

**UNIVERSIDAD DE SALAMANCA**  
**DEPARTAMENTO DE ESTADÍSTICA**

---



**MODELOS DE CLASES LATENTES EN TABLAS POCO  
OCUPADAS: UNA CONTRIBUCIÓN BASADA EN  
BOOTSTRAP**

**CARLOMAGNO ARAYA ALPÍZAR**

**2010**

# MODELOS DE CLASES LATENTES EN TABLAS POCO OCUPADAS: UNA CONTRIBUCIÓN BASADA EN BOOTSTRAP

---

Memoria para optar al Grado de Doctor por el  
Departamento de Estadística de la Universidad de  
Salamanca, presenta:

**Carlomagno Araya Alpízar**

Salamanca

2010



*Universidad de Salamanca*  
*Departamento de Estadística*

---

**M<sup>a</sup> PURIFICACIÓN GALINDO VILLARDÓN**

**y**

**ROSA A. SEPÚLVEDA CORREA**

*Profesoras del Departamento de Estadística*  
*de la Universidad de Salamanca*

CERTIFICAN: Que Don **Carlomagno Araya Alpízar**, ha realizado, en el Departamento de Estadística de la Universidad de Salamanca, bajo su dirección, el trabajo que, para optar al Grado de Doctor, presenta con el título: “Modelos de Clases Latentes en Tablas poco Ocupadas: una Contribución Basada en Bootstrap” y para que conste, firman el presente certificado en Salamanca, en Diciembre de 2010.

## AGRADECIMIENTOS

---

A la Universidad de Costa Rica, por el esfuerzo que, desde las Instancias administrativas hasta mis colegas de la Sede de Occidente, han realizado en pro de mi formación.

Al Ministerio de Ciencias y Tecnologías (MICIT) por la financiación de una proporción de los gastos de manutención.

A la doctora **M<sup>a</sup>. Purificación Galindo Villardón**, directora de este trabajo, por su constante apoyo, tanto en la parte académica como en la personal, y por no dejar que su talento opaque su calidad humana.

A la doctora **Rosa A. Sepúlveda Correa**, codirectora de este trabajo, por todas sus aportaciones técnicas al mismo, pero sobre todo por haberme brindado su amistad.

A todo el **Personal del Departamento de Estadística de la Universidad de Salamanca** por demostrar que la calidad académica no tiene que ser excluyente con la calidad humana.

*A la Memoria de mi padre*

*Marcelo*

*A Angie, Andrea y Rosalina mis fuentes de amor, comprensión  
e inspiración.*

*A mis hermanos*

## **ABREVIATURAS**

MCL	Modelos de clases latentes.
EBA	Estadístico de bondad de ajuste.
EM	Esperanza-Maximización.
AIC	Criterio de información de Akaike.
BIC	Criterio de información bayesiana.
CR	Cressie-Read.
BNP	Bootstrap no paramétrico.
BP	Bootstrap paramétrico
MP	Método de patrones.
TPO	Tablas poco ocupadas.
VIH	Virus de la Inmunodeficiencia Humana

# NOTACIÓN

$p$	Número de variables manifiestas.
$X_j$	J-ésima variable manifiesta.
$I_j$	Número de categorías de la variable $X_j$ .
$x_j$	Valor observado de la j-ésima variable manifiesta.
$\mathbf{x}$	Conjunto de patrones observados para los $n$ individuos.
$\mathbf{x}_i$	Patrón de respuesta para el i-ésimo individuo.
$T$	Número de clases latentes para la variable latente $Y$ .
$\pi_{\mathbf{X}}(\mathbf{x})$	Probabilidad conjunta de las variables manifiestas.
$\mathbf{Y}$	Vector de variables latentes.
$\pi_{\mathbf{X}/Y(c)}(\mathbf{x})$	Probabilidad condicional de obtener un determinado patrón de respuesta para un individuo perteneciente a la latente $c$ .
$\pi_{X_j/Y(c)}(x_j)$	Probabilidad de que la variable j-ésima tome un determinado valor para un individuo de la clase $c$ .
$\pi_{\mathbf{X},Y}(\mathbf{x}, c)$	Probabilidad conjunta de obtener un determinado patrón de respuesta y estar en la clase latente $c$ .
$\pi_{Y(c)}$	Probabilidad de la clase latente $c$ , conocida como probabilidad a priori.
$\pi_{Y/X(\mathbf{x})}(c)$	Probabilidad de que un individuo con determinado patrón de respuesta se encuentre en la clase latente $c$ , conocida como probabilidad <i>a posteriori</i> .
$f_{\mathbf{x}_h}$	Frecuencia observada para el patrón de respuesta $h$ .
$\hat{f}_{\mathbf{x}_h}$	Frecuencia estimada para el patrón de respuesta $h$ bajo un modelo de clases latentes.
$n$	Tamaño de la muestra.
$R$	Número posible de patrones de respuesta.
$\mathbf{X}'_h$	Patrón de respuesta $h$ .
$q$	Número de variables latentes.
$Y_q$	Variable latente q-ésima.
$G^2$	Estadístico Razón de verosimilitud.
$FT^2$	Estadístico Freeman-Tukey.
$Q^2$	Estadístico Pearson.
$\chi^2$	Distribución teórica Chi-Cuadrado.
$H_0$	Hipótesis nula.

$H_1$	Hipótesis alternativa.
$\mathbf{x}_b^*$	b-ésima remuestra Bootstrap.
$B$	Número de remuestras Bootstrap.
$A$	Número de muestras de Patrones.
$\theta$	Estadístico de contraste.
$\hat{\theta}$	Estadístico de contraste calculado con los datos de la muestra.
$\hat{\theta}_b^*$	Estadístico Bootstrap de la $b$ -ésima remuestra.
$\hat{\theta}_a^*$	Estadístico de la $a$ -ésima muestra de patrones.
$\bar{\theta}^*$	Media de las réplicas Bootstrap.
$Var(\theta^*)$	Varianza del estadístico Bootstrap $\theta^*$ .
$F_B(\hat{\theta}^*)$	Función de distribución empírica del estadístico $\theta$ .
$gl$	Grados de libertad del modelo de clases latentes.
$gl_{cor}$	Grados de libertad corregidos.
$Z$	Variable normal estándar.
$pv$	Nivel de significación calculado con la distribución muestral.
$w$	Número de celdas con frecuencias esperadas nulas.
$\delta$	Constante sumada a la frecuencia observada de las celdas.



---

# ÍNDICE

---

<b>INTRODUCCIÓN</b>	<b>1</b>
<b>CAPÍTULO 1. EL MODELO DE CLASES LATENTES</b>	
1.1 INTRODUCCIÓN	7
1.2 PLANTEAMIENTO DEL MODELO	8
1.3 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO	13
1.4 BONDAD DE AJUSTE DEL MODELO	15
1.5 CRITERIOS DE SELECCIÓN DEL MODELO	22
<b>CAPÍTULO 2. EL MÉTODO BOOTSTRAP</b>	
2.1 INTRODUCCIÓN	25
2.2 FUNDAMENTOS TEÓRICOS	26
2.3 VERSIONES DEL MÉTODO BOOTSTRAP	27
2.3.1 BOOTSTRAP NO PARAMÉTRICO	28
2.3.2 BOOTSTRAP PARAMÉTRICO	31
2.4 CÁLCULO DEL NIVEL DE SIGNIFICACIÓN	33
2.5 ESTIMACIÓN DEL SESGO Y ERROR ESTÁNDAR	34
2.6 PROBLEMAS DEL MÉTODO BOOTSTRAP	35
<b>CAPÍTULO 3. MÉTODOS UTILIZADOS PARA LA SELECCIÓN DE MODELOS</b>	
3.1 INTRODUCCIÓN	39
3.2 AÑADIR UNA CONSTANTE A CADA CELDA	40
3.3 ANÁLISIS DE RESIDUOS	46
3.4 ESTIMACIÓN ROBUSTA CON $\beta$ -VEROSIMILITUD	50
3.5 UTILIZAR CELDAS SUFICIENTEMENTE OCUPADAS	52
3.6 AGRUPACIÓN DE CATEGORÍAS DE VARIABLES	54
3.7 ESTIMACIÓN DE LA DISTRIBUCIÓN EMPÍRICA	55
3.7.1 INTRODUCCIÓN	55
3.7.2 BOOTSTRAP NO PARAMÉTRICO	56
3.7.3 BOOTSTRAP PARAMÉTRICO	57
3.7.4 COMPARACIÓN ENTRE BOOTSTRAP PARAMÉTRICO Y NO PARAMÉTRICO	61
3.7.5 CÁLCULO DEL NIVEL DE SIGNIFICACIÓN	65

3.7.6	GRADOS DE LIBERTAD EN TABLAS POCO OCUPADAS	66
3.8	BOOTSTRAP NO PARAMÉTRICO CORREGIDO	68
<b>CAPÍTULO 4. PROPUESTA DEL MÉTODO DE PATRONES</b>		
4.1	INTRODUCCIÓN	71
4.1.1	DESCRIPCIÓN DEL MÉTODO	72
4.1.2	PROMEDIO ARITMÉTICO Y VARIANZA DE $\hat{\theta}^*$	75
4.1.3	CÁLCULO DEL NIVEL DE SIGNIFICACIÓN	76
4.1.4	ILUSTRACIÓN DEL MÉTODO DE PATRONES	77
4.2	PROBLEMA DE DENSIDAD DE LOS PATRONES	81
4.3	CONSIDERACIONES FINALES	82
<b>CAPÍTULO 5. PROGRAMA BOOTLATENTCLASS</b>		
5.1	INTRODUCCIÓN	84
5.2	ENTRADA AL PROGRAMA	85
5.3	ANÁLISIS DE CLASES LATENTES	89
5.4	BOOTSTRAP NO PARAMÉTRICO	105
5.5	BOOTSTRAP NO PARAMÉTRICO CORREGIDO	110
5.6	BOOTSTRAP PARAMÉTRICO	113
5.7	MÉTODO DE PATRONES	115
5.8	SIMULACIÓN DE DATOS	118
5.9	CONSIDERACIONES FINALES	121
5.10	COMPARACIÓN DEL BOOTLATENTCLASS CON OTROS PROGRAMAS	122
<b>CAPÍTULO 6. APLICACIÓN DE LAS PROPUESTAS</b>		
6.1	INTRODUCCIÓN	126
6.2	APLICACIÓN A DATOS SIMULADOS	127
6.3	APLICACIONES A DATOS REALES	135
6.3.1	CONSUMO DE DROGAS EN ADOLESCENTES	136
6.3.2	PRUEBAS DE DIAGNÓSTICO DEL VIH	146
6.3.3	DIAGNÓSTICO DE LA ENFERMEDAD DE ALZHEIMER	151

**CAPÍTULO 7. TABLAS ESTADÍSTICAS DE VALORES  
CRÍTICOS PARA EL MÉTODO DE PATRONES**

7.1	INTRODUCCIÓN	157
7.2	USO DE LAS TABLAS ESTADÍSTICAS	186
7.3	APLICACIÓN A DATOS SIMULADOS	186
7.4	CONSIDERACIONES FINALES	190

<b>CONCLUSIONES</b>	<b>191</b>
---------------------	------------

<b>BIBLIOGRAFÍA</b>	<b>194</b>
---------------------	------------

**ANEXOS: DATOS SIMULADOS DEL APARTADO 6.2**

DATOS 1	202
DATOS 2	203
DATOS 3	204
DATOS 4	205
DATOS 5	207

---

# INTRODUCCIÓN

---

El contexto general de esta investigación, se enmarca en el estudio del problema que puede surgir en la aplicación de los Modelos de Clases Latentes, cuando se incumplen las propiedades asintóticas de los estadísticos de bondad de ajuste, situación que se presenta en las tablas de contingencia poco ocupadas, conocidas como “sparse data” (AGRESTI & YANG, 1987). Los datos “sparse” se presentan a menudo en conjuntos de datos pequeños o cuando el número posible de patrones de respuesta es grande, ya que la mayoría de los patrones de respuestas tienen frecuencias cero o tienden a cero.

Se han propuesto algunas soluciones para hacer frente al problema. Una de ellas es encontrar el modelo de clases latentes más apropiado utilizando el método Bootstrap Paramétrico. Básicamente, el método consiste en simular conjuntos de datos adicionales (o remuestras), utilizando una distribución de probabilidad conocida. Los nuevos conjuntos de datos son generados con el objetivo de encontrar la distribución de probabilidad empírica de los estadísticos de bondad de ajuste. A continuación, los estadísticos obtenidos de la muestra son evaluados comparándolos con la distribución empírica, aplicando reglas de decisión propuestas por LANGEHEINE et al. (1996) y von DAVIER (1997).

El método Bootstrap No-Paramétrico se lleva a cabo por medio de la distribución obtenida directamente de los datos. La idea consiste en generar observaciones a partir de la distribución de una muestra aleatoria independiente obtenida de la población en estudio. La diferencia entre el Bootstrap Paramétrico y No Paramétrico está en función de que el remuestreo se produzca sobre una distribución teórica o una distribución empírica (EFRON, 1979).

La realización de esta investigación se sustenta en tres metas que en cierta forma definen la originalidad del tema:

1. Determinar el poder del Método Bootstrap No Paramétrico en la diagnosis de modelos de clases latentes.

2. Proponer el Bootstrap No Paramétrico Corregido y el Método de Patrones para el diagnóstico de modelos de clases latentes en tablas poco ocupadas.
3. Construir un software utilizando el lenguaje de computación técnico MATLAB que permita la diagnosis de modelos de clases latentes en tablas poco ocupadas.

Para las metas antes mencionadas, no existen aportes metodológicos en la bibliografía consultada. Hasta ahora, se ha realizado propuestas de cómo abordar el problema de las tablas poco ocupadas en los modelos de clases latentes utilizando principalmente el método Bootstrap Paramétrico.

Centrado el tema de investigación y las metas que debemos alcanzar, nos hemos planteado los siguientes objetivos:

1. Estudiar exhaustivamente las bases teóricas de los modelos de clases latentes para variables manifiestas binarias.
2. Analizar los métodos Bootstrap No Paramétrico y Paramétrico, respecto a sus cualidades para determinar la distribución empírica de los estadísticos de bondad ajuste.
3. Plantear las bases teóricas y metodológicas de nuestra propuesta, el Método de Patrones y Bootstrap No Paramétrico Corregido, para el diagnóstico de modelos de clases latentes en tablas poco ocupadas.
4. Construir un programa en el lenguaje técnico MATLAB, que permite la diagnosis de modelos con Bootstrap y Método de Patrones.
5. Comparar el Método Bootstrap y el Método de Patrones en la determinación de modelos de clases latentes y tablas poco ocupadas con datos reales y simulados.

La presente memoria se estructura en siete capítulos. En el **Capítulo 1** presentamos los fundamentos teóricos del Modelo de Clases Latentes para variables manifiestas binarias, necesarios para la comprensión de la problemática de la selección de modelos en tablas poco ocupadas, la cual se desarrolla con mayor profundidad en los capítulos siguientes.

El **Capítulo 2** está centrado en la presentación de los fundamentos teóricos del método Bootstrap, que deriva la distribución de los estimadores empíricamente, utilizando para esto la información de la muestra original. De esta forma, el enfoque Bootstrap permite eludir las limitaciones del enfoque estadístico clásico, que resultan principalmente de asumir supuestos, que en muchas ocasiones prácticas resultan poco realistas.

Para el **Capítulo 3** se considerarán un conjunto de métodos que han propuesto los investigadores para la selección de Modelos de Clases Latentes en tablas poco ocupadas. Seguidamente, se detalla con mayor amplitud el uso de los métodos Bootstrap: No Paramétrico, No Paramétrico Corregido y Paramétrico, para la diagnosis de modelos. Se presenta el cálculo del nivel de significación y una comparación entre el Bootstrap Paramétrico y el No Paramétrico.

El **Capítulo 4** ofrece una solución al problema del diagnóstico de modelos de clases latentes en tablas poco ocupadas mediante el Método de Patrones. Este método es una propuesta innovadora para el diagnóstico de modelo de clases latentes con conjuntos de datos que tienen problemas de poca densidad (en inglés, “sparseness”). No se encuentra, ni ha sido propuesto en ningún otro texto que aborda esta problemática.

En tanto, el **Capítulo 5** consiste en una presentación del programa BootLatentClass, que ilustra el uso de los métodos Bootstrap y de Patrones. Puesto que los algoritmos que conforman la aplicación son originales, este capítulo en su conjunto representa una aportación de este trabajo al diagnóstico de modelos de clases latentes en tablas poco ocupadas.



El **Capítulo 6** se dedica a la aplicación de los métodos de diagnóstico de modelos con datos reales y simulados. Primero, se comparan los resultados de todos los métodos conocidos y nuestras propuestas: Método de Patrones y el Bootstrap No Paramétrico Corregido, utilizando cinco conjuntos de datos simulados. Seguidamente, se hace una comparación entre los Bootstrap Paramétrico, Bootstrap No Paramétrico y el Método de Patrones, con datos sobre el consumo de drogas en adolescentes, métodos de detección del síndrome de inmunodeficiencia adquirida y diagnóstico de la enfermedad de Alzheimer. Cerramos esta memoria presentando en el **Capítulo 7**, un conjunto de tablas estadísticas de valores críticos para el Método de Patrones.

Finalmente se presentan las conclusiones y las referencias bibliográficas que han servido de base y de inspiración para los desarrollos realizados en este trabajo.

# **CAPÍTULO 1**

---

## **EL MODELO DE CLASES LATENTES**

---

## 1.1 INTRODUCCIÓN

Los métodos estadísticos basados en modelos de variables latentes juegan un importante papel en el análisis de datos multivariantes. Bajo esta denominación genérica se agrupa una amplia variedad de modelos, que se diferencian entre sí por la naturaleza tanto de las variables latentes<sup>1</sup> como de las variables manifiestas (o indicadoras)<sup>2</sup>. Así, según que las variables indicadoras y las latentes sean continuas, discretas, nominales u ordinales se puede considerar una amplia tipología de modelos de variables latentes.

Es evidente que un tratamiento exhaustivo de todos estos tipos de modelos excede los objetivos de esta investigación, de forma que centraremos nuestro planteamiento en el análisis de clases latentes que supone, tanto las variables manifiestas como las latentes categóricas.

Las primeras nociones del Modelo de Clases Latentes (MCL) fueron introducidas por LAZARSFELD & HENRY (1968). Los problemas de estimación e identificación han sido tratados por ANDERSON (1954) y McHUGH (1956), GOODMAN (1974) conectó estos modelos con la teoría moderna de las tablas de contingencia. GOODMAN (1974) y HABERMAN (1979) realizaron importantes aportaciones teóricas, sobre todo en lo que se refiere a la formulación del mismo y a la estimación de los parámetros. Se puede citar a distintos autores que han desarrollado esta metodología, como AGRESTI (1984), ANDERSEN (1991), BARTHOLOMEW (1987), CLOGG (1995), McCUTCHEON (1987), MAGIDSON & VERMUNT (2002, 2004, 2005), DIAS & VERMUNT (2006) y FORMANN (1982). Un aporte importante es la utilización de los métodos Biplot basados en Modelos Lineales Generalizados, para diagnosticar la dependencia local en un Modelo de Clases Latentes (SEPÚLVEDA et al., 2008).

---

<sup>1</sup> Las *variables latentes* son elaboraciones teóricas o constructos y se asume que subyacen a las variables manifiestas.

<sup>2</sup> Una variable manifiesta puede registrarse u observarse de manera directa.

Por otra parte, existen múltiples aplicaciones de los MCL. Por ejemplo, se han presentado en la literatura de investigación diferentes tipos de consumidores de alcohol (CHASSIN et al., 2002; BUCHOLZ et al., 1996), de niños con patrones agresivos (TREMBLAY et al., 2004), relación entre calidad de servicio, satisfacción y confianza con intención de recompra (RONDÁN et al., 2007), análisis de Síndrome de Tourette (GRADOS & MATHEWS, 2008), identificación del cromosoma 1q23-25 (HOLLIDAY et al., 2009) y, estimación de la prevalencia de enfermedades (CHU et al., 2010).

En este capítulo se plantean los fundamentos teóricos de los modelos de clases latentes para variables manifiestas binarias, como son la estimación de los parámetros, los estadísticos de bondad de ajuste, los criterios de selección de los modelos, entre otros.

## 1.2 PLANTEAMIENTO DEL MODELO

Matemáticamente el MCL se esboza partiendo de una matriz que contiene los resultados de  $p$  variables categóricas directamente observadas, llamadas variables manifiestas, que serán denotadas como  $X_j$ , que conforman un vector columna de  $p$  componentes  $\mathbf{X}' = (X_1, \dots, X_p)$  sobre una muestra total de  $n$  individuos.

$$\mathbb{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Denotaremos la observación de un individuo  $i$  sobre una variable  $j$  como  $x_{ij}$  y  $\mathbf{x}_i$  será el vector para todas las variables sobre el individuo  $i$ , conocido también como patrón de respuesta.

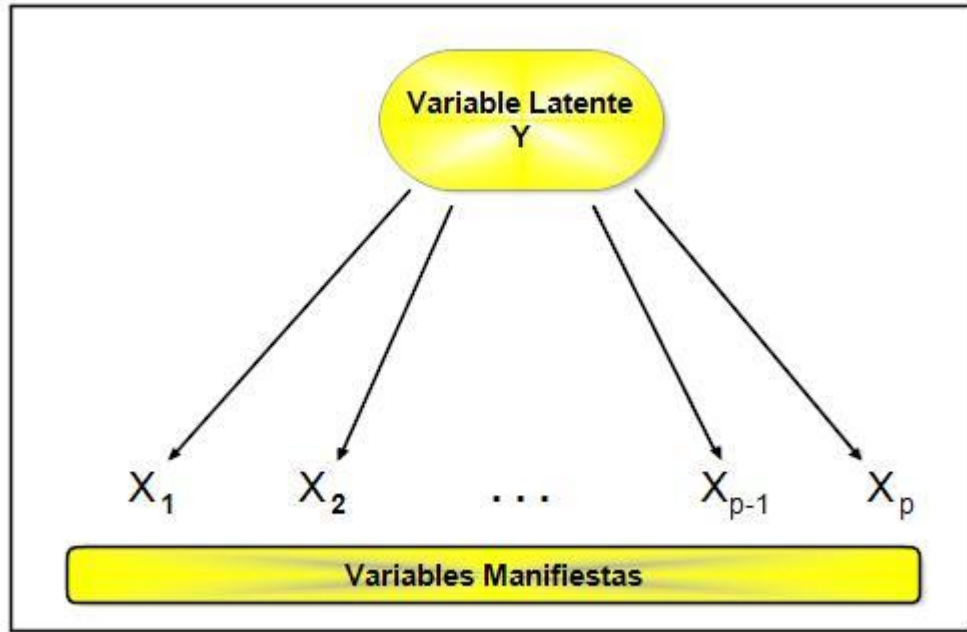
Las variables latentes son representadas por  $Y$ , se expresan mediante el vector  $\mathbf{Y}' = (Y_1, \dots, Y_q)$ , donde  $q < p$ . Para el planteamiento teórico del modelo es suficiente considerar sólo una variable latente, ya que modelos con más de una

variable latente,  $q > 1$ , pueden ser desarrollados considerando  $q=1$  bajo adecuadas restricciones a los parámetros del modelo (GOODMAN, 1974.). Por esta razón, se presenta el MCL considerando una única variable latente  $Y$  con  $T$  categorías o clases latentes. Las  $p$  variables manifiestas  $\mathbf{X}' = (X_1, \dots, X_p)$  se consideran indicadoras de la variable latente  $Y$ .

El vector  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$  denota un determinado patrón de respuesta en el cual cada una de las  $x_j$  toma diferentes valores dependiendo de las categorías de la correspondiente variable manifiesta. Estas variables conforman una tabla de contingencia múltiple con  $\prod_{j=1}^p I_j$  patrones de respuesta, tal que cada  $X_j$  contiene  $I_j$  categorías.

La representación gráfica del modelo de clases latentes (bajo el supuesto de *independencia local*), se encuentra en la **Figura 1.1**. Las *variables manifiestas*  $X_j$  no están conectadas directamente y la dirección de las flechas, indica que la variable latente ( $Y$ ) explica toda la posible asociación entre las variables manifiestas.

Toda la asociación observada entre las variables manifiestas, está medida o explicada por las variables latentes. En otras palabras, la *independencia local* implica en esta situación que la relación entre las variables indicadoras, viene explicada por la pertenencia de cada individuo a una clase latente concreta, teniendo presente que cada clase latente tendrá probabilidades condicionadas de respuesta a las variables manifiestas, diferentes a las probabilidades condicionadas asociadas a otra clase latente distinta, y que los individuos que pertenecen a la misma clase latente, tendrán la misma probabilidad de responder a las variables manifiestas en cualquier combinación de categorías de las mismas. Este hecho sirve para diferenciar a los individuos pertenecientes a diferentes grupos y poder caracterizar tanto la variable latente como las clases latentes.



**Figura 1.1.** Representación gráfica de un Modelo de Clases Latentes con una variable latente y p variables manifiestas.

El incumplimiento del principio de *Independencia local* trae como consecuencia que los estadísticos de bondad de ajuste del modelo (EBA): Pearson ( $Q^2$ ), Razón de verosimilitud ( $G^2$ ), Cressie-Read (CR) y Freeman-Tukey ( $FT^2$ ) sean demasiado grandes, los valores de las estimaciones de los parámetros del modelo se distorsionen (UEBERSAX, 2000), los errores estándar para los estimadores sean grandes, y los estimadores de las diferencias de sus varianzas sean no consistentes (QU et al., 1996). Una de las razones que determina la falta de ajuste de un MCL, está asociada a la violación del principio de *independencia local* (MAGIDSON & VERMUNT, 2002).

Así, a partir del principio de *independencia local*, citado con anterioridad, la densidad condicional, de que un determinado individuo con un patrón de respuesta  $\mathbf{x}$  pertenezca a una clase latente, está dada por,

$$\pi_{X/Y(c)}(\mathbf{x}) = \prod_{j=1}^p \pi_{X_j/Y(c)}(x_j) \quad (1.1)$$

donde:

$$\pi_{X_j/Y(c)}(x_j) = P(X_j = x_j/Y = c)$$

$$x_j = 1, \dots, I_j$$

$$c = 1, \dots, T$$

La distribución conjunta de  $\mathbf{X}$  e  $Y$  está dada por,

$$\pi_{X,Y}(\mathbf{x}, c) = \pi_Y(c) \pi_{X/Y(c)}(\mathbf{x}) \quad (1.2)$$

donde  $\pi_Y(c) = P(Y = c)$  representa la proporción de elementos que se encuentran en la clase latente  $c$ , también conocida como probabilidad *a priori*.

Utilizando las expresiones anteriores, el MCL se expresa como,

$$\pi_X(\mathbf{x}) = \sum_{c=1}^T \pi_Y(c) \prod_{j=1}^p \pi_{X_j/Y(c)}(x_j) \quad (1.3)$$

donde  $\pi_{X_j/Y(c)}(x_j)$  = probabilidad de respuesta condicional de cada una de las variables manifiestas dentro de la clase latente  $c$ , para  $c = 1, \dots, T$ ;  $j = 1, \dots, p$ ;  $x_j = 1, \dots, I_j$ .

Los parámetros de un modelo de clases latentes, es decir las probabilidades de clase ( $\pi_Y(c)$ ) y las probabilidades condicionales ( $\pi_{X_j/Y(c)}(x_j)$ ), están sujetos a las siguientes restricciones,

$$\sum_{c=1}^T \pi_Y(c) = 1 \quad \sum_{x_j=1}^{I_j} \pi_{X_j/Y(c)}(x_j) = 1 \quad (1.4)$$

De la exposición anterior, hay que distinguir dos aspectos importantes de las clases latentes: el número de clases y su tamaño relativo. El número de clases ( $T$ ) de la variable latente ( $Y$ ) representa el número de tipologías definidas por el modelo para los valores de las variables observadas en la tabla de contingencia. El tamaño relativo de cada clase latente también proporciona información significativa para la interpretación de las probabilidades de las clases latentes,

indicando el tipo de distribución de la población de elementos en las diferentes clases.

Es evidente, que un tratamiento exhaustivo de todas las formas de diseño de un modelo de clases latentes, excede los objetivos de esta investigación. De forma que centraremos nuestro planteamiento en un modelo de clases latentes con variables manifiestas binarias (o dicotómicas). Supongamos que hay  $p$  variables manifiestas binarias  $X_1, X_2 \dots X_p$  con  $X_j = 0$  ó  $1$  para toda  $j$ , las cuales conforman un vector  $p$ -dimensional llamado patrón de respuesta. Las  $p$  variables se consideran indicadoras de una variable latente  $Y$  con  $T$  clases latentes.

Pues bien, puesto que la asunción de *independencia local* supone que las variables manifiestas son estadísticamente independientes para los individuos que tengan la misma posición en la variable latente, la probabilidad condicionada de observar un patrón de respuesta  $\mathbf{x}$  podrá expresarse como el producto de las probabilidades de respuesta condicionadas para cada una de las diferentes variables manifiestas, esto es:

$$\pi_{X_j/Y(c)}(\mathbf{x}) = \pi_{jc}^{x_j} (1 - \pi_{jc})^{1-x_j} \quad (1.5)$$

donde  $\pi_{jc}$  es la probabilidad condicional de obtener una respuesta positiva en la variable  $X_j$  para un individuo de la clase latente  $c$ . Así, el MCL lo podemos reescribir como:

$$\pi_{\mathbf{x}}(\mathbf{x}) = \sum_{c=1}^T \pi_Y(c) \prod_{j=1}^p \pi_{jc}^{x_j} (1 - \pi_{jc})^{1-x_j} \quad (1.6)$$

Para realizar las estimaciones de los parámetros del modelo, es decir, las probabilidades de clase, conjuntas y condicionales, se utilizan procedimientos iterativos basados en estimaciones de máxima verosimilitud. Los más conocidos son el algoritmo de Newton-Raphson (HABERMAN, 1979) y el algoritmo Esperanza-Maximización (EM) (DEMPSTER et al., 1977). En la siguiente sección se estudian con algún detalle los algoritmos, principalmente el algoritmo EM, que lo utilizaremos posteriormente para el ajuste de un modelo de clases latentes.



### 1.3 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Se utilizan distintos métodos de estimación de los parámetros de un modelo de clases latentes, entre los cuales los más conocidos son el algoritmo de Newton-Raphson y el algoritmo Esperanza-Maximización (EM).

El algoritmo de Newton-Raphson se basa en el desarrollo de series de Taylor y requiere una estimación inicial para los valores que maximizan la función de verosimilitud. A continuación, dicha función se aproxima en un entorno de la mencionada estimación inicial mediante un polinomio de segundo grado, de forma que la segunda estimación será aquella que arroje el máximo valor del polinomio.

La función de verosimilitud se vuelve a aproximar en un entorno de la segunda estimación, obteniéndose la tercera estimación como el valor que maximiza el polinomio, y así sucesivamente. A través de este procedimiento se van generando una secuencia de estimaciones, de forma que, finalmente, las estimaciones convergerán al máximo, siempre y cuando la función sea apropiada y/o la estimación inicial sea buena. Para mayores detalles sobre este algoritmo de estimación puede consultarse ANDERSEN (1991).

Un enfoque alternativo (propuesto por DEMPSTER et al., 1977) es estimar el modelo con el algoritmo EM. Es probablemente el procedimiento más utilizado en el análisis de clases latentes. Básicamente, el algoritmo EM es un método que permite encontrar los estimadores máximos verosímiles de los parámetros de la distribución subyacente de un conjunto de datos, cuando los datos son incompletos o existen datos faltantes. En un MCL los “datos faltantes” son las clases a las cuales los individuos pertenecen y los “estimadores” de estos valores son las probabilidades *a posteriori*,  $\pi_{Y/X}(c)$ . Consecuentemente, la estimación de las probabilidades  $\pi_{Y/X}(c)$  constituye el paso **E** del algoritmo y la inserción de estos valores para obtener los estimadores mejorados constituyen el paso **M** (SEPÚLVEDA, 2004).

El algoritmo EM es un procedimiento iterativo y cada iteración está compuesta por dos pasos:

Paso 1. Paso de esperanza (E). Se calculan los valores esperados dados los datos observados y los “actuales” parámetros del modelo, que constituyen las probabilidades *a posteriori*,  $\pi_{Y/X(x_h)}(c)$ .

Paso 2. Paso de maximización (M). Se maximiza la función de verosimilitud (L) de los datos a partir de los valores esperados calculados en el paso anterior. Para el caso de variables manifiestas binarias, el logaritmo de la función de verosimilitud vendrá dado por,

$$\begin{aligned} \ln(L) &= \sum_{h=1}^R f_{x_h} \ln \left[ \sum_{c=1}^T \pi_Y(c) \prod_{j=1}^p \pi_{jc}^{x_j} (1 - \pi_{jc})^{1-x_j} \right] \\ &= \sum_{h=1}^R f_{x_h} \ln[\pi_X(\mathbf{x})] \end{aligned} \quad (1.7)$$

donde:  $f_{x_h}$  = Frecuencia observada del patrón de respuesta h.

$\pi_X(\mathbf{x})$  = Probabilidad del patrón de respuesta h.

$R$  = Número posible de patrones de respuesta ( $2^p$ ).

La estimación de las probabilidades constituye el *Paso 1* del algoritmo y la inserción de estos valores para obtener los estimadores mejorados constituyen el paso de maximización. Este proceso iterativo de estimación finalizará cuando se alcance un número predeterminado de iteraciones o cuando la diferencia entre las estimaciones máximo verosímiles de una iteración y las de la iteración inmediatamente anterior sea menor a una cantidad fijada previamente y denominada nivel de tolerancia, por ejemplo,  $10^{-6}$ . Es indudable que la opción de detener el proceso dado un número máximo de iteraciones, tiene el inconveniente que el algoritmo de estimación, puede converger en un número menor o mayor de iteraciones que las preestablecidas.

En general, el algoritmo EM es sencillo tanto en la teoría como en el cálculo, y los valores iniciales elegidos aleatoriamente son suficientes para llegar a una solución. El número de iteraciones necesarias para la convergencia del método será menor si estos valores se encuentran “cercaños” a los alcanzados por los estimadores máximos verosímiles.

Un problema que puede presentarse al utilizar el algoritmo EM es la no identificabilidad de un modelo de clases latentes. Un modelo es identificable cuando existe una solución única y basta con que haya más de una solución para que el modelo no sea identificable. El diagnóstico se hace usando diferentes valores iniciales. Si con esos valores iniciales, el modelo proporciona el mismo valor de la función de verosimilitud pero distintas estimaciones de los parámetros, el modelo es no identificable.

Una vez solucionado el problema de estimaciones de los parámetros de modelo, resulta conveniente analizar en qué medida el modelo de clases latentes ajustado a los datos es el adecuado. Para esta cuestión, se utilizan los estadísticos de bondad de ajuste que serán estudiados en la siguiente sección.

#### 1.4 BONDAD DE AJUSTE DEL MODELO

Las técnicas de bondad de ajuste se usan frecuentemente para decidir si una muestra observada  $\mathbb{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ , de patrones de respuesta binarios pueden ser considerados como un conjunto de realizaciones independientes de una función de distribución dada  $F_0(\mathbb{X})$ ; es decir, se usan para contrastar la hipótesis  $H_0: F(\mathbb{X}) = F_0(\mathbb{X})$ .

En cuanto a la evaluación de la bondad de ajuste del modelo de clases latentes, se puede diagnosticar comparando las frecuencias observadas para cada patrón de respuesta, con las frecuencias estimadas si la muestra siguiese el modelo teórico hipotético, valiéndonos de los estadísticos de bondad de ajuste (EBA). Aunque, evidentemente una forma muy simple de realizar esta comparación es la que consiste en inspeccionar los residuos, que se definen como

la diferencia entre las frecuencias observadas y las frecuencias esperadas bajo el modelo.

Los estadísticos de bondad de ajuste frecuentemente aplicados son: el estadístico  $Q^2$  de Pearson, la razón de verosimilitud ( $G^2$ ), el estadístico de Cressie-Read (CR) y el estadístico de Freeman-Tukey ( $FT^2$ ). Para el cálculo de los EBA resulta trascendental definir las frecuencias esperadas dadas por la expresión:

$$\hat{f}_{\mathbf{x}_h} = n\hat{p}_{\mathbf{x}_h} = n \left[ \sum_{c=1}^T \hat{\pi}_Y(c) \prod_{j=1}^p \hat{\pi}_{X_j/Y(c)}(x_j) \right] \quad (1.8)$$

Evidentemente, si las frecuencias observadas se acercan a las esperadas se obtendría un valor pequeño del estadístico de bondad de ajuste y el modelo no se rechaza. Por el contrario, cuando existan considerables diferencias entre las frecuencias observadas y esperadas, el valor del estadístico de bondad de ajuste sería grande y la bondad de ajuste del modelo es mala, rechazándose el modelo hipotético.

El estadístico  $Q^2$  de Pearson se emplea para decidir cuándo un modelo de clases latentes es apropiado. Se calcula mediante la fórmula:

$$Q^2 = \sum_{h=1}^R \frac{(f_{\mathbf{x}_h} - \hat{f}_{\mathbf{x}_h})^2}{\hat{f}_{\mathbf{x}_h}} \quad (1.9)$$

donde:  $f_{\mathbf{x}_h}$  = Frecuencia observada del patrón de respuesta  $\mathbf{x}_h$ .  
 $\hat{f}_{\mathbf{x}_h}$  = Frecuencia esperada del patrón  $\mathbf{x}_h$ .  
 $R$  = Número posible de patrones de respuesta ( $2^p$ ).

Dado que el principal problema de estudio de la tesis doctoral, es la distribución de probabilidad de los estadísticos de bondad de ajuste, nos parece conveniente demostrar que estos estadísticos tienen una distribución teórica Chi-Cuadrado.

Bajo esta suposición, el estadístico  $Q^2$  de Pearson posee asintóticamente una distribución  $\chi^2$  con  $gl$  grados de libertad cuando  $H_0$  sea verdadera. Para demostrar el anterior supuesto, consideremos una muestra aleatoria de tamaño  $n$  proveniente de una distribución multinomial, donde  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R$  representan los distintos patrones de respuesta y los términos  $f_{x_1}, f_{x_2}, \dots, f_{x_R}$  son las frecuencias observadas de cada patrón, respectivamente. De tal manera que, la suma de las frecuencias es igual al tamaño de la muestra,  $n = f_{x_1} + f_{x_2} + \dots + f_{x_R}$ . Además, se tiene que la frecuencia esperada del patrón  $\mathbf{x}_h$  es,  $\hat{f}_{x_h} = n\hat{p}_{x_h}$  tal que  $\hat{p}_{x_1} + \hat{p}_{x_2} + \dots + \hat{p}_{x_R} = 1$ . Como demostración que el estadístico  $Q^2$  tiene una distribución teórica esperada Chi-Cuadrado, resulta suficiente demostrarlo para  $R = 2$  patrones de respuesta (BARTOSZYNSKI & NIEWIADOMSKA, 1996).

$$\begin{aligned} Q^2 &= \sum_{h=1}^R \frac{(f_{x_h} - \hat{f}_{x_h})^2}{\hat{f}_{x_h}} \\ &= \frac{(f_{x_1} - \hat{f}_{x_1})^2}{\hat{f}_{x_1}} + \frac{(f_{x_2} - \hat{f}_{x_2})^2}{\hat{f}_{x_2}} \end{aligned}$$

Considerando que la probabilidad del segundo patrón está dada por  $\hat{p}_{x_2} = 1 - \hat{p}_{x_1}$  y su frecuencia observada por  $f_{x_2} = n - f_{x_1}$ , tenemos que,

$$\begin{aligned} Q^2 &= \frac{(f_{x_1} - n\hat{p}_{x_1})^2}{n\hat{p}_{x_1}} + \frac{[(n - f_{x_1}) - n(1 - \hat{p}_{x_1})]^2}{n(1 - \hat{p}_{x_1})} \\ &= \left[ \frac{f_{x_1} - n\hat{p}_{x_1}}{\sqrt{n\hat{p}_{x_1}\hat{q}_{x_1}}} \right]^2 = [Z_1]^2 \end{aligned}$$

Lo anterior está fundamentado en la aproximación de la distribución Binomial por la distribución Normal. La distribución Binomial tiene una esperanza  $\mu = n\hat{p}_{x_1}$  y una desviación típica  $\sigma = \sqrt{n\hat{p}_{x_1}\hat{q}_{x_1}}$ . Por el llamado Teorema Integral de De Moivre–Laplace, si  $f_{x_h} \rightarrow \infty$  y  $n \rightarrow \infty$  se verifica que  $Z_1$  converge hacia una variable aleatoria con distribución normal estándar  $N(0,1)$ , esto significa que tiene media cero ( $\mu = 0$ ) y desviación típica uno ( $\sigma = 1$ ). Una demostración clara del

teorema exige cálculos un poco largos y delicados que sobrepasan los marcos de este trabajo. La demostración del teorema puede encontrarse en FELLER (1950).

La distribución  $\chi^2$  está en estrecha relación con la distribución normal<sup>3</sup>. Sea  $Z_1$  una variable aleatoria con distribución  $N(0,1)$ , entonces la variable aleatoria  $Y = Z_1^2$  posee una distribución  $\chi^2$  con 1 grado de libertad. De esta manera, la distribución del estadístico de bondad de ajuste  $Q^2$  de Pearson es  $\chi^2$  con los grados de libertad definidos por el modelo de clases latentes. Las condiciones necesarias para poder aplicar el estadístico, es que tanto la frecuencia observada de los patrones como el tamaño de la muestra sean suficientemente grandes para que la utilización teórica  $\chi^2$  sea válida. Sin embargo, no es suficiente un tamaño de muestra grande, ya que las observaciones pueden tener una distribución donde unos pocos patrones de respuestas son observados y en consecuencia provoca una invalidación del uso de la distribución  $\chi^2$ .

Otra alternativa para analizar la bondad de ajuste del modelo de clases latentes es la Razón de Verosimilitud ( $G^2$ ). La razón se define como:

$$G^2 = 2 \sum_{h=1}^R f_{x_h} \log \left( \frac{f_{x_h}}{\hat{f}_{x_h}} \right) \quad (1.10)$$

donde:  $f_{x_h}$  = Frecuencia observada del patrón de respuesta  $x_h$ .

$\hat{f}_{x_h}$  = Frecuencia esperada del patrón  $x_h$ .

$R$  = Número posible de patrones de respuesta.

Análogamente al estadístico  $Q^2$  de Pearson, se parte del supuesto de que la Razón de Verosimilitud tiene una distribución teórica esperada Chi-Cuadrado. Una forma de comprobar la validez del supuesto, es considerar que las frecuencias observadas de los patrones tienden a ser grandes  $f_{x_h} \rightarrow \infty$  y un tamaño de muestra

---

<sup>3</sup> La Distribución  $\chi^2$  fue descubierta en 1876 por R. Helmert (como distribución de la suma de cuadrados de variables aleatorias independientes con distribución  $N(0,1)$ ) y vuelta a hallar en 1900 por K. Pearson, fundador en Inglaterra de una escuela de Estadística de altos rendimientos; por eso esta distribución se denomina de Helmert o de Helmert-Pearson.

tal que  $n \rightarrow \infty$ , cuando el modelo hipotético de clases latentes es verdadero ( $H_0$ ), el estadístico  $G^2$  es asintóticamente equivalente a  $Q^2$ .

Esta proposición se puede demostrar reescribiendo la Razón de Verosimilitud como,

$$G^2 = 2 \sum_{h=1}^R f_{x_h} \log \left( 1 + \frac{f_{x_h} - \hat{f}_{x_h}}{\hat{f}_{x_h}} \right)$$

considerando,

$$k = \frac{f_{x_h} - \hat{f}_{x_h}}{\hat{f}_{x_h}}$$

y utilizando la serie de Taylor,

$$\log(1 + k) = k + \frac{k^2}{2} + \dots \Rightarrow |k| < 1,$$

definimos el estadístico  $G^2$  como,

$$G^2 \approx \sum_{h=1}^R f_{x_h} \frac{f_{x_h} - \hat{f}_{x_h}}{\hat{f}_{x_h}} + \sum_{h=1}^R \frac{(f_{x_h} - \hat{f}_{x_h})^2}{\hat{f}_{x_h}} \times \frac{f_{x_h}}{\hat{f}_{x_h}}$$

por consiguiente, si  $f_{x_h} \rightarrow \infty$

$$\frac{f_{x_h}}{\hat{f}_{x_h}} \rightarrow 1$$

y además

$$E \left[ \sum_{h=1}^{\infty} (f_{x_h} - \hat{f}_{x_h}) \right] = 0$$

entonces, se cumple

$$G^2 \approx \sum_{h=1}^R \frac{(f_{x_h} - \hat{f}_{x_h})^2}{\hat{f}_{x_h}}$$

En definitiva, hemos probado que cuando el modelo hipotético de clases latentes es verdadero, la diferencia entre el estadístico de Pearson ( $Q^2$ ) y la razón de verosimilitud ( $G^2$ ) converge a cero.

Por otra parte, FREEMAN & TUKEY (1950) introdujeron el estadístico de bondad de ajuste  $FT^2$  para probar  $H_0$ , dado por la expresión,

$$FT^2 = 4 \sum_{h=1}^R \left( \sqrt{f_{x_h}} - \sqrt{\hat{f}_{x_h}} \right)^2 \quad (1.11)$$

donde:  $f_{x_h}$  = Frecuencia observada del patrón de respuesta  $x_h$ .

$\hat{f}_{x_h}$  = Frecuencia esperada del patrón  $x_h$ .

El estadístico  $FT^2$  tiene una media y una varianza que tienden a ser similares a las respectivas en la distribución Chi-Cuadrado (READ, 1993). Este es una variante de la clase de estadísticos definidos como,

$$FT^2(b_0, b_1, \dots, b_r) = \sum_{h=1}^R \left\{ \sum_{j=1}^k b_j \left( \sqrt{\frac{f_{x_h}}{\hat{f}_{x_h}}} \right)^j \right\} \left( \sqrt{f_{x_h}} - \sqrt{\hat{f}_{x_h}} \right)^2 \quad (1.12)$$

para

$$\sum_{j=1}^R b_j = 4$$

donde  $k$  es un entero no negativo.

En resumen, los estadísticos mencionados anteriormente siguen una distribución teórica  $\chi^2$  cuando las tablas múltiples no presentan el problema de frecuencias pequeñas en los patrones de respuestas. En caso contrario, si las tablas son poco ocupadas (o “sparse data” en inglés) el supuesto de la distribución teórica no se cumple y será necesario utilizar otros métodos con el fin de evaluar la bondad de ajuste del modelo. *Este problema de las frecuencias pequeñas en las celdas de la tabla múltiple y las posibles soluciones para la evaluación de un modelo de clases latentes, son la esencia del presente proyecto, por lo cual serán abordados más adelante en la presente investigación.*

Una alternativa de solución al problema de celdas con frecuencias bajas, es presentada por READ & CRESSIE (1988). La familia de estadísticos de divergencia tiene la siguiente expresión,



$$CR(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{h=1}^R f_{x_h} \left[ \left( \frac{f_{x_h}}{\hat{f}_{x_h}} \right)^\lambda - 1 \right] \quad \lambda \neq 0 \quad (1.13)$$

donde:  $f_{x_h}$  = Frecuencia observada del patrón de respuesta  $x_h$ .

$\hat{f}_{x_h}$  = Frecuencia esperada del patrón  $x_h$ .

Dependiendo del valor de  $\lambda$  (lambda) este estadístico toma diferentes formas (**Tabla 1.1**). Si  $\lambda = 0$ , se obtiene el estadístico de Razón de Verosimilitud ( $G^2$ ); si  $\lambda = 1$  se tiene el estadístico de Pearson ( $Q^2$ ) cuando  $\lambda = -1/2$  resulta el estadístico de Freeman Tukey.

Read y Cressie recomiendan considerar  $\lambda = 2/3$ , ya que cuando se trabaja con “grandes” tablas de datos, tiene un comportamiento mejor que los estadísticos  $Q^2$  y  $G^2$  (VERMUNT, 1997a). Los estadísticos de la familia (1.13) tienen la misma distribución asintótica y ésta es  $\chi^2$  para un determinado  $\lambda$  dado  $H_0$  con grados de libertad representados por (1.14).

Estadístico	Cálculo	$\lambda$
$Q^2$	$\sum_{h=1}^R \frac{(f_{x_h} - \hat{f}_{x_h})^2}{\hat{f}_{x_h}}$	1
$CR$	$\frac{2}{\lambda(\lambda + 1)} \sum_{h=1}^R f_{x_h} \left[ \left( \frac{f_{x_h}}{\hat{f}_{x_h}} \right)^\lambda - 1 \right]$	$\frac{2}{3}$
$G^2$	$2 \sum_{h=1}^R f_{x_h} \log \left( \frac{f_{x_h}}{\hat{f}_{x_h}} \right)$	0
$FT^2$	$4 \sum_{h=1}^R \left( \sqrt{f_{x_h}} - \sqrt{\hat{f}_{x_h}} \right)^2$	$-\frac{1}{2}$

**Tabla 1.1.** Estadísticos de bondad de ajuste y su cálculo en términos del estadístico Read-Cressie.

El número de grados de libertad ( $gl$ ) de los estadísticos de bondad de ajuste para la distribución  $\chi^2$ , se obtiene a partir de la diferencia entre el número de

celdas de la tabla múltiple menos el número de parámetros a estimar en el modelo, o de igual forma,

$$gl = (2^p - 1) - [T * p + (T - 1)] \quad (1.14)$$

donde:  $p$  = Número de variables manifiestas binarias.

$T$  = Número de clases latentes para la variables latente Y.

Si algunas de las frecuencias esperadas ( $\hat{f}_{x_h}$ ) son ceros estructurales<sup>4</sup> o ceros aleatorios<sup>5</sup>, no podrán ser estimados los parámetros, aunque realmente existan. Si una celda de una tabla múltiple contiene un cero estructural, el correspondiente parámetro no existe, mientras que si se trata de un cero aleatorio, dicho parámetro sí que existe, pero no puede estimarse a partir del conjunto de datos observados. En este sentido, CLOGG & GOODMAN (1984) mostraron que el número de grados de libertad pasaría a ser igual al número de celdas sin ceros, menos el número de parámetros estimables.

## 1.5 CRITERIOS DE SELECCIÓN DEL MODELO

Estadísticamente, pueden existir cientos de modelos para un sólo conjunto de datos, que se ajusten con la misma calidad. A partir de la teoría de la información, es posible desarrollar una forma de seleccionar el modelo más adecuado. El objetivo no es descubrir el modelo “verdadero”, sino aquel que proporciona mayor información sobre la realidad. Por un lado, las frecuencias estimadas deben ser parecidas a las observadas y, por otro, el modelo debe ser tan reducido como sea posible. La idea principal es que, dados dos modelos con igual valor en la función verosimilitud, el mejor modelo es el que tiene el menor número de parámetros.

---

<sup>4</sup> Los *ceros estructurales* surgen en aquellas casillas en las que es teóricamente imposible obtener observaciones, es decir, en las que se sabe a priori que dicha casilla contendrá un valor nulo, puesto que el valor esperado de la misma es cero.

<sup>5</sup> Cuando el número de observaciones con el que se construye una tabla de contingencia es pequeño, o es suficientemente grande pero el número de celdas de la tabla es elevado como consecuencia de la inclusión en la misma de muchas variables o de variables (aunque sean pocas) que poseen múltiples categorías, puede ser relativamente frecuente encontrar en dichas tablas una o varias celdas (o patrones de respuestas) con frecuencia observada nula. Cuando esto ocurre, se estará ante un cero aleatorio si el valor observado en dicha casilla es cero pero el valor esperado de la misma es positivo.

Los contrastes más conocidos basados en la teoría de la información son el criterio de información de Akaike (AIC) (AKAIKE, 1987) y el criterio de información bayesiano (BIC) (SCHWARZ ,1978). El criterio AIC (en inglés, "An Information Criterion"), se basa en la medida de información de KULLBACK-LEIBLER (1951), la cual permite interpretar la distancia entre dos distribuciones (en nuestro caso, la observada a partir de la muestra y la teórica) a partir del log-verosimilitud de un modelo, en otras palabras, penalizando al modelo según su grado de complejidad y determinando hasta qué punto un modelo concreto se desvía de la realidad. Su expresión es:

$$AIC = -2 \ln(\ell) + 2m \quad (1.15)$$

donde:  $\ell$  = Valor de la función de verosimilitud.

$m$  = Número de parámetros desconocidos.

Cuando el número de parámetros de un modelo aumenta, el criterio de información de Akaike también, por tanto, escoger el modelo que tiene el mínimo criterio de información de Akaike supone elegir el modelo con el menor número de parámetros posible.

SCHWARZ (1978) propone el criterio bayesiano para desarrollar una medida consistente asintóticamente basada en el logaritmo de la función de verosimilitud ( $\ell$ ), el número de parámetros independientes a ser estimados ( $m$ ) y el tamaño muestral, llamada BIC (en inglés "Bayesian Information Criteria") que se calcula de la siguiente forma:

$$BIC = -2 \ln(\ell) + m \times \ln(n) \quad (1.16)$$

En general, para ambos valores de los estadísticos AIC y BIC, cuanto menores sean mejor será el modelo, porque mayor información contendrá.

En el próximo capítulo, vamos a estudiar los Métodos Bootstrap, que son usados para encontrar la distribución empírica de un estadístico cuando el objetivo de un estudio es de tipo inferencial y no conocemos su distribución de probabilidad o los supuestos de los modelos probabilísticos no se cumplen.

## **CAPÍTULO 2**

---

### **EL MÉTODO BOOTSTRAP**

---

## 2.1 INTRODUCCIÓN

En términos generales, es conocido que los modelos estadísticos clásicos incorporan distintos supuestos teóricos (por ejemplo: normalidad e independencia) que establecen restricciones sobre las variables aleatorias analizadas o sobre la distribución esperada de los estadísticos. La violación de los supuestos incide sobre la estimación de la probabilidad asociada al estadístico, fundamento de la decisión estadística. La pérdida de potencia y la subestimación de la probabilidad de error tipo I constituyen aspectos fundamentales en la aplicación de la estadística.

Sobre este tópico, el método Bootstrap<sup>1</sup> deriva la distribución de los estimadores empíricamente, utilizando para esto la información de la muestra original. De esta forma, el enfoque Bootstrap permite eludir las limitaciones del enfoque estadístico clásico, que resultan principalmente de asumir supuestos, que en muchas ocasiones prácticas resultan poco realistas. En este sentido, el método Bootstrap es una metodología de inferencia estadística. Este proceder ha sido considerado por la American Statistical Association como “el único gran descubrimiento en estadística desde 1970” (KOTZ & JOHNSON, 1992).

Entre las aplicaciones del procedimiento Bootstrap están: problemas de estimación, ajuste de modelos de regresión, contraste de hipótesis, análisis de componentes principales, entre otras, presentadas en EFRON & TIBSHIRANI (1993).

En este capítulo, se realiza una presentación de los fundamentos teóricos del método Bootstrap. Una vez descritos formalmente, se proponen para la solución del problema de la selección de modelos de clases latentes en tablas poco ocupadas.

---

<sup>1</sup> El término Bootstrap, es una expresión inglesa que significa “levantarse tirando hacia arriba de las propias correas de las botas”. Refleja el aspecto fundamental de esta técnica, su autosuficiencia. Se podría traducir al español como remuestreo, si bien hay que reconocer que esta palabra no está incluida dentro del Diccionario de la Real Academia Española.

## 2.2 FUNDAMENTOS TEÓRICOS

En primer lugar, debemos comentar que el término remuestreo, es aplicado a aquellas técnicas de simulación empleadas en la teoría de probabilidades y la inferencia estadística que, a partir de los datos observados, generan nuevas muestras simuladas de igual tamaño que la muestra original, con el propósito de examinar los resultados obtenidos en esas remuestras y seguidamente, estimar inductivamente la forma de la distribución muestral de los estadísticos, en lugar de partir de una distribución teórica supuesta *a priori*. En este sentido, el procedimiento Bootstrap es útil para la descripción de la distribución muestral de aquellos estimadores con propiedades muestrales desconocidas o difícilmente obtenibles por medios matemáticos (EFRON, 1979).

La mayoría de la estadística convencional, se apoya en la suposición de que los estadísticos resultantes del análisis de los datos, como la media aritmética, desviación típica, razón de verosimilitud, etc., se distribuyen de acuerdo con un cierto modelo probabilístico, por ejemplo: la distribución normal, la distribución Chi-Cuadrado ( $\chi^2$ ), la distribución F de Fisher, etc., entre los más utilizados. Bajo el supuesto de un modelo hipotético para el estadístico es posible contrastar hipótesis. A modo de ejemplo, para el contraste de hipótesis de bondad de ajuste utilizando el estadístico  $Q^2$  de Pearson, se parte de la suposición de que la distribución muestral del estadístico  $Q^2$  es  $\chi^2$ .

Cuando la distribución límite de los estadísticos de contraste no es la esperada, no pueden utilizarse los valores críticos tabulados para contrastar hipótesis. Basándonos en el método Bootstrap, la idea de fondo es construir un modelo de probabilidad para determinados estadísticos a partir de la información proporcionada por la muestra, obviando los supuestos sobre la distribución teórica que son la base para hacer inferencias sobre la población. En este sentido, supongamos que  $X$  es una variable aleatoria (v.a.) y se tiene una muestra aleatoria de tamaño  $n$  ( $x_1, x_2, \dots, x_n$ ), con función de distribución de probabilidades dada por  $F(x) = Pr(X \leq x)$ . Lo cual se escribe  $(x_1, x_2, \dots, x_n) \sim F(\mathbf{x})$  o simplemente  $x \sim F(\mathbf{x})$ .

Cuando el valor del parámetro  $\theta$  de una población es desconocido, y en consecuencia se desea utilizar el estimador  $\hat{\theta} = f(x_1, x_2, \dots, x_n, \theta)$ , la distribución de  $\hat{\theta}$  es aproximada generando un conjunto de resultados independientes  $\hat{\theta}_b^*$  para  $b = 1, 2, \dots, B$  y construyendo la distribución empírica  $F_{\hat{\theta}}$ . El número de remuestras Bootstrap ( $B$ ) está determinado por el objetivo del estudio. Si es utilizado el Bootstrap para la estimación de la distribución empírica de  $\hat{\theta}$ ,  $B$  tiene que ser muy grande. En caso de que el método Bootstrap sea usado para pruebas de hipótesis, el número de remuestras Bootstrap puede ser muy pequeño (EFRON & TIBSHIRANI, 1993).

Como resultado de la aplicación del método Bootstrap, cada remuestra será muy probablemente, algo diferente de la muestra original; con lo cual, un estadístico  $\hat{\theta}^*$ , calculado a partir de una de esas remuestras tomará un valor diferente del que produce otra remuestra y del  $\hat{\theta}$  observado.

La afirmación fundamental del Bootstrap es que una distribución de frecuencias de esos  $\hat{\theta}^*$ , calculadas a partir de las remuestras, es una estimación de la distribución muestral de  $\hat{\theta}$  (MOONEY & DUVAL, 1993). Como se advierte de lo anterior, el método Bootstrap no representa un estadístico en sí mismo, sino un enfoque sobre cómo usar los estadísticos de cara a hacer inferencias acerca de los parámetros.

### 2.3 VERSIONES DEL MÉTODO BOOTSTRAP

El método Bootstrap puede ser paramétrico o no, en función de que el remuestreo se produzca sobre una distribución teórica o una distribución empírica. Las principales diferencias entre las distintas variaciones del método Bootstrap se hallan: en la forma de tomar las remuestras, en las variables aleatorias y en el objetivo de la aplicación (estimación o contraste de hipótesis). Sin pretender revisar todas las variaciones del método Bootstrap en este apartado son presentados únicamente el método Bootstrap No Paramétrico (BNP) y Paramétrico (BP). Además, los procedimientos de cálculo del nivel de

significación, sesgo y error estándar, para finalmente comentar algunos problemas de este método.

### 2.3.1 BOOTSTRAP NO PARAMÉTRICO

El Bootstrap No Paramétrico (en inglés, “Naive Bootstrap”) corresponde al descrito por EFRON (1979). Se basa en que la distribución de la muestra es la mejor representación de la distribución de la población y remuestrear la muestra nos permite ver lo que sucedería si remuestreásemos la población. Los datos observados son remuestreados con reemplazamiento para generar nuevas muestras que serán usadas para ajustar el modelo. A manera de ejemplo, si la muestra original es  $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ , dos posibles muestras Bootstrap pueden ser las siguientes:

$$\mathbf{x}_1^* = (x_3, x_2, x_4, x_1, x_3)$$

$$\mathbf{x}_2^* = (x_5, x_5, x_3, x_2, x_5)$$

Cualquier dato de la muestra original puede aparecer ninguna, una o varias veces repetido.

En general, se tiene una muestra aleatoria de tamaño  $n$ , en donde  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  son los valores observados de dicha muestra. Muestreando  $B$  veces con reemplazamiento la muestra original  $(x_1, x_2, \dots, x_n)$  se obtienen las remuestras Bootstrap las cuales están denotadas por  $\mathbf{x}_b^* = (x_{1b}, x_{2b}, \dots, x_{nb})$  para  $b = 1, 2, \dots, B$ . De esta manera, a partir de la muestra aleatoria se define la función de distribución empírica  $\hat{F}$  como la distribución discreta que asigna una probabilidad  $1/n$  a cada elemento de la muestra original.

En relación al caso de contraste de hipótesis, el Bootstrap No Paramétrico (BNP) implica desarrollar un proceso en el que distinguimos los siguientes pasos:

1. Dada una muestra de tamaño  $n$ , calcular el estadístico de interés ( $\hat{\theta}$ ).



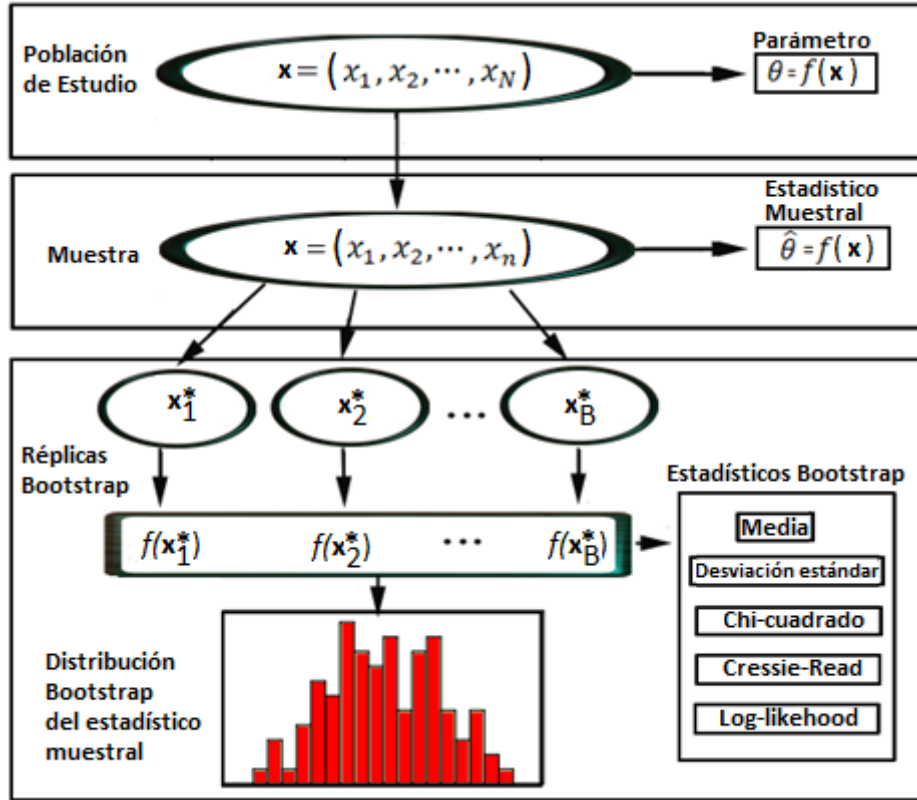
2. Generar una muestra Bootstrap de tamaño  $n$  mediante muestreo con reemplazamiento de la muestra original, asignando a cada observación una probabilidad  $1/n$  de selección. Esta es denominada “remuestra” y se representa como  $\mathbf{x}_b^* = (x_{1b}, x_{2b}, \dots, x_{nb})$ .
3. Para la muestra obtenida se calcula el valor del estadístico  $\hat{\theta}_b^*$  que se utiliza como estimador del parámetro poblacional  $\theta$ .
4. Repetimos los pasos 2 y 3 hasta obtener  $B$  estimaciones  $\hat{\theta}_b^*$ .
5. Construir la distribución Bootstrap del estadístico asignando una probabilidad de  $1/B$  a cada punto  $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*)$ , que representa una aproximación de la verdadera función de probabilidad para ese estadístico.
6. Determinar la proporción de réplicas mayores que el estimador original  $(\hat{\theta})$ .
7. Tomar la decisión sobre la hipótesis nula.

De manera esquemática, el proceso de estimación de la función de distribución empírica mediante el Bootstrap No Paramétrico, puede representarse mediante la **Figura 2.1**.

Nótese que los valores correspondientes a cada remuestra ( $\mathbf{x}_b^*$ ) nos permiten obtener la función de distribución empírica  $F_{\hat{\theta}}$ , que constituye la estimación máximo verosímil no paramétrica de la función de distribución  $F_{\theta}$  (EFRON, 1979). Por tanto, es factible establecer  $F_{\theta} = F_{\hat{\theta}}$ , significando que la función de distribución del estadístico  $\theta$  se estima a partir de la función de distribución empírica, constituyendo el aspecto fundamental del Bootstrap No Paramétrico.

En términos matemáticos, la función de distribución empírica  $F_{\hat{\theta}}$  se puede definir de la siguiente manera. Supongamos una muestra de observaciones

$(x_1, x_2, \dots, x_n)$  para una v.a.  $X$ . A continuación, por medio del método Bootstrap No Paramétrico se obtienen remuestras Bootstrap, denotadas por  $\mathbf{x}_b^* = (x_1, x_2, \dots, x_n)$  donde  $b = 1, 2, \dots, B$ . Para cada remuestra se calcula el estimador  $\hat{\theta}_b^*$  y se ordenan de menor a mayor,  $(\hat{\theta}_1^* < \hat{\theta}_2^* < \dots < \hat{\theta}_B^*)$ .



**FIGURA 2.1.** Algoritmo Bootstrap No-Paramétrico.

La función de distribución empírica  $F_{\hat{\theta}}$  es la proporción de  $\hat{\theta}_i^*$  menor o igual a un valor  $\hat{\theta}^*$ , es decir:

$$F_B(\hat{\theta}^*) = \frac{1}{n} \sum_{i \leq B} I(\hat{\theta}_i^* \leq \hat{\theta}^*)$$

$F_B(\hat{\theta}^*)$  es constante entre dos valores consecutivos, siendo  $1/B$  la diferencia entre un valor y el consecutivo. En efecto, si la probabilidad acumulada hasta  $\hat{\theta}_k^*$  es:

$$P(\theta \leq \hat{\theta}_k^*) = \frac{k}{B}$$

y hasta  $\hat{\theta}_{k+1}^*$  es:

$$P(\theta \leq \hat{\theta}_k^*) = \frac{k+1}{B}$$

entonces

$$\Delta F_k(\hat{\theta}^*) = \frac{k+1}{B} - \frac{k}{B}$$

Finalmente,  $\Delta F_k(\hat{\theta}^*)$  es el valor que toma la distribución empírica acumulada para  $\hat{\theta}_k^*$ .

Por otra parte, la ventaja del Bootstrap No Paramétrico con respecto a otras variantes del método Bootstrap, consiste en mantener la estructura de los datos observados, sin recurrir a consideraciones paramétricas, es decir, no es necesario asumir que los datos se ajustan a un determinado modelo estadístico.

En tanto, entre las desventajas del Bootstrap No Paramétrico, puede mencionarse que falla por la inexactitud inherente de la muestra<sup>2</sup> y la presencia de valores muestrales atípicos<sup>3</sup>. Puesto que se apoya en la capacidad de la muestra para reflejar o representar los aspectos relevantes de la población de la cual fue extraída, la calidad de la muestra resulta crucial, sea porque no ha sido extraída por medio de un procedimiento que asegure cierta representatividad o sea porque su tamaño es demasiado pequeño. En estas situaciones, puede ser dudoso que la información ofrecida por la muestra permita estimar las propiedades desconocidas de la población (GIL, 2003).

### 2.3.2 BOOTSTRAP PARAMÉTRICO

En ocasiones, el investigador conoce la función de distribución correspondiente a la variable aleatoria, aunque se desconozcan los parámetros de la misma, punto que nos conduce al Bootstrap Paramétrico (BP)<sup>4</sup>. Entonces, se

---

<sup>2</sup> Si algunos valores de la variable de estudio, son posibles en la población pero no se presentan en la muestra original, cuando se realiza un Bootstrap No Paramétrico en estas muestras, nunca se obtendrán remuestras con valores diferentes a los obtenidos con la muestra original.

<sup>3</sup> Puesto que el Bootstrap asume que la distribución de la muestra, representa la distribución de la población, si la muestra es inusual también lo serán los estimadores Bootstrap derivados de ella.

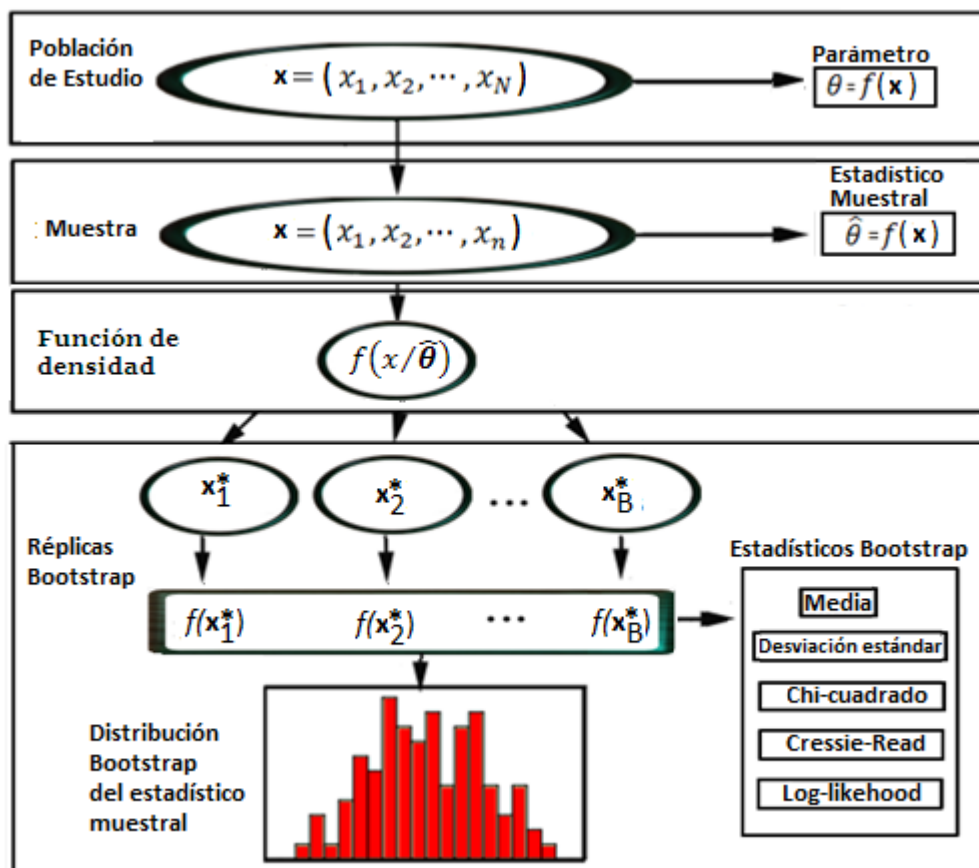
<sup>4</sup> La palabra “paramétrica” alude al hecho que se fijan completamente la distribución de probabilidad excepto por el valor de uno o varios parámetros que deben ser estimados.

requiere la especificación de la distribución estadística de la cual provienen los datos. Esto implica suponer un modelo paramétrico predeterminado, a partir del cual se realiza la simulación para generar nuevos datos.

La **Figura 2.2** muestra el proceso de estimación de estadísticos mediante Bootstrap Paramétrico. Se tiene una muestra  $(x_1, x_2, \dots, x_n)$  para una v.a  $X$ , con función de densidad  $f(x/\hat{\theta})$ , donde  $\hat{\theta}$  es el estimador del parámetro  $\theta$ . Esto puede expresarse como,

$$(x_1^*, x_2^*, \dots, x_n^*) \sim f(x/\hat{\theta}) \quad (2.1)$$

En otras palabras, si conocemos que la función de distribución poblacional  $F$  pertenece a una determinada familia paramétrica  $F_\theta$ , bastaría construir la distribución del estadístico utilizando como función de distribución  $F_{\hat{\theta}}$ .



**FIGURA 2.2.** Proceso de estimación del Bootstrap Paramétrico.

## 2.4 CÁLCULO DEL NIVEL DE SIGNIFICACIÓN

En este apartado, se expone la metodología para el cálculo del nivel de significación (es decir, p-valor) para pruebas de hipótesis utilizando el método Bootstrap. Sea  $X$  una variable aleatoria y  $\hat{\theta} = f(X)$  una función que define el estadístico de contraste de una prueba sobre  $X$ . Dado que  $X$  es una variable aleatoria,  $\hat{\theta}$  será también una variable aleatoria que tiene su propia función de probabilidad. La probabilidad que  $\hat{\theta}$  pueda ser mayor o igual a un valor determinado  $k_0$ , definido por la distribución muestral se expresa como,

$$pv = P[\hat{\theta} \geq k_0] \quad (2.2)$$

Si  $\hat{\theta}$  es el valor obtenido en la muestra original de datos, la prueba de significación consistirá en calcular cuán inusual es  $\hat{\theta}$  en relación a la distribución muestral. El p-valor ( $pv$ ) del contraste sería,

$$pv = P[\hat{\theta}^* \geq \hat{\theta}] = \frac{\text{Número de veces } \{\hat{\theta}^* \geq \hat{\theta}\} + 1}{B + 1} \quad (2.3)$$

donde:  $\hat{\theta}$  = Estadístico de bondad de ajuste para la muestra original.  
 $\hat{\theta}^*$  = Estadístico de bondad de ajuste para cualquiera de las muestras Bootstrap.  
 $B$  = Número de remuestras Bootstrap.

y la regla de decisión para rechazar la hipótesis nula es,

$$\text{Rechazar } H_0 \text{ si } pv \leq \alpha \quad (2.4)$$

donde  $\alpha$  es el nivel de significación establecido “a priori”, frecuentemente es fijado en 0,10, 0,05 y 0,01. Por lo tanto, el cálculo del nivel de significación se fundamenta en estimar la distribución muestral del estadístico de contraste  $\theta$  bajo la hipótesis nula. De esta manera, el nivel de significación unilateral es,

simplemente, la proporción de muestras simuladas en las cuales el valor de  $\theta$  es mayor o igual al valor obtenido en la muestra original  $\hat{\theta}$ .

## 2.5 ESTIMACIÓN DEL SESGO Y ERROR ESTÁNDAR

El sesgo de un estimador es la diferencia entre el valor estimado ( $\hat{\theta}$ ) y el valor verdadero ( $\theta$ ), se expresa como,

$$Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta \quad (2.5)$$

y se puede aproximar mediante Bootstrap como la diferencia entre la media del estimador obtenida con un gran número de muestras Bootstrap y el estimador obtenido de la muestra original (MANLY, 1997), esto es:

$$Sesgo(\hat{\theta}) = \bar{\theta}^* - \hat{\theta} \quad (2.6)$$

donde  $\bar{\theta}^*$  es la media de las réplicas Bootstrap, que corresponde al estimador Bootstrap del estadístico  $\theta$ .

$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* \quad (2.7)$$

donde:  $B$  = Número de remuestras Bootstrap.

$\hat{\theta}_b^*$  = Estimador en la b-ésima remuestra.

La estimación Bootstrap del error estándar ( $ES(\theta^*)$ ), se realiza a partir de la desviación estándar de la distribución muestral generada empíricamente del remuestreo (EFRON & TIBSHIRANI, 1993) y se obtiene mediante la raíz cuadrada de  $Var(\theta^*)$ , definida como,

$$Var(\theta^*) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2 \quad (2.8)$$

donde:  $B$  = Número de remuestras Bootstrap.  
 $\bar{\theta}^*$  = Media aritmética del estadístico de las  $B$  remuestras.  
 $\hat{\theta}_b^*$  = Valor obtenido del estadístico para la  $b$ -ésima remuestra.

El estimador Bootstrap del error estándar, es independiente de la complejidad de la función matemática que defina al estimador y es un procedimiento de carácter general. En la mayor parte de casos, el número  $B$  de muestras Bootstrap adecuado para la estimación del error estándar puede oscilar entre 50 y 200 (EFRON & TIBSHIRANI, 1986).

## 2.6 PROBLEMAS DEL MÉTODO BOOTSTRAP

En algunos casos en particular, los estadísticos Bootstrap no son exactos ya que pueden tener una varianza sustancial. Esta varianza puede tener dos orígenes distintos (EFRON & TIBSHIRANI, 1993), por un lado está el error debido a que analizamos una muestra y no la población entera; es decir, el error estadístico o la variabilidad de muestreo, y por otra parte, el error debido a que no podemos realizar infinitas remuestras, es decir, el error de simulación o la variabilidad del remuestreo. El primer objetivo de un análisis mediante métodos *Bootstrap* es reducir al máximo ambos tipos de errores.

El error de simulación se puede reducir aumentando el número de réplicas. La teoría nos dice que los mejores estimadores se obtienen con infinitas réplicas, algo que en la práctica es imposible e innecesario. La cuestión de interés es el número de réplicas Bootstrap necesarias para obtener un estadístico de suficiente precisión para nuestro propósito. Sabemos que conforme aumenta el número de réplicas se reduce el error de simulación. En cualquier caso es necesario comprobar que estamos trabajando con un número suficiente, y hay dos maneras de hacerlo: la primera sería repetir el proceso Bootstrap varias veces para comprobar si convergen en el mismo resultado; una segunda manera sería siguiendo los estadísticos de interés (media, Pearson, Cressie-Read, etc.) conforme aumenta el número de réplicas hasta que se alcanza un nivel de estabilidad que consideremos suficiente (MANLY, 1997).

El método puede fallar debido a sus propiedades asintóticas. Estas se refieren a la probabilidad del método de converger hacia un determinado valor según aumenta el número de réplicas; se ha observado que en determinadas situaciones esta convergencia se produce más lentamente de lo deseable y el método falla si se detiene en un número insuficiente de réplicas.

Puesto que el único supuesto del Bootstrap No Paramétrico es que la distribución de la muestra conserva las propiedades estadísticas de la distribución de la población, el Bootstrap fallará cuando la distribución muestral no sea representativa de la distribución poblacional. Este es el caso, cuando se tienen valores atípicos en la muestra original, tal que se asume que la distribución de la muestra representa la distribución de la población, si la muestra es inusual también lo serán los estimadores Bootstrap derivados de ella. Un claro ejemplo de este problema es para el caso de poblaciones con distribuciones muy asimétricas donde es fácil que los valores extremos puedan ser utilizados repetidamente, provocando resultados y conclusiones erróneos.

Otro problema son los datos incompletos (en inglés, missing value). El Bootstrap No Paramétrico asume que las observaciones de la muestra  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  provienen de la distribución  $F$ , pero en algunos casos la información obtenida es incompleta, lo que conlleva dificultades para la interpretación de los resultados y la garantía de la consistencia de éstos.

El método Bootstrap No Paramétrico se apoya en la analogía entre la muestra observada y la población de la que fue extraída, la calidad de la muestra es decisiva. En este sentido, los resultados del Bootstrap se ven afectados cuando la muestra no se extrae por un procedimiento de muestreo aleatorio simple y cuando el tamaño de ésta es demasiado pequeño (MOONEY & DUVAL, 1993). En ambas situaciones, puede cuestionarse que la información ofrecida por la muestra permita reconstruir adecuadamente la población estudiada.

En la literatura se han abordado diferentes versiones del método Bootstrap según el objetivo inferencial y el tipo de datos disponibles, como por ejemplo, el



Bootstrap suavizado<sup>5</sup> y doble, que resultan de mucho interés en otros campos del conocimiento, pero no se presentan de forma explícita en este texto, dado que no son utilizados para la selección de un modelo de clases latentes en tablas poco ocupadas.

El siguiente capítulo, se dedica al estudio de los métodos de diagnósticos de modelos de clases latentes en tablas poco ocupadas, como son el de agregar una constante a cada celda de la tabla múltiple, unión de las categorías de la variable, análisis de residuos y los métodos Bootstrap.

---

<sup>5</sup> Es una modificación simple del Bootstrap, no sólo hace remuestreo, sino que agrega una pequeña cantidad de ruido aleatorio a cada observación. Generalmente, este ruido es una variable aleatoria continua con media cero y varianza pequeña.

## **CAPÍTULO 3**

---

# **MÉTODOS UTILIZADOS PARA LA SELECCIÓN DE MODELOS**

---

### 3.1 INTRODUCCIÓN

Los estadísticos de bondad de ajuste (EBA) son un criterio comúnmente utilizado para la selección de un Modelo de Clases Latentes (MCL). Estos estadísticos tienen, bajo ciertas condiciones una distribución de probabilidad teórica  $\chi^2$ . La suposición es válida basándonos en el Teorema Integral de De Moire-Laplace<sup>1</sup>, si tanto la frecuencias observadas de los patrones y el tamaño de la muestra son grandes ( $f_{x_h} \rightarrow \infty, n \rightarrow \infty$ ). Pero cuando se trabaja con tablas poco ocupadas, se representa el problema de posible incremento en el riesgo tipo I. Las frecuencias pequeñas nos darán valores altos en los EBA y nos llevará con mayor probabilidad a rechazar el modelo aunque sea el apropiado para describir el conjunto de datos.

Aunque la mayoría de los investigadores coinciden al afirmar que las tablas poco ocupadas (frecuencias observadas bajas de los patrones de respuesta) no deben ser abordadas en la forma tradicional porque el riesgo Tipo I se incrementa considerablemente, existe discrepancia a la hora de fijar qué se entiende por frecuencia esperada baja. CRAMER en 1946 entiende por frecuencias bajas aquellas que son menores de 10, FISHER (1941) y COCHRAN (1954) fijan ese valor en 5 y KENDALL (1952) y TATE & HYER (1973) en 20. En la práctica, el valor 5 es el más utilizado.

COLLINS et al. (1993) sugiere que la frecuencia esperada es baja si es menor o igual a  $n/R$ , donde  $n$  es el tamaño muestral y  $R = 2^p$  siendo  $p$  el número de variables manifiestas binarias.

HABERMAN (1988) afirma que dado cualquier mínimo esperado por celda bajo la hipótesis nula, es posible calcular el estadístico de Pearson ( $Q^2$ ) si se tienen celdas con frecuencias pequeñas, siempre y cuando la tabla múltiple sea suficientemente grande.

---

<sup>1</sup> Se trata de un caso particular del Teorema Central del Límite. Establece que la distribución Binomial del número de éxitos en  $n$  pruebas independientes de Bernoulli con probabilidad de éxito  $p$  en cada intento es, aproximadamente, una distribución normal de media  $np$  y desviación típica  $\sqrt{npq}$ , si  $n$  es suficientemente grande y se satisfacen determinadas condiciones.

Todo esto pone de manifiesto que se trata de un problema abierto. Por esta razón, dedicaremos este capítulo a presentar las alternativas recogidas en la literatura para abordar este problema.

### 3.2 AÑADIR UNA CONSTANTE A CADA CELDA

La alternativa de sumar una pequeña constante  $\delta$  a la frecuencia observada de cada patrón de respuesta, tiene su origen en los Modelos log-lineales<sup>2</sup>. HABERMAN (1979), BARTHOLOMEW (1987), HAGENAARS & McCUTCHEON (2002) y FORMANN (1982) representan algunas fuentes para entender las conexiones entre el Modelo de Clases Latentes y los Modelos Log-Lineales. En este sentido, HAGENAARS & McCUTCHEON (2002) afirman (en inglés para no dejar duda de la traducción):

*“Adding small constants to the observed frequencies will, at least in regular loglinear models without latent variables, prevent the occurrence of estimated zero cells, but the parameter estimates may be very much dependent on the particular constants chosen”.*

GOODMAN (1974) recomendó utilizar este procedimiento cuando se tienen frecuencias bajas o nulas, mediante la suma de  $\delta = 1/2$  a la frecuencia de cada celda. GRIZZLE et al. (1969) y JOHNSON & KOCH (1970) propusieron sumar a todas las frecuencias de la tabla  $1/R$ , donde  $R$  es el número total de posibles patrones de respuestas. En este sentido, AGRESTI (1990) propone realizar un análisis de la sensibilidad con diferentes valores de  $\delta$ , para medir el efecto sobre las estimaciones de los parámetros y sobre los estadísticos de bondad de ajuste (EBA).

---

<sup>2</sup> El Modelo Log-Lineal es un método estadístico que tiene por objeto estudiar la clasificación de las variables categóricas o cualitativas. En esencia es un modelo de regresión lineal múltiple entre las variables categóricas y el logaritmo neperiano de la frecuencia de los patrones de respuesta, de la forma:

$$\log m_{X,Y}(x, c) = \mu + \lambda_{i_1}^{X_1} + \dots + \lambda_{i_p}^{X_p} + \lambda_c^Y + \lambda_{i_1 c}^{X_1 Y} + \dots + \lambda_{i_p c}^{X_p Y}$$

Esta alternativa resulta útil solo si hay pocas celdas con frecuencias bajas debido al incremento del tamaño de la muestra. Incluso en estos casos, sumar una pequeña constante, a menudo produce un gran efecto conservador sobre el resultado del ajuste (AGRESTI, 1990; AGRESTI & YANG, 1987). Para HAGENAARS & McCUTCHEON (2002) los parámetros estimados pueden ser muy dependientes de la constantes elegidas.

Con el objetivo de medir el efecto en los parámetros del modelo al sumar una constante a las celdas en tablas poco ocupadas, se trabajó con los resultados para cuatro pruebas de diagnóstico del VIH (ALVORD et al, 1988), aplicados a un total de 428 sujetos<sup>3</sup>. Cada variable  $X_j$  toma el valor “0” si el sujeto no es portador del virus del VIH (negativo) y el valor “1” para el sujeto que presenta el virus (positivo). En la **Tabla 3.1**, se presentan los patrones de respuestas y las frecuencias observadas.

Por otra parte, en la **Tabla 3.2** son mostrados los resultados de las probabilidades totales y condicionales para un modelo con tres clases latentes utilizando diferentes valores de  $\delta$ . En general, no existe mayor diferencia en las probabilidades totales de las clases latentes ( $\pi_Y(c)$ ), al sumar a las frecuencias observadas de los patrones las constantes recomendadas por diversos autores. La diferencia con respecto a las probabilidades de los datos originales no supera el 0,03. Por otra parte, a través de las probabilidades condicionales ( $\pi_{X_j/Y(c)}$ ), se puede afirmar que la composición de las clases latentes tampoco se ve afectada significativamente por la constante utilizada.

---

<sup>3</sup> En el apartado 6.3.2 se describen con mayor amplitud las pruebas para el diagnóstico del VIH.

$X_1$	$X_2$	$X_3$	$X_4$	$f_{x_h}$
0	0	0	0	170
0	0	0	1	15
0	0	1	0	0
0	0	1	1	0
0	1	0	0	6
0	1	0	1	0
0	1	1	0	0
0	1	1	1	0
1	0	0	0	4
1	0	0	1	17
1	0	1	0	0
1	0	1	1	83
1	1	0	0	1
1	1	0	1	4
1	1	1	0	0
1	1	1	1	128

**TABLA 3.1.** Patrones de respuesta y frecuencias observadas para las cuatro pruebas de diagnóstico del VIH.

Con este conjunto de datos existe duda sobre la identificabilidad intrínseca del modelo, como lo afirma SEPÚLVEDA (2004) de la siguiente manera: “Por ejemplo, un modelo que considera cuatro variables manifiestas binarias y tres clases latentes es no identificable, ya que el rango de la matriz Hessiana es de 13 y se necesita estimar 14 parámetros, por lo que debe considerarse una restricción adicional sobre uno de los parámetros del modelo”.

Constante	Parámetros	Clase 1	Clase 2	Clase 3
1/4	$\pi_Y(c)$	<b>0,45</b>	<b>0,06</b>	<b>0,49</b>
	$\pi_{X_1/Y(c)}$	0,00	1,00	1,00
	$\pi_{X_2/Y(c)}$	0,04	0,20	0,61
	$\pi_{X_3/Y(c)}$	0,00	0,02	1,00
	$\pi_{X_4/Y(c)}$	0,08	0,80	1,00
1/2	$\pi_Y(c)$	<b>0,45</b>	<b>0,06</b>	<b>0,49</b>
	$\pi_{X_1/Y(c)}$	0,00	1,00	1,00
	$\pi_{X_2/Y(c)}$	0,04	0,21	0,61
	$\pi_{X_3/Y(c)}$	0,01	0,04	1,00
	$\pi_{X_4/Y(c)}$	0,08	0,79	1,00
3/4	$\pi_Y(c)$	<b>0,44</b>	<b>0,08</b>	<b>0,48</b>
	$\pi_{X_1/Y(c)}$	0,00	0,88	1,00
	$\pi_{X_2/Y(c)}$	0,04	0,23	0,61
	$\pi_{X_3/Y(c)}$	0,01	0,14	1,00
	$\pi_{X_4/Y(c)}$	0,07	0,77	1,00
1	$\pi_Y(c)$	<b>0,43</b>	<b>0,10</b>	<b>0,47</b>
	$\pi_{X_1/Y(c)}$	0,00	0,84	0,99
	$\pi_{X_2/Y(c)}$	0,04	0,24	0,62
	$\pi_{X_3/Y(c)}$	0,01	0,19	1,00
	$\pi_{X_4/Y(c)}$	0,07	0,76	1,00
1/R	$\pi_Y(c)$	<b>0,45</b>	<b>0,06</b>	<b>0,49</b>
	$\pi_{X_1/Y(c)}$	0,00	1,00	1,00
	$\pi_{X_2/Y(c)}$	0,03	0,20	0,61
	$\pi_{X_3/Y(c)}$	0,00	0,02	1,00
	$\pi_{X_4/Y(c)}$	0,08	0,80	1,00
$\sqrt{n}/R$	$\pi_Y(c)$	<b>0,42</b>	<b>0,11</b>	<b>0,47</b>
	$\pi_{X_1/Y(c)}$	0,00	0,80	0,99
	$\pi_{X_2/Y(c)}$	0,04	0,26	0,62
	$\pi_{X_3/Y(c)}$	0,01	0,23	1,00
	$\pi_{X_4/Y(c)}$	0,06	0,74	1,00
Datos originales $\delta = 0$	$\pi_Y(c)$	<b>0,46</b>	<b>0,08</b>	<b>0,46</b>
	$\pi_{X_1/Y(c)}$	<b>0,03</b>	<b>1,00</b>	<b>1,00</b>
	$\pi_{X_2/Y(c)}$	<b>0,04</b>	<b>0,14</b>	<b>0,64</b>
	$\pi_{X_3/Y(c)}$	<b>0,00</b>	<b>0,45</b>	<b>0,99</b>
	$\pi_{X_4/Y(c)}$	<b>0,08</b>	<b>1,00</b>	<b>1,00</b>

**TABLA 3.2.** Efecto en los parámetros al sumar una constante a las frecuencias observadas bajas.

Por lo anteriormente comentado, vamos a simular un conjunto de datos que tengan el problema de frecuencias bajas o nulas en los distintos patrones de respuesta (en inglés, “sparse data”), mediante el sistema LEM (VERMUNT, 1997b) con cinco variables manifiestas, 3 clases latentes y un tamaño de muestra de 320 individuos; aplicando posteriormente la misma metodología de sumar una constante a la frecuencias observadas de los patrones de respuesta. Utilizaremos el siguiente programa:

---

```

lat 1
man 5
dim 3 2 2 2 2 2
mod X A|X B|X C|X D|X E|X
sim 320 MCLv5c3.dat
sta X [0.70 0.20 0.10]
sta A|X [1.00 0.00 0.31 0.69 0.97 0.03]
sta B|X [0.78 0.22 0.12 0.98 0.74 0.26]
sta C|X [0.00 1.00 0.20 0.80 0.60 0.40]
sta D|X [0.90 0.10 0.00 1.00 0.50 0.50]
sta E|X [0.05 0.95 0.60 0.40 0.73 0.27]

```

---

Los datos resultantes de la simulación se muestran a continuación:

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$f_{x_h}$
0	0	0	0	0	2
0	0	0	0	1	2
0	0	0	1	0	5
0	0	0	1	1	1
0	0	1	0	0	14
0	0	1	0	1	166
0	0	1	1	0	5
0	0	1	1	1	15
0	1	0	0	0	1
0	1	0	0	1	1
0	1	0	1	0	5
0	1	0	1	1	5
0	1	1	0	0	3
0	1	1	0	1	34
0	1	1	1	0	9
0	1	1	1	1	12

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$f_{x_h}$
1	0	0	0	0	1
1	0	0	0	1	0
1	0	0	1	0	0
1	0	0	1	1	0
1	0	1	0	0	0
1	0	1	0	1	0
1	0	1	1	0	3
1	0	1	1	1	2
1	1	0	0	0	0
1	1	0	0	1	0
1	1	0	1	0	3
1	1	0	1	1	2
1	1	1	0	0	0
1	1	1	0	1	0
1	1	1	1	0	14
1	1	1	1	1	15

**TABLA 3.3.** Datos simulados para una muestra de 320 individuos, 5 variables manifiestas binarias y tres clases latentes.

En la **Tabla 3.4** se presentan los resultados de las probabilidades totales ( $\pi_Y(c)$ ) y condicionales ( $\pi_{X_j/Y(c)}$ ), al sumarle las constantes a la frecuencia observada de los patrones. Las diferencias en las  $\pi_Y(c)$  tienen un valor máximo de 7% con respecto a los datos originales; la constante que muestra los valores más similares es  $\delta = 1/R$ , tal que los resultados son: 0,19, 0,72 y 0,09, a diferencia de 0,19, 0,73 y 0,07 obtenidos con los valores reales ( $\delta = 0$ ).



Constante	Parámetros	Clase 1	Clase 2	Clase 3
1/4	$\pi_Y(c)$	<b>0,18</b>	<b>0,71</b>	<b>0,11</b>
	$\pi_{X_1/Y(c)}$	0,63	0,00	0,16
	$\pi_{X_2/Y(c)}$	0,87	0,17	0,47
	$\pi_{X_3/Y(c)}$	0,86	0,99	0,37
	$\pi_{X_4/Y(c)}$	0,99	0,07	0,67
	$\pi_{X_5/Y(c)}$	0,52	0,93	0,32
1/2	$\pi_Y(c)$	<b>0,18</b>	<b>0,69</b>	<b>0,13</b>
	$\pi_{X_1/Y(c)}$	0,64	0,00	0,21
	$\pi_{X_2/Y(c)}$	0,87	0,17	0,48
	$\pi_{X_3/Y(c)}$	0,86	0,99	0,40
	$\pi_{X_4/Y(c)}$	0,99	0,07	0,64
	$\pi_{X_5/Y(c)}$	0,53	0,93	0,35
3/4	$\pi_Y(c)$	<b>0,18</b>	<b>0,67</b>	<b>0,15</b>
	$\pi_{X_1/Y(c)}$	0,64	0,00	0,25
	$\pi_{X_2/Y(c)}$	0,87	0,17	0,48
	$\pi_{X_3/Y(c)}$	0,86	0,99	0,41
	$\pi_{X_4/Y(c)}$	0,98	0,07	0,62
	$\pi_{X_5/Y(c)}$	0,53	0,93	0,37
1	$\pi_Y(c)$	<b>0,18</b>	<b>0,66</b>	<b>0,17</b>
	$\pi_{X_1/Y(c)}$	0,64	0,00	0,27
	$\pi_{X_2/Y(c)}$	0,87	0,17	0,48
	$\pi_{X_3/Y(c)}$	0,86	0,99	0,42
	$\pi_{X_4/Y(c)}$	0,98	0,07	0,61
	$\pi_{X_5/Y(c)}$	0,53	0,93	0,39
1/R	$\pi_Y(c)$	<b>0,19</b>	<b>0,72</b>	<b>0,09</b>
	$\pi_{X_1/Y(c)}$	0,63	0,00	0,08
	$\pi_{X_2/Y(c)}$	0,87	0,17	0,43
	$\pi_{X_3/Y(c)}$	0,86	1,00	0,32
	$\pi_{X_4/Y(c)}$	1,00	0,07	0,64
	$\pi_{X_5/Y(c)}$	0,51	0,93	0,35
$\sqrt{n}/R$	$\pi_Y(c)$	<b>0,18</b>	<b>0,69</b>	<b>0,13</b>
	$\pi_{X_1/Y(c)}$	0,64	0,00	0,22
	$\pi_{X_2/Y(c)}$	0,87	0,17	0,48
	$\pi_{X_3/Y(c)}$	0,86	0,99	0,40
	$\pi_{X_4/Y(c)}$	0,99	0,07	0,64
	$\pi_{X_5/Y(c)}$	0,53	0,93	0,35
Datos originales $\delta = 0$	$\pi_Y(c)$	<b>0,19</b>	<b>0,73</b>	<b>0,07</b>
	$\pi_{X_1/Y(c)}$	<b>0,62</b>	<b>0,00</b>	<b>0,08</b>
	$\pi_{X_2/Y(c)}$	<b>0,87</b>	<b>0,17</b>	<b>0,44</b>
	$\pi_{X_3/Y(c)}$	<b>0,85</b>	<b>0,99</b>	<b>0,30</b>
	$\pi_{X_4/Y(c)}$	<b>1,00</b>	<b>0,07</b>	<b>0,71</b>
	$\pi_{X_5/Y(c)}$	<b>0,52</b>	<b>0,93</b>	<b>0,25</b>

**TABLA 3.4.** Efecto en los parámetros al sumar una constante a las frecuencias observadas para datos simulados.

Por otra parte, las  $\pi_{X_j/Y(c)}$  calculadas al utilizar las diversas constantes no presentan diferencias significativas respecto a las de los datos originales, por esta razón, la composición de las 3 clases latentes no se modifica sustancialmente.

De todo lo anterior se desprende que, la opción de sumar las constantes recomendadas por autores a las celdas es una solución válida, que no produce diferencias significativas respecto a los parámetros del modelo de clases latentes calculados con los datos originales. De tal forma que, las apreciaciones de AGRESTI (1990) y HAGENAARS & McCUTCHEON (2002) tienen que ser investigadas con mayor experimentación, para comprobar o refutar la validez de la afirmación que los parámetros del modelo son muy dependientes de la constante seleccionada.

### 3.3 ANÁLISIS DE RESIDUOS

El análisis de los residuos lo encontramos en MENEZES & LASAOSA (2007), con el objetivo de determinar la bondad de ajuste del Modelo de Clases Latentes (MCL). Los estadísticos de bondad (EBA) de ajuste miden la discrepancia a nivel global entre las frecuencias observadas y las estimadas por el modelo, pero no proporcionan dónde están las mayores diferencias. Basta con que una sola celda de la tabla múltiple proporcione una gran diferencia, para obtener valores de los EBA que nos llevarían a rechazar el modelo hipotético de clases latentes.

Las autoras proponen realizar un análisis de residuos, para comprobar si un modelo subestima o sobrestima las frecuencias observadas de los patrones de respuesta. En concreto, se considerará un modelo de regresión lineal donde la variable dependiente son las frecuencias observadas  $f_{x_h}$  y la variable independiente (o explicativa) es la frecuencia esperada del patrón de respuesta, denotada como  $\hat{f}_{x_h}$ , de esta manera se tiene que,

$$f_{x_h} = \alpha + \beta \hat{f}_{x_h} + \varepsilon_{x_h} \quad (3.1)$$

donde:  $f_{x_h}$  = Frecuencia observada del patrón de respuesta  $x_h$ .

$\hat{f}_{x_h}$  = Frecuencia esperada del patrón de respuesta  $x_h$ .

$\alpha$  = Constante de regresión lineal.

$\beta$  = Coeficiente de regresión lineal.

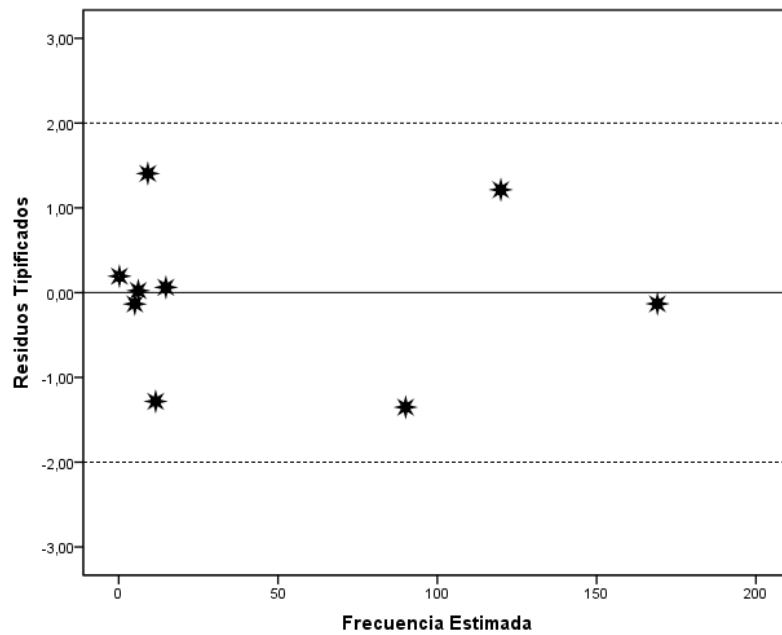
El término  $\varepsilon_i$  es una perturbación estocástica agregada al modelo para recoger todos los posibles errores de medida tanto en las variables  $f_{x_h}$  y  $\hat{f}_{x_h}$  así como los errores en la especificación lineal del modelo. Es decir recogerá todos aquellos factores que por error no se han incluido en el modelo y que pueden afectar a la variable dependiente. Nótese que, en definitiva, el término de perturbación  $\varepsilon_{x_h}$  indica en qué medida las variables  $f_{x_h}$  y  $\hat{f}_{x_h}$  se apartan de la relación lineal.

El diagnóstico del MCL, se efectúa contrastando los coeficientes del modelo de regresión lineal. La formulación de la hipótesis nula se realiza de la siguiente forma: para la constante de regresión como  $H_0: \alpha = 0$  y para el coeficiente de regresión como  $H_0: \beta = 1$ . Además, mediante el examen de los residuos estandarizados versus frecuencias observadas del modelo (3.1), se puede apreciar qué patrones de respuesta influyen de manera desproporcionada en los parámetros de la distribución de los residuos, y en definitiva son los responsables de la posible falta de ajuste del modelo. También es necesario, verificar el contraste de normalidad de los residuos y comparar los promedios y varianza de los residuales entre modelos alternativos. El diagnóstico se efectúa de manera gráfica representando el valor de los residuos tipificados frente a las predicciones, con el objetivo de constatar la linealidad y la igualdad de las varianzas.

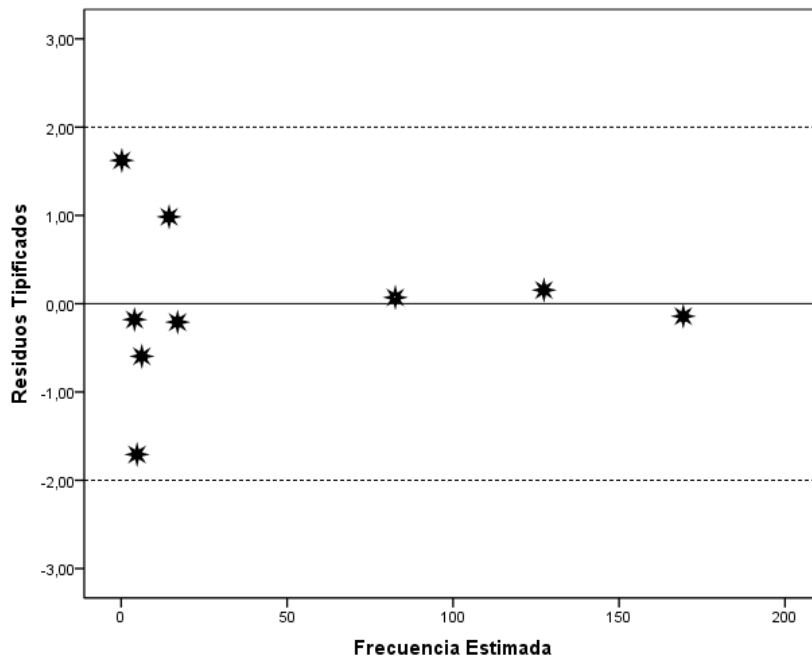
A manera ilustrativa del método expuesto, vamos a utilizar nuevamente los datos para el diagnóstico del VIH (ALVORD et al, 1988). Bajo el supuesto de independencia local, hemos ajustado un modelo con dos y tres clases latentes. Analizando las **Figuras 3.1**, podemos concluir que el modelo con tres clases latentes tiene un mejor ajuste a los datos; los residuos de los patrones de respuesta observados se distribuyen cercanos a cero y tienen una varianza más pequeña.

El mismo ejercicio de análisis de residuos lo hicimos con los datos de la **Tabla 3.3**. En la **Figura 3.2** se observa claramente que los residuos tipificados del modelo con tres clases latentes tienen menor varianza y se concentran

alrededor de cero; indicando que es efectivamente el modelo más apropiado para el conjunto de datos.

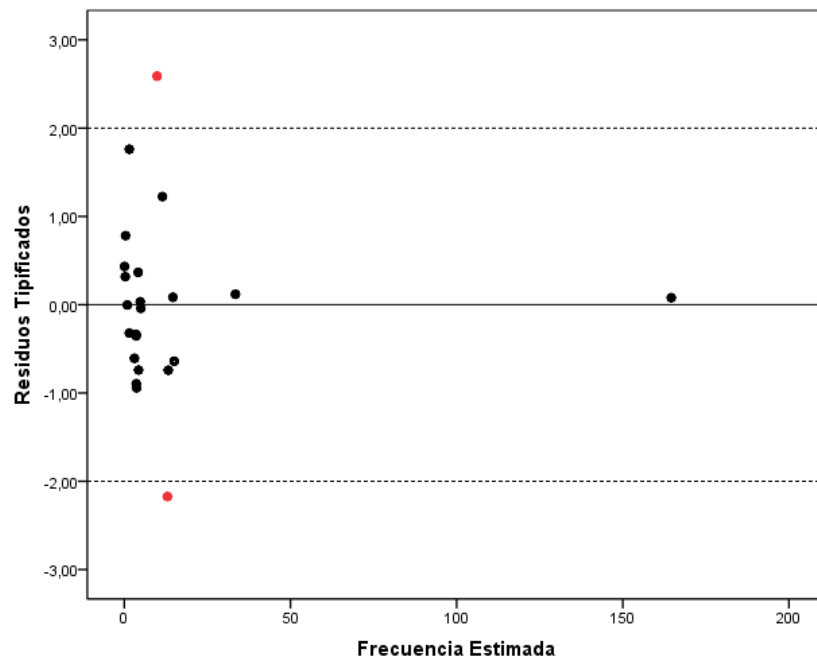


### Dos Clases

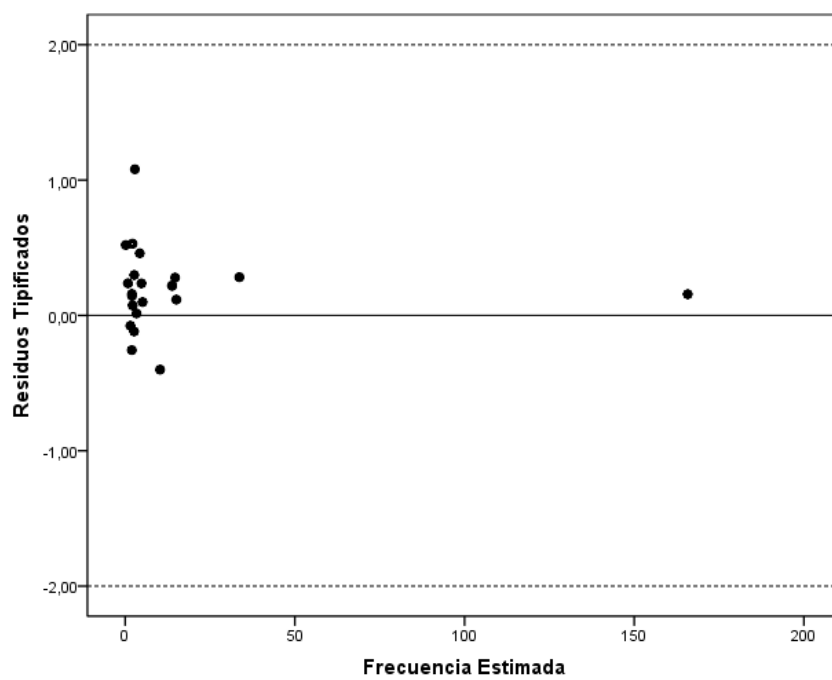


### Tres Clases

**Figura 3.1.** Diagnóstico de los residuos tipificados para el modelo con dos y tres clases latentes (datos del VIH).



Dos Clases



Tres Clases

**Figura 3.2.** Residuos tipificados para el modelo con dos y tres clases latentes (datos simulados).

Finalmente, esta alternativa tiene una ventaja práctica, implica poco esfuerzo computacional en comparación con otros métodos, que son utilizados para el diagnóstico de modelos de clases latentes en tablas poco ocupadas.

### 3.4 ESTIMACIÓN ROBUSTA CON $\beta$ -VEROSIMILITUD

YU & NOBORU (2006) proponen un método que calcula los parámetros de un modelo de mixtura por medio del algoritmo EM, utilizando la función  $\beta$ -verosimilitud en lugar de log-verosimilitud, para tablas poco ocupadas. La función  $\beta$ -verosimilitud para una muestra de  $n$  individuos está dada por la expresión,

$$l_{\beta}(\mathbf{X}; \boldsymbol{\theta}) = \frac{1}{n\beta} \sum_{h=1}^R P(\mathbf{x}_h; \boldsymbol{\theta})^{\beta} - \frac{1}{1+\beta} \sum_{h=1}^R P(\mathbf{x}_h; \boldsymbol{\theta})^{1+\beta} \quad (3.2)$$

donde:  $n$  = Tamaño de la muestra.

$\beta$  = Parámetro de calibración  $0 \leq \beta \leq 1$ .

$\mathbf{x}' = (x_1, x_2, \dots, x_p)$  = Vector de respuestas para cada individuo.

$\boldsymbol{\theta}$  = Parámetros del modelo de mixtura  $(\pi_1 \dots, \pi_T, \phi_1, \dots, \phi_C)$ .

Las etapas Esperanza (**E**) y Maximización (**M**) del algoritmo original EM se modifican como sigue,

**Etapas Esperanza.** Primer paso (paso E), en la  $(t + 1)$ -ésima iteración, calcular,

$$Q_{\beta}(\boldsymbol{\theta}; \theta^{(t)}) = \sum_{h=1}^R \frac{f_{\mathbf{x}_h}}{n\beta} \left( \prod_{c=1}^T (\pi_c P_c(\mathbf{x}_h; \phi_c))^{z_{c\mathbf{x}_h}^{(t)}} \right)^{\beta} - \frac{1}{1+\beta} \sum_{h=1}^R \left( \prod_{c=1}^T (\pi_c P_c(\mathbf{x}_h; \phi_c))^{z_{c\mathbf{x}_h}^{(t)}} \right)^{1+\beta} \quad (3.3)$$

donde:  $\phi_c$  = Parámetros en la clase latente  $c$ .

$P_c(\mathbf{x}_h; \phi_c) = \pi_Y(c) \pi_{X/Y(c)}(\mathbf{x}_h)$  = Probabilidad conjunta de  $\mathbf{X}$  e  $Y$ .

$\pi_c = \pi_Y(c)$  = Probabilidad *a priori* de la clase latente  $c$ ,

tal que  $\sum_{c=1}^T \pi_c = 1$ .

$f_{\mathbf{x}_h}$  = Frecuencia observada del patrón de respuesta.

$z_{c\mathbf{x}_h}$  = Probabilidad a posteriori de  $\mathbf{x}_h$  en la clase latente  $c$ .

$T$  = Número de clases latentes,  $c = 1, 2, \dots, T$ .

$\theta^{(t)}$  = Parámetros del modelo en la  $t$ -ésima iteración.

La probabilidad a posteriori sobre  $\mathbf{x}_h$  en la  $t$ -ésima iteración  $Z_{c\mathbf{x}_h}^{(t)}$ , toma la forma,

$$z_{c\mathbf{x}_h}^{(t)} = \frac{\pi_c^{(t)} P_c(\mathbf{x}_h; \phi_c^{(t)})}{\sum_{c=1}^T \pi_c^{(t)} P_c(\mathbf{x}_h; \phi_c^{(t)})} \quad (3.4)$$

**Etapla Maximización.** El segundo paso (paso M) del algoritmo EM busca maximizar la esperanza calculada en el primer paso. Esto es encontrar:

$$\theta^{(t+1)} = \arg \max_{\theta} Q_{\beta}(\theta; \theta^{(t)}) \quad (3.5)$$

Estos dos pasos son repetidos hasta alcanzar la convergencia de los parámetros,  $\theta^{(t+1)} = \theta^{(t)}$ .

YU & NOBORU (2006), probaron de forma experimental este procedimiento de estimación. Según los autores, los estimadores  $\beta$ -verosimilitud son más exactos en comparación al método usual de maximizar el logaritmo de la verosimilitud, especialmente con tablas poco ocupadas y tamaños de muestra pequeños.

Seguidamente, vamos a probar la validez de la anterior afirmación y la exactitud de las estimaciones de los parámetros usando  $\beta$ -verosimilitud con los datos de la **Tabla 3.3**. Como puede observarse en la Tabla 3.5, *resulta interesante confirmar la exactitud de la maximización de la función  $\beta$ -verosimilitud utilizando un factor de calibración  $\beta=0,50$* . En todas las estimaciones de los parámetros los resultados tienen un error muy pequeño. Por ejemplo, las probabilidades totales de las clases latentes son casi iguales, a diferencia de los resultados obtenidos con la maximización de la función Log-Verosimilitud.

Probabilidades	Parámetros del modelo (*)			Estimación $\beta$ – Verosimilitud			Estimación Log – Verosimilitud		
	c=1	c=2	c=3	c=1	c=2	c=3	c=1	c=2	c=3
$\pi_Y(c)$	<b>0,70</b>	<b>0,20</b>	<b>0,10</b>	<b>0,72</b>	<b>0,19</b>	<b>0,09</b>	<b>0,61</b>	<b>0,28</b>	<b>0,12</b>
$\pi_{X_1/Y(c)}$	0,00	0,69	0,03	0,00	0,63	0,08	0,00	0,46	0,00
$\pi_{X_2/Y(c)}$	0,22	0,98	0,26	0,17	0,87	0,43	0,00	0,77	0,99
$\pi_{X_3/Y(c)}$	1,00	0,80	0,40	1,00	0,86	0,33	0,98	0,73	0,97
$\pi_{X_4/Y(c)}$	0,10	1,00	0,50	0,07	1,00	0,64	0,06	0,95	0,00
$\pi_{X_5/Y(c)}$	0,95	0,40	0,27	0,93	0,51	0,35	0,92	0,48	0,93

(\*) Utilizados para simular el conjunto de datos con el programa LEM.

**TABLA 3.5.** Comparación de los parámetros estimados según el tipo de maximización de la función de verosimilitud.

Podemos afirmar que la estimación de las probabilidades totales y condicionales a través de la maximización de la función  $\beta$ -verosimilitud produce resultados con un nivel de error muy pequeño. Además, hemos construido un programa en MATLAB (MATHWORKS, 2010) que permite ajustar modelos de clases latentes con esta metodología, esto resulta un aporte importante de la tesis, ya que los autores de la propuesta no lo habían plasmado en un software (o por lo menos, no encontramos información de que lo hayan publicado).

### 3.5 UTILIZAR CELDAS SUFICIENTEMENTE OCUPADAS

Las celdas con frecuencias muy pequeñas o cero, tienen una contribución muy importante en los estadísticos de bondad de ajuste (EBA), al provocar un aumento del error Tipo I. Por esta razón, los EBA están muy distorsionados a medida que el número de celdas de la tabla múltiple tienen frecuencias observadas pequeñas.

JÖRESKOG & MOUSTAKI (2001) proponen resolver el problema, estimando el modelo de clases latentes sólo sobre los patrones de respuesta con frecuencias mayores o iguales a una cantidad preestablecida  $\omega$ ; calculando los EBA para aquellos patrones cuya frecuencias observadas sean superiores al valor  $\omega$ .



Esta manera de abordar el problema, puede dar una buena aproximación a la distribución  $\chi^2$ , pero a expensas de descartar los datos de las celdas que no cumplen la condición (JÖRESKOG & MOUSTAKI, 2001). En la literatura, se han sugerido los valores 1, 2, 5 y 10 para  $\omega$ . Sin embargo, *la arbitrariedad en la elección de  $\omega$  no garantiza que la distribución asintótica de los EBA sea Chi-Cuadrado*.

Con el propósito de evaluar la metodología anterior, hemos utilizado los datos de la **Tabla 3.3** y calculado los parámetros del modelo de clases latentes. En la **Tabla 3.6** se puede observar que hay una pequeña diferencia entre efectuar los cálculos con todos los patrones posibles ( $R=32$ ,  $n=320$ ) o los que tienen frecuencias mayores o iguales a 1 ( $n=23$ ). Sin embargo, al fijar la cuota mínima de las frecuencias de los patrones en 5 ( $n=299$ ), los parámetros del modelo son significativamente diferentes, principalmente en las probabilidades totales de las clases latentes.

Probabilidades	Parámetros del modelo			Patrones con Frecuencias Observadas					
	(*)			$(f_{x_h} \geq 1)$			$(f_{x_h} \geq 5)$		
	c=1	c=2	c=3	c=1	c=2	c=3	c=1	c=2	c=3
$\pi_Y(c)$	<b>0,70</b>	<b>0,20</b>	<b>0,10</b>	<b>0,73</b>	<b>0,20</b>	<b>0,07</b>	<b>0,52</b>	<b>0,23</b>	<b>0,25</b>
$\pi_{X_1/Y(c)}$	0,00	0,69	0,03	0,00	0,60	0,08	0,00	0,42	0,00
$\pi_{X_2/Y(c)}$	0,22	0,98	0,26	0,17	0,87	0,41	0,22	0,87	0,00
$\pi_{X_3/Y(c)}$	1,00	0,80	0,40	0,99	0,84	0,30	1,00	0,78	1,00
$\pi_{X_4/Y(c)}$	0,10	1,00	0,50	0,07	1,00	0,69	0,02	1,00	0,19
$\pi_{X_5/Y(c)}$	0,95	0,40	0,27	0,93	0,52	0,22	1,00	0,49	0,77

(\*) Utilizados para simular el conjunto de datos con el programa LEM.

**TABLA 3.6.** Comparación de los parámetros estimados según el valor mínimo fijado para las frecuencias de los patrones de respuesta.

De todo lo anterior se desprende que, utilizar solamente los patrones de frecuencias con información ( $f_{x_h} \geq 1$ ), da una buena aproximación de los parámetros sin tener la necesidad de descartar los datos de la muestra original.

### 3.6 AGRUPACIÓN DE CATEGORÍAS DE VARIABLES

Antes de abordar el tema, es importante aclarar que la agrupación de categorías, solamente es válida, cuando las variables manifiestas tienen más de dos categorías de respuesta, o sea, son “politómicas”. Para los propósitos de la presente investigación, la estrategia de agrupar no tiene sentido práctico, ya que solamente nos limitamos al problema de ajuste de modelos de clases latentes, cuando se tienen variables manifiestas con dos categorías, llamadas “dicotómicas”. Por esta razón, no pretendemos detallar exhaustivamente el método de agrupación de categorías.

Una manera habitual de abordar este problema, es combinar las categorías de las variables de tal manera que se garantice un tamaño mínimo en las frecuencias observadas y de esta forma, reducir el número de patrones con frecuencias nulas o bajas. Los estadísticos pueden calcularse sobre la base de este conjunto de categorías reducidas de las variables, resolviéndose el problema de convergencia de los estadísticos de bondad de ajuste para los modelos de clases latentes (MCL). Aún cuando este tipo de estrategia afecta a la estimación de los parámetros del MCL, esta técnica puede ser la más adecuada y en algunos casos la única viable (AGRESTI, 1990).

Para decidir la agrupación de categorías de una variable manifiesta, hay que considerar que exista una forma natural de combinarlas y que la pérdida de información no sea muy significativa. El método de agrupación de las categorías de las variables fue propuesto por KASS (1980) para el análisis de segmentación; llamando a los tres tipos de predictores (variables) como monótonos, libres y flotantes.

Primero, si una variable es medida en escala ordinal (*monótona*), sólo categorías contiguas en la escala pueden ser agrupadas. Un ejemplo de predictor monótono adecuado es el nivel de estudios. Si esta variable tuviera como valores “primarios”, “secundarios” y “universitarios”, el procedimiento permitiría la fusión de las categorías primera y segunda o segunda y tercera, y descartaría la

posibilidad de formar un grupo compuesto por sujetos con estudios primarios y universitarios.

Segundo, en una variable medida en escala nominal (*libres*), tiene sentido la agrupación de cualquier par de categorías. Sea, por ejemplo, la variable situación ocupacional con los valores “ocupado”, “parado”, “jubilado”, “ama de casa”, “estudiante” y “otros”. De cara a la formación de grupos, la categoría “ocupado” podría formar grupo con “parados” y/o “estudiantes”. La primera categoría es contigua, pero la segunda no lo es. Este procedimiento se aplica por regla general a variables de tipo nominal.

Tercero, si todas las categorías de una variable están medidas en escala ordinal, salvo una de ellas, cuya posición en la escala es desconocida (categoría *flotante*); la variable manifiesta será tratada como ordinal y la categoría flotante puede quedar sola o combinada con cualquier otra categoría. Si la variable nivel de estudios tuviera el valor “No sabe, no responde”, con este procedimiento, los sujetos que no contestasen podrían agruparse con cualquiera de las tres categorías establecidas.

Los métodos presentados en los apartados anteriores, no resuelven cuestiones como: el cálculo de nivel de significación para contrastar los modelos de clases latentes y la distribución de probabilidad de los estadísticos de bondad de ajuste. Por tanto, el problema del diagnóstico de modelos de clases latentes en tablas poco ocupadas está abierto a nuevas metodologías. Como consecuencia de lo anterior, vamos a abordarlo desde la perspectiva de los métodos Bootstrap.

### **3.7 ESTIMACIÓN DE LA DISTRIBUCIÓN EMPÍRICA**

#### **3.7.1 INTRODUCCIÓN**

El problema de las tablas poco ocupadas, se puede superar mediante la generación de la distribución empírica de muestreo del estadístico mediante el método Bootstrap. EFRON (1979) propuso el método Bootstrap para producir una

aproximación de la distribución desconocida  $F_\theta$  de un estadístico de contraste  $\theta$ , donde  $\hat{\theta}$  es una función de los datos observados  $(x_1, \dots, x_n)$  y de un vector de parámetros  $\boldsymbol{\varphi}$ , de esta forma se tiene que  $\hat{\theta} = f(x_1, \dots, x_n, \boldsymbol{\varphi})$ .

En nuestro caso, consideremos que  $\theta$  es un estadístico de bondad de ajuste. La distribución empírica  $\hat{F}_{\hat{\theta}^*}$  de  $\theta$  es aproximada simulando datos adicionales  $(\mathbf{x}_{1b}^*, \dots, \mathbf{x}_{nb}^*)$  y un conjunto de resultados independientes  $\hat{\theta}_b^*$  para  $b = 1, \dots, B$ , donde  $B$  es el número de remuestras Bootstrap y  $\mathbf{x}_{1b}^*$  es un patrón de respuesta.

Entre las diferentes versiones del método Bootstrap, tenemos el simplista (“Naive”, No Paramétrico) y el paramétrico. Mientras que “naive” utiliza remuestreo de la muestra original, el Bootstrap paramétrico utiliza una hipotética distribución con características conocidas. En este apartado, analizamos las características de los métodos Bootstrap en función del análisis de clases latentes y realizamos una comparación entre ellos.

### 3.7.2 BOOTSTRAP NO PARAMÉTRICO

Vamos a considerar que  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  representa una muestra aleatoria de tamaño  $n$  de patrones de respuesta. Cada patrón  $\mathbf{x}_h$  está formado por las respuestas obtenidas de las variables manifiestas binarias. A título ilustrativo, considérese el patrón 10110.

Pensemos en un estadístico de bondad de ajuste poblacional  $\theta$ , con distribución de probabilidad desconocida, estimado con la información de la muestra como  $\hat{\theta}$ . El método Bootstrap No Paramétrico (BNP) remuestrea la población definida por el modelo empírico  $\mathbf{F}_{\hat{\theta}}$  para estimar la distribución muestral de  $\hat{\theta}$ , asignando una probabilidad  $1/n$  sobre cada patrón  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ . De esta forma, se obtiene una remuestra  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  tomada de la muestra original con reemplazamiento.

El valor del estadístico de bondad de ajuste  $\hat{\theta}_b^*$ , es calculado usando la misma fórmula que para  $\hat{\theta}$ , pero utilizando la  $b$ -ésima remuestra Bootstrap. Dada una colección de  $B$  remuestras Bootstrap independientes, se obtienen los valores de los estadísticos de bondad de ajuste  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$  y se define la distribución empírica del estadístico  $\hat{\theta}$  a partir del histograma<sup>4</sup> construido a partir de las  $B$  réplicas.

Es importante destacar, que no existen muchas referencias bibliográficas sobre la aplicación del Bootstrap No Paramétrico para el diagnóstico de modelos de clases latentes en tablas poco ocupadas. Una de ellas, es DIAS (2005) que comparó los promedios y medianas de las estimaciones de los parámetros del modelo de clases latentes usando el Bootstrap No Paramétrico y Paramétrico.

### 3.7.3 BOOTSTRAP PARAMÉTRICO

El Bootstrap Paramétrico (conocido como Monte Carlo), consiste en simular conjuntos de datos adicionales, mediante un modelo de clases latentes *hipotetizado* como el apropiado para los datos de la muestra. La estrategia consiste primero, en calcular los parámetros del modelo para la muestra y segundo, se generan nuevos conjuntos de datos ( $\mathbf{x}_{1b}^*, \dots, \mathbf{x}_{nb}^*$ ), con el propósito de encontrar la distribución empírica del estadístico de bondad de ajuste. Este procedimiento de simulación de datos y reestimación de parámetros se repite  $B$  veces, o hasta alcanzar un criterio de precisión para que se cumpla la aproximación.

El algoritmo de Bootstrap Paramétrico para variables manifiestas binarias consiste en las siguientes 5 etapas:

**Etapas 1.** Ajustar el MCL supuesto como el apropiado a los datos originales. Esto produce las estimaciones de los parámetros y los EBA.

**Etapas 2.** Calcular el número de elementos para cada clase latente utilizando los resultados calculados anteriormente (etapa 1).

---

<sup>4</sup> El análisis gráfico no proporciona una medida analítica de la bondad de ajuste, sólo permite comprobar visualmente si un histograma se parece o no a una determinada función de densidad.

**Etap 3.** Generar los datos para las clases latentes dados los parámetros de la Etapa 1 utilizando la distribución multinomial. Primero, suponiendo una matriz de datos  $X(i, j)$  donde  $i = 1, 2, \dots, n$  son individuos y  $j = 1, 2, \dots, p$  las variables binarias. Segundo, a partir del modelo estimado en la etapa 1 se tienen las probabilidades condicionales  $\pi_{X_j/Y(c)}$  y los tamaños de las clases latentes  $\pi_Y(c)$ . Donde  $j = 1, 2, \dots, p$  es el número de variables (ítems) y  $c = 1, 2, \dots, T$  es el número de clases latentes.

Con el objetivo de generar los patrones de respuesta correspondientes a cada clase latente se debe seguir el siguiente procedimiento:

- A. Para la clase latente  $c$ , calcular el número de patrones de respuesta  $n_c$  utilizando la distribución multinomial.
- B. Generar un número aleatorio ( $\#rand$ ) con distribución uniforme<sup>5</sup> entre 0 y 1.
- C. Comparar el número simulado con los parámetros del modelo  $\pi_{X_j/Y(c)}$ . Si  $\#rand > \pi_{X_j/Y(c)}$  el valor Bootstrap simulado para la posición  $j$  y  $c$  es igual a 1, entonces  $X(i, j) = 1$ . En caso contrario  $X(i, j) = 0$ . A continuación se debe regresar al **paso B** hasta completar el patrón de respuesta.
- D. Repetir los **pasos B y C** hasta generar todos los patrones de respuesta para la clase latente  $c$ .

---

<sup>5</sup> Es una distribución de probabilidad cuyos valores tienen la misma probabilidad. Para variables continuas, se dice que una variable aleatoria  $X$  continua tiene una distribución uniforme en el intervalo  $[a, b]$  si la función de densidad de probabilidad es,

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{para el resto} \end{cases}$$

- E. Regresar al Paso **A**, hasta completar los patrones de respuestas de todas las clases latentes, tal que, el tamaño de la muestra Bootstrap sea  $n = n_1 + \dots + n_T$  donde  $c = 1, 2, \dots, T$ .

**Etapas 4.** Volver a la Etapa 3 hasta completar las  $B$  remuestras Bootstrap Paramétricas.

**Etapas 5.** Ajustar el modelo hipotético a las  $B$  muestras. Los valores replicados de los estadísticos de bondad de ajuste obtenidos mediante el proceso de Bootstrap proporcionan una valoración de la verdadera distribución del estadístico.

**Etapas 6.** Comparar los estadísticos de bondad de ajuste de la muestra original para el  $(1 - \alpha)$  B-ésimo percentil, con los estadísticos de bondad de ajuste Bootstrap ordenados de menor a mayor de la siguiente manera: si el estadístico de bondad de ajuste es mayor que el  $(1 - \alpha)$  **B**-ésimo percentil, rechazar el modelo; si es menor, no rechazar el modelo.

Un acercamiento en la utilización del Bootstrap Paramétrico, la encontramos en COLLINS et al. (1993). Ellos estudiaron por medio de una simulación de datos con problemas de “sparse” (tablas poco ocupadas) la distribución de probabilidad empírica de los estadísticos de bondad de ajuste: Razón de Verosimilitud ( $G^2$ ), estadístico Pearson ( $Q^2$ ) y Cressie-Read ( $CR$ ). Hallando que la media de la distribución empírica del estadístico  $Q^2$  es igual a los grados de libertad, es decir, al valor esperado de la distribución  $\chi^2$ . Pero la desviación típica es considerablemente mayor que la encontrada con los estadísticos Razón de Verosimilitud y Cressie-Read. Por esta razón, recomiendan no utilizar el estadístico de Pearson ( $Q^2$ ) para modelos de clases latentes en presencia de tablas poco ocupadas. Concluyen que ninguno de los EBA es una opción clara para la evaluación de un modelo de clases latentes. El promedio de los estadísticos difiere del valor esperado con la distribución teórica  $\chi^2$ , especialmente la Razón de Verosimilitud y el estadístico Cressie-Read. La desviación típica también es diferente a la raíz cuadrada de dos

veces los grados de libertad  $(\sqrt{2 gl})^6$ , resultando que  $Q^2$  tiene la mayor variabilidad de estos tres estadísticos.

Posteriormente, LANGEHEINE et al. (1996) concluyen que las tablas poco ocupadas, usualmente no causan problemas en la estimación de los parámetros, pero la evaluación del ajuste del modelo de clases latentes, se puede dificultar por el hecho de que la verdadera distribución del estadístico de bondad de ajuste, es una aproximación inadecuada de la distribución teórica  $\chi^2$ . Además de lo anterior, *proponen* utilizar la siguiente regla de decisión para el diagnóstico de modelos de clases latentes en tablas poco ocupadas, a través del Bootstrap Paramétrico:

*Rechazar el modelo de clases latentes para un nivel de significación ( $\alpha$ ), si al menos ( $\alpha B$ ) remuestras Bootstrap Paramétricas, tienen un valor de la Razón de Verosimilitud ( $G^2$ ) mayor al observado con los datos originales con un nivel de confianza del  $(1 - \alpha)\%$ .*

Asimismo los autores concluyen que los resultados de los estadísticos Bootstrap pueden variar en tablas poco ocupadas. Por esta razón, proponen rechazar un modelo si, por lo menos, un estadístico de bondad de ajuste es rechazado por el procedimiento Bootstrap. Demostraron que el poder de discrepancia es mayor para el estadístico de Pearson ( $Q^2$ ) y seguidamente por los estadísticos Cressie-Read (CR) y Razón de Verosimilitud ( $G^2$ ). La diferencia entre los estadísticos de bondad de ajuste, tiende a ser más grande, a medida que aumenta la cantidad de celdas vacías en la tabla múltiple.

Von DAVIER (1997) demostró que los *estadísticos* de bondad de ajuste Pearson y Cressie-Read, proporcionan resultados similares aun cuando, un alto porcentaje de los patrones de respuesta tenga frecuencias observadas bajas o nulas. Por tanto, afirma que no hay necesidad de aplicar cualquier otro estadístico de bondad de ajuste. En contraste, los estadísticos Freeman-Tukey y Razón de Verosimilitud pueden conducir a decisiones incorrectas en diversas condiciones de tablas poco ocupadas. Él propone una regla de decisión para rechazar un modelo

---

<sup>6</sup> Este término es la varianza teórica de una variable con distribución de probabilidad  $\chi^2$ .



de clases latentes similar a la propuesta por LANGEHEINE et al. (1996), pero más general, para cualquier estadístico de bondad de ajuste.

### 3.7.4 COMPARACIÓN ENTRE BOOTSTRAP PARAMÉTRICO Y NO PARAMÉTRICO

Respecto a la confianza que merecen los métodos Bootstrap, BOLLEN & STINE (1992) demuestran que es un error tratar de estimar las frecuencias de las tablas poco ocupadas remuestreando los datos de la muestra original. Demuestran que el Bootstrap No Paramétrico, generará una distribución de probabilidad con un valor esperado de la Razón de Verosimilitud dos veces más grande que el valor teórico según la distribución  $\chi^2$ ,  $E(G^2) = 2gl$ . En este mismo sentido, von DAVIER (1997) afirma que la distribución empírica del estadístico de bondad de ajuste  $\hat{F}$ , puede estar demasiado lejos de la verdadera distribución  $F$ .

Con el propósito de mostrar el problema del Bootstrap No Paramétrico, BOLLEN & STINE (1992) consideraron el estadístico  $Z$  para el contraste de la hipótesis nula  $\mu = 0$ , para una muestra aleatoria  $\{x_1, x_2, \dots, x_n\}$  proveniente de una distribución normal con media  $\mu$  y varianza conocida  $\sigma^2 = 1$ . Entonces, la distribución del estadístico de contraste

$$Z = \frac{\bar{x}}{\sigma/\sqrt{n}} = \sqrt{n}\bar{x} \quad (3.6)$$

es normal estándar con media cero y varianza 1. Así,  $Z^2$  tiene una distribución  $\chi^2$  con 1 grado de libertad. A través del Bootstrap No Paramétrico son generadas  $B$  remuestras  $\{x_1^*, x_2^*, \dots, x_n^*\}$ , se calcula el promedio  $\bar{x}^*$  para cada remuestra y el estadístico  $Z^{*2} = n\bar{x}^{*2}$ . De esta forma se obtienen el conjunto de valores  $\{Z_1^{*2}, Z_2^{*2}, \dots, Z_B^{*2}\}$ . El valor esperado de las réplicas Bootstrap para la media es

$$E^*(\bar{x}^*) = \bar{x} \quad (3.7)$$

y la varianza es

$$Var^*(\bar{x}^*) = v^2/n \quad (3.8)$$

donde

$$v^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.9)$$

Con estos resultados, podemos calcular el valor esperado de las réplicas Bootstrap de  $Z^{*2}$  y descubrir la fuente del problema de este esquema de remuestreo Bootstrap No Paramétrico. Por medio de los resultados de las ecuaciones (3.6), (3.7) y (3.8), se deriva que

$$\begin{aligned} E^*(Z^{*2}) &= nE^*(\bar{x}^{*2}) \\ &= n E^*(\bar{x}^* - \bar{x} + \bar{x})^2 \\ &= nE^*[(\bar{x}^* - \bar{x})^2] + n\bar{x}^2 \\ &= n Var^*(\bar{x}^*) + n\bar{x}^2 \\ &= v^2 + Z^2 \end{aligned} \quad (3.10)$$

Teniendo en cuenta que la ilustración es para una muestra aleatoria  $\{x_1, x_2, \dots, x_n\}$  con una distribución de probabilidad normal estándar, implica que  $E(Z^2) = 1$ , mientras que por la ecuación (3.10) resulta que el valor esperado de  $Z^{*2}$  es

$$\begin{aligned} E(v^2 + Z^2) &= \frac{1}{n} \sum_{i=1}^n [E(x_i - \bar{x})^2] + E(Z^2) \\ &= \frac{1}{n} E \left[ \sum_{i=1}^n x_i^2 - n\bar{x}^2 \right] + E(Z^2) \\ &= \frac{n-1}{n} + 1 \\ &= 2 - 1/n \end{aligned} \quad (3.11)$$

Finalmente, se deduce que la media de  $Z^2$  es 1 pero el promedio de  $Z^{*2}$  es aproximadamente 2.

A pesar de la argumentación de lo anterior, ALBANESE & KNOTT (1994) obtuvieron resultados similares con el Bootstrap Paramétrico y No Paramétrico para modelos de clases latentes con variables manifiestas binarias. Asimismo,

DIAS (2005) comparó el promedio y mediana de las estimaciones de los parámetros para el modelo de clases latentes, concluyendo que las diferencias entre los métodos Bootstrap son pequeñas.

El procedimiento Bootstrap No Paramétrico proporciona resultados correctos en muchas ocasiones, pero también puede fallar según las características de la muestra original. ARAYA et al. (2009) simularon cuatro conjuntos de datos binarios, correspondientes a tamaños de muestra de 50, 100, 200 y 500 individuos, respectivamente, para un modelo con 5 variables manifiestas y 3 clases latentes (con probabilidades de clase de 0,50, 0,30 y 0,20); y cuatro conjuntos de datos binarios, correspondientes a tamaños de muestras de 300, 500, 1000 y 2000 individuos, respectivamente, correspondientes a un modelo con 8 variables manifiestas y 5 clases latentes (con probabilidades de clase de 0,40, 0,30, 0,15, 0,10 y 0,05). Para todos los modelos estudiados fueron simuladas 500 remuestras Bootstrap.

Los autores probaron empíricamente, que las estimaciones obtenidas utilizando el Bootstrap Paramétrico presentan menor diferencia promedio respecto al valor calculado con la muestra original. Obsérvese por ejemplo en la **Tabla 3.7**, para un tamaño de muestra de 500 individuos el estadístico  $Q^2$  resultante es 17,85. En tanto, la media aritmética del Bootstrap Paramétrico es 14,12 y 31,74 para el Bootstrap No Paramétrico. En general, las desviaciones estándar de los estadísticos de bondad de ajuste de las remuestras Bootstrap Paramétrico son significativamente más pequeñas en comparación al Bootstrap No Paramétrico (véase las columnas 5 y 8).

El modelo de tres clases latentes es correctamente confirmado por el Bootstrap Paramétrico y el Bootstrap No Paramétrico. La diferencia se encuentra en que, el Bootstrap No Paramétrico aporta *p-valores* más altos en los 4 estadísticos de bondad de ajuste. Por ejemplo, obsérvese que el *p-valor* de la Razón de Verosimilitud de la muestra tamaño 200 para el Bootstrap Paramétrico resulta 0,667, en tanto, el Bootstrap No Paramétrico proporciona un valor de 0,980.

Estadístico	Tamaño de la muestra	Muestra Original	Paramétrico			No Paramétrico		
			Media	Desviación estándar	P-Valor	Media	Desviación estándar	P-Valor
Pearson	50	13,91	14,98	6,13	0,535	22,35	7,56	0,912
	100	8,92	14,53	5,12	0,904	20,32	6,12	0,992
	200	12,67	16,08	6,49	0,675	28,57	9,22	0,980
	<b>500</b>	17,85	14,12	5,68	0,202	31,74	9,72	0,914
Cressie-Read	50	13,91	13,32	5,65	0,375	20,78	6,69	0,856
	100	9,33	14,26	4,86	0,886	20,06	5,86	0,990
	200	12,88	15,95	6,34	0,667	28,34	8,91	0,982
	500	18,17	14,20	5,68	0,204	32,15	9,82	0,934
Razón de Verosimilitud	50	17,01	15,13	5,93	0,329	23,23	6,87	0,828
	100	11,13	16,06	5,34	0,862	22,55	6,46	0,992
	<b>200</b>	<b>14,05</b>	<b>17,35</b>	<b>6,78</b>	<b>0,667</b>	<b>30,81</b>	<b>9,53</b>	<b>0,980</b>
	500	19,03	14,80	6,01	0,202	34,17	10,76	0,932
Freeman-Tukey	50	25,67	28,07	11,18	0,555	37,67	11,52	0,866
	100	15,71	23,37	8,22	0,854	32,22	9,80	0,992
	200	17,00	22,57	9,25	0,707	39,79	13,01	0,984
	500	19,90	16,24	7,38	0,249	38,05	13,10	0,940

**TABLA 3.7.** Estadísticos para el modelo con cinco variables manifiestas binarias y tres clases latentes.

Con respecto a los conjuntos de datos correspondientes a un modelo con cinco clases latentes, el Bootstrap Paramétrico determinó (correctamente) que el modelo con dos clases latentes no era apropiado, en tanto, el Bootstrap No Paramétrico falló al no rechazar el modelo erróneo (**Tabla 3.8**). Las medias aritméticas de los estadísticos de bondad de ajuste del Bootstrap Paramétrico son semejantes a las alcanzadas con la muestra original, además presentan una menor desviación estándar en comparación al Bootstrap No Paramétrico. Estos resultados están en concordancia con la argumentación de BOLLEN & STINE (1992) sobre las limitaciones del Bootstrap No Paramétrico, tal que las medias aritméticas con todos los tamaños de muestras simulados (300, 500, 1000 y 2000) son aproximadamente dos veces las esperadas; es decir aquellas conseguidas con la muestra original.

En la siguiente sección, se presenta la forma de decidir cuándo rechazar o no rechazar un modelo hipotético de clases latentes, utilizando la significación calculada (*p-valor*) para el Bootstrap No Paramétrico y Paramétrico.

Estadístico	Tamaño de la muestra	Muestra Original	Paramétrico			No Paramétrico		
			Media	Desviación estándar	P-Valor	Media	Desviación estándar	P-Valor
Pearson	300	289,62	238,85	25,41	0,038	520,99	57,27	1,000
	500	313,12	237,89	23,84	0,008	535,99	65,50	1,000
	1000	416,80	241,15	31,32	0,008	641,45	62,87	1,000
	2000	545,37	238,19	27,71	0,002	775,64	74,88	1,000
Cressie-Read	300	261,60	221,18	20,27	0,026	428,45	36,13	1,000
	500	300,57	230,76	20,38	0,006	484,51	44,02	1,000
	1000	379,74	239,56	28,37	0,008	584,32	47,24	1,000
	2000	502,97	237,76	26,21	0,002	718,05	58,85	1,000
Razón de Verosimilitud	300	267,55	233,35	19,20	0,028	390,37	27,19	1,000
	500	324,24	255,55	20,13	0,006	487,65	31,65	1,000
	1000	365,70	264,88	28,05	0,008	580,41	39,41	1,000
	2000	483,56	253,75	26,41	0,002	707,09	47,89	1,000
Freeman-Tukey	300	387,37	337,89	29,02	0,050	536,87	37,49	1,000
	500	458,58	370,09	29,76	0,006	679,29	41,38	1,000
	1000	448,57	358,97	38,50	0,012	765,69	54,48	1,000
	2000	567,09	307,57	33,91	0,002	866,03	58,08	1,000

**TABLA 3.8.** Estadísticos para el modelo con ocho variables manifiestas binarias y cinco clases latentes.

### 3.7.5 CÁLCULO DEL NIVEL DE SIGNIFICACIÓN

El fundamento del método Bootstrap para evaluar el grado de significación (es decir, *p-valor*)<sup>7</sup> de un modelo de clases latentes es el siguiente:

Supóngase un conjunto de datos obtenidos a través de una muestra aleatoria de  $n$  individuos y  $p$  variables manifiestas binarias  $\mathbb{X} = (X_1, X_2, \dots, X_n)$ . De esta forma, cada  $\mathbf{X}'_i = (X_1, \dots, X_p)$  representa el vector columna de respuestas sobre el  $i$ -ésimo individuo. Ahora, si consideramos un modelo hipotético de clases latentes y  $\theta = f(\mathbb{X})$  es la función que define el estadístico de bondad de ajuste, entonces  $\theta$  es una variable aleatoria que tiene su propia distribución de probabilidad, tal que  $\mathbb{X}$

<sup>7</sup> El *p-valor* es la probabilidad de que el estadístico de prueba tome un valor mayor o igual al valor observado del estadístico cuando la hipótesis nula es verdadera. Es decir, es el menor nivel de significación con el que se puede rechazar la hipótesis nula.

es una matriz aleatoria. La probabilidad de que  $\theta$  pueda ser mayor o igual a un valor determinado  $k_0$ , definido por la distribución muestral de  $\theta$ , se expresa como,

$$pv = P[\theta \geq k_0] \quad (3.12)$$

Si  $\hat{\theta}$  es el valor obtenido en la muestra original, la prueba de significación consistirá en calcular cuán inusual es  $\hat{\theta}$  en relación a la distribución muestral de  $\theta$ . De esta manera, el nivel de significación de la prueba estadística es,

$$pv = P[\theta \geq \hat{\theta}] \quad (3.13)$$

$$pv = \frac{\text{Número de veces que } \{\theta \geq \hat{\theta}\} + 1}{B + 1}$$

La regla de decisión sobre el modelo hipotético consistirá en,

$$\text{Rechazar el modelo de clases latentes si } pv \leq \alpha$$

siendo  $\alpha$  el nivel de significación establecido “a priori”.

Por tanto, el cálculo del nivel de significación se fundamenta en estimar la distribución muestral del estadístico de bondad de ajuste bajo el modelo de clases latentes hipotético. De esta forma, el nivel de significación unilateral es, sencillamente, la proporción de muestras simuladas en las cuales el valor de  $\theta$  es mayor o igual al calculado en la muestra original  $\hat{\theta}$ . Los valores grandes de  $\theta$  proporcionan evidencia para rechazar el modelo de clases latentes.

### 3.7.6 GRADOS DE LIBERTAD EN TABLAS POCO OCUPADAS

Un problema con los modelos de clases latentes en tablas poco ocupadas es la determinación de los grados de libertad y el número efectivo de parámetros para ajustar el modelo. En tablas poco ocupadas no todos los patrones de respuestas

tienen frecuencias mayores que cero, por lo cual, se presentan problemas de selección y comparación de modelos de clases latentes (COLLINS et al., 1993).

En estas situaciones, los grados de libertad, varían en función de las dimensiones de la tabla y del número de celdas vacías, que produce una disminución gradual de los grados de libertad en la medida que se incrementan los patrones de respuesta con frecuencias observadas nulas. En cierto modo, los grados de libertad no son fijados por el modelo, sino que pasan a depender de las características de los datos y se convierten en variables aleatorias.

Los grados de libertad ( $gl$ ) corregidos de un modelo de clases latentes se obtienen eliminando las celdas vacías. Para un modelo de clases latentes con variables binarias los grados de libertad se calculan como,

$$gl = (2^p - 1) - [T * p + (T - 1)] \quad (3.14)$$

donde:  $T$  = Número de clases latentes.

$p$  = Número de variables manifiestas binarias.

Se elimina un grado de libertad por cada celda de la tabla múltiple con frecuencia nula de la siguiente manera,

$$gl_{cor} = gl - \Delta \quad (3.15)$$

donde:  $\Delta$  = Número de celdas con frecuencias nulas.

Por ejemplo, considere un modelo de clases latentes con cinco variables manifiestas binarias y dos clases latentes. Además, supóngase que dos patrones de respuestas tienen frecuencia cero. Los grados de libertad deberán ser corregidos de la siguiente forma:

$$gl = (2^5 - 1) - [2 * 5 + (2 - 1)] = 20$$

Veinte son los grados de libertad sin corregir, al tener dos celdas con frecuencias nulas ( $\Delta = 2$ ), los grados de libertad corregidos del modelo de clases latentes son dieciocho.

$$gl_{cor} = 20 - 2 = 18$$

### 3.8 BOOTSTRAP NO PARAMÉTRICO CORREGIDO

En este apartado vamos a proponer un nuevo método que lo llamaremos “Bootstrap No Paramétrico Corregido” (BNPC). EL BNPC resuelve el problema del Bootstrap No Paramétrico (BNP) en la estimación de la distribución de probabilidad empírica, causado por los patrones de respuesta que no son observados en la muestra original. En el BNP, un patrón de respuesta que no se observa en la muestra, tiene probabilidad cero de ser seleccionado en las remuestras Bootstrap y, en consecuencia, la distribución de probabilidad empírica puede ser una mala aproximación de la verdadera distribución.

En general, si tenemos  $p$  variables dicotómicas, entonces  $R = 2^p$  es el número posible de patrones de respuesta. En la muestra original de datos, la proporción de patrones observado del total  $R$  la simbolizaremos como  $p_1$ . Esto implica que,  $p_2 = 1 - p_1$  es la proporción que no se presentaron y son la principal causa de la mala estimación de la distribución de probabilidad empírica del Bootstrap No Paramétrico.

La solución propuesta consiste en seleccionar  $n_1 = np_1$  patrones de respuesta de la muestra original y  $n_2^* = n(1 - p_1) = np_2$  de la población de patrones. Vamos a considerar que  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_1}\}$  simboliza la muestra que proviene de los datos originales y  $(\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_{n_2}^*)$  representa los patrones de respuesta tomadas de  $R$ , de modo que  $n = n_1 + n_2^*$  constituyen una remuestra BNPC. Así, la proporción de patrones de respuesta en cada remuestra es mayor que la proporción observada en la muestra original, teniéndose mayor cantidad de información para estimar la distribución de probabilidad empírica del estadístico de bondad de ajuste  $\theta$ . De esta forma, se obtiene el conjunto de estadísticos Bootstrap  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$  y el análisis de los  $\hat{\theta}_b^*$  da información sobre la distribución muestral de  $\theta$ .

El Bootstrap No Paramétrico Corregido se ejecuta a través del procedimiento siguiente:

1. Ajustar el modelo hipotetizado a los datos de la muestra original.



2. Generar  $B$  remuestras del mismo tamaño que la muestra original mediante la simulación de  $n_1$  y  $n_2^*$  patrones de respuesta, de modo que  $n_1 + n_2^*$  constituyen el tamaño de la remuestra.
3. Ajustar el modelo hipotetizado a cada remuestra, obteniéndose  $B$  resultados del estadístico  $\hat{\theta}^*$ .
4. Construir el histograma a partir de los  $B$  estadísticos mediante la asignación de una probabilidad de  $1/B$  cada valor  $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ .
5. Rechazar el modelo si el  $p$ -valor<sup>8</sup> del Bootstrap es menor al nivel de significación fijado  $\alpha$ .

En el **Capítulo 6**, se compara el Bootstrap No Paramétrico Corregido (propuesto) con los otros métodos de diagnóstico de modelos en tablas poco ocupadas, considerando datos simulados.

A continuación, vamos a exponer una de nuestras *propuestas innovadoras para el diagnóstico de modelos de clases latentes en tablas poco ocupadas, el Método de Patrones*. Primero, hacemos una descripción teórica del Método de Patrones, y a continuación presentamos las fórmulas de cálculo del promedio aritmético, desviación típica y nivel de significación. Al final, se presenta un ejemplo ilustrativo sobre los pasos a seguir con nuestra propuesta, para la selección de un modelo de clases latentes utilizando la distribución de probabilidad empírica de los estadísticos de bondad de ajuste.

---

<sup>8</sup> Véase el apartado (3.7.5) sobre el cálculo del nivel de significación observado ( $p$ -valor).

## **CAPÍTULO 4**

---

### **PROPUESTA: MÉTODO DE PATRONES**

---

## 4.1 INTRODUCCIÓN

En este capítulo proponemos un procedimiento alternativo para conocer la distribución de probabilidad empírica de los estadísticos de bondad de ajuste (EBA), en tablas de frecuencias de patrones de respuesta poco ocupadas, que llamaremos *Método de Patrones*. El problema de los patrones de respuesta con frecuencias nulas, fue abordado por DIAS & VERMUNT (2006), adjunto copia del texto en inglés para un mejor criterio.

*Bootstrap methods for measuring classification uncertainty in latent class analysis*

*José G. Dias and Jeroen K. Vermunt*

*In the nonparametric bootstrap, because the sampling is from the empirical distribution, a data pattern that is not observed in the sample has probability zero of being selected into the bootstrap samples and, consequently,  $\hat{\mathbf{F}}_{\text{nonpar}}$  may be too far from the true distribution  $\mathbf{F}$ . The same problem can, however, also occur in the parametric bootstrap, namely, when certain parameter estimates are on the boundary of the parameter space. In such a case, the resampling will not show any variability within the component concerned, although zero estimated cell frequencies are very rare.*

---

En el presente capítulo se exponen los fundamentos metodológicos de la propuesta, el cálculo del promedio aritmético y la varianza, además del cálculo del nivel de significación para la selección de modelos de clases latentes utilizando los EBA. Al final del capítulo, se hace una ilustración comparativa entre los métodos Bootstrap y el *Método de Patrones*.

El *Método de Patrones* (MP) es una propuesta innovadora para el diagnóstico de modelo de clases latentes con conjuntos de datos que tienen problemas de poca densidad (en inglés, “sparseness”). La idea básica del *Método de Patrones* es tratar al número de patrones posibles ( $R = 2^p$ ) como si fuera la población, y aplicar el muestreo aleatorio simple con reemplazamiento sobre esa población de patrones, con el objetivo de simular muestras de igual tamaño que la muestra original, para construir la distribución de probabilidad empírica del estadístico de bondad de

ajuste, a diferencia del Método Bootstrap No Paramétrico que genera muestras de los datos originales o del Bootstrap Paramétrico que supone un modelo hipotético de clases latentes para reproducir nuevos conjuntos de datos.

#### 4.1.1 DESCRIPCIÓN DEL MÉTODO

Una característica básica del *Método de Patrones* es sustituir la función subyacente de la distribución de probabilidad desconocida  $F(\mathbf{X})$  por un estimador de la misma. Asimismo se emplea el muestreo con reemplazamiento de los patrones de respuesta, para obtener, un gran número de muestras aleatorias sobre las que se realiza la estimación. La distribución de probabilidad empírica  $\hat{F}(\mathbf{X})$ , obtenida a partir de los patrones de respuesta, asigna una probabilidad de  $1/R$  a cada patrón de respuesta  $\mathbf{X}_r$ , para  $r = 1, 2, \dots, R$ , donde  $R$  es el número total de patrones ( $R = 2^p$ ) para  $p$  variables manifiestas binarias (véase **Figura 4.1**).

Considerando la población de patrones de respuesta  $R$ , se obtiene un patrón en forma aleatoria con probabilidad  $1/R$ . Iterando este proceso  $n$  veces, se obtiene un conjunto de datos que constituyen la denominada muestra aleatoria de patrones, que representaremos mediante la notación  $(\mathbf{x}_{1a}^*, \mathbf{x}_{2a}^*, \dots, \mathbf{x}_{na}^*)^1$ , utilizando  $a$  para denotar la  $a$ -ésima muestra de patrones, para  $a = 1, 2, \dots, A$ . El valor  $A$  es el número total de muestras aleatorias de patrones que se consideran aceptables para hacer la estimación de la distribución de probabilidad empírica del estimador de bondad de ajuste<sup>2</sup>.

Así, aunque cada muestra aleatoria de patrones tendrá el mismo número de elementos que la muestra original, mediante el muestreo aleatorio con reemplazamiento, las muestras pueden tener la mayoría de patrones de la muestra original, y otros nuevos que son parte de la población de patrones que no fueron

---

<sup>1</sup> Para hacer la diferencia entre los patrones de respuesta  $\mathbf{X}_{ib}^*$  obtenidos con los métodos Bootstrap de los calculados con el Método de Patrones, denotaremos a estos últimos como  $\mathbf{X}_{ia}^*$ .

<sup>2</sup> Una consecuencia directa es que a medida que aumenta  $A$ , mejor será la estimación que podemos hacer sobre la distribución empírica del estadístico de bondad de ajuste, y por tanto, implicaría el aumento de la potencia de la prueba de hipótesis sobre el modelo de clases latentes.

observados por medio de la recolección de los datos, debido a que el tamaño de la muestra puede ser más pequeño que el número posible de patrones de respuesta<sup>3</sup>. Por ejemplo, si tenemos datos para una muestra de 500 personas, considerando 10 variables manifiestas binarias, son posibles un total de  $R = 2^{10} = 1024$  patrones de respuesta, luego es evidente que no todos los posibles patrones se van a presentar en la muestra.

Para cada una de estas muestras aleatorias de patrones, se obtiene el estadístico de bondad de ajuste de interés, dado un modelo de clases latentes propuesto como el apropiado para la muestra original. Con el propósito de diferenciar el estadístico de bondad de ajuste calculado sobre los valores de la muestra original  $\hat{\theta}$ , lo denotaremos mediante,

$$\hat{\theta}_a^* = (\mathbf{x}_{1a}^*, \mathbf{x}_{2a}^*, \dots, \mathbf{x}_{na}^*) \quad (4.1)$$

para  $i = 1, 2, \dots, n$  y  $a = 1, 2, \dots, A$ .

El subíndice  $a$  en  $\mathbf{x}_{ia}^*$  indica una muestra aleatoria de patrones cualquiera y  $\hat{\theta}_a^*$  es el estadístico de bondad de ajuste para la muestra  $a$ .

Repitiendo el procedimiento el número de veces que se determine satisfactorio, obtenemos la estimación de la distribución de probabilidad empírica de  $\hat{\theta}^*$ . Si consideramos la población teórica de patrones como el conjunto de valores posibles, partiendo de un conjunto de variables manifiestas binarias, tendremos que el estadístico de bondad de ajuste ligado a los sucesos que ocurren tomará un conjunto de valores y la frecuencia relativa de estos valores irá tendiendo a un límite que constituye la definición de distribución de probabilidad empírica.

---

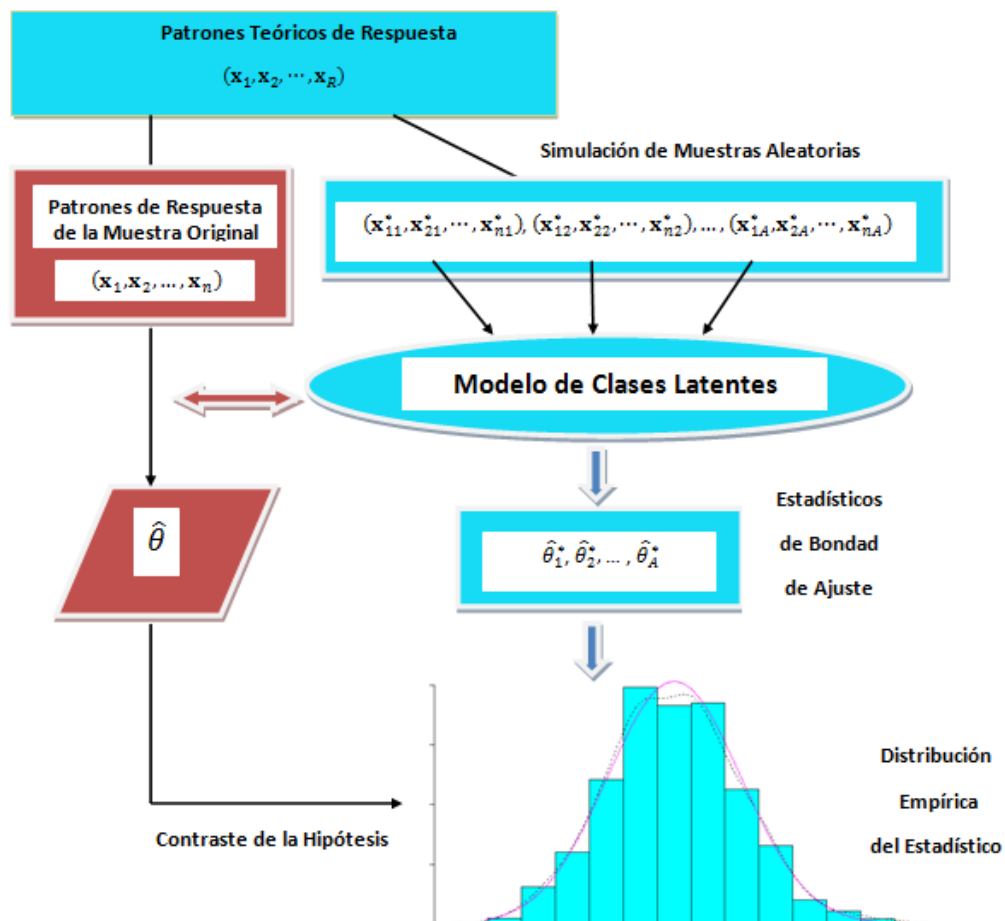
<sup>3</sup> Hay que distinguir entre ceros aleatorios en las frecuencias de los patrones de respuesta, cuando su aparición puede explicarse por azar, y los ceros estructurales cuando su aparición es imposible, en cuyo caso la tabla se denomina incompleta. Por ejemplo, cuando se cruzan las variables “sexo” y “si presenta el ciclo menstrual”, no puede existir un patrón de respuesta indicando un “varón con dolores debido al ciclo menstrual”.

Una vez generadas las distintas muestras aleatorias de patrones, es factible estimar la distribución de probabilidad empírica<sup>4</sup> de  $\hat{\theta}^*$ , asignando una probabilidad  $1/A$  a cada punto,

$$\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_A^* \quad (4.2)$$

Esta distribución se convierte así en un estimador inmediato para la función de distribución de  $\theta$  y puede usarse especialmente, para contrastar la validez del modelo de clases latentes hipotético como el apropiado para los datos originales.

En resumen, podemos visualizar el proceso de construcción de la distribución empírica del estadístico de bondad de ajuste por medio del *Método de Patrones*, desde un punto de vista esquemático de la siguiente forma:



**Figura 4.1.** Esquema del Método de Patrones

<sup>4</sup> Para representar la distribución empírica del estadístico de bondad de ajuste se utiliza el histograma.

#### 4.1.2 PROMEDIO ARITMÉTICO Y VARIANZA DE $\hat{\theta}^*$

El promedio aritmético del estadístico de bondad de ajuste, se obtiene a través de las  $A$  muestras aleatorias generadas con el *Método de Patrones*, utilizando la siguiente fórmula,

$$\bar{\theta}^* = \frac{1}{A} \sum_{a=1}^A \hat{\theta}_a^* \quad (4.3)$$

donde:  $\hat{\theta}_a^*$  = Estimador de bondad de ajuste para la  $a$ -ésima muestra aleatoria de patrones.

$A$  = Número de muestras de patrones.

En tanto, la estimación de la varianza de  $\hat{\theta}^*$ , es calculada usando la distribución muestral generada empíricamente por el *Método de Patrones* y se obtiene mediante la expresión,

$$\widehat{Var}(\hat{\theta}^*) = \frac{1}{A-1} \sum_{a=1}^A (\hat{\theta}_a^* - \bar{\theta}^*)^2 \quad (4.4)$$

donde:  $\hat{\theta}_a^*$  = Estimador del estadístico de bondad de ajuste de la  $a$ -ésima muestra aleatoria de patrones de respuesta.

$\bar{\theta}^*$  = Estimador del promedio aritmético para todas las muestras.

Se espera que la varianza estimada del estadístico de bondad de ajuste sea pequeña, lo cual estaría reflejando mucha similitud entre los valores  $\hat{\theta}_a^*$  obtenidos en la  $A$  muestras aleatorias.

### 4.1.3 CÁLCULO DEL NIVEL DE SIGNIFICACIÓN

El conjunto de estos estadísticos  $\hat{\theta}^*$ , constituyen una distribución de probabilidad empírica con la que comparar el estadístico de bondad de ajuste calculado a partir de los datos originales. Para realizar la prueba, se calcula el *p-valor* o nivel de significación empírico del contraste, que es la probabilidad calculada de asumir que  $H_0$  (el modelo de clases latentes) es apropiado para los datos.

Se deben ordenar los  $A$  resultados de los estadísticos de bondad de ajuste de menor a mayor,

$$\hat{\theta}_1^* \leq \hat{\theta}_2^* \leq \dots \leq \hat{\theta}_A^* \quad (4.5)$$

Seguidamente, el *p-valor* del contraste se calcula como,

$$pv = P[\hat{\theta}^* \geq \hat{\theta}] \quad (4.6)$$

$$pv = \frac{\text{Número de Muestras } \{\hat{\theta}^* \geq \hat{\theta}\} + 1}{A + 1}$$

donde:  $\hat{\theta}$  = Estadístico de bondad de ajuste para la muestra original.  
 $\hat{\theta}^*$  = Estadístico de bondad de ajuste para cualquiera de las muestras aleatorias de patrones.  
 $A$  = Número de muestras aleatorias de patrones.

El *p-valor* se estima como la proporción de muestras aleatorias simuladas en las cuales el valor de  $\hat{\theta}^*$  es mayor o igual al calculado con la muestra original  $\hat{\theta}$ . Los valores grandes de  $\hat{\theta}^*$  son la evidencia para no rechazar la hipótesis nula, es decir, el modelo de clases latentes propuesto como el apropiado para los datos de la muestra. Además, un *p-valor* pequeño implica que los datos de la muestra original son inverosímiles (o poco probables) bajo la hipótesis nula. La regla de decisión para un nivel de significación  $\alpha$ , consiste en rechazar  $H_0$  (o sea, el modelo hipotético) si  $p - valor < \alpha$ .



#### 4.1.4 ILUSTRACIÓN DEL MÉTODO DE PATRONES

Partimos de un estudio simulado con una muestra aleatoria de 124 individuos, los cuales recogen información sobre 4 variables manifiestas binarias (1=sí, 0=no). Para la simulación de los datos se utilizó el programa LEM (VERMUNT, 1997b) a través de las siguientes instrucciones:

```
Simulating data
lat 1
man 4
dim 2 2 2 2 2
mod X A|X B|X C|X D|X
sim 124 Datos.dat
sta X [0.85 0.15]
sta A|X [1.00 0.00 0.31 0.69]
sta B|X [0.78 0.22 0.12 0.98]
sta C|X [0.00 1.00 0.20 0.80]
sta D|X [0.90 0.10 0.00 1.00]
```

El programa LEM generó el archivo de datos partiendo de los parámetros definidos por las probabilidades de las clases y de las probabilidades condicionales. La siguiente **Tabla 4.1**, presenta las frecuencias observadas de los patrones de respuesta.

$X_1$	$X_2$	$X_3$	$X_4$	Frecuencia $f_{x_h}$
0	0	0	1	1
0	0	1	0	71
0	0	1	1	10
0	1	0	1	1
0	1	1	0	18
0	1	1	1	8
1	0	1	1	1
1	1	0	1	3
1	1	1	1	11

**TABLA 4.1.** Frecuencias observadas para los distintos patrones de respuesta simulados.

El 56,3% del total de *patrones teóricos posibles* ( $R = 16$ ) son observados en la muestra simulada. De todo lo anterior se desprende que, estamos en presencia de una tabla poco ocupada porque 7 (43,7%) patrones tienen frecuencias observadas cero. Como hipótesis nula ( $H_0$ ), vamos a suponer que es un modelo con dos clases latentes. Bajo esta suposición, se calculan los estadísticos de bondad de ajuste, presentados a continuación (**Tabla 4.2**).

Estadístico	Resultado	<i>P-Valor</i> (*)
Razón de Verosimilitud	2,64	0,85
Pearson	3,85	0,70
Freeman-Tukey	2,48	0,87
Cressie-Read	3,27	0,77

(\*) Calculado haciendo caso omiso a la corrección de los grados de libertad, debido a las celdas vacías, que es lo correcto.

**TABLA 4.2.** Estadísticos de Bondad de Ajuste.

Los grados de libertad corregidos del MCL son -1 (véase apartado 3.7.6), hecho que no hace posible la utilización de la distribución de probabilidad  $\chi^2$  para evaluar la bondad de ajuste del modelo hipotético. Por esta razón, debemos utilizar un método de estimación de la distribución de probabilidad empírica de los estadísticos. Utilizaremos la propuesta metodológica, el *Método de Patrones*.

Los patrones de respuesta posibles para cuatro variables manifiestas binarias es  $R = 2^4=16$ , como se presentan seguidamente:

Patrones	$X_1$	$X_2$	$X_3$	$X_4$
$\mathbf{x}_1$	0	0	0	0
$\mathbf{x}_2$	0	0	0	1
$\mathbf{x}_3$	0	0	1	0
$\mathbf{x}_4$	0	0	1	1
$\mathbf{x}_5$	0	1	0	0
$\mathbf{x}_6$	0	1	0	1
$\mathbf{x}_7$	0	1	1	0
$\mathbf{x}_8$	0	1	1	1
$\mathbf{x}_9$	1	0	0	0
$\mathbf{x}_{10}$	1	0	0	1
$\mathbf{x}_{11}$	1	0	1	0
$\mathbf{x}_{12}$	1	0	1	1
$\mathbf{x}_{13}$	1	1	0	0
$\mathbf{x}_{14}$	1	1	0	1
$\mathbf{x}_{15}$	1	1	1	0
$\mathbf{x}_{16}$	1	1	1	1

**TABLA 4.3.** Población de patrones de respuesta teóricos para cuatro variables manifiestas binarias.

El procedimiento de estimación de la distribución de probabilidad empírica de los estadísticos, continua a través de la generación de A muestras aleatorias de patrones (por ejemplo, A=500) de igual tamaño que la muestra original ( $n=124$ ), mediante el muestreo aleatorio simple con reemplazamiento de  $(X_1, X_2, \dots, X_R)$ , donde cada patrón de respuesta tiene una probabilidad de inclusión y conocida de  $1/R$ . Por razones de espacio, solamente se presentan las dos primeras muestras aleatorias y la última.

a=1						a=2						a=500				
$X_1$	$X_2$	$X_3$	$X_4$	$f_{x_h}$		$X_1$	$X_2$	$X_3$	$X_4$	$f_{x_h}$		$X_1$	$X_2$	$X_3$	$X_4$	$f_{x_h}$
0	0	0	0	5		0	0	0	0	9		0	0	0	0	4
0	0	0	1	8		0	0	0	1	7		0	0	0	1	10
0	0	1	0	6		0	0	1	0	12		0	0	1	0	3
0	0	1	1	11		0	0	1	1	6		0	0	1	1	8
0	1	0	0	9		0	1	0	0	7		0	1	0	0	8
0	1	0	1	9		0	1	0	1	5		0	1	0	1	8
0	1	1	0	9		0	1	1	0	12	.....	0	1	1	0	7
0	1	1	1	11		0	1	1	1	7		0	1	1	1	11
1	0	0	0	8		1	0	0	0	8		1	0	0	0	9
1	0	0	1	7		1	0	0	1	8		1	0	0	1	8
1	0	1	0	8		1	0	1	0	9		1	0	1	0	18
1	0	1	1	6		1	0	1	1	8		1	0	1	1	8
1	1	0	0	8		1	1	0	0	3		1	1	0	0	8
1	1	0	1	3		1	1	0	1	8		1	1	0	1	5
1	1	1	0	10		1	1	1	0	6		1	1	1	0	3
1	1	1	1	6		1	1	1	1	9		1	1	1	1	6

**TABLA 4.4.** Muestras aleatorias utilizando el Método de Patrones.

A continuación, se procede a calcular para cada muestra aleatoria de patrones, los estadísticos de bondad de ajuste bajo la hipótesis nula, un modelo con 2 clases latentes,  $\hat{\theta}^*(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_{500}^*)$ . Estos valores se muestran en la **Tabla 4.5**.

Estadístico	Valor			
	$\hat{\theta}_1^*$	$\hat{\theta}_2^*$		$\hat{\theta}_{500}^*$
Razón de Verosimilitud	5,63	2,20		5,98
Pearson	5,44	2,19	...	6,05
Freeman-Tukey	5,77	2,21		6,00
Cressie-Read	5,49	2,19		6,01

**TABLA 4.5.** Estadísticos de bondad de ajuste para las muestras de patrones.

A partir de este conjunto de resultados de los estadístico de bondad de ajuste, se estima el promedio aritmético (fórmula 4.3) y el error estándar por medio de la raíz cuadrada de la varianza (fórmula 4.4). Además, se calcula el *p-valor* (fórmula 4.6) para cada estadístico, debiéndose ordenar los valores de menor a mayor  $\hat{\theta}_1^* \leq \hat{\theta}_2^* \leq \dots \leq \hat{\theta}_A^*$ .

Estadístico	Promedio Aritmético	Desviación Típica	P-valor
Razón de Verosimilitud	6,21	3,51	0,88
Pearson	6,00	3,32	0,70
Freeman-Tukey	6,43	3,79	0,91
Cressie-Read	6,04	3,35	0,79

**TABLA 4.6.** Promedio aritmético, desviación típica y significaciones calculadas de los estadísticos.

A un nivel de significación de  $\alpha = 0,05$ , tomamos la decisión de no rechazar el modelo de dos clases latentes propuesto como apropiado para los datos originales (**Ta**bla 4.1). *La decisión es coherente con los parámetros dados al programa LEM para generar el archivo de datos.*

## 4.2 PROBLEMA DE DENSIDAD DE LOS PATRONES

Un problema para estimar la distribución de probabilidad empírica de los estadísticos de bondad de ajuste ( $\theta$ ) con tablas múltiples poco ocupadas, consiste en que la  $\hat{F}(\theta)$  puede ser una mala aproximación de la verdadera distribución  $F(\theta)$ , debido a las celdas con frecuencias pequeñas o nulas. Utilizando el *Método de Patrones* esta complicación tiene poco efecto negativo en la estimación de  $\hat{F}(\theta)$  en comparación a los métodos Bootstrap, esto lo demostramos con los datos del apartado anterior. Debemos considerar que solamente el 56,3% de los patrones de respuesta se presentaron en la muestra original de 124 individuos; esto nos indica que la tabla múltiple es bastante “rala”, en el sentido de que 7 patrones no fueron observados. Al tomar 500 muestras con los métodos Bootstrap y Patrones, estos fueron los resultados sobre la densidad de patrones de respuesta:

Método	Densidad de Patrones		Valores	
	Media Aritmética	Desviación Típica	Mínimos	Máximos
Bootstrap No Paramétrico	0,49	0,05	0,38	0,56
Bootstrap Paramétrico	0,51	0,06	0,31	0,75
Método de Patrones	1,00	0,01	0,94	1,00

**TABLA 4.7.** Densidad de los patrones de respuesta utilizando los métodos Bootstrap y el Método de Patrones.

El porcentaje de patrones observados usando el Bootstrap No Paramétrico, como es lo esperado, no supera el 56% encontrado en la muestra original, presentando una media aritmética y desviación estándar de 49% y 5%, respectivamente. En tanto, el Bootstrap Paramétrico alcanza un valor máximo del 75% y una media aritmética de 51%. La gran diferencia la exhibe el *Método de Patrones*, tal que la media de patrones de respuesta observados en las muestras simuladas es del 100%, para una desviación típica muy cercana a cero, resultando que el porcentaje mínimo observado fue del 94%.

En conclusión, el *Método de Patrones* no se encuentra afectado por el problema de datos “faltantes”, que son aquellos patrones de respuesta que tienen frecuencias nulas en la muestra original y que producen malas aproximaciones de la distribución de probabilidad empírica de los estadísticos de bondad de ajuste.

### 4.3 CONSIDERACIONES FINALES

Basándonos en el *Método de Patrones*, la idea de fondo sigue siendo la de construir un modelo de distribución para los estadísticos de bondad de ajuste, con el propósito de hacer inferencias sobre el modelo de clases latentes apropiado para los datos de la muestra original, aunque el modo de proceder es distinto al Bootstrap Paramétrico y No Paramétrico.

Es conveniente resaltar, que el *Método de Patrones* no parte de cualquier supuesto paramétrico relativo a la distribución de los datos de la muestra ni de la distribución de probabilidad teórica  $\chi^2$  que tienen los estadísticos de bondad de ajuste.

Los métodos citados en el capítulo 3, como: el sumar una constante a las frecuencias observadas de los patrones (3.2), el utilizar las celdas suficientemente ocupadas (3.5) y la maximización de la función  $\beta$ -verosimilitud (3.4), pueden complementarse con el *Método de Patrones*. De hecho, podemos utilizar el *Método de Patrones* usando la suma de una constante a todas las celdas y maximizando la función  $\beta$ -verosimilitud.

En el próximo capítulo, se presenta el programa BootLatentClass que nos sirve para poner en práctica los métodos Bootstrap y nuestras aportaciones, el *Método de Patrones y el Bootstrap No Paramétrico Corregido*. El software que se ha creado es una aplicación realizada por completo en MATLAB®. La simplicidad de manejo se ha conseguido mediante el uso de sucesivos interfaces gráficos, muy intuitivos, a través de los cuales se controlan todos los aspectos de la aplicación.

# **CAPÍTULO 5**

---

## **PROGRAMA BOOTLATENTCLASS**

---

## 5.1 INTRODUCCIÓN

En este capítulo se describe un aporte original que consiste en la construcción del programa *BootLatentClass* implementado en MATLAB® R2010. El programa informático es el producto del desarrollo teórico de los modelos de clases latentes (MCL) para variables manifiestas binarias, de los métodos Bootstrap Clásicos, del *Método Bootstrap No Paramétrico Corregido* y del *Método de Patrones*.

Puesto que los algoritmos que conforman la aplicación son originales, este capítulo en su conjunto representa la contribución importante de nuestro trabajo al diagnóstico de MCL en tablas poco ocupadas.

La idea del programa *BootLatentClass*, está acorde a las políticas de la dirección del Departamento de Estadística de la Universidad de Salamanca, que tiene como norma que las contribuciones de las tesis doctorales, queden plasmadas, cuando procede, en un programa de cómputo que permitan poner de manifiesto su relevancia práctica.

Seguidamente, se presenta con detalle la rutina *BootLatentClass*, la cual está compuesta por 5 programas y 7 funciones, que interactúan para la simulación de datos binarios y el diagnóstico de modelos. Para ejecutar la aplicación, resulta suficiente copiar todos sus componentes en la ruta de trabajo de MATLAB® y digitar “*BootLatentClass*” en la ventana de comandos.

Para que el programa resultase más sencillo al usuario final, se ha creado un entorno basado en ventanas gráficas, que facilitan la utilización del software, sin tener que hacer uso de la línea de comandos (como se puede ver en la **Figura 5.1**). Para ello se ha usado la herramienta GUIDE que incorpora MATLAB® en sus últimas versiones, que permite una interacción, a modo de diálogo, entre el ordenador y el usuario.

Al final del capítulo, se hace una comparación de los resultados obtenidos con los programas: WINMIRA (von DAVIER, 2001), LEM (VERMUNT, 1997b), LATENT GOLD (MAGIDSON & VERMUNT, 2005) y BOOTLATENTCLASS,



respecto a los estadísticos de bondad de ajuste y los parámetros del modelo de clases latentes.

El programa **WINMIRA** tiene implementados el Análisis de clases latentes, modelos de Rasch (o Teoría de Respuesta a los Ítems), modelos Mixtos Rasch y modelos híbridos. Particularmente, el *Modelo de Rasch*, permite la medición conjunta de personas e ítems en una misma dimensión o constructo. Significa que los parámetros de las personas y de los ítems se expresan en las mismas unidades. En tanto, **LEM** posee aplicaciones para modelos usando frecuencias de celdas, modelos de regresión, Modelos Path, modelos de clases latentes, modelos Path con variables latentes, análisis histórico de eventos y clases latentes extendido. El *Análisis Path*, consiste en una técnica analítica de experimentación y exploración dentro de la investigación sociológica para la manipulación de variables específicas con objeto de determinar e inferir relaciones causales entre esas variables. Finalmente, el programa **LATENT GOLD** puede utilizarse para realizar diversos tipos de análisis con modelos de clases latentes. Un caso en particular, es el modelo de regresión de clases latentes, que es usado para pronosticar una variable dependiente en función de un conjunto de variables manifiestas y covariables<sup>1</sup>.

## 5.2 ENTRADA AL PROGRAMA

El programa empieza con un menú en la ventana principal, esta ventana es representada mediante una interfaz gráfica que se muestra en la **Figura 5.1**. La rutina BootLatentClass está compuesta por un conjunto de 6 módulos:

- Análisis de Clases latentes Clásico con maximización por Log-Verosimilitud y Beta-Verosimilitud. Además de la opción de ajustar el modelos solo con datos de las celdas ocupadas.

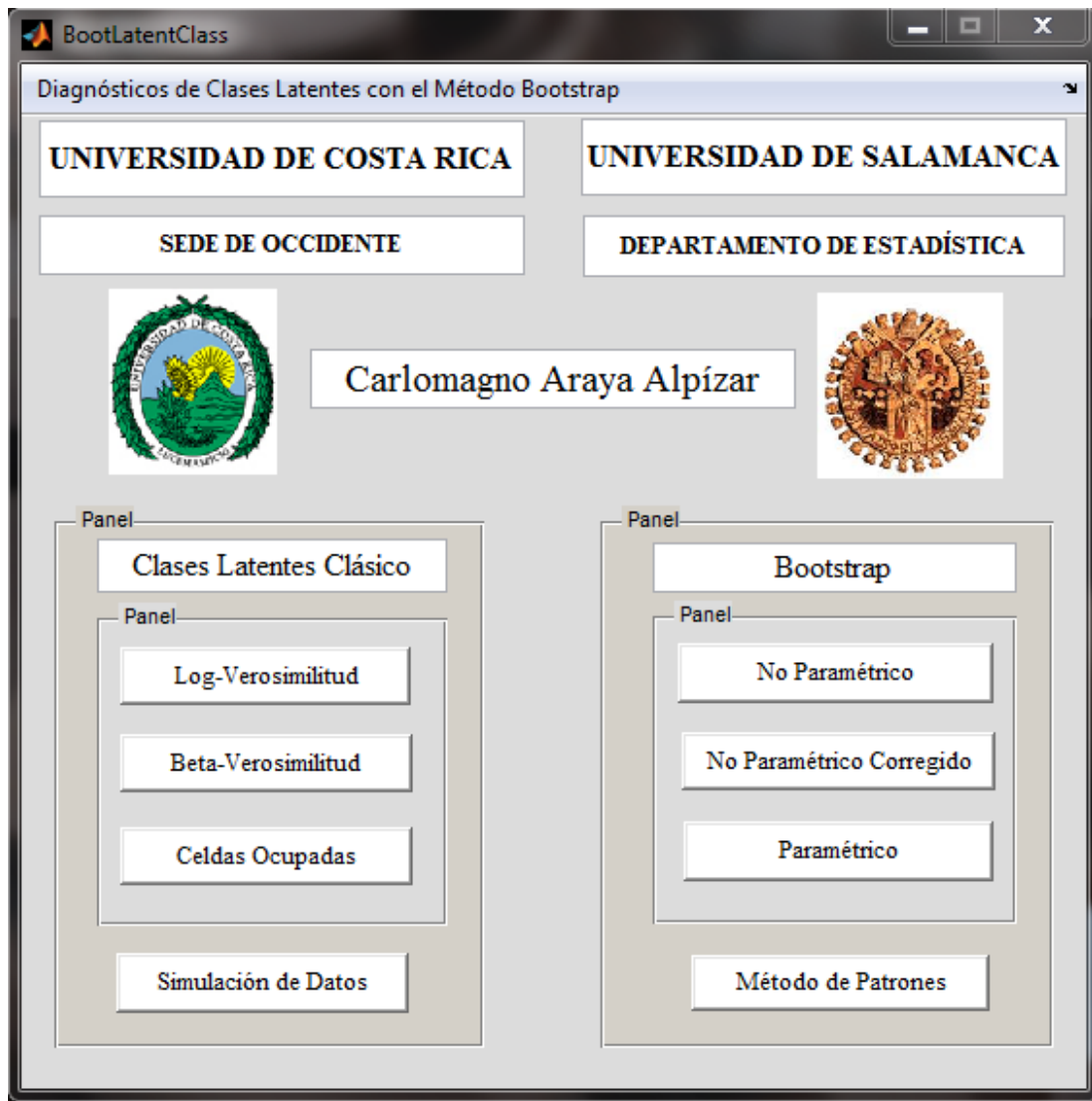
---

**Order Latent GOLD® 4.5**

Latent GOLD 4.5 is a powerful [latent class and finite mixture](#) program. Latent GOLD contains separate modules for estimating three different model structures – LC Cluster models, DFactor models, and LC Regression models – which are useful in somewhat different application areas. Latent GOLD 4.5 comes in either a Basic or Advanced version. To learn about Latent GOLD:

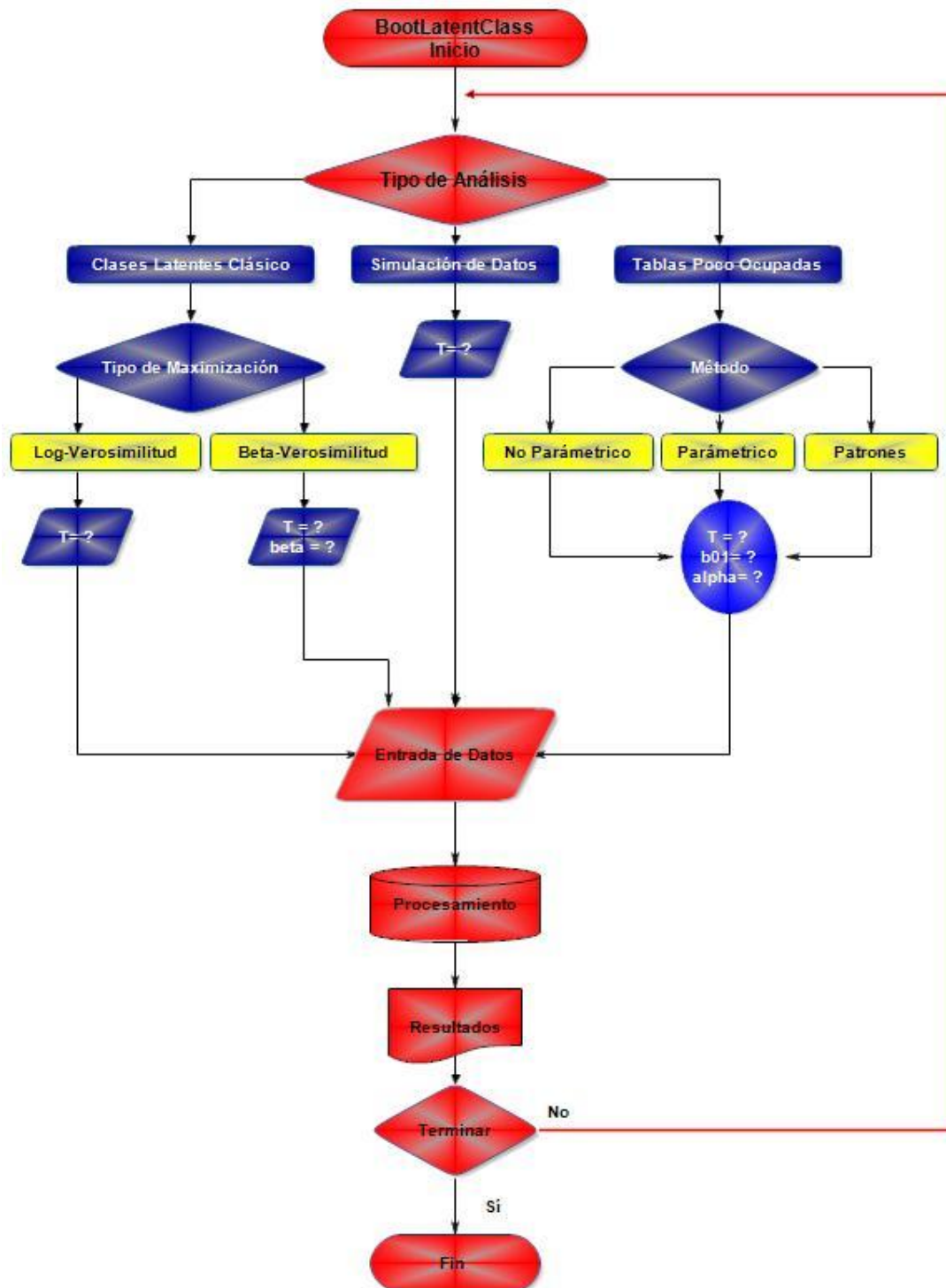
- Diagnóstico de modelos de clases latentes en tablas poco ocupadas utilizando los métodos Bootstrap: No Paramétrico, No Paramétrico Corregido y Paramétrico.
- El Método de Patrones, que es una propuesta del presente trabajo.
- Simulación de datos bajo un modelo de clases latentes.

Tras escoger algún módulo, aparece un cuadro de diálogo, indicando que estamos listos para empezar a trabajar con *BootLatentClass*. A continuación, describimos brevemente los módulos en cuestión, los detalles que tienen interés desde el punto de vista computacional y teórico, para aquellas personas que podrían tener interés en implementar todo el programa *BootLatentClass* o un módulo en concreto.



**Figura 5.1.** Interfaz gráfica del programa *BootLatentClass*.

A continuación, en la **Figura 5.2** presentamos el diagrama de flujo con la secuencia de pasos para ejecutar el programa; está compuesto por un conjunto de símbolos y flechas dirigidas, que muestran de una manera gráfica la solución de un problema relacionado con el ajuste de un modelo de clases latentes para variables manifiestas binarias. Los parámetros de entrada al programa son:  $T$ = número de clases latentes,  $\alpha$ = nivel de significación,  $bo1$ = número de remuestras,  $w$ = mínima frecuencia por celda ocupada y  $\beta$ = factor de calibración entre 0 y 1 utilizado en la función de maximización  $\beta$ -verosimilitud.



**Figura 5.2.** Diagrama de flujo de *BootLatentClass*.

### 5.3 ANÁLISIS DE CLASES LATENTES

En relación al módulo de Análisis de Clases Latentes, es necesaria la introducción del conjunto de datos de la muestra. La matriz de datos originales ( $\mathbb{X}$ ) debe estar formada por el conjunto de respuestas a las variables binarias, o sea, ceros y unos (no contiene títulos para las variables)<sup>2</sup>. La matriz  $\mathbb{X}$  puede representarse como,

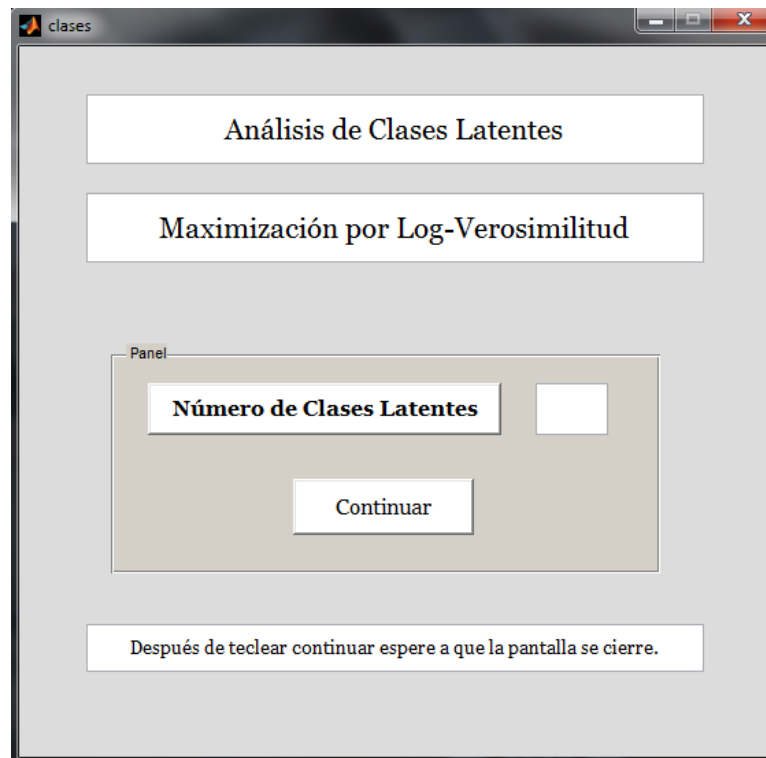
$$\mathbb{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

El fichero que contiene la matriz de datos originales debe tener el nombre: `archivo.dat`, donde “archivo” es el identificador de la base de datos suministrado por el usuario. Si, por ejemplo, el identificador de la base de datos es “datos”, la correspondiente al archivo datos originales debe estar disponible en la ruta de trabajo de MATLAB®, con el nombre “datos.dat”.

Como datos de entrada al programa, únicamente debemos introducir el número de clases latentes usando maximización por Log-Verosimilitud (**Figura 5.3**), el parámetro de calibración  $0 \leq \beta \leq 1$  para la maximización por Beta-Verosimilitud (**Figura 5.4**) y el valor mínimo de la celda (**Figura 5.5**). Los resultados son mostrados en la pantalla y grabados en el disco duro, en un archivo con el nombre de “Clases Latentes Log.txt”, “Clases Latentes Beta.txt” y “Clases Latentes Ocupadas.txt”, respectivamente. Este archivo contiene la siguiente información: porcentajes de patrones de respuestas observados, estadísticos de bondad de ajuste, grados de libertad del modelo, probabilidades totales de las clases y probabilidades condicionales.

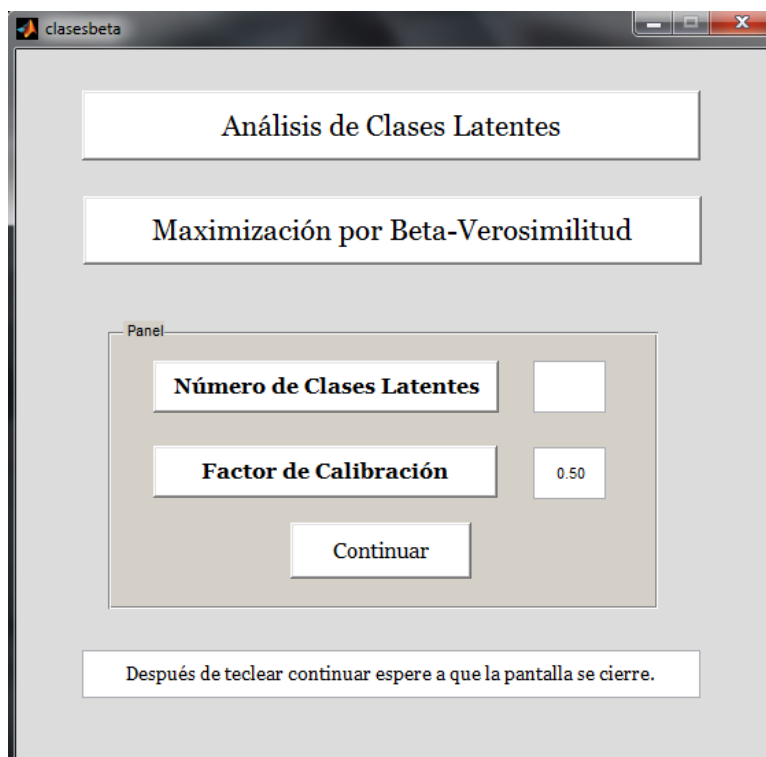
---

<sup>2</sup> No se permiten valores perdidos o faltantes. En el presente contexto, siempre que se hable de datos faltantes, se estará haciendo referencia a un patrón de respuesta.



The screenshot shows a window titled 'clases' with a light gray background. At the top, there is a button labeled 'Análisis de Clases Latentes'. Below it is another button labeled 'Maximización por Log-Verosimilitud'. In the center, there is a smaller panel titled 'Panel' with a light beige background. Inside this panel, there is a label 'Número de Clases Latentes' next to a white text input field. Below the input field is a button labeled 'Continuar'. At the bottom of the main window, there is a text box containing the instruction: 'Después de teclear continuar espere a que la pantalla se cierre.'

**Figura 5.3.** Entrada de datos a la rutina de clases latentes maximización por log-verosimilitud.



The screenshot shows a window titled 'clasesbeta' with a light gray background. At the top, there is a button labeled 'Análisis de Clases Latentes'. Below it is another button labeled 'Maximización por Beta-Verosimilitud'. In the center, there is a smaller panel titled 'Panel' with a light beige background. Inside this panel, there are two labels: 'Número de Clases Latentes' next to a white text input field, and 'Factor de Calibración' next to a white text input field containing the value '0.50'. Below these input fields is a button labeled 'Continuar'. At the bottom of the main window, there is a text box containing the instruction: 'Después de teclear continuar espere a que la pantalla se cierre.'

**Figura 5.4.** Entrada de datos a la rutina de clases latentes maximización por beta-verosimilitud.

**Figura 5.5.** Entrada de datos a la rutina de clases latentes utilizando celdas con ocupadas.

La sintaxis y la secuencia de pasos del programa que hace posible calcular los parámetros del modelo de clases latentes es la siguiente:

Pasos	Código del programa: Log-Verosimilitud
Entrada de datos	<pre> clc; disp('Modelos de Clases Latentes') [filename, pathname]=uigetfile('*.dat', 'Seleccione un archivo .dat'); if ~isequal(filename,0)   ~isequal(pathname,0) file = fullfile(pathname,filename); x= textread(file); end [tm,p]=size(x);R=2^p; </pre>
Introducir número de clases	<pre> T=input('Número de clases latentes: '); </pre>
Cálculo de los patrones de respuesta	<pre> [M,np] = patterns(p); </pre>

Construcción tablas de frecuencia	<code>[tabla] = frecuen(x,M);</code> <code>x=tabla;</code>
Parámetros iniciales	<code>[c00,p00] = inicio(p,T);</code>
Estimación de los parámetros	<code>[bondad,p0,c0,sparse]=cmaa(x,T,c00,p00);</code>
Salida de resultados	<code>diary('Clases latentes Log.txt');</code> <code>disp('Análisis de clases latentes: maximización log- verosimilitud');</code> <code>disp(' ');</code> <code>original;</code>

Pasos	Código del programa: Beta-Verosimilitud
Entrada de datos	<code>clc;</code> <code>disp('Modelos de Clases Latentes')</code> <code>[filename, pathname]=uigetfile('*.dat', 'Seleccione un archivo .dat');</code> <code>if ~isequal(filename,0)   ~isequal(pathname,0)</code> <code>file = fullfile(pathname,filename);</code> <code>x= textread(file);</code> <code>end</code> <code>[tm,p]=size(x);</code> <code>R=2^p;</code>
Introducir número de clases	<code>T=input('Número de clases: ');</code>
Introducir parámetro de calibración	<code>beta=input('Parámetro de calibración: ');</code>
Cálculo de los patrones de respuesta	<code>[M,np] = patterns(p);</code>
Construcción tablas de frecuencia	<code>[tabla] = frecuen(x,M);</code> <code>x=tabla;</code>
Parámetros iniciales	<code>[c00,p00] = inicio(p,T);</code>



Estimación de los parámetros	<code>[bondad,p0,c0,sparse]=cmaabeta(x,T,c00,p00,beta);</code>
Salida de resultados	<code>diary('Clases Latentes Beta.txt'); disp('Análisis de Clases latentes: maximización beta-verosimilitud'); disp(' '); original; clear;</code>

Pasos	Código del programa: Celdas Ocupadas
Entrada de datos	<code>clc; disp('Modelos de Clases Latentes') [filename, pathname]=uigetfile('*.dat', 'Seleccione un archivo .dat'); if ~isequal(filename,0)   ~isequal(pathname,0) file = fullfile(pathname,filename); x= textread(file); end [tm,p]=size(x);</code>
Introducir número de clases	<code>T=input('Number of latent classes: '); w=input('Valor minimo de las frecuencias de los patrones: ');</code>
Cálculo de los patrones de respuesta	<code>[M,np] = patterns(p);</code>
Construcción tablas de frecuencia	<code>[tabla] = frecuen(x,M);</code>
Eliminación de patrones con frecuencias menores a w	<code>[p2,pp]=size(tabla); f0=tabla(:,pp); s1 = sum(tabla&gt;w); s2=(2^p)-s1/(p+1); k=find(f0&lt;w); tabla(k,:)=[]; f1=tabla(:,pp); n1=sum(f1);</code>
Grados de libertad	<code>gl1=(p2-1)-(T*p+(T-1));</code>

Grados de libertad corregidos	<code>gl2=gl1-s2;</code>
Parámetros iniciales	<code>x=tabla;</code> <code>[c00,p00] = inicio(p,T);</code>
Estimación de los parámetros	<code>[bondad,p0,c0]=cmaaajmlog(x,T,c00,p00);</code>
Salida de resultados	<pre> diary('Clases Latentes Ocupadas.txt'); format bank; diary on; disp('Análisis de clases latentes: utilizar celdas suficientemente ocupadas'); disp(' '); disp(['Número mínimo de las frecuencias por celda: ',num2str(w)]); disp(['Tamaño de la muestra original: ',num2str(tm)]); disp(['Tamaño de la muestra después de aplicar el mínimo por celda: ',num2str(n1)]); disp(['Número de Clases Latentes: ',num2str(T)]); if (gl2&gt;0)     for i=1:4         eba=bondad(i);         sig(i)=chi2df(eba,gl);     end     else         sig=[];         disp('Se quedó sin grados de libertad, no puede utilizar la distribución');         disp('de probabilidad Chi-Cuadrado para evaluar la bondad de ajuste del MCL');     end disp(' '); disp('Estadísticas de Bondad de Ajuste'), disp('          G2          Q2          FT          CR'), disp(bondad); disp(sig); disp(['Grados de libertad: ',num2str(gl1)]); disp(['Grados de libertad corregidos: ',num2str(gl2)]); </pre>

```

disp(' ');
disp('Probabilidades totales de las clases');
disp(c0');
disp('Probabilidades condicionadas (variables x
clases)');
disp(p0);
diary off;
clear;

```

Los programas “Log-Verosimilitud”, “Beta-Verosimilitud” y “Celdas Ocupadas” utilizan las funciones: patterns, frecuen, inicio, cmaa, cmaabeta y cmaajmlog, además del programa “original”. La función “patterns” construye los patrones de respuesta asociados al número de variables manifiestas binarias, el código es el siguiente:

**Código de la función: patterns**

```

function [M,np] = patterns(p)
    np=2^p; c=zeros(1,p);
    estado=-ones(1,p); global M;
    M=ones(1,p);
    for j=1:p
        M(1,j)=0;
    end
    for i=2:np,
        c=c+1;
        for j=1:p,
            if c(j)==2^(p-j),
                estado(j)=-estado(j);
                c(j)=0;
            end
        end
        for j=1:p,
            if estado(j)==-1,
                M(i,j)=0;
            else,
                M(i,j)=1;
            end
        end
    end
end

```

La distribución de frecuencias de los patrones de respuesta los construye la función “frecuen”. La sintaxis de la función es la siguiente:

**Código de la función: frecuen**

```
function [tabla] = frecuen(x,M)
    [n,p]=size(x);
    r=zeros(n,1);
    j=0; xx=[];
    for i=1:n
        if r(i)==0,
            j=j+1;
            e(j)=0;
            xx=[xx;x(i,:)];
            for k=i:n,
                if xx(j,:)==x(k,:);
                    e(j)=e(j)+1;
                    r(k)=1;
                end
            end
        end
    end
    [n2,p2]=size(xx);
    [n,p]=size(M);
    pas=[M(:,1:p),zeros(n,1)];
    for i=1:n
        for j=1:n2
            if pas(i,1:p)==xx(j,:)
                pas(i,:)=[pas(i,1:p),e(j)];
            end
        end
    end
    tabla=pas;
    clear pas
end
```

Los parámetros iniciales del modelo de clases se calculan usando la función “inicio” por medio del siguiente código:

Código de la función: inicio
<pre>function [c00,p00] = inicio(p,T)     p00=rand(p,T);     s=sum(p00')';     p00=diag(1./s)*p00;     c00 = rand(1,T);     c00 = c00./ sum(c00); end</pre>

El problema de calcular los parámetros del modelo y calcular los estadísticos de bondad de ajuste, se realiza utilizando las funciones “cmaa”, “cmaabeta” y “cmaajmlog”.

Pasos	Código de la función: cmaa
Definición de constantes	<pre>function [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00) c0=c00;p0=p00; [p2,pp]=size(x); p=pp-1; a=x(:,1:p); f0=x(:,pp);</pre>
Corrección del problema de las tablas poco ocupadas	<pre>s1 = sum(x&gt;0); s2=(2^p)-s1((p+1)); nc=1-s2/(2^p); if (nc&lt;1)     f0=x(:,pp)+(1/(2^p)); end</pre>
Grados de libertad	<pre>gl1=((2^p)-1)-(T*p+(T-1));</pre>
Grados de libertad corregidos	<pre>gl2=gl1-s2;</pre>
Etapas de esperanza	<pre>n=sum(f0); iter=0; error=1; while error &gt; 0.0000001     iter=iter+1;</pre>

	<pre> q=ones(p2,T); for j=1:T;     for k=1:p2;         for i=1:p             if a(k,i)==1                 q(k,j)=q(k,j)*p0(i,j);             else                 q(k,j)=q(k,j)*(1-p0(i,j));             end         end     end end q2=zeros(p2,T); for i=1:T     q2(:,i)=q(:,i)*c0(i); end </pre>
Etapa de maximización	<pre> q3=sum(q2'); q3=q3'; for i=1:p2     for j=1:T         h2(i,j)=(1./q3(i))*q2(i,j)*f0(i);     end end </pre>
Cálculo de los parámetros del modelo	<pre> t0=0; t0=sum(h2); c0=t0/n; q4=zeros(p,T); for i=1:T;     for j=1:p         for k=1:p2;             if a(k,j)==1                 q4(j,i)=q4(j,i)+h2(k,i);             end         end     end end p0=zeros(p,T); for i=1:T     for j=1:p         p0(j,i)=q4(j,i)/t0(i);     end end q5=log(q3); </pre>

	<pre> L2=q5'*f0; if iter&gt;1     error=abs(L-L2); end h=h2; L=L2; end </pre>
Cálculo de los estadísticos de bondad de ajuste	<pre> f1=n*q3; f1(find(f1==0))=1; g2 = f0./f1; g2(find(g2==0))=1; g2 = 2 * sum(f0.*log(g2)); pearson=sum(((f0-f1).^2)./f1); ft1=f1.^(1/2); ft2=f0.^(1/2); ft3=ft1-ft2; ftukey=4*sum(ft3.^2); cread=(2/((2/3)*((2/3)+1)))*sum(f0.*(((f0./f1).^(2/3))-1)); </pre>
Vector de estadísticos	<pre> bondad=[g2 pearson ftukey cread nc]; sparse=[s2 nc gl1 gl2]; </pre>
Clasificación de los individuos	<pre> [Y,I]=max(h'); I=I'; </pre>
Final de la función	<pre> end </pre>

Pasos	Código de la función: cmaabeta
Definición de constantes	<pre> function [bondad,p0,c0,sparse]=cmaabeta(x,T,c00,p00,beta) c0=c00;p0=p00; [p2,pp]=size(x); p=pp-1; a=x(:,1:p); f0=x(:,pp); n=sum(f0); </pre>
Corrección del problema de las tablas poco ocupadas	<pre> s1 = sum(x&gt;0); s2=(2^p)-s1((p+1)); nc=1-s2/(2^p); if (nc&lt;1) </pre>

	<pre> f0=x(:,pp)+(1/(2^p)); end </pre>
Grados de libertad	<pre> gl1=((2^p)-1)-(T*p+(T-1)); </pre>
Grados de libertad corregidos	<pre> gl2=gl1-s2; </pre>
Etapa de esperanza	<pre> iter=0; error=1; while error &gt; 0.0000001     iter=iter+1;     q=ones(p2,T);     for j=1:T;         for k=1:p2;             for i=1:p                 if a(k,i)==1                     q(k,j)=q(k,j)*p0(i,j);                 else                     q(k,j)=q(k,j)*(1-p0(i,j));                 end             end         end     end     q2=zeros(p2,T);     for i=1:T         q2(:,i)=q(:,i)*c0(i);     end end </pre>
Etapa de maximización	<pre> q3=sum(q2'); q3=q3'; for i=1:p2     for j=1:T         h2(i,j)=(1./q3(i))*q2(i,j)*f0(i);     end end </pre>
Cálculo de los parámetros del modelo	<pre> t0=0; t0=sum(h2); c0=t0/n; q4=zeros(p,T); for i=1:T;     for j=1:p         for k=1:p2;             if a(k,j)==1 </pre>



	<pre>                                 q4(j,i)=q4(j,i)+h2(k,i);                                 end                                 end                                 end                                 end                                 p0=zeros(p,T);                                 for i=1:T                                     for j=1:p   p0(j,i)=q4(j,i)/t0(i);                                     end                                 end                                 q5=(q3).^(beta);                                 L2=q5'*(f0/(n*beta));                                 if iter&gt;1                                     error=abs(L-L2);                                 end                                 h=h2;                                 L=L2;                                 end </pre>
Cálculo de los estadísticos de bondad de ajuste	<pre> f1=n*q3; f1(find(f1==0))=1; g2 = f0./f1; g2(find(g2==0))=1; g2 = 2 * sum(f0.*log(g2)); pearson=sum(((f0-f1).^2)./f1); ft1=f1.^(1/2); ft2=f0.^(1/2); ft3=ft1-ft2; ftukey=4*sum(ft3.^2); cread=(2/((2/3)*((2/3)+1)))*sum(f0.*(((f0./f1).^(2/3))-1)); </pre>
Vector de estadísticos	<pre> bondad=[g2 pearson ftukey cread nc]; sparse=[s2 nc gl1 gl2]; </pre>
Clasificación de los individuos	<pre> [Y,I]=max(h'); I=I'; </pre>
Final de la función	<pre> end </pre>

Pasos	Código de la función: cmaajmlog
Definición de constantes	<pre>function [bondad,p0,c0]=cmaajmlog(x,T,c00,p00) c0=c00;p0=p00; [p2,pp]=size(x); p=pp-1; a=x(:,1:p); f0=x(:,pp); n=sum(f0);</pre>
Etapa de esperanza	<pre>iter=0; error=1; while error &gt; 0.0000001     iter=iter+1;     q=ones(p2,T);     for j=1:T;         for k=1:p2;             for i=1:p                 if a(k,i)==1                     q(k,j)=q(k,j)*p0(i,j);                 else                     q(k,j)=q(k,j)*(1-p0(i,j));                 end             end         end     end     q2=zeros(p2,T);     for i=1:T         q2(:,i)=q(:,i)*c0(i);     end</pre>
Etapa de maximización	<pre>q3=sum(q2');    q3=q3'; for i=1:p2     for j=1:T         h2(i,j)=(1./q3(i))*q2(i,j)*f0(i);     end end</pre>
Cálculo de los parámetros del modelo	<pre>t0=0; t0=sum(h2); c0=t0/n; q4=zeros(p,T); for i=1:T;     for j=1:p         for k=1:p2;             if a(k,j)==1                 q4(j,i)=q4(j,i)+h2(k,i);</pre>

	<pre> end end end end p0=zeros(p,T); for i=1:T     for j=1:p         p0(j,i)=q4(j,i)/t0(i);     end end q5=log(q3);    L2=q5'*f0; if iter&gt;1     error=abs(L-L2); end h=h2;L=L2; end </pre>
Cálculo de los estadísticos de bondad de ajuste	<pre> f1=n*q3; f1(find(f1==0))=1; g2 = f0./f1; g2(find(g2==0))=1; g2 = 2 * sum(f0.*log(g2)); pearson=sum(((f0-f1).^2)./f1); ft1=f1.^(1/2); ft2=f0.^(1/2); ft3=ft1-ft2; ftukey=4*sum(ft3.^2); cread=(2/((2/3)*((2/3)+1)))*sum(f0.*(((f0./f1).^(2/3))-1)); </pre>
Vector de estadísticos	<pre> bondad=[]; bondad=[g2 pearson ftukey cread nc]; </pre>
Clasificación de los individuos	<pre> [Y,I]=max(h'); I=I'; </pre>
Final de la función	<pre> end </pre>

Finalmente, el programa “Original” se encarga de imprimir los resultados.

Pasos	Código del programa: Original
Tamaño de muestra y advertencia sobre la densidad de los datos	<pre> format bank; disp(['Tamaño de la muestra: ',num2str(tm)]); disp(' '); disp(['Patrones de respuesta con frecuencias cero: ',num2str(sparse(1))]); if (sparse(2)&lt;1)     disp('Cuidado, se presenta poca densidad de los datos');     disp(['Porcentajes de patrones de respuestas observados: ',num2str(sparse(2))]); end </pre>
Grados de libertad brutos y corregidos	<pre> disp(' '); disp(['Grados de libertad: ',num2str(sparse(3))]); gl2=sparse(4); disp(['Grados de libertad corregidos: ',num2str(gl2)]); </pre>
Cálculo del nivel de significación observado	<pre> if (gl2&gt;0)     for i=1:4         eba=bondad(i);         sig2(i)=chi2df(eba,gl2);     end else     sig2=[];     disp('Se quedó sin grados de libertad, no puede utilizar la distribución');     disp('de probabilidad Chi-Cuadrado para evaluar la bondad de ajuste del MCL'); end disp(' '); </pre>
Presentación de los resultados de los estadísticos de bondad de ajuste	<pre> disp('Estadísticas de Bondad de Ajuste de la muestra original'), disp('      G2      Q2      FT      CR'), disp(bondad); disp(sig2); disp(' '); disp('Probabilidades totales de las clases'); disp(c0); disp('Probabilidades condicionadas (variables x </pre>

```

clases)');
disp(p0);
diary off;

```

Este programa utiliza la función “chi2df”, que calcula el nivel de significación observado (p-valor) de los estadísticos de bondad de ajuste, suponiendo que tienen una distribución de probabilidad Chi-Cuadrado. En tablas poco ocupadas estos niveles de significación, no son válidos y por esa razón son necesarios los métodos Bootstrap y Patrones.

#### Código de la función: chi2df

```

function pv = chi2df(eba,gl)
    if nargin < 2,
        error('Usage: pv = chi2cdf(eba,gl)');
    end
    eba(eba < 0) = 0 ;
    pv = 1-(gammainc(eba/2, gl/2));
end

```

## 5.4 BOOTSTRAP NO PARAMÉTRICO

La forma del fichero de entrada que contiene la matriz de datos, para el diagnóstico de modelos utilizando el Bootstrap No Paramétrico, es similar a la descrita para el caso de análisis de clases latentes del apartado 5.3.

Después de haber pulsado el botón de la opción “No Paramétrico”, debe introducirse como datos de entrada la siguiente información: número de clases latentes, número de remuestras Bootstrap y el nivel de significación (**Figura 5.5**). Tras haber introducido estos datos, el programa *BootLantetClass* inicia el proceso de cálculo.

Los resultados son mostrados en la pantalla y grabados en el disco duro en un archivo llamado “NoParametrico.txt”. El fichero contiene en la primera parte, los resultados obtenidos de la muestra original, o sea, los porcentajes de patrones de respuestas observados, los estadísticos de bondad de ajuste (EBA), los grados de

libertad del modelo, las probabilidades totales de las clases y las probabilidades condicionales.

En la segunda parte de los resultados se visualiza, la decisión sobre el modelo propuesto como el apropiado según los EBA, donde  $H_0$  significa que no debemos rechazar el modelo y  $H_1$  implica lo contrario, o sea, se debe rechazar. Además, se muestra para los EBA: el promedio aritmético y la desviación típica respecto al estadístico calculado con la muestra original. En tanto, para los patrones de respuestas se da el promedio aritmético y la desviación típica. Al final son presentados los valores de los EBA ordenados para cada remuestra Bootstrap. Además en un archivo en Excel, “Noparametrico.xls”, se guardan los percentiles de los EBA.

**Figura 5.6.** Entrada de datos a la rutina del Bootstrap No Paramétrico.

El código del programa que realiza el ajuste del modelo con Bootstrap No Paramétrico se presenta a continuación:

Pasos	Código del programa: Bootstrap No Paramétrico
Entrada de datos	<pre> clc; disp('Bootstrap No Paramétrico ') [filename, pathname]=uigetfile('*.dat', 'Seleccione un archivo .dat'); if ~isequal(filename,0)   ~isequal(pathname,0)     file = fullfile(pathname,filename);     x= textread(file); end x0=x; [tm,p]=size(x); </pre>
Entrada del número de clases	<pre> T=input('Número de clases latentes: '); </pre>
Entrada del número de remuestras	<pre> bol=input('Número de Remuestras Bootstrap: '); </pre>
Entrada de nivel de significación	<pre> alpha=input('Nivel de significación: '); </pre>
Cálculo de los patrones de respuesta	<pre> [M,np] = patterns(p); </pre>
Cálculo de las frecuencias de patrones	<pre> [tabla] = frecuen(x,M); x=tabla; </pre>
Cálculo de los parámetros iniciales	<pre> [c00,p00] = inicio(p,T); </pre>
Cálculo de parámetros del modelo	<pre> [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00); st0=bondad;c00=c0;p00=p0; </pre>
Salida de resultados de la muestra original	<pre> diary('Bootstrap Parametrico.txt'); disp('Diagnóstico de modelos: Bootstrap Paramétrico'); disp(' '); original; diary off; diary off; </pre>

Cálculo de los parámetros y estadístico de bondad de ajuste para cada remuestra Bootstrap	<pre> [R] = multisize(c0,tm); p000=p00'; stat=[]; %Matriz de resultados for b=1:bol     x=[]; %Matriz de datos bootstrap     for k=1:T         h=R(k);         [B]= clase(h,p000,k);         x=[x;B]; %Construye la matriz total de datos     end     [tabla] = frecuen(x,M);     x=tabla;     [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00);     stat=[stat;bondad]; disp(b); end </pre>
Decisión sobre el modelo propuesto	<pre> decision; </pre>
Presentación de los resultados	<pre> resultados; clear; </pre>

Este programa utiliza las funciones: patterns, frecuen, inicio y cmaa. Además, los programas “original”, “decision” y “resultados”. Todas las funciones fueron presentadas en el apartado anterior (5.3). Hemos descrito el programa “original”, faltarían por describir los programas “decision” y “resultados”.

Pasos	Código del programa: decision
Cálculo de los percentiles de los estadísticos	<pre> [stl]=sort(stat); c1=1:1: bol; c1=c1'; A=c1./bol; C=[A stl]; C( :,6)=[]; in=bol*0.01; D=C(in:in:bol,: )'; k1=bol*0.25;k2=bol*0.75; E1=C(k1:k1:k2,: )'; k1=bol*0.90;k2=bol*0.05;k3=bol*0.95; E2=C(k1:k2:k3,: )'; k1=bol*0.96;k2=bol*0.01;k3=bol*0.99; E3=C(k1:k2:k3,: )'; E=[E1 E2 E3]; </pre>



Contraste de la hipótesis del modelo	<pre> H=''; for i=1:4     eba=st1(:,i);     punto=st0(i);     if (punto &gt;= eba(cr)),         HH='H1 '; H=[H HH];     else         HH='H0 '; H=[H HH];     end; end </pre>
Cálculo del nivel de significación observado	<pre> pv=[1 1 1 1]; for i=1:4     eba=st1(:,i); punto=st0(i);     for j=1:bol         if (punto &gt;= eba(j)),             pv(i)=1-((j+1)/(bol+1)); j=bol;         end     end end </pre>
Cálculo de la desviación típica respecto al estadístico original	<pre> dt=[]; for i=1:5     eba=stat(:,i); k0=st0(i);     for j=1:bol         d=(eba(i)-k0)^2/k0;     end     dt(i)=sqrt(d); end; </pre>

Pasos	Código del programa: resultados
Presentación del nivel de significación calculado	<pre> diary on; disp('Significación calculada de los estadísticos - p-valor'), disp('      G2      Q2      FT      CR      Densidad'), disp(pv); </pre>
Decisión sobre el modelo hipotético	<pre> disp('Decisión sobre el modelo: H0: No rechazar H1: Rechazar'), disp('G2 Q2 FT CR'), disp(H); disp(' '); </pre>

Estadísticas descriptivas de los estadísticos de bondad de ajuste	<pre> disp('Estadísticas Descriptivas'), disp('Promedio Aritmético'), disp('      G2      Q2      FT      CR      Densidad'), disp(mean(stat)); disp('Desviación Estándar'), disp('      G2      Q2      FT      CR      Densidad'), disp(std(stat,0,1)); disp('Desviación Estándar respecto al estadístico original'), disp('      G2      Q2      FT      CR      Densidad'), disp(dt); </pre>
Estadísticas de bondad de ajuste ordenadas	<pre> disp('Estadístico de bondad de ajuste Ordenados'), disp('      G2      Q2      FT      CR      Densidad'), disp(stl); diary off; </pre>

## 5.5 BOOTSTRAP NO PARAMÉTRICO CORREGIDO

La forma del fichero de entrada que contiene la matriz de datos, para el diagnóstico de modelos utilizando el Bootstrap No Paramétrico Corregido, es igual a la detallada para el caso de análisis de clases latentes del apartado 5.3.

Después de haber pulsado el botón de la opción “No Paramétrico Corregido”, debe introducirse como datos de entrada la siguiente información: número de clases latentes, número de remuestras Bootstrap y el nivel de significación (**Figura 5.7**). Tras haber introducido estos datos, el programa *BootLatentClass* inicia el proceso de cálculo.

Los resultados son mostrados en la pantalla y grabados en el disco duro en un archivo llamado “BNPCorregido.txt”. *El fichero* contiene en la primera parte, los resultados obtenidos de la muestra original, o sea, los porcentajes de patrones de respuestas observados, los estadísticos de bondad de ajuste (EBA), los grados de libertad del modelo, las probabilidades totales de las clases y las probabilidades condicionales.

bnpcorregido

Diagnóstico de Modelos de Clases Latentes

Bootstrap No Paramétrico Corregido

Panel

Número de clases latentes	
Número de remuestras	300
Nivel de significación	0.05

Continuar

Después de teclar continuar espere a que la pantalla se cierre.

**Figura 5.7.** Entrada de datos a la rutina del Bootstrap No Paramétrico Corregido.

En la segunda parte de los resultados se visualiza, la decisión sobre el modelo propuesto como el apropiado según los EBA, donde  $H_0$  significa que no debemos rechazar el modelo y  $H_1$  implica lo contrario, o sea, se debe rechazar. Además, se muestra para los EBA: el promedio aritmético y la desviación típica respecto al estadístico calculado con la muestra original. En tanto, para los patrones de respuestas se da el promedio aritmético y la desviación típica. Al final son presentados los valores de los EBA ordenados para cada remuestra Bootstrap. Además en un archivo en Excel, “BNPCorregido.xls”, se guardan los percentiles de los EBA.

El código del programa que realiza el ajuste del modelo con Bootstrap No Paramétrico Corregido se presenta a continuación:

Pasos	Código del programa: Bootstrap No Paramétrico Corregido
Entrada de datos	<pre> clc; disp('Bootstrap No Paramétrico Corregido') [filename, pathname]=uigetfile('*.dat', 'Seleccione un archivo .dat'); if ~isequal(filename,0)   ~isequal(pathname,0)     file = fullfile(pathname,filename); x= textread(file); end [tm,p]=size(x);R=2^p; </pre>
Número de clases	<pre> T=input('Número de clases latentes: '); </pre>
Número de remuestras	<pre> bol=input('Número de Remuestras Bootstrap: '); </pre>
Nivel de significación	<pre> alpha=input('Nivel de significación: '); </pre>
Cálculo de los patrones de respuesta	<pre> [M,np] = patterns(p); </pre>
Cálculo de las frecuencias de patrones	<pre> x0=x; [tabla] = frecuen(x,M); x=tabla; </pre>
Parámetros iniciales	<pre> [c00,p00] = inicio(p,T); </pre>
Parámetros del modelo	<pre> [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00); st0=bondad; c00=c0; p00=p0; </pre>
Salida de resultados de la muestra original	<pre> diary('Bootstrap No Parametrico Corregido.txt'); disp('Diagnóstico de modelos: Bootstrap No Paramétrico Corregido'); disp(' '); original; diary off; disp('Probabilidades condicionadas (variables x clases)'); disp(p0); diary off; </pre>
Cálculo de las muestras para la corrección	<pre> tpo=st0(:,5); tm1=ceil(tm*tpo); tm2=tm-tm1; </pre>

Cálculo de los parámetros y estadístico de bondad de ajuste para cada remuestra Bootstrap	<pre> stat=[]; for b=1:bo1     clear x tabla;     x1=[];x2=[];     n1=ceil((tm1-1)*rand(tm1,1));     x1=x0(n1,:);     n2=ceil(np*rand(tm2,1));     x2=M(n2,:);     x=[x1;x2];     [tabla] = frecuen(x,M);     x=tabla;     [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00);     stat=[stat;bondad];disp(b); end </pre>
Decisión sobre el modelo	<pre> decision; </pre>
Presentación de los resultados	<pre> resultados; clear; </pre>

Este programa utiliza las funciones: patterns, frecuen, inicio y cmaa. También los programas “original”, “decision” y “resultados”, todos ya fueron presentados en el apartado anterior (5.3).

## 5.6 BOOTSTRAP PARAMÉTRICO

La introducción de los datos para la opción seleccionada “Paramétrico”, es semejante a la comentada en el apartado 5.3. Una vez introducidos los valores de entrada: número de clases latentes, número de remuestras Bootstrap y el nivel de significación (**Figura 5.8**), el programa inicia el proceso de cálculo, desplegando los resultados por pantalla y guardándolos en un archivo llamado “Parametrico.txt”.

parametrico

Diagnóstico de Modelos de Clases Latentes

Bootstrap Paramétrico

Panel

Número de clases latentes	
Número de remuestras	300
Nivel de significación	0.05

Continuar

Después de teclear continuar espere a que la pantalla se cierre.

**Figura 5.8.** Entrada de Datos a la rutina del Bootstrap Paramétrico

Primero, se presenta para la muestra original la siguiente información: los porcentajes de patrones de respuestas observados, los estadísticos de bondad de ajuste, los grados de libertad del modelo, las probabilidades totales de las clases y las probabilidades condicionales.

Segundo, para el diagnóstico con el método Bootstrap Paramétrico se exhibe la decisión sobre el modelo propuesto basado en los EBA, que consiste en  $H_0$ : No Rechazar y  $H_1$ : Rechazar. Seguidamente, para cada EBA se despliegan: el promedio aritmético, la desviaciones estándar y la desviación estándar respecto al estadístico original. Para los patrones de respuestas se da el promedio aritmético y la desviación estándar. Al final del listado de resultados son presentados los EBA obtenidos en cada remuestra Bootstrap ordenados según el valor en forma ascendente. La tabla de percentiles de los estadísticos de bondad de ajuste es guardada en un archivo en formato tipo Excel con el nombre “Parametrico.xls”.

Respecto al código de esta aplicación es muy semejante al utilizado para el Bootstrap No Paramétrico, la diferencia está en la sintaxis coloreada en la página anterior (pág. 115). Para el Bootstrap Paramétrico hay que cambiar esas instrucciones por las siguientes:

Pasos	Código del programa: Bootstrap Paramétrico
Cálculo de los parámetros y estadístico de bondad de ajuste para cada remuestra Bootstrap	<pre> for b=1:bol     x=[];     for k=1:T         h=R(k); [B]= clase(h,p000,k);         x=[x;B];     end     [tabla] = frecuen(x,M);    x=tabla;     [bondad,p0,c0,g1]=cmaa(x,T,c00,p00);     stat=[stat;bondad]; end </pre>

Las funciones “clase”, “frecuen” y “cmaa” ya han sido descritas con anterioridad.

## 5.7 MÉTODO DE PATRONES

Respecto al formato del fichero de datos de la muestra original obsérvese el apartado 5.3. El usuario debe proporcionar al programa utilizando el Método de Patrones, la siguiente información: número de clases latentes y número de remuestras de la población de patrones de respuesta (**Figura 5.9**).

Al igual que los módulos anteriores, los resultados son mostrados en la pantalla y guardados en un archivo llamado “Patrones.txt”. La primera parte del listado de resultados corresponde a los obtenidos con los datos originales. Cuando el número de patrones observados es inferior al número de patrones posibles, el programa da el siguiente mensaje: “*Cuidado, se presenta poca densidad de los datos*”.

Posteriormente, el programa proporciona los siguientes resultados: el porcentaje de patrones de respuestas observados, los estadísticos de bondad de ajuste de la muestra original, los grados de libertad del modelo, las probabilidades totales de las clases y las probabilidades condicionales.

**Figura 5.9.** Entrada de datos a la rutina del Método de Patrones.

A continuación se muestran los resultados de la simulación. Primero, se presenta la decisión sobre el modelo propuesto como válido, donde  $H_0$  implica no rechazar y  $H_1$  significa rechazar. Seguidamente, para los estadísticos de bondad de ajuste se presentan los siguientes estadísticos: promedio aritmético y desviaciones estándar respecto al estadístico original. En tanto, para los patrones de respuestas se suministra el promedio aritmético y las desviaciones estándar. Al final, son suministrados los resultados de los estadísticos ordenados y también, son guardados en un archivo tipo Excel llamado “Patrones.xls”, los percentiles de los estadísticos de bondad de ajuste. El código y la secuencia de pasos del programa para el Método de Patrones, es el siguiente:



Pasos	Código del programa: Método de Patrones
Entrada de datos	<pre> clc; disp('Método de los Patrones ') [filename, pathname]=uigetfile('*.dat', 'Seleccione un archivo .dat'); if ~isequal(filename,0)   ~isequal(pathname,0)     file = fullfile(pathname,filename);    x= textread(file); end x0=x; [tm,p]=size(x); </pre>
Número de clases	<pre> T=input('Número de clases latentes: '); </pre>
Número de muestras	<pre> bol=input('Número de muestras de patrones: '); </pre>
Nivel de significación	<pre> alpha=input('Nivel de significación: '); </pre>
Patrones de respuesta	<pre> [M,np] = patterns(p); </pre>
Frecuencias de patrones	<pre> [tabla] = frecuen(x,M); x=tabla; </pre>
Parámetros iniciales	<pre> [c00,p00] = inicio(p,T); </pre>
Parámetros del modelo	<pre> [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00); st0=bondad;c00=c0;p00=p0; </pre>
Resultados de la muestra original	<pre> delete('Metodo de Patrones.txt'); diary('Metodo de Patrones.txt'); disp('Diagnóstico de modelos: Método de Patrones'); disp(' '); original; diary off; </pre>
Parámetros y estadístico de bondad de ajuste para cada muestra de patrones	<pre> for b=1:bol     clear x tabla;     x=[];     pos=ceil(np*rand(tm,1));     x=M(pos,:);     [tabla] = frecuen(x,M);     x=tabla;     [bondad,p0,c0,g1]=cmaa(x,T,c00,p00);     stat=[stat;bondad]; end </pre>

Decisión sobre el modelo	decision;
Presentación de los resultados	resultados; clear;

## 5.8 SIMULACIÓN DE DATOS

Cuando se quiere simular datos bajo el supuesto de un modelo hipotético de clases latentes, la matriz de entrada de los datos tiene que contener los parámetros del modelo. La matriz debe contener  $T$  columnas (clases o grupos). En la primera fila, se introducen los tamaños de las clases  $\pi_Y(c)$  para  $c = 1, 2, \dots, T$ . A partir de la segunda fila ( $j = 2, 3, \dots, p$ ), se tiene que introducir para cada variable indicadora las probabilidades condicionales ( $\pi_{X_j/Y(c)}$ ) de la siguiente forma:

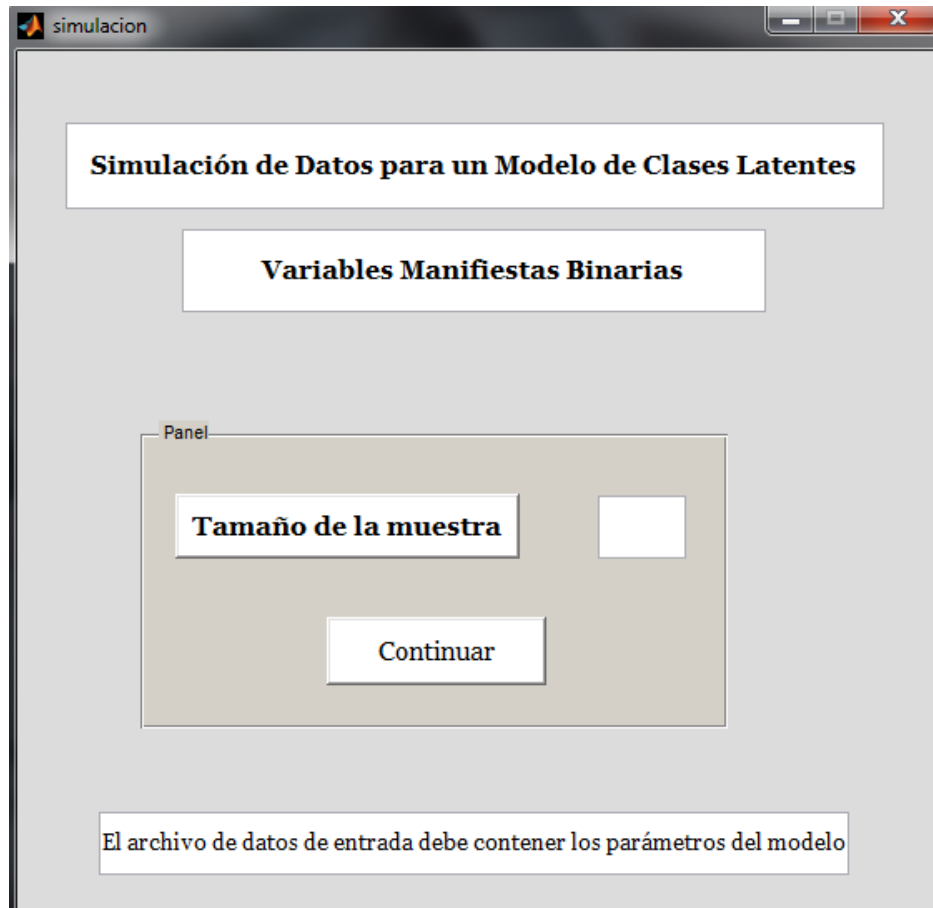
$$\begin{bmatrix} \pi_Y(1) & \pi_Y(2) & \cdots & \pi_Y(T) \\ \pi_{X_1/Y(1)} & \pi_{X_1/Y(2)} & \cdots & \pi_{X_1/Y(T)} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{X_p/Y(1)} & \pi_{X_p/Y(2)} & \cdots & \pi_{X_p/Y(T)} \end{bmatrix}$$

Considérese el siguiente ejemplo de una matriz de datos para 4 clases latentes y 7 variables manifiestas binarias:

0.40	0.30	0.25	0.05	→	Tamaño de las clases latentes
0.94	0.11	0.67	0.36		
0.97	0.59	0.15	0.48	}	Probabilidades condicionales
0.50	0.94	0.60	0.15		
0.38	0.85	0.30	0.20		
0.50	0.77	0.70	0.95		
0.08	0.31	0.97	0.17		
0.50	0.19	0.80	0.40		

Como datos de entrada al programa, además de la matriz anteriormente mencionada (guardada en un archivo \_\_\_\_\_.dat), solamente debemos introducir el

tamaño de la muestra sobre el campo vacío (o en blanco) que se activará (**Figura 5.10**) y dar un “click” izquierdo al ratón sobre “continuar”. El fichero de datos, resultado de la simulación, es grabado en el disco duro con el nombre de “datos.dat”.



**Figura 5.10.** Entrada de datos a la rutina de simulación de datos

La sintaxis y la secuencia de pasos del programa que hace posible simular conjuntos de datos es la siguiente:

Pasos	Código del programa: Simulación de Datos
Entrada de datos	<pre> clc;clear; disp('Simulación de datos de clases latentes '); [filename, pathname]=uigetfile('*.dat','Seleccione un archivo .dat'); if ~isequal(filename,0)   ~isequal(pathname,0)     file = fullfile(pathname,filename);     x= textread(file); end </pre>

	<code>[p,T]=size(x);</code>
Introducir el tamaño de la muestra	<code>tm=input('Digitar el tamaño de la muestra: ');</code>
Cálculo del número de individuos por cada clase latente	<code>c0=x(1,:);</code> <code>p00=x(2:p,:);</code> <code>[R] = multisize(c0,tm);</code> <code>T=length(c0); z=[];</code> <code>for k=1:T</code> <code>    h=R(k);</code> <code>    [B]= clase(h,p00,k);</code> <code>    z=[z;B];</code> <code>end</code>
Guarda el archivo de datos	<code>disp('El archivo es guardado con el nombre:</code> <code>datos.dat');</code> <code>save datos.dat z -ascii;</code>

Las funciones para el algoritmo de simulación, “multisize” y “clase”, son programadas como funciones externas, se puede apreciar seguidamente el código fuente de las mismas. Primero, la función “multisize” calcula el número de elementos que son ubicados en cada clase latente dado el tamaño de la muestra que es introducido, considerando para ello, los tamaños de las clases representados en la primera fila de la matriz de parámetros.

Código de la función: multisize
<pre>function [R] = multisize(c0,tm)     d=length(c0);     if (sum(c0)&lt;.99999   sum(c0)&gt;1.000001)         error('probabilidad suma más que uno');     end     if ~isscalar(tm) error('tm si no entero'),     end     if ~isvector(c0) error('tm es un vector'),     end     R=binornd(tm,c0(1));     for i=2:d-1         N=tm-sum(R);         If N==0             R=[R 0];</pre>

```

        else
            z=c0(i)/sum(c0(i:d));
            R=[R binornd(N,z)];
        end
    end
R=[R tm-sum(R)];
end

```

La función “**clase**” construye los datos simulados partiendo de un generador de número aleatorios entre 0 y 1, este número es comparado con la celda correspondiente en la matriz de parámetros iniciales, para decidir el valor que toma la variable manifiesta binaria.

#### Código de la función: clase

```

Function [B] = clase(h,p00,k)
    for i=1:h
        for j = 1:length(p00)
            u = rand(1,1);
            if u< p00(k,j)
                B(i,j) = 1;
            else
                B(i,j) = 0;
            end
        end
    end
end
end

```

## 5.9 CONSIDERACIONES FINALES

La ventana del menú no desaparece cuando es seleccionada una opción de análisis. Tras la ejecución de una opción en particular, puede comenzar otro desde el menú principal. Mientras no cierre la ventana del menú el cuadro permanece activo. Para salir del programa *BootLatentClass* basta teclear los comandos exit ó quit y pulsar la tecla Enter. Otra opción es abrir el menú desplegable File y ejecutando la orden Exit MATLAB®. Cuando se desea interrumpir una tarea antes de su término, se hace con la combinación ctrl-C.

En general, la implementación en MATLAB® resulta de utilidad por las limitaciones de los programas comerciales. Es un aporte único para la simulación de datos y el diagnóstico de un modelo de clases latentes usando el *Método de Patrones*. Sin embargo, al tratarse MATLAB® de un lenguaje interpretado o compilado en tiempo de ejecución<sup>3</sup>, las aplicaciones desarrollados con MATLAB® no son muy eficientes respecto al tiempo de ejecución, lo que hace poco adecuados para realizar simulaciones con los métodos Bootstrap y Patrones, cuando los tamaños de la muestra y el número de remuestras son grandes.

## 5.10 COMPARACIÓN DEL BOOTLATENTCLASS CON OTROS PROGRAMAS

Mediante la aplicación del Modelo de Clases Latentes, se obtienen resultados semejantes a los programas comerciales como LEM (VERMUNT, 1997b), WINMIRA (von DAVIER, 2001) y LATENTGOLD (MAGIDSON & VERMUNT, 2005). En este sentido, con objeto de medir la eficiencia del programa desarrollado, hemos llevado a cabo pruebas numéricas que confirman la precisión del programa *BootLatentClass*. Un ejemplo, para poner a prueba la anterior afirmación es el conjunto de datos simulados con el programa LEM de la **Tabla 5.1**, correspondientes a 808 elementos para cuatro variables manifiestas binarias (0=no, 1=sí). La quinta columna representa las frecuencias observadas de los patrones de respuestas.

La comparación de los resultados obtenidos para un modelo con 2 clases latentes se presenta en la **Tabla 5.2**. Se observa que los programas LEM y *BootLatentclass*, proporcionan resultados completamente iguales tanto en los estadísticos de bondad de ajuste, como en los parámetros de distribuciones de las clases. Por otra parte, las diferencias entre el programa *BootLatentClass* respecto a WINMIRA y LATENTGOLD, es muy pequeña en todos los valores que nos sirven de comparación.

---

<sup>3</sup> Un lenguaje compilado es una serie de órdenes que se pasan a un intérprete para que las ejecute. Por esta razón, MATLAB® no cumplen la definición de lenguaje porque no son ejecutables por ellos mismos.

$X_1$	$X_2$	$X_3$	$X_4$	Frecuencia ( $f_{x_h}$ )
0	0	0	0	56
0	0	0	1	71
0	0	1	0	10
0	0	1	1	30
0	1	0	0	1
0	1	0	1	13
0	1	1	0	3
0	1	1	1	20
1	0	0	0	89
1	0	0	1	145
1	0	1	0	28
1	0	1	1	116
1	1	0	0	14
1	1	0	1	80
1	1	1	0	19
1	1	1	1	113

**TABLA 5.1.** Datos simulados para cuatro variables manifiestas y dos clases latentes.

En el próximo capítulo ponemos en práctica las alternativas de diagnósticos de los MCL con datos simulados y reales. De tal modo, que sobre la base de los resultados experimentales obtenidos con el programa BootLatentClass, analizaremos la propiedades de los métodos que han sido propuestos para solucionar el problema de las tablas poco ocupadas y en particular, haremos énfasis en las ventajas y desventajas del Bootstrap No Paramétrico Corregido y *Método de Patrones*.

PROGRAMA	Estadístico de Bondad de Ajuste				Tamaño de las Clases	
	Pearson	Razón de Verosimilitud	Cressie-Read	Freeman-Tukey	1	2
LEM	1,6668 (0,9477)	1,8280 (0,9348)	1,7135 (0,9441)	***	0,6235	0,3765
WINMIRA	1,76 (0,9402)	1,95 (0,9239)	1,82 (0,9357)	2,08 (0,9119)	0,6189	0,3811
LATENGOLD	1,8942 (0,93)	2,1400 (0,91)	1,9648 (0,92)	***	0,6158	0,3142
BOOTLATENTCLASS	1,6668 (0,9477)	1,8280 (0,9348)	1,7135 (0,9441)	1,9370 (0,9254)	0,6235	0,3765

\*\*\* No calculan este estadístico.

**TABLA 5.2.** Comparación de resultados de programas comerciales y *BootLatentClass*.



## **CAPÍTULO 6**

---

# **APLICACIÓN DE LAS PROPUESTAS**

---

## 6.1 INTRODUCCIÓN

Para la ilustración de los métodos Bootstrap y *Método de Patrones* expuestos en los capítulos precedentes, se ha propuesto un análisis de conjuntos de datos reales y simulados, con la finalidad de medir la eficiencia de los métodos de diagnóstico de clases latentes en tablas poco ocupadas.

En una primera fase se estudian varios conjuntos de datos simulados para diferentes números de variables manifiestas binarias y número de clases latentes, con el propósito de verificar la efectividad de los métodos estudiados, en poder detectar el modelo con los cuáles los datos fueron construidos.

En la segunda fase se realizaron análisis con datos sobre el consumo de drogas en adolescentes<sup>1</sup>, pruebas sobre la detección del síndrome de inmunodeficiencia adquirida e identificación del síndrome de Alzheimer. Los datos sobre consumo de drogas forman parte de estudio más extenso, patrocinado por un conjunto de instituciones de Costa Rica, durante noviembre del 2006, como el Instituto Costarricense sobre Drogas y la Caja Costarricense de Seguro Social. El conjunto de datos sobre las pruebas de detección del VIH, se tomó de la revista: “Aids Research and Human Retroviruses”, publicado por ALVORD et al, 1988. En tanto, los datos sobre los síntomas para la identificación del síndrome de Alzheimer, fueron publicados por WASH (2006), en la revista Metodološki Zvezki.

Para los procedimientos de estimación de la distribución de probabilidad empírica de los estadísticos de bondad de ajuste, tanto con datos reales y simulados, se utilizan 300 remuestras usando los Métodos Bootstrap y *Método de Patrones*.

---

<sup>1</sup> El análisis es un extracto del artículo publicado en la “Revista de Matemática: Teoría y Aplicaciones”, en enero 2010. Al documento original se le ha agregado los resultados del Bootstrap No Paramétrico y *Método de Patrones*.

## 6.2. APLICACIÓN A DATOS SIMULADOS

Este apartado tiene interés ya que, con la *experimentación* que consiste en generar conjuntos de datos (o muestras aleatorias) dado un modelo de clases latentes, podemos analizar el comportamiento de los estadísticos de bondad de ajuste y de los métodos de diagnóstico en situaciones de tablas poco ocupadas. El proceso de verificación de la simulación consiste en comprobar que el modelo simulado es confirmado como válido o rechazado por los métodos de diagnóstico según si cumple con los requisitos del diseño para los que se elaboró.

En concreto, se simularon cinco conjuntos de datos binarios, correspondientes a diferentes tamaños de muestra y por lo tanto, con diversos grados de densidad (en inglés, “sparseness”)<sup>2</sup> de la tabla de patrones de respuesta, estimado por el factor  $n/R$ , donde  $n$  es el tamaño de la muestra original y  $R$  el número de patrones de respuesta. El detalle de los conjuntos de datos simulados se encuentra en la **Tabla 6.1**.

La metodología consiste en determinar si los modelos de clases latentes para los conjuntos de datos: “Datos 1”, “Datos 2” y “Datos 4” son correctamente confirmados por los métodos de diagnóstico estudiados. En tanto, para los datos: “Datos 3” y “Datos 5” serán evaluados modelos utilizando 2 clases latentes, con el objetivo de analizar cuales métodos rechazan el modelo, que de antemano, sabemos que es incorrecto.

---

<sup>2</sup> “Although such sparseness usually does not cause problems in estimating a model, evaluation of model fit by means of chi-square-based statistics will be invalidated because the true distribution of these statistics is badly approximated by the theoretical  $\chi^2$  distribution” (HAGENAARS & McCUTCHEON (2002), p. 336).

Conjuntos de Datos (1)	Tamaño de la muestra	Número de variables binarias	Total de patrones de respuesta	Número de patrones de respuesta observados	Número de clases latentes	Densidad de patrones	
						Estimada ( $n/k$ )	Real (2)
Datos 1	160	5	32	8	2	5	0,25
Datos 2	640	6	64	50	3	10	0,78
Datos 3	960	6	64	45	4	15	0,70
Datos 4	1.280	7	128	60	3	10	0,47
Datos 5	1.920	7	128	49	4	10	0,38

(1) En el **Anexo 1** se presentan los parámetros del modelo y las tablas de frecuencias de los patrones de respuesta para cada conjunto de datos.

(2) Calculado mediante la proporción de patrones de respuestas observados en los archivos de datos simulados.

**TABLA 6.1.** Simulación de 5 conjuntos de datos para modelos de clases latentes según la densidad de patrones.

En la **Tabla 6.2** podemos observar que, el modelo con 2 clases latentes para el conjunto de datos “**Datos 1**” es confirmado correctamente por todos los métodos, presentándose diferencias solamente en el valor de la significación observada ( $p$ -valor). La utilización de únicamente celdas con información ( $f_{x_h} \geq 1$ ), en general, proporciona los  $p$ -valores máximos en los estadísticos de bondad de ajuste (EBA). El Bootstrap Paramétrico es conservador, en el sentido, que los  $p$ -valores de los EBA son los más pequeños; solo en el estadístico de Pearson es similar al proporcionado por la maximización por la función de  $\beta$ -verosimilitud. El método de análisis de residuos confirma el modelo de 2 clases latentes, véase en la **Figura 6.1** que los tipificados se concentran alrededor de cero entre -1,50 y 1,50; además el contraste de hipótesis sobre la constante y el coeficiente de regresión validan el modelo hipotetizado (**Tabla 6.3**).

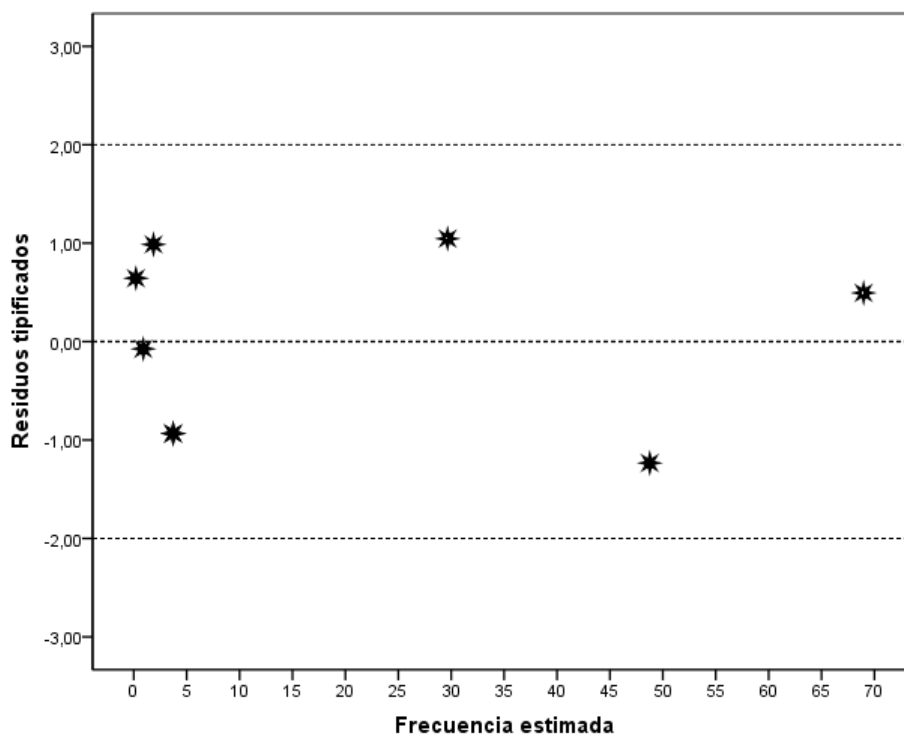
Conjunto de datos <b>Datos 1</b>	Significación observada (p-valor)						
	Original (1)	Bootstrap No Paramétrico		Bootstrap Paramétrico	Método de Patrones	Celdas ocupadas	Beta (3)
		Naive (2)	Corregido				
Razón de Verosimilitud	1,00	0,53	1,00	0,18	1,00	1,00	1,00
Estadístico de Pearson	0,30	0,66	0,38	0,29	0,22	1,00	0,30
Freeman-Tukey	1,00	0,54	1,00	0,21	1,00	1,00	1,00
Cressie-Read	0,93	0,59	0,96	0,21	0,86	1,00	0,93

(1) Calculado haciendo caso omiso a la corrección de los grados de libertad, debido a las celdas vacías, que es lo correcto. El número de grados de libertad corregidos por las celdas vacías es -4; esto imposibilita el uso de la distribución teórica  $\chi^2$ .

(2) "Naive" significa que es el Bootstrap no corregido.

(3) Maximización de la función  $\beta$ -Verosimilitud ( $\beta = 0,50$ ).

**TABLA 6.2.** Significación observada para los estadísticos de bondad de ajuste según el método de diagnóstico.



**FIGURA 6.1.** Representación de las frecuencias estimadas respecto a los residuos tipificados (Datos 1).

$H_0: \alpha = 0$ $H_1: \alpha \neq 0$					
Archivo de datos	Constante de regresión	Error estándar	Estadístico $t$	Grados de libertad	$p$ -valor*
Datos 1	0,170	0,438	0,389	6	0,711
Datos 2	0,397	0,356	1,115	47	0,271
Datos 3	2,406	2,130	1,129	43	0,265
Datos 4	0,143	0,420	0,340	58	0,735
Datos 5	10,636	3,828	2,779	47	0,008
$H_0: \beta = 1$ $H_1: \beta \neq 1$					
Archivo de datos	Coefficiente de regresión	Error estándar	Estadístico $t$	Grados de libertad	$p$ -valor*
Datos 1	1,005	0,014	0,36	6	0,733
Datos 2	0,976	0,018	1,33	47	0,189
Datos 3	0,972	0,079	0,35	43	0,725
Datos 4	1,003	0,009	0,33	58	0,740
Datos 5	0,931	0,08	0,86	47	0,393

**TABLA 6.3.** Contraste de hipótesis sobre la constante y el coeficiente de regresión lineal.

De igual forma, los métodos de diagnóstico determinan correctamente el modelo con 3 clases latentes en los datos “Datos 2”. Se observa en la **Tabla 6.4**, que las significaciones calculadas ( $p$ -valores) son semejantes, con excepción de los métodos Bootstrap No Paramétrico que brinda probabilidades de 1. El Bootstrap Paramétrico y *Método de Patrones* dan probabilidades muy similares. En la **Figura 6.2**, se observa que los residuos tipificados se encuentran entre -2 y 2, que ratifican la validez del modelo ajustado. Por esta misma razón, los contrastes sobre los parámetros de la regresión lineal no son rechazados (**Tabla 6.3**), como así debe ocurrir, cuando el modelo seleccionado es el apropiado para los datos.

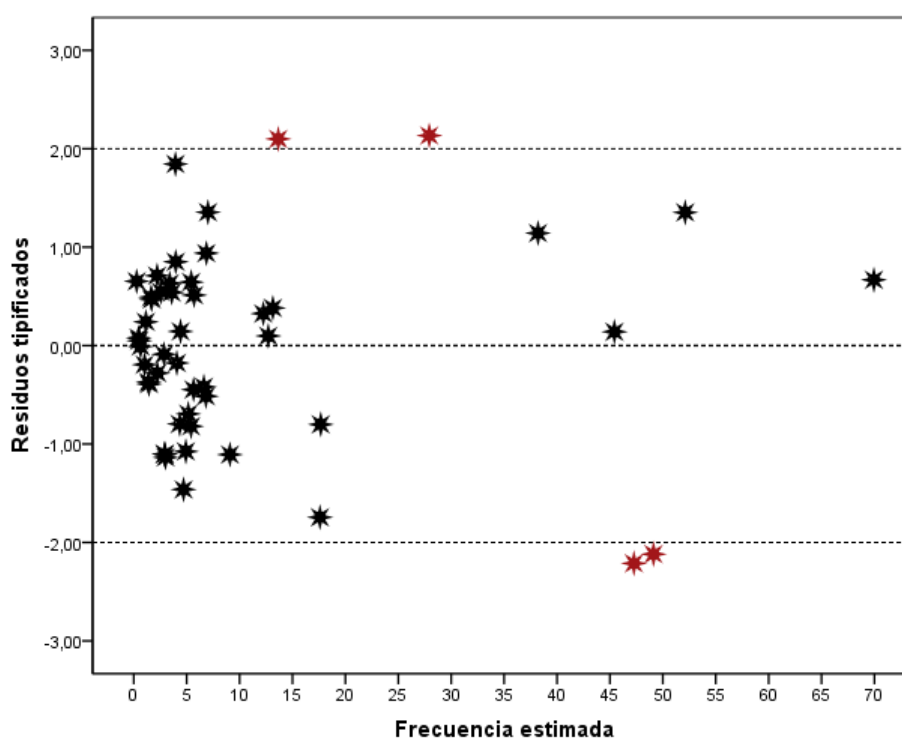
Conjunto de datos <b>Datos 2</b>	Significación observada (p-valor)						
	Original (1)	Bootstrap No Paramétrico		Bootstrap Paramétrico	Método de Patrones	Celdas Ocupadas	Beta (3)
		Naive (2)	Corregido				
Razón de Verosimilitud	0,02	1,00	0,92	0,36	0,30	0,26	0,23
Estadístico de Pearson	0,02	1,00	0,85	0,34	0,27	0,64	0,24
Freeman-Tukey	0,00	1,00	0,88	0,37	0,14	0,92	0,06
Cressie-Read	0,03	1,00	0,90	0,35	0,32	0,36	0,30

(1) Calculado usando el número de grados de libertad corregidos por las celdas vacías ( $gl = 29$ ).

(2) “Naive” significa que es el Bootstrap no corregido.

(3) Maximización de la función  $\beta$ -Verosimilitud ( $\beta = 0,50$ ).

**TABLA 6.4.** Significación observada para los estadísticos de bondad de ajuste según el método de diagnóstico.



**FIGURA 6.2.** Representación de las frecuencias estimadas respecto a los residuos tipificados (Datos 2).

Respecto a los datos “**Datos 3**”, es de esperar que los métodos de diagnóstico rechazaran el modelo ajustado de 2 clases latentes porque realmente fueron simulados con 4 clases latentes. Así lo hacen todos los métodos excepto el Bootstrap No Paramétrico “Naive” y el análisis de residuos. Aunque los residuos

tipificados varían de forma aceptable entre -2 y 2 (**Figura 6.3**), la hipótesis sobre el coeficiente de regresión, definida como  $H_0: \beta = 1$ , no es rechazada, siendo lo correcto lo contrario (**Tabla 6.3**).

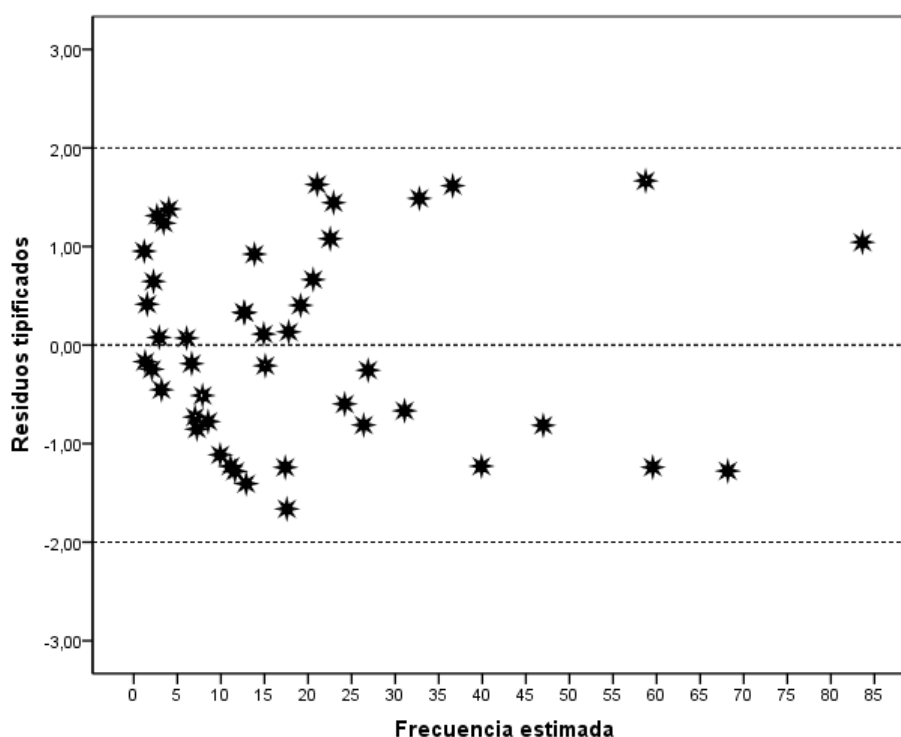
Conjunto de datos <b>Datos 3</b>	Significación observada (p-valor)						
	Original (1)	Bootstrap No Paramétrico		Bootstrap Paramétrico	Método de Patrones	Celdas ocupadas	Beta (3)
		Naive (2)	Corregido				
Razón de Verosimilitud	0,00	0,82	0,00	0,00	0,00	0,00	0,00
Estadístico de Pearson	0,00	0,82	0,00	0,00	0,00	0,00	0,00
Freeman-Tukey	0,00	0,84	0,00	0,00	0,00	0,00	0,00
Cressie-Read	0,00	0,83	0,00	0,00	0,00	0,00	0,00

(1) Calculado usando el número de grados de libertad corregidos por las celdas vacías ( $gl = 31$ ).

(2) "Naive" significa que es el Bootstrap no corregido.

(3) Maximización de la función  $\beta$ -Verosimilitud ( $\beta = 0,50$ ).

**TABLA 6.5.** Significación observada para los estadísticos de bondad de ajuste según el método de diagnóstico.



**FIGURA 6.3.** Representación de las frecuencias estimadas respecto a los residuos tipificados (Datos 3).



En relación a los datos “**Datos 4**”, todos los métodos de diagnóstico suministraron resultados correctos, ya que los niveles de *p-valores* nos indican que no debemos rechazar el modelo con 3 clases latentes (**Tabla 6.6**). Aunque el Bootstrap Paramétrico es más conservador, al proporcionar *p-valores* para los estadísticos de bondad de ajuste inferiores a 1, como así lo hacen todos los demás métodos de diagnóstico de modelos. La **Figura 6.4**, muestra que algunos residuos tipificados pueden considerarse grandes al dar resultados mayores a 3, a pesar de esta situación, las hipótesis  $H_0: \alpha = 0$  y  $H_0: \beta = 1$  sobre los parámetros de la regresión lineal entre las frecuencias esperadas y observadas de los patrones de respuesta, no deben ser rechazadas según los *p-valores* de la **Tabla 6.3**.

Conjunto de datos <b>Datos 4</b>	Significación observada (p-valor)						
	Original (1)	Bootstrap No Paramétrico		Bootstrap Paramétrico	Método de Patrones	Celdas ocupadas	Beta (3)
		Naive (2)	Corregido				
Razón de Verosimilitud	0,07	1,00	1,00	0,86	1,00	1,00	1,00
Estadístico de Pearson	0,03	1,00	1,00	0,76	1,00	1,00	1,00
Freeman-Tukey	0,01	1,00	1,00	0,85	1,00	1,00	1,00
Cressie-Read	0,07	1,00	1,00	0,84	1,00	1,00	1,00

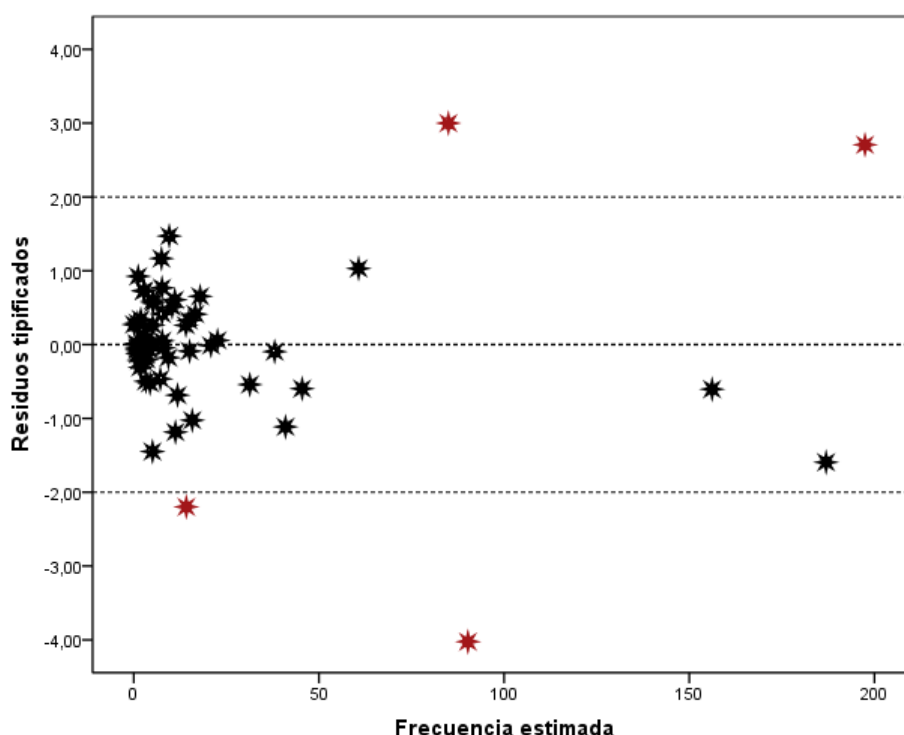
(1) Calculado usando el número de grados de libertad corregidos por las celdas vacías ( $gl = 36$ ).

(2) “Naive” significa que es el Bootstrap no corregido.

(3) Maximización de la función  $\beta$ -Verosimilitud ( $\beta = 0,50$ ).

**TABLA 6.6.** Significación observada para los estadísticos de bondad de ajuste según el método de diagnóstico.

Por último, al igual que lo observado con los resultados de los datos “**Datos 3**”, para los datos “**Datos 5**” los únicos métodos que no rechazan el modelo de 2 clases latentes son el Bootstrap No Paramétrico “Naive” (**Tabla 6.7**) y el análisis de residuos (**Tabla 6.3** y **Figura 6.5**), tal que el verdadero modelo para estos datos consiste en 4 clases. Los otros métodos brindan *p-valores* iguales a cero.



**FIGURA 6.4.** Representación de las frecuencias estimadas respecto a los residuos tipificados (Datos 4).

Conjunto de datos <b>Datos 5</b>	Significación observada (p-valor)						
	Original (1)	Bootstrap No Paramétrico		Bootstrap Paramétrico	Método de los Patrones	Celdas ocupadas	Beta (3)
		Naive (2)	Corregido				
Razón de Verosimilitud	0,00	0,68	0,00	0,00	0,00	0,00	0,00
Estadístico de Pearson	0,00	0,68	0,00	0,00	0,00	0,00	0,00
Freeman-Tukey	0,00	0,64	0,00	0,00	0,00	0,00	0,00
Cressie-Read	0,00	0,71	0,00	0,00	0,00	0,00	0,00

(1) Calculado usando el número de grados de libertad corregidos por las celdas vacías ( $gl = 33$ ).

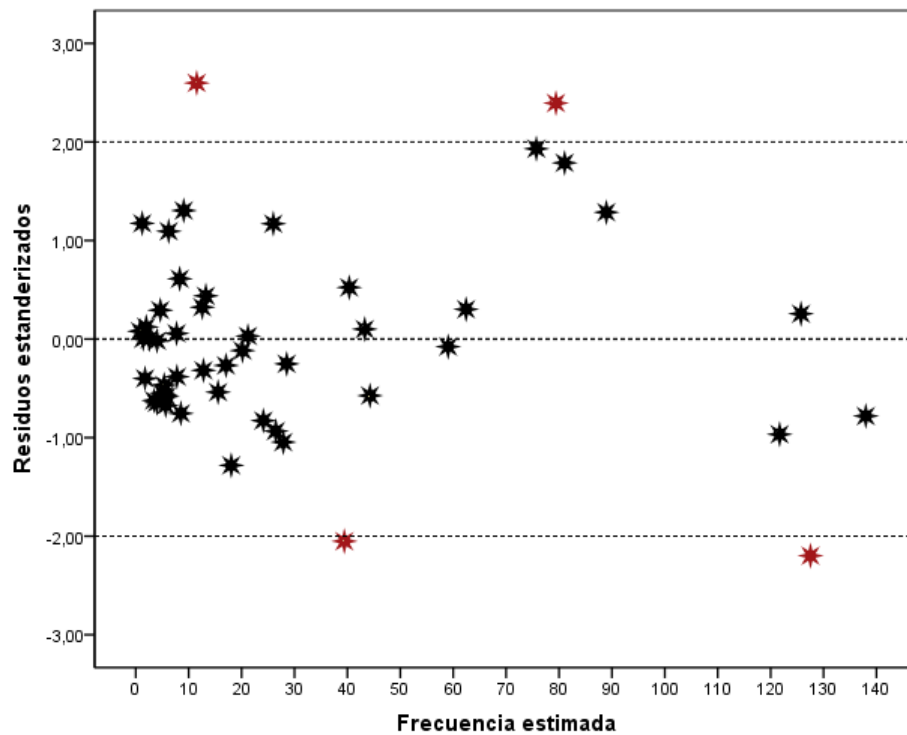
(2) “Naive” significa que es el Bootstrap no corregido.

(3) Maximización de la función  $\beta$ -Verosimilitud ( $\beta = 0,50$ ).

**TABLA 6.7.** Significación observada para los estadísticos de bondad de ajuste según el método de diagnóstico.

Como conclusión del análisis de los conjuntos de datos simulados, podemos afirmar que el Bootstrap No Paramétrico “Naive” y el análisis de residuos no funcionan para el diagnóstico de modelos de clases latentes en tablas poco

ocupadas, tal que no rechazaron los modelos incorrectos para los datos “**Datos 3**” y “**Datos 5**”. Además, el Bootstrap Paramétrico brinda los *p-valores* (o niveles de significaciones empíricos) más pequeños cuando el modelo es el apropiado para los datos.



**FIGURA 6.5.** Representación de las frecuencias estimadas respecto a los residuos tipicados (Datos 5).

### 6.3. APLICACIONES A DATOS REALES

La aplicación a datos reales tiene cuatro objetivos. *Primero*, demostrar que hacer caso omiso a los supuestos de la distribución de los estadísticos de bondad de ajuste, puede implicar la decisión de rechazar un modelo siendo el correcto (Error tipo I, rechazar  $H_0$  siendo verdadera) o no rechazar el modelo propuesto como válido siendo en la realidad incorrecto (Error tipo II, no rechazar  $H_0$  siendo falso). *Segundo*, comparar las medias aritméticas y las desviaciones típicas de los estadísticos. *Tercero*, comparar la significación calculada (p-valor) obtenida con los métodos Bootstrap y Patrones. *Cuarto*, tomar la decisión de rechazar o no rechazar el modelo de clases latentes propuesto como el apropiado para los datos.

### 6.3.1 CONSUMO DE DROGAS EN ADOLESCENTES

El objetivo del ejemplo es determinar a través del análisis de clases latentes, conglomerados de jóvenes en relación con sus patrones de consumo de drogas, para producir un conocimiento que brinde elementos que permitan direccionar las estrategias de prevención e intervención que actualmente se desarrollan en el país (Costa Rica).

Los datos corresponden a un estudio, llamado “*Factores que inciden en el consumo de drogas, población juvenil. Región Central de Occidente*” de Costa Rica, elaborado durante el mes de noviembre del 2006. El estudio fue patrocinado por el Instituto Costarricense sobre Drogas, la Asociación Ramonense Pro Bienestar de la Comunidad, la Caja Costarricense de Seguro Social, el Ministerio de Educación, el Instituto de Alcoholismo y Farmacodependencia y la Universidad de Costa Rica. La población en estudio está formada por 13.428 jóvenes perteneciente a 135 centros de enseñanza y 17 Equipos Básicos de Atención Integral de Salud (EBAIS). La muestra de 7.553 jóvenes se seleccionó utilizando un muestreo por conglomerados. Esta encuesta se diseñó principalmente para proporcionar información sobre el consumo de drogas, actividades de tiempo libre, participación de los jóvenes en actividades de la comunidad, condiciones de vida y niveles de información de los jóvenes.

El módulo de interés son las variables respecto a “*Cuándo fue que, por primera vez utilizó drogas*”; se tienen así 12 drogas (variables o ítems) con respuestas binarias (0=no, 1=Sí)<sup>3</sup>. El listado de variables del cuestionario utilizado en la presente investigación es el siguiente:

- $X_1$  Fumó cigarrillos.
- $X_2$  Bebió cerveza.
- $X_3$  Bebió vino.
- $X_4$  Consumió bebidas fuertes (whisky, vodka, ginebra, guaro, etc.).

---

<sup>3</sup> Los jóvenes respondieron a qué edad comenzaron a consumir las drogas, para aquellos que marcaron una edad se codificó como 1 y en caso contrario, los que no han utilizados drogas como 0.

$X_5$	Fumó marihuana.
$X_6$	Consumió cocaína.
$X_7$	Utilizó pastillas (estimulantes, tranquilizantes).
$X_8$	Uso inhalante.
$X_9$	Utilizó alucinógenos.
$X_{10}$	Uso heroína.
$X_{11}$	Uso éxtasis.
$X_{12}$	Consumió crack.

En la **Tabla 6.8** se presentan las distribuciones de frecuencias marginales y bivariantes del consumo de drogas. Tanto en filas como en columnas las drogas se identifican con números de 1 a 12 según el orden expuesto en el párrafo anterior. En la diagonal de la matriz se presenta la distribución de frecuencias de las drogas; por ejemplo, 2.441 jóvenes tienen el hábito de fumar cigarrillos y 166 han consumido cocaína.

Las frecuencias bivariantes nos revelan por ejemplo, que 2.300 personas tienen el hábito de fumar y beber cerveza. También que 147 jóvenes consumen cocaína y marihuana. Se observa una asimetría de la matriz, determinada por la droga cinco (fumó marihuana), que resulta como una frontera entre las drogas débiles (o legales) y las fuertes (las prohibidas).

Por otra parte, considerando que se tiene 12 variables manifiestas binarias, el número posible de patrones de respuesta es 4.096. El tamaño de la muestra es 7.553 personas, sin embargo, el 97% de los patrones (es decir, 3.992) no fueron observados. En la **Tabla 6.9**, se presentan los patrones de respuestas observados, que equivalen aproximadamente a un 3% (137) del total de patrones. Por ejemplo el patrón (000000000000) nos indica que 1.514 personas no consumen ninguna de las doce drogas y por el contrario, 59 jóvenes consumen todas las drogas, que es el patrón (111111111111).

	DROGAS												
DROGAS		1	2	3	4	5	6	7	8	9	10	11	12
	1	2.441											
	2	2.300	4.573										
	3	3.239	4.202	5.588									
	4	831	2.840	2.886	3.016								
	5	498	576	580	554	601							
	6	145	155	154	154	147	166						
	7	125	129	131	127	109	94	139					
	8	121	133	137	128	110	95	92	145				
	9	98	103	102	104	98	84	87	86	110			
	10	79	80	82	83	79	79	75	76	78	89		
	11	82	91	91	92	85	80	82	81	81	81	99	
	12	96	100	101	102	97	92	87	88	81	83	85	109

**TABLA 6.8.** Matriz de frecuencias marginales y bivariantes sobre consumo de drogas.

Debido a que nos encontramos con una tabla de frecuencias con muchas celdas vacías, la selección del modelo adecuado para clasificar a los jóvenes en grupos homogéneos en relación al consumo de drogas, se llevó a cabo considerando los métodos Bootstrap y el Método de Patrones, con el propósito de determinar la distribución empírica de los estadísticos de bondad de ajuste. Para encontrar el modelo, se procedió al ajuste de múltiples modelos utilizando los métodos de diagnóstico en tablas poco ocupadas (los métodos Bootstrap y de Patrones) incrementando consecutivamente el número de clases latentes. Los modelos en que  $T = 1, 2, \dots, 7$  no son adecuados<sup>4</sup>.

<sup>4</sup> El modelo con 7 clases latentes proporciona los siguientes p-valores asociados a los EBA por medio del Bootstrap Paramétrico (B=300 remuestras): Razón de Verosimilitud 0,033, Estadístico de Pearson 0,033, Freeman-Tukey 0.0267 y Cressie-Read 0,033.

**TABLA 6.9.** Distribución de los patrones de respuestas observados sobre consumo de drogas.

Se procedió a estimar la distribución empírica de los estadísticos con el modelo compuesto por 8 clases latentes, los resultados se muestran en la **Tabla 6.11**. En general, el promedio aritmético y la desviación típica de los estadísticos de bondad de ajuste, son menores a través del Bootstrap Paramétrico, en comparación con los resultados obtenidos con el Bootstrap No Paramétrico. El Estadístico de Pearson presenta la diferencia más grande entre los métodos Bootstrap respecto al promedio aritmético y la desviación típica.

Estadístico	Resultado	P-valor Original (*)
Razón de Verosimilitud	453,83	0,00
Estadístico de Pearson	3.0184,24	0,00
Freeman-Tukey	428,82	0,00
Cressie-Read	3.627,85	0,00

(\*) Calculado usando el número de grados de libertad corregidos por las celdas vacías ( $gl = 33$ ).

**TABLA 6.10.** Estadísticos de la muestra original para un modelo con ocho clases latentes: consumo de drogas en adolescentes.

El *Método de Patrones* proporciona un valor promedio del estadístico Cressie-Read de 4001,57, relativamente similar al obtenido para la muestra original, con una desviación típica de 85,83, que resulta mucho menor al obtenido por el Bootstrap No Paramétrico y Paramétrico, 2236,42 y 287,50 respectivamente.

En cuanto, al contraste del modelo de 8 clases latentes hipotetizado para los datos de la muestra original, el Bootstrap No Paramétrico no lo rechaza con *p-valor*es altos (**Tabla 6.11**). El Bootstrap Paramétrico no lo rechaza únicamente a través del estadístico Freeman-Tukey, si trabajamos con niveles de confianza superiores al 95%. Este resultado se obtiene al aplicar nuestra propuesta de programa, *BootLatentClass*. A través del software WINMIRA los *p-valor*es de los cuatro estadísticos son aproximadamente 0,047. El *Método de Patrones*, no rechaza el modelo usando los estadísticos de Razón de Verosimilitud y Freeman-Tukey; pero sí a través de los estadísticos Pearson y Cressie-Read.

Con el propósito de tener más información para tomar la decisión sobre el modelo hipotetizado, se procedió a realizar un análisis de regresión lineal simple entre las frecuencias observadas y estimadas de los patrones de respuesta. La **Figura 6.6** presenta las frecuencias estimadas contra los residuos tipificados. La gran mayoría de los residuos se encuentran en el rango comúnmente aceptados como válidos, entre -3 y 3. Solamente se tienen dos residuos atípicos con valores superiores a 3, relacionados a los patrones de respuesta (11111000000) y



(1111111111), aunque no necesariamente estas observaciones alejadas son influyentes en los resultados.

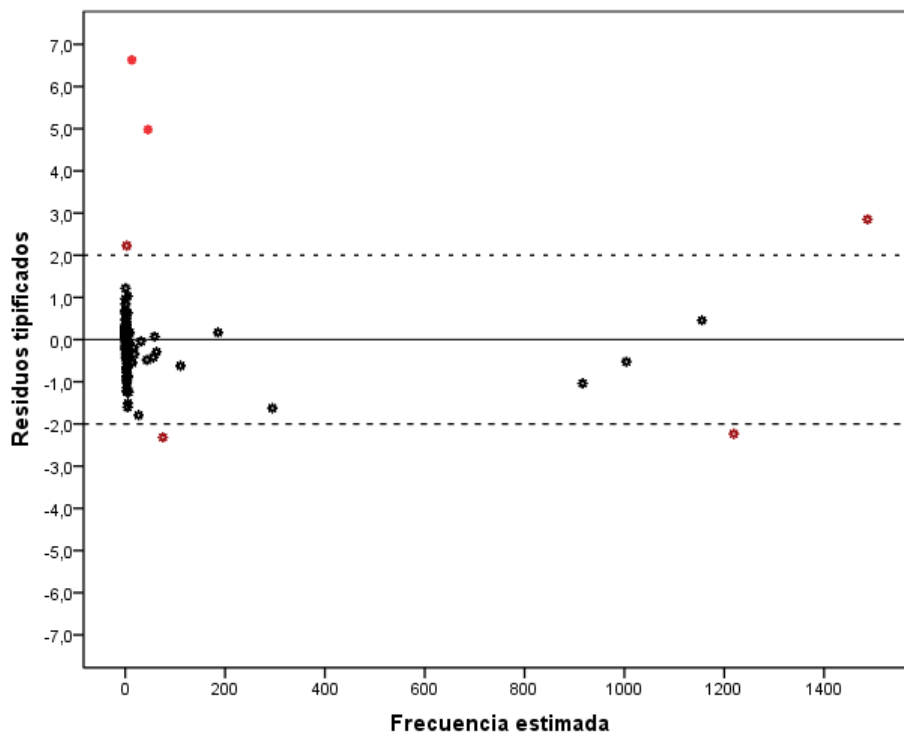
Estadístico	Promedio Aritmético	Desviación Típica	P-Valor
<b>Bootstrap No Paramétrico</b>			
Razón de Verosimilitud	529,93	61,48	0,88
Estadístico de Pearson	57.484,96	64.783,36	0,60
Freeman-Tukey	511,01	55,85	0,89
Cressie-Read	4.846,13	2.236,42	0,61
<b>Bootstrap Paramétrico</b>			
Razón de Verosimilitud	355,77	27,54	0,000
Estadístico de Pearson	3.351,38	3.738,76	0,010
Freeman-Tukey	383,04	26,97	0,046
Cressie-Read	919,85	287,50	0,000
<b>Método de Patrones</b>			
Razón de Verosimilitud	4.596,08	94,15	1,00
Estadístico de Pearson	4.002,15	91,83	0,00
Freeman-Tukey	6.667,16	148,33	1,00
Cressie-Read	4.001,57	85,83	0,00

**TABLA 6.11.** Estadísticos según los métodos de diagnóstico:  
consumo de drogas en adolescentes.

Por todo lo anterior, considerando los resultados de los *p-valores* de los estadísticos de bondad de ajuste calculados con los métodos de diagnósticos en tablas poco ocupadas, el análisis de residuos y la conformación de las clases latentes (**Tabla 6.12**)<sup>5</sup>; *concluimos que el modelo con 8 clases latentes nos permite explicar la dimensionalidad del consumo de drogas en adolescentes, a un*

<sup>5</sup> Cada clase latente estará caracterizada por aquellas variables para las que estas probabilidades condicionadas son mayores.

nivel de significación menor al 5%. A continuación detallamos la composición y caracterización demográfica de las clases latentes.



**Figura 6.6.** Análisis de residuos tipificados según regresión lineal simple para el modelo con ocho clases latentes.

La *primera clase latente (C1)* con una probabilidad de pertenencia de 0,298, se puede interpretar como los jóvenes abstemios, o sea, los que no consumen ningún tipo de droga, ya que las probabilidades condicionales son pequeñas para todas las drogas.

En tanto, los sujetos que componen la *segunda clase (C2)* son los consumidores de las drogas legales: cigarrillos (0,909) y cerveza (0,998), con vino (0,998) y vodka (0,983). Los jóvenes tienen una probabilidad de pertenecer a la clase de 0,185. Estas personas constituyen el grupo más tolerado por las sociedades.

La *tercera clase latente (C3)* se puede interpretar como la correspondiente al grupo de jóvenes “alcohólicos”, debido a que solamente tienen el hábito de

consumir bebidas alcohólicas: cerveza (0,895), vino (0,997) y vodka (0,774); la probabilidad de pertenecer a esta clase es 0,154.

Las personas de la *cuarta clase latente (C4)*, está formada por los que tienen el hábito de fumar (0,856), tomar cerveza (0,9228) y beber vino (0,833) y tiene una probabilidad de pertenencia de 0,137. Nótese que la diferencia con respecto a la *segunda clase*, consiste en que los jóvenes de la cuarta clase latente no consumen vodka o bebidas alcohólicas fuertes.

Los jóvenes de la *quinta clase (C5)* son aquellos que solamente tienen el hábito de tomar vino, la probabilidad de pertenencia a esta clase es 0,102 y la probabilidad condicional asociada al vino es de 0,733.

Drogas		C1	C2	C3	C4	C5	C6	C7	C8
		Probabilidades totales de las clases latentes							
		0,298	0,185	0,154	0,137	0,102	0,094	0,019	0,010
Cigarrillos	$X_1$	0,025	0,909	0,005	0,856	0,015	0,022	0,904	0,894
Cerveza	$X_2$	0,029	0,998	0,895	0,922	0,458	0,783	0,967	0,908
Vino	$X_3$	0,340	0,998	0,997	0,833	0,733	0,923	0,936	0,934
Vodka	$X_4$	0,009	0,983	0,774	0,438	0,015	0,067	0,962	0,947
Marihuana	$X_5$	0,002	0,228	0,042	0,035	0,002	0,002	0,835	0,950
Cocaína	$X_6$	0,002	0,004	0,002	0,002	0,002	0,002	0,556	0,963
Pastillas	$X_7$	0,002	0,007	0,002	0,004	0,002	0,002	0,327	0,963
Inhalantes	$X_8$	0,002	0,006	0,004	0,007	0,002	0,006	0,298	0,974
Alucinógenos	$X_9$	0,002	0,002	0,002	0,002	0,002	0,002	0,204	0,998
Heroína	$X_{10}$	0,002	0,002	0,002	0,002	0,002	0,002	0,091	0,987
Éxtasis	$X_{11}$	0,002	0,002	0,002	0,002	0,002	0,002	0,147	0,998
Crack	$X_{12}$	0,002	0,002	0,002	0,002	0,002	0,002	0,218	0,998

**TABLA 6.12.** Parámetros del modelo clásico de ocho clases latentes: probabilidades totales y condicionales.

La *sexta clase latente (C6)* con probabilidad de 0,094, son jóvenes “alcohólicos leves”, solamente tienen el hábito de tomar cerveza (0,783) y vino (0,923). En tanto, la *séptima clase latente (C7)* tiene probabilidades de pertenencia de 0,019. Están compuestas por jóvenes que consumen todas las

drogas legales y marihuana (0,835). Las probabilidades condicionales de las drogas legales son: cigarrillos 0,904, cerveza 0,967, vino 0,936 y vodka 0,962.

Finalmente, los casos más graves son sujetos *politoxicómanos* que definen la *octava clase latente (C8)* y la probabilidad de que un adolescente pertenezca a la clase es de 0,0100. Se observa que todas las drogas tienen probabilidades altas con relación a la clase. La politoxicomanía o policonsumo se produce cuando una persona se administra una variada gama de drogas. Hemos clasificado a los adolescentes en ocho clases latentes y con ello, se obtuvo una segmentación de los jóvenes en grupos. Trataremos ahora de describir las características demográficas de los jóvenes pertenecientes a cada uno de estos.

Con relación al sexo de los jóvenes encuestados, los varones representan la mayoría en las clases anteriormente analizadas, solamente en la clase latente 1 y 5 predominan las mujeres con un 58,4% y 52,4% respectivamente. En tanto, la clase ocho (o sea, los que consumen todas las drogas) está constituida por un 79,5% de hombres.

Los adolescentes que poseen mayor probabilidad de pertenecer a la clase 5 son los más pequeños con una edad media de 13,31 años y seguidamente los que no consumen ningún tipo de droga (clase 1) con un promedio de 14,7 años. Por otra parte, los sujetos de la clase 4 que tienen el hábito de fumar y tomar cerveza son las personas que promedian los 16,57 años. Podemos concluir que las medias de edad de los adolescentes que forman los ocho grupos no todas son iguales ( $p\text{-valor}=0,000$ ).

En todas las clases latentes se observa que los adolescentes están estudiando, pero en los *politoxicómanos* el 13,7% ha dejado de estudiar. En relación con el último año aprobado la mayoría están cursando la enseñanza secundaria, solamente el 60,7% de los jóvenes de la clase 6 (vino – cerveza) son de la enseñanza primaria.

El porcentaje de jóvenes que han trabajado o realizaron tareas en casa en los últimos 12 meses es pequeño (menos del 7%) en los ocho grupos. Los que

trabajaron y estudiaron tiende a aumentar, presentándose el porcentaje más grande en la clase 7, o sea, los que consumen las drogas legales y marihuana (38,5%). En tanto, el menor porcentaje se tiene en la clase 1 (“sanos”) con un 20.7%.

En la clase 1, 3, 4, 6 y 7 los adolescentes dedicaron los últimos 12 meses solamente a estudiar: 57,0%, 69,5%, 74,0%, 53,5% y 58,9% respectivamente; por el contrario, en la clase 8 el 46.2% de las personas dedicaron el tiempo exclusivamente a estudiar. Es importante mencionar que los que no trabajaron y no hicieron nada en especial en los últimos 12 meses constituyen el 20,8% de los jóvenes en la clase latente 8, que resulta significativamente mayor que en todos los otros grupos.

Tal y como se advierte a partir de lo previamente expuesto y de los resultados reseñados, el modelo de clases latentes nos permite llegar a las siguientes conclusiones: hay un grupo importante de jóvenes que no son consumidores de ningún tipo de droga (30%), un pequeño número de personas son adictos a todas las drogas (1%), existen un conjunto jóvenes que solamente consumen bebidas alcohólicas (35%) y la mayoría de personas utilizan drogas legales (67%).

Los rasgos característicos de las diferentes clases (o segmentos) contribuyen a comprender ciertos aspectos cualitativos del consumo y utilización de las drogas, explorar su heterogeneidad y clasificar a las personas en distintas categorías para poder caracterizarlos según variables demográficas. Así por ejemplo, los hábitos de consumo son diferentes en ambos sexos, especialmente es importante destacar que las mujeres representan la mayoría en los abstemios y los que han probado vino. Los “policonsumidores” son más frecuentes a medida que aumenta la edad de los jóvenes.

Podemos afirmar que este trabajo presenta un modelo alternativo para estudiar el problema del consumo de drogas, al utilizar el modelo de clases latentes y así incorporar un enfoque multidimensional al análisis de los resultados, a diferencias de los trabajos que se han publicado a nivel nacional, que tienen limitaciones importantes: por lo general se limitan a estudiar el problema

utilizando métodos univariados como distribuciones de frecuencias y pueden no reflejar el proceso que subyace al problema multivariante de la drogadicción.

### 6.3.2 PRUEBAS DE DIAGNÓSTICO DEL VIH

El conjunto de datos sobre las pruebas de detección del VIH, se tomó de la revista: “Aids Research and Human Retroviruses”, publicado por ALVORD et al., 1988. Los datos se presentan en la **Tabla 6.13**. Estos corresponden a los resultados proporcionados por cuatro pruebas de diagnóstico del VIH, aplicados a un total de 428 sujetos. La descripción de cada una de las pruebas de diagnóstico es la siguiente:

Variable	Test	Descripción
$X_1$	RIA-ag121	Radioimmunoassay of antigen ag121
$X_2$	RIA-p24	Radioimmunoassay of VIH p24
$X_3$	RIA-gp120	Radioimmunoassay of VIH gp120
$X_4$	ELISA	Enzyme-linked immunosorbent assay

Cada variable aleatoria  $X_j$  toma el valor 0 para un resultado negativo y 1 para el resultado positivo.

- La prueba RIA-ag121 detecta la presencia de un antígeno específico, ag12, para el virus del VIH.
- La prueba RIA-p24 y RIA-gp120 (glucoproteína de superficie) detectan la presencia de proteínas específicas que son componentes estructurales del VIH.
- La prueba de ELISA (enzimoinmunoanálisis) detecta directamente la presencia del virus. Se basa en la captura sobre una base antigénica, específica del virus de la inmunodeficiencia humana, de los anticuerpos anti-VIH presentes en la muestra de suero. Presenta una gran sensibilidad (superior del 98% con las técnicas desarrolladas actualmente) pero el número de falsos positivos en grupos de bajo riesgo puede ser significativo (LÓPEZ-BERNALDO et al., 2007).

Las cuatro pruebas (para nuestro interés de análisis, son las variables manifiestas binarias) de diagnóstico son aplicadas a cada sujeto de la población de estudio. Las personas pueden o no presentar el VIH, por lo que el “verdadero” estado del sujeto, que representamos por la variable latente  $Y$ , puede tomar dos valores: el valor 1 si el sujeto es portador del virus o el valor 0 si el sujeto no es portador del virus del VIH.

$X_1$	$X_2$	$X_3$	$X_4$	$f_{x_h}$
0	0	0	0	170
0	0	0	1	15
0	0	1	0	0
0	0	1	1	0
0	1	0	0	6
0	1	0	1	0
0	1	1	0	0
0	1	1	1	0
1	0	0	0	4
1	0	0	1	17
1	0	1	0	0
1	0	1	1	83
1	1	0	0	1
1	1	0	1	4
1	1	1	0	0
1	1	1	1	128

**TABLA 6.13.** Patrones de respuesta para las cuatro pruebas de diagnóstico del VIH.

En la tabla anterior, se puede observar que solamente el 56% ( $R = 16$ ) de los patrones de respuesta son observados en la muestra, por tanto, estamos en presencia de una tabla poco ocupada. Los grados de libertad corregidos por las celdas vacías son -1 y -6 para los modelos con 2 y 3 clases latentes, respectivamente.

En la publicación citada, se ajusta primero un modelo con dos clases latentes (VIH positivo y VIH negativo) pero es rechazado. A continuación se presenta los *p-valores*, haciendo caso omiso a la poca densidad de los datos (sin remuestreo) y calculando la distribución de probabilidad empírica de los estadísticos con los Métodos Bootstrap y Patrones son los siguientes:

Estadístico	Resultado	Valores de la significación calculada ( <i>P-valores</i> )			
		Sin Remuestreo (*)	Estimando la distribución del estadístico		
			Bootstrap No Paramétrico	Bootstrap Paramétrico	Método de Patrones
Razón de Verosimilitud	16,23	0,0040	0,70	0,00	0,02
Estadístico de Pearson	17,11	0,0059	0,70	0,01	0,01
Freeman-Tukey	17,56	0,006	0,69	0,00	0,01
Cressie-Read	16,42	0,0065	0,71	0,01	0,02

(\*) Calculado haciendo caso omiso a la corrección de los grados de libertad, debido a las celdas vacías, que es lo correcto. El número de grados de libertad corregidos por las celdas vacías es -1; esto imposibilita el uso de la distribución teórica  $\chi^2$ .

**TABLA 6.14.** Valores de la significación calculada según método de diagnóstico para un modelo con dos clases latentes.

Prueba de Diagnóstico del VIH.

El Bootstrap Paramétrico y el Método de Patrones son similares en cuanto a la decisión de rechazar el modelo de dos clases latentes. Sin embargo, el Bootstrap No Paramétrico proporciona resultados discordantes, es decir, nos lleva a aceptar  $H_0$ .

Posteriormente los autores del artículo (ALVORD et al., 1988), ajustaron el modelo con tres clases latentes que fue aceptado como el apropiado, pero cometieron el error de no corregir los grados de libertad. Lo correcto es determinar la validez de modelo con los métodos Bootstrap y de Patrones. La **Tabla 6.15** presenta los estadísticos de bondad de ajuste de la muestra original.

Los tres métodos de diagnóstico confirman que el modelo apropiado debe tener tres clases latentes. Sin embargo, el promedio de los estadísticos de bondad de ajuste del *Método de Patrones* son más similares a los obtenidos con la muestra original de datos (**Tabla 6.16**) y además, la media aritmética de la proporción de patrones observados en las muestras es 1,00, mayor que 0,54 y 0,58 conseguidos con el Bootstrap No Paramétrico y Paramétrico, respectivamente. Por esta razón,



las conclusiones se deben hacer utilizando los *p-valores del Método de Patrones*, al tener mayor cantidad de información para estimar la distribución de probabilidad empírica de los estadísticos de bondad de ajuste.

Estadístico	Resultado	P-Valor (*)
Razón de Verosimilitud	3,06	0,34
Estadístico de Pearson	4,49	0,26
Freeman-Tukey	3,66	0,31
Cressie-Read	3,71	0,31

(1) Calculado haciendo caso omiso a la corrección de los grados de libertad, debido a las celdas vacías, que es lo correcto. El número de grados de libertad corregidos por las celdas vacías es -6; esto imposibilita el uso de la distribución teórica  $\chi^2$ .

**TABLA 6.15.** Estadísticos de la muestra original para un modelo con tres clases latentes. Prueba de diagnóstico del VIH.

Método	Promedio Aritmético	Desviación Típica	P-Valor
<b>Bootstrap No Paramétrico</b>			
Razón de Verosimilitud	3,18	3,12	0,37
Estadístico de Pearson	6,14	7,76	0,34
Freeman-Tukey	2,95	2,53	0,40
Cressie-Read	4,56	5,33	0,34
<b>Bootstrap Paramétrico</b>			
Razón de Verosimilitud	1,93	1,83	0,68
Estadístico de Pearson	3,05	4,00	0,77
Freeman-Tukey	1,96	1,88	0,64
Cressie-Read	2,38	2,51	0,77
<b>Método de Patrones</b>			
Razón de Verosimilitud	3,63	2,63	0,90
Estadístico de Pearson	3,61	2,60	0,83
Freeman-Tukey	3,65	2,66	0,87
Cressie-Read	3,61	2,61	0,88

**TABLA 6.16.** Estadísticos según los métodos de diagnóstico para el modelo con tres clases latentes. Prueba de detección del VIH.

Los parámetros del modelo con tres clases latentes se muestran en la **Tabla 6.17**. La *clase latente 1* posee una probabilidad de pertenencia del 0,06 y está compuesta por los resultados positivos en las pruebas RIA-ag121 (1,00) y ELISA (1,00). En tanto, la *clase latente 2* tiene probabilidades condicionales similares a la clase latente 1 en dos pruebas, pero la prueba RIA-gp120 (0,98) también exhibe su importancia. Estas dos clases latentes están formadas por todos los pacientes que dan positivo en la prueba del VIH. Los autores llaman a la primera clase latente, “VIH incierto” (en inglés, HIV uncertain). La *clase latente 3*, que tiene una probabilidad de pertenencia de 0,46, está compuesta por todos los pacientes que dan un resultado negativo en la prueba del virus de la inmunodeficiencia humana.

Variable	c=1	c=2	c=3
	Probabilidades totales de clase		
	0,06	0,48	0,46
RIA-ag121	1,00	1,00	0,03
RIA-p24	0,06	0,64	0,04
RIA-gp120	0,42	0,98	0,00
ELISA	1,00	1,00	0,08

**TABLA 6.17.**Parámetros del modelo con tres clases latentes.

Prueba de diagnóstico del VIH.

Concluimos que aunque al ignorar los supuestos de la distribución de probabilidad de los estadísticos de bondad de ajuste, se llega a la misma conclusión respecto al hacerlo mediante la metodología apropiada, lo importante es destacar, el *gran error de interpretación, al proporcionar conclusiones utilizando los p-valores obtenidos de manera incorrecta de la distribución de probabilidad Chi-Cuadrado*.

### 6.3.3 DIAGNÓSTICO DE LA ENFERMEDAD DE ALZHEIMER

Para este apartado, utilizamos datos publicados en la revista Metodološki Zvezki, a través de un artículo de CATHAL D. WASH (2006), sobre síntomas para la identificación del síndrome de Alzheimer.

El *diagnóstico de la enfermedad de Alzheimer* se realiza sobre la base de síntomas neurológicos y neuropsicológicos. Los investigadores del Hospital de St. James de Dublín y del Instituto Nacional Clínico Nacional de Irlanda, estudiaron la enfermedad utilizando síntomas neuropsicológicos, como son las alucinaciones y el trastorno afectivo. Para garantizar la homogeneidad de la muestra clínica y su validez, examinaron un conjunto de 240 individuos que no tenían diagnosticada la enfermedad. El denominador común en los pacientes, era que presentaban insuficiencia renal leve, síntoma que no está relacionado directamente con la enfermedad de Alzheimer.

El paciente con Enfermedad de Alzheimer va a presentar una combinación de síntomas cognitivos, conductuales y neurológicos. Las variables manifiestas son un conjunto de indicadores binarios para cada individuo, siendo la presencia o ausencia de los síntomas de la enfermedad: alucinaciones ( $X_1$ ), alteración de la actividad ( $X_2$ ), agresividad ( $X_3$ ), agitación ( $X_4$ ), alteración del ritmo diurno ( $X_5$ ) y el trastorno afectivo ( $X_6$ ).

La distribución de los patrones de respuestas observados es representada en la **Tabla 6.18**. A este respecto, solamente el 61% de los patrones fueron observados en la muestra original de datos, esto implica, que la tabla tiene el problema de frecuencias nulas o ceros y por esta razón, estamos en presencia de una tabla poco ocupada. El diagnóstico de modelos de clases latentes apropiado, no se puede hacer asumiendo que los estadísticos de bondad de ajuste tienen una distribución de probabilidad Chi-Cuadrado y mucho menos, sin corregir los grados de libertad del modelo, como lo ha hecho el autor del artículo.

El procedimiento correcto para la selección del modelo de clases latentes, es efectuar el diagnóstico con los métodos Bootstrap y de Patrones. Aunque, el autor

del artículo WALSH, parte de un modelo predeterminado como el únicamente válido, el compuesto por tres clases latentes, él justifica que: “*un modelo de dos clases latentes no es clínicamente interesante, porque presenta sólo dos grupos de pacientes con síntomas altos y otros bajos*”<sup>6</sup>.

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
0	0	0	0	0	0	18
0	0	0	0	0	1	25
0	0	0	0	1	0	1
0	0	0	0	1	1	6
0	0	0	1	0	0	3
0	0	0	1	0	1	9
0	0	0	1	1	0	1
0	0	0	1	1	1	3
0	0	1	0	0	0	2
0	0	1	0	0	1	4
0	0	1	0	1	0	1
0	0	1	0	1	1	1
0	0	1	1	0	1	2

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
0	0	1	1	1	1	3
0	1	0	0	0	0	20
0	1	0	0	0	1	35
0	1	0	0	1	0	2
0	1	0	0	1	1	11
0	1	0	1	0	0	3
0	1	0	1	0	1	24
0	1	0	1	1	1	11
0	1	1	0	0	0	1
0	1	1	0	0	1	9
0	1	1	0	1	0	3
0	1	1	0	1	1	2
0	1	1	1	0	1	14

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
0	1	1	1	1	0	1
0	1	1	1	1	1	6
1	0	0	0	0	0	1
1	0	0	0	0	1	1
1	0	0	1	0	1	1
1	0	1	0	0	1	1
1	1	0	0	0	0	2
1	1	0	0	0	1	5
1	1	0	0	1	1	2
1	1	0	1	0	1	1
1	1	1	0	0	1	1
1	1	1	0	1	1	1
1	1	1	1	1	1	3

**TABLA 6.18.** Datos sobre la prevalencia de los patrones de síntomas de la Enfermedad de Alzheimer.

Haciendo caso omiso a la sugerencia de WALSH, la primera opción es evaluar con los métodos Bootstrap y *Método de Patrones*, si el modelo compuesto por dos clases latentes es el correcto. En la **Tabla 6.19** se presentan los estadísticos de bondad de ajuste y sus niveles significación, para el modelo señalado anteriormente.

Utilizando los métodos de diagnóstico de modelos en tablas poco ocupadas, se calculan las medias aritméticas, desviación típica y significación observada (**Tabla 6.20**). Los tres métodos de diagnóstico determinan que el modelo definido por dos clases latentes es el apropiado. Las medias de los estadísticos del Bootstrap

<sup>6</sup> “A 2 class model is not clinically interesting, because it shows only a high and low symptom group. The movement beyond 3 classes is not justified by an improvement in fit”.

No Paramétrico son mayores que las respectivas en el Bootstrap Paramétrico y *Método de Patrones*; debido a que las medias de las distribuciones empíricas estimadas con el Bootstrap No Paramétrico son aproximadamente dos veces más grandes que las obtenidas con la muestra original.

Estadístico	Resultado	P-Valor (*)
Razón de Verosimilitud	36,87	0,06
Estadístico de Pearson	42,47	0,02
Freeman-Tukey	41,82	0,02
Cressie-Read	38,63	0,04

(\*) Calculado usando el número de grados de libertad corregidos por las celdas vacías ( $gl = 25$ ).

**TABLA 6.19.** Estadísticos de la muestra para un modelo con dos clases latentes.  
Enfermedad de Alzheimer.

Método	Promedio Aritmético	Desviación Típica	P-Valor
<b>Bootstrap No Paramétrico</b>			
Razón de Verosimilitud	66,04	12,02	0,98
Estadístico de Pearson	74,37	17,18	0,97
Freeman-Tukey	89,02	17,02	0,98
Cressie-Read	66,82	13,50	0,98
<b>Bootstrap Paramétrico</b>			
Razón de Verosimilitud	42,33	8,05	0,59
Estadístico de Pearson	48,19	17,86	0,50
Freeman-Tukey	57,60	10,88	0,68
Cressie-Read	42,48	10,91	0,51
<b>Método de Patrones</b>			
Razón de Verosimilitud	50,74	10,25	0,87
Estadístico de Pearson	46,56	8,89	0,56
Freeman-Tukey	60,59	15,64	0,97
Cressie-Read	46,99	8,88	0,77

**TABLA 6.20.** Estadísticos según método de diagnóstico para un modelo con dos clases latentes. Enfermedad del Alzheimer.

Los resultados del Bootstrap Paramétrico y *Método de Patrones* son similares, tanto en la media aritmética como en la desviación típica. En cuanto al nivel de confianza del modelo, es recomendable concluir utilizando los niveles de

significación, porque son calculados con mayor cantidad de información – densidad de patrones de respuesta - que el Bootstrap Paramétrico y No Paramétrico.

Este modelo no era del agrado de WALSH. Evidentemente, ajustar un modelo para un número mayor de clases latentes, será válido con toda seguridad pero con más cantidad de parámetros. A este respecto, comenta WALSH, que desde una perspectiva clínica, las clases latentes incluyen una clase de baja y otra de alta prevalencia de los síntomas, la tercera clase está compuesta por aquellos que presentan alucinaciones. Los argumentos en favor del modelo con 3 clases latentes es que hay un ajuste ligeramente mejor, entre las frecuencias de los patrones observados y los estimados.

A pesar de lo comentado, los parámetros del modelo con dos clases latentes se presentan en la **Tabla 6.21**. La *clase latente 1* está compuesta por los pacientes que presentan síntomas graves de la Enfermedad de Alzheimer relacionados con la alteración de la actividad (0,81) y principalmente los síntomas de trastorno afectivo (0,96), representa el 44% de total personas estudiadas en la muestra.

Variables	c=1	c=2
	Probabilidades totales de clase	
	<b>0,44</b>	<b>0,56</b>
Alucinaciones	0,09	0,07
Alteración de la actividad	<b>0,81</b>	0,53
Agresividad	0,39	0,10
Agitación	0,65	0,12
Alteración del ritmo diurno	0,38	0,13
Trastorno afectivo	<b>0,96</b>	0,59

**TABLA 6.21.** Parámetros del modelo con dos clases latentes.  
Enfermedad del Alzheimer.

La *clase latente 2* formada por el 56% de los pacientes, tiene síntomas leves de la enfermedad en la alteración de la actividad (0,53) y trastorno afectivo (0,59). Las personas que padecen la enfermedad de Alzheimer leve suelen verse saludables, pero la verdad es que tienen dificultades para entender el mundo que las rodea (MOLLOY & CALDWELL, 2002).

Dado el interés del autor por el modelo con 3 clases latentes, vamos a compararlo respecto al ajustado (2 clases) a través de los criterios de información. En la **Tabla 6.22**, puede apreciarse el aumento en los valores de estos estadísticos en la solución de tres clases respecto a la de dos clases latentes. Presentándose mayor diferencia en Akaike Condicionado y Bayesiano. De esta manera, los criterios de información apuntan a la solución de dos clases como la más plausible.

Criterio de Información	Número de clases latentes	
	Dos	Tres
Akaike (AIC)	1.524,85	1.527,15
Bayesiano (BIC)	1.570,10	1.596,77
Akaike Condicionado (CAIC)	1.583,10	1.616,77

**TABLA 6.22.** Criterios de información según modelo de clases latentes.  
Enfermedad del Alzheimer.

*A manera de conclusión, afirmamos que el autor “pasa de lado” los supuestos sobre la distribución de probabilidad de los estadísticos de bondad de ajuste y selecciona el modelo con tres clases latentes, fundamentado en el conocimiento clínico de la enfermedad de Alzheimer.*

En el **capítulo 7**, propondremos un conjunto de tablas de valores críticos según el *Método de Patrones* para el diagnóstico de modelos en tablas poco ocupadas, que es un aporte de nuestro proyecto de tesis doctoral.

## **CAPÍTULO 7**

---

### **TABLAS ESTADÍSTICAS DE VALORES CRÍTICOS PARA EL MÉTODO DE PATRONES**

---



## 7.1 INTRODUCCIÓN

En este trabajo proponemos tablas de valores críticos para los estadísticos de bondad de ajuste, con fin de contrastar un modelo hipotético de clases latentes en tablas poco ocupadas. Esto es posible, ya que el *Método de Patrones* tiene un carácter general, no depende de los patrones de respuesta observados en la muestra, solamente toma el tamaño de la muestra para la simulación de la distribución empírica de los EBA.

El proceso de simulación de las tablas de valores críticos para los EBA, se realizó con 500 muestras aleatorias aplicando el *Método de Patrones*, para tratar que los valores tendieran a un límite de convergencia. Las tablas se construyen utilizando el siguiente procedimiento:

1. Selección del número de variables manifiestas binarias ( $p$ ).
2. Selección del número de clases latentes ( $T$ ).
3. Calcular el número de patrones posibles ( $R = 2^p$ ).
4. Construir la matriz de patrones de respuesta  $M$ .
5. Selección del tamaño de la muestra ( $n$ ), que determina el grado de densidad de patrones  $n/R$ .
6. Generar una muestra aleatoria simple de tamaño  $n$  con reemplazamiento sobre la población de patrones ( $M$ ).
7. Sobre esta muestra (paso 6), calcular el EBA  $\hat{\theta}_1^*$ .
8. Los pasos 6 y 7 se repiten 500 veces, de esta manera se obtienen los EBA para todas las muestras  $(\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_{500}^*)$ .
9. Ordenar los valores obtenidos de menor a mayor  $\hat{\theta}_1^* \leq \hat{\theta}_2^* \leq \dots \leq \hat{\theta}_{500}^*$ .
10. Calcular los percentiles  $p_m$ , donde  $m=0,25, 0,50, 0,75, 0,95, 0,96, 0,97, 0,98$  y  $0,99$ .

Todo el proceso, fue implementado mediante un programa escrito en MATLAB<sup>®</sup>, que a continuación se especifica.

Pasos	Código del programa: tablas
Entrada del número de clases	<code>clear; clc T=input('Número de clases latentes: ');</code>
Entrada del número de muestras de patrones	<code>bol=input('Número de muestras de patrones: ');</code>
Entrada del número de variables	<code>p=input('Número de variables manifiestas: ');</code>
Matriz de patrones (M)	<code>[M,np] = patterns(p);</code>
Tamaños de muestra	<code>C=2^p; vm=[]; for i=1:10     n=i*C; vm(i)=n; end</code>
Construcción de la tabla de valores críticos	<code>G=[]; for i=1:10     tm=vm(i);     [c00,p00] = inicio(p,T);     [E]= tablasfuncion(c00,p00,C,tm,M,T,bol);     G=[G;E]; end</code>
Guardar la tabla en un archivo Excel	<code>xlswrite('Tabla.xls', G, 'Hoja1');</code>

El programa para generar las tablas de valores críticos, utiliza las funciones “patterns”, “inicio” y “tablasfuncion”. Las funciones “patterns” y “inicio”, se presentaron en las páginas 105 y 106, respectivamente. En tanto, la función “tablasfuncion” construye la tabla de valores críticos para un valor de  $n/R$ . El código Matlab es el siguiente:

Pasos	Código de la función: <code>tablasfuncion</code>
Inicio de función	<code>function [E]= tablasfuncion (c00,p00,C,tm,M,T,b01)</code>
Cálculo de los estadísticos de bondad de ajuste para los $b01=500$ muestras de patrones	<pre> stat=[]; for b=1:b01     clear x tabla;     pos=ceil(C*rand(tm,1));     x=M(pos,:);     [tabla] = frecuen(x,M);     x=tabla;     [bondad,p0,c0,sparse]=cmaa(x,T,c00,p00);     stat=[stat;bondad]; end </pre>
Cálculo de los percentiles	<pre> [st1]=sort(stat); c1=1:1:b01;c1=c1'; A=c1./b01; C=[A st1];C(:,6)=[]; k1=b01*0.25;k2=b01*0.75; E1=C(k1:k1:k2,: )'; k1=b01*0.90;k2=b01*0.05;k3=b01*0.95; E2=C(k1:k2:k3,: )'; k1=b01*0.96;k2=b01*0.01;k3=b01*0.99; E3=C(k1:k2:k3,: )'; E=[E1 E2 E3]; end </pre>

Debido a que los tamaños de muestra pueden ser muy variados, hemos decidido representar en la tabla el factor  $n/R$ , donde  $n$  es el tamaño de la muestra original y  $R$  el número de patrones de respuesta. Por esta razón, representamos los valores de  $n/R$  que manifiesten aproximadamente el problema de la densidad de frecuencias para la selección de modelos de clases latentes. Los valores tabulados de  $n/R$  están comprendidos entre 1 y 10. Sin embargo, debe considerarse con precaución el valor crítico cuando  $n/R$  representada en la tabla sea muy diferente al valor real.

A continuación son presentadas las tablas de valores críticos, ordenadas según el número de variables manifiestas binarias y el número de clases. Por limitaciones de espacio y tiempo de procesamiento, se construyen las tablas solamente para 4, 5, 6, 7, 8, 9 y 10 variables.

**TABLA 1.** Valores Críticos para el Método de Patrones:  
4 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	8,2	10,9	13,9	16,6	19,2	20,3	21,5	22,7	24,1
	Q2	7,0	9,2	12,2	15,9	18,8	19,8	20,4	21,7	27,5
	FT	12,9	17,1	21,3	25,9	30,7	31,4	33,0	34,0	37,8
	CR	7,0	9,3	11,8	15,2	17,3	18,4	19,4	20,5	23,9
2	G2	7,1	10,2	13,7	17,2	19,7	20,4	20,8	22,1	24,8
	Q2	6,1	8,8	11,9	14,9	16,6	17,4	18,0	19,6	21,0
	FT	9,8	14,6	19,4	26,1	30,0	30,9	32,1	33,9	38,4
	CR	6,1	9,0	11,8	14,6	16,7	17,1	18,1	19,9	20,5
3	G2	5,1	7,6	11,0	14,5	17,4	18,2	18,6	20,2	21,5
	Q2	4,8	7,0	9,5	12,7	14,4	14,9	15,2	16,6	17,6
	FT	5,6	9,4	14,6	19,9	24,1	25,9	28,1	29,0	32,6
	CR	4,8	7,1	9,7	13,0	14,6	15,2	15,7	17,0	18,3
4	G2	6,2	9,0	12,7	16,2	18,6	19,5	19,8	21,3	23,4
	Q2	6,0	8,5	11,6	14,7	16,9	17,2	18,0	19,3	20,5
	FT	6,5	9,6	14,8	21,4	24,8	26,0	27,6	29,1	32,2
	CR	6,0	8,7	11,7	14,8	16,8	17,4	18,4	18,9	20,8
5	G2	6,3	8,9	12,5	16,5	19,3	20,1	21,3	22,8	25,7
	Q2	6,1	8,7	11,8	15,3	18,0	18,6	19,5	21,2	23,8
	FT	6,5	9,3	13,5	19,7	23,7	25,2	27,3	28,6	33,6
	CR	6,1	8,7	11,9	15,4	17,8	18,6	19,6	21,4	23,6
6	G2	6,1	9,0	12,5	16,7	20,1	20,9	21,9	23,2	25,1
	Q2	6,0	8,7	11,9	15,6	18,2	18,9	21,0	22,1	23,1
	FT	6,3	9,2	13,2	18,0	23,7	24,6	26,1	30,2	33,0
	CR	6,0	8,9	12,0	15,9	18,2	18,6	21,1	21,5	24,1
7	G2	4,3	6,5	9,6	12,8	15,5	16,0	16,4	17,8	19,1
	Q2	4,2	6,3	9,1	12,0	14,7	15,2	15,8	16,4	18,4
	FT	4,3	6,7	10,2	13,9	16,7	17,4	18,7	20,2	22,9
	CR	4,3	6,4	9,3	12,3	14,9	15,3	15,7	16,7	18,1
8	G2	4,3	6,5	9,8	12,5	14,7	15,5	15,9	16,7	17,8
	Q2	4,3	6,4	9,5	12,2	14,1	14,4	15,2	16,3	18,3
	FT	4,3	6,7	10,0	13,0	15,0	16,3	17,1	17,7	18,7
	CR	4,2	6,4	9,5	12,3	14,1	14,5	15,2	16,3	18,4
9	G2	4,6	6,7	9,5	13,0	14,6	15,0	16,2	17,5	19,5
	Q2	4,5	6,5	9,2	12,6	14,0	14,5	15,7	16,7	18,4
	FT	4,6	6,9	9,6	13,3	15,3	15,8	17,7	18,2	20,7
	CR	4,5	6,6	9,3	12,7	14,2	14,6	15,5	17,3	18,6
10	G2	4,1	6,2	8,8	11,9	13,8	14,3	14,9	15,2	17,1
	Q2	4,2	6,1	8,6	11,7	13,1	13,7	14,6	15,3	16,5
	FT	4,2	6,3	9,0	12,3	14,4	14,9	15,2	15,8	17,3
	CR	4,2	6,1	8,7	11,7	13,1	13,8	14,6	15,2	16,5

**TABLA 2.** Valores Críticos para el Método de Patrones:  
5 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	22,0	25,7	29,9	33,6	36,5	37,1	38,4	40,4	42,3
	Q2	18,1	21,8	26,0	30,3	33,3	35,0	35,8	38,2	39,3
	FT	34,8	40,2	46,8	53,0	56,5	58,2	59,7	63,2	67,5
	CR	18,4	21,6	25,5	29,2	32,3	33,5	34,5	35,9	36,8
2	G2	20,3	25,6	30,2	34,6	37,6	38,3	39,2	40,4	42,9
	Q2	17,3	21,1	26,0	30,3	33,5	33,8	35,0	36,9	39,9
	FT	28,7	37,1	44,7	52,4	56,0	56,8	57,6	60,1	62,0
	CR	17,7	21,5	26,3	30,2	32,9	33,5	34,7	36,3	37,9
3	G2	19,4	23,9	28,5	34,5	37,6	39,1	39,5	41,5	47,8
	Q2	17,6	21,1	25,3	29,8	33,2	34,0	34,8	36,7	39,8
	FT	23,3	30,2	37,6	46,2	53,5	55,3	56,6	58,3	70,4
	CR	17,9	21,7	25,7	30,2	33,3	34,4	35,5	36,5	40,3
4	G2	18,3	23,0	27,9	33,9	36,8	37,6	39,0	39,7	42,8
	Q2	17,2	21,2	25,6	30,9	34,4	34,7	35,6	37,2	38,6
	FT	19,8	25,9	34,8	42,2	47,7	48,8	51,3	53,5	58,0
	CR	17,5	21,4	25,8	31,1	34,3	34,7	35,3	36,9	38,7
5	G2	15,4	19,2	24,0	28,2	31,9	32,8	34,2	36,9	39,1
	Q2	14,9	18,0	22,3	26,8	28,9	30,2	31,6	33,3	37,0
	FT	16,1	20,4	26,3	34,1	38,4	40,2	42,1	44,3	49,2
	CR	15,0	18,2	22,5	26,8	29,3	30,5	31,8	33,2	37,2
6	G2	14,9	18,5	23,8	28,8	31,5	32,2	34,0	34,9	36,7
	Q2	14,3	17,8	22,6	27,5	30,1	30,8	31,1	32,0	33,2
	FT	15,3	19,0	25,7	31,5	36,3	38,1	39,6	40,7	43,8
	CR	14,4	18,0	22,9	27,8	30,2	30,5	31,4	32,4	33,6
7	G2	15,1	18,8	23,1	27,7	30,6	31,0	32,2	33,2	35,7
	Q2	14,8	18,4	22,3	26,2	29,0	29,3	30,5	31,4	33,1
	FT	15,5	19,5	24,1	29,5	33,1	34,8	36,2	38,3	42,5
	CR	14,8	18,4	22,5	26,6	29,1	29,7	30,4	31,7	33,1
8	G2	14,4	18,2	22,6	26,5	28,9	29,4	30,0	31,0	34,2
	Q2	14,2	17,8	21,8	25,8	28,1	28,5	29,1	30,4	32,6
	FT	14,8	18,8	23,6	28,0	30,6	31,2	31,7	34,2	36,7
	CR	14,2	18,0	22,0	25,9	28,1	28,5	29,3	30,1	32,8
9	G2	14,1	17,7	22,1	26,0	29,5	30,1	31,0	32,5	35,8
	Q2	13,8	17,0	21,5	25,4	27,9	29,5	30,3	31,5	34,8
	FT	14,4	18,1	22,8	27,7	30,6	31,1	32,6	33,6	38,7
	CR	13,9	17,1	21,5	25,5	28,2	29,5	30,5	31,6	34,7
10	G2	14,5	17,6	21,5	26,2	29,6	30,8	31,7	34,1	38,8
	Q2	14,3	17,2	21,0	25,4	28,3	29,8	31,1	31,6	37,0
	FT	14,8	18,0	22,4	26,9	31,1	31,6	33,7	37,7	40,7
	CR	14,3	17,2	21,2	25,6	28,4	29,9	30,9	31,8	37,3

**TABLA 3.** Valores Críticos para el Método de Patrones:  
5 variables manifiestas y 3 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	17,6	20,7	24,2	27,6	30,1	31,6	32,2	33,5	33,8
	Q2	14,8	18,1	21,8	25,7	29,4	29,9	30,7	32,6	35,6
	FT	27,4	32,4	37,8	43,1	47,5	48,3	49,9	51,2	53,2
	CR	14,8	17,8	21,0	24,3	27,5	27,9	28,7	29,6	31,7
2	G2	16,6	20,6	25,6	30,9	33,6	34,9	35,5	36,9	38,4
	Q2	14,2	17,4	21,7	26,9	29,5	29,8	32,1	32,9	36,2
	FT	23,3	30,1	37,6	45,7	50,8	51,3	52,3	54,7	59,9
	CR	14,4	17,8	21,9	26,8	29,3	29,9	31,2	32,9	34,4
3	G2	15,7	19,8	24,9	29,7	32,5	34,0	35,0	36,3	38,2
	Q2	14,3	18,2	22,1	26,7	28,9	29,5	30,6	32,1	33,6
	FT	19,0	25,2	32,8	40,9	45,4	45,9	47,4	50,5	54,8
	CR	14,6	18,3	22,2	26,7	28,8	29,6	30,5	31,7	33,6
4	G2	16,3	20,3	25,4	29,7	33,8	35,3	36,0	36,9	38,5
	Q2	15,3	18,8	23,3	27,6	30,6	31,2	32,4	34,5	36,9
	FT	17,5	23,2	31,1	38,2	42,1	44,1	47,5	49,1	51,5
	CR	15,4	19,0	23,4	27,5	30,7	31,5	32,5	34,4	37,0
5	G2	14,2	17,6	21,9	27,0	31,7	32,2	33,3	35,8	39,6
	Q2	13,5	16,8	21,1	25,8	29,6	30,5	31,4	33,6	36,8
	FT	14,7	18,8	24,3	31,8	36,7	39,2	41,7	43,7	45,2
	CR	13,6	16,8	21,1	25,7	29,8	30,4	32,0	33,1	36,9
6	G2	12,0	15,4	20,5	24,6	27,4	28,3	28,7	30,7	32,2
	Q2	11,7	14,8	19,1	23,3	25,5	26,1	27,4	27,9	29,1
	FT	12,3	15,9	21,9	27,9	30,6	32,2	33,2	36,4	42,1
	CR	11,8	15,0	19,3	23,5	25,6	26,3	27,5	28,2	29,2
7	G2	10,7	14,4	18,0	22,8	25,6	26,6	27,8	29,5	31,4
	Q2	10,6	14,0	17,2	21,6	24,6	25,3	26,5	27,7	30,7
	FT	11,0	14,9	18,8	24,3	28,2	29,0	31,3	32,6	34,8
	CR	10,6	14,1	17,4	21,6	24,3	25,4	26,5	28,3	30,3
8	G2	11,2	14,6	17,8	21,9	24,4	25,0	25,7	26,7	27,3
	Q2	11,0	14,2	17,3	21,2	24,0	24,6	25,2	25,8	26,8
	FT	11,4	14,9	18,4	22,8	25,5	25,9	26,7	27,7	29,3
	CR	10,9	14,3	17,4	21,3	24,1	24,5	25,6	26,0	26,9
9	G2	10,8	14,0	18,1	23,1	25,1	25,6	26,5	28,9	31,6
	Q2	10,6	13,6	17,4	22,5	24,4	25,3	26,4	28,4	30,3
	FT	11,0	14,4	18,6	24,1	26,8	27,1	27,7	31,9	34,3
	CR	10,6	13,7	17,4	22,5	24,4	25,3	26,6	28,2	30,5
10	G2	10,8	13,9	17,4	21,6	25,2	25,6	27,2	30,2	32,8
	Q2	10,6	13,5	17,1	21,1	24,0	25,6	26,4	29,7	31,7
	FT	11,0	14,1	17,7	22,3	25,4	26,9	28,8	30,9	34,7
	CR	10,6	13,6	17,0	21,2	24,3	25,4	26,6	29,7	31,6

**TABLA 4.** Valores Críticos para el Método de Patrones:  
6 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	51,7	57,3	63,0	68,2	71,0	72,2	73,3	74,6	76,8
	Q2	42,8	48,3	54,0	60,2	63,7	65,1	67,2	69,1	72,4
	FT	81,9	90,3	99,1	106,6	110,7	112,5	113,0	116,1	118,6
	CR	43,1	48,4	53,4	58,9	61,9	63,2	65,1	66,4	69,8
2	G2	51,5	58,0	66,4	73,0	77,3	78,3	79,9	82,3	86,9
	Q2	44,2	49,9	56,9	63,5	68,7	70,0	72,7	74,3	77,5
	FT	73,3	82,8	96,2	108,6	114,0	116,7	119,2	123,6	129,7
	CR	44,6	50,3	57,5	63,3	68,1	69,3	71,5	72,8	76,5
3	G2	45,1	53,1	61,5	69,5	73,7	75,0	78,1	79,1	83,7
	Q2	40,3	47,2	54,7	61,7	64,7	66,1	68,0	69,7	74,2
	FT	56,2	68,5	80,9	92,5	100,8	105,9	109,8	114,0	115,5
	CR	40,7	47,5	55,1	62,2	65,8	66,3	67,2	70,3	74,2
4	G2	43,3	50,1	57,8	64,1	68,0	69,1	71,0	73,8	76,4
	Q2	40,6	45,6	52,6	58,5	62,4	63,1	65,4	67,9	72,8
	FT	48,7	57,8	70,2	79,9	87,1	89,1	90,5	94,9	99,2
	CR	40,8	46,2	53,3	58,8	61,9	63,7	65,4	68,2	71,8
5	G2	42,7	48,8	55,2	63,7	66,0	67,0	69,2	71,7	76,0
	Q2	40,9	46,1	52,1	58,3	62,4	64,0	65,5	68,2	71,1
	FT	45,2	53,0	62,4	72,9	78,4	79,8	82,0	86,8	97,1
	CR	41,1	46,5	52,4	59,0	62,6	64,4	65,1	68,3	70,6
6	G2	42,2	48,6	56,0	63,5	67,7	68,7	69,8	72,1	75,1
	Q2	40,6	46,6	54,3	61,0	64,9	65,9	66,8	68,3	72,1
	FT	44,1	51,1	59,6	70,2	74,7	75,8	77,8	81,7	87,1
	CR	41,0	46,9	54,5	60,8	65,2	65,8	67,4	69,0	72,6
7	G2	41,8	47,6	54,3	62,4	68,3	69,0	70,7	74,1	78,3
	Q2	40,6	46,1	52,3	60,7	64,4	65,8	67,8	70,0	73,0
	FT	43,7	50,2	57,2	67,1	73,8	78,3	79,7	82,0	87,5
	CR	40,8	46,4	52,7	60,4	65,2	66,1	67,4	71,1	73,1
8	G2	40,5	47,6	53,7	61,5	65,2	68,5	68,7	74,6	78,4
	Q2	39,7	45,8	51,9	59,3	63,0	64,4	66,5	71,2	75,1
	FT	41,8	49,2	56,2	64,6	69,6	71,6	74,6	80,4	86,7
	CR	39,7	46,1	52,3	60,0	63,2	64,4	66,8	71,9	75,5
9	G2	41,0	47,6	55,2	62,2	67,1	69,0	70,0	72,2	76,3
	Q2	40,4	46,8	54,1	59,7	64,7	66,3	68,4	70,0	73,3
	FT	42,2	49,2	57,0	64,9	70,7	72,9	74,5	75,9	82,3
	CR	40,5	46,5	54,3	59,9	64,8	67,0	68,3	69,8	72,8
10	G2	40,9	47,4	54,4	61,8	65,0	66,1	67,5	69,9	74,2
	Q2	40,3	46,6	53,1	60,3	63,0	63,3	64,8	68,2	72,9
	FT	41,7	48,4	56,2	63,7	67,9	69,4	70,9	73,6	78,5
	CR	40,5	46,5	53,5	60,5	63,1	63,7	64,7	69,2	72,8

**TABLA 5.** Valores Críticos para el Método de Patrones:  
6 variables manifiestas y 3 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	50,8	56,1	62,0	67,9	71,9	73,0	74,5	75,9	79,3
	Q2	42,5	48,3	55,2	62,0	65,8	67,3	68,8	70,6	72,0
	FT	80,0	87,5	96,5	105,6	112,7	113,8	114,7	117,4	123,1
	CR	42,5	47,8	53,6	60,0	63,2	64,5	65,4	68,1	69,6
2	G2	48,2	54,2	62,5	68,9	74,2	75,3	76,5	79,1	82,9
	Q2	41,0	46,8	53,4	59,5	64,1	65,4	68,0	70,4	75,2
	FT	68,7	78,0	91,2	102,0	109,3	111,0	112,5	118,6	121,8
	CR	41,9	47,0	53,7	59,4	63,8	65,3	67,5	68,9	73,0
3	G2	42,3	48,6	56,1	63,7	67,0	67,8	69,3	72,5	74,4
	Q2	38,0	43,6	50,0	56,2	59,5	62,0	62,8	63,8	67,7
	FT	51,8	62,8	73,9	85,9	91,4	93,6	95,7	100,4	107,7
	CR	38,5	44,0	50,2	56,5	60,7	61,6	62,2	63,4	67,6
4	G2	41,2	47,9	56,0	61,3	66,4	67,4	69,0	70,9	73,3
	Q2	38,3	44,6	50,9	56,2	62,2	63,0	63,9	64,8	66,3
	FT	46,3	56,6	66,3	77,1	83,6	85,1	87,9	91,9	94,1
	CR	38,8	44,9	51,3	56,6	61,3	63,4	64,4	64,8	66,2
5	G2	37,5	43,2	49,1	55,9	59,9	61,2	63,1	65,0	69,8
	Q2	35,5	41,2	46,6	53,0	55,9	56,4	59,3	60,8	64,2
	FT	39,6	46,9	55,6	63,9	70,5	72,7	74,5	81,7	83,6
	CR	35,6	41,4	46,9	53,2	56,2	57,2	58,6	61,9	64,7
6	G2	41,6	47,9	56,0	62,0	67,0	68,1	69,4	70,4	72,1
	Q2	40,5	46,2	54,0	60,6	63,5	65,7	68,4	70,6	74,0
	FT	43,4	50,7	60,1	69,1	73,8	75,3	76,7	81,8	84,3
	CR	40,8	46,5	54,0	60,4	63,8	65,7	68,4	69,3	72,4
7	G2	35,3	41,0	47,3	53,8	59,0	61,2	62,4	64,9	68,7
	Q2	34,1	39,7	45,3	52,6	57,6	58,8	60,0	61,3	65,1
	FT	36,7	42,5	50,8	58,2	63,8	64,7	67,3	70,2	78,5
	CR	34,3	40,0	45,4	52,3	57,2	59,2	60,3	61,4	65,0
8	G2	36,5	41,9	47,9	54,4	58,0	60,8	61,7	64,4	68,3
	Q2	35,4	40,7	46,4	52,7	56,4	57,3	59,1	59,9	64,3
	FT	37,5	43,3	49,6	56,6	61,1	65,3	67,2	71,1	77,6
	CR	35,6	40,8	46,5	52,8	56,5	58,4	59,5	60,8	65,0
9	G2	34,6	40,0	46,1	52,1	57,5	58,5	59,2	60,6	63,9
	Q2	33,9	38,9	45,2	50,8	55,0	55,6	56,7	58,2	61,9
	FT	35,4	41,0	47,8	54,4	60,4	62,0	63,3	64,2	68,5
	CR	34,1	39,1	45,2	51,1	55,4	55,8	57,1	58,5	62,0
10	G2	36,3	41,9	48,8	55,4	59,1	60,1	61,5	62,6	65,7
	Q2	35,4	41,3	48,0	54,2	58,4	59,2	60,5	60,7	63,3
	FT	37,1	42,9	50,2	57,6	61,3	62,0	63,8	64,6	71,0
	CR	35,6	41,5	47,8	54,4	58,2	59,1	60,2	60,9	63,0



**TABLA 6.** Valores Críticos para el Método de Patrones:  
6 variables manifiestas y 4 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	49,5	54,6	59,9	66,3	69,3	70,3	71,6	73,6	74,7
	Q2	41,7	46,9	52,6	60,0	63,1	64,5	66,6	69,3	70,9
	FT	78,2	86,5	94,8	102,7	109,0	111,3	111,9	113,2	115,7
	CR	41,8	46,5	51,7	57,6	61,0	61,5	63,9	65,0	68,3
2	G2	47,2	54,3	61,4	68,0	70,9	73,5	75,1	78,2	82,1
	Q2	40,2	46,3	53,1	60,4	64,8	65,6	68,2	69,0	71,2
	FT	66,9	78,8	90,1	100,8	107,2	109,6	112,0	116,7	124,3
	CR	40,7	46,7	53,4	59,6	63,2	64,0	66,5	67,7	71,9
3	G2	35,9	43,0	49,6	56,0	60,3	61,1	62,2	65,1	67,8
	Q2	31,9	37,9	44,0	49,7	53,0	54,7	56,1	58,4	60,7
	FT	44,6	54,1	65,7	74,5	83,1	85,0	88,4	89,6	94,6
	CR	32,5	38,3	44,6	50,5	53,7	54,8	56,1	58,6	61,0
4	G2	34,5	40,6	47,3	52,1	56,1	57,8	58,6	61,3	64,5
	Q2	32,3	37,6	43,1	48,9	51,8	52,6	53,5	55,7	57,9
	FT	38,3	46,7	56,2	65,5	70,0	71,3	74,5	76,9	82,4
	CR	32,4	38,0	43,5	49,0	51,6	53,3	54,3	55,9	58,7
5	G2	34,2	40,0	46,6	52,5	57,8	59,0	60,2	61,5	64,4
	Q2	32,8	38,2	43,7	49,2	53,3	55,0	57,2	58,1	60,9
	FT	35,9	43,0	51,3	61,3	66,9	69,0	69,8	73,6	79,8
	CR	32,9	38,2	43,9	49,5	53,9	55,1	57,1	58,2	60,3
6	G2	30,8	35,9	42,4	49,6	53,5	54,3	55,9	58,6	60,1
	Q2	29,8	34,9	40,4	46,5	50,4	51,1	52,0	54,5	56,7
	FT	31,7	37,6	46,3	55,0	60,2	62,0	64,8	67,0	70,0
	CR	30,0	34,9	40,7	47,2	50,7	51,5	52,9	54,9	56,7
7	G2	30,6	36,4	43,0	50,2	53,0	53,8	55,3	56,7	60,2
	Q2	29,7	35,6	41,3	48,1	51,1	52,2	53,1	55,3	58,4
	FT	31,3	37,7	45,3	54,2	58,9	60,0	60,7	62,7	63,8
	CR	29,7	35,8	41,6	48,5	51,4	52,4	53,5	55,1	58,4
8	G2	28,0	33,4	38,9	44,8	48,7	50,3	51,3	53,4	58,1
	Q2	27,3	32,7	37,7	43,3	46,2	47,6	50,1	51,7	55,7
	FT	28,5	34,4	40,4	47,4	51,9	52,7	56,2	59,4	64,1
	CR	27,6	32,8	38,0	43,4	46,3	48,3	49,8	51,8	56,0
9	G2	28,3	33,2	38,8	45,5	49,2	50,3	51,5	53,3	55,7
	Q2	27,8	32,5	37,7	43,3	47,5	48,8	49,9	51,3	54,9
	FT	29,0	33,9	40,0	47,5	52,0	52,9	54,2	57,0	60,2
	CR	27,7	32,4	37,8	43,9	47,7	48,7	49,6	51,4	54,8
10	G2	29,4	35,3	42,2	49,1	52,1	53,1	55,0	57,5	59,7
	Q2	29,1	34,8	41,4	48,0	51,2	52,4	54,1	56,1	58,2
	FT	30,2	35,8	43,2	50,5	53,8	54,8	55,6	59,9	63,4
	CR	29,1	34,8	41,3	48,1	51,2	52,3	53,8	55,7	58,2

**TABLA 7.** Valores Críticos para el Método de Patrones:  
7 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	120,1	128,4	136,6	143,2	148,9	150,9	151,9	153,8	156,3
	Q2	101,0	110,2	118,9	130,0	135,5	136,7	138,5	141,5	149,8
	FT	190,6	203,0	216,5	225,3	231,5	234,2	238,2	240,0	244,5
	CR	100,9	109,1	117,1	125,5	131,5	132,4	133,4	135,8	141,7
2	G2	116,6	127,9	138,9	149,8	155,6	156,8	158,8	160,9	164,0
	Q2	99,7	110,1	120,7	130,1	135,3	137,1	139,4	141,7	145,8
	FT	167,2	184,8	203,6	221,2	230,4	233,3	238,6	243,0	246,8
	CR	101,1	110,9	121,0	130,2	135,1	136,4	138,2	140,4	144,4
3	G2	114,4	123,8	135,0	146,8	152,2	155,0	157,2	159,2	161,1
	Q2	104,3	112,7	122,4	131,7	139,2	141,9	142,5	143,8	147,6
	FT	141,0	158,5	176,4	195,3	205,3	208,1	211,7	218,8	223,1
	CR	105,4	113,4	123,3	131,9	139,2	140,5	142,9	143,3	146,0
4	G2	103,3	114,4	126,2	135,2	141,0	142,7	144,7	147,0	151,8
	Q2	97,3	106,3	116,9	125,3	131,3	134,6	135,7	140,1	142,0
	FT	117,9	134,2	150,1	164,4	171,9	175,1	176,5	184,4	191,9
	CR	97,4	106,6	117,0	125,9	132,1	133,9	135,7	137,9	140,6
5	G2	103,9	112,1	122,3	134,2	140,6	143,1	145,1	148,0	153,9
	Q2	98,8	107,0	116,8	126,4	132,4	133,6	137,4	142,2	149,8
	FT	111,6	124,4	138,3	155,4	163,7	167,1	169,3	171,7	177,4
	CR	99,3	107,3	117,2	126,5	131,8	134,2	138,0	141,7	148,3
6	G2	103,3	114,9	124,1	135,5	142,9	144,6	146,9	151,2	155,1
	Q2	99,4	110,0	119,5	129,3	136,9	139,6	144,1	144,7	152,5
	FT	109,2	122,2	135,2	150,0	158,0	160,9	164,8	168,9	178,3
	CR	99,8	110,5	119,2	129,7	137,0	139,7	142,8	145,7	149,7
7	G2	101,9	111,3	122,5	132,5	138,6	141,3	145,2	148,4	151,0
	Q2	98,7	107,2	118,1	127,8	133,9	136,0	138,4	141,5	145,1
	FT	106,8	117,2	129,9	141,2	150,1	153,2	154,5	160,3	167,4
	CR	99,3	107,8	118,5	128,0	133,7	134,9	139,3	141,7	144,4
8	G2	106,0	116,5	127,3	140,2	145,6	146,6	148,9	154,5	158,5
	Q2	103,6	113,5	124,2	135,5	142,2	143,7	145,5	152,0	154,2
	FT	109,3	121,3	132,7	147,2	154,9	157,4	161,2	163,0	166,7
	CR	103,9	113,5	123,8	136,5	141,5	142,6	145,9	149,2	154,0
9	G2	102,8	112,7	123,4	133,4	137,9	139,4	140,9	145,1	149,4
	Q2	100,4	110,5	120,4	129,6	135,4	136,4	139,1	140,0	143,2
	FT	105,8	116,6	127,9	138,7	143,3	145,9	149,6	153,3	157,5
	CR	100,3	110,6	120,8	129,5	135,0	136,8	139,4	140,2	143,1
10	G2	105,7	116,6	127,0	137,2	142,3	145,1	145,5	147,9	152,0
	Q2	103,3	114,4	124,1	134,0	139,3	140,4	142,5	144,9	147,8
	FT	108,6	120,3	131,7	142,9	149,5	151,3	154,1	157,4	162,8
	CR	103,6	114,6	123,9	133,9	139,9	140,4	142,1	144,2	147,0

**TABLA 8.** Valores Críticos para el Método de Patrones:  
7 variables manifiestas y 3 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	109,4	120,1	130,8	141,2	145,3	146,4	149,3	153,5	157,6
	Q2	93,1	102,3	113,1	122,3	127,8	129,2	130,9	133,2	137,0
	FT	157,0	174,6	191,0	207,0	218,2	220,9	223,4	229,8	234,5
	CR	94,3	103,2	113,8	122,2	127,0	129,3	130,2	131,7	136,4
2	G2	108,7	118,0	130,5	139,5	145,1	146,6	150,1	154,3	156,5
	Q2	93,5	101,8	112,6	121,2	126,7	130,0	132,3	133,6	137,1
	FT	154,5	171,7	190,3	205,7	214,1	219,1	222,5	228,1	237,3
	CR	94,3	102,9	113,3	121,2	126,7	129,4	130,8	133,5	137,0
3	G2	99,2	110,5	120,5	131,7	139,3	142,6	144,2	147,4	151,2
	Q2	89,8	98,0	108,3	118,5	125,2	128,0	128,7	134,4	139,2
	FT	125,6	142,2	159,4	176,1	186,5	192,3	196,4	201,8	204,0
	CR	90,4	99,1	109,1	118,8	125,2	126,9	129,6	133,5	137,5
4	G2	100,1	110,5	122,8	133,3	141,2	143,0	145,0	151,5	162,5
	Q2	93,4	103,5	112,7	124,5	131,2	133,0	135,7	139,9	148,4
	FT	114,1	130,5	147,8	163,5	172,0	175,7	182,7	185,5	202,4
	CR	94,6	103,9	113,4	124,4	132,3	133,6	135,3	139,0	148,7
5	G2	102,2	112,0	122,9	132,2	138,4	139,9	142,6	147,5	153,1
	Q2	97,1	107,3	117,6	125,1	129,7	132,0	134,6	138,4	143,5
	FT	109,5	123,3	138,6	153,4	165,3	169,0	172,3	177,1	182,2
	CR	97,2	107,7	117,4	125,3	130,3	130,6	134,9	136,5	143,7
6	G2	101,1	112,4	121,7	131,0	135,4	137,4	139,4	142,9	147,0
	Q2	98,0	108,0	116,7	125,0	130,8	132,1	133,6	135,8	141,5
	FT	106,6	120,5	132,2	145,6	151,2	152,6	156,9	159,9	163,2
	CR	98,0	108,6	116,5	125,4	131,1	132,0	133,7	136,0	140,4
7	G2	95,3	105,1	114,7	124,5	131,1	133,3	136,2	137,6	143,7
	Q2	92,9	101,5	111,2	120,6	127,4	128,4	130,2	130,8	135,6
	FT	98,5	109,6	121,3	133,5	140,8	143,5	145,0	152,7	154,9
	CR	92,9	101,8	111,5	121,0	127,1	128,7	129,5	130,9	138,2
8	G2	93,4	103,5	115,2	123,2	129,9	131,4	133,4	135,8	140,0
	Q2	91,2	100,7	111,5	120,3	125,1	128,1	130,0	132,3	134,7
	FT	96,7	107,7	119,7	130,6	137,6	138,8	143,7	145,1	150,8
	CR	91,4	101,1	111,9	120,2	125,7	127,9	129,3	131,9	135,1
9	G2	96,1	106,6	116,4	124,9	133,0	134,3	135,6	137,3	139,7
	Q2	93,7	104,0	113,6	121,6	129,2	130,8	132,9	134,7	137,1
	FT	98,9	110,1	121,0	131,2	139,9	140,7	142,3	145,7	149,6
	CR	94,0	104,2	113,7	121,8	129,1	130,7	132,3	134,2	136,5
10	G2	93,6	102,4	111,8	121,0	127,2	129,1	130,4	132,7	133,9
	Q2	91,5	100,3	109,8	118,8	124,2	125,4	125,9	129,4	130,7
	FT	96,1	105,3	115,7	126,7	133,5	135,4	137,0	138,2	139,6
	CR	91,9	100,5	109,9	119,0	124,1	125,8	126,9	129,1	130,8

**TABLA 9.** Valores Críticos para el Método de Patrones:  
7 variables manifiestas y 4 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	114,3	122,1	130,8	139,5	143,3	145,0	146,4	149,1	152,1
	Q2	97,4	105,4	116,9	128,7	134,7	136,1	138,1	141,2	148,1
	FT	180,3	193,1	205,0	217,4	223,4	223,7	226,6	229,6	236,9
	CR	97,0	103,9	113,8	123,7	128,6	129,9	131,6	132,9	136,3
2	G2	107,4	116,1	126,1	135,7	142,7	144,1	146,0	148,0	152,6
	Q2	91,2	99,4	109,0	118,8	123,9	125,5	126,7	127,9	130,9
	FT	153,6	169,1	185,0	199,8	209,7	212,5	215,6	221,9	226,8
	CR	92,6	100,5	109,2	118,6	123,0	125,1	127,4	128,1	130,7
3	G2	95,9	106,7	116,0	127,4	136,7	137,7	141,6	144,7	149,6
	Q2	87,3	95,9	104,1	114,3	122,4	124,4	127,9	130,9	133,3
	FT	120,1	135,3	153,1	169,3	182,3	187,2	193,1	198,2	208,7
	CR	88,0	96,9	104,5	114,6	122,2	125,4	127,3	129,1	133,4
4	G2	91,5	100,9	111,3	121,3	127,4	130,7	132,3	134,4	138,1
	Q2	85,2	93,4	103,0	110,7	118,1	120,1	123,0	126,0	127,8
	FT	104,0	116,7	132,1	149,5	156,1	157,7	162,7	165,6	171,1
	CR	85,7	94,2	103,2	111,6	118,3	120,9	123,4	125,5	128,6
5	G2	89,6	100,0	110,0	119,6	126,6	128,2	130,7	133,9	137,8
	Q2	85,5	94,3	103,8	112,0	118,3	119,8	122,5	123,9	125,5
	FT	96,2	110,0	124,7	139,9	148,4	151,7	154,6	158,7	166,0
	CR	86,1	95,0	104,4	112,3	118,5	120,1	122,6	124,4	127,4
6	G2	85,4	94,7	105,7	116,7	121,0	123,0	124,4	127,9	135,1
	Q2	82,3	91,3	101,3	111,4	118,2	120,6	122,4	124,8	128,8
	FT	89,7	102,6	115,5	125,6	132,8	134,4	139,3	142,6	144,8
	CR	82,9	91,8	101,7	112,2	117,4	119,7	121,5	124,0	129,4
7	G2	88,2	98,3	107,9	118,2	127,6	130,4	133,2	136,4	138,9
	Q2	85,5	95,7	104,2	114,8	123,9	125,8	128,0	129,7	134,8
	FT	91,8	103,3	114,2	126,6	139,3	142,0	144,1	147,5	149,7
	CR	85,9	95,7	104,1	114,1	123,6	125,2	128,8	130,9	135,2
8	G2	84,7	92,8	102,9	112,7	117,3	120,3	122,9	126,5	128,4
	Q2	82,3	90,5	99,1	108,7	116,2	116,9	119,7	121,7	124,3
	FT	87,5	96,1	107,3	119,0	125,7	127,8	131,4	136,7	142,1
	CR	82,9	90,7	99,6	109,2	115,6	117,5	120,1	121,6	124,1
9	G2	85,6	94,2	104,4	112,5	116,9	118,3	121,0	122,6	126,4
	Q2	83,8	91,9	101,6	110,0	114,1	114,5	116,4	118,8	120,9
	FT	88,7	97,2	107,7	116,5	122,9	125,0	127,6	129,0	132,1
	CR	84,0	91,8	102,0	110,3	113,6	114,9	116,0	119,2	121,7
10	G2	82,9	91,9	103,5	109,4	113,5	114,6	117,0	123,1	126,8
	Q2	80,7	90,3	100,5	106,9	110,6	113,1	114,1	119,8	123,0
	FT	84,5	94,1	107,0	114,0	118,4	119,9	124,2	128,3	132,5
	CR	80,8	90,5	100,9	107,3	111,2	112,3	114,7	120,0	123,7

**TABLA 10.** Valores Críticos para el Método de Patrones:  
7 variables manifiestas y 5 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	97,4	104,0	111,8	119,1	124,2	125,4	125,9	126,6	130,6
	Q2	84,7	92,5	100,7	109,8	115,0	116,6	118,6	120,5	123,6
	FT	134,5	143,6	154,6	165,5	170,8	171,7	172,8	176,9	181,6
	CR	84,4	91,1	98,5	107,0	111,0	112,3	112,9	115,3	117,6
2	G2	92,0	101,9	111,9	121,0	125,5	126,0	128,3	131,3	133,9
	Q2	78,6	86,4	96,2	107,0	110,3	111,6	112,7	114,3	115,7
	FT	122,6	136,2	151,9	163,8	168,5	170,7	173,7	176,9	187,4
	CR	79,6	88,1	96,7	107,0	110,3	112,4	113,1	114,7	115,4
3	G2	87,8	96,8	106,7	116,8	122,8	124,1	127,0	131,4	138,3
	Q2	78,6	86,1	95,2	103,8	109,5	110,4	114,2	117,4	119,7
	FT	106,2	119,0	134,4	149,7	157,3	159,7	162,4	172,3	181,1
	CR	79,9	87,1	96,3	104,8	109,6	111,6	114,2	117,7	121,4
4	G2	82,4	91,1	101,6	111,6	116,9	117,7	119,4	122,7	126,8
	Q2	77,3	85,1	94,2	103,3	109,3	111,2	112,2	114,9	116,7
	FT	92,6	104,4	118,9	133,0	138,5	140,4	142,2	147,6	151,0
	CR	77,8	85,7	94,8	103,2	109,1	110,2	112,7	114,4	117,4
5	G2	79,0	87,0	96,7	106,2	110,8	113,1	116,6	118,6	126,7
	Q2	75,3	82,6	91,9	99,4	103,5	105,3	106,8	110,8	119,6
	FT	84,3	95,5	106,9	119,0	127,4	131,7	135,8	138,6	147,3
	CR	75,8	83,0	92,2	100,3	104,1	106,2	107,8	112,3	120,7
6	G2	82,6	90,6	100,9	111,9	116,6	118,9	120,9	124,7	129,5
	Q2	79,0	87,7	96,6	106,3	110,0	111,8	115,4	118,0	126,3
	FT	86,5	96,8	109,3	122,7	128,7	132,1	137,0	138,9	146,3
	CR	79,3	87,5	97,3	107,0	110,7	111,6	115,5	117,4	125,9
7	G2	77,9	88,1	97,8	107,2	113,8	115,4	116,4	117,3	120,1
	Q2	76,7	85,8	94,5	102,6	107,8	110,9	112,8	115,4	117,3
	FT	80,4	92,2	102,6	114,8	121,2	122,3	125,0	128,6	131,3
	CR	76,9	86,0	94,8	103,0	109,4	111,5	112,3	114,4	117,0
8	G2	77,5	86,4	95,2	104,4	110,2	111,2	111,8	114,4	117,7
	Q2	76,2	83,7	92,7	101,1	106,3	107,2	108,2	111,0	113,5
	FT	79,9	89,8	99,4	110,5	116,3	117,6	118,7	122,8	125,3
	CR	76,5	84,1	93,1	101,1	106,9	107,7	108,2	112,1	114,5
9	G2	76,8	84,9	95,3	102,9	108,2	109,7	110,6	113,7	117,0
	Q2	75,9	83,3	92,6	100,5	105,0	106,8	108,6	109,7	114,3
	FT	78,6	87,8	98,4	107,2	113,4	114,3	117,1	119,8	123,3
	CR	76,0	83,3	93,0	100,5	105,6	107,6	108,5	109,0	116,0
10	G2	82,1	90,8	99,3	107,7	113,7	114,4	116,7	119,4	122,4
	Q2	80,6	89,0	97,0	105,4	109,3	112,3	113,4	115,1	121,6
	FT	84,2	93,0	102,2	111,5	118,3	120,5	122,8	124,7	128,3
	CR	80,9	89,0	97,4	105,5	110,8	112,2	113,0	116,4	121,1

**TABLA 11.** Valores Críticos para el Método de Patrones:  
8 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	260,5	273,9	284,4	296,2	303,1	304,7	305,8	309,6	313,1
	Q2	221,2	235,4	250,7	263,1	271,3	273,5	277,9	280,5	290,8
	FT	413,9	432,3	447,8	464,0	473,2	476,4	480,2	483,4	493,4
	CR	220,2	232,8	245,3	257,2	263,8	265,7	268,7	271,9	276,1
2	G2	256,9	273,5	287,7	302,6	317,1	319,5	321,2	328,4	332,6
	Q2	223,3	237,1	252,2	265,1	278,8	282,0	283,9	285,5	287,4
	FT	368,0	395,2	419,2	446,7	460,0	465,6	470,6	485,4	497,3
	CR	224,5	238,4	252,1	264,0	277,8	280,2	282,8	285,2	288,7
3	G2	243,1	257,8	274,4	290,2	303,8	304,8	307,6	314,6	319,0
	Q2	221,7	235,1	248,3	264,4	274,5	277,5	281,3	286,2	292,3
	FT	304,9	331,3	357,2	383,3	403,1	404,6	410,3	416,4	426,5
	CR	222,3	235,6	248,9	265,4	274,1	277,3	282,2	285,1	290,5
4	G2	233,3	250,0	267,7	282,7	295,7	297,7	300,3	309,4	315,7
	Q2	218,7	234,5	249,8	263,0	273,1	275,7	280,0	285,5	291,0
	FT	269,1	291,1	319,1	345,0	362,3	365,6	376,6	383,4	386,2
	CR	219,4	234,8	250,6	262,7	272,5	278,5	281,3	286,3	290,8
5	G2	233,8	249,9	264,6	278,1	293,1	294,7	296,4	299,3	302,8
	Q2	222,5	237,0	251,6	264,5	274,3	277,9	279,4	284,7	289,4
	FT	256,0	275,7	297,1	318,1	333,7	337,9	347,8	351,8	366,5
	CR	223,9	237,5	252,1	265,5	275,2	278,3	279,9	282,9	288,6
6	G2	229,4	243,7	257,8	272,6	280,9	283,3	287,4	291,7	298,9
	Q2	221,5	234,8	248,7	263,0	269,5	273,8	275,1	278,9	287,6
	FT	242,9	260,6	279,9	299,5	308,9	311,7	316,0	329,3	337,7
	CR	221,9	235,2	248,6	262,8	270,3	271,3	274,6	280,2	288,3
7	G2	230,0	245,1	259,4	272,6	282,1	283,7	287,7	289,9	294,5
	Q2	224,1	237,5	251,1	263,2	272,5	274,4	278,4	284,0	287,3
	FT	240,8	257,1	275,4	292,8	304,2	306,5	312,0	314,6	321,1
	CR	224,0	238,2	252,1	262,7	271,7	275,5	278,0	283,5	286,6
8	G2	223,4	238,3	251,6	266,0	272,4	274,7	277,0	281,7	286,7
	Q2	217,4	231,4	244,5	259,2	267,8	268,5	271,1	273,8	281,8
	FT	232,0	248,4	264,6	278,5	288,3	289,9	296,2	299,9	302,6
	CR	217,7	231,9	245,0	259,3	267,2	269,2	270,7	273,5	279,4
9	G2	222,5	236,4	254,6	270,0	276,7	279,2	283,1	287,4	295,4
	Q2	218,5	231,5	249,8	262,0	270,5	271,5	275,5	285,4	293,0
	FT	228,7	243,9	264,2	281,3	290,8	294,6	296,1	300,6	307,3
	CR	218,3	232,4	249,7	261,4	270,2	271,4	273,9	284,5	291,4
10	G2	225,6	241,8	256,4	270,5	277,5	280,0	284,6	288,3	292,4
	Q2	222,4	236,7	251,9	264,4	273,2	276,4	281,3	282,9	287,9
	FT	232,8	248,6	264,9	281,1	288,6	289,9	294,2	297,6	305,2
	CR	222,2	236,8	251,6	265,3	273,5	275,7	277,8	283,5	287,5

**TABLA 12.** Valores Críticos para el Método de Patrones:  
8 variables manifiestas y 3 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	260,1	272,3	284,9	292,6	296,6	299,2	303,0	306,2	313,3
	Q2	221,9	235,7	249,9	262,3	269,7	271,3	273,5	277,5	285,6
	FT	414,5	431,4	448,5	459,2	464,4	468,9	474,8	480,7	487,9
	CR	220,8	232,7	245,9	255,3	260,6	262,0	264,3	269,8	275,6
2	G2	251,2	265,5	281,3	295,5	304,1	306,2	308,8	310,7	316,1
	Q2	218,4	231,1	244,6	257,8	265,6	268,1	269,9	276,8	281,3
	FT	359,6	384,3	411,4	434,7	446,8	452,8	455,8	458,1	459,8
	CR	219,5	231,3	244,9	257,1	265,6	267,2	270,0	273,9	278,1
3	G2	236,9	251,7	264,8	280,1	292,4	295,2	300,2	303,4	307,0
	Q2	214,8	226,8	240,4	254,2	262,4	264,1	268,2	274,8	281,1
	FT	298,3	324,4	345,9	367,8	388,8	393,6	399,4	404,9	415,5
	CR	215,9	228,1	240,5	255,3	262,2	266,6	270,4	274,4	280,0
4	G2	225,4	242,2	258,8	272,7	285,2	289,5	292,1	296,6	299,8
	Q2	212,4	226,9	241,1	254,8	263,9	265,8	270,3	273,3	277,0
	FT	259,0	281,7	306,3	332,2	351,5	359,3	365,1	373,6	381,2
	CR	213,2	227,9	241,6	254,6	264,6	268,2	271,8	273,1	276,2
5	G2	219,9	236,6	251,5	264,6	272,9	277,3	279,5	286,2	293,9
	Q2	209,5	225,0	238,9	251,4	260,5	263,0	266,1	269,3	275,8
	FT	239,5	260,9	282,2	303,0	316,3	320,0	323,5	327,5	336,0
	CR	210,5	225,4	239,6	251,3	260,0	262,1	266,0	269,3	278,2
6	G2	223,0	235,1	253,0	264,6	271,5	275,4	277,1	283,0	298,5
	Q2	214,2	226,0	242,1	254,0	262,5	265,3	266,7	273,1	277,6
	FT	234,6	253,1	273,8	289,8	297,4	306,8	311,3	318,6	339,1
	CR	215,1	226,7	243,2	254,4	261,5	264,1	266,2	272,1	278,2
7	G2	218,4	232,2	247,9	260,4	271,3	274,1	280,3	283,4	291,3
	Q2	212,2	225,3	239,7	252,6	262,1	266,2	270,3	277,5	284,1
	FT	227,1	244,0	262,6	277,8	291,5	295,6	300,5	303,4	312,8
	CR	212,5	225,5	240,4	253,1	263,4	265,7	270,9	275,5	282,3
8	G2	216,7	231,2	244,6	255,7	265,3	267,8	271,0	273,5	277,0
	Q2	211,8	225,0	238,7	251,2	258,4	260,4	263,5	265,1	270,2
	FT	224,1	240,6	255,5	269,9	279,4	281,9	283,8	288,9	296,9
	CR	212,1	225,6	238,7	250,6	258,9	261,6	262,8	265,2	270,1
9	G2	214,7	232,2	245,5	261,7	270,3	271,7	274,9	278,2	284,5
	Q2	210,3	226,5	241,0	254,6	264,8	266,8	268,9	270,6	280,5
	FT	222,5	240,0	255,8	274,3	281,5	282,4	285,0	290,2	300,4
	CR	210,4	226,8	241,1	255,4	264,5	266,6	268,5	271,5	279,5
10	G2	216,1	229,0	245,0	259,1	266,3	268,7	271,1	275,2	278,6
	Q2	211,6	225,4	239,7	252,6	262,0	265,1	267,6	270,3	274,1
	FT	222,0	236,3	254,2	267,9	276,2	278,5	280,8	285,0	289,5
	CR	211,5	225,6	240,5	252,8	261,0	265,0	266,6	269,8	273,3

**TABLA 13.** Valores Críticos para el Método de Patrones:  
8 variables manifiestas y 4 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	233,5	244,7	255,4	266,2	271,5	273,8	276,9	280,4	282,7
	Q2	198,9	211,6	223,4	237,7	244,4	245,7	246,9	249,1	252,3
	FT	337,5	352,9	367,2	382,4	390,7	393,4	395,6	401,7	405,6
	CR	198,7	210,3	220,8	232,2	238,0	239,2	241,7	244,9	245,8
2	G2	226,8	240,4	255,5	268,2	279,8	281,0	284,7	288,0	295,0
	Q2	193,8	208,0	220,6	233,9	245,1	248,7	250,2	252,8	259,4
	FT	306,6	327,9	352,7	371,2	385,9	390,1	394,0	400,5	414,5
	CR	196,5	209,5	221,3	234,3	244,6	248,1	249,8	252,3	257,4
3	G2	222,7	237,2	251,1	264,9	273,0	275,8	279,6	281,6	291,2
	Q2	201,5	213,6	225,6	238,9	247,5	251,7	252,9	257,6	263,6
	FT	273,9	295,4	318,5	340,3	351,9	354,8	360,2	364,3	368,3
	CR	202,8	214,9	227,5	240,5	247,8	251,0	252,7	256,7	263,5
4	G2	210,7	225,2	240,9	255,2	262,5	265,5	269,4	273,8	281,9
	Q2	198,0	210,4	223,7	236,5	244,9	246,2	250,8	251,8	260,7
	FT	238,1	258,8	281,2	301,9	317,1	320,3	323,2	330,7	346,2
	CR	198,9	211,3	225,4	238,5	244,2	247,1	249,5	253,7	263,9
5	G2	210,3	225,5	240,1	256,1	264,5	266,0	270,3	273,5	281,4
	Q2	200,1	214,6	227,6	240,6	247,8	249,5	257,4	260,6	265,9
	FT	229,9	249,0	266,8	287,5	300,3	305,8	311,3	317,7	327,0
	CR	201,9	214,8	228,7	241,6	248,7	252,5	254,5	260,5	264,6
6	G2	209,6	224,5	239,9	252,5	260,1	263,8	266,3	270,0	273,8
	Q2	201,5	215,3	229,8	243,5	251,0	252,2	254,2	259,5	263,8
	FT	221,7	239,9	258,3	274,9	284,2	285,7	292,3	294,7	302,7
	CR	202,1	216,1	230,2	243,2	250,9	252,6	255,0	260,8	263,2
7	G2	202,8	221,4	232,2	243,5	250,8	255,1	258,2	263,1	270,4
	Q2	197,5	213,9	224,6	235,8	244,5	247,3	250,1	254,7	256,6
	FT	213,2	231,6	245,9	261,5	269,4	271,3	274,5	279,7	289,7
	CR	198,0	214,9	225,6	236,3	244,1	246,7	249,5	256,2	257,7
8	G2	204,0	217,3	231,8	245,2	251,3	252,9	254,3	259,7	265,0
	Q2	198,7	211,2	225,0	237,7	245,0	247,2	249,8	253,3	262,5
	FT	211,4	226,9	241,7	257,4	264,7	267,5	270,8	274,9	280,8
	CR	199,4	211,7	225,0	238,4	245,0	247,0	249,1	253,5	259,6
9	G2	203,1	216,1	229,9	244,8	252,3	255,7	259,6	264,6	268,9
	Q2	197,8	211,1	224,9	238,5	245,1	247,9	252,1	257,6	267,2
	FT	209,4	223,2	240,1	256,2	264,1	271,3	273,9	276,5	280,5
	CR	198,3	211,6	224,9	238,9	246,1	247,1	251,8	257,9	265,0
10	G2	204,9	219,4	233,9	247,0	253,8	255,4	257,5	262,2	269,5
	Q2	200,9	214,7	229,0	240,4	247,9	249,1	250,4	254,3	268,0
	FT	210,4	226,3	241,5	255,8	263,7	266,4	271,4	275,2	281,1
	CR	201,2	215,2	229,3	241,3	248,3	250,1	251,4	253,4	266,8



**TABLA 14.** Valores Críticos para el Método de Patrones:  
8 variables manifiestas y 5 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	229,0	238,7	249,0	258,4	265,3	266,4	267,8	270,3	275,9
	Q2	196,2	206,7	218,5	229,9	239,2	240,3	242,9	246,8	249,5
	FT	329,2	344,2	358,0	372,6	378,2	381,3	385,8	388,8	393,0
	CR	195,5	205,0	215,0	225,3	233,5	234,5	236,4	238,3	241,1
2	G2	224,9	239,6	255,2	269,3	279,1	282,8	285,1	287,9	295,4
	Q2	194,5	207,3	221,5	235,1	247,2	248,6	251,5	254,9	259,5
	FT	307,6	329,1	353,3	375,6	387,2	392,0	397,8	404,3	411,1
	CR	196,8	209,0	222,4	235,3	246,2	248,3	251,4	252,5	255,4
3	G2	209,4	222,1	240,0	252,9	264,9	265,9	267,8	269,9	273,9
	Q2	189,7	200,4	215,4	228,8	239,3	240,4	241,5	243,7	247,0
	FT	256,8	276,9	302,2	325,3	335,1	340,8	347,8	350,3	356,8
	CR	191,3	202,1	216,9	229,2	237,4	241,5	242,3	243,1	246,8
4	G2	207,0	221,8	237,7	253,4	261,5	262,9	264,3	268,8	280,7
	Q2	193,2	207,0	221,0	235,3	241,2	243,0	245,1	249,8	255,3
	FT	234,3	253,6	277,2	299,8	315,0	318,8	322,1	327,1	333,0
	CR	194,4	208,0	222,2	235,8	242,0	244,1	245,6	248,3	259,5
5	G2	197,3	211,2	224,6	238,0	245,3	246,2	249,9	254,0	260,0
	Q2	186,3	200,0	212,9	224,7	230,5	232,7	235,9	238,7	248,0
	FT	214,6	233,4	249,8	269,6	279,5	281,9	285,4	292,1	301,5
	CR	187,4	200,9	213,8	224,9	231,6	233,4	235,5	239,9	247,7
6	G2	199,4	213,6	226,3	241,4	247,1	247,7	249,5	255,7	264,6
	Q2	192,4	204,1	218,2	231,6	237,0	238,5	240,0	246,7	253,2
	FT	210,4	228,3	244,3	263,5	269,0	270,9	274,2	278,2	291,7
	CR	192,8	204,9	218,2	232,2	237,5	238,7	239,2	245,5	253,7
7	G2	197,8	209,4	224,7	236,2	241,8	242,9	245,7	248,3	254,6
	Q2	191,4	202,3	216,9	227,4	232,2	234,5	238,1	241,4	250,8
	FT	206,6	220,3	237,6	252,2	260,6	262,4	265,5	267,4	271,5
	CR	192,7	202,9	217,5	227,1	233,4	234,7	236,7	241,0	248,8
8	G2	190,32	203,43	216,66	227,42	235,77	237,01	239,23	241,47	245,9
	Q2	185,02	197,92	211,16	221,79	228,87	230,22	234,05	237,34	242,63
	FT	198,17	212,36	226,88	237,8	246,49	248,74	251,07	255,06	261,81
	CR	185,63	198,5	211,43	222,11	228,81	230,76	233,49	237,36	241,23
9	G2	197,3	210,1	224,2	237,0	244,2	247,1	252,1	259,2	264,4
	Q2	192,7	206,0	218,8	230,0	238,9	243,7	247,6	250,9	264,1
	FT	203,4	217,4	232,4	247,0	254,3	258,3	265,7	268,7	277,6
	CR	193,3	206,5	218,8	231,0	238,6	241,6	246,7	251,0	263,6
10	G2	189,8	203,6	216,7	230,9	238,7	243,0	244,7	247,7	249,1
	Q2	186,1	199,7	212,2	225,5	233,8	236,4	239,3	241,3	243,9
	FT	195,1	210,6	224,1	237,8	249,1	251,5	254,8	256,5	258,8
	CR	186,6	200,2	212,4	225,3	233,5	237,6	239,5	242,0	244,0

**TABLA 15.** Valores Críticos para el Método de Patrones:  
8 variables manifiestas y 6 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	227,0	238,4	251,3	262,3	267,0	269,9	273,1	275,8	283,6
	Q2	194,6	209,4	222,4	235,3	244,8	246,2	250,1	255,1	264,9
	FT	328,8	344,6	361,2	377,5	383,5	387,7	391,4	394,6	401,4
	CR	193,8	206,7	218,4	229,2	236,5	238,6	241,0	245,0	254,6
2	G2	209,1	223,8	239,1	251,3	258,0	259,8	263,1	266,4	275,2
	Q2	178,9	191,1	205,3	218,8	224,1	226,8	229,7	233,5	238,8
	FT	285,3	307,8	327,9	347,4	356,1	361,0	364,5	371,2	376,2
	CR	181,3	193,6	206,1	219,6	224,9	227,0	230,0	235,0	241,8
3	G2	196,9	211,3	227,7	240,2	248,6	250,3	251,8	255,1	265,0
	Q2	177,7	188,8	202,9	213,9	222,9	224,5	226,3	228,6	234,5
	FT	242,5	264,4	286,7	307,2	319,3	322,6	327,9	331,4	351,0
	CR	179,4	190,7	205,5	215,5	224,0	226,4	227,4	229,7	234,4
4	G2	193,6	206,7	220,9	234,7	241,9	245,6	248,8	252,2	255,5
	Q2	180,7	192,8	206,1	217,1	225,6	228,3	229,7	231,4	235,7
	FT	218,9	238,9	258,0	276,8	291,0	297,1	299,4	305,1	308,8
	CR	181,5	194,1	207,2	218,2	226,8	227,7	228,8	233,0	237,2
5	G2	192,3	207,1	220,1	235,4	243,4	244,9	249,0	254,7	260,2
	Q2	182,7	196,0	208,5	221,6	228,9	231,1	234,5	238,2	247,4
	FT	210,2	227,3	245,9	266,0	278,1	283,0	287,0	293,1	300,1
	CR	183,2	197,1	209,1	223,0	228,7	231,3	234,3	237,9	248,5
6	G2	189,1	202,7	217,5	232,9	241,3	243,5	249,4	252,5	257,6
	Q2	181,4	195,1	208,2	221,3	232,2	236,1	237,9	239,4	247,1
	FT	201,0	215,7	234,0	254,1	265,7	269,3	273,0	276,7	284,0
	CR	182,6	195,4	209,1	222,5	232,0	236,0	238,0	240,7	247,1
7	G2	193,2	206,5	222,1	233,4	242,7	244,0	247,6	253,3	259,5
	Q2	186,8	198,7	214,1	225,0	233,1	235,6	237,8	240,9	247,6
	FT	201,7	217,0	234,2	248,7	261,5	263,5	270,6	277,6	288,2
	CR	187,4	199,8	214,4	225,9	233,7	235,7	237,6	241,3	249,4
8	G2	184,5	198,5	214,1	225,1	232,4	234,4	237,9	240,8	245,8
	Q2	179,7	193,5	208,0	219,0	226,5	227,9	230,7	232,8	241,8
	FT	190,6	206,9	223,6	237,8	246,9	248,7	252,7	255,5	258,8
	CR	180,4	193,5	208,8	219,9	226,9	228,0	230,9	233,7	240,9
9	G2	187,9	200,9	212,6	226,1	235,1	236,3	240,1	243,4	247,4
	Q2	182,4	195,5	207,6	220,1	228,8	229,9	232,7	234,7	245,0
	FT	192,8	207,4	221,3	236,4	246,9	249,4	251,5	254,1	261,6
	CR	183,5	196,5	207,7	220,7	228,6	230,3	232,9	235,6	244,7
10	G2	192,9	205,4	219,3	230,7	237,7	239,9	240,7	246,7	250,6
	Q2	188,6	201,6	214,9	226,8	233,0	234,5	236,5	239,7	243,9
	FT	198,0	211,0	226,6	238,0	246,1	247,5	250,5	254,4	262,4
	CR	188,9	202,1	215,0	226,3	233,4	234,5	236,4	239,6	243,8

**TABLA 16.** Valores Críticos para el Método de Patrones:  
9 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	540,3	559,1	576,9	590,2	601,3	603,0	609,2	619,3	620,3
	Q2	469,3	488,5	509,9	528,2	543,0	543,4	554,4	556,3	556,7
	FT	804,3	829,7	853,4	873,2	884,2	891,2	894,5	910,0	911,6
	CR	464,7	481,9	499,4	517,0	524,9	530,3	535,0	541,9	542,1
2	G2	531,4	556,8	581,1	602,2	608,4	612,0	615,1	624,8	629,6
	Q2	464,6	484,5	510,4	528,5	542,5	545,5	552,8	558,0	561,3
	FT	739,3	772,5	813,0	843,0	859,8	862,2	870,0	872,9	878,7
	CR	466,7	485,9	508,9	528,5	541,4	541,7	546,2	550,9	557,7
3	G2	509,0	530,0	549,1	573,2	587,6	589,4	593,8	612,2	623,3
	Q2	463,6	483,9	501,3	519,8	530,3	536,9	542,7	566,4	584,1
	FT	634,0	665,1	694,5	733,2	751,2	751,9	765,4	768,1	799,1
	CR	468,0	485,2	502,3	520,2	531,5	535,1	542,0	562,5	579,3
4	G2	494,6	518,0	537,4	553,8	570,5	572,6	578,6	585,9	586,9
	Q2	462,8	484,7	504,6	523,6	533,5	536,5	538,7	545,2	554,5
	FT	568,3	598,8	626,0	654,1	669,8	671,8	695,7	697,3	722,0
	CR	463,6	487,6	506,4	522,0	534,1	535,8	541,8	542,8	551,4
5	G2	484,7	506,7	529,8	552,7	563,9	564,9	568,1	582,2	588,5
	Q2	465,1	485,8	501,5	527,2	538,2	543,3	550,1	551,0	559,8
	FT	527,5	558,3	590,4	619,5	633,0	638,5	640,5	647,9	674,0
	CR	464,9	485,3	503,4	526,3	537,2	540,3	547,3	552,5	558,4
6	G2	483,9	509,1	529,4	551,6	564,2	566,2	568,5	571,8	577,6
	Q2	468,3	491,2	511,1	532,2	544,9	545,2	545,9	551,5	557,8
	FT	516,6	548,8	572,1	590,5	607,9	609,5	627,9	629,5	636,9
	CR	469,4	492,0	512,5	533,4	542,6	543,1	545,4	551,3	554,7
7	G2	483,7	503,8	523,6	544,1	553,1	557,8	562,3	566,1	572,0
	Q2	469,8	490,1	508,1	529,5	535,7	538,1	548,0	550,5	554,7
	FT	505,4	528,2	554,1	582,7	595,1	596,1	596,3	598,8	611,6
	CR	470,7	490,7	509,4	528,8	533,8	538,9	546,6	549,0	555,5
8	G2	478,3	497,5	523,0	542,2	553,5	554,2	557,6	562,6	567,9
	Q2	465,5	484,9	508,8	528,3	541,0	543,7	546,9	550,0	552,9
	FT	495,5	518,5	548,7	572,3	581,2	584,1	586,9	593,1	597,1
	CR	466,8	485,7	509,9	528,8	539,0	542,3	546,6	549,5	551,6
9	G2	474,5	495,2	515,9	532,0	544,0	550,6	555,2	556,6	564,1
	Q2	462,5	484,1	502,8	520,2	533,0	535,5	539,1	544,4	552,8
	FT	489,9	512,8	535,2	558,4	574,4	575,5	577,5	578,3	583,4
	CR	463,6	484,5	504,9	519,6	531,3	536,2	540,0	546,4	550,8
10	G2	469,2	494,6	517,7	535,4	543,8	545,6	551,2	571,3	584,9
	Q2	461,2	482,5	506,5	527,0	536,5	537,5	540,1	556,0	573,8
	FT	481,5	510,9	537,1	553,4	564,4	567,9	571,9	592,6	605,4
	CR	461,3	484,0	508,0	526,0	535,3	536,6	539,8	557,5	573,5

**TABLA 17.** Valores Críticos para el Método de Patrones:  
9 variables manifiestas y 3 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	526,1	541,0	557,5	572,8	583,4	584,0	585,2	592,1	594,5
	Q2	451,9	469,2	491,2	511,0	524,0	526,1	528,6	530,4	534,9
	FT	781,6	803,5	822,8	848,4	855,7	861,8	863,3	870,1	875,1
	CR	447,4	462,9	481,5	497,7	508,9	511,2	514,2	517,2	518,3
2	G2	518,3	541,7	569,5	585,3	603,2	606,6	609,7	611,5	618,4
	Q2	450,8	469,1	493,8	516,2	530,8	532,5	536,1	538,3	539,5
	FT	715,9	754,0	796,9	824,9	847,7	849,8	854,9	860,8	868,8
	CR	452,0	473,7	496,8	515,9	529,3	530,6	534,7	537,7	541,0
3	G2	499,4	518,9	535,7	557,9	572,8	581,3	589,9	595,3	605,5
	Q2	453,6	468,1	491,3	505,8	522,6	523,9	526,0	540,8	558,5
	FT	620,2	650,4	677,3	713,1	731,7	736,9	757,2	771,9	779,1
	CR	455,7	471,2	491,7	504,4	525,1	525,5	528,6	542,3	556,2
4	G2	478,9	497,2	519,2	545,4	555,6	557,8	564,2	570,0	577,9
	Q2	449,6	467,7	487,3	507,3	526,3	528,0	528,7	535,9	552,3
	FT	546,3	576,2	609,1	636,0	660,4	661,7	663,7	665,7	670,7
	CR	450,4	467,9	488,2	510,2	521,0	527,3	530,7	533,2	549,0
5	G2	467,3	484,8	512,3	534,7	549,6	553,6	554,1	562,0	567,7
	Q2	448,2	466,5	487,9	507,0	519,0	521,4	528,8	534,7	545,0
	FT	507,9	538,6	568,5	599,9	626,7	632,8	637,7	638,4	642,4
	CR	450,1	465,9	487,6	508,4	520,3	523,4	529,1	534,7	542,3
6	G2	471,4	491,2	509,2	526,6	536,1	538,9	542,3	546,8	547,7
	Q2	454,6	474,7	491,5	506,2	515,8	518,6	521,3	528,1	537,4
	FT	499,6	526,1	550,4	572,2	580,2	589,1	591,5	596,3	615,6
	CR	455,9	475,1	491,6	508,2	516,9	518,2	522,0	528,5	537,4
7	G2	463,2	483,1	508,1	523,1	533,2	534,1	538,7	543,0	546,3
	Q2	448,6	470,6	490,7	505,7	517,3	518,6	524,3	531,3	532,8
	FT	483,8	508,1	540,0	561,5	569,8	570,6	573,1	575,2	585,9
	CR	449,5	470,9	492,4	507,3	518,9	519,8	521,2	525,5	531,4
8	G2	473,6	493,3	515,1	539,2	554,0	559,4	565,5	569,0	574,7
	Q2	461,8	483,1	501,9	528,7	534,4	537,8	542,9	553,7	555,2
	FT	494,3	514,4	538,8	570,2	584,9	593,1	595,8	601,7	613,1
	CR	463,0	482,1	501,1	527,9	536,5	537,8	543,6	554,2	556,0
9	G2	459,9	479,4	502,0	525,0	534,8	536,6	538,8	543,5	547,3
	Q2	449,5	467,6	494,1	514,1	523,6	525,9	526,3	531,5	535,6
	FT	473,1	499,2	519,2	544,8	560,5	561,7	563,8	574,6	577,2
	CR	449,4	469,8	493,6	514,6	519,7	523,6	527,3	529,7	533,3
10	G2	454,8	476,9	493,3	515,1	529,1	533,0	538,2	542,4	554,9
	Q2	447,2	466,2	484,9	507,3	519,0	524,6	527,8	531,5	534,4
	FT	467,9	492,7	512,6	533,4	549,6	553,7	556,9	562,6	582,5
	CR	447,6	467,3	485,1	505,8	517,9	524,0	525,3	531,6	537,5

**TABLA 18.** Valores Críticos para el Método de Patrones:  
9 variables manifiestas y 4 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	511,7	526,9	543,0	555,3	570,5	571,5	571,7	575,0	581,4
	Q2	435,6	456,0	475,2	494,4	505,0	507,3	509,8	510,5	527,9
	FT	757,0	780,9	803,1	821,8	837,2	838,7	847,2	850,3	861,4
	CR	434,9	451,2	468,2	482,9	492,6	495,4	498,2	499,3	511,7
2	G2	506,5	528,9	553,6	572,2	582,9	584,1	588,2	593,7	598,1
	Q2	440,8	457,1	485,7	501,9	514,8	516,3	521,2	522,7	537,6
	FT	697,1	738,0	774,9	802,4	816,0	819,5	820,5	828,6	844,7
	CR	444,4	460,7	486,5	501,6	513,7	517,3	517,8	518,9	531,8
3	G2	477,7	501,1	521,5	539,3	553,1	558,2	558,7	566,2	574,2
	Q2	434,5	454,2	472,1	489,2	500,0	505,8	508,3	509,3	516,4
	FT	588,4	624,2	662,9	687,3	707,6	710,7	723,7	734,9	743,3
	CR	435,1	457,1	474,0	491,5	501,5	505,2	510,8	513,6	518,0
4	G2	465,2	485,2	512,7	530,0	540,2	546,1	548,3	557,1	568,9
	Q2	437,5	453,6	478,2	496,5	505,2	508,8	513,3	515,0	526,8
	FT	534,6	561,9	599,0	623,7	639,5	642,7	653,7	662,0	666,3
	CR	438,2	454,5	479,7	499,1	505,2	508,9	511,9	516,0	530,0
5	G2	453,9	474,4	494,1	517,0	528,4	529,5	534,0	538,4	550,9
	Q2	432,6	451,7	470,8	489,2	507,1	508,0	509,7	512,2	519,6
	FT	495,5	521,4	550,4	586,2	596,8	601,1	606,9	613,0	616,2
	CR	434,6	453,7	471,5	489,9	506,2	507,7	508,7	514,3	519,3
6	G2	455,4	479,6	502,5	519,2	530,7	532,1	533,4	538,0	548,3
	Q2	438,2	462,7	481,3	497,4	511,0	513,6	517,2	520,3	525,1
	FT	484,9	513,9	539,4	562,3	572,6	573,7	583,1	597,5	600,9
	CR	438,5	463,5	483,1	497,2	510,4	515,5	515,9	520,5	524,6
7	G2	449,1	472,6	496,8	514,7	524,6	525,2	530,2	533,1	543,0
	Q2	437,2	456,3	480,9	498,6	505,8	507,2	514,0	520,6	531,7
	FT	470,1	499,2	525,5	548,6	563,3	564,1	566,5	572,1	584,1
	CR	438,0	457,1	481,7	499,7	506,4	508,2	513,1	516,4	530,9
8	G2	450,8	472,0	490,3	513,4	522,5	524,3	526,5	532,4	536,5
	Q2	438,9	457,3	476,1	498,4	509,9	510,9	514,5	516,9	519,6
	FT	469,1	492,7	514,7	534,8	551,8	553,4	558,3	564,3	573,1
	CR	439,6	459,0	475,7	499,8	509,1	511,5	514,8	517,7	518,0
9	G2	460,8	483,7	500,6	524,7	540,0	548,9	553,8	558,4	563,0
	Q2	449,5	470,5	490,6	511,5	530,3	536,1	538,0	542,5	551,6
	FT	477,9	498,0	520,2	544,0	562,7	574,4	580,7	582,0	590,1
	CR	450,9	470,7	490,1	512,2	531,8	533,0	538,2	545,1	551,5
10	G2	444,1	464,9	485,5	509,7	516,8	525,1	526,3	531,1	533,2
	Q2	436,9	454,7	475,0	495,0	509,0	509,4	514,3	518,1	539,2
	FT	456,5	481,0	500,5	526,6	532,6	539,1	548,2	551,2	558,2
	CR	436,5	455,4	476,1	497,8	511,1	511,8	514,1	516,7	533,2

**TABLA 19.** Valores Críticos para el Método de Patrones:  
9 variables manifiestas y 5 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	506,7	521,1	539,0	554,0	561,6	566,5	567,6	570,3	574,9
	Q2	435,8	452,8	474,4	491,6	506,2	507,1	511,8	519,3	527,5
	FT	748,9	769,2	794,1	815,4	824,7	826,8	831,0	840,5	843,0
	CR	432,6	446,9	465,3	480,5	491,9	495,1	500,2	501,4	508,4
2	G2	499,4	522,4	542,4	563,7	574,9	576,7	584,2	594,3	602,1
	Q2	430,2	454,9	475,0	496,3	504,8	515,0	519,2	522,7	533,7
	FT	692,2	724,3	760,9	789,3	808,2	818,2	821,7	829,4	841,1
	CR	433,2	457,8	475,0	495,7	502,8	509,7	514,7	519,3	531,1
3	G2	475,6	497,7	520,7	542,0	550,0	550,4	554,8	557,6	561,8
	Q2	430,0	451,3	472,7	488,4	498,1	500,7	503,8	505,6	507,7
	FT	590,3	628,6	660,8	692,1	708,4	711,0	713,5	717,8	731,8
	CR	433,2	453,3	472,8	491,4	497,1	500,9	505,3	505,6	507,1
4	G2	454,8	476,9	496,8	524,4	534,7	537,3	539,6	546,9	549,9
	Q2	428,6	445,4	463,9	488,4	496,7	503,0	505,0	508,5	521,5
	FT	518,0	552,7	582,3	612,6	632,5	640,3	643,0	653,8	656,0
	CR	430,3	446,8	465,9	489,3	499,2	501,1	504,4	513,2	520,0
5	G2	449,2	470,7	489,9	513,1	526,5	528,1	529,8	541,2	544,1
	Q2	429,6	443,8	465,2	487,5	499,7	501,6	503,9	507,0	513,2
	FT	490,7	516,9	546,0	571,8	591,3	601,0	604,9	613,0	613,5
	CR	430,5	446,5	466,0	489,0	498,9	501,7	505,3	507,7	515,8
6	G2	449,2	469,5	487,6	503,6	515,8	518,2	520,8	522,7	526,7
	Q2	433,0	452,5	470,1	486,8	497,6	499,9	500,2	504,8	509,2
	FT	476,6	503,7	524,7	545,1	558,2	562,3	567,8	576,9	582,9
	CR	434,2	451,7	470,3	485,9	498,4	499,8	502,3	503,0	509,6
7	G2	437,9	457,5	480,8	498,9	512,6	516,1	517,0	521,9	537,1
	Q2	423,1	443,9	465,1	483,4	495,4	498,6	501,6	503,3	514,2
	FT	458,8	481,6	514,2	535,5	546,6	552,9	555,5	561,2	569,6
	CR	424,6	442,4	466,0	483,1	492,6	499,7	502,4	503,7	516,3
8	G2	445,6	463,8	482,6	503,6	516,8	520,6	521,6	524,6	528,3
	Q2	432,9	450,7	469,6	491,0	502,6	503,8	506,6	511,6	516,3
	FT	460,8	485,2	505,9	527,2	546,3	548,9	550,1	562,7	570,6
	CR	433,7	452,4	471,1	490,8	503,7	505,1	505,8	510,7	517,8
9	G2	435,3	453,7	474,0	497,7	509,2	511,2	513,7	514,4	520,9
	Q2	425,5	443,6	463,3	482,8	497,6	498,6	500,3	502,5	506,3
	FT	448,6	470,0	492,1	517,2	533,8	535,1	539,3	543,2	545,2
	CR	426,5	444,2	463,1	485,5	496,5	497,8	500,3	501,8	505,6
10	G2	431,5	454,0	478,3	500,1	509,4	511,0	519,3	529,1	550,3
	Q2	422,5	442,1	467,7	487,5	496,4	500,0	507,0	517,1	533,0
	FT	444,5	467,7	495,2	518,4	530,5	531,1	535,1	548,5	574,3
	CR	423,5	442,8	468,4	487,9	497,5	500,7	507,8	517,7	536,6

**TABLA 20.** Valores Críticos para el Método de Patrones:  
9 variables manifiestas y 6 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	496,9	509,9	526,7	539,5	548,7	552,2	552,7	561,1	567,0
	Q2	424,8	442,2	460,9	479,5	494,3	498,5	500,7	504,8	515,3
	FT	734,0	754,3	776,4	797,0	810,4	811,4	818,2	821,1	822,9
	CR	422,4	438,5	453,7	469,4	479,9	482,0	486,3	492,6	498,8
2	G2	487,2	509,6	528,0	546,4	556,5	557,9	562,7	564,4	575,4
	Q2	419,3	440,6	458,9	478,9	486,1	487,4	491,9	495,1	510,1
	FT	671,8	708,5	735,3	764,2	777,8	782,4	786,5	796,3	805,4
	CR	423,0	443,5	461,1	478,8	485,6	487,0	489,3	494,5	504,9
3	G2	453,2	473,7	496,7	512,3	523,4	528,0	528,2	539,9	565,8
	Q2	409,1	428,4	446,8	463,1	472,6	476,9	477,9	488,6	498,9
	FT	557,3	595,4	628,2	653,3	672,4	673,7	684,7	700,4	729,3
	CR	411,7	430,9	449,6	465,4	477,7	479,2	481,9	487,0	500,8
4	G2	438,9	458,5	478,5	495,1	506,9	508,4	509,3	510,5	529,3
	Q2	409,8	426,7	447,9	461,9	470,6	477,6	480,8	483,9	484,2
	FT	503,8	532,6	557,3	584,5	600,4	609,4	611,5	616,4	637,4
	CR	411,9	428,5	449,5	462,2	472,5	477,3	478,8	485,4	487,7
5	G2	427,1	448,8	473,5	492,6	500,4	501,9	510,2	514,9	526,9
	Q2	408,6	425,5	448,5	468,9	479,7	485,5	488,8	490,2	502,7
	FT	465,6	491,0	527,8	551,3	565,3	569,3	570,4	577,9	591,5
	CR	409,8	427,1	450,1	468,9	479,3	484,3	487,9	491,1	504,5
6	G2	438,9	464,0	481,6	500,9	517,0	518,6	520,5	524,7	536,0
	Q2	421,7	445,2	464,9	484,0	492,0	494,8	499,1	511,3	518,5
	FT	467,0	496,1	517,8	548,1	565,2	567,6	571,6	578,8	587,4
	CR	423,5	445,8	465,0	482,6	493,5	495,4	498,3	508,1	519,1
7	G2	423,8	445,7	465,1	482,3	493,2	496,4	502,6	509,8	511,9
	Q2	410,8	431,1	451,1	467,9	478,8	481,0	483,8	497,5	502,0
	FT	448,7	467,7	488,7	516,7	524,7	528,9	531,7	535,0	540,4
	CR	412,4	432,9	451,9	467,6	479,7	480,8	487,1	494,6	499,1
8	G2	426,0	445,8	464,7	485,2	488,9	489,8	491,5	498,1	507,3
	Q2	415,1	432,2	452,0	470,6	478,5	479,4	479,6	484,6	500,1
	FT	443,4	463,3	484,5	507,4	514,0	516,9	519,0	529,8	540,1
	CR	416,2	433,9	452,9	470,8	477,9	478,6	480,7	483,3	498,9
9	G2	435,5	457,1	476,3	504,5	511,5	512,3	513,5	517,3	525,6
	Q2	424,6	446,6	465,8	489,9	501,4	505,1	506,8	510,1	511,2
	FT	450,6	471,7	494,7	525,3	534,0	535,9	538,5	544,6	544,9
	CR	424,7	447,1	465,2	492,4	500,0	503,7	507,7	509,2	511,6
10	G2	416,9	437,2	458,4	476,1	485,9	487,3	489,0	491,8	512,0
	Q2	409,5	426,0	447,9	464,9	475,6	479,8	485,2	488,5	494,1
	FT	428,6	451,4	472,5	493,5	501,4	502,3	503,8	509,6	534,4
	CR	409,6	427,8	448,2	466,0	476,8	478,0	484,6	485,5	496,7

**TABLA 21.** Valores Críticos para el Método de Patrones:  
10 variables manifiestas y 2 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	1121,9	1140,6	1169,4	1189,8	1201,4	1204,1	1204,8	1206,8	1219,9
	Q2	968,8	996,1	1026,6	1049,3	1066,2	1066,8	1072,6	1084,8	1087,6
	FT	1692,4	1725,3	1759,2	1792,8	1807,0	1810,3	1814,0	1822,8	1826,8
	CR	955,8	980,5	1009,2	1028,4	1039,1	1041,6	1047,7	1055,2	1065,1
2	G2	1096,1	1132,3	1167,8	1193,1	1209,0	1212,6	1215,9	1223,5	1226,6
	Q2	962,1	991,3	1024,4	1051,8	1063,4	1065,9	1074,4	1077,8	1090,8
	FT	1532,2	1589,9	1648,8	1688,4	1712,9	1717,6	1722,0	1723,4	1726,0
	CR	962,6	993,7	1023,3	1049,0	1058,9	1062,0	1066,8	1077,9	1083,1
3	G2	1058,4	1092,5	1119,4	1148,8	1168,3	1179,7	1182,5	1192,3	1199,4
	Q2	968,6	991,8	1021,9	1044,2	1071,2	1077,2	1080,2	1083,4	1087,7
	FT	1323,8	1375,0	1421,0	1469,8	1501,1	1502,9	1506,9	1525,5	1561,1
	CR	969,7	997,6	1022,2	1045,2	1067,8	1077,9	1080,3	1083,1	1087,5
4	G2	1021,1	1055,0	1087,6	1120,2	1134,7	1139,7	1141,5	1149,7	1158,2
	Q2	959,2	991,0	1020,1	1047,2	1061,3	1067,6	1070,8	1075,8	1083,0
	FT	1173,4	1222,8	1281,1	1322,4	1347,4	1353,6	1356,5	1360,5	1382,4
	CR	959,7	992,0	1019,8	1048,0	1063,6	1067,1	1070,1	1080,4	1081,7
5	G2	1009,2	1042,2	1073,8	1099,2	1116,3	1127,0	1136,9	1140,2	1152,1
	Q2	960,0	993,0	1024,9	1052,0	1075,5	1077,5	1081,6	1095,1	1101,4
	FT	1110,2	1153,5	1190,9	1232,4	1255,0	1266,1	1269,1	1273,6	1296,5
	CR	963,1	992,8	1028,0	1051,3	1071,7	1077,0	1080,8	1093,7	1099,9
6	G2	1005,7	1033,1	1064,2	1087,3	1096,6	1103,4	1108,4	1114,9	1131,7
	Q2	971,3	999,2	1022,0	1050,8	1059,6	1064,5	1070,6	1075,8	1089,5
	FT	1076,6	1107,4	1147,8	1181,1	1196,8	1206,7	1216,6	1222,1	1225,6
	CR	971,8	998,2	1022,8	1052,1	1060,6	1061,8	1067,4	1071,2	1090,1
7	G2	993,0	1026,1	1054,7	1082,1	1097,3	1103,5	1107,4	1110,2	1125,1
	Q2	968,3	993,7	1021,5	1047,0	1066,1	1070,6	1073,3	1078,5	1101,9
	FT	1045,2	1082,4	1117,0	1151,7	1165,2	1166,8	1175,5	1188,7	1197,4
	CR	968,9	995,1	1025,1	1048,3	1061,8	1067,5	1073,2	1078,0	1099,1
8	G2	990,3	1018,2	1047,4	1077,5	1089,7	1095,7	1101,6	1112,1	1116,9
	Q2	966,3	992,2	1021,7	1047,5	1062,3	1064,0	1071,7	1073,7	1096,4
	FT	1034,0	1062,2	1097,9	1133,0	1145,6	1148,4	1155,5	1167,8	1185,5
	CR	967,8	993,1	1021,0	1050,3	1061,1	1066,1	1070,9	1078,4	1093,8
9	G2	983,3	1012,4	1046,2	1063,3	1077,1	1078,8	1081,3	1085,1	1108,0
	Q2	963,6	990,2	1022,6	1046,9	1058,9	1059,6	1060,4	1068,4	1079,2
	FT	1016,4	1049,2	1085,6	1106,8	1120,8	1125,4	1132,3	1135,0	1160,5
	CR	963,5	990,6	1022,7	1044,5	1056,3	1057,5	1059,2	1066,5	1082,4
10	G2	987,7	1013,9	1047,1	1070,8	1092,1	1097,1	1098,4	1101,1	1104,9
	Q2	968,2	996,0	1024,4	1048,8	1061,7	1063,0	1068,9	1072,8	1082,5
	FT	1017,1	1049,7	1082,4	1110,2	1136,2	1138,2	1142,1	1146,0	1148,6
	CR	970,7	995,7	1025,9	1049,8	1062,7	1066,9	1070,3	1074,9	1083,2



**TABLA 22.** Valores Críticos para el Método de Patrones:  
10 variables manifiestas y 3 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	1103,9	1127,1	1149,3	1172,1	1185,6	1187,2	1188,1	1190,2	1209,7
	Q2	952,7	983,7	1011,5	1040,6	1059,7	1063,0	1066,9	1070,6	1073,1
	FT	1666,2	1702,3	1732,4	1760,1	1772,9	1778,1	1784,9	1794,5	1805,7
	CR	943,4	966,9	991,3	1016,1	1030,9	1035,1	1036,9	1040,1	1051,8
2	G2	1087,5	1114,4	1150,3	1172,3	1188,5	1194,9	1203,5	1213,9	1225,7
	Q2	948,8	974,9	1009,9	1029,1	1041,5	1049,9	1051,9	1054,4	1063,4
	FT	1523,5	1564,5	1619,8	1665,9	1685,6	1699,3	1708,8	1713,9	1749,2
	CR	952,5	976,1	1009,5	1027,9	1039,6	1046,0	1051,8	1056,7	1065,6
3	G2	1039,6	1073,0	1102,2	1129,5	1143,7	1151,8	1153,2	1178,6	1188,5
	Q2	952,0	975,8	1006,0	1028,0	1044,2	1047,1	1053,9	1066,3	1077,7
	FT	1304,6	1356,5	1403,6	1446,0	1469,3	1474,0	1497,3	1507,7	1545,0
	CR	954,1	977,0	1007,4	1025,2	1043,1	1047,8	1050,7	1068,1	1077,4
4	G2	1008,5	1034,3	1064,9	1094,7	1112,4	1116,4	1120,4	1130,0	1147,3
	Q2	940,9	973,3	998,7	1027,8	1037,9	1041,9	1044,5	1051,9	1065,8
	FT	1160,0	1203,1	1257,4	1297,5	1317,7	1318,9	1326,9	1338,9	1354,6
	CR	945,9	974,7	999,2	1028,7	1040,4	1043,7	1046,5	1051,0	1069,1
5	G2	995,2	1020,7	1046,2	1073,7	1099,1	1100,4	1111,6	1123,6	1134,7
	Q2	945,6	972,3	999,9	1031,5	1039,4	1045,3	1049,3	1066,5	1073,6
	FT	1092,0	1131,9	1168,7	1197,1	1224,2	1238,7	1260,1	1278,6	1295,2
	CR	949,1	972,6	999,7	1030,7	1042,1	1046,2	1054,1	1068,0	1070,8
6	G2	980,8	1010,4	1036,0	1063,6	1083,4	1087,6	1094,7	1097,2	1107,7
	Q2	943,5	971,7	1001,4	1025,6	1038,3	1046,6	1050,7	1059,7	1064,6
	FT	1054,8	1082,6	1111,1	1158,4	1177,7	1184,2	1191,7	1195,5	1215,3
	CR	945,9	972,6	1001,0	1025,2	1040,1	1046,7	1049,6	1058,6	1066,2
7	G2	975,3	1003,0	1035,7	1067,3	1089,1	1095,4	1099,4	1113,0	1114,3
	Q2	945,5	970,7	1006,1	1034,9	1050,2	1055,6	1066,8	1073,6	1084,7
	FT	1025,5	1059,7	1098,3	1134,9	1152,4	1171,8	1172,7	1179,1	1194,1
	CR	946,1	973,7	1005,4	1037,0	1053,1	1056,7	1067,8	1072,5	1084,0
8	G2	970,9	1001,1	1028,3	1057,3	1073,2	1076,0	1086,5	1094,3	1098,6
	Q2	946,8	972,0	1001,2	1028,6	1051,5	1055,4	1057,2	1064,5	1067,7
	FT	1010,4	1044,7	1073,9	1113,3	1118,9	1124,3	1141,5	1149,5	1155,5
	CR	947,4	974,3	1002,4	1031,0	1051,8	1053,5	1056,6	1064,5	1068,0
9	G2	968,0	998,0	1030,8	1052,8	1066,9	1081,3	1088,7	1093,3	1102,3
	Q2	945,7	977,8	1005,2	1029,3	1050,3	1055,5	1059,4	1062,8	1070,8
	FT	996,7	1033,8	1069,4	1097,7	1111,3	1121,8	1137,4	1139,8	1145,1
	CR	947,3	977,8	1006,6	1031,4	1043,1	1053,3	1062,2	1064,5	1073,5
10	G2	960,7	999,8	1034,0	1060,0	1085,2	1092,2	1094,5	1098,9	1135,8
	Q2	941,8	980,6	1015,0	1039,0	1061,6	1065,3	1079,8	1087,1	1107,4
	FT	991,3	1033,9	1068,9	1093,5	1111,1	1131,6	1139,0	1147,6	1169,5
	CR	944,0	981,6	1014,4	1038,7	1062,6	1065,2	1079,2	1083,8	1104,8

**TABLA 23.** Valores Críticos para el Método de Patrones:  
10 variables manifiestas y 4 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	1092,3	1118,1	1143,8	1164,1	1173,5	1179,9	1183,1	1188,9	1200,7
	Q2	940,3	975,0	1003,8	1027,3	1049,2	1052,5	1054,0	1076,2	1092,7
	FT	1652,6	1686,8	1722,9	1750,2	1765,3	1770,6	1772,2	1783,1	1790,6
	CR	929,3	959,7	985,6	1005,0	1021,1	1022,8	1029,4	1044,1	1057,9
2	G2	1074,2	1104,7	1135,4	1168,9	1181,4	1185,9	1191,9	1198,4	1221,6
	Q2	936,6	964,7	994,3	1020,4	1037,3	1042,7	1044,2	1044,5	1063,4
	FT	1507,9	1553,6	1606,2	1651,7	1675,6	1687,9	1693,5	1710,2	1741,7
	CR	940,2	967,9	995,5	1021,9	1035,6	1037,0	1041,6	1044,7	1056,7
3	G2	1034,7	1064,4	1097,7	1127,7	1145,1	1147,0	1153,2	1163,1	1191,2
	Q2	942,1	969,4	996,6	1024,7	1042,8	1044,9	1057,0	1072,5	1096,0
	FT	1296,7	1346,6	1395,9	1445,3	1467,2	1469,3	1474,6	1481,3	1521,2
	CR	945,5	973,0	998,0	1028,5	1042,4	1044,3	1053,4	1069,5	1091,9
4	G2	994,4	1027,3	1062,2	1090,0	1109,2	1110,3	1112,4	1113,8	1149,1
	Q2	932,9	962,3	995,2	1021,5	1035,0	1038,3	1042,2	1052,3	1067,5
	FT	1139,6	1186,4	1245,7	1284,3	1310,0	1314,9	1320,0	1340,4	1358,3
	CR	933,8	963,9	995,0	1022,3	1032,9	1038,2	1041,0	1050,9	1067,8
5	G2	982,7	1009,7	1036,8	1065,7	1079,2	1087,3	1091,3	1100,8	1108,5
	Q2	933,4	963,9	987,5	1011,9	1026,9	1029,2	1036,2	1043,2	1051,7
	FT	1079,7	1123,5	1153,7	1185,8	1218,0	1224,2	1226,2	1248,7	1264,7
	CR	936,1	965,8	989,3	1012,7	1026,2	1029,2	1038,6	1043,7	1051,0
6	G2	974,7	1008,4	1035,7	1060,7	1077,3	1078,2	1083,4	1090,6	1102,2
	Q2	940,5	971,0	997,7	1032,1	1040,8	1045,7	1050,5	1053,4	1055,3
	FT	1039,2	1087,1	1117,7	1147,5	1160,0	1167,3	1173,0	1188,2	1193,1
	CR	942,2	973,4	999,7	1027,7	1040,3	1046,0	1048,4	1051,7	1056,8
7	G2	975,7	1004,7	1034,9	1053,1	1064,7	1069,2	1071,1	1075,8	1083,2
	Q2	945,1	974,9	1002,0	1022,3	1030,1	1032,2	1038,6	1039,7	1047,3
	FT	1025,4	1062,6	1095,1	1120,6	1130,1	1134,0	1140,9	1150,6	1166,4
	CR	947,7	975,6	1002,9	1022,8	1034,0	1034,7	1035,7	1037,6	1045,5
8	G2	962,6	988,6	1016,6	1047,2	1062,3	1066,3	1071,0	1080,4	1086,7
	Q2	936,6	964,1	996,2	1023,9	1033,5	1039,2	1043,5	1047,4	1063,7
	FT	1003,2	1031,7	1068,4	1101,0	1117,6	1121,0	1125,0	1132,7	1162,7
	CR	937,3	964,9	995,2	1023,2	1037,2	1039,2	1041,5	1044,3	1062,6
9	G2	954,2	978,9	1007,9	1037,0	1054,4	1056,6	1065,3	1079,9	1092,0
	Q2	933,2	955,4	983,9	1013,4	1030,8	1042,3	1044,6	1053,7	1062,2
	FT	988,4	1015,0	1047,6	1079,5	1099,4	1100,5	1105,6	1123,3	1144,4
	CR	935,3	956,2	985,8	1015,8	1031,5	1040,2	1042,9	1056,0	1060,8
10	G2	955,7	987,8	1013,7	1041,6	1055,4	1061,4	1063,9	1067,2	1075,1
	Q2	938,7	966,3	991,9	1025,4	1036,2	1038,5	1048,0	1051,9	1057,7
	FT	984,4	1019,9	1047,1	1076,8	1093,3	1094,6	1102,4	1106,8	1115,3
	CR	937,9	967,3	992,6	1021,5	1035,3	1040,2	1045,2	1047,8	1054,3

**TABLA 24.** Valores Críticos para el Método de Patrones:  
10 variables manifiestas y 5 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	1078,4	1107,6	1129,6	1151,9	1168,2	1174,9	1178,0	1179,0	1186,5
	Q2	932,4	963,5	989,9	1008,7	1042,1	1042,7	1045,2	1050,3	1063,9
	FT	1636,4	1669,1	1702,6	1736,4	1754,4	1765,7	1768,5	1769,4	1773,2
	CR	922,1	948,3	971,5	992,3	1017,1	1019,6	1020,6	1023,1	1035,8
2	G2	1074,1	1102,4	1125,0	1157,5	1182,8	1194,3	1199,6	1208,9	1213,0
	Q2	933,8	961,5	987,4	1017,1	1037,6	1042,5	1049,1	1053,7	1056,7
	FT	1510,0	1545,7	1588,2	1635,6	1688,4	1695,2	1699,8	1718,0	1729,3
	CR	937,4	964,7	988,0	1015,8	1037,3	1043,8	1052,5	1054,3	1057,9
3	G2	1033,0	1061,4	1095,5	1128,4	1146,1	1147,3	1154,1	1170,9	1198,3
	Q2	942,7	964,0	995,6	1023,4	1041,6	1053,0	1061,1	1075,5	1097,1
	FT	1296,7	1345,7	1395,2	1441,0	1467,9	1479,5	1482,8	1505,0	1527,4
	CR	947,0	967,0	997,2	1024,5	1043,2	1056,0	1059,4	1066,9	1096,8
4	G2	984,9	1010,4	1045,6	1071,2	1084,6	1090,9	1096,0	1103,2	1135,1
	Q2	922,2	942,5	973,4	1006,3	1016,4	1023,0	1039,1	1050,9	1052,2
	FT	1132,6	1171,2	1216,2	1254,1	1277,8	1284,9	1291,9	1311,7	1353,6
	CR	924,8	945,0	977,4	1006,4	1015,9	1021,5	1036,5	1047,0	1059,7
5	G2	964,5	998,7	1029,4	1054,6	1073,3	1076,3	1082,2	1096,1	1110,6
	Q2	917,7	948,4	978,1	997,2	1019,3	1028,3	1033,8	1047,4	1050,2
	FT	1055,4	1112,8	1148,6	1183,7	1201,7	1209,4	1216,2	1228,0	1232,6
	CR	921,1	951,3	979,7	997,8	1018,7	1029,8	1035,9	1044,6	1048,5
6	G2	977,3	1003,7	1029,2	1051,4	1063,6	1075,5	1078,2	1084,5	1107,8
	Q2	939,1	965,4	990,7	1018,7	1026,4	1028,2	1035,1	1038,1	1060,9
	FT	1043,2	1076,7	1112,9	1141,7	1164,2	1172,9	1180,1	1187,6	1216,6
	CR	939,4	966,6	990,0	1017,2	1026,0	1032,7	1034,4	1039,5	1062,1
7	G2	956,0	983,3	1012,8	1038,4	1052,8	1055,9	1062,8	1069,8	1091,4
	Q2	927,2	954,5	980,3	1002,7	1023,2	1025,0	1036,0	1037,9	1056,1
	FT	1002,0	1039,4	1074,4	1106,0	1128,5	1132,6	1137,1	1142,6	1156,4
	CR	929,7	956,3	979,7	1004,1	1023,9	1025,6	1036,8	1037,8	1056,8
8	G2	943,8	973,8	1003,2	1031,1	1044,4	1050,7	1058,1	1064,3	1076,2
	Q2	915,1	946,2	976,0	1004,4	1018,2	1021,6	1026,5	1030,1	1048,8
	FT	982,7	1019,3	1050,3	1078,2	1098,5	1103,6	1117,7	1126,9	1134,6
	CR	917,8	947,5	976,2	1005,1	1018,1	1020,1	1027,6	1031,7	1051,0
9	G2	942,2	972,1	1001,8	1020,8	1037,2	1041,6	1044,0	1045,6	1051,9
	Q2	920,9	950,5	975,7	999,3	1016,5	1022,1	1024,3	1027,1	1039,6
	FT	977,8	1006,2	1040,1	1063,4	1077,5	1078,9	1079,2	1086,9	1094,2
	CR	922,4	952,5	977,6	1000,3	1017,6	1017,9	1024,5	1025,4	1034,4
10	G2	944,1	974,4	1003,4	1032,1	1044,3	1046,7	1050,2	1054,5	1055,7
	Q2	925,3	954,9	982,3	1008,0	1014,4	1021,6	1022,7	1031,1	1032,5
	FT	972,6	1003,9	1038,2	1072,9	1091,6	1092,1	1092,9	1097,1	1097,3
	CR	927,5	955,7	983,5	1008,7	1018,1	1019,6	1026,4	1029,5	1033,5

**TABLA 25.** Valores Críticos para el Método de Patrones:  
10 variables manifiestas y 6 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	1058,2	1082,8	1108,4	1126,5	1141,7	1148,3	1153,1	1156,0	1159,4
	Q2	909,9	938,1	971,4	996,2	1005,8	1010,8	1014,2	1015,5	1033,4
	FT	1597,3	1633,8	1670,5	1695,8	1715,4	1722,2	1727,0	1733,6	1745,4
	CR	902,1	927,4	954,7	973,5	986,1	989,4	993,6	999,0	1011,4
2	G2	1038,4	1070,1	1097,4	1122,3	1144,1	1148,9	1152,3	1162,1	1166,7
	Q2	902,8	928,5	955,5	982,6	997,9	1000,7	1001,0	1009,6	1012,2
	FT	1452,2	1504,1	1546,4	1588,4	1624,6	1627,9	1634,6	1647,0	1655,9
	CR	905,9	933,3	957,3	984,6	1000,3	1002,9	1004,0	1011,6	1014,7
3	G2	996,1	1025,7	1059,3	1081,4	1099,4	1101,1	1108,0	1115,1	1116,4
	Q2	908,8	934,0	956,7	985,2	995,6	1001,2	1011,4	1012,5	1039,3
	FT	1247,6	1288,1	1347,1	1389,9	1407,4	1412,9	1421,0	1427,5	1452,8
	CR	911,1	937,3	959,7	988,4	997,1	1004,5	1005,7	1015,8	1032,3
4	G2	955,9	989,2	1026,0	1053,7	1068,6	1070,6	1079,8	1080,8	1089,5
	Q2	896,0	927,3	959,6	985,0	993,9	1000,3	1000,8	1007,3	1022,7
	FT	1101,3	1149,3	1202,1	1242,1	1268,9	1273,0	1276,6	1294,7	1312,2
	CR	899,4	929,2	960,2	989,3	994,8	1001,2	1003,5	1016,0	1017,7
5	G2	952,2	982,3	1014,2	1036,7	1052,7	1061,8	1063,1	1071,1	1083,7
	Q2	907,1	934,2	963,5	988,1	1001,2	1002,4	1009,9	1021,0	1040,1
	FT	1049,7	1088,7	1127,8	1166,8	1185,2	1190,8	1193,9	1198,6	1209,3
	CR	910,3	937,2	964,6	985,8	1002,1	1002,8	1009,3	1022,1	1038,0
6	G2	926,0	959,4	991,7	1024,5	1035,4	1038,2	1044,6	1052,0	1070,1
	Q2	892,6	924,9	957,2	982,1	998,4	1004,8	1006,0	1012,0	1025,9
	FT	988,3	1025,4	1072,1	1104,9	1119,2	1140,6	1148,7	1151,3	1171,8
	CR	893,9	926,1	957,6	984,8	1000,4	1001,9	1007,4	1009,7	1026,7
7	G2	926,4	957,8	986,1	1013,7	1031,9	1044,1	1047,4	1051,7	1054,0
	Q2	896,4	929,0	953,0	987,4	1003,5	1007,1	1016,3	1024,2	1026,7
	FT	978,0	1009,8	1045,6	1079,3	1095,9	1108,2	1114,3	1116,9	1132,1
	CR	897,9	929,1	953,1	985,8	1004,5	1008,5	1017,1	1024,0	1025,5
8	G2	924,0	955,8	985,6	1011,8	1034,8	1043,3	1046,8	1057,1	1060,4
	Q2	898,7	932,3	956,8	985,7	1005,9	1011,7	1016,3	1018,2	1020,4
	FT	967,8	996,8	1029,6	1058,1	1088,7	1094,3	1113,8	1119,2	1124,8
	CR	900,9	933,4	958,5	985,8	1005,1	1015,7	1019,2	1020,4	1022,2
9	G2	912,5	942,1	974,5	1003,5	1024,5	1026,4	1027,1	1033,9	1045,2
	Q2	892,9	920,4	951,5	980,1	997,4	1005,1	1008,3	1010,0	1025,7
	FT	942,6	977,0	1014,1	1040,4	1065,2	1071,1	1074,8	1076,5	1085,0
	CR	893,5	921,2	952,6	980,4	999,4	1006,2	1007,6	1010,8	1025,7
10	G2	916,4	941,6	971,4	999,8	1023,4	1024,6	1030,0	1039,5	1041,0
	Q2	897,9	923,0	953,5	977,3	998,3	1003,5	1006,1	1008,5	1016,6
	FT	944,1	972,7	1002,8	1036,7	1055,1	1061,3	1064,3	1078,0	1082,5
	CR	899,2	922,9	954,2	978,6	997,8	1005,5	1008,5	1009,3	1018,2

**TABLA 26.** Valores Críticos para el Método de Patrones:  
10 variables manifiestas y 7 clases latentes.

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
1	G2	1045,8	1068,9	1093,6	1120,9	1133,6	1135,3	1138,2	1142,2	1143,2
	Q2	900,2	926,4	959,1	990,5	1005,5	1012,7	1014,3	1018,7	1024,8
	FT	1578,5	1611,6	1643,6	1679,3	1695,6	1698,1	1704,4	1709,7	1723,9
	CR	892,0	913,8	943,9	970,5	983,3	986,9	989,2	992,4	1003,0
2	G2	1035,7	1060,5	1094,2	1116,1	1127,0	1127,9	1130,7	1133,8	1150,7
	Q2	897,1	921,2	949,0	973,7	986,4	986,7	989,5	991,7	1007,1
	FT	1449,1	1493,8	1538,2	1577,6	1605,7	1606,8	1608,5	1613,0	1631,6
	CR	903,1	925,4	953,8	976,5	984,3	986,3	990,7	994,5	1001,7
3	G2	977,9	1009,9	1052,8	1070,0	1084,6	1087,3	1101,4	1111,3	1122,7
	Q2	886,4	917,3	947,9	970,0	975,3	977,3	979,6	1001,6	1014,6
	FT	1224,1	1279,2	1333,3	1380,1	1403,0	1410,3	1419,6	1435,6	1445,1
	CR	890,8	922,4	953,8	971,7	979,2	980,5	981,9	1002,0	1020,1
4	G2	942,2	973,6	1005,1	1033,4	1048,0	1049,9	1052,3	1054,4	1083,2
	Q2	884,6	910,3	939,4	967,7	979,5	980,5	981,7	989,6	1004,3
	FT	1084,8	1127,1	1177,5	1213,0	1245,2	1246,2	1252,0	1256,3	1280,1
	CR	887,6	911,9	943,3	973,0	979,3	982,0	982,7	984,0	1008,5
5	G2	932,0	965,3	991,1	1017,4	1038,1	1041,0	1049,0	1056,7	1067,7
	Q2	885,3	917,5	940,5	968,6	985,4	986,8	993,3	996,9	1013,0
	FT	1024,8	1062,4	1101,9	1139,5	1164,0	1165,4	1178,0	1205,7	1227,1
	CR	890,1	919,8	943,6	970,0	987,4	989,9	995,6	1002,3	1012,1
6	G2	913,3	946,1	974,5	1007,5	1026,1	1030,8	1037,2	1040,6	1045,6
	Q2	875,6	907,6	938,9	965,8	981,9	991,1	994,3	1000,5	1001,4
	FT	976,9	1014,9	1053,7	1085,6	1112,2	1117,2	1122,8	1124,2	1146,0
	CR	880,7	909,6	938,8	969,5	986,8	989,0	998,7	999,3	1005,4
7	G2	931,4	963,7	992,0	1024,2	1039,7	1045,0	1048,8	1060,6	1070,9
	Q2	904,9	930,1	962,4	988,3	1010,7	1012,7	1015,5	1028,1	1047,8
	FT	977,7	1019,1	1049,4	1074,6	1110,2	1115,0	1121,4	1123,3	1153,0
	CR	905,4	932,6	963,9	987,8	1009,4	1014,0	1015,5	1028,5	1045,6
8	G2	910,8	940,7	972,0	993,6	1023,4	1028,2	1032,3	1034,7	1057,9
	Q2	880,9	915,3	946,7	970,0	994,1	994,8	1002,6	1013,3	1025,5
	FT	947,5	982,9	1015,3	1047,3	1071,8	1075,7	1085,4	1089,9	1102,9
	CR	883,8	917,4	947,0	969,1	992,1	998,4	1003,9	1011,9	1028,7
9	G2	898,7	931,5	963,4	992,5	1008,4	1014,8	1018,7	1020,3	1025,7
	Q2	881,1	910,2	938,8	961,1	981,4	987,7	994,1	995,1	1006,3
	FT	935,5	964,4	998,5	1030,9	1052,4	1055,7	1057,8	1066,7	1080,7
	CR	882,2	910,8	940,8	965,0	987,2	988,8	994,7	996,2	1004,5
10	G2	909,9	938,6	970,8	1000,5	1010,4	1018,0	1020,0	1023,9	1029,2
	Q2	892,0	921,5	950,5	977,2	990,0	993,7	998,2	1003,4	1013,2
	FT	935,1	969,5	1003,3	1039,0	1046,7	1048,4	1056,1	1057,1	1064,2
	CR	892,4	922,8	952,5	979,2	991,0	993,6	1001,8	1006,2	1009,5

## 7.2 USO DE LAS TABLAS ESTADÍSTICAS

En las tablas de valores críticos, se presenta en la fila superior las probabilidades acumuladas, en la primera columna la razón de  $n/R$  y seguidamente los EBA. Donde se cruzan la fila y la columna corresponde al valor del estadístico de bondad de ajuste  $\theta$ . El nivel de la significación teórico se calcula como,

$$p_{valor} = 1 - P(\theta < \theta^*)$$

Para utilizar las tablas de valores críticos se debe seguir el siguiente procedimiento:

1. Considérese el número de variables manifiestas binarias y establezca la hipótesis nula ( $H_0$ ).
2. Ajuste el modelo de clases latentes para obtener los valores de los estadísticos de bondad de ajuste.
3. Calcúlese la razón  $n/R$  y búsquelo en la tabla.
4. Para cada estadístico de bondad de ajuste (EBA) determine el  $p$ -valor.
5. Rechazar  $H_0$  si el nivel de significación fijado ( $\alpha$ ) es mayor al  $p$ -valor.

## 7.3 APLICACIÓN A DATOS SIMULADOS

Supóngase que un investigador tiene el siguiente conjunto de datos (**Tabla 7.1**), resultado de haber pasado un instrumento de recolección de información a 165 personas. Una parte del cuestionario, consta de 5 preguntas con respuesta dicotómica, cero (0) significa una respuesta negativa y uno (1) positiva.

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$f_{x_h}$
0	0	0	0	0	7
0	0	1	0	0	4
0	1	0	0	0	4
0	1	0	0	1	5
0	1	0	1	0	5
0	1	0	1	1	18
0	1	1	0	0	1
0	1	1	0	1	1
0	1	1	1	0	1
0	1	1	1	1	8
1	0	0	0	1	1
1	0	1	0	0	8
1	0	1	0	1	11
1	0	1	1	0	3
1	0	1	1	1	6
1	1	0	0	1	4
1	1	0	1	0	3
1	1	0	1	1	10
1	1	1	0	0	22
1	1	1	0	1	18
1	1	1	1	0	10
1	1	1	1	1	15

**TABLA 7.1.** Datos simulados para cinco variables binarias y un tamaño de muestra de 165 elementos.

El investigador tiene la inquietud de verificar la hipótesis nula definida por un modelo con 2 clases latentes. Para ello calcula los estadísticos de bondad de ajuste, los resultados fueron los siguientes:

Estadístico	Valor
Razón de Verosimilitud	66,82
Estadístico de Pearson	169,12
Freeman-Tukey	67,37
Cressie-Read	108,23

**TABLA 7.2.** Estadísticos de bondad de ajuste para el modelo con dos clases latentes.

Tal que, estamos en presencia de una tabla poco ocupada, solamente el 69% de patrones de respuestas fueron observados, no debe hacer uso de los valores críticos de la distribución de probabilidad Chi-Cuadrado. Una solución al problema, es contrastar la hipótesis nula utilizando los valores críticos del *Método de Patrones*.

Usando la **Tabla 2** (pág. 161) se tiene que seleccionar el valor de  $n/R$  más próximo al calculado  $165/32 = 5,15 \cong 5$ . De esta manera, se posiciona en  $n/R = 5$  y se determina que los cuatro estadísticos obtenidos en la muestra (**Tabla 7.2**) son mayores a los proporcionados en la tabla, como se observa a continuación,

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
5	G2	15,4	19,2	24,0	28,2	31,9	32,8	34,2	36,9	39,1
	Q2	14,9	18,0	22,3	26,8	28,9	30,2	31,6	33,3	37,0
	FT	16,1	20,4	26,3	34,1	38,4	40,2	42,1	44,3	49,2
	CR	15,0	18,2	22,5	26,8	29,3	30,5	31,8	33,2	37,2

esto implica que los *p-valores* son muy pequeños o tienden a cero (menores a 0.01), por lo cual, debe rechazar el modelo con 2 clases latentes.

La opción es contrastar el modelo con 3 clases latentes, los resultados de los estadísticos de bondad de ajuste son:



Estadístico	Valor
Razón de Verosimilitud	7,41
Estadístico de Pearson	7,05
Freeman-Tukey	8,53
Cressie-Read	6,99

**TABLA 7.3.** Estadísticos de bondad de ajuste para el modelo con tres clases latentes.

Mediante la **Tabla 3** (pág. 162) podemos observar que los estadísticos calculados para la muestra son inferiores a los representados,

n/R	EBA	Probabilidad Acumulada								
		0,25	0,50	0,75	0,90	0,95	0,96	0,97	0,98	0,99
5	G2	14,2	17,6	21,9	27,0	31,7	32,2	33,3	35,8	39,6
	Q2	13,5	16,8	21,1	25,8	29,6	30,5	31,4	33,6	36,8
	FT	14,7	18,8	24,3	31,8	36,7	39,2	41,7	43,7	45,2
	CR	13,6	16,8	21,1	25,7	29,8	30,4	32,0	33,1	36,9

Por tanto, los *p-valores* de los 4 estadísticos son mayores a  $1 - 0,25 = 0,75$  . Entonces, tomamos la decisión de no rechazar el modelo propuesto, con 3 clases latentes.

## 7.4 CONSIDERACIONES FINALES

Las tablas de valores críticos para el Método de Patrones, permiten que, cualquier investigador que tiene interés en ajustar un modelo de clases latentes en tablas poco ocupadas, lo pueda hacer, sin necesidad de tener amplios conocimientos en estadística y en informática; como así lo requiere el método de Bootstrap Paramétrico (en caso que no tenga a disposición de un software), ya que debe saber simular conjuntos de datos para un determinado número de variables manifiestas y de clases latentes.

En vista de lo anterior, teniendo en cuenta la sustentación teórica presentada sobre el Método de Patrones, nos permiten presentar las tablas de valores críticos como un método sencillo y una alternativa práctica al Bootstrap Paramétrico para el diagnóstico de modelos de clases latentes en tablas poco ocupadas. Este hecho facilita el estudio de problemas en el marco de clases latentes y ayuda al proceso de inferencia del modelo hipotetizado.

---

## **CONCLUSIONES**

---

1. Se presenta el método *Bootstrap No Paramétrico Corregido*, que representa una contribución innovadora al problema del diagnóstico del modelo de clases latentes en tablas poco ocupadas.
2. Se desarrolla un nuevo método para el diagnóstico de modelos de clases latentes en tablas poco ocupadas, *Método de Patrones*, cuya fiabilidad se ha probado tanto con datos reales como con datos simulados.
3. Se ha resuelto el problema de patrones de respuestas faltantes, para la estimación de la función de distribución empírica de los estadísticos de bondad de ajuste, utilizando el *Método de Patrones*, dado que la proporción de patrones observados en las muestras es siempre mayor que la misma proporción en la muestra original.
4. Se demuestra empíricamente que, el *Método de Patrones* necesita de menos supuestos teóricos para generar la distribución empírica de los estadísticos de bondad de ajuste, en tablas poco ocupadas, si lo comparamos con el enfoque clásico (Bootstrap Paramétrico), permitiendo seleccionar con mejor criterio, el modelo apropiado para los datos.
5. Se construyen tablas de valores críticos para el diagnóstico de modelos de clases latentes con variables manifiestas binarias, utilizando el *Método de Patrones*.

6. Se elabora un programa en Matlab llamado BootLatentclass, que permite el diagnóstico de modelos de clases para variables manifiestas binarias, en tablas poco ocupadas. Además, se pueden simular conjuntos de datos binarios para modelos de clases latentes.
7. Se aportan las pruebas experimentales que demuestran que el Bootstrap No Paramétrico “Naive” es inadecuado para el diagnóstico de Modelos de Clases Latentes en tablas poco ocupadas.

---

## **BIBLIOGRAFÍA**

---

- AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data*. John Wiley and Sons, New York.
- AGRESTI, A. (1990). *Categorical Data Analysis*. Wiley, New York.
- AGRESTI, A. & YANG, M.C. (1987). An Empirical Investigation of some Effects of Sparseness in Contingency Tables. *Computational Statistics and Data Analysis*, 5(1), 9-21.
- AKAIKE, H. (1987). Factor Analysis and AI. *Psychometrika*, 52(3), 317-332.
- ALBANESE, M.T. & KNOTT, M. (1994). Bootstrapping Latent Variable Models for Binary Response. *British Journal of Mathematical and Statistical Psychology*, 47, 235-246.
- ALVORD, W.G; DRUMMOND, J.E.; ARTHUR, L.O.; BIGGAR, R.J.; GOEDERT, J.J; LEVINE, P.H.; MURPHY, E.L.; WEISS, S.H.; BLATTNER, W.A. (1988). A Method for Predicting Individual HIV Infection Status in the Absence of Clinical Information. *Aids Research and Human Retroviruses*, 4(4), 295-304.
- ANDERSEN, E.B. (1991). *The Analysis of Categorical Data*. Springer-Verlag, Berlin.
- ANDERSON, T.W. (1954). On Estimation of Parameters in Latent Structure Analysis. *Psychometrika*, 19(1), 1-10.
- ARAYA, C.; SEPÚLVEDA, R.; GALINDO, M.P. (2009). *Modelo de Clases Latentes con Variables Binarias: una comparación entre el Método Bootstrap Paramétrico y No Paramétrico*. En Actas del XII Conferencia Nacional de Biometría (23 al 25 septiembre de 2009).
- BARTHOLOMEW, D.J. (1987). *Latent Variable Models and Factor Analysis*. 2th ed., Oxford University Press, London.
- BARTOSZYNSKI, R. & NIEWIADOMSKA, M. (1996). *Probability and Statistical Inference*. John Wiley & Sons, New York.
- BOLLEN, K.A. & STINE, R.A. (1992). Bootstrapping Goodness-of-Fit Measures in Structural Equation Models. *Sociological Methods and Research*, 21(2), 205-229.

- BUCHOLZ, K.; HEATH, A.; REICH, T.; HESSELBROCK, V.; KRARNER, J.; NURNBERGER, J.; SCHUCKIT, M. (1996). Can we Subtype Alcoholism?. A Latent Class Analysis of Data from Relatives of Alcoholics in a Multicenter Family Study of Alcoholism. *Alcoholism Clinical and Experimental Research*, 20(8), 1462-1471.
- CHASSIN, L.; PITTS, S.C.; PROST, J. (2002). Binge Drinking Trajectories from Adolescence to Emerging Adulthood in a High Risk Sample: Predictors and Substance Abuse Outcomes. *Journal of Consulting and Clinical Psychology*, 70(1), 67-78.
- CHU, H.; ZHOU, Y; COLE, S.R.; IBRAHIM, J.G. (2010). On the Estimation of Disease Prevalence by Latent Class Models for Screening Studies using two Binary Screening Tests with Categorical Disease Status Verified in Test Positives only. *Statistics in Medicine*, 29(11), 1206-1218.
- CLOGG, C.C. (1995). Latent Class Models: Recent Developments and Prospects for the Future. In: Clogg, C.C.; Arminger, G. & Sobel, M.E. (eds.) *Handbook of Statistical Modeling in the Social Sciences*, Plenum, New York, 311-359.
- CLOGG, C.C. & GOODMAN, L. A. (1984). Latent Structure Analysis of a set of Multidimensional Contingency Tables. *Journal of the American Statistical Association*, 79 (388), 762-771.
- COCHRAN, W.G. (1954). Some Methods for Strengthening the Common  $\chi^2$  Tests. *Biometrics*, 10(4), 417-451.
- COLLINS, L.M.; FIDLER, P.L.; WUGALTER, S.E.; LONG, J.D. (1993). Goodness-of-Fit Testing for Latent Class Models. *Multivariate Behavioral Research*, 28(3), 375-389.
- CRAMER, H. (1946). *Mathematical Models of Statistics*. Princeton University Press. Princeton, N.J.
- de MENEZES L.M. & LASAOSA, A. (2007). Comparing Fits from Latent Trait and Latent Class Models Applied to Sparse Binary Data: an illustration with human resource management data. *Journal of Applied Statistics*, 34(3), 303-319.
- DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. (1977). Maximum likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1), 1-38.



- DIAS, J.G. (2005). Bootstrapping Latent Class Models. In: Weihs. C. and Gaul, W. (eds.) *Classification - The Ubiquitous Challenge*, Springer-Verlag, Berlin, 121-128.
- DIAS, J.G. & VERMUNT, J.K. (2006). Bootstrap Methods for Measuring Classification Uncertainty in Latent Class Analysis. In: A. Rizzi and M Vichi (eds.) *Proceedings in Computational Statistics*, 31-41, Springer, Heidelberg.
- EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1-26.
- EFRON, B. & TIBSHIRANI, R.J. (1986). Bootstrap Methods for Standards Errors, Confidence Intervals, and other Measures of Statistical Accuracy. *Statistical Science*, 1(1), 54-77.
- EFRON, B. & TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.
- FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*, Vol. 1. Wiley, New York.
- FISHER, R.A. (1941). *Statistical Methods for Research Workers*. 8th ed., Oliver and Boyd, London.
- FORMANN, A.K. (1982). Linear Logistic Latent Class Analysis. *Biometrical Journal*, 24(2), 171-190.
- FREEMAN, M.F. & TUKEY, J.W. (1950). Transformations Related to the Angular and the Square Root. *Annals Mathematical Statistics*, 21(4), 607-611.
- GIL, J. (2003). Aplicación del Método Bootstrap al Contraste de Hipótesis en la Investigación Educativa. *Revista de Educación*, 336, 251-265.
- GOODMAN, L.A. (1974). Exploratory Latent Structure Analysis using both Identifiable and Unidentifiable Models. *Biometrika*, 61(2), 215-231.
- GRADOS, M.A. & MATHEWS, C.A. (2008). Latent Class Analysis of Gilles de la Tourette Syndrome using Comorbidities: Clinical and Genetic Implications. *Biological Psychiatry*, 64(3), 219-225.
- GRIZZLE, J.E.; STARMER, C.F.; KOCH, G.G. (1969). Analysis of Categorical Data by Linear Models. *Biometrics*, 25(3), 489-502.
- HABERMAN, S.J. (1979). *Analysis of Qualitative Data: New developments*, Vol. 2. Academic Press, New York.

- HABERMAN, S.J. (1988). A Warning on the use of Chi-square Statistics with Frequency Tables with Small Expected Cell Counts. *Journal of the American Statistical Association*, 83, 555-560.
- HAGENAARS, J.A. & McCUTCHEON, A.L. (2002). *Applied Latent Class Analysis*. Cambridge University Press, Cambridge.
- HOLLIDAY, E.G.; McLEAN, D.E.; NYHOLT, D.R.; MOWRY, B.J. (2009). Susceptibility Locus on Chromosome 1q23-25 for a Schizophrenia Subtype Resembling Deficit Schizophrenia Identified by Latent Class Analysis. *Archives of General Psychiatry*, 66(10), 1058-1067.
- JOHNSON, W.D. & KOCH, G.G. (1970). Analysis of Qualitative Data: Linear Functions. *Health Services Research*, 5(4), 358-369.
- JÖRESKOG, K.G. & MOUSTAKI, I. (2001). Factor Analysis for Ordinal Variables: A Comparison of three Approaches. *Multivariate Behavioral Research*, 36(3), 347-387.
- KASS, G.V. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29(2), 119-127.
- KENDALL, M. G. (1952). *The Advanced Theory of Statistics*, Vol. 1. 5th ed., Charles Griffin and Co, London.
- KOTZ, S. & JOHNSON, N.L. (1992). *Breakthroughs in Statistics*, Vol. 1-2. Springer-Verlag, New York.
- KULLBACK, S. & LEIBLER, R.A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79-86.
- LANGHEINE, R.; PANNEKOEK, J.; van de POL, F. (1996). Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis. *Sociological Methods & Research*, 24(4), 492-516.
- LAZARSFELD, P.F., & HENRY, N.W. (1968). *Latent Structure Analysis*. Houghton Mill, Boston.
- LÓPEZ-BERNALDO, J.C.; DELGADO, R.; GARCÍA, F.; EIROS, J.M.; ORTIZ, R. (2007). Diagnóstico Microbiológico de la Infección por el VIH. *Enfermedades Infecciosas y Microbiología Clínica*, 25(10), 632-638.
- MAGIDSON, J. & VERMUNT, J.K. (2004). Latent class models. In: Kaplan, D. (ed.) *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Sage Publications, Thousand Oakes, 175-198.

- MAGIDSON, J. & VERMUNT, J.K. (2002). Latent Class Cluster Analysis. In: Hagenaars, J.A. & McCutcheon, A.L. (eds.) *Applied Latent Class Analysis*, Cambridge University Press, London, 89-106.
- MANLY, F.J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2th ed., Chapman & Hall, London.
- MATHWORKS (2010). *MATLAB: The Language of Technical Computing*. Massachusetts.
- MCCUTCHEON, A.C. (1987). *Latent Class Analysis*. Sage Publications, Beverly Hills.
- McHUGH, R.B. (1956). Efficient Estimation and Local Identification in Latent Class Analysis. *Psychometrika*, 21, 331-347.
- MOLLOY, W. & CALDWELL, P. (2002). *La Enfermedad de Alzheimer: Una Guía Práctica para Cuidados y Familiares*. Editorial Paidós Ibérica, Barcelona.
- MOONEY, C.Z. & DUVAL, R.D. (1993). *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Sage, Newbury Park, CA.
- QU, Y.; TAN, M.; KUTNER, M.H. (1996). Random Effects Models in Latent Class Analysis for Evaluating Accuracy of Diagnostics test. *Biometrics*, 52(3), 797-810.
- READ, C. (1993). Freeman-Tukey Chi-Squared Goodness-of-fit Statistics. *Statistics & Probability Letters*, 18(4), 271-278.
- READ, T. & CRESSIE, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York.
- RONDÁN, F.J.; SÁNCHEZ, M.J.; VILLAREJO, A.F (2007). *Análisis de Clases Latentes en la relación entre Calidad de Servicio, Satisfacción y Confianza con la intención de recompra*. Tesis doctoral. Universidad de la Rioja.
- SCHWARZ, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.
- SEPÚLVEDA, R. (2004). *Contribuciones al Análisis de Clases Latentes en Presencia de Dependencia Local*. Tesis Doctoral, Universidad de Salamanca.
- SEPÚLVEDA, R.; VICENTE-VILLARDÓN, J.L.; GALINDO, M.P. (2008). The Biplot as a Diagnostic tool of Local Dependence in Latent Class Models. A medical Application. *Statistics in Medicine*, 27(11), 1855-1869.

- TATE, M.W. & HYER, L.A. (1973). Inaccuracy of the  $\chi^2$  Test of Goodness of Fit when Expected Frequencies are small. *Journal of the American Statistical Association*, 68 (344), 836-841.
- TREMBLAY, R.E.; NAGIN, D.S.; SEGUIN, J.R.; ZOCCOLILLO, M.; ZELAZO, P.D.; BOIVIN, M.; PERUSSE, D.; JAPEL, C. (2004). Physical Aggression during early Childhood: Trajectories and Predictors. *Pediatrics*, 114(1), e43-e50.
- UEBERSAX, J. (2000). A practical guide to local dependence in latent class models. Internet WWW page, at URL: <http://ourworld.compuserve.com/homepages/jsuebersax/condep.htm>.
- VERMUNT, J.K. (1997a). *Log-linear Models for Event Histories* (Advanced Quantitative Techniques in the Social Sciences), Vol 8. Sage Publications, Thousand Oakes.
- VERMUNT, J.K. (1997b). *LEM 1.0: A General Program for the Analysis of Categorical Data*. Tilburg University, Tilburg.
- VERMUNT, J.K. & MAGIDSON, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Statistical Innovations Inc., Belmont Massachusetts.
- von DAVIER (2001). *WINMIRA: A program System for Analyses with the Rasch Model, with the Latent Class Analysis and with the Mixed Rasch Model*. Institute for Science Education (IPN), Kiel.
- von DAVIER, M. (1997). Bootstrapping Goodness-of-Fit *Statistics* for Sparse Categorical Data - Results of a Monte Carlo Study. *Methods of Psychological Research*, 2(2), 29-48.
- WALSH, C.D (2006). Latent Class Analysis Identification of Syndromes in Alzheimer's Disease: A Bayesian Approach. *Metodološki Zvezki*, 3(1), 147-162.
- YU, F. & NOBORU, M. (2006). Robust Estimation for Mixture of Probability Tables based on Beta-Likelihood. In: *Proceedings of the Sixth SIAM International Conference on Data Mining*, Bethesda, MD, 519-523.

---

# **ANEXOS**

---

ANEXO 1  
DATOS SIMULADOS DEL APARTADO 6.3.

DATOS 1

Parámetros	Clase latente	
	c=1	c=2
$\pi_Y(c)$	<b>0.540</b>	<b>0.460</b>
$\pi_{X_1/Y(c)}$	1.000	0.028
$\pi_{X_2/Y(c)}$	0.971	0.036
$\pi_{X_3/Y(c)}$	0.613	0.000
$\pi_{X_4/Y(c)}$	1.000	0.081
$\pi_{X_5/Y(c)}$	1.000	0.000

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$f_{x_h}$
0	0	0	0	0	<b>70</b>
0	0	0	0	1	<b>0</b>
0	0	0	1	0	<b>3</b>
0	0	0	1	1	<b>0</b>
0	0	1	0	0	<b>0</b>
0	0	1	0	1	<b>0</b>
0	0	1	1	0	<b>0</b>
0	0	1	1	1	<b>0</b>
0	1	0	0	0	<b>1</b>
0	1	0	0	1	<b>0</b>
0	1	0	1	0	<b>0</b>
0	1	0	1	1	<b>0</b>
0	1	1	0	0	<b>0</b>
0	1	1	0	1	<b>0</b>
0	1	1	1	0	<b>0</b>
0	1	1	1	1	<b>0</b>

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$f_{x_h}$
1	0	0	0	0	<b>3</b>
1	0	0	0	1	<b>0</b>
1	0	0	1	0	<b>1</b>
1	0	0	1	1	<b>0</b>
1	0	1	0	0	<b>0</b>
1	0	1	0	1	<b>0</b>
1	0	1	1	0	<b>0</b>
1	0	1	1	1	<b>3</b>
1	1	0	0	0	<b>0</b>
1	1	0	0	1	<b>0</b>
1	1	0	1	0	<b>0</b>
1	1	0	1	1	<b>31</b>
1	1	1	0	0	<b>0</b>
1	1	1	0	1	<b>0</b>
1	1	1	1	0	<b>0</b>
1	1	1	1	1	<b>48</b>

## DATOS 2

Parámetros	Clase latente		
	c=1	c=2	c=3
$\pi_Y(c)$	<b>0.75</b>	<b>0.20</b>	<b>0.05</b>
$\pi_{X_1/Y(c)}$	0.96	0.36	0.60
$\pi_{X_2/Y(c)}$	0.60	0.78	1.00
$\pi_{X_3/Y(c)}$	0.50	0.64	0.80
$\pi_{X_4/Y(c)}$	0.92	0.37	0.60
$\pi_{X_5/Y(c)}$	0.93	0.86	0.10
$\pi_{X_6/Y(c)}$	0.60	0.79	1.00

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
0	0	0	0	0	0	<b>0</b>
0	0	0	0	0	1	<b>2</b>
0	0	0	0	1	0	<b>0</b>
0	0	0	0	1	1	<b>1</b>
0	0	0	1	0	0	<b>0</b>
0	0	0	1	0	1	<b>0</b>
0	0	0	1	1	0	<b>0</b>
0	0	0	1	1	1	<b>4</b>
0	0	1	0	0	0	<b>0</b>
0	0	1	0	0	1	<b>0</b>
0	0	1	0	1	0	<b>1</b>
0	0	1	0	1	1	<b>8</b>
0	0	1	1	0	0	<b>0</b>
0	0	1	1	0	1	<b>1</b>
0	0	1	1	1	0	<b>3</b>
0	0	1	1	1	1	<b>4</b>
0	1	0	0	0	0	<b>1</b>
0	1	0	0	0	1	<b>4</b>
0	1	0	0	1	0	<b>3</b>
0	1	0	0	1	1	<b>14</b>
0	1	0	1	0	0	<b>1</b>
0	1	0	1	0	1	<b>5</b>
0	1	0	1	1	0	<b>1</b>
0	1	0	1	1	1	<b>7</b>
0	1	1	0	0	0	<b>0</b>
0	1	1	0	0	1	<b>9</b>
0	1	1	0	1	0	<b>2</b>
0	1	1	0	1	1	<b>14</b>
0	1	1	1	0	0	<b>1</b>
0	1	1	1	0	1	<b>7</b>
0	1	1	1	1	0	<b>5</b>
0	1	1	1	1	1	<b>16</b>

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
1	0	0	0	0	0	<b>0</b>
1	0	0	0	0	1	<b>0</b>
1	0	0	0	1	0	<b>2</b>
1	0	0	0	1	1	<b>10</b>
1	0	0	1	0	0	<b>3</b>
1	0	0	1	0	1	<b>6</b>
1	0	0	1	1	0	<b>40</b>
1	0	0	1	1	1	<b>44</b>
1	0	1	0	0	0	<b>2</b>
1	0	1	0	0	1	<b>0</b>
1	0	1	0	1	0	<b>4</b>
1	0	1	0	1	1	<b>3</b>
1	0	1	1	0	0	<b>1</b>
1	0	1	1	0	1	<b>5</b>
1	0	1	1	1	0	<b>32</b>
1	0	1	1	1	1	<b>45</b>
1	1	0	0	0	0	<b>0</b>
1	1	0	0	0	1	<b>1</b>
1	1	0	0	1	0	<b>6</b>
1	1	0	0	1	1	<b>18</b>
1	1	0	1	0	0	<b>4</b>
1	1	0	1	0	1	<b>6</b>
1	1	0	1	1	0	<b>54</b>
1	1	0	1	1	1	<b>70</b>
1	1	1	0	0	0	<b>0</b>
1	1	1	0	0	1	<b>5</b>
1	1	1	0	1	0	<b>3</b>
1	1	1	0	1	1	<b>13</b>
1	1	1	1	0	0	<b>7</b>
1	1	1	1	0	1	<b>13</b>
1	1	1	1	1	0	<b>42</b>
1	1	1	1	1	1	<b>87</b>

## DATOS 3

Parámetros	Clase latente			
	c=1	c=2	c=3	c=4
$\pi_Y(c)$	<b>0.50</b>	<b>0.30</b>	<b>0.15</b>	<b>0.05</b>
$\pi_{X_1/Y(c)}$	1.00	0.36	0.60	0.20
$\pi_{X_2/Y(c)}$	0.98	0.38	0.00	0.09
$\pi_{X_3/Y(c)}$	0.50	1.00	0.40	0.11
$\pi_{X_4/Y(c)}$	0.62	0.98	0.60	0.00
$\pi_{X_5/Y(c)}$	0.43	0.00	1.00	0.30
$\pi_{X_6/Y(c)}$	0.60	0.09	0.98	0.07

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
0	0	0	0	0	0	<b>13</b>
0	0	0	0	0	1	<b>2</b>
0	0	0	0	1	0	<b>8</b>
0	0	0	0	1	1	<b>18</b>
0	0	0	1	0	0	<b>0</b>
0	0	0	1	0	1	<b>0</b>
0	0	0	1	1	0	<b>0</b>
0	0	0	1	1	1	<b>20</b>
0	0	1	0	0	0	<b>6</b>
0	0	1	0	0	1	<b>0</b>
0	0	1	0	1	0	<b>2</b>
0	0	1	0	1	1	<b>11</b>
0	0	1	1	0	0	<b>94</b>
0	0	1	1	0	1	<b>9</b>
0	0	1	1	1	0	<b>0</b>
0	0	1	1	1	1	<b>18</b>
0	1	0	0	0	0	<b>0</b>
0	1	0	0	0	1	<b>0</b>
0	1	0	0	1	0	<b>1</b>
0	1	0	0	1	1	<b>0</b>
0	1	0	1	0	0	<b>0</b>
0	1	0	1	0	1	<b>0</b>
0	1	0	1	1	0	<b>0</b>
0	1	0	1	1	1	<b>0</b>
0	1	1	0	0	0	<b>0</b>
0	1	1	0	0	1	<b>0</b>
0	1	1	0	1	0	<b>0</b>
0	1	1	0	1	1	<b>0</b>
0	1	1	1	0	0	<b>48</b>
0	1	1	1	0	1	<b>5</b>
0	1	1	1	1	0	<b>0</b>
0	1	1	1	1	1	<b>0</b>

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$f_{x_h}$
1	0	0	0	0	0	<b>7</b>
1	0	0	0	0	1	<b>1</b>
1	0	0	0	1	0	<b>3</b>
1	0	0	0	1	1	<b>18</b>
1	0	0	1	0	0	<b>0</b>
1	0	0	1	0	1	<b>3</b>
1	0	0	1	1	0	<b>1</b>
1	0	0	1	1	1	<b>35</b>
1	0	1	0	0	0	<b>2</b>
1	0	1	0	0	1	<b>1</b>
1	0	1	0	1	0	<b>1</b>
1	0	1	0	1	1	<b>18</b>
1	0	1	1	0	0	<b>56</b>
1	0	1	1	0	1	<b>7</b>
1	0	1	1	1	0	<b>1</b>
1	0	1	1	1	1	<b>25</b>
1	1	0	0	0	0	<b>25</b>
1	1	0	0	0	1	<b>20</b>
1	1	0	0	1	0	<b>21</b>
1	1	0	0	1	1	<b>26</b>
1	1	0	1	0	0	<b>39</b>
1	1	0	1	0	1	<b>54</b>
1	1	0	1	1	0	<b>26</b>
1	1	0	1	1	1	<b>40</b>
1	1	1	0	0	0	<b>18</b>
1	1	1	0	0	1	<b>29</b>
1	1	1	0	1	0	<b>15</b>
1	1	1	0	1	1	<b>20</b>
1	1	1	1	0	0	<b>76</b>
1	1	1	1	0	1	<b>49</b>
1	1	1	1	1	0	<b>39</b>
1	1	1	1	1	1	<b>29</b>



## DATOS 4

Parámetros	Clase latente		
	c=1	c=2	c=3
$\pi_Y(c)$	<b>0.50</b>	<b>0.30</b>	<b>0.20</b>
$\pi_{X_1/Y(c)}$	0.04	0.01	1.00
$\pi_{X_2/Y(c)}$	0.67	0.19	0.60
$\pi_{X_3/Y(c)}$	1.00	0.24	0.80
$\pi_{X_4/Y(c)}$	1.00	0.05	0.30
$\pi_{X_5/Y(c)}$	0.50	0.00	1.00
$\pi_{X_6/Y(c)}$	0.08	0.11	0.87
$\pi_{X_7/Y(c)}$	1.00	0.19	0.78

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$f_{x_h}$
0	0	0	0	0	0	0	<b>155</b>
0	0	0	0	0	0	1	<b>30</b>
0	0	0	0	0	1	0	<b>20</b>
0	0	0	0	0	1	1	<b>4</b>
0	0	0	0	1	0	0	<b>0</b>
0	0	0	0	1	0	1	<b>0</b>
0	0	0	0	1	1	0	<b>0</b>
0	0	0	0	1	1	1	<b>0</b>
0	0	0	1	0	0	0	<b>12</b>
0	0	0	1	0	0	1	<b>3</b>
0	0	0	1	0	1	0	<b>1</b>
0	0	0	1	0	1	1	<b>0</b>
0	0	0	1	1	0	0	<b>0</b>
0	0	0	1	1	0	1	<b>0</b>
0	0	0	1	1	1	0	<b>0</b>
0	0	0	1	1	1	1	<b>0</b>
0	0	1	0	0	0	0	<b>44</b>
0	0	1	0	0	0	1	<b>9</b>
0	0	1	0	0	1	0	<b>7</b>
0	0	1	0	0	1	1	<b>1</b>
0	0	1	0	1	0	0	<b>0</b>
0	0	1	0	1	0	1	<b>0</b>
0	0	1	0	1	1	0	<b>0</b>
0	0	1	0	1	1	1	<b>0</b>
0	0	1	1	0	0	0	<b>3</b>
0	0	1	1	0	0	1	<b>79</b>
0	0	1	1	0	1	0	<b>0</b>
0	0	1	1	0	1	1	<b>10</b>
0	0	1	1	1	0	0	<b>0</b>
0	0	1	1	1	0	1	<b>94</b>
0	0	1	1	1	1	0	<b>0</b>

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$f_{x_h}$
0	1	0	0	0	0	0	<b>38</b>
0	1	0	0	0	0	1	<b>9</b>
0	1	0	0	0	1	0	<b>3</b>
0	1	0	0	0	1	1	<b>1</b>
0	1	0	0	1	0	0	<b>0</b>
0	1	0	0	1	0	1	<b>0</b>
0	1	0	0	1	1	0	<b>0</b>
0	1	0	0	1	1	1	<b>0</b>
0	1	0	1	0	0	0	<b>2</b>
0	1	0	1	0	0	1	<b>1</b>
0	1	0	1	0	1	0	<b>0</b>
0	1	0	1	0	1	1	<b>0</b>
0	1	0	1	1	0	0	<b>0</b>
0	1	0	1	1	0	1	<b>0</b>
0	1	0	1	1	1	0	<b>0</b>
0	1	0	1	1	1	1	<b>0</b>
0	1	1	0	0	0	0	<b>13</b>
0	1	1	0	0	0	1	<b>3</b>
0	1	1	0	0	1	0	<b>0</b>
0	1	1	0	0	1	1	<b>0</b>
0	1	1	0	1	0	0	<b>0</b>
0	1	1	0	1	0	1	<b>0</b>
0	1	1	0	1	1	0	<b>0</b>
0	1	1	0	1	1	1	<b>0</b>
0	1	1	1	0	0	0	<b>1</b>
0	1	1	1	0	0	1	<b>206</b>
0	1	1	1	0	1	0	<b>0</b>
0	1	1	1	0	1	1	<b>18</b>
0	1	1	1	1	0	0	<b>0</b>
0	1	1	1	1	0	1	<b>183</b>
0	1	1	1	1	1	0	<b>0</b>



## DATOS 5

Parámetros	Clase latente			
	c=1	c=2	c=3	c=4
$\pi_Y(c)$	<b>0.40</b>	<b>0.30</b>	<b>0.20</b>	<b>0.10</b>
$\pi_{X_1/Y(c)}$	1.00	0.50	0.60	0.50
$\pi_{X_2/Y(c)}$	1.00	0.00	0.40	0.00
$\pi_{X_3/Y(c)}$	0.25	1.00	0.78	0.00
$\pi_{X_4/Y(c)}$	0.00	1.00	0.85	0.03
$\pi_{X_5/Y(c)}$	0.50	0.79	1.00	0.78
$\pi_{X_6/Y(c)}$	0.87	0.00	1.00	0.60
$\pi_{X_7/Y(c)}$	0.45	0.58	0.00	1.00

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$f_{x_h}$
0	0	0	0	0	0	0	<b>0</b>
0	0	0	0	0	0	1	<b>13</b>
0	0	0	0	0	1	0	<b>0</b>
0	0	0	0	0	1	1	<b>14</b>
0	0	0	0	1	0	0	<b>0</b>
0	0	0	0	1	0	1	<b>36</b>
0	0	0	0	1	1	0	<b>2</b>
0	0	0	0	1	1	1	<b>39</b>
0	0	0	1	0	0	0	<b>0</b>
0	0	0	1	0	0	1	<b>0</b>
0	0	0	1	0	1	0	<b>0</b>
0	0	0	1	0	1	1	<b>0</b>
0	0	0	1	1	0	0	<b>0</b>
0	0	0	1	1	0	1	<b>0</b>
0	0	0	1	1	1	0	<b>13</b>
0	0	0	1	1	1	1	<b>4</b>
0	0	1	0	0	0	0	<b>0</b>
0	0	1	0	0	0	1	<b>0</b>
0	0	1	0	0	1	0	<b>0</b>
0	0	1	0	0	1	1	<b>0</b>
0	0	1	0	1	0	0	<b>0</b>
0	0	1	0	1	0	1	<b>0</b>
0	0	1	0	1	1	0	<b>15</b>
0	0	1	0	1	1	1	<b>0</b>
0	0	1	1	0	0	0	<b>27</b>
0	0	1	1	0	0	1	<b>29</b>
0	0	1	1	0	1	0	<b>0</b>
0	0	1	1	0	1	1	<b>0</b>
0	0	1	1	1	0	0	<b>104</b>
0	0	1	1	1	0	1	<b>121</b>
0	0	1	1	1	1	0	<b>64</b>

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$f_{x_h}$
0	1	0	0	0	0	0	<b>0</b>
0	1	0	0	0	0	1	<b>0</b>
0	1	0	0	0	1	0	<b>0</b>
0	1	0	0	0	1	1	<b>0</b>
0	1	0	0	1	0	0	<b>0</b>
0	1	0	0	1	0	1	<b>0</b>
0	1	0	0	1	1	0	<b>1</b>
0	1	0	0	1	1	1	<b>0</b>
0	1	0	1	0	0	0	<b>0</b>
0	1	0	1	0	0	1	<b>0</b>
0	1	0	1	0	1	0	<b>0</b>
0	1	0	1	0	1	1	<b>0</b>
0	1	0	1	1	0	0	<b>0</b>
0	1	0	1	1	0	1	<b>0</b>
0	1	0	1	1	1	0	<b>12</b>
0	1	0	1	1	1	1	<b>0</b>
0	1	1	0	0	0	0	<b>0</b>
0	1	1	0	0	0	1	<b>0</b>
0	1	1	0	0	1	0	<b>0</b>
0	1	1	0	0	1	1	<b>0</b>
0	1	1	0	1	0	0	<b>0</b>
0	1	1	0	1	0	1	<b>0</b>
0	1	1	0	1	1	0	<b>4</b>
0	1	1	0	1	1	1	<b>0</b>
0	1	1	1	0	0	0	<b>0</b>
0	1	1	1	0	0	1	<b>0</b>
0	1	1	1	0	1	0	<b>0</b>
0	1	1	1	0	1	1	<b>0</b>
0	1	1	1	1	0	0	<b>0</b>
0	1	1	1	1	0	1	<b>0</b>
0	1	1	1	1	1	0	<b>46</b>

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$f_{x_h}$
0	0	1	1	1	1	1	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	1	6
1	0	0	0	0	1	0	0
1	0	0	0	0	1	1	15
1	0	0	0	1	0	0	0
1	0	0	0	1	0	1	31
1	0	0	0	1	1	0	5
1	0	0	0	1	1	1	53
1	0	0	1	0	0	0	0
1	0	0	1	0	0	1	0
1	0	0	1	0	1	0	0
1	0	0	1	0	1	1	0
1	0	0	1	1	0	0	0
1	0	0	1	1	0	1	1
1	0	0	1	1	1	0	21
1	0	0	1	1	1	1	1
1	0	1	0	0	0	0	0
1	0	1	0	0	0	1	0
1	0	1	0	0	1	0	0
1	0	1	0	0	1	1	0
1	0	1	0	1	0	0	0
1	0	1	0	1	0	1	0
1	0	1	0	1	1	0	16
1	0	1	0	1	1	1	0
1	0	1	1	0	0	0	31
1	0	1	1	0	0	1	32
1	0	1	1	0	1	0	0
1	0	1	1	0	1	1	0
1	0	1	1	1	0	0	84
1	0	1	1	1	0	1	134
1	0	1	1	1	1	0	75
1	0	1	1	1	1	1	0

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$f_{x_h}$
0	1	1	1	1	1	1	0
1	1	0	0	0	0	0	14
1	1	0	0	0	0	1	21
1	1	0	0	0	1	0	123
1	1	0	0	0	1	1	120
1	1	0	0	1	0	0	16
1	1	0	0	1	0	1	16
1	1	0	0	1	1	0	133
1	1	0	0	1	1	1	123
1	1	0	1	0	0	0	0
1	1	0	1	0	0	1	0
1	1	0	1	0	1	0	0
1	1	0	1	0	1	1	0
1	1	0	1	1	0	0	0
1	1	0	1	1	0	1	0
1	1	0	1	1	1	0	19
1	1	0	1	1	1	1	0
1	1	1	0	0	0	0	4
1	1	1	0	0	0	1	4
1	1	1	0	0	1	0	59
1	1	1	0	0	1	1	32
1	1	1	0	1	0	0	10
1	1	1	0	1	0	1	3
1	1	1	0	1	1	0	59
1	1	1	0	1	1	1	40
1	1	1	1	0	0	0	0
1	1	1	1	0	0	1	0
1	1	1	1	0	1	0	0
1	1	1	1	0	1	1	0
1	1	1	1	1	0	0	0
1	1	1	1	1	0	1	0
1	1	1	1	1	1	0	75
1	1	1	1	1	1	1	0