UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ANALISIS DE TABLAS DE CONTINGENCIA MULTIDIMENSIONALES MEDIANTE EL USO DE LOS MODELOS LOGLINEALES

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

JOSE JAVIER RODRIGUEZ CHAVEZ

México, D. F. 1989

I N D I C E

INTRODUCCION

En este trabajo se presenta un panorama general acerca de los modelos loglineales para el análisis de tablas de contingencia multivariadas y su utilidad para la investigación en Ciencias Sociales, con el propósito de que pueda ser aprovechado por investigadores y alumnos de éstas disciplinas.

Dado que a través de gran parte del trabajo se efectúan pruebas de hipótesis, en el primer capítulo se busca ubicar a las mismas dentro del método científico, además de dar una descripción de cada uno de los pasos para efectuarlas.

En el segundo capítulo, se hace una introducción a lo que es una tabla de contingencia, partiendo de las más simples que son las de 2 X 2 hasta las tablas multidimensionales, asimismo se incluye la prueba ji-cuadrada y la noción de independencia estadística. Una vez que han sido expuestos tales conceptos se procede a explicar las diferentes pruebas de hipótesis que pueden hacerse respecto a las relaciones existentes entre las variables que intervienen en el análisis. En este capítulo las hipótesis se plantean en términos de probabilidades. Para ejemplificar ésto se efectúa un análisis de asociación entre tres variables.

En el tercer capítulo se realiza la introducción a los modelos loglineales a través de una descripción de cómo es construido un modelo a partir de la noción de independencia estadística. En seguida se muestran los diferentes modelos correspondientes a las diferentes hipótesis planteadas en el segundo capítulo, tanto para el modelo bidimensional como para el tridimensional.

Otro punto importante es el que se refiere al ajuste de modelos y estimación de parámetros. En este capítulo se desarrolla un ejemplo del cálculo de los estimadores de los parámetros del modelo de asociación parcial. Por otra parte se describe el proceso iterativo, que se usa cuando los valores esperados no pueden calcularse en forma directa. Para concluir se hacen una serie de consideraciones acerca de la forma de identificar el modelo que describe mejor las distribuciones observadas.

En el capítulo IV se incorpora un análisis basado en información tomada de la encuesta "Migración Interna, Estructura Ocupacional y Movilidad Social en el Area Metropolitana de la Ciudad de México", levantada por El Colegio de México en coolaboración con el Instituto de Investigaciones Sociales de la UNAM en 1970.

En este trabajo el autor se propuso explorar algunas relaciones entre variables que permitan caracterizar a las trabajadoras manuales no calificadas que se seleccionaron de la encuesta citada.

Al tomarlo para aplicar el modelo se pretende destacar que de manera muy directa es posible obtener resultados a los que por otras vías se llega de manera muy laboriosa e incluso en ocasiones no es posible lograr por la imposibilidad de analizar todas las combinaciones que en términos lógicos pueden expresar las relaciones entre las variables que se estudian.

. Por último se presentan las conclusiones y recomendaciones acerca de los conceptos expuestos a loa largo del trabajo.

CAPITULO I

Pruebas de Hipótesis

Uno de los retos de las matemáticas es el desarrollo de modelos que permitan hacer análisis en campos en los que se consideraba que no había forma de aplicación por tratarse de conceptos no-cuantitativos, o de variables no métricas. Al respecto B. Russell[1] escribió: "Quizás en su última perfección toda ciencia será matemática; pero, mientras tanto, existen vastos campos en los que las matemáticas apenas pueden aplicarse, y en ellos han de ser realizadas algunas de las más importantes hazañas de la ciencia moderna". El trabajo aquí presentado pretende mostrar parte de lo que se ha desarrollado con la ayuda de una de las ramas de las matemáticas, (concretamente la estadística) en el campo de las ciencias sociales.

1.1 Hipótesis.

Por considerarlo importante para el área en la que se ha aplicado este estudio, estimo conveniente empezar por definir el término de hipótesis.

Hipótesis, como su nombre lo indica, es un punto de partida, o un supuesto sobre algún hecho observable susceptible de ser verificado[2]; considerado dentro del proceso de

---

[1] Russell, Bertrand. "La perspectiva científica". Ariel. España. 2da. edición. 1949.

[2] Bunge, Mario. "La Investigación Científica". Ed. Ariel. México. 1984.

investigación es un concepto mucho más general que el equivalente usado en estadística, por ello trataremos de hacer la distinción entre las diferentes connotaciones que tiene dicho término.

Comencemos con el significado del término "hipótesis" empleado en el método de la ciencia.

Etimológicamente, la palabra hipótesis proviene del griego, y significa someter o poner abajo. Esta, a su vez, tiene raíces semejantes a las de la palabra latina suposición (de sub-ponere).

"Ambos términos significan la aceptación provisional de una afirmación acerca de algún hecho, o de alguna relación funcional, como cierta, aun cuando no tenga base experimental adecuada y suficiente para ello"[3].

Pero para que esta proposición lógica o hipótesis pueda considerarse como tal, debe ser susceptible de verificación, y para ello se debe averiguar que es lo que si se puede verificar .

Al respecto Bunge dice lo siguiente : "Cuando un enunciado verificable posee un grado de generalidad suficiente habitualmente se lo llama hipótesis científica. O lo que es equivalente, cuando una proposición general (particular o universal) puede verificarse sólo de manera indirecta -esto es, por el examen de algunas de sus consecuencias- es conveniente llamarla "hipótesis científica"; y añade "las

[3]Bunge, Mario. op. cit.

hipótesis científicas son puntos de partida de cadenas deductivas cuyos últimos eslabones -los más próximos a los sentidos, en el caso de la ciencia fáctica- deben pasar la prueba de la experiencia".

Para intentar aclarar un poco más lo que se entiende por hipótesis, Bunge trata de dar una definición al decir: "Un enunciado fáctico general susceptible de ser verificado puede llamarse hipótesis".

Ahora bien ,una vez que hemos aceptado ésta definición de hipótesis, veamos cómo éstas proposiciones son sometidas a prueba . Volviendo a Bunge tenemos que "El método científico es el conjunto de procedimientos por los cuales : a) se plantean los problemas científicos y b) se ponen a prueba las hipótesis científicas".

Según el autor antes mencionado, la forma como deben ser puestas a prueba las hipótesis, ha de responder al tipo de suposición de que se trate, es decir, "si la hipótesis se refiere a objetos ideales como números, funciones, figuras, fórmulas lógicas o suposiciones filosóficas, su verificación consistirá en la prueba de su coherencia -o incoherencia- con enunciados previamente aceptados. En cambio si el enunciado se refiere a la naturaleza o a la sociedad, puede ocurrir, o bien que podamos averiguar su valor de verdad con la ayuda de la razón, o que debamos recurrir, además a la experiencia". Y es aquí donde se halla la conexión entre la estadística y el método científico, pues dice que el analisis lógico y matemático comprobará la validez de los enunciados (hipótesis) que son analíticos en determinado contexto . Y esto se da debido a que cuando la ayuda de la lógica no es suficiente para la comprobación de la hipótesis, es necesario recurrir a otra alternativa, y una de ellas lo constituye el confrontar los supuestos formulados con los datos empíricos, y así llegar a una conclusión[4].

Aquí la estadística juega un papel importante, no sólo como herramienta -en el sentido de que la recolección y el análisis de los datos deben hacerse conforme a sus reglas- sino además como el método que nos ayuda a reafirmar o a reformular nuestra teoría, o a descubrir planteamientos erróneos.

Entonces, cuando hablamos de confrontar los supuestos con los datos empíricos estamos haciendo referencia al método deductivo de la ciencia. Al respecto Stinchcombe[5] dice que el hecho de construir teorías científicas implica que se deben tener presentes los requerimientos lógicos necesarios para verificar dichas teorías con los hechos. La inferencia estadística incorpora estos requerimientos en sus procedimientos.

Stinchcombe escribe : "La inferencia estadística comienza con una proposición teórica que dice que una clase de

---

[4]Bunge, Mario. op. cit.

[5]Stinchcombe, Arthur. "La construcción de teorías sociales". Ed. Nueva Visión. Buenos Aires. 1970. 341 pp.

fenómenos se conectará de cierta manera con otra clase de fenómenos", y continúa, "De una proposición teórica derivamos, por deducción lógica y mediante definiciones operacionales de los conceptos, una proposición empírica". Una proposición empírica es la que formula lo siguiente : "Si realizamos tales y tales observaciones, se obtendrán tales y tales resultados ".

Esto significa que los conceptos deberán ser expresados en términos observacionales.

En estadística una hipótesis es un enunciado respecto a uno o más parámetros de una población[6]. Por ejemplo una hipótesis estadística sería la siguiente: el salario promedio de los trabajadores de la industria automotriz es de $480,000 mensuales.

1.2 Pruebas de hipótesis.

En esta sección trataremos de ver qué es lo que se entiende por hipótesis en estadística, -de hecho más concretamente por pruebas de hipótesis[7]-, así como algunos de los casos particulares en los que se emplean estas pruebas. Tal vez el término "prueba de hipótesis" sea resultado de la traducción de <u>test of hypothesis</u> y no tiene el mismo sentido que se le da en inglés; en español la palabra "prueba"

---

[6]Ostle, Bernard. Estadística aplicada. Limusa. The Iowa state University Press. 1965.

[7]De entre las diversas traducciones se encuentran las siguientes: "pruebas de significación" (véase "La controversia sobre las pruebas de significación", FLACSO. México. 1983.) o "docimasia de hipótesis" (en Mood y Graybill. "Introducción a la teoría de la estadística". Ed. Aguilar. Madrid. 1972).

corresponde a dos términos diferentes: test y proof (someter a prueba y comprobar). Por consiguiente a continuación describiremos el sentido que tiene en estadística este término y los pasos que se realizan para su falsación[8]. El establecimiento de una hipótesis estadística surge:

a) como una necesidad ante la imposibilidad de estudiar íntegramente a todos los miembros de una población, y por consiguiente conocer los parámetros de las distribuciones poblacionales. Es por ello que la estadística debe hacer uso de la inferencia como una vía para poder conocer algunos aspectos de dicha población.

Por ejemplo podría interesarnos conocer la distribución de una variable y una forma de hacerlo es suponer conocida la forma de la función de dicha distribución y entonces plantear hipótesis acerca de sus parámetros.

b) como una forma de realizar pruebas de independencia o asociación estadística entre variables. Tales pruebas son las que se utilizan en el presente trabajo. Tambien son conocidas como pruebas de no interacción entre las variables. Este último término esta tomado de una técnica de la estadística llamada análisis de varianza, debido a que existe mucha semejanza entre el tipo de análisis ahí realizado y el análisis de asociación practicado por los modelos

[8]La noción de "falsación" la define Karl Popper en su libro "La lógica de la investigación científica", como "únicamente decimos que una teoría esta falsada si hemos aceptado enunciados básicos que la contradigan". Ed. Tecnos, Madrid. 1962. p. 83.

logilneales, que se presentarán más adelante[9].

Aquí es pertinente hacer la aclaración de que en estadística lo que frecuentemente se hace es someter a prueba la hipótesis de que las variables son independientes[10] y la lejanía de la independencia se considera como evidencia de relación entre variables[11]. Pero la discusión lógica de este problema queda fuera de los propósitos del presente trabajo. Por lo anterior consideraremos que el hecho de concluir que no hay indicios de independencia estadística se entenderá como : "hay relación entre las variables en cuestión ". Ahora veamos cómo se plantea y se somete a prueba en estadística una hipótesis de esta naturaleza.

1. Formulación de las hipótesis estadísticas, esto es de la hipótesis nula Ho y de la hipótesis alternativa (Ha). La hipótesis nula es una aseveración acerca del valor de un parámetro poblacional, y la hipótesis alternativa es el complemento lógico de la hipótesis nula. Aquí es importante señalar que a) éstas hipótesis no surgen de la estadística sino de la teoría manejada por el investigador, la cuestión es plantear la hipótesis de investigación en términos estadísticos

---

[9] Everitt, B.S. "The Analysis of Contingency Tables". London Chapman and Hall LTD; New York, A Halsted Press Book, John Wiley & Sons Inc. 1977. 128 págs.

[10] En el capítulo 2 sección 2.3 se presenta la definición de independencia estadística entre dos variables.

[11] Sin embargo esta definición de relación no especifica de qué tipo de asociación se habla sino que la postula como negación de la independencia.

y esto no se realiza en forma directa, sino que pasa por un proceso de objetivación y operacionalización de los conceptos, y de traducción del lenguaje natural en que se expresan las proposiciones a uno formal y b) algunos autores como Hoel[12] sugieren que Ho sea planteada en términos de que sea rechazada y optar por Ha, lo cual esta de acuerdo con la corriente del falsacionismo descrita por Popper (op. cit.).

2. Elección de un estadístico de prueba (con su distribución estadística asociada) para contrastar Ho. El investigador debe plantear su hipótesis en términos de una variable aleatoria en que intervenga el parámetro sobre el que se hace la hipótesis y cuya distribución sea conocida. En el caso de las tablas de contingencia las distribuciones mas usuales son la distribución normal y la distribución ji-cuadrada.

3. Especificación de a) un nivel de significación (probabilidad de que se tome una decisión equivocada al rechazar la Ho cuando es cierta), también llamado alfa: que es la probabilidad de cometer el error tipo I y b) del tamaño de la muestra (n). El procedimiento para determinar cuántos elementos deben componer la muestra depende del tipo de análisis a realizar. En el caso de las tablas de contingencia debe de ser lo suficientemente grande como para no tener muchas casillas

---

[12]En "Estadística elemental" sugiere que la hipótesis sea planteada de tal forma que se busque rechazar Ho. Véase Hoel, Paul G. "Elementary Statistics". John Wiley & Sons, Inc. New York. 1960.

vacías[13]. El otro error que se puede cometer es el de no rechazar una hipótesis falsa (error tipo II). En la práctica no es posible determinar si una muestra[14] es representativa o no por lo que se requiere un procedimiento de prueba que minimice la probabilidad de cometer uno de los dos tipos de error. Una forma de reducir alfa es aumentando la probabilidad de cometer el error tipo II (beta), y así equilibrar ambas probabilidades.

4. Con base en los puntos anteriores, definir la región de rechazo de la hipótesis nula. Es decir, cuál es el rango de valores del estadístico que llevará a rechazar Ho.

5. Cálculo del valor del estadístico para la prueba estadística, utilizando los datos obtenidos de la muestra. Si dicho valor se encuentra dentro de la región de rechazo, la decisión que se toma es la de rechazar la hipótesis nula Ho; si, por el contrario, dicho valor se encuentra fuera de la región de rechazo, la decisión que se toma es que no hay elementos suficientes para rechazar Ho al nivel de significación elegido.

En las pruebas de hipótesis la forma como se llega a una conclusión es realizando la evaluación de la probabilidad de

---

[13]Para mayores detalles véase: Cortés, Fernando. "Tamaño de la muestra y análisis de asociación". Revista Mexicana de Sociología no. 4. México. 1982.

[14]Cuando hablemos de muestra nos referiremos al conjunto de elementos de una población seleccionados con la intención de estimar los valores verdaderos de los parámetros en la población. Dicha selección deberá hacerse mediante alguno de los procedimientos aleatorios del muestreo.

observar cierto hecho asumiendo como verdadera una hipótesis (en este caso la hipótesis nula Ho). Si esa probabilidad es baja (con respecto a un cierto valor que puede atribuírsele en atención a la aleatoriedad del muestreo), se considera que la hipótesis debe rechazarse y se dice que el hecho observado es estadísticamente significativo. En el análisis de contingencia, la hipótesis nula plantea que hay independencia entre las variables que intervienen en la tabla. El estadístico de prueba es una medida de discrepancia entre los valores observados y los esperados supuesta la hipótesis Ho. Bajo la suposición de Ho, esta medida tiene propiedades distribucionales conocidas, por lo que se puede calcular la probabilidad de que se obtengan determinados valores.

En la estadística una hipótesis, -en este caso sobre un modelo para estimar los valores esperados en una tabla de contingencia- siempre debe ser susceptible de someterse a prueba, y para ello es necesario que ésta sea planteada en términos de hipótesis sobre funciones estimables; en el desarrollo de este trabajo trataremos con hipótesis tales como: la variable 1 es independiente de la variable 2. Esto debe plantearse como una hipótesis estadística y someterse a prueba mediante una función construida a partir de las observaciones. En el capítulo III veremos cómo se construyen éstas funciones.

CAPITULO II

Tablas de contingencia

2.1 Cuadros y tablas.

Para el análisis estadístico de una muestra se cuenta con la ayuda de diferentes tipos de tablas. Existen tablas cuya finalidad es mostrar solamente el desglose de los atributos de una variable; otras que muestran dicha clasificación con respecto a otra variable sin más intención que la de la comparación. Inclusive algunas de ellas podrían simplemente verse como cuadros de presentación de la información, y que no se construyen con la intención de analizar la relación entre variables; son un simple reporte, tal es el caso del cuadro 2.1. Por lo tanto es importante hacer la diferencia entre un cuadro descriptivo o comparativo, y una tabla para determinar si hay o no asociación.

Cuadro 2.1
Centroamérica: Población rural y urbana (miles), 1980.

	rural	urbana
Costa Rica	1198	1015
El Salvador	2667	2130
Guatemala	4471	2791
Honduras	2207	1484
Nicaragua	1249	1484

Fuente: CELADE (En: El crecimiento desigual en Centroamérica 1950-2000, Mayorga Quiroz, Román, El Colegio de México, 1983).

Si el interés es el de medir asociación entre variables no métricas entonces el investigador construye la tabla para confrontar la distribución con su proposición empírica. A esas tablas se les conoce como tablas de contingencia, y en la estadística se encuentran diversos métodos para su estudio.

2.2 Tablas de contingencia.

Dentro de la estadística encontramos un método que es útil para el análisis de variables cualitativas[1]. Este método consiste en analizar las frecuencias que resultan de clasificar a los miembros de una población (sean cosas o personas), según las características que nos sean de interés, por ejemplo la ocupación, el estado civil, el nivel de ingresos, etc. si son personas; no. de ocupados y valor de la producción en establecimientos industriales; superficie sembrada y tipo de cultivo, si se trata de predios, etc.

Cuando la clasificación se construye con dos o más variables consideradas simultáneamente, se conoce como cruce de variables. En el cuerpo de la tabla encontraremos el número de observaciones, de acuerdo con los criterios de clasificación cuyo cruce define las casillas de la tabla. El número que resulta en cada casilla es una frecuencia absoluta y se le conoce como

[1]Entenderemos por variables cualitativas tanto a variables que originalmente se hayan medido en escala nominal u ordinal como aquéllas que a través de transformaciones se hayan llevado a esas escalas.

frecuencia observada. Ahora bien las categorías[2] de las variables en las que estaremos interesados deberán ser exhaustivas y mutuamente excluyentes. Exhaustivas en el sentido de que no quedará sin clasificarse ningún elemento de la muestra, y mutuamente excluyentes en cuanto a que un miembro de la población no puede caer simultáneamente en 2 categorías de una misma variable.

Las tablas resultantes de ésta operación son conocidas como tablas de contingencia. El nombre proviene del interés en determinar si la forma en que es clasificada una variable es contingente del modo de clasificar a la otra[3]. Usualmente se construyen con el propósito de estudiar la relación entre las variables de clasificación.

2.2.1 Tablas de 2 X 2.

Cuando se construye una tabla de dos variables, el propósito es determinar si existe independencia entre los dos criterios de clasificación, (al hablar de independencia estamos haciendo referencia a independencia estadística[4]). Por ejemplo podemos estar interesados en saber si la forma cómo se clasifica a un grupo de trabajadoras, de acuerdo a su ocupación, es independiente a las formas de clasificarlas según su edad o su

---

[2]Se usa la palabra "categorías" en general porque no siempre son valores.

[3]Beaver R. y Mendenhall W. Introducción a la probabilidad y la estadística. Duxbury Press, Belmont California. 1971. 408 págs.

[4]Véase la sección 2.3 en la pág. 18.

estado civil. Dicho de otra manera, ver si las mujeres jóvenes se distribuyen de la misma manera que las no jóvenes en las diferentes ocupaciones, o en caso contrario detectar si hay un mayor número de mujeres jóvenes en alguna ocupación en particular. En el punto 2.4 se tocará este tema con detalle.

Cuando se realiza una clasificación a partir de dos variables, se dice que la tabla resultante es de dos entradas o de doble clasificación o de dos dimensiones, y si las variables son dicotómicas diremos que se trata de una tabla de 2 X 2, (esta notación hace referencia al número de categorías de las variables y a las casillas que contendrá la tabla). En el cuadro 2.2 se muestra una tabla con tales características.

CUADRO 2.2

Mujeres que trabajaban al momento de la encuesta de la fase A, según estrato y situación migratoria, área metropolitana, 1970.

	SITUACION MIGRATORIA		
ESTRATO	nativas	no nativas	total
alto	395	248	643
bajo	259	593	852
total	654	841	1495

Fuente: Muñoz, Oliveira y Stern. "Migración y desigualdad en la ciudad de México". El Colegio de México, Instituto de Investigaciones Sociales, UNAM. México. 1977.

2.2.2 Tablas de contingencia de r x c.

En el cuadro 2.3 se muestra la forma general de una tabla de contingencia de 2 dimensiones. Al mismo tiempo se introduce la notación que se usará a lo largo del presente trabajo[5].

Para generalizar se puede suponer que la variable que forma los renglones tiene r categorías, y la que determina las columnas tiene c categorías. En este caso se habla de una tabla de contingencia de r x c.

Cuadro 2.3

variable 2 (c columnas)

		1	2	...	j	...	c	total
	1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2.}$
	⋮	⋮	⋮		⋮		⋮	⋮
variable 1 (r renglones)	i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
	⋮	⋮	⋮		⋮		⋮	⋮
	r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r.}$
total		$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	$n_{..} = N$

La tabla representa una muestra de N observaciones clasificadas con respecto a dos variables. Por convención la variable dependiente se coloca en los renglones (r), y la independiente en las columnas (c).

En cada casilla se encuentra el número de elementos de la

[5]De aquí en adelante se usará la notacion del libro de Everitt, op. cit.

muestra que poseen simultáneamente el atributo de la variable renglón y el atributo de la variable columna, en cuya intersección está la casilla. En la intersección de la categoría i de la variable 1 y de la j de la variable 2, está la frecuencia observada que representaremos por n_{ij}.

A la frecuencia total de la i-ésima categoría de la variable renglón se le denota por $n_{i.}$ y al total de la j-ésima categoría de la variable columna por $n_{.j}$. Con $n_{..}$ se indica el total de casos que se presentan en la tabla.

Dado que éstos totales se colocan al margen de la tabla, se hace referencia a ellos como totales marginales.

El total marginal de la categoría i-ésima es:

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{ic} = \sum_{j=1}^{c} n_{ij} \qquad (2.1)$$

El total marginal de la categoría j-ésima es:

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{rj} = \sum_{i=1}^{r} n_{ij} \qquad (2.2)$$

La suma de las frecuencias de todas las casillas es:

$$\begin{aligned} n_{..} = {} & n_{11} + n_{12} + \dots + n_{1c} + \\ & n_{21} + n_{22} + \dots + n_{2c} + \dots \\ & n_{r1} + n_{r2} + \dots + n_{rc} = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} = N \end{aligned}$$

2.3 Independencia estadística.

En el punto 2.2 mencionamos que al someter a prueba una hipótesis lo que estamos haciendo es comparar los resultados obtenidos de la muestra con los resulatdos teóricos que se obtendrían en caso de haber independencia en la forma de clasificar a las variables. Ahora veremos cómo se obtienen tales frecuencias teóricas. Si P_{ij} representa la probabilidad de que un elemento de la población posea simultáneamente el atributo i-ésimo de la variable renglón y el atributo j-ésimo de la variable columna, la frecuencia esperada F_{ij} estaría dada por la relación:

$$F_{ij} = N\, P_{ij}$$

Sea $P_{i.}$ la probabilidad de que un elemento de la <u>población</u> pertenezca a la i-ésima categoría de la variable renglón, y sea $P_{.j}$ la probabilidad de que pertenezca a la j-ésima categoría de la variable columna. Entonces si dos variables son independientes estadísticamente se debe de cumplir que[6]:

$$P_{ij} = (P_{i.})(P_{.j}) \quad \text{para todos los valores de } i, j. \qquad (2.4)$$

Si el supuesto de independencia es cierto entonces:

$$F_{ij} = N(P_{i.})(P_{.j}) \qquad (2.5)$$

Como esto está basado en probabilidades sobre la población, es necesario que estimemos tales valores. Dichas probabilidades se estiman por su estimador de máxima verosimilitud[7] a partir de los totales marginales de las frecuencias observadas:

---

[6]Mood y Graybill. op. cit. pág. 51.

[7]Mood y Graybill, op. cit. pág. 206.

$$\hat{p}_{i.} = \frac{n_{i.}}{N} \quad \text{y} \quad \hat{p}_{.j} = \frac{n_{.j}}{N} \qquad (2.6)$$

Donde $\hat{p}_{i.}$ es el estimador de $P_{i.}$ y $\hat{p}_{.j}$ es el estimador de $P_{.j}$.

Una vez obtenidos estos valores estimados podemos estimar la frecuencia esperada en la ij-ésima casilla E_{ij}, esto es, el número de casos que habría en esa casilla si hubiera independencia entre las dos clasificaciones. Sustituyendo en 2.5 encontramos que:

$$\begin{aligned} E_{ij} &= N(\hat{p}_{i.})(\hat{p}_{.j}) \\ &= N \frac{n_{i.}}{N} \frac{n_{.j}}{N} = \frac{(n_{i.})(n_{.j})}{N} \end{aligned} \qquad (2.7)$$

Por lo anterior se debe de cumplir que:

$n_{i.} = E_{i.}$ y $n_{.j} = E_{.j}$ [8]

Esto significa que los totales marginales esperados deben de coincidir con los totales marginales observados. Como ejemplo veamos el siguiente cuadro[9]:

---

[8] $$\sum_{j=1}^{c} E_{ij} = E_{i.} = (n_{i.})(n_{..})/n_{..} = n_{i.}$$

[9] La información que se presenta a lo largo de este trabajo se tomó de la tesis de Rosa María Igartúa: "Algunas características de las trabajadoras manuales no calificadas de la Ciudad de México". Tesis de maestría. FLACSO. México. 1983. Las variables que usaremos son: ocupación, edad, número de hijos nacidos vivos, estado civil, nivel de instrucción, y situación migratoria.

Cuadro 2.4
Frecuencias observadas y frecuencias esperadas bajo la hipótesis de independencia de la muestra clasificada por ocupación y estado civil.

	ESTADO CIVIL				
OCUPACION	casada	divorciada	viuda	soltera	total
vendedoras ambulantes	14(7.2)	7(5.6)	11(4.2)	2(17.0)	34
servicio no doméstico	20(10.9)	13(8.7)	11(6.4)	8(26.0)	52
servicio doméstico	65(88.6)	70(70.0)	53(51.9)	233(210.5)	421
producción	48(40.2)	26(31.7)	11(23.5)	106(95.5)	191
total	147	116	86	349	698

Nota: las frecuencias esperadas se presentan entre paréntesis.

En el cuadro 2.4 se encuentra la información referente a las trabajadoras manuales no calificadas de la encuesta de migración fase A, clasificadas por ocupación y estado civil[10].

Si la ocupación fuera independiente del estado civil se esperaría que el número de vendedoras ambulantes casadas fuera:

$$E_{11} = \frac{(n_{1.})(n_{.1})}{N} = \frac{(34)(147)}{698} = 7.2$$

De igual manera se obtienen los valores esperados del resto de las casillas.

La noción de independencia puede verse también como una relación de momios[11]. Si este cociente se aleja de la unidad

---

[10] Humberto Muñoz, Orlandina de Oliveira y Claudio Stern, "Migración Interna, Estructura Ocupacional y Movilidad Social en el Area Metropolitana de la Cd. de México". El Colegio de México, Instituto de Investigaciones Sociales, UNAM. México. 1977.

[11]Knoke y Burke en su libro "Loglinear Models" definen a los momios (en inglés odds) como la razón entre la frecuencia que resulta de pertenecer a una categoría y la frecuencia que resulta de no pertenecer a ésa categoría. Su interpretación es la

significa que los componentes de la razón de momios (numerador y denominador) se alejan entre sí. Este hecho indica un alejamiento de la independencia. Veamos el caso de una tabla de 2 X 2.

Para el cuadro 2.5 los momios de la variable situación migratoria son:

$$\frac{n_{11}}{n_{21}} = \frac{130}{112} = 1.16 \qquad \frac{n_{12}}{n_{22}} = \frac{226}{260} = 0.86$$

. En las jóvenes respecto a las no jóvenes, los momios de situación migratoria son 1.16 y 0.86 respectivamente.

$$\frac{n11}{n12} = \frac{130}{226} = 0.57 \qquad \frac{n21}{n22} = \frac{112}{260} = 0.43$$

Y en las nativas respecto a las no nativas los momios de edades son 0.57 y 0.43.

Cuadro 2.5
Frecuencias observadas de la muestra clasificada por edad y situación migratoria.

SITUACION MIGRATORIA	EDAD 15-24	25-64	total
nativas	130	226	356
no nativas	112	260	372
total	242	486	728

Para comparar estos momios podemos hacerlo mediante un

---

probabilidad de que un individuo seleccionado aleatoriamente caiga en la categoría de interés en lugar de hacerlo en otra categoría.

cociente. Lo que deberíamos esperar en caso de que las variables fueran independientes es que los momios fueran parecidos. Es decir si la edad no influye en la situacion migratoria, la relación entre las nativas y las no nativas debe ser la misma entre las jóvenes y entre las no jóvenes. Si esto sucede el resultado de la división debe estar cercano a la unidad. Al cociente se le llama razón de momios y queda de la siguiente manera:

$$\frac{\frac{n_{11}}{n_{21}}}{\frac{n_{12}}{n_{22}}} = \frac{(n_{11})(n_{22})}{(n_{21})(n_{12})} = \frac{1.16}{0.86} = 1.35$$

El resultado indica una relación débil: el valor de la razón de momios es cercano a la unidad. En este caso se podría concluir que la edad no influye de una manera determinante en la situación migratoria.

2.4 Prueba ji-cuadrada.

Si nuestro propósito es determinar si las variables bajo estudio son estadísticamente independientes o no, podemos recurrir a una prueba que nos ayuda para tal efecto: la prueba ji-cuadrada. Este estadístico de prueba, sugerido por Pearson en 1904, confronta la hipótesis de independencia estadística con las observaciones. Es decir, la idea central es comparar las frecuencias observadas con las frecuencias que se esperarían si las variables fueran independientes. Si dichas frecuencias son muy parecidas o iguales, concluiremos que las variables son independientes[12]. En caso contrario, podemos decir que se rechaza la hipótesis de independencia estadística entre las mismas. Pero aquí surge una pregunta: ¿Hasta que punto se considerará que las frecuencias son muy parecidas o iguales?. La respuesta se encuentra en el hecho de que las diferencias encontradas entre ambas cantidades pueden deberse a factores aleatorios al momento de obtener la muestra. El límite para establecer si se toma una u otra decisión, está dado por el valor que proporciona la distribución teórica ji-cuadrada (en tablas).

Para realizar la prueba debemos usar una medida construida a partir de las observaciones que nos permita efectuar una confrontación entre las frecuencias esperadas y observadas, (esta medida se usará como estadístico de prueba). Tal indicador se denota por una X^2 [13] para diferenciarlo de la distribución ji-cuadrada teórica (nos referiremos a ésta como "X^2 de tablas"). A continuación se describe su construcción.

$$X^2 = \sum_i \sum_j \frac{(n_{ij}-E_{ij})^2}{E_{ij}} \qquad (2.8)$$

El hecho de elevar al cuadrado las diferencias obedece a que se pueden encontrar diferencias negativas, que al momento de sumarse con las positivas se contrarresten e impidan lograr lo que se está buscando: medir la discrepancia entre n_{ij} y E_{ij};

[12]En realidad la conclusión de la prueba es "no hay evidencia para rechazar la hipótesis de independencia".

[13]Usaremos X^2 en lugar de la letra griega ji para evitar dificultades tipográficas.

además al elevar al cuadrado se otorga mayor impacto a las cantidades más grandes. Pero como no puede considerarse de igual importancia una diferencia de tamaño 2 respecto a 100 que respecto a 10, es necesario relativizar éstas diferencias. Para tal efecto se divide entre el valor esperado de la casilla correspondiente. Se toma E_{ij} porque este valor es siempre diferente de cero. Si alguna E_{ij} fuera igual a cero, esto significaría que al menos uno de los totales marginales es cero, y por lo tanto esa categoría debe ser eliminada de la tabla.

El procedimiento de la prueba consiste en suponer que si las variables son independientes entonces los valores observado y esperado en cada casilla de la tabla serán muy parecidos. Esto provocaría que los valores de las diferencias fueran pequeños, o mejor dicho, menores a los encontrados si las variables estuvieran relacionadas.

Esto significa que si Ho es cierta (Ho:las variables son independientes), X^2 proporcionará un valor menor al que le correspondería si Ho fuera falsa.

Ahora bien la decisión de rechazar o no la hipótesis de independencia depende de la probabilidad de obtener por azar un valor como el dado por X^2. Mientras más estrictos más probable será no rechazar Ho, (puesto que estaríamos pidiendo que para considerar independencia, los valores observados y los esperados deberían ser muy próximos).

Esta probabilidad -también conocida como nivel de significancia o de significación- se considera usualmente de 0.05

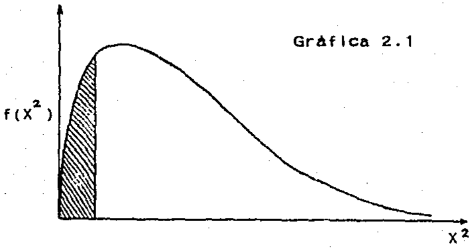
o 0.01 y nos permite fijar la región crítica para la prueba[14]. En este punto cabe mencionar que el estadístico de prueba que resulta de comparar los valores esperados contra los observados

$$\frac{n_{ij} - E_{ij}}{E_{ij}}$$ se distribuye aproximadamente como una normal

estandarizada (con media 0 y varianza 1).

Por lo tanto la relación 2.8 tiene una distribución aproximada ji-cuadrada con (r-1)(c-1) grados de libertad.

Visto gráficamente:



la zona sombreada corresponde a los valores de X^2 que nos llevarían a no rechazar Ho.

2.4.1 Grados de libertad

Ahora bien, como el valor de X^2 crece conforme aumenta el número de categorías, (es decir, mientras más grande es el número de casillas en la tabla, mayor es el número de sumandos de X^2 y por consiguiente también su valor) el valor teórico de X^2 cambia

---

[14]De aquí en adelante se considerará un nivel de significancia de 0.05 para todas las pruebas.

su zona de rechazo conforme aumenta el número de casillas. Este hecho hace que existan diferentes valores de X^2 para un mismo nivel de significancia.

Al parámetro que incorpora el número de casillas en la tabla se le conoce como "grados de libertad". Este nombre se deriva del número de casillas que se pueden modificar en el valor de su frecuencia sin alterar el valor de los totales marginales. El concepto de grados de libertad es un concepto matemático[15]: Es el nombre dado al número de observaciones linealmente independientes que ocurren en una suma de cuadrados. Por ejemplo, en una tabla de 2 X 2 sólo se tiene la libertad de fijar la frecuencia de 1 casilla. Las casillas restantes quedan automáticamente determinadas por la diferencia entre este valor y los totales marginales. Veamos el caso del cuadro 2.5, si la frecuencia n_{11} fuera 150 en lugar de 130, n_{12} tomaría el valor 356-150=206. De igual forma las otras 2 frecuencias quedarían automáticamente determinadas. En una tabla de 3 X 3 podemos modificar (3-1)(3-1)=4 casillas sin alterar los totales marginales. Entonces en el caso de 2 variables la forma general de calcular los grados de libertad es (r-1)(c-1).

Cabe mencionar que la distribución de X^2 es solamente aproximada a la distribución teórica, que descansa en los supuestos de que las frecuencias observadas se distribuyen como una distribución multinomial y de que las frecuencias esperadas no son demasiado pequeñas (digamos no menores de 5). Si esto se

---

[15]Chou, Ya-Lun. Análisis Estadístico. Ed. Interamericana. 1969.

cumple se puede demostrar que X^2 se aproxima a una distribución ji-cuadrada teórica.

Es por ello que X^2 puede compararse con el valor de la distribución teórica dado en tablas.

Si calculamos este estadístico de prueba con los datos del cuadro 2.4: $X^2 = 21.64$,y lo comparamos con su correspondiente valor en tablas con (4-1)(4-1)=9 grados de libertad y un nivel de significancia de 5% : 16.91, vemos que es mayor, lo que nos conduce a rechazar la hipótesis de independencia, entre la ocupación y el estado civil.

2.5 Tablas multidimensionales.

Hasta el momento sólo hemos visto el caso en el que intervienen en el análisis dos variables, pero en la investigación social es más usual la situación en la que se introduce una tercera, o inclusive más variables. Lazarsfeld[16] estudia las posibles relaciones que resultan de incluir en el análisis una variable control y plantea la necesidad de analizar todas las que se pueden dar en conjunto. A este tipo de relaciones las llama relaciones marginales y condicionales.

Este hecho acarrea un grado de complejidad mayor para el análisis de las relaciones que privan entre las variables consideradas. Para ello se deben emplear otros métodos diferentes a los mencionados hasta ahora.

---

[16] Boudon R., Lazarsfeld P. "Metodología de las Ciencias Sociales", en La interpretación de las relaciones estadísticas como propiedad de investigación. LAIA. España. 1974.

Si reconocemos que no es fácil determinar la naturaleza de la relación entre dos variables con sólo ver la tabla, podemos imaginar en qué grado se complica la situación cuando intervienen un número mayor de ellas en forma simultánea.

Para ejemplificar lo anterior podemos analizar las variables ocupación, estado civil y edad. En primera instancia podemos hacerlo en forma bidimensional, es decir, revisar lo que ocurre con la ocupación y el estado civil, con la ocupación y la edad, y con el estado civil y la edad; dado que ya en el punto 2.4 se encontró que había asociación entre la ocupación y el estado civil, sólo falta hacer las pruebas de independencia entre la ocupación y la edad, y entre el estado civil y la edad. Si analizamos la primera relación encontramos que el valor de X^2 no es significativo (estadísticamente la ocupación y la edad son independientes), en cambio en la segunda relación si encontramos asociación.

Veamos qué sucede si las tres variables se combinan en un solo cuadro tridimensional (ver cuadro 2.6). Aquí se presenta una nueva situación. Cuando se trabajaba con tablas de dos dimensiones, el interés radicaba en probar únicamente la hipótesis de independencia entre las dos variables que aparecen en cada tabla. Sin embargo en una tabla tridimensional surgen nuevas combinaciones de hipótesis, por ejemplo: a) puede interesarnos probar si la edad y el estado civil son independientes del estrato ocupacional, o bien b) que haya interacción entre la ocupación y una de las variables restantes, corresponde al renglón i, columna j y tabla k.

$$Ho: P_{ijk} = P_{i..}\ P_{.j.}\ P_{..k} \qquad (2.9)$$

P_{ijk} es la probabilidad de que una observación de la población caiga en la ijk-ésima casilla, y $P_{i..}$, $P_{.j.}$ y $P_{..k}$ son las probabilidades marginales de las variable renglón, columna y control respectivamente. Esta prueba es la equivalente a la prueba de independencia en una tabla de dos dimensiones, y como el procedimiento de prueba es similar en todas las tablas, necesitamos calcular los valores estimados de las frecuencias esperadas en el caso de que Ho sea cierta, después comparar estos valores con las frecuencias observadas usando el estadístico X^2 y por último comparar con su correspondiente valor en tablas.

De igual manera, la forma de encontrar los estimadores de los valores esperados, es similar a la de las tablas bidimensionales. Tales valores están dados por la relación:

$$E_{ijk} = N\ (p_{i..})(p_{.j.})(p_{..k}) \qquad (2.10)$$

donde $\hat{p}_{i..}$, $\hat{p}_{.j.}$ y $\hat{p}_{..k}$ son los estimadores de las probabilidades $P_{i..}$, $P_{.j.}$ y $P_{..k}$.

Se ha demostrado que los mejores estimadores son los de máxima verosimilitud[17], los cuales están calculados a partir de los totales marginales de cada variable, como se indica a continuación:

$$\hat{p}_{i..} = \frac{n_{i..}}{N} \qquad \hat{p}_{.j.} = \frac{n_{.j.}}{N} \qquad \hat{p}_{..k} = \frac{n_{..k}}{N} \qquad (2.11)$$

---

[17] Insesgados, eficientes, suficientes y de varianza mínima. Para mayores detalles véase Mood y Graybill, op.cit. pág. 213.

o c) que la posible relación entre la ocupación y el estado civil sea una asociación espúria, es decir, que ésta se deba a una tercera variable, como puede ser la edad o el nivel de instrucción. O aún más simple: d) que las 3 variables sean mutuamente independientes.

La hipótesis más simple para una tabla multidimensional es aquella que plantea la independencia mutua de las variables. En la siguiente sección se describen las cuatro hipótesis posibles para una tabla tridimensional.

Cuadro 2.6
Frecuencias observadas de la muestra de mujeres clasificadas por ocupación, estado civil y edad.

EDAD											
jóvenes						no jóvenes					
	EDO. CIVIL						EDO. CIVIL				
OCUPACION	cas	div	vda	solt		OCUPACION	cas	div	vda	solt	
vend ambul	1	1	0	2	4	vend ambul	13	6	11	0	30
serv no dom	2	3	0	6	11	serv no dom	18	10	11	2	41
serv dom	23	27	2	200	252	serv dom	42	43	51	33	169
producción	20	11	1	92	124	producción	28	15	10	14	67
total	46	42	3	300	391	total	101	74	83	49	307

2.5.1 Hipótesis de independencia mutua.

Veamos cómo se plantea la hipótesis de independencia mutua en una tabla tridimensional, en términos de probabilidades: Denotaremos con i,j,k los subíndices de las categorías correspondientes a las variables renglón, columna y tabla (control) respectivamente. Así p_{ijk} es la probabilidad estimada de que una observación tenga la categoría i de la ocupación, j del estado civil y k de la edad, i.e. esté en la casilla que

Si sustituimos éstos estimadores en la relación (2.10) encontramos que:

$$E_{ijk} = N \frac{n_{i..}}{N} \frac{n_{.j.}}{N} \frac{n_{..k}}{N}$$

$$E_{ijk} = \frac{(n_{i..})(n_{.j.})(n_{..k})}{N} \qquad (2.12)$$

Con estos valores ya podemos calcular X^2 mediante la siguiente fórmula:

$$X^2 = \sum_i \sum_j \sum_k \frac{(n_{ijk} - E_{ijk})^2}{E_{ijk}} \qquad (2.13)$$

Una vez encontrado este valor lo único que resta es determinar el número de grados de libertad de X^2.

Para el caso en que la prueba de hipótesis es de independencia entre las tres variables, el número de grados de libertad esta determinado por la relación[18]:

$$g.l. = rcl - r - c - l + 2 \qquad (2.14)$$

Si la hipótesis que nos interesa probar no es la de independencia mutua, tanto los grados de libertad como los valores esperados se calcularán de otra manera (dependiendo del tipo de prueba que queramos realizar). En la sección 3.5 se describe la forma general para el cálculo de los grados de libertad.

Retomando la información del cuadro 2.6 podemos anticipar

---

[18] $g.l. = (rcl-1)-[(r-1)+(c-1)+(l-1)] = rcl-r-c-l+2$. Véase Everitt, op. cit.

que la prueba de independencia mutua será rechazada, debido a que ya habíamos encontrado evidencia estadística de no independencia entre la ocupación y el estado civil de las trabajadoras[19].

A pesar de ello desarrollaremos el procedimiento de los cálculos, empleando la fórmula 2.12 .

El primer paso consiste en encontrar los valores esperados de cada una de las casillas de la tabla. Si las variables fueran independientes, en la casilla correspondiente a las mujeres casadas entre 10 y 29 años y que trabajan como vendedoras ambulantes, deberíamos esperar el siguiente número de casos:

$n_{1..} = 34$

$n_{.1.} = 147$ $\qquad E_{111} = \dfrac{(34)(147)(391)}{(698)(698)}$

$n_{..1} = 391$

$n_{...} = N = 698$

Si obtenemos de ésta manera todos los valores y utilizamos la fórmula (2.13) llegamos al valor $X^2 = 791.82$, con los siguientes grados de libertad:

$$g.l. = (4)(2)(4) - 4 - 2 - 4 + 2$$
$$= 32 - 10 + 2 = 24$$

El valor de la distribución teórica X^2 con 24 grados de libertad es 33.9, lo cual significa que rechazamos la hipótesis de independencia mutua, como ya se había previsto.

Por consiguiente debemos hacer un análisis más profundo de la información. Si la hipótesis de independencia mutua ha sido rechazada, significa que hay interacción al menos entre dos de

---

[19]Véase el apartado 2.4

las variables. Ya vimos que la ocupación y el estado civil estadísticamente no son independientes. Al introducir una tercera variable, en este caso la edad, podríamos esperar que esta relación se mantuviera, o incluso se acentuara o desvaneciera. Se podría pensar que las mujeres que se encuentran en la rama de la producción son las jóvenes y solteras. La proposición empírica en este caso es que si una mujer tiene entre 10 y 29 años y aún es soltera, tendrá mayor probabilidad de ser absorbida por la rama de la producción, que por los servicios.

Pero puede suceder lo contrario; es posible que la no independencia entre la ocupación y el estado civil sea solamente una relación espúria. Es decir, ésta supuesta asociación puede hallar su explicación en una variable que aún no se ha considerado al usar una tabla de dos dimensiones. El hacer el estudio de las variables en forma simultánea, nos ayuda a detectar este tipo de situaciones.

Sabemos que hay relación entre la ocupación y el estado civil, pero no sabemos nada todavía acerca de la relación entre la ocupación y la edad, ni entre ésta y el estado civil.

A continuación plantearemos las otras hipótesis que se pueden desprender -en términos de probabilidades- del análisis tridimensional de variables cualitativas. Estas hipótesis pueden hacerse extensivas a más de tres dimensiones.

2.5.2. Hipótesis de independencia parcial.

Una posible razón para rechazar la hipótesis de independencia mutua es que dos de las variables estén asociadas y la tercera sea independiente de ambas. Sería el caso, por ejemplo, de que la edad fuera independiente del estado civil y de la ocupación.

La hipótesis de independencia parcial puede plantearse de la siguiente manera:

$$Ho(1): P_{ijk} = P_{..k} P_{ij.} \qquad (2.15)$$

(Las variables 1 y 2 estan asociadas y ambas son independientes de la variable 3).

De manera similar pueden plantearse las hipótesis equivalentes, en las que las variables independientes pueden ser tanto la variable renglón como la variable columna.

$$Ho(2): P_{ijk} = P_{i..} P_{.jk} \qquad (2.17)$$

(Las variables 2 y 3 estan asociadas y ambas son independientes de la variable 1).

$$Ho(3): P_{ijk} = P_{.j.} P_{i.k} \qquad (2.18)$$

(Las variables 1 y 3 estan asociadas y ambas son independientes de la variable 2).

La hipótesis Ho(1) significa que la probabilidad de que una observación sea clasificada en la ijk-ésima casilla de la tabla, está dada por el producto de $P_{..k}$ -probabilidad de que ésta caiga en la k-ésima categoría de la variable edad- y $P_{ij.}$ -la probabilidad de que ésta esté en la ij-ésima casilla de la clasificación renglón, columna.

Si la edad (k) es independiente tanto de la ocupación (i) como del estado civil (j) se debe de cumplir que:

$$P_{i.k} = P_{i..}\ P_{..k} \qquad y \qquad P_{.jk} = P_{.j.}\ P_{..k}$$

	edad (k)			
ocupación (i)		$P_{i.k}$		$P_{i..}$
		$P_{..k}$		

	edad (k)			
edo. civil (j)		$P_{.jk}$		$P_{.j.}$
		$P_{..k}$		

Por lo que los valores esperados que estimaríamos en caso de que la hipótesis de independencia parcial fuera cierta según Ho(1) serían:

$$E_{ijk} = N\, p_{..k}\, p_{ij.} \tag{2.19}$$

como en el caso de dos variables, se tiene que:

$$\hat{p}_{..k} = \frac{n_{..k}}{N} \qquad y \qquad \hat{p}_{ij.} = \frac{n_{ij.}}{N} \tag{2.20}$$

Como vemos, para calcular los estimadores necesitamos los marginales de la variable edad y los marginales de las variables ocupación y estado civil sumados sobre la variable edad. De tal forma podemos sustituir los valores en la relación (2.19).

$$E_{ijk} = N\, \frac{n_{..k}}{N}\, \frac{n_{ij.}}{N}$$

$$E_{ijk} = \frac{(n_{..k})(n_{ij.})}{N} \tag{2.21}$$

Con éstos valores podemos calcular el estadístico X^2 con (rcl-cl-r+1) grados de libertad y comparar el resultado con su correspondiente valor en tablas.

Para ejemplificar esta prueba utilicemos los datos de la tabla 2.6. Sustituyendo los marginales correspondientes en la relación 2.21 como se muestra a continuación se obtienen las frecuencias esperadas bajo la hipótesis de independencia parcial.

$$E_{111} = \frac{n_{..1} \quad n_{11.}}{N} = \frac{(391)(14)}{698} = 7.84$$

Una vez encontrados los valores esperados podemos calcular la estadística X^2 mediante la relación 2.13. El resultado del cálculo nos da un valor de 291.46; si lo comparamos con el 32.6 de la X^2 de tablas con 21 grados de libertad rechazamos la hipótesis de independencia parcial.

Cuadro 2.7
Valores esperados bajo la hipótesis de independencia parcial

	EDAD									
	Jóvenes					no Jóvenes				
	ESTADO CIVIL					ESTADO CIVIL				
OCUPACION	cas	div	vda	solt	total	cas	div	vda	solt	total
v. ambul.	7.8	3.9	6.2	1.1	19.0	6.2	3.1	4.8	0.9	15.0
s. n. d.	11.2	7.3	6.2	4.5	29.1	8.8	5.7	4.8	3.5	22.9
s. dom.	36.4	39.2	29.7	130.5	235.8	28.6	30.8	23.3	102.5	185.2
prod.	26.9	14.6	6.2	59.4	107.0	21.1	11.4	4.8	46.6	84.0
total	82.3	65.0	48.2	195.5	391.0	64.7	51.0	37.8	153.5	307.0

Si la prueba de independencia parcial resulta ser no significativa (i.e. no se rechaza Ho), se puede aglutinar[20] (reunir mediante suma) la tabla eliminando la variable que

[20]La traducción corresponde al verbo "collapse" en inglés. El equivalente en español sería aglutinar en lugar de colapsar que emplean algunas traducciones.

resultó ser independiente de las dos restantes y llegaremos a los mismos resultados[21]. En el caso tridimensional nos quedará solamente una tabla, a la que aplicando la prueba X^2 encontraremos que es significativa.

2.5.3 Hipótesis de independencia condicional.

Otra alternativa que se presenta al rechazar la hipótesis de independencia mutua es la posibilidad de que exista independencia entre dos variables dentro de cada nivel de la tercera variable, pero que cada una de ellas esté asociada por separado con ésta última. En tal caso las hipótesis a plantear, considerando todas las posibles combinaciones, se expresarían en los siguientes términos:

$$Ho(1): P_{ijk} = \frac{P_{i.k}\, P_{.jk}}{P_{..k}}$$ (las variables 1 y 2 son condicionalmente independientes en cada nivel k de la tercera variable)

$$Ho(2): P_{ijk} = \frac{P_{ij.}\, P_{.jk}}{P_{.j.}}$$ (las variables 1 y 3 son condicionalmente independientes en cada nivel j de la segunda variable)

$$Ho(3): P_{ijk} = \frac{P_{ij.}\, P_{i.k}}{P_{i..}}$$ (las variables 2 y 3 son condicionalmente independientes en cada nivel i de la primera variable)

Si el caso que nos interesa es el planteado por Ho(1), las

---

[21]Si se elimina la edad, la tabla resultante será la del cuadro 2.4.

frecuencias teóricas que esperaríamos que ocurrieran están dadas por

$$F_{ijk} = N \frac{P_{i.k}\, P_{.jk}}{P_{..k}} \tag{2.23}$$

Para estimar éstas frecuencias podemos sustituir a las probabilidades por sus correspondientes estimadores de máxima verosimilitud:

$$E_{ijk} = N \frac{\hat{p}_{i.k}\, \hat{p}_{.jk}}{\hat{p}_{..k}} = N \frac{\frac{n_{i.k}}{N}\, \frac{n_{.jk}}{N}}{\frac{n_{..k}}{N}}$$

$$= \frac{(n_{i.k})(n_{.jk})}{n_{..k}} \tag{2.24}$$

De igual manera podemos estimar los valores esperados para las hipótesis 2 y 3. El estadístico de prueba es el especificado en (2.13). El cálculo de los grados de libertad para estas tres hipótesis es[22]:

$$Ho(1):\ g.l. = (rcl-1)-(rl-1)-(cl-1)-(l-1) \tag{2.25}$$

$$Ho(2):\ g.l. = (rcl-1)-(rc-1)-(cl-1)-(c-1) \tag{2.26}$$

$$Ho(3):\ g.l. = (rcl-1)-(rc-1)-(rl-1)-(r-1) \tag{2.27}$$

Las restricciones[23] sobre los totales marginales para las

---

[22]Everitt, op. cit.

[23]En la sección 3.6 se brindan mayores detalles.

diferentes hipótesis son[24]:

$$Ho(1):\ n_{i..} = E_{i..} \qquad n_{i.k} = E_{i.k} \tag{2.28}$$
$$n_{.j.} = E_{.j.} \qquad n_{.jk} = E_{.jk}$$
$$n_{..k} = E_{..k}$$

$$Ho(2):\ n_{i..} = E_{i..} \qquad n_{ij.} = E_{ij.} \tag{2.29}$$
$$n_{.j.} = E_{.j.} \qquad n_{.jk} = E_{.jk}$$
$$n_{..k} = E_{..k}$$

$$Ho(3):\ n_{i..} = E_{i..} \qquad n_{ij.} = E_{ij.} \tag{2.30}$$
$$n_{.j.} = E_{.j.} \qquad n_{i.k} = E_{i.k}$$
$$n_{..k} = E_{..k}$$

En este punto debemos hacer notar que no debemos conglomerar (sumar o agregar) las tablas sobre la variable que condiciona dado que hay asociación entre ésta y las dos restantes.

2.5.4 Hipótesis de no interacción de segundo orden.

En esta sección revisaremos otra de las causas (y la última) por las que pudo haberse rechazado la hipótesis de independencia mutua: la presencia de interacción de segundo orden. Esta se manifiesta cuando la relación entre dos variables cualesquiera, cambia de grado o de dirección en los diferentes niveles de la tercera variable. Como en los puntos 2.5.1 a 2.5.3, a continuación presentamos cómo puede expresarse la hipótesis que establece la no asociación de segundo orden en términos de probabilidades:

---

[24] Recuérdese que los marginales esperados son iguales a los observados (pág. 18).

$$Ho: \frac{P_{rcl}\,P_{ijl}}{P_{icl}\,P_{rjl}} = \frac{P_{rck}\,P_{ijk}}{P_{ick}\,P_{rjk}} \qquad (2.31)$$

La expresión de la izquierda puede verse como la medida de asociación producto cruzado[25] o razón de momios, de las dos primeras variables en la l-ésima categoría de la tercera variable. Asimismo la expresión de la derecha representa la medida de asociación de las mismas dos variables pero ahora en la k-ésima categoría de la tercera variable.

Si el cociente

$$\frac{\dfrac{P_{rcl}}{P_{icl}}}{\dfrac{P_{rjl}}{P_{ijl}}} = m$$

(var 2) c j

var 1			c	j	
	r		P_{rcl}	P_{rjl}	
	i		P_{icl}	P_{ijl}	

(var 3) categ. l

(var 2) c j

var 1			c	j	
	r		P_{rck}	P_{rjk}	
	i		P_{ick}	P_{ijk}	

(var 3) categ. k

quiere decir que la razón de momios en la categoría c es m veces la razón de momios en la categoría j, en el nivel l de la tercera variable. Si m=1 entonces podemos decir que no hay relación de primer orden entre las variables 1 y 2 en el nivel l de la

---

[25]Knoke y Burke. op. cit. Véase sección de momios.

tercera variable. La hipótesis de no interacción de segundo orden, establece que sea cual sea la relación entre las variables 1 y 2, ésta será la misma para cada nivel de la tercera variable.

$$Ho: \frac{\frac{P_{rci}}{P_{ici}}}{\frac{P_{rji}}{P_{iji}}} = \frac{\frac{P_{rck}}{P_{ick}}}{\frac{P_{rjk}}{P_{ijk}}} \qquad (2.32)$$

Un caso particular de esta hipótesis es la hipótesis de independencia condicional, ya que en ella se postula que dos variables cualesquiera son condicionalmente independientes, para todos los niveles de la tercera variable. Es decir la relación que priva entre las variables 1 y 2 no cambia en ningún nivel de la variable 3, y así sucede para cualquiera de las tres parejas susceptibles de prueba. Por lo tanto la hipótesis de no interacción de segundo orden implica que la interacción entre cualquier par se mantiene constante en los diferentes niveles o categorías de la variable restante. Pero aquí surge un problema, los estimadores de los valores esperados bajo el supuesto de Ho no se pueden obtener por el producto de totales marginales como se había venido haciendo, sino que es necesario recurrir a un método iterativo, el cual veremos en el siguiente capítulo.

2.6 Estadística de razón de verosimilitud.

En las pruebas realizadas en esta sección se ha utilizado el estadístico X . Pero existe otro estadístico que es L^2 (ji-cuadrada de razón de verosimilitud)[26], cuyo valor se obtiene por la expresión:

$$L^2 = 2\sum\sum n_{ij} \left(\ln \frac{n_{ij}}{E_{ij}}\right) \qquad (2,33)$$

Bajo el supuesto de independencia este estadístico tiene una distribución X^2 con un número de grados de libertad equivalente al de las hipótesis vistas anteriormente. La construcción de este estadístico nos permite hacer particiones aditivas para comparar modelos anidados[27]. Esta propiedad nos será de gran utilidad más adelante en la selección del mejor modelo de ajuste de la información.

---

[26] Bishop, Fienberg y Holland. "Discrete Multivariate Analysis: Theory and Practice". Cambridge, Massachussetts, Institute of Technology. 1975.

[27] Fienberg. op. cit.

CAPITULO III

Modelos loglineales

3.1 Introducción

Hasta el momento hemos visto cómo se desarrolla una prueba de hipótesis para analizar una tabla de contingencia y cómo se plantea en forma evaluable mediante las observaciones.

En este capitulo abordaremos un nuevo tópico, y es el que se refiere al ajuste de modelos loglineales y estimación de sus parámetros.

El término de modelo al que se hará referencia es el resultado de una teoría o trabajo conceptual acerca de las observaciones, mientras que los parámetros del mismo son los representantes de los efectos que tienen las variables particulares o combinaciones de variables para determinar los valores esperados de las observaciones.

Los modelos llamados lineales tienen como característica expresar los valores esperados en cada casilla en términos de una combinación lineal[1] de parámetros.

---

[1] Si una variable se puede obtener a traves de un subconjunto de k variables mediante una expresión del tipo:
$a_1x_1 + a_2x_2 + \ldots + a_kx_k$ con $a_i \neq 0$, $i=1,2,\ldots,k$
se dice que esa variable es una combinación lineal de las k

Las ventajas que ofrecen estos modelos son que permiten una aproximación sistemática al análisis de tablas multidimensionales complejas, y proveen estimaciones de la magnitud de los efectos de interés, lo que permite evaluar la importancia relativa de los mismos.

Por lo expuesto anteriormente cabe señalar que si nuestro interés radica en el análisis de tablas de dos dimensiones únicamente, deberán utilizarse las técnicas apropiadas y suficientes para ello[2].

3.2 Modelos loglineales.

En esta sección se hará la exposición para tablas de contingencia de 2 dimensiones con el propósito de extender posteriormente el modelo a tres o más dimensiones. En la sección 3.3 se verá el modelo tridimensional.

La construcción del modelo está basada en la idea de independencia vista en el capítulo II. La hipótesis establecía que si dos variables son independientes se debe de cumplir que:

$$P_{ij} = P_{i.}\, P_{.j} \quad \text{para todos los valores } i, j . \tag{3.1}$$

variables que aparecen en la expresión.

[2]Véase "Análisis cuantitativo de variables cualitativas". F. Cortés y R. M. Rubalcava. El Colegio de México. México. 1987.

Esta relación plantea una estructura particular o MODELO para los datos, y significa que para la población, la probabilidad de que una observación caiga en la ij-ésima casilla es simplemente el producto de las correspondientes probabilidades marginales.

Una cuestión ahora es cómo expresar esta probabilidad, (o una función de ella), como la suma de las marginales,(o alguna función de ellas), para tener una función lineal. Tomando los logaritmos naturales[3] de 3.1 se tiene que:

$$\ln P_{ij} = \ln P_{i.} + \ln P_{.j} \tag{3.2}$$

Si ésta misma relación la expresamos en términos de las frecuencias teóricas, nos queda[4] :

$$\ln F_{ij} = \ln F_{i.} + \ln F_{.j} - \ln N \tag{3.3}$$

Sumando sobre i,

$$\sum_{i=1}^{r} \ln F_{ij} = \sum_{i=1}^{r} \ln F_{i.} + r \ln F_{.j} - r \ln N \tag{3.4}$$

Sobre j,

---

[3]También se les conoce como logaritmos neperianos. Su base es el número e.

[4]
$$\ln p_{ij} = \ln \frac{F_{ij}}{N} = \ln \frac{(F_{i.})(F_{.j})}{N} = \ln F_{i.} + \ln F_{.j} - \ln N$$

$$\sum_{j=1}^{c} \ln F_{ij} = c \ln F_{i.} + \sum_{j=1}^{c} \ln F_{.j} - c \ln N \qquad (3.5)$$

Y por último sobre i y sobre j,

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij} = c \sum_{i=1}^{r} \ln F_{i.} + r \sum_{j=1}^{c} \ln F_{.j} - rc \ln N \qquad (3.6)$$

Con estas ecuaciones se puede llegar a demostrar que la relación 3.3 se puede expresar de la siguiente manera

$$\ln F_{ij} = \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij}}{rc} + \frac{\sum_{j=1}^{r} \ln F_{ij}}{r} - \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij}}{rc} + \frac{\sum_{i=1}^{c} \ln F_{ij}}{c} - \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij}}{rc} \qquad (3.7)$$

Esta expresión es una combinación lineal de los logaritmos de las frecuencias[5] . De ahí el nombre de modelo LOGLINEAL.

Para abreviar la representación del modelo se utilizan

[5]En este caso la variable F_{ij} esta expresada como suma de probabilid

simbolos u para indicar a qué parámetro se hace referencia[6]. En el caso de 3.7 ésta puede reducirse a

$$\ln F_{ij} = u + u_{1(i)} + u_{2(j)} \qquad (3.8)$$

$$i = 1,\ldots,r \qquad j = 1,\ldots,c$$

El simbolo u representa el "efecto medio global" (se puede ver en el 1er. sumando de 3.7 que es un promedio), $u_{1(i)}$ representa el "efecto principal" de la i-ésima categoria de la variable 1, $u_{2(j)}$ representa el "efecto principal" de la j-ésima categoria de la variable 2.

Al efecto del renglón se le resta el efecto medio global y lo que queda es el "efecto principal" del renglón i-ésimo. Lo mismo sucede con las columnas.

$$u = \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij}}{r\,c}$$

$$u_{1(i)} = \frac{\sum_{i=1}^{c} \ln F_{ij}}{c} - \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij}}{r\,c} \qquad (3.9)$$

$$u_{2(j)} = \frac{\sum_{j=1}^{r} \ln F_{ij}}{r} - \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \ln F_{ij}}{r\,c}$$

---

[6]Se usará la notación del libro de Everitt. op. cit.

Al igual que en la obtención de los valores esperados de una tabla, es preciso estimar las frecuencias teóricas F_{ij}, bajo el supuesto de independencia. Una vez logrado esto, estimar los parámetros del modelo para someter a prueba si éste se ajusta o no a los datos empíricos.

El modelo propuesto en 3.8 es adecuado para el caso en que se espera independencia estadística . Si esa situación no se postula se debe incorporar un término que corresponde al efecto de asociación entre las variables. Dicho de otro modo, cuando se trata de dos variables se debe incluir un parámetro adicional que mida el efecto de interacción.

Entonces, si se supone que no hay independencia, el modelo a proponer sería el siguiente:

$$\ln F_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \qquad (3.10)$$

y $u_{12(ij)}$ queda como el efecto de interacción entre las categorías i y j de las variables 1 y 2 respectivamente.

La utilidad de la estimación de estos parámetros radica en que con ellos podemos identificar cuáles son las categorías responsables de cualquier alejamiento de la independencia, en la distribución conjunta de las dos variables.

Ahora estamos en condiciones de plantear las hipótesis de independencia en términos de los efectos medidos a través de los parámetros. De manera que si se propone que las variables son independientes se esperará que el término $u_{12(ij)}=0$ para todos los valores de i y j. Esto es equivalente a probar que el término de interacción en 3.10 sea cero, o dicho en otras palabras, que la expresión 3.8 proporciona un modelo adecuado de ajuste para los datos.

El modelo propuesto en 3.10 se conoce como el modelo saturado[7] para una tabla bidimensional . Este modelo ajusta perfectamente a los datos ya que los valores esperados bajo este modelo coinciden exactamente con las frecuencias observadas.

El número de parámetros en el modelo es igual al número de casillas de la tabla.

Veamos cómo se llega al sistema de ecuaciones que permitirá estimar los parámetros.

En el caso de las variables dicotómicas, como la condición migratoria en el ejemplo, cada parámetro es el inverso aditivo de su término complementario:

$u_1 = u_{1(1)} = - u_{1(2)}$

$u_2 = u_{2(1)} = - u_{2(2)}$

$u_{12} = u_{12(11)} = u_{12(22)} = - u_{12(12)} = - u_{12(21)}$

En una tabla de 2 X 2 hay 9 parámetros a estimar y sólo 4 casillas (ver cuadro 3.1). Es decir hay más ecuaciones que incógnitas. Esto indica que el sistema de ecuaciones podría no tener solución única. Si consideramos las restricciones mencionadas arriba, sólo quedan 4 variables linealmente

---

[7]Se dice que un modelo loglineal es saturado si contiene a todos los efectos posibles (global, principales y de interacciones).

independientes: u, u_1, u_2 y u_{12} con lo cual llegamos a un sistema de 4 ecuaciones con 4 incógnitas sin ningún grado de libertad.

Los parámetros nos permitirán llegar en cada casilla a frecuencias estimadas que coincidirán exactamente con las observadas.

Cuadro 3.1

Frecuencias Teóricas para el modelo saturado

$F_{11}=u+u_{1(1)}+u_{2(1)}+u_{12(11)}$	$F_{12}=u+u_{1(1)}+u_{2(2)}+u_{12(12)}$
$F_{21}=u+u_{1(2)}+u_{2(1)}+u_{12(21)}$	$F_{22}=u+u_{1(2)}+u_{2(2)}+u_{12(22)}$

3.3 Modelo loglineal para 3 variables

Hemos mencionado anteriormente que el interés en el uso de los modelo loglineales se centra en la inclusión de 3 o más variables en el análisis. Veamos ahora el modelo loglineal saturado que corresponde a una tabla tridimensional.

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \tag{3.11}$$

En este modelo podemos ver que además de los parámetros de efecto principal de cada una de las variables y los parámetros que miden el efecto de interacción de primer orden de cada pareja de variables, se incluye un parámetro que mide el efecto de interacción de segundo orden entre las 3 variables (asociación simultánea entre las 3 variables, que en el modelo de 2 variables no existía).

Si queremos plantear nuestras hipótesis a través de un modelo loglineal, lo que tenemos que hacer es eliminar del modelo saturado (3.11) aquellos términos que correspondan a relaciones que consideramos que no deben aparecer.

Por ejemplo, si nuestra hipótesis es que no hay asociación ente las tres variables simultáneamente, entonces se debería de cumplir que $u_{123(ijk)}=0$ y el modelo que correspondería a esta hipótesis es:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} \qquad (3.12)$$

Otra hipótesis posible es la que plantea no interacción entre una pareja cualquiera de variables. Tomemos por ejemplo el cruce entre las variables ocupación(i), número de hijos(j) y edad(k). Si se postula que las variables ocupación y edad son condicionalmente independientes (la independencia es condicional a la presencia del número de hijos), el término que se debe igualar a cero es el $u_{13(ik)}$. El modelo a proponer quedaría como (3.11) sin los términos $u_{13(ik)}$, y $u_{123(ijk)}$. El término $u_{123(ijk)}$ debe excluirse también debido a que el desarrollo matemático de los modelos loglineales considera sólo modelos jerárquicos . Bishop, Fienberg y Holland[8] escriben sobre el principio de jerarquía: "Los modelos jerárquicos son aquellos que incluyen a todos los términos de orden inferior que corresponden a cualquier término de orden superior. Es decir si el modelo incluye al término u_{123}, también se deben incluir los términos u_1, u_2, u_3, u_{12}, u_{13}, u_{23}. En contraparte, para cualquier término u que sea igual a cero, todos sus términos correspondientes de orden superior deben también igualarse a cero. Entonces si $u_{13}=0$ debemos tener $u_{123}=0$; sin embargo esto no significa que éste término sea necesariamente igual a cero. Si la hipótesis estadística plantea que un término de interacción de primer orden es cero, se debe someter a prueba el modelo que incluya a todos los términos de interacción (a excepción del término de segundo orden), y observar posteriormente si ese modelo ajusta o no a los datos.

Esta limitante viene determinada por las restricciones de los procedimientos de máxima verosimilitud utilizados para estimar los parámetros[9]. Esto no representa problema por el hecho de que la mayoría de las tablas pueden describirse por una serie de modelos jerárquicos.

En el capítulo IV veremos cómo se realiza la selección del modelo con mejor ajuste.

---

[8] Bishop, Fienberg y Holland. op. cit.

[9] Bishop, Fienberg y Holland. op. cit.

3.3.1 Modelo loglineal de independencia mutua

Del modelo propuesto en 3.11 puede desprenderse también la hipótesis de independencia mutua, que es la que especifica que no hay asociación entre ninguna de las variables. En el ejemplo esto implicaría que la ocupación no está relacionada ni con la edad ni con el número de hijos, asimismo que el número de hijos no está asociado a la edad. En términos de los parámetros la hipótesis quedaría expresada así:

$$Ho: u_{12}=0 \ , \ u_{13}=0 \ , \ u_{23}=0 \ , \ u_{123}=0 \ . \qquad (3.13)$$

Lo que llevaría al siguiente modelo:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \qquad (3.14)$$

Este modelo sólo incluye el efecto medio global u, y el efecto principal de cada una de las variables. Si se elige éste como el modelo que mejor ajusta a los datos esto implicaría que las diferencias entre las frecuencias de cada celda simplemente reflejarían diferencias entre los totales marginales de cada variable, no debidas a interacción. Esto se debe a que las frecuencias observadas deben coincidir con las frecuencias esperadas bajo la hipótesis de independencia.

3.3.2 Modelo loglineal de independencia condicional

Otra posibilidad del tipo de relación que puede haber entre las tres variables es que la interacción entre las variables 1 y 2 sea nula en cada uno de los niveles o categorías de la 3a. variable. Por ejemplo que la ocupación y el número de hijos sean independientes, tanto en el caso de las mujeres jóvenes como en el caso de las no jóvenes.

Con este modelo de independencia condicional podemos someter a prueba dicha hipótesis. En este caso se postularía que Ho: $u_{12(ij)}=0$, -que a su vez haría desaparecer el término $u_{123(ijk)}$ -, como hemos mencionado por tratarse de modelos jerárquicos. Esta hipótesis se vería reflejada en el modelo que se describe a continuación:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)} \quad (3.15)$$

Este modelo propone que los logaritmos de los valores esperados de las frecuencias pueden expresarse como una combinación lineal de los parámetros de efecto medio global, efecto principal de cada una de las variables, y de los parámetros del efecto de interacción de primer orden entre las variables 1 y 3 y entre las variables 2 y 3. Para el caso que nos ocupa podemos decir que cada casilla está compuesta por una combinación de: el efecto promedio de todas las variables, los efectos de la ocupación, el No. de hijos y la edad, más las interacciones entre la ocupación y la edad, y entre el No. de hijos y la edad.

3.3.3 Modelo loglineal de independencia parcial

Un último modelo que podríamos estar interesados en proponer, es el que incluye solamente un efecto de interacción de primer orden. Esto se logra haciendo algunos parámetros iguales a cero. Una de las hipótesis que este modelo postula es:

$$Ho:\ P_{ijk} = P_{i..}P_{.jk}.$$

(donde i=ocupación, j=no. de hijos, k=edad)
y se plantea en los siguientes términos:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} \qquad (3.16)$$

La hipótesis implícita es que de entre el conjunto de posibles interacciones entre las variables, sólo debe estar presente la correspondiente al no. de hijos(j) y la edad(k).

Este modelo incluye el mínimo número de posibles interacciones. Ahora bien, si durante el análisis encontramos que este modelo ajusta bien a los datos , podemos concluir que las interacciones que hayan sido excluidas son redundantes, por lo que la dimensión de la tabla deberá reducirse de acuerdo a las interacciones que se excluyan. Bishop, Fienberg y Holland demuestran que una tabla tridimensional puede ser aglutinada sobre la variable que sea independiente de las otras dos. La tabla resultante brindará las mismas conclusiones a las que se llegaría en caso de hacer el análisis sobre todas las variables.

Esto significa que en caso de no rechazar la hipótesis de asociación parcial, podemos realizar el análisis con menos variables, lo que trae consigo una simplificación en los cálculos a realizar más adelante.

3.4 Ajuste de modelos loglineales y estimación de parámetros.

Hasta el momento hemos visto cómo al proponer un modelo loglineal, lo que hacemos es someter a prueba una hipótesis particular acerca de la relación que hay entre las variables que intervienen en una tabla de contingencia. Una vez determinada la hipótesis se procede a estimar los valores esperados en cada casilla, usando las probabilidades que corresponden a esa hipótesis, y entonces se comparan estos valores teóricos con las frecuencias empíricas, en una medida resumen como lo son las estadísticas L^2 o X^2.

Pero pensando en el caso en el que ya hayamos encontrado el modelo que explica mejor a las frecuencias observadas- o dicho en los términos ya usados, el que mejor ajusta a las observaciones-, los parámetros estimados nos sirven para medir el efecto que cada variable tiene sobre la frecuencia observada en cada casilla. Esto permite medir la magnitud del impacto de cada variable y de los efectos de interacción (en la sección 3.7 se explica este punto con más detalle).

Los estimadores de los parámetros en el modelo ajustado se obtienen como funciones de los logaritmos de los valores esperados estimados E_{ijk}, y la forma de tales estimadores es muy similar a los calculados en los modelos de análisis de varianza.

3.4.1 Ejemplo 1

Para ejemplificar la estimación e interpretación de los parámetros, usemos los datos del cuadro 3.2 .

Cuadro 3.2
Mujeres trabajadoras por ocupación y no. de hijos nacidos vivos.

OCUPACION	NO. DE HIJOS NACIDOS VIVOS 0 a 2	3 o +	TOTAL
vendedoras ambulantes	5	29	34
servicio no doméstico	21	31	52
servicio doméstico	309	112	421
producción	142	49	191
TOTAL	477	221	698

Si proponemos el modelo:

$$\ln F_{ij} = u + u_{1(i)} + u_{2(j)} \qquad (3.17)$$

debemos calcular los valores esperados bajo la hipótesis propuesta implícitamente. Como el modelo no incluye el término que mide la interacción entre las 2 variables, se supone que la hipótesis es de independencia mutua.

Eso significa que debemos usar la relación 2.7, para estimar cada valor.

Cuadro 3.3
Valores esperados del cuadro 3.2 bajo el supuesto de independencia entre la ocupación y el no. de hijos.

OCUPACION	NO. DE HIJOS NACIDOS VIVOS 0 a 2	3 o +	total
vendedoras ambulantes	23.23	10.77	34
servicio no doméstico	35.54	16.46	52
servicio doméstico	287.70	133.30	421
producción	130.53	60.47	191
TOTAL	477.00	221.00	698

Por ejemplo para estimar el valor F_{11} se tiene:

$$E_{11} = \frac{n_{1.}\ n_{.1}}{n_{..}} = \frac{34 \times 477}{698} = 23.23$$

En el cuadro 3.3 se presentan los valores esperados bajo la hipótesis de independencia mutua.

En el cuadro 3.4 se presentan los logaritmos naturales de los valores esperados del cuadro 3.2 .

De acuerdo con la fórmula 3.9 el estimador del efecto principal de la primera variable para la primera categoria toma el valor:

$$\hat{u}_{1(1)} = \left(\frac{1}{2}\right)(\ln 23.23+\ln 10.77)-\left(\frac{1}{8}\right)(\ln 23.23+\ldots+\ln 60.47)$$

$$= 2.76 - 3.92 = -1.16$$

Cuadro 3.4
Logaritmos de los valores esperados bajo el supuesto de independencia entre la ocupación y el no. de hijos. ($\ln E_{ij}$).

	NO. DE HIJOS NACIDOS VIVOS		
OCUPACION	0 a 2	3 y +	promedio
vendedoras ambulantes	3.14	2.38	2.76
servicio no doméstico	3.57	2.80	3.18
servicio doméstico	5.66	4.89	5.27
producción	4.87	4.10	4.48
promedio	4.31	3.54	3.92

Podemos notar que el estimador del parámetro es el desvio del promedio de los logaritmos de la categoria con respecto al promedio global de los logaritmos.

Los estimadores para cada categoria de las 2 variables se presentan en el cuadro 3.5 .

Aquí podemos mencionar dos caracteristicas de los estimadores:

i) que los estimadores de los efectos principales sólo reflejan las magnitudes o impactos de los totales marginales. Es decir la magnitud más alta dentro de los estimadores corresponde al total marginal más grande de cada variable y

ii) que la suma de los estimadores para cada variable es cero. El signo indica si el total marginal se encuentra por arriba o por debajo del promedio de los logaritmos para la variable.

Cuadro 3.5
Estimadores de los efectos principales para los datos del cuadro 3.2

OCUPACION	$u_{1(i)}$	NHNV	$u_{2(j)}$
vendedoras ambulantes	-1.167	0-2	0.385
servicio no doméstico	-0.742	3 y +	-0.385
servicio doméstico	1.350		
producción	0.559		
suma	0.000		0.000

Si hacemos $Z_{ij} = \ln E_{ij}$ podemos expresar en forma abreviada los promedios de los logaritmos así:

$$\bar{z}_{i.} = \frac{\sum_{j=1}^{c} \ln E_{ij}}{c} \tag{3.18}$$

$\bar{z}_{i.}$ representa el promedio de los logaritmos de los valores esperados del renglón i-ésimo.

De igual forma los estimadores de los efectos principales pueden escribirse como:

$$\hat{u} = \bar{z}_{..} \tag{3.19}$$

$$\hat{u}_{1(i)} = \bar{z}_{i.} - \bar{z}_{..} \tag{3.20}$$

$$\hat{u}_{2(j)} = \bar{z}_{.j} - \bar{z}_{..} \tag{3.21}$$

y el efecto de interacción entre las dos variables como:

$$\hat{u}_{12(ij)} = \tilde{z}_{i.} - \tilde{z}_{.j} + \tilde{z}_{..} \qquad (3.22)$$

3.4.2 Ejemplo 2

Para ejemplificar el cálculo de los parámetros de interacción de un modelo, veamos el caso de la hipótesis que postula que la ocupación (v1) es independiente del no. de hijos (v2) y de la edad (v3). De esta hipótesis se desprende el modelo:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} \qquad (3.23)$$

Bajo ésta hipótesis calculamos los valores esperados de las frecuencias del cuadro 3.5.1 los cuales se muestran en el cuadro 3.6 .

Cuadro 3.5.1
Mujeres trabajadoras clasificadas por ocupación, no. de hijos nacidos vivos y edad.

	EDAD			
	10 a 29		30 o +	
NHNV / OCUPACION	0 a 2	3 o +	0 a 2	3 o +
vendedoras ambulantes	3	1	2	28
servicio no doméstico	9	2	12	29
servicio doméstico	233	19	76	93
producción	109	15	33	34
total	354	37	123	184

Cuadro 3.6
Valores esperados bajo la hipótesis de independencia parcial para los datos del cuadro 3.5.1 , (E_{ijk}) .

	EDAD			
	10 a 29		30 o +	
NHNV	0 a 2	3 o +	0 a 2	3 o +
OCUPACION				
vendedoras ambulantes	17.2	1.8	6.0	9.0
servicio no doméstico	26.4	2.8	9.2	13.7
servicio doméstico	213.5	22.3	74.2	111.0
producción	96.9	10.1	33.7	50.3

y de éstos obtenemos los logaritmos naturales, mostrados en el cuadro 3.7 ,

Cuadro 3.7
Logaritmos naturales de los valores esperados del cuadro 3.6, (z_{ijk}).

	EDAD			
	10 a 29		30 o +	
NHNV	0 a 2	3 o +	0 a 2	3 o +
OCUPACION				
vendedoras ambulantes	2.84	0.59	1.79	2.20
servicio no doméstico	3.27	1.03	2.22	2.62
servicio doméstico	5.36	3.10	4.31	4.71
producción	4.57	2.31	3.52	3.92

Ya que tenemos los z_{ijk}, podemos pasar a calcular los estimadores de los parámetros que componen al modelo. Para el análisis tridimensional los estimadores están dados por las siguientes relaciones:

El efecto medio global:

$$\hat{u} = \bar{z}_{...} \qquad (3.24)$$

Los efectos principales de cada variable

$$\hat{u}_{1(i)} = \bar{z}_{i..} - \bar{z}_{...} \quad (3.25)$$

$$\hat{u}_{2(j)} = \bar{z}_{.j.} - \bar{z}_{...} \quad (3.26)$$

$$\hat{u}_{3(k)} = \bar{z}_{..k} - \bar{z}_{...} \quad (3.27)$$

y los efectos de interacción de 1er. orden

$$\hat{u}_{12(ij)} = \bar{z}_{ij.} - \bar{z}_{i..} - \bar{z}_{.j.} + \bar{z}_{...} \quad (3.28)$$

$$\hat{u}_{13(ik)} = \bar{z}_{i.k} - \bar{z}_{i..} - \bar{z}_{..k} + \bar{z}_{...} \quad (3.29)$$

$$\hat{u}_{23(jk)} = \bar{z}_{.jk} - \bar{z}_{.j.} - \bar{z}_{..k} + \bar{z}_{...} \quad (3.30)$$

Como el único efecto de interacción de 1er orden en el modelo 3.23 es el u_{23}, utilicemos la relación 3.30 para estimar sus valores

$$\hat{u}_{23(11)} = \bar{z}_{.11} - \bar{z}_{.1.} - \bar{z}_{..1} + \bar{z}_{...}$$

$$\hat{u}_{23(21)} = \bar{z}_{.21} - \bar{z}_{.2.} - \bar{z}_{..1} + \bar{z}_{...}$$

$$\hat{u}_{23(12)} = \bar{z}_{.12} - \bar{z}_{.1.} - \bar{z}_{..2} + \bar{z}_{...}$$

$$\hat{u}_{23(22)} = \bar{z}_{.22} - \bar{z}_{.2.} - \bar{z}_{..2} + \bar{z}_{...}$$

Los promedios de los logaritmos se calculan extendiendo a tres variables la relación definida en (3.18) y aplicándola a los datos del cuadro 3.7.

$$\bar{z}_{...} = 3.02 \qquad \bar{z}_{..1} = 3.48$$
$$\bar{z}_{1..} = 1.86 \qquad \bar{z}_{..2} = 2.56$$
$$\bar{z}_{2..} = 2.28 \qquad \bar{z}_{.11} = 4.01$$
$$\bar{z}_{3..} = 4.37 \qquad \bar{z}_{.12} = 1.76$$
$$\bar{z}_{4..} = 3.58 \qquad \bar{z}_{.21} = 2.96$$
$$\bar{z}_{.1.} = 2.88 \qquad \bar{z}_{.22} = 3.36$$
$$\bar{z}_{.2.} = 3.16$$

Con estos valores, si aplicamos las formulas 3.24, 3.25, 3.26, 3.27 y 3.30, podemos evaluar cada uno de los estimadores de los parámetros del modelo propuesto en 3.23 ,

$$\hat{u} = 3.02 \qquad \hat{u}_{2(1)} = -0.137 \qquad \hat{u}_{23(11)} = 0.66$$
$$\hat{u}_{1(1)} = -1.167 \qquad \hat{u}_{2(2)} = 0.137 \qquad \hat{u}_{23(21)} = -0.66$$
$$\hat{u}_{1(2)} = -0.472 \qquad \hat{u}_{3(1)} = 0.464 \qquad \hat{u}_{23(12)} = -0.66$$
$$\hat{u}_{1(3)} = 1.350 \qquad \hat{u}_{3(2)} = -0.464 \qquad \hat{u}_{23(22)} = 0.66$$
$$\hat{u}_{1(4)} = 0.559$$

Veamos cómo se descompone cada uno de los valores esperados de la tabla en efectos lineales, por ej. el E_{111}

$$\ln E_{111} = u + u_{1(1)} + u_{2(1)} + u_{3(1)} + u_{23(11)}$$

$$= 3.02 - 1.167 - 0.137 + 0.66 + 0.464 = 2.84$$

Aplicando a los logaritmos la función inversa se observa que podemos recuperar el valor esperado E_{111} bajo Ho:

$$E_{111} = \exp(\ln E_{111}) = \exp(2.84) = 17.1$$

Cuadro 3.8
Estimadores del efecto de interacción entre la edad (v3) y el no. de hijos (v2).

$\hat{u}_{23(11)} = 0.66$	$\hat{u}_{23(12)} = -0.66$	$\hat{u}_{23(1.)} = 0.0$
$\hat{u}_{23(21)} = -0.66$	$\hat{u}_{23(22)} = 0.66$	$\hat{u}_{23(2.)} = 0.0$
$\hat{u}_{23(.1)} = 0.000$	$\hat{u}_{23(.2)} = 0.000$	

Como los estimadores tienen la propiedad de que la suma sobre cualquier variable es cero, no es necesario efectuar el cálculo de todos ellos, esto tiene que ver también con el cálculo de los grados de libertad; en el ejemplo (ver cuadro 3.8) podemos ver que sólo es necesario calcular $u_{23(11)}$, los demás estimadores se pueden obtener por diferencia.

El valor $\hat{u}_{23(11)} = 0.66$ indica que hay una asociación positiva entre las categorías que corresponden a las mujeres "jóvenes" y con "pocos hijos". Esto significa que en esa casilla hay más casos de mujeres con ésas características que las que se esperarían si fuera cierta la hipótesis Ho. En este aspecto cabe hacer mención de que el hecho de obtener un parámetro con un valor diferente de cero puede deberse a la aleatoriedad del muestreo. Por lo que es necesario someter a prueba la hipótesis de que el parámetro es significativamente diferente de cero. Para ver cómo se realiza tal prueba véase la sección 3.7 .

3.5 Ceros en la tabla

Cuando se realiza un cruce entre dos o más variables existe la posibilidad de que haya casillas que no contengan casos. Esto puede deberse a dos causas: a) que la combinación entre las categorías correspondientes a esa casilla sea tal que no tenga posibilidad de ocurrir empíricamente; a ese tipo de ceros se les conoce como ceros a priori o ceros estructurales, y b) que el número de celdas resultantes del cruce sea grande en relación al número de observaciones, o visto de otra manera, que el tamaño de la muestra sea insuficiente para el número de casillas de la tabla; éstos son los llamados ceros muestrales. Everitt recomienda para eliminar éstos ceros aumentar el tamaño de la muestra, que sería la forma más natural, mas cuando ésto no sea posible podría ser necesario incrementar con una constante todas las frecuencias de la tabla, por ejemplo con 0.5, antes de empezar el análisis. Algunos autores como Fienberg (1969) desarrollan un método para calcular el valor de la constante que debe sumarse a cada casilla.

Cuando se tienen ceros estructurales en una tabla entonces ya no es necesario calcular los parámetros de ésas celdas, puesto que de antemano sabemos que las frecuencias son cero. Pero si debemos modificar los grados de libertad.

Cuando se presenta el caso de los ceros estructurales se debe restar a los grados de libertad el número de casillas que tienen ceros estructurales. Los nuevos grados de libertad serán:

$$g. \text{ de } l. = N_1 - N_2 - N_3$$

donde N_1 = número total de casillas, N_2 = número de parámetros que se deben estimar y N_3 = número de casillas con ceros estructurales.

3.5.1 Forma general para el cálculo de los grados de libertad

En general el número de grados de libertad se puede calcular de dos maneras:

1) Se calcula el número de parámetros independientes que son iguales a cero. Por ej: si se postula que $u_{123} = u_{12} = 0$ se suman los parámetros asociados con cada uno de éstos términos (ver tabla en la siguiente página):

   $(i-1)(j-1)(k-1) + (i-1)(j-1) = k(i-1)(j-1)$.

2) Se determina el número de parámetros estimados Tp, y se resta este número del total de casillas estimadas Te. Por ej: si el modelo propuesto es $u + u_1 + u_2 + u_3$, el número de parámetros estimados es : $1 + (i-1) + (j-1) + (k-1)$. Entonces los grados de libertad serán:

   $Te - Tp = ijk - [1 + (i-1) + (j-1) + (k-1)]$

Tabla 1
Grados de libertad asociados con cada término u en 3 dimensiones

término u	Número de parámetros independientes				
u_{123}	$(I-1)(J-1)(K-1)$	=	IJK	$-(IJ+JK+IK)+(I+J+K)$	-1
u_{12}	$(I-1)(J-1)$	=	IJ	$-(I+J)$	$+1$
u_{13}	$(I-1)(K-1)$	=	IK	$-(I+K)$	$+1$
u_{23}	$(J-1)(K-1)$	=	JK	$-(J+K)$	$+1$
u_1	$(I-1)$	=		I	-1
u_2	$(J-1)$	=		J	-1
u_3	$(K-1)$	=		K	-1
u	1	=			1
total	IJK				

3.6 Totales marginales fijos

Como se ha visto en secciones anteriores, las diversas hipótesis que se han mostrado se plantean en términos de probabilidades. Estas a su vez implican que conjuntos particulares de totales marginales de valores esperados sean iguales a los correspondientes totales marginales de los valores observados. En forma equivalente, si lo vemos en los términos de los modelos loglineales, los parámetros incluidos en el modelo nos dirán cuáles serán las restricciones marginales que deberán cumplir los valores esperados. Veamos el caso del modelo tridimensional sin ningún término de interacción:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \qquad (3.31)$$

este modelo establece las siguientes restricciones:

$$E_{i..} = n_{i..} \quad , \quad E_{.j.} = n_{.j.} \quad , \quad E_{..k} = n_{..k} \qquad (3.32)$$

sin imponer ninguna restricción sobre los marginales de 2 variables cualesquiera.

Cuando el modelo es de dependencia condicional de un solo par de variables, por ejemplo:

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} \quad (3.33)$$

se tiene que además de las restricciones anteriores se debe de cumplir que:

$$E_{.jk} = n_{.jk} \quad (3.34)$$

Es importante hacer esta consideración debido a que en ciertos casos, por restricciones del diseño de la muestra[10], se deben de mantener fijos determinados totales marginales. Esto implica que se deben incluir en el modelo los términos correspondientes a esa restricción. Bishop (1969) brinda mayores detalles al respecto.

Por ejemplo, en el modelo propuesto en 3.23 se incluye el término $u_{23(jk)}$. Esto implica que se deben de cumplir las condiciones 3.32 y 3.34. Si calculamos $E_{.jk}$ para $j=1,2$ y $k=1,2$ podemos comprobar que se satisfacen dichas condiciones. Para tal efecto debemos sumar sobre la variable ocupación los valo-

[10]Por ejemplo en ciertos análisis psicológicos se busca comparar dos muestras de igual tamaño.

res mostrados en el cuadro 3.5.1. De esta forma encontraremos que

$$E_{.11} = 354 = n_{.11},$$

$$E_{.12} = 123 = n_{.12},$$

$$E_{.21} = 37 = n_{.21}$$

$$E_{.22} = 184 = n_{.22}\ .$$

3.7 Obtención de los valores esperados mediante el proceso iterativo.
· Modelo de no interacción de segundo orden.

En los ejemplos 1 y 2 desarrollados en la sección 3.4 vimos que los valores esperados de las casillas se calculaban con base en la hipótesis estadística previamente establecida. Esto se puede lograr cuando la hipótesis sea de independencia parcial, pero hay casos en los que la obtención de tales frecuencias no puede hacerse por vía directa, y es preciso recurrir a otros métodos para llegar a ellos. Esto se debe a que para tales hipótesis las ecuaciones de máxima verosimilitud no tienen solución explícita.

Varios son los autores que han propuesto caminos alternativos para este problema. Los primeros en hacerlo fueron Deming y Stephan (1940), quienes desarrollaron un método que consiste en ajustar los valores esperados en forma iterativa.

Otro método es el Newton-Raphson, aportado por Bock(1972) y Haberman (1974).

El proceso que describiremos a continuación será el propuesto por Deming y Stephan y para ejemplificar tal procedimiento utilizaremos los datos del cuadro 3.5.1.

En el caso del análisis tridimensional, el único modelo para el cual no es posible obtener los valores esperados en forma directa es el que plantea la hipótesis de no interacción de segundo orden.

$$Ho: u_{123(ijk)} = 0$$

$$\ln F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} \qquad (3.35)$$

Como ya hemos probado antes otras hipótesis sobre la misma información, consideremos este modelo sólo como una vía para ejemplificar la obtención de los valores esperados iterativamente.

Lo que hace el proceso iterativo es ajustar el valor de cada celda utilizando el valor obtenido en el paso anterior, y para el arranque supone que en cada celda la frecuencia E_{ijk} toma el valor uno, hasta satisfacer la primera de las restricciones marginales. Por lo expuesto en la sección 3.6, se observa que los marginales de los valores esperados $(E_{ij.}, E_{i.k}, E_{.jk})$ deben ser iguales a los correspondientes marginales de los valores observados $(n_{ij.}, n_{i.k}, n_{.jk})$. De aquí que el primer valor del proceso esté dado por

$$E_{ijk}^{(1)} = \frac{(E_{ijk}^{(0)})(n_{ij.})}{E_{ij.}} \qquad (3.36)$$

De donde se desprende que $E_{ij.} = n_{ij.}$

Enseguida se busca satisfacer la segunda restricción ($E_{i.k}=n_{i.k}$) con los valores esperados obtenidos en el primer paso

$$E_{(ijk)}^{(2)} = \frac{(E_{ijk}^{(1)})(n_{i.k})}{E_{i.k}} \qquad (3.37)$$

de manera análoga a la relación (3.34) se tiene que:

$E_{i.k} = n_{i.k}$

Por último, el ciclo se completa usando los valores obtenidos en (3.37) para satisfacer la tercera restricción ($E_{.jk}= n_{.jk}$), con la siguiente relación:

$$E_{ijk}^{(3)} = \frac{(E_{ijk}^{(2)})(n_{.jk})}{E_{.jk}} \qquad (3.38)$$

también aquí se cumple que $E_{.jk} = n_{.jk}$

Para iniciar un nuevo ciclo se toman los valores obtenidos en (3.38) y se sustituyen en (3.36). El proceso se detiene hasta que las diferencias entre el último y el penúltimo valores esperados sean menores a una pequeña cantidad preestablecida, digamos de 0.01. Entonces para encontrar los valores esperados bajo la hipótesis implícita en el modelo propuesto en (3.35), tenemos que para todas las casillas:

$E_{ijk}^{(o)}=1$; $i=1,4$, $j=1,2$, $k=1,2$.

Retomando el ejemplo del cruce entre las variables ocupación, edad y número de hijos tenemos que para encontrar los valores esperados que satisfagan al modelo 3.35 debemos empezar por evaluar la relación 3.36 . Para ello necesitamos disponer de los siguientes marginales:

$E^{(o)}_{11.} = E_{111} + E_{112} = 1+1 = 2$ $\quad E^{(o)}_{12.} = 2$

$E^{(o)}_{21.} = 2$ $\quad E^{(o)}_{22.} = 2$

$E^{(o)}_{31.} = 2$ $\quad E^{(o)}_{32.} = 2$

$E^{(o)}_{41.} = 2$ $\quad E^{(o)}_{42.} = 2$

$n_{11.} = 4$	$n_{12.} = 30$	$n_{1.1} = 5$	$n_{1.2} = 29$	$n_{.11} = 354$
$n_{21.} = 11$	$n_{22.} = 41$	$n_{2.1} = 21$	$n_{2.2} = 31$	$n_{.21} = 123$
$n_{31.} = 252$	$n_{32.} = 169$	$n_{3.1} = 309$	$n_{3.2} = 112$	$n_{.12} = 37$
$n_{41.} = 124$	$n_{42.} = 67$	$n_{4.1} = 142$	$n_{4.2} = 49$	$n_{.22} = 184$

Con estos valores podemos proceder a evaluar la relacion 3.36.

$$E_{111}^{(1)} = \frac{(1)(4)}{2} = 2$$

$$E_{211}^{(1)} = \frac{(1)(11)}{2} = 5.5$$

$$E_{311}^{(1)} = \frac{(1)(252)}{2} = 126$$

$$E_{411}^{(1)} = \frac{(1)(124)}{2} = 62$$

$$E_{121}^{(1)} = \frac{(1)(30)}{2} = 15$$

$$E_{221}^{(1)} = \frac{(1)(41)}{2} = 20.5$$

$$E_{321}^{(1)} = \frac{(1)(169)}{2} = 84.5$$

$$E_{421}^{(1)} = \frac{(1)(67)}{2} = 33.5$$

$$E_{112}^{(1)} = \frac{(1)(4)}{2} = 2$$

$$E_{212}^{(1)} = \frac{(1)(11)}{2} = 5.5$$

$$E_{312}^{(1)} = \frac{(1)(252)}{2} = 126$$

$$E_{412}^{(1)} = \frac{(1)(124)}{2} = 62$$

$$E_{122}^{(1)} = \frac{(1)(30)}{2} = 15$$

$$E_{222}^{(1)} = \frac{(1)(41)}{2} = 20.5$$

$$E_{322}^{(1)} = \frac{(1)(169)}{2} = 84.5$$

$$E_{422}^{(1)} = \frac{(1)(67)}{2} = 33.5$$

Para el paso 2 necesitamos calcular los marginales sobre J, (EI.J) sustituyendo en 3.37 los valores obtenidos en el primer paso, por ejemplo:

$$E^{(1)}_{1.1} = E^{(1)}_{111} + E^{(1)}_{121} = 2 + 15 = 17$$

$$E^{(2)}_{111} = \frac{(2)(5)}{17} = 0.60 \qquad E^{(2)}_{312} = \frac{(126)(112)}{210.5} = 67$$

$$E^{(2)}_{211} = \frac{(5.5)(21)}{26} = 4.4 \qquad E^{(2)}_{412} = \frac{(62)(49)}{95.5} = 31.8$$

$$E^{(2)}_{311} = \frac{(126)(309)}{210.5} = 185 \qquad E^{(2)}_{122} = \frac{(15)(29)}{17} = 25.6$$

$$E^{(2)}_{411} = \frac{(62)(142)}{95.5} = 92.2 \qquad E^{(2)}_{222} = \frac{(20.5)(31)}{26} = 24.4$$

$$E^{(2)}_{121} = \frac{(15)(5)}{17} = 4.4 \qquad E^{(2)}_{322} = \frac{(84.5)(112)}{210.5} = 45$$

$$E^{(2)}_{221} = \frac{(20.5)(21)}{26} = 16.6 \qquad E^{(2)}_{422} = \frac{(33.5)(49)}{95.5} \; 17.2$$

$$E^{(2)}_{321} = \frac{(84.5)(309)}{210.5} = 124$$

$$E^{(2)}_{421} = \frac{(33.5)(142)}{95.5} = 49.8$$

$$E^{(2)}_{112} = \frac{(2)((29)}{17} = 3.4$$

$$E^{(2)}_{212} = \frac{(5.5)(31)}{26} = 6.6$$

Para el paso 3 necesitamos los marginales sobre I, $(E_{.jk})$, sustituyendo en 3.38 los valores obtenidos en el segundo paso:

$$E^{(2)}_{.11} = E^{(2)}_{111} + E^{(2)}_{211} + E^{(2)}_{311} + E^{(2)}_{411} = 0.60 + 4.4 + 185 + 92.2$$

$$= 282.2$$

$$E^{(3)}_{111} = \frac{(0.6)(354)}{282.2} = 0.7$$

$$E^{(3)}_{211} = \frac{(4.4)(354)}{282.2} = 5.5$$

$$E^{(3)}_{311} = \frac{(185)(354)}{282.2} = 232.1$$

$$E^{(3)}_{411} = \frac{(92.2)(354)}{282.2} = 115.7$$

$$E^{(3)}_{121} = \frac{(4.4)(123)}{194.8} = 2.8$$

$$E^{(3)}_{221} = \frac{(16.6)(123)}{194.8} = 10.5$$

$$E^{(3)}_{321} = \frac{(124)(123)}{194.8} = 78.3$$

$$E^{(3)}_{421} = \frac{(49.8)(123)}{194.8} = 31.4$$

$$E^{(3)}_{112} = \frac{(3.4)(37)}{108.8} = 1.2$$

$$E^{(3)}_{212} = \frac{(6.6)(37)}{108.8} = 2.2$$

$$E^{(3)}_{312} = \frac{(67)(37)}{108.8} = 22.8$$

$$E^{(3)}_{412} = \frac{(31.8)(37)}{108.8} = 10.8$$

$$E^{(3)}_{122} = \frac{(25.6)(184)}{112.2} = 42$$

$$E^{(3)}_{222} = \frac{(24.4)(184)}{112.2} = 40$$

$$E^{(3)}_{322} = \frac{(45)(184)}{112.2} = 73.8$$

$$E^{(3)}_{422} = \frac{(17.2)(184)}{112.2} = 28.2$$

Con este cálculo de todos los valores E_{ijk} completamos el primer ciclo.

El segundo ciclo comienza evaluando nuevamente la ecuación 3.36 pero con los valores obtenidos en el primer ciclo. El proceso se detiene cuando la diferencia entre los valores del último y penúltimo ciclo es de 0.01 o menos (esta cantidad se fija de antemano). En el cuadro 3.9 se pueden observar los valores esperados calculados bajo este procedimiento. (Estos valores se obtuvieron en 3 ciclos).

Cuadro 3.9
Valores esperados bajo la hipótesis de no interacción de segundo orden.

	EDAD			
	10 a 29		30 o +	
NHNV	0 a 2	3 o +	0 a 2	3 o +
OCUPACION				
vendedoras ambulantes	2.2	1.8	2.8	27.2
servicio no doméstico	9.1	1.9	11.9	29.1
servicio doméstico	230.3	21.7	78.7	90.3
producción	112.4	11.6	29.6	37.4

Con esta tabla podemos comprobar que los valores esperados bajo la hipótesis señalada cumplen con las restricciones sobre los totales marginales fijadas anteriormente. Por ejemplo $E_{11.} = 2.2 + 1.8 = 4$. Este valor coincide con $n_{11.} = 4$.
$E_{1.1} = 2.2 + 2.8 = 5$ y $n_{1.1} = 5$.

También $E_{.11} = 2.2 + 9.1 + 230.3 + 112.4 = 354$ y $n_{.11} = 354$. (De igual modo el resto de los valores).

3.8 Programas disponibles

Como hemos visto en los puntos anteriores el desarrollo de los cálculos para encontrar los estimadores de los parámetros puede volverse una tarea tediosa y complicada, máxime si el análisis incluye más de tres variables. En cierta medida éstas técnicas multivariadas no serían factibles sin el desarrollo de las computadoras. Gracias a la capacidad que han logrado alcanzar éstas máquinas se han podido utilizar los diversos procedimientos del análisis multivariado.

Afortunadamente en el caso de los modelos loglineales se cuenta en la actualidad con una diversidad de programas que realizan estos cálculos. Algunos de ellos se mencionan a continuación: GLIM, LOGLIN, ECTA, HILOGLINEAR (que es parte del paquete estadístico SPSS[11]) y el P4F (que es parte del paquete estadístico Biomedical, BMD[12]), que fué el programa utilizado para desarrollar este trabajo[13]. Como este programa forma parte de un paquete estadístico, sólo deben indicarse algunas especificaciones. El paquete esta dividido en párrafos, los cuales estan indicados por una diagonal.

El primero de ellos:

---

[11]Norusis, Marija J. Statistical Package for Social Sciences. Chicago Illinois, SPSS Inc. 1986.

[12]Biomedical Computer Programs. P. Series. Berkeley, University of California Press. 1977, 880 págs.

[13]Todos los procesos se desarrollaron en un equipo PDP-11/70 en la Unidad de Cómputo de El Colegio de México.

/PROBLEM

sirve para dar el título o comentarios acerca de lo que se desarrolla en el programa:

En el siguiente párrafo, indicado como:

/INPUT

se deben especificar el número de variables que intervienen, el nombre del archivo de donde se tomarán los datos, el formato con el que serán leídas las variables y el número de categorías que tiene cada una de ellas.

En el párrafo

/VARIABLE

se dan los nombres de las variables.A continuación, en el párrafo

/TABLE

se indican los índices con los que se identificarán a las variables, al momento de especificar un modelo.

Si existen casillas con frecuencia cero se puede especificar que a todas las casillas se sume una constante, con la instrucción ADD.

En el párrafo

/FIT

se especifica el modelo (o los modelos) que se deben analizar.

Como opciones, se puede solicitar en el párrafo

/PRINT

el cálculo los valores esperados, los estimadores de los parámetros u, y los estimadores divididos entre su desviación estándar (llamados Zc).

Si se especifica la instrucción ASSOCIATION el programa realiza las pruebas de hipótesis de asociación parcial e independencia condicional de todos los modelos, calculando el valor de X y L con su correspondiente probabilidad.

En el anexo I se muestra el programa con las instrucciones utilizadas para los cálculos de este trabajo.

3.9 Significación de los efectos

En la introducción del presente capítulo mencionamos que una de las ventajas del uso de los modelos loglineales es que nos brindan una magnitud de los efectos de interacción, y para reconocer cuáles son las categorías responsables del rechazo de la hipótesis planteada podemos asimismo realizar pruebas de hipótesis sobre cada uno de éstos valores. En este sentido debemos mencionar que el hecho de encontrar asociación entre variables, depende de la asociación que haya entre las categorías de las mismas[14], de modo que una sola categoría significativa por parte de cada variable basta para considerar que hay interacción.

---

[14] Parzen en su libro "Modern Probability Theory and its Aplications" da la siguiente definición de independencia: Sean X_1 y X_2 2 variables aleatorias conjuntamente distribuidas, con funciones F_{X1} y F_{X2} respectivamente y función de distribución conjunta $F_{X1,X2}$. Se dice que las variables X_1 y X_2 son independientes si para cualesquiera 2 conjuntos de Borel de números reales B_1 y B_2 los eventos $[X_1$ está en $B_1]$ y $[X_2$ está en $B_2]$ son independientes, es decir si
$P[X_1$ está en B_1 y X_2 está en $B_2] = P[X_1$ está en $B_1]P[X_2$ está en $B_2]$

Por lo tanto si consideramos que no hay asociación entre dos variables, podemos plantear que:

$$Ho: \; u_{12(ij)} = 0$$

En este caso debemos usar un estadístico de prueba que contenga el valor de ese parámetro. Por el teorema central del límite sabemos que si a los valores de una variable (sin importar su distribución) se les resta su promedio y se les divide entre su desviación estándar, la variable resultante tendrá una distribución normal con media cero y desviación estándar uno. Si llamamos Zc a la nueva variable entonces:

$$Zc = \frac{\hat{u}_{12(ij)} - u_{12(ij)}}{\sqrt{V\,(u_{12(ij)})}}$$

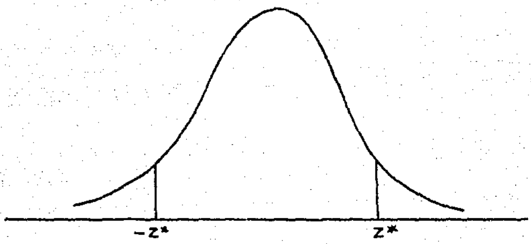
Suponiendo Ho cierta entonces $u_{12(ij)}=0$ y Zc queda como

$$Zc = \frac{\hat{u}_{12(ij)}}{\sqrt{V\,(u_{12(ij)})}}$$

Entonces para cada valor se hará una comparación con el valor respectivo Z de tablas con nivel de significación con el siguiente criterio: si $Zc > Z^{*}$ o $Zc < -Z^{*}$ consideramos que $u_{12(ij)}$ es significativamente diferente de cero con un nivel de confianza $1-\alpha$, lo que significa que las categorías correspondientes a ese parámetro son las responsables de la asociación.

En la gráfica 3.1 se muestra la zona de rechazo para los valores de Zc .

GRAFICA 3.1



Para el ejemplo de la página 80 mostramos a continuación los valores Zc:

Cuadro 3.10
Estimadores estandarizados de los parámecros de interacción (Zc).

$Z_{12(11)} = -1.55$	$Z_{13(11)} = -3.38$	$Z_{23(11)} = 11.7$
$Z_{12(21)} = -2.27$	$Z_{13(21)} = -0.05$	$Z_{23(12)} = -11.7$
$Z_{12(31)} = 2.95$	$Z_{13(31)} = 3.98$	$Z_{23(21)} = -11.7$
$Z_{12(41)} = 3.85$	$Z_{13(41)} = 3.05$	$Z_{23(22)} = 11.7$
$Z_{12(12)} = 1.55$	$Z_{13(12)} = 3.38$	
$Z_{12(22)} = 2.27$	$Z_{13(22)} = 0.05$	
$Z_{12(32)} = -2.95$	$Z_{13(32)} = -3.98$	
$Z_{12(42)} = -3.85$	$Z_{13(42)} = -3.05$	

Si comparamos cada uno de estos valores contra Z de tablas al 5% ($Z^*_{.025}$ = 1.96) se puede apreciar que en los tres términos de interacción (u_{12}, u_{13}, u_{23}) hay efectos significativos (mayores que 1.96 o menores que -1.96).

El término $Z_{12(21)}$=-2.27 por ser negativo indica que hay menos mujeres mayores de 30 años que trabajan en el servicio doméstico de las que se esperarían en caso de ser cierta Ho. Cuando el signo es positivo como el caso del término $Z_{12(31)}$ significa que hay más mujeres que trabajan en el servicio doméstico y con pocos hijos de las que se esperarían bajo el modelo propuesto en Ho.

Este cuadro brinda una información muy valiosa ya que no sólo indica cuáles son las variables que están asociadas, sino que además nos dice cuales categorías son las responsables de la asociación.

CAPITULO IV

Análisis de tablas de contingencia multidimensionales mediante el uso de los modelos loglineales

4.1 Introducción

Una vez que han sido expuestos los elementos básicos del análisis loglineal para tablas de contingencia se presenta en este capitulo un procedimiento general para el uso e interpretación de los modelos loglineales, tomando un conjunto de variables para explicar la posible relación o interacción entre ellas. El propósito de este capitulo no es el de hacer un análisis profundo de las variables que intervienen en el cuadro que se propone, ni de contradecir posibles conclusiones a las que hayan llegado otros investigadores mediante el uso de otros procedimientos. Unicamente se intenta contrastar las proposiciones teóricas con las empiricas con la ayuda de la herramienta que constituyen los modelos loglineales; además de mostrar cómo éstos pueden ayudar en el proceso de investigación.

En el cuadro (4.1) se desglosan las frecuencias de un grupo de trabajadoras manuales no calificadas tomado de la encuesta "Migración Interna, Estructura Ocupacional y Movilidad Social en el Area Metropolitana de la Cd. de México", realizada por El Colegio de México y el Instituto de Investigaciones Sociales de la UNAM en 1970[1] (ver cuadro 1). Dicha muestra fue clasificada por ocupación, edad, estado civil, número de hijos nacidos vivos, y nivel de instrucción. Estas variables fueron seleccionadas con el fin de ubicar las características de las mujeres que se insertan en los diferentes sectores laborales [2]. El cuadro 4.1 nos muestra la distribución de las 698 mujeres que constituyen la muestra. Las categorías de la variable ocupación se agruparon en vendedoras ambulantes, servicio no doméstico, servicio doméstico, y producción. La edad se consideró de 10 a 29 años, para las jóvenes y de 30 años y más para las no jóvenes. El número de hijos nacidos vivos se consideró de 0 a 2 como pocos hijos y de 3 o más como no pocos hijos. Por último el estado civil se clasificó en 4 categorías: casadas, viudas, divorciadas y solteras.

---

[1] Esta encuesta se realizó entre los años de 1969 y 1970 por el Centro de Estudios Económicos y Demográficos de El Colegio de México, junto con el Instituto de Investigaciones Sociales de la UNAM, bajo la dirección de Humberto Muñoz, Orlandina de Oliveira y Claudio Stern. Dicha encuesta comprendió 2 etapas de estudio: En la fase 'A' se captó información individual general con fines descriptivos . En la fase 'B' se buscó la faceta analítico explicativa. La muestra analizada en este trabajo pertenece a la fase 'A'. Algunos resultados fueron publicados en el libro "Migración y desigualdad social en la ciudad de México". El Colegio de México. Instituto de Investigaciones Sociales, UNAM. México. 1977.

[2] La muestra del cuadro 4.1 comprende sólo a trabajadoras que desempeñan labores consideradas como manuales, y que no necesitan de mucha preparación. Esta información fué tomada de la tesis "Algunas características de las trabajadoras manuales no calificadas de la ciudad de México". FLACSO. México. 1983.

Cuadro 4.1
Trabajadoras manuales no calificadas por ocupación, edad, número de hijos nacidos vivos y estado civil.

EDO CIVIL	NHNV	EDAD	OCUPACION (1)	(2)	(3)	(4)
casada	0-2	10-29	0	1	14	10
		30 Y +	1	3	10	7
	3 o +	10-29	1	1	9	10
		30 Y +	12	15	32	21
divorciada	0-2	10-29	1	2	19	8
		30 Y +	0	3	17	8
	3 o +	10-29	0	1	8	3
		30 Y +	6	7	26	7
viuda	0-2	10-29	0	0	1	1
		30 Y +	1	4	16	5
	3 o +	10-29	0	0	1	1
		30 Y +	10	7	35	5
soltera	0-2	10-29	2	6	199	91
		30 Y +	0	2	33	13
	3 o +	10-59	0	0	1	1
		30 Y +	0	0	0	1
		TOTAL	34	52	421	698

Nota: (1) vendedoras ambulantes, (2) servicio no doméstico
(3) servicio doméstico, (4) producción.

El interés radica en encontrar cuál es la relación, si la hay, entre las variables demográficas con las que han sido caracterizadas las mujeres que trabajan en los sectores antes señalados.

Si realizamos cruces de 2 variables podemos a simple vista o con ayuda de porcentajes, encontrar posibles relaciones entre ellas. Pero si incluimos una tercera o más variables al análisis, el grado de dificultad aumenta en forma considerable. Tal es el caso del cuadro 4.1, si tratamos de ver cual es la relación entre éstas variables con solo mirar la tabla, encontraremos que es un asunto muy complejo. Un análisis multidimensional nos ayudará a estudiar en forma simultánea las relaciones que privan entre las variables, y en este sentido el modelo loglineal nos brinda una alternativa para el análisis de este tipo de tablas[3].

4.2 Análisis de los resultados.

Una vez que ya hemos definido las variables, y las hipótesis de trabajo, procederemos a la elaboración del programa (ver anexo A) que realiza todos los cálculos descritos en el capitulo III. Debemos recordar que estos procesos multivariados han cobrado auge gracias al desarrollo de las computadoras y por consiguiente de los paquetes estadisticos. La

---

[3] Aqui es pertinente hacer una observación: el modelo loglineal no expresa causalidad. El análisis que se realiza mediante éstos modelos nos permite observar cómo interactúan en la malla propuesta las variables, más no a una variable como función de otras. El modelo que corresponde a este tipo de análisis es el modelo Logit, y su equivalente para variables métricas es el modelo de regresión lineal múltiple.

computadora nos ofrece la oportunidad de analizar todos los modelos posibles, por lo que es recomendable establecer las hipótesis previamente, y así contrastar sólo estas hipótesis con los datos, mediante los modelos correspondientes.

El primer paso consiste en efectuar las pruebas de hipótesis sobre los efectos de interacción con el propósito de decidir sobre la inclusión o eliminación de los mismos, y así determinar cuál es el término de interacción de mayor orden que debe de permanecer en el modelo. La salida de uno de estos términos significaría que no hay interacción entre las variables a que hace referencia.

Al revisar los resultados del programa P4F (ver anexo A) podemos observar la tabla de frecuencias, y las marginales para 1,2 y 3 variables. Posteriormente encontramos una tabla que nos muestra los valores tanto de la X^2 de Pearson como del estadístico X^2_L de razón de verosimilitud. Con ellos podemos efectuar la prueba de hipótesis correspondiente a probar que los términos de los efectos de interacción y los términos de los efectos principales son iguales a cero. Así el valor de X^2 obtenido de sustraer del modelo el término de interacción OE (que equivale al término $u_{12(ij)}$), corresponde a probar la hipótesis Ho: $u_{12(ij)}=0$. En términos de los modelos logilineales tendríamos la siguiente relación:

$$\ln F_{ijkl} = u+u_{1(i)}+u_{2(j)}+u_{3(k)}+u_{4(l)}+u_{13(ik)}+u_{14(il)}+u_{23(jk)}+ u_{24(jl)}+u_{34(kl)} \qquad (4.1)$$

Como los modelos son jerárquicos se omiten los términos de orden superior (i.e. las relaciones entre 3 y 4 variables).

Podemos observar que el cálculo de X^2 y X^2_L (págs. 111 y 112) se efectúa para cada una de las combinaciones de términos de interacción. El renglón que incluye el término OE contiene el valor de X^2 y su respectivo nivel de significancia para efectuar la prueba directamente. Si observamos los valores de X^2 calculados para cada uno de los modelos que no incluyen a los términos indicados en el modelo 4.1, vemos que son significativos todos menos el correspondiente a la ocupación y la edad OE [4] (marcado con *, pág 112), lo cual nos indica que éstas variables, a diferencia del resto, probablemente son independientes[5].

Las parejas de variables que se esperaría fueran dependientes (por ser variables demográficas obviamente relacionadas) son las parejas que forman la edad con el no. de hijos, la edad con el estado civil, y el no. de hijos con el estado civil. Pero podemos observar que ningún término de 2do. orden es significativo. Esto quiere decir que no hay asociación cuando intervienen 3 variables conjuntamente. La asociación que existía entre la ocupación y el estado

---

[4] La probabilidad de que X tome ése valor es mayor que 5%, por lo tanto cae en la zona de rechazo.

[5] El programa proporciona todas las combinaciones de interacciones excepto la de mayor orden (que corresponde al modelo saturado).

civil desaparece cuando se introduce la variable edad o cuando se introduce la variable No. de hijos. Con esto encontramos que no hay evidencia empirica para pensar que la ocupación depende de la edad, el No. de hijos, y el estado civil simultáneamente como se suponia al inicio.

Con esta información estamos en condiciones de seleccionar el modelo que mejor explica a las frecuencias observadas. Un modelo a proponer seria por ejemplo el descrito por la relación 4.1 .

Otro modelo podria ser el siguiente:

$$\ln F_{ijkl} = u+u_{1(i)}+u_{2(j)}+u_{3(k)}+u_{4(l)}+u_{12(ij)}+u_{13(ik)} \qquad (4.2)$$

si observamos el valor de X^2 de tablas con 33 grados de libertad (14.82) encontramos que es significativo. Es decir es un buen modelo.

Ahora la pregunta es ¿cuál de los dos es el mejor modelo? Cuando se presenta el caso de que varios modelos son significativos a un cierto nivel de confianza podemos seleccionar: a) el mejor modelo de ajuste, o dicho de otra manera el que mas se acerque a la información de la tabla original, b) el que lo haga con el menor número de parámetros. Un elemento fundamental para decidir, será la concordancia con la argumentación del investigador.

4.3 Selección del mejor modelo de ajuste.

Cuando se trata de escoger el mejor modelo de ajuste de la información se busca seleccionar aquel que contenga el menor número de términos posibles, de esta manera determinamos cuáles son las variables reelevantes y cuáles no lo son. Pero puede suceder que contemos con varios modelos significativos de los cuales unos tengan más términos que otros. Para ello debemos efectuar otra prueba de hipótesis que consiste en probar si la inclusión de un término de interacción hace que el ajuste mejore significativamente. Por ejemplo para el modelo que incluye a todos los términos de interacción de primer orden la razón de verosimilitud vale: $L^2=55.60$. Si restamos un valor de otro lo que hacemos es efectuar una prueba de hipótesis sobre la significación de incluir ese término de interacción,

$L^2_m - L^2_n = 55.60 - 14.85 = 7.78$ y g. l. = 36 - 33 = 3 .

La diferencia entre estos dos estadísticos tiene una distribución asintótica ji-cuadrada con un número de grados de libertad igual a la diferencia entre los grados de libertad de ambos estadísticos.

La diferencia la tenemos que confrontar con el valor de la X^2 de tablas: $X^2_3=7.81$ calculada no es significativa estadísticamente, esto quiere decir que la inclusión del término u_{12} no aporta información reelevante, por lo tanto nos quedamos con el modelo que se presenta en la relación 4.1 .

Una vez que ya se ha escogido el mejor modelo podemos pedir al paquete que calcule los parámetros de los términos de interacción de dicho modelo. La primera tabla que se muestra en el anexo A contiene los valores esperados bajo la hipótesis expresada en el modelo, a los que les hemos llamado E_{ij}, éstos valores se incluyen en el cuadro 3.2 . En seguida se muestran los residuales estandarizados, cuyo cálculo es la diferencia entre el valor esperado y el valor observado dividida entre la raiz cuadrada del valor esperado.

$$\frac{(O_{ij}-E_{ij})}{\sqrt{E_{ij}}} \tag{4.3}$$

La suma de éstos términos al cuadrado nos da el valor de X^2. A continuación el programa presenta los valores de los estimadores de los parámetros loglineales, es decir los u_{ij}.

Por último el paquete nos brinda los valores de los estimadores antes mencionados divididos entre la desviación estándar de los errores. Con ellos podemos efectuar la prueba para determinar cuáles son las categorías que muestran evidencias de la posible independencia o asociación entre variables. En el caso que nos ocupa vemos que todos los términos de interacción de segundo orden son cero (lo cual coincide con la hipótesis planteada inicialmente en 4.1. Por lo tanto sólo analizaremos los valores estandarizados de los marginales de dos variables.

Para la pareja de variables ocupación y edad ninguno de éstos valores resulta significativo. Este resultado era de esperarse ya que como habíamos visto el término u_{12} resultó ser no significativo. Al hacer la prueba de hipótesis sobre el valor u_{12} vimos que no se rechazaba. No así para el resto de los términos. Como podemos observar ningún valor esta fuera del rango del valor de Z^* que deja como zona critica el 5% de la distribución (-1.96,1.96).

Al revisar el resto de las tablas, podemos notar cuáles son las casillas en las que la diferencia entre el valor esperado y el observado no se debe a factores aleatorios.

Por ejemplo en la tabla 4.2 que nos muestra los efectos de interacción divididos entre su desviación estándar para las variables ocupación y número de hijos, los valores significativos corresponden a la catogoria de vendedoras ambulantes. El valor indicado con el asterisco señala que en la muestra hubo significativamente menos mujeres de las esperadas en caso de independencia. El valor negativo se encuentra en la casilla de las mujeres con pocos hijos, y el positivo en las que tienen 3 o más hijos. Esto significa que las mujeres que tienen 3 o más hijos se distribuyen preferentemente entre las que se encuentran trabajando como vendedoras ambulantes.

Cuadro 4.2
Estimadores de los parámetros de interacción entre la ocupación y el NHNV, divididos entre su desviación estándar.

NHNV	OCUPACION vendedoras ambulantes	servicio no doméstico	servicio doméstico	producción
0 - 2	-2.119^*	0.213	1.418	1,581
3 o +	2.119^*	-0.213	-1.418	-1,581

En el caso de los valores obtenidos para las variables ocupación y estado civil, encontramos que solamente una casilla es la que muestra evidencia de la posible no independencia entre ellas, y es la que corresponde a las mujeres casadas que trabajan en el servicio doméstico. En el cuadro 4.3 se muestran las cantidades mencionadas.

En lo que toca al resto de las variables podemos confirmar la relación entre las variables demográficas. El número de hijos está íntimamente ligado a la edad, así como al estado civil. De igual forma se aprecia que las mujeres que alguna vez han estado unidas (casadas, divorciadas, viudas) tienen más hijos que las solteras. Este hecho nos ayuda a reclasificar las variables; por ejemplo podemos dicotomizar el estado civil en mujeres casadas y mujeres alguna vez unidas.

Cuadro 4.3
Estimadores de los parámetros de interacción de las variables ocupación y estado civil, divididos entre su desviación estándar.

	OCUPACION			
ESTADO CIVIL	vendedoras ambulantes	servicio no doméstico	servicio doméstico	producción
casada	0.175	0.595	-2.169*	0.974
divorciada	-0.024	0.597	-0.195	-0.531
viuda	0.970	0.970	0.288	0.041
soltera	-1.006	-1.149	1.438	1.529

También podemos percatarnos de que hay más mujeres casadas con 3 o más hijos, y de que la distribución de mujeres solteras se concentra más hacia la categoría de pocos hijos.

El número de mujeres divorciadas es estadísticamente igual tanto para las que tienen pocos hijos como para las que tienen 3 o más, y el número de mujeres solteras con pocos hijos es visiblemente superior al de las solteras que se encuentran en la categoria complementaria.

Las conclusiones que podemos obtener de los resultados anteriores pueden resumirse de la siguiente manera:

Las cuatro variables analizadas no están relacionadas simultáneamente, es decir la ocupación no está determinada por la edad, el número de hijos y el estado civil al mismo tiempo.

Es cierto que hay relación entre la ocupación y el estado civil y entre la ocupación y el número de hijos, pero esto no significa que haya relación entre las 3 simultáneamente. Esto se explica porque la relación (Edo. civ.- Núm. de hijos) es la misma para las categorias de la variable producción . Por eso la relación no puede trasladarse a una conclusión de tipo tridimensional, por ejemplo, no se puede decir que las mujeres solteras jóvenes y con pocos hijos se insertan preferentemente en algún sector de la producción.

4.2 Análisis con 5 variables

Otro punto interesante de analizar es la inclusión del nivel de instrucción. La ventaja con la que contamos es que podemos observar si la relación entre las variables anteriormente analizadas cambia, ya que podemos realizar el análisis en forma simultánea. Es de esperarse que el nivel alcanzado por la mujer en su formación sea un factor importante en la actividad que vaya a desempeñar.

Considerando las pocas oportunidades de educación que existían en nuestro país y en particular las que hubieran tenido las mujeres al momento de la encuesta, clasificamos a las mujeres de la muestra en mujeres sin primaria y mujeres con primaria o más estudios.

Además de esto se agruparon en una sola clase a las mujeres alguna vez unidas (casadas, viudas, divorciadas) y en otra a las mujeres solteras. Esto se hizo tomando en cuenta los estimadores estandarizados del término de interacción correspondiente a las variables ocupación y estado civil, denotado como $u_{14(ll)}$: se observó que un grupo significativo era el constituído por las mujeres solteras y otro el resto de la muestra.

También se hizo un reagrupamiento respecto al número de hijos, se dividió la clasificación en mujeres sin hijos y mujeres con 1 o más hijos.

Al analizar los valores de X^2_L (ver anexo B, pág. 121), encontramos que se sigue manteniendo la independencia entre la ocupación y el resto de las variables, exceptuando la relación que se presenta entre ésta y el nivel de instrucción. Esto significa que al menos en la muestra, la educación formal recibida está asociada con el sector en el que trabaja la mujer, en el caso de las consideradas manual no calificada.

Más adelante haremos la prueba de hipótesis sobre los parámetros en cada una de las casillas para encontrar el nivel de instrucción responsable de la determinación del sector ocupacional.

Con respecto a la ausencia de relación entre el nivel de instrucción y el número de hijos podemos decir que se trata de un caso particular. En diversos estudios demográficos[6] se ha encontrado que la pauta de comportamiento de la mujer con respecto al número de hijos tiene relación con el nivel de instrucción. Sin embargo quizá se deba tomar en consideración la condición urbana de la mujer, o el nivel de instrucción del esposo. En lo que toca a la muestra se puede observar que no hay elementos que apoyen ésta relación.

Un elemento más para confirmar lo comentado en el punto 4.2 es que las variables demográficas mantienen la relación esperada.

De igual forma las posibles relaciones entre 3 variables no se confirman, es decir no hay términos de interacción de segundo orden y tampoco los hay de orden superior.

Una vez que se han hecho éstas observaciones podemos proceder a la selección del mejor modelo. Tal modelo es el que incluye a los siguientes términos:

$$\ln F_{ijklm} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{5(m)} + u_{15(im)} + u_{23(jk)} + u_{24(jl)} + u_{34(kl)} + u_{45(lm)} \quad (4.4)$$

---

[6] Por ejemplo ver Zambrano, Jorge H. "La relación entre la fecundidad y el grado de escolaridad en el medio rural mexicano y en la Cd. de México". El Colegio de México. Centro de Estudios Económicos y Demográficos. México. 1977. Tesis de maestría.

Con él podemos estimar los valores de los parámetros. El cuadro 4.5 nos muestra los valores de los estimadores de los parámetros en los que hay interacción. En el caso de la ocupación y el nivel de instrucción, nos indica que hay un mayor número de mujeres en el sector de la producción cuyos estudios son de primaria o más; en cambio las mujeres que no tienen primaria se encuentran preferentemente en el resto de los sectores.

Cuadro 4.4
Frecuencias observadas de la muestra clasificada por ocupación y nivel de instrucción.

	OCUPACION			
NIVEL DE INSTRUCCION	vendedoras ambulantes	servicio doméstico	no servicio doméstico	producción
< primaria	27	40	354	79
primaria o más estudios	7	15	67	111

Si observamos el cuadro 4.6 con los estimadores estandarizados vemos que hay 4 valores significativos (indicados con *). Las mujeres que no tienen primaria se concentran significativamente en el servicio doméstico, y las que tienen estudios de primaria o más se inclinan hacia la producción.

Cuadro 4.5
Valores estimados del parámetro de interacción entre la ocupación y el nivel de instrucción.

NIVEL DE INSTRUCCION	OCUPACION vendedoras ambulantes	servicio doméstico	no servicio doméstico	producción
< primaria	0.101	0.088	0.391	-0.580
primaria o más estudios	-0.101	-0.088	-0.391	0.580

Cuadro 4.6
Valores estimados del parámetro de interacción entre la ocupación y el nivel de instrucción divididos entre su desviación estándar.

NIVEL DE INSTRUCCION	OCUPACION vendedoras ambulantes	servicio no doméstico'	servicio doméstico	producción
< primaria	0.755	0.734	5.074*	-7.535*
primaria o más estudios	-0.755	-0.734	-5.074 *	7.535*

CONCLUSIONES

El desarrollo del análisis multivariado aunado al progreso vertiginoso de las computadoras han permitido una mayor profundidad en el estudio de los fenómenos sociales. Considerando que la gama de posibilidades crece con el uso de los programas de análisis multivariado es conveniente conocer las opciones, alcances y límites que éstos nos brindan. El modelo presentado en los capítulos anteriores constituye un ejemplo de las múltiples aplicaciones que se pueden realizar en este campo.

A lo largo de este trabajo se han presentado los conceptos básicos sobre las cuales descansa la teoría del análisis loglineal. Hemos visto que es suficiente con que las variables involucradas en el análisis sean de escala nominal. Y por supuesto el nivel de medición puede ser superior.

Otro requerimiento es que el número de observaciones sea mayor que el número de celdas. Además que las categorías de las variables sean exhaustivas y mutuamente excluyentes. Así como es recomendable que la frecuencia mínima en cada casilla sea de 5 observaciones.

Una vez cumplidos estos requisitos podemos confiar en llegar a conclusiones válidas acerca de la relación estadística entre las variables bajo estudio.

El análisis loglineal nos permite hacer la selección del mejor modelo de ajuste y por consiguiente la evaluación de hipótesis estadísticas referentes a la interacción o independencia de variables. Es decir, el modelo plantea hipótesis de asociación o de independencia entre factores y posibilita determinar cuáles son aquellas relaciones que estadísticamente son signifivativas y cuáles no lo son.

Una vez que se ha seleccionado el mejor modelo, el que mejor ajusta a la información, y el que mejor corresponda a la argumentación del estudio, podemos adentrarnos a las tablas resultantes de los efectos de interacción que hayan permanecido en él. Con ello es posible detectar (a través de las pruebas de hipótesis correspondientes) las categorías responsables de la asociación. Al mismo tiempo podemos observar el sentido y la magnitud de la lejanía respecto a la independencia entre las variables. Por ejemplo cuando se hablaba de la relación entre la ocupación y el nivel de instrucción se pudo constatar que la asociación era producida por las frecuencias que se concentraban en la casilla correspondiente a las trabajadoras ocupadas en el sector de la producción, con nivel de instrucción de primaria o más.

Con este análisis vimos que es posible llegar a las conclusiones acerca de la interacción simultánea de las variables. Este análisis no es equivalente a realizar un análisis vibariado (con todas las posibles parejas de variables) cuyas conclusiones se trasladen a las del tipo multidimensional.

Respecto a la información de las trabajadoras manuales no calificadas observamos que las mujeres que se emplean en el sector productivo no cumplen con un patrón o conjunto de características sociodemográficas para cada sector. Si bien es cierto que las variables demográficas (edad, número de hijos, estado civil) juegan un papel importante en la selección de las mujeres en edad de trabajar, también es cierto que la asociación entre éstas variables y la ocupación se debe a la interacción que guardan las variables demográficas, y no propiamente a la relación que éstas puedan tener con la ocupación.

BIBLIOGRAFIA

Agresti, Alan. Analysis of Ordinal Categorical Data.John Wiley & Sons. University of Florida. 1984. 287 pp.

Bishop, Fienberg y Holland. Discrete Multivariate Analysis: Theory and Practice. Cambridge, Massachussetts. Institute of Technology. London. 1975.

Beaver R. y Mendenhall W. Inotroducción a la probabilidad y la estadística. Duxbury Press, Belmont California. 1971. 408 págs.

Blalock, Hubert M. Estadística Social. Fondo de Cultura Económica. México. 2a. ed. 1978.

BMDP. Biomedical Computer Program Series. Berkeley University of California Press. 1977. 880 pp.

Boudon, R., Lazarsfeld, P. Metodologia de las Ciencias Sociales, en la interpretación de las relaciones estadisticas como propiedad de investigación. España. 1974.

Bunge, Mario. La investigación científica. Ed. Ariel. México. 1984.

Cortés, Fernando. Tamaño de la Muestra y Análisis de Asociación. Revista Mexicana de Sociología. No. 4. México. 1982.

Cortés, F., Rubalcava, R.M. Análisis cuantitativo de variables cualitativas. El Colegio de México. México, 1987.

Everitt, B.S. The Analysis of Contingency Tables. London Chapman and Hall LTD; New York, A Halsted Press ook, John Wiley & Sons Inc. 1977. 128 pp.

Fienberg, S. The Analysis of Cross-Classified Categorical Data.1977.

García Ferrando M. Socioestadística. Centro de Investigaciones Sociológicas. Madrid. 1982. 461 pp.

Hogg y Craig. Estadística matemática. John Wiley & Sons.

Igartúa, R.M. Algunas caracteristicas de las trabajadoras manuales no calificadas de la cd. de México. Tesis. FLACSO. México. 1983.

Knoke, D. y Burke, P. Loglinear Models , en Quantitative Applications in the Social Sciences. Sage publications. Beverly Hills. 1980. 80 pp.

Mayorga quiroz, Román. El crecimiento desigual en Centroamérica 1950-2000. El Colegio de México. México. 1983.

Mood y Graybill. Introducción a la teoría de la estadística. Ed. Aguilar. Madrid. 1972.

Muñoz, Olivelra y Stern. Migración y desigualdad social en la ciudad de México. El Colegio de México, Instituto de Investigaciones Sociales, UNAM. México. 1977.

Parzen, Emanuel. Modern Probability Theory and Its Aplications. John Wiley & Sons. New York. 1963.

Popper, Karl. La lógica de la investigación científica. Ed. Tecnos. Madrid. 1a. ed. 1962.

Russell, Bertrand. La perspectiva Científica. Ariel. España. 2a. edición. 1949.

Stinchcombe, Arthur. La construcción de teorías sociales. Ed. Nueva Visión. Buenos Aires. 1970. 341 pp.

Taborga, Huascar. Como hacer una tesis. Editorial Grijalbo. México. 1a. ed. 1980. 220 pp.

Zambrano Jorge. La relación entre la fecundidad y el grado de escolaridad en el medio rural mexicano y la cd. de México. Tesis. El Colegio de México. Centro de Estudios Económicos y Demográficos. México. 1977. Tesis de maestría.

Zeisel, Hans. Dígalo con Números. Fondo de Cultura Económica. México. 2a. ed. 1980. 262 pp.

A N E X O A

```
TWO-WAY FREQUENCY TABLES -- MEASURES OF ASSOCIATION
MULTIWAY FREQUENCY TABLES -- LOGLINEAR MODELS (INCLUDING STRUCTURAL ZEROS
Copyright (c)1985 Regents of the University of California

BMDP Statistical Software, Inc.   Phone: (213)475-5700
1964 Westwood Blvd. Suite 202, Los Angeles, CA 90025

Program Revised:  October 1986  (PDP-11 Version 5.0C)
Manual Revised :  1983, 1985 Reprint

PDP-11 Version of BMDP4F has been converted by:
     Software Development Inc.    Phone: (206)827-8784
     P.O. Box 3732, Bellevue, WA 98009

Licensed Site  : (5-600) El Colegio de Mexico - Unidad de Computo



BMDP Program Instructions:
--------------------------

/PROBLEM
     TITLE='MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
            OCUPACION, EDAD, NUMERO DE HIJOS Y ESTADO CIVIL'.
/INPUT
     VARIABLES=4.
     FILE='SUBGPO12.DAT'.
     FORMAT='(4F4.0)'.
     TABLE=4,2,2,4.
/VARIABLES
     NAMES='OCUP','EDAD','NHNV','EDOCIV'.
/TABLE
     INDICES='OCUP','EDAD','NHNV','EDOCIV'.
     SYMBOLS='O','E','N','C'.
     DELTA=0.5.
/CATEGORY
     NAMES(1)='V AMBUL','SERV NO DOM','SERV DOM','PRODUCCION'.
     NAMES(2)='10-29','30 Y +'.
     NAMES(3)='0-2','3 Y +'.
     NAMES(4)='CASADA','DIVORCIADA','VIUDA','SOLTERA'.
/FIT
     MODEL=EN,EC,NC.
     ASSOCIATION=3.
     ALL.
/PRINT
     EXPECTED.STANDARDIZED.LAMBDA.MARGINAL=3.
/END
```

MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
OCUPACION, EDAD, NUMERO DE HIJOS Y ESTADO CIVIL

****
* TABLE PARAGRAPH 1 *
****

***** OBSERVED FREQUENCY TABLE 1

EDOCIV	NHNV	EDAD	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
CASADA	0-2	10-29	0	1	14	10	25
		30 Y +	1	3	10	7	21
		TOTAL	1	4	24	17	46
	3 Y +	10-29	1	1	9	10	21
		30 Y +	12	15	32	21	80
		TOTAL	13	16	41	31	101
DIVORCIA	0-2	10-29	1	2	19	8	30
		30 Y +	0	3	17	8	28
		TOTAL	1	5	36	16	58
	3 Y +	10-29	0	1	8	3	12
		30 Y +	6	7	26	7	46
		TOTAL	6	8	34	10	58
VIUDA	0-2	10-29	0	0	1	0	1
		30 Y +	1	4	16	5	26
		TOTAL	1	4	17	5	27
	3 Y +	10-29	0	0	1	1	2
		30 Y +	10	7	35	5	57
		TOTAL	10	7	36	6	59

SOLTERA	0-2	10-29	2	6	199	91	298
		30 Y +	0	2	33	13	48
		TOTAL	2	8	232	104	346
	3 Y +	10-29	0	0	1	1	2
		30 Y +	0	0	0	1	1
		TOTAL	0	0	1	2	3

TOTAL OF THE OBSERVED FREQUENCY TABLE IS 698.

***** MARGINAL SUBTABLE -- TABLE 1

OCUP

V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
34	52	421	191	698

***** MARGINAL SUBTABLE -- TABLE 1

EDAD

10-29	30 Y +	TOTAL
391	307	698

***** MARGINAL SUBTABLE -- TABLE 1

NHNV

0-2	3 Y +	TOTAL
477	221	698

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV

CASADA	DIVORCIA	VIUDA	SOLTERA	TOTAL
147	116	86	349	698

***** MARGINAL SUBTABLE -- TABLE 1

EDAD	OCUP				
	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
10-29	4	11	252	124	391
30 Y +	30	41	169	67	307
TOTAL	34	52	421	191	698

***** MARGINAL SUBTABLE -- TABLE 1

NHNV	OCUP				
	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
0-2	5	21	309	142	477
3 Y +	29	31	112	49	221
TOTAL	34	52	421	191	698

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV	OCUP				
	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
CASADA	14	20	65	48	147
DIVORCIA	7	13	70	26	116
VIUDA	11	11	53	11	86
SOLTERA	2	8	233	106	349
TOTAL	34	52	421	191	698

***** MARGINAL SUBTABLE -- TABLE 1

NHNV	EDAD		
	10-29	30 Y +	TOTAL
0-2	354	123	477
3 Y +	37	184	221
TOTAL	391	307	698

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV	EDAD 10-29	30 Y +	TOTAL
CASADA	46	101	147
DIVORCIA	42	74	116
VIUDA	3	83	86
SOLTERA	300	49	349
TOTAL	391	307	698

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV	NHNV 0-2	3 Y +	TOTAL
CASADA	46	101	147
DIVORCIA	58	58	116
VIUDA	27	59	86
SOLTERA	346	3	349
TOTAL	477	221	698

***** MARGINAL SUBTABLE -- TABLE 1

NHNV	EDAD	OCUP V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
0-2	10-29	3	9	233	109	354
	30 Y +	2	12	76	33	123
	TOTAL	5	21	309	142	477
3 Y +	10-29	1	2	19	15	37
	30 Y +	28	29	93	34	184
	TOTAL	29	31	112	49	221

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV	EDAD	OCUP				
		V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
CASADA	10-29	1	2	23	20	46
	30 Y +	13	18	42	28	101
	TOTAL	14	20	65	48	147
DIVORCIA	10-29	1	3	27	11	42
	30 Y +	6	10	43	15	74
	TOTAL	7	13	70	26	116
VIUDA	10-29	0	0	2	1	3
	30 Y +	11	11	51	10	83
	TOTAL	11	11	53	11	86
SOLTERA	10-29	2	6	200	92	300
	30 Y +	0	2	33	14	49
	TOTAL	2	8	233	106	349

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV	NHNV	OCUP				
		V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
CASADA	0-2	1	4	24	17	46
	3 Y +	13	16	41	31	101
	TOTAL	14	20	65	48	147
DIVORCIA	0-2	1	5	36	16	58
	3 Y +	6	8	34	10	58
	TOTAL	7	13	70	26	116
VIUDA	0-2	1	4	17	5	27
	3 Y +	10	7	36	6	59
	TOTAL	11	11	53	11	86
SOLTERA	0-2	2	8	232	104	346
	3 Y +	0	0	1	2	3
	TOTAL	2	8	233	106	349

***** MARGINAL SUBTABLE -- TABLE 1

EDOCIV	NHNV	EDAD 10-29	30 Y +	TOTAL
CASADA	0-2	25	21	46
	3 Y +	21	80	101
	TOTAL	46	101	147
DIVORCIA	0-2	30	28	58
	3 Y +	12	46	58
	TOTAL	42	74	116
VIUDA	0-2	1	26	27
	3 Y +	2	57	59
	TOTAL	3	83	86
SOLTERA	0-2	298	48	346
	3 Y +	2	1	3
	TOTAL	300	49	349

*****DELTA= 0.500 IS ADDED TO EACH CELL FOR ALL ANALYSES

MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
OCUPACION, EDAD, NUMERO DE HIJOS Y ESTADO CIVIL

***** THE RESULTS OF FITTING ALL K-FACTOR MARGINALS.
THIS IS A SIMULTANEOUS TEST THAT ALL K+1 AND
HIGHER FACTOR INTERACTIONS ARE ZERO.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ
0-MEAN	63	1626.91	0.00000	4157.67
1	55	806.74	0.00000	1097.33
2	33	14.82	0.99732	17.08
3	9	2.66	0.97616	2.82
4	0	0.	1.	0.

A SIMULTANEOUS TEST THAT ALL K-FACTOR INTERACTIONS ARE
SIMULTANOUSLY ZERO.
THE CHI-SQUARES ARE DIFFERENCES IN THE ABOVE TABLE.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ
1	8	820.17	0.00000	3060.34
2	22	791.92	0.00000	1080.25
3	24	12.15	0.97913	14.47
4	9	2.66	0.97616	2.62

***** ASSOCIATION OPTION SELECTED FOR ALL TERMS OF ORDER LESS THAN
OR EQUAL TO 3

		PARTIAL ASSOCIATION			MARGINAL ASSOCIATION	
EFFECT	D.F.	CHISQUARE	PROB	ITER	D.F.	CHISQUARE
O.	3	510.90	0.0000			
E.	1	9.69	0.0019			
N.	1	91.71	0.0000			
C.	3	207.88	0.0000			
OE. (*)	3	7.78	0.0507	8	3	56.05
ON.	3	10.10	0.0177	8	3	63.69
OC.	9	21.64	0.0101	8	9	81.31
EN.	1	25.17	0.0000	8	1	201.08
EC.	3	112.11	0.0000	6	3	294.10
NC.	3	178.04	0.0000	7	3	365.35
OEN.	3	1.48	0.6865	4	3	3.14
OEC.	9	2.43	0.9827	5	9	3.35
ONC.	9	5.79	0.7609	6	9	6.83
ENC.	3	1.42	0.7001	7	3	1.32

MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
OCUPACION, EDAD, NUMERO DE HIJOS Y ESTADO CIVIL

```
****************
*   MODEL  1   *
****************
```

MODEL	D.F.	LIKELIHOOD-RATIO CHI-SQUARE	PROB	PEARSON CHI-SQU/
EN,EC,NC.	51	628.96	0.0000	597.19

***** EXPECTED VALUES USING ABOVE MODEL

EDOCIV	NHNV	EDAD	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
CASADA	0-2	10-29	6.5	6.5	6.5	6.5	26.0
		30 Y +	6.0	6.0	6.0	6.0	24.0
		TOTAL	12.5	12.5	12.5	12.5	50.0
	3 Y +	10-29	6.0	6.0	6.0	6.0	24.0
		30 Y +	20.3	20.3	20.3	20.3	81.0
		TOTAL	26.3	26.3	26.3	26.3	105.0
DIVORCIA	0-2	10-29	8.0	8.0	8.0	8.0	32.0
		30 Y +	7.5	7.5	7.5	7.5	30.0
		TOTAL	15.5	15.5	15.5	15.5	62.0
	3 Y +	10-29	3.5	3.5	3.5	3.5	14.0
		30 Y +	12.0	12.0	12.0	12.0	48.0
		TOTAL	15.5	15.5	15.5	15.5	62.0
VIUDA	0-2	10-29	1.1	1.1	1.1	1.1	4.3
		30 Y +	6.7	6.7	6.7	6.7	26.7
		TOTAL	7.7	7.7	7.7	7.7	31.0
	3 Y +	10-29	0.7	0.7	0.7	0.7	2.7
		30 Y +	15.1	15.1	15.1	15.1	60.3

		TOTAL	15.8	15.8	15.8	15.8	63.0
SOLTERA	0-2	10-29	74.9	74.9	74.9	74.9	299.7
		30 Y +	12.6	12.6	12.6	12.6	50.3
		TOTAL	87.5	87.5	87.5	87.5	350.0
	3 Y +	10-29	1.1	1.1	1.1	1.1	4.3
		30 Y +	0.7	0.7	0.7	0.7	2.7
		TOTAL	1.8	1.8	1.8	1.8	7.0

STANDARDIZED DEVIATES = (OBS - EXP)/SQRT(EXP) FOR ABOVE MODEL

EDOCIV	NHNV	EDAD	OCUP			
			V AMBUL	SERV NO	SERV DOM	PRODUCCI
CASADA	0-2	10-29	-2.4	-2.0	3.1	1.6
		30 Y +	-1.8	-1.0	1.8	0.8
	3 Y +	10-29	-1.8	-1.8	1.4	1.8
		30 Y +	-1.7	-1.1	2.7	0.3
DIVORCIA	0-2	10-29	-2.3	-1.9	4.1	0.2
		30 Y +	-2.6	-1.5	3.7	0.4
	3 Y +	10-29	-1.6	-1.1	2.7	0.0
		30 Y +	-1.6	-1.3	4.2	-1.3
VIUDA	0-2	10-29	-0.6	-0.6	0.4	-0.6
		30 Y +	-2.0	-0.8	3.8	-0.5
	3 Y +	10-29	-0.2	-0.2	1.0	1.0
		30 Y +	-1.2	-2.0	5.3	-2.5
SOLTERA	0-2	10-29	-8.4	-7.9	14.4	1.9
		30 Y +	-3.4	-2.8	5.9	0.3

3 Y +	10-29	-0.6	-0.6	0.4	0.4
	30 Y +	-0.2	-0.2	-0.2	1.0

ASYMPTOTIC STANDARD ERRORS OF THE PARAMETER ESTIMATES ARE COMPUTED BY INVERTING THE INFORMATION MATRIX.

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

THETA(MEAN) 1.6941

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDAD	
10-29	30 Y +
-0.310	0.310

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

EDAD	
10-29	30 Y +
-4.786	4.786

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NHNV	
0-2	3 Y +
0.379	-0.379

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

NHNV	
0-2	3 Y +
5.709	-5.709

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NHNV	EDAD 10-29	30 Y +
0-2	0.324	-0.324
3 Y +	-0.324	0.324

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

NHNV	EDAD 10-29	30 Y +
0-2	5.337	-5.337
3 Y +	-5.337	5.337

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDOCIV

CASADA	DIVORCIA	VIUDA	SOLTERA
0.422	0.264	-0.622	-0.063

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

EDOCIV

CASADA	DIVORCIA	VIUDA	SOLTERA
4.265	2.604	-3.923	-0.400

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDOCIV	EDAD 10-29	30 Y +
CASADA	0.026	-0.026
DIVORCIA	0.019	-0.019
VIUDA	-0.923	0.923
SOLTERA	0.878	-0.878

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

EDOCIV	EDAD	
	10-29	30 Y +
CASADA	0.282	-0.282
DIVORCIA	0.199	-0.199
VIUDA	-5.978	5.978
SOLTERA	9.640	-9.640

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDOCIV	NHNV	
	0-2	3 Y +
CASADA	-0.663	0.663
DIVORCIA	-0.290	0.290
VIUDA	-0.463	0.463
SOLTERA	1.415	-1.415

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

EDOCIV	NHNV	
	0-2	3 Y +
CASADA	-7.326	7.326
DIVORCIA	-3.128	3.128
VIUDA	-4.311	4.311
SOLTERA	9.225	-9.225

Approximate array space used: 2420(23%)
Elapsed Time: 1.99 minutes.

TWO-WAY FREQUENCY TABLES -- MEASURES OF ASSOCIATION
MULTIWAY FREQUENCY TABLES -- LOGLINEAR MODELS (INCLUDING STRUCTURAL ZERO!

Licensed Site : (5-600) El Colegio de Mexico - Unidad de Computo

BMDP Program Instructions:

/FINISH

ANEXO B

```
/PROGRAM
    TITLE='MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
           OCUPACION, EDAD, NUMERO DE HIJOS, ESTADO CIVIL Y
           NIVEL DE INSTRUCCION'.
/INPUT
    VARIABLES=5.
    FILE='SUBGPO14.DAT'.
    FORMAT='(4F4.0)'.
    TABLE=4,2,2,2,2.
 /VARIABLES
    NAMES='OCUP','EDAD','NHNV','EDOCIV','NIVINSTR'.
 /TABLE
    INDICES='OCUP','EDAD','NHNV','EDOCIV','NIVINSTR'.
    SYMBOLS='O','E','N','C','I'.
    ASSOCIATION.
    DELTA=0.5.
    ITERATION=100.
 /CATEGORY
    NAMES(1)='V AMBUL','SERV NO DOM','SERV DOM','PRODUCCION'.
    NAMES(2)='15-22','23 Y +'.
    NAMES(3)='0','1 O +'.
    NAMES(4)='ALG VEZ UNIDA','SOLTERA'.
    NAMES(5)='<PRIM','PRIM Y +'.
 /FIT
    MODEL=OI,EN,EC,EI,NC,CI.
    ASSOCIATION=5.
 /PRINT
    EXPECTED.STANDARDIZED.LAMBDA.
 /END
```

```
************************
* TABLE PARAGRAPH    1  *
************************
```

OBSERVED FREQUENCY TABLE 1

NIVINSTR	EDOCIV	NHNV	EDAD	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TO'
<PRIM	ALG VEZ	O	15-22	0	0	1	2	3
			23 Y +	0	3	14	4	21
			TOTAL	0	3	15	6	24
		1 O +	15-22	0	1	13	7	21
			23 Y +	27	33	148	40	248
			TOTAL	27	34	161	47	269
	SOLTERA	O	15-22	0	2	120	12	134
			23 Y +	0	1	51	13	65
			TOTAL	0	3	171	25	199
		1 O +	15-22	0	0	1	0	1
			23 Y +	0	0	6	1	7
			TOTAL	0	0	7	1	8
PRIM Y +	ALG VEZ	O	15-22	0	0	0	0	0
			23 Y +	0	0	2	2	4
			TOTAL	0	0	2	2	4
		1 O +	15-22	0	0	1	5	6
			23 Y +	5	7	9	25	46
			TOTAL	5	7	10	30	52

	23 Y +	1	2	11	19	33
	TOTAL	2	6	54	76	137
1 O +	15-22	0	0	0	0	0
	23 Y +	0	0	1	3	4
	TOTAL	0	0	1	3	4

TOTAL OF THE OBSERVED FREQUENCY TABLE IS 697.

DELTA= 0.500 IS ADDED TO EACH CELL FOR ALL ANALYSES

MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
OCUPACION, EDAD, NUMERO DE HIJOS, ESTADO CIVIL Y NIVEL DE INSTRUCCION

THE RESULTS OF FITTING ALL K-FACTOR MARGINALS.
THIS IS A SIMULTANEOUS TEST THAT ALL K+1 AND HIGHER FACTOR INTERACTIONS ARE ZERO.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ
0-MEAN	63	1850.34	0.00000	3683.05
1	56	1173.50	0.00000	1671.55
** THE LOGLINEAR ALGORITHM DID NOT CONVERGE WHEN FITTING THE MODEL BELOW				
2	38	26.81	0.91725	30.11
3	16	10.41	0.84442	11.92
4	3	1.10	0.77786	1.14
5	0	0.	1.	0.

A SIMULTANEOUS TEST THAT ALL K-FACTOR INTERACTIONS ARE SIMULTANOUSLY ZERO.
THE CHI-SQUARES ARE DIFFERENCES IN THE ABOVE TABLE.

K-FACTOR	D.F.	LR CHISQ	PROB.	PEARSON CHISQ
1	7	676.85	0.00000	2011.50
2	18	1146.88	0.00000	1641.44
3	22	16.21	0.80557	18.19
4	13	9.31	0.74899	10.79
5	3	1.10	0.77786	1.14

ASSOCIATION OPTION SELECTED FOR ALL TERMS OF ORDER LESS THAN OR EQUAL TO 5

EFFECT	D.F.	PARTIAL ASSOCIATION CHISQUARE	PROB	ITER
O.	3	510.73	0.0000	
E.	1	34.96	0.0000	
N.	1	1.32	0.2508	

C.	1	0.00	0.9670	
I.	1	129.84	0.0000	
OE.	3	3.61	0.3068	20
ON.	3	3.57	0.3115	20
OC.	3	7.01	0.0715	20
OI. (*)	3	121.50	0.0000	20
EN.	1	15.20	0.0001	20
EC.	1	30.52	0.0000	20
EI.	1	4.59	0.0322	20
NC.	1	343.54	0.0000	14
NI.	1	0.22	0.6412	13
CI.	1	9.55	0.0020	12
OEN.	3	0.13	0.9880	9
OEC.	3	3.48	0.3231	9
OEI.	3	0.46	0.9279	9
ONC.	3	3.21	0.3608	10
ONI.	3	1.91	0.5904	11
OCI.	3	1.20	0.7519	11
ENC.	1	2.35	0.1249	10
ENI.	1	0.00	1.0000	10
ECI.	1	0.17	0.6829	9
NCI.	1	0.11	0.7378	18
OENC.	3	5.44	0.1422	6
OENI.	3	0.23	0.9734	5
OECI.	3	1.06	0.7868	4
ONCI.	3	0.67	0.8808	11
ENCI.	1	0.34	0.5625	9
OENCI.	3	1.10	0.7779	

MODELO LOGLINEAL PARA EL ANALISIS DE LAS VARIABLES:
OCUPACION, EDAD, NUMERO DE HIJOS, ESTADO CIVIL Y NIVEL DE INSTRUCCION

****
* MODEL 1 *
****

MODEL	D.F.	LIKELIHOOD-RATIO CHI-SQUARE	PROB	PEARSON CHI-SQUARE
OI,EN,EC,EI,NC,CI,	48	100.74	0.0000	92.62

EXPECTED VALUES USING ABOVE MODEL

NIVINSTR	EDOCIV	NHNV	EDAD	V AMBUL	SERV NO	SERV DOM	PRODUCCI	TOTAL
<PRIM	ALG VEZ	0	15-22	0.4	0.6	5.0	1.2	7.2
			23 Y +	1.3	1.9	15.3	3.6	22.1
			TOTAL	1.8	2.5	20.3	4.7	29.3
		1 O +	15-22	1.3	1.9	15.1	3.5	21.8
			23 Y +	15.0	21.3	173.4	40.2	249.9
			TOTAL	16.3	23.2	188.5	43.7	271.7
	SOLTERA	0	15-22	8.0	11.4	92.8	21.5	133.7
			23 Y +	4.1	5.9	47.8	11.1	68.9
			TOTAL	12.2	17.3	140.8	32.6	202.6
		1 O +	15-22	0.3	0.4	2.9	0.7	4.2
			23 Y +	0.5	0.7	5.7	1.3	8.2
			TOTAL	0.7	1.1	8.6	2.0	12.4
PRIM Y +	ALG VEZ	0	15-22	0.1	0.2	0.7	1.2	2.2
			23 Y +	0.2	0.3	1.5	2.4	4.5
			TOTAL	0.3	0.5	2.2	3.6	6.7
		1 O +	15-22	0.3	0.5	2.2	3.6	6.7

		23 Y +	2.6	3.8	16.9	27.3	50.6
		TOTAL	3.0	4.3	19.1	30.9	57.3
SOLTERA	O	15-22	5.5	7.9	35.2	57.1	105.7
		23 Y +	1.8	2.7	11.9	19.3	35.7
		TOTAL	7.3	10.6	47.1	76.3	141.4
	1 O +	15-22	0.2	0.3	1.1	1.8	3.3
		23 Y +	0.2	0.3	1.4	2.3	4.2
		TOTAL	0.4	0.6	2.5	4.1	7.6

STANDARDIZED DEVIATES = (OBS - EXP)/SQRT(EXP) FOR ABOVE MODEL

NIVINSTR	EDOCIV	NHNV	EDAD	OCUP			
				V AMBUL	SERV NO	SERV DOM	PRODUCCI
<PRIM	ALG VEZ	O	15-22	0.1	-0.1	-1.6	1.2
			23 Y +	-0.7	1.2	-0.2	0.5
		1 O +	15-22	-0.7	-0.3	-0.4	2.1
			23 Y +	3.2	2.6	-1.9	0.0
	SOLTERA	O	15-22	-2.7	-2.6	2.9	-1.9
			23 Y +	-1.8	-1.8	0.5	0.7
		1 O +	15-22	0.5	0.2	-0.8	-0.2
			23 Y +	0.0	-0.2	0.3	0.2
PRIM Y +	ALG VEZ	O	15-22	1.1	0.8	-0.3	-0.6
			23 Y +	0.6	0.3	0.8	0.1
		1 O +	15-22	0.3	-0.0	-0.5	1.0
			23 Y +	1.8	1.9	-1.8	-0.3
	SOLTERA	O	15-22	-1.7	-1.6	1.4	0.1
			23 Y +	-0.3	-0.1	-0.1	0.0

1 0 +	15-22	0.8	0.5	-0.6	-1.0
	23 Y +	0.6	0.3	0.1	0.8

ASYMPTOTIC STANDARD ERRORS OF THE PARAMETER ESTIMATES ARE COMPUTED BY INVERTING THE INFORMATION MATRIX.

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

THETA(MEAN) 0.9936

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

OCUP

V AMBUL	SERV NO	SERV DOM	PRODUCCI
-1.046	-0.684	1.110	0.620

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

OCUP

V AMBUL	SERV NO	SERV DOM	PRODUCCI
-7.541	-5.677	14.381	7.727

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDAD

15-22	23 Y +
-0.338	0.338

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

EDAD

15-22	23 Y +
-5.920	5.920

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NHNV	
0	1 0 +
0.257	-0.257

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

NHNV	
0	1 0 +
3.242	-3.242

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NHNV	EDAD	
	15-22	23 Y +
0	0.331	-0.331
1 0 +	-0.331	0.331

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

NHNV	EDAD	
	15-22	23 Y +
0	4.167	-4.167
1 0 +	-4.167	4.167

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDOCIV	
ALG VEZ	SOLTERA
-0.110	0.110

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD ERROR

EDOCIV

ALG VEZ	SOLTERA
-1.311	1.311

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDOCIV	EDAD 15-22	23 Y +
ALG VEZ	-0.445	0.446
SOLTERA	0.445	-0.445

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

EDOCIV	EDAD 15-22	23 Y +
ALG VEZ	-5.694	5.694
SOLTERA	5.694	-5.694

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

EDOCIV	NHNV 0	1 O +
ALG VEZ	-1.139	1.139
SOLTERA	1.139	-1.139

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

EDOCIV	NHNV 0	1 O +
ALG VEZ	-15.279	15.279
SOLTERA	15.279	-15.279

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NIVINSTR	
<PRIM	PRIM Y +
0.433	-0.433

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

NIVINSTR	
<PRIM	PRIM Y +
6.712	-6.712

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NIVINSTR	OCUP			
	V AMBUL	SERV NO	SERV DOM	PRODUCCI
<PRIM	0.101	0.088	0.391	-0.580
PRIM Y +	-0.101	-0.088	-0.391	0.580

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

NIVINSTR	OCUP			
	V AMBUL	SERV NO	SERV DOM	PRODUCCI
<PRIM	0.725	0.734	5.074	-7.235
PRIM Y +	-0.725	-0.734	-5.074	7.235

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NIVINSTR	EDAD	
	15-22	23 Y +
<PRIM	-0.105	0.105
PRIM Y +	0.105	-0.105

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

NIVINSTR	EDAD	
	15-22	23 Y +
<PRIM	-2.079	2.079
PRIM Y +	2.079	-2.079

ESTIMATES OF THE LOG-LINEAR PARAMETERS (LAMBDA) IN THE MODEL ABOVE

NIVINSTR	EDOCIV	
	ALG VEZ	SOLTERA
<PRIM	0.235	-0.235
PRIM Y +	-0.235	0.235

RATIO OF THE LOG-LINEAR PARAMETER ESTIMATES TO ITS STANDARD EI

NIVINSTR	EDOCIV	
	ALG VEZ	SOLTERA
<PRIM	4.499	-4.499
PRIM Y +	-4.499	4.499

Approximate array space used: 3008(29%)
Elapsed Time: 4.55 minutes.

TWO-WAY FREQUENCY TABLES -- MEASURES OF ASSOCIATION
MULTIWAY FREQUENCY TABLES -- LOGLINEAR MODELS (INCLUDING STRUCTURAL ZERO!

BMDP Program Instructions:

/FINISH

No more instructions to process...

------- Program Terminated -------