



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

MODELOS GRÁFICOS LOG-LINEALES

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRÍA EN CIENCIAS

PRESENTA:  
GRACIELA MARTÍNEZ SÁNCHEZ

DIRECTORA DE LA TESIS:  
DRA. GUILLERMINA ESLAVA GÓMEZ  
FACULTAD DE CIENCIAS, UNAM

MÉXICO, D.F. FEBRERO 2016



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Agradecimientos

A la Dra. Guillermina Eslava, por el tiempo y la dedicación brindada durante todo este proceso.

A los sinodales Dra. Silvia Ruiz-Velasco, Dr. Ricardo Ramírez, M. en C. Antonio Soriano y M. en C. Salvador Zamora, que con sus observaciones enriquecieron este trabajo.

Por último, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por haberme otorgado una beca para llevar a cabo estos estudios.

# Modelos gráficos log-lineales

## Resumen

El objetivo de esta tesis es presentar los modelos gráficos log-lineales, así como una aplicación de éstos a un conjunto de observaciones provenientes de la Encuesta Nacional de Adicciones 2011. En estos modelos, se define una gráfica donde los vértices corresponden a variables categóricas, y la asociación entre cualesquiera dos de éstas se representa por medio de una arista. Las relaciones de independencia condicional para un conjunto de variables categóricas se interpretan a partir de las tres propiedades de Markov de la gráfica asociada al modelo.

Las variables categóricas seleccionadas para la aplicación son el consumo de alcohol, tabaco y drogas, así como de las variables sexo, ingreso del hogar y escolaridad. Donde las principales variables de interés son las correspondientes al consumo de sustancias.

Se ajustaron tres modelos gráficos log-lineales. En el primer modelo, se consideró sólo población adulta en los grupos de edad de 18 a 65 años con 10,294 observaciones. Para este modelo, fue posible mostrar la equivalencia entre las 3 propiedades de Markov utilizando el teorema de Pearl y Paz. Además, con base en la descomposición de su gráfica asociada, y usando el teorema de Hammersley y Clifford se obtuvo la factorización para su función de probabilidad conjunta.

En el segundo y tercer modelo se eliminó la variable drogas, esto debido a la baja prevalencia de consumo en la muestra. Para el segundo modelo ajustado, sólo se consideró la subpoblación que consume drogas con 231 observaciones. En el tercer modelo ajustado, se utilizó una muestra de 10,063 adultos que no consume drogas. Los dos modelos obtenidos son gráficos descomponibles.

Posteriormente, con el primer modelo ajustado, se definieron tres regresiones logísticas con variables respuesta tabaco, drogas y alcohol. Las variables explicativas para cada regresión están determinadas por la gráfica inducida por la variable respuesta.

Finalmente, se obtuvieron las probabilidades ajustadas considerando un modelo gráfico y una regresión logística. Se observó que estas probabilidades coincidían si la variable respuesta estaba contenida en una sola clase generadora, como fue en el caso de las variables drogas y tabaco, a diferencia de la variable alcohol.

# Índice de figuras

1.1.	Diagrama de la gráfica $\mathcal{G} = (V, E)$ , con vértices $V = \{1, 2, 3, 4, 5\}$ y conjunto de aristas $E = \{(1, 2), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5)\}$ . . . . .	4
1.2.	Gráfica no dirigida $\mathcal{G} = (V, E)$ , con conjunto de vértices $V = \{1, 2, 3, 4, 5, 6\}$ y conjunto de aristas $E = \{(1, 2), (1, 5), (2, 3), (2, 5), (3, 4), (3, 5)\}$ . . . . .	5
1.3.	Unión e intersección de las gráficas $\mathcal{G}_1 = (V, E)$ y $\mathcal{G}_2 = (W, F)$ . . . . .	6
1.4.	Gráfica no dirigida, $\mathcal{G} = (V, E)$ . Donde $V = \{1, 2, 3, 4, 5, 6, 7\}$ y $E = \{(1, 2), (1, 3), (2, 3), (2, 5), (3, 5), (3, 6), (4, 7), (5, 6), (5, 7), (6, 7)\}$ . . . . .	9
1.5.	Ejemplos de gráficas principales. . . . .	10
1.6.	Descomposición de una gráfica en subgráficas principales . . . . .	11
1.7.	Factorización de la gráfica $\mathcal{G} = (V, E)$ , donde $V = \{1, 2, 3, 4\}$ y $E = \{(1, 2), (1, 3), (1, 4)\}$ . . . . .	12
1.8.	Un ordenamiento de la gráfica $\mathcal{G}$ (figura 1.4). . . . .	13
1.9.	Ilustración del algoritmo glotón . . . . .	14
1.10.	Algoritmo de Búsqueda de Máxima cardinalidad. Los vértices en rojo, indican los vértices con vecinos etiquetados, en tanto que los vértices en verde indican los vértices con el mayor número de vecinos etiquetados. . . . .	15
1.11.	Numeración obtenida por el algoritmo de máxima cardinalidad. . . . .	16
1.12.	<i>Junction tree</i> . . . . .	17
2.1.	Gráficas asociadas a modelos jerárquicos con diferentes conjuntos de clases generadoras para modelos jerárquicos de 3 variables . . . . .	27
2.2.	Diagrama de la gráfica $\mathcal{G}$ , asociada al modelo jerárquico descomponible con clases generadoras $\{\{1, 2\}, \{1, 3\}, \{1, 4\}\}$ . . . . .	28
2.3.	Modelos gráficos descomponibles. <b>(a)</b> Gráfica asociada al modelo jerárquico $M_0$ , con clases generadoras $\mathcal{C}_0 = \{\{A, B, C\}, \{A, C, D\}\}$ ; <b>(b)</b> Gráfica asociada al modelo jerárquico anidado $M_1$ , con clases generadoras $\mathcal{C}_1 = \{\{B, C\}, \{A, C, D\}\}$ . . . . .	36
2.4.	Modelo gráfico, con clases generadoras $\{\{1, 2\}, \{1, 3\}\}$ . . . . .	44
2.5.	Gráficas inducidas por la frontera de cada vértice en la gráfica $\mathcal{G} = (V, E)$ , donde $V = \{1, 2, 3, 4\}$ y $E = \{(1, 2), (1, 3), (1, 4), (2, 3), (3, 4)\}$ . . . . .	44
3.1.	Gráfica de mosaico por parejas . . . . .	52
3.2.	Gráfica $\mathcal{G} = (V, E)$ asociada al modelo gráfico log-lineal descomponible. Con clases generadoras $\mathcal{C} = \{\{A, E, I, S\}, \{A, D, S, T\}\}$ y conjunto separador $\mathcal{S} = \{A, S\}$ . . . . .	56

3.3. Gráfica del modelo log-lineal descomponible para la subpoblación que consume drogas. . . . .	59
3.4. Modelo seleccionado para la población que no consume drogas. . . . .	62
3.5. Comparación de modelos gráficos. <b>(a)</b> Modelo gráfico para la subpoblación que consume drogas y <b>(b)</b> modelo gráfico para la subpoblación que no consume drogas. . . . .	63
3.6. Subgráficas de <b>(a)</b> $\mathcal{G} = (V, E)$ inducidas por la frontera del consumo de sustancias: <b>(b)</b> tabaco, $\mathcal{G}_{bd(T)}$ , con clase generadora $\{S, D, A\}$ ; <b>(c)</b> drogas, $\mathcal{G}_{bd(D)}$ , con clase generadora $\{A, T, S\}$ y <b>(d)</b> alcohol, $\mathcal{G}_{bd(A)}$ , con clases generadoras $\{E, S, I\}, \{S, T, D\}$ . . . . .	64
A.1. Intervalos al 95 % de confianza para el consumo de alcohol. . . . .	81
A.2. Intervalos al 95 % de confianza para el consumo de tabaco. . . . .	81
A.3. Intervalos de confianza al 95 % por el método logit para el consumo de drogas médicas. . . . .	82
A.4. Intervalos de confianza al 95 % por el método logit para el consumo de drogas ilegales. . . . .	82

# Índice de cuadros

2.1. Factores de riesgo asociados a afecciones cardíacas. Donde $A$ , fumador; $B$ , trabajo mentalmente estresante; $C$ , trabajo físicamente estresante; $D$ , presión sanguínea sistólica alta; $E$ , cociente de lipo-proteínas beta y alfa alta; $F$ , historia familiar de afecciones cardíacas . . . . .	19
2.2. Clases generadoras para modelos log-lineales jerárquicos con 3 variables. . . . .	26
2.3. Modelos log-lineales jerárquicos para los datos de la tabla 2.1. . . . .	29
2.4. Primera iteración del algoritmo IPF para el modelo jerárquico log-lineal $(ABC, ACD)$ . . . . .	32
3.1. Codificación de las 6 variables seleccionadas para el ajuste de modelos gráficos log-lineales. . . . .	48
3.2. Estimación del porcentaje de la población para el consumo de sustancias de último año. . . . .	49
3.3. Descripción de las categorías agregadas para las variables escolaridad e ingreso. . . . .	51
3.4. Razón de momios marginales para las variables sexo ( $S$ ), escolaridad ( $E$ ), ingreso del hogar ( $I$ ), alcohol ( $A$ ), tabaco ( $T$ ) y drogas ( $D$ ). . . . .	51
3.5. Estadísticas de bondad de ajuste para diferentes modelos log-lineales jerárquicos. Donde $S$ , denota el sexo del individuo; $E$ , nivel escolaridad; $I$ , ingreso por hogar; $A$ , consumo de alcohol; $T$ , consumo de tabaco; $D$ , consumo de drogas. . . . .	53
3.6. Selección de modelos jerárquicos usando la opción <i>stepwise</i> . . . . .	54
3.7. Modelo gráfico log-lineal seleccionado. . . . .	54
3.8. Razón de momios condicionales y marginales para vértices no adyacentes en el modelo. Las expresiones para $\theta_{IJ}$ , $\hat{\theta}_{IJ}$ y $\hat{\theta}_{IJ(R)}$ se muestran en la sección 3.4. . . . .	56
3.9. Devianzas de inclusión para las aristas en el modelo gráfico log-lineal seleccionado con clases generadoras: $\{\{A, S, E, I\}, \{A, D, S, T\}\}$ . . . . .	58
3.10. Frecuencias observadas y ajustadas para la tabla de contingencia con variables ingreso por hogar ( $I$ ), escolaridad ( $E$ ), sexo ( $S$ ), alcohol ( $A$ ), tabaco ( $T$ ) y drogas ( $D$ ). El modelo ajustado corresponde a uno gráfico descomponible con clases generadoras $\{\{I, S, A, E\}, \{D, S, A, T\}\}$ . . . . .	58
3.11. Modelos seleccionados para la subpoblación que consume drogas ( $D = 1$ ). . . . .	59

3.12. Frecuencias observadas y ajustadas para la tabla de contingencia con variables ingreso del hogar ( $I$ ), escolaridad ( $E$ ), sexo ( $S$ ), alcohol ( $A$ ) y tabaco ( $T$ ), para la población que consume drogas ( $D = 1$ ). El modelo gráfico log-lineal correspondiente tiene clases generadoras $\{\{A, S, T\}, \{S, E\}, \{E, I\}\}$ .	61
3.13. Modelo seleccionado para la subpoblación que no consume drogas ( $D = 2$ ).	61
3.14. Frecuencias observadas y ajustadas para la muestra que no consume drogas ( $D = 2$ ). Donde ingreso por hogar ( $I$ ), escolaridad ( $E$ ), sexo ( $S$ ), alcohol ( $A$ ) y tabaco ( $T$ ),	63
3.15. Ajuste de una regresión logística con variable respuesta tabaco ( $T = 1$ ) y variables explicativas sexo ( $S$ ), alcohol ( $A$ ) y drogas ( $D$ ).	65
3.16. Ajuste del modelo logístico con variable respuesta drogas ( $D = 1$ ) y variables explicativas sexo ( $S$ ), tabaco ( $T$ ) y alcohol ( $A$ ).	66
3.17. Ajuste del modelo logístico para la variable respuesta alcohol ( $A = 1$ ) y variables explicativas ingreso ( $I$ ), tabaco ( $T$ ), drogas ( $D$ ) y escolaridad ( $E$ ).	68
3.18. Probabilidades ajustadas con variable respuesta tabaco ( $T = 1$ ) y variables explicativas drogas ( $D$ ), alcohol ( $A$ ) y sexo ( $S$ ). Los valores de $p_{obs}$ , $\hat{p}_{graph}$ y $\hat{p}_{logit}$ se obtuvieron a partir de las fórmulas en (3.17), (3.14) y (3.11), respectivamente.	70
3.19. Probabilidades ajustadas con variable respuesta drogas ( $D = 1$ ) y variables explicativas alcohol ( $A$ ), tabaco ( $T$ ) y sexo ( $S$ ). Los valores de $p_{obs}$ , $\hat{p}_{graph}$ y $\hat{p}_{logit}$ se obtuvieron a partir de las fórmulas en (3.18), (3.15) y (3.12), respectivamente.	70
3.20. Probabilidades ajustadas con variable respuesta alcohol ( $A = 1$ ) y variables explicativas drogas ( $D$ ), tabaco ( $T$ ), ingreso ( $I$ ), escolaridad ( $E$ ) y sexo ( $S$ ). Los valores de $p_{obs}$ , $\hat{p}_{graph}$ y $\hat{p}_{logit}$ se obtuvieron a partir de las fórmulas en (3.19), (3.16) y (3.13), respectivamente.	71
A.1. Descripción de los estratos por tipo de urbanidad.	76
A.2. Estados que integran cada una de las 8 regiones.	76
A.3. Estructura del diseño muestral para la Encuesta Nacional de Adicciones 2011.	77

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Modelos gráficos probabilísticos</b>	<b>3</b>
1.1. Conceptos básicos de teoría de gráficas . . . . .	3
1.2. Independencia condicional . . . . .	5
1.2.1. Probabilidad condicional . . . . .	5
1.2.2. Modelos de independencia condicional . . . . .	7
1.2.3. Propiedades de Markov . . . . .	8
1.3. Modelos gráficos descomponibles . . . . .	10
1.3.1. Árboles de conglomerados . . . . .	16
<b>2. Modelos gráficos log-lineales</b>	<b>18</b>
2.1. Notación . . . . .	18
2.2. Distribución del número de observaciones en la tabla de contingencia . . . . .	20
2.3. Modelos log-lineales . . . . .	21
2.3.1. Interpretación de los parámetros . . . . .	24
2.3.2. Modelos log-lineales jerárquicos . . . . .	26
2.4. Modelos gráficos loglineales . . . . .	27
2.4.1. Estimación . . . . .	30
2.4.2. Bondad de Ajuste . . . . .	33
2.4.3. Selección del modelo . . . . .	37
2.5. Modelo log-lineal expresado como un GLM . . . . .	40
2.6. Relación de un modelo gráfico log-lineal y el modelo de regresión logística . . . . .	42
<b>3. Aplicación</b>	<b>46</b>
3.1. ENA 2011 . . . . .	46
3.2. Análisis exploratorio . . . . .	47
3.3. Modelos jerárquicos . . . . .	53
3.4. Modelos gráficos log-lineales . . . . .	54
3.4.1. Subpoblación que consume drogas . . . . .	59
3.4.2. Subpoblación que no consume drogas . . . . .	61
3.5. Regresiones logísticas asociadas al modelo gráfico log-lineal . . . . .	64
3.5.1. Comparación de probabilidades ajustadas . . . . .	69
<b>Conclusiones</b>	<b>73</b>

<i>ÍNDICE GENERAL</i>	VIII
<b>A. Diseño muestral de la ENA 2011</b>	<b>75</b>
<b>B. Código en R</b>	<b>83</b>
B.1. Esqueleto del diseño muestral ENA 2011 . . . . .	83
B.2. Ajuste de modelos gráficos log-lineales . . . . .	91
<b>Referencias</b>	<b>99</b>

# Introducción

El estudio de modelos gráficos log-lineales tiene sus inicios a principio de 1980's. Como antecedente, se puede mencionar a [Darroch et al. \(1980\)](#), quienes describen de manera detallada la relación entre una gráfica y un modelo log-lineal.

Posteriormente, el concepto de modelos gráficos se generaliza para distribuciones continuas. En particular, si tiene una distribución normal multivariada, se le conocen como *modelos gráficos Gaussianos*. Recientemente, esta área ha sido ampliamente estudiada. Asimismo, es posible modelar relaciones de dependencia condicional entre variables continuas y discretas, a estos modelos se les conocen como *modelos gráficos mixtos*.

En un modelo gráfico log-lineal se estudian las relaciones de independencia condicional y marginal para variables categóricas, por medio de propiedades de gráficas no dirigidas. Este tipo de análisis es adecuado cuando el número de parámetros es grande.

Algunos libros dedicados a modelos gráficos son [Lauritzen \(1996\)](#), [Whitaker \(1990\)](#) y [Edwards \(2000\)](#). Recientemente, en algunos libros clásicos de análisis de datos categóricos ha incluido alguna sección dedicada a modelos gráficos log-lineales, *e.g.* [Agresti \(2003\)](#) y [Christensen \(1990\)](#). En el caso de modelos gráficos gaussianos, [Anderson \(2003\)](#) incluye una sección dedicada a estos modelos.

En el presente trabajo, la información está organizada en 3 capítulos. El primero comprende la teoría necesaria para identificar y analizar un modelo gráfico probabilístico. Inicialmente, se definen algunos conceptos básicos de teoría de gráficas no dirigidas que se utilizarán más adelante. Enseguida, se establece la relación de independencia condicional de un vector aleatorio y su gráfica asociada mediante las tres propiedades de Markov. En este capítulo es de particular interés estudiar los modelos gráficos descomponibles. En estos modelos, la densidad conjunta del vector aleatorio puede expresarse en términos de la descomposición de su gráfica asociada. Para ello, se presentan los algoritmos *voraces* y de *búsqueda de máxima cardinalidad* que se utilizan para obtener la descomposición de la gráfica asociada al modelo gráfico descomponible.

El segundo capítulo se enfoca a presentar el ajuste de un modelo gráfico log-lineal. Dentro de este capítulo se introduce la notación empleada por [Lauritzen \(1996\)](#), la cual tiene la ventaja de ser más sencilla para tablas de contingencia de alta dimensión. Posteriormente, se presentan los modelos log-lineales, en particular, se muestran los posibles modelos asociados a una tabla de contingencia de dimensión 3. Posteriormente, se presentan los modelos log-lineales jerárquicos y los modelos gráficos log-lineales. Para estos últimos, se obtiene su ajuste por máxima verosimilitud, y se presentan estadísticas de bondad de ajuste del modelo. Adicionalmente, se muestra la definición de un modelo log-lineal como un modelo lineal generalizado (GLM), esto con el propósito de introdu-

cir los GLM que posteriormente se definen para las regresiones logísticas. Por último, se presenta la relación entre un modelo gráfico log-lineal y una regresión logística.

En el tercer capítulo se ilustra el uso de modelos gráficos para datos de la Encuesta Nacional de Adicciones (ENA) 2011. La base de datos utilizada para esta sección se encuentra disponible en la página electrónica del [Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz](#) (INSP), la cual consiste de 16,249 personas entrevistadas. En esta encuesta, uno de los dominios de estudio es la población adulta (18 a 65 años) y adolescente (12 a 17 años). Debido al diseño de la muestra, existe una correlación entre estas dos subpoblaciones, lo cual viola el supuesto de independencia de un modelo log-lineal. Por tal razón, se consideró sólo población adulta y se eliminaron datos faltantes. La muestra utilizada para el ajuste de modelos gráficos log-lineales quedó conformada por 10,294 adultos.

Se propusieron como posibles factores de riesgo las variables sexo, ingreso del hogar y escolaridad. Las variables de interés o de respuesta son el consumo de alcohol, tabaco y drogas. Debido a la dimensión del conjunto de variables y desconocimiento del tema, se ajustaron modelos jerárquicos y gráficos con base en los métodos de selección *stepwise*. En ambos casos, se observó que la tabla generada por las 6 variables era poco densa, esto debido a la baja prevalencia de consumo de drogas, por lo que se decidió hacer un análisis condicional a subpoblaciones que consumen o no drogas.

Finalmente, con el modelo gráfico obtenido con la muestra completa, se ajustaron regresiones logísticas para cada variable de interés donde los términos de la regresión están determinados por la gráfica asociada.

# Capítulo 1

## Modelos gráficos probabilísticos

El objetivo de este capítulo es presentar algunos elementos básicos de la teoría de modelos gráficos probabilísticos. Para ello en la primera sección se dan algunas definiciones de teoría de gráficas, en particular, sólo consideraremos gráficas no dirigidas. En la segunda sección, se introduce el concepto de independencia condicional para gráficas asociadas a un vector aleatorio. Finalmente, la última sección se enfoca a estudiar un subconjunto de modelos gráficos, denominados descomponibles.

### 1.1. Conceptos básicos de teoría de gráficas

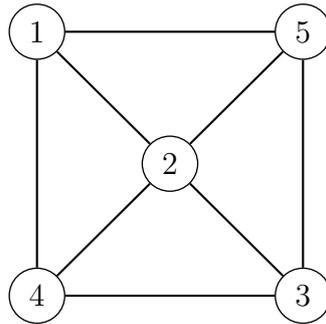
En esta sección se presenta algunas definiciones que se utilizarán más adelante para estudiar los modelos gráficos probabilísticos. Para una definición más formal ver [Bondy y Murty \(2008\)](#).

Una gráfica  $\mathcal{G} = (V, E)$  es un ente matemático definido por dos conjuntos, un conjunto de vértices,  $V$ , y un conjunto de aristas,  $E$ , conformado por un subconjunto de  $V \times V$ . Usualmente, los vértices en  $V$  toma valores en el conjunto de los naturales  $\{1, 2, \dots, |V|\}$ .

Por otra parte, se dice que existe una arista dirigida o *arco* entre los vértices  $\alpha$  y  $\beta$ , si el conjunto de aristas contiene a la pareja ordenada  $(\alpha, \beta)$ ; en este caso, el vértice  $\alpha$  es *padre* de  $\beta$ , y  $\beta$  es un *hijo* del vértice  $\alpha$ . En otro caso, existe una arista no dirigida o *línea* entre los vértices  $\alpha$  y  $\beta$ , si el conjunto de aristas contiene a las parejas  $(\alpha, \beta)$  y  $(\beta, \alpha)$ . La gráfica es *no dirigida* si todas sus aristas son no dirigidas ([Whitakker, 1990](#), p. 58).

Para este trabajo, sólo consideramos gráficas simples no dirigidas, esto es, gráficas donde no existen múltiples aristas para un mismo par de vértices, y no existen *bucles*, es decir, no hay aristas de la forma  $(\alpha, \alpha)$ , con  $\alpha \in V$ . Usualmente para representar una gráfica no dirigida  $\mathcal{G}$ , se utiliza un diagrama como se muestra en la figura 1.1, donde los vértices son representados por un círculo y las aristas por líneas.

Un vértice  $\alpha$  es *vecino* o *adyacente* al vértice  $\beta$ , si existe una arista entre ellos; en tal caso, se usa la notación  $\alpha \sim \beta$ . Si  $\alpha$  no es vecino de  $\beta$ , se denota por  $\alpha \not\sim \beta$ . En la figura 1.1, el vértice 1 es vecino de los vértices 2, 4 y 5, pero no es vecino de 3.



**Figura 1.1:** Diagrama de la gráfica  $\mathcal{G} = (V, E)$ , con vértices  $V = \{1, 2, 3, 4, 5\}$  y conjunto de aristas  $E = \{(1, 2), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5)\}$ .

Un *camino* de longitud  $n$  de un vértice  $\alpha$  a un vértice  $\beta$  es una sucesión de distintos vértices,  $\alpha = \alpha_0, \dots, \alpha_n = \beta$ , tales que  $(\alpha_{j-1}, \alpha_j) \in E$ , con  $j = 1, 2, \dots, n$ .

Un *ciclo* de longitud  $n$  es un camino donde el vértice inicial de la sucesión es igual al vértice final de la sucesión,  $\alpha_0 = \alpha_n$ . En este caso, cada par de vértices en la sucesión son adyacentes, y será un ciclo no enlazado (*chordless*), si no existen vértices adyacentes en la sucesión, salvo los vértices sucesivos; es decir, todo vértice en el ciclo tiene a lo más dos vértices adyacentes.

Dos vértices  $\alpha$  y  $\beta$  están *conectados* si existe un camino del vértice  $\alpha$  al vértice  $\beta$ , y será una gráfica *conectada* si todos los pares de vértices de  $V$  están conectados.

Un subconjunto de vértices,  $S \subset V$ , *separa* a los vértices  $\alpha$  y  $\beta$ , si todos los caminos que unen ambos vértices contienen al menos un vértice del conjunto separador  $S$ . Un subconjunto  $S \subset V$ , *separa* a dos subconjuntos de vértices de  $V$ ,  $A$  y  $B$ , si separa a todo par de vértices  $\alpha \in A$  y  $\beta \in B$ .

La *frontera* de un subconjunto  $A$ , denotada por  $\text{bd}(A)$ , es igual al conjunto de vértices que no pertenecen a  $A$  pero tienen un vértice adyacente para algún vértice en  $A$ , es decir

$$\text{bd}(A) = \{\beta \in V \setminus A \mid \exists \alpha \in A, \text{ tal que } (\alpha, \beta) \in E\}.$$

La *cerradura* de  $A$ , denotada por  $\text{cl}(A)$  es igual a la unión de  $A$  y su frontera,

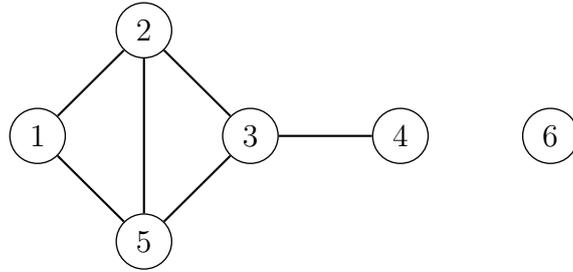
$$\text{cl}(A) = A \cup \text{bd}(A).$$

La subgráfica inducida por  $A$ ,  $\mathcal{G}_A$ , es la gráfica con conjunto vértices igual a  $A$  y subconjunto de aristas  $E_A$ . Donde  $E_A$  es el conjunto de aristas contenidas en  $E$  tales que conectan vértices en  $A$ .

Una gráfica o subgráfica es *completa* si todos sus vértices son vecinos mutuamente, y un *clique* es un subconjunto de vértices tales que su gráfica inducida es completa. Si el clique es tal que al agregar un vértice, la nueva gráfica inducida no es completa, entonces el clique es *maximal*. Usualmente, al conjunto de cliques se le denota por  $\mathcal{C}$ .

En la gráfica siguiente se ilustra cada uno de los conceptos presentados. A partir de ésta, se puede observar que existen múltiples caminos entre los vértices 1 y 3, siendo uno igual  $\{1, 2, 5, 3\}$ . El conjunto  $\{2, 5\}$  separa a los conjuntos  $\{1\}$  y  $\{3, 4\}$ , y el vértice 3 separa a los vértices 2 y 4. La frontera para el conjunto  $\{1, 5\}$ ,  $\text{bd}(\{1, 5\})$ , es igual al conjunto  $\{2, 3\}$ , y la frontera para el vértice 6,  $\text{bd}(\{6\})$ , es igual al conjunto vacío, por lo

tanto, la gráfica no es conectada. Los vértices  $\{1, 5, 3, 2, 1\}$  forman un ciclo enlazado, ya que el vértice 5 tiene 3 vecinos, los vértices 1, 2 y 3; mientras, los vértices  $\{1, 5, 2\}$  forman un ciclo no enlazado de longitud 3. El conjunto  $\{2, 5\}$  es un clique de la gráfica, pero no es maximal ya que al agregar el vértice 1, la gráfica inducida por  $\{1, 2, 5\}$  es completa. Los cliques maximales para esta gráfica están dados por los subconjuntos  $\{1, 2, 5\}$ ,  $\{2, 3, 4\}$ ,  $\{3, 4\}$  y  $\{6\}$ . Es fácil notar que la gráfica no es completa, ya que 6 no es vecino de los demás vértices.



**Figura 1.2:** Gráfica no dirigida  $\mathcal{G} = (V, E)$ , con conjunto de vértices  $V = \{1, 2, 3, 4, 5, 6\}$  y conjunto de aristas  $E = \{(1, 2), (1, 5), (2, 3), (2, 5), (3, 4), (3, 5)\}$ .

La unión e intersección de dos gráficas simples no dirigidas,  $(V, E)$  y  $(W, F)$ , es igual a la unión e intersección de sus conjuntos de vértices y aristas:

$$(V, E) \cup (W, F) = (V \cup W, E \cup F) \quad (V, E) \cap (W, F) = (V \cap W, E \cap F)$$

Para ilustrar la anterior, consideremos las gráficas  $\mathcal{G}_1 = (V, E)$  y  $\mathcal{G}_2 = (W, F)$ , dadas en las figuras 1.3a y 1.3b, respectivamente. La unión e intersección para las gráficas  $\mathcal{G}_1$  y  $\mathcal{G}_2$  están dadas por las figuras 1.3c y 1.3d.

## 1.2. Independencia condicional

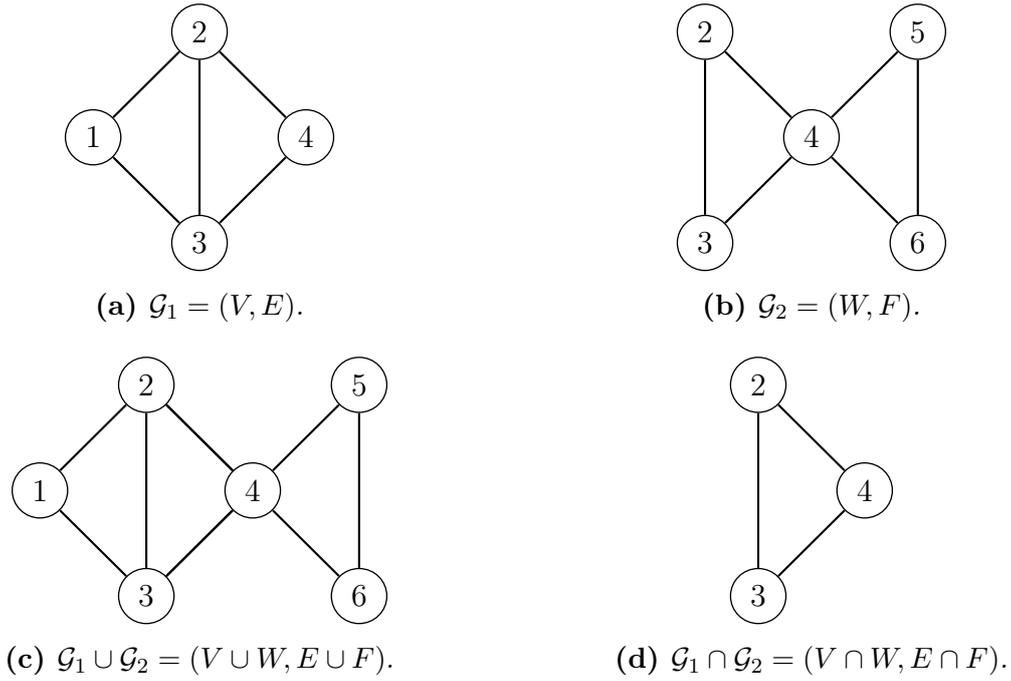
Un elemento esencial en modelos gráficos probabilísticos es el concepto de independencia condicional. En Lauritzen (1996, sec. 3) se estudia de manera detallada la relación entre teoría de gráficas e independencia condicional, la cual define un modelo gráfico probabilístico. Algunos de los principales resultados concernientes a estos modelos se presentan a continuación.

### 1.2.1. Probabilidad condicional

Sean  $X$  y  $Y$  variables aleatorias, se dice que  $X$  y  $Y$  son condicionalmente independientes dada la variable aleatoria  $Z$  si su función de verosimilitud puede expresarse como

$$\mathcal{L}(X, Y, Z) = \mathcal{L}(X|Y, Z)\mathcal{L}(X|Z). \quad (1.1)$$

En este caso, usaremos la notación  $X \perp\!\!\!\perp Y|Z$  ó  $X \perp\!\!\!\perp_P Y|Z$ , si se desea enfatizar que depende sobre una medida de probabilidad  $P$ . Si  $X$  es condicionalmente independiente de  $Y$  dado



**Figura 1.3:** Unión e intersección de las gráficas  $\mathcal{G}_1 = (V, E)$  y  $\mathcal{G}_2 = (W, F)$ . Donde  $V = \{1, 2, 3, 4\}$ ,  $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$ ,  $W = \{2, 3, 4, 5, 6\}$  y  $F = \{(2, 3), (2, 4), (3, 4), (4, 5), (4, 6), (5, 6)\}$ .

$Z$ ,  $X \perp\!\!\!\perp Y | Z$ , entonces la función de densidad conjunta de  $X$  y  $Y$  admite la factorización siguiente.

$$f(x, y, z)f(z) = f(x, z)f(y, z). \quad (1.2)$$

En este caso, la función de densidad puede expresarse como el producto de dos funciones,  $a(x, z)$  y  $b(y, z)$ ,

$$f(x, y, z) = a(x, z)b(y, z).$$

Otras propiedades para variables aleatorias condicionalmente independientes son:

- (C1) Si  $X \perp\!\!\!\perp Y | Z$  entonces  $Y \perp\!\!\!\perp X | Z$ .
- (C2) Si  $X \perp\!\!\!\perp Y | Z$  y  $U = g(Y)$  entonces  $X \perp\!\!\!\perp U | Z$ .
- (C3) Si  $X \perp\!\!\!\perp Y | Z$  y  $U = g(Y)$  entonces  $X \perp\!\!\!\perp Y | (Z, U)$ .
- (C4) Si  $X \perp\!\!\!\perp Y | Z$  y  $X \perp\!\!\!\perp W | (Y, Z)$  entonces  $X \perp\!\!\!\perp (Y, W) | Z$ .

Donde  $g(\cdot)$  es una función medible sobre el espacio muestral de  $X$ . Además, si la función de distribución conjunta tiene función de densidad estrictamente positiva, se cumple que

- (C5) Si  $X \perp\!\!\!\perp Y | (Z, W)$  y  $X \perp\!\!\!\perp Z \perp\!\!\!\perp (Y, W)$  entonces  $X \perp\!\!\!\perp (Y, Z) | W$ .

En cierto modo, la relación de independencia condicional puede ser interpretada como una codificación de lo que es *relevante* (Lauritzen, 1996, p. 30). Si a la relación  $A \perp\!\!\!\perp B | C$  le damos la interpretación: dado que conocemos  $C$ ,  $A$  es irrelevante para inferir sobre  $B$ , las propiedades (C1)-(C5) se traducen a:

- (I1) Dado que conocemos  $Z$ , inferir sobre  $X$  es irrelevante para inferir  $Y$ ; entonces inferir sobre  $Y$  es irrelevante para inferir  $X$ .
- (I2) Dado que se tiene información acerca de  $Z$ , inferir sobre  $X$  es irrelevante para inferir  $Y$ ; entonces  $X$  es irrelevante para inferir cualquier función que dependa de  $Y$ .
- (I3) Dado que se tiene información acerca de  $Z$ , inferir sobre  $X$  es irrelevante para inferir  $Y$ ; entonces  $X$  sigue siendo irrelevante después de haber inferido sobre cualquier función de  $Y$ .
- (I4) Dado que se tiene información acerca de  $Z$ , inferir sobre  $X$  es irrelevante para inferir  $Y$ ; si después de haber inferido para  $Y$ ,  $X$  es irrelevante para inferir  $W$ , entonces  $X$  es irrelevante para inferir  $(Y, W)$ .

A partir de las propiedades (C1)-(C4) se pueden establecer axiomas para analizar independencia condicional en modelos más generales, los cuales denominaremos como *grafoides*.

### 1.2.2. Modelos de independencia condicional

Lauritzen (2006, pp. 6-7) define una relación  $\perp_\sigma$  como *grafoide*, si para  $A, B, C, D$  subconjuntos disjuntos de un conjunto finito  $V$  satisface lo siguiente:

- (S1) Si  $A \perp_\sigma B|C$ , entonces  $B \perp_\sigma A|C$ .
- (S2) Si  $A \perp_\sigma B|C$  y  $D \subseteq B$ , entonces  $A \perp_\sigma D|C$ .
- (S3) Si  $A \perp_\sigma B|C$  y  $D \subseteq B$ , entonces  $A \perp_\sigma B|(C \cup D)$ .
- (S4) Si  $A \perp_\sigma B|C$  y  $A \perp_\sigma D|(B \cup C)$ , entonces  $A \perp_\sigma (B \cup D)|C$ .
- (S5) Si  $A \perp_\sigma B|(C \cup D)$  y  $A \perp_\sigma C|(B \cup D)$  entonces  $A \perp_\sigma (B \cup C)|D$ .

En caso de que sólo se cumplan las condiciones (S1)-(S4), se le denomina *semi-grafoide*. Por ejemplo, para la relación  $A \perp_G B|S$ , la cual denota la separación de la gráfica  $\mathcal{G}$  en  $A$  y  $B$  por medio de  $S$ , es fácil verificar que las propiedades (S1)-(S5) se satisfacen, por lo tanto, la relación  $\perp_G$  es un *grafoide*.

Ahora, si consideramos  $V$  un conjunto finito y  $\{X_v : v \in V\}$  una colección de variables aleatorias. Entonces para  $A, B$  y  $C$  subconjuntos de  $V$ , podemos definir los vectores aleatorios marginales  $X_A = \{X_v : v \in A\}$ ,  $X_B = \{X_v : v \in B\}$  y  $X_C = \{X_v : v \in C\}$ , tales que  $X_A \perp\!\!\!\perp X_B | X_C$ . En este caso, se puede probar que la relación  $\perp\!\!\!\perp$  es semi-grafoide. Además, si la función de densidad conjunta de  $X$  es estrictamente positiva, entonces la relación de independencia condicional  $\perp\!\!\!\perp$  es un grafoide. Con el propósito de usar la notación de modelos gráficos, se tiene la equivalencia.

$$X_A \perp\!\!\!\perp X_B | X_C \Leftrightarrow A \perp\!\!\!\perp B | C.$$

De acuerdo con esto, un vector aleatorio  $X = (X_1, X_2, \dots, X_n)$  se puede representar mediante una gráfica  $\mathcal{G} = (V, E)$ , con  $V = \{1, 2, \dots, n\}$  y donde  $(i, j)$  no es una arista

de  $E$  si y sólo si  $X_i \perp\!\!\!\perp X_j | X_{V \setminus \{i,j\}}$ , es decir, existe una arista entre los vértices  $(i, j)$  si existe una relación de dependencia entre ambos. En particular, [Whitaker \(1990, p. 60\)](#) denomina a esta gráfica como *gráfica independiente condicional*. De esta manera, usando las propiedades de separación de una gráfica  $\mathcal{G}$  se identifican independencias condicionales para un vector aleatorio  $X$ . En general, para una gráfica  $\mathcal{G}$ , esta puede presentar 3 tipos de independencia condicional, las cuales se describen a continuación.

### 1.2.3. Propiedades de Markov

Sea  $\mathcal{G} = (V, E)$  una gráfica simple no dirigida, donde  $V$  es el conjunto de vértices asociados a un vector aleatorio  $X$  y  $E$  el conjunto de aristas. Decimos que la relación de independencia condicional  $\perp\!\!\!\perp$  satisface

(P) *La propiedad de Markov por parejas* con respecto a  $\mathcal{G}$ , si

$$\alpha \not\sim \beta \Rightarrow \alpha \perp\!\!\!\perp \beta | V \setminus \{\alpha, \beta\}. \quad (1.3)$$

Esto implica que para vértices no adyacentes, éstos son condicionalmente independiente dado el resto de las variables. En la figura 1.4, el vértice 4 no es adyacente del vértice 5, entonces se cumple que  $4 \perp\!\!\!\perp 5 | \{1, 2, 3, 6, 7\}$ .

(L) *La propiedad de Markov Local* con respecto a  $\mathcal{G}$ , si para toda  $\alpha \in V$ ,

$$\alpha \perp\!\!\!\perp V \setminus \text{cl}(\alpha) | \text{bd}(\alpha). \quad (1.4)$$

Es decir, todo vértice es condicionalmente independiente de sus vértices no adyacentes dada su frontera. En la figura 1.4, la frontera de 5, es igual al conjunto  $\text{bd}(5) = \{2, 3, 6, 7\}$ , y el complemento de la cerradura igual  $V \setminus \text{cl}(5) = \{1, 4\}$ , entonces se cumple que  $5 \perp\!\!\!\perp \{1, 4\} | \{2, 3, 6, 7\}$ .

(G) *La propiedad de Markov Global* con respecto a  $\mathcal{G}$ , si para  $A, B, S$  subconjuntos disjuntos de  $V$  tales que  $S$  separa  $A$  de  $B$  en  $\mathcal{G}$ , entonces

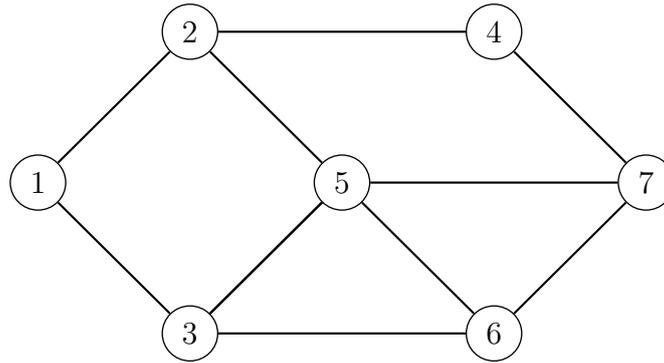
$$A \perp\!\!\!\perp B | S. \quad (1.5)$$

En la gráfica 1.4 se observa que los subconjuntos  $\{1, 3\}$  y  $\{4, 7\}$  están separados por  $\{2, 5, 6\}$ . Por lo tanto,  $\{1, 3\} \perp\!\!\!\perp \{4, 7\} | \{2, 5, 6\}$ .

Para las tres propiedades enunciadas, [Lauritzen \(1996, p. 33\)](#) demuestra que para toda gráfica no dirigida  $\mathcal{G} = (V, E)$ , y vector aleatorio  $X$ , la propiedad de Markov Global (G) implica la propiedad de Markov local (L), y ésta a su vez implica la propiedad de Markov por parejas (P).

Además, si la función densidad de  $X$  es estrictamente positiva, entonces las tres propiedades de Markov son equivalentes [Lauritzen \(1996, p. 34\)](#). A este resultado se le conoce como el teorema de *Pearl y Paz*.

Alternativamente, [Matúš \(1992\)](#) prueba la equivalencia de las tres propiedades de Markov utilizando gráficas duales. En esta prueba, cada propiedad tiene asociada una



**Figura 1.4:** Gráfica no dirigida,  $\mathcal{G} = (V, E)$ . Donde  $V = \{1, 2, 3, 4, 5, 6, 7\}$  y  $E = \{(1, 2), (1, 3), (2, 3), (2, 5), (3, 5), (3, 6), (4, 7), (5, 6), (5, 7), (6, 7)\}$ .

clase de gráfica dual. Bajo los supuestos de la función de densidad, demuestra que las tres clases de gráficas duales son equivalentes.

Por otra parte, existe una relación entre la función de densidad conjunta de  $X$  y su gráfica asociada  $\mathcal{G}$ , la cual está dada por la definición siguiente.

Sea  $f$  la función de densidad de  $X$  y  $\mathcal{G} = (V, E)$  su gráfica asociada. Se dice que  $f$  admite una factorización (F), si para todos los subconjuntos completos  $a \subset V$ , existen funciones  $\psi_a(x_a)$  estrictamente positivas que dependen de  $x$  sólo a través de  $x_a$  tales que la función de densidad está dada por:

$$f(x) = \prod_{a \in \mathcal{A}} \psi_a(x_a), \tag{1.6}$$

donde  $\mathcal{A}$  es el conjunto de cliques de la gráfica  $\mathcal{G}$ .

En la figura 1.4, los cliques de la gráfica son iguales a los conjuntos  $\mathcal{A} = \{\{3, 5, 6\}, \{5, 6, 7\}, \{1, 2\}, \{1, 3\}, \{2, 5\}, \{2, 4\}, \{4, 7\}\}$ , de modo que la función de densidad conjunta,  $f$ , admite la factorización:

$$f(x_1, x_2, x_3, x_4, x_6, x_7) = \psi_{356}(x_3, x_5, x_6) \psi_{567}(x_5, x_6, x_7) \psi_{12}(x_1, x_2) \\ \times \psi_{13}(x_1, x_3) \psi_{25}(x_2, x_5) \psi_{24}(x_2, x_4) \psi_{47}(x_4, x_7).$$

A partir de la definición anterior, Lauritzen (1996, pág. 34) demuestra que para toda factorización (F) con respecto a  $\mathcal{G}$ , ésta satisface la propiedad de Markov Global (G) con respecto a  $\mathcal{G}$ . En consecuencia, la factorización también satisface la propiedad de Markov local (L) y por parejas (P).

Un resultado más general e importante en modelos gráficos, prueba que si la función de densidad de  $X$ ,  $f$ , es estrictamente positiva, entonces la implicación es válida en ambos sentidos, es decir, si la función de densidad de  $X$  es estrictamente positiva, entonces ésta satisface la propiedad de Markov por parejas en  $\mathcal{G}$ , si y solo si admite una factorización con respecto  $\mathcal{G}$ . A este resultado se le conoce como el *teorema de Hammersley y Clifford*. Hammersley y Clifford demostraron este teorema por primera vez para variables aleatorias discretas. Posteriormente, se han dado múltiples pruebas de este teorema que generalizan para variables aleatorias continuas, ver *e.g.* Besag (1974) y Clifford (1990).

### 1.3. Modelos gráficos descomponibles

Dentro de los modelos gráficos existe un subconjunto de estos, denominados *descomponibles*. Estos modelos tienen la ventaja de que su factorización puede expresarse en función de densidades marginales, las cuales están asociadas a la descomposición de su gráfica. Para ello, primero definamos la descomposición de una gráfica  $\mathcal{G}$ .

Sea  $\mathcal{G} = (V, E)$  una gráfica condicionalmente independiente, se dice que una partición de  $V$ ,  $(A, B, S)$ , es una *descomposición* de  $\mathcal{G}$  si se cumple que:

1.  $A \perp_{\mathcal{G}} B | S$ ,
2.  $S$  es completo

La descomposición es propia si  $A \neq \emptyset$  y  $B \neq \emptyset$ . En este caso, los componentes de la descomposición están dadas por las subgráficas inducidas por  $\mathcal{G}_{A \cup S}$  y  $\mathcal{G}_{B \cup S}$ . Una gráfica se dice que es *principal* si no existe una descomposición propia (Lauritzen, 2006, p. 25). En particular, la gráfica principal para un gráfica completa es ella misma. En la figura 1.5 se muestran ejemplos de gráficas principales.

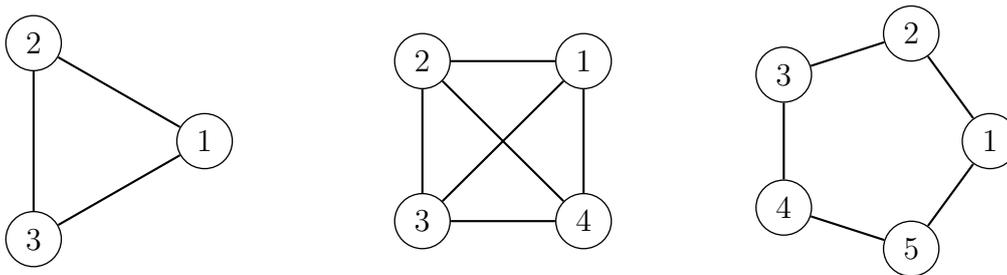
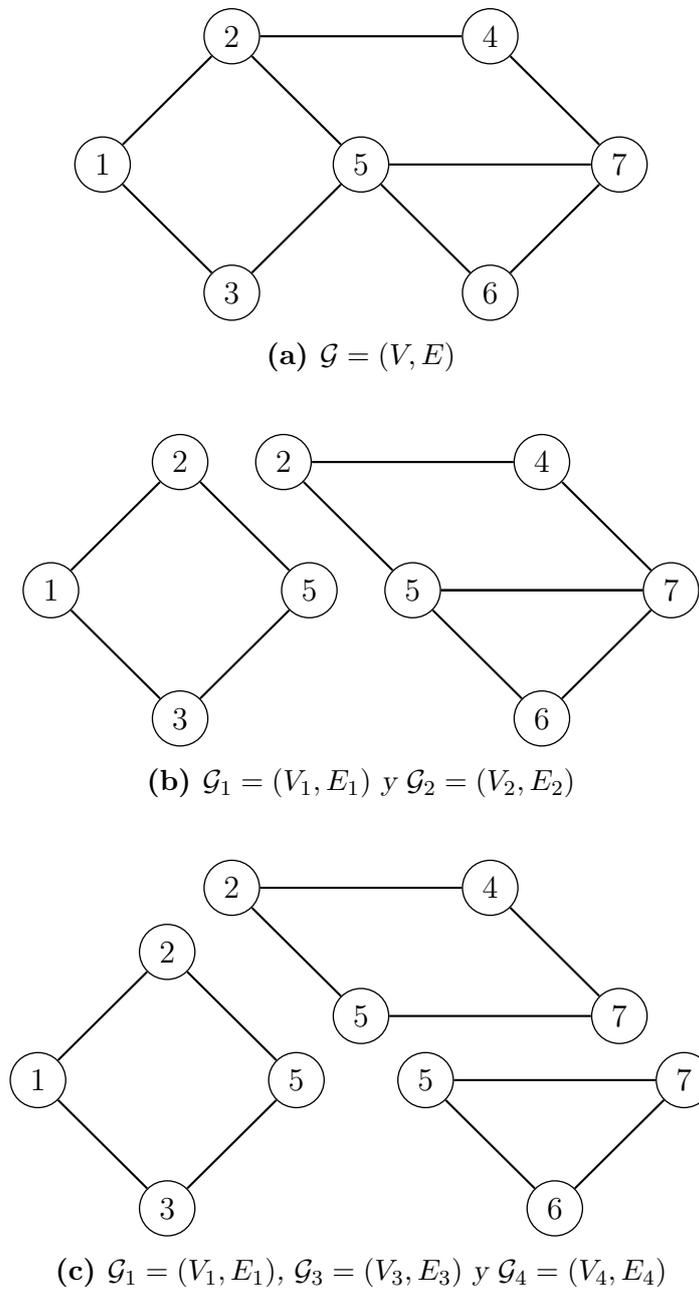


Figura 1.5: Ejemplos de gráficas principales.

Cabe mencionar que para gráficas donde hay más de dos sub-gráficas principales, éstas se pueden obtener al descomponer la gráfica de manera recursiva. Por ejemplo, en la figura 1.6a, se observa que para la primera descomposición de  $\mathcal{G}$ , el conjunto separador es igual a  $S_1 = \{2, 5\}$ , los componentes de la primera descomposición están dadas por  $\mathcal{G}_1 = (V_1, E_1)$  y  $\mathcal{G}_2 = (V_2, E_2)$ , donde  $V_1 = \{1, 2, 3, 5\}$  y  $V_2 = \{2, 4, 5, 6, 7\}$  (figura 1.6b). Ahora para estas dos subgráficas, se observa que  $\mathcal{G}_1$  es una sub-gráfica principal, mientras que para la sub-gráfica  $\mathcal{G}_2$ , esta tiene una descomposición con conjunto separador  $S_2 = \{5, 7\}$ , donde los componentes de la descomposición están dadas por las sub-gráficas  $\mathcal{G}_3 = (V_3, E_3)$  y  $\mathcal{G}_4 = (V_4, E_4)$ , donde  $V_3 = \{2, 4, 5, 7\}$  y  $V_4 = \{5, 6, 7\}$  (figura 1.6c). Para las sub-gráficas obtenidas en la segunda descomposición se observa que todas son principales, por lo tanto, podemos concluir que la descomposición de la gráfica  $\mathcal{G}$  tiene tres sub-gráficas principales dadas por  $\mathcal{G}_1, \mathcal{G}_3$  y  $\mathcal{G}_4$ , con conjuntos separadores  $S_1 = \{2, 5\}$  y  $S_2 = \{5, 7\}$ .

Una gráfica es *descomponible* si todas sus sub-gráficas principales son cliques. Es claro que en la gráfica anterior  $\mathcal{G}$  no es descomponible, ya que al menos dos de sus sub-gráficas principales no son completas.



**Figura 1.6:** Descomposición de una gráfica en subgráficas principales. (a) Gráfica  $\mathcal{G} = (V, E)$  con  $V = \{1, 2, 3, 4, 5, 6, 7\}$ , (b) primera descomposición de la gráfica  $\mathcal{G} = (V, E)$ , donde el conjunto separador es  $S_1 = \{2, 5\}$  y subgráficas principales  $\mathcal{G}_1 = (V_1, E_1)$  y  $\mathcal{G}_2 = (V_2, E_2)$  con  $V_1 = \{1, 2, 3, 5\}$  y  $V_2 = \{2, 4, 5, 6, 7\}$ , (c) segunda descomposición de  $\mathcal{G} = (V, E)$ , con conjunto separador  $S_2 = \{5, 7\}$  y subgráficas principales  $\mathcal{G}_1 = (V_1, E_1)$ ,  $\mathcal{G}_3 = (V_3, E_3)$  y  $\mathcal{G}_4 = (V_4, E_4)$ , con  $V_1 = \{1, 2, 3, 5\}$ ,  $V_3 = \{2, 4, 5, 7\}$  y  $V_4 = \{5, 6, 7\}$

**Propiedades de Markov en descomposición de gráficas**

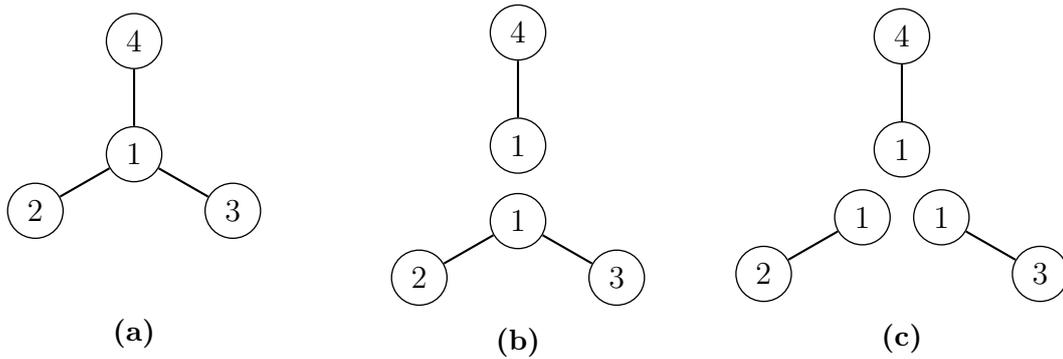
Sea  $(A, B, S)$  una descomposición de  $\mathcal{G}$  y  $P$  una distribución de probabilidad. Se dice que  $P$  se factoriza con respecto a  $\mathcal{G}$ , si y sólo si, se satisface las siguientes condiciones:

- i)  $P_{AUS}$  y  $P_{BUS}$  se factorizan con respecto a  $\mathcal{G}_{AUS}$  y  $\mathcal{G}_{BUS}$ ,
- ii)  $f(x)f(x_S) = f(x_{AUS})f(x_{BUS})$ .

Este resultado se puede aplicar de manera recursiva para gráficas con más de dos sub-gráficas principales. De manera que para gráficas descomponibles, la factorización de su función de densidad  $f$ , con respecto a  $\mathcal{G}$  está dada por:

$$f(x) \prod_{S \in \mathcal{S}} f(x_S)^{\nu(S)} = \prod_{c \in \mathcal{C}} f(x_C), \quad (1.7)$$

donde  $\mathcal{C}$  es el conjunto de cliques de  $\mathcal{G}$ ,  $\mathcal{S}$  el conjunto de separadores y  $\nu(S)$  el número de veces que aparece el conjunto separador  $S$  en la descomposición (Lauritzen, 2006, p. 26). Para ilustrar lo anterior, consideremos la descomposición de la gráfica  $\mathcal{G}$  en la figura 1.7.



**Figura 1.7:** (a) Descomposición de la gráfica  $\mathcal{G} = (V, E)$ , donde  $V = \{1, 2, 3, 4\}$  y  $E = \{(1, 2), (1, 3), (1, 4)\}$ . (b) Primera descomposición de  $\mathcal{G}$ , con conjunto separador  $S = \{1\}$  y componentes  $\mathcal{G}_1 = (V_1, E_1)$  y  $\mathcal{G}_2 = (V_2, E_2)$ , donde  $V_1 = \{1, 4\}$  y  $V_2 = \{1, 2, 3\}$ . (c) Subgráficas principales de  $\mathcal{G}$ :  $\mathcal{G}_1 = (V_1, E_1)$ ,  $\mathcal{G}_3 = (V_3, E_3)$  y  $\mathcal{G}_4 = (V_4, E_4)$ , donde  $V_3 = \{1, 2\}$  y  $V_4 = \{1, 3\}$ .

A partir de esta gráfica, podemos ver que el conjunto separador,  $S = \{1\}$ , aparece 2 veces en la descomposición. En la primera descomposición, el conjunto  $S$  separa a los vértices  $\{2, 3\}$  del vértice 4 (figura 1.7b); en la segunda descomposición, el mismo conjunto  $S$  separa los vértices  $\{2\}$  y  $\{3\}$ , obteniéndose así todas las sub-gráficas principales de  $\mathcal{G}$  (figura 1.7c). En este caso, usando (1.7), la factorización para  $X$  está dada por:

$$f(x_1, x_2, x_3, x_4) = \frac{f(x_1, x_2)f(x_1, x_3)f(x_1, x_4)}{[f(x_1)]^2}.$$

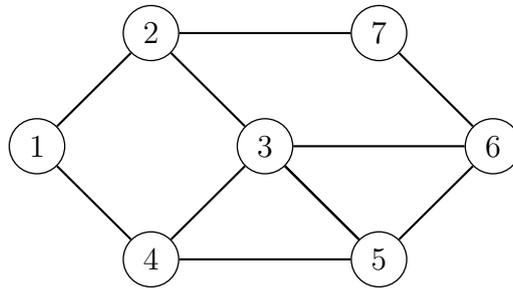
Cuando la cardinalidad del conjunto de vértices es grande, la descomposición de la gráfica es un problema  $NP$ -complejo, es decir, conforme el número de vértices crece, el número de operaciones necesarias para obtener la descomposición de la gráfica crece de manera exponencial. Afortunadamente, existen algunos algoritmos de búsqueda basados en el concepto de *triangulación* y *ordenamiento* que permite evaluar si una gráfica es

descomponible, y en algunos casos obtener los cliques y conjuntos separadores de la descomposición.

Un ordenamiento de vértices  $V = \{1, 2, \dots, |V|\}$  de una gráfica no dirigida  $\mathcal{G}$  se dice que es perfecto si se cumple que:

$$\forall j = 2, \dots, |V|, \quad \text{bd}(j) \cap \{1, \dots, j-1\} \text{ es completa en } \mathcal{G}. \quad (1.8)$$

Por ejemplo, en la figura 1.8 se muestra una ordenamiento de una gráfica  $\mathcal{G}$ . Para el vértice  $j = 5$ ,  $\text{bd}(5) = \{3, 4, 6\}$ , en consecuencia,  $\text{bd}(5) \cap \{1, 2, 3, 4\} = \{3, 4\}$ , el cual es un conjunto completo en  $\mathcal{G}$ ; sin embargo, para  $j = 7$ ,  $\text{bd}(7) = \{2, 6\}$ , de modo que  $\text{bd}(7) \cap \{1, 2, 3, 4, 5, 6\} = \{2, 6\}$ . Este último conjunto no es completo en  $\mathcal{G}$ , ya que 2 no es adyacente a 6, por lo tanto, el ordenamiento no es perfecto.



**Figura 1.8:** Un ordenamiento de la gráfica  $\mathcal{G}$  (figura 1.4).

Una gráfica no dirigida  $\mathcal{G}$  se dice que es *triangulada* si no contiene ciclos de longitud mayor o igual a 4 como sub-gráfica principal. El resultado principal para gráficas descomponibles está dado por la siguiente proposición.

Para una gráfica no dirigida  $\mathcal{G}$ , las siguientes condiciones son equivalentes:

- i)  $\mathcal{G}$  es triangulada.
- ii)  $\mathcal{G}$  es descomponible.
- iii) Todas las sub-gráficas principales de  $\mathcal{G}$  son cliques máximas.
- iv)  $\mathcal{G}$  admite un ordenamiento perfecto.
- v) Para todo par de vértices  $\alpha$  y  $\beta$  en  $V$ , su conjunto separador minimal es completo.

Con base en este resultado, se han propuesto algoritmos para evaluar si una gráfica es descomponible. Entre los más utilizados se encuentra el algoritmo glotón y el algoritmo de búsqueda de máxima cardinalidad, descritos en Lauritzen (2002, pp. 28-30).

---

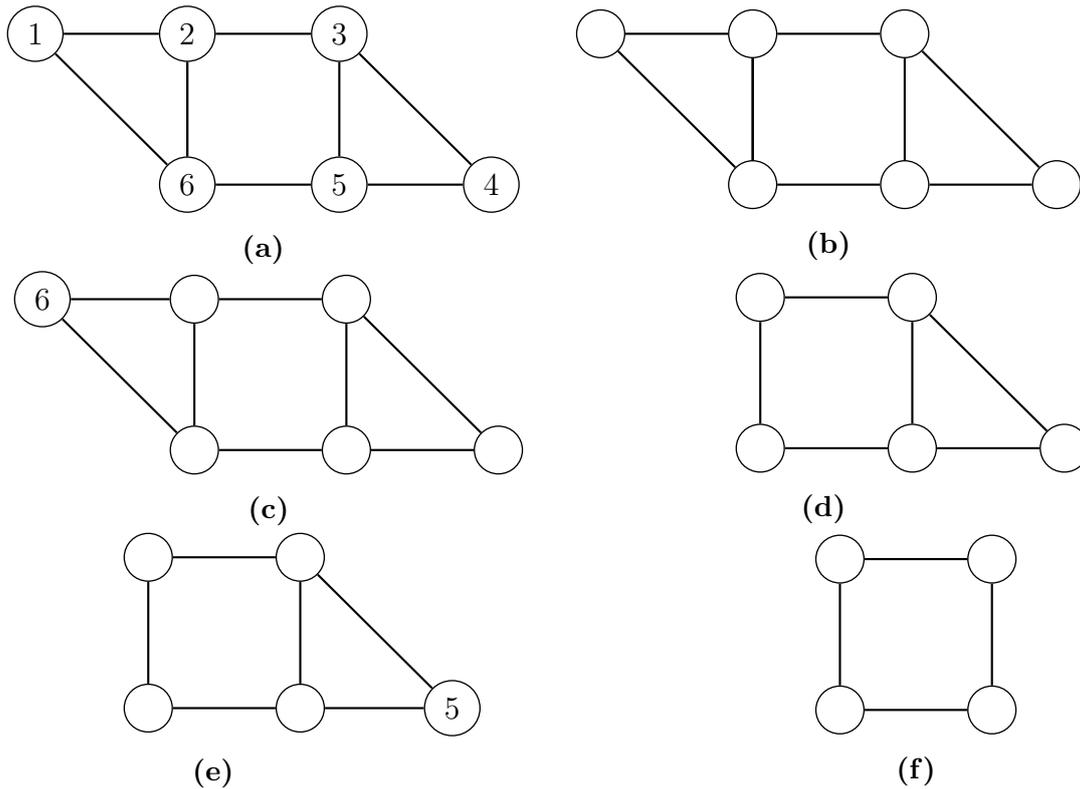
### Algoritmo glotón

---

1. Se selecciona un nodo  $v^*$  en  $\mathcal{G}$ , tal que  $\text{bd}(v^*)$  es completa.  
*Si no existe  $v^*$ , entonces  $\mathcal{G}$  no es triangulada (termina el algoritmo).*
  2. Se forma la sub-gráfica  $\mathcal{G}_{V \setminus v^*}$  y etiqueta  $v^* = |V'|$ , donde  $|V'|$  es el número de nodos en  $\mathcal{G}_{V \setminus v^*}$ .
  3. Se repite el paso 1 con  $\mathcal{G} = \mathcal{G}_{V \setminus v^*}$ .
  4. Si el algoritmo continua hasta que  $v^* = 1$ , entonces la gráfica es triangulada.
- 

Los algoritmos glotonos son métodos búsqueda exhaustiva con complejidad igual a  $O(|V|^2)$ . Para ilustrar este algoritmo, consideremos la gráfica dada por la figura 1.9a.

De acuerdo al algoritmo, primero debemos considerar la gráfica sin etiquetar, como se muestra en la figura 1.9b. Enseguida, seleccionamos un nodo arbitrario y lo etiquetamos con la cardinalidad del conjunto de vértices (figura 1.9c). Si la frontera del vértice es completa, removemos el vértice de  $\mathcal{G}$  (figura 1.9d). Nuevamente, seleccionamos un vértice arbitrario tal que su frontera es completa y lo etiquetamos de manera descendente (figura 1.9e). Para el nuevo vértice etiquetado, evaluamos si su frontera es completa y repetimos el paso 2 del algoritmo (figura 1.9f). En este último paso del algoritmo, la gráfica obtenida no tiene vértices tales que su frontera sea completa, por lo tanto no es triangulada.



**Figura 1.9:** Ilustración del algoritmo glotón para la gráfica  $\mathcal{G} = (V, E)$ , con conjunto de vértices  $V = \{1, 2, 3, 4, 5, 6\}$  y aristas  $E = \{(1, 2), (1, 6), (2, 3), (2, 6), (3, 4), (3, 5), (4, 5), (5, 6)\}$ .

---

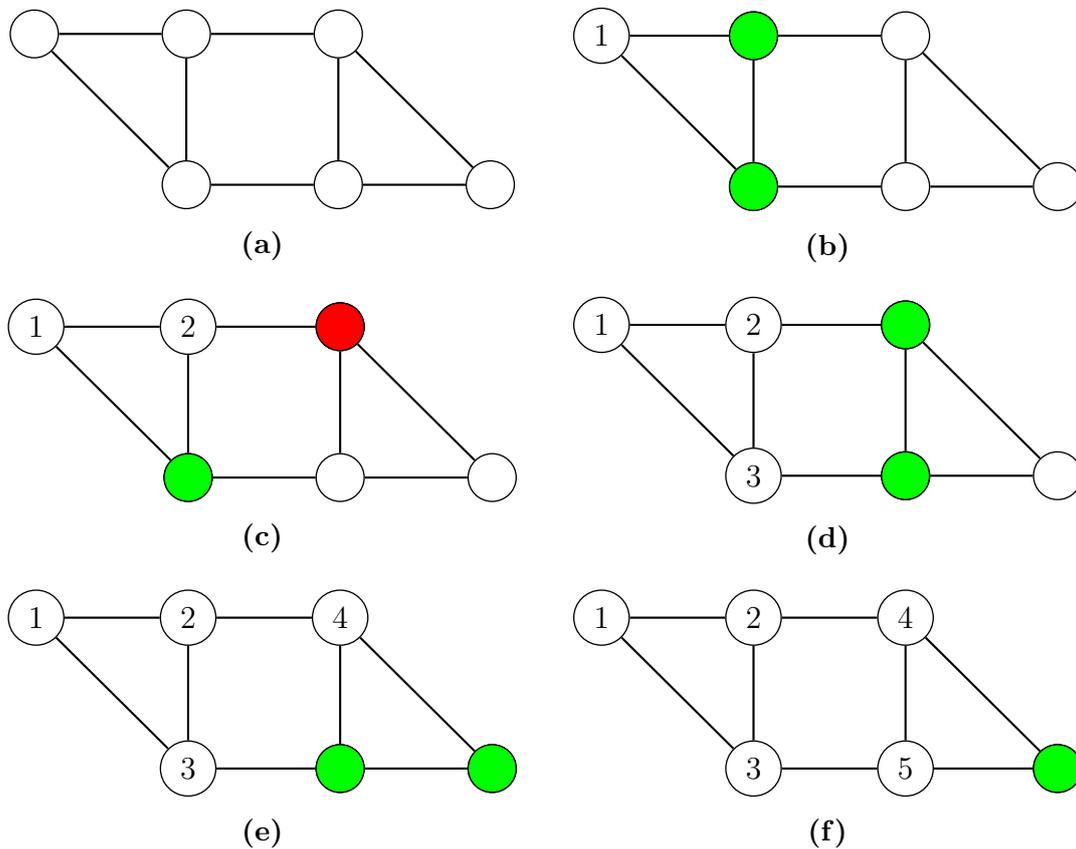
### Algoritmo de búsqueda de máxima cardinalidad (MCS)

---

1. Se selecciona un nodo arbitrario  $v_0 \in \mathcal{G}$  y se etiqueta  $v_0 = 1$ .
  2. Para cada  $j = 1, \dots, n - 1$ .
    - i) Se selecciona un nodo de  $V \setminus \{1, 2, \dots, j\}$  tal que, tenga el mayor número de vecinos etiquetados. En caso de empates, se resuelve de manera arbitraria.
    - ii) Se etiqueta el nodo seleccionado como  $j + 1$ .
    - iii) Si  $bd(j + 1) \cap \{1, 2, \dots, j\}$  no es completa, entonces  $\mathcal{G}$  no es triangulada (termina).
    - iv) De lo contrario, repita i).
  3. Si el algoritmo continúa hasta que sólo queda uno de los vértices por etiquetar, la gráfica es triangulada y la numeración es perfecta.
-

La complejidad para el algoritmo de búsqueda de máxima cardinalidad es  $O(|E|+|V|)$  y fue propuesto por [Tarjan y Yannakakis \(1984\)](#).

En la figura 1.10 se ilustra el método de máxima cardinalidad. En el paso 1 del algoritmo, etiquetamos de manera arbitraria un nodo inicial (figura 1.10b), los vértices en verde indican aquellos con vecinos etiquetados. Para el paso 2 del algoritmo, seleccionamos el vértice con mayor número de vecinos etiquetados, en este caso, existe un empate, por lo tanto, seleccionamos de manera arbitraria alguno de los vértices en verde y etiquetamos de manera ascendente como se muestra en la figura 1.10c. Una vez etiquetado, observamos que  $bd(2) \cap 1 = \{1\}$  es un subconjunto completo, por lo tanto, podemos seleccionar el siguiente vértice con el mayor número de vecinos etiquetados (figura 1.10d). Para el vértice 3,  $bd(3) = \{1, 2\}$ , por consiguiente,  $bd(3) \cap \{1, 2\} = \{1, 2\}$ , el cual es un subconjunto completo en  $\mathcal{G}$ . Nuevamente, etiquetamos el vértice con el mayor número de vecinos etiquetados (figura 1.10e). En el último paso del algoritmo (figura 1.10f), se observa que para el último vértice etiquetado 5,  $bd(5) = \{3, 4\}$  y  $bd(5) \cap \{1, 2, 3, 4\} = \{3, 4\}$ , el cual no es un conjunto completo en  $\mathcal{G}$ , por lo tanto, no es descomponible.



**Figura 1.10:** Algoritmo de Búsqueda de Máxima cardinalidad. Los vértices en rojo, indican los vértices con vecinos etiquetados, en tanto que los vértices en verde indican los vértices con el mayor número de vecinos etiquetados.

### Búsqueda de cliques en una gráfica triangulada

Para una búsqueda de máxima cardinalidad (MCS) y un ordenamiento  $V = \{1, 2, 3, \dots, |V|\}$ , se define la función

$$B_\lambda = \text{bd}(\lambda) \cap \{1, 2, \dots, \lambda - 1\}, \quad (1.9)$$

y  $|B_\lambda| = \pi_\lambda$ . Al vértice  $\lambda$  se le denomina vértice escalón (*ladder*), si

$$\lambda = |V| \quad \text{o} \quad \pi_{\lambda+1} < \pi_\lambda + 1.$$

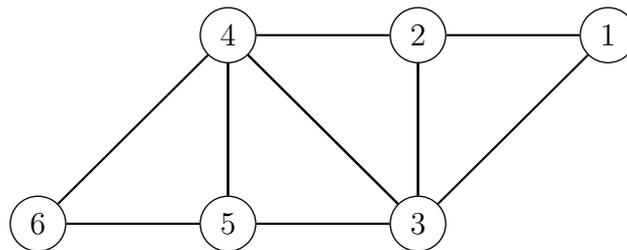
Si  $\Lambda$  es el conjunto de vértices escalón, los cliques de la gráfica están dados por:

$$C_\lambda = \{\lambda\} + B_\lambda, \quad \lambda \in \Lambda. \quad (1.10)$$

En la figura 1.11, se muestra la numeración obtenida con el algoritmo de máxima cardinalidad para una gráfica  $\mathcal{G}$ . En este caso, se obtuvo para

$$\begin{aligned} \lambda = 1, & \quad B_1 = \{\emptyset\}, \\ \lambda = 2, & \quad B_2 = \{1\}, \\ \lambda = 3, & \quad B_3 = \{1, 2\}, \\ \lambda = 4, & \quad B_4 = \{2, 3\}, \\ \lambda = 5, & \quad B_5 = \{3, 4\}, \\ \lambda = 6, & \quad B_6 = \{4, 5\}. \end{aligned}$$

De modo que  $\pi_\lambda = (0, 1, 2, 2, 2, 2)$ , esto implica que los vértices escalón son  $\{3, 4, 5, 6\}$ . Por lo tanto, los cliques de la gráfica están dados por  $\{\{1, 2, 3\}, \{2, 3, 4\}, \{3, 4, 5\}, \{4, 5, 6\}\}$ .



**Figura 1.11:** Numeración obtenida por el algoritmo de máxima cardinalidad.

#### 1.3.1. Árboles de conglomerados

Sea  $\mathcal{A}$  una colección finita de subconjuntos de  $V$ . Un árbol de conglomerado (*junction tree*),  $\tau$ , es una gráfica no dirigida, donde el conjunto de vértices es igual a los subconjuntos de  $\mathcal{A}$ , y el conjunto de aristas igual a las intersecciones de los subconjuntos en  $\mathcal{A}$ . Si los conjuntos en  $\mathcal{A}$  no son comparables por pares, entonces existe un árbol de conglomerados en  $\mathcal{A}$  si y sólo si  $\mathcal{A} = \mathcal{C}$ , donde  $\mathcal{C}$  es el conjunto de cliques de una gráfica triangulada.

Partiendo de esta definición, el árbol de conglomerados se puede construir directamente del algoritmo MCS de la siguiente manera.

Para todo vértice  $\lambda > \lambda_{\min}$ , en donde  $\lambda_{\min}$  corresponde al vértice escalón minimal de un ordenamiento de vértices, existe un conjunto  $S_\lambda$  tal que satisface lo siguiente

$$S_\lambda = C_\lambda \cap (\cup_{\lambda' < \lambda} C_{\lambda'}) = C_\lambda \cap C_{\lambda^*}, \quad \text{para algún } \lambda^* < \lambda, \quad (1.11)$$

a  $S_\lambda$  se le conoce como un conjunto separador de la gráfica. Por ejemplo, en el ordenamiento mostrado en la figura 1.11, los vértices escalón son  $(3, 4, 5, 6)$ , con cliques asociados iguales a

$$C_3 = \{1, 2, 3\},$$

$$C_4 = \{2, 3, 4\},$$

$$C_5 = \{3, 4, 5\},$$

$$C_6 = \{4, 5, 6\}.$$

Por lo tanto,  $\lambda_{\min} = 3$ . Para toda  $\lambda > 3$ , los conjuntos separadores de la gráfica son iguales a:

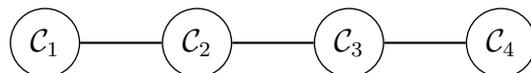
$$S_4 = C_4 \cap C_3 = \{2, 3\},$$

$$S_5 = C_5 \cap (C_3 \cup C_4) = C_5 \cap C_3 = \{3, 4\},$$

$$S_6 = C_6 \cap (C_3 \cup C_4 \cup C_5) = C_6 \cap C_5 = \{4, 5\}.$$

En la práctica, los elementos del conjunto de cliques se etiquetan de acuerdo con los números naturales. En este caso, se tendría el conjunto de cliques  $\mathcal{C}_1 = \{1, 2, 3\}$ ,  $\mathcal{C}_2 = \{2, 3, 4\}$ ,  $\mathcal{C}_3 = \{3, 4, 5\}$  y  $\mathcal{C}_4 = \{4, 5, 6\}$ . De manera análoga, se tiene los conjuntos separadores  $\mathcal{S}_1 = \{2, 3\}$ ,  $\mathcal{S}_2 = \{3, 4\}$  y  $\mathcal{S}_3 = \{4, 5\}$ .

El *junction tree* se define como la unión de  $C_\lambda$  y conjuntos  $C_{\lambda^*}$ , tales que  $\lambda^* < \lambda$ . En la figura 1.12 se muestra el árbol de conglomerados para la gráfica dada en la figura 1.11.



**Figura 1.12:** Junction tree.

Es importante observar que pueden existir distintos valores de  $\lambda^*$ , tales que se obtenga el mismo conjunto separador. Para ello, se define la multiplicidad del conjunto separador como:

$$\nu(s) = |\{\lambda \in \Lambda : S_\lambda = S\}|,$$

es decir,  $\nu(s)$  es el número de veces en que el conjunto separador  $S$  aparece en la descomposición de la gráfica.

En el capítulo siguiente, se presentan los modelos loglineales gráficos. Estos últimos son una clase de modelos gráficos probabilísticos, donde las variables aleatorias asociados a los vértices en  $V$  toman valores en un conjunto discreto.

# Capítulo 2

## Modelos gráficos log-lineales

Los modelos loglineales permiten identificar algunas relaciones de asociación entre las variables categóricas. En particular, el objetivo de este trabajo es estudiar algunos modelos log-lineales que pueden ser representados por medio de una gráfica  $\mathcal{G}$ , denominados *modelos gráficos log-lineales*. Estos modelos tienen la ventaja de identificar relaciones de independencia condicional usando propiedades de teoría de gráficas.

En la primera sección, introduciremos la notación empleada por [Lauritzen \(1996\)](#) para modelos gráficos log-lineales. En la segunda sección, se presentan algunos modelos log-lineales que usualmente se utilizan, los cuales tienen una gráfica asociada  $\mathcal{G}$ . En la tercera sección, se presenta el ajuste por máxima verosimilitud para un modelo log-lineal gráfico. Finalmente en la última sección, se describe la relación entre un modelo gráfico log-lineal y un modelo de regresión logística.

### 2.1. Notación

Sea  $I = (I_\delta)_{\delta \in V}$  un conjunto de variables aleatorias discretas asociadas a los vértices de la gráfica  $\mathcal{G} = (V, E)$ . En particular, cada variable  $I_\delta$  puede identificarse por el vértice  $\delta$  en la gráfica  $\mathcal{G}$ . Asimismo definimos el conjunto  $I_\delta$ , como el conjunto de valores o niveles de categoría que toma la variable  $I_\delta$ . Cada celda  $i$  en la tabla de contingencia está dada por los valores que toman simultáneamente las variables, es decir

$$i \in I = \prod_{\delta \in V} I_\delta,$$

donde  $I$  es el conjunto de todas las posibles combinaciones de  $I$ . En tanto,  $n(i)$  denota el número de observaciones clasificadas de la celda  $i$ , las cuales se asumen son provenientes de una variable aleatoria de conteo,  $N(i)$ . La esperanza del número de observaciones o frecuencia de la celda  $i$  se denota por  $m(i) = E(N(i))$ . Asimismo, la dimensión de la tabla de contingencia es igual a la cardinalidad del conjunto de vértices,  $|V|$ .

Ahora bien, si sólo consideramos información marginal, es decir, sólo nos interesa el número de observaciones clasificadas para un subconjunto de variables  $a \subset V$ , definimos la celda marginal

$$i_a \in I_a = \prod_{\delta \in a} I_\delta,$$

donde  $I_a$  es el conjunto de todas las posibles combinaciones para el vector marginal  $I_a$ . Si  $\mathbf{n}' = (n(i))_{i \in I}$  es el conjunto de observaciones para cada celda  $i$ , entonces el número de observaciones para la celda  $i_a$  es igual a la suma de  $n(i)$ 's correspondientes a todas las celdas con valor  $i_a$  para los elementos en  $a$ , esto es

$$n_a(i_a) = \sum_{j:j_a=i_a} n(j).$$

Análogamente, la frecuencia marginal esperada para  $i_a$  está dada por

$$m_a(i_a) = \sum_{j:j_a=i_a} m(j).$$

En el caso de que se tenga plenamente identificado el conjunto  $a$ , las expresiones  $n_a(i_a)$  y  $m_a(i_a)$  se reducen a  $n(i_a)$  y  $m(i_a)$ , respectivamente.

Para ilustrar la notación anterior, consideremos la siguiente tabla de contingencia de seis dimensiones, tomada del libro de [Edwards \(2000, pág. 22\)](#). Los datos corresponden a un estudio relacionado a los factores de riesgo asociados a las afecciones cardíacas de una población de 1,841 trabajadores del transporte en Checoslovaquia.

				B		si		
				no		sí		
				A	no	sí	no	sí
F	E	D	C					
No	No	No	No	44	40	112	67	
			Sí	129	145	12	63	
		Sí	No	35	12	80	33	
		Sí	Sí	109	67	7	9	
	Si	No	No	23	32	70	66	
			Sí	50	80	7	13	
		Sí	No	24	25	73	57	
		Sí	Sí	51	63	7	16	
Sí	No	No	No	5	7	21	9	
			Sí	9	17	1	4	
		Sí	No	4	3	11	8	
		Sí	Sí	14	17	5	2	
	Si	No	No	7	3	14	14	
			Sí	9	16	2	3	
		Sí	No	4	0	13	11	
		Sí	Sí	5	14	4	4	

**Tabla 2.1:** Factores de riesgo asociados a afecciones cardíacas. Donde A, fumador; B, trabajo mentalmente estresante; C, trabajo físicamente estresante; D, presión sanguínea sistólica alta; E, cociente de lipo-proteínas beta y alfa alta; F, historia familiar de afecciones cardíacas

En este ejemplo, hay seis variables:  $A, B, C, D, E$  y  $F$ , donde el conjunto de vértices asociados es igual a  $V = \{A, B, C, D, E, F\}$ , con categorías

$$I_A = I_B = I_C = I_D = I_E = I_F = \{\text{sí, no}\}.$$

De modo que  $I = (I_A, I_B, I_C, I_D, I_E, I_F)$  denota el vector aleatorio formado por las variables de clasificación. Mientras que,  $i = (i_A, i_B, i_C, i_D, i_E, i_F)$  es una combinación observada de  $I$ , con un total de  $2^6$  posibles resultados. Ahora, si definimos el subconjunto  $a = \{A, B, E, F\} \subset V$ , entonces  $I_a = \{I_A, I_B, I_E, I_F\}$  es el vector aleatorio marginal definido por las variables aleatorias en  $a$ . En ese caso, si  $i_a = (\text{sí}, \text{sí}, \text{sí}, \text{sí})$  es una configuración de  $I_a$ ,  $n_a(i_a) = 14 + 3 + 11 + 4 = 32$  indica el número de individuos tales que: son fumadores, tienen estrés físico y mental, y además tienen historia familiar de afecciones. Usando la notación punto clásica (Agresti, 2003; Bishop et al., 2007), la expresión equivalente a  $n_a(i_a)$  es igual a

$$n_{ab..ef} = \sum_{c \in I_C} \sum_{d \in I_D} n_{abcdef}.$$

En esta notación, la letra  $a$  indica el valor que toma la variable  $A$ . Con este ejemplo se puede observar que la notación es manejable cuando la dimensión de la tabla de contingencia se incrementa.

## 2.2. Distribución del número de observaciones en la tabla de contingencia

Como se mencionó anteriormente, el número de observaciones en una celda  $i$ ,  $n(i)$ , provienen de una variable de conteo,  $N(i)$ . Generalmente, esta variable aleatoria de conteo se asume que es observada bajo los siguientes esquemas de muestreo (Lauritzen, 1996, pp. 69-71).

1. *Poisson*. Bajo este esquema, se asume que el número de observaciones,  $n(i)$ , de la celda  $i$  en la tabla de contingencia, son una realización de variables aleatorias independientes con una distribución Poisson,  $N(i)$ , con media  $\lambda(i) = m(i)$ . En este caso, la distribución conjunta de las observaciones,  $\{N(i)\}_{i \in I}$  es igual a:

$$P(N(i) = n(i), i \in I) = \prod_{i \in I} \frac{m(i)^{n(i)}}{n(i)!} e^{-m(i)}. \quad (2.1)$$

Bajo este esquema, se tiene la restricción  $m(i) \geq 0$ .

2. *Multinomial*. En este esquema, se asume que el número de total de observaciones,  $|n|$ , es fijo. En tal caso, el número de observaciones en la tabla de contingencia se suponen independientes, donde la probabilidad de que un individuo pertenezca a la celda  $i$  es igual a  $p(i)$ . La distribución conjunta para  $\{N(i)\}_{i \in I}$  es multinomial:

$$P(N(i) = n(i), i \in I) = \frac{|n|!}{\prod_{i \in I} n(i)!} \prod_{i \in I} p(i)^{n(i)}. \quad (2.2)$$

Donde  $p(i) \geq 0$  y  $\sum_{i \in I} p(i) = 1$ . De modo que  $\sum_{i \in I} m(i) = |n|$ , la frecuencia esperada en la celda  $i$  es igual a  $m(i) = |n|p(i)$ , con  $m(i) \geq 0$ .

3. *Multinomial restringido* o *Multinomial producto*. Bajo este esquema, se tiene que para un subconjunto  $b \subset \Delta$ , los totales marginales  $n(i_b)$  están fijos. En este caso, cada vector marginal correspondiente a los posibles valores que puede tomar  $i_b$  tiene una distribución multinomial. Por ejemplo, en una tabla de dimensión 2, si fijamos el número de observaciones por renglones, entonces cada renglón tiene una distribución multinomial. La probabilidad de pertenecer a una celda  $i_a$  de  $i_b$  es  $p(i_a|i_b)$  y la distribución conjunta para  $\{N(i)\}_{i \in I}$  está dada por

$$P(N(i) = n(i), i \in I) = \prod_{i_b \in I_b} \left\{ \frac{n(i_b)!}{\prod_{i_a \in I_a} n(i)!} \prod_{i_a \in I_a} p(i_a|i_b)^{n(i)} \right\}. \quad (2.3)$$

La frecuencia en la celda  $i$  es igual a  $m(i) = n(i_b)p(i_a|i_b)$ , donde  $n(i_b) = m(i_b)$ . De manera análoga al caso multinomial se tiene las restricciones  $\sum_{i_a} p(i_a|i_b) = 1$ , con  $p(i_a|i_b) \geq 0$  y  $m(i) \geq 0$ .

Si en el esquema Poisson condicionamos a que el total de observaciones  $|N|$  sea igual a  $|n|$ , entonces se obtiene una distribución multinomial. De manera similar, si en el esquema multinomial condicionamos a que  $N(i_b) = n(i_b)$ , obtenemos el esquema multinomial restringido.

## 2.3. Modelos log-lineales

Un modelo log-lineal se define de la siguiente forma. Dada una especificación de  $p(i)$ , existen funciones  $u_a(i_a)$  que dependen sólo de  $a$  a través de  $i_a$ , tales que el logaritmo de la frecuencia esperada,  $m(i)$ , en la celda  $i$  es igual a:

$$\log m(i) = \sum_{a \subseteq \Delta} u_a(i_a). \quad (2.4)$$

Donde  $\Delta$  es el conjunto de variables categóricas. En particular, [Lauritzen \(2002, p. 8\)](#) denomina a éstos como modelos log-lineales de interacción. El término log-lineal se deriva del hecho de que el logaritmo de  $m(i)$  está dado en función de  $u_a(i_a)$ , de manera lineal. Si  $a = \emptyset$ , entonces  $u_a$  es un valor constante. A las funciones  $u_a$  se les denomina las interacciones de las variables  $I_a$  en  $a$  de orden  $|a| - 1$ . Si el orden es igual a 1, entonces se les denomina efectos principales ([Darroch et al., 1980](#)). Las interacciones del modelo están determinadas por las relaciones de dependencia de las variables en  $a$ , y como se verá más adelante, son interpretables en términos de los momios.

En el caso de tres variables aleatorias,  $\{I_1, I_2, I_3\}$ , y bajo el supuesto de que el número de observaciones en la tabla de contingencia tiene una distribución multinomial es posible ajustar alguno de los siguientes modelos log-lineales.

1. *Modelo de independencia*. Si  $I_1, I_2$  y  $I_3$  son variables aleatorias independientes, la probabilidad de pertenecer a la celda  $i$  es igual a:

$$p(i) = p(i_1, i_2, i_3) = p(i_1)p(i_2)p(i_3).$$

En este caso, el modelo log-lineal asociado a  $p(i)$  está dado por la siguiente expresión:

$$\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3). \quad (2.5)$$

Para probar lo anterior, sólo basta expandir el logaritmo de  $p(i)$ :

$$\log p(i) = \log p(i_1) + \log p(i_2) + \log p(i_3),$$

donde  $p(i) = m(i)/|n|$ ,  $p(i_1) = m(i_1)/|n|$ ,  $p(i_2) = m(i_2)/|n|$  y  $p(i_3) = m(i_3)/|n|$ . De modo que la frecuencia en la celda  $i$ ,  $m(i)$ , está determinada por una constante y de funciones que sólo dependen de  $i_1$ ,  $i_2$  y  $i_3$ .

2. *Modelo de independencia marginal.* Supongamos que  $I_1$  es marginalmente independiente de  $I_2$  y  $I_3$ , la probabilidad de pertenecer a la celda  $i$  es igual a:

$$p(i) = p(i_1, i_2, i_3) = p(i_1)p(i_2, i_3).$$

En este caso, el modelo log-lineal asociado a  $p(i)$  tiene la forma:

$$\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{23}(i_2, i_3). \quad (2.6)$$

Para probarlo, basta ver que:

$$p(i_1)p(i_2, i_3) = \frac{m(i_1)}{|n|} \frac{m(i_2, i_3)}{|n|},$$

bajo el modelo en (2.6), la frecuencia para  $m(i_1)$  y  $m(i_2, i_3)$  son iguales a:

$$\begin{aligned} m(i_2, i_3) &= \sum_{i_1} m(i) = \exp\{u + u_2(i_2) + u_3(i_3) + u_{23}(i_2, i_3)\} \sum_{i_1} \exp\{u_1(i_1)\}, \\ m(i_1) &= \sum_{i_2, i_3} m(i) = \exp\{u_1(i_1)\} \sum_{i_2, i_3} \exp\{u + u_2(i_2) + u_3(i_3) + u_{23}(i_2, i_3)\}. \end{aligned}$$

por lo tanto, se sigue que

$$\begin{aligned} p(i_1)p(i_2, i_3) &= \frac{1}{|n|^2} \exp\{u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{23}(i_2, i_3)\} \sum_{i_1, i_2, i_3} m(i) \\ &= \frac{1}{|n|} \exp\{u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{23}(i_2, i_3)\} |n| \\ &= p(i). \end{aligned}$$

En donde bajo el supuesto de distribución multinomial, se satisface que la suma de los valores observados es igual a la suma de valores observados. De manera análoga, se definen los modelos de independencia marginal para las variables  $I_2$  ó  $I_3$ .

3. *Modelo de independencia condicional.* Si  $I_1$  y  $I_2$  son condicionalmente independientes dado  $I_3$ , la probabilidad de la celda  $i$  es igual a:

$$p(i) = p(i_1, i_2 | i_3)p(i_3) = p(i_1 | i_3)p(i_2 | i_3)p(i_3) = \frac{p(i_1, i_3)p(i_2, i_3)}{p(i_3)},$$

En este caso, el modelo log-lineal asociado a  $p(i)$  está dado por:

$$\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{13}(i_1, i_3) + u_{23}(i_2, i_3), \quad (2.7)$$

Para probarlo, sólo basta ver que:

$$\frac{p(i_1, i_3)p(i_2, i_3)}{p(i_3)} = \frac{1}{|n|} \frac{m(i_1, i_3)m(i_2, i_3)}{m(i_3)} = \frac{1}{|n|} \frac{\sum_{i_2} m(i) \sum_{i_1} m(i)}{\sum_{i_1, i_2} m(i)}, \quad (2.8)$$

donde

$$\begin{aligned} \sum_{i_2} m(i) &= \exp\{u + u_1(i_1) + u_3(i_3) + u_{13}(i_1, i_3)\} \sum_{i_2} \exp\{u_2(i_2) + u_{23}(i_2, i_3)\}, \\ \sum_{i_1} m(i) &= \exp\{u + u_1(i_2) + u_3(i_3) + u_{23}(i_2, i_3)\} \sum_{i_1} \exp\{u_1(i_1) + u_{13}(i_1, i_3)\}, \\ \sum_{i_1, i_2} m(i) &= \exp\{u + u_3(i_3)\} \sum_{i_1, i_2} \exp\{u_1(i_1) + u_2(i_2) + u_{13}(i_1, i_3) + u_{23}(i_2, i_3)\}, \end{aligned}$$

si sustituimos las ecuaciones anteriores en (2.8), se obtiene la condición para  $p(i)$ :

$$\frac{p(i_1, i_3)p(i_2, i_3)}{p(i_3)} = \frac{1}{|n|} m(i) = p(i). \quad (2.9)$$

De manera similar, se pueden definir los modelos de independencia condicional dadas las variables  $I_1$  o  $I_2$ .

4. *Modelo de asociación homogénea.* Bajo este modelo, todo par de variables son condicionalmente independientes dada una tercera variable. El modelo log-lineal asociado tiene todos los términos de primer orden:

$$\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{12}(i_1, i_2) + u_{13}(i_1, i_3) + u_{23}(i_2, i_3), \quad (2.10)$$

Si a la expresión anterior le aplicamos la función exponencial, la probabilidad de la celda  $i$  tiene la forma:

$$p(i) = \phi(i_1, i_2)\psi(i_2, i_3)\omega(i_1, i_3),$$

Donde  $\phi(i_1, i_2)$ ,  $\psi(i_2, i_3)$ ,  $\omega(i_1, i_3)$  son funciones que depende de dos variables. En este caso, no existe una expresión cerrada de  $p(i)$  en términos de sus densidades marginales (Agresti, 2003, p. 320).

5. *Modelo saturado.* Si  $I_1$ ,  $I_2$  y  $I_3$  son conjuntamente dependientes, el modelo log-lineal asociado está dado por:

$$\begin{aligned} \log m(i) &= u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{12}(i_1, i_2) \\ &\quad + u_{13}(i_1, i_3) + u_{23}(i_2, i_3) + u_{123}(i_1, i_2, i_3). \end{aligned} \quad (2.11)$$

En este modelo se obtiene un ajuste perfecto para la tabla de contingencia.

Así como en análisis de varianza (ANOVA), en los modelos log-lineales existen parámetros redundantes. Para obtener una de las soluciones, se agregan restricciones sobre los parámetros. Usualmente se utiliza alguna de las siguientes restricciones:

- i) *Variables dummy*. Para cada variable categórica  $a \in \Delta$ , se define un nivel de categoría como referencia. En este caso, el parámetro asociado es cero.
- ii) *Efecto de codificación*. Con esta restricción, la suma sobre todos los valores en las interacciones son igual a cero.

Por ejemplo, supongamos un modelo de independencia para dos variables binarias

$$\log m(i) = u + u_1(i_1) + u_2(i_2).$$

En forma matricial, el modelo log-lineal puede expresarse como:

$$\begin{bmatrix} \log m(1,1) \\ \log m(1,0) \\ \log m(0,1) \\ \log m(0,0) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} [u \quad u_1(1) \quad u_1(0) \quad u_2(1) \quad u_2(0)]^T.$$

La matriz con entradas cero y uno se le conoce como la matriz diseño del modelo. Si no hay restricciones sobre los parámetros, el rango de esta matriz es incompleto.

Para las variables dummy, definimos el nivel 1 de cada variable como de referencia,  $u_1(1) = u_2(1) = 0$ . En este caso, la matriz de diseño está dada por:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Por otra parte, si utilizamos el efecto de codificación,  $\sum_{i_1} u_1(i_1) = \sum_{i_2} u_2(i_2) = 0$ . Entonces la matriz de diseño queda expresada como

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & -1 \end{bmatrix}$$

### 2.3.1. Interpretación de los parámetros

En general para modelos log-lineales, la interpretación de las interacciones en el modelo es más fácil si se da en función de momios (en el caso de efectos principales) y de la razón de momios (interacciones de orden mayor o igual a 2). Para ello, un momio se define como el cociente de la probabilidad de éxito y su probabilidad de fracaso (Agresti, 2003, p. 44),

$$\Omega = \frac{p}{1-p}.$$

Para dos eventos supongamos que sus momios correspondientes están dados por  $\Omega_1$  y  $\Omega_2$ . Una manera de comparar éstos, es utilizando la razón de momios:

$$\theta = \frac{\Omega_1}{\Omega_2}.$$

En el modelo de independencia (2.5), la interacción máxima está dada por sus efectos principales. En este caso, estos están asociados a los momios para cada variable. Por ejemplo, supongamos que  $I_1$  es una variable binaria; dado el modelo de independencia, su momio es igual:

$$\Omega_1 = \frac{p(1, i_2, i_3)}{p(0, i_2, i_3)} = \frac{m(1, i_2, i_3)}{m(0, i_2, i_3)} = \exp\{u_1(1) - u_1(0)\}.$$

Dependiendo de las restricciones del modelo, el valor de  $\exp\{u_1(1) - u_1(0)\}$  indica qué tan probable es que la variable  $I_1$  tome el valor 1, con respecto al valor 0 (Agresti, 2003, p. 315). Como se observa, el momio no depende de los valores que tomen las demás variables.

En otro caso, si consideramos el modelo de independencia marginal en (2.6), las interacciones de primer orden están asociados a la razón de momios condicionales (Agresti, 2003, p. 321). Por ejemplo, si  $I_2$  y  $I_3$  son variables binarias; dado un valor de  $I_1$ , la razón de momios condicional es igual a

$$\theta_{(i_1)I_2I_3} = \frac{p(i_1, 1, 1)/p(i_1, 0, 1)}{p(i_1, 1, 0)/p(i_1, 0, 0)} = \frac{m(i_1, 1, 1)m(i_1, 0, 0)}{m(i_1, 0, 1)m(i_1, 1, 0)}.$$

Si a la expresión anterior le aplicamos la función logaritmo, obtenemos que:

$$\log \theta_{(i_1)I_2I_3} = u_{23}(1, 1) + u_{23}(0, 0) - u_{23}(0, 1) - u_{23}(1, 0),$$

en este caso, los parámetros  $u_{23}(i_2, i_3)$  determinan la asociación entre las variables  $I_2$  y  $I_3$ , las cuales no dependen del valor de  $I_1$ . Lo anterior refleja la independencia marginal de  $I_1$  y  $\{I_2, I_3\}$ .

De manera análoga, en el modelo de independencia condicional (2.7), las interacciones  $u_{13}(i_1, i_3)$  y  $u_{23}(i_2, i_3)$  están relacionadas con la razón de momios condicionales  $\theta_{I_1(i_2)I_3}$  y  $\theta_{(i_1)I_2I_3}$ . En este modelo, la razón de momios condicional para un valor fijo de  $I_3$ ,  $\theta_{I_1I_2(i_3)}$ , es igual a 1, esto proviene del hecho que  $I_1$  y  $I_2$  son condicionalmente independientes dado  $I_3$  (Christensen, 1990, p. 85). Adicionalmente, si la razón de momios condicionales  $\theta_{I_1(i_2)I_3}$  y  $\theta_{(i_1)I_2I_3}$  son iguales a 1, esto indica que  $I_1 \perp\!\!\!\perp I_3 | I_2$  y  $I_3 \perp\!\!\!\perp I_2 | I_1$ . En este caso, el modelo a ajustar sería el de independencia.

En un modelo saturado (2.11), la interacción de segundo orden,  $u_{123}(i_1, i_2, i_3)$  está determinado por el cambio entre la razón de momios condicionales. Por ejemplo si  $I_1, I_2, I_3$  son variables binarias, el cambio en la razón de momios condicionales con respecto a  $I_3$  está dada por:

$$\frac{\theta_{I_1I_2(1)}}{\theta_{I_1I_2(0)}} = \frac{m(1, 1, 1)m(0, 0, 1)}{m(0, 1, 1)m(1, 0, 1)} \bigg/ \frac{m(1, 1, 0)m(0, 0, 0)}{m(0, 1, 0)m(1, 0, 0)}.$$

Si aplicamos la función logaritmo, obtenemos lo siguiente:

$$\begin{aligned} \log \frac{\theta_{I_1I_2(1)}}{\theta_{I_1I_2(0)}} &= [u_{123}(0, 0, 0) + u_{123}(1, 1, 0) - u_{123}(0, 1, 0) - u_{123}(1, 0, 0)] \\ &\quad - [u_{123}(0, 0, 1) + u_{123}(1, 1, 1) - u_{123}(0, 1, 1) + u_{123}(1, 0, 1)]. \end{aligned}$$

Si el cociente de la razón de momios es igual a 1, entonces la interacción de segundo orden es cero, esto indica que el modelo que se debe ajustar es el de asociación homogénea. En el caso más general, donde las variables toman más de dos valores, la interpretación de las interacciones están en términos de la *razón de momios locales* (Agresti, 2003, p. 55). Para términos de interacción mayor, la interpretación consiste en obtener de manera sucesiva cocientes razones de momios. Como se observa, la interpretación se dificulta.

### 2.3.2. Modelos log-lineales jerárquicos

Un modelo log-lineal jerárquico es un modelo log-lineal, donde el conjunto de interacciones satisface lo siguiente, si  $u_a(i_a)$  está determinado como nulo en el modelo, es decir, no existe interacciones de orden  $|a| - 1$ , entonces para cualquier subconjunto  $b \supseteq a$ , las interacciones  $u_b(i_b)$  son nulas. La expresión para un modelo jerárquico está dada por:

$$\log m(i) = \sum_{a \in K} u_a(i_a), \tag{2.12}$$

en donde  $K$  es el conjunto de subconjuntos de interacciones maximales del modelo. A estos últimos se les denomina como las clases generadoras del modelo (Darroch et al., 1980). La razón para incluir interacciones de orden menor es que, de lo contrario, las interpretaciones para términos de orden mayor se modifican de acuerdo a la codificación de las variables. En el caso de los modelos jerárquicos, se obtienen las mismas interpretaciones sin importar la codificación de las variables (Agresti, 2003, p. 316). Además, bajo

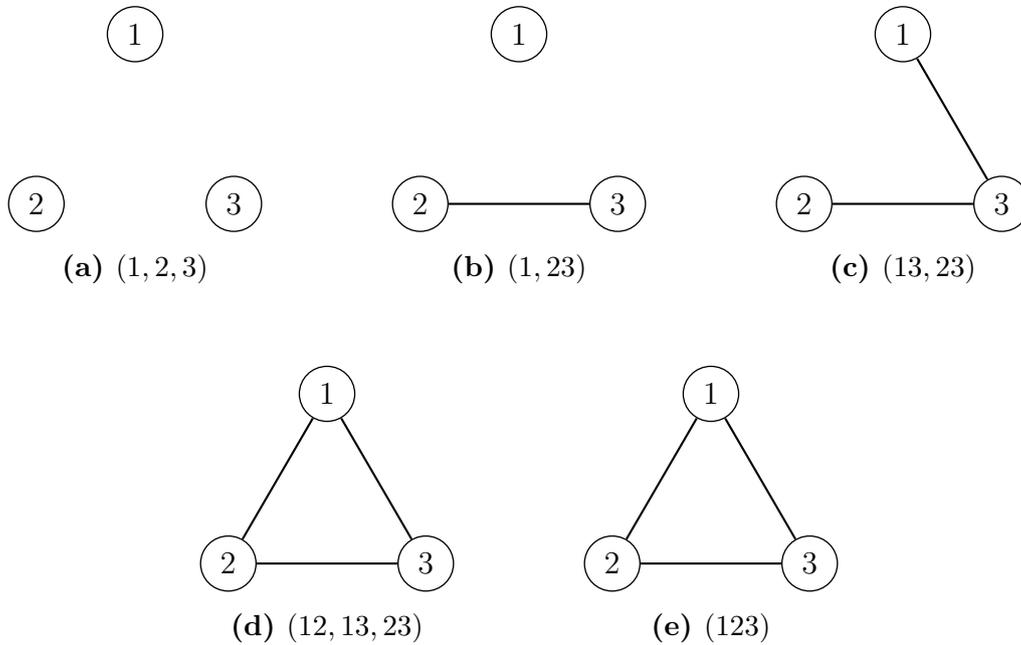
Para los modelos log-lineales definidos en la sección anterior, estos son modelos jerárquicos, donde el conjunto de clases generadoras para cada modelo están dados en la tabla siguiente. En la tercera columna, se muestra la notación empleada por Agresti (2003).

Modelo	Clases generadoras	Símbolo	Ecuación
Independencia	$\{\{1\}, \{2\}, \{3\}\}$	(1, 2, 3)	(2.5)
Independencia marginal	$\{\{1\}, \{2, 3\}\}$	(1, 23)	(2.6)
Independencia condicional	$\{\{1, 3\}, \{2, 3\}\}$	(13, 23)	(2.4)
Asociación homogénea	$\{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$	(12, 13, 23)	(2.10)
Saturado	$\{1, 2, 3\}$	(123)	(2.11)

**Tabla 2.2:** Clases generadoras para modelos log-lineales jerárquicos con 3 variables.

A partir de las interacciones de un modelo jerárquico, se define la *gráfica de interacciones*, ver e.g. Whitakker (1990); Lauritzen (1996). También conocida como *gráfica dependiente* (Edwards et al., 2012). En esta gráfica, cada vértice  $\delta \in V$  corresponde a la variable  $I_\delta$ , y para cualquier par de vértices  $\alpha, \beta$ , existe una arista entre ellos,  $(\alpha, \beta)$ , si existe alguna interacción  $u_a(i_a)$  en el modelo log-lineal tal que el conjunto  $a$  contiene a las variables  $I_\alpha$  e  $I_\beta$  (Whitakker, 1990, p. 209). Para esta gráfica es fácil verificar que para toda arista  $(\alpha, \beta)$ , satisface que  $\alpha \not\perp\!\!\!\perp \beta | V \setminus \{\alpha, \beta\}$ , es decir, la gráfica de interacciones es una gráfica condicional independiente. De esta manera, la gráfica dependiente asociada a un modelo jerárquico se puede definir simplemente con las interacciones de primer orden.

En la figura 3.6 se muestran las gráficas asociadas a los modelos jerárquicos para 3 variables. A partir de ésta, se observa que los modelos jerárquicos (12, 13, 23) y (123) tienen la misma gráfica asociada, sin embargo, sólo una gráfica corresponde a un modelo jerárquico. Para ello, a continuación se define los modelos gráficos log-lineales, los cuales son un subconjunto de modelos jerárquicos.



**Figura 2.1:** Gráficas asociadas a diferentes conjuntos de clases generadoras: **a)** (1, 2, 3),  $\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3)$ ; **b)** (1, 23),  $\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{23}(i_2, i_3)$ ; **c)** (13, 23),  $\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{13}(i_1, i_3) + u_{23}(i_2, i_3)$ ; **d)** (12, 13, 23),  $\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{12}(i_1, i_2) + u_{23}(i_2, i_3) + u_{13}(i_1, i_3)$ ; **e)** (123),  $\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{12}(i_1, i_2) + u_{13}(i_1, i_3) + u_{23}(i_2, i_3) + u_{123}(i_1, i_2, i_3)$ .

## 2.4. Modelos gráficos loglineales

Un modelo log-lineal jerárquico es *gráfico*, si el conjunto de clases generadoras del modelo,  $\mathcal{A}$ , corresponden al conjunto de cliques maximales de la gráfica de interacciones,  $\mathcal{G} = (V, E)$ , esto es,  $\mathcal{A} = \mathcal{C}$  (Lauritzen, 1996, p. 89).

De acuerdo con lo anterior, el modelo de asociación homogénea (2.10) no es gráfico, ya que el clique maximal de su gráfica de interacciones es el conjunto  $\{1, 2, 3\}$ , el cual difiere de las clases generadoras  $\mathcal{C} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$  de su gráfica de interacciones asociada. En cambio, el modelo saturado (2.11) sí es un modelo gráfico log-lineal.

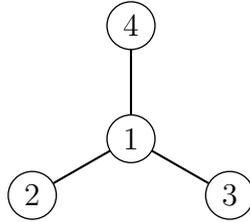
La interpretación en un modelo log-lineal se basa en las propiedades de Markov. Por ejemplo, en la gráfica asociada al modelo gráfico log-lineal (13, 23), se puede identificar la relación  $I_1 \perp\!\!\!\perp I_2 | I_3$ . El modelo gráfico log-lineal más sencillo de ajustar es el de independencia, en este caso, la gráfica asociada tiene conjunto de aristas vacío.

Dentro de los modelos gráficos, existe un subconjunto de éstos que tiene la propiedad de que su gráfica asociada es descomponible. En este caso, usando el teorema de Hammersley-Clifford, la factorización para  $p(i)$  está dada por la descomposición de  $\mathcal{G}$ .

Un ejemplo de un modelo gráfico descomponible es el siguiente:

$$\log m(i) = u + u_1(i_1) + u_2(i_1) + u_3(i_3) + u_4(i_4) + u_{12}(i_1, i_2) + u_{13}(i_1, i_3) + u_{14}(i_1, i_4).$$

En este modelo, las clases generadoras son iguales a los subconjuntos  $\{\{1, 2\}, \{1, 3\}, \{1, 4\}\}$ , donde la gráfica de interacciones se muestra en la figura 2.2.



**Figura 2.2:** Diagrama de la gráfica  $\mathcal{G}$ , asociada al modelo jerárquico descomponible con clases generadoras  $\{\{1, 2\}, \{1, 3\}, \{1, 4\}\}$ .

Como se vio en el primer capítulo, la factorización asociada a esta gráfica está dada por:

$$p(i) = \frac{p(i_1, i_2)p(i_1, i_3)p(i_1, i_4)}{p(i_1)^2}.$$

En este ejemplo, se observa que el problema de estimación se reduce en densidades marginales a lo más de dimensión 2. Dada la variable  $I_1$ , las variables  $I_2$ ,  $I_3$  y  $I_4$  son condicionalmente independientes.

En resumen, la relación de contención entre los conjuntos de distintos tipos de modelos (M) es:

$$M. \text{ Descomponibles} \subset M. \text{ gráficos} \subset M. \text{ jerárquicos} \subset M. \text{ log-lineales}$$

En la tabla 2.3 se muestran algunos modelos loglineales jerárquicos para los datos de la tabla 2.1. Alguno de éstos son jerárquicos pero no gráficos, y otros son gráficos pero no descomponibles. En la sección siguiente, se presenta la estimación por máxima-verosimilitud de modelos jerárquicos, para posteriormente obtener la estimación para modelos gráficos log-lineales.

Cabe mencionar que los modelos gráficos log-lineales se pueden restringir en modelos más sofisticados, por ejemplo, para datos pareados (*matched pairs*), ver *e.g.* Agresti (2003, pp. 423-430). En este caso, la tabla de contingencia se conforma por la clasificación cruzada de dos muestras diferentes para una misma variable de clasificación. El estudio de estos modelos desde una perspectiva de modelos gráficos se puede encontrar en Ramírez-Aldana (2010).

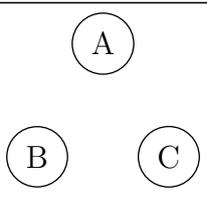
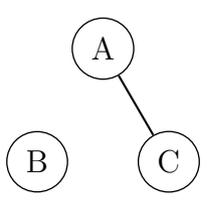
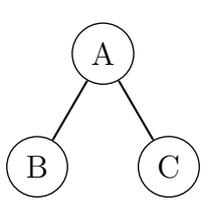
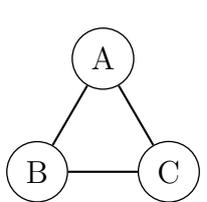
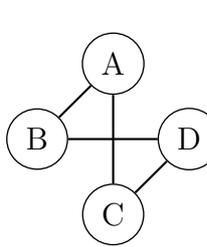
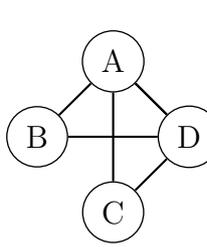
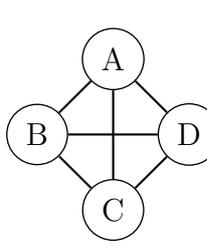
Modelo	Gráfico	Descomponible	Gráfica
$(A, B, C)$	Sí	No	
$(B, AC)$	Sí	Sí	
$(AC, AB)$	Sí	Sí	
$(AB, BC, AC)$	No	No	
$(AB, AC, CD, BD)$	Sí	No	
$(ABD, ACD)$	Sí	Sí	
$(ABCD)$	Sí	Sí	

Tabla 2.3: Modelos log-lineales jerárquicos para los datos de la tabla 2.1.

### 2.4.1. Estimación

Para el ajuste de un modelo gráfico log-lineal, primero obtendremos sus estimadores máximo-verosímil usando el hecho de que es un modelo jerárquico. Si las observaciones en la tabla de contingencia tienen una distribución Poisson (2.1), su función de verosimilitud está dada por:

$$L(m) = \prod_{i \in I} \frac{m(i)^{n(i)}}{n(i)!} e^{-m(i)}. \quad (2.13)$$

Por otro lado, si las observaciones tienen una distribución multinomial (2.2), donde  $p(i) = m(i)/|n|$ . Su función de verosimilitud es igual a:

$$L(m) = \frac{|n|}{\prod_{i \in I} n(i)!} \prod_{i \in I} \binom{m(i)}{|n|}^{n(i)}. \quad (2.14)$$

Observemos que bajo esta distribución

$$\prod_{i \in I} e^{-m(i)} = e^{-|n|},$$

por lo tanto, su función de verosimilitud puede expresarse de la siguiente manera.

$$\begin{aligned} L(m) &= \frac{\prod_{i \in I} e^{-m(i)}}{e^{-|n|}} \frac{|n|}{\prod_{i \in I} n(i)!} \prod_{i \in I} \binom{m(i)}{|n|}^{n(i)} \\ &= \frac{|n|}{e^{-|n|} \prod_{i \in I} |n|^{n(i)}} \prod_{i \in I} \frac{m(i)^{n(i)} e^{-m(i)}}{n(i)!} \\ &\propto \prod_{i \in I} \frac{m(i)^{n(i)} e^{-m(i)}}{n(i)!}. \end{aligned}$$

Esto indica que la función de verosimilitud para la distribución multinomial es proporcional a la función de verosimilitud de la distribución Poisson. De manera análoga, para la distribución multinomial restringida (2.3), si usamos el hecho de que  $p(i_a|i_b) = m(i)/n(i_b)$ , entonces el kernel de la función de verosimilitud está dada por:

$$\begin{aligned} L(m) &= \prod_{i_b \in I_b} \left\{ \frac{n(i_b)!}{\prod_{i_a \in I_a} n(i)!} \prod_{i_a \in I_a} \binom{m(i)}{n(i_b)}^{n(i)} \right\} \\ &= \frac{\prod_{i_b \in I_b} n(i_b)!}{e^{-|n|} \prod_{i \in I} n(i_b)^{n(i)}} \prod_{i \in I} \frac{m(i)^{n(i)} e^{-m(i)}}{n(i)!} \\ &\propto \prod_{i \in I} \frac{m(i)^{n(i)} e^{-m(i)}}{n(i)!}. \end{aligned}$$

En este caso, para los tres esquemas de muestreo, al maximizar sus funciones de log-verosimilitud se obtendrán los mismos estimadores máximo-verosímiles.

Para obtener el estimador máximo verosímil  $\hat{m}(i)$ , es más fácil si maximizamos con respecto al logaritmo de la verosimilitud; para los tres esquemas de recolección, sus funciones de log-verosimilitud son proporcionales

$$l(m) = \sum_{i \in I} n(i) \log m(i) - \sum_{i \in I} m(i), \quad (2.15)$$

en donde  $m$  es el vector formado por las frecuencias esperadas para las celdas. En particular, si la frecuencia del logaritmo para la celda  $i$  está dada por un modelo jerárquico (2.12), se tiene que

$$l(m) = \sum_{a \in \mathcal{A}} \sum_{i_a \in I_a} n(i_a) u_a(i_a) - \sum_{i \in I} \exp \left\{ \sum_{a \in \mathcal{A}} u_a(i_a) \right\}. \quad (2.16)$$

Derivando en (2.16) con respecto a los parámetros en el modelo,  $u_a(i_a)$ , e igualando a cero obtenemos las ecuaciones de verosimilitud:

$$\hat{m}(i_a) = n(i_a), \quad \forall i_a \in I_a, \quad a \in \mathcal{A}. \quad (2.17)$$

Bajo un esquema multinomial, esto implica que el estimador para  $p(i_a)$  está dado por:

$$\hat{p}(i_a) = \frac{n(i_a)}{|n|}. \quad (2.18)$$

Lauritzen (1996, p. 82) prueba que, bajo los tres esquemas de muestreo, el estimador máximo verosímil  $\hat{m}(i)$  es único si satisface las ecuaciones de verosimilitud (2.17). Este resultado nos permite identificar las ecuaciones que deben ser resueltas para obtener  $\hat{m}(i)$ , sin embargo, no sugiere cómo resolverlas. Usualmente, se utilizan métodos iterativos para resolver las ecuaciones de verosimilitud, tal como el algoritmo *IPS* (por sus siglas en inglés Iterative Proportional Scaling), también conocido como algoritmo *IPF* (Iterative Proportional Fitting), el cual fue propuesto por Deming y Stephan (1940).

---

### Algoritmo IPS

---

1. Sea  $\hat{m}^{(0)}(i)$  un valor inicial, generalmente igual a 1.
2. La función de iteración para  $\hat{m}^{(t)}(i)$  está dada por

$$\hat{m}^{(t)}(i) = \hat{m}^{(t-1)}(i) \frac{n(i_{a_j})}{\hat{m}^{(t-1)}(i_{a_j})}, \quad j = 1, 2, \dots, k.$$

Donde  $i_{a_j}$  es el vector marginal definido por el elemento  $a_j$  dentro de la clase generadora del modelo jerárquico. Para cada iteración en el algoritmo, existen  $k$  pasos intermedios, correspondiente a cada elemento en la clase generadora.

3. El algoritmo se detiene cuando  $\hat{m}(i_{a_j}) = n(i_{a_j})$ .
- 

Edwards et al. (2012, p. 35) mencionan que si las clases generadoras provienen de algún ordenamiento perfecto (p. 16), el algoritmo converge en una iteración. De otro modo, el algoritmo converge en dos iteraciones para modelos gráficos. Otro criterio que se utiliza

para detener el algoritmo es fijar un error. Si el valor absoluto de la diferencia entre los valores estimados y observados es menor al error, entonces el algoritmo se detiene.

Si el modelo es gráfico, las estadísticas suficientes corresponderán al número de observaciones marginales determinadas por los cliques de su gráfica asociada.

Para ilustrar este algoritmo, consideremos un modelo jerárquico con 4 variables,  $\{A, B, C, D\}$  para los datos de la tabla 2.1. Usando la notación punto, el modelo jerárquico queda expresado como:

$$\begin{aligned} \log m(a, b, c, d) = & u + u_A(a) + u_B(b) + u_C(c) + u_D(d) + u_{AB}(ab) + u_{AC}(ac) \\ & + u_{BC}(bc) + u_{CD}(cd) + u_{AD}(ad) + u_{ABC}(abc) + u_{ACD}(acd). \end{aligned}$$

En este ejemplo existen los elementos de clase generadora son  $\{\{A, B, C\}, \{A, C, D\}\}$ . Las estadísticas suficientes del modelo están dadas por  $n(a, b, c, \cdot)$  y  $n(a, \cdot, c, d)$ . De manera que la primera iteración del algoritmo IPF tiene dos pasos:

$$\begin{aligned} m_1^{(1)}(a, b, c, d) &= \widehat{m}^{(0)}(a, b, c, d) \frac{n(a, b, c, \cdot)}{\widehat{m}^{(0)}(a, b, c, \cdot)}, \\ m_2^{(1)}(a, b, c, d) &= \widehat{m}_1^{(1)}(a, b, c, d) \frac{n(a, \cdot, c, d)}{\widehat{m}_1^{(1)}(a, \cdot, c, d)}. \end{aligned}$$

En la tabla 2.4 se ilustra la primera iteración del algoritmo IPS para el modelo jerárquico con clases generadoras  $\{\{A, B, C\}, \{A, C, D\}\}$ .

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	$n(i)$	$m^{(0)}(i)$	$m_1^{(1)}(i)$	$m_2^{(1)}(i)$
sí	sí	sí	sí	79	1	73	80.02
			no	67	1	73	65.97
		no	sí	197	1	188	195.59
			no	179	1	188	180.40
	no	sí	sí	217	1	197	215.97
			no	177	1	197	178.02
		no	sí	22	1	22.5	23.41
			no	23	1	22.5	21.59
no	sí	sí	sí	82	1	61	75.02
			no	40	1	61	46.97
		no	sí	258	1	209.5	255.82
			no	161	1	209.5	163.18
	no	sí	sí	156	1	76	162.97
			no	109	1	76	102.03
		no	sí	43	1	36	45.18
			no	31	1	36	28.81

**Tabla 2.4:** Primera iteración del algoritmo IPF para el modelo jerárquico log-lineal  $(ABC, ACD)$ .

Para modelos gráficos descomponibles, el teorema de Hammersley-Clifford nos asegura que la distribución de  $p(i)$  puede factorizarse como:

$$p(i) = \frac{\prod_{C \in \mathcal{C}} p(i_C)}{\prod_{S \in \mathcal{S}} p(i_S)^{\nu(S)},}$$

donde  $\mathcal{C}$  es el conjunto de cliques de la gráfica asociada  $\mathcal{G}$ ,  $\mathcal{S}$  el conjunto de separadores y  $\nu(S)$  la multiplicidad del conjunto separador  $S$ . En este caso, la frecuencia esperada en la celda  $i$  está dada por la siguiente expresión

$$m(i) = \frac{\prod_{C \in \mathcal{C}} m(i_C)}{\prod_{S \in \mathcal{S}} m(i_S)^{\nu(S)}}.$$

Por lo tanto, un estimador para la frecuencia esperada está dada por

$$\hat{m}(i) = \frac{\prod_{C \in \mathcal{C}} n(i_C)}{\prod_{S \in \mathcal{S}} n(i_S)^{\nu(S)}}.$$

En donde los estimadores obtenidos por máxima-verosimilitud para  $m(i_C)$  y  $m(i_S)$  son  $n(i_C)$  y  $n(i_S)$ .

### 2.4.2. Bondad de Ajuste

Una vez ajustado un modelo log-lineal  $M$ , es importante evaluar si dicho modelo es consistente con los datos. Para esto, se define la devianza del modelo como dos veces el logaritmo del cociente de verosimilitudes:

$$d = -2 \log \frac{L(\hat{m})}{L(\hat{m}_s)} = 2 [\ell(\hat{m}_s) - \ell(\hat{m})],$$

donde  $\hat{m}$  maximiza la log-verosimilitud de los datos bajo el modelo  $M$ , y  $\hat{m}_s$  maximiza la log-verosimilitud bajo el modelo *saturado*. Este último, tiene el mismo número de parámetros que observaciones, por lo tanto, el ajuste a los datos es perfecto,  $\hat{m}_s(i) = n(i)$ . Usualmente, a esta estadística se le conoce como la *devianza residual* del modelo.

Como se vio en la sección 2.4.1, bajo los tres esquemas de muestreo, el kernel de la función de log-verosimilitud es igual a:

$$\sum_{i \in I} n(i) \log m(i) - \sum_{i \in I} m(i).$$

En este caso, la devianza para el modelo  $M$  está dada por:

$$\begin{aligned} d &= 2 [\ell(\hat{m}_s) - \ell(\hat{m})] \\ &= 2 \left[ \sum_{i \in I^+} n(i) \log \frac{n(i)}{\hat{m}(i)} - \sum_{i \in I} (n(i) - \hat{m}(i)) \right], \end{aligned}$$

donde

$$I^+ = \{i \in I \mid n(i) > 0\}.$$

Bajo el supuesto de la distribución multinomial se tiene que  $\sum_{i \in I} \hat{m}(i) = |n|$ . Por lo tanto, la expresión para la devianza se reduce a

$$d = 2 \sum_{i \in I^+} n(i) \log \frac{n(i)}{\hat{m}(i)}.$$

Dado el modelo  $M$ , la estadística  $d$  tiene asintóticamente una distribución  $\chi^2_{(p-q)}$ , donde los grados de libertad  $p - q$  es igual a la diferencia del número de parámetros libres (linealmente independientes) del modelo saturado y el modelo  $M$  (Edwards et al., 2012, p. 36). De esta manera, usando la estadística de devianza se puede probar la hipótesis nula de que los parámetros que están en el modelo saturado y que no están contenidos en  $M$  son iguales a cero. Para un nivel de significancia  $\alpha$  se rechaza la hipótesis nula si

$$d \geq \chi^2_{(p-q, 1-\alpha)},$$

donde  $\chi^2_{(p-q, 1-\alpha)}$  es igual al cuantil  $1 - \alpha$  de una  $\chi^2$  con grados de libertad  $p - q$ . Si no se rechaza la hipótesis nula, entonces el modelo  $M$  es consistente con los datos.

Asimismo, la devianza nos permite comparar la bondad de ajuste para dos modelos anidados  $M_1 \subset M_0$ , es decir,  $M_1$  es un caso particular de  $M_0$ . La devianza entre estos dos modelos es igual a la diferencia de sus devianzas residuales:

$$d_{01} = d_0 - d_1 = -2 \log \frac{L(\hat{m}_0)}{L(\hat{m}_1)} = 2 \sum_{i \in I^+} n(i) \log \frac{\hat{m}_0(i)}{\hat{m}_1(i)}.$$

Asintóticamente, la estadística  $d_{01}$  tiene una distribución  $\chi^2_{(p_0-p_1)}$ , donde  $p_0$  y  $p_1$  son iguales al número de parámetros libres de los modelos  $M_0$  y  $M_1$ , respectivamente, con  $p_1 < p_0$ . En este caso, la devianza se utiliza para probar si los términos en el modelo  $M_0$  que no están contenidos en  $M_1$  son iguales a cero. Si no se rechaza la hipótesis nula, el modelo  $M_1$  ajusta a los datos tanto como el modelo  $M_0$ , y por lo tanto, resulta un modelo con más parsimonia que el modelo  $M_0$ .

De manera alternativa, la estadística de Pearson  $X^2$  se utiliza para evaluar el ajuste del modelo, la cual corresponde a una aproximación de la devianza, cuya expresión está dada por

$$X^2 = \sum_{i \in I^+} \frac{[n(i) - \hat{m}(i)]^2}{\hat{m}(i)}.$$

Bajo el modelo  $M$ , la estadística de Pearson tiene la misma distribución asintótica que la devianza. Para valores cercanos de  $m(i)$  a  $n(i)$ , las estadísticas de devianza y de Pearson son equivalentes.

Otro de los elementos a considerar en bondad de ajuste, es determinar los grados de libertad de la distribución. Para modelos jerárquicos, la siguiente regla nos permite calcular de manera recursiva la dimensión del modelo (Lauritzen, 1996, p. 86).

1. Los subespacios formados por los factores del modelo tienen dimensión igual a

$$\dim M_F = |I_a| = \prod_{\delta \in a} |I_\delta|.$$

Si  $a = \{\emptyset\}$ , entonces  $\dim M_F = 1$ .

2. Si el modelo  $M$  tiene clases generadoras  $\mathcal{A} = A_1 \cup A_2$ , entonces la dimensión del modelo es igual a:

$$\dim M_{\mathcal{A}} = \dim M_{A_1} + \dim M_{A_2} - \dim M_{A_1 \cap A_2}.$$

Por ejemplo, para el modelo de asociación homogénea (2.10), su conjunto de clases generadoras está dado por  $\mathcal{A} = \{A_1, A_2, A_3\}$ , con  $A_1 = \{1, 2\}$ ,  $A_2 = \{1, 3\}$  y  $A_3 = \{2, 3\}$ . En este caso, la dimensión del modelo  $M_{\mathcal{A}}$  está dada de la siguiente manera:

$$\begin{aligned} \dim M_{\mathcal{A}} &= \dim M_{A_1} + \dim M_{A_2 \cup A_3} - \dim M_{A_1 \cap (A_2 \cup A_3)} \\ &= \dim M_{A_1} + (\dim M_{A_2} + \dim M_{A_3} - \dim M_{A_2 \cap A_3}) \\ &\quad - (\dim M_{A_1 \cap A_2} + \dim M_{A_1 \cap A_3} - \dim M_{A_1 \cap A_2 \cap A_3}) \\ &= |I_1||I_2| + |I_1||I_3| + |I_2||I_3| - |I_3| - |I_1| - |I_2| + 1. \end{aligned}$$

En consecuencia, la devianza del modelo se distribuye  $\chi^2$  con grados de libertad igual a

$$\begin{aligned} p - q &= \dim M_s - \dim M_{\mathcal{A}} \\ &= |I_1||I_2||I_3| - (|I_1||I_2| + |I_1||I_3| + |I_2||I_3| - |I_3| - |I_1| - |I_2| + 1) \\ &= (|I_1| - 1)(|I_2| - 1)(|I_3| - 1), \end{aligned}$$

donde  $\dim_{M_s}$  es la dimensión del modelo saturado e igual al número de celdas en la tabla de contingencia. Cabe mencionar que los grados de libertad de la distribución  $\chi^2$  están restringidos a que  $\hat{m}(i) > 0$ , de lo contrario, se debe corregir la dimensión del modelo. Para esto, primero debemos analizar qué mecanismo está generando que  $\hat{m}(i) = 0$ .

Para una tabla de contingencia con celdas vacías  $n(i) = 0$ , el estimador de la frecuencia esperada en la celda  $i$  es igual  $\hat{m}(i) = 0$ . En este caso, estas celdas vacías pueden ser de dos tipos (Agresti, 2003, pp. 391-392):

- i) *Ceros muestrales*. En este caso, aunque no se haya observado individuos en la celda  $i$ , la frecuencia esperada en la celda es positiva,  $m(i) > 0$ . De manera que para tamaños de muestra grande, es posible observar  $n(i) > 0$ .
- ii) *Ceros estructurales*. En este caso, por definición  $m(i) = 0$ , es decir, no está permitido que haya observaciones en la celda  $i$ , por ejemplo, para un estudio de cáncer, si clasificamos de acuerdo al género y tipo de cáncer (hígado, próstata, ovario, etc), es imposible observar pacientes hombres con cáncer de ovario.

Para corregir los grados de libertad de la estadística de prueba, Bishop et al. (2007, p. 115) proponen contar el número de celdas con ceros estimados,  $z_e$ , y el número de parámetros que no pueden ser estimados,  $z_p$ . De tal manera que la corrección para los grados de libertad está dada por:

$$g.l.' = g.l. - z_e + z_p.$$

No obstante, si la tabla es poco densa, es decir, las frecuencias estimadas  $\hat{m}(i)$  son pequeñas, la aproximación a la distribución  $\chi^2$  es pobre (Edwards et al., 2012, p. 37).

Una alternativa para evaluar el ajuste del modelo, es probar independencias condicionales en el modelo. De esta manera, si en la tabla de contingencia hay celdas vacías, la estadística de devianza para independencia condicional estará calculada con base en las tablas marginales, las cuales son más densas con respecto a la tabla completa (Edwards et al., 2012, p. 37). En este caso, el análisis se enfoca a modelos gráficos descomponibles.

Sea  $M_1 \subset M_0$ , dos modelos gráficos descomponibles tales que sólo difieren por una arista  $(\alpha, \beta)$ . En este caso, existe un sólo clique de  $M_0$ , denotado por  $C^*$  tal que contiene a la arista  $(\alpha, \beta)$ . De manera que la hipótesis nula de independencia condicional  $I_\alpha \perp\!\!\!\perp I_\beta | I_{V \setminus \{\alpha, \beta\}}$  está calculada con base en la tabla marginal generada por el clique  $C^*$ .

Por ejemplo, para los datos de la tabla 2.1 consideremos los modelos  $M_0$  y  $M_1$  con conjunto de cliques  $\mathcal{C}_0 = \{\{A, B, C\}, \{A, C, D\}\}$  y  $\mathcal{C}_1 = \{\{B, C\}, \{A, C, D\}\}$ . Sus gráficas asociadas están dadas en la figura 2.3.



**Figura 2.3:** Modelos gráficos descomponibles. (a) Gráfica asociada al modelo jerárquico  $M_0$ , con clases generadoras  $\mathcal{C}_0 = \{\{A, B, C\}, \{A, C, D\}\}$ ; (b) Gráfica asociada al modelo jerárquico anidado  $M_1$ , con clases generadoras  $\mathcal{C}_1 = \{\{B, C\}, \{A, C, D\}\}$ .

A partir de esta última figura, se observa que los modelos  $M_0$  y  $M_1$  difieren por la arista  $(A, B)$ , la cual está contenida en el clique  $C^* = \{A, B, C\}$  de  $M_0$ . En ambos modelos su conjunto separador es igual  $\{A, C\}$ , considerando que ambos son modelos descomponibles, las frecuencias estimadas para la celda  $i$  bajo los modelos  $M_0$  y  $M_1$  son iguales a

$$\hat{m}_0(a, b, c, d) = \frac{n(a, b, c, \cdot)n(a, \cdot, c, d)}{n(a, \cdot, c, \cdot)}, \quad \hat{m}_1(a, b, c, d) = \frac{n(\cdot, b, c, \cdot)n(a, \cdot, c, d)}{n(a, \cdot, c, \cdot)}.$$

En este caso, la devianza entre estos dos modelos está dada por:

$$d_{01} = 2 \sum_{i \in C^*} n(i) \log \frac{\hat{m}_0(i)}{\hat{m}_1(i)} = 2 \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} n(a, b, c) \log \frac{n(a, b, c)}{n(b, c)}.$$

Para valores grandes de  $d_{01}$  se rechaza la hipótesis nula, es decir,  $I_\alpha \not\perp\!\!\!\perp I_\beta | I_{V \setminus \{\alpha, \beta\}}$ . Por lo tanto, se concluye que la arista  $(A, B)$  es significativa en el modelo  $M_0$ .

En particular, si  $M_0$  es el modelo saturado, entonces hay necesariamente un único clique correspondiente al conjunto de vértices y correspondiente también al conjunto que contiene a la arista  $(\alpha, \beta)$  (Lauritzen, 1996, p. 101). En este caso, la estadística de devianza para probar independencia condicional queda expresada como:

$$d = 2 \sum_{i \in I} n(i) \log \frac{n(i_{V \setminus \{\alpha, \beta\}})n(i)}{n(i_{V \setminus \{\alpha\}})n(i_{V \setminus \{\beta\}})}. \quad (2.19)$$

Lo anterior se deriva del hecho de que el modelo es descomponible y al quitar la arista  $(\alpha, \beta)$ , se obtiene una subgráfica descomponible con dos cliques,  $V \setminus \{\alpha\}$  y  $V \setminus \{\beta\}$ , donde el conjunto separador es  $(\alpha, \beta)$ .

En un caso más general, para probar la hipótesis nula de independencia condicional  $I_\alpha \perp\!\!\!\perp I_\beta | I_W$ , con  $W \subset V$ , es equivalente a probar la significancia de la arista  $(\alpha, \beta)$  en el modelo saturado definido por las variables  $\{I_\alpha, I_\beta, I_W\}$ , por ejemplo, si  $W$  es un conjunto de dos variables binarias,  $I_W$  corresponde a una variable aleatoria con 4 posibles valores. En este caso, si definimos a  $S$  como una nueva variable que toma valores en un conjunto igual a las combinaciones formadas por las variables en  $W$ , la estadística de prueba puede reexpresarse como  $I_\alpha \perp\!\!\!\perp I_\beta | I_S$ , la cual está definida en una tabla de dimensión tres (Edwards et al., 2012, p. 39). La devianza queda expresada como:

$$\begin{aligned} d &= 2 \sum_{i \in I} n(i_\alpha, i_\beta, i_S) \log \frac{n(i_\alpha, i_\beta, i_S) n(i_S)}{n(i_\alpha, i_S) n(i_\beta, i_S)} \\ &= \sum_{i_S \in I_S} 2 \sum_{i_\alpha \in I_\alpha} \sum_{i_\beta \in I_\beta} n(i_\alpha, i_\beta, i_S) \log \frac{n(i_\alpha, i_\beta, i_S) n(i_S)}{n(i_\alpha, i_S) n(i_\beta, i_S)} \\ &= \sum_{i_S \in I_S} d_{i_S} \end{aligned}$$

donde  $d_{i_S}$  es la devianza del modelo de independencia condicional de  $I_\alpha$  y  $I_\beta$  para un nivel de  $I_S$ .

Si ahora consideramos una sucesión de modelos gráficos descomponibles  $\mathcal{G}_0 \subset \dots \subset \mathcal{G}_k = \mathcal{G}$  tales que sólo difieren por una arista, la devianza del modelo  $\mathcal{G}$  puede particionarse como (Lauritzen, 1996, p. 103):

$$d = -2 \log \frac{L(\mathcal{G}_0)}{L(\mathcal{G})} = \sum_{j=1}^k d_j. \quad (2.20)$$

Finalmente, para determinar la dimensión para un modelo gráfico descomponible, el resultado 4.35 mostrado en Lauritzen (1996, p. 105), afirma que

$$\dim M_{\mathcal{C}} = \sum_{\mathcal{C} \in \mathcal{C}} \prod_{\delta \in \mathcal{C}} |I_\delta| - \sum_{S \in \mathcal{S}} \nu(S) \prod_{\delta \in S} |I_\delta|,$$

donde  $\mathcal{C}$  es el conjunto de cliques del modelo,  $\mathcal{S}$  el conjunto de separadores y  $\nu(S)$  la multiplicidad del conjunto  $S$ .

### 2.4.3. Selección del modelo

En general, los métodos de selección para modelos gráficos consisten en agregar o eliminar aristas de una gráfica asociada  $\mathcal{G}$ . Edwards (2000, sec. 6) clasifica estos métodos en tres categorías.

1. Selección *stepwise*. A partir de un modelo inicial, las aristas son agregadas o eliminadas sucesivamente hasta satisfacer algún criterio.

2. Métodos de búsqueda global, los cuales consisten de algoritmos para encontrar los modelos más simples consistentes con los datos.
3. Métodos de selección basados en el criterio de verosimilitud penalizada. Dado un conjunto de modelos elegibles, se selecciona aquel que minimice la función de verosimilitud penalizada.

Usualmente, para la selección de modelo se utiliza alguna combinación de los métodos anteriores. Para el primer tipo, en cada paso, la inclusión o exclusión de las aristas elegibles está basada en pruebas de significancia. La selección *stepwise* puede realizarse en alguna de las dos direcciones:

- i) *Backward*. En este caso, las aristas son eliminadas sucesivamente del modelo inicial. En cada paso se obtienen las devianzas de exclusión para las aristas elegibles, es decir, la devianza entre modelos sucesivos que sólo difieren por la arista eliminada, lo cual es equivalente a probar la significancia de dicha arista. El siguiente paso es eliminar la arista con el mayor *p-value*, dado un nivel de significancia especificado. El procedimiento se detiene cuando todas las aristas son significativas.
- ii) *Forward*. En este caso, las aristas son agregadas sucesivamente, dado un modelo inicial. En cada paso, se obtienen las devianzas de inclusión, estos es, la devianza entre los modelos con la arista añadida y el modelo inicial, lo que es equivalente a probar la significancia de la arista agregada. Se selecciona el modelo con la arista más significativa, es decir, con el *p-value* más pequeño.

En ambos métodos es posible utilizar la estadística de Pearson para evaluar la significancia de las aristas en cada paso. Para la selección *backward*, ésta satisface el *principio de coherencia*, es decir, para dos modelos anidados,  $M_1 \subset M_0$ , si  $M_0$  es rechazado entonces  $M_1$  también debe ser rechazado (Edwards, 2000, p. 165).

Una variante de estos métodos es *headlong search* (Edwards et al., 2012, p. 44). En este método, para cada paso, el orden en que son probadas las aristas es aleatoria, y se selecciona la primera arista en satisfacer el criterio de inclusión o exclusión. Este tipo de selección es conveniente cuando el número de parámetros es grande.

Adicionalmente, Whitakker (1990, p.252) propone una combinación de los métodos de selección *stepwise*, el cual consiste en dos pasos.

1. Obtener la gráfica independiente. Para esto, se calculan las devianzas de exclusión para cada arista del modelo saturado y se eliminan aquellas con *p-values* mayores a un nivel de significancia especificado. A la gráfica obtenida la denotamos por  $\mathcal{G}^1$ .
2. Con la gráfica  $\mathcal{G}^1$ , obtenemos las devianzas de inclusión de las aristas eliminadas en el paso anterior. Todas las aristas significativas se agregan a la gráfica  $\mathcal{G}^1$  y obtenemos la gráfica  $\mathcal{G}^2$ .

Entre los métodos de búsqueda global, se puede mencionar el algoritmo EH, propuesto por Edwards y Havranek (1985), y descrito en Edwards (2000, p. 167). Para una sucesión

de modelos ajustados, éstos se clasifican, con base en la devianza del modelo, en aceptados y rechazados. Usando el principio de coherencia, para un modelo rechazado, todos sus submodelos también son rechazados. En este caso, para cada etapa de la búsqueda se hace la siguiente partición: i) débilmente aceptados, si el modelo contiene un submodelo aceptado, ii) débilmente rechazado, si el modelo contiene uno o más submodelos rechazados y iii) indeterminado, si el modelo tiene submodelos por clasificar. El siguiente paso del algoritmo es clasificar todos los submodelos minimales o máximos en el conjunto indeterminado. El proceso se detiene hasta que todos los submodelos son clasificados en débilmente aceptados y rechazados. En este caso, el modelo seleccionado es el submodelo minimal aceptado.

Una variante de este método, es el propuesto por [Whitakker \(1990, p. 253\)](#). Un modelo  $M$  es *adecuado* si la devianza del modelo  $M$  es pequeña. En este caso, la selección del modelo consiste en obtener el modelo adecuado minimal, es decir, si  $M$  es adecuado, entonces no contiene ningún modelo anidado  $N$  que también sea adecuado.

Finalmente, los métodos basados en el criterio de verosimilitud son procedimientos que se emplean para seleccionar modelos que no son anidados. Supongamos un conjunto de modelos  $M_j$ ,  $j = 1, \dots, R$ . El criterio de verosimilitud penalizada consiste en seleccionar el modelo  $M_j$  que minimice la expresión

$$-2 \log L(\hat{m}_j(i)) + kp_j, \quad j = 1, \dots, R,$$

donde  $p_j$  es el número de parámetros libres del modelo  $M_j$  y  $k$  una constante positiva de penalización.

El criterio de información de Akaike (AIC) es un caso particular, donde la constante de penalización es  $k = 2$ . Otra alternativa es utilizar el criterio de información Bayesiana (BIC), donde la constante de penalización es igual al logaritmo del tamaño de muestra,  $k = \log |n|$ . En ambos casos, la constante  $k$  penaliza aquellos modelos con mayor número de parámetros, siendo el BIC más severo, por lo que tenderá a seleccionar modelos más simples. La elección de la constante de penalización dependerá de la aproximación asintótica que se desee utilizar. En el caso del AIC, éste se aproxima asintóticamente a la divergencia de Kullback-Leibler entre el modelo verdadero y estimado. Para el BIC, corresponde a seleccionar, asintóticamente, el modelo con mayor probabilidad posterior ([Edwards et al., 2012, p. 43](#)).

Asimismo, el criterio de verosimilitud penalizada se puede utilizar para evaluar la significancia de cada arista en la selección *stepwise*. Para la dirección *backward*, en cada paso, la arista que se elimina es la que tenga la diferencia más pequeña de AIC ó BIC entre modelos sucesivos. De manera análoga, con dirección *forward*, la arista agregada en cada paso será aquella con la mayor diferencia de AIC ó BIC.

Cabe mencionar que para obtener modelos gráficos descomponibles, en cada etapa de selección se agrega o elimina la arista tal que la gráfica obtenida sea descomponible. De esta manera, este tipo de método es menos eficiente con respecto a seleccionar un modelo solamente gráfico; sin embargo, los modelos descomponibles tienen la ventaja de ser fácilmente interpretados.

En conclusión, los métodos de selección son útiles para explorar los datos cuando no se tiene ningún supuesto sobre la asociación entre las variables, además, permiten identificar si el modelo propuesto está sobre-ajustado, es decir, si hay variables que no son significativas para el ajuste. De manera recíproca, pueden identificar interacciones que no estaban consideradas en el modelo propuesto y que resulten en un mejor ajuste.

## 2.5. Modelo log-lineal expresado como un GLM

Otra manera para obtener los estimadores máximo-verosímiles  $\{\widehat{m}(i)\}_{i \in I}$  es expresar el modelo log-lineal como un modelo lineal generalizado (GLM). Para un estudio más detallado ver *e.g.* Agresti (2003). En un GLM se tienen tres elementos:

1. *Componente aleatorio* que consiste de una variable respuesta  $y$  con observaciones independientes  $y = (y_1, \dots, y_n)$ , donde  $y$  es una variable aleatoria con una distribución de probabilidad perteneciente a la familia exponencial.
2. *Componente sistemático* que consiste de un vector  $\eta = (\eta_1, \dots, \eta_n)$  y un conjunto de variables explicativas  $(x_1, \dots, x_p)$  tales que existe una combinación lineal de éstas que es igual a  $\eta_i$ , esto es

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j,$$

donde  $\beta = (\beta_1, \dots, \beta_p)$  es un vector de parámetros y  $x_{ij}$  es el valor observado de la variable  $j$  para el individuo  $i$ . Al vector  $\eta$  se le conoce como el predictor lineal del modelo.

3. Una función *liga*  $g(\cdot)$  que relaciona la esperanza del componente aleatorio,  $\mu_i$ , con el sistemático.

$$g(\mu_i) = \eta_i = \sum_j x_{ij}\beta_j$$

Para el componente aleatorio, se dice que la variable aleatoria  $y$  pertenece a la familia exponencial si su función de densidad puede expresarse como:

$$f(y_i; \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right), \quad (2.21)$$

donde  $\theta_i$  se le conoce como el parámetro canónico y  $\phi$  es un parámetro de dispersión conocido. La función  $b(\theta)$  se le conoce como la función cumulante, donde  $b'(\theta) = E(Y_i)$  y  $\text{var}(Y_i) = a(\phi)b''(\theta)$ . Si la función liga  $g(\cdot)$  transforma la media en el parámetro canónico se le denomina *liga canónica*.

En un modelo log-lineal, si el número de observaciones en cada celda de la tabla de contingencia se supone tiene una distribución Poisson de media  $m(i)$ , entonces se puede especificar un GLM de la siguiente forma.

$$\log m(i) = \sum_{j=1}^p x_{ij}\beta_j, \quad (2.22)$$

donde el componente aleatorio está dado por la variable aleatoria Poisson  $N(i)$ , la cual tiene una distribución perteneciente a la familia exponencial

$$P(N(i) = n(i)) = \exp \{n(i) \log m(i) - m(i) - \log n(i)!\},$$

donde  $\theta = \log m(i)$ ,  $b(\theta) = \exp(\theta)$  y  $a(\phi) = 1$ . En un modelo log-lineal la función liga es canónica.

Para el componente sistemático, el vector de parámetros está determinado por el conjunto de interacciones del modelo y la matriz  $x_{ij}$  por la matriz diseño, definida como:

$$x_{ij} = \begin{cases} 1 & \text{si } \beta_j \text{ está en } m(i) \\ 0 & \text{en otro caso} \end{cases}$$

Por ejemplo, en la sección 2.3 se mostró que el modelo log-lineal de independencia para  $I_1$  y  $I_2$  se puede expresar de manera matricial como:

$$\begin{bmatrix} \log m(1,1) \\ \log m(1,0) \\ \log m(0,1) \\ \log m(0,0) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} [u \quad u_1(1) \quad u_1(0) \quad u_2(1) \quad u_2(0)]^T.$$

En este caso, el vector de parámetros es  $\beta = [u, u_1(1), u_1(0), u_2(1), u_2(0)]^T$ . Para obtener una solución, es necesario especificar alguna restricción sobre los parámetros, por ejemplo, usando variables dummy con  $u_1(1) = u_2(1) = 0$  valores de referencia, la matriz diseño es

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Por otra parte, si utilizamos la restricción  $\sum_{i_1} u_1(i_1) = \sum_{i_2} u_2(i_2) = 0$ , entonces la matriz diseño queda expresada como

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & -1 \end{pmatrix}$$

Al definir un modelo log-lineal como un GLM, el procedimiento para obtener  $\{\hat{m}(i)\}$  se hace por métodos iterativos como Newton-Rhapson o Fisher-Score, *e.g.* Agresti (2003, pp. 342-343). No obstante, estos métodos son imprecisos para dimensiones grandes de  $\beta$ .

Una de las ventajas de utilizar un GLM para el ajuste de un modelo log-lineal es que se puede cuantificar las interacciones y hacer un análisis de residuos.

Otra relación existente entre un modelo log-lineal y un GLM, es una regresión logística. En este modelo, la variable respuesta es binaria. Algunos autores como Agresti (2003, sec 8.5), Von-Eye y Mung (2013, sec. 15.2) han descrito esta relación para casos muy particulares. En modelos gráficos log-lineales, la relación con la regresión logística es fácil de identificar utilizando la gráfica asociada. En la sección siguiente, se describe el procedimiento para definir regresiones logísticas basadas en un modelo gráfico log-lineal.

## 2.6. Relación de un modelo gráfico log-lineal y el modelo de regresión logística

En modelos gráficos log-lineales el interés principal es describir la estructura de asociación entre las variables categóricas. No obstante, si alguna de estas variables categóricas es binaria, se puede calcular su momio condicional dado el resto de las variables. De esta manera, podría ser de interés plantear una regresión logística, donde la variable binaria sea la variable respuesta, y las variables restantes como explicativas. Por ejemplo, supongamos el siguiente modelo de independencia condicional:

$$\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_{12}(i_1, i_2) + u_{13}(i_1, i_3)$$

Para este modelo, supongamos que  $I_1$  es la variable respuesta. Con base en el modelo gráfico log-lineal, el momio condicional de  $I_1$  dado los niveles fijos de  $I_2$  e  $I_3$  puede expresarse como:

$$\begin{aligned} \log \frac{p(I_1 = 1|I_2 = i_2, I_3 = i_3)}{p(I_1 = 0|I_2 = i_2, I_3 = i_3)} &= \log \frac{m(1, i_2, i_3)}{m(0, i_2, i_3)} \\ &= \log m(1, i_2, i_3) - \log m(0, i_2, i_3) \\ &= [u + u_1(1) + u_2(i_2) + u_3(i_3) + u_{12}(i_1, i_2) + u_{13}(1, i_3)] \\ &\quad - [u + u_1(0) + u_2(i_2) + u_3(i_3) + u_{12}(0, i_2) + u_{13}(0, i_3)] \\ &= [u_1(1) - u_1(0)] + [u_{12}(1, i_2) - u_{12}(0, i_2)] + \\ &\quad [u_{13}(1, i_3) - u_{13}(0, i_3)]. \end{aligned}$$

En este caso, observamos que los términos dentro del primer paréntesis son constantes; en el segundo, sólo dependen del valor que toma la variable  $I_2$ , y en el tercer paréntesis, los términos sólo dependen de  $I_3$ . Por lo tanto, la regresión logística asociada al momio condicional de  $I_1$  tiene como variables explicativas a  $I_2$  y  $I_3$ :

$$\text{logit } p(I_1 = 1|I_2 = i_2, I_3 = i_3) = \log \frac{p(I_1 = 1|I_2 = i_2, I_3 = i_3)}{p(I_1 = 0|I_2 = i_2, I_3 = i_3)} = \beta_0 + \beta_2(i_2) + \beta_3(i_3).$$

En este caso, los parámetros en la regresión logística se interpretan de la siguiente manera. Supongamos que  $I_2$  e  $I_3$  son variables binarias que indican ausencia (= 0) o presencia (= 1) de alguna característica del individuo. Dado un valor fijo de  $I_3$ , el momio para un individuo con la característica  $I_2$  es

$$\frac{p(I_1 = 1|I_2 = 1, I_3 = i_3)}{p(I_1 = 0|I_2 = 1, I_3 = i_3)} = \exp [\beta_0 + \beta_2 + \beta_3(i_3)].$$

De manera análoga, el momio para un individuo que no presenta la característica  $I_2$  es

$$\frac{p(I_1 = 1|I_2 = 0, I_3 = i_3)}{p(I_1 = 0|I_2 = 0, I_3 = i_3)} = \exp [\beta_0 + \beta_3(i_3)].$$

De manera que la razón de momios queda expresada en función del coeficiente  $\beta_2$

$$\frac{p(I_1 = 1|I_2 = 1, I_3 = i_3)/p(I_1 = 0|I_2 = 1, I_3 = i_3)}{p(I_1 = 1|I_2 = 0, I_3 = i_3)/p(I_1 = 0|I_2 = 0, I_3 = i_3)} = \exp(\beta_2).$$

Para este ejemplo, la probabilidad de ocurrencia para la variable respuesta,  $I_1$ , dada las variables explicativas es igual a:

$$p(I_1 = 1|I_2 = i_2, I_3 = i_3) = \frac{\exp[\beta_0 + \beta_2(i_2) + \beta_3(i_3)]}{1 + \exp[\beta_0 + \beta_2(i_2) + \beta_3(i_3)]}. \quad (2.23)$$

Si consideramos el modelo gráfico, la probabilidad para el mismo evento está dada por:

$$\begin{aligned} p(I_1 = 1|I_2 = i_2, I_3 = i_3) &= \frac{p(I_1 = 1, I_2 = i_2, I_3 = i_3)}{p(I_2 = i_2, I_3 = i_3)} \\ &= \frac{m(1, i_2, i_3)}{\sum_{i_1} m(i_1, i_2, i_3)}. \end{aligned}$$

Al ser un modelo gráfico descomponible, la frecuencia en la celda  $i$ ,  $m(i)$ , puede expresarse en términos de la descomposición de la gráfica, esto es

$$m(i_1, i_2, i_3) = \frac{m(i_1, i_2, \cdot)m(i_1, \cdot, i_3)}{m(i_1, \cdot, \cdot)}.$$

Por lo tanto, la probabilidad estimada bajo el modelo gráfico es

$$p(I_1 = 1|I_2 = i_2, I_3) = \frac{m(1, i_2, \cdot)m(1, \cdot, i_3)/m(1, \cdot, \cdot)}{\sum_{i_1} m(i_1, i_2, \cdot)m(i_1, \cdot, i_3)/m(i_1, \cdot, \cdot)}.$$

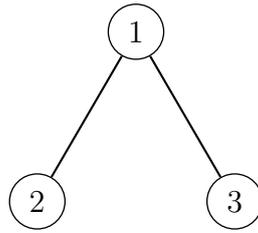
Por otra lado, dado el modelo log-lineal (12, 13), el momio condicional de  $I_2$  dado los valores de  $I_1$  y  $I_3$  puede expresarse como:

$$\begin{aligned} \log \frac{p(I_2 = 1|I_1 = i_1, I_3 = i_3)}{p(I_2 = 0|I_1 = i_1, I_3 = i_3)} &= \log \frac{m(i_1, 1, i_3)}{m(i_1, 0, i_3)} \\ &= \log m(i_1, 1, i_3) - \log m(i_1, 0, i_3) \\ &= [u + u_1(i_1) + u_2(1) + u_3(i_3) + u_{12}(i_1, 1) + u_{13}(i_1, i_3)] \\ &\quad - [u + u_1(i_1) + u_2(0) + u_3(i_3) + u_{12}(i_1, 0) + u_{13}(i_1, i_3)] \\ &= [u_2(1) - u_2(0)] + [u_{12}(i_1, 1) - u_{12}(i_1, 0)]. \end{aligned}$$

En este caso, el momio condicional sólo depende de la variable  $I_1$ . Por lo tanto, la regresión logística asociada a la variable respuesta  $I_2$  sólo tiene como variable explicativa a  $I_1$ :

$$\text{logit } p(I_2 = 1|I_1 = i_1) = \beta_0 + \beta_1(i_1) \quad (2.24)$$

Con estos dos ejemplos, es fácil notar que los términos para las regresiones logísticas con variables respuesta  $I_1$  y  $I_2$ , están dados por sus fronteras,  $\text{bd}(1) = \{2, 3\}$  y  $\text{bd}(2) = \{1\}$  (ver figura 2.4).

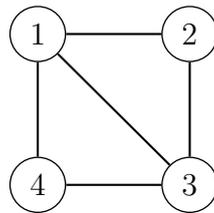


**Figura 2.4:** Modelo gráfico, con clases generadoras  $\{\{1, 2\}, \{1, 3\}\}$ .

Más aún, a partir del modelo gráfico es posible identificar que términos de interacción están presentes en la regresión logística. Por ejemplo, si ahora suponemos el siguiente modelo gráfico con clases generadoras  $\{\{1, 2, 3\}, \{1, 3, 4\}\}$ , su expresión está dada por:

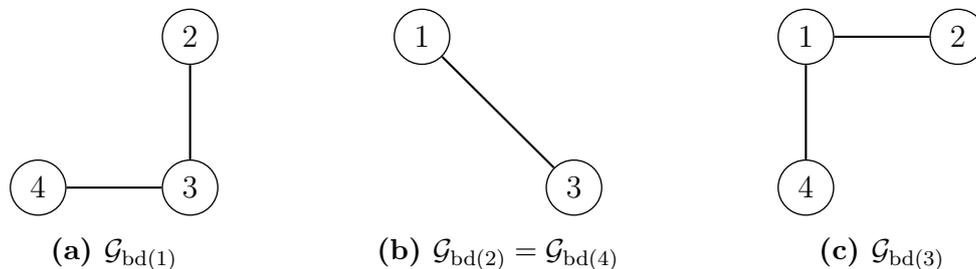
$$\log m(i) = u + u_1(i_1) + u_2(i_2) + u_3(i_3) + u_4(i_4) + u_{12}(i_1, i_2) + u_{13}(i_1, i_3) + u_{14}(i_1, i_4) + u_{23}(i_2, i_3) + u_{34}(i_3, i_4) + u_{123}(i_1, i_2, i_3) + u_{134}(i_1, i_3, i_4).$$

La gráfica asociada a este modelo log-lineal está dada por el siguiente diagrama.



Si suponemos que las 4 variables son binarias, los términos para cada regresión logística se pueden identificar por medio de las interacciones dentro de los cliques de la gráfica inducida por su frontera. Por ejemplo, en la figura 2.5a se muestra la gráfica inducida por la frontera de la variable  $I_1$ . En este caso, se observa que los cliques de la gráfica son  $\{\{2, 3\}, \{3, 4\}\}$ . Por lo tanto, la regresión logística asociada a la variable  $I_1$  tiene dos términos de interacción:

$$\text{logit } p(I_1 = 1 | I_2 = i_2, I_3 = i_3, I_4 = i_4) = \beta_0 + \beta_2(i_2) + \beta_3(i_3) + \beta_4(i_4) + \beta_{23}(i_2 i_3) + \beta_{34}(i_3 i_4).$$



**Figura 2.5:** Gráficas inducidas por la frontera de cada vértice en la gráfica  $\mathcal{G} = (V, E)$ , donde  $V = \{1, 2, 3, 4\}$  y  $E = \{(1, 2), (1, 3), (1, 4), (2, 3), (3, 4)\}$ .

De manera análoga, las regresiones logísticas correspondientes a las variables  $I_2$ ,  $I_3$  e  $I_4$  quedan determinadas por las siguientes expresiones.

$$\text{logit } p(I_2 = 1 | I_1 = i_1, I_3 = i_3) = \beta_0 + \beta_1(i_1) + \beta_3(i_3) + \beta_{13}(i_1 i_3),$$

$$\text{logit } p(I_3 = 1 | I_1 = i_1, I_2 = i_2, I_4 = i_4) = \beta_0 + \beta_1(i_1) + \beta_2(i_2) + \beta_4(i_4) + \beta_{12}(i_1 i_2) + \beta_{14}(i_1 i_4),$$

$$\text{logit } p(I_4 = 1 | I_1 = i_1, I_3 = i_3) = \beta_0 + \beta_1(i_1) + \beta_3(i_3) + \beta_{13}(i_1 i_3).$$

En este caso, se observa que las regresiones logísticas con variables respuesta  $I_2$  ó  $I_4$  tienen los mismos términos, es decir, las probabilidades de ocurrencia dependen de las mismas variables explicativas. Como se verá en el siguiente capítulo, numéricamente las probabilidades estimadas considerando un modelo gráfico y una regresión logística son equivalentes si la gráfica inducida por la frontera de la variable respuesta es completa.

Cabe mencionar, que la relación entre un modelo gráfico log-lineal y una regresión logística se puede generalizar para modelos log-lineales. El estudio de esta relación se reporta en [Ramírez-Aldana y Eslava-Gómez \(2015\)](#).

# Capítulo 3

## Aplicación

En este capítulo se muestra el uso de los modelos gráficos log-lineales para modelar las relaciones de asociación para 6 variables categóricas seleccionadas de la Encuesta Nacional de Adicciones (ENA 2011). Estas variables corresponden al consumo de tabaco, alcohol y drogas, las cuales se suponen como respuesta, y de las variables ingreso del hogar, escolaridad y sexo como posibles variables explicativas.

La información utilizada para esta sección corresponde a individuos entrevistados en la ENA 2011. La base de datos correspondiente consiste de 16,249 registros y 1007 campos. En el apéndice A de este trabajo se da una breve descripción de los cuestionarios aplicados.

Para el ajuste de los modelos gráficos log-lineales se utilizó una muestra de 10,294 individuos en los grupos de edad de 18 a 65 años. Se usaron los paquetes `gRbase` y `gRim` del software R para la estimación de los modelos.

### 3.1. ENA 2011

La ENA 2011 fue un estudio coordinado por diferentes organismos de salud pública. La elaboración de los cuestionarios aplicados fueron llevados a cabo por el Instituto Nacional de Psiquiatría “Ramón de la Fuente Muñiz”, los cuales están orientados a recolectar información acerca de los hogares e individuos.

El diseño muestral para la ENA 2011 estuvo a cargo del Instituto Nacional de Salud Pública. La muestra seleccionada se realizó con un diseño de 4 etapas de selección y estratificado por entidad federativa y tipo de urbanidad (rural, urbano y metropolitano). Cada etapa de selección se describe en el apéndice A de este trabajo.

La ENA 2011 es una encuesta con representatividad nacional, por tipo de urbanidad y para ocho regiones del país. Su objetivo principal es estimar las prevalencias de consumo de alcohol, tabaco, y drogas. Asimismo, se busca evaluar las tendencias de consumo para la población adolescente (de 12 a 17 años) y para la población adulta (18 a 65 años), así como identificar algunos factores relacionados con el abuso y consumo de sustancias para incidir en políticas gubernamentales para el control de sustancias.

## 3.2. Análisis exploratorio

### Población de estudio

Se consideró la información recolectada por el cuestionario de individuos, con un total de 16,249 personas entrevistadas. De esta muestra, se descartaron datos faltantes, eliminándose 3,396 registros.

Debido al diseño de la muestra, existen hogares en donde se entrevistó un adulto y un adolescente, en consecuencia, no se puede afirmar que las observaciones recolectadas en un mismo hogar sean independientes. En este caso, para reducir la posible asociación entre las observaciones de un mismo hogar, se consideró solo población adulta en el grupo de edad de 18 a 65 años. El tamaño de muestra utilizado para este trabajo fue de 10,294 adultos.

### Variables de interés

De acuerdo con el reporte de la ENA 2011, la encuesta tiene el propósito de estudiar el uso, abuso y dependencia de sustancias psicoactivas. Para ello, el cuestionario se enfoca en la estimación de las prevalencias de consumo de las sustancias de alcohol, tabaco y drogas.

De acuerdo con el tiempo de consumo, la prevalencia de consumo se clasifica en tres tipos.

- a) *Prevalencia alguna vez en la vida.* Índice que considera a la población que reportó haber consumido alguna sustancia (alcohol, tabaco, drogas) por lo menos una vez en la vida.
- b) *Prevalencia último año.* Índice que considera a la población que reportó haber tomado alguna sustancia por lo menos una vez en el año previo a la entrevista.
- c) *Prevalencia último mes.* Índice que considera a la población que reportó haber tomado alguna sustancia por lo menos una vez en los treinta días previos a la entrevista.

Para la aplicación, en este trabajo, sólo se consideró la prevalencia de último año para las sustancias de alcohol, tabaco, drogas médicas y drogas ilegales, recolectadas con las siguientes preguntas dadas en el cuestionario de individuos:

- *Tabaco.* ¿Fumó tabaco en los últimos 12 meses?
- *Drogas médicas.* ¿Consumió en los últimos 12 meses opiáceos, tranquilizantes, sedantes, barbitúricos, anfetaminas o estimulantes?
- *Drogas ilegales.* ¿Consumió en los últimos 12 meses marihuana, cocaína, crack, alucinógenos, inhalables, heroína, estimulante tipo anfetamínico?
- *Alcohol.* ¿Consumió en los últimos 12 meses alcohol?

Asimismo, se seleccionaron las variables ingreso por hogar, edad, sexo y escolaridad del entrevistado como posibles variables asociadas al consumo de sustancias. En la base de datos, la variable ingreso está reportada por región económica: A, B y C. Se observó que la diferencia entre los ingresos por región no eran significativas por lo que se decidió ignorar información acerca de la región económica. La descripción para las seis variables categóricas seleccionadas se muestra en la tabla siguiente.

Variable	Descripción	Categorías
sexo	Género de la persona entrevistada	1: masculino 2: femenino
escolaridad	Último grado escolar que completó en la escuela.	01: Primaria incompleta (1 a 5 años) 02: Primaria completa (6 años) 03: Secundaria incompleta (1 a 2 años) 04: Secundaria completa o equivalente (6 años) 05: Bachillerato incompleto (1 a 2 años) 06: Bachillerato completo o equivalente (3 años) 07: Estudios universitarios incompletos (1 a 3 años) 08: Estudios universitarios completos (4 a 5 años) 09: Estudios de posgrado (2 a 4 años)
ingreso	Ingreso mensual del hogar	1: Menos de un salario mínimo 2: 1 salario mínimo 3: 2 salarios mínimos 4: $>2$ y $\leq 4$ salarios mínimos 5: $>4$ y $\leq 6$ salarios mínimos 6: $>6$ y $\leq 8$ salarios mínimos 7: $> 8$ salarios mínimos
tabaco	Consumo de tabaco en los últimos 12 meses.	1: sí 2: no
drog_med	Consumo de drogas médicas sin prescripción en los últimos 12 meses.	1: sí 2: no
drog_ileg	Consumo de drogas ilegales en los últimos 12 meses	1: sí 2: no
alcohol	Consumo de alcohol en los últimos 12 meses.	1: sí 2: no

**Tabla 3.1:** Codificación de las 6 variables seleccionadas para el ajuste de modelos gráficos log-lineales.

## Objetivos

Dentro de los objetivos que se plantearon en esta sección están:

1. Ilustrar el uso de modelos gráficos para 6 variables categóricas seleccionadas de la Encuesta Nacional de Adicciones 2011. Con esto, identificar posibles independencias condicionales o marginales entre los factores de riesgo asociados a el consumo de sustancias.
2. Ilustrar la relación que existe entre un modelo gráfico y las 3 regresiones logísticas asociadas a cada sustancia de consumo. Así como comparar las probabilidades estimadas obtenidas bajo ambos modelos.

## Efecto del diseño muestral

Con la muestra de 10,294 adultos se evaluó el efecto del diseño muestral mediante la variable `ponde_indiv_final`, reportada en la base de datos de individuos. Esta variable corresponde al factor de expansión o peso muestral de la persona entrevistada. Esto tiene el propósito de evaluar si el efecto del diseño debe ser considerado en el ajuste de un modelo gráfico log-lineal.

Se obtuvieron las prevalencias estimadas del consumo de último año ignorando el efecto del diseño muestral, esto al suponer un muestro aleatorio simple, y considerando los pesos muestra del diseño. A continuación se reportan los valores obtenidos.

	Alcohol	Tabaco	Drogas médicas	Drogas ilegales
$100 \times \hat{p}^*$	52.78	22.70	0.97	1.45
$[\hat{\text{var}}(\hat{p}^*)]^{1/2}$	0.27	0.25	0.06	0.07
IC 95 %	52.25 - 53.31	22.21 - 23.19	0.85 - 1.09	1.31 - 1.59
cve	0.51	1.01	6.19	4.82
$100 \times \hat{p}^{**}$	56.31	23.70	0.86	1.63
$[\hat{\text{var}}(\hat{p}^{**})]^{1/2}$	0.885	0.72	0.13	0.21
IC 95 %	54.57 - 58.03	22.30 - 25.10	0.60 - 1.10	1.21 - 2.04
cve	1.52	3.00	14.97	12.98
deff	3.27	2.92	1.99	2.87

Nota: valores obtenidos con la muestra de 10,294 adultos.

\* Proporción estimada ignorando el efecto del diseño muestral.

\*\* Proporción estimada considerando el diseño muestral.

**Tabla 3.2:** Estimación del porcentaje de la población para el consumo de sustancias de último año.

En esta tabla se observa que para la variable alcohol, los intervalos de confianza considerando e ignorando el efecto del diseño no se interceptan, por lo tanto, las estimaciones

calculadas a partir de ignorar el diseño serán incorrectas. Para el resto de las variables, los intervalos de confianza ignorando el efecto de diseño interceptan con los intervalos de confianza considerando el efecto de diseño. Las gráficas para los intervalos de confianza se muestra en el apéndice A de este trabajo.

De acuerdo con el coeficiente de variación **cve**, la estimación puntual más precisa considerando el diseño fue la prevalencia del consumo de alcohol con un valor de 1.5%. Las sustancias drogas médicas e ilegales reportaron un **cve** mayor a 10%, esto indica una estimación imprecisa (Särndal et al., 1991, pp. 41-42).

Con relación al deff estimado, éste se define como el cociente entre las desviaciones estándar estimadas del diseño muestral y un muestreo aleatorio simple, es decir

$$\text{deff} = \frac{[\hat{\text{var}}(\hat{p}^{**})]^{1/2}}{[\hat{\text{var}}(\hat{p}^*)]^{1/2}}.$$

En la práctica, este valor se utiliza como referencia para medir la eficiencia del diseño con respecto a un muestreo aleatorio simple. Para diseños complejos, usualmente el deff es mayor a 1. En particular, para la ENA 2011, se observaron deff mayores a 1 para las 4 tipos de sustancias.

En resumen, considerando los pesos muestrales la sustancia que se estima con mayor precisión es la variable alcohol. Esta sustancia tiene el mayor consumo en la muestra. Al ignorar el diseño muestral, el intervalo de confianza para la variable alcohol no intercepta con el intervalo obtenido con diseño, por lo tanto, las estimaciones obtenidas no serán correctas. Dicho esto, se concluye que el efecto del diseño podría ser incorporado al ajuste del modelo. Sin embargo, los algoritmos que se utilizan para la selección de modelo no consideran los pesos muestrales, por lo que se decidió ignorar.

### Definición de las variables categóricas

De acuerdo a las estimaciones obtenidas, las prevalencias de consumo de drogas médicas e ilegales son menores que el 2%. Al realizar la clasificación cruzada para estas variables se obtendrá una tabla con celdas vacías. Para obtener una tabla menos dispersa, se agregó en una sola variable el consumo de drogas médicas e ilegales, la cual denominaremos como drogas.

Por otro parte, para simplificar el número de celdas posibles, las categorías de la variable escolaridad se agregaron en dos niveles: alta y baja. Para esto, se utilizó el grado promedio de escolaridad de la población, reportado por el Censo de población y vivienda 2010, con un valor igual a 8.6 años, lo cual es equivalente a secundaria incompleta.

Análogamente, las categorías de la variable ingreso del hogar se agregaron en dos niveles: alta y baja. Para esta variable, se consideró la información de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2010. En este caso, el ingreso del hogar se reporta en deciles, por lo que se utilizó el cuantil 0.5 como criterio para clasificar, siendo igual a 7,422 o equivalente a 4.13 salarios mínimos diarios. La codificación para estas variables se muestra en la tabla 3.3.

En adelante, las variables categóricas se denotarán por sus iniciales: *A*, alcohol; *D*, drogas; *E*, escolaridad; *I*, ingreso; *S*, sexo y *T*, tabaco.

Variable	Codificación	Categorías agregadas
escolaridad	1 : baja	01: Primaria incompleta (1 a 5 años) 02: Primaria completa (6 años) 03: Secundaria incompleta (1 a 2 años)
	2 : alta	04: Secundaria completa o equivalente (6 años) 05: Bachillerato incompleto (1 a 2 años) 06: Bachillerato completo o equivalente (3 años) 07: Estudios universitarios incompletos(1 a 3 años) 08: Estudios universitarios completos (4 a 5 años) 09: Estudios de posgrado (2 a 4 años)
ingreso	1 : baja	1: Menos de un salario mínimo 2: 1 salario mínimo 3: 2 salarios mínimos 4: $> 2$ y $\leq 4$ salarios mínimos
	2 : alto	5: $>4$ y $\leq 6$ salarios mínimos 6: $> 6$ y $\leq 8$ salarios mínimos 7: $> 8$ salarios mínimos

**Tabla 3.3:** Descripción de las categorías agregadas para las variables escolaridad e ingreso.

Una vez definidas las variables binarias se calcularon las razones de momios marginales. En la tabla 3.4, la entrada  $(i, j)$  con  $i, j = 1, 2, \dots, 6$  muestra la razón de momios marginales de las variables  $(i, j)$ .

	$S$	$E$	$I$	$A$	$T$	$D$
$S$		0.83	0.65	3.38	2.05	3.34
$E$	1.20		6.98	0.52	0.80	0.74
$I$	1.54	0.14		0.40	0.67	0.69
$A$	0.30	1.92	2.50		5.33	4.76
$T$	0.49	1.25	1.49	0.19		5.62
$D$	0.30	1.35	1.45	0.21	0.18	

**Tabla 3.4:** Razón de momios marginales para las variables sexo ( $S$ ), escolaridad ( $E$ ), ingreso del hogar ( $I$ ), alcohol ( $A$ ), tabaco ( $T$ ) y drogas ( $D$ ).

A partir de esta tabla, se observa que la razón de momios entre alcohol y tabaco es 5.33, esto indica que un individuo tiene una mayor probabilidad de consumir alcohol que tabaco. De manera análoga, se tiene la misma relación entre las variables alcohol y drogas, donde su razón de momios es 4.76. Finalmente, se observa que la razón de momios entre las variables tabaco y drogas es 5.62. En este caso, se observa que el alcohol es la sustancia preferida con respecto al tabaco y las drogas.

De acuerdo con las razones de momios, las variables que tienen un mayor grado de asociación son escolaridad e ingreso, alcohol y tabaco, y tabaco y drogas.

En la figura 3.1 se muestra la gráfica de mosaico por parejas. En la diagonal se muestra la distribución de las observaciones para cada variable categórica. La interpretación para las gráficas de mosaicos debajo de la diagonal está dada de la siguiente forma. Si seleccionamos las variables sexo y escolaridad, su gráfica de mosaico se muestra en la entrada (2, 1). Para esta gráfica, la primera columna muestra la distribución de la variable escolaridad dentro de la categoría 1 de la variable sexo. Donde el color azul oscuro indica la categoría 1 de la variable escolaridad y el color azul claro indica la categoría 2. De manera análoga, la segunda columna muestra la distribución de la variable escolaridad dentro de la categoría 2 de la variable sexo. De esta manera, se observa que existe una mayor proporción dentro de la población masculina (= 1) con escolaridad alta (= 2) comparada con la población femenina. Para las gráficas de mosaico por arriba de la diagonal, la interpretación es simétrica con respecto a las variables.

En un caso más general, si utilizamos la gráfica de mosaico con las 6 variables, las interpretaciones resultan difíciles. Más adelante, se ajustarán modelos gráficos log-lineales para interpretar la asociación entre las variables categóricas.

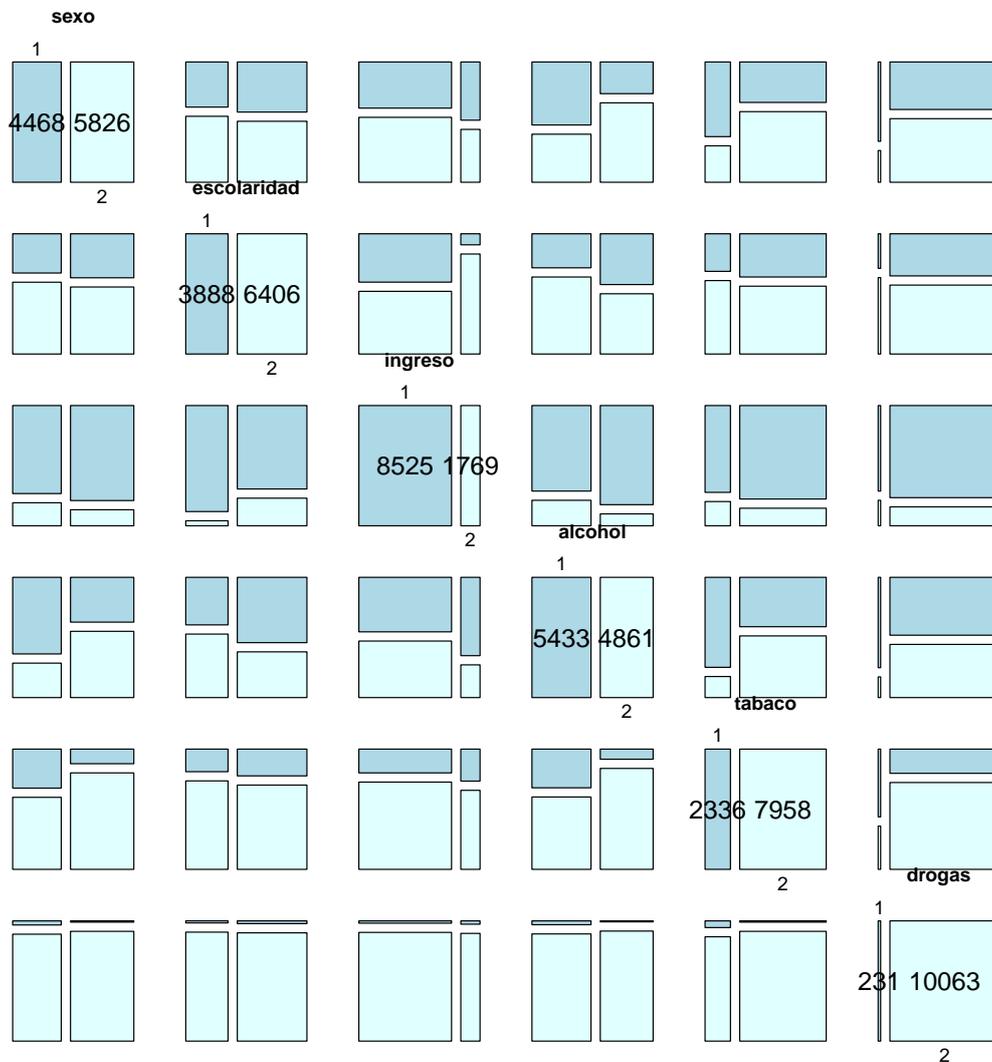


Figura 3.1: Gráfica de mosaico por parejas

### 3.3. Modelos jerárquicos

Adicional al ajuste de modelos log-lineales gráficos, se ajustaron modelos jerárquicos. Esto es con el objetivo de mostrar las ventajas de usar modelos gráficos log-lineales para el análisis de variables categóricas. Para esto, se definieron distintos órdenes de interacción, para luego seleccionar, de acuerdo con el Akaike (AIC), el orden de interacción óptimo, es decir, aquel grado que minimice el AIC. Para el ajuste de modelos jerárquicos se utilizó la rutina `dmod` de la librería `gRim`. En la tabla siguiente se reporta las estadísticas de devianza y  $X^2$  de Pearson, así como los grados de libertad y AIC para cada modelo jerárquico ajustado.

<sup>1</sup> Modelo	<sup>2</sup>  K	Devianza	$X^2$	g.l.	AIC
(S, E, I, A, T, D)	6	3,822.89	4,954.525	57	64,472.31
(SE, SI, ..., TD)	15	111.86	111.65	42	60,991.27
(SEI, SEA, ..., ATD)	20	19.20	16.32	22	<b>60,938.62</b>
(SEIA, SEIT, ..., IATD)	15	5.58	-	7	60,955.00
(SEIAT, ..., EIATD)	6	0.00	-	1	60,961.42

<sup>1</sup>Conjunto de clases generadoras maximales.

<sup>2</sup>Cardinalidad del conjunto de clases generadoras maximales.

**Tabla 3.5:** Estadísticas de bondad de ajuste para diferentes modelos log-lineales jerárquicos. Donde  $S$ , denota el sexo del individuo;  $E$ , nivel escolaridad;  $I$ , ingreso por hogar;  $A$ , consumo de alcohol;  $T$ , consumo de tabaco;  $D$ , consumo de drogas.

A partir de la tabla anterior, se observa que el modelo jerárquico de interacciones de segundo orden minimiza el AIC, en donde su conjunto de clases generadoras consiste de 20 subconjuntos de dimensión 3.

Una vez seleccionado el modelo jerárquico con interacciones de segundo orden, se hizo una selección de modelo con el comando `stepwise` del paquete `gRim`. Como se vio en la sección anterior, aunque un modelo jerárquico no necesariamente es gráfico, siempre se puede representar por medio de una gráfica. Con el método `stepwise`, la selección de un modelo jerárquico está basado en agregar o eliminar aristas de la gráfica. Al utilizar este método, la búsqueda del modelo es más eficiente con respecto a métodos específicos para modelos log-lineales.

En la tabla 3.6 se reporta los modelos obtenidos para la selección `stepwise` en direcciones `backward` y `forward`. En dirección `backward` el modelo inicial es el modelo jerárquico con conjunto de interacciones de segundo orden. En este caso, el modelo seleccionado no es gráfico, y por lo tanto, tampoco descomponible. En dirección `forward`, el modelo inicial es el modelo de independencia. El modelo seleccionado es gráfico y descomponible.

De acuerdo con el valor del AIC, el mejor ajuste se obtuvo con el modelo jerárquico. Las clases generadoras de este modelo son:

$$\{\{S, E, I\}, \{S, E, A\}, \{S, I, A\}, \{S, A, T\}, \{S, A, D\}, \{S, T, D\}, \{E, I, A\}, \{A, T, D\}\}$$

Las interacciones de segundo orden para los 10 subconjuntos del modelo son significativas. Sin embargo, como se puede notar, resulta difícil identificar relaciones de asociación para variables que no estén en una misma clase generadora. En la sección siguiente, se ilustra el uso de modelos gráficos para interpretar relaciones de asociación entre las variables.

Método de selección	Clases generadoras del modelo	AIC	Gráfico	Descomp.
<i>backward</i>	$\{\{S, E, I\}, \{S, E, A\}, \{S, I, A\}, \{S, A, T\}, \{S, A, D\}, \{S, T, D\}, \{E, I, A\}, \{A, T, D\}\}$	60,921.68	No	No
<i>forward</i>	$\{\{I, S, A, E\}, \{D, S, A, T\}\}$	60,923.99	Sí	Sí

**Tabla 3.6:** Selección de modelos jerárquicos usando la opción *stepwise*.

La expresión para el modelo seleccionado es igual a

$$\begin{aligned} \log m(i) = & u + u_A(a) + \dots + u_T(t) + u_{AD}(ad) + \dots + u_{AT}(at) + \dots + u_{ST}(st) + \\ & u_{SEI}(sei) + u_{SEA}(sea) + u_{SIA}(sia) + u_{SAT}(sat) + u_{SAD}(sad) + \\ & u_{STD}(std) + u_{EIA}(eia) + u_{ATD}(atd). \end{aligned} \tag{3.1}$$

### 3.4. Modelos gráficos log-lineales

Debido a que no hay información preliminar aportada por investigadores especialistas en el tema, no fue posible identificar alguna estructura de asociación entre las variables. Cabe mencionar que dentro de los modelos gráficos hay dos aspectos a considerar: la estimación de los parámetros o de los valores ajustados y de la estimación de la estructura.

Para ajustar un modelo gráfico log-lineal se utilizaron métodos automatizados de selección de modelo. Estos últimos, están basados en dos criterios: i) dirección de búsqueda y ii) el tipo de modelo gráfico seleccionado.

Con el primer criterio, se realizó una selección *stepwise* en direcciones *backward* y *forward*. Para el segundo criterio, en cada paso de la selección *stepwise* se tiene la opción de que la inclusión o eliminación de las aristas estén restringidas a que la gráfica final sea la de un modelo gráfico descomponible, es decir, se pueden obtener dos tipos de modelos: gráficos y gráficos descomponibles. Al combinar estos dos criterios de búsqueda se obtienen 4 procedimientos para la selección de modelo. Para datos de la ENA 2011, en la tabla 3.7 se reporta el modelo obtenido. En este caso, al aplicar los 4 procedimientos se obtuvo el mismo modelo, el cual fue gráfico y descomponible.

Modelo	Tipo	Devianza	g.l.	AIC	BIC
$\{\{I, S, A, E\}, \{D, S, A, T\}\}$	Gráfico descomponible	32.58	36	60,293	61,119.46

**Tabla 3.7:** Modelo gráfico log-lineal seleccionado.

Para el modelo ajustado, se realizó la prueba de significancia del modelo. La estadística para esta prueba está dada por la devianza residual,  $d = 32.58$ . Al comparar este valor con el valor en tablas de una ji-cuadrada con 36 grados de libertad ( $50.99, \alpha = 0.05$ ), existe evidencia para no rechazar la hipótesis de que el modelo describe a los datos.

Otra manera de evaluar el ajuste del modelo es probar la hipótesis nula  $H_0$  de que el modelo de independencia (ver tabla 3.6) describe a los datos, contra la hipótesis  $H_1$  de que el modelo ajustado describe a los datos. En este caso, el valor de la estadística de prueba es la diferencia de devianzas  $3,822.89 - 32.58 = 3,790.31$ . Al comparar éste con el valor en tablas de una ji-cuadrada con 21 grados de libertad ( $32.67, \alpha = 0.05$ ), se rechaza la hipótesis nula, es decir, existe evidencia para suponer que el modelo ajustado describe a los datos.

Finalmente, la expresión del modelo seleccionado es

$$\begin{aligned} \log m(i) = & u + u_A(a) + \cdots + u_T(t) + u_{AE}(ae) + \cdots + u_{ST}(st) + \\ & u_{Ade}(ade) + \cdots + u_{IST}(ist) + u_{ISAE}(isae) + u_{DSAT}(dsat). \end{aligned} \quad (3.2)$$

Su gráfica asociada  $\mathcal{G} = (V, E)$  se muestra en la figura 3.2. Al ser un modelo gráfico descomponible, el teorema de Hammersley y Clifford nos permite obtener la factorización de la distribución conjunta mediante la descomposición de la gráfica del modelo. Para esto, se utilizó el algoritmo MCS con el comando `rip` de la librería `gRbase` de R. Usando la notación de punto, la frecuencia estimada en la celda  $(a, d, e, i, s, t)$  es

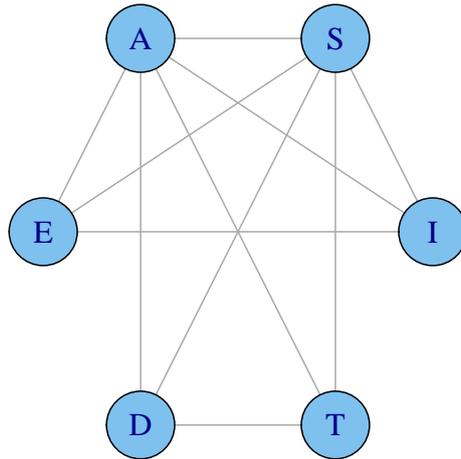
$$\hat{m}(a, d, e, i, s, t) = \frac{n(a, \cdot, e, i, s, \cdot) n(a, d, \cdot, \cdot, s, t)}{n(a, \cdot, \cdot, \cdot, s, t)},$$

en donde el conjunto de clases generadoras es igual a  $\mathcal{C} = \{\{A, E, I, S\}, \{A, D, S, T\}\}$ , y conjunto separador  $\mathcal{S} = \{A, S\}$ . Se observa que la estimación se reduce a obtener frecuencias marginales en tablas de contingencia de dimensión 4, lo cual representa una reducción de 2 dimensiones.

### Propiedades de Markov en el modelo gráfico log-lineal

Para identificar las relaciones de asociación para las variables alcohol, tabaco y drogas se utilizaron las tres propiedades de Markov. En este caso, para la gráfica del modelo se observó lo siguiente.

1. *Propiedad de Markov por parejas.* Esto es, para cualquier pareja de vértices no adyacentes  $\alpha \not\sim \beta$ ,  $\alpha \perp\!\!\!\perp \beta | V \setminus \{\alpha, \beta\}$ .
  - a) La variable drogas es condicionalmente independiente de ingreso dada las variables tabaco, sexo, alcohol y escolaridad,  $D \perp\!\!\!\perp I | \{T, S, A, E\}$ . Esto se puede interpretar como: dado que conocemos las variables tabaco, sexo, alcohol y escolaridad, tener información sobre ingreso del hogar es irrelevante para obtener información acerca de la variable drogas, y viceversa.
  - b) La variable drogas es condicionalmente independiente de escolaridad dada las variables tabaco, ingreso, alcohol y sexo,  $D \perp\!\!\!\perp E | \{T, I, A, S\}$ .



**Figura 3.2:** Gráfica  $\mathcal{G} = (V, E)$  asociada al modelo gráfico log-lineal descomponible. Con clases generadoras  $\mathcal{C} = \{\{A, E, I, S\}, \{A, D, S, T\}\}$  y conjunto separador  $\mathcal{S} = \{A, S\}$ .

- c) El consumo de tabaco es condicionalmente independiente del ingreso dada las variables sexo, alcohol, escolaridad y drogas,  $T \perp\!\!\!\perp I \mid \{S, A, E, D\}$ .
- d) El consumo de tabaco es condicionalmente independiente de la escolaridad dada las variables ingreso, sexo, alcohol y drogas,  $T \perp\!\!\!\perp E \mid \{I, S, A, D\}$ .

Una consecuencia de la propiedad de Markov por parejas es la posibilidad de identificar que parejas de variables tienen una razón de momio condicional igual a 1. En la siguiente tabla, se reportan la razón de momios marginales y condicionales para vértices no adyacentes en la gráfica del modelo.

Pareja	$\theta_{IJ}$	$\hat{\theta}_{IJ}$	$\hat{\theta}_{IJ(R)}$
$(D, E)$	0.74	0.85	1
$(D, I)$	0.69	0.77	1
$(T, E)$	0.80	0.80	1
$(T, I)$	0.67	0.72	1

**Tabla 3.8:** Razón de momios condicionales y marginales para vértices no adyacentes en el modelo. Las expresiones para  $\theta_{IJ}$ ,  $\hat{\theta}_{IJ}$  y  $\hat{\theta}_{IJ(R)}$  se muestran en la sección 3.4.

En la primera y segunda columna se observa que las razones de momios marginales observadas y estimadas bajo el modelo gráfico son menores que 1, esto indica que la ocurrencia de la primera variable es menos verosímil comparada con la segunda, por ejemplo, consumir drogas es menos verosímil que tener escolaridad baja. En la última columna se observan los momios condicionales para las parejas de vértices no adyacentes dado el resto de las variables, siendo iguales a 1.

2. *Propiedad de Markov local.* Para cualquier variable, ésta es condicionalmente independiente de sus vértices no adyacentes dada su frontera,  $\alpha \perp\!\!\!\perp V \setminus \text{cl}(\alpha) \mid \text{bd}(\alpha)$ .

- a) Para la variable drogas, su frontera está dada por  $\text{bd}(D) = \{A, S, T\}$ . En este caso, se tiene que el consumo de drogas es condicionalmente independiente de la escolaridad e ingreso del hogar, dadas las variables alcohol, sexo y tabaco,  $D \perp\!\!\!\perp \{E, I\} | \{A, S, T\}$ . Es decir, dado que conocemos las variables alcohol, sexo y tabaco, inferir sobre la variable droga es irrelevante para inferir sobre las variables escolaridad e ingreso.
- b) Para la variable tabaco, su conjunto frontera es  $\text{bd}(T) = \{A, S, D\}$ . Entonces se tiene que el consumo de tabaco es condicionalmente independiente de escolaridad e ingreso del hogar dada las variables alcohol, sexo y drogas,  $T \perp\!\!\!\perp \{E, I\} | \{A, S, D\}$ .
- c) En el caso de la variable alcohol, se observa que su frontera está dada por el resto de las variables, por lo tanto, no existe una relación de independencia local.

3. *Propiedad de Markov Global.* Si  $(A, B, S)$  es una partición de  $V$  tal que  $S$  separa a  $A$  y  $B$ , entonces  $A \perp\!\!\!\perp B | S$ . En el modelo ajustado, la única relación de independencia global es

$$\{E, I\} \perp\!\!\!\perp \{T, D\} | \{A, S\}. \quad (3.3)$$

Esta relación se interpreta como: dado que conocemos las variables alcohol y sexo, que para obtener información conjunta sobre las variables tabaco y drogas es irrelevante conocer conjuntamente sobre la escolaridad e ingreso, y viceversa.

Por otra parte, el modelo gráfico log-lineal está definido para todas las celdas en la tabla de contingencia, es decir, todas las combinaciones de categorías son observables. Por lo tanto,  $p(i) > 0$  para toda  $i$ . En este caso, por el teorema de Pearl-Paz, las tres propiedades son equivalentes

$$\begin{aligned} & D \perp\!\!\!\perp I | \{T, S, A, E\}, \\ & D \perp\!\!\!\perp E | \{T, I, A, S\}, \\ & T \perp\!\!\!\perp I | \{S, A, E, D\}, \\ & T \perp\!\!\!\perp E | \{I, S, A, D\} \end{aligned} \iff D \perp\!\!\!\perp \{E, I\} | \{A, S, T\}, \iff \{E, I\} \perp\!\!\!\perp \{T, D\} | \{A, S\}. \quad (3.4)$$

Para evaluar el grado de asociación entre las variables, se obtuvieron las devianzas de inclusión para las aristas del modelo, y se muestran en la tabla 3.9. Con base en la devianza, se observa que la asociación más fuerte es el ingreso del hogar y la escolaridad, con una devianza de 718.38 y 4 grados de libertad. La asociación más débil es el consumo de drogas con la variable sexo, con una devianza de 19.825 y 4 grados de libertad.

En la tabla 3.10, se reportan las frecuencias observadas y estimadas para el modelo gráfico seleccionado. Se observa que el 12.5% del total de celdas son vacías, es decir, la tabla es poco densa. Esto se debe a que la variable drogas tiene una baja prevalencia en la muestra. Una manera de obtener una estimación más robusta es eliminar la variable drogas del análisis, sin embargo, se eliminaría el efecto que ésta tiene con las demás variables. Por ello, se decidió hacer un análisis para dos submuestras, uno para individuos que consume drogas y otro para individuos que no consumen.

Arista	Devianza	g.l.	AIC
$(E, I)$	718.38	4	710.4
$(A, T)$	673.80	4	665.8
$(S, T)$	344.10	4	336.3
$(E, A)$	194.50	4	186.5
$(I, A)$	142.60	4	134.3
$(T, D)$	84.60	4	76.6
$(S, E)$	57.85	4	49.9
$(S, I)$	25.69	4	17.7
$(A, D)$	21.12	4	13.1
$(S, D)$	19.83	4	11.8

**Tabla 3.9:** Devianzas de inclusión para las aristas en el modelo gráfico log-lineal seleccionado con clases generadoras:  $\{\{A, S, E, I\}, \{A, D, S, T\}\}$ .

**Tabla 3.10:** Frecuencias observadas y ajustadas para la tabla de contingencia con variables ingreso por hogar ( $I$ ), escolaridad ( $E$ ), sexo ( $S$ ), alcohol ( $A$ ), tabaco ( $T$ ) y drogas ( $D$ ). El modelo ajustado corresponde a uno gráfico descomponible con clases generadoras  $\{\{I, S, A, E\}, \{D, S, A, T\}\}$ .

				<i>Frecuencia observada</i>				<i>Frecuencia ajustada</i>			
				$T$				$1$		$2$	
				$D$	$1$	$2$	$1$	$2$	$1$	$2$	
$S$	$I$	$E$	$A$	$1$	$2$	$1$	$2$	$1$	$2$	$1$	$2$
1	1	1	1	33	401	11	526	32.45	385.63	12.60	540.32
1	1	2	2	3	90	4	425	4.14	87.38	4.14	426.34
1	1	2	1	43	523	21	796	46.22	549.25	17.95	769.58
1	1	2	2	5	109	7	544	5.28	111.31	5.28	543.13
1	2	1	1	4	28	2	28	2.07	24.62	0.80	34.50
1	2	1	2	0	4	0	20	0.19	4.02	0.19	19.60
1	2	2	1	23	272	6	365	22.26	264.50	8.64	370.60
1	2	2	2	3	29	0	143	1.39	29.29	1.39	142.93
2	1	1	1	4	122	5	458	5.51	136.29	5.76	441.44
2	1	2	2	1	87	5	1,537	2.35	91.47	7.51	1,528.68
2	1	2	1	13	304	12	913	11.62	287.39	12.15	930.84
2	1	2	2	4	84	7	1,428	2.19	85.46	7.01	1,428.33
2	2	1	1	0	10	0	33	0.40	9.95	0.42	32.23
2	2	2	2	0	4	0	43	0.07	2.64	0.22	44.08
2	2	2	1	5	108	6	358	4.46	110.37	4.67	357.50
2	2	2	2	0	20	4	251	0.40	15.43	1.27	257.91

### 3.4.1. Modelos gráficos log-lineales para la subpoblación que consume drogas

Para la subpoblación que consume drogas ( $D = 1$ ), su muestra está conformada por 231 observaciones. Para obtener un modelo gráfico log-lineal, se utilizaron los 4 métodos de selección descritos en la sección anterior. Donde el conjunto de vértices es  $V_1 = \{A, E, I, S, T\}$ .

En la tabla 3.11 se reportan los dos modelos obtenidos. En este caso, éstos difieren por la arista  $(A, I)$ . Con base en los valores de AIC y BIC, el modelo a seleccionar es el gráfico descomponible con clases generadoras  $\{\{E, I\}, \{E, S\}, \{A, S, T\}\}$ .

Tipo de modelo	Clases generadoras	Devianza	g.l.	AIC	BIC
Gráfico	$\{\{E, I\}, \{E, S\}, \{A, I\}, \{A, S, T\}\}$	16.08	20	1,314.16	1,355.47
Descomponible	$\{\{E, I\}, \{E, S\}, \{A, S, T\}\}$	14.45	19	1,313.79	1,351.65

**Tabla 3.11:** Modelos seleccionados para la subpoblación que consume drogas ( $D = 1$ ).

Al realizar la prueba de significancia del modelo, el valor de la estadística de prueba es  $d = 14.45$ , que, al comparar con el valor en tablas de una ji-cuadrada con 19 grados de libertad ( $30.15, \alpha = 0.05$ ), indica que hay evidencia para no rechazar la hipótesis de que el modelo describe a los datos.

Por otra parte, al ser el modelo seleccionado descomponible, la factorización para la frecuencia estimada en la celda  $(a, s, e, i, t)$  está dada por

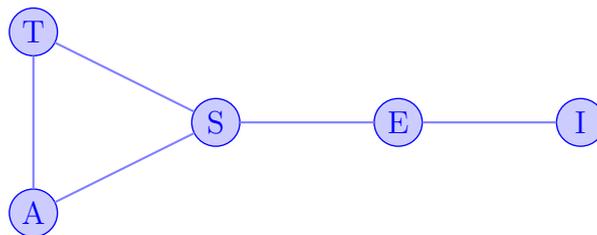
$$\hat{m}(a, e, i, s, t) = \frac{n(a, ., ., s, t)n(., e, s, .)n(., e, i, ., .)}{n(., ., ., s, .)n(., e, ., ., .)},$$

donde el conjunto de cliques es  $\mathcal{C} = \{\{A, S, T\}, \{E, S\}, \{E, I\}\}$ , y con conjunto de separadores  $\mathcal{S} = \{\{E\}, \{S\}\}$ .

La expresión del modelo gráfico log-lineal seleccionado es

$$\log m(i) = u + u_A(a) + u_E(e) + u_I(i) + u_S(s) + u_T(t) + u_{AS}(as) + u_{AT}(at) + u_{EI}(ei) + u_{ES}(es) + u_{ST}(st) + u_{AST}(ast). \tag{3.5}$$

La gráfica para este modelo se muestra a continuación.



**Figura 3.3:** Gráfica del modelo log-lineal descomponible para la subpoblación que consume drogas,  $\mathcal{G}_1 = (V_1, E_1)$ . Donde  $V_1 = \{A, I, E, S, T\}$  y  $E_1 = \{(A, T), (A, S), (S, T), (S, E), (E, I)\}$

En este modelo, se observaron las siguientes relaciones de independencia condicional para las variables de tabaco ( $T$ ) y alcohol ( $A$ ).

1. *Propiedad de Markov por parejas.*

- a) La variable alcohol es condicionalmente independiente de escolaridad dada las variables tabaco, sexo e ingreso, es decir,  $A \perp\!\!\!\perp E | \{T, S, I\}$ .
- b) La variable alcohol es condicionalmente independiente del ingreso dada las variables tabaco, sexo y escolaridad,  $A \perp\!\!\!\perp I | \{T, S, E\}$ .
- c) La variable tabaco es condicionalmente independiente de escolaridad dada las variables alcohol, sexo e ingreso, es decir,  $T \perp\!\!\!\perp E | \{A, S, I\}$ .
- d) La variable tabaco es condicionalmente independiente del ingreso dada las variables alcohol, sexo y escolaridad,  $T \perp\!\!\!\perp I | \{A, S, E\}$ .

2. *Propiedad de Markov local*

- a) Para la variable alcohol, ésta es condicionalmente independientemente de escolaridad e ingreso dadas las variables tabaco y sexo,  $A \perp\!\!\!\perp \{E, I\} | \{T, S\}$ .
- b) La variable tabaco es condicionalmente independiente de escolaridad e ingreso dadas las variables tabaco y sexo,  $T \perp\!\!\!\perp \{E, I\} | \{A, S\}$ .

3. *Propiedad de Markov global*

- a) Las variables tabaco y alcohol son condicionalmente independientes de escolaridad e ingreso dada la variable sexo,  $\{T, A\} \perp\!\!\!\perp \{E, I\} | S$ .
- b) Las variables tabaco, alcohol y sexo son condicionalmente independientes del ingreso dada la variable escolaridad,  $\{T, A, S\} \perp\!\!\!\perp I | E$ .

En este caso,  $p > 0$ , por lo tanto, se satisface la equivalencia entre las tres propiedades de Markov

$$\begin{array}{l}
 A \perp\!\!\!\perp E | \{T, S, I\}, \\
 A \perp\!\!\!\perp I | \{T, S, E\}, \\
 T \perp\!\!\!\perp E | \{A, S, I\}, \\
 T \perp\!\!\!\perp I | \{A, S, E\}
 \end{array}
 \iff
 \begin{array}{l}
 A \perp\!\!\!\perp \{E, I\} | \{T, S\}, \\
 T \perp\!\!\!\perp \{E, I\} | \{A, S\}
 \end{array}
 \iff
 \{A, T\} \perp\!\!\!\perp \{E, I\} | S. \quad (3.6)$$

En la tabla 3.12 se reportan las frecuencias observadas y estimadas para la población que consume drogas. A partir de esta tabla, se observa que hay un 25% de celdas que están vacías; sin embargo, las frecuencias estimadas son distintas de cero, por lo tanto no es necesario hacer una corrección en los grados de libertad.

**Tabla 3.12:** Frecuencias observadas y ajustadas para la tabla de contingencia con variables ingreso del hogar ( $I$ ), escolaridad ( $E$ ), sexo ( $S$ ), alcohol ( $A$ ) y tabaco ( $T$ ), para la población que consume drogas ( $D = 1$ ). El modelo gráfico log-lineal correspondiente tiene clases generadoras  $\{\{A, S, T\}, \{S, E\}, \{E, I\}\}$ .

			<i>Frecuencia observada</i>				<i>Frecuencia ajustada</i>					
			A		1		2		1		2	
			T	1	2	1	2	1	2	1	2	
S	I	E										
1	1	1	33	11	3	4	32.62	12.67	3.48	3.48		
1	1	2	43	21	5	7	47.49	18.44	5.07	5.07		
1	2	1	4	2	0	0	2.97	1.15	0.32	0.32		
1	2	2	23	6	3	0	19.93	7.74	2.13	2.13		
2	1	1	4	5	1	5	4.58	4.79	1.04	3.33		
2	1	2	13	12	4	7	11.97	12.52	2.72	8.71		
2	2	1	0	0	0	0	0.42	0.44	0.09	0.30		
2	2	2	5	6	0	4	5.03	5.25	1.14	3.65		

### 3.4.2. Modelos gráficos log-lineales para la subpoblación que no consumen drogas.

Por último, para personas adultas que no consumen drogas, el tamaño de muestra fue igual a 10,063. Para los 4 métodos de selección, se obtuvo el mismo modelo gráfico descomponible, el cual se reporta a continuación.

Tipo de modelo	Clases generadoras	Devianza	g.l.	AIC	BIC
Descomponible	$\{\{S, T, A\}, \{I, S, A, E\}\}$	10.97	12	57,399.73	57,536.85

**Tabla 3.13:** Modelo seleccionado para la subpoblación que no consume drogas ( $D = 2$ ).

Para evaluar el ajuste del modelo, se probó la significancia del modelo. El valor para esta estadística de prueba es  $d = 10.97$ , que, comparada con el valor en tablas de una ji-cuadrada con 12 grados de libertad (21.02,  $\alpha = 0.05$ ), nos lleva a no rechazar la hipótesis de que el modelo ajustado describe a los datos.

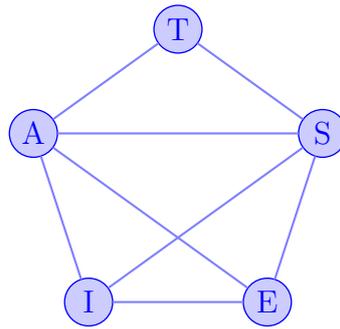
La frecuencia estimada en la celda  $(a, s, e, i, t)$  es

$$\hat{m}(a, e, i, s, t) = \frac{n(a, \dots, s, t)n(a, e, i, s, \cdot)}{n(a, \dots, s, \cdot)}, \tag{3.7}$$

donde el conjunto de cliques es igual a  $\mathcal{C} = \{\{A, S, T\}, \{A, E, I, S\}\}$ , y el conjunto de separador es  $\mathcal{S} = \{A, Sx\}$ . La expresión del modelo seleccionado es

$$\log m(i) = u + u_A(a) + u_E(e) + u_I(i) + u_S(s) + u_T(t) + u_{AE}(ae) + u_{AS}(as) + u_{AT}(at) + u_{EI}(ei) + u_{ES}(es) + u_{SI}(si) + u_{ST}(st) + u_{AST}(ast) + u_{ISAE}(isae).$$

La gráfica asociada a este modelo se muestra en la figura 3.5b.



**Figura 3.4:** Modelo seleccionado para la población que no consume drogas.

Usando las propiedades de Markov se tiene las siguientes relaciones de independencia condicional para las variables de alcohol y tabaco.

1. Para la propiedad de Markov por parejas se tiene que
  - a) La variable tabaco es condicionalmente independiente de ingreso dada las variables alcohol, sexo y escolaridad,  $T \perp\!\!\!\perp I \{A, S, E\}$ .
  - b) La variable tabaco es condicionalmente independiente de escolaridad dada las variables alcohol, sexo e ingreso,  $T \perp\!\!\!\perp E \{A, S, I\}$ .
  - c) Para la variable alcohol, no existe una independencia condicional por parejas, es decir, está relacionada de manera directa con las demás variables.
2. Para la propiedad de Markov local se tiene lo siguiente
  - a) La variable tabaco es condicionalmente independiente de ingreso y escolaridad dada las variables alcohol y sexo,  $T \perp\!\!\!\perp \{I, E\} \{A, S\}$ .
3. Con la propiedad de Markov global, la única relación de independencia condicional es

$$T \perp\!\!\!\perp \{I, E\} \{A, S\}.$$

Como se observa en la figura 3.5b, la gráfica resultante para la subpoblación que no consume drogas difiere de la gráfica completa sólo por las aristas de tabaco-alcohol y tabaco-escolaridad, por lo que hay pocas relaciones de independencia condicional y marginal. En este caso, la única relación de independencia condicional local satisface la propiedad de Markov global.

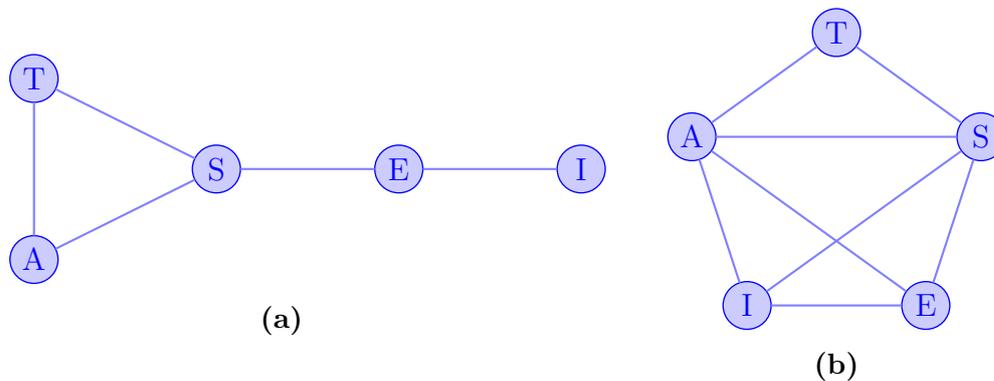
$$\begin{aligned} T \perp\!\!\!\perp I \{A, S, E\}, \\ T \perp\!\!\!\perp E \{A, S, I\} \end{aligned} \iff T \perp\!\!\!\perp \{I, E\} \{A, S\} \tag{3.8}$$

En la tabla 3.14 se muestran las frecuencias observadas y estimadas para la submuestra que no consume drogas ( $D = 2$ ). A partir de esta tabla, se observa que es densa.

			<i>Frecuencia observada</i>				<i>Frecuencia ajustada</i>			
			1		2		1		2	
<i>A</i>			1	2	1	2	1	2	1	2
<i>T</i>			1	2	1	2	1	2	1	2
<i>S</i>	<i>I</i>	<i>E</i>								
1	1	1	401	526	90	425	386.07	540.93	87.60	427.40
1	1	2	523	796	109	544	549.32	769.68	111.07	541.93
1	2	1	28	28	4	20	23.32	32.68	4.08	19.92
1	2	2	272	365	29	143	265.29	371.71	29.26	142.74
2	1	1	122	458	87	1,537	136.83	443.17	91.69	1,532.31
2	1	2	304	913	84	1,428	287.10	929.90	85.36	1,426.64
2	2	1	10	33	4	43	10.14	32.86	2.65	44.35
2	2	2	108	358	20	251	109.93	356.07	15.30	255.70

**Tabla 3.14:** Frecuencias observadas y ajustadas para la muestra que no consume drogas ( $D = 2$ ). Donde ingreso por hogar ( $I$ ), escolaridad ( $E$ ), sexo ( $S$ ), alcohol ( $A$ ) y tabaco ( $T$ ),

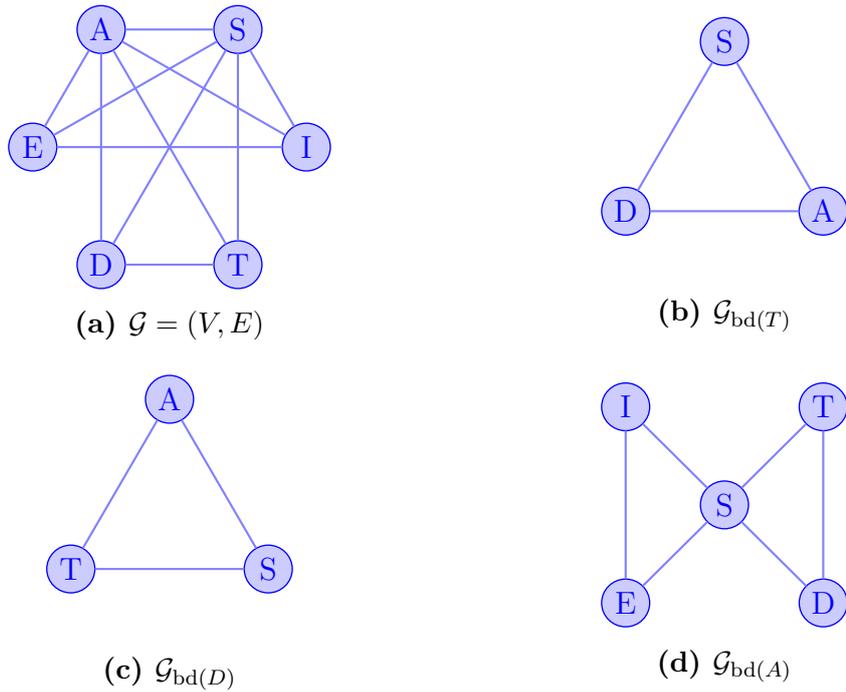
Finalmente, en la figura 3.5 se muestran las gráficas obtenidas para las dos subpoblaciones. En esta figura, se observa que comparado con el modelo gráfico log-lineal obtenido para la subpoblación que consume drogas ( $D=1$ ), la gráfica para la subpoblación que no consume drogas tiene 3 aristas adicionales. Esto indicaría que es más fácil identificar aquellas variables que están asociadas al consumo de tabaco y alcohol para la subpoblación que consume drogas.



**Figura 3.5:** Comparación de modelos gráficos. (a) Modelo gráfico para la subpoblación que consume drogas y (b) modelo gráfico para la subpoblación que no consume drogas.

### 3.5. Regresiones logísticas asociadas al modelo gráfico log-lineal

Para ilustrar la relación entre un modelo gráfico log-lineal y una regresión logística utilizaremos los datos de la muestra completa. Supondremos a las variables alcohol, tabaco y drogas como de respuesta. En cada regresión logística, las variables explicativas corresponderán al conjunto de interacciones, dentro de los cliques, de la gráfica inducida por la frontera de la variable respuesta. En la figura 3.6 se muestran las gráficas inducidas por las variables respuesta tabaco, drogas y alcohol.



**Figura 3.6:** Subgráficas de (a)  $\mathcal{G} = (V, E)$  inducidas por la frontera del consumo de sustancias: (b) tabaco,  $\mathcal{G}_{bd(T)}$ , con clase generadora  $\{S, D, A\}$ ; (c) drogas,  $\mathcal{G}_{bd(D)}$ , con clase generadora  $\{A, T, S\}$  y (d) alcohol,  $\mathcal{G}_{bd(A)}$ , con clases generadoras  $\{\{E, S, I\}, \{S, T, D\}\}$ .

A partir de esta gráfica, si suponemos a la variable tabaco como de respuesta, su regresión logística tiene como variables explicativas a las variables alcohol ( $A$ ), drogas ( $D$ ) y sexo ( $S$ ). La expresión para esta regresión logística es:

$$\log \left( \frac{P[T = 1]}{1 - P[T = 1]} \right) = \alpha + \beta_S(S = 2) + \beta_A(A = 2) + \beta_D(D = 2) + \beta_{SA}(S = 2)(A = 2) + \beta_{SD}(S = 2)(D = 2) + \beta_{AD}(A = 2)(D = 2) + \beta_{ASD}(A = 2)(S = 2)(D = 2). \quad (3.9)$$

En donde para cada variable explicativa, se definió la categoría 1 como de referencia. La estimación para este modelo se muestra en la tabla 3.15.

Coeficiente	$\hat{\beta}$	$\sqrt{\hat{\text{var}}(\hat{\beta})}$	$z$	$p\text{-value}$	$\exp(\hat{\beta})$
Intercepto	0.946	0.186	5.09	<0.001	
$\beta_S(S = 2)$	-0.990	0.352	-2.82	0.005	0.37
$\beta_A(A = 2)$	-0.946	0.465	-2.03	0.042	0.39
$\beta_D(D = 2)$	-1.283	0.190	-6.75	< 0.001	0.28
$\beta_{SA}(S = 2)(A = 2)$	-0.173	0.754	-0.23	0.819	0.84
$\beta_{SD}(S = 2)(D = 2)$	0.152	0.357	0.43	0.670	1.16
$\beta_{AD}(A = 2)(D = 2)$	-0.302	0.472	-0.64	0.523	0.74
$\beta_{ASD}(A = 2)(S = 2)(D = 2)$	-0.220	0.763	-0.29	0.773	0.80

Devianza nula: 11,025.7 con 10,293 grados de libertad  
Devianza residual 9,540.7 con 10,286 grados de libertad  
AIC: 9,557

**Tabla 3.15:** Ajuste de una regresión logística con variable respuesta tabaco ( $T = 1$ ) y variables explicativas sexo ( $S$ ), alcohol ( $A$ ) y drogas ( $D$ ).

Las interpretaciones para los coeficientes en la regresión logística (3.9) están dados en función de los momios de la variable respuesta. Por ejemplo para valores fijos de sexo y alcohol, el momio para un individuo que no consume drogas ( $D = 2$ ) es  $\exp(\hat{\beta}_D) = 0.28$  veces el momio para alguien que consume drogas ( $D = 1$ ), lo cual indica que el consumo de drogas incrementa la probabilidad de consumir tabaco. Mientras que, para los términos de interacción, se observa que la razón de momios para tabaco disminuye en 0.84 para una mujer ( $S = 2$ ) que no consume alcohol ( $A = 2$ ) con respecto a un hombre que consume alcohol.

Con base en la estadística de cociente de verosimilitudes, se probó la hipótesis nula de que todos los términos en la regresión sean iguales a cero, es decir,

$$H_0 : \boldsymbol{\beta} = (\beta_A, \dots, \beta_{SAD}) = (0, \dots, 0) \text{ vs } H_a : \boldsymbol{\beta} \neq (0, \dots, 0).$$

El valor para esta estadística de prueba es la diferencia de devianzas de una regresión logística con coeficiente constante y la regresión ajustada, cuyo valor es 1,485. Al comparar con el valor en tablas de una ji-cuadrada con 7 grados de libertad (14.067,  $\alpha = 0.05$ ), se rechaza la hipótesis de que los términos de regresión sean cero, es decir, podemos suponer que la regresión es significativa.

En cuanto a los términos de la regresión, se probaron las siguientes hipótesis

$$H_0 : \beta_j = 0 \text{ vs } H_a : \beta_j \neq 0.$$

Para esto, se utilizó la estadística de Wald definida como

$$z_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\text{var}}(\hat{\beta}_j)}}, \text{ con } j \in i_a.$$

En este caso, el valor de esta estadística se compara contra el valor en tablas de una normal con media cero y varianza uno (1.96,  $\alpha = 0.05$ ). De acuerdo con la tabla 3.15, las interacciones en la regresión logística son no significativas. Sin embargo, al probar la hipótesis

$$H_0 : (\beta_{SD}, \beta_{SA}, \beta_{AD}, \beta_{SAD}) = (0, 0, 0, 0) \text{ vs } H_a : (\beta_{SD}, \beta_{SA}, \beta_{AD}, \beta_{ASD}) \neq (0, 0, 0, 0).$$

El valor de esta estadística de prueba, dada por el cociente de verosimilitudes, es 12.09. Al comparar éste con el valor en tablas de una ji-cuadrada con 4 grados de libertad (9.48,  $\alpha = 0.05$ ), se rechaza la hipótesis nula, lo cual indica que existen términos de interacción que son distintos de cero. Para descartar las interacciones en el modelo que no son significativas, se realizó una selección *stepwise*. Los coeficientes descartados son  $\beta_{ASD}$ ,  $\beta_{SD}$  y  $\beta_{AD}$ .

Por otra parte, si suponemos a la variable drogas como respuesta, su correspondiente regresión logística tiene como variables explicativas a las variables alcohol, tabaco y sexo. Además, tendrá términos de interacción de primero y segundo orden. La expresión para esta regresión logística es

$$\begin{aligned} \log \left( \frac{P[D = 1]}{1 - P[D = 1]} \right) = & \alpha + \beta_S(S = 2) + \beta_T(T = 2) + \beta_A(A = 2) + \beta_{TA}(T = 2)(A = 2) \\ & + \beta_{SA}(S = 2)(A = 2) + \beta_{ST}(S = 2)(T = 2) + \beta_{AST}(A = 2)(S = 2)(T = 2). \end{aligned}$$

Donde las categorías 1 de cada variable se definieron como de referencia. La estimación de este modelo se muestra en la tabla siguiente.

Coeficiente	$\hat{\beta}$	$\sqrt{\hat{\text{var}}(\hat{\beta})}$	$z$	$p\text{-value}$	$\exp(\hat{\beta})$
Intercepto	-2.475	0.102	-24.13	<0.001	
$\beta_S(S = 2)$	-0.733	0.240	3.05	0.002	0.481
$\beta_T(T = 2)$	-1.283	0.190	-6.75	<0.001	0.277
$\beta_A(A = 2)$	-0.574	0.325	-1.76	0.077	0.563
$\beta_{TA}(T = 2)(A = 2)$	-0.303	-0.472	0.64	0.523	0.739
$\beta_{SA}(S = 2)(A = 2)$	0.152	0.357	0.43	0.670	1.164
$\beta_{ST}(S = 2)(T = 2)$	0.118	0.599	0.20	0.844	1.125
$\beta_{AST}(A = 2)(S = 2)(T = 2)$	-0.220	-0.763	0.29	0.773	0.802
Devianza nula: 2,210.9 con 10,293 grados de libertad					
Devianza residual: 2,000.4 con 10,286 grados de libertad					
AIC: 2,016.4					

**Tabla 3.16:** Ajuste del modelo logístico con variable respuesta drogas ( $D = 1$ ) y variables explicativas sexo ( $S$ ), tabaco ( $T$ ) y alcohol ( $A$ ).

Con base en esta tabla, se observa que el momio para el sexo femenino ( $S = 2$ ) es  $\exp(\beta_S) = 0.481$  veces que el momio para masculino, es decir, existe una mayor proba-

bilidad de que un hombre consuma drogas con respecto a una mujer. De igual manera, para la variable tabaco, se tiene que el momio de un individuo que no consume tabaco ( $T = 2$ ) es  $\exp(\beta_T) = 0.27$  veces el momio de un individuo que consume tabaco. En cuanto los términos de interacción, se observa que el momio de que una persona no consuma alcohol ni tabaco es  $\exp(\beta_{TA}) = 0.74$  veces el momio de alguien que consuma tabaco ó alcohol, es decir, un individuo disminuye su probabilidad de consumir drogas si sólo consume una sola sustancia.

Para este modelo, se probó su significancia utilizando la estadística del cociente de verosimilitudes. La diferencia entre la devianza nula y la residual es 210.5, que, comparada con el valor en tablas de una ji-cuadrada con 7 grados de libertad (14.06,  $\alpha = 0.05$ ), indica que el modelo es significativo.

Con la estadística de Wald, los únicos coeficientes significativos de la regresión logística son  $\beta_S$  y  $\beta_T$ . Para probar si es necesario agregar términos de interacción se probó la hipótesis nula de que los términos de interacción sean iguales a cero,

$$H_0 : (\beta_{TA}, \beta_{SA}, \beta_{ST}, \beta_{AST}) = (0, 0, 0, 0) \text{ vs } H_a : (\beta_{TA}, \beta_{SA}, \beta_{ST}, \beta_{AST}) \neq (0, 0, 0, 0)$$

En este caso, la diferencia entre las devianzas de la regresión sin términos de interacción y la regresión logística ajustada es 1.24. Al comparar este valor con el valor en tablas de una ji-cuadrada con 4 grados de libertad (9.48,  $\alpha = 0.05$ ), no se rechaza la hipótesis nula, por lo que podemos suponer que la regresión logística sin términos de interacción es consistente con los datos.

Finalmente, si consideramos el alcohol como variable respuesta, su correspondiente regresión logística tiene como variables explicativas el consumo de tabaco y drogas, así como el sexo, ingreso del hogar y escolaridad. La expresión para este modelo está dada por la siguiente ecuación.

$$\begin{aligned} \log \left( \frac{P[A = 1]}{1 - P[A = 1]} \right) = & \alpha + \beta_S(S = 2) + \beta_I(I = 2) + \beta_T(T = 2) + \beta_D(D = 2) + \beta_E(E = 2) \\ & + \beta_{TS}(T = 2)(S = 2) + \beta_{DS}(D = 2)(S = 2) + \beta_{ES}(E = 2)(S = 2) \\ & + \beta_{IS}(I = 1)(S = 1) + \beta_{TD}(T = 2)(D = 2) + \beta_{IE}(I = 2)(E = 2) \\ & + \beta_{TDS}(T = 2)(D = 2)(S = 2) + \beta_{IES}(I = 2)(E = 2)(S = 2). \end{aligned}$$

Donde la primera categoría de cada variable explicativa se definió como de referencia. El ajuste para esta regresión logística se muestra en la tabla 3.17. En este caso, aunque existen efectos e interacciones que no son significativas. Al hacer la prueba de significancia del modelo, la diferencia entre la devianza nula y la devianza residual es 1,953. Comparada con el valor en tablas de una ji-cuadrada con 13 grados de libertad (22.36,  $\alpha = 0.05$ ), la hipótesis nula se rechaza, por lo que podemos suponer que el modelo es significativo.

De acuerdo con la tabla 3.17, se observa que para el coeficiente de la variable sexo, la razón de momios es  $\exp(\beta_S) = 0.406$ , al comparar las categorías alta y baja. Es decir, un individuo con escolaridad alta tiene una menor probabilidad de consumir drogas con respecto a un individuo con escolaridad baja. En cuanto los términos de interacción, se puede mencionar la escolaridad y sexo, donde se observa que la razón de momios se

Coeficiente	$\hat{\beta}$	$\sqrt{\hat{\text{var}}(\hat{\beta})}$	$z$	$p\text{-value}$	$\exp(\hat{\beta})$
Intercepto	2.032	0.321	6.33	< 0.001	
$\beta_S(S = 2)$	-1.260	0.599	-2.10	0.035	0.284
$\beta_I(I = 2)$	0.253	0.256	0.99	0.323	1.287
$\beta_T(T = 2)$	-0.901	0.467	-1.93	0.054	0.406
$\beta_D(D = 2)$	-0.562	0.326	-1.72	0.084	0.570
$\beta_E(E = 2)$	0.141	0.075	1.90	0.057	1.151
$\beta_{TS}(T = 2)(S = 2)$	-0.257	0.765	-0.34	0.737	0.773
$\beta_{DS}(D = 2)(S = 2)$	0.183	0.606	0.30	0.762	1.201
$\beta_{ES}(E = 2)(S = 2)$	0.639	0.098	6.52	< 0.001	1.893
$\beta_{IS}(I = 2)(S = 2)$	0.644	0.340	1.90	0.058	1.904
$\beta_{TD}(T = 2)(D = 2)$	-0.345	0.474	-0.73	0.467	0.708
$\beta_{IE}(I = 2)(E = 2)$	0.325	0.274	1.19	0.236	1.384
$\beta_{TDS}(T = 2)(D = 2)(S = 2)$	-0.105	0.775	-0.14	0.891	0.899
$\beta_{IES}(I = 2)(E = 2)(S = 2)$	-0.466	0.365	-1.28	0.202	0.627

Devianza nula: 14,239 con 10,293 grados de libertad  
Devianza residual: 12,286 con 10,280 grados de libertad  
AIC: 12,314

**Tabla 3.17:** Ajuste del modelo logístico para la variable respuesta alcohol ( $A = 1$ ) y variables explicativas ingreso ( $I$ ), tabaco ( $T$ ), drogas ( $D$ ) y escolaridad ( $E$ ).

incrementa en  $\exp(\beta_{ES}) = 1.893$  al comparar la razón de momios de una mujer que tiene escolaridad alta con respecto a un hombre de escolaridad baja.

Además, se observa que aunque el ingreso no es significativo, la interacción de ésta con sexo está en el umbral de serlo. Esto puede ser debido a que el término de interacción desaparece el efecto del ingreso, a este fenómeno se le conoce como la paradoja de Simpson, ver *e.g.* Edwards (2000, pp. 8-11). Al haber muchos coeficientes no significativos, se realizó una selección *stepwise*, eliminándose los términos de interacción  $\beta_{TDS}(tds)$  y  $\beta_{IES}(ies)$ , así como las interacciones  $\beta_{TD}(td)$ ,  $\beta_{DS}(ds)$  y  $\beta_{IE}(ie)$ .

En resumen, para las tres regresiones logísticas basadas en el modelo gráfico de la muestra completa, la mayoría de los términos de interacción son no significativas. Sin embargo, con el propósito de mostrar la relación que existe entre el modelo gráfico y una regresión logística, a continuación se muestran las probabilidades estimadas considerando un modelo gráfico y una regresión logística para cada variable respuesta.

### 3.5.1. Comparación de probabilidades ajustadas

Bajo un modelo de regresión logística, la probabilidad ajustada para una observación  $y_i$  dada las variables explicativas es:

$$\hat{P}(y_i = 1|x_i) = \frac{\exp(\sum_{j=1}^p x_{ij}\hat{\beta}_j)}{1 + \exp(\sum_{j=1}^p x_{ij}\hat{\beta}_j)}, \text{ con } i = 1, \dots, n, \quad (3.10)$$

donde  $x_i = (x_{i1}, \dots, x_{ip})$  es el vector formado por los valores que toman las variables explicativas para la  $i$ -ésima observación. El valor  $\hat{\eta}_i = \sum x_{ij}\hat{\beta}_j$  es igual el predictor lineal ajustado bajo el modelo.

Para las variables tabaco, drogas y alcohol, sus respectivos predictores lineales son:

$$\eta_T = \beta + \beta_S(s) + \beta_D(d) + \beta_A(a) + \beta_{SD}(sd) + \beta_{SA}(sa) + \beta_{DA}(da) + \beta_{SDA}(sda), \quad (3.11)$$

$$\eta_D = \beta + \beta_A(a) + \beta_T(t) + \beta_S(s) + \beta_{AS}(as) + \beta_{AT}(at) + \beta_{ST}(st) + \beta_{AST}(ast), \quad (3.12)$$

$$\begin{aligned} \eta_A = & \beta + \beta_S(s) + \beta_I(i) + \beta_D(d) + \beta_T(t) + \beta_E(e) + \beta_{SI}(is) + \beta_{SE}(se) + \beta_{ST}(st) \\ & + \beta_{SD}(sd) + \beta_{IE}(ie) + \beta_{TD}(td) + \beta_{ISE}(ise) + \beta_{STD}(std). \end{aligned} \quad (3.13)$$

Por otra lado, para el modelo gráfico log-lineal obtenido con la muestra completa, la expresión para la probabilidad estimada de la celda  $(a, d, e, i, s, t)$  es

$$\hat{p}(a, d, e, i, s, t) = \frac{\hat{p}(a, \cdot, e, i, s, \cdot)\hat{p}(a, d, \cdot, \cdot, s, t)}{\hat{p}(a, \cdot, \cdot, \cdot, s, \cdot)}.$$

En este caso, una estimación para la probabilidad condicional de observar la variable tabaco dada las variables alcohol, drogas y sexo se obtiene de la siguiente forma:

$$\begin{aligned} \hat{P}[T = 1|A = a, D = d, S = s] &= \frac{\hat{p}(a, d, \cdot, \cdot, s, 1)}{\sum_t \hat{p}(a, d, \cdot, \cdot, s, t)} \\ &= \frac{n(a, d, \cdot, \cdot, s, 1)}{\sum_t n(a, d, \cdot, \cdot, s, t)}, \end{aligned} \quad (3.14)$$

en donde  $n(t, a, d, s)$  es la estadística suficiente para  $\hat{p}(t, a, d, s)$ .

Análogamente, una estimación para la probabilidad condicional de observar la variable drogas dada las variables alcohol, tabaco y sexo es:

$$\hat{P}[D = 1|A = a, T = t, S = s] = \frac{n(a, 1, \cdot, \cdot, s, t)}{\sum_d n(a, d, \cdot, \cdot, s, t)}. \quad (3.15)$$

Finalmente, una estimación para la probabilidad de observar la variable alcohol dada las demás variables es

$$\begin{aligned} \hat{P}[A = 1|D = d, E = e, I = i, S = s, T = t] &= \frac{\hat{p}(1, d, e, i, s, t)}{\hat{p}(\cdot, d, e, i, s, t)} \\ &= \frac{n(1, \cdot, e, i, s, \cdot)n(1, d, \cdot, \cdot, s, t)/n(1, \cdot, \cdot, \cdot, s, \cdot)}{\sum_a n(a, \cdot, e, i, s, \cdot)n(a, d, \cdot, \cdot, s, t)/n(1, \cdot, \cdot, \cdot, s, \cdot)}. \end{aligned} \quad (3.16)$$

Para comparar contra los valores observados, calculamos sus correspondientes probabilidades condicionales de tabaco, drogas y alcohol, dada sus variables explicativas. Las expresiones para estas probabilidades son:

$$P[T = 1|A = a, D = d, S = s] = \frac{\sum_{e,i} n(1, a, d, s, e, i)}{\sum_{t,e,i} n(t, a, d, s, e, i)}, \quad (3.17)$$

$$P[D = 1|A = a, T = t, S = s] = \frac{\sum_{e,i} n(t, a, 1, s, e, i)}{\sum_{d,e,i} n(t, a, d, s, e, i)}, \quad (3.18)$$

$$P[A = 1|T = t, D = d, S = s, E = e, I = i] = \frac{n(t, 1, d, s, e, i)}{\sum_a n(t, a, d, s, e, i)}. \quad (3.19)$$

En las tablas 3.18, 3.19 y 3.20 se muestran las probabilidades observadas  $p_{obs}$ , y las probabilidades ajustadas con un modelo gráfico y una regresión logística,  $\hat{p}_{graph}$  y  $\hat{p}_{logit}$ .

$D$	$A$	$S$	No. de obs.	$p_{obs}$	$\hat{p}_{graph}$	$\hat{p}_{logit}$
1	1	1	103	0.7203	0.7203	0.7203
1	1	2	22	0.4889	0.4889	0.4889
1	2	1	11	0.5000	0.5000	0.5000
1	2	2	5	0.2381	0.2381	0.2381
2	1	1	1,224	0.4165	0.4165	0.4165
2	1	2	544	0.2359	0.2359	0.2359
2	2	1	232	0.1701	0.1701	0.1701
2	2	2	195	0.0565	0.0565	0.0565

**Tabla 3.18:** Probabilidades ajustadas con variable respuesta tabaco ( $T = 1$ ) y variables explicativas drogas ( $D$ ), alcohol ( $A$ ) y sexo ( $S$ ). Los valores de  $p_{obs}$ ,  $\hat{p}_{graph}$  y  $\hat{p}_{logit}$  se obtuvieron a partir de las fórmulas en (3.17), (3.14) y (3.11), respectivamente.

$A$	$T$	$S$	No. de obs.	$p_{obs}$	$\hat{p}_{graph}$	$\hat{p}_{logit}$
1	1	1	103	0.0776	0.0776	0.0776
1	1	2	22	0.0389	0.0389	0.0389
1	2	1	40	0.0228	0.0228	0.0228
1	2	2	23	0.0129	0.0129	0.0129
2	1	1	11	0.0453	0.0453	0.0453
2	1	2	5	0.0250	0.0250	0.0250
2	2	1	11	0.0096	0.0096	0.0096
2	2	2	16	0.0049	0.0049	0.0049

**Tabla 3.19:** Probabilidades ajustadas con variable respuesta drogas ( $D = 1$ ) y variables explicativas alcohol ( $A$ ), tabaco ( $T$ ) y sexo ( $S$ ). Los valores de  $p_{obs}$ ,  $\hat{p}_{graph}$  y  $\hat{p}_{logit}$  se obtuvieron a partir de las fórmulas en (3.18), (3.15) y (3.12), respectivamente.

$D$	$T$	$I$	$E$	$S$	No. de obs.	$p_{obs}$	$\hat{p}_{graph}$	$\hat{p}_{logit}$
1	1	1	1	1	33	0.9167	0.8868	0.8842
1	1	1	1	2	4	0.8000	0.7015	0.6842
1	1	1	2	1	43	0.8958	0.8975	0.8979
1	1	1	2	2	13	0.7647	0.8414	0.8254
1	1	2	1	1	4	1.0000	0.9158	0.9076
1	1	2	1	2	0	*	0.8561	0.8416
1	1	2	2	1	23	0.8846	0.9413	0.9400
1	1	2	2	2	5	1.0000	0.9186	0.9096
1	2	1	1	1	11	0.7333	0.7526	0.7561
1	2	1	1	2	5	0.5000	0.4343	0.4049
1	2	1	2	1	21	0.7500	0.7728	0.7812
1	2	1	2	2	12	0.6316	0.6341	0.5975
1	2	2	1	1	2	1.0000	0.8086	0.7996
1	2	2	1	2	0	*	0.6603	0.6253
1	2	2	2	1	6	1.0000	0.8616	0.8642
1	2	2	2	2	6	0.6000	0.7866	0.7598
2	1	1	1	1	401	0.8167	0.8153	0.8131
2	1	1	1	2	122	0.5837	0.5984	0.5973
2	1	1	2	1	523	0.8275	0.8315	0.8337
2	1	1	2	2	304	0.7835	0.7708	0.7639
2	1	2	1	1	28	0.8750	0.8597	0.8485
2	1	2	1	2	10	0.7143	0.7905	0.7844
2	1	2	2	1	272	0.9037	0.9003	0.8993
2	1	2	2	2	108	0.8438	0.8773	0.8733
2	2	1	1	1	526	0.5531	0.5590	0.5558
2	2	1	1	2	458	0.2296	0.2241	0.2289
2	2	1	2	1	796	0.5940	0.5863	0.5904
2	2	1	2	2	913	0.3900	0.3946	0.3931
2	2	2	1	1	28	0.5833	0.6377	0.6170
2	2	2	1	2	33	0.4342	0.4223	0.4213
2	2	2	2	1	365	0.7185	0.7217	0.7198
2	2	2	2	2	358	0.5878	0.5809	0.5798

\* No se observaron casos en el denominador.

**Tabla 3.20:** Probabilidades ajustadas con variable respuesta alcohol ( $A = 1$ ) y variables explicativas drogas ( $D$ ), tabaco ( $T$ ), ingreso ( $I$ ), escolaridad ( $E$ ) y sexo ( $S$ ). Los valores de  $p_{obs}$ ,  $\hat{p}_{graph}$  y  $\hat{p}_{logit}$  se obtuvieron a partir de las fórmulas en (3.19), (3.16) y (3.13), respectivamente.

Como se puede observar en las tablas 3.18 y 3.19, las probabilidades estimadas considerando un modelo gráfico y una regresión logística son iguales. Mas aún, coinciden con las probabilidades observadas, lo cual tiene sentido ya que se está ajustando un modelo saturado. Si en el análisis, se hubieran eliminado las interacciones que no son significativas, el ajuste para las regresiones logísticas de tabaco y drogas no hubiera sido perfecto.

Con respecto a la variable alcohol, se puede observar en la tabla 3.20, que aunque las probabilidades estimadas en ambos modelos son muy parecidas, no son idénticas. Asimismo, en ninguno de estos dos modelos se obtuvo un ajuste perfecto, a diferencia de las variables tabaco y drogas.

Una forma de seleccionar alguna de las regresiones logísticas es utilizar el AIC. Para discernir entre un modelo gráfico y una regresión logística no hay criterios para evaluar cuál describe mejor a los datos. Sin embargo, esto no implica que el análisis bajo estos dos tipos de modelos sean excluyentes. Por ejemplo, al ajustar un modelo gráfico, se pueden identificar con mayor facilidad las variables explicativas que deben ser incluidas en la regresión logística. Mas aún, identifica términos de interacción, los cuales en la práctica son difíciles de incluir en un análisis. Además, este tipo de análisis no está restringido a variables categóricas, existen modelos gráficos más generales, como son los mixtos. En estos modelos, el conjunto de vértices están asociados tanto a variables discretas como a continuas. De esta manera, un modelo gráfico puede apoyar el análisis con una regresión logística y viceversa.

También es interesante notar que para cuando se considera a la variable de drogas o la de tabaco, el ajuste es perfecto. Estos modelos tienen como característica común que ambas variables están contenidas en un sólo clique. Lo anterior, tiene como consecuencia que la gráfica inducida por su frontera sea también completa. Esto podría suponer que las probabilidades ajustadas considerando un modelo gráfico y una regresión logística serán iguales y tendrán un ajuste perfecto si la variable respuesta está en un sólo clique. Formalmente, no se conoce ningún resultado que pruebe esta conjetura.

# Conclusiones

El propósito principal de este trabajo es ilustrar con datos de la ENA 2011 el uso de modelos gráficos log-lineales para el análisis de variables categóricas. Para ello, se propusieron como variables de interés el consumo de alcohol, tabaco y drogas, así como su posible asociación con las variables sexo, ingreso y escolaridad.

Inicialmente, para las variables ingreso y escolaridad, se habían propuesto colapsar sus niveles en categorías baja, media y alta; con esta definición, tiene más sentido considerarlas como variables ordinales. Sin embargo, las funciones del paquete `gRim` de `R` que se utilizaron para el ajuste, no tienen la opción de definir variables ordinales, por lo que se decidió dicotomizar a estas dos variables. Para esto, se utilizó la mediana del ingreso del hogar y el grado promedio de escolaridad.

En un primer ajuste, se hizo una selección *stepwise* para obtener un modelo jerárquico log-lineal. En este caso, a pesar de que este modelo fue más consistente con los datos comparado al modelo gráfico loglineal, al ser un modelo con un número considerable de clases generadoras (15), las interacciones resultan difícil de interpretar y sólo se enfocan para las variables dentro de cada clase generadora que sólo incluye tres variables.

Posteriormente, con las 6 variables, se obtuvo un modelo gráfico descomponible. En este caso, al utilizar el teorema de Hammersley y Clifford, los estimadores para la frecuencias en cada celda pueden factorizarse en productos de menor dimensión. De modo que el problema de estimación se redujo 2 dimensiones.

Para identificar las variables que están relacionadas con el consumo de sustancias, utilizamos las 3 propiedades de Markov en la gráfica asociada. Debido a la definición del modelo, no hubo ninguna restricción sobre las observaciones en la tabla de contingencia, esto es, para cualquier celda  $i$ , su probabilidad es estrictamente positiva,  $p(i) > 0$ . De modo que fue posible mostrar la equivalencia entre las 3 propiedades de Markov.

En particular, para mostrar las variables que no están relacionadas directamente con las variables de consumo de sustancias, utilizamos la propiedad de Markov local. De acuerdo con la gráfica del modelo seleccionado, se observó el consumo de tabaco está asociado directamente con las variables de alcohol, sexo y drogas, mientras que con las variables de escolaridad e ingreso, se tiene una relación indirecta. Una consecuencia de esto, es que para estimar la variable tabaco, es suficiente tener información acerca de las variables sexo, drogas y tabaco. En el caso de la variable alcohol, se observa que su conjunto frontera es igual al resto de las variables. De modo que para inferir sobre esta sustancia se requiere información de las demás variables, lo cual no representa una reducción de la información.

Para evaluar el grado de asociación, se obtuvieron las devianzas de inclusión de las

aristas en el modelo. Los valores más grande de esta estadística indican una mayor asociación. En este caso, se observó que para la variable alcohol, ésta tiene una asociación fuerte con las variables tabaco y escolaridad. Adicionalmente, la variable tabaco tiene una asociación fuerte con la variable sexo. De las tres variables correspondientes al consumo de sustancias, la mayor asociación es el consumo de alcohol y tabaco, seguida por el consumo de tabaco y drogas, y finalmente el consumo de alcohol y drogas.

Por otra parte, debido a la baja prevalencia del consumo de drogas, se identificaron celdas vacías en la tabla de contingencia de dimensión 6 y correspondiente a 10,294 observaciones. Para obtener estimaciones menos imprecisas y no ignorar el efecto de las drogas, se hizo un análisis por separado, uno para la muestra que consume drogas y otro donde no hay consumo, el número de observaciones para cada submuestra fue de 231 y 10,063, respectivamente.

Los modelos gráficos log-lineales obtenidos para estas dos subpoblaciones, son descomponibles. Para la población que consume drogas se observó una gráfica con menos aristas comparada con la población que no consume drogas. Esto indica que el consumo de drogas puede ser estimado con mayor facilidad.

En el caso de la población que no consume drogas, la tabla de contingencia para estos datos fue densa, a diferencia de la población que consume drogas, la cual aún presentaba celdas vacías. Sin embargo, en ambos modelos la devianza disminuyó comparada con el modelo obtenido con la muestra completa.

Adicionalmente, al ser las variables de consumo de sustancias binarias, se planteó definir regresiones logísticas. Donde la variable respuesta es la sustancia consumida, y los coeficientes de la regresión estuvieran determinadas por las interacciones (dentro de los cliques) del modelo gráfico seleccionado con la muestra completa.

Al comparar las probabilidades estimadas bajo un modelo gráfico y una regresión logística, se observó que éstas coincidían si la variable respuesta estaba contenida en un sólo clique, como fue el caso de la variable tabaco y drogas, y a diferencia de la variable alcohol.

Con las 6 variables definidas para el ajuste de modelos gráficos, resultó fácil identificar las relaciones de independencia condicional con base en la gráfica. Sin embargo, cuando el número de variables es considerablemente grande, resulta conveniente interpretar las independencias condicionales en términos de los conjuntos de variables formadas por la descomposición de la gráfica.

Finalmente, un aspecto que no se consideró para el ajuste de modelos gráficos fue el diseño de la muestra. Sin embargo, si expresamos el modelo log-lineal como un modelo lineal generalizado, es posible incorporar información del diseño. No obstante, la selección de un modelo gráfico considerando pesos muestrales, y basado en un modelo lineal generalizado, no garantiza que el modelo seleccionado sea gráfico. De modo que se puede plantear como un problema que se puede abordar en un trabajo posterior.

# Apéndice A

## Diseño muestral de la Encuesta Nacional de Adicciones 2011

### Marco muestral

La selección de las unidades muestrales se realizó en 4 etapas, las cuales se definieron de la siguiente manera.

- i) En la primera etapa, el universo de selección para las unidades primarias de muestreo (UPM) lo conforma el agregado de las AGEB's (Área geo-estadística Básica) de las localidades listadas en el Censo 2005 y las localidades del Censo 2010 no listadas en el Censo 2005 (localidades nuevas).
- ii) En la segunda etapa, las unidades secundarias de muestreo (USM) están determinadas por el conjunto de manzanas o segmentos de viviendas definidas en cada AGEB.
- iii) En la tercera etapa, las unidades terciarias de muestreo (UTM) lo conforman los hogares en México.
- iv) Finalmente las unidades últimas de muestreo (UUM) están delimitadas por la población en el rango de edad de 12 a 65 años.

### Unidades de análisis

Como menciona el reporte de la ENA 2011, las unidades de análisis o grupos de dominio son las siguientes:

1. *Hogar*. Conjunto de personas relacionadas por algún parentesco o no, que habitualmente duermen en una vivienda bajo el mismo techo, beneficiándose de un ingreso común, aportado por uno o más de los miembros del hogar.
2. *Adolescentes*. Integrantes de los hogares en el grupo de edad de 12 a 17 años de edad.
3. *Adultos*. Integrantes de los hogares mayores de 17 y menores a 65 años.

## Diseño muestral

El diseño de la muestra para la ENA 2011 es probabilístico con 4 etapas de muestro y estratificado por tipo de entidad federativa y tipo de urbanidad. Para esta última, en la tabla siguiente se muestra la descripción de cada estrato.

Estrato	Definición
Rural	Localidades con menos de 2,500 habitantes.
Urbano	Localidades no incluidas en los estratos rural y metropolitano.
Metropolitano	Capitales de los estados Localidades con más de 100 mil habitantes Localidades con 2,500 o más habitantes de los municipios que formaban las áreas metropolitanas en el año 2000.

**Tabla A.1:** Descripción de los estratos por tipo de urbanidad.

En este caso, las AGEB's del país quedaron clasificadas en 96 estratos. Se definieron como dominios de estudio 8 regiones geográficas, descritas a continuación.

Región	Estados
1. Norcentral	Coahuila, Chihuahua, Durango
2. Noroccidental	Baja California, Baja California Sur, Sonora, Coahuila
3. Nororiental	Nuevo León, Tamaulipas, San Luis Potosí
4. Occidental	Zacatecas, Aguascalientes, Jalisco, Colima y Nayarit
5. Centro	Puebla, Tlaxcala, Morelos, Edo. de México, Hidalgo, Querétaro, Guanajuato
6. Ciudad de México	Distrito Federal
7. Centro Sur	Veracruz, Oaxaca, Guerrero, Michoacán
8. Sur	Yucatán, Quintana Roo, Campeche, Chiapas, Tabasco

**Tabla A.2:** Estados que integran cada una de las 8 regiones.

Las localidades nuevas se agruparon en 8 estratos, definidas por región. La estratificación fue hecha con el fin de obtener estimaciones para dominios de estudio de región y tipo de urbanidad. En la tabla siguiente se desglosa por región el número de unidades de muestreo seleccionadas para cada etapa de selección. A partir de esta, se observa que existe una diferencia de 12 estratos entre la base de datos y el reporte de la encuesta.

Región	Estratos	UPM	Est. 1UPM <sup>3</sup>	UTM <sup>1</sup>	UUM	Factor <sup>2</sup>
Noroccidental	13	56	3	1,812	2,014	6,683,452
Norcentral	11	98	1	2,979	3,206	5,449,864
Nororiental	10	56	1	1,700	1,537	7,471,910
Occidental	16	56	7	1,963	2,086	8,323,237
Cd. de México	17	56	4	1,580	1,433	6,601,452
Centro	20	56	10	1,800	1,873	24,177,124
Centro sur	13	56	1	1,832	1,972	13,263,120
Sur	16	57	5	2,128	1,905	7,651,051
<b>Nacional</b>	<b>116</b>	<b>491</b>	<b>32</b>	<b>15,571</b>	<b>16,249</b>	<b>79,621,211</b>

<sup>1</sup> Viviendas en muestra

<sup>2</sup> Población estimada en los grupos de edad de 12 a 65 años

<sup>3</sup> Número de estratos con una unidad primaria de muestreo

**Tabla A.3:** Estructura del diseño muestral para la Encuesta Nacional de Adicciones 2011.

## Tamaño de muestra

De acuerdo con el reporte de la ENA 2011, se consideró que el tamaño de muestra seleccionada por región debería permitir obtener estimaciones con errores similares a los utilizados en el diseño de la Encuesta Nacional de Adicciones 2008.

El tamaño de muestra para cada una de las 8 regiones se determinó con la siguiente fórmula:

$$n = \frac{z_{\alpha/2}^2(1-p)deff}{r^2pTRh}$$

En donde:

- $n$  Tamaño de muestra de viviendas
- $p$  Proporción a estimar.
- $z_{\alpha/2}$  Cuantil de una distribución normal asociado al nivel de confianza deseado,  $1 - \alpha = 0.90$ .
- $r$  Error relativo máximo de estimación en el 90 % de estimaciones.
- $deff$  Efecto de diseño. El deff asumido fue de 3.27.
- $TR$  Tasa de respuesta esperada.
- $h$  Promedio de personas por vivienda.

Como consecuencia de los supuestos, se determinó un tamaño de muestra de 1,994 viviendas por región. A nivel regional se espera una tasa de respuesta igual a 75 % y una proporción a estimar de 13.7 %. El promedio de personas por vivienda fue de 1.29 y un error relativo de 17 %.

## Esquema de selección

Una vez definidos los estratos, se implementó el siguiente proceso de selección:

- a) El tamaño de muestra por región (aproximadamente 2,000 viviendas) se distribuye de manera proporcional al número de viviendas en los estratos.
- b) El tamaño de muestra por estado se distribuyó de manera proporcional en los estratos urbano, rural y metropolitano del estado y en el estrato de localidades nuevas se asignó un tamaño de muestra de 1 ó 2 localidades dependiendo del porcentaje de la población que vive en las localidades nuevas de la región.

Posterior a la asignación del tamaño de muestra por estrato, las viviendas se seleccionaron como se describe a continuación.

### *Estrato urbano y metropolitano*

1. En la primera etapa se seleccionaron como unidades primarias  $m_j$  AGEB's con probabilidad proporcional al número de viviendas en la AGEB.
2. En la segunda etapa para cada AGEB seleccionada se eligieron 6 manzanas con probabilidad proporcional al número de viviendas en la manzana.
3. Posteriormente, en cada manzana seleccionado se realizó un muestreo sistemático con arranque aleatorio, seleccionándose 6 viviendas. La selección de viviendas se realizó a campo a partir de un croquis y listado de viviendas elaborados por el equipo de campo.

### *Estrato rural*

1. En la primera etapa, se seleccionaron como unidades primarias  $m_j$  AGEB's con probabilidad proporcional al número de viviendas en la AGEB.
2. Posteriormente, se seleccionaron 3 localidades con probabilidad proporcional al número de viviendas en la localidad.
3. En la tercera etapa, dado que no existen mapas de manzanas disponibles para las localidades rurales, se formaron pseudo-manzanas en campo, las cuales estuvieron integradas por aproximadamente 50 viviendas. Luego, se seleccionó una pseudo-manzana con un muestreo sistemático por cada localidad y en ella se generaron conglomerados de aproximadamente 12 viviendas, para finalmente, seleccionar un conglomerado de 12 viviendas con muestreo aleatorio simple.

### *Estrato de localidades de nueva creación*

1. En la primera etapa se seleccionaron como unidades primarias,  $m_j$  localidades con probabilidad proporcional al número de viviendas.
2. En la segunda etapa, dado que no existen mapas de manzanas disponibles para estas localidades, se construyeron en oficina pseudo-manzanas que estuvieron constituidas por aproximadamente 20 viviendas de modo aproximado. Mediante un muestreo

sistemático, se seleccionaron tres pseudo-manzanas por localidad y, por último, se seleccionaron en campo 12 viviendas en cada pseudo-manzana mediante un muestreo sistemático.

Finalmente, una vez seleccionada una vivienda a través de cualquiera de los tres procedimientos descritos con anterioridad, y siempre que la composición del hogar lo permitió, se seleccionó mediante el uso de muestreo aleatorio simple a un adolescente con edad entre 12 a 17 años y a un adulto con edad entre 18 y 65 años.

## **Cuestionarios de recolección de información**

De acuerdo al reporte del ENA 2011, los cuestionarios se diseñaron tomando como base el cuestionario de la ENA 2008 y se programaron para su captura simultánea en una plataforma electrónica que fue instalada en computadoras portátiles para su aplicación a través de entrevistas cara a cara.

### **Cuestionario de hogar**

El informante adecuado para el cuestionario del hogar fue el jefe de hogar, ama de casa u otro miembro de la familia de 18 años o más, sin impedimentos para poder responder y que conociera las características de la vivienda y sus residentes habituales. Los temas del cuestionario de hogar se pueden clasificar en tres rubros principales:

- Datos de identificación y de control. Se registró información referente a la ubicación de cada vivienda seleccionada: entidad, municipio, localidad, AGEB, manzana, número de vivienda, domicilio y el registro de hasta cuatro visitas del entrevistador por informante.
- Características de la vivienda: total de cuartos, disponibilidad de cocina exclusiva y número de focos.
- Características de los miembros del hogar. Se identificó el número de miembros del hogar y para cada uno se registró la siguiente información: nombre (listado de personas), edad, fecha de nacimiento, sexo, parentesco y verificación de residencia.

### **Cuestionario individual**

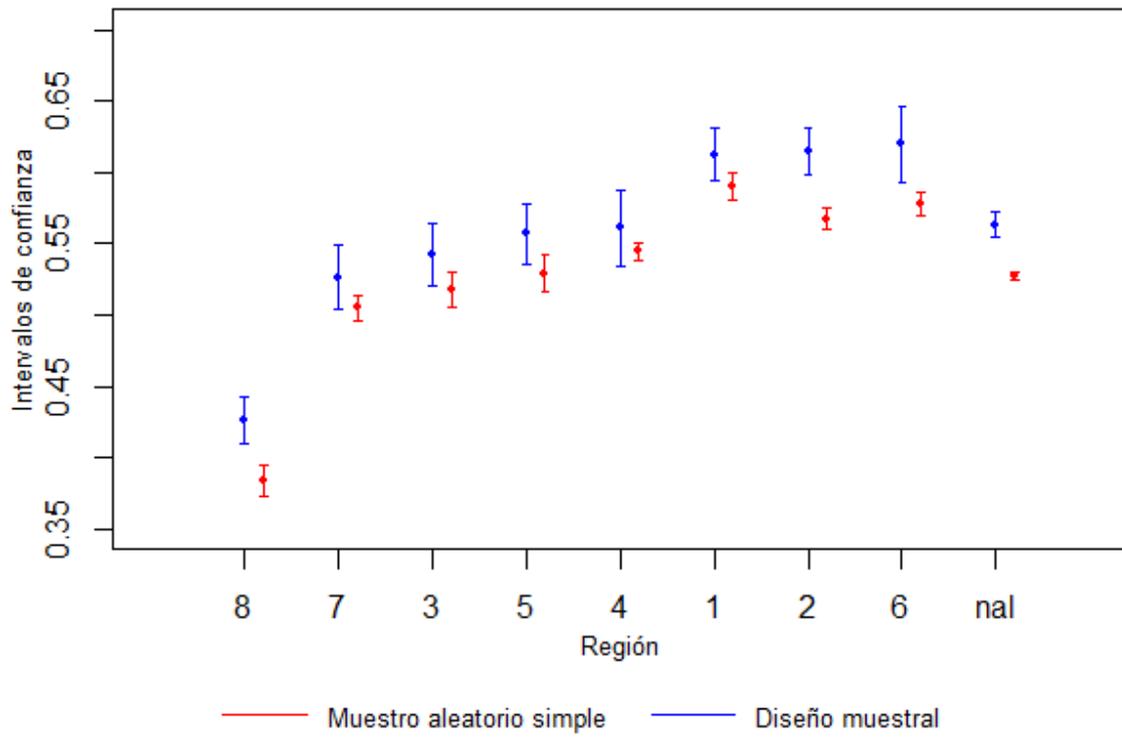
El informante adecuado fue el individuo seleccionado aleatoriamente en los grupos de edad antes mencionados. Se administró un cuestionario individual por informante seleccionado, esto es, una persona entre 12 y 17 años y/o entre 18 y 65 años de edad, cumplidos al momento de la visita. El cuestionario individual consistió de 359 preguntas y recabó información sobre los siguientes temas:

- Datos socio-demográficos de la persona seleccionada: sexo, edad, estado civil, escolaridad, datos sobre hijos, ocupación e ingreso escolar.

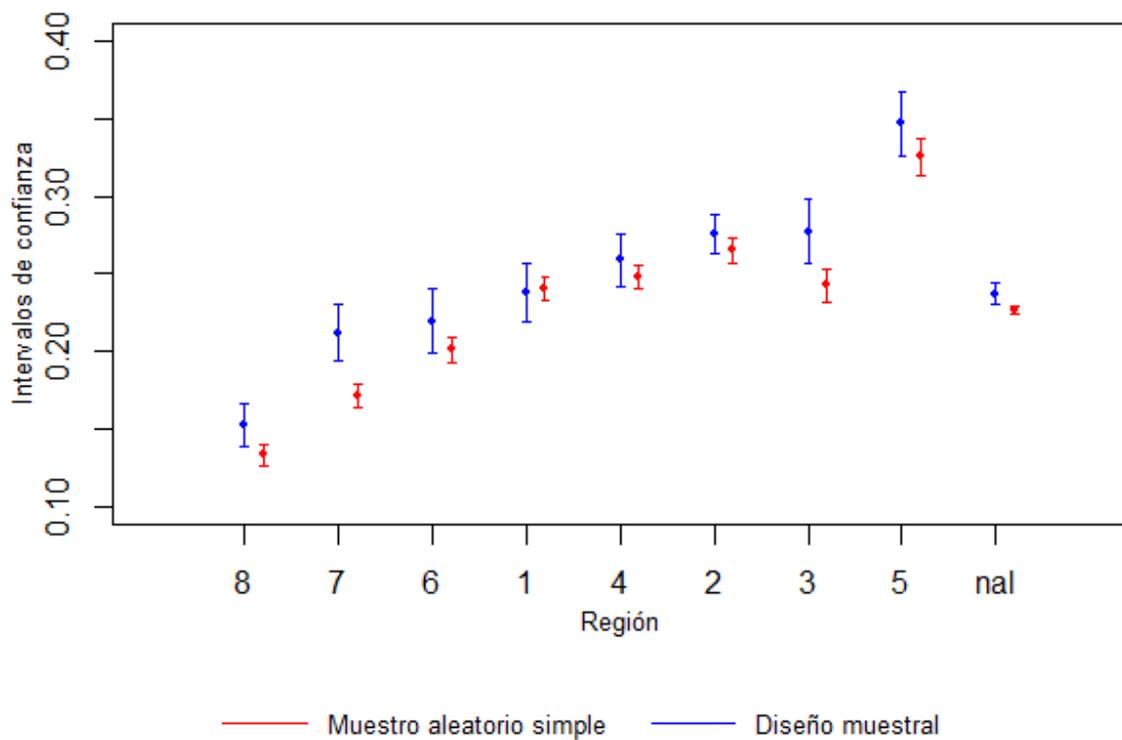
- Consumo de tabaco.
- Tolerancia social.
- Percepción de riesgo.
- Consumo de drogas (médicas e ilegales).
- Dependencia al consumo de drogas.
- Consumo de alcohol.
- Dependencia al consumo de alcohol.
- Consumo de alcohol y drogas en períodos establecidos.
- Tratamiento por consumo de alcohol o drogas.
- Escala de salud mental.
- Embarazo.
- Violencia por parte de la pareja.
- Conducta antisocial, inseguridad y delincuencia.
- Percepción social de la comunidad en relación al uso de drogas.
- Conocimientos sobre el VIH/SIDA.
- Migración hacia los Estados Unidos.

Para propósitos de comparar el efecto del diseño muestral, a continuación se muestra la prevalencia estimada de consumo a nivel nacional y región para las sustancias de alcohol, tabaco, drogas médicas y drogas ilegales.

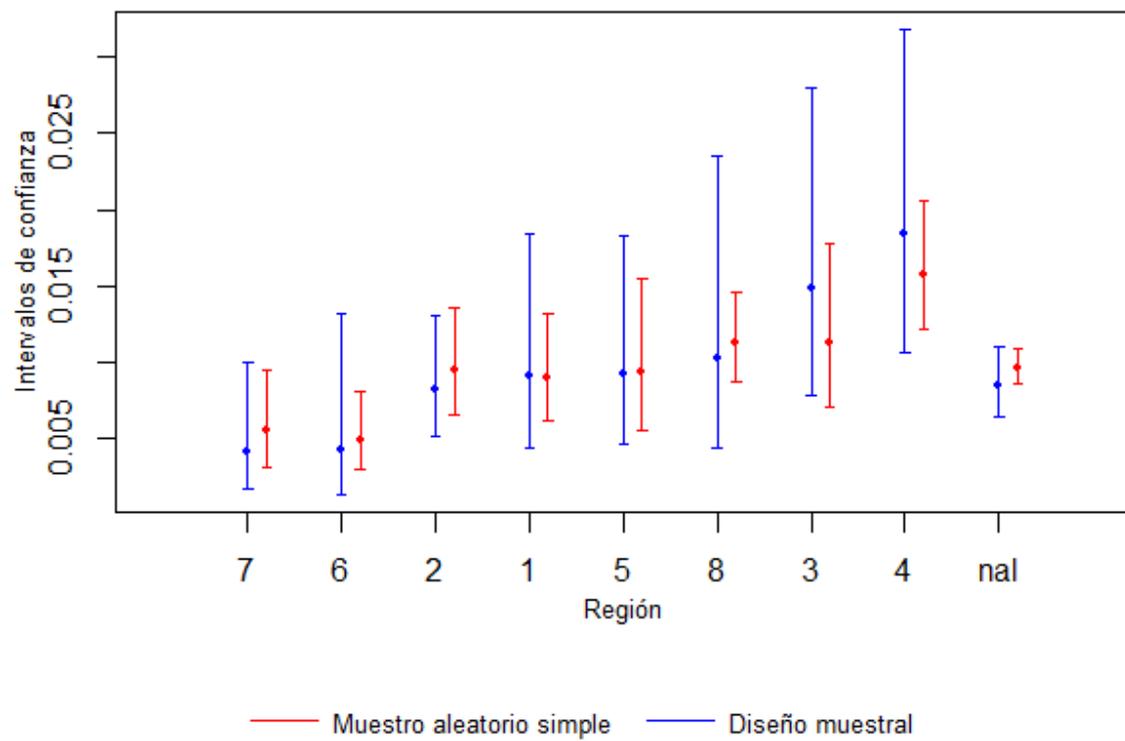
**Figura A.1:** Intervalos al 95 % de confianza para el consumo de alcohol.



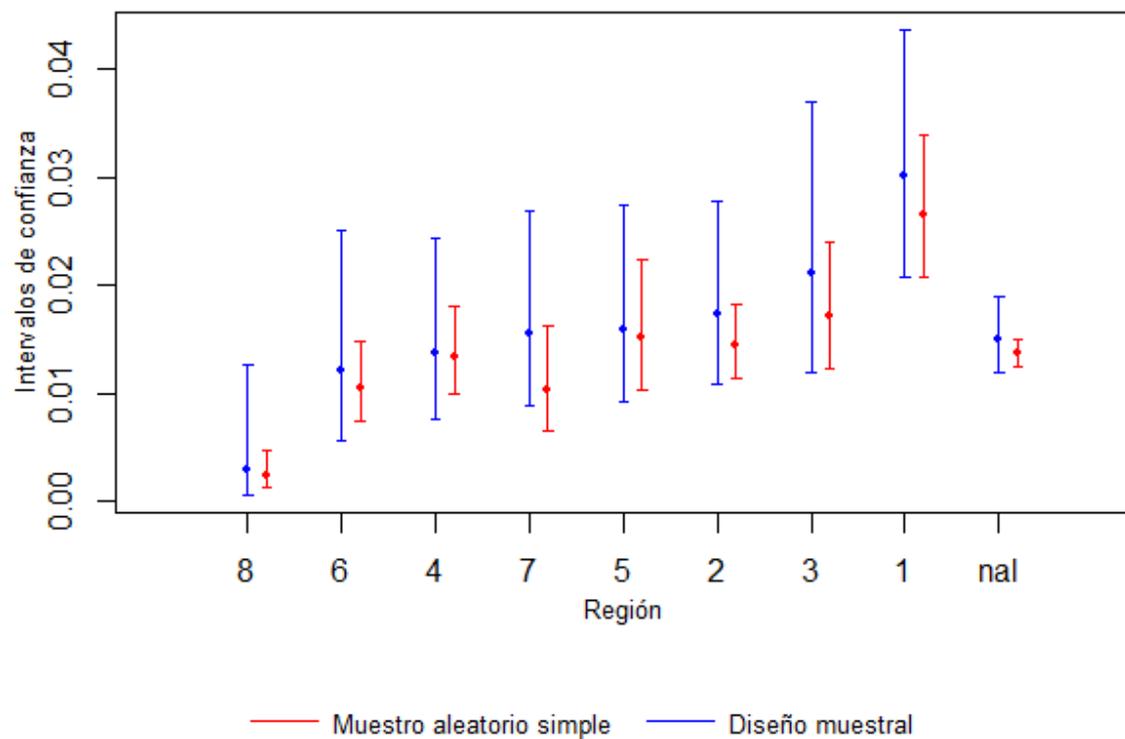
**Figura A.2:** Intervalos al 95 % de confianza para el consumo de tabaco.



**Figura A.3:** Intervalos de confianza al 95% por el método logit para el consumo de drogas médicas.



**Figura A.4:** Intervalos de confianza al 95% por el método logit para el consumo de drogas ilegales.



# Apéndice B

## Código en R

### B.1. Esqueleto del diseño muestral ENA 2011

```
#importar base de datos, archivo en stata
library(foreign)
ena2011<-read.dta("tbl_individual_seleccionados_2012_01_30.dta")

names(ena2011)
summary(ena2011$est_dis)

# est_u estratos de urbanidad
# est_d estratificacion del disenio
# regiong grupos de dominio
# code_upm llave de la unidad primaria de muestreo
# ponde_indiv_final peso muestral

#agregar unidades ultimas de muestro por region
table_UUM=aggregate.data.frame(ena2011$ponde_indiv_final,
as.data.frame(ena2011$regiong), length)
#aproximacion al tamaño de poblacion por region
table_FACT=aggregate.data.frame(ena2011$ponde_indiv_final,
as.data.frame(ena2011$regiong), sum)

labels<-levels(ena2011$regiong)
levels(ena2011$regiong)<-c(1,2,3,4,5,6,7,8)

#crear llave por region
ena2011$EST1<-paste0(ena2011$regiong,"_",ena2011$est_dis)

# numero de personas entrevistadas por estrato
table1<-aggregate.data.frame(ena2011$est_d,as.data.frame(ena2011$EST1),
length)

names(table1)[1]<-"est_d"
table1$est_d<-as.character(table1$est_d)

aux<-matrix(unlist(strsplit(table1$est_d,"_")),ncol=2,byrow=T)
```

```

table1$region<-aux[,1]
table1$estrato<-aux[,2]

# Numero de estratos por region
table_EST<-aggregate.data.frame(table1$region,as.data.frame(table1$
  region),length)

#llave unica de upm
ena2011$upm1<-paste0(ena2011$regiong,"_",ena2011$code_upm)

#personas entrevistadas por upm
table2<-aggregate.data.frame(ena2011$code_upm,as.data.frame(ena2011$
  upm1),length)

table2$Region<-substr(table2[["ena2011$upm1"]],1,1)

table_UPM<-aggregate.data.frame(table2$Region,as.data.frame(table2$
  Region),length)

#numero de upm por estrato
#crear una llave para el estrato y upm

ena2011$est_upm<-paste0(ena2011$est_dis,"_",ena2011$code_upm)

table3<-aggregate.data.frame(ena2011$code_upm,as.data.frame(ena2011$est
  _upm),length)
names(table3)[1]<-"est_upm"

table3$est_upm<-as.character(table3$est_upm)

aux<-matrix(unlist(strsplit(table3$est_upm,"_")),ncol=2,byrow=T)
table3$est<-aux[,1]

# se crea una tabla para el numero de upm por estrato
table_EST_UPM<-aggregate.data.frame(table3$est,data.frame(table3$est),
  length)
# tabla del esqueleto muestral
names(table_EST_UPM)[1]<-"estrato"

table_EST_UPM$upm1<-0
table_EST_UPM$upm1[table_EST_UPM$x==1]<-1
table_EST_UPM<-merge(table_EST_UPM,table1,"estrato")

table_UPM1<-aggregate.data.frame(table_EST_UPM$upm1,
  data.frame(table_EST_UPM$region),sum)

tabular<-table_EST
tabular$UPM<-table_UPM$x
tabular$UPM1<-table_UPM1$x # numero de upm por estrato
tabular$UUM<-table_UUM$x
tabular$Factor<-table_FACT$x

names(tabular)[1:2]<-c("Region","Estratos")

```

```

tabular$Region<-labels
nm<-nrow(tabular)

tabular[nm+1,1]<-"Nacional"
tabular[nm+1,2]<-sum(tabular[1:nm,2])
tabular[nm+1,3]<-sum(tabular[1:nm,3])
tabular[nm+1,4]<-sum(tabular[1:nm,4])
tabular[nm+1,5]<-sum(tabular[1:nm,5])
tabular[nm+1,6]<-sum(tabular[1:nm,6])

table(ena2011$upm_adicional) # variable binaria que indica si la upm
recolectada es una upm adicional

# -----ESTIMACIONES BAJO DISEÑO
# seleccion de las variables de interes

ena2011<-data.frame(ena2011$a002, ena2011$a003, ena2011$a008a, ena2011$a015a,
ena2011$a015b, ena2011$a015c, ena2011$ccp40, ena2011$ccp61,
ena2011$ccp72, ena2011$ccp106, ena2011$code_upm, ena2011$est_u, ena2011$est_dis,
ena2011$regiong, ena2011$ponde_indiv_final)

names(ena2011)<-c("sexo", "edad", "escolaridad", "ingreso_a", "ingreso_b",
"ingreso_c", "tabaco", "drogmed", "drogileg", "alcohol", "upm", "urb", "est_dis", "region", "factor")

# ----- se definieron datos faltantes -----
# definimos los datos faltantes para escolaridad

levels(ena2011$escolaridad)

ena2011$escolaridad[ena2011$escolaridad=="no responde"]<-NA

ena2011$escolaridad<-factor(ena2011$escolaridad, levels=levels(ena2011$escolaridad)[-10])

# se agrego el ingreso por region en una sola columna

ena2011$ingreso_a[ena2011$ingreso_a=="-999"] <-0
ena2011$ingreso_a[ena2011$ingreso_a=="9"] <-0
ena2011$ingreso_b[ena2011$ingreso_b=="-999"] <-0
ena2011$ingreso_b[ena2011$ingreso_b=="9"] <-0
ena2011$ingreso_c[ena2011$ingreso_c=="-999"] <-0
ena2011$ingreso_c[ena2011$ingreso_c=="9"] <-0

ena2011$ingreso2<-rowSums(ena2011[,c(4,5,6)])
ena2011$ingreso2[ena2011$ingreso2==0]<-NA

ena2011$ingreso_a<-NULL
ena2011$ingreso_b<-NULL
ena2011$ingreso_c<-NULL

levels(ena2011$sexo)<-c("hombre", "mujer")

```

```

# se restringe a la pob adulta

pobAdulta<-ena2011[ena2011$edad>17,]
pobAdulta$edad<-NULL

summary(pobAdulta)

levels(pobAdulta$alcohol) <-c(1,0)
levels(pobAdulta$tabaco) <-c(1,0)
levels(pobAdulta$drogmed) <-c(1,0)
levels(pobAdulta$drogileg) <-c(1,0)

pobAdulta$alcohol <-as.numeric(as.character(pobAdulta$alcohol))
pobAdulta$tabaco <-as.numeric(as.character(pobAdulta$tabaco))
pobAdulta$drogmed <-as.numeric(as.character(pobAdulta$drogmed))
pobAdulta$drogileg <-as.numeric(as.character(pobAdulta$drogileg))

# se evalua el efecto de disenio

library(survey)

# aproximacion para el disenio con la primera etapa de seleccion

disenio_ena<-svydesign(id=~upm,strata=~est_dis,weights=~factor,
data=pobAdulta,nest=F,na.rm=T)

# para los estratos con 1 UPM, se estimo su varianza con el
# promedio de los demas estratos
options(survey.lonely.psu="average")

#----- ALCOHOL
# a nivel nacional
mean_alcnal <-svymean(~alcohol, disenio_ena, na.rm=TRUE, deff=TRUE)

# a nivel region
mean_alcreg <-svyby(~alcohol, ~region, disenio_ena, svymean, na.rm=TRUE
, deff=TRUE)

# a nivel region con un disenio SI
mean_alcreg$meanSI <-aggregate(alcohol~region,data=pobAdulta,FUN=mean)
[,2]

# a nivel nacional con un disenio SI
mean(pobAdulta$alcohol)

##### tabla para las estimaciones puntuales e intervalos para alcohol
tabular<-matrix(0,ncol=6,nrow=9)
tabular[-9,2:5]<-as.matrix(mean_alcreg[,2:5])
tabular<-data.frame(tabular)

colnames(tabular)<-c("region","p.dis","se.dis","deff","p.si","se.si")
tabular$region<-c(mean_alcreg$region,"nal")

```

```

tabular[9,2:4]<-c(0.5630569,0.0088456,3.2742)
tabular[9,5]<-mean(pobAdulta$alcohol)

# calculo de la varianza bajo un SI

tabular$se.si<-tabular$se.dis/tabular$def

# ordenar por proporcion bajo disenio
index<-order(tabular$p.dis[-9])
tabular<-tabular[c(index,9),]

# grafica para alcohol
library(plotrix)
par(xaxt = "n",mar=c(5.1,4.1,4.1,2.1)+c(2,0,-4,-1),xpd=T,mgp=c(3,1,0)+c
(-1,0,0))

plotCI(1:9,tabular$p.dis,tabular$se.dis,pt.bg=par("bg"),pch=19,cex=.5,
xlim=c(0,10),ylim=c(0.35,0.70),col="blue",ylab="Intervalos de
confianza",xlab="Region",sfra=.005,cex.lab=.8)
par(xaxt="s")
axis(1,at=1:9,labels=tabular$region)

plotCI(seq(1:9)+.2, tabular$p.si,tabular$se.si,sfra=0.005,pt.bg=par("bg
"),pch=19,cex=.5,add=TRUE,col="red")

legend("bottom",inset=c(0,-.4),legend=c("Muestro aleatorio simple","
Disenio muestral"),
cex=.8,lty=c(1,1),col=c("red","blue"),text.width = 3,horiz=T,
x.intersp=0.5, xjust=0,yjust=0,box.col="white")

#----- TABACO -----
# a nivel nacional
mean_tabnal <-svymean(~tabaco, disenio_ena, na.rm=TRUE, deff=TRUE)

# a nivel region
mean_tabreg <-svyby(~tabaco, ~region, disenio_ena, svymean, na.rm=TRUE,
deff=TRUE)

# a nivel region con un disenio SI
mean_tabreg$meanSI<-aggregate(tabaco~region,data=pobAdulta,FUN=mean)
[,2]

# a nivel nacional con un disenio SI
mean(pobAdulta$tabaco)

##### tabla para las estimaciones puntuales e intervalos para alcohol
tabular<-matrix(0,ncol=6,nrow=9)
tabular[-9,2:5]<-as.matrix(mean_tabreg[,2:5])
tabular<-data.frame(tabular)
colnames(tabular)<-c("region","p.dis","se.dis","deff","p.si","se.si")
tabular$region<-c(mean_tabreg$region,"nal")

mean_tabnal

```

```

tabular[9,2:4]<-c(0.2370497,0.0071618,2.9197)
tabular[9,5]<-mean(pobAdulta$tabaco)

# calculo de la varianza bajo un SI
tabular$se.si<-tabular$se.dis/tabular$deff

# ordenar por proporcion bajo disenio
index<-order(tabular$p.dis[-9])
tabular<-tabular[c(index,9),]

# grafica para tabaco
library(plotrix)
par(xaxt = "n",mar=c(5.1,4.1,4.1,2.1)+c(2,0,-3,-1),xpd=T,mgp=c(3,1,0)+c
(-1,0,0))

plotCI(1:9,tabular$p.dis,tabular$se.dis,pt.bg=par("bg"),pch=19,cex=.5,
xlim=c(0,10),ylim=c(.10,.40),col="blue",ylab="Intervalos de
confianza",xlab="Region",sfra=.005,cex.lab=0.8)
par(xaxt="s")
axis(1,at=1:9,labels=tabular$region)

plotCI(seq(1:9)+.2, tabular$p.si,tabular$se.si,sfra=0.005, pt.bg=par("
bg"),pch=19,cex=.5,add=TRUE,col="red")

legend("bottom",inset=c(0,-.5),legend=c("Muestro aleatorio simple","
Disenio muestral"),cex=.8,lty=c(1,1),col=c("red","blue"),text.width
= 3,horiz=T, box.col="white",x.intersp=0.5, xjust=0,yjust=0)

#----- DROGAS MEDICAS

# Se USO EL METODO LOGIT
# a nivel nacional
mean_drogmednal<-svymean(~drogmed,disenio_ena,deff=T)

# a nivelregion
mean_drogmedreg<-svyby(~drogmed,~region,disenio_ena,svymean, na.rm=TRUE
,deff=TRUE)
mean_drogmedreg$meanSI<-aggregate(drogmed~region,data=pobAdulta,FUN=
mean)[,2]

##### tabla para las estimaciones puntuales e intervalos para drogas
medicas

tabular<-matrix(0,ncol=6,nrow=9)
tabular[-9,2:5]<-as.matrix(mean_drogmedreg[,2:5])

tabular<-data.frame(tabular)
colnames(tabular)<-c("region","p.dis","se.dis","deff","p.si","se.si")
tabular$region<-c(mean_tabreg$region,"nal")

mean_drogmednal
tabular[9,2:4]<-c(0.0084383,0.0011655,2.0134)
tabular[9,5]<-mean(pobAdulta$drogmed)

```

```

# calculo de la varianza bajo un SI
tabular$se.si<-tabular$se.dis/tabular$deff

# ordenar por proporcion bajo disenio
index<-order(tabular$p.dis[-9])
tabular<-tabular[c(index,9),]

# definicion de limite inferior y superior con el metodo logit
y.lower <- log(tabular$p.dis/(1-tabular$p.dis))-
          1.96*(tabular$se.dis/(tabular$p.dis*(1-tabular$p.dis)))
tabular$lim.inf.dis <- exp(y.lower)/(1+exp(y.lower))
y.upper <- log(tabular$p.dis/(1-tabular$p.dis))+
          1.96*(tabular$se.dis/(tabular$p.dis*(1-tabular$
          p.dis)))
tabular$lim.sup.dis <- exp(y.upper)/(1+exp(y.upper))

# limites inferior y superior ignorando disenio
y.lower <- log(tabular$p.si/(1-tabular$p.si))-
          1.96*(tabular$se.si/(tabular$p.si*(1-tabular$p.si)))
tabular$lim.inf.si <- exp(y.lower)/(1+exp(y.lower))
y.upper <- log(tabular$p.si/(1-tabular$p.si))+
          1.96*(tabular$se.si/(tabular$p.si*(1-tabular$p.si)))
tabular$lim.sup.si <- exp(y.upper)/(1+exp(y.upper))

# grafica para drogas medicas
library(plotrix)
par(xaxt = "n",mar=c(5.1,4.1,4.1,2.1)+c(2,0,-3,-1),xpd=T,mgp=c(3,1,0)+c
(-1,0,0))

plotCI(1:9,tabular$p.dis,ui=tabular$lim.sup.dis,li=tabular$lim.inf.dis,
pt.bg=par("bg"),pch=19,cex=.5,xlim=c(0,10),col="blue",ylab="
Intervalos de confianza",xlab="Región",sfra=.005,cex.lab=0.8)
par(xaxt="s")
axis(1,at=1:9,labels=tabular$region)

plotCI(seq(1:9)+.2,tabular$p.si,ui=tabular$lim.sup.si,li=tabular$lim.
inf.si,sfra=0.005,pt.bg=par("bg"),pch=19,cex=.5,add=TRUE,col="red")

legend("bottom",inset=c(0,-.5),legend=c("Muestro aleatorio simple","
Diseño muestral"),cex=.8,lty=c(1,1),col=c("red","blue"),text.width =
3,horiz=T, box.col="white",x.intersp=0.5, xjust=0,yjust=0)

##----- DROGAS ILEGALES

# a nivel nacional
mean_drogilegnal <-svymean(~drogileg,disenio_ena,deff=T)

# a nivelregion
mean_drogilegreg<-svyby(~drogileg,~region,disenio_ena,svymean, na.rm=
TRUE, deff=TRUE)
mean_drogilegreg$meanSI<-aggregate(drogileg~region,data=pobAdulta,FUN=

```

```

mean)[,2]

#### tabla para las estimaciones puntuales e intervalos para drogas
medicas

tabular<-matrix(0,ncol=6,nrow=9)
tabular[-9,2:5]<-as.matrix(mean_drogilegreg[,2:5])

tabular<-data.frame(tabular)
colnames(tabular)<-c("region","p.dis","se.dis","deff","p.si","se.si")
tabular$region<-c(mean_tabreg$region,"nal")

mean_drogilegnal
tabular[9,2:4]<-c(0.0150577,0.0017574,2.5825)
tabular[9,5]<-mean(pobAdulta$drogileg)

# calculo de la varianza bajo un SI
tabular$se.si<-tabular$se.dis/tabular$deff

# ordenar por proporcion bajo disenio
index<-order(tabular$p.dis[-9])
tabular<-tabular[c(index,9),]

# definicion de limite inferior y superior con el metodo logit
y.lower <- log(tabular$p.dis/(1-tabular$p.dis))-
1.96*(tabular$se.dis/(tabular$p.dis*(1-tabular$p.dis)))
tabular$lim.inf.dis <- exp(y.lower)/(1+exp(y.lower))
y.upper <- log(tabular$p.dis/(1-tabular$p.dis))+
1.96*(tabular$se.dis/(tabular$p.dis*(1-tabular$p.dis)))
tabular$lim.sup.dis <- exp(y.upper)/(1+exp(y.upper))

# limites inferior y superior ignorando disenio
y.lower <- log(tabular$p.si/(1-tabular$p.si))-
1.96*(tabular$se.si/(tabular$p.si*(1-tabular$p.si)))
tabular$lim.inf.si <- exp(y.lower)/(1+exp(y.lower))
y.upper <- log(tabular$p.si/(1-tabular$p.si))+
1.96*(tabular$se.si/(tabular$p.si*(1-tabular$p.si)))
tabular$lim.sup.si <- exp(y.upper)/(1+exp(y.upper))

# grafica para drogas medicas
library(plotrix)
par(xaxt = "n",mar=c(5.1,4.1,4.1,2.1)+c(2,0,-3,-1),xpd=T,mgp=c(3,1,0)+c
(-1,0,0))

plotCI(1:9,tabular$p.dis,ui=tabular$lim.sup.dis,li=tabular$lim.inf.dis,
pt.bg=par("bg"),pch=19,cex=.5,xlim=c(0,10),col="blue",ylab="
Intervalos de confianza",xlab="Región",sfra=.005,cex.lab=0.8)
par(xaxt="s")
axis(1,at=1:9,labels=tabular$region)

plotCI(seq(1:9)+.2, tabular$p.si, ui=tabular$lim.sup.si,li=tabular$lim.
inf.si,sfra=0.005,pt.bg=par("bg"),pch=19,cex=.5,add=TRUE,col="red")

```

```
legend("bottom",inset=c(0,-.5),legend=c("Muestro aleatorio simple","
  Diseño muestral"),cex=.8,lty=c(1,1),col=c("red","blue"),text.width =
  3,horiz=T,
box.col="white",x.intersp=0.5, xjust=0,yjust=0)
```

## B.2. Ajuste de modelos gráficos log-lineales

```
#-----SE ANALIZO PREVALENCIA DE ULTIMO ANIO-----

# existe una baja prevalencia de drogas, se agregaron en una variable
# tanto drogas medicas e ilegales

pobAdulta$drogas<-"no"
pobAdulta$drogas[pobAdulta$drogmed=="sí" |pobAdulta$drogileg=="sí"]<-"s
  í"
pobAdulta$drogas<-factor(pobAdulta$drogas,levels=c("sí","no"))

pobAdulta$drogmed <-NULL
pobAdulta$drogileg <-NULL

# para escolaridad se considero la siguiente particion

# nivel 1: Secundaria incompleta
# nivel 2: Al menos secundaria completa

nivel1<-c("primaria incompleta (1 a 5 años)","primaria completa (6 años
  )",
"secundaria incompleta (1 a 2 años)")

nivel2<-c("secundaria completa o equivalente (3 años)",
"bachillerato incompleto (1 a 2 años)",
"bachillerato completo o equivalente (aproximadamente 3 años)",
"estudios universitarios incompletos (1 a 3 años)",
"estudios universitarios completos (4 a 5 años)",
"estudios de posgrado (2 a 4 años)" )

pobAdulta$esc <-1
pobAdulta[pobAdulta$escolaridad %in% nivel2,"esc"] <-2
pobAdulta$escolaridad <-as.factor(pobAdulta$esc)
pobAdulta$esc <-NULL

# se colapso la variable ingreso de la siguiente forma

#nivel 1: <1 salario a 4 salarios

nivel2<-c("5","6","7")

pobAdulta$ing<-1
pobAdulta[pobAdulta$ingreso2 %in% nivel2, "ing"]<-2
pobAdulta$ingreso2<-NULL
```

```

pobAdulta$ingreso<-as.factor(pobAdulta$ing)

pobAdulta$ing      <-NULL
pobAdulta$sexo     <-factor(as.numeric(pobAdulta$sexo))
pobAdulta$tabaco   <-factor(as.numeric(pobAdulta$tabaco))
pobAdulta$alcohol  <-factor(as.numeric(pobAdulta$alcohol))
pobAdulta$drogas   <-factor(as.numeric(pobAdulta$drogas))

### se forma la tabla de contingencia para 6 las seis variables
seleccionadas

tabla<-table(sexo,escolaridad,ingreso,alcohol,tabaco,drogas)

#grafica de mosaico para las variables categoricas
attach(pobAdulta)
library(vcd) # paqueteria para graficos de datos categoricos

# grafica de mosaico
cotabplot(~sexo+ingreso+escolaridad+alcohol+tabaco|drogas,data=tabla)

##----- MODELOS JERARQUICOS

# paquete para el ajuste de modelos graficos
library(gRim)

# paquetes para visualizar graficas
library(gRbase)
library(igraph)
library(grid)
library(Rgraphviz)

# para seis variables se ajustaron modelos jerarquicos
# con orden de interaccion de distinto orden

h<-dmod(~.^1,data=tabla) # efectos principales
h$fitinfo$dev
h<-dmod(~.^2,data=tabla) # primer orden
h$fitinfo$dev
h<-dmod(~.^3,data=tabla) # segundo orden
h$fitinfo$dev
h<-dmod(~.^4,data=tabla) # tercer orden
h$fitinfo$dev
h<-dmod(~.^5,data=tabla)
h$fitinfo$dev

# para un modelo jerarquico se selecciono de un modelo de orden 3 y se
hizo
# seleccion retrospectiva

h.init<-dmod(~.^3,data=tabla) # modelo inicial con interacciones de 2do
orden
h.init$fitinfo$dev

```

```

#seleccion de modelo en direccion backward
hbwd<-stepwise(h.init,type="unrestricted",direction="backward",details
=1)
hbwd
hbwd$glist

#seleccion de modelo con direccion forward
hinit<-dmod(~.^1,data=tabla) # modelo de independencia

hfwd<-stepwise(hinit,type="unrestricted",direction="forward")
hfwd
hfwd$glist

# ----- MODELOS GRAFICOS LOG-LINEALES

# definicion de un modelo jerarquico saturado con 6 variables
saturated<-dmod(~.^.,data=tabla)

# clases generadoras del modelo
saturated$glist

# ----- SELECCION DE MODELO EN DIRECCION BACKWARD

# restringido a modelos graficos descomponibles
bwdmodel<-stepwise(saturated,direction="backward",criterion="test",
type="decomposable")
bwdmodel

# grafica del modelo log-lineal
bwd.p<-as(bwdmodel.test,"igraph")
plot(bwd.p,layout=layout.circle,vertex.size=35)

# clases generadoras del modelo seleccionado
bwdmodel$glist

# seleccion de modelo sin restricciones
bwdmodel.unrest<-stepwise(saturated,direction="backward",criterion="test"
,
type="unrestricted")
bwdmodel.unrest

#----- SELECCION DE MODELO EN DIRECCION FORWARD

# definimos el modelo de independencia

hinit<-dmod(~.^1,data=tabla)

# seleccion de modelo sin restricciones
fwdmodel.unrest<-stepwise(hinit,direction="forward",criterion="test",
type="unrestricted")
fwdmodel.unrest

# seleccion de modelo para un obtener un modelo grafico descomponible

```

```

fwdmodel<-stepwise(hinit,direction="forward",criterion="test",
type="decomposable",details=1)
fwdmodel

#----- GRAFICA DEL MODELO LOG-LINEAL SELECCIONADO
model.graph<-ug(~I*S*A*E+D*S*A*T,result = "igraph")
plot(model.graph,layout=layout.grid,vertex.size=35)

#-----SIGNIFICANCIA DE LAS ARISTAS EN EL MODELO
testdelete(bwdmodel.aic,edge=c("sexo","escolaridad"))
testdelete(bwdmodel.aic,edge=c("sexo","ingreso"))
testdelete(bwdmodel.aic,edge=c("sexo","tabaco"))
testdelete(bwdmodel.aic,edge=c("sexo","drogas"))
testdelete(bwdmodel.aic,edge=c("escolaridad","ingreso"))
testdelete(bwdmodel.aic,edge=c("escolaridad","alcohol"))
testdelete(bwdmodel.aic,edge=c("ingreso","alcohol"))
testdelete(bwdmodel.aic,edge=c("alcohol","tabaco"))
testdelete(bwdmodel.aic,edge=c("alcohol","drogas"))
testdelete(bwdmodel.aic,edge=c("tabaco","drogas"))

# ----- DESCOMPOSICION DE LA GRAFICA

uG<-ug(~I*S*A*E+D*S*A*T)

descomp<-mpd(uG)
descomp

#----- REGRESIONES LOGISTICAS ASOCIADAS AL MODELO GRAFICO LOG-
LINEAL

# graficas inducidas por la frontera de las variables respuesta

#-                ALCOHOL

model.alcohol<-subGraph(c("E","S","I","T","D"),uG)
model.alcohol<-as(model.alcohol,"igraph")
plot(model.alcohol,vertex.size=35)

# regresion logistica con variable respuesta alcohol
alc<-glm(alcohol~ingreso+tabaco+drogas+escolaridad+sexo+sexo*tabaco+
sexo*drogas+sexo*escolaridad+sexo*ingreso+tabaco*drogas+escolaridad*
ingreso
+tabaco*drogas*sexo+sexo*escolaridad*ingreso,data=pobAdulta,family=
binomial)

summary(alc)

## se compara con las probabilidades ajustadas con el modelo grafico

# frecuencias ajustadas por el modelo grafico seleccionada
fitdata <- data.frame(bwdmodel$fitinfo$fit)
# frecuencias observadas
obsdata <- data.frame(tabla)

```

```

# se define la tabla para la variable alcohol
data_alcohol<-aggregate(Freq~sexo+escolaridad+ingreso+tabaco+drogas+
  alcohol,
data=fitdata,FUN=sum)
pfit_alcohol<-aggregate(Freq~sexo+escolaridad+ingreso+tabaco+drogas,
data=data_alcohol,FUN=sum)

pfit_alcohol$pfit<-data_alcohol$Freq[1:32]/pfit_alcohol[,6]

### numero de observaciones marginal de t,d,s,e,i

pfit_alcohol$Freq<-aggregate(Freq~sexo+escolaridad+ingreso+tabaco+
  drogas+alcohol,
data=obsdata,FUN=sum)$Freq[1:32]

### probabilidades ajustadas con la regresion logistica

pfit_alcohol$pfit_logit<-1-predict.glm(alc,newdata=pfit_alcohol[,1:5],
type="response")
pfit_alcohol$p_rel<-pfit_alcohol$Freq/aggregate(Freq~sexo+escolaridad
+ingreso+tabaco+drogas,data=obsdata,FUN= sum)$Freq

pfit_alcohol<-pfit_alcohol[,c(5,4,3,2,1,6,9,7,8)] # reordenando
  columnas

#-          TABACO

#grafica inducida por la frontera de tabaco

model.tabaco<-subGraph(c("D","A","S"),uG)
model.tabaco<-as(model.tabaco,"igraph")
plot(model.tabaco,vertex.size=35)

# regresion logistica
tabaco.glm<-glm(tabaco~sexo+alcohol+drogas+sexo*alcohol+sexo*drogas+
  alcohol*sexo+
alcohol*sexo*drogas,data=pobAdulta,family=binomial)

summary(tabaco)

# compararan con las frecuencias estimadas con el modelo grafico

data_tabaco<-aggregate(Freq~sexo+alcohol+drogas+tabaco,data=fitdata,FUN
  = sum)
pfit_tabaco      <-aggregate(Freq~sexo+alcohol+drogas,data=data_tabaco,
  FUN=sum)

pfit_tabaco$pfit  <-data_tabaco[1:8,5]/pfit_tabaco$Freq
pfit_tabaco$Freq  <-aggregate(Freq~sexo+alcohol+drogas+tabaco,
data=obsdata,FUN=sum)[1:8,5]

# prediccion con modelos logisticos

```

```

pfit_tabaco$pfit_logit<-1-predict.glm(tabaco.glm,newdata=pfit_tabaco
[,1:3],
type="response")
pfit_tabaco$p_rel      <-pfit_tabaco$Freq/aggregate(Freq~sexo+alcohol+
  drogas ,
  data=obsdata ,FUN=sum)$Freq
pfit_tabaco<-pfit_tabaco[,c(3,2,1,4,7,5,6)]

#----- DROGAS

# grafica inducida por la frontera de drogas
model.drugs<-subGraph(c("A","S","T"),uG)
model.drugs<-as(model.drugs,"igraph")
plot(model.drugs ,vertex.size=35)

#regresion logistica con variable respuesta drogas
drugs<-glm(drogas~sexo+tabaco+alcohol+tabaco*alcohol+tabaco*sexo+sexo*
  alcohol+
  sexo*tabaco*alcohol ,data=pobAdulta ,family=binomial)

summary(drugs)

#se compara con las probabilidades ajustadas con el modelo grafico

data_drugs<-aggregate(Freq~sexo+tabaco+alcohol+drogas ,data=fitdata ,FUN=
  sum)
pfit_drugs<-aggregate(Freq~sexo+tabaco+alcohol ,data=fitdata ,FUN=sum)

pfit_drugs$pfit<-data_drugs$Freq[1:8]/pfit_drugs$Freq
pfit_drugs$Freq<-aggregate(Freq~sexo+tabaco+alcohol+drogas ,
  data=obsdata ,FUN=sum)[1:8,5]
pfit_drugs$prel<-pfit_drugs$Freq/aggregate(Freq~sexo+tabaco+alcohol ,
  data=obsdata ,FUN=sum)$Freq

pfit_drugs$plogit<-1-predict.glm(drugs ,newdata=pfit_drugs[,1:3] ,type="
  response")

pfit_drugs<-pfit_drugs[,c(3,2,1,5,6,4,7)]

#imprimir las tablas en LaTeX
library(xtable)

pfit_drugs_table      <-xtable(pfit_drugs)
pfit_alcohol_table    <-xtable(pfit_alcohol)
pfit_tabaco_table     <-xtable(pfit_tabaco)

digits(pfit_drugs_table)  <-4
digits(pfit_alcohol_table) <-4
digits(pfit_tabaco_table) <-4

# datos ajustados y observados para la tabla de contingencia con 6

```

```

variables

obs<-aggregate(Freq~drogas+tabaco+alcohol+escolaridad+ingreso+sexo,
data=obsdata ,FUN=sum)

matrix(obs$Freq,ncol=4,byrow=T)

#-----MODELOS GRAFICOS PARA SUBPOBLACIONES
detach(pobAdulta)

#-----POBLACION QUE CONSUME DROGAS
ena2011drug<-pobAdulta[pobAdulta$drogas==1,]
attach(ena2011drug)

# definicion de la tabla de contigencia con 5 variables
tabledrug<-table(sexo,escolaridad,tabaco,alcohol,ingreso)
detach(ena2011drug)

#----- SELECCION DE MODELO

#          FORWARD
# restringido a modelos descomponibles
modeldrug.null<-dmod(~.^1,data=tabledrug)
fwdmodel.drug<-stepwise(modeldrug.null,direction="forward",type="
decomposable")
fwdmodel.drug$glist

fwdmodel.drug.graph<-ug(~E*I+A*S*T+E*S)
fwdmodel.drug.graph<-as(fwdmodel.drug.graph,"igraph")

plot(fwdmodel.drug.graph,vertex.size=35)

# sin restricciones
fwdmodel.drug.unrest<-stepwise(modeldrug.null,direction="forward",
type="unrestricted")
fwdmodel.drug.unrest
fwdmodel.drug.unrest$glist
fwdmodel.drug$glist

plot(fwdmodel.drug)
fwdmodel.drug.unrest.graph<-ug(~E*I+A*S*T+E*S+A*I)
fwdmodel.drug.unrest.graph<-as(fwdmodel.drug.unrest.graph,"igraph")
plot(fwdmodel.drug.unrest.graph,vertex.size=35)

# ---          SUBPOLACION QUE NO CONSUME DROGAS
ena2011nodrug<-pobAdulta[pobAdulta$drogas==2,]
attach(ena2011nodrug)

tablenodrug<-table(sexo,escolaridad,tabaco,alcohol,ingreso)
detach(ena2011drug)

#----- SELECCION DE MODELO

```

```
### FORWARD
modelnodrug.null<-dmod(~.^1,data=tablenodrug)

# seleccion de modelos graficos descomponibles
fwdmodel.nodrug<-stepwise(modelnodrug.null,direction="forward",
type="decomposable")
fwdmodel.nodrug
plot(fwdmodel.nodrug)
fwdmodel.nodrug$glist

# sin restricciones
fwdmodel.nodrug.unrest<-stepwise(modelnodrug.null,direction="forward",
type="unrestricted")
fwdmodel.nodrug.unrest
plot(fwdmodel.nodrug.unrest)
fwdmodel.nodrug.unrest$glist

#--                BACKWARD

modelnodrug.sat<-dmod(~.^.,data=tablenodrug)

bwdmodel.nodrug<-stepwise(modelnodrug.sat,direction="backward",
type="decomposable")
bwdmodel.nodrug
plot(bwdmodel.nodrug)
bwdmodel.nodrug$glist

# sin restricciones para no consumo de drogas
bwdmodel.nodrug.unrest<-stepwise(modelnodrug.sat,direction="backward",
type="unrestricted")
bwdmodel.nodrug.unrest
plot(bwdmodel.nodrug.unrest)
bwdmodel.nodrug.unrest$glist
```

# Bibliografía

- Agresti, A. (2003). *Categorical Data Analysis*. Wiley, New York, 2a edición.
- Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 3a. edición.
- Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B*, 34:192–236.
- Bishop, Y. M., Fienberg, S. E. y Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*. Springer Verlag, New York.
- Bondy, J. y Murty, U. S. R. (2008). *Graph Theory*. Springer.
- Christensen, R. (1990). *Log-Linear Models and Logistic Regression*. Springer Verlag, New York, 2a edición.
- Clifford, P. (1990). Markov fields in statistics. En Grimmet, G. R., editor, *Disorder in Physical Systems: A Volume in Honor of John M. Hammersley*, pages 19–32. Oxford University Press.
- Darroch, J. N., Lauritzen, S. L. y Speed, T. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, 8:522–539.
- Edwards, D. (2000). *Introduction to Graphical Modelling*. Springer, New York.
- Edwards, D. y Havranek, T. (1985). A fast procedure for model search in multidimensional contingency tables. *Biometrika*, 72:339–351.
- Edwards, D. y Kreiner, S. (1983). The analysis of contingency tables by graphical models. *Biometrika*, 70:553–565.
- Edwards, D., Lauritzen, S. y Højsgaard, S. L. (2012). *Graphical Models with R*. Springer Verlag, New York.
- Højsgaard, S. R. (2015a). R package grbase. <http://cran.r-project.org/web/packages/gRbase/gRbase.pdf>.
- Højsgaard, S. R. (2015b). R package grim. <http://cran.r-project.org/web/packages/gRim/gRim.pdf>.

- Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz (2012). Encuesta nacional de adicciones 2011: Reporte alcohol. [http://www.conadic.salud.gob.mx/pdfs/ENA\\_2011\\_ALCOHOL.pdf](http://www.conadic.salud.gob.mx/pdfs/ENA_2011_ALCOHOL.pdf).
- Instituto Nacional de Salud Pública (2012). Encuesta nacional de adicciones 2011: Bases de datos y documentación. <http://encuestas.insp.mx/ena2011>.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Lauritzen, S. L. (2002). Lectures of contingency tables. Edición electrónica.
- Lauritzen, S. L. (2006). Elements of graphical models. Lectures from the XXVIth International Probability Summer School in Saint Flour.
- Matúš, F. (1992). On equivalence of Markov properties over undirected graphs. *Journal of Applied Probability*, 55:745–749.
- Ramírez-Aldana, R. (2010). *Restricted or Coloured Graphical Log-Linear Models*. PhD thesis, Facultad de Ciencias, UNAM.
- Ramírez-Aldana, R. y Eslava-Gómez, G. (2015). *On the connection between log-linear and logistic models*. Comunicación personal.
- Särndal, C. E., Swensson, B. y Wretman, J. (1991). *Model Assisted Survey Sampling*. Springer, New York.
- Tarjan, R. E. y Yannakakis, M. (1984). Simple linear-time algorithms to test chordality of graph, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal on Computing*, 13:566–579.
- Von-Eye, A. y Mung, E. (2013). *Log Linear Modelling: Concepts, Interpretation and Application*. Wiley.
- Whitakker, J. (1990). *Graphical Models in Applied Statistics*. Wiley, Chichester, England.