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:

$$\Pi_i \Pi_j \Pi_k \frac{e^{-\mu_{ijk}} \mu_{ijk}^{n_{ijk}}}{n_{ijk}!}$$

Por lo tanto

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

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 (obviando $\mathcal{C}(n_{ijk})$)

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

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:

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

$$
\frac{\partial \ell}{\partial \lambda} \;=\; n - \sum_i \sum_j \sum_k \exp\left[\lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}\right] = 0
$$

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

Luego:

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

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

$\{n_{i++} = \widehat{\mu}_{i++}, i = 1, \ldots, I\}$, $\{n_{+j+} = \widehat{\mu}_{+j+}, j = 1, \ldots, J\}$, $\{n_{++k} = \widehat{\mu}_{++k}, k = 1, \ldots, K\}$ y que

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

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

con lo que $\widehat{\mu}_{i+k} = n_{i+k}$ y $\widehat{\mu}_{+jk} = n_{+jk}$.

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

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

Birch mostró que hay una única solución $\{\widehat{\mu}_{ijk}\}$ que satisface el modelo bajo ciertas condiciones 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.

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) logeneralizó sin hacer 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 estimadas μ_{ijk} de manera de satisfacer las restricciones:

$$\widehat{\mu}_{ij+} \ = \ n_{ij+} \qquad \widehat{\mu}_{i+k} = n_{i+k} \qquad \widehat{\mu}_{+jk} = n_{+jk}$$

Es decir

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

Por lo tanto

$$\begin{aligned} \widehat{\mu}_{ijk} \ &= \ \left(\frac{n_{ij+}}{\widehat{\mu}_{ij+}}\right) \widehat{\mu}_{ijk} \\ \widehat{\mu}_{ijk} \ &= \ \left(\frac{n_{i+k}}{\widehat{\mu}_{i+k}}\right) \widehat{\mu}_{ijk} \end{aligned}$$

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

Veamos a mano un ejemplo sencillo para ver como funciona.

Esto sugiere un método iterativo que comienza con ciertos valores iniciales $\widehat{\mu}_{ijk}^{(o)}$'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

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

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.

Renuniendo todos estos resultados, veamos qué 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 $\widehat{\mu}_{ijk}$ 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 o la deviance:

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

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(n_{ijk} - \widehat{\mu}_{ijk})^2}{\widehat{\mu}_{ijk}}$$

$$= \frac{(716 - 739.9)^2}{739.9} + \ldots \frac{(22 - 20.65)^2}{20.65} = 8.73$$

$$G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk} \log \left[\frac{n_{ijk}}{\widehat{\mu}_{ijk}} \right]$$

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

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:

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

obtenemos los siguientes valores:

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

$(\mathbf{A},\mathbf{BC}),(B,AC),(C,AB)$

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

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 *mala* o *buena*, 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 Mala y Buena y las seis columnas Bajo–N, Bajo–R, Medio–N, Medio–R, Alto–N, Alto–R.

		Adversidad Escolar						
		Bajo		Medio		Alto		
	Riesgo	N	R	N	R	N	R	Total
Conducta	Buena	16	7	15	34	5	3	80
en Clase	Mala	1	1	3	8	1	3	17
	Total	17	8	18	42	6	6	97

Cuadro 8: Valores Observados

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

Los grados de libertad de este modelo son:

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

Los estadísticos de bondad de ajuste dan:

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

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

por lo que no dan evidencias contra el modelo supuesto.

		Adversidad		Escolar				
		Bajo		Medio		Alto		
	Riesgo	N	R	N	R	N	R	Total
Conducta	Buena	14.02	6.60	14.85	34.64	4.95	4.95	80
en Clase	Mala	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

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.

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

Cuadro 10: 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 marginal Adversidad–Riesgo resulta:

La tabla de valores esperados estimados con este modelo es

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

$\widehat{\mu}_{+jk}$		Adversidad			
		Bajo	Medio	Alto	$\widehat{\mu}_{+j+}$
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

que son significativos al 1 %.

Los residuos de Pearson son

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 sobrerepresentados y los de riesgo (R) están subrepresentados.

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