada y poco "Moderna", nada "Investigadora", entre nada y poco "Innovadora", que no "Contribuye con el desarrollo del País", y entre nada o poco "Reconocida internacionalmente".

Por último, con respecto a la valoración de la formación y servicios que brinda la universidad, esta clase tiene mayores probabilidades de sentirse satisfecho con la actualidad del plan de estudios y con la organización de los cursos de su especialidad. Se caracteriza por considerar buena la formación académica y la calidad de los docentes, entre regular y buena la difusión de investigaciones de su especialidad, y como bueno el Sistema de Bibliotecas, la Infraestructura, la inserción al mercado laboral y el vínculo con el sector privado. De esta manera concluimos que esta clase agrupa a los egresados que se encuentran en general satisfechos y valoran los aspectos de la formación pero tienen una muy mala imagen de la universidad a nivel institucional.

En resumen, si lo vemos por ejes evaluados se observa que con respecto a Estudios Generales, las clases 1 y 4 tienen una mayor valoración de la importancia y utilidad de pasar por esta etapa de estudios, la clase 3 tiene una valoración buena, y la clase 2 tiene sus respuestas difuminadas entre las opciones 3, 4 y 5.

Con respecto a las capacidades desarrolladas por la formación, la clase 1 y 4 consideran estas capacidades entre desarrolladas y totalmente desarrolladas por la formación, la clase 2 tiene una mayor variación de respuesta según la capacidad observada, pero en general tiene una valoración regular sobre el desarrollo de estas, incluso indicando capacidades como nada desarrolladas en algunos casos, y la clase 3 considera principalmente como desarrolladas estas capacidades.

Con respecto a la Imagen Institucional de la universidad, la clase 4 tiene una visión muy negativa de la universidad con respecto a los atributos presentados, mientras que la clase 1 tiene una visión muy positiva. En el caso de la clase 3 el puntaje principal es 4 y de la clase 2 se distribuye principalmente entre las opciones 3 y 4.

Por último, con respecto a los aspectos de la formación y servicios que brinda la universi-

Indicador	Clases Latentes			
	Clase 1	Clase 2	Clase 3	Clase 4
Prevalencia	0.3312	0.2021	0.3574	0.1093
Indicador	Clase 1	Clase 2	Clase 3	Clase 4
EEGG1 Nada Importante-1	0.0053	0.0453	0.0073	0.0000
	2 0.0170	0.1189	0.0518	0.0310
	3 0.0992	0.1725	0.2063	0.1643
	4 0.2656	0.2765	0.4297	0.3159
	Totalmente Importante-5	0.6129	0.3869	0.3050
EEGG2 Nada útil-1	0.0128	0.0807	0.0135	0.0108
2 0.0338	0.1480	0.0949	0.0607	
3 0.1460	0.2397	0.2863	0.2203	
4 0.3399	0.2758	0.4097	0.3731	
Totalmente útil-5	0.4675	0.2557	0.1956	0.3352

Cuadro 4.3: Importancia y Utilidad que se le da a Estudios Generales

dad, las clases 3 y 4 califican todos los aspectos como buenos principalmente, la clase 1 como bueno o muy bueno, y la clase 2 entre regular y bueno.

Observamos que la clase 1 es una clase que valora bastante bien todos los aspectos evaluados, la clase 2 es una clase de valoración entre regular y baja en general, la clase 3 es una clase que valora todos los aspectos como buenos en general, y la clase 4 es una clase de valoración alta pero con una muy mala apreciación de la universidad como institución y los valores que representa.

De esta manera, podemos etiquetar a cada clase de la siguiente manera:

1. Clase 1: Muy alta valoración en general (33 %).
2. Clase 2: Baja valoración en general (20 %).
3. Clase 3: Mediana valoración (36 %).
4. Clase 4: Alta valoración con mala imagen de la universidad a nivel institucional (11 %).

En la siguiente Sección analizaremos el efecto que pueden tener ciertas covariables en la probabilidad de un individuo de pertenecer a cada una de las clases definidas.

4.4. Factores asociados a los constructos de valoración

Si bien solo se ha utilizado data de la encuesta para el cálculo de las clases latentes, las covariables que utilizaremos para medir el impacto en la pertenencia a las clases serán obtenidas tanto de la encuesta como de la data interna de la universidad. De esta manera, para el presente análisis se considerarán las siguientes covariables:

- AÑO DE EGRESO: Año que egreso de la universidad (numérica)
- SEXO: Sexo del egresado (categórica)

Indicador		Clases Latentes			
		Clase 1	Clase 2	Clase 3	Clase 4
CF1	Nada desarrollado-1	0.0022	0.0190	0.0000	0.0000
	2	0.0037	0.0557	0.0056	0.0000
	3	0.0436	0.3190	0.1543	0.0939
	4	0.4876	0.3947	0.6892	0.5356
	Totalmente desarrollado-5	0.4629	0.2115	0.1509	0.3706
CF2	Nada desarrollado-1	0.0055	0.0671	0.0000	0.0022
	2	0.0132	0.1620	0.0280	0.0174
	3	0.1423	0.3312	0.2848	0.1388
	4	0.3848	0.3133	0.5121	0.4888
	Totalmente desarrollado-5	0.4542	0.1265	0.1751	0.3528
CF3	Nada desarrollado-1	0.0059	0.1040	0.0010	0.0022
	2	0.0190	0.2557	0.0558	0.0343
	3	0.1926	0.3880	0.4202	0.2322
	4	0.4584	0.1960	0.4565	0.5240
	Totalmente desarrollado-5	0.3241	0.0563	0.0665	0.2073
CF4	Nada desarrollado-1	0.0036	0.1096	0.0000	0.0065
	2	0.0259	0.2485	0.0825	0.0439
	3	0.2358	0.3599	0.4440	0.2599
	4	0.4719	0.2194	0.4218	0.5185
	Totalmente desarrollado-5	0.2628	0.0625	0.0517	0.1712
CF5	Nada desarrollado-1	0.0000	0.0274	0.0000	0.0000
	2	0.0007	0.0572	0.0040	0.0045
	3	0.0340	0.2155	0.1117	0.1061
	4	0.2580	0.4002	0.5838	0.3914
	Totalmente desarrollado-5	0.7073	0.2996	0.3005	0.4980
CF6	Nada desarrollado-1	0.0000	0.0476	0.0000	0.0000
	2	0.0050	0.1742	0.0198	0.0085
	3	0.0940	0.3692	0.2860	0.1692
	4	0.4550	0.3296	0.5885	0.5512
	Totalmente desarrollado-5	0.4461	0.0794	0.1057	0.2711
CF7	Nada desarrollado-1	0.0030	0.0856	0.0000	0.0000
	2	0.0214	0.1898	0.0587	0.0336
	3	0.1688	0.3525	0.3549	0.2212
	4	0.4094	0.2948	0.4890	0.4830
	Totalmente desarrollado-5	0.3974	0.0773	0.0973	0.2622

Cuadro 4.4: Desarrollo de capacidades en la formación (1)

Indicador		Clases Latentes			
		Clase 1	Clase 2	Clase 3	Clase 4
CF8	Nada desarrollado-1	0.0852	0.3641	0.0807	0.0732
	2	0.1324	0.2834	0.2492	0.1538
	3	0.3472	0.2270	0.4010	0.3992
	4	0.3013	0.0955	0.2383	0.3116
	Totalmente desarrollado-5	0.1338	0.0300	0.0308	0.0622
CF9	Nada desarrollado-1	0.0309	0.2182	0.0241	0.0154
	2	0.0999	0.2960	0.1572	0.0912
	3	0.2887	0.2891	0.4001	0.2804
	4	0.3509	0.1520	0.3587	0.4322
	Totalmente desarrollado-5	0.2297	0.0447	0.0600	0.1807
CF10	Nada desarrollado-1	0.0015	0.0333	0.0000	0.0000
	2	0.0014	0.1091	0.0105	0.0045
	3	0.0599	0.3283	0.1864	0.0821
	4	0.4358	0.4102	0.6895	0.5395
	Totalmente desarrollado-5	0.5014	0.1190	0.1136	0.3739
CF11	Nada desarrollado-1	0.0000	0.0131	0.0000	0.0000
	2	0.0007	0.0586	0.0053	0.0000
	3	0.0178	0.1986	0.0956	0.0481
	4	0.2753	0.4697	0.6532	0.4585
	Totalmente desarrollado-5	0.7061	0.2601	0.2458	0.4934
CF12	Nada desarrollado-1	0.0022	0.0310	0.0013	0.0023
	2	0.0022	0.1174	0.0221	0.0110
	3	0.0547	0.2372	0.2081	0.1443
	4	0.3212	0.4125	0.5477	0.3864
	Totalmente desarrollado-5	0.6197	0.2019	0.2208	0.4561
CF13	Nada desarrollado-1	0.0015	0.0571	0.0000	0.0000
	2	0.0097	0.1229	0.0364	0.0204
	3	0.1082	0.3495	0.2715	0.2299
	4	0.3777	0.3377	0.5555	0.4439
	Totalmente desarrollado-5	0.5030	0.1328	0.1367	0.3058

Cuadro 4.5: Desarrollo de capacidades en la formación (2)

Indicador		Clases Latentes			
		Clase 1	Clase 2	Clase 3	Clase 4
VIII1	Nada Cercana-1	0.0325	0.2018	0.0335	0.0395
	2	0.1127	0.3346	0.1651	0.1186
	3	0.3960	0.3725	0.5399	0.4535
	4	0.3632	0.0862	0.2459	0.3288
	Muy Cercana-5	0.0956	0.0049	0.0156	0.0595
VII2	Nada cercana-1	0.0028	0.0818	0.0007	0.6077
	2	0.0099	0.1258	0.0053	0.2909
	3	0.0507	0.2680	0.1706	0.0899
	4	0.2933	0.3052	0.5622	0.0070
	Totalmente cercana-5	0.6434	0.2192	0.2612	0.0045
VII3	Nada cercana-1	0.0000	0.0719	0.0000	0.7370
	2	0.0000	0.0869	0.0000	0.2400
	3	0.0172	0.2374	0.1001	0.0186
	4	0.2067	0.3584	0.5391	0.0044
	Totalmente cercana-5	0.7761	0.2454	0.3608	0.0000
VII4	Nada cercana-1	0.0007	0.0464	0.0000	0.3222
	2	0.0031	0.1458	0.0037	0.4995
	3	0.0778	0.4625	0.2307	0.1573
	4	0.4371	0.2750	0.6473	0.0210
	Totalmente cercana-5	0.4813	0.0703	0.1182	0.0000
VII5	Nada cercana-1	0.0000	0.0516	0.0000	0.5571
	2	0.0041	0.1635	0.0124	0.3041
	3	0.0570	0.3643	0.2089	0.1218
	4	0.2871	0.2790	0.5682	0.0127
	Totalmente cercana-5	0.6517	0.1416	0.2105	0.0044
VII6	Nada cercana-1	0.0000	0.0465	0.0000	0.3242
	2	0.0041	0.1878	0.0056	0.4715
	3	0.0841	0.5065	0.3197	0.1868
	4	0.4991	0.2221	0.6124	0.0174
	Totalmente cercana-5	0.4127	0.0371	0.0623	0.0000
VII7	Nada cercana-1	0.0000	0.0540	0.0000	0.5703
	2	0.0015	0.1370	0.0026	0.3422
	3	0.0260	0.3637	0.1785	0.0829
	4	0.2941	0.3270	0.6172	0.0024
	Totalmente cercana-5	0.6783	0.1183	0.2017	0.0022
VII8	Nada cercana-1	0.0010	0.0603	0.0023	0.4793
	2	0.0063	0.1618	0.0116	0.3475
	3	0.0813	0.3232	0.2428	0.1587
	4	0.3170	0.3253	0.5254	0.0102
	Totalmente cercana-5	0.5944	0.1294	0.2180	0.0044

Cuadro 4.6: Valoración que se tiene de la Imagen Institucional de la Universidad

Indicador		Clases Latentes			
		Clase 1	Clase 2	Clase 3	Clase 4
VFS1	Nada satisfecho-1	0.0103	0.0972	0.0098	0.0128
	2	0.0279	0.2570	0.0300	0.0392
	3	0.2350	0.4952	0.4513	0.2726
	4	0.5671	0.1320	0.4821	0.5687
	Totalmente satisfecho-5	0.1596	0.0186	0.0268	0.1067
VFS2	Nada satisfecho-1	0.0085	0.0894	0.0022	0.0000
	2	0.0317	0.2752	0.0344	0.0352
	3	0.2578	0.4931	0.4820	0.2962
	4	0.5494	0.1311	0.4630	0.5730
	Totalmente satisfecho-5	0.1526	0.0111	0.0185	0.0956
VFS3	Muy Malo-1	0.0000	0.0178	0.0000	0.0000
	Malo-2	0.0014	0.0609	0.0033	0.0000
	Regular-3	0.0329	0.3637	0.1154	0.0340
	Bueno-4	0.4666	0.4803	0.7447	0.5976
	Muy Bueno-5	0.4991	0.0772	0.1366	0.3685
VFS4	Muy Malo-1	0.0000	0.0262	0.0000	0.0000
	Malo-2	0.0000	0.0893	0.0041	0.0000
	Regular-3	0.0640	0.4257	0.1827	0.1110
	Bueno-4	0.5015	0.3819	0.6844	0.6111
	Muy Bueno-5	0.4345	0.0770	0.1289	0.2779
VFS5	Muy Malo-1	0.0102	0.1082	0.0000	0.0022
	Malo-2	0.0331	0.2997	0.0628	0.0572
	Regular-3	0.2166	0.3946	0.4801	0.3342
	Bueno-4	0.5252	0.1769	0.4164	0.4706
	Muy Bueno-5	0.2149	0.0206	0.0407	0.1358
VFS6	Muy Malo-1	0.0013	0.0337	0.0012	0.0069
	Malo-2	0.0096	0.0863	0.0159	0.0112
	Regular-3	0.0892	0.2455	0.2263	0.1648
	Bueno-4	0.4181	0.4358	0.5603	0.4413
	Muy Bueno-5	0.4819	0.1987	0.1962	0.3758
VFS7	Muy Malo-1	0.0022	0.0410	0.0024	0.0022
	Malo-2	0.0173	0.1332	0.0302	0.0158
	Regular-3	0.1397	0.3646	0.2610	0.2237
	Bueno-4	0.4734	0.3660	0.5925	0.5119
	Muy Bueno-5	0.3674	0.0952	0.1139	0.2464

Cuadro 4.7: Valoración con aspectos de la formación y servicios que brinda la universidad (1)

Indicador		Clases Latentes			
		Clase 1	Clase 2	Clase 3	Clase 4
VFS8	Muy Malo-1	0.0038	0.0721	0.0035	0.0000
	Malo-2	0.0178	0.1549	0.0270	0.0351
	Regular-3	0.1457	0.4014	0.2474	0.2077
	Bueno-4	0.5063	0.2998	0.6282	0.5567
	Muy Bueno-5	0.3263	0.0717	0.0939	0.2006
VFS9	Muy Malo-1	0.0042	0.0765	0.0034	0.0022
	Malo-2	0.0251	0.1826	0.0254	0.0308
	Regular-3	0.1404	0.4260	0.2926	0.2157
	Bueno-4	0.4898	0.2560	0.6027	0.5605
	Muy Bueno-5	0.3405	0.0590	0.0759	0.1908

Cuadro 4.8: Valoración con aspectos de la formación y servicios que brinda la universidad (2)

Covariables	Resultados						
	Obs.	Parámetros	χ^2	Grados de libertad	AIC	BIC	G^2
No	4439	515	1.455179e+33	3924	322181.6	325476.7	246592.5
Si	4127	548	1.633219e+33	3579	298903.1	302369.4	229203.9

Cuadro 4.9: Modelo con 4 clases latentes

- DESEMPEÑO: Coeficiente de rendimiento académico (Craest) (numérica)
- ESCALA: Escala de pago del egresado durante los años de su carrera (numérica)
- ESPOS: Si ha llevado estudios de posgrado (categórica)
- JEFE: Si es jefe de hogar al momento del censo (categórica)
- TIPOEEGG: Si el egresado ha pasado por Estudios Generales Letras o Estudios Generales Ciencias (categórica)
- SECTOR: Para quien trabaja al momento del censo (categórica)
- ESTUDIO: Número de años que le tomo al egresado concluir sus estudios (numérica)
- ETRAB: Si considera que su empleo al momento del censo solo puede ser realizado por personas de su misma especialidad (categórica)

Observando los estadísticos de ajuste del modelo con covariables (Cuadro 4.9), vemos que nuestro modelo tiene mejores indicadores de ajuste con las covariables que sin ellas (Cuadro 4.2). De igual manera, el número de observaciones disminuye debido a que en el caso de ingreso de covariables el paquete poLCA remueve por default los casos "missing".

El Cuadro 4.10 presenta los coeficientes estimados para las covariables en relación con la clase de referencia, en este caso, la clase 1, la cual etiquetamos como la clase de "Muy alta valoración en general". Observando el estadístico t y el valor p concluimos que no todas las

Covariable	Clase 2 vs. 1			Clase 3 vs. 1			Clase 4 vs. 1		
	β	t value	p-value	β	t value	p-value	β	t value	p-value
Intercepto	-0.00001	-0.231	0.817	0.00001	0.354	0.724	0.00000	0.150	0.881
Año egreso	-0.00071	-1.923	0.055	-0.00051	-1.690	0.091	-0.00052	-1.115	0.265
Sexo _{mujer}	-0.18946	-2.287	0.022	0.17340	2.300	0.022	-0.02401	-0.346	0.729
Craest	0.01663	1.414	0.157	0.02452	2.415	0.016	0.00945	0.639	0.523
Escala	0.14756	3.867	0.000	0.04275	1.243	0.214	-0.13635	-2.704	0.007
Posgrado _{si}	0.00557	0.069	0.945	0.00790	0.110	0.912	0.05065	0.679	0.497
Jefehogar _{si}	0.05076	0.616	0.538	-0.06891	-0.893	0.372	0.01809	0.274	0.784
EEGG _{letras}	0.01893	0.294	0.769	-0.44243	-5.670	0.000	-0.23595	-4.348	0.000
Sector	0.09215	1.403	0.161	0.00455	0.075	0.940	0.11577	1.474	0.141
Años estudio	-0.05209	-1.661	0.097	-0.01105	-0.489	0.625	-0.03900	-0.962	0.336
Funciones _{si}	-0.00281	-0.039	0.969	-0.11608	-1.372	0.170	-0.15216	-3.086	0.002

Cuadro 4.10: Valores estimados para las Covariables

Covariable	Odds		
	Clase 2 vs. 1	Clase 3 vs. 1	Clase 4 vs. 1
Año egreso	-	-	-
Sexo _{mujer}	0.83	1.19	-
Craest	-	1.02	-
Escala	1.16	-	0.87
Posgrado _{si}	-	-	-
Jefehogar _{si}	-	-	-
EEGG _{letras}	-	0.64	0.79
Sector	-	-	-
Años estudio	-	-	-
Funciones _{si}	-	-	0.86

Cuadro 4.11: Odds de las covariables con efecto significativo

covariables resultan significativas a un nivel de 0.05. El Cuadro 4.11 resume los odds de las covariables que son significativas al 0.05.

Siguiendo lo desarrollado en la Sección 3.7, los odds de pertenecer a la clase de baja valoración con respecto a la clase 1 aumentan en 16% a medida que aumentan los pagos en una escala y las demás covariables se mantienen constantes; mientras que los odds de pertenecer a la clase 4 de alta valoración con mala imagen institucional con respecto a la clase 1 disminuyen más bien en un 13% bajo esta misma variación de la escala de pagos.

Con respecto al sexo del egresado, se encontró que se incrementa en un 19% el odds de pertenecer a la clase de mediana valoración con respecto a la clase 1, y que disminuye en 17% los odds de pertenecer a la clase de baja valoración con respecto a la clase 1 cuando el egresado es de sexo femenino. Esto nos indica dos cosas; por un lado, que hay menos mujeres que valoran alto la universidad en comparación al número de mujeres que la valoran de manera regular, y que hay más mujeres que valoran alto la universidad que número de mujeres que tienen una baja valoración. Si comparamos los odds entre estas dos clases (mala y regular) $0.83/1.19=0.70$ se confirma que hay mayor probabilidad que las mujeres valoren de manera regular la universidad que mal, en relación a los hombres. Es decir; las mujeres

tendrían una mejor valoración de los aspectos evaluados.

En esta misma línea, se encontró que el odds de pertenecer a la clase de mediana valoración con respecto a la clase 1, se incrementa en un 2% por cada punto adicional del puntaje estandarizados de rendimiento académico (CRAEST).

Por otro lado, en el caso de los egresados que pasaron por EEGLL, se identifica un decremento del 36% en la clase de mediana valoración; en otras palabras, tales egresados tuvieron una leve mejor valoración de la universidad que las de sus pares que pasaron por EEGCC. Note además que los odds de pertenecer a la clase de alta valoración con mala imagen institucional con respecto a la clase 1 disminuyó en un 21% para este mismo grupo de egresados que pasaron por EEGLL.

Por último, los odds de pertenecer a la clase de alta valoración con una mala imagen a nivel institucional con respecto a la clase 1 disminuyeron en un 14% para los egresados que consideran que su trabajo al momento del censo solo podría ser realizado por personas de su misma especialidad.

En resumen, se observó que el incremento en la escala de pagos influyó negativamente en la valoración de la universidad; mientras que el sexo, el desempeño académico y el tipo de Estudios Generales influenciaron en un leve cambio de la valoración de la universidad. Finalmente, la escala de pago, el tipo de Estudios Generales, y las consideraciones que tienen de sus funciones en su trabajo al momento del censo, tuvieron ciertamente una influencia sobre la imagen que tienen los egresados de la universidad a nivel institucional.

Capítulo 5

Conclusiones

5.1. Conclusiones

El análisis de clases latentes es un procedimiento estadístico que nos permite modelar y agrupar una gran cantidad de variables manifiestas categóricas en dos o más clases de cierto constructo latente. Una de las extensiones de este análisis es el modelo de regresión de clases latentes, el cual incorpora covariables que buscan explicar las probabilidades de pertenencia de un individuo a las distintas clases.

En el presente trabajo se buscó fundamentar y aplicar el modelo de regresión de clases latentes propuesto por [Dayton and Macready \(1988\)](#) y desarrollado por [Bandeem-Roche et al. \(1997\)](#) a la data del Sistema de Seguimiento a Egresados de una universidad privada del país, buscando identificar relaciones entre las variables observadas de valoración al asumir que estas asociaciones se explicaban por una variable latente categórica.

Tanto en el segundo y tercer capítulo de esta tesis se desarrollaron los fundamentos y la teoría de estimación clásica para los modelos de clases latentes y de regresión de clases latentes, asumiendo que las variables manifiestas a emplearse en la determinación de las clases eran de carácter nominal. Una contribución de este trabajo consistió en obtener una caracterización explícita de los errores estándar de estimación asintóticos mediante el cálculo de la matriz Hessiana tanto para el caso dicotómico como politómico, desarrollos poco estudiados en la literatura existente.

Para la aplicación se utilizaron los resultados de la encuesta aplicada a los egresados de los años 2009, 2010, 2011 y 2012 (5138 casos en total). En base a esta data se buscó identificar constructos de valoración, indagando si esta evaluación se veía afectada por características personales de los egresados.

Tomando en cuenta los métodos para analizar el ajuste del modelo como el AIC, el BIC, el estadístico Chi-Cuadrado de Pearson y la razón de verosimilitud, se decidió que un modelo de cuatro clases latentes se ajustaba mejor a la data. De esta manera se identificaron las siguientes clases: Clase 1: Muy alta valoración en general (33 %). Clase 2: Baja valoración en general (20 %). Clase 3: Mediana Valoración (36 %). Clase 4: Alta valoración con mala imagen de la universidad a nivel institucional (11 %).

Con respecto a las covariables que explican la pertenencia a una clase en relación con la clase de referencia, en nuestro caso la clase de muy alta valoración de la universidad, se observó, marginalmente, que solo el incremento en la escala de pagos influenció negativamente en un fuerte cambio de la valoración de la universidad; mientras que el desempeño académico

influyó muy ligeramente en una menor valoración de esta. Tanto los hombres como los egresados que pasaron por EEGGCC tuvieron también una menor valoración de la universidad. Finalmente, la mejor imagen que percibe un egresado de la universidad a nivel institucional se vio que estaba asociada a un incremento en la escala de pagos. Los egresados que pasaron por EEGLL y que consideran sus funciones como exclusivas de su especialidad mostraron en general tener una mejor imagen de la universidad a nivel institucional.

5.2. Sugerencias para investigaciones futuras

Una primera reflexión está orientada a la consideración dada a las variables manifiestas. En el presente trabajo las variables manifiestas han sido tratadas como politómicas nominales, sin considerar el orden que puede tener la escala de respuesta.

Quinn (2004) aborda este problema introduciendo un constructo latente con indicadores ordinales y continuos. El diseña un modelo que unifica el análisis factorial estándar y los modelos de teoría de respuesta al ítem para data ordinal; a través de un algoritmo de cadenas de Markov de Monte Carlo para el ajuste del modelo. La aplicación de su modelo a una data de riesgo político económico se encuentra implementado en el paquete "MCMCpack" de R.

Por otro lado, si bien en esta tesis los parámetros han sido estimados de la manera más clásica y conocida, a través del método de máxima verosimilitud y el algoritmo EM, es posible considerar otros tipos de estimaciones, como la Bayesiana. Chung (2003) muestra un caso donde es imposible calcular la inversa de la matriz Hessiana y calcular el error estándar. Ante esto él plantea el uso de inferencia Bayesiana y métodos MCMC mostrando en (Chung et al., 2006) un estudio sobre el uso y actitudes hacia la marihuana para ilustrar esta metodología.

Otra consideración que se puede tener en cuenta para investigaciones futuras es asumir la clase latente como una variable ordinal y no nominal como aquí se ha asumido. Cha (2011) propone un modelo de regresión de clases latentes ordenadas (OLCR), que extiende el modelo de clases latente ordinal propuesto por Croon (1990) (OLCA). El manifiesta que dado que las variables manifiestas son ordenadas, es natural pensar que la variable latente también lo sea. Tal concepción podría resultar particularmente útil en el campo de la evaluación educativa, ya que los evaluados pertenecientes a la clase más alta se podrían pensar que tienen un mejor desempeño en la prueba. Este modelo asume que $P(\mathbf{X} \leq x | j_1) \geq P(\mathbf{X} \leq x | j_2) \geq P(\mathbf{X} \leq x | j_3) \geq P(\mathbf{X} \leq x | j_4)$, donde los j_k representan las clases latentes ordenadas.

Por último, otras consideraciones deben ser hechas sobre la relación que se establecen entre las covariables y las variables manifiestas. En el presente modelo las covariables solo afectan a la probabilidad de pertenencia a la clase, de manera que la determinación de las clases latentes se hace de igual manera que en el análisis de clases latentes gracias a la propiedad de marginalización. Existen, sin embargo, otros modelos, donde se permite que las covariables afecten también a las probabilidades de respuesta π_{ij} . Vermunt and Magidson (2013), en el programa comercial Latent Gold, implementan una variedad de modelos de clases latentes, incorporando covariables tanto en las probabilidades de pertenencia a la clase η como en las de respuesta a una categoría manifiesta π_{ij} , permitiendo adicionalmente que las variables manifiestas sean de distinta naturaleza. Chung (2003) señala, sin embargo, que dejar que los π_{ij} se vean afectadas por las covariables no es del todo deseable ya que esto incrementa el número de parámetros a estimar y destruye la propiedad de marginalización,

resultando en un modelo más difícil de interpretar. Un estudio comparativo con los modelos implementados en el programa Latent Gold sería por tanto de gran interés.



Apéndice A

Sintaxis del poLCA

A.1. Sintaxis del poLCA

Para la estimación de clases latentes y su extensión con covariables, utilizaremos el paquete poLCA, implementado en el programa estadístico R ([Linzer and Lewis, 2011](#)).

En esta Sección se presenta la sintaxis del paquete poLCA para el cálculo de 4 clases latentes, explicando brevemente cada comando utilizado.

- Análisis de Clases Latentes

```
res4CL= poLCA(cbind(pseegg1,pseegg2,pesf7_1,pesf7_2,pesf7_3,pesf7_4,pesf7_5,
pesf7_6,pesf7_7,pesf7_8,pesf7_9,pesf7_10,pesf7_11, pesf7_12,pesf7_13,vpii1,
vpri2_1,vpri2_2,vpri2_3,vpri2_4,vpri2_5,vpri2_6,vpri2_7,psfe1,psfe2,psfe3_1,
psfe3_2,psfe3_3,psfe3_4,psfe3_5,psfe3_6,psfe3_7) ~ 1,
maxiter=50000,na.rm=TRUE,probs.start=NULL,nclass=4, data=base2,nrep=10)
```

donde:

- maxiter: número máximo de iteraciones del algoritmo de estimación
- na.rm: indica el tratamiento a dar a la data pérdida. En el caso que sea "TRUE", se remueven los casos perdidos.
- prob.start: genera valores iniciales para los π_{ij} . En el caso que sea "NULL", los valores iniciales son generados aleatoriamente.
- nclass: número de clases latentes que asume el modelo
- data: especifica la base de datos usada
- nrep: el número de veces que repite este comando, seleccionando el mejor modelo.
- Regresión de Clases Latentes

```
res4= poLCA(cbind(pseegg1,pseegg2,pesf7_1,pesf7_2,pesf7_3,pesf7_4,pesf7_5,
pesf7_6,pesf7_7,pesf7_8,pesf7_9,pesf7_10,pesf7_11, pesf7_12,pesf7_13,vpii1,
vpri2_1,vpri2_2,vpri2_3,vpri2_4,vpri2_5,vpri2_6,vpri2_7,psfe1,psfe2,psfe3_1,
psfe3_2,psfe3_3,psfe3_4,psfe3_5,psfe3_6,psfe3_7) ~ a.o+sexo+craest+escala+
estegre1+jefe_hogar+tipoeegg+col2are4+anosestudio+pef1234,
maxiter=50000,na.rm=TRUE, probs.start=NULL,nclass=4,data=base2,nrep=10)
```

Bibliografía

- Agresti, A. (1990). *Categorical data analysis*, Wiley, New York.
- Bandeen-Roche, K., Miglioretti, D., Zeger, S. and Rathouz, P. (1997). Latent variable regression for multiple discrete outcomes, *Journal of the American Statistical Association* **92(440)**: 1375–1386.
- Bartholomew, D., Knott, M. and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis*, Wiley, NN.
- Blaydes, L. and D.A., L. (2007). The political economy of women s support for fundamentalist islam, *World Politics* **60(4)**: 576–609.
- Breen, R. (2000). Why is support for extreme parties underestimated by surveys? a latent class analysis, *British Journal of Political Science* **Vol. 30, No. 2**: 375–382.
- Cha, J. (2011). *Application of Ordered Latent Class Regression in Educational Assessment*, PhD thesis, Columbia University.
- Chung, H. (2003). *Latent-Class Modeling with Covariates*, PhD thesis, The Pennsylvania State University-Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy.
- Chung, H., Flaherty, B. and Schafer, J. (2006). Latent class logistic regression: Application to marijuana use and attitudes among high school seniors, *Journal of the Royal Statistical Society Series A*, **169**: 723–743.
- Clogg, C. F. (1981). Latent structure models of mobility, *American Journal of Sociology* **86(4)**: 836–868.
- Collins, L. and Lanza, S. (2010). *Latent Class and Latent Transition Analysis*, Wiley, New Jersey.
- Dayton, M. and Macready, G. (1988). Concomitant-variable latent-class models, *Journal of the American Statistical Association* **83(401)**: 173–178.
- Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm, *Journal of the Royal Statistical Society* **29(1)**: 1–38.
- Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data, *Journal of the American Statistical Association* **87(418)**: 476–486.
- Kamakura, W., Wedel, M. and Agrawal, J. (1994). Concomitant variable latent class models for conjoint analysis, *International Journal of Research in Marketing* **11**: 451–464.
- Lazarsfeld, P. and Henry, N. (1968). *Latent structure Analysis*, Houghton Mifflin, Boston.
- Linzer, D. and Lewis, J. (2011). polca: An r package for polytomous variable latent class analysis, *Journal of Statistical Software* **42(10)**: 1–29.

- Morrison, E., Rudd, E., Zumeta, W. and Nerad, M. (2011). What matters for excellence in phd programs: Latent constructs of doctoral program quality used by early career social scientists, *The Journal of Higher Education* **82(5)**: 535–563.
- Quinn, K. (2004). Bayesian factor analysis for mixed ordinal and continuous responses, *Political Analysis* **12(4)**: 338–353.
- Vermunt, J. and Magidson, J. (2013). *Technical Guide for Latent Gold 5.0: Basic, Advanced, and Syntax*, Belmont, MA: Statistical Innovations Inc.
- Yamaguchi, K. (2000). Multinomial logit latent-class regression models: An analysis of the predictor of gender role attitude among japanese women, *Chicago Journals* **105(6)**: 1702–1740.

