

Modelos Log–Lineales

Hasta los años 60 las tablas de contingencia de 2×2 eran analizadas calculando estadísticos tipo χ^2 para testear independencia. Cuando las tablas involucraban más variables se solía repetir este análisis para las subtablas para determinar las interacciones o asociaciones entre las variables. A partir de los 70 con los trabajos de Goodman y la difusión de estos en libros como el de Bishop, Finberg y Holland (1975) y Haberman (1975) hubo un cambio sustancial en el tratamiento de estos problemas, en particular con la inclusión de los modelos log–lineales.

Un modelo loglineal puede ser visto como un caso particular del GLM.

Los modelos log–lineales se usan con frecuencia para analizar la relación de tres o más variables categóricas en una tabla de contingencia.

El objetivo de estos modelos es estudiar asociación, es por ello que no se hace distinción entre variables de respuesta y covariables. Cuando interesa estudiar algunas variables como independientes y otras como dependientes es más adecuado un modelo lineal generalizado. Lo mismo es cierto si las variables estudiadas son continuas y no se pueden discretizar apropiadamente.

La estrategia básica consiste en ajustar un modelo a las frecuencias observadas en la tabla cruzada. Los modelos contienen componentes que reflejan las distintas asociaciones entre las variables. Los modelos son representados por las frecuencias esperadas. Los patrones de asociación entre las variables pueden describirse en términos de los odds y los odds ratios.

Por ahora supondremos que observamos $Y = (Y_{11}, \dots)$ un vector multinomial con $n = y_{++}$ y $\mu = n\boldsymbol{\pi}$.

Modelos para tablas bidimensionales

Comenzamos por considerar el caso más sencillo de tablas de contingencia. Luego, los conceptos que aquí veremos se extienden a tablas más complejas. Podríamos tener un tabla como en el ejemplo que vimos en nuestras primeras clases:

S:Sexo	C: Cree en la vida postmortem		Total
	Si	No	
Mujer	435	147	582
Hombre	375	134	509
Total	810	281	1091

En general, tendremos:

A	B		Total
	Si	No	Total
Si	n_{11}	n_{12}	n_{1+}
No	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	$n_{++} = n$

Modelo de Independencia

Bajo el modelo de independencia tenemos que $\pi_{ij} = \pi_{i+} \pi_{+j} \quad \forall i \forall j$, por lo tanto si μ_{ij} denota la esperanza de la casilla (i, j) , entonces

$$\begin{aligned} \log \mu_{ij} &= \log n + \log \pi_{i+} + \log \pi_{+j} \\ \log \mu_{ij} &= \lambda + \lambda_i^A + \lambda_j^B \end{aligned}$$

donde λ_i^A y λ_j^B representan el efecto de la fila i y de la columna j , respectiva-

mente.

La interpretación de los parámetros es más sencilla para respuestas binarias. Por ejemplo, en una tabla de $I \times 2$, donde las columnas corresponden a la respuesta dicotómica Y , para cada fila i el logit para la probabilidad π_i de que $Y = 1$ es

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \log\left(\frac{\mu_{i1}}{\mu_{i2}}\right)$$

Bajo independencia

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \lambda_1^B - \lambda_2^B$$

\implies no depende de i , es decir no depende de la fila.
Esto corresponde al caso en que

$$\text{logit}(\pi_i) = \alpha$$

por lo tanto la chance de clasificar en una columna particular es constante a lo largo de las filas.

Identificabilidad y Restricciones sobre los parámetros

En una tabla de 2×2 , por ejemplo, el modelo independiente especifica 5 parámetros, por lo tanto está sobreespecificado.

La siguiente tabla muestra tres conjuntos de parámetros diferentes para los datos de *creencia* que dan los mismos valores estimados para las frecuencias esperadas.

Como en el caso lineal, podemos imponer restricciones a los parámetros de manera de obtener unicidad, por ejemplo, pidiendo que para el primer nivel de cada factor el parámetro sea 0 o bien pidiendo que la suma de los parámetros dentro de un factor sea 0, esto sería en nuestro ejemplo de 2×2

$$\lambda_1^A + \lambda_2^A = 0 \quad \lambda_1^B + \lambda_2^B = 0$$

En general, trabajaremos con esta restricción. De todos modos, lo que todos

Frecuencias Observadas	Frecuencias		Log Frecuencias	
	Observadas	Ajustadas	Conjunto 1	Conjunto 2
435	147	432.1	149.9	6.069
375	134	377.9	131.1	5.935
Parámetro				
λ			Conjunto 1	Conjunto 2
			4.876	6.6069
λ_1^A			0.134	0
λ_2^A			0	-0.134
λ_1^B			1.059	0
λ_2^B			0	-1.059
			Conjunto 3	Conjunto 3
				5.472
				0.067
				-0.067
				0.529
				-0.529

cumplirán es que la diferencia entre dos efectos principales es la misma. En nuestro ejemplo, tenemos que $\lambda_1^B - \lambda_2^B = 1,059$ para los tres conjuntos de parámetros.

Modelo Saturado

Cuando las variables son dependientes satisfacen un modelo más complejo

$$\log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

donde los parámetros λ_{ij}^{AB} reflejan la asociación entre A y B . Este modelo describe perfectamente cualquier conjunto de frecuencias y es el modelo más general para una tabla de contingencia bivariada. El caso de independencia corresponde a $\lambda_{ij}^{AB} = 0$

Existe una relación directa entre los *log odds ratios* y los parámetros de asociación λ_{ij}^{AB}

$$\begin{aligned} \log \theta &= \log \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} \right) = \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21} \\ &= (\lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB}) + (\lambda + \lambda_2^A + \lambda_2^B + \lambda_{22}^{AB}) \\ &\quad - (\lambda + \lambda_1^A + \lambda_2^B + \lambda_{12}^{AB}) + (\lambda + \lambda_2^A + \lambda_1^B + \lambda_{21}^{AB}) \\ &= \lambda_{11}^{AB} + \lambda_{22}^{AB} - \lambda_{12}^{AB} - \lambda_{21}^{AB} \\ &= 4\lambda_{11}^{AB}, \end{aligned}$$

donde la última igualdad resulta de las condiciones impuestas:

$$\sum_{i=1}^J \lambda_{ij}^{AB} = 0 \quad \forall j, \quad \sum_{j=1}^J \lambda_{ij}^{AB} = 0 \quad \forall i$$

Los λ_{ij}^{AB} determinan los log odds ratios. Cuando $\lambda_{ij}^{AB} = 0$ los odds ratios valen 1 y A e B son independientes.

En la tabla de *creencia* el odd ratio es

$$\theta = \frac{435 \times 134}{147 \times 375} = 1.057$$

y $\log \theta = 0.056$, por lo tanto

$$\lambda_{11}^{AB} + \lambda_{22}^{AB} - \lambda_{12}^{AB} - \lambda_{21}^{AB} = 0.056 \quad (*)$$

Los parámetros de asociación se pueden ajustar de manera que el primero de cada fila y el primero de cada columna sea 0 o que la suma sobre cada fila y la suma sobre cada columna sea 0.

Cualquiera de estas combinaciones satisfará (*).

El modelo saturado tiene $IJ = 1 + (I - 1) + (J - 1) + (I - 1)(J - 1)$ parámetros no redundantes, es decir tiene tantos parámetros como observaciones, dando un ajuste perfecto.

En la práctica se trata de usar modelos no saturados en tanto su ajuste *suaviza* a los datos y dan origen a interpretaciones más simples.

Los modelos log-lineales que hemos visto son **modelos jerárquicos**. Decimos que un modelo es **jerárquico** cuando incluye todos los términos de orden menor que están presentes en un término de orden mayor. Así, si el modelo contiene λ_{ij}^{AB} , entonces también están presentes en el modelo λ_i^A y λ_j^B . Estos son los modelos más frecuentes.

Como en ANOVA cuando hay interacciones, debemos ser cuidadosos en interpretar los efectos principales cuando hay términos de orden mayor. En general, para cada variable la atención se restringe a los términos de orden mayor.

Modelos para tablas tridimensionales

Los diferentes modelos que veremos representan distintos patrones de independencia y asociación.

Supongamos que tenemos tres variables categóricas A , B y C que tienen valores posibles:

$A: 1, 2, \dots, I$

$B: 1, 2, \dots, J$

$C: 1, 2, \dots, K$

Para desplegar los casos observados deberemos combinar tablas bidimensionales, como la que sigue.

Asumiremos que en una tabla como ésta, un individuo puede clasificar con una probabilidad π_{ijk} en la casilla ijk . Si las n unidades experimentales son independientes, entonces el vector $(n_{111}, \dots, n_{ijk}, \dots, n_{IJK})$ tiene distribución

Raza de la Víctima	Raza del Defendido	Pena de Muerte	Porcentaje de Si
		Si	No
Blanca	Blanca	53	414
	Negra	11	37
Negra	Blanca	0	16
	Negra	4	139
Total	Blanca	53	430
	Negra	15	176
			11.0
			7.9
			Tabla Parcial
			Tabla Parcial
			Tabla Marginal

Cuadro 1: Pena de Muerte por raza del defenido y raza de la víctima

multinomial de parámetros n y $\boldsymbol{\pi} = (\pi_{111}, \dots, \pi_{ijk}, \dots, \pi_{1JK})'$. La única restricción que se impone al vector de probabilidades $\boldsymbol{\pi}$ es que sume 1 y el EMV será el vector de componentes

$$\hat{\pi}_{ijk} = \frac{n_{ijk}}{n}$$

Calculemos el odds ratio en las tablas parciales. Cuando la víctima es blanca

tenemos

$$\theta_B = \frac{53 \times 37}{11 \times 414} = 0.4306$$

por otro lado si la raza de la víctima es negra

$$\theta_N = \frac{0 \times 139}{4 \times 16} = 0$$

Sin embargo, si consideramos la tabla colapsada obtenemos

$$\theta_{Total} = \frac{53 \times 176}{15 \times 430} = 1.44$$

es decir que la conclusión a partir de este valor sería la opuesta que a partir de los odds ratios de las parciales. Este cambio de dirección en la asociación de dos variables al considerar una tercera se conoce como **paradoja de Simpson**. Por esta razón debemos tener mucho cuidado antes de colapsar una tabla, tratando de entender cual es la asociación entre las variables en primera instancia y a partir de ella decidir si es razonable colapsar o no.

Otro Ejemplo

Supongamos que tenemos los siguientes datos de sobrevivencia de pacientes sometidos a cirugía en dos hospitales, A y B donde vivos significa que el paciente sobrevivió 6 semanas a la cirugía.

Hospital	Muertos	Vivos
A	63	2037
B	16	784

Cuadro 2: Sobrevivencia a una cirugía según hospital

A partir de esta tabla obtenemos que

$$\theta = \frac{63 \times 784}{16 \times 2037} = 1.515464$$

con lo que pareceríamos conveniente el hospital B. Sin embargo, si tenemos en cuenta una tercera variable **C: Estado inicial del paciente** la información sería

Hospital	Malas condiciones		Buenas Condiciones	
	Muertos	Vivos	Muertos	Vivos
A	6	594	57	1443
B	8	692	8	92

Cuadro 3: Sobrevida a una cirugía según hospital y estado inicial

Si analizamos la información teniendo en cuenta el estado del paciente al ser intervenido vemos que

$$\theta_{MC} = \frac{6 \times 692}{8 \times 594} = 0.8737374 \quad \theta_{BC} = \frac{57 \times 92}{8 \times 1443} = 0.454262$$

Es decir, el hospital A es siempre preferible, pero es sensiblemente mejor en caso en que el paciente esté en malas condiciones iniciales. El hospital A tiene mayor porcentaje de muertos en general, pero menor porcentaje de muertos al considerar los grupos de buenas/malas condiciones. Estamos otra vez ante la paradoja de Simpson.

Modelos de dependencia

Mutua o Completa Independencia

El modelo más simple es aquel en que

$$P(A = i, B = j, C = k) = P(A = i).P(B = j).P(C = k) \quad \forall i, j, k$$

de manera que si

$$\begin{aligned} \alpha_i &= P(A = i) & i &= 1, \dots, I \\ \beta_j &= P(B = j) & j &= 1, \dots, J \\ \delta_k &= P(C = k) & k &= 1, \dots, K \end{aligned}$$

$$\pi_{ijk} = \alpha_i \beta_j \delta_k$$

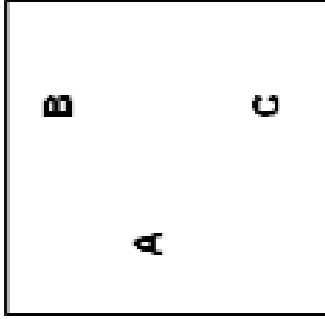
Como la suma de los α 's, de los β 's y de los δ 's es 1 tenemos en total $(I - 1) + (J - 1) + (K - 1)$ parámetros a estimar.

Además, bajo este modelo los vectores marginales tienen distribución:

$$\begin{aligned}(n_{1++}, \dots, n_{I++}) &\sim M(n, \alpha_1, \alpha_2, \dots, \alpha_I) \\ (n_{+1+}, \dots, n_{+J+}) &\sim M(n, \beta_1, \beta_2, \dots, \beta_J) \\ (n_{++1}, \dots, n_{++K}) &\sim M(n, \delta_1, \delta_2, \dots, \delta_K)\end{aligned}$$

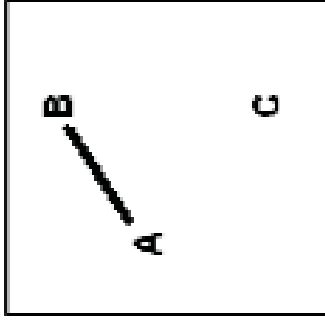
por lo tanto, cada vector de parámetros podría estimarse por separado. Más aún, el EMV será

$$\begin{aligned}\hat{\alpha}_i &= \frac{n_{i++}}{n} \\ \hat{\beta}_j &= \frac{n_{+j+}}{n} \\ \hat{\delta}_k &= \frac{n_{++k}}{n}\end{aligned}$$



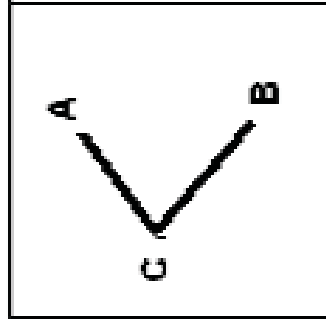
(1)

Mutua Independencia



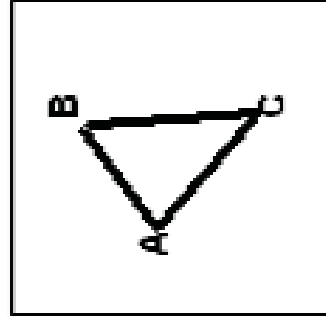
(2)

Dos independientes de una tercera



(3)

Independencia Condicional



(4)

Asociacion Homogenea

Gráficamente, este modelo se representa como en el plot (1) del gráfico que se presenta más arriba.

En este gráfico no hay conexión entre los tres nodos, lo que indica que no hay relación entre las tres variables. En la notación de modelos log–lineales jerárquicos este modelo se representa como (A, B, C) .

En términos de los odds ratios en este modelo significa las tablas marginales $A \times B$, $A \times C$ y $B \times C$ tienen odds ratios iguales a 1.

El modelo log–lineal correspondiente es

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C \quad (1)$$

Independencia conjunta

En el gráfico tenemos a A y B conectadas entre sí, pero no conectadas con C . Esto indica que C es conjuntamente independientes de A y B . Que los nodos A y B estén conectados indica que están posiblemente relacionados, pero no

necesariamente. De manera que el modelo de **mutua independencia** es un caso particular de este modelo que indicaremos como (AB, C) . Bajo este modelo tenemos que

$$\pi_{ijk} = \pi_{ij+} \pi_{++k} \quad \forall i, j, k$$

Si este modelo se cumple, A y C son independientes en la tabla marginal y B y C también son independientes en la tabla marginal. También podemos escribirlo como

$$\pi_{ijk} = \theta_{ij} \delta_k \quad \forall i, j, k$$

donde

$$\sum_i \sum_j \theta_{ij} = 1 \quad \sum_k \delta_k = 1$$

El número de parámetros es: $(IJ - 1) + (K - 1)$.

El EMV de estas probabilidades son: $\hat{\theta}_{ij} = \frac{n_{ij+}}{n}$ y $\hat{\delta}_k = \frac{n_{++k}}{n}$ y por lo tanto los valores esperados son

$$\hat{\mu}_{ijk} = \frac{n_{ij+}n_{++k}}{n}$$

Esto corresponde al concepto de independencia habitual entre la variable C y una nueva variable formada por la IJ combinaciones de A y B .

El modelo log–lineal jerárquico correspondiente resulta

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} \quad (2)$$

Independencia Condicional

Ahora consideremos la relación entre A y B controlando por C . Si A y B son independientes en la tabla *parcial* correspondiente al nivel k de C , decimos

que A y B son condicionalmente independientes en el nivel k de C .

Notemos $\pi_{ij|k} = \frac{\pi_{ijk}}{\pi_{++k}}$ la distribución conjunta de A y B en el nivel k de C . Luego, la independencia condicional de A y B al nivel k de C equivale a

$$\pi_{ij|k} = \pi_{i+|k}\pi_{+j|k} \quad \forall i, j$$

Por lo tanto, diremos que A y B son condicionalmente independientes dado C si la condición anterior vale para todo k .

Entonces, tenemos que

$$\pi_{ijk} = \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}} \quad \forall i, j, k$$

Este modelo de dependencia corresponde al gráfico en el que A y C están conectados y también lo están B y C .

El modelo de **mutua independencia** es un caso particular de este modelo. Independencia condicional de A y B corresponde al modelo log–lineal:

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (3)$$

En la nomenclatura de los modelos log–lineales este modelo se llama (AC, BC).

Independencia marginal vs. Independencia Condicional

Consideremos el siguiente ejemplo en que se registraron las variables Sexo, Ingreso y Carrera

	Ingreso		
Carrera	Sexo	Bajo	Alto
Social	Mujer	18	12
	Hombre	12	8
Ciencias	Mujer	2	8
	Hombre	8	32
Total	Mujer	20	20
	Hombre	20	40

Tenemos que $\theta_{Social} = \frac{18 \times 8}{12 \times 12} = 1$ y $\theta_{Ciencias} = \frac{2 \times 32}{8 \times 8} = 1$, es decir que hay independencia en cada nivel de carrera, sin embargo en la tabla marginal $\theta = \frac{20 \times 40}{20 \times 20} = 2$ y por lo tanto no hay independencia marginal.

Por otro lado, en los odds de Ciencias son 6 veces más grandes en Hombres que en Mujeres dado Ingreso y los odds condicionales de Ingreso Alto son 6 veces más altos en Ciencias que en Sociales dado Sexo. Ciencias tiene relativamente más hombres y Ciencias tiene relativamente ingresos más altos.

La *independencia condicional* y la *independencia marginal* se verifican simultáneamente cuando alguna independencia más fuerte es válida.

Tenemos la siguiente relación:

Mutua Independencia entre A , B y C



B independiente conjuntamente de A y C



A y B condicionalmente independientes A y B marginalmente independientes

Cuando tenemos tres factores podemos tener tres, dos o un par de variables condicionalmente independientes de acuerdo a que tengamos el modelo (1), (2) o (3).

Asociación Homogénea

En efecto, los términos de la forma λ_{ts}^{XY} identifican a las variables condicionalmente dependientes.

Para permitir que las tres pares de variables sean condicionalmente dependientes debemos agregar al modelo anterior (AC, BC) una conexión entre A y B :

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} \quad (4)$$

que corresponde al modelo (AB, AC, BC), conocido como modelo de

asociación homogénea, pues en este modelo los odds ratios condicionales entre dos variables son idénticos para cada nivel de la tercera variable.

Dada la tabla parcial $A-B$ para cada nivel k de C podemos describir la asociación parcial mediante los odds ratios condicionales como

$$\theta_{ij(k)} = \frac{\pi_{ijk}\pi_{i+1,j+1,k}}{\pi_{i,j+1,k}\pi_{i+1,j,k}} \quad 1 \leq i \leq I-1 \quad 1 \leq j \leq J-1$$

Probaremos que si el modelo (4) vale, entonces

$$\log \theta_{ij(k)} = \lambda_{ij}^{AB} + \lambda_{i+1,j+1}^{AB} - \lambda_{i+1,j}^{AB} - \lambda_{i,j+1}^{AB}$$

es decir que

$$\theta_{ij(1)} = \theta_{ij(2)} = \dots = \theta_{ij(k)} \quad \forall i, j$$

Lo mismo es cierto para $\theta_{ik(j)}^{AC}$ y para $\theta_{jk(i)}^{BC}$.

Luego, la asociación entre dos variables es idéntica para cada nivel de la tercera variable.

Modelo saturado o Con Interacción Triple

El modelo más general para tres variables es

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad (5)$$

En este caso las tres variables son condicionalmente dependientes, pero además los odds ratios de cualquier par de variables puede variar a lo largo de los niveles de la tercera. Identificamos este modelo como (ABC) .

Condiciones para asociación marginal y parcial idénticas

El siguiente teorema establece condiciones para que los odds ratios entre A y B sean los mismos en la tabla marginal que en las tablas parciales.

Cuando esto es así podemos estudiar la asociación entre A y B de manera más sencilla colapsando la tabla sobre C .

Teorema: En una tabla tridimensional una variable es colapsable con respecto a la interacción entre las otras dos variables si y sólo si es al menos condicionalmente independiente de otra dada la tercera.

Bishop, Fienberg y Holland (1975)

En otra palabras, A y B tienen la misma asociación marginal que parcial si A y C son condicionalmente independientes (vale el modelo (AB, BC)) o si B y C son condicionalmente independientes (vale el modelo (AB, AC))

Cuando colapsar en una tabla

Dicho de otra manera el resultado que hemos visto nos dice que si tenemos tres variables, A , B y C podemos colapsar en C si se cumplen las dos condiciones siguientes:

1. No hay interacción ABC , es decir $\lambda_{ijk}^{ABC} = 0$ para todo i, j, k .
2. La interacción AC o BC es nula, es decir $\lambda_{ik}^{AC} = 0$ para todo i, k o $\lambda_{jk}^{BC} = 0$ para todo j, k .

Sin embargo, del Teorema enunciado más arriba (probado por Bishop, Fienberg y Holland (1975, pp 39, 47)) sólo probaremos la suficiencia. Whittemore (1978) mostró contraejemplos del sólo si, es decir que no son necesarias. De hecho mostró los siguientes ejemplos:

TABLE 4

F_3	1		2		3		Sum over F_3
F_2	1	2	1	2	1	2	1
F_1	4	2	2	1	1	4	7
1	2	1	4	2	1	4	7

$\mu_{123}^{(3)} = \mu_{12}^{(2)} = \mu_{12,s}^{(2)} = 0; \mu_{23}^{(2)} \neq 0, \mu_{13}^{(2)} \neq 0.$

TABLE 5

F_3	1		2		3		Sum over F_3
F_2	1	2	1	2	1	2	1
F_1	75	24	25	8	20	16	120
1	20	16	60	48	16	32	96

$\mu_{123}^{(3)} = 0; \mu_{12}^{(2)} = \mu_{12,s}^{(2)} \neq 0; \mu_{13}^{(2)} \neq 0, \mu_{23}^{(2)} \neq 0.$

Estadísticos suficientes

Supongamos que n_{ijk} tienen distribución de Poisson con media μ_{ijk} (lo mismo se obtendría bajo la distribución multinomial). La distribución conjunta será:

$$\prod_i \prod_j \prod_k \frac{e^{-\mu_{ijk}} \mu_{ijk}^{n_{ijk}}}{n_{ijk}!}$$

Por lo tanto

$$\ell(\boldsymbol{\mu}) = \sum_i \sum_j \sum_k n_{ijk} \log \mu_{ijk} - \sum_i \sum_j \sum_k \mu_{ijk}$$

Si consideramos el modelo saturado

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} + \lambda_{ik}^{AC} + \lambda_{ijk}^{ABC}$$

queda

$$\begin{aligned} \ell(\boldsymbol{\mu}) = & n\lambda + \sum_i n_{i++} \lambda_i^A + \sum_j n_{+j+} \lambda_j^B + \sum_k n_{++k} \lambda_k^C + \sum_i \sum_j n_{ij+} \lambda_{ij}^{AB} \\ & + \sum_i \sum_k n_{i+k} \lambda_{ik}^{AC} + \sum_j \sum_k n_{+jk} \lambda_{jk}^{BC} + \sum_i \sum_j \sum_k n_{ijk} \lambda_{ijk}^{ABC} \\ & - \sum_i \sum_j \sum_k \exp[\lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} + \lambda_{ik}^{AC} + \lambda_{ijk}^{ABC}] \quad (1) \end{aligned}$$

Como la distribución de Poisson es una familia exponencial los coeficientes de los parámetros son los estadísticos suficientes. En el caso del modelo saturado no hay reducción de los datos en tanto los n_{ijk} son los coeficientes de λ_{ijk}^{ABC} y los demás se pueden obtener a partir de estos.

Para los modelos más sencillos algunos parámetros son 0, por lo que (1) se simplifica y se puede obtener cierta reducción. Por ejemplo en el modelo (A, B, C) , los estadísticos suficientes serán los coeficientes de λ_i^A , λ_j^B y λ_k^C , es decir n_{i++} , n_{+j+} y n_{++k} .

En la tabla que sigue damos los estadísticos minimales suficientes para distintos modelos. En general, en los modelos reducidos los estadísticos suficientes minimales son las sumas marginales correspondientes a los términos de mayor orden presentes en el modelo.

Ecuaciones de máxima verosimilitud

Para un modelo particular veamos como se calculan los EMV. Asumamos el modelo (AC, BC) . Para obtener los estimadores de los parámetros derivamos

Modelos	Estadísticos Suficientes Minimales		
(A,B,C)	n_{i++}	n_{+j+}	n_{++k}
(AB,C)		n_{ij+}	n_{++k}
(AB,BC)		n_{ij+}	n_{+jk}
(AB,BC,AC)	n_{ij+}	n_{+jk}	n_{i+k}

Cuadro 4: Estadísticos Suficientes Minimales

e igualamos a 0, entonces:

$$\begin{aligned} \ell(\boldsymbol{\mu}) = & n\lambda + \sum_i n_{i++} \lambda_i^A + \sum_j n_{+j+} \lambda_j^B + \sum_k n_{++k} \lambda_k^C \\ & + \sum_{i,k} n_{i+k} \lambda_{ik}^{AC} + \sum_{j,k} n_{+jk} \lambda_{jk}^{BC} \\ & - \sum_{i,j,k} \sum \exp [\lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}] \end{aligned}$$

$$\frac{\partial \ell}{\partial \lambda} = n - \sum_{i,j,k} \sum \exp [\lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}] = 0$$

$$= n - \sum_i \sum_j \sum_k \mu_{ijk} = 0$$

Luego:

$$\hat{\mu}_{+++} = n$$

De la misma manera, obtenemos

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda_i^A} &= n_{i++} - \sum_j \sum_k \mu_{ijk} = 0 \\ &= n_{i++} - \mu_{i++} = 0 \end{aligned}$$

Los valores estimados tienen los mismos totales que los observados. Así, obtenemos que

$$\{n_{i++} = \mu_{i++}, i = 1, \dots, I\}, \{n_{+j+} = \mu_{+j+}, j = 1, \dots, J\}, \{n_{++k} = \mu_{++k}, k = 1, \dots, K\}$$
 y que

$$\frac{\partial \ell}{\partial \lambda_{ik}^{AC}} = n_{i+k} - \mu_{i+k} = 0$$

$$\frac{\partial \ell}{\partial \lambda_{jk}^{BC}} = n_{+jk} - \mu_{+jk} = 0$$

Para este modelo los estadísticos suficientes minimales coinciden los EMV de las frecuencias esperadas marginales.

Este es un resultado más general que se cumple para los otros modelos estudiados hasta ahora.

Birch(1963) mostró que las ecuaciones de máxima verosimilitud en un modelo log-lineal igualan los estadísticos suficientes minimales con sus valores esperados.

Por ejemplo, en el modelo saturado:

$$E(n_{ijk}) = \mu_{ijk}$$

con lo cual

$$\bar{\mu}_{ijk} = n_{ijk}$$

Birch mostró que hay una única solución $\{\bar{\mu}_{ijk}\}$ que satisface el modelo y que iguala a los datos en sus estadísticos suficientes minimales. Por lo tanto si logramos esa solución, debe ser la solución de las ecuaciones de MV.

Birch mostró que los EMV son los mismos para el muestreo multinomial que para el muestreo Poisson independiente. Probó sus resultados bajo el supuesto de que en todas las casillas hay un número positivo de observaciones, sin embargo Haberman (1973, 1974) lo demostró sin este supuesto.

En realidad con tres variables, las ecuaciones pueden resolverse en forma directa, salvo para el modelo de *asociación homogénea*, en el que no hay una solución explícita.

¿Cómo ajustamos en este caso? Una posibilidad es hacer una matriz de diseño que contenga los efectos principales y las asociaciones y ajustar una regresión

de Poisson mediante Newton–Raphson.

Otra posibilidad es el Método Iterativo Proporcional.

Método Iterativo Proporcional (IPF)

Este método ajusta proporcionalmente las frecuencias esperadas μ_{ijk} de manera de satisfacer las restricciones:

$$\hat{\mu}_{ij+} = n_{ij+}$$

$$\hat{\mu}_{i+k} = n_{i+k}$$

$$\hat{\mu}_{+jk} = n_{+jk}$$

Es decir

$$1 = \frac{n_{ij+}}{\hat{\mu}_{ij+}} = \frac{n_{i+k}}{\hat{\mu}_{i+k}} = \frac{n_{+jk}}{\hat{\mu}_{+jk}}$$

Por lo tanto

$$\hat{\mu}_{ijk} = \left(\frac{n_{ij+}}{\hat{\mu}_{ij+}} \right) \hat{\mu}_{ijk}$$

$$\hat{\mu}_{ijk} = \left(\frac{n_{i+k}}{\hat{\mu}_{i+k}} \right) \hat{\mu}_{ijk}$$

$$\widehat{\mu}_{ijk} = \left(\frac{n_{+jk}}{\widehat{\mu}_{+jk}} \right) \widehat{\mu}_{ijk}$$

Esto sugiere un método iterativo que comienza con ciertos valores iniciales $\widehat{\mu}_{ijk}^{(0)}$, s que se van actualizando de la siguiente forma. Dados estimadores $\widehat{\mu}_{ijk}^{(3t)}$ las actualizaciones se obtienen

$$\begin{aligned} \widehat{\mu}_{ijk}^{(3t+1)} &= \left(\frac{n_{ij+}}{\widehat{\mu}_{ij+}^{(3t)}} \right) \widehat{\mu}_{ijk}^{(3t)} \\ \widehat{\mu}_{ijk}^{(3t+2)} &= \left(\frac{n_{i+k}}{\widehat{\mu}_{i+k}^{(3t+1)}} \right) \widehat{\mu}_{ijk}^{(3t+1)} \\ \widehat{\mu}_{ijk}^{(3(t+1))} &= \left(\frac{n_{+jk}}{\widehat{\mu}_{+jk}^{(3t+2)}} \right) \widehat{\mu}_{ijk}^{(3t+2)} \end{aligned}$$

El proceso se detiene cuando

$$\widehat{\mu}_{ijk}^{(3t)} \doteq \widehat{\mu}_{ijk}^{(3t+1)} \doteq \widehat{\mu}_{ijk}^{(3t+2)} \doteq \widehat{\mu}_{ijk}^{(3(t+1))}$$

Se suele tomar

$$\widehat{\mu}_{ijk}^{(0)} = 1 \quad \forall i, j, k$$

Este método converge más lentamente que N-R, pero es muy sencillo en tanto opera directamente sobre las μ_{ijk} y no es necesario introducir una matriz de diseño.

La ventaja de N-R es que automáticamente nos da una matriz de covarianza de los coeficientes con la que podemos computar estimadores de sus desvíos standard y de funciones de los coeficientes. Esto también se puede hacer con IFP.

Veamos que pasa con nuestros modelos

Clase 1: independencia total o mutua: (A, B, C)

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C$$

- Las variables son independientes
- A y B son independientes
- A y C son independientes
- B y C son independientes

Además:

- Las variables son condicionalmente independientes
- Las variables son marginalmente independientes
- La asociación marginal entre A y B es idéntica a la asociación parcial entre A y B . Idem $A-C$ y $B-C$.

- La asociación marginal=asociación parcial= nula entre cualquier par de variables.

Ejemplo: Christensen (1997)

Como parte de un estudio longitudinal una muestra de 3182 individuos sin enfermedad cardiovascular fue seguida durante 4 años y medio. Un total de 2121 de estos individuos no tuvieron actividad física en forma regular y no desarrollaron enfermedad cardiovascular. Los individuos fueron clasificados de acuerdo a tres factores: **Tipo de Personalidad**, **Nivel de Colesterol** y **Presión Diastólica**. Una personalidad es de tipo I si presenta signos de stress, inquietud e hiperactividad. Un individuo con personalidad de tipo II es relajado, pausado y de actividad normal. Nivel de Colesterol y Presión Diastólica fueron categorizados en alto y normal. Los datos se muestran en la siguiente tabla:

		Presión Diastólica	
Personalidad	Colesterol	Normal	Alta
I	Normal	716	79
	Alto	207	25
II	Normal	819	67
	Alto	186	22

Cuadro 5: Valores Observados

Si ajustamos los datos suponiendo un modelo de independencia total o completa (A, B, C) obtenemos la siguiente tabla:

Notemos que los $\hat{\mu}_{ijk}$ verifican que $\hat{\mu}_{i++} = n_{i++}$, $\hat{\mu}_{+j+} = n_{+j+}$ y $\hat{\mu}_{++k} = n_{++k}$ para todo i, j, k . Por ejemplo: $n_{1++} = 716 + 79 + 207 + 25 = 1027$ y $\hat{\mu}_{1++} = 739.9 + 74.07 + 193.7 + 19.39 = 1027.06$. La diferencia se atribuye a errores de redondeo.

Para testear la bondad del ajuste podemos usar el estadístico de Pearson o la deviance:

$\hat{\mu}_{ijk}$	Presión Diastólica		
Personalidad	Colesterol	Normal	Alta
I	Normal	739.9	74.07
	Alto	193.7	19.39
II	Normal	788.2	78.90
	Alto	206.3	20.65

Cuadro 6: Valores Esperados Estimados

$$\begin{aligned}
 &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}} \\
 &= \frac{(716 - 739.9)^2}{739.9} + \dots + \frac{(22 - 20.65)^2}{20.65} = 8.73 \\
 G^2 &= 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K n_{ijk} \log \left[\frac{n_{ijk}}{\hat{\mu}_{ijk}} \right]
 \end{aligned}$$

$$= 2 \left\{ 716 \log \frac{716}{739.9} + \dots + 22 \log \frac{22}{20.65} \right\} = 8.723$$

Grados de libertad:

$$df = IJK - [1 + (I - 1) + (J - 1) + (K - 1)] = IJK - I - J - K + 2$$

En nuestro caso nos queda: $df = 2 * 2 * 2 - 2 - 2 - 2 + 2 = 4$

Como $\chi^2_{4,0.05} = 9.49$ no rechazamos la hipótesis de independencia. Sin embargo, el p-valor es 0.07, por lo tanto no hay clara evidencia de independencia completa entre los tres factores tipo de personalidad, colesterol y presión diastólica.

Si computamos los residuos de Pearson como:

$$r_{ijk} = \frac{n_{ijk} - \hat{\mu}_{ijk}}{\sqrt{\hat{\mu}_{ijk}}}$$

obtenemos los siguientes valores:

$\hat{\mu}_{ijk}$	Presión Diastólica	
	Colesterol Normal	Alta
I	Normal	0.573
	Alto	1.274
II	Normal	-1.340
	Alto	0.297

Cuadro 7: Residuos de Pearson

Si bien no hay ninguna estructura especial en los residuos, parecería que hay más individuos con alta presión y alto colesterol en la personalidad I que los que espera una hipótesis de independencia y lo mismo ocurre con los individuos con colesterol normal y presión normal entre los de personalidad II. Dentro de la personalidad II hay menos individuos que tienen sólo un factor alto que los esperados bajo independencia.

Clase 2: Una variable independiente de otras dos:

(A, BC), (B, AC), (C, AB)

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$$

B y C dado A pueden ser condicionalmente dependientes.

- Dado C, A y B son condicionalmente independientes
- Dado B, A y C son condicionalmente independientes

Además:

- La tabla se puede colapsar en cualquier dirección.
- Asociación marginal = Asociación parcial para cualquier par de variables.
- Asociación marginal A-B = Asociación parcial A-B=nula.
Idem para A-C.

Por lo tanto:

- A y B son independientes en la marginal $A-B$.
- A y C son independientes en la marginal $A-C$.

Ejemplo

Everitt (1977) considera una muestra de 97 niños escolarizados de 10 años que fueron clasificados según los siguientes factores:

Conducta en Clase (A), Riesgo hogareño (B) y Adversidad Escolar (C).

La Conducta en Clase fue clasificada por los maestros como *descarriada* o *no descarriada*, el Riesgo Hogareño como *riesgo (R)* o *no riesgo (N)* y *Adversidad Escolar* en *baja*, *media* o *alta*. Los datos se hallan en la siguiente tabla.

Testear el modelo $\pi_{ijk} = \pi_{i++} \pi_{+jk}$ o equivalentemente

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$$

También equivale a testear independencia entre las dos filas Descarriado y No

Riesgo	Adversidad Escolar						
	Bajo		Medio		Alto		
	N	R	N	R	N	R	
Conducta No Descarriado	16	7	15	34	5	3	80
en Clase Descarriado	1	1	3	8	1	3	17
Total	17	8	18	42	6	6	97

Cuadro 8: Valores Observados

Descarriado y las seis columnas Bajo–N, Bajo–R, Medio–N, Medio–R, Alto–N, Alto–R.

La tabla de valores esperados bajo este modelo se ve en el Cuadro 9.

Los grados de libertad de este modelo son:

$$df = IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (J - 1)(K - 1)] \\ = IJK - I - JK + 1$$

Los estadísticos de bondad de ajuste dan:

$$X^2 = 6.19 \text{ (p-valor=0.288)} \quad G^2 = 5.56 \text{ (p-valor=0.351)}$$

Riesgo	Adversidad						Escolar		
	Bajo		Medio		Alto		N	R	Total
	N	R	N	R	N	R			
Conducta No Descarriado	14.02	6.60	14.85	34.64	4.95	4.95	80		
en Clase Descarriado	2.98	1.40	3.15	7.36	1.05	1.05	17		
Total	17	8	18	42	6	6	97		

Cuadro 9: Valores Esperados Estimados

$$df = 2 * 3 * 2 - 2 - 3 * 2 + 1 = 5$$

por lo que no dan evidencias contra el modelo supuesto.

Es decir, no tenemos razones para dudar sobre que la Conducta en Clase es independiente del Riesgo y la Adversidad.

Si nuestro objetivo inicial hubiera sido explicar la conducta en clase en términos del riesgo de la casa y de la adversidad en la escuela, habríamos sido desafortunados ya que la conducta es independiente de estas variables.

Por otra lado, bajo este modelo examinar la relación entre Riesgo y Adversidad

es sencillo, ya que si la conducta en clase es independiente de las otras dos variables podríamos estudiar la relación de Riesgo y Adversidad a través de la tabla marginal ya que bajo H_0 no puede ocurrir la paradoja de Simpson.

El modelo de independencia para esta tabla marginal es

$$\pi_{+jk} = \pi_{+j+} \pi_{++k} \quad \forall j, k$$

que corresponde a

$$\log \mu_{jk} = \lambda + \lambda_j^B + \lambda_k^C$$

La tabla marginal Adversidad–Riesgo resulta:

La tabla de valores esperados estimados con este modelo es

$$X^2 = 10.78 \quad G^2 = 10.86 \quad df = (2 - 1)(3 - 1) = 2$$

que son significativos al 1 %.

Los residuos de Pearson son

n_{+jk}		Adversidad			n_{+j+}
		Bajo	Medio	Alto	
Riesgo	N	17	18	6	41
	R	8	42	6	56
	n_{++k}	25	60	12	97

Cuadro 10: Tabla Marginal

$\widehat{\mu}_{+jk}$		Adversidad			$\widehat{\mu}_{+j+}$
		Bajo	Medio	Alto	
Riesgo	N	10.57	25.36	5.07	41
	R	14.43	34.64	6.93	56
	$\widehat{\mu}_{++k}$	25	60	12	97

Cuadro 11: Valores Esperados

		Adversidad		
		Bajo	Medio	Alto
Riesgo	N	1.98	-1.46	0.35
	R	-1.69	1.25	-0.35

Cuadro 12: Residuos

Parece haber cierta estructura en los residuos. En las escuelas de alta Adversidad los residuos son pequeños. Sin embargo, en las escuelas de baja adversidad (buenas) los alumnos sin riesgo (N) están sobrerrepresentados y los de riesgo (R) están subrepresentados.

Notemos que en realidad, antes de hacer el segundo paso de nuestro análisis sí aconsejable estudiar los residuos del primer modelo como para ver si el modelo se sostiene.

Clase 3: Independencia Condicional:

(AC, BC), (AB, AC), (AB, BC)

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

Sólo un par de variables son condicionalmente independientes.

- Dado C, A y B son condicionalmente independientes
- La tabla se puede colapsar en A y B.
- Asociación marginal A–C = Asociación parcial A–C
- Asociación marginal B–C = Asociación parcial B–C
- Asociación parcial A–B es nula. Asociación marginal = ??????

Este es un modelo en el que la asociación marginal A–B puede ser espúrea si uno ignora la variable C.

Grados de Libertad

$$\begin{aligned}
 df &= IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(K - 1) \\
 &\quad + (J - 1)(K - 1)] \\
 &= K(I - 1)(J - 1)
 \end{aligned}$$

Ejemplo

Volvemos al ejemplo del estudio longitudinal. Examinaremos el modelo en el que Nivel de Colesterol y Presión Diastólica son independientes dada la personalidad. Esto corresponde al modelo

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$$

Testear la validez de este modelo es testear simultáneamente de independencia entre las tablas:

n_{1kj} Colesterol	Personalidad I	
	Presión Diastólica Normal	Alto
Normal	716	79
Alto	207	25

n_{2kj} Colesterol	Personalidad II	
	Presión Diastólica Normal	Alto
Normal	819	67
Alto	186	22

donde cada tabla tiene $(2 - 1)(2 - 1) = 1$ grados de libertad.
Los estadísticos de bondad de ajuste obtenidos son:

$$X^2 = 2.188 \quad G^2 = 2.062$$

Los grados de libertad de los estadísticos de bondad de ajuste serán $df = 2(2 - 1)(2 - 1) = 2$ y por lo tanto obtenemos un ajuste razonable ya que el p-valor de cada uno es 0.3348743 y 0.3566501, respectivamente.

Veamos la salida de LEM para este ejemplo.

LEM: log-linear and event history analysis with missing data.
Developed by Jeroen Vermunt (c), Tilburg University, The Netherlands.
Version 1.0 (September 18, 1997).

```
*** INPUT ***
* A = Personalidad
* B = Colesterol
* C = Presión Diastólica
```

```
man 3
dim 2 2 2
lab A B C
mod {AB,AC}
dat [ 716 79
      207 25
      819 67
      186 22]
```

```
*** STATISTICS ***
```

```
Number of iterations = 2
Converge criterion   = 0.00000000000
X-squared            = 2.1876 (0.3349)
```

L-squared = 2.0626 (0.3565)
 Cressie-Read = 2.1439 (0.3423)
 Dissimilarity index = 0.0062
 Degrees of freedom = 2
 Log-likelihood = -3195.22085
 Number of parameters = 5 (+1)
 Sample size = 2121.0
 BIC(L-squared) = -13.2567
 AIC(L-squared) = -1.9374
 BIC(log-likelihood) = 6428.7399
 AIC(log-likelihood) = 6400.4417

*** FRECUENCIAS ***

A	B	C	observed	estimated	std. res.
1	1	1	716.000	714.494	0.056
1	1	2	79.000	80.506	-0.168
1	2	1	207.000	208.506	-0.104
1	2	2	25.000	23.494	0.311
2	1	1	819.000	813.921	0.178
2	1	2	67.000	72.079	-0.598
2	2	1	186.000	191.079	-0.367
2	2	2	22.000	16.921	1.235

*** LOG-LINEAR PARAMETERS ***

* TABLE ABC [or P(ABC)] *

effect	beta	std err	z-value	exp(beta)	Wald	df	prob
main	4.8147			123.3082			
A							
1	0.0495	0.0410	1.207	1.0507			
2	-0.0495			0.9517	1.46	1	0.228
B							
1	0.6702	0.0268	24.991	1.9546			
2	-0.6702			0.5116	624.56	1	0.000
C							
1	1.1518	0.0379	30.423	3.1640			
2	-1.1518			0.3161	925.54	1	0.000
AB							
1 1	-0.0544	0.0268	-2.029	0.9471			
1 2	0.0544			1.0559			
2 1	0.0544			1.0559			
2 2	-0.0544			0.9471	4.11	1	0.043
AC							
1 1	-0.0602	0.0379	-1.591	0.9416			
1 2	0.0602			1.0621			
2 1	0.0602			1.0621			
2 2	-0.0602			0.9416	2.53	1	0.112

Clase 4: Asociación Homogénea: (AB, AC, BC)

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

Todos los pares de variables son condicionalmente dependientes dada la tercera. La asociación parcial entre dos variables no varía con los distintos niveles de la tercera.

- ¿Podemos colapsar en alguna dirección?

Grados de Libertad

$$\begin{aligned} df &= IJK - [1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1)] \\ &\quad + (I - 1)(K - 1) + (J - 1)(K - 1) \\ &= (I - 1)(J - 1)(K - 1) \end{aligned}$$

Ejemplo: Fienberg(1980)

Los siguientes observaciones corresponden a datos sobre accidentes de autos pequeños. Las variables consideradas son: **severidad de las heridas del conductor, tipo de accidente y si el conductor fue eyectado o no**. La tabla de valores observados y esperados usando el modelo de Asociación Homogénea son:

n_{ijk}	Herido Conductor Eyectado	Tipo de Accidente			
		Colisión		Vuelco	
	No	No Severo	Severo	No Severo	Severo
No		350	150	60	112
Sí		26	23	19	80

Cuadro 13: Valores Observados

Los estadísticos de bondad de ajuste resultan

$$X^2 = 0.04323 \quad G^2 = 0.04334 \quad df = (2 - 1)(2 - 1)(2 - 1) = 1$$

$\hat{\mu}_{ijk}$	Herido	Tipo de Accidente			
		Colisión		Vuelco	
Conductor Eyectado	No	No Severo	Severo	No Severo	Severo
	No	350.5	149.5	59.51	112.5
	Sí	25.51	23.49	19.49	79.51

Cuadro 14: Valores Estimados

con un p-valor 0.8352928 y 0.8350864, respectivamente. Por lo tanto, el modelo de igualdad de los odds ratios ajusta muy bien.

Sugiero comprobar que los modelos de independencia e independencia condicional no dan un buen ajuste.

Nos falta considerar una cosa importante:

- ¿Cómo elegir un modelo log-lineal adecuado para un conjunto de datos?

Raza de la Víctima	Raza del Defendido	Pena de Muerte	
		Si	No
Blanca	Blanca	53	414
	Negra	11	37
Negra	Blanca	0	16
	Negra	4	139
Total		53	430
		Negra	15 176

Cuadro 15: Pena de Muerte por raza del defenido y raza de la víctima

¿Cómo elegir un modelo log–lineal adecuado para un conjunto de datos?

Consideremos nuevamente el ejemplo de **raza víctima – raza defenido – veredicto**. Los datos se muestran en el Cuadro 15.

En la siguiente tabla mostramos los resultados correspondientes a tests de bondad de ajuste de diferentes modelos loglineales.

Como sabemos, dado df fijo un G^2 mayor refleja un peor ajuste. La tabla anterior muestra que los modelos (D, V, P) , (VP, D) , (DP, V) y (DP, VP) dan un ajuste muy pobre. La característica común de todos ellos es que la asociación $D - -V$ no está presente. Esto sugiere una asociación importante entre estas dos variables.

De los restantes 4 modelos no saturados (VP, DV) y (DP, VP, DV) dan un ajuste adecuado (al menos para los niveles acostumbrados 5% y 10%). En el modelo (DP, VP, DV) todas las variables son condicionalmente dependientes, mientras que en el (VP, DV) la raza del defendido y el veredicto son independientes dada la raza de la víctima.

Inferencia sobre asociación condicional

Los tests respecto a las asociaciones condicionales comparan modelos loglineales.

Supongamos que tenemos dos modelos M_0 y M_1 , donde M_0 está anidado en

Modelo	G^2	df	p-valor
(D,V,P)	137.93	4	0.000
(VP,D)	131.68	3	0.000
(DP,V)	137.71	3	0.000
(DV,P)	8.63	3	0.043
(DP,VP)	131.46	2	0.000
(DP,DV)	7.91	2	0.019
(VP,DV)	1.88	2	0.390
(DP,VP,DV)	0.70	1	0.402
(DVP)	0	0	—

Cuadro 16: Tests de bondad de ajuste

M_1 , teniendo un término menos que M_1 .

El test de cociente de verosimilitud $-2(\ell_0 - \ell_1)$ es, como ya vimos, idéntico a la diferencia de las deviances $G^2(M_0) - G^2(M_1)$. Esto tiene sentido cuando M_1 da un buen ajuste.

Así por ejemplo, supongamos que para el modelo (AB, AC, BC) consideramos la hipótesis de que A-B son condicionalmente independientes. Esto es

$$H_0 : \lambda_{ij}^{AB} = 0$$

para las $(I - 1) * (J - 1)$ parámetros de asociación AB.

El estadístico del test es

$$G^2((AC, BC)) - G^2((AB, AC, BC))$$

con $(I - 1) * (J - 1)$ grados de libertad.

En el caso de la situación anterior, por ejemplo, el test para independencia condicional entre veredicto y raza del defendido compara el modelo (VP, DV) con el (DP, VP, DV).

El estadístico correspondiente es

$$G^2(VP, DV) - G^2(DP, VP, DV) = 1.88 - 0.70 = 1.18$$

con $df = 1$ y $p\text{-valor} = 0.277356$.

Finalmente, podríamos considerar una secuencia de modelos anidados tales como

$$(D, V, P) \quad (P, DV) \quad (VP, DV) \quad (DP, VP, DV) \quad (DVP)$$

Para asegurar un nivel global en nuestra decisión que no exceda a 0.10 realizamos cada test con un nivel $1 - (0.90)^{0.25} = 0.026$. Para un grado de libertad el valor crítico de una χ^2 es 4.96. En la siguiente tabla se muestran los resultados de las comparaciones

De la tabla concluimos que primero aceptamos el modelo de asociación homogénea (DP, VP, DV) y luego aceptamos (VP, DV) dado (DP, VP, DV) . En tercera instancia el modelo (P, DV) es rechazado.

Por otro lado, tal como vimos el modelo (VP, DV) da un buen ajuste global (en un test no condicional), por lo que parece razonable elegirlo para representar los datos.

Modelo	G^2	diferencia	df
(D,V,P)	137.93		4
(DV,P)	8.63	129.80	1
(VP,DV)	1.88	6.25	1
(DP,VP,DV)	0.70	1.18	1
(DVP)	0	0.70	1
		0	0

Cuadro 17: Tests de bondad de ajuste

Por supuesto, la selección en una secuencia de modelos anidados, puede depender de la secuencia elegida. Así, por ejemplo, si la secuencia es (D, V, P) , (P, DV) , (DP, DV) , (DP, VP, DV) el modelo resultante será (DP, VP, DV) .

Veamos otro Ejemplo.

Ejemplo: Cigarrillo–Alcohol–Marihuana

Los datos corresponden a un estudio realizado en la Wright State University School of Medicine and the United Health Services en Dayton, Ohio, realizado a fin de relacionar el uso de alcohol (A), marihuana (M) y cigarrillos (C) en los estudiantes. Se realizó una encuesta entre los estudiantes del último año y se les preguntó si alguna vez habían consumido alguna de las tres cosas.

Table 7.3. Alcohol (A), Cigarette (C), and Marijuana (M) Use for High School Seniors

Alcohol Use	Cigarette Use	Marijuana Use	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

Table 7.4. Fitted Values for Loglinear Models Applied to Table 7.3

Alcohol Use	Cigarette Use	Marijuana Use	Loglinear Model				
			(A, C, M)	(AC, M)	(AM, CM)	(AC, AM, CM)	(ACM)
Yes	Yes	Yes	540.0	611.2	909.24	910.4	911
	No	No	740.2	837.8	438.84	538.6	538
No	Yes	Yes	282.1	210.9	45.76	44.6	44
	No	No	386.7	289.1	555.16	455.4	456
No	Yes	Yes	90.6	19.4	4.76	3.6	3
	No	No	124.2	26.6	142.16	42.4	43
No	Yes	Yes	47.3	118.5	0.24	1.4	2
	No	No	64.9	162.5	179.84	279.6	279

(AM, CM)

$$AC = 1.0 = \frac{909.24 \times 0.24}{45.76 \times 4.76} = \frac{438.84 \times 179.84}{555.16 \times 142.16}$$

marginal AC

$$2.7 = \frac{(909.24 + 438.84)(0.24 + 179.84)}{(45.76 + 555.16)(4.76 + 142.16)}$$

Table 7.5. Estimated Odds Ratios for Loglinear Models in Table 7.4

Model	Conditional Association			Marginal Association		
	AC	AM	CM	AC	AM	CM
(A, C, M)	1.0	1.0	1.0	1.0	1.0	1.0
(AC, M)	17.7	1.0	1.0	17.7	1.0	1.0
(AM, CM)	1.0	61.9	25.1	2.7	61.9	25.1
(AC, AM, CM)	7.8	19.8	17.3	17.7	61.9	25.1
(ACM) level 1	13.8	24.3	17.5	17.7	61.9	25.1
(ACM) level 2	7.7	13.5	9.7			

Procesemos el ejemplo con LEM.

Table 7.7. Goodness-of-Fit Tests for Loglinear Models Relating Alcohol (A), Cigarette (C), and Marijuana (M) Use

Model	G^2	X^2	df	P -value*
(A, C, M)	1286.0	1411.4	4	<0.001
(A, CM)	534.2	505.6	3	<0.001
(C, AM)	939.6	824.2	3	<0.001
(M, AC)	843.8	704.9	3	<0.001
(AC, AM)	497.4	443.8	2	<0.001
(AC, CM)	92.0	80.8	2	<0.001
(AM, CM)	187.8	177.6	2	<0.001
(AC, AM, CM)	0.4	0.4	1	0.54
(ACM)	0.0	0.0	0	—

* P -value for G^2 statistic.