14
2ej.

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Escuela Nacional de Estudios Profesionales ACATLAN

ANALISIS DE DATOS CATEGORICOS MEDIANTE MODELOS LOGLINEALES Y MODELOS LOGIT.

T E S I S
Que para obtener el Título de
A C T U A R I O
p r e s e n t a

EDUARDO MENA SOLORZANO

Asesor de Tesis Dr. Ignacio Méndez

Acatlán Edo. de México 1994

VNIVERSDAD NACIONAL AVPNMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLAN"

DIVISION DE MATEMATICAS E INGENIERIA
PROGRAMA DE ACTUARIA Y M.A.C.

SR. EDUARDO MENA SOLORZANO
Alumno de la carrera de Actuaría
P r e s e n t e .

De acuerdo a su solicitud presentada con fecha 15 de diciembre de 1993, me complace notificarle que esta Jefatura tuvo a bien asignarle el siguiente tema de tesis: "ANALISIS DE DATOS CATEGORICOS MEDIANTE MODELOS LOGLINEALES Y MODELOS LOGIT", el cual se desarrollará como sigue:

INSTRODUCCION
CAP. I Modelos Loglineales.
CAP. II Estimación de los Parámetros.
CAP. III Selección de Modelos.
CAP. IV Modelos Logit
CONCLUSIONES
BIBLIOGRAFIA

Asimismo fué designado como Asesor de Tesis el Dr. Ignacio Méndez Ramírez, Profesor del I.I.M.A.S. -- U.N.A.M.

Ruego a usted tomar nota que en cumplimiento de lo especificado en la Ley de Profesiones, deberá presentar servicio social durante un tiempo mínimo de seis meses como requisito básico para sustentar examen profesional, así como de la disposición de la Coordinación de la Administración Escolar en el sentido de que se imprima en lugar visible de los ejemplares de la tesis el título del trabajo realizado. Esta comunicación deberá imprimirse en el interior de la tesis.

A T E N T A M E N T E
"POR MI RAZA HABLARA EL ESPIRITU"
Acatlán, Edo. Méx. abril 29 de 1994.

ACT. LAURA MA. RIVERA BECERRA
Jefe del Programa de Actuaría
y M.A.C.

E.N.E.P. ACATLAN

LMRB'cg.

PREFACIO

El presente trabajo trata acerca de algunas técnicas apropiadas para el análisis estadístico con varias variables de tipo categórico, centrándose en particular en el planteamiento de modelos que resultan lineales en una escala logarítmica. Los tópicos que se abordan se refieren al entendimiento de los modelos como una prueba de hipótesis estadística, a la estimación de los parámetros de los mismos y la selección de un modelo o hipótesis apropiado a las características estructurales de un conjunto de datos determinado.

Las ideas expuestas a continuación no representan de ninguna manera una investigación original, se pretende acaso a lo largo del presente documento reunir y adaptar algunos de los desarrollos de diversos autores que han ahondado en la especificación de técnicas para el análisis multivariado discreto, obteniendo con ello para el presente trabajo una tónica de carácter teórico donde no se realiza investigación aplicada a situación alguna o fenómeno particular.

El enfoque de la obra abarca fundamentalmente el entendimiento y ajuste de modelos loglineales de tipo jerárquico para tablas completas, aunque se introduce en una clase de modelos de mayor amplitud denominada Modelos Lineales Generalizados donde se plantea una mayor gama de posibilidades de modelos incluso los de carácter no jerárquico y dentro de la cual los modelos loglineales son apenas un subconjunto de alternativas posibles. Se realiza finalmente una somera incursión hacia algunos modelos que han sido utilizados para el análisis de datos categóricos donde destacan los modelos logit o logísticos por su relación con los modelos loglineales.

Reconocimientos

Agradezco en primera instancia al Dr. Ignacio Mendez por aceptar la dirección del presente trabajo de tesis y por sus valiosas aportaciones para la mejora del mismo.

A Celanese Mexicana por las facilidades bindadas para llevar a efecto este proyecto, a Conrado, José Luis y Gustavo por su apoyo y comprensión.

Así mismo, agradezco a mis sinodales y en particular a Ricardo Aparicio por sus minuciosas observaciones y sugerencias que hicieron que mejorara la calidad de este trabajo.

A Refugio Rivera por su amistad y tiempo brindado en el desarrollo del trabajo.

A mi familia por su paciencia y apoyo incondicional.

A MIS PADRES:

SIN QUIENES MI FORMACIÓN
PROFESIONAL NO HUBIESE
SIDO POSIBLE.

PARA LUCY, LALO Y VINI:

EN QUIENES FUNDAMENTO MIS METAS
Y CON QUIENES CIFRO MIS MAYORES
ANHELOS

A MIS HERMANOS:

POR SU APOYO Y COMPRENSIÓN,
Y POR CONSTRUIR UNIDA
NUESTRA VIDA

A MIS AMIGOS:

LOS HERMANOS DEL ALMA
CON QUIENES HE COMPARTIDO
LOS ALTIBAJOS DE LA VIDA

A DANIEL BUQUET:

SER CABAL Y COMPROMETIDO CON
SU TIEMPO, EJEMPLO DE VIDA
PARA TODA LA HUMANIDAD.

C O N T E N I D O

INTRODUCCION

Las técnicas estadísticas que se utilizan para el análisis de datos de tipo categórico han evolucionado enormemente en los últimos treinta años, partiendo desde el establecimiento de algunas medidas de asociación entre dos variables hasta el planteamiento actual de modelos multivariados que implican hipótesis más complejas acerca de las características estructurales de los datos, puede decirse que se cuenta hoy en día con herramientas equiparables a las desarrolladas para el análisis de regresión o el análisis de varianza.

El análisis multivariado discreto dirigido en particular a variables de tipo categórico implica, por una parte, el estudio de los efectos conjuntos que producen más de dos variables, así como las interacciones de un subconjunto de ellas en presencia de las restantes variables y, en segunda instancia, implica el manejo de variables que son medidas o agrupadas en categorías y que usualmente el registro de observaciones se realiza en las denominadas tablas cruzadas o tablas de contingencia.

El presente material hace especial énfasis en una clase de modelos que permite realizar análisis multivariado con variables de tipo categórico denominada modelos loglineales; se detalla en la construcción de los mismos, su interpretación, así como el cálculo de sus parámetros y en las medidas de bondad de ajuste a los datos. Adicionalmente, se introduce el enfoque de los modelos lineales generalizados en el caso categórico y finalmente se muestran algunas técnicas que permiten aprovechar la información cuando alguna de las variables involucradas es de respuesta, con particular referencia al planteamiento de modelos logit.

CONCEPTOS PRELIMINARES

Uno de los primeros problemas a los que se enfrenta un investigador cuando decide utilizar el análisis estadístico para el estudio de una situación determinada, es el elegir una técnica apropiada. Para tal efecto es importante considerar dos aspectos de las variables que habrán de utilizarse para el registro de las observaciones.

Por una parte conviene distinguir el tipo de variables a tratar, en cuanto a si éstas son continuas o discretas y mejor aún en lo referente a la escala de medición de las mismas sea esta nominal, ordinal, intervalar o de razón. En una segunda instancia, es importante distinguir la relación entre las variables en cuanto a plantear si se tienen variables de respuesta y de explicación en el estudio (variables dependientes e independientes).

En el cuadro I.1 sin pretender ser exhaustivo se ubican las principales técnicas de análisis estadístico y se hace especial énfasis en las técnicas para variables de tipo categórico.

Cuadro I.1: Técnicas estadísticas en el estudio de asociaciones

VARIABLES RESPUESTA	*EXPLICACION*	*TECNICA RECOMENDADA*
Continua	*Continuas*	*Análisis de Regresión*
Continua	*Categóricas*	*Análisis de Varianza*
Continua	*Ambas*	*Análisis de Covarianza* *Regresión con variables indicadoras*
Categórica	*Continuas*	*Análisis Discriminante* *Modelos de Regresión Logística*
Categórica	*Categóricas*	*Modelos Loglineales* *Modelos Logit* *Modelos Gráficos* *Regresión Logística* *Regresión Logística Generalizada*
Categóricas	*Ambas*	*En los casos de respuesta con más de dos categorías y variables de explicación continuas el problema suele redefinirse intercambiando los roles entre variables.*

NATURALEZA DE LAS VARIABLES CATEGORICAS

En la escala de medición nominal los individuos de una categoría difieren de los de la otra en una escala que no revela en sí misma una magnitud, es decir, no se cuenta con algún orden para clasificar una categoría con respecto a la otra en el sentido de determinar que una es mayor que la otra, de aqui que se denomine a esta clase de variables como cualitativas, por ejemplo, marcas de detergentes o de automóviles.

En la escala de medición ordinal en contraste con la nominal es posible establecer una magnitud de orden (mayor o menor) para cada nivel o categoría de la variable. A su vez, la diferencia entre una variable de tipo ordinal y una intervalar radica en la claridad para establecer la magnitud de orden entre las categorías, requiriéndose en el caso intervalar de conocer una distancia absoluta para medir la diferencia (ejemplo de variable ordinal: niños, adolescentes y adultos; ejemplo de intervalar: edad en años cumplidos, 0,1,2, etc.).

Las técnicas abordadas en el presente trabajo son apropiadas para el uso de variables nominales, ordinales, intervalares que cuenten con relativamente pocos valores y con variables de intervalo continuo agrupadas en un pequeño número de categorías.

FORMAS DE ANALISIS PARA TABLAS DE CONTINGENCIA

En términos generales se cuenta con tres caminos para describir tablas de contingencia:

1) Modelos que no requieren distinguir entre variables de respuesta y de explicación.

2) Modelos que requieren la distinción.

3) Medidas de asociación.

La estrategia (3) describe algunos aspectos importantes de de relaciones aisladas entre las variables de la tabla, pero tales medidas no están basadas en un modelo para los datos, además de no describir completamente las relaciones de asociación múltiple entre todas las variables o entre subconjuntos de ellas en presencia de las restantes.

En la estrategia (1) los modelos loglineales constituyen una alternativa para un vasto planteamiento de hipótesis acerca de las características estructurales de los datos donde no se establecen relaciones de causa-efecto.

En la estrategia (2) se cuenta con los modelos de dósis-respuesta, los modelos logit y modelos logit acumulativos entre otras técnicas disponibles.

Es importante resaltar que para una misma tabla de contingencia, los modelos loglineales presentan diferentes patrones de asociación que los modelos logit cuando el número de categorías en la variable de respuesta excede de dos.

En el presente trabajo se hará énfasis principalmente en la estrategia (1) y se introducirá en modelos para la estrategia (2), quedando fuera del alcance del mismo las medidas de asociación debido a que no representan una alternativa de modelización de datos.

PROCESO PARA EL ESTABLECIMIENTO DE MODELOS

Cuando se cuenta con un conjunto de datos y se desea establecer un modelo adecuado para los mismos, se sugiere considerar los siguientes tres pasos generales:

1) Especificación de las ecuaciones plausibles y distribuciones de probabilidad para describir las principales características de la variable de respuesta (planteamiento de un modelo).

2) Estimación de los parámetros del modelo.

3) Realizar inferencia estadística mediante la prueba de hipótesis que consideren la adecuacía del ajuste del modelo a los datos.

En el presente trabajo se considera el análisis de datos agrupados en tablas de contingencia, para lo cual, se realiza el planteamiento de modelos loglineales (capítulo I) y modelos logit (capítulo IV) como formas de especificación del primer aspecto mencionado.

El segundo y tercer aspectos son abordados desde el punto de vista de los modelos lineales generalizados, lo que permite su aplicación tanto a los modelos loglineales como a los modelos logit (capítulos II y III).

BOSQUEJO HISTORICO

El análisis de datos categóricos mediante el uso de tablas de contingencia se remonta a finales de la última decada del siglo pasado cuando Quetelet estableció la primera medida de asociación que se ha registrado. A principios de siglo, Pearson y Yule formularon los primeros adelantos, Yule en 1912 aportó la medida de asociación denominada razón de productos cruzados, mientras que Pearson construyó la popular prueba de independencia para dos variables basada en la distribución Ji-cuadrada. Fisher en 1922, sin proponer una prueba concreta, estableció la necesidad de considerar efectos denominados de interacción de segundo orden, es decir, planteó la necesidad de realizar análisis multivariado.

Bartlett en 1935 a partir de las premisas de Fisher y de la medida propuesta por Yule definió el concepto de interacción de segundo orden en una tabla de 2x2x2 estableciendo con ello las bases del análisis multivariado; sin embargo, las limitaciones técnicas de la época dificultaron la realización de cálculos y con ello el desarrollo de técnicas multivariadas.

Con la llegada de las computadoras, el análisis multivariado tomó un nuevo aire y en las décadas de los 40 y 50 aparecieron dentro de la literatura estadística los modelos aditivos y los modelos resultantes de particiones de la Ji-cuadrada. Destacan los trabajos* de Berkson en 1944, de Neyman en 1949, de Cochran en 1954 y de Lancaster en 1951. No obstante, estos modelos han perdido relevancia y se han convertido en un vasto conjunto de medidas de asociación en desuso hoy en día.

---

* Remitase a la bibliografía de [Bishop et.al.,1975], [Fienberg,1977] y [Agresti, 1984 y 1990]

A mediados de la década de los 50 surgieron algunos métodos de estimación, máxima verosimilitud, mínima Ji-cuadrada modificada e información mínima discriminante, a partir de las primicias de Lancaster y Roy con su grupo de alumnos de Carolina del Norte.

Son de apreciarse entre otros de la época, los trabajos* de Goodman y Kruskal en 1954, de Roy y Mitra en 1956, de Cox en 1958 y de Mantel y Haenzel en 1959.

Los mejores avances para el análisis de tablas de contingencia multidimensionales se ubicaron en la década de los 60 donde destacan los trabajos de Birch en 1963 y 1964, de Goodman en 1968, de Mosteller en 1968, de Grizzle, Starmer y Kock en 1969. En esta época aparecieron los principios para la obtención de estimadores de máxima verosimilitud y mínima Ji-cuadrada modificada, sentandose las bases para el desarrollo de los modelos loglineales.

En la década de los 70 se plantearon avances en el estudio de modelos lineales generalizados y se mejoró en las técnicas de obtención de estimadores mediante ajuste iterativo proporcional y la aplicación del método derivado de Newton-Raphson, conviene señalar aquí los trabajos* de Goodman en 1970 y 1979, de Haberman en 1974 y 1979, de Nelder y Wedderburn en 1972, de Bishop, Fienberg y Holland en 1975 y de Upton en 1978.

En la década de los 80 y principios de los 90 se mejoran las técnicas para la determinación de la bondad de ajuste de los modelos, se elaboran paquetes computacionales muy completos para el manejo de datos categóricos basados en la potencialidad de los modelos lineales generalizados y se mejoran los modelos aprovechando la información para variables de tipo ordinal, aquí resaltan los trabajos de Mc Cullagh en 1980, de Agresti en 1984 y 1990, de Goodman en 1986 y Dobson en 1990.

---

* Remitase a la bibliografía de [Bishop et.al.,1975], [Fienberg,1977] y [Agresti, 1984 y 1990]

ORGANIZACION DE LOS CAPITULOS

El presente trabajo consta de cuatro capítulos, el primero de ellos, se refiere a los modelos loglineales y el cuarto a técnicas apropiadas cuando se requiere la distinción entre variables de respuesta y explicación. El segundo y tercer capítulos aunque hacen énfasis en los modelos loglineales ofrecen un marco teórico aplicable también a los modelos del cuarto capítulo.

Capítulo I: Modelos loglineales

Se define el concepto de tablas de contingencia y se consideran los modelos muestrales aplicables para la recolección de datos; posteriormente, se describe la prueba de independencia de Pearson y se analiza la razón de productos cruzados para dar base a la construcción del modelo loglineal en dos dimensiones; se ofrece la interpretación de tales modelos y el cálculo de grados de libertad. Finalmente, se procede a la generalización a tres y más dimensiones mediante la definición del concepto de jerarquía, la interpretación de los parámetros en el modelo y el plegamiento (colapsamiento) de tablas de contingencia.

Capítulo II: Estimación de los parámetros

Se presentan algunos estadísticos de prueba que permiten la construcción de estimadores, se ofrece una introducción a la teoría de los modelos lineales generalizados, para sentar las bases en la obtención de estimadores de máxima verosimilitud y de mínimos cuadrados. Se presentan además, algunos procedimientos empleados para el cálculo de los estimadores tales como el Ajuste Iterativo Proporcional y el derivado de Newton-Raphson.

Capítulo III: Selección de modelos

Se analizan las distribuciones muestrales para estadísticos de máxima verosimilitud en un modelo lineal generalizado, y se procede a la construcción de la devianza como el estadístico a emplear en la determinación de la adecuacía de un modelo. Se plantea el cálculo de grados de libertad y se aborda el tema de bondad de ajuste interna del modelo mediante el análisis de residuos, la asignación de la magnitud de una interacción y la detección de datos discrepantes. Finalmente se presentan algunas consideraciones para la prueba de hipótesis tales como cálculos iniciales, procedimientos por pasos "stepwise" y otros aspectos prácticos en la selección de un modelo.

Capítulo IV: Modelos logit

Se consideran los modelos de dosis-respuesta como una clase de modelos lineales generalizados y se presentan en particular los modelos uniforme, probit, logit y valor extremo. Posteriormente, se analizan los modelos de regresión logística y la relación entre los modelos logit y loglineal. Finalmente se considera la estimación de máxima verosimilitud, la devianza, así como la bondad de ajuste de estos modelos.

CAPITULO I: MODELOS LOGLINEALES

Los modelos loglineales constituyen una estructura teórica acerca de los valores observados de manera simultánea de dos o más variables de tipo categórico, donde los parámetros del modelo representan los efectos que tienen las variables o combinaciones de ellas sobre la determinación de los valores tomados por las observaciones. La complejidad intrínseca en los datos es reflejada por el número de parámetros en el modelo loglineal que describe adecuadamente la estructura de la tabla de contingencia, así entonces, cuando la estructura es simple habrá pocos parámetros en el modelo. Esta idea puede parecer un tanto obscura en este momento, no obstante, los conceptos implicados serán detallados en las secciones subsecuentes.

En una tabla de contingencia bidimensional la idea de modelo loglineal no es del todo nueva, se encuentra implícita en la prueba de independencia de Pearson, así los modelos loglineales en varias dimensiones pueden ser considerados como una adecuada generalización de la prueba citada.

En el presente trabajo para la representación de los modelos se utilizará la notación propuesta por Birch y una terminología análoga a la utilizada en el Análisis de Varianza para definir conceptos tales como "interacción", término referido a la asociación entre variables de una tabla de contingencia, "interacción de primer orden" asociación entre pares de variables, "interacción de segundo orden" asociación entre tripletas de variables y así sucesivamente.

En el presente capítulo se describirán las características más relevantes acerca de la estructura de los modelos, dejando el problema de la estimación de parámetros y selección de un mejor modelo para capítulos posteriores. El capítulo inicia abordando el concepto de tabla de contingencia y la notación utilizada para describir una tabla bidimensional, la cual puede generalizarse facilmente a más dimensiones; seguidamente, se hace referencia a los modelos muestrales que suelen utilizarse para la obtención de los datos de una tabla de contingencia. Posteriormente, se introduce el concepto de independencia en dos dimensiones y la prueba al respecto sugerida por Pearson, así como la medida de Yule denominada razón de productos cruzados.

Se continua con la construcción del modelo loglineal bidimensional a partir del concepto de independencia y se ofrece una interpretación de dicho modelo. Finalmente se proporcionan las propiedades estructurales de los modelos de tres y más dimensiones.

Conviene resaltar que unicamente se estudian las propiedades teóricas de los modelos de tipo jerárquico debido a que los modelos que no son de esta clase carecen de una interpretación adecuada; sin embargo, la teoría de los modelos lineales generalizados soporta perfectamente el manejo de modelos de tipo no jerárquico.

1.1 TABLAS DE CONTINGENCIA

En el análisis de datos categóricos es posible clasificar a los elementos de una población o muestra en diferentes clases según las características que se deseen estudiar. Cada individuo es descrito por un conjunto determinado de atributos y al examinar los atributos para cada uno de ellos sus frecuencias son registradas en las casillas de una tabla de contingencia.

Una tabla de contingencia es entonces un conjunto ordenado de casillas donde cada una de estas refleja el conteo de casos en que ocurre simultaneamente que los individuos cumplen con un determinado atributo para cada una de las variables involucradas en la tabla. Un ejemplo de tabla de contingencia de 2x3 se dá en la tabla 1.1.

Tabla 1.1: Muestra de individuos por sexo y edad

	0-12 años	*13-17 años*	*18 ó más años*
Masculino	*327*	*278*	*417*
Femenino	*335*	*270*	*423*

Se dice que la tabla de contingencia es de 2x3 debido a que en ella participan dos variables; una de ellas (sexo) con dos categorías y la otra (edad) con tres.

Para obtener una adecuada clasificación es preciso que las variables posean la característica de tener categorías exhaustivas y mutuamente excluyentes, es decir, para cada unidad de análisis de la muestra o población debe existir una única categoría de cada variable en donde pueda ser registrada.

1.1.1 NOMENCLATURA

La forma general para una tabla de contingencia en dos dimensiones está dada en la tabla 1.2

Tabla 1.2: Forma general de una tabla bidimensional

		Variable 2 1	2	...	J	Total
Variable 1	1	x_{11}	x_{12}		x_{1J}	x_{1+}
	2	x_{21}	x_{22}		x_{2J}	x_{2+}
	...					
	I	x_{I1}	x_{I2}		x_{IJ}	x_{I+}
Total		x_{+1}	x_{+2}		x_{+J}	$x_{++}= N$

La tabla 1.2 se denomina también tabla de contingencia IxJ, donde la clasificación de los elementos de una muestra o población con N observaciones se realiza de acuerdo a dos variables, una de ellas con I categorías y la segunda con J.

La frecuencia observada (x_{ij}) es el número de individuos cuyos atributos corresponden a la i-ésima categoría de la primer variable con la j-ésima categoría de la segunda variable con $i=1,2,...,I$ y $j=1,2,...,J$.

El conteo total para una categoría de alguna variable se denomina marginal, así x_{i+} representa la marginal para la i-ésima categoría de la primer variable y x_{+j} lo es para la j-ésima categoría de la segunda variable donde se cumple lo siguiente:

$$x_{i+} = \sum_{j=1}^{J} x_{ij} \quad \text{para } i=1,2,...,I \qquad x_{+j} = \sum_{i=1}^{I} x_{ij} \quad \text{para } j=1,2,...J$$

$$N = x_{++} = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij} \quad \text{para } i=1,2,\ldots,I \text{ y } j=1,2,\ldots,J$$

Bajo un muestreo multinomial se considera que cada individuo elegido al azar tiene una probabilidad p_{ij} de caer en la casilla (i,j), $i=1,2,\ldots,I$ $j=1,2,\ldots J$ y debido a que las categorías de las variables son exhaustivas y mutuamente excluyentes resulta:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = \sum_{i=1}^{I} p_{i+} = \sum_{j=1}^{J} p_{+j} = 1$$

donde:

$$p_{i+} = \sum_{j=1}^{J} p_{ij} \quad , \quad i=1,2,\ldots,I$$

$$p_{+j} = \sum_{i=1}^{I} p_{ij} \quad , \quad j=1,2,\ldots,J$$

Los correspondientes conteos esperados para cada casilla (m_{ij}) se determinan por $m_{ij} = Np_{ij}$. Puede verse que $m_{ij} = \mathcal{E}(x_{ij})$ bajo el supuesto de que las frecuencias observadas (x_{ij}) tienen una distribución multinomial con parámetros p_{ij} , $i=1,2,\ldots,I$, $j=1,2,\ldots,J$ (ver [Mood & Graybill,1963], cap. 3).

1.1.2 TABLAS DE CONTINGENCIA COMPLETAS

Se dice que una tabla de contingencia es completa si hay una probabilidad diferente de cero de que un conteo ocurra en toda casilla del arreglo.

Existen dos posibilidades de que en una tabla se presente la ausencia de conteos, una denominada cero estructural, es decir, que debido a razones lógicas de diseño una casilla dada deba permanecer siempre vacia, ej. hombres con problemas de aborto. Otra llamada ceros aleatorios, los cuales son debidos exclusivamente a causa del azar (baja probabilidad de ocurrencia y/o tamaño de muestra muy reducido), por ejemplo personas mayores de 100 años en una localidad pequeña.

Se considera una tabla de contingencia incompleta si presenta casillas con ceros estructurales.

En una tabla completa con ceros aleatorios, en general, pueden obtenerse los conteos esperados para las casillas.

En el presente trabajo se consideran unicamente tablas completas, dado que los modelos requeridos cuando se presentan ceros estructurales en la tabla (tablas incompletas) implican un mayor grado de complejidad teórica. Los lectores interesados en el manejo adecuado de este tipo de tablas pueden remitanse a [Bishop et.al.,1975] cap 5.

1.2 ESQUEMAS MUESTRALES

Para la recolección aleatoria de una muestra cuyo registro se realizará en una tabla de contingencia, se cuenta con tres modelos fundamentales, Poisson, Multinomial y Producto-multinomial.

1) Poisson

Se considera un proceso Poisson para cada una de las casillas de la tabla durante un periodo fijo de tiempo, sin tener un conocimiento previo acerca del número total de observaciones que se obtendrán. Cada proceso conducirá a una frecuencia en la correspondiente casilla.

2) Multinomial

Se considera una muestra fija de tamaño N y cada elemento es clasificado entonces de acuerdo a los valores de las variables.

3) Producto-multinomial

Se determinan previamente los valores de las marginales para alguna(s) de las variables de la tabla, es decir, se consideran muestras de tamaño fijo para cada una de las categorías de las variables elegidas y se clasifican los elementos de acuerdo a los valores observados en las categorías del resto de las variables.

Desde el punto de vista de tener o no variables de respuesta además de variables de explicación, se tienen dos situaciones en general en que se recomienda el uso de alguno de los esquemas señalados:

1) Las variables son todas de respuesta.

En este caso se recomienda utilizar el modelo Poisson o Multinomial dependiendo de las características del estudio.

2) Se cuenta con al menos una variable de explicación.

En esta situación se recomienda el uso del modelo producto multinomial donde las marginales fijas corresponden a las categorías de las variables de explicación.

En capítulos Posteriores se mostrará que los tres esquemas muestrales conducen a los mismos estimadores de máxima verosimilitud para los conteos esperados de las casillas.

1.3 PRUEBA DE INDEPENDENCIA DE PEARSON

En una tabla de contingencia de IxJ un aspecto importante es determinar si las variables son o no independientes. Suponga que en la población de donde fué extraída la muestra, la probabilidad de que una observación pertenezca al cruce de la i-ésima categoría de la primer variable con la j-ésima de la segunda variable es p_{ij} , p_{i+} representa la probabilidad marginal en la población de que la observación pertenezca a la i-ésima categoría de la primer variable y similarmente p_{+j} representa la probabilidad marginal de pertenencia a la j-ésima categoría de la segunda variable. Entonces si las variables son independientes (estocásticamente) por la ley de probabilidad debe ocurrir que:

$$p_{ij} = p_{i+}p_{+j} \qquad (1.3\text{-}1)$$

En términos de las frecuencias esperadas de las casillas, la independencia puede expresarse como:

$$m_{ij} = Np_{i+}p_{+j} \qquad (1.3\text{-}2)$$

Si se Considera una distribución multinomial para las casillas de la tabla, los parámetros p_{i+} y p_{+j} pueden estimarse por máxima verosimilitud, puede probarse además, que se obtienen los mejores estimadores bajo ese criterio (ver [Hogg & Craig,1968]). Las expresiones para tales estimadores son

$$\hat{p}_{i+} = x_{i+} / N \qquad y \qquad \hat{p}_{+j} = x_{+j} / N \tag{1.3-3}$$

Mediante estas expresiones puede estimarse entonces la frecuencia esperada para cada una de las casillas. A partir de (1.3-2) se deduce que:

$$\hat{m}_{ij} = N\, \hat{p}_{i+}\, \hat{p}_{+j} = x_{i+}\, x_{+j} / N \tag{1.3-4}$$

Cuando las variables son independientes las frecuencias estimadas y las observadas no deben diferir significativamente.

La prueba de independencia de Pearson está basada en este hecho y considera la distribución χ^2 para la construcción del estadístico de prueba. La prueba de hipótesis de independencia es entonces:

H_0: $p_{ij} = p_{i+}\, p_{+j}$ para todo i,j $i=1,2,...,I$ y $j=1,2,...J$

v.s. (1.3-5)

H_1: Otra alternativa.

El estadístico de prueba que sugiere Pearson es:

$$\lambda = \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \hat{m}_{ij})^2 / \hat{m}_{ij} \tag{1.3-6}$$

Si el estadístico λ calculado con los valores observados, resulta menor que el valor de la distribución teórica de una χ^2 con $(I-1)(J-1)$ grados de libertad y un nivel de significancia α para la prueba, se dice entonces que no se cuenta con evidencia suficiente para descartar el supuesto de la hipótesis nula, es decir, de la hipótesis de independencia entre las variables.

1.4 RAZON DE PRODUCTOS CRUZADOS

Una medida acerca de la independencia de dos variables en una tabla de contingencia de 2x2 es la denominada **razón de productos cruzados o razón de momios (odds ratio).**

$$\alpha = p_{11}\ p_{22}\ /\ p_{12}\ p_{21} \qquad (1.4\text{-}1)$$

Esta medida posee las siguientes propiedades:

1) Es invariante ante el intercambio del orden de las variables en la tabla (cambio de renglones por columnas).

2) Es invariante ante la multiplicación de constantes en renglones y/o columnas. Suponga que se multiplican las probabilidades en el renglón 1 por $r_1 > 0$, en el renglón 2 por $r_2 > 0$, en la columna 1 por $c_1 > 0$ y en la columna 2 por $c_2 > 0$, la nueva razón de productos cruzados α' sería:

$$\alpha' = (r_1c_1p_{11})(r_2c_2p_{22})\ /\ (r_1c_2p_{12})(r_2c_1p_{21})$$
$$\alpha' = p_{11}p_{22}/p_{12}p_{21} = \alpha$$

3) La razón de productos cruzados puede interpretarse como el cociente de "momios"

$$\alpha = (p_{11}/p_{12})\ /\ (p_{21}/p_{22})$$

Donde p_{11}/p_{12} es el momio de pertenecer a la primer columna dado que se está en el primer renglón y p_{21}/p_{22} es el momio correspondiente al segundo renglón.

4) α puede ser utilizado en tablas de IxJ (y multidimensionales) particionando la tabla en subtablas de 2x2.

5) Si $\alpha = 1$ en una tabla de 2x2 o todas las posibles $\alpha=1$ en una tabla multidimensional, las variables de la tabla son independientes.

6) El estimador de máxima verosimilitud $\hat{\alpha}$ está dado por:

$$\hat{\alpha} = x_{11}\ x_{22}\ /\ x_{12}\ x_{21}$$

7) $|\ln(\alpha)|$ puede tomarse como una medida de alejamiento de la independencia donde si las variables son independientes $\ln(\alpha) = 0$ y es invariante ante el intercambio de renglones y/o columnas.

En efecto, si α' es la razón de productos cruzados para una tabla, al realizar el intercambio de renglones (columnas) resulta que $\alpha' = 1/\alpha$ pero

$$|\ln(\alpha')| = |\ln(1/\alpha)| = |-\ln(\alpha)| = |\ln(\alpha)|$$

8) $\ln(\hat{\alpha})$ tiene distribución asintótica Normal con media $\ln(\alpha)$ y $\mathrm{var}[\ln(\hat{\alpha})] = (1/x_{11}) + (1/x_{12}) + (1/x_{21}) + (1/x_{22})$

1.5 MODELO BIDIMENSIONAL

En la sección 1.3 se trató acerca de la prueba propuesta por Pearson, ahora se presenta una forma alternativa denominada modelo loglineal para la prueba de hipótesis de independencia. Primeramente se construye un modelo loglineal en base a la hipótesis de independencia estocástica (modelo loglineal con ausencia del término de interacción de primer orden) y posteriormente se ofrece una interpretación de los posibles modelos en dos dimensiones.

Si se considera una tabla de contingencia bidimensional la hipótesis de independencia implica que para cada casilla de la tabla la expresión para la probabilidad de que un individuo pertenezca a la misma, queda dada por la siguiente expresión:

$$p_{ij} = p_{i+}\, p_{+j} \qquad (1.5\text{-}1)$$

Esta relación especifica una estructura particular para los datos e indica que en la población la probabilidad de que una observación sea registrada en la casilla ij-ésima es simplemente el producto de las probabilidades marginales. Tomando logaritmo natural en (1.5-1) resulta

$$\ln p_{ij} = \ln p_{i+} + \ln p_{+j} \qquad (1.5\text{-}2)$$

Si se considera que $m_{ij} = N\, p_{ij}$, $m_{i+} = N\, p_{i+}$ y $m_{+j} = N\, p_{+j}$ resulta que

$$\ln m_{ij} = \ln m_{i+} + \ln m_{+j} - \ln N \qquad (1.5\text{-}3)$$

sumando sobre i en (1.5-3) resulta

$$\sum_{i=1}^{I} \ln m_{ij} = \sum_{i=1}^{I} \ln m_{i+} + I \ln m_{+j} - I \ln N \qquad (1.5\text{-}4)$$

sumando sobre j en (1.5-3) se tiene

$$\sum_{j=1}^{J} \ln m_{ij} = J \ln m_{i+} + \sum_{j=1}^{J} \ln m_{+j} - J \ln N \qquad (1.5\text{-}5)$$

sumando sobre i y j la misma expresión resulta

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \ln m_{ij} = J \sum_{i=1}^{I} \ln m_{i+} + I \sum_{j=1}^{J} \ln m_{+j} - IJ \ln N \qquad (1.5\text{-}6)$$

En base a cálculos algebraicos simples se puede probar que la ecuación (1.5-3) puede reescribirse de la siguiente forma

$$\ln m_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} \qquad (1.5\text{-}7)$$

donde

$$\mu = (1/IJ) \sum_{i} \sum_{j} \ln m_{ij} \qquad (1.5\text{-}8)$$

$$\mu_{1(i)} = (1/J) \sum_{j} \ln m_{ij} - (1/IJ) \sum_{i} \sum_{j} \ln m_{ij} \qquad (1.5\text{-}9)$$

$$\mu_{2(j)} = (1/I) \sum_{i} \ln m_{ij} - (1/IJ) \sum_{i} \sum_{j} \ln m_{ij} \qquad (1.5\text{-}10)$$

con objeto de facilitar la notación se define $\ell_{ij} = \ln m_{ij}$ para $i=1,2,\ldots,I$; $j=1,2,\ldots,J$. $\ell_{i+} = \sum_{j} \ln m_{ij}$, $\ell_{+j} = \sum_{i} \ln m_{ij}$ y $\ell_{++} = \sum_{i} \sum_{j} \ln m_{ij}$ donde resulta que las ecuaciones (1.5-7) a (1.5-10) pueden reescribirse de la siguiente forma:

$$\ell_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} \qquad (1.5\text{-}11)$$

donde

$$\mu = (1/IJ)\ \ell_{++} \qquad (1.5\text{-}12)$$
$$\mu_{1(i)} = (1/J)\ \ell_{i+} - (1/IJ)\ \ell_{++} \qquad (1.5\text{-}13)$$
$$\mu_{2(j)} = (1/I)\ \ell_{+j} - (1/IJ)\ \ell_{++} \qquad (1.5\text{-}14)$$

la ecuación (1.5-11) representa un modelo loglineal bajo el supuesto de independencia de las variables. Utilizando la terminología análoga al Análisis de Varianza acerca de los parámetros, se denomina a μ "media general", a $\mu_{1(i)}$ "efecto principal de la variable 1" y a $\mu_{2(j)}$ "efecto principal de la variable 2".

Examinando las ecuaciones (1.5-13) y (1.5-14) puede deducirse fácilmente que

$$\sum_{i} \mu_{1(i)} = 0 \qquad (1.5\text{-}15)$$

$$\sum_{j} \mu_{2(j)} = 0 \qquad (1.5\text{-}16)$$

En la obtención del modelo especificado en (1.5-11) se han involucrado las frecuencias teóricas m_{ij}. En la práctica es necesario primeramente estimarlas y seguidamente probar la bondad de ajuste del modelo lo cual puede verse en los capítulos II y III.

Un modelo de interés que resulta de extender el modelo loglineal (1.5-11) para el caso en el que las variables no son independientes es el siguiente:

$$\ell_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{12(ij)} \qquad (1.5\text{-}17)$$

donde μ, $\mu_{1(i)}$ y $\mu_{2(j)}$ están dadas por las mismas expresiones (1.5-12) a (1.5-14) y por consiguiente se obtiene

$$\mu_{12(ij)} = \ell_{ij} - (\mu + \mu_{1(i)} + \mu_{2(j)})$$
$$\mu_{12(ij)} = \ell_{ij} - (\ell_{i+} / J) - (\ell_{+j} / I) + (\ell_{++} / IJ) \qquad (1.5\text{-}18)$$

$\mu_{12(ij)}$ representa el efecto de interacción entre los niveles i y j de las variables 1 y 2 respectivamente, es también conocido como efecto de interacción de primer orden o efecto conjunto de dos factores.

De (1.5-18) se deduce facilmente que

$$\sum_{j} \mu_{12(ij)} = 0 \qquad (1.5\text{-}19)$$

$$\sum_{i} \mu_{12(ij)} = 0 \qquad (1.5.20)$$

En términos de los parámetros de interacción, la hipótesis de independencia especifica que $\mu_{12(ij)} = 0$ para todos los valores de i y de j. Probar la hipótesis de independencia es por tanto equivalente a probar que todos los términos de interacción en (1.5-17) son cero, o en otras palabras, que el modelo por (1.5-11) proporciona un ajuste adecuado a los datos.

1.5.1 GRADOS DE LIBERTAD

El número de parámetros independientes en un modelo es conocido bajo el nombre de grados de libertad, las restricciones (1.5-15), (1.5-16), (1.5-19) y (1.5-20) implican una reducción en el número de parámetros para cada término μ en el modelo loglineal. Así por ejemplo, hay I valores diferentes de $\mu_{1(i)}$, uno para cada categoría de la primer variable, pero la restricción (1.5-15) reduce el número de parámetros independientes a I-1, similarmente la restricción (1.5-16) reduce a J-1 los parámetros independientes asociados con $\mu_{2(j)}$ y las restantes resticciones conducen a que el arreglo $\{\mu_{12(ij)}\}$ con un total de IxJ parámetros tenga solo (I-1)(J-1) independientes.

Tabla 1.5.1-1: Grados de libertad en el modelo bidimensional

μ-término	*Grados de libertad*
μ	*1*
μ_1	*(I-1)*
μ_2	*(J-1)*
μ_{12}	*(I-1) (J-1)*
Total	*IJ*

Puede notarse que el número total de parámetros independientes es igual al número de casillas en la tabla.

1.5.2 MODELO SATURADO

Un modelo loglineal en cualquier dimensión se le denomina **modelo saturado** cuando el número de parámetros independientes presentes en el modelo es igual al número de casillas en la tabla. Es claro que el modelo (1.5-17) es el modelo saturado para una tabla bidimensional.

Si se considera una tabla de contingencia bidimensional, en general, son dos los modelos de interés:

1) Modelo saturado: implica la existencia de una relación entre las variables de la tabla.

2) Modelo de independencia de variables (modelo con $\mu_{12} = 0$)

1.5.3 MODELO DESCRITO POR PROBABILIDADES

El modelo descrito por los conteos esperados difiere en un solo término de un modelo descrito por probabilidades. En efecto, se tiene que $m_{ij} = N\, p_{ij}$ y de aquí se deduce que

$$
\begin{aligned}
\ln p_{ij} &= \ln m_{ij} - \ln N \\
&= (\mu - \ln N) + (\mu_{1(i)} + \mu_{2(j)} + \mu_{12(ij)}) \\
&= \mu' + \mu_{1(i)} + \mu_{2(j)} + \mu_{12(ij)} \qquad (1.5.3\text{-}1)
\end{aligned}
$$

La media general es el único término afectado pero las ecuaciones (1.5-12) a (1.5-14) y (1.5-18) para determinar los valores de $\mu_{1(i)}$, $\mu_{2(j)}$ y $\mu_{12(ij)}$ no se alteran.

Por facilidad a lo largo del presente trabajo se utilizarán los conteos esperados para definir los modelos.

1.5.4 RELACION CON LA RAZON DE PRODUCTOS CRUZADOS

Los términos μ en el modelo loglineal pueden ser expresados en función de la razón de productos cruzados, tal relación permite mostrar la consistencia en la definición del modelo loglineal sobre todo cuando se realiza la generalización a dimensiones mayores. Considerando el modelo para una tabla de 2x2 los términos μ son expresados de la siguiente forma:

De (1.5-13) resulta

$$\begin{aligned}\mu_{1(1)} &= \frac{1}{2}\,\ell_{1+} - \frac{1}{4}\,\ell_{++}\\ &= \frac{1}{2}\,(\ell_{11} + \ell_{12}) - \frac{1}{4}\,(\ell_{11} + \ell_{12} + \ell_{21} + \ell_{22})\\ &= \frac{1}{4}\,(\ell_{11} + \ell_{12} - \ell_{21} - \ell_{22})\\ &= \frac{1}{4}\,\ln\,(\,m_{11}m_{12}\,/\,m_{21}m_{22}) \qquad (1.5.4\text{-}1)\end{aligned}$$

De (1.5-14) se tiene

$$\begin{aligned}\mu_{2(1)} &= \frac{1}{2}\,\ell_{+1} - \frac{1}{4}\,\ell_{++}\\ &= \frac{1}{2}\,(\ell_{11} + \ell_{21}) - \frac{1}{4}\,(\ell_{11} + \ell_{12} + \ell_{21} + \ell_{22})\\ &= \frac{1}{4}\,(\ell_{11} + \ell_{21} - \ell_{12} - \ell_{22})\\ &= \frac{1}{4}\,\ln\,(\,m_{11}m_{21}\,/\,m_{12}m_{22}) \qquad (1.5.4\text{-}2)\end{aligned}$$

De (1.5-18) resulta

$$\begin{aligned}\mu_{12(11)} &= \ell_{11} - \frac{1}{2}\,\ell_{+1} - \frac{1}{2}\,\ell_{1+} + \frac{1}{4}\,\ell_{++}\\ &= \ell_{11} - \frac{1}{2}\,(\ell_{11} + \ell_{12} + \ell_{11} + \ell_{21})\\ &\qquad + \frac{1}{4}\,(\ell_{11} + \ell_{12} + \ell_{21} + \ell_{22})\end{aligned}$$

$$= \frac{1}{4} (\ell_{11} + \ell_{22} - \ell_{12} - \ell_{21})$$

$$= \frac{1}{4} \ln (m_{11}m_{22} / m_{12}m_{21}) \qquad (1.5.4\text{-}3)$$

En tablas de mayor tamaño pueden expresarse los términos μ en función de razones de productos cruzados, así por ejemplo en una tabla 2xJ se obtienen las siguientes expresiones:

De (1.5-13) se tiene

$$\mu_{1(1)} = \frac{1}{J} \ell_{1+} - \frac{1}{2J} \ell_{++}$$

$$= \frac{1}{J} \sum_j \ell_{1j} - \frac{1}{2J} \sum_j (\ell_{1j} + \ell_{2j})$$

$$= \frac{1}{2J} \sum_j (\ell_{1j} + \ell_{2j}) \qquad (1.5.4\text{-}4)$$

y si se vuelve a la escala original se tiene

$$\exp (\mu_{1(1)}) = \left[\prod_j (m_{1j} / m_{2j}) \right]^{\frac{1}{2J}} \qquad (1.5.4\text{-}5)$$

De (1.5-14) se tiene

$$\mu_{2(1)} = \frac{1}{2} \ell_{+1} - \frac{1}{2J} \ell_{++}$$

$$= \frac{1}{2} (\ell_{11} + \ell_{21}) - \frac{1}{2J} \sum_j (\ell_{1j} + \ell_{2j})$$

$$= \frac{1}{2J} \sum_{j=2}^{J} (\ell_{11} + \ell_{21} - \ell_{1j} - \ell_{2j}) \qquad (1.5.4\text{-}6)$$

al volver a la escala original se tiene

$$\exp\ (\mu_{2(1)}) = \left[\prod_{j=2}^{J} (m_{11}\ m_{21}\ /\ m_{1j}\ m_{2j}) \right]^{\frac{1}{2J}} \qquad (1.5.4\text{-}7)$$

De (1.5-18) se tiene

$$\begin{aligned} \mu_{2(1)} &= \ell_{11} - \frac{1}{2}\ \ell_{+1} - \frac{1}{J}\ \ell_{1+} + \frac{1}{2J}\ \ell_{++} \\ &= \frac{1}{2}\ (\ell_{11} - \ell_{21}) + \frac{1}{2J} \sum_{j} (\ell_{2j} + \ell_{1j}) \\ &= \frac{1}{2J} \sum_{j=2}^{J} (\ell_{11} - \ell_{21} + \ell_{2j} - \ell_{1j}\ ) \qquad (1.5.4\text{-}8) \end{aligned}$$

si se vuelve a la escala original se tiene

$$\exp\ (\mu_{12(11)}) = \left[\prod_{j=2}^{J} (m_{11}\ m_{2j}\ /\ m_{21}\ m_{1j}) \right]^{\frac{1}{2J}} \qquad (1.5.4\text{-}9)$$

1.5.5 PLEGAMIENTO DE TABLAS

Cuando en una tabla de contingencia bidimensional las variables son independientes y al menos una de ellas tiene más de dos categorías, puede realizarse un plegamiento (colapsamiento) de las casillas de la tabla sin que las propiedades estructurales se pierdan. A continuación se presenta una justificación más formal de este planteamiento.

Lema 1.5.5-1

En una tabla de contingencia bidimensional IxJ sean r y s dos categorías cualesquiera de una variable (sin pérdida de generalidad suponga que es la primera)

Si $m_{r1}/m_{s1} = m_{r2}/m_{s2} = \ldots = m_{rJ}/m_{sJ}$

Entonces $m_{r+}/m_{s+} = m_{rj}/m_{sj}$ *para* $j=1,2,\ldots J$.

La demostración del lema se realiza mediante el principio de inducción matemática.

Independencia y Proporcionalidad

En una tabla de contingencia de 2x2 puede notarse que para (1.5.4-3) el término $\mu_{12(ij)} = 0$ si y sólo si $m_{11}/m_{21} = m_{12}/m_{22}$, haciendo desarrollos semejantes para los restantes términos μ_{12} puede notarse la misma doble implicación, de donde se deduce que la independencia de las variables implica la igualdad de proporciones m_{1j}/m_{2j} $j=1,2$ y recíprocamente.

Así mismo en una tabla de 2xJ a partir de la expresión (1.5.4-9) con desarrollos análogos para todos términos μ_{12} se desprende que el modelo de independencia se ajusta a los datos si y sólo si $m_{11}/m_{21} = m_{12}/m_{22} = \ldots = m_{1J}/m_{2J}$.

Independencia en una tabla IxJ

En una tabla de contingencia bidimensional la independencia entre renglones y columnas puede expresarse de dos formas equivalentes:

$$m_{ij} = m_{i+}\, m_{+j} \;/\; N \tag{1.5.5-1}$$
$$\ell_{ij} = \mu + \mu_{1(i)} + \mu_{2(j)} \tag{1.5.5-2}$$

A partir de (1.5.5-2) para dos renglones cualesquiera en la misma columna resulta que

$$\ell_{rj} - \ell_{sj} = \mu_{1(r)} - \mu_{1(j)}$$

retornando a la escala original se tiene

$$\begin{aligned} m_{rj} \;/\; m_{sj} &= \exp\left(\mu_{1(r)} - \mu_{1(s)} \right) \\ &= \exp\left[\frac{1}{J}\,\ell_{r+} - \frac{1}{IJ}\,\ell_{++} - \frac{1}{J}\,\ell_{s+} + \frac{1}{IJ}\,\ell_{++} \right] \\ &= \exp\left[\frac{1}{j} \sum_{j} \left(\ell_{rj} - \ell_{sj} \right) \right] \\ &= \prod_{j} \left(m_{rj} \;/\; m_{sj} \right)^{1/J} \\ &= m_{r+} \;/\; m_{s+} \end{aligned} \tag{1.5.5-3}$$

Debido al lema 1.5.5-1.

Del modelo de independencia se deducen las siguientes conclusiones:

1) No se pierde la independencia si se pliegan algunas categorías para cualquiera de las variables.

2) Los parámetros μ_1 pueden determinarse a partir de las marginales $\{m_{i+}\}$ y los parámetros μ_2 a partir de $\{m_{+j}\}$.

Ejemplo:

Considere las siguientes tablas

Tabla 1

5	20
10	40

Tabla 2

5	4	16
10	8	32

Tabla 3

5	13	7
10	20	20

En la tabla 2 las proporciones m_{1j} / m_{2j} son constantes para todo j de aquí que el modelo de independencia ajuste a la tabla, si se pliegan las dos últimas columnas se obtiene la tabla 1, se conservan constantes las proporciones y por tanto la independencia. Puede notarse por otra parte que la tabla 3 no conserva constantes las proporciones y por tanto no ajusta el modelo de independencia, plegando sus dos últimas categorías se obtiene la tabla 1 la cual como se sabe si se ajusta al modelo de independencia.

Como consecuencia final se tiene que la independencia en una tabla expandida implica la independencia en una tabla condensada, mientras el recíproco no siempre es cierto.

1.6 MODELO EN TRES DIMENSIONES

Al aumentar el número de variables se requiere utilizar tablas de contingencia multidimensionales y como consecuencia de las relaciones entre los datos, una mayor variedad de resultados se obtiene; tales relaciones se ven reflejadas en los parámetros de los modelos loglineales multivariados.

En esta sección se analizan los siguientes tópicos:

1) Construcción del modelo loglineal asociado a una tabla de contingencia tridimensional.

2) Interpretación de los parámetros del modelo saturado.

3) Interpretación de modelos no saturados como descripción de hipótesis.

4) Determinación de las condiciones para el plegamiento de tablas.

Notación

Considerando una tabla de contingencia de tres entradas, la notación utilizada en las secciones anteriores puede ser generalizada. Así, una celda (i,j,k) presenta una frecuencia observada x_{ijk} , una probabilidad p_{ijk} de que la observación sea registrada en esa celda y un conteo esperado m_{ijk} donde $i=1,2,\ldots,I$; $j=1,2,\ldots,J$ y $k=1,2,\ldots,K$.

El subíndice "+", al igual que en secciones anteriores, indica la suma a través de todas las casillas con una posición común en el subíndice, por ejemplo $m_{+jk} = \sum_{i=1}^{I} m_{ijk}$ y $N=m_{+++}$ es ahora el número total de observaciones en la tabla.

Con objeto de facilitar el entendimiento en la construcción del modelo tridimensional se utilizará primeramente una tabla 2x2x2 para posteriormente generalizar hacia la tabla general IxJxK.

1.6.1 CONSTRUCCION DEL MODELO TRIDIMENSIONAL

Modelo 2x2x2

El modelo tridimensional más simple es el asociado a una tabla de contingencia de 2x2x2 cuya forma general se presenta a continuación:

Tabla 1.6.1-1: Forma general de una tabla 2x2x2

		Primera categoría de la variable 3				Segunda categoría de la variable 3	
		Variable 2				Variable 2	
		1	2			1	2
Variable 1:	1	m_{111}	m_{121}	Variable 1:	1	m_{112}	m_{122}
	2	m_{211}	m_{221}		2	m_{212}	m_{222}

El modelo loglineal asociado a la tabla puede generarse a partir de los modelos loglineales particulares asociados a cada una de las subtablas identificadas en la tabla 1.6.1-1, así entonces los dos modelos quedan como sigue:

$$\ell_{ijk} = u^{(k)} + u^{(k)}_{1(i)} + u^{(k)}_{2(j)} + u^{(k)}_{12(ij)} \quad k=1,2 \quad (1.6.1\text{-}1)$$

con k=1,2 y con sus respectivas restricciones lineales

$$\sum_{i} u^{(k)}_{1(i)} = \sum_{j} u^{(k)}_{2(j)} = \sum_{i} u^{(k)}_{12(ij)} = \sum_{j} u^{(k)}_{12(ij)} = 0 \quad (1.6.1\text{-}2)$$

La combinación de los modelos descritos en (1.6.1-1) conducen a la obtención del modelo saturado, así entonces los parámetros quedan dados de la siguiente forma:

Media general

$$\mu = \frac{1}{K}\sum_{k} u^{(k)} \qquad (1.6.1\text{-}3)$$

Efecto principal de la primer variable

$$\mu_{1(i)} = \frac{1}{K}\sum_{k} u_{1(i)}^{(k)} \qquad (1.6.1\text{-}4)$$

Efecto principal de la segunda variable

$$\mu_{2(j)} = \frac{1}{K}\sum_{k} u_{2(j)}^{(k)} \qquad (1.6.1\text{-}5)$$

Interacción entre las variables 1 y 2

$$\mu_{12(ij)} = \frac{1}{K}\sum_{k} u_{12(ij)}^{(k)} \qquad (1.6.1\text{-}6)$$

Las desviaciones de estos parámetros dependen de la tercer variable y se definen como sigue:

Efecto principal de la variable 3

$$\mu_{3(k)} = u^{(k)} - \mu \qquad (1.6.1\text{-}7)$$

Interacciones de primer orden con la variable 3

$$\mu_{13(ik)} = u_{1(i)}^{(k)} - \mu_{1(i)} \qquad (1.6.1\text{-}8)$$

$$\mu_{23(jk)} = u_{2(j)}^{(k)} - \mu_{2(j)} \qquad (1.6.1\text{-}9)$$

Interacción de segundo orden o efecto de tres factores

$$\mu_{123(ijk)} = u_{12(ij)}^{(k)} - \mu_{12(ij)} \qquad (1.6.1\text{-}10)$$

Finalmente, el modelo loglineal saturado para la tabla 2x2x2 queda determinado por

$$\ell_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{13(ik)} + \\ + \mu_{23(jk)} + \mu_{123(ijk)} \qquad (1.6.1\text{-}11)$$

Debido a las restricciones (1.6.1-2) se deduce que

$$\sum_i \mu_{1(i)} = \sum_j \mu_{2(j)} = \sum_k \mu_{3(k)} = \sum_i \mu_{12(ij)} = \sum_j \mu_{12(ij)} = 0$$

$$\sum_i \mu_{13(ik)} = \sum_k \mu_{13(ik)} = \sum_j \mu_{23(jk)} = \sum_k \mu_{23(jk)} = 0$$

$$\sum_i \mu_{123(ijk)} = \sum_j \mu_{123(ijk)} = \sum_k \mu_{123(ijk)} = 0 \qquad (1.6.1\text{-}12)$$

Modelo IxJxK

Para una tabla general en tres dimensiones un procedimiento similar al realizado para obtener el modelo (1.6.1-11) se puede utilizar con la diferencia de obtener el promedio de las K tablas de IxJ. Si se considera que $i=1,2,\ldots,I$; $j=1,2,\ldots,J$ y $k=1,2,\ldots,K$ las expresiones (1.6.1-1) a la (1.6.1-12) no se alteran y el modelo saturado (1.6.1-11) es completamente aplicable.

Tabla 1.6.1-2: Grados de libertad en modelo tridimensional

término μ	*Grados de libertad*
μ	*1*
μ_1	*(I-1)*
μ_2	*(J-1)*
μ_3	*(K-1)*
μ_{12}	*(I-1) (J-1)*
μ_{13}	*(I-1) (K-1)*
μ_{23}	*(J-1) (K-1)*
μ_{123}	*(I-1) (J-1) (K-1)*
Total	*IJK*

Los grados de libertad asociados con cada parámetro del modelo saturado se muestran en la tabla 1.6.1-2.

Puede notarse que el número de parámetros independientes en el modelo saturado es igual al número de casillas en la tabla.

1.6.2 RELACION CON LA RAZON DE PRODUCTOS CRUZADOS

En la sección anterior se realizó la construcción del modelo loglineal mediante la partición de la tabla a través de las categorías de la tercera variable, tal acción podría indicar una aparente diferencia en la definición de los parámetros en caso de ser utilizada una partición diferente de la tabla. Por ejemplo, podrían generarse las definiciones de los parámetros para la primera o segunda variables como desviaciones del promedio de los parámetros de las restantes variables. Una forma de mostrar la unicidad de los resultados independientemente de la partición elegida, consiste en expresar los parámetros del modelo en función de las razones de productos cruzados.

En la sección 1.5.4 se mostró que todo término en el modelo bidimensional puede ser expresado en función de razones de productos cruzados, ahora mediante un rearreglo de las casillas de la tabla tridimensional también pueden obtenerse expresiones de este tipo para los parámetros del modelo. En [Bishop,Fienberg & Holland,1975], pag. 33-35 pueden encontrarse expresiones para los parámetros del modelo asociado a una tabla de 2x2x2 y puede demostrarse que independientemente de la partición seleccionada para la tabla de 2x2x2 se obtienen los siguientes expresiones únicas:

$$\mu_{12(11)} = \frac{1}{8} \ln \left(\alpha^{(1)} \alpha^{(2)} \right) \qquad (1.6.2\text{-}1)$$

$$\mu_{123(111)} = \frac{1}{8} \ln \left(\alpha^{(1)} / \alpha^{(2)} \right) \qquad (1.6.2\text{-}2)$$

Puede notarse que $\mu_{12(11)} = 0$ implica que $\mu_{123(1jk)} = 0$ dado que si $\alpha^{(1)} \alpha^{(2)} = 1$ entonces también $\alpha^{(1)} / \alpha^{(2)} = 1$. Este hecho muestra las bases para definir el principio jerárquico analizado en la siguiente sección.

1.6.3 INTERPRETACION DE PARAMETROS Y EL PRINCIPIO JERARQUICO

La interpretación de los parámetros en el modelo (1.6.1-11) está en función de los factores que intervienen en cada uno de ellos. Así entonces,

μ	representa la media general del modelo. Promedio de las medias generales de las subtablas de la partición.
μ_1, μ_2 y μ_3	Efectos principales de las variables 1, 2 y 3 respectivamente.
μ_{12}, μ_{13} y μ_{23}	Interacciones de primer orden o Efectos conjuntos de dos factores. Representa el efecto de interacción entre las variables 1 y 2, 1 y 3 y 2 y 3
μ_{123}	Efecto conjunto de tres factores o Interacción de segundo orden. Representa la interacción entre todas las variables.

Una interpretación alternativa para el término μ_{123} es el que es un indicador de la diferencia en la magnitud de los efectos de dos factores entre las subtablas de la partición, lo que implica que si cualquier efecto conjunto de dos factores (interacción) permanece constante entre todas las subtablas de la partición entonces el efecto conjunto de tres factores resulta ser cero.

Por ejemplo si

$$u^{(1)}_{12(ij)} = u^{(2)}_{12(ij)} = \ldots = u^{(K)}_{12(ij)}$$

entonces de (1.6.1-6) se deduce que

$$\mu_{12(ij)} = u^{(k)}_{12(ij)} \text{ para todo } k$$

y por tanto de (1.6.1-10) resulta que

$$\mu_{123(ijk)} = 0 \text{ para todo } k.$$

Orden relativo de los términos

Si un término μ tiene un conjunto A de subíndices y un segundo término tiene un conjunto B de subíndices y $B \subset A$, se dice entonces que el primer término es de orden relativo mayor que el segundo. Por ejemplo, μ_{12} es de orden relativo mayor que μ_2.

Principio Jerárquico

Se define como familia de modelos jerárquicos al conjunto de modelos en los cuales ocurre que si se tiene un término μ igual a cero entonces todos los términos de orden relativo superior también deben ser igual a cero. Por ejemplo, si $\mu_{12} = 0$ entonces $\mu_{123} = 0$.

1.6.4 INTERPRETACION DE LOS MODELOS

A continuación se presenta la interpretación de los modelos no saturados en tres dimensiones teniendo en cuenta que en todos los casos se hará bajo la consideración del principio jerárquico.

1) Asociación parcial (entre pares de variables)

Cuando se considera que $\mu_{123} = 0$, el modelo

$$\ell_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{13(ik)} + \mu_{23(jk)} \qquad (1.6.4\text{-}1)$$

indica la asociación parcial entre cada par de variables. Este modelo puede entenderse también como el indicio de que no existe el efecto conjunto de las tres variables (interacción de segundo orden).

A través de los estadísticos de bondad de ajuste se determina si un modelo es adecuado a la estructura de una tabla de contingencia en particular, como puede notarse en el capítulo III. Así entonces, el confrontar la bondad de ajuste del modelo (1.6.4-1) equivale a probar la hipótesis de inexistencia de interacción de segundo orden entre las variables.

2) Independencia condicional

El modelo que resulta de considerar algún término de dos factores igual a cero, por ejemplo $\mu_{12} = 0$ (y por jerarquía $\mu_{123} = 0$)

$$\ell_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{13(ik)} + \mu_{23(jk)} \qquad (1.6.4\text{-}2)$$

establece que las variables 1 y 2 son independientes para cada nivel de la variable 3, aunque cada una de las primeras está relacionada con la última variable, es decir, las variables 1 y 2 son condicionalmente independientes.

3) Completa independencia de una variable (independencia marginal)

Si se considera que dos interaciones de primer orden son iguales a cero, por ejemplo, $\mu_{12} = \mu_{13} = 0$ (y $\mu_{123} = 0$) el modelo

$$\ell_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{23(jk)} \qquad (1.6.4\text{-}3)$$

indica que la variable 1 es completamente independiente de las otras dos, aunque estas últimas se encuentren relacionadas.

4) Completa independencia de todas las variables

En el caso de considerar que no hay interacciones de primer orden y por consiguiente tampoco de segundo, el modelo

$$\ell_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} \qquad (1.6.4\text{-}4)$$

señala que todas las variables son completamente independientes.

5) Modelos no comprensibles

Si se considera el hecho de retirar alguno ó algunos términos más del modelo de completa independencia (1.6.4-4) se obtiene un modelo denominado no comprensible. Suponga que $\mu_3 = 0$ entonces el modelo resultante

$$\ell_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} \qquad (1.6.4\text{-}5)$$

nos conduce a la siguiente aseveración:
De la expresión (1.6.4-5) resulta que

$$m_{ijk} = \exp\ (\mu + \mu_{1(i)} + \mu_{2(j)})$$

$$m_{ij+} = \sum_{k} \exp\ (\mu + \mu_{1(i)} + \mu_{2(j)})$$

de donde finalmente resulta que

$$m_{ijk} = m_{ij+} / K$$

lo cual indica que se tiene la misma tabla IxJ para todo k. Entonces sumando sobre cada variable no incluida en el modelo se puede describir la estructura condensada mediante un modelo "comprensible" que incluya las variables restantes de la tabla, resultando un arreglo de menor dimensión.

6) Modelos no jerárquicos

Un modelo de este tipo ocurre cuando no se cumple con el principio jerárquico, es decir, cuando se considera algún parámetro igual a cero pero alguno(s) de los términos de orden relativo superior aparece en el modelo. Aunque estos modelos presentan la dificultad de interpretación , son defendidos por algunos autores que argumentan el concepto de "sinergismo" lo que significa puede existir la respuesta cuando dos factores se presentan juntos pero no cuando cada uno se presenta por separado.
Este tipo de modelos resulta adecuado a situaciones en que se desea encontrar alguna ecuación que ajuste mejor a los datos sin que importe tanto su interpretabilidad.

1.6.5 PLEGAMIENTO DE TABLAS

En la sección 1.5.5 se mostró que cuando las variables son independientes en una tabla bidimensional puede realizarse un plegamiento de categorías sin perder las propiedades estructurales de la tabla. Ahora, en el caso tridimensional puede realizarse el plegamiento a través de alguna variable si ésta es al menos condicionalmente independiente de las dos restantes.

A continuación se muestran las condiciones para poder obtener estimadores válidos de los términos de dos factores μ_{12}, μ_{13} y μ_{23} de la tabla bidimensional resultante de la suma de categorías de alguna de las variables $\{m_{ij+}\}$, $\{m_{i+j}\}$ o $\{m_{+jk}\}$.

Cuando para un modelo loglineal asociado con una tabla de contingencia tridimensional ocurre que un término de dos factores (digamos μ_{12}) es igual al que se obtiene con un modelo loglineal bidimensional asociado a la tabla de marginales $\{m_{ij+}\}$ entonces es posible la reducción en la dimensión de la tabla original, la condición que posibilita tal igualdad es que $\mu_{13} = 0$ o $\mu_{23} = 0$.

De manera más formal, considérese la tabla tridimensional $\{m_{ijk}\}$ y a partir del modelo saturado correspondiente (1.6.1-11) se desprende que

$$m_{+jk}=e^{(\mu+\mu_{2(j)}+\mu_{3(k)}+\mu_{23(jk)})}\sum_{i}e^{(\mu_{1(i)}+\mu_{12(ij)}+\mu_{13(ik)}+\mu_{23(jk)})} \tag{1.6.5-1}$$

Ahora, si se describe la tabla $\{m_{+jk}\}$ mediante el modelo saturado

$$\ln m_{+jk} = \upsilon + \upsilon_{2(j)} + \upsilon_{3(k)} + \upsilon_{23(jk)}$$

resulta que

$$m_{+jk} = \exp(\upsilon + \upsilon_{2(j)} + \upsilon_{3(k)} + \upsilon_{23(jk)}) \tag{1.6.5-2}$$

se tiene entonces que en general $\upsilon_{23(jk)} \neq \mu_{23(jk)}$.

Ahora si $u_{23(jk)} = \mu_{23(jk)}$ para todo j,k entonces se dice que la tabla tridimensional es **PLEGABLE** con respecto a μ_{23}.
Puede probarse que la variable 1 es plegable respecto a μ_{23} únicamente si la tabla IxJxK es descrita por un modelo loglineal con $\mu_{123} = 0$ y además μ_{12} o $\mu_{13} = 0$ o ambas.
En [Bishop, Fienberg & Holland,1975], pag 39 Teorema 2.4-1 se presenta tal demostración. A continuación se reproduce el enunciado del citado teorema.

Teorema 1.6.5-1

En una tabla rectangular tridimensional, una variable es plegable con respecto a la interacción entre las otras dos variables, si y sólo si ésta es al menos condicionalmente independiente de una de las otras dos variables dada la tercera.

1.7 MODELO EN CUATRO Y MAS DIMENSIONES

La generalización realizada del modelo de dos dimensiones al de tres, consistió en promediar los modelos bidimensionales asociados con cada subtabla definida para cada categoría de la tercer variables. Similarmente para un arreglo de cuatro dimensiones puede obtenerse la definición de modelo loglineal a partir de promediar los modelos tridimensionales asociados con cada subtabla correspondiente a cada una de las L categorías de la cuarta variable.
Así entonces para cada subtabla $\{m_{ijk1}\},\{m_{ijk2}\},\ldots,\{m_{ijkL}]$ con $i=1,2,\ldots,I$; $j=1,2,\ldots,J$ y $k=1,2,\ldots,K$ puede definirse

$$\ell_{ijkl}=w^{(l)}+w^{(l)}_{1(i)}+w^{(l)}_{2(j)}+w^{(l)}_{3(k)}+ w^{(l)}_{12(ij)}+w^{(l)}_{13(ik)}+w^{(l)}_{23(jk)}+w^{(l)}_{123(ijk)} \qquad (1.7\text{-}1)$$

para $l=1,2,\ldots,L$. Los parámetros del modelo tetradimensional quedan dados por

$$\mu = \frac{1}{L}\sum_{l} w^{(l)} \qquad (1.7\text{-}2)$$

$$\mu_{1(i)} = \frac{1}{L}\sum_{l} w^{(l)}_{1(i)} \qquad (1.7\text{-}3)$$

$$\mu_{2(j)} = \frac{1}{L}\sum_{l} w^{(l)}_{2(j)} \qquad (1.7\text{-}4)$$

$$\mu_{3(k)} = \frac{1}{L}\sum_{l} w^{(l)}_{3(k)} \qquad (1.7\text{-}5)$$

$$\mu_{12(ij)} = \frac{1}{L}\sum_{l} w^{(l)}_{12(ij)} \qquad (1.7\text{-}6)$$

$$\mu_{13(ik)} = \frac{1}{L} \sum_{l} w_{13(ik)}^{(l)} \qquad (1.7\text{-}7)$$

$$\mu_{23(jk)} = \frac{1}{L} \sum_{l} w_{23(jk)}^{(l)} \qquad (1.7\text{-}8)$$

$$\mu_{123(ijk)} = \frac{1}{L} \sum_{l} w_{123(ijk)}^{(l)} \qquad (1.7\text{-}9)$$

Los parámetros relativos a la cuarta variable quedan como sigue

$$\mu_{4(l)} = w^{(l)} - \mu \qquad (1.7\text{-}10)$$
$$\mu_{14(l)} = w_{1(i)}^{(l)} - \mu_{1(i)} \qquad (1.7\text{-}11)$$
$$\mu_{24(l)} = w_{2(j)}^{(l)} - \mu_{2(j)} \qquad (1.7\text{-}12)$$
$$\mu_{34(l)} = w_{3(k)}^{(l)} - \mu_{3(k)} \qquad (1.7\text{-}13)$$
$$\mu_{124(l)} = w_{12(ij)}^{(l)} - \mu_{12(ij)} \qquad (1.7\text{-}14)$$
$$\mu_{134(l)} = w_{13(ik)}^{(l)} - \mu_{13(ik)} \qquad (1.7\text{-}15)$$
$$\mu_{234(l)} = w_{23(jk)}^{(l)} - \mu_{23(jk)} \qquad (1.7\text{-}16)$$
$$\mu_{1234(l)} = w_{123(ijk)}^{(l)} - \mu_{123(ijk)} \qquad (1.7\text{-}17)$$

Puede entonces escribirse el modelo en cuatro dimensiones como

$$\begin{aligned} \ell_{ijkl} = {} & \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{4(l)} + \mu_{12(ij)} + \mu_{13(ik)} + \\ & + \mu_{14(il)} + \mu_{23(jk)} + \mu_{24(jl)} + \mu_{34(kl)} + \mu_{123(ijk)} + \\ & + \mu_{124(ijl)} + \mu_{134(ikl)} + \mu_{234(jkl)} + \mu_{1234(ijkl)} \qquad (1.7\text{-}18) \end{aligned}$$

Debido a que los términos μ son definidos como desviaciones del promedio, se nota que la suma sobre cualquiera de los subíndices de los parámetros es igual a cero; así, por ejemplo,

$$\sum_{i} \mu_{1234(ijkl)} = \sum_{j} \mu_{1234(ijkl)} = \sum_{k} \mu_{1234(ijkl)} = \sum_{l} \mu_{1234(ijkl)} = 0$$

Un proceso similar al utilizado para la obtención de modelos de 4 dimensiones puede ser utilizado para dimensiones mayores; así entonces, el modelo saturado en s dimensiones contiene 2^s términos μ de los cuales se tienen $\binom{s}{0}$ términos con cero subíndices, $\binom{s}{1}$ con un subíndice y en general $\binom{s}{r}$ términos con r subíndices.

1.7.1 UTILIDAD DE LOS MODELOS

Los modelos jerárquicos se dividen en dos clases fundamentales, aquellos con todos los efectos de dos factores presentes y aquellos con al menos una interacción de primer orden ausente.

La hipótesis más frecuentemente buscada es la independencia de variables, pero ella requiere que al menos un término de dos factores esté ausente del modelo, de aquí que los modelos con todas las interacciones de primer orden presentes sean utilizados fundamentalmente para la obtención de estimadores de las casillas de manera más confiable; no tanto como modelos para prueba de hipótesis, también suelen ser utilizados en la detección de datos discrepantes o atípicos (outliers).

Los modelos con al menos un efecto de dos factores ausente son referidos preferentemente a la prueba de hipótesis de independencia entre variables, aunque también permiten inspecciones para determinar si el tamaño de una tabla puede ser reducido mediante el plegamiento de alguna(s) de las categorías de sus variables sin que se distorsionen los términos de interacción de mayor interés.

1.7.2 PLEGAMIENTO DE TABLAS

El teorema 1.6.1-1 establece, para el caso tridimensional, que una variable de la tabla puede ser plegada respecto al efecto conjunto de las restantes dos, siempre que la primera no esté relacionada con alguna o ambas de las restante variables, es decir, que se requiere que al menos un efecto de dos factores sea igual a cero. Tal situación puede ser generalizada para el caso con s dimensiones.

Definición de Plegabilidad

Se dice que las variables sobre las cuales son sumadas sus casillas son **PLEGABLES** con respecto a un conjunto de términos μ específicos, cuando los parámetros de los términos μ en el modelo original son idénticos a aquellos términos μ correspondientes en el modelo loglineal para el arreglo reducido.

Teorema 1.7.2-1

Suponga que las variables en un arreglo s-dimensional son divididas en tres grupos mutuamente excluyentes. Un grupo es PLEGABLE con respecto a los términos μ del segundo grupo pero no con respecto a los términos μ del tercer grupo, si y sólo si, los primeros dos grupos son independientes uno del otro. (es decir, los términos μ con subíndices comunes en ambos grupos son iguales a cero)

Demostración

Puede mirarse cada grupo de variables como una variable compuesta. Entonces, al aplicar el enunciado del teorema 1.6.1-1 para las tres variables compuestas el resultado es inmediato.

Por ejemplo, considere un modelo en cuatro dimensiones y suponga que $\mu_{12} = 0$, de acuerdo con el principio jerárquico, esto implica que μ_{123}, μ_{124} y μ_{1234} sean también iguales a cero

$$\ell_{ijkl} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{4(l)} + \mu_{13(ik)} + \mu_{14(il)} + \\ + \mu_{23(jk)} + \mu_{24(jl)} + \mu_{34(kl)} + \mu_{134(ikl)} + \mu_{234(jkl)}$$

Sea A la suma de los elementos del conjunto de términos que no incluye a las variables 1 y 2, B la suma de términos que incluyen a la variable 1 pero no a la variable 2, C la suma de términos que incluye a la variable 2 pero no a la 1 y D la suma de términos que incluye a ambas variables, es decir,

$$A = \mu + \mu_3 + \mu_4 + \mu_{34}$$
$$B = \mu_1 + \mu_{13} + \mu_{14} + \mu_{134}$$
$$C = \mu_2 + \mu_{23} + \mu_{24} + \mu_{234}$$
$$D = \mu_{12} + \mu_{123} + \mu_{124} + \mu_{1234}$$

Es claro entonces que

$$\ell_{ijkl} = A + B + C$$
$$m_{ijkl} = \exp(A + B + C) \qquad (1.7.2\text{-}1)$$

Por otra parte, si se suma (1.7.2-1) sobre la primer variable, sobre la segunda y sobre ambas, se obtiene:

$$m_{+jkl} = \exp(A + C) \sum_i \exp(B),$$

$$m_{i+kl} = \exp(A + B) \sum_j \exp(C)$$

$$m_{++kl} = \exp(A) \sum_i \exp(B) \sum_j \exp(C)$$

de donde puede deducirse que

$$m_{ijkl} = m_{i+kl}\ m_{+jkl} \ /\ m_{++kl} \qquad (1.7.2\text{-}2)$$

Cuando el arreglo original IxJxKxL es partido en KL tablas de dos entradas relativas a las variables 1 y 2, puede notarse que las marginales de cada tabla son elemento de $\{m_{+jkl}\}$ y $\{m_{i+jk}\}$ además que el total de cada tabla es un elemento de $\{m_{++kl}\}$. Mediante lo anterior, puede observarse que la expresión (1.7.2-2) describe la independencia en cada tabla y de acuerdo con ello, puede decirse que las variables 1 y 2 son condicionalmente independientes.

1.7.3 IMPLICACIONES DEL PLEGAMIENTO DE TABLAS

La independencia al menos condicional de dos variables nos implica a través del principio jerárquico que un término de dos factores y sus respectivos términos de orden relativo superior sean también igual a cero. Si una variable es plegable con respecto a ciertos términos μ, esto significa que puede ser eliminada de la tabla mediante la suma sobre todas sus categorías o condensada mediante la combinación de algunas de sus categorías sin que tales términos μ se vean alterados. A partir de lo anterior se deducen dos importantes implicaciones.

1) Si todos los efectos de dos factores estás presentes en el modelo, el plegamiento de la tabla implica la alteración de todos los términos μ.

2) Si alguna variable es independiente de todas las otras puede ser eliminada de la tabla mediante la suma de sus categorías sin que se alteren los términos μ del modelo.

Conviene distinguir que para el ejemplo expuesto en la sección 1.7.2 el teorema 1.7.2-1 capacita para obtener estimadores válidos para los términos μ que involucren a la variable 1 cuando ella es plegada, es decir, pueden estimarse μ_1, μ_{13}, μ_{14} y μ_{134} válidamente pero no pueden obtenerse estimadores adecuados para los términos μ donde no interviene la variable 1.

CAPITULO II: ESTIMACION DE LOS PARAMETROS

En el capítulo anterior se mostró la especificación de los modelos loglineales como una alternativa para describir la estructura de una tabla de contingencia. En el presente capítulo se muestra la forma de obtener estimadores de máxima verosimilitud y de mínimos cuadrados para los parámetros del modelo.

En una primera instancia se presentan las distribuciones muestrales más utilizadas para la recolección de datos y se realiza una introducción a la teoría de los Modelos Lineales Generalizados, posteriormente se establecen los estadísticos suficientes para los esquemas muestrales y se prosigue con la obtención de los estimadores de los parámetros del modelo.

Finalmente se discuten el Método de Ajuste Iterativo Proporcional y el Método de Newton-Raphson así como las técnicas para el cálculo de los estimadores de los parámetros.

2.1 DISTRIBUCIONES MUESTRALES

En la sección 1.2 se mencionó que se tienen fundamentalmente tres distribuciones muestrales para la recolección de datos de una tabla de contingencia: Poisson, Multinomial y Producto Multinomial. A continuación se muestra que tales distribuciones pertenecen a la Familia Exponencial de Distribuciones (o sólo Familia Exponencial) y, por tanto, es aplicable la teoría de los Modelos Lineales Generalizados para la obtención de los estimadores.

2.1.1 FAMILIA EXPONENCIAL DE DISTRIBUCIONES

Se considera una variable aleatoria Y cuya función de densidad de probabilidad depende de un solo parámetro θ, la distribución pertenece a la familia exponencial si ésta puede escribirse en la forma

$$f(y;\theta) = s(y)\ t(\theta)\ e^{a(y)b(\theta)} \qquad (2.1.1\text{-}1)$$

donde a, b, s y t son funciones conocidas.

La ecuación (2.1.1-1) puede ser reescrita en la forma

$$f(y,\theta) = \exp[\ a(y)b(\theta) + c(\theta) + d(y)\ ] \qquad (2.1.1\text{-}2)$$

donde $s(y) = \exp d(y)$ y $t(\theta) = \exp c(\theta)$.

Si $a(y) = y$ la distribución en (2.1.1-2) se dice estar en la forma canónica y $b(\theta)$ es denominado el parámetro natural de la distribución.

Si existen otros parámetros en adición al parámetro de interés θ formando parte de las funciones a, b, c y d son considerados como constantes conocidas (parámetros de "ruido").

De manera más general puede verse que si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes todas ellas con la misma distribución dada por (2.1.1-2), su función de densidad de probabilidad conjunta queda dada por

$$\begin{aligned} f(y_1, y_2, \ldots, y_n) &= \prod_{i=1}^{n} \exp[\ b(\theta)\ a(y_i) + c(\theta) + d(y_i)\ ] \\ &= \exp[\ b(\theta) \sum_{i=1}^{n} a(y_i) + n\ c(\theta) + \sum_{i=1}^{n} d(y_i)\ ] \end{aligned} \qquad (2.1.1\text{-}3)$$

El término $\sum a(y_i)$ se dice que es un estadístico suficiente para $b(\theta)$, lo que significa que en cierto sentido $\sum a(y_i)$ resume toda la información disponible acerca del parámetro θ. Vease por ejemplo [Hogg & Craig,1968], cap 10 o [Cox & Hinkley,1974] cap 2.

2.1.2 DISTRIBUCION POISSON

Sea Y una variable aleatoria con función de densidad de probabilidad Poisson

$$g(y;\lambda) = \lambda^{y} e^{-\lambda} / y! \qquad (2.1.2\text{-}1)$$

$$g(y;\lambda) = \exp[\, y \ln \lambda - \lambda - \ln y! \,] \qquad (2.1.2\text{-}2)$$

puede notarse que la distribución Poisson está en la forma canónica con parámetro natural $\ln \lambda$ y pertenece a la familia exponencial.

Si se considera para una tabla de contingencia el no tener restricciones sobre el tamaño total de la muestra, cada casilla en la tabla tiene una distribución Poisson independiente. Suponiendo una tabla IxJxK y teniendo en cuenta que $\mathcal{E}(x_{ijk}) = m_{ijk}$ la función de densidad de probabilidad conjunta para las x_{ijk} está dada por

$$f(\{x_{ijk}\}) = \prod_{ijk} m_{ijk}^{x_{ijk}} e^{-m_{ijk}} / x_{ijk}! \qquad (2.1.2\text{-}3)$$

$$f(\{x_{ijk}\}) = \prod_{ijk} \exp[\, x_{ijk} \ln m_{ijk} - m_{ijk} - \ln x_{ijk}\,! \,]$$

$$= \exp\left\{\sum_{ijk} x_{ijk}\, \ln m_{ijk} - \sum_{ijk} m_{ijk} - \sum_{ijk} \ln x_{ijk}!\right\}$$

(2.1.2-4)

Considerando que para todo modelo $\sum \hat{m}_{ijk} = n$ y $\sum_{ijk} \ln x_{ijk}!$ es una constante, el Kernel de la función es

$$\sum_{ijk} x_{ijk}\, \ln m_{ijk} \qquad (2.1.2\text{-}5)$$

Puede notarse que (2.1.2-4) responde a las condiciones de la familia exponencial de distribuciones.

2.1.3 DISTRIBUCION MULTINOMIAL

Si se tiene la condición de tener fijo el tamaño total de la muestra, tal restricción impuesta sobre una serie de distribuciones Poisson independientes conduce a la denominada distribución multinomial.

De la propiedad aditiva para variables aleatorias con distribución Poisson, su suma $N = \sum x_{ijk}$ también tiene distribución Poisson con parámetro $m_{+++} = \sum m_{ijk}$, de aquí que la distribución conjunta de las Y_{ijk} condicional a N es

$$f(\{x_{ijk}\} \mid N) = \prod_{ijk}\left[m_{ijk}^{x_{ijk}}\, e^{-m_{ijk}} / x_{ijk}!\right] / \left[m_{+++}^{N}\, e^{-m_{+++}} / N!\right]$$

$$= \left[N! \Big/ \prod_{ijk} x_{ijk}! \right] \prod_{ijk} \left(m_{ijk} \;/\; m_{+++} \right)^{x_{ijk}} \qquad (2.1.3\text{-}1)$$

Dado que $m_{+++}^{N} = \prod_{ijk} m_{+++}^{x_{ijk}}$ y $e^{-m_{+++}} = \prod_{ijk} e^{-m_{ijk}}$

Ahora considerando que para todo modelo $\sum m_{ijk} = N$

$$f(\{x_{ijk}\} \mid N) = \left[N! \Big/ \prod_{ijk} x_{ijk}! \right] \prod_{ijk} \left(m_{ijk} \;/\; N \right)^{x_{ijk}} \qquad (2.1.3\text{-}2)$$

La función (2.1.3-2) puede ser reescrita como sigue

$$f(\{x_{ijk}\} \mid N) = \exp\left\{ \ln\left[N! \Big/ \prod_{ijk} x_{ijk}! \right] + \sum_{ijk} x_{ijk} \ln m_{ijk} - N \ln N \right\} \qquad (2.1.3\text{-}3)$$

el primero y tercer términos resultan ser constantes para todo modelo, de aquí que el Kernel de la función sea nuevamente igual a la ecuación (2.1.2-5).

A partir de (2.1.3-3) puede deducirse que la distribución multinomial satisface las condiciones de la Familia exponencial de distribuciones.

2.1.4 DISTRIBUCION PRODUCTO MULTINOMIAL

En algunas situaciones de investigaciones comparativas es usual tener varios grupos con un número total de individuos en cada grupo determinado por un diseño muestral, tal situación implicará el mantener fijas algunas marginales.

Considere en tres dimensiones que se mantiene fija $C_{12} = \{x_{ij+}\}$, la distribución de C_{12} es multinomial con función de densidad de probabilidad conjunta

$$f(\{x_{ijk}\}) = \left[N! / \prod_{ij} x_{ij+}! \right] \prod_{ij} (\, m_{ij+} \,/\, N \,)^{x_{+jk}} \qquad (2.1.4\text{-}1)$$

dado que la distribución de C_{12} es el producto de IJ multinomiales independientes.

Ahora la distribución condicional conjunta quedará dada por el cociente de las expresiones (2.1.3-2) y (2.1.4-1) donde resulta que

$$f(\{x_{ijk}\} \mid \{m_{ij+} = x_{ij+}\}) =$$

$$\prod_{ij} \left[\left[x_{ij+}! \;/\; \prod_{k} x_{ijk}! \right] \prod_{k} (m_{ijk} \;/\; x_{ij+})^{x_{ijk}} \right] \qquad (2.1.4\text{-}2)$$

tal expresión puede reescribirse en la forma siguiente:

$$f(\{x_{ijk}\} \mid \{m_{ij+} = x_{ij+}\}) =$$

$$\exp\left[\ln \prod_{ij} x_{ij+}! \;-\; \ln \prod_{ijk} x_{ijk}! \;+\; \sum_{ijk} x_{ijk} \ln m_{ijk} \;-\; \ln \prod_{ijk} x_{ij+}^{-x_{ijk}} \right]$$

(2.1.4-3)

Donde puede notarse nuevamente que el Kernel de la función es igual a la expresión (2.1.2-5) dado que las x's son constantes en un modelo determinado.

De (2.1.4-3) se deduce que la distribución pertenece a la familia exponencial.

2.2 MODELOS LINEALES GENERALIZADOS

La idea de Modelos Lineales Generalizados expresada en [Nelder & Wedderburn,1972], implica la unicidad de una gran cantidad de métodos estadísticos que involucran una combinación lineal de parámetros.

En términos generales tales modelos pueden resumirse de la siguiente forma:

Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias independientes, cada una de ellas con una distribución perteneciente a la familia exponencial, suponga que su distribución es expresada en forma canónica dependiente de un parámetro θ_i

$$f(y_i;\theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)] \qquad (2.2\text{-}1)$$

La distribución de todas las Y's son de la misma forma (por ejemplo todas Poisson o Normal) obteniéndose una función de densidad de probabilidad conjunta para $Y_1, Y_2, \ldots, Y_n$

$$f(y_1, y_2, \ldots, y_n; \theta_1, \theta_2, \ldots, \theta_n) =$$

$$\exp\left[\sum_i y_i b(\theta_i) + \sum_i c(\theta_i) + \sum_i d(y_i)\right] \qquad (2.2\text{-}2)$$

Usualmente los parámetros θ_i no son directamente de interés, en vez de ello se considera para un modelo lineal generalizado un conjunto de parámetros $\beta_1, \beta_2, \ldots, \beta_p$ (p<n) tal que la combinación lineal de las β's es igual a alguna función del valor esperado de cada Y_i, esto es,

$$g(\mu_i) = \underline{X}_i' \, \underline{\beta} \qquad (2.3\text{-}3)$$

donde $\mu_i = \mathcal{E}(Y_i)$, g es una función monótona y diferenciable denominada función de liga, $\underline{X}_i$ es un vector px1 de variables de explicación y $\underline{\beta}$ es un vector de parámetros $\underline{\beta}' = [\ \beta_1, \beta_2, \ldots, \beta_p\ ]$

Así el modelo lineal generalizado tiene tres componentes

1) Las variables aleatorias de respuesta $Y_1, Y_2, \ldots, Y_n$ independientes con la misma distribución perteneciente a la familia exponencial.

2) Un conjunto de parámetros $\underline{\beta}$ pxl y variables de explicación X nxp.

3) Una función de liga monótona y diferenciable tal que

$$g(\mu_i) = \underline{X}_i' \ \underline{\beta} \text{ donde } \mu_i = \mathcal{E}(Y_i)$$

El ejemplo más popular es el caso del modelo lineal

$$\underline{y} = X \ \underline{\beta} + \underline{e}$$

donde los elementos de $\underline{e}$ son tales que $e_i \sim N(0,\sigma^2)$, lo que implica que $y_i \sim N(\mu_i,\sigma^2)$, la función de liga $g(\mu_i) = \mu_i = \underline{X}_i' \ \underline{\beta}$ y es sabido que la distribución Normal pertenece a la familia exponencial.

2.2.1 CASO PARTICULAR EL MODELO LOGLINEAL

Considerando una tabla de contingencia de dos o más dimensiones, si cada una de las casillas de la tabla es etiquetada como Y_i, $i=1,2,...,n$ las tres principales distribuciones presentan las siguientes características:

2.2.1.1 DISTRIBUCION POISSON

Si no se tienen restricciones sobre las frecuencias y_i o sobre los parámetros λ_i la función de densidad de probabilidad conjunta queda dada por

$$f(\underline{y};\underline{\lambda}) = \prod_i \lambda_i^{y_i} e^{-\lambda_i} \; / \; y_i! \qquad (2.2.1.1\text{-}1)$$

y se sabe que el valor esperado para cada y_i es

$$\mathscr{E}(Y_i) = \lambda_i \qquad (2.2.1.1\text{-}2)$$

2.2.1.2 DISTRIBUCION MULTINOMIAL

Si se considera fijo el total de observaciones en la tabla, $N = \sum_i y_i$ y se tiene que $\sum_i \theta_i = 1$, la función de densidad de probabilidad conjunta queda dada por

$$f(\underline{y};\underline{\theta} \mid N) = N! \prod_i \theta_i^{y_i} \; / \; y_i! \qquad (2.2.1.2\text{-}1)$$

De donde puede probarse que

$$\mathcal{E}(Y_i) = N\,\theta_i \qquad (2.2.1.2\text{-}2)$$

Ver por ejemplo [Bishop,Fienberg & Holland,1975], sec 13.4.

2.2.1.3 DISTRIBUCION PRODUCTO MULTINOMIAL

Si se considera una tabla tridimensional IxJxK y si se mantienen fijos los totales por renglón en cada nivel $\{y_{i+k}\}$, la función de densidad de probabilidad conjunta queda dada por

$$f(\underline{y};\underline{\theta}\ |\{y_{i+k}\}) = \prod_{ik} y_{i+k}!\ \prod_{j} \theta_{ijk}^{y_{ijk}} \,/\, y_{ijk}! \qquad (2.2.1.3\text{-}1)$$

con $\sum_{j} \theta_{ijk} = 1$ para cada combinación de i y k.

A partir de (2.2.1.2-2) puede concluirse que

$$\mathcal{E}(Y_{ijk}) = y_{i+k}\,\theta_{ijk} \qquad (2.2.1.3\text{-}2)$$

Si sólo se mantiene fijo el total en cada nivel o subtabla $\{y_{++k}\}$ la distribución queda dada por

$$f(\underline{y};\underline{\theta}\ |\{y_{++k}\}) = \prod_{k} y_{++k}!\ \prod_{i}\prod_{j} \theta_{ijk}^{y_{ijk}} \,/\, y_{ijk}! \qquad (2.2.1.3\text{-}3)$$

con $\sum_{i}\sum_{j} \theta_{ijk} = 1$ para cada nivel de k. Puede concluirse que

$$\mathcal{E}(Y_{ijk}) = y_{++k}\,\theta_{ijk} \qquad (2.2.1.3\text{-}2)$$

2.2.1.4 MODELO LOGLINEAL

El planteamiento de hipótesis que pueden ser formuladas como modelos multiplicativos en los cuales los valores esperados de las casillas quedan dados por el producto de probabilidades marginales y frecuencias marginales fijas, permite el planteamiento de modelos loglineales como caso particular de los modelos lineales generalizados. Por ejemplo, si se considera la hipótesis de independencia mutua y la distribución multinomial, de (2.2.1.2-2) se tiene que

$$\mathcal{E}(Y_{ijk}) = N\ \theta_{i++}\ \theta_{+j+}\ \theta_{++k} \qquad (2.2.1.4\text{-}1)$$

Tales hipótesis sugieren el uso de modelos lineales generalizados con función de liga logarítmica

$$\eta_i = \ln \mathcal{E}(Y_i) = \underline{X}_i'\ \underline{\beta} \qquad i=1,2,\ldots,n$$

donde (2.2.1.4-1) puede ser expresada por

$$\eta_{ijk} = \ln \mathcal{E}(Y_{ijk}) = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} \qquad (2.2.1.4\text{-}2)$$

En analogía con el análisis de varianza el modelo saturado para $\mathcal{E}(Y_{ijk}) = N\ \theta_{ijk}$ puede expresarse como

$$\begin{aligned}\eta_{ijk} &= \ln \mathcal{E}(Y_{ijk}) \\ &= \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \\ &\quad + \mu_{13(ik)} + \mu_{23(jk)} + \mu_{123(ijk)} \qquad (2.2.1.4\text{-}3)\end{aligned}$$

En expresiones para las frecuencias esperadas de las casillas para el modelo multinomial o producto multinomial algunos términos son constantes fijas ej. N en (2.2.1.2-2) y y_{i+k} en (2.2.1.3-2). Tal situación implica que los correspondientes parámetros deben considerarse siempre en el modelo. Así por ejemplo, el modelo saturado correspondiente a (2.2.1.3-2) es (2.2.1.4-3) donde la expresión

$$\mu + \mu_{1(i)} + \mu_{3(k)} + \mu_{13(ik)} \qquad (2.2.1.3\text{-}4)$$

corresponde a la marginal fija y_{i+k}, la cual no debe omitirse del modelo.

2.3 ESTADISTICOS SUFICIENTES

En la sección anterior se mostró que cuando $\sum m_{ijk} = N$ el Kernel de la función de verosimilitud es idéntico para las tres distribuciones muestrales. Sin embargo, puede mirarse una demostración más completa de que los estimadores de máxima verosimilitud son idénticos para los tres esquemas en [Bishop, Fienberg & Holland,1975] Teorema 13.4-1.

Una generalización de los resultados obtenidos para una dimensión mayor puede realizarse si se considera un conjunto de subíndices θ dividido en dos subcunjuntos θ_1 y θ_2 y una configuración C_{θ_1} fija que conduce a una expresión de distribución producto multinomial

$$f(\{x_\theta\} \mid \{m_{\theta_1} = x_{\theta_1}\}) = \prod_{\theta_1} \left[\left(x_{\theta_1}! / \prod_{\theta_2} x_\theta! \right) \prod_{\theta_2} \left(m_\theta / x_{\theta_1} \right)^{x_\theta} \right] \qquad (2.3\text{-}1)$$

Así el Kernel queda dado por

$$\sum_\theta x_\theta \ln m_\theta \qquad (2.3\text{-}2)$$

independientemente de que la distribución pertenezca a cualquiera de los tres esquemas muestrales siempre que $\sum \hat{m}_\theta = N$.

A partir de la ecuación (2.1.1.3) se puede deducir que el Kernel de la función permite la obtención de los estadísticos suficientes para el modelo, dado que los tres esquemas de distribución pertenecen a la familia exponencial de distribuciones.

2.3.1 ESTIMADORES SUFICIENTES EN TRES DIMENSIONES

Teniendo en cuenta que si $\sum \hat{m}_{ijk} = N$ en el caso tridimensional, se tiene que el Kernel de la función es idéntico para los tres esquemas de distribución y si se considera el modelo loglineal saturado

$$\ln m_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{13(ik)} + \mu_{23(jk)} + \mu_{123(ijk)} \qquad (2.3.1\text{-}1)$$

entonces sustituyendo (2.3.1-1) en (2.1.2-5) se tiene que

$$\sum_{ijk} x_{ijk} \ln m_{ijk} = N\mu + \sum_{i} x_{i++}\, \mu_{1(i)} + \sum_{j} x_{+j+}\, \mu_{2(j)} + \sum_{k} x_{++k}\, \mu_{3(k)} + \sum_{ij} x_{ij+}\, \mu_{12(ij)} + \sum_{ik} x_{i+k}\, \mu_{13(ik)} + \sum_{jk} x_{+jk}\, \mu_{23(jk)} + \sum_{ijk} x_{ijk}\, \mu_{123(ijk)} \qquad (2.3.1\text{-}2)$$

y a partir del resultado (2.1.1-3) se deduce que los términos x's adyacentes a los términos μ representan los estadísticos suficientes para los parámetros desconocidos del modelo.

Si se postula el modelo sin el efecto conjunto de tres factores ($\mu_{123(ijk)} = 0$ para todo i,j,k) el último término de la expresión (2.3.1-2) desaparece y los elementos N, x_{i++}, x_{+j+}, x_{++k}, x_{ij+}, x_{i+k} y x_{+jk} son los estadísticos suficientes para el modelo.

2.3.2 ESTADISTICOS SUFICIENTES MINIMALES

El modelo loglineal asociado a una determinada hipótesis puede ser inspeccionado para obtener el conjunto minimal de estadísticos suficientes. A continuación se muestra una forma empírica de llegar a él.

Si se considera el modelo sin el efecto conjunto de tres factores ($\mu_{123} = 0$) y por simplicidad se elimina el segundo conjunto de subíndices, entonces:

$$\ell n\, m_{ijk} = \mu + \mu_1 + \mu_2 + \mu_3 + \mu_{12} + \mu_{13} + \mu_{23} \qquad (2.3.2\text{-}1)$$

puede ser reescrito en familias de términos agrupados

$$\ell n\, m_{ijk} = \mu + (\mu_1 + \mu_2 + \mu_{12}) + (\mu_1 + \mu_3 + \mu_{13}) + (\mu_2 + \mu_3 + \mu_{23}) - (\mu_1 + \mu_2 + \mu_3) \qquad (2.3.2\text{-}2)$$

lo cual puede a su vez escribirse en una notación más general

$$\ell n\, m_{ijk} = \mu + U_{12} + U_{13} + U_{23} - (\mu_1 + \mu_2 + \mu_3)$$

y la expresión (2.3.1-2) queda de la siguiente forma

$$\sum_{ijk} x_{ijk}\, \ell n\, m_{ijk} = N\mu + \left[\sum_{ij} x_{ij+}\, U_{12} + \sum_{ik} x_{i+k}\, U_{13} + \sum_{jk} x_{+jk}\, U_{23} \right] - \left[\sum_{i} x_{i++}\, \mu_1 + \sum_{j} x_{+j+}\, \mu_2 + \sum_{k} x_{++k}\, \mu_3 \right] \qquad (2.3.2\text{-}3)$$

Los términos en el primer conjunto de paréntesis cuadrados conducen al conjunto minimal de estadísticos suficientes debido a que los términos en el segundo conjunto son redundantes, así entonces U_{12}, U_{13} y U_{23} tienen sus correspondientes configuraciones $C_{12} = \{x_{ij+}\}$, $C_{13} = \{x_{i+k}\}$ y $C_{23} = \{x_{+jk}\}$ como el conjunto minimal de estadísticos suficientes.

Si se considera el modelo con un efecto conjunto de dos factores ($\mu_{23} = \mu_{123} = 0$) se tiene que

$$\begin{aligned} \ln m_{ijk} &= \mu + \mu_1 + \mu_2 + \mu_3 + \mu_{12} + \mu_{13} \\ &= \mu + (\mu_1 + \mu_2 + \mu_{12}) + (\mu_1 + \mu_3 + \mu_{13}) - \mu_1 \end{aligned} \tag{2.3.2-4}$$

lo cual implica que (2.3.2-3) se determine por

$$\sum_{ijk} x_{ijk} \ln m_{ijk} = N\mu + \left[\sum_{ij} x_{ij+} U_{12} + \sum_{ik} x_{i+k} U_{13} \right] - \left[\sum_{i} x_{i++} \mu_1 \right] \tag{2.3.2-5}$$

El conjunto minimal de estadísticos suficientes queda dado por sus correspondientes configuraciones:

$$C_{12} = \{x_{ij+}\} \text{ y } C_{13} = \{x_{i+k}\}$$

A continuación se muestran las configuraciones suficientes minimales para los cuatro modelos comprensibles y jerárquicos.

Tabla 2.3.2-1: Configuraciones suficientes

Términos ausentes	*Configuraciones suficientes*
μ_{123}	C_{12}, C_{13} y C_{23}
μ_{23}, μ_{123}	C_{12} y C_{13}
$\mu_{13}, \mu_{23}, \mu_{123}$	C_{12} y C_{3}
$\mu_{12}, \mu_{13}, \mu_{23}, \mu_{123}$	C_{1}, C_{2} y C_{3}

En general, cuando se cuenta con un modelo multidimensional, para obtener el conjunto minimal de estadísticos suficientes, se seleccionan los términos μ de orden t donde t es la interacción conjunta de orden mayor en el modelo. Si todos los posibles términos de orden t están incluidos en el modelo la selección termina, de otra forma, se examinan los términos de orden t-1 y se seleccionan aquellos cuyo orden relativo menor no esté incluido en los términos de orden t.

El proceso continua en los niveles inferiores donde se eligien los términos no incluidos en los ordenes relativos superiores.

2.3.3 RESULTADOS DE BIRCH

Los dos resultados de Birch* permiten a partir del conjunto minimal de estadísticos suficientes obtener estimadores de máxima verosimilitud.

2.3.3.1 RESULTADO 1

Si x_{θ_i} es un elemento del conjunto de estadísticos suficientes, entonces el estimador de máxima verosimilitud $\hat{m}_{\theta_i}$ de m_{θ_i} es x_{θ_i}

Para el modelo (2.3.2-1), con $\mu_{123} = 0$ resulta que los estimadores de máxima verosimilitud para m_{ij+}, m_{i+k} y m_{+jk} son

$$\hat{m}_{ij+} = x_{ij+} \ , \ \hat{m}_{i+k} = x_{i+k} \text{ y } \hat{m}_{+jk} = x_{+jk} \qquad (2.3.3.1\text{-}1)$$

2.3.3.2 RESULTADO 2

Existe una solución única que conduce a estimadores positivos de máxima verosimilitud $\hat{m}_{\theta}$ para toda casilla elemental de la tabla. Tal solución satisface tanto las restricciones del modelo como la condición $\hat{m}_{\theta_i} = x_{\theta_i}$ para todo θ_i tal que x_{θ_i} es un elemento del conjunto de estadísticos suficientes.

Para el modelo (2.3.1-1) su correspondiente hipótesis de no existencia del efecto conjunto de tres factores equivale a plantear la igualdad de los efectos conjuntos de segundo orden:

$$(m_{ijk}\, m_{rsk}) \ / \ (m_{rjk}\, m_{isk}) = (m_{ijt}\, m_{rst}) \ / \ (m_{rjt}\, m_{ist})$$
$$\text{para } i \neq r,\ j \neq s \text{ y } k \neq t \qquad (2.3.3.2\text{-}1)$$

El segundo resultado de Birch establece que existe un único conjunto de estimadores para las casillas de la tabla que satisface las condiciones (2.3.3.1-1) y (2.3.3.2-1).

2.3.3.3 IMPLICACIONES PARA EL ESQUEMA PRODUCTO MULTINOMIAL

Cuando se consideran los esquemas muestrales Poisson o Multinomial los resultados de Birch establecen que si C_{θ_1} es un estadístico suficiente para el modelo, entonces, los estimadores para las casillas satisfacen la condición $\hat{m}_{\theta_1} = x_{\theta_1}$, lo que significa que las configuraciones son fijas por el modelo.

Recíprocamente en el esquema producto multinomial la configuración C_{θ_2} es fijada por el esquema, entonces para lograr la consistencia bajo este esquema deben considerarse aquellos modelos para los cuales C_{θ_2} es fija, lo que a su vez implica que deben incluirse los términos μ_{θ_2} en el modelo.

En resumen, se puede concluir a partir de los resultados de Birch, que los estimadores de máxima verosimilitud son los mismos para los tres esquemas muestrales siempre que se incluyan en el modelo los parámetros correspondientes a las marginales fijas.

2.4 METODOS DE ESTIMACION PARA MODELOS LINEALES

Se tienen dos formas que son las más empleadas para la estimación de los parámetros de un modelo lineal, el método de máxima verosimilitud y el de mínimos cuadrados. Para el primero de ellos los valores de los estimadores de los parámetros son derivados a partir del valor máximo de la función de verosimilitud o equivalentemente de la función de log-verosimilitud, mientras que para el segundo, los valores de los estimadores son obtenidos a partir de la minimización de la suma de cuadrados de la diferencia entre las variables de respuesta y sus valores esperados.

2.4.1 METODO DE MAXIMA VEROSIMILITUD

Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias con función de densidad de probabilidad conjunta $f(\underline{y};\underline{\theta})$ donde $\underline{y}$ es de nx1 y $\underline{\theta}$ de px1, su funcion de verosimilitud $L(\underline{\theta};\underline{y})$ y Ω el espacio parametral para $\underline{\theta}$. El estimador de máxima verosimilitud de $\underline{\theta}$ es el valor $\underline{\hat{\theta}}$ que maximiza la función de verosimilitud, esto es,

$$L(\underline{\hat{\theta}};\underline{y}) \geq L(\underline{\theta};\underline{y}) \text{ para todo } \underline{\theta} \in \Omega \qquad (2.4.1\text{-}1)$$

Equivalentemente $\underline{\hat{\theta}}$ es también el valor que maximiza la función de log-verosimilitud $\ell(\underline{\theta};\underline{y}) = \ell n\, L(\underline{\theta};\underline{y})$ dada la monotonía de la función logaritmo natural, esto es,

$$\ell(\underline{\hat{\theta}};\underline{y}) \geq \ell(\underline{\theta};\underline{y}) \text{ para todo } \underline{\theta} \in \Omega \qquad (2.4.1\text{-}2)$$

Usualmente el estimador $\underline{\hat{\theta}}$ es obtenido mediante la derivación de la función de log-verosimilitud con respecto a cada θ_j lo cual conduce a su obtención como la solución al sistema de ecuaciones

$$\partial \ell(\underline{\theta};\underline{y})/\partial\theta_j = 0 \text{ para } j=1,2,\ldots,p \qquad (2.4.1\text{-}3)$$

$\hat{\underline{\theta}}$ será el máximo de la función siempre que $\partial^2 \ell(\underline{\theta};\underline{y})/\partial\theta_j\partial\theta_k$ evaluada en el punto $\underline{\theta} = \hat{\underline{\theta}}$ sea una matriz definida negativa.

Es importante notar la propiedad de invarianza de los estimadores máximo verosímiles, la cual implica que si $g(\underline{\theta})$ es una función de $\underline{\theta}$, entonces el estimador de $g(\underline{\theta})$ es $g(\hat{\underline{\theta}})$, otras propiedades de estos estimadores son la consistencia, la suficiencia y la eficiencia asintótica.

2.4.1.1 ESTIMACION MAXIMO VEROSIMIL EN EL MODELO LOGLINEAL

Si las casillas de la tabla se etiquetan como $Y_1, Y_2, \ldots, Y_n$ y se considera a partir de las ecuaciones (2.1.2-4),(2.1.3-3) y (2.1.4-3) que la función de verosimilitud para los tres esquemas muestrales puede ser expresada como el kernel (2.3-2) más una constante digamos C, esto es, si $m_i = \mathcal{E}(Y_i)$ entonces

$$\ell = \sum_i y_i \ln m_i + C \qquad (2.4.1.1\text{-}1)$$

Para algunos modelos loglineales pueden obtenerse los estimadores maximizando directamente la función de log-verosimilitud sujeta a las restricciones del modelo muestral, sin embargo, en ocasiones deberá emplearse algún procedimiento iterativo para la obtención de los estimadores tales como el ajuste iterativo proporcional o el derivado de Newton-Raphson. Tales procedimientos estiman $\underline{\beta}$ a partir del modelo loglineal

$$\eta_i = \ln \mathcal{E}(Y_i) = \underline{X}_i' \underline{\beta}$$

A partir del estimador de $\underline{\beta}$ se determinan los valores $\hat{m}_i$ y considerando la propiedad de invarianza, los valores ajustados $\hat{m}_i$ serán los estimadores máximo verosímiles de los valores esperados de las casillas $\mathcal{E}(Y_i)$.

2.4.2 ESTIMACION POR MINIMOS CUADRADOS

Sean $Y_1, Y_2, \ldots, Y_n$ variables aleatorias con valores esperados $\mu_i = \mathcal{E}(Y_i)$ y sean las μ_i' funciones de los parámetros $\beta_1, \beta_2, \ldots \beta_p$ $(p<n)$ y considere el modelo

$$Y_i = \mu_i + e_i \quad i=1,2,\ldots,n \qquad (2.4.2.1)$$

El método de mínimos cuadrados consiste en encontrar los estimadores $\hat{\underline{\beta}}$ que minimizan la suma de cuadrados de los errores, esto es, $\hat{\underline{\beta}}$ es el vector de valores que minimiza

$$S = \sum_i e_i^2 = \sum_i [Y_i - \mu_i(\underline{\beta})]^2 = (\underline{Y} - \underline{\mu})'(\underline{Y} - \underline{\mu}) \qquad (2.4.2.2)$$

Usualmente $\hat{\underline{\beta}}$ se obtiene mediante la derivación de S con respecto a cada β_j, lo cual conduce a la solución de las ecuaciones simultáneas

$$\partial S / \partial \beta_j = 0 \qquad j=1,2,\ldots,p \qquad (2.4.2.3)$$

Verificando que la solución corresponda realmente al mínimo (i.e. confirmando que la matriz de segundas derivadas sea definida positiva).

Una característica de este método es la aparente falta de supuestos al respecto de la distribución de las variables aleatorias; sin embargo, para obtener la distribución muestral de los estimadores $\hat{\underline{\beta}}$ es necesario el realizar suposiciones acerca de la distribución de las Y_i.

2.5 TECNICAS PARA LA ESTIMACION DE PARAMETROS

Para algunos modelos los estimadores de las casillas se pueden obtener de forma directa como funciones de los estadísticos suficientes; sin embargo, para otros modelos se requiere utilizar algunas técnicas iterativas para la estimación de los parámetros.

En la presente sección se muestra una técnica para obtener estimadores de forma directa, cuando ellos existen, aunque dadas las aplicaciones de computo existentes son de mayor relevancia las técnicas iterativas para la obtención de los estimadores, las cuales suelen converger hacia las mismas estimaciones.

Las técnicas presentadas son el ajuste iterativo proporcional y la técnica derivada del método Newton-Raphson para modelos lineales generalizados.

2.5.1 ESTIMADORES DIRECTOS

Considere el modelo loglineal en dos dimensiones bajo el supuesto de independencia ($\mu_{12}=0$)

$$\ln m_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} \qquad (2.5.1\text{-}1)$$

las configuraciones suficientes minimales para tal modelo son $C_1=\{x_{i+}\}$ y $C_2=\{x_{+j}\}$, a partir de los resultados de Birch se obtiene que $\hat{m}_{i+}=x_{i+}$ y $\hat{m}_{+j}=x_{+j}$. A partir de (2.5.1-1) se tiene que

$$m_{i+} = \exp(\mu + \mu_{1(i)})\sum_j \exp(\mu_{2(j)})$$

$$m_{+j} = \exp(\mu + \mu_{2(j)})\sum_i \exp(\mu_{1(i)})$$

$$N = e^{\mu}\left[\sum_{i} \exp(\mu_{1(i)})\right]\left[\sum_{j} \exp(\mu_{2(j)})\right] \qquad (2.5.1\text{-}2)$$

de las ecuaciones (2.5.1-2) se deduce que

$$m_{ij} = \exp\,(\mu + \mu_{1(i)} + \mu_{2(j)}) = m_{i+}\, m_{+j} \,/\, N \qquad (2.5.1\text{-}3)$$

Así entonces se deduce que los estimadores de máxima verosimilitud para los valores esperados de las casillas de la tabla son

$$\hat{m}_{ij} = x_{i+}\, x_{+j} \,/\, N \qquad (2.5.1\text{-}4)$$

Un procedimiento similar nos permite obtener en forma directa los estimadores para las casillas de la tabla en tres o más dimensiones cuando el modelo específico tenga asociadas una o dos configuraciones suficientes minimales y en algunos casos particulares con más de dos.

Considere en tres dimensiones el modelo loglineal bajo el supuesto de que $\mu_{23}=\mu_{123}=0$ entonces

$$m_{ijk} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{12(ij)} + \mu_{13(ik)}$$

Se sabe de (2.3.2-5) que las configuraciones suficientes minimales son $C_{12}=\{x_{ij+}\}$ y $C_{13}=\{x_{i+k}\}$ las cuales por los resultados de Birch son estimadores de

$$m_{ij+} = \exp(\mu+\mu_{1(i)}+\mu_{2(j)}+\mu_{12(ij)})\sum_{k} \exp(\mu_{3(k)}+\mu_{13(ik)}) \qquad (2.5.1\text{-}5)$$

$$m_{i+k} = \exp(\mu+\mu_{1(i)}+\mu_{3(k)}+\mu_{13(ik)})\sum_j \exp(\mu_{2(j)}+\mu_{12(ij)})$$

(2.5.1-6)

ambas tienen en común al vector $C_1=\{x_{i++}\}$ el cual es estimador de

$$m_{i++} = \exp(\mu+\mu_{1(i)})\sum_j \exp(\mu_{2(j)}+\mu_{12(ij)})\sum_k \exp(\mu_{3(k)}+\mu_{13(ik)})$$

(2.5.1-7)

A partir de (2.5.1-5) y (2.5.1-6) y (2.5.1-7) se obtiene

$$m_{ijk} = (m_{ij+}\ m_{i+k}) \ /\ m_{i++} \qquad (2.5.1\text{-}8)$$

De donde se obtiene que los estimadores de máxima verosimilitud para los valores esperados de las casillas de la tabla son

$$\hat{m}_{ijk} = x_{ij+}\ x_{i+k} \ /\ x_{i++} \qquad (2.5.1\text{-}9)$$

En [Bishop,Fienberg & Holland,1975] pag.78-82, se prueban los siguientes dos teoremas que permiten la determinación de la existencia de estimadores directos para un modelo determinado.

Teorema 2.5.1-1

Si el conjunto minimal de estadísticos suficientes para un modelo consiste unicamente de dos configuraciones, los estimadores pueden obtenerse de forma directa.

Teorema 2.5.1-2

Si cada una de las tres configuraciones tiene un conjunto diferente de variables en común con las restantes dos configuraciones, los estimadores directos no existen. Recíprocamente la estimación de manera directa es posible con tres configuraciones si:

1) Al menos dos no tienen variables en común,

2) Si se tiene un conjunto de variables en común a las tres configuraciones, cuando ellas son removidas la condición (1) es aplicable.

Una importante implicación es que los estimadores no pueden ser calculados de forma directa a menos que cuando menos un efecto de dos factores esté ausente.

2.5.2 AJUSTE ITERATIVO PROPORCIONAL

Un procedimiento que permite calcular los estimadores para los valores esperados de las casillas de la tabla particularmente útil cuando estos no pueden ser calculados de manera directa es el Ajuste Iterativo Proporcional, el cual tiene la propiedad de que cuando es usado para modelos que no requieren iteración conduce a los estimadores directos al final del primer ciclo. Así entonces, si se cuenta con un computador y una aplicación con este algoritmo no es necesario determinar si existen o no los estimadores de forma directa, ya que el procedimiento iterativo conduce siempre a los estimadores correctos.

La presente técnica fue derivada de un procedimiento llamado ajuste proporcional "clásico" atribuido a Deming-Stephan*, en el cual las casillas internas de la tabla son proporcionalmente ajustadas a partir de un conjunto de marginales obtenidas por una fuente distinta.

* Deming,W.E & Stephan,F.F. (1940). On a least squares adjustement of a sampled frecuency table when the expected marginal totals are known. Ann. Math. Statist. 11, 427-444.

A través del ajuste iterativo proporcional de las configuraciones suficientes pueden obtenerse los estimadores de máxima verosimilitud para los valores esperados de las casillas elementales bajo el supuesto de algún modelo jerárquico. Este método cuenta además con las siguientes propiedades:

1) Converge siempre al conjunto único de estimadores de máxima verosimilitud.

2) La exactitud en el cálculo de los estimadores puede regularse en el grado deseble mediante alguna regla de paro, en vez de sólo lograr exactitud en uno o dos estadísticos sumarios.

3) Los estimadores dependen unicamente de las configuraciones suficientes, no requiriéndose de alguna condición adicional para la estimación de las casillas sin registro de observaciones (ceros muestrales).

4) Cualquier conjunto de valores iniciales puede ser elegido para obtener los estimadores.

5) Si los estimadores directos existen, el procedimiento conduce a los estimadores exactos en el primer ciclo.

2.5.2.1 AJUSTE EN EL MODELO LOGLINEAL TRIDIMENSIONAL

Como implicación de los teoremas 2.5.1-1 y 2.5.1-2 el modelo con $\mu_{123} = 0$ es el único modelo jerárquico que no cuenta con estimación de manera directa y a partir del cual puede ejemplificarse el procedimiento y su convergencia.

PROCEDIMIENTO

Las configuraciones suficientes minimales para el modelo son C_{12}, C_{13} y C_{23} a partir de las cuales se ajustarán los estimadores. Se elige un conjunto de estimadores preliminares $\{\hat{m}^{(0)}_{ijk}\}$ como datos de arranque (se sugiere $\hat{m}^{(0)}_{ijk} = 1$ ya que $ln(\hat{m}^{(0)}_{ijk}) = 0$) y se realizan ajustes sucesivos para C_{12}, C_{13} y C_{23} de la siguiente forma:

Para C_{12}: $\hat{m}^{(1)}_{ijk} = \hat{m}^{(0)}_{ijk}\ x_{ij+}\ /\ \hat{m}^{(0)}_{ij+}$

Para C_{13}: $\hat{m}^{(2)}_{ijk} = \hat{m}^{(1)}_{ijk}\ x_{i+k}\ /\ \hat{m}^{(1)}_{i+k}$

Para C_{23}: $\hat{m}^{(3)}_{ijk} = \hat{m}^{(2)}_{ijk}\ x_{+jk}\ /\ \hat{m}^{(2)}_{+jk}$ (2.5.2.1-1)

Se repite el ciclo de tres pasos (iteración) hasta lograr una convergencia satisfactoria. Se puede considerardo una regla de paro en ciclo r-ésimo, cuando para algún δ (por ejemplo, δ=0.001) ocurra que:

$$\left|\hat{m}^{(3r)}_{ijk} - \hat{m}^{(3r-3)}_{ijk}\right| < \delta \text{ para todo } i,j,k \qquad (2.5.2.1\text{-}2)$$

CONVERGENCIA

[Brown,1959]* demostró que la función de verosimilitud de las $\{\hat{m}_{ijk}\}$ es una función monótona decreciente y que por tanto converge, lo cual permite deducir que pueden obtenerse siempre estimadores para las casillas con una aproximación adecuada. Una demostración no rigurosa de que los estimadores son únicos está dada en [Bishop, Fienberg & Holland,1975] pag. 85-86 y es como sigue:

Se define el logaritmo de la razón de verosimilitud

$$D^{(t)}_{123} = \sum_{ijk} \left(x_{ijk}\, \ln x_{ijk} - x_{ijk}\, \ln \hat{m}^{(t)}_{ijk}\right) \qquad (2.5.2.1\text{-}3)$$

---

* Brown,D.T. [1959] A note on a approximations to discrete probability distributions. Information and Control 2, 386-392.

dado que $\sum_{ijk} x_{ijk} \ln \hat{m}_{ijk}$ se minimiza para $\hat{m}_{ijk} = x_{ijk}$. Se deduce que $D_{123}^{(t)} \geq 0$ dado que $\hat{m}_{ijk}^{(t)} \geq 0$ y $\hat{m}_{+++}^{(t)} = x_{+++}^{(t)}$.

Si se procede a ajustar las configuraciones como en (2.5.2.1-1) resulta

$$\hat{m}_{ijk}^{(3r)} = \hat{m}_{ijk}^{(3r-1)} x_{+jk} / \hat{m}_{+jk}^{(3r-1)} \qquad (2.5.2.1\text{-}4)$$

y de aqui que

$$x_{ijk} \ln \hat{m}_{ijk}^{(3r)} = x_{ijk} \ln \hat{m}_{ijk}^{(3r-1)} + x_{ijk} \ln x_{+jk} + \\ - x_{ijk} \ln \hat{m}_{+jk}^{(3r-1)} \qquad (2.5.2.1\text{-}5)$$

De (2.5.2.1-3) y (2.5.2.1-5) resulta que

$$D_{123}^{(3r)} = D_{123}^{(3r-1)} - \sum_{jk} (x_{+jk} \ln x_{+jk} - x_{+jk} \ln \hat{m}_{+jk}^{(3r-1)})$$

$$D_{123}^{(3r)} = D_{123}^{(3r-1)} - D_{23}^{(3r-1)} \qquad (2.5.2.1\text{-}6)$$

$D_{23}^{(3r-1)}$ es obtenido de C_{23} y $\{\hat{m}_{+jk}^{(t)}\}$, por lo que es no negativo y de donde se concluye que

$$D_{123}^{(3r)} \leq D_{123}^{(3r-1)}$$

Similarmente se obtiene que

$$D_{123}^{(3r-1)} = D_{123}^{(3r-2)} - D_{23}^{(3r-2)} \qquad (2.5.2.1\text{-}7)$$

$$D_{123}^{(3r-2)} = D_{123}^{(3r-3)} - D_{23}^{(3r-3)} \qquad (2.5.2.1\text{-}8)$$

De las ecuaciones (2.5.2.1-6), (2.5.2.1-7) y (2.5.2.1-8) se deduce que

$$D_{123}^{(3r-3)} \geq D_{123}^{(3r-2)} \geq D_{123}^{(3r-1)} \geq D_{123}^{(3r)} \text{ para } r=1,2,3$$

con igualdad al paso t si y sólo si $\hat{m}_{ijk}^{(t)} = \hat{m}_{ijk}^{(t-1)}$ para todo i,j,k. La función está acotada inferiormente y converge al límite del valor para $\{\hat{m}_{ijk}\}$.

2.5.2.2 ALGORITMO GENERAL

Considere el conjunto minimal de configuraciones C_{θ_q}, $q=1,2,\ldots,s$ con las correspondientes marginales x_{θ_q}, se elige un conjunto inicial de estimadores $\hat{m}_\theta^{(0)}$ y se procede a ajustar cada una de las configuraciones. Al final del ciclo r se tiene

$$\hat{m}_\theta^{(sr+1)} = \hat{m}_\theta^{(sr)}\ x_{\theta_1} \ /\ \hat{m}_{\theta_1}^{(sr)\prime} \qquad (2.5.2.2\text{-}1)$$

En general al t-ésimo paso resulta

$$\hat{m}_\theta^{(t)} = \hat{m}_\theta^{(t-1)}\ x_{\theta_q} \ /\ \hat{m}_{\theta_q}^{(t-1)\prime} \qquad (2.5.2.2\text{-}2)$$

donde t-q es un múltiplo de s.

Después de ajustar t configuraciones, la razón de verosimilitud queda dada por

$$D_\theta^{(t)} = \sum_\theta x_\theta \ln x_\theta - \sum_\theta x_\theta \ln \hat{m}_\theta^{(t)} \qquad (2.5.2.2\text{-}3)$$

la cual es una función monótona decreciente como se mostró en tres dimensiones.

Si se considera $\hat{m}_\theta^{(t)}=1$ se obtiene el máximo valor $D_\theta^{(0)}$ y parece razonable que valores iniciales que den un valor menor $D_\theta^{(t)}$ podrían ser desventajosos en la velocidad de convergencia. Una regla de paro podría ser cuando para toda casilla ocurra que

$$\left| \hat{m}_\theta^{(st)} - \hat{m}_\theta^{(st-s)} \right| < \Delta m_\theta \qquad (2.5.2.2\text{-}4)$$

para algún incremento Δm_θ apropiado (0.001, 0.0001 etc.)

Otra regla de paro a partir de la razón de verosimilitud es

$$D_\theta^{(t-s)} - D_\theta^{(t)} < \Delta_D \qquad (2.5.2.2\text{-}5)$$

2.5.3 METODO DE NEWTON-RAPHSON

El ajuste iterativo proporcional es una técnica especializada para modelos loglineales de tipo jerárquico, mientras que la presente técnica permite la aproximación de los estimadores de máxima verosimilitud para un modelo lineal generalizado, donde el modelo loglineal puede ser ajustado por ser un caso especial de esta clase de modelos.

Si se considera un modelo lineal generalizado basado en un conjunto de variables aleatorias independientes $Y_1, Y_2, \ldots, Y_n$ todas ellas con una misma distribución perteneciente a la familia exponencial, la función de log-verosimilitud es

$$\ell(\underline{\theta};\underline{y}) = \sum_i y_i b(\theta_i) + \sum_i c(\theta_i) + \sum_i d(y_i) \qquad (2.5.3\text{-}1)$$

además

$$\mathcal{E}(Y_i) = \mu_i = -\,c'(\theta_i) \;/\; b'(\theta_i) \qquad (2.5.3\text{-}2)$$

y

$$\eta_i = g(\mu_i) = \underline{X}_i' \; \underline{\beta} \qquad (2.5.3\text{-}3)$$

donde g es una función monótona y diferenciable.

Una propiedad de la familia exponencial es que garantiza las condiciones de regularidad que permiten que el máximo global de la función de log-verosimilitud quede dado por la solución al sistema de ecuaciones simultaneas

$$\partial\ell/\partial\theta_i = 0 \qquad (2.5.3\text{-}4)$$

o equivalentemente a la solución del sistema de ecuaciones

$$\partial\ell/\partial\beta_j = 0 \qquad (2.5.3\text{-}5)$$

Ver por ejemplo [Cox & Hinkley,1974], cap 9).

Puede probarse que (ver [Dobson,1990], sec. 3.2.3 pag. 28-30 y anexos A y B)

$$\partial \ell/\partial \beta_j = U_j = \sum_i \left[(Y_i-\mu_i)x_{ij} \;/\; \mathrm{var}(Y_i)\right] \left(\partial \mu_i/\partial \eta_i\right) \qquad (2.5.3\text{-}6)$$

En general las ecuaciones $U_j=0$ son no lineales y deberán resolverse mediante iteración numérica.

En particular si el método de Newton-Raphson es usado, la m-ésima aproximación quedará dada por

$$\underline{b}^{(m)} = \underline{b}^{(m-1)} - \left[\partial^2 \ell/\partial \beta_j \partial \beta_k\right]^{-1}_{\beta=b^{(m-1)}} \underline{U}^{(m-1)} \qquad (2.5.3\text{-}7)$$

donde $\left[\partial^2 \ell/\partial \beta_j \partial \beta_k\right]^{-1}_{\beta=b^{(m-1)}}$ es la matriz de segundas derivadas de ℓ evaluadas en el punto $\underline{\beta}=\underline{b}^{(m-1)}$ y $\underline{U}^{(m-1)}$ es el vector de primeras derivadas con $U_j = \partial \ell/\partial \beta_j$ evaluadas en el mismo punto.

Esta es la versión multidimensional del método de Newton Raphson para encontrar la solución a una ecuación de una sola variable ($f(x)=0$) donde la m-ésima iteración queda dada por

$$x^{(m)} = x^{(m-1)} - [f(x^{(m-1)})/ f'(x^{(m-1)})] \qquad (2.5.3\text{-}8)$$

Una variación del método para la ecuación (2.5.3-7) puede realizarse utilizando la función de log-verosimilitud mediante la serie de Taylor evaluada en el punto $\underline{\beta}=\underline{\beta}^*$. La ecuación utilizada es

$$\ell(\underline{\beta};\underline{y}) = \ell(\underline{\beta}^*;\underline{y}) + (\underline{\beta}-\underline{\beta}^*)'\underline{U} + \frac{1}{2}(\underline{\beta}-\underline{\beta}^*)'H(\underline{\beta}-\underline{\beta}^*) \qquad (2.5.3\text{-}9)$$

Donde $\underline{U}$ es un vector px1 con elementos $U_j = \partial \ell / \partial \beta_j$ y H es una matriz de pxp con elementos $\partial^2 \ell / \partial \beta_i \partial \beta_k$ evaluados ambos en el punto $\underline{\beta} = \underline{\beta}^*$.

CAPITULO III: SELECCION DE MODELOS

En los capítulos previos se trataron los temas acerca de la especificación de los modelos y la estimación de los parámetros, en el presente capítulo se muestran algunas consideraciones para realizar una selección apropiada de un modelo que responda a la estructura de los datos de la tabla de contingencia.

En una primera instancia se presenta la distribución muestral para estimadores de máxima verosimilitud en un modelo lineal generalizado y el estadístico derivado de la razón de verosimilitud. Posteriormente se discuten algunas propiedades de las medidas de bondad de ajuste y los grados de libertad asociados al modelo.

Se detalla acerca de la bondad de ajuste interna de un modelo, de la detección de datos discrepantes mediante el análisis de residuos y de la evaluación de la magnitud de una interacción.

Finalmente, se tratan algunas estrategias para la comparación de modelos, algunos cálculos iniciales, procedimientos por pasos "stepwise" y aspectos prácticos para la selección de un mejor modelo.

3.1 DISTRIBUCION MUESTRAL DE ESTADISTICOS MAXIMOVEROSIMILES

Sea Y una variable aleatoria con función de densidad de probabilidad $f(y,\theta)$ y función de log-verosimilitud $\ell(\theta,y)$. Si se define $U = \partial\ell/\partial\theta$ entonces

$$\mathcal{E}(U) = \int \left[\partial\ell(\theta,y)/\partial\theta\right] f(y,\theta)\, dy \qquad (3.1\text{-}1)$$

si se considera que

$$\frac{\partial\ell(\theta,y)}{\partial\theta} = \frac{\partial \ln f(y,\theta)}{\partial\theta} = \frac{1}{f(y,\theta)} \frac{\partial f(y,\theta)}{\partial\theta} \qquad (3.1\text{-}2)$$

bajo ciertas condiciones de regularidad de $f(y,\theta)$ resulta

$$\mathcal{E}(U) = \int \left[\partial f(\theta,y)/\partial\theta\right] dy = \frac{\partial}{\partial\theta} \int f(y,\theta)\, dy = 0 \qquad (3.1\text{-}3)$$

Por otra parte, a partir de la igualdad (3.1-2) se obtiene

$$\frac{\partial}{\partial\theta} \int \left[\partial\ell(\theta,y)/\partial\theta\right] f(y,\theta)\, dy = (\partial^2/\partial\theta^2) \int f(y,\theta)\, dy$$

El término de la derecha es igual a cero y el de la izquierda puede reescribirse de la siguiente forma

$$\int \left[\partial^2\ell(\theta,y)/\partial\theta^2\right] f(y,\theta)dy + \int \left[\partial\ell(\theta,y)/\partial\theta\right]\left[\partial f(y,\theta)/\partial\theta\right]dy = 0$$

sustituyendo (3.1-2) en el segundo sumando se tiene

$$-\int \left[\partial^2\ell(\theta,y)/\partial\theta^2\right] f(y,\theta)dy = \int \left[\partial\ell(\theta,y)/\partial\theta\right]^2 f(y,\theta)\, dy$$

De aquí que

$$\mathcal{E}\left[-\partial^2\ell(\theta,y)/\partial\theta^2\right] = \mathcal{E}\left\{\left[\partial\ell(\theta,y)/\partial\theta\right]^2\right\}$$

lo que finalmente implica que

$$\mathrm{var}(U) = \mathcal{E}(U^2) = \mathcal{E}(-U') \tag{3.1-4}$$

dado que $\mathcal{E}(U)=0$, donde U' denota $\partial U/\partial\theta$.

Considere ahora las condiciones de un modelo lineal generalizado. Sean $Y_1,Y_2,\ldots,Y_n$ variables aleatorias independientes todas con la misma distribución perteneciente a la familia exponencial, sea $f(\underline{y};\underline{\theta})$ la función de densidad de probabilidad conjunta y g una función de liga monótona tal que $g(\mu_i) = \underline{X}_i'\underline{\beta}$ donde $\mu_i=\mathcal{E}(Y_i)$ y sea, además, $\ell(\underline{\theta};\underline{y})$ la función de log-verosimilitud.

Dado que las variables aleatorias tienen la misma distribución puede notarse que

$$\ell(\underline{\theta};\underline{y}) = \sum_i \ell_i(\theta;y_i) \tag{3.1-5}$$

se define

$$U_j = \partial\ell(\underline{\theta};\underline{y})/\partial\theta_j = \sum_i (\partial\ell_i(\theta;y_i)/\partial\theta_j) \tag{3.1-6}$$

por el mismo argumento que (3.1.2) resulta que

$$\mathcal{E}(\partial\ell_j(\theta;y_j)/\partial\theta_j) = 0$$

y, por lo tanto,

$$\mathcal{E}(U_j) = 0 \text{ para todo } j \qquad (3.1\text{-}7)$$

Considerando que las ecuaciones simultáneas $\partial\ell/\partial\underline{\theta} = \underline{0}$ son equivalentes a $\partial\ell/\partial\underline{\beta} = \underline{0}$ (ver [Cox & Hinkley,1974] cap 9), si se define $U_i = \partial\ell/\partial\beta_i$ para $i=1,2,\ldots,p$ para modelos lineales generalizados resulta que $\mathcal{E}(U_i) = 0$ para todo i.

Sea $\mathcal{J}$ la matriz de varianza-covarianza de las U_i, conocida también como matriz de información $\mathcal{J} = \mathcal{E}(\underline{U}\ \underline{U}')$. Por el teorema del límite central, al menos asintóticamente, $\underline{U} \sim N(\underline{0},\mathcal{J})$ y por tanto $\underline{U}'\mathcal{J}^{-1}\underline{U} \sim \chi^2_p$ siempre que $\mathcal{J}$ sea no singular y su inversa exista.

Si se supone que la función de log-verosimilitud tiene un máximo en $\hat{\underline{\beta}}$ y que tal estimador tiene un valor cercano al parámetro real $\underline{\beta}$, entonces, a partir de la aproximación de la serie de Taylor para $U(\underline{\beta})$ en el punto $\underline{\beta}=\hat{\underline{\beta}}$ se tiene que

$$U(\underline{\beta}) \cong U(\hat{\underline{\beta}}) + H(\hat{\underline{\beta}})\ (\underline{\beta} - \hat{\underline{\beta}}) \qquad (3.1\text{-}8)$$

donde $H(\hat{\underline{\beta}})$ denota la matriz de segundas derivadas de la función de log-verosimilitud evaluada en el punto $\underline{\beta}=\hat{\underline{\beta}}$.

Notese además que los elementos de la matriz de información

$$\mathcal{J}_{jk} = \mathcal{E}(U_jU_k) = \mathcal{E}[(\partial\ell/\partial\theta_j)(\partial\ell/\partial\theta_k)]$$

Por argumentos semejantes a los realizados para obtener (3.1-4), se tiene que

$$\mathcal{E}[(\partial\ell/\partial\theta_j)(\partial\ell/\partial\theta_k)] = \mathcal{E}[-\partial^2\ell/\partial\theta_j\partial\theta_k]$$

y de aquí que los elementos de la matriz de información queden dados por

$$\mathcal{I}_{jk} = \mathcal{E}[-\partial^2 \ell / \partial\theta_j \partial\theta_k] \qquad (3.1\text{-}9)$$

de donde se deduce finalmente que, al menos asintóticamente,

$$\mathcal{I} = \mathcal{E}(\underline{U}\ \underline{U}') = \mathcal{E}(-H). \qquad (3.1\text{-}10)$$

A partir de (3.1-8) y (3.1-10) para muestras grandes se tiene que

$$U(\underline{\beta}) \cong U(\hat{\underline{\beta}}) - \mathcal{I}\ (\underline{\beta} - \hat{\underline{\beta}})$$

Pero $U(\hat{\underline{\beta}}) = \underline{0}$ dado que $\hat{\underline{\beta}}$ es el punto donde la función de log-verosimilitud tiene su máximo, entonces

$$\mathcal{I}^{-1}\ U(\underline{\beta}) \cong (\underline{\beta} - \hat{\underline{\beta}}).$$

Si a $\mathcal{I}^{-1}$ se le considera constante y dado que $\mathcal{E}(\underline{U}) = \underline{0}$,

$$\mathcal{E}(\hat{\underline{\beta}} - \underline{\beta}) \cong \mathcal{I}^{-1}\ \mathcal{E}(\underline{U}) = \underline{0} \qquad (3.1\text{-}11)$$

así entonces $\hat{\underline{\beta}}$ resulta un estimador insesgado de $\underline{\beta}$ al menos asintóticamente y la matriz de varianza para $\hat{\underline{\beta}}$ es

$$\mathcal{E}[(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})'] \cong \mathcal{I}^{-1}\ \mathcal{E}(\underline{U}\ \underline{U}')\ \mathcal{I}^{-1} = \mathcal{I}^{-1} \qquad (3.1\text{-}12)$$

dado que $\mathcal{I} = \mathcal{E}(\underline{U}\ \underline{U}')$ y la simetría de $\mathcal{I}^{-1}$, se tiene que, al menos para muestras grandes el estadístico de Wald queda dado por la siguiente expresión

$$(\underline{\hat{\beta}} - \underline{\beta})' \, \mathcal{J} \, (\underline{\hat{\beta}} - \underline{\beta}) \sim \chi^2_p \qquad (3.1\text{-}13)$$

o equivalentemente

$$(\underline{\hat{\beta}} - \underline{\beta}) \sim N(\underline{0}, \mathcal{J}^{-1}) \qquad (3.1\text{-}14)$$

3.2 ADECUACIA DE UN MODELO LINEAL GENERALIZADO

Lo adecuado de un modelo para describir un conjunto de datos puede ser establecido comparando la verosimilitud del modelo saturado o maximal respecto a la verosimilitud del modelo en cuestión, para lo cual puede ser utilizada como medida de bondad de ajuste la razón de verosimilitud

$$\lambda = L(\underline{\hat{\beta}}_{max};\underline{y}) \;/\; L(\underline{\hat{\beta}};\underline{y}) \qquad (3.2\text{-}1)$$

o equivalentemente

$$ln\ \lambda = \ell(\underline{\hat{\beta}}_{max};\underline{y}) - \ell(\underline{\hat{\beta}};\underline{y}) \qquad (3.2\text{-}2)$$

donde $\underline{\hat{\beta}}_{max}$ y $\underline{\hat{\beta}}$ son los estimadores de máxima verosimilitud para el modelo maximal y el modelo en cuestión respectivamente, $L(\underline{\hat{\beta}}_{max};\underline{y})$ y $L(\underline{\hat{\beta}};\underline{y})$ las funciones de verosimilitud, $\ell(\underline{\hat{\beta}}_{max};\underline{y})$ y $\ell(\underline{\hat{\beta}};\underline{y})$ las funciones de log-verosimilitud. Si el modelo de interés describe adecuadamente los datos se espera que $L(\underline{\hat{\beta}};\underline{y})$ tendrá un valor cercano a $L(\underline{\hat{\beta}}_{max};\underline{y})$, así entonces valores grandes de $ln\ \lambda$ sugerirán que el modelo en cuestión no es adecuado.

Para determinar una región crítica se requiere conocer la distribución muestral de $ln\ \lambda$. Suponiendo que el modelo de interés involucra p parámetros denotados por $\underline{\beta}$, entonces, una aproximación mediante la serie de Taylor para $\ell(\underline{\hat{\beta}};\underline{y})$ conduce a la expresión

$$\ell(\underline{\beta};\underline{y}) \approx \ell(\underline{\hat{\beta}};\underline{y})+(\underline{\beta} - \underline{\hat{\beta}})'U(\underline{\hat{\beta}})+\frac{1}{2}(\underline{\beta} - \underline{\hat{\beta}})'H(\underline{\hat{\beta}})(\underline{\beta} - \underline{\hat{\beta}}) \qquad (3.2.3)$$

donde $U(\underline{\hat{\beta}})$ es el vector $[\partial\ell/\partial\beta_j]$ evaluado en el punto $\underline{\beta}=\underline{\hat{\beta}}$ y

$H(\hat{\underline{\beta}})$ es la matriz de segundas derivadas $[\partial^2 \ell / \partial\beta_j \partial\beta_k]$ evaluada en el mismo punto.

Dado que $U(\hat{\underline{\beta}}) = \underline{0}$ y para muestras grandes $\mathcal{J} = \mathcal{E}(-H)$, la ecuación (3.2.3) puede reescribirse como

$$\ell(\hat{\underline{\beta}};\underline{y}) - \ell(\underline{\beta};\underline{y}) \approx \frac{1}{2}(\hat{\underline{\beta}} - \underline{\beta})' \mathcal{J} (\hat{\underline{\beta}} - \underline{\beta})$$

dado que el término de la derecha de la ecuación es el estadístico de Wald (3.1-13), al menos asintóticamente,

$$2[\ell(\hat{\underline{\beta}};\underline{y}) - \ell(\underline{\beta};\underline{y})] \sim \chi^2_p \qquad (3.2\text{-}4)$$

Se define el estadístico devianza a partir de la razón de verosimilitud de la siguiente forma:

$$G^2 = 2\ \ln \lambda = 2[\ell(\hat{\underline{\beta}}_{max};\underline{y}) - \ell(\hat{\underline{\beta}};\underline{y})] \qquad (3.2\text{-}5)$$

la ecuación para G^2 puede ser reescrita com sigue

$$G^2 = 2\ [\ell(\hat{\underline{\beta}}_{max};\underline{y}) - \ell(\underline{\beta}_{max};\underline{y})] - 2\ [\ell(\hat{\underline{\beta}};\underline{y}) - \ell(\underline{\beta};\underline{y})]$$
$$+ 2\ [\ell(\underline{\beta}_{max};\underline{y}) - \ell(\underline{\beta};\underline{y})] \qquad (3.2\text{-}6)$$

El primer término entre paréntesis cuadrados se distribuye χ^2_n por (3.2-4), el segundo se distribuye χ^2_p y el tercero es una constante positiva la cual será cercana a cero si el modelo en cuestión es adecuado a los datos.

Si las variables aleatorias definidas por el primero y segundo términos en (3.2-6) son independientes y el tercer termino constante es cercano a cero entonces

$$G^2 \sim \chi^2_{n-p} \qquad (3.2\text{-}7)$$

Así entonces, si el modelo en cuestión es bueno G^2 tendrá distribución χ^2, en cambio si no lo es G^2 tenderá a ser mayor que el valor de la distribución χ^2correspondiente, de hecho tendrá aproximadamente un distribución χ^2 no central.

En general (3.2-7) no proporciona una excelente aproximación para las diferentes distribuciones muestrales, más que asintóticamente; pero en el caso de la distribución Normal, el resultado es exacto.

3.2.1 DEVIANZA PARA LOS MODELOS LOGLINEALES

Si se consideran las frecuencias de las casillas de la tabla etiquetadas como $x_1, x_2, \ldots, x_n$ y los valores esperados para las mismas $m_i = \mathcal{E}(x_i)$; $i=1,2,\ldots n$ para los tres esquemas muestrales se tiene bajo la condición $\sum_i x_i = \sum_i \hat{m}_i$, que la función de log-verosimilitud es

$$\ell(\underline{\hat{\beta}};\underline{x}) = \sum_i x_i \ln \hat{m}_i + C \qquad (3.2.1\text{-}1)$$

donde C es una constante que depende sólo del esquema muestral y $\hat{m}_i$ es el estimador de máxima verosimilitud para m_i.

De acuerdo con el segundo resultado de Birch para los tres esquemas muestrales, el modelo saturado considera que $\hat{m}_i = x_i$ lo que conduce a la función de verosimilitud

$$\ell(\underline{\hat{\beta}_{max}};\underline{x}) = \sum_i x_i \ln x_i + C \qquad (3.2.1\text{-}2)$$

De aquí que el estadístico devianza sea

$$G^2 = 2[\ell(\underline{\hat{\beta}_{max}};\underline{x}) - \ell(\underline{\hat{\beta}};\underline{x})] = 2\sum_i x_i \ln(x_i/\hat{m}_i) \qquad (3.2.1\text{-}3)$$

3.3 MEDIDAS DE BONDAD DE AJUSTE

Las principales medidas de bondad de ajuste para un modelo loglineal son la X^2 de Pearson y el estadístico de la razón de verosimilitud G^2, los cuales se distribuyen asintóticamente χ^2 con los apropiados grados de libertad.

Si se considera una tabla de contingencia $\{x_i\}$ y sus estimadores de los valores esperados $\{\hat{m}_i\}$, el estadístico de Pearson es

$$X^2 = \sum_i (x_i - \hat{m}_i)^2 / \hat{m}_i \qquad (3.3\text{-}1)$$

El estadístico de la razón de verosimilitud o devianza está dado por la expresión (3.2.1-3).

3.3.1 PARTICIONAMIENTO PARA MODELOS ANIDADOS

El estadístico de la razón de verosimilitud permite un particionamiento simple a partir de la ecuación (3.2.1-3). Si se consideran dos modelos M_1 y M_2, se dice que los modelos están anidados si M_2 contiene únicamente un subconjunto de los parámetros μ contenidos en M_1.

$$\begin{aligned} G^2(M_2) &= 2[\ell(\underline{\hat{\beta}}_{max};\underline{x}) - \ell(\underline{\hat{\beta}}_2;\underline{x})] \\ &= 2[\ell(\underline{\hat{\beta}}_1;\underline{x}) - \ell(\underline{\hat{\beta}}_2;\underline{x})] + 2[\ell(\underline{\hat{\beta}}_{max};\underline{x}) - \ell(\underline{\hat{\beta}}_1;\underline{x})] \\ &= G^2(M_2|M_1) + G^2(M_1) \qquad (3.3\text{-}2) \end{aligned}$$

puede notarse entonces que se cumple la condición

$$G^2(M_1) \le G^2(M_2) \qquad (3.3\text{-}3)$$

Si se considera (ver [Bishop,Fienberg & Holland,1975],cap 4) que para todo modelo jerárquico

$$\sum_i x_i \, \ln \, \hat{m}_i = \sum_i \hat{m}_i \, \ln \, \hat{m}_i \qquad (3.3\text{-}4)$$

y, más aún, si los términos μ de M_2 son subconjunto de los de M_1, se tiene

$$\sum_i \hat{m}_i^{(1)} \, \ln \, \hat{m}_i^{(2)} = \sum_i \hat{m}_i^{(2)} \, \ln \, \hat{m}_i^{(2)} \qquad (3.3\text{-}5)$$

La G^2 condicional resulta a partir de (3.3-2)

$$\begin{aligned} G^2(M_2|M_1) &= G^2(M_2) - G^2(M_1) \\ &= \sum_i x_i \, \ln(x_i/\hat{m}_i^{(2)}) - \sum_i x_i \, \ln(x_i/\hat{m}_i^{(1)}) \\ &= \sum_i x_i \, \ln \, \hat{m}_i^{(1)} - \sum_i x_i \, \ln \, \hat{m}_i^{(2)} \\ &= \sum_i \hat{m}_i^{(1)} \, \ln \, \hat{m}_i^{(1)} - \sum_i \hat{m}_i^{(2)} \, \ln \, \hat{m}_i^{(2)} \\ &= \sum_i \hat{m}_i^{(1)} \, \ln \, \hat{m}_i^{(1)} - \sum_i \hat{m}_i^{(1)} \, \ln \, \hat{m}_i^{(2)} \end{aligned}$$

Dado que $G^2(M_1)$ y $G^2(M_2)$ se distribuyen asintóticamente con χ^2 con ν_1 y ν_2 grados de libertad respectivamente, entonces $G^2(M_2|M_1)$ se distribuye asintóticamente χ^2con $(\nu_2-\nu_1)$ grados de libertad.

Utilizando el estadístico X^2 de Pearson, la diferencia entre dos modelos anidados $X^2(M_2)-X^2(M_2)$ no tiene la misma forma y tal diferencia no siempre es no negativa, un estadístico más apropiado para comparar modelos bajo este criterio sería

$$X^2(M_2|M_1) = \sum_i (\hat{m}_i^{(1)}-\hat{m}_i^{(2)})^2 \ / \ \hat{m}_i^{(2)} \qquad (3.3\text{-}7)$$

[Haberman,1977] muestra que $X^2(M_2|M_1)$ y $G^2(M_2|M_1)$ tienen similar distribución asintótica.

3.4 GRADOS DE LIBERAD

La mayoría de los estadísticos utilizados para detrminar la bondad de ajuste en el modelo tienen una distribución asintótica χ^2, donde el número de grados de libertad asociados a un determinado modelo depende de la estructura de la tabla y el número de parámetros independientes.

En general se tienen dos formas de calcular los grados de libertad para un modelo.

1) Se calcula el número de parámetros independientes iguales a cero.

Por ejemplo, en el caso tridimensional si se considera $\mu_{23}=\mu_{123}=0$ los grados de libertad se obtienen sumando el número de parámetros independientes para cada término μ ausente del modelo

$$(J-1)(K-1) + (I-1)(J-1)(K-1) = I(J-1)(K-1)$$

2) Se determina el número de parámetros estimados en el modelo (T_p) y se substrae de el total de casillas estimadas en la tabla (T_e)

Por ejemplo, si se considera el mismo caso anterior, para el modelo loglineal $\mu + \mu_1 + \mu_2 + \mu_3 + \mu_{12} + \mu_{13}$ los grados de libertad son

$$T_e-T_p = IJK - [1+(I-1)+(J-1)+(K-1)+(I-1)(J-1)+(I-1)(K-1)]$$

En la tabla 1.6.1-2 se muestra el número de parámetros independientes asociados con cada término μ en el modelo tridimensional. Puede notarse que es inmediata la generalización a una dimension mayor.

3.4.1 CEROS EN LA TABLA

En el primer capítulo se mencionó que en una tabla de contingencia puede ocurrir alguna frecuencia observada igual a cero de manera muestral o de forma estructural. Cuando se tiene el caso de los ceros muestrales, los grados de libertad se calculan de la misma forma que con la tabla completa, dado que habrán de calcularse los estimadores para las casillas con este tipo de observación.

En el caso de los ceros estructurales los grados de libertad sí son afectados por el número de casillas que presentan esta característica, así entonces el total de grados de libertad para un modelo con ceros estructurales se calcula como sigue:

$g\ell = T_1 - T_2 - T_3$

donde T_1 = número de casillas en la tabla

T_2 = número de parámetros a estimar

y T_3 = número de casillas con ceros estructurales

3.5 BONDAD DE AJUSTE INTERNA

En ocasiones bajo ciertas particularidades de los datos es interesante detectar posibles valores discrepantes o atípicos ("outliers"), es decir, valores observados que difieren considerablemente de los restantes valores esperados.

En particular, cuando se tiene un número grande de casillas en la tabla puede no resultar suficiente el obtener las medidas para determinar la bondad de ajuste de un modelo, en tales casos, es deseable evaluar el ajuste para cada casilla en particular. Hay dos razones por las que se recomienda hacer este análisis.

Por una parte, puede ocurrir que para un modelo que considere muchos parámetros, la medida de bondad de ajuste caiga dentro de límites aceptables aun cuando algunas casillas presenten un mal ajuste. La detección de tales casillas podría implicar la existencia de datos importantes para el entendimiento de la estructura de la tabla y sugerir, tal vez, algún posible particionamiento de la misma.

Por otra parte puede presentarse el caso de tener desviaciones positivas y negativas entre los valores esperados y los observados que hagan que las mediadas de bondad de ajuste caigan en rangos aceptables. Sin embargo, tal situación podría sugerir la búsqueda de un modelo más apropiado para los datos.

Además del examen de bondad de ajuste para cada casilla particular de la tabla, puede obtenerse una medida para la significación de los efectos, es decir, cuando se desea probar si dos variables específicas están relacionadas, se detrmina una magnitud de la significación del efecto de asociación representado por un parámetro del modelo, así entonces, puede resultar útil el determinar la magnitud relativa de los parámetros del modelo y con ello mejorar el entendimiento del mismo.

3.5.1 ANALISIS DE RESIDUOS

Una de las mejores técnicas para determinar la bondad de ajuste de cada casilla de la tabla respecto a un particular modelo es el análisis de residuos.

Dependiendo de la técnica disponible para el cálculo de los parámetros del modelo se tendrá información para obtener diferentes clases de residuos, por ejemplo, cuando se cuenta con el ajuste iterativo proporcional pueden analizarse residuos estandarizados componentes de los estadísticos X^2 o G^2. Cuando se dispone de una técnica derivada de Newton-Raphson puede tenerse información acerca de la varianza de las observaciones y por consiguiente poder calcular residuos ajustados.

Residuos estandarizados componentes de X^2

$$z_i^{(1)} = (x_i-\hat{m}_i)\ /(\hat{m}_i)^{1/2} \qquad (3.5.1\text{-}1)$$

$$z_i^{(2)} = 2\ x_i\ \ln(x_i/\hat{m}_i) \qquad (3.5.1\text{-}2)$$

Los estadísticos z_i son tales que cuando el modelo es apropiado $\sum_i z_i^2$ se distribuye asintóticamente como χ_ν^2 donde ν son los grados de libertad apropiados para la estimación de $\{\hat{m}_i\}$.

Una guia para determinar una desviación grande de la casilla respecto al modelo consiste en comparar cada z_i contra el valor teórico $\chi^2_{(\nu,\alpha)}$ dividido entre el número de casillas en la tabla donde α es algún nivel de significancia para la prueba.

Otras estrategias que permiten determinar la adecuacía interna del modelo mediante el uso de residuos estandarizados son:

1) Comparar los z_i contra el valor teórico de la distribución Normal Estandar.

2) Graficar los valores z_i contra los valores de cada parámetro del modelo; así, si el modelo describe apropiadamente

los datos no deberán mostrarse patrones definidos en la gráfica, si el modelo es inadecuado podrían detectarse patrones curvilíneos o alguna otra clase de tendencia sistemática.

3) Graficar los z_i contra los valores ajustados de las casillas $(\hat{m}_i)$. Al igual que en el punto anterior no deberán presentarse patrones definidos en el gráfico.

3.5.1.1 DETECCION DE VALORES DISCREPANTES

Cuando se tiene la sospecha de que la casilla (j) puede presentar un valor significativamente discrepante con respecto a un modelo dado, se sugiere seguir la siguiente estrategia de prueba.

Modelo (M_1): el ajustado originalmente a la tabla

Modelo (M_2):mismo modelo M_1 pero considerando la casilla (j) como si fuera un cero estructural.

La diferencia de bondad de ajuste, es decir, $G^2(M_1)-G^2(M_2)$ se distribuye asintóticamente como χ^2 con un grado de libertad. Así, si se obtiene un valor significativo respecto al valor teórico de la distribución se estará en presencia de un dato discrepante.

3.5.2 SIGNIFICACION DE LOS EFECTOS

Una de las principales ventajas de los modelos loglineales es la posibilidad de asignar una magnitud a los efectos de interacción entre las variables, es decir, en el estudio de un determinado conjunto de datos se desea determinar si una relación particular entre variables es significativa o no lo es.

La estrategia consiste en determinar si una o más categorías de una variable son responsables del rechazo de alguna hipótesis planteada, se realizan pruebas de hipótesis para cada cruce de las variables implicadas y cabe mencionar que con el hecho de encontrar significancia en al menos una de las pruebas para cada cruce se considerará que existe interacción.

Si por ejemplo se plantea la hipótesis

$$H_0\text{: } \mu_{12(ij)}=0 \text{ para todo } i,j$$

puede construirse un estadístico de prueba

$$z_{ij} = (\hat{\mu}_{12(ij)} - \mu_{12(ij)}) \;/\; var(\hat{\mu}_{12(ij)}) \qquad (3.5.2\text{-}1)$$

donde por el teorema del límite central z_{ij} tiene una distribución asintótica Normal estandar.

En caso de ser apropiada la suposición, z_{ij} se simplifica y toma la siguiente forma

$$z_{ij} = \hat{\mu}_{12(ij)} \;/\; var(\hat{\mu}_{12(ij)}) \qquad (3.5.2\text{-}2)$$

el cual puede ser comparado contra los valores de la distribución teórica de forma que si $|z_{ij}| > z_\alpha$ se considera que $\mu_{12(ij)}$ es significativamente diferente de cero con un nivel de confianza $1-\alpha$.

3.6 CONSIDERACIONES PARA LA PRUEBA DE HIPOTESIS

A continuación se presentan algunas sugerencias y consideraciones para cuando se desea realizar una prueba de hipótesis acerca de un conjunto de datos, lo cual es equivalente a realizar una selección de un modelo adecuado a los datos.

El investigador al trabajar con cuatro o mas variables se enfrenta a una enorme gama de posibilidades de elección de un modelo adecuado a los datos (113 modelos jerárquicos en cuatro dimensiones) y ante la inexistencia de un método inequívoco para elegir el mejor modelo se sugiere tomar en cuenta las siguientes consideraciones:

1) El invertigador debe tener bién definidos los propositos de la investigación, es decir, previamente apreciará cuales son las variables de mayor interés, así como las interrelaciones relevantes de acuerdo a su supuesto teórico que sospecha su existencia o por conocimiento de investigaciones previas, tal situación mejorará definitivamente la conducción en la selección de un modelo.

2) Deben tenerse presentes las limitaciones impuestas por el diseño muestral empleado para la recolección de datos, así como las aleatoriedades que suelen presentarse con tamaños de muestra muy reducidos o la presencia de ceros estructurales en la tabla. Tales situaciones conducen a incluir de manera definitiva algunos términos μ en el modelo asociados a las marginales fijas o a verificar el cálculo adecuado de los grados de libertad.

3) Se debe considerar seriamente la inclusión de cuatro ó más variables en la hipótesis ya que con ello la elección de un modelo adecuado puede resultar una labor bastante ardua, debido a que las alternativas de modelos diferentes crece geométricamente con el aumento en el número de variables y puede finalmente resultar muy costoso e impráctico ajustar todos los modelos posibles, esto sin contar el esfuerzo que implica el verificar la bondad de ajuste interna.

Sin embargo esta problemática no siempre se puede descartar, y en tal caso convendrá seguir alguna de las estrategias propuestas en la sección 3.6.1.

Otra inconveniencia en el manejo de múltiples variables es el hecho de enfrentarse a la dificultad de interpretar en la realidad la presencia de efectos conjuntos multifactores.

4) Debe tenerse presente que el probar hipótesis a priori y realizar a posteriori la selección de un modelo apropiado, todo ello con los mismos datos conduce a un consecuente efecto sobre el nivel de significancia de los estadísticos de prueba y por tanto debe tenerse en consideración el manejo del error tipo II en las pruebas, sin embargo, dadas las aplicaciones de computo disponibles poco puede hacerse para mejorar el control de este tipo de error. [Knoke & Burke,1980] sugieren algunas consideraciones prácticas para evitar que se incremente el error tipo II y éstas son el aumento del tamaño de la muestra o el acceder a un aumento en la probabilidad de error tipo I. La segunda alternativa trae consigo el dilema de la inclusión potencial de relaciones multifactores en el modelo cuando estas no debiesen ser incluidas por ser el reflejo de simples variaciones de carácter muestral. Los mismos autores insisten en que quizás la segunda opción sea la más factible y se opte por un modelo que ajuste a los datos con una probabilidad de error tipo I del 10% o incluso mayor.

3.6.1 ALGUNAS ESTRATEGIAS DE SELECCION

Si bién es cierto que el conocimiento teórico y la perspicacia del investigador son las mejores herramientas para la selección del modelo, en ocasiones se requiere que algunas rutinas mecánicas proporcionen mayor información acerca de las relaciones entre las variables.

Desafortunadamente no se cuenta con una estrategia única que permita hacer una selección óptima, por ello diversos autores sugieren diferentes formas de enfocar la selección de un modelo. No obstante, la mayoría proponen realizar algunos cálculos de rutina utilizando algún criterio sobre el nivel de significancia para obtener estadísticos sumarios que permitan en una primera instancia elegir uno o varios modelos base que a su vez den cabida al análisis de modelos muy parecidos a estos, es decir, modelos que difieren en pocos términos μ.

Dadas las facilidades que presentan las aplicaciones de cómputo actuales, es posible realizar una serie de calculos preliminares basados en la razón de verosimilitud (G^2), la cual puede obtenerse para una determinada clase de modelos y comparar mediante la propiedad de la partición condicional, modelos que difieran en algún término de interacción.

[Bishop, Fienberg & Holland,1975] proponen tres estrategias de cálculos iniciales basadas en el particionamiento estructural de la G^2 para la obtención de uno o varios modelos de arranque.

1) Ajuste de modelos con términos uniformes.

Si se considera un modelo de dimensión S entonces se ajustan los siguientes modelos uniformes:

(1) Modelo de completa independencia

(2) Modelo con todos los términos de interacción conjunta de dos variables.

...

(S-1) Modelo con todos los términos con (S-1) subíndices.

Se propone elegir dos modelos consecutivos de forma tal que uno de ellos no ajuste adecuadamente a los datos mientras que el otro sí lo hace, considerando un determinado nivel de significancia fijo. Una vez elegida la pareja de modelos se procede a analizar la bondad de ajuste de modelos intermedios.

2) Examinar los estadísticos G^2 para modelos cuyos estimadores puedan calcularse de manera directa, tal ajuste es útil cuando se tiene la sospecha de que una o más interacciones de primer orden pueden excluirse del modelo. La principal desventaja de esta estrategia se da cuando ninguno de los modelos directos ajusta a los datos, lo cual dificulta la decisión de un próximo paso a seguir.

3) Ajuste de modelos de asociación parcial, es decir, modelos con un efecto conjunto de dos factores (y terminos de orden relativo superior) igual a cero. La ventaja de ahondar en el estudio de este tipo de modelos es el poder determinar de alguna forma si se debe o no incluir cada uno de los términos de interacción. El siguiente paso en esta estrategia consiste en la construcción de un modelo que resuma la información de aquellos modelos de asociación parcial que presentan un buén ajuste.

Tomando como base modelos de arranque como se propone en las estrategias (1) y (3) pueden emplearse procedimientos por pasos "stepwise" (hacia adelante o hacia atrás) con el objeto de evaluar el efecto al agregar o retirar un término de interacción en el modelo base, de tal forma que se obtenga al final un modelo con un mínimo de parámetros que ajuste adecuadamente a los datos.

Cabe señalar que los procedimientos "stepwise" estarán limitados por el modelo que se utilice como base y que, además, cuando se utiliza un procedimiento hacia adelante no necesariamente converge al mismo modelo al que se llegaría con un procedimiento hacia atrás.

CAPITULO IV: MODELOS LOGIT

En capítulos anteriores se revisaron con mayor énfasis los modelos loglineales para los cuales, como puede notarse, no se requiere el distinguir entre variables de respuesta y variables de explicación, debido a que las frecuencias esperadas para las casillas de la tabla son modeladas en términos de asociación entre todas las variables; en contraste con ello, en el presente capítulo se presenta una serie de técnicas alternativas para el análisis de tablas de contingencia cuando una de las variables debe considerarse como respuesta.

En particular se describirán algunos modelos pertenecientes a la clase de modelos lineales generalizados en los cuales la variable de respuesta es dicotómica.

En una primera instancia se presentan algunos modelos denominados de dosis-respuesta tales como el de distribución de tolerancia uniforme, modelo probit, modelo logit y modelo de valor extremo, posteriormente se adentra en los modelos de regresión logística y en su relación con los modelos loglineales. Finalmente, se discuten algunos aspectos relativos a la bondad de ajuste y selección de un modelo apropiado.

4.1 MODELOS DE DOSIS-RESPUESTA

Considere una variable de respuesta dicotómica con los términos genéricos de "éxito" y "fracaso" para sus categorías y defínase la variable aleatoria

$$Z = \begin{cases} 1 \text{ si el resultado es éxito} \\ 0 \text{ si es fracso} \end{cases}$$

con $Pr(Z=1)=\pi$ y $Pr(Z=0)=1-\pi$. Si se tienen n variables aleatorias $Z_1, Z_2, \ldots, Z_n$ independientes con $Pr(Z_i=1)=\pi_i$ la función (Bernoulli) de probabilidad conjunta queda dada por:

$$\prod_{j=1}^{n} \pi_j^{z_j} (1-\pi_j)^{1-z_j} = \exp(\sum_{j=1}^{n} z_j \ln (\pi_j/(1-\pi_j)) + \sum_{j=1}^{n} (1-\pi_j) \qquad (4.1\text{-}1)$$

la cual es miembro de la familia exponencial y su liga natural es $\ln (\pi_j/(1-\pi_j)$.

Considerando el caso en que los π_j's son iguales, se define la variable aleatoria $Y = \sum_{j=1}^{n} z_j$ la cual tiene como se sabe una distribución binomial $b(n,\pi)$

$$Pr(Y=y) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \qquad y=0,1,2\ldots \qquad (4.1\text{-}2)$$

Finalmente, si se considera el caso de N variables independientes $Y_1, Y_2, \ldots, Y_N$ correspondientes al número de éxitos en N diferentes grupos o estratos, se sabe que la función de distribución conjunta pertenece a la familia exponencial (N variables aleatorias independientes con distribución binomial) cuya función de logverosimilitud es

$$\ell(\underline{\pi};\underline{y}) = \sum_{i=1}^{N}\left[y_i \ln(\pi_i/(1-\pi_i) + N_i \ln(1-\pi_i) + \ln\binom{N_i}{y_i}\right] \quad (4.1\text{-}3)$$

Con los modelos lineales generalizados se desea describir la proporción de éxitos $P_i = Y_i / N_i$ en cada subgrupo en términos de otras variables de explicación que caracterizan al subgrupo, lo cual se realiza modelando las probabilidades π_i como

$$g(\pi_i) = \underline{X}_i' \ \underline{\beta} \quad (4.1\text{-}4)$$

donde g es una función de liga $\underline{X}_i'$ el vector de variables de explicación y $\underline{\beta}$ el vector de parámetros.

El modelo lineal más simple es

$$\pi = \underline{X}_i' \ \underline{\beta} \quad (4.1\text{-}5)$$

el cual es usado en algunas aplicaciones prácticas con el incoveniente de que $\underline{X}_i' \ \underline{\beta}$ en ocaciones puede tomar valores fuera del intervalo [0,1].

Para asegurar que π esté restringido al intervalo [0,1] se suele utilizar una función de distribución acumulativa

$$\pi = g^{-1}(\underline{X}_i' \ \underline{\beta}) = \int_{-\infty}^{t} f(s)\ ds$$

donde $f(s) \geq 0$ y $\int_{-\infty}^{\infty} f(s)\ ds = 1$, la función de densidad de probabilidad f(s) se denomina función de tolerancia.

4.1.1 FUNCION DE TOLERANCIA UNIFORME

Uno de los primeros modelos semejantes a los del análisis de regresión empleando una variable de respuesta binomial es el que utiliza la distribución uniforme como función de tolerancia.

Para describir la probabilidad de éxito π como una función de dosis x se considera el modelo

$$g(\pi) = \beta_0 + \beta_1 x \tag{4.1.1-1}$$

y la distribución uniforme en el intervalo $[c_1, c_2]$

$$f(s) = (1/c_2-c_1) \text{ si } c_1 \leq s \leq c_2 \text{ y } f(s)=0 \text{ en otro caso.}$$

entonces

$$\pi = \int_{c_1}^{x} f(s)\, ds = (x-c_1)/(c_2-c_1) \quad \text{con } c_1 \leq x \leq c_2 \tag{4.1.1-2}$$

teniendose finalmente un modelo de la forma

$$\pi = \beta_0 + \beta_1 x \tag{4.1.1-3}$$

donde

$$\beta_0 = -c_1 \,/\, (c_2-c_1) \text{ y } \beta_1 = 1 \,/\, (c_2-c_1) \tag{4.1.1-4}$$

Este modelo lineal es equivalente a usar la función identidad como función de liga (g) sujeta a las restricciones sobre x, β_0 y β_1 correspondientes a $c_1 \leq x \leq c_2$. Tales condiciones extra implican que las técnicas estandar para estimar β_0 y β_1 para modelos lineales generalizados no puedan ser utilizadas directamente. Este modelo no es muy utilizado en la práctica.

4.1.2 MODELO PROBIT

Uno de los principales modelos empleados en la experimentación biológica es el que utiliza como distribución de tolerancia la distribución Normal y se denomina Probit

$$f(s) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right] \qquad (4.1.2\text{-}1)$$

de donde resulta que

$$\pi = \int_{-\infty}^{x} f(s)\, ds = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad (4.1.2\text{-}2)$$

Φ denota la función acumulativa de probabilidad para una distribución Normal estandar $N(0,1)$ y de aquí que

$$\Phi^{-1}(\pi) = \beta_0 + \beta_1 x \qquad (4.1.2\text{-}3)$$

donde $\beta_0=-\mu/\sigma$ y $\beta_1=1/\sigma$, así entonces la función de liga g es la inversa de la función acumulativa de probabilidad Normal estandar Φ^{-1}.

4.1.3 MODELO LOGIT O LOGISTICO

Un modelo que presenta resultados muy similares al modelo probit pero que computacionalmente resulta ser más fácil es el modelo logit cuya distribución de tolerancia es

$$f(s) = \beta_1 \exp(\beta_0+\beta_1 s) \;/\; [1+\exp(\beta_0+\beta_1 s)]^2 \qquad (4.1.3\text{-}1)$$

de aquí que

$$\pi = \int_{-\infty}^{x} f(s)\ ds = \exp(\beta_0+\beta_1 x) \ / \ [1+\exp(\beta_0+\beta_1 x)] \quad (4.1.3\text{-}2)$$

donde la función de liga queda dada por

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0+\beta_1 x \quad (4.1.3\text{-}3)$$

$\ln(\pi/(1-\pi))$ se denomina función logit y su interpretación es como el logaritmo natural de momios.

4.1.4 MODELO DE VALOR EXTREMO

Si se considera como función de tolerancia

$$f(s) = \beta_1 \exp[(\beta_0+\beta_1 s) - \exp(\beta_0+\beta_1 s)] \quad (4.1.4\text{-}1)$$

resulta que

$$\pi = \int_{-\infty}^{x} f(s)\ ds = 1 - \exp[-\exp(\beta_0+\beta_1 x)] \quad (4.1.4\text{-}2)$$

de aquí que

$$\ln[-\ln(1-\pi)] = \beta_0+\beta_1 x \quad (4.1.4\text{-}3)$$

así entonces la función de liga $\ln[-\ln(1-\pi)]$ se denomina función log log complementaria. Esta función puede resultar útil en el estudio de modelos de sobrevivencia.

4.2 REGRESION LOGISTICA

Los modelos de tipo logit describen los efectos de un conjunto de variables de explicación sobre una variable de respuesta. Al igual que los modelos de regresión para variables de respuesta cuantitativas, los modelos logit no describen la asociación o patrones de interacción entre las variables de explicación, motivo por el cual en situaciones multivariadas resultan más fáciles de formular que los modelos loglineales.

El modelo logístico simple $\ln(\pi/(1-\pi)) = \beta_0+\beta_1 x$ descrito en la sección 4.1.3 es un caso particular del modelo general de regresión logística

$$\mathrm{logit}(\pi_i) = \ln(\pi_i/(1-\pi_i)) = \underline{X}_i'\underline{\beta} \qquad (4.2\text{-}1)$$

donde $\underline{X}_i'$ es el vector de mediciones asociadas a las covariables y variables indicadoras correspondientes a los niveles de factor y $\underline{\beta}$ es el vector de parámetos.

Este modelo es ampliamente usado para análisis multivariados que involucran una respuesta binaria y proporcionan una poderosa técnica análoga al análisis de varianza o regresión múltiple utilizados con respuesta continua.

4.3 CORRESPONDENCIA ENTRE LOS MODELOS LOGIT Y LOGLINEAL

Considere un modelo logit en el caso tridimensional en el cual la tercer variable es dicotómica y se utiliza como variable de respuesta. Sean además, $\{\pi_{ijk}\}$ las probabilidades de las casillas de la tabla y $\{m_{ijk}\}$ los valores esperados para las mismas.

El logit para la tercer variable en cada combinación de las dos primeras es

$$\begin{aligned} \ln[\pi_{ij2}/(1-\pi_{ij2})] &= \ln[\pi_{ij2}/\pi_{ij1}] \\ &= \ln[\pi_{ij2}] - \ln[\pi_{ij1}] \qquad (4.3\text{-}1) \end{aligned}$$

Así entonces el modelo con efectos aditivos resulta

$$\ln[\pi_{ij2}/\pi_{ij1}] = w + w_{1(i)} + w_{2(j)} \qquad (4.3\text{-}2)$$

donde $\sum_i w_{1(i)} = \sum_j w_{2(j)} = 0$; $\{w_{1(i)}\}$ pertenece a la asociación parcial entre la primera y la tercer variable y $\{w_{2(j)}\}$ a la asociación entre la segunda y la tercera, el modelo (4.3-2) supone la ausencia de interacción conjunta de los tres factores.

Para observar la correspondencia entre los modelos logit y loglineal considere el modelo loglineal con la ausencia del efecto conjunto de las tres variables.

$$\ln m_{ijk}=\mu+\mu_{1(i)}+\mu_{2(j)}+ \mu_{3(k)}+\mu_{12(ij)}+\mu_{13(ik)}+\mu_{23(jk)} \qquad (4.3\text{-}3)$$

Considere entonces el logit para la tercer variable

$$
\begin{aligned}
\ln[m_{ij2}/m_{ij1}] &= \ln m_{ij2} - \ln m_{ij1} \\
&= \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(2)} + \mu_{12(ij)} + \mu_{13(i2)} + \mu_{23(j2)} \\
&- \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(1)} + \mu_{12(ij)} + \mu_{13(i1)} + \mu_{23(j1)} \\
&= [\mu_{3(2)}-\mu_{3(1)}] + [\mu_{13(i2)}-\mu_{13(i1)}] + [\mu_{23(j2)}-\mu_{23(j1)}] \\
&= 2[\mu_{3(2)} + \mu_{13(i2)} + \mu_{23(j2)}] \qquad (4.3\text{-}4)
\end{aligned}
$$

Puede notarse que (4.3-4) es de la forma (4.3-2) pudiendose definir

$$w = 2\,\mu_{3(2)},\ w_{1(i)} = 2\,\mu_{13(i2)} \text{ y } w_{2(j)} = 2\,\mu_{23(j2)} \qquad (4.3\text{-}5)$$

para conformar entonces el modelo logit.

INTERPRETACION DE LOS PARAMETROS

A partir de (4.3-4) los parámetros de (4.3-2) satisfacen las restricciones

$$\sum_i w_{1(i)} = \sum_i 2\,\mu_{13(i2)} = 0 \text{ y } \sum_j w_{2(j)} = \sum_j 2\,\mu_{23(j2)} = 0$$

de donde resulta que

$$\sum_i \sum_j \ln[m_{ij2}/m_{ij1}] = 2IJ\,\mu_{3(2)} = IJ\,w$$

Asi entonces,

$$w = \frac{1}{IJ} \sum_i \sum_j \ln[m_{ij2}/m_{ij1}]$$

puede ser interpretado como la media de los IJ logits, también $w_{1(i)}$ puede considerarse como la desviación de la media debido a la localización de la casilla en el renglón i y $w_{2(j)}$ como la desviación de la media debido a la localización de la casilla en la columna j.

GRADOS DE LIBERTAD

Considerando la tabla IxJx2 los grados de libertad para las pruebas de bondad de ajuste pueden calcularse como el número de logits en la tabla menos el número de parámetros linealmente independientes.

$$gl = IJ - [1 + (I-1) + (J-1)] = (I-1)(J-1)$$

4.4 ESTIMACION MAXIMO VEROSIMIL Y DEVIANZA

Para cualquiera de los modelos de dosis-respuesta y el modelo logístico general, los estimadores de máxima verosimilitud de los parámetros $\underline{\beta}$ y en consecuencia de las probabilidades $\pi_i=g(\underline{X}_i'\underline{\beta})$ se obtienen mediante la maximización de la función de log-verosimilitud.

$$\ell(\underline{\pi};\underline{y}) = \sum_i \left[y_i \ln \pi_i + (N_i-y_i) \ln (1-\pi_i) + \ln\binom{N_i}{y_i} \right] \qquad (4.4\text{-}1)$$

utilizando los métodos descritos en el capítulo 2.

La estimación de máxima verosimilitud es posible aun cuando $N_i=1$ y/o $y_i=0$ lo que no necesariamente es cierto para el criterio de mínimos cuadrados.

La medida de bondad de ajuste del modelo utilizando la razón de verosimilitud es el estadístico

$$G^2 = 2[\ell(\underline{\hat{\pi}}_{max};\underline{y}) - \ell(\underline{\hat{\pi}};\underline{y})]$$

donde $\underline{\hat{\pi}}_{max}$ es el vector de estimadores de máxima verosimilitud correspondientes al modelo maximal y $\underline{\hat{\pi}}$ es el vector de estimadores para el modelo de interés.

Sin pérdida de generalidad para el modelo maximal se consideran las π_i's como parámetros a ser estimados y se tiene entonces que

$$\partial\ell/\partial\pi_i = (y_i/\pi_i) - (N_i-y_i)/(1-\pi_i)$$

para el i-ésimo elemento de $\underline{\hat{\pi}}_{max}$ la solución de la ecuación $\partial\ell/\partial\pi_i = 0$ es y_i/N_i, la proporcion de éxitos para el i-ésimo estrato.

De donde resulta que

$$\ell(\underline{\hat{\pi}}_{max};\underline{y}) = \sum_i \left[y_i \, \ln(y_i/N_i) + (N_i-y_i) \, \ln \, (1-y_i/N_i) + \ln\binom{N_i}{y_i} \right]$$

y por lo tanto

$$G^2 = 2\sum_i \left[y_i \ln(y_i/(N_i\hat{\pi}_i)) + (N_i-y_i)\ln((N_i-y_i)/(N_i-N_i\hat{\pi}_i)) \right] \qquad (4.4\text{-}2)$$

puede notarse que G^2 tiene la forma

$$G^2 = 2 \sum o \ln \frac{o}{e}$$

donde o denota las frecuencias observadas de las casillas y_i y (N_i-y_i) mientras que e representa las correspondientes frecuencias esperadas o valores ajustados $N_i\hat{\pi}_i$ y $(N_i-N_i\hat{\pi}_i)$.

Si el modelo ajusta adecuadamente a los datos entonces $G^2 \sim \chi^2_{N-p}$ donde p es el número de parámetros β a ser estimados.

4.5 BONDAD DE AJUSTE

Algunas estrategias utilizadas para los modelos loglineales resultan aplicables con variables de respuesta binarias, tales como pruebas para la inclusión de términos en el modelo y selecciones "stepwise" cuando se tienen suficientes variables de explicación.

El examen gráfico de los residuos es útil para determinar la adecuacía del modelo propuesto y las estrategias planteadas para los modelos loglineales a este respecto resultan también válidas.

Una definición simple de residuo estandarizado es

$$r_i = (p_i-\hat{\pi}_i) \;/\; [\hat{\pi}_i(1-\hat{\pi}_i)/N_i]^{1/2} \tag{4.5-1}$$

donde $p_i=y_i/N_i$ es la proporción observada y $\hat{\pi}_i$ es la proporción estimada bajo el modelo propuesto. Los r_i's tienen aproximadamente media cero y varianza uno y son la raíz cuadrada de los sumandos del estadístico X^2. La graficación contra covariables podría mostrar patrones sistemáticos, pero no obstante su distribución puede no ser muy aproximada a la Normal.

Pierce & Schafer* en 1986 probaron que la raíz cuadrada (con signo) de los sumandos de la Devianza son residuos apropiados para el propósito de diagnóstico.

Así los estadísticos

$$d_i = 2\left[y_i \ln(y_i/(N_i\hat{\pi}_i))+(N_i-y_i)\ \ln(N_i-y_i)/(N_i-N_i\hat{\pi}_i)\right] \tag{4.5-2}$$

se distribuyen al menos asintóticamente Normal estandar si el modelo es apropiado.

---

* Pierce, D.A. & Schafer, D.W. (1986) Residuals in generalised linear models. J. Amer. Statist. Assoc., 81, 977-986.

ANEXO 1: HERRAMIENTAS DE COMPUTO DISPONIBLES

En la actualidad se cuenta con una variedad de paquetes de cómputo que permiten el ajuste de modelos con variables de tipo categórico, donde pueden destacarse los siguientes:

GLIM

Paquete orientado a ajustar modelos lineales generalizados donde en particular los modelos para variables categóricas son un subconjunto de los modelos posibles.

SAS

Paquete estadístico de proposito general que ofrece para el análisis de variables categóricas las rutinas de CATMOD, FREQ y LOGISTIC.

BMDP

Otro paquete estadístico de proposito general que ofrece los programas BMDP-LR y BMDP-4R para regresión logistica y modelos con datos categóricos respectivamente.

SPSS

Al igual que SAS y BMDP es un paquete de proposito general que proporciona fundamentalmente la rutinas LOGLINEAR, HILOGLINEAR, CROSSTABS y PROBIT para el manejo de modelos loglineales, loglineales jerárquicos, Tablas de contingencia y modelo Probit respectivamente.

Conclusiones

Acerca del objetivo:

A lo largo del presente trabajo se han expuesto algunas ideas acerca del planteamiento de modelos que permiten realizar análisis estadístico con varias variables de tipo categórico, donde se hizo especial énfasis en los modelos loglineales jerárquicos para tablas completas, en los modelos de dosis-respuesta, en los modelos de regresión logísticos para una respuesta binaria y en la teoría de los modelos lineales generalizados orientada a modelos con variables categóricas.

De ninguna menera se ha pretendido ser exhaustivo en el planteamiento de los temas, cabe insistir en que el trabajo expuesto en los capítulos anteriores es en buena medida una adaptación de las ideas planteadas por diversos autores que han abundado en la especificación de técnicas para el análisis multivariado, pero donde se trató de realizar una unificación en la notación y en la secuencia de exposición y aglutinamiento de los temas.

Conviene aclarar que el objetivo del presente trabajo, fue el elaborar una fuente bibliográfica en lengua hispánica que reuniera los elementos básicos para introducir al lector en el estudio del análisis multivariado con variables de tipo categórico.

Dado el contexto teórico del presente trabajo, es más apropiado hablar acerca de los alcances y limitaciones en cuanto al material reunido y el material que no fue contemplado en el mismo.

Acerca de los modelos loglineales:

En particular el planteamiento del modelo loglineal bidimensional se realizó en base a los trabajos de Everitt, Reynolds, Bishop, Fienberg y Holland. El tema de los modelos en varias dimensiones se basó fundamentalmente en los textos de Agresti, Goodman, Haberman, Reynolds, Upton, Bishop, Fienberg y Holland. Algunas ideas de Little, Knoke y Burke, Healy ,Cox y Hinkley son también incluidas aunque de mamera más puntual, así como las ideas de los textos en lengua hispánica.

Dentro del contexto de los modelos de tipo jerárquico para tablas completas el primer capítulo del trabajo puede considerarse relativamente completo, dado que contempla la construcción e interpretación del modelo como alternativas de prueba de hipótesis estadísticas. Aunque conviene distinguir las limitaciones del mismo, en cuanto a no abordar el problema de las tablas incompletas y la falta de una mayor profundidad en la prueba de unicidad para la construcción de modelos multivariados.

Acerca de los modelos lineales generalizados

Los temas relacionados con los modelos lineales generalizados se basaron en los textos de Dobson, Searle, Silvey, Nelder y Wedderburn.

Es aquí, donde el alcance del presente trabajo se ve reducido a contemplar los elementos meramente básicos de una teoría por demás compleja y extensa, que soporta un amplio espectro de posibilidades al manejar planteamientos generales para distribuciones pertenecientes a la familia exponencial, y que plantea perspectivas de nuevos desarrollos equiparables a los realizados hoy en día para el análisis de regresión y el análisis de varianza.

Acerca de la estimación de los parámetros

Para los aspectos relativos a los estadísticos suficientes y la estimación de los parámetros del modelo, se utilizaron los textos de Birch, Silvey, Hogg y Craig, Bishop, Fienberg y Holland, Agresti y Dobson. Algunas ideas de Little, Knoke y Burke, Healy ,Cox y Hinkley son también incluidas aunque de mamera más puntual, así como las ideas de los textos en lengua hispánica.

La teoría de los modelos lineales generalizados con su supuesto básico de emplear distribuciones pertenecientes a la familia exponecial, sienta las bases teóricas para soportar la distribución de los estimadores de los parámetros de los modelos contemplados en el presente trabajo; sin embargo, se trató de abundar en las propiedades de los estimadores de los parámetros del modelo loglineal en particular en lo referente a los estimadores de máxima verosimilitud para modelos de tipo jerárquico. Las principales limitaciones giran en torno a la estimación por mínimos cuadrados.

Existen además algunos criterios de estimación tales como el de mínima χ^2 modificada, información mínima discriminante y el bayesiano que quedaron también fuera del alcance del trabajo y que muestran alternativas para futuros desarrollos.

Acerca de la selección del modelo

Para los aspectos relativos a las consideraciones para la prueba de hipótesis, la bondad de ajuste interna y la selección de un modelo, la consulta tuvo base en los textos de Bishop, Fienberg y Holland, Agresti, Dobson, Draper y Smith y Haberman.

En relación a estos tópicos, se trató de realizar una recopilación relativamente exhaustiva, limitada a la bibliografía a mi alcance, que contenía recomendaciones para realizar la selección de un mejor modelo. Tal vez, la mayor limitación a este respecto, sea la falta de una aplicación práctica donde se ejemplificara la utilización de las técnicas referidas.

Acerca de los modelos con respuesta binaria

A éste respecto se abordaron los modelos de dosis-respuesta y regresión logística donde destacan los textos de Dobson, Fienberg y Agresti.

El centro de este tópico son los modelos logit, debido a su relación con la estructura de los modelos loglineales, auque se intenta recopilar algunos modelos de dosis-respuesta por ser los primeros desarrollos importantes equivalentes al análisis de regresión, utilizando una respuesta binaria.

Este tópico se limitó al estudio de modelos con respuesta binaria, lo cual redujo enormemente el alcance del mismo, conviene señalar que Agresti resalta una gama muy amplia de modelos pertenecientes a la clase de los lineales generalizados para variables categóricas, donde se puede aprovechar la información de las variables ordinales y que han quedado fuera del alcance del presente trabajo.

Acerca del alcance de las técnicas estudiadas

Cabe notar que el desarrollo de las técnicas multivariadas en particular las orientadas a variables de tipo discreto han evolucionado enormemente en los años recientes; tal situación hace que el alcance del presente trabajo sea más limitado. Sin embargo, el planteamiento del modelo loglineal jerárquico, y los modelos de regresión logística apoyados con las técnicas de estimación y las estrategias para la selección de un modelo permiten una plataforma básica para el entendimiento del concepto de "modelización" como la forma más actual de análisis estadístico multivariado.

También conviene destacar que el vasto desarrollo de aplicaciones de cómputo que se tiene a disposición hoy en día y que aunado al abaratamiento de los microcomputadores, permite a los investigadores adentrarse más en la aplicación del análisis estadístico con varias variables.

Se puede concluir que la modelización como forma de análisis estadístico multivariado es perfectamente factible con variables de tipo categórico y que dados los desarrollos teóricos hechos hasta la fecha y las facilidades de cómputo existentes, las investigaciones prácticas en materia social basadas en encuestas y las de otra índole como la biológica cuentan ya con una herramienta más apropiada dadas las características de las variables discretas.

BIBLIOGRAFIA

Agresti, A. (1984).
Analysis of Ordinal Categorical Data
New York: Wiley

Agresti, A. (1990).
Categorical Data Analysis
New York: Wiley

Birch , M.W. (1963)
Maximum likelihood in three-way contingency tables
J.Roy. Statist. Soc. Ser. B25 220-233

Bishop Y.V., Fienberg S.E. & Holland P.W. (1975).
Discrete Multivariate Analysis
Cambridge MA: MIT Press.

Brown, D.T. (1959)
A note on aproximations to discrete probability distributions
Information and Control 2, 386-392.

Castillo, A. (1988)
El modelo loglineal para el análisis de tablas de contingencia una herramienta útil y sencilla
Monografías y Manuales en Estadística y Cómputo Vol. 7 Num. 8
México: Colegio de Posgraduados, Chapingo

Cox D.R. & Hinkley D.V, (1974)
Theoretical Statistics
London: Chapman and Hall

Deming W.E. & Stephan F.F (1940)
On a least squares adjustment of a sampled frequency table when the expected marginal totals are Known
Ann. Math. Statist. 11, 427-444

Dobson, A.J. (1990).
An Introduction to Generalized Linear Models
London: Chapman and Hall

Draper, N.R. & Smith H. (1981)
Applied Regression Analysis 2nd Ed.
New York: Wiley

Everitt, B.S. (1977).
The Analysis of Contingency Tables
London: Chapman and Hall

Fienberg, S.E. (1977)
The Analysis of Cross-Classified Categorical Data
Second Edition (1979)
Cambridge Mass.: MITT Press

Godoy , J.E. (1980)
La caracterización de las funciones de probabilidad
México: ENEP Acatlán UNAM

Gokhale D.V. & Kullback S. (1978)
The information in Contingency Tables
New York: Marcel Dekker

Goodman, L.A. (1978)
Analyzing Quantitative Categorical Data: Log-Linear Models and Latent Structure Analysis, Reimpresión de varios artículos)
Cambridge Mass.

Haberman, S.J. (1974).
The Analysis of Frequency Data
Chicago: University of Chicago Press

Haberman, S.J. (1977)
Maximum likelihood estimation in exponential response models
Ann. Statist. 5, 815-841

Haberman, S.J. (1978).
Analysis of Qualitative Data Vol 1 Introductory Topics
New York: Academic

Haberman, S.J. (1979).
Analysis of Qualitative Data Vol 2 New Developments
New York: Academic

Healy,M.R.J.(1988)
GLIM: An Introduction
Oxford: Claredon

Hogg R.V. & Craig A.T. (1968)
Introduction to Matematical Statistics. Second Ed.
New York: Macmillan

Knoke, D & Burke P.J. (1980)
Log-linear Models
Sage University Parer Series on Qualitative Applications in the Social Science. Series No. 07-020
Beverly Hills and London: Sage publications

Little J.A. (1978)
Generalized Linear Models for Cross-Clasified Data from WFS
World Fertility Survey, Tecnical Bulletins No. 5/TECIL834
Netherlands

Márquez J. y Méndez I. (1978)
Modelos logarítmico lineales para análisis de variables categóricas múltiples, Comunicación tec. Vol. 5 Serie Azul No. 35 México:IIMAS-UNAM

Martínez M. y Gutiérrez N.I. (1981)
Análisis de Datos Categóricos
Traducción de Stegman Ch.Z. (1980) Categórical Data Analysis
Trab. inedito presentado a la Asoc. Americana para la Investigación Educativa,en Los Angeles CA. U:S.A.

Nelder, J.A. (1974)
Log linear models for contingency tables: a generalization of classical least squares
Appl. Statist. 23, 323-329

Nelder J.A. & Wedderburn R.W.M. (1972)
Generalizaed Linear Models
J. R. Statist. Soc. A. 135, 370-384.

Pierce, D.A. & Schafer D.W. (1986)
Residuals in generalized linear models
J. Amer. Statist. Assoc. 81, 977-986

Reynolds H.T. (1977)
The analysis of Cross-Clasification
New York: Free Press

Searle, S.R. (1971)
Linear Models
New York: Wiley

Silvey, S.D. (1975)
Statistical Inference
London: Chapman and Hall

Simon, G. (1973)
Additivity of information in exponential family probability laws
J. Amer. Statist. Assoc. 68, 478-482

Takeuchi K., Yanai H. & Murkherjee B.N. (1982)
The foundations of Multivariate Analysis
New Delhi: Wiley Eastern Limited

Upton, G.J.(1978).
The Analysis of Cross-Tabulated Data
New York: Wiley