

Modelo 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.

El modelo loglineal puede ser visto como un caso del GLM para datos con distribución Poisson. Los modelos log-lineales se usan con frecuencia para analizar la relación de dos, tres o más variables categóricas en una tabla de contingencia.

Todas las variables que se analizan son vistas como variables de respuesta, es decir que no se hace distinción entre variables independientes y dependientes. Es por ello, que estos modelos sólo estudian asociación. Cuando interesa estudiar algunas variables como independientes y otras como dependientes es más adecuado el modelo logístico. Lo mismo es cierto si las variables estudiadas son continuas y no se pueden discretizar apropiadamente.

La estrategia básica en el modelado ajusta las frecuencias observadas en la tabla cruzada. Los modelos son representados por las frecuencias esperadas y podrán ser descriptos por las restricciones que imponen a las asociaciones o interacciones entre las variables. Los patrones de asociación entre las variables pueden describirse en términos de los odds y los odds ratios.

Modelos para tablas bidimensionales

Comenzamos por considerar el caso más sencillo de tablas de contingencia. Los conceptos que aquí veremos se extienden a tablas más complejas.

Consideremos el siguiente ejemplo en el que tenemos dos variables:

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

En general tendremos:

S	C		Total
	Si	No	
Mujer	y_{11}	y_{12}	y_{1+}
Hombre	y_{21}	y_{22}	y_{2+}
Total	y_{+1}	y_{+2}	$y_{++} = n$

Dada una tabla de contingencia hay varios esquemas de muestreo que pueden conducir a los datos tal como los hemos observado y que influirán en el modelo de probabilidad que usaremos. En este caso

tenemos dos factores S y C cada uno con dos niveles. En general, tendremos un factor *fila* con I niveles y un factor *columna* con J niveles que corresponde a

	$Y = 1$	$Y = 2$			$Y = J$
$X = 1$	y_{11}	y_{12}			y_{1J}
$X = 2$	y_{21}	y_{22}			y_{2J}
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot	\cdot
$X = I$	y_{I1}	y_{I2}			y_{IJ}

Modelo Multinomial

En el primer esquema asumimos que todos los datos han sido recolectados muestreando 1091 individuos que fueron clasificados de acuerdo con el **sexo** y la **creencia en la vida después de la muerte**. Vemos las dos variables como respuesta y nos interesa su distribución conjunta.

La distribución *conjunta* de X y Y se describe en forma completa a través de las probabilidades Π_{ij} :

$$\Pi_{ij} = P(X = i, Y = j) \quad i = 1, \dots, I, \quad j = 1, \dots, J.$$

La distribución *marginal* de cada variable la calculamos como es habitual:

Π_{i+} = probabilidad de que la variable fila tome el valor i

Π_{+j} = probabilidad de que la variable columna tome el valor j

$$\Pi_{i+} = P(X = i) \quad \text{y} \quad \Pi_{+j} = P(Y = j)$$

Cuando hay sólo dos variables la hipótesis de interés es si las variables son independientes. Podemos escribir esta hipótesis como

$$H_o : \Pi_{ij} = \Pi_{i+} \Pi_{+j} \quad \forall i \forall j$$

Podríamos estimar por máxima verosimilitud las probabilidades y realizar un test de cociente de verosimilitud, para esto necesitamos asumir una distribución subyacente.

Si cada una de las n observaciones es clasificada en forma independientemente en una de las $I \times J$ celdas de la tabla con probabilidad Π_{ij} , entonces si Y_{ij} es la v.a. que representa el número de individuos clasificados en la celda (i, j) e y_{ij} es el valor observado en dicha celda, tenemos que $\mathbf{Y} = (Y_{11}, \dots, Y_{IJ})'$ tiene distribución multinomial. Para $I = J = 2$ sería

$$P(\mathbf{Y} = \mathbf{y}) = \frac{n!}{y_{11}! y_{12}! y_{21}! y_{22}!} \Pi_{11}^{y_{11}} \Pi_{12}^{y_{12}} \Pi_{21}^{y_{21}} \Pi_{22}^{y_{22}},$$

Salvo constantes, el log-likelihood en el caso general queda:

$$\ell = \sum_{i=1}^I \sum_{j=1}^J y_{ij} \log \Pi_{ij}$$

La maximización de ℓ debe contemplar que $\sum_{i=1}^I \sum_{j=1}^J \Pi_{ij} = 1$.

Encontramos que el EMV es

$$\widehat{\Pi}_{ij} = \frac{Y_{ij}}{n}$$

Bajo el supuesto de independencia tenemos que $\Pi_{ij} = \Pi_{i+} \Pi_{+j}$ si calculamos el EMV bajo esta hipótesis obtenemos

$$\widehat{\Pi}_{i+} = \frac{Y_{i+}}{n} \quad \text{y} \quad \widehat{\Pi}_{+j} = \frac{Y_{+j}}{n}$$

Combinando estos estimadores las probabilidades estimadas resultan

$$\widehat{\Pi}_{ij} = \widehat{\Pi}_{i+} \widehat{\Pi}_{+j}$$

y las frecuencias esperadas estimadas

$$\widehat{\mu}_{ij} = \frac{y_{i+} y_{+j}}{n}$$

que es el resultado conocido.

En el ejemplo queda

$$\widehat{\mu} = (27.2, 9970.8, 19.8, 266.2)$$

El ajuste de las frecuencias estimadas puede realizarse a través del estadístico de Pearson o a la deviance que asintóticamente tienen $(I - 1)(J - 1)$ grados de libertad. La deviance en este caso resulta

$$D = 2 \sum_{i=1}^I \sum_{j=1}^J y_{ij} \log \frac{y_{ij}}{n \widehat{\Pi}_{ij}} = 2 \sum_{i=1}^I \sum_{j=1}^J y_{ij} \log \frac{y_{ij}}{\widehat{\mu}_{ij}}$$

Modelo Poisson

Otra posibilidad es tratar a los 4 datos de la tabla como si provinieran de realizaciones independientes de v.a. con distribución Poisson.

Un proceso que sustentaría este modelo es que en cada celda los individuos llegan aleatoriamente al lugar donde se los clasifica. En este caso n no está prefijado y todos los valores de la tabla son aleatorios.

Bajo el supuesto de independencia entre las observaciones la conjunta queda

$$P(\mathbf{Y} = \mathbf{y}) = \prod_{i=1}^I \prod_{j=1}^J \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}!}$$

Podríamos suponer un modelo log–lineal: $\log \mu_{ij} = \eta_{ij}$ y recordemos la expresión de la deviance cuando tenemos observaciones y_1, \dots, y_n

$$D = 2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{\widehat{\mu}_i} - (y_i - \widehat{\mu}_i) \right)$$

y que se reduce a

$$D = 2 \sum_{i=1}^n y_i \log \frac{y_i}{\widehat{\mu}_i}$$

cuando el modelo tiene intercept. Esta expresión coincide con la del modelo multinomial.

En términos de la componente sistemática podríamos establecer modelos: el modelo nulo, un modelo aditivo en el que suponemos que las tasas dependen en forma aditiva del sexo y de la creencia, pero no en una combinación de ambas y un modelo saturado.

El modelo aditivo supone:

$$\log \mu_{ij} = \nu + \alpha_i + \beta_j$$

donde α y β representan los efectos del sexo y de la creencia respectivamente. Notemos que en modelo multinomial bajo independencia queda

$$\begin{aligned} \mu_{ij} &= n \Pi_{i+} \Pi_{+j} \\ \log \mu_{ij} &= \log n + \log \Pi_{i+} + \log \Pi_{+j} \end{aligned}$$

por lo tanto ambos modelos tienen la misma estructura.

Las coincidencias entre ambos modelos no son casuales sino que se pueden justificar por el siguiente resultado:

Supongamos que X_1, X_2, \dots, X_k son v.a. independientes con distribución $P(\lambda_i)$ de manera que $E(X_i) = \lambda_i \quad i = 1, \dots, k$. Entonces la distribución condicional de

$$\mathbf{X} = (X_1, \dots, X_k) |_{\sum_{i=1}^k X_i = n} \sim M(n, \Pi_1, \dots, \Pi_k)$$

donde

$$\Pi_k = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_k}$$

Este resultado es importante ya que nos dice que la distribución conjunta de $\mathbf{X} = (X_1, \dots, X_k)$ se puede factorizar como un producto de

$$n = \sum_{i=1}^k X_i \sim P(\lambda_1 + \dots + \lambda_k)$$

$$\mathbf{X} = (X_1, \dots, X_k) |_n \sim M(n, \Pi_1, \dots, \Pi_k)$$

donde n tiene información sobre $\lambda_1 + \dots + \lambda_k$, pero no tiene información sobre $\mathbf{\Pi} = (\Pi_1, \dots, \Pi_k)$.

Por lo tanto, este resultado más el hecho de que en presencia de intercept el modelo Poisson ajusta el total, implica que la inferencia sobre $\mathbf{\Pi}$ no depende de si (X_1, \dots, X_k) provienen de k variables Poisson o de una multinomial.

Modelo Producto–Multinomial

En un tercer caso podríamos asumir que se tomó una muestra de 582 mujeres y otra muestra independiente de 509 hombres a los que se clasificó según su creencia. En este caso nos centramos en la distribución condicional de la creencia dado cada nivel de sexo.

En este esquema los totales por fila están fijos y tenemos n_1 individuos de sexo femenino y n_2 individuos de sexo masculino.

Este tipo de muestreo también es apropiado cuando los totales de las filas no están fijos, pero nos interesa $P(Y|X)$ y no $P(X)$, esto en general ocurre cuando Y es el resultado de interés y X es una variable explicativa que no nos interesa modelar.

Si Π_i es la probabilidad de que el individuo crea en la vida después de la muerte para el nivel i de sexo tendremos:

$$\Pi_i = P(C = 1|S = i) = \frac{\Pi_{i1}}{\Pi_{i+}}$$

Las variables de interés serán: $Y_{i1} \sim Bi(n_i, \Pi_i)$, $i = 1, 2$, que cuentan el número de individuos que sí creen en cada sexo.

La distribución conjunta de (Y_{11}, Y_{21}) es

$$P((Y_{11}, Y_{21}) = (y_{11}, y_{21})) = \frac{n_1!}{y_{11}!y_{12}!} \Pi_1^{y_{11}} (1 - \Pi_1)^{y_{12}} \frac{n_2!}{y_{21}!y_{22}!} \Pi_2^{y_{21}} (1 - \Pi_2)^{y_{22}}$$

La hipótesis de interés es la de **homogeneidad**, es decir que la probabilidad de creencia es la misma en los dos sexos:

$$H_0 : \Pi_1 = \Pi_2$$

Si hiciéramos un ajuste por regresión logística a los datos usando creencia como respuesta y sexo como predictora observaríamos que el test de cociente de verosimilitud para H_0 coincide con la deviance del la multinomial y de Poisson.

Esto, nuevamente, no es una coincidencia sino que los modelos son equivalentes. El producto de las dos binomiales lo podríamos haber obtenido comenzando con el supuesto de que Y_{ij} son $P(\mu_{ij})$ independientes y condicionando a los totales Y_{i+} que tienen distribución

$P(\sum_j \mu_{ij})$, para $i = 1, 2$. Repitiendo el análisis anterior, tomando el cociente de la conjunta y el producto de las dos marginales de los totales llegamos a la distribución producto–binomial. También podríamos pensar en que tenemos una distribución original multinomial y que la factorizamos en dos (I) binomiales (multinomiales) independientes y otro factor que es el total de las filas.

Notemos que bajo independencia $\Pi_{ij} = \Pi_{i+}\Pi_{+j}$, entonces la probabilidad condicional

$$\Pi_{j|i} = \frac{\Pi_{ij}}{\Pi_{i+}} = \Pi_{+j}$$

es decir no depende de la fila i , con lo cual homogeneidad e independencia son equivalentes.

La importancia de estos resultados es que el análisis que hagamos es independiente del esquema de muestreo y depende de los parámetros de interés.

1. Si lo que nos interesa son funciones de las $\Pi_{j|i}$ podemos tratar los datos como si proviniesen de un producto–multinomial
2. Si el total está fijo podemos tratar a las columnas y a las filas como respuestas y testear independencia a través de un modelo multinomial
3. o bien podemos tratar a todas las casillas como aleatorias y ajustar un modelo log–lineal usando una regresión Poisson.

Modelo de Independencia

Como ya vimos bajo independencia

$$\begin{aligned}\log \mu_{ij} &= \log n + \log \Pi_{i+} + \log \Pi_{+j} \\ \log \mu_{ij} &= \lambda + \lambda_i^X + \lambda_j^Y\end{aligned}$$

donde λ_i^X y λ_j^Y representan el efecto de la fila i y de la columna j , respectivamente.

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

$$\begin{aligned}\log \left(\frac{\Pi}{1 - \Pi} \right) &= \log \left(\frac{\mu_{i1}}{\mu_{i2}} \right) \\ &= \lambda_1^Y - \lambda_2^Y\end{aligned}$$

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

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

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

Restricciones sobre los parámetros

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

La siguiente tabla muestra tres conjuntos de parámetros diferentes para los datos de *creencia*.

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

$$\lambda_1^X + \lambda_2^X = 0 \quad \lambda_1^Y + \lambda_2^Y = 0$$

Lo que todos cumplirán es que la diferencia entre dos efectos principales es la misma. En nuestro ejemplo, tenemos que $\lambda_1^Y - \lambda_2^Y = 1.059$ para los tres conjuntos de parámetros.

Modelo Saturado

Cuando las variables son dependientes satisfacen un modelo más complejo

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

donde los parámetros λ_{ij}^{XY} reflejan la asociación entre X y Y . 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 $\lambda_{ij}^{XY} = 0$.

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

$$\begin{aligned} \log \theta &= \log \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} \right) = \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21} \\ &= (\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + (\lambda + \lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}) \\ &\quad - (\lambda + \lambda_1^X + \lambda_2^Y + \lambda_{12}^{XY}) + (\lambda + \lambda_2^X + \lambda_1^Y + \lambda_{21}^{XY}) \\ &= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY} \end{aligned}$$

Los λ_{ij}^{XY} determinan los log odds ratios. Cuando $\lambda_{ij}^{XY} = 0$ los odds ratios valen 1 y X e Y 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

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

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 Poisson, dando un ajuste perfecto.

En la práctica se trata de usar modelos no saturados en tanto su ajuste "suaviza" 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}^{XY} , entonces también están presentes en el modelo λ_i^X y λ_j^Y . Estos son los modelo 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 X , Y y Z que tienen valores posibles:

A : 1, 2, ..., I

B : 1, 2, ..., J

C : 1, 2, ..., K

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

Si las n unidades en la muestra son i.i.d. el vector $(Y_{111}, Y_{112}, \dots, Y_{IJK})$ tiene distribución multinomial con parámetros n y $\mathbf{\Pi} = (\Pi_{111}, \Pi_{112}, \dots, \Pi_{IJK})$

La única restricción que se imponga a $\mathbf{\Pi}$ es que sume 1 y el EMV será el vector de componentes

$$\widehat{\Pi}_{ijk} = \frac{Y_{ijk}}{n}$$

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

$$\alpha_i = P(A = i) \quad i = 1, \dots, I$$

$$\beta_j = P(B = j) \quad j = 1, \dots, J$$

$$\delta_k = P(C = k) \quad k = 1, \dots, K$$

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

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

$$(Y_{1++}, \dots, Y_{I++}) \sim M(n, \alpha_1, \alpha_2, \dots, \alpha_I)$$

$$(Y_{+1+}, \dots, Y_{+J+}) \sim M(n, \beta_1, \beta_2, \dots, \beta_J)$$

$$(Y_{++1}, \dots, Y_{++K}) \sim M(n, \delta_1, \delta_2, \dots, \delta_K)$$

por lo tanto cada vector de parámetros podría estimarse en forma independiente uno de otro. Más aún, el EMV será

$$\widehat{\alpha}_i = \frac{Y_{i++}}{n}$$

$$\widehat{\beta}_j = \frac{Y_{+j+}}{n}$$

$$\widehat{\delta}_k = \frac{Y_{++k}}{n}$$

Graficamente, este modelo se representa como en el plot (1) del siguiente gráfico. 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 este modelo se representa como (A, B, C) .

En términos de los odds ratios este modelo significa que 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 independiente 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

$$\Pi_{ijk} = \alpha\beta_{ij} \delta_k \quad \forall i, j, k$$

donde

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

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

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

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

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 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 .

Equivalentemente, 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

Recordemos el ejemplo de primer Sexo–Ingreso–Carrera en el vimos que puede haber independencia condicional, pero independencia marginal

		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 idéntica

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.

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 las variables A o B 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 .

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, pasusado 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	Alto
I	Normal	716	79
	Alto	207	25
II	Normal	819	67
	Alto	186	22

Table 1: Valores Observados

Si ajustamos los datos suponiendo un modelo de independencia total o completa obtenemos

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

Table 2: Valores Esperados Estimados

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

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

$$\begin{aligned}
 X^2 &= \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \frac{(n_{ijk} - \widehat{\mu}_{ijk})^2}{\widehat{\mu}_{ijk}} \\
 &= \frac{(716 - 739.9)^2}{739.9} + \dots + \frac{(22 - 20.65)^2}{20.65} = 8.730
 \end{aligned}$$

y la deviance D :

$$\begin{aligned} D &= 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K n_{ijk} \log \left[\frac{n_{ijk}}{\widehat{\mu}_{ijk}} \right] \\ &= 2 \left\{ 716 \log \frac{716}{739.9} + \dots + 22 \log \frac{22}{20.65} \right\} = 8.723 \end{aligned}$$

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 $\chi_{4,0.05}^2 = 9.49$ 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:

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

obtenemos

$\widehat{\mu}_{ijk}$		Presión Diastólica	
Personalidad	Colesterol	Normal	Alto
I	Normal	-0.879	0.573
	Alto	0.956	1.274
II	Normal	1.097	-1.340
	Alto	-1.413	0.297

Table 3: 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. Esto responde al sentido común. 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:

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

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

En este caso, eventualmente un par de variables son condicionalmente dependientes: B y C dado A .

- 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

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

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

Table 4: Valores Observados

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

$\hat{\mu}_{ijk}$		Adversidad Escolar						Total
		Bajo		Medio		Alto		
		N	R	N	R	N	R	
Conducta en Clase	No Descarriado	14.02	6.60	14.85	34.64	4.95	4.95	80
	Descarriado	2.98	1.40	3.15	7.36	1.05	1.05	17
Total		17	8	18	42	6	6	97

Table 5: Valores Esperados Estimados

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:

$$X^2 = 6.19 \quad D = 5.56 \quad df = 2 * 3 * 2 - 2 - 3 * 2 + 1 = 5$$

Como $\chi_{5,0.05}^2 = 11.07$ al nivel 5% el test no da 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, seríamos 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 podemos estudiar la relación de Riesgo y Adversidad a través de la tabla marginal sin temer que ocurra la paradoja de Simpson.

La tabla marginal Adversidad–Riesgo resulta:

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

Table 6: Tabla Marginal

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 de valores esperados estimados con este modelo es

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

Table 7: Valores Esperados

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

que son significativos al 1%.

Los residuos de Pearson son

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

Table 8: 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 deberíamos 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 es condicionalmente independiente

- Dado C , A y B son condicionalmente independientes
- La tabla se puede colapsar en A y B , pero no en C .
- 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 importante 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 un test simultáneo de independencia entre las tablas:

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

n_{1kj}	Personalidad II	
	Presión Diastólica	
Colesterol	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:

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

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 muy buen ajuste ya que el p-valor de cada uno es 0.3348743 y 0.3566501, respectivamente.

Veamos la salida de LEM para este ejemplo.

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.

- No se puede colapsar en ninguna 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 estimados usando el modelo de Asociación Homogénea son:

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

Table 9: Valores Observados

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

Table 10: Valores Estimados

Los estadísticos de bondad de ajuste resultan

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

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

Veremos en la clase práctica que los modelos de independencia e independencia condicional no dan un buen ajuste.

Nos falta considerar dos cosas importantes

- Cómo se estimaron los valores estimados en el último ejemplo?
- Cómo elegir un modelo log-lineal adecuado para un conjunto de datos?

Estadísticos suficientes y minimales

Supongamos que n_{ijk} tienen distribución de Poisson con media μ_{ijk} . 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

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

queda

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

Como la distribución de Poisson es una familia exponencial los coeficientes de los parámetros son los estadísticos suficientes minimales. 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, como por ejemplo para el modelo (AB, AC, BC) los estadísticos minimales suficientes serán

$$\{n_{ij+}\}, \{n_{i+k}\}, \{n_{+jk}\}$$

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.

Estos resultados coinciden con las deducciones que hicimos para el modelo multinomial algunas clases atrás.

Ecuaciones de máxima verosimilitud

Para un modelo particular veamos como se calculan los EMV. Asumamos el modelo (AC, BC) , entonces:

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

Para obtener los estimadores de los parámetros derivamos e igualamos a 0, por ejemplo:

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= n - \sum_i \sum_j \sum_k \exp \left[\lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \right] \\ &= n - \sum_i \sum_j \sum_k \mu_{ijk} \end{aligned}$$

Luego:

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

De la misma manera, obtenemos

$$\frac{\partial \ell}{\partial \lambda_i^A} = \frac{n_{i++} - \sum_j \sum_k \mu_{ijk}}{n_{i++} - \mu_{i++}}$$

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

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda_{ik}^{AC}} &= n_{i+k} - \mu_{i+k} \\ \frac{\partial \ell}{\partial \lambda_{jk}^{BC}} &= n_{+jk} - \mu_{+jk} \end{aligned}$$

Para este modelo vemos que los estadísticos suficientes minimales son 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

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

Birch mostró que hay una única solución $\{\widehat{\mu}_{ijk}\}$ que satisface el modelo y que iguala a los datos en sus estadísticos suficientes mini-

males. Por lo tanto si logramos esa solución, debe ser la solución de las ecuaciones de MV.

Por ejemplo, en el caso de (AC, BC) el modelo en términos de las Π_{ijk} es:

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

o equivalentemente

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

Por las ecuaciones de MV, sabemos que $\widehat{\mu}_{i+k} = n_{i+k}$ y $\widehat{\mu}_{+jk} = n_{+jk}$ y entonces $\widehat{\mu}_{++k} = n_{++k}$. Luego:

$$\widehat{\mu}_{ijk} = \frac{\widehat{\mu}_{i+k}\widehat{\mu}_{+jk}}{\widehat{\mu}_{++k}} = \frac{n_{i+k}n_{+jk}}{n_{++k}} \quad \forall i, j, k$$

Esta solución satisface el modelo y las ecuaciones, por lo tanto $\widehat{\mu}_{ijk}$ es el EMV pues como consecuencia del resultado de Birch es la única solución 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, vimos que las ecuaciones puede

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 (IPF)

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

$$\mu_{ij+} = n_{ij+} \quad (1)$$

$$\mu_{i+k} = n_{i+k} \quad (2)$$

$$\mu_{+jk} = n_{+jk} \quad (3)$$

Primero se ajusta de manera que se cumpla (1), luego se multiplica por un factor adecuado como para que se cumpla (2) y finalmente se ajusta para que se cumpla (3). Repitiendo estos pasos partiendo de un punto inicial adecuado se converge a la solución que cumple las tres igualdades.

En la iteración t para este modelo pasaríamos a la del paso $t + 1$:

$$\begin{aligned}\mu_{ijk}^{(t+1/3)} &= \mu_{ijk}^{(t+0/3)} \left(\frac{n_{ij+}}{\mu_{ij+}^{(t+0/3)}} \right) \\ \mu_{ijk}^{(t+2/3)} &= \mu_{ijk}^{(t+1/3)} \left(\frac{n_{i+k}}{\mu_{i+k}^{(t+1/3)}} \right) \\ \mu_{ijk}^{(t+3/3)} &= \mu_{ijk}^{(t+2/3)} \left(\frac{n_{+jk}}{\mu_{+jk}^{(t+2/3)}} \right)\end{aligned}$$

Este método converge más lentamente que N–R, pero 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.