



UNIVERSIDAD NACIONAL
AUTÓNOMA

Unidad Académica de los Ciclos Profesionales
y de Postgrado del CCH.

Instituto de Investigación en Matemáticas Aplicadas
y en Sistemas.

ANÁLISIS Y MODELOS PARA DATOS POLITÓMICOS Y
CATEGÓRICOS MULTIVARIADOS

T E S I S
que para obtener el grado de
MAESTRA EN ESTADÍSTICA E
INVESTIGACIÓN DE OPERACIONES
presenta la Actuaría
SILVIA RUIZ VELASCO ACOSTA

BIBLIOTECA
JUAN A. ESCALANTE H.
UNIDAD ACADÉMICA DE
LOS CICLOS PROFESIONAL

México, D.F.

Y DE POSGRADO / CCH

Diciembre 1984

UNAM



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Table of Contents

1. INTRODUCCION	1
1.1 Analisis Probit	2
1.2 Modelo Logistico Lineal	4
2. MODELOS PARA DATOS POLITOMICOS Y CATEGORICOS MULTIVARIADOS.	11
2.1 Modelos de Categoria de Respuesta Ordenada.	11
2.2 Modelos de respuesta categorizada no ordenada.	25
2.3 Diferencias entre los modelos de respuesta ordenada y no ordenada.	38
2.4 Modelos para el caso Multivariado	38
2.4.1 Modelos provenientes de distribuciones marginales	39
2.4.2 Modelos de distribucion condicional	45
3. ESTIMACION	48
3.1 Maxima Verosimilitud	52
3.1.1 El Metodo de Newton-Raphson.	53
3.1.2 El Metodo de Derivadas Fijas (aproximación de Newton)	54
3.1.3 El Metodo de Falsas Posiciones.	55
3.1.4 El Metodo de Casi-verosimilitud.	58
3.1.5 Escalamiento Iterativo Generalizado	63
3.2 Estimadores de Minima Ji-cuadrada	66
3.3 Otros Metodos de estimacion	67
3.4 Analisis de Residuales y Bondad de Ajuste	69
4. APLICACIONES	71
4.1 Aplicacion en Modelos de Categoria de Respuesta Ordenada	71
4.1.1 Modelos de Mc. Cullagh	75
4.1.2 Probit Politomico.	78
4.1.3 Primer Ejemplo	80
4.2 Ajuste en el caso de categorias no ordenadas	85
4.3 Ajuste en el caso Multivariado.	92
5. Conclusiones y Generalizaciones	96
6. BIBLIOGRAFIA	101
7. APENDICE	104
7.1 Demostracion de identificabilidad de los parametros en el modelo de Aitchson y Benett	104
7.2 logit independiente por maximización de utilidad	105
7.3 Demostracion de la unicidad de la funcion discriminadora (escalamiento iterativo)	106

1. INTRODUCCION

Los modelos de respuesta politómica y categórica multivariada provienen esencialmente de generalizaciones de modelos para respuesta dicotómica, también llamada binaria ó quantal.

En los modelos de respuesta dicotómica, si para el i -ésimo individuo representamos la observación o respuesta por la variable aleatoria y_i , podemos considerar sin pérdida de generalidad los dos posibles valores de y_i como 1,0 y escribir

$$E(y_i) = P(y_i = 1) = p_i$$

$$P(y_i = 0) = 1 - p_i$$

Es común llamar al evento $y_i = 1$ éxito y al evento $y_i = 0$ fracaso, ó equivalentemente considerar que el evento $y_i = 1$, corresponde a que el individuo presente una respuesta predeterminada y el evento $y_i = 0$ corresponde a que el individuo no la presente.

Suponemos que tales observaciones son realizadas en n individuos comunmente supuestos independientes. El problema es evaluar cualquier dependencia de p_i en variables explicativas las cuales representan, por ejemplo, grupos de individuos o variables cuantitativas asociadas.

Entre este tipo de modelos se encuentran el de análisis probit y el modelo logístico.

1.1 Analisis Probit

El modelo de análisis probit tiene su origen en el bioensayo o ensayo biológico, en este análisis originalmente suponemos un estímulo en k diferentes dosis y cada dosis se aplica a n_i individuos ($i = 1, \dots, k$). Al transcurrir cierto período de tiempo determinado de antemano, se observa en cada individuo si se presentó o no una respuesta predeterminada, usualmente la respuesta es la muerte.

Suponemos que cada individuo tiene una distribución probabilística de tolerancia a la dosis o a una función de ella, es decir existe una variable subyacente, en este caso la tolerancia, con una distribución continua que determina la respuesta binaria. Esta suposición esencialmente lo que nos permite es interpretabilidad, suponemos que un individuo va a tolerar la dosis hasta un cierto umbral, es decir no va a presentar respuesta, y al pasar ese umbral va a morir, es decir la respuesta se va a presentar.

El modelo relaciona entonces la proporción esperada p que responde con la dosis x (medida en una escala apropiada), si además suponemos que la distribución de tolerancia a la dosis es normal tenemos

$$\int_{-\infty}^x \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt$$

Gaddum(1933) propuso estandarizar la variable bajo la integral (Normit) y posteriormente Bliss(1935) propuso añadir 5 unidades a la variable estandarizada con el único propósito de evitar números negativos (Probit). El modelo que tenemos es entonces

$$p_i = (\alpha + \beta x_i)$$

Este modelo se puede emplear, también en problemas que se estudian en enseñanza, demografía, etc.. En enseñanza, por ejemplo, en las pruebas de inteligencia en donde se supone que la inteligencia es una variable continua subyacente y lo que se mide es la respuesta a cierto estímulo (usualmente llamados items).

Estos modelos también pueden emplearse en economía y en este caso no se supone la existencia de una variable subyacente, sino que se utiliza el enfoque de maximización de utilidad, es decir un individuo va a presentar respuesta si la utilidad que le representa es mayor a la de no presentarla, (ver Marshak, 1960, Quandt, 1968). Un ejemplo de este enfoque sería el que presentan Domencich y Mc. Fadden (1975), en este ejemplo consideran la decisión de una persona entre manejar al ir al trabajo o no; si la utilidad que representa manejar desde el punto de vista costo, tiempo, etc., es mayor a la de no hacerlo (emplear otro medio de transporte) entonces se obtendrá respuesta, este enfoque será tratado más adelante.

Fisher(1935) resolvió el problema que representan, desde el punto de vista de estimación de parámetros, los casos extremos, es decir que todos los individuos presenten la respuesta o que

ninguno la presente.

Finney(1971)propuso un método de estimación para los parámetros, α, β , presentando además una discusión más amplia sobre el tema.

1.2 Modelo Logístico Lineal

El modelo logístico lineal es en muchos aspectos la manera más sencilla de representar la dependencia de la probabilidad de éxito en variables explicativas (de tal forma que se satisfaga la condición $0 < p < 1$). El modelo es el siguiente

$$p_i = \frac{e^{a_i \beta}}{1 + e^{a_i \beta}}$$

$$1-p_i = \frac{1}{1 + e^{a_i \beta}}$$

donde a_i es un vector renglón de constantes conocidas y β es una columna de parámetros desconocidos, Cox(1970).

Equivalentemente este modelo se puede escribir como:

$$= \log \left(\frac{p_i}{1-p_i} \right) = a_i \beta = \sum_s a_{is} \beta_s$$

o

$$\lambda = a \beta$$

λ_i es la llamada transformación logística de la probabilidad p_i y

$\lambda = \alpha$ es llamado el modelo logístico lineal, también llamado log odds (log momios).

Posibles aplicaciones apropiadas de este modelo son : Si suponemos dos grupos de individuos etiquetados 1 y 2 ,de tamaño n_1 y n_2 respectivamente y de cada individuo se obtiene respuesta binaria, suponemos además que cada individuo responde independientemente con probabilidad de éxito dependiendo solo del grupo e igual a $p^{(1)}$ y $p^{(2)}$ respectivamente, entonces el modelo logístico lineal correspondiente es:

$$\lambda^{(1)} = \alpha, \quad \lambda^{(2)} = \alpha + \Delta$$

en este caso Δ representa la diferencia entre los grupos en escala logística.

$$\begin{aligned} \Delta &= \lambda^{(2)} - \lambda^{(1)} \\ &= \log \left(\frac{p^{(2)} (1-p^{(1)})}{(1-p^{(2)}) p^{(1)}} \right) \end{aligned}$$

e es el cociente de los momios de éxito contra fracaso en los 2 grupos, esto es una reparametrización del modelo original

$$\log \left(\frac{p_i}{1-p_i} \right) = \alpha + \beta x_i$$

con $x_1 = 0$ y $x_2 = 1$.

Es conveniente notar que este modelo esta saturado en parámetros, esto es el número de probabilidades binomiales

independientes es igual al número de parámetros.

Si suponemos que para comparar dos tratamientos tenemos k conjuntos de observaciones similares al anterior, de tal forma que diferentes conjuntos pueden corresponder a los niveles de un factor o a diferentes bloques de un diseño experimental. Entonces posibles modelos logísticos lineales son:

$$\begin{array}{ll} (1) & (2) \\ \lambda_j & = \alpha_j & \lambda_j & = \alpha_j + \Delta_j \end{array}$$

δ

$$\begin{array}{ll} (1) & (2) \\ \lambda_j & = \alpha_j & \lambda_j & = \alpha_j + \Delta_j \end{array}$$

δ

$$\begin{array}{ll} (1) & (2) \\ \lambda_j & = \alpha_j & \lambda_j & = \alpha_j + \Delta_j \end{array}$$

Si $P_j^{(1)}$ y $P_j^{(2)}$ representan la probabilidad de éxito en los dos grupos para la j -ésima tabla entonces, en el primer caso se tiene una reparametrización del modelo en la cual las probabilidades $P_j^{(1)}$ y $P_j^{(2)}$ son arbitrarias, en el segundo caso lo que se tiene son diferencias arbitrarias entre los conjuntos de datos pero una diferencia logística constante entre grupos y en el último los conjuntos son homogéneos. De estas reparametrizaciones la más utilizada es la segunda, ya que el parámetro de interés es Δ , mientras que los parámetros de estorboso (nuisance) $\alpha_1, \dots, \alpha_k$ toman en cuenta las diferencias entre los conjuntos, esta reparametrización es el análogo directo de un modelo de bloques al azar, el parámetro Δ representa el efecto de tratamiento y

$\alpha_1, \dots, \alpha_k$ los efectos de bloque.
1 k

Otro ejemplo del uso de este modelo involucra la dependencia de la probabilidad de éxito en una variable explicativa x .

$$\lambda_i = \alpha + \beta x_i$$

$$p_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}$$

Este tipo de modelos es utilizado por ejemplo en el caso estímulo-respuesta Berkson (1944, 1951, 1953, 1955), o en el caso en que tenemos g grupos ordenados y es razonable esperar que cualquier cambio en la probabilidad de éxito es monótono con el orden del grupo. Si además pueden ser asignados a los g grupos puntajes x_1, \dots, x_g de tal forma que una relación suave entre la probabilidad de éxito y el valor de x sea razonable.

En el caso en que existan más de dos categorías de respuesta es decir la variable respuesta sea politómica (varias categorías), es necesario tomar en cuenta el orden, a diferencia del caso dicotómico en el que el orden no afecta.

En el caso de categorías ordenadas podemos suponer que existe una variable subyacente, y que las categorías van a estar determinadas por diferentes umbrales cotas o fronteras, de tal forma que si el individuo sobrepasa el i -ésimo umbral, va a pertenecer a la categoría $i+1$, es decir un individuo pertenece a la categoría j si el valor de la variable subyacente es mayor que

el umbral de la categoría $j-1, j-2, \dots, 1$.

Este tipo de modelos tienen aplicación en ensayo biológico, por ejemplo, al aplicar una cierta dosis (estímulo) las diferentes respuestas que podemos observar pueden ser vivo moribundo y muerto. Suponemos que cada individuo tiene una tolerancia a la dosis (o a una función de ella) y que es por lo tanto necesaria más dosis para que un individuo muera que para que quede moribundo, además un individuo para estar muerto antes tuvo que estar vivo y moribundo.

También es posible utilizar estos modelos en el caso de que no exista esa variable subyacente si sabemos que los datos tienen un comportamiento, que es posible modelar con un modelo de este estilo.

La suposición de una variable subyacente, como ya se había mencionado, es más que nada para facilitar la interpretación, por lo que es posible ajustar estos modelos sin esta suposición, ya que puede modelar adecuadamente la realidad.

Estos modelos al igual que en caso dicotómico tienen diversas aplicaciones, por ejemplo en el caso de sociología en un estudio de opinión, en donde la categoría de la variable respuesta podría ser: muy de acuerdo, de acuerdo, indiferente, en desacuerdo, muy en desacuerdo. En el campo psicológico estos modelos se podrían utilizar en estudios de caracterizaciones de personalidad, por ejemplo.

Al igual que en el caso dicotómico, estos modelos tiene

aplicación importante en el campo económico, son utilizados en un por David y Lega(1975), para tratar de explicar el precio de una casa por el tamaño de la casa, la edad del jefe de familia, el ingreso del jefe de familia y el número de años de educación del jefe de familia. El precio de la casa es observado solamente como perteneciente a una de tres categorías :

$$y = \begin{cases} 0 < \$28,999 \\ 1 & \$29,000 - \$54,999 \\ 2 & > \$55,000 \end{cases}$$

en este caso no suponemos la variable subyacente y los umbrales son conocidos de antemano.

En el caso de modelos de categoría de respuesta no ordenada, a diferencia de lo anterior, las aplicaciones son mas comunes en el campo económico o sociológico que en el campo biológico.

Una aplicación de estos modelos es el comparar tendencias en poblaciones definidas por ciertas características, ya sean demográficas o de otro estilo.

El enfoque de maximización de utilidad se generaliza a este tipo de modelos, ya que de k categorías posibles de respuesta, el individuo va a elegir aquella que le proporcione mayor utilidad. Como ejemplo puede discutirse una extensión del presentado en caso binario, en este caso, el individuo va elegir entre manejar, ir en autobús o ir en bicicleta, en base a ciertas variables como costo, tiempo, salud, etc..

Ahora bien el caso de modelos para datos multivariados, es decir cuando existen dos o mas respuestas categoricas, también los modelos provienen de generalizaciones del caso binario. La primera generalización sería el considerar dos variables respuesta binarias .

Podemos suponer la existencia de variables subyacentes, suponiendo que un mismo estímulo afecta a varios sistemas y que se puede obtener la respuesta de cada uno de ellos por separado, además cada variable subyacente va a tener una distribución de tolerancia a la(s) dosis.

Un ejemplo muy conocido lo presentan Ashford y Sowden(1970) en el cual los datos provienen de un conjunto de trabajadores de minas de carbón, agrupados en 9 grupos de edad, la cual es considerada como el estímulo y se obtienen como respuesta la presencia de dos sintomas respiratorios: silbilancia y falta de respiración. Es claro que la suposición de variables subyacentes auxilia para la interpretabilidad y es posible que el modelo represente la realidad sin la necesidad de esta suposición.

Al igual que en caso univariado este tipo de modelos también tienen aplicación en problemas sociológicos, psicológicos y/o económicos.

2. MODELOS PARA DATOS POLITOMICOS Y CATEGORICOS MULTIVARIADOS.

Tratando primeramente el caso de respuesta politómica. Los modelos pueden ser:

1. de categorías ordenadas de respuesta.
2. de categorías no ordenadas de respuesta.

2.1 Modelos de Categoría de Respuesta Ordenada.

Dentro de estos modelos se puede hablar de los que resultan una generalización de modelos establecidos para el caso de variables respuesta dicotómica. Estas son generalizaciones de los modelos probit y logit principalmente.

En el caso de Probit la primera generalización fue propuesta por Aitchinson y Silvey (1959). Esta generalización nació como la solución a un problema particular, el cual consistía en estimar el tiempo medio que un insecto (*Petrobios Leach*) pasa en cada estado de los posibles que puede pasar en su vida.

Las suposiciones en el modelo propuesto son :

1. en la vida del insecto hay k estados, y
2. un insecto observado está necesariamente en uno de esos estados, además el estado k es alcanzado por todos los insectos, por lo que no es necesario estimar parámetros desconocidos en dicho estado.

Además suponen que las observaciones son hechas en m tiempos diferentes ($x_i, i=1, \dots, m$).

El tiempo utilizado por un insecto en cada estado i puede ser

visto como una variable aleatoria y considerando que el tiempo que pasa cada insecto en cada estado es independiente, el problema es estimar la media de esa variable, esto es equivalente a estimar para el estado j la suma de las medias. Si suponemos que la variable aleatoria tiene una distribución normal, la suma de ellas también va a tener una distribución normal, en donde la varianza de esa distribución la podemos considerar constante para todos los estados o como una función de la media de los estados (esencialmente la media por una constante) o en el caso más complejo diferente para todos los estados.

La generalización de este problema al análisis probit con varias categorías de respuesta sería el considerar que los sujetos (insectos) pueden corresponder a más de dos categorías (en este ejemplo estados) como por ejemplo vivo, moribundo, muerto, a diferentes dosis de algun(os) estímulo(s), tiempo en este caso.

Las condiciones que debe satisfacer un experimento para ser considerado dentro de este método de análisis son:

1. Las categorías deben ser ordenadas, mutuamente exclusivas y exhaustivas.
2. las reacciones de los sujetos a dosis crecientes deben ser sistemáticas, en el sentido de que si la dosis x coloca al individuo en la clase i , entonces una dosis mayor es requerida para colocar al individuo en la clase j donde $j > i$.

Ashford(1959) presenta una generalización del análisis probit considerando que un individuo puede ser clasificado en más de dos categorías de acuerdo a una medición continua y la categoría va a ser establecida si la medición sobrepasa ciertos valores (cotas, umbrales o fronteras). Este es el tipo de respuesta conocido como semi-quantal.

Esencialmente el método es el mismo que el descrito por Aitchinson y Silvey(1957) con la consideración de una reacción subyacente continua.

La escala de respuesta es subdividida en intervalos ordenados y mutuamente exclusivos, considerandose una respuesta multinomial. Entonces los puntos $l_1 = l_1, \dots, l_{k-1}$ dividen a la escala de respuesta en k clases C_1, \dots, C_k , que corresponden respectivamente a los intervalos $(-\infty, l_1), \dots, (l_{k-1}, \infty)$. Al inicio del ensayo cada sujeto pertenece a la clase C_1 .

Si suponemos que la variable subyacente tiene una distribución normal, la probabilidad de que un individuo este en la categoría C_j después de la aplicación de la dosis x_i puede escribirse como:

$$P_{(j,i)} = Q_{(j-1,i)} - Q_{(j,i)}$$

$$= (1/\sqrt{2}\sigma) \int_{(g(x_i) - l_{j-1})/\sigma}^{(g(x_i) - l_j)/\sigma} e^{-t^2/2} dt$$

$$= (1/\sqrt{2}\sigma) \int_{(1 - g(x_i))/\sigma}^{(1 - g(x_i))/\sigma} e^{-t^2/2} dt$$

en donde $g(x) = E(y)$, x representa el estímulo y y la reacción subyacente, es decir suponemos que existe una relación cuantitativa subyacente entre la dosis y la respuesta. La relación más común es $E(y) = \alpha + \beta x$.

En este caso l representa los puntos de corte o fronteras, que definen las categorías y son supuestos conocidos.

Algo más general, pero dentro de este mismo tipo de generalización, es lo propuesto por Gurland, Lee, Dahm (1960), los cuales proponen utilizar una función monótona como función de distribución de la variable subyacente, en particular logit o normit (probit).

La generalización del análisis probit para el caso de k categorías es la siguiente: se tienen m grupos con n_1, \dots, n_m individuos expuestos a dosis x_1, \dots, x_m de cierto estímulo. Al final de un tiempo dado de observación los n_i individuos expuestos a la dosis x_i se encuentran distribuidos en k categorías, de tal forma que r_{ih} pertenecen a la categoría h -ésima.

Las proporciones esperadas se expresan como :

$$p_{ih} = r_{ih} / n_i \quad h=1, \dots, k-1 .$$

$$p_{ik} = 1 - \sum_{h=1}^{k-1} p_{ih}$$

y las proporciones esperadas como $P_{ij} = E(p_{ij})$ y considerando una

distribución normal de tolerancia a las dosis. Entonces

$$P_{i1} = \Phi(\alpha_1 + \beta x_i)$$

$$P_{i1} + P_{i2} = \Phi(\alpha_2 + \beta x_i)$$

en general

$$\sum_{j=1}^h P_{ij} = \Phi(\alpha_j + \beta x_i) \quad h=2, \dots, k-1$$

donde $\beta = 1/\sigma$ y $\alpha^{(j)} = \mu^{(j)}/\sigma$. Si suponemos que las posibles respuestas son tres (por ejemplo vivo, moribundo y muerto), estamos suponiendo una distribución de tolerancia a las dosis letales como $N(\mu^{(1)}, \sigma^2)$ y una distribución $N(\mu^{(2)}, \sigma^2)$ a las dosis moribundas, más aún $\mu^{(1)} > \mu^{(2)}$. Además debe ser un parámetro común para evitar que sea factible $P_{i1} + P_{i2} < P_{i1}$ lo cual obviamente no es permisible. Aitchinson y Silvey no consideraron esta restricción, en cambio Ashford lo hace implícitamente al considerar que existe un orden en los valores de las fronteras.

En este modelo las probabilidades se expresan como:

$$P_{i2} = \Phi(\alpha^{(2)} + \beta x_i) - P_{i1}$$

o en general

$$P_{ij} = \int_{\alpha^{(j-1)} + x_i}^{\alpha^{(j)} + \beta x_i} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Lo cual corresponde a la definición de Ashford considerando $g(x)$ definida como $\alpha + \beta x$, los modelos difieren en que Ashford considera conocidas las fronteras $1^{(j)}$ y Gurland et. al. no; la equivalencia es $\alpha^{(j)} = \alpha + 1^{(j)}$.

Si en vez de considerar una distribución de la variable subyacente normal, consideramos una logística la definición de P_{ij} queda como:

$$P_{11} = \left(1 + e^{-\alpha^{(1)} - \beta x_1} \right)^{-1}$$

$$P_{11} + P_{12} = \left(1 + e^{-\alpha^{(2)} - \beta x_1} \right)^{-1}$$

en general

$$P_{ij} = \left(1 + e^{-\alpha^{(j)} - \beta x_i} \right)^{-1} - \left(1 + e^{-\alpha^{(j-1)} - \beta x_i} \right)^{-1}$$

Bock(1975) menciona también esta clase de modelos, las aclaraciones importantes hechas por él son:

1. El considerar que las categorías tienen valores

1.....m se justifica en el sentido de que los valores guardan un orden pero no supone igual longitud para los intervalos que definen las categorías. Estos valores pueden ser empleados mas que nada como etiquetas.

2. Al considerar varias poblaciones ó subpoblaciones, estamos suponiendo que $\sigma_j = \sigma$ para todas las subpoblaciones.

Todos los modelos anteriores, con excepción del de Aitchinson y Silvey tiene tras de si la existencia de una variable subyacente y probablemente no observada, por ejemplo, tolerancia, la cual se supone tiene una distribución continua en la población. Entonces las categorías son consideradas intervalos contiguos en la escala continua. Los puntos de corte, cotas, fronteras o umbrales pueden ser considerados tanto conocidos como desconocidos. Cuando las fronteras son desconocidas este tipo de modelos se conoce como modelos de umbral (threshold), las suposiciones de este modelo son:

1. Una variable aleatoria discreta definida por $z_i = j$, cuando el individuo i pertenece a la categoría j , la cual esta relacionada a una variable respuesta continua no observable t_i de la manera

$$n_{j-1} < t_i < n_j \Leftrightarrow z_i = j,$$

donde $j=1, \dots, m$, $n_0 = -\infty$, $n_m = \infty$ y n_1, \dots, n_{m-1} desconocidas.

2. La variable respuesta subyacente satisface el modelo lineal

$$t_i = x_i' u_i + e_i \quad i=1, \dots, n$$

- $x_{i,ixr}$ representa los valores de r variables explicativas,
- u_{ixr} es un vector de parámetros desconocidos,
- e_i es un error aleatorio.

Si suponemos que los e_i 's son independientes e idénticamente distribuidos con una función de distribución continua y específica, la probabilidad P_{ij} de que el i -ésimo individuo pertenezca a la categoría j está dada por:

$$P_{ij} = P(n_{j-1} < t_i < n_j) = F(n_j - x_i' t) - F(n_{j-1} - x_i' t).$$

Dentro de este tipo de modelos de categoría de respuesta ordenada que no son generalizaciones de probit o logit están los propuestos por Mc. Cullagh (1980). Todos los modelos propuestos por Mc. Cullagh suponen, en principio, la existencia de una variable subyacente continua no observable, es decir modelos de tipo umbral, por lo que las categorías pueden ser pensadas como intervalos contiguos en alguna escala continua.

Distintos modelos difieren en las suposiciones concernientes a la variable latente, sin embargo los modelos descritos muchas veces no hacen referencia a la existencia de la variable latente y ésta no se requiere para la interpretación del modelo. Las

conclusiones son de cualquier forma posibles en ambos casos a pesar de que si una variable latente es directamente considerada, las afirmaciones que se pueden hacer son mas concisas e incisivas.

La forma general de los modelos de Mc. Cullagh es:

$$\text{liga}(\delta_{ij}) = \alpha_j - \beta x_i^t$$

donde liga es una función creciente monótona que mapea el intervalo $(0,1) \rightarrow (-\infty, \infty)$ y y_1, y_2 son los indicadores y que tiene por objeto relacionar la probabilidad acumulada con un modelo lineal, α_j representa las fronteras desconocidas.

En particular Mc. Cullagh presenta dos modelos explícitamente. El primero es el modelo de momios (odds) proporcionales en el cual supone que las k categorías ordenadas de la respuesta tienen probabilidades P_{1i}, \dots, P_{ki} cuando la covariable toma el valor x_i . Si Y es la respuesta que toma valores en el rango $1, \dots, k$ con las probabilidades anteriores y $K_j(x)$ los momios que especifican $Y = j$ dado el valor de la variable(s) o factor(es) explicativo(s). Entonces el modelo de momios proporcionales establece

$$K_j(x) = K_j \exp(-\beta x^t) \quad (1 \leq j < k),$$

donde β es un vector de parámetros desconocidos. El cociente de los momios correspondientes es independiente de j y solo depende de las diferencias entre los valores de la variable explicativa

$$K_j(x) / K_j(x_1) = \exp(\beta_j (x - x_1)) \quad 1 \leq j < k$$

El momio para el evento $y \leq j$ es el cociente

$$\delta_{ji} / (1 - \delta_{ji})$$

donde $\delta_{ji} = P_{1i} + \dots + P_{ji}$. Entonces :

$$\log[\delta_{ji} / (1 - \delta_{ji})] = \alpha_j - \beta_j x_i \quad 1 \leq j < k$$

donde $\alpha_j = \log K_j$.

Es importante notar que en el caso de una respuesta con sólo dos categorías el modelo anterior es equivalente al modelo logístico lineal usual para datos binarios propuesto por Cox(1970) y en este caso particular es también equivalente al modelo log-lineal.

El segundo modelo propuesto por Mc. Cullagh es el de los modelos de riesgos proporcionales (proportional hazards). El nombre de estos modelos proviene de la función de riesgo instantáneo, utilizada en el análisis de supervivencia, definida como la probabilidad de falla instantánea en el tiempo t condicionada a sobrevivir hasta el tiempo t . Para un individuo con variable explicativa x el modelo de riesgos proporcionales es

$$\lambda(t;x) = \lambda_0(t) \exp(-\beta x)$$

donde $\lambda(t)$ es la función de riesgo cuando $x=0$ y β es un vector de parámetros desconocidos.

La función de supervivencia $S(t;x)$, que es la probabilidad de sobrevivir más allá del tiempo t dado x , satisface

$$-\log[S(t;x)] = \int_0^t \lambda(s) \exp(-\beta x) ds,$$

donde $\lambda(t) = \int_0^t \lambda(s) ds.$

Además, para dos individuos con valores de la variables explicativa x_1 y x_2 respectivamente la función de supervivencia satisface :

$$\log[S(t;x_2)] / \log[S(t;x_1)] = \exp[\beta (x_2 - x_1)],$$

es decir, tanto el cociente del logaritmo de la función de supervivencia como el cociente de la función de riesgo, dependen únicamente de la diferencia $x_2 - x_1$ y es constante para toda t .

Para datos discretos, el modelo se convierte en:

$$-\log(1 - \delta_{ji}) = \exp(\alpha_j - \beta x_{j1}),$$

donde $1 - \delta_{ji}$ es la probabilidad complementaria o la probabilidad de sobrevivir más allá de la categoría j dado los valores de x_{j1} .

El modelo linealizado es:

$$\log[-\log(1 - \delta_{ji})] = \alpha_j - \beta x_i^t$$

este modelo es el que se conoce como la transformación complementaria log-log, o doble logaritmo complementario.

Como se mencionó antes los modelos de Mc. Cullagh corresponden a una estructura $\text{liga}(\delta_{ji}) = \alpha_j - \beta x_i^t$, la cual en los modelos anteriores corresponde a la función logística y doble logaritmo complementario respectivamente. Esto es posible generalizarlo utilizando como función liga cualquier función creciente monótona que mapee el intervalo $(0,1) \rightarrow (-\infty, \infty)$. Es claro que los modelos descritos anteriormente (logit politómico, probit politómico) se pueden generalizar a esta expresión.

En los modelos de Mc. Cullagh los parámetros a estimar son α_j y el parámetro de regresión β . Los parámetros α_j son generalmente de poco interés (nuisance) y usualmente se denominan puntos de corte o umbrales (es decir son los puntos que definen las categorías), mientras que el parámetro de regresión β describe como los momios o cualquier otra cantidad de interés esta relacionada con los valores de x .

Todos los modelos de esta forma describen un orden estocástico estricto. Esto es, si tomamos dos subpoblaciones con valores de la variable explicativa x_1 y x_i se tiene que :

$$\text{liga}(\delta_{j1}) - \text{liga}(\delta_{j1}) = \beta (x_1 - x_i)^t = \Delta$$

como la liga es una función monótona creciente se sigue que:

$$\begin{array}{ccc} \delta & > & \delta & & j & \delta \\ j_1 & & j_1 & & & \\ \delta & < & \delta & & j & \\ j_1 & & j_1 & & & \end{array}$$

de acuerdo al signo de Δ .

En general los modelos para categorías de respuesta ordenadas no son invariantes bajo cualquier permutación de las categorías. Intuitivamente parecería razonable pensar que los modelos de este tipo son en algún sentido invariantes bajo la permutación especial que consiste en tomar el orden inverso de las categorías pero no bajo cualquier permutación arbitraria. Estas ideas están fundamentadas en el concepto conocido como invarianza palindrómica (Mc. Cullagh, 1972) pero esto depende fuertemente de la aplicación particular.

En los modelos cuya función liga es la función logística, probit o Cauchy inversa se tiene esta invarianza palindrómica y al utilizar esta permutación los estimadores de los parámetros varían en signo y las α_j en orden. Para los modelos cuya función liga es doble logaritmo (log-log) o doble logaritmo complementario esta invarianza no siempre se da, depende mucho de la estructura de los datos. La falta de invarianza puede o no ser vista como defecto del modelo.

Un atractivo para cualquier modelo de este tipo es que puede

ser reparametrizado en forma tal que la relación pueda aplicarse variando las condiciones y sea consistente, en lo posible, con las leyes físicas o biológicas. Variar las condiciones significa una redefinición de las categorías de las respuestas, agrupando o dividiendo las categorías, esto es posible hacerlo ya que el parámetro o los parámetros de interés no dependen de la interpretación de las categorías actuales de la respuesta, a pesar de que los estimadores de los puntos de corte en general van a ser afectados. Esta propiedad permite probar la consistencia de varias fuentes de información y si es justificable combinar información de fuentes separadas. Todos los modelos de este estilo tienen esta propiedad, en cambio los modelos log-lineales no la tienen.

Una desventaja de este tipo de modelos es que no existe manera de utilizar el orden de una variable explicativa cuando dicho orden existe, sin embargo hay varios métodos alternativos, por ejemplo utilizar puntajes, partir la estadística ² en componentes lineales, cuadráticos o de más alto orden. Y en el caso en que no fuera posible utilizar puntajes los estimadores podrían estar sujetos a la propiedad de monotonía $\alpha_1 < \dots < \alpha_p$.

Dado que el modelo lineal general presentado por Mc. Cullagh no es la única estructura que describe un orden estocástico estricto, Mc. Cullagh sugiere el utilizar por ejemplo puntajes en un modelo log-lineal para partir la interacción en dos componentes, uno que describa orden estocástico estricto y el otro componente de orden mayor la interacción.

En particular en el caso de dos muestras (dos subpoblaciones diferentes) donde la respuesta tiene k categorías ordenadas, scores $1, 2, \dots, k$ son asignados a las categorías. El modelo log-lineal con efectos principales solamente implica que el cociente de logaritmos-cruzados de las cuatro celdas adyacentes es cero, mientras que la inclusión del término lineal-lineal implica que este cociente es constante sobre la tabla entera.

Sin embargo, el cociente cruzado para celdas adyacentes es poco recomendable, aún en casos donde el modelo ajusta bien, lo cual sucede seguido. La principal desventaja es que las posibles inferencias basadas en esa estadística están limitadas a las categorías usadas en la muestra, además que diferentes pero no menos arbitrarias categorías, producirían cocientes-cruzados completamente diferentes los cuales en general no serían constantes en toda la tabla, es decir existiría carencia de invarianza bajo agrupación de categorías adyacentes. Mientras que por otra parte, conclusiones basadas en los modelos proporcionales ya sean momios o riesgos proporcionales pueden ser establecidas sin referencia a las categorías utilizadas en la muestra.

2.2 Modelos de respuesta categoriaca no ordenada.

En cuanto a los modelos de respuesta no ordenada, de acuerdo a Amemiya(1975) podemos clasificarlos en tres tipos:

El primero de estos tipos es llamado por Amemiya (1981) logit independiente, el cual es también conocido como logit condicional. Este método se genera suponiendo que dado cualquier

par de respuestas la selección es hecha de acuerdo al modelo logístico dicotómico. Entonces se tiene en el caso de tres categorías.

$$\frac{P_{11}}{(P_{11} + P_{31})} = \frac{1}{1 + \exp(\beta_1 x_1)}$$

$$\frac{P_{21}}{(P_{21} + P_{31})} = \frac{1}{1 + \exp(\beta_2 x_1)}$$

$$\frac{P_{11}}{(P_{11} + P_{21})} = \frac{1}{1 + \exp(\beta_3 x_1)}$$

es claro que P_{11}, P_{21}, P_{31} están unívocamente determinadas por las dos primeras relaciones y de el hecho de que $P_{11} + P_{21} + P_{31} = 1$ entonces se tiene que $\beta_3 = \beta_1 - \beta_2$ y esto nos lleva a lo que se conoce como el modelo logístico no ordenado, en donde

$$P_{11} = \frac{\exp(\beta_1 x_1)}{(1 + \exp(\beta_1 x_1) + \exp(\beta_2 x_1))}$$

$$P_{21} = \frac{\exp(\beta_2 x_1)}{(1 + \exp(\beta_1 x_1) + \exp(\beta_2 x_1))}$$

$$P_{31} = \frac{1}{(1 + \exp(\beta_1 x_1) + \exp(\beta_2 x_1))}$$

obviamente en el caso de más de tres categorías respuesta

$$P_{ji} = \frac{\exp(\beta_j x_i)}{1 + \sum_{s=1}^{k-1} \exp(\beta_s x_i)} \quad)$$

$$P_{ki} = \frac{1}{1 + \sum_{s=1}^{k-1} \exp(\beta_s x_i)} \quad)$$

donde k representa el número de categorías.

El segundo tipo de modelos consiste en considerar que la selección de categorías se realiza secuencialmente, por ejemplo, en el caso particular de tres categorías, una vez que se ha determinado si la elección es la primera categoría o no lo es, entonces dado que no pertenece a la primera categoría se determina si la elección es la segunda o la tercera. Estos modelos pueden ser tanto el secuencial probit ó secuencial logit. Lo importante de estos modelos es que la estimación de los parámetros se puede realizar como la estimación sucesiva de modelos dicotómicos lo cual desde el punto de vista computacional resulta muy bueno tanto por facilidad, como por economía.

El modelo secuencial probit queda expresado como:

$$P_{1i} = \Phi(\beta_1 x_i)$$

$$P_{li} = \prod_{j=1}^{l-1} [1 - \Phi(\beta_j x_i)] \Phi(\beta_l x_i), \quad l=2, \dots, k-1$$

$$P_{ki} = \prod_{j=1}^{k-1} [1 - \Phi(\beta_j x_i)]$$

Mientras que el modelo secuencial logit queda definido como:

$$P_{1i} = \frac{e^{\beta_1 x_i}}{1 + e^{\beta_1 x_i}},$$

$$P_{li} = \frac{e^{\beta_l x_i}}{1 + e^{\beta_l x_i}} \left[\prod_{j=1}^{l-1} \frac{e^{\beta_j x_i}}{1 + e^{\beta_j x_i}} \right], \quad l=2, \dots, k-1$$

$$P_{ki} = \frac{1}{1 + e^{\beta_k x_i}}.$$

La tercera clase de modelos es la generada por maximización de utilidad, es decir suponemos que el individuo elige la categoría que representa para el mayor utilidad o beneficio.

Dentro de este tipo de modelos esta el propuesto por Aitchinson y Bennett(1970), este enfoque tiene mayor interpretabilidad en el contexto de los problemas en economía que el comunmente conocido. Para facilidad se describira el modelo en el caso de respuesta binaria y de ahí se hará la generalización.

En este caso el modelo puede generarse como el resultado de dos experimentos, en donde la categoría de la respuesta va a estar determinada por los efectos que produce cada experimento. Si consideramos que las categorías son respuesta y no respuesta, se va a obtener respuesta del individuo si la utilidad estimada del experimento 1 es mayor que la del experimento 2.

Esto es más claro con un ejemplo. Supongamos que se está estimando la demanda de un cierto bien y que las posibles respuestas para el estímulo que representa el ingreso de una persona son comprar o no comprar. En este caso el primero de los experimentos estima el grado de uso ó disfrute, y_1 , proveniente de la compra del bien tomando en cuenta la consecuencia de quedar corto de dinero, mientras que el segundo experimento mide bajo las condiciones de no haber comprado el bien, la comodidad y beneficios, y_2 , proveniente de la retención del ingreso pero sin poseer el bien.

La compra va a ser hecha si y solo si $y_1 > y_2$. Entonces suponiendo que para una persona con ingreso x las variables aleatorias y_1, y_2 son independientes y con distribuciones $N(\alpha_1 + \beta_1 x, \sigma_1^2)$ y $N(\alpha_2 + \beta_2 x, \sigma_2^2)$ respectivamente. La probabilidad de compra es:

$$\begin{aligned}
 P(y_1 > y_2) &= \frac{1}{\sigma_1} \int_{-\infty}^{\infty} \frac{[(y_2 - \alpha_2 - \beta_2 x)/\sigma_2] dy_2}{\sigma_2} \\
 &= \Phi\left[\frac{(\alpha_1 - \alpha_2 + (\beta_1 - \beta_2)x)/\sigma\sqrt{2}}{\sigma\sqrt{2}}\right] \\
 &= \Phi(A+Bx)
 \end{aligned}$$

que corresponde exactamente al modelo probit con $A = (\alpha_1 - \alpha_2)/(\sigma\sqrt{2})$ y $B = (\beta_1 - \beta_2)/(\sigma\sqrt{2})$, donde y_1, y_2 son los indicadores de la respuestas 1,2, la elección de la respuesta es conocida como indicador máximo.

La generalización del modelo a k categorías de respuesta

consiste en que el sujeto con estímulo x_i realiza k experimentos independientes, el resultado, y_j , del j -ésimo experimento, es el indicador de la respuesta j y suponemos esta distribuido como $N(\alpha_j + \beta_j x_i, \sigma_j^2)$, $j=1, \dots, k$. El sujeto entonces elige la categoría que corresponde al máximo indicador. Suponemos que no hay empates o que la probabilidad de empates es negligible.

Es importante hacer notar que la suposición de independencia de las y_j no es muy razonable siendo que un mismo sujeto esta involucrado en todos los experimentos, sin embargo el considerar que las y_j son independientes representa simplicidad y tratabilidad en el modelo, mientras que el no suponerlo permite mayor realismo pero induce intratabilidad además viene a acentuar un problema del que más tarde se hablará que es identificabilidad.

Entonces, suponiendo independencia y definiendo P_{ji} = probabilidad de la categoría j , dado x_i

$$\begin{aligned}
 P_{ji} &= P(y_j = \max(y_1, \dots, y_k) = P(y_j > y_s \quad \forall s \neq j) \\
 &= 1/\sigma_j \int_{-\infty}^{\infty} [(y_j - \alpha_j - \beta_j x_i)/\sigma_j] \\
 &\quad \prod_{s \neq j} [(y_s - \alpha_s - \beta_s x_i)/\sigma_s] dy_j \\
 &= \phi(v) \Phi(v + A_j + B_j x_i) \prod_{s \neq j, k} \Phi(v + (A_s - A_j) + (B_s - B_j) x_i) dv
 \end{aligned}$$

donde $A_j = (\alpha_j - \alpha_k)/\sigma_j$, $B_j = (\beta_j - \beta_k)/\sigma_j$ $j=1, \dots, k-1$.

Un problema realmente importante en esta generalización es el

de identificabilidad de los parámetros. En el caso de dos categorías, para un solo ensayo binomial el éxito o respuesta identifica completamente el modelo, pero para cualquier valor específico de θ a esta probabilidad de tener éxito corresponden muchas parejas (A,B) que satisfacen:

$$A + Bx = \Phi^{-1}(\theta)$$

es decir A,B no son identificables, sin embargo para un modelo consistente de dos ensayos binomiales a diferentes estímulos x_1 y x_2 los parámetros A,B son identificables ya que las ecuaciones

$$P_{11} = \Phi(A + Bx_1) = \theta_1$$

$$P_{12} = \Phi(A + Bx_2) = \theta_2$$

tienen solución única, por ser el jacobiano diferente de cero

$$J = (x_2 - x_1) \phi(A+Bx_1) \phi(A+Bx_2) \neq 0$$

Este resultado se puede generalizar al caso de k respuestas, en este caso para que los parámetros sean identificables el modelo debe consistir, al menos, de dos ensayos multinomiales con probabilidades de las categorías $P_{11}, \dots, P_{k-1,1}$, $x_1 \neq x_2$ los parámetros $A_1, \dots, A_{k-1}, B_1, \dots, B_{k-1}$ son identificables si el jacobiano es diferente de cero

El valor de este jacobiano es

$$(x_2 - x_1)^{k-1} \prod_{j=1}^{k-1} m_{ij}$$

donde M_i es la matriz que tiene elementos m_{ij} de la forma :

$$m_{ic} = \delta P / \delta A_i \quad \text{si } i \neq j$$

$$= \int_{-\infty}^{\infty} \varphi(v) \Phi(v + A_i + B_i x) \prod_{j \neq i, k} \Phi(v + (A_i - A_j) + (B_i - B_j) x) dv$$

y

$$m_{ii} = d_i - \sum_{k \neq i} m_{ik}$$

con

$$d_i = \int_{-\infty}^{\infty} \varphi(v) \varphi(v + A_i + B_i x) \prod_{j=1, k} \Phi(v + (A_i - A_j) + (B_i - B_j) x) dv$$

Además $\frac{\delta P_i}{\delta B_c} = x \frac{\delta P_i}{\delta A_c}$

basta ver que $\prod_{j=1}^{k-1} m_{ij}$ y $\prod_{j=1}^{k-1} m_{ij}$ son diferentes de cero para probar identificabilidad, esto se hace por inducción, el primer paso de la inducción está dado por el caso de dos categorías. La prueba se anexa al final.

Aunque los parámetros $\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k$ del modelo general no son identificables, las diferencias estandarizadas de las α_i 's y las β_i 's son identificables.

Dentro de este mismo enfoque Mc. Fadden [1974] deriva el modelo de logit independiente considerando como categoría de respuesta la que genera la utilidad máxima. Si suponemos que un individuo i tiene utilidades asociadas dadas por:

$$U_{ij} = \mu_{ij} + e_{ij}$$

donde μ_{ij} es una función no estocástica de variables explicativas y parámetros desconocidos y e_{ij} una variable aleatoria no observable. Mc. Fadden (1974) probó que el logit independiente es derivado de esta forma si y solo si los e_{ij} son independientes y su función de distribución es dada por $\exp(-e_{ij})$. La prueba para el caso de tres categorías se anexa al final.

Bock(1975) trata estos modelos desde otro punto de vista, ya que permite que exista cierta estructura en las categorías de la variable respuesta. Suponiendo que existen k categorías dadas por un vector de proceso latente continuo (y_1, \dots, y_k) y el sujeto elige el estímulo correspondiente a la componente con el valor más grande para él en ese momento. Dado que el proceso tiene una distribución normal multivariada $F(y)$ en la población de sujetos, la probabilidad de que un sujeto seleccionado alatormente elija el estímulo x_1 es:

$$\begin{aligned} P(x_1 > x_2 \cup x_3 \cup \dots \cup x_k) &= P(y_1 > y_2 \cup \dots \cup y_k) \\ &= P((y_1 - y_2) > 0 \dots (y_1 - y_k) > 0) \\ &= \int_0 \int_0 \dots \int_0 G(y_1 - y_j) d(y_1 - y_j) \quad j=2,3,\dots,k \end{aligned}$$

donde G es la distribución de el proceso de diferencias derivado de $F(y)$.

Si suponemos que $F(y)$ es normal multivariada, $G(y_1 - y_j)$ es

normal multivariada ya que

$$Y \sim N(\mu, \Sigma) \quad CY \sim N(C\mu, C\Sigma C')$$

donde C es la matriz de contrastes

$$C = \begin{vmatrix} 1 & 0 & \dots & -1 \\ 0 & 1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{vmatrix}$$

Esta suposición, como ya se menciono antes tiene la dificultad práctica de la evaluación de la integral.

Un caso especial es en el que las varianzas en son iguales y las covarianzas son también iguales. Entonces

$$C C' = \begin{vmatrix} 1 & 1/2 & \dots & 1/2 \\ 1/2 & 1 & \dots & 1/2 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 1/2 & \dots & 1 \end{vmatrix}$$

Es decir el proceso de diferencia tiene varianza constante y correlación 1/2.

Para este caso existen aproximaciones de las cuales la más simple puede ser derivada de la distribución logística.

$$\Psi(t_1, \dots, t_p) = 1 / (1 + e^{-t_1} + \dots + e^{-t_p})$$

ya que la distribución marginal de U es logística, con vector de medias cero y varianzas $\sigma^2 = \pi^2/3$. La covarianza para cualquier par de variables es $\pi^2/6$ y la correlación es

$$= (\pi^2/6) \cdot (3/\pi^2) = 1/2$$

Para valores "moderados" de t aproxima razonablemente bien a la distribución normal multivariada con esas medias, varianzas y covarianzas.

A manera de aproximación $t_1 = \mu_1 - \mu_2, t_2 = \mu_2 - \mu_1, \dots, t_k = \mu_k - \mu_1$ ($p=k-1$). Entonces

$$P(y_1 > y_2 \dots y_k) = 1 / (1 + e^{-(\mu_1 - \mu_2)} + e^{-(\mu_2 - \mu_1)} + \dots + e^{-(\mu_k - \mu_1)})$$

$$= e^{\mu_1} / (e^{\mu_1} + \dots + e^{\mu_k})$$

similarmente para 2, 3, ..., k.

Definimos el logit multinomial, suponiendo n grupos de sujetos en k categorías, para el grupo j como:

$$z_j = \begin{vmatrix} z_{j1} \\ z_{j2} \\ \vdots \\ z_{jk} \end{vmatrix}$$

Entonces las probabilidades de las categorías para el grupo j son :

$$P_{j1} = \frac{e^{z_{j1}}}{D_j}, \quad P_{j2} = \frac{e^{z_{j2}}}{D_j}, \dots, \quad P_{jk} = \frac{e^{z_{jk}}}{D_j}$$

donde $D_j = \sum_j e^{z_{j1}} + \dots + e^{z_{jk}}$.

Y el modelo lineal multivariado

$$Z_{nxk} = X_{nxq} B_{qxt} A_{txk}$$

donde X es una matriz de valores conocidos de q variables independientes asociadas con cada uno de los grupos de sujetos. B contiene parámetros desconocidos y A es una matriz que incorpora la estructura de las k categorías de respuesta, es decir se permite la posibilidad de una estructura entre grupos y entre categorías de respuesta.

Ahora bien si los grupos de sujetos tienen una estructura cruzada o anidada, X será la matriz modelo de un diseño, y generalmente de rango incompleto r. n. En este caso será necesario reparametrizar el modelo, $X=KL$, donde L satisface la condición de estimabilidad

$$\text{rango } \begin{vmatrix} X \\ L \end{vmatrix} = \text{rango}(X) = r.$$

L puede ser elegido por el investigador y K calculada de

$$K = XL^{-1} (LL^{-1})$$

Como A juega un papel similar a X , pero con respecto a las categorías de respuesta, es muy probablemente de rango incompleto y puede ser factorizada como.

$$A = S T$$

$$t_{xk} \quad t_{xs} \quad s_{xk}$$

donde $\text{rango}(s) = s$ y las columnas de S son linealmente dependientes de las de A . La matriz T puede ser obtenida $T = (S^t S)^{-1} S^t A$ o generada directamente.

$$\text{Entonces } Z = K(LBS)T = K \Gamma T$$

donde K es llamado prefactor, Γ parámetros y T el postfactor.

2.3 Diferencias entre los modelos de respuesta ordenada y no ordenada.

Cuando se cuenta con categorías de respuesta ordenada y utilizamos modelos para categorías de respuesta no-ordenada, estamos perdiendo información que de otra forma podríamos incorporar en el análisis, además el ajuste no nos garantiza que se cumpla el orden estocástico estricto, por otro lado hacer lo contrario es igual de riesgoso, ya que estamos implantando un modelo con restricciones adicionales, que no tiene por que cumplir.

2.4 Modelos para el caso Multivariado

Existen esencialmente dos maneras diferentes de definir estos modelos:

1. especificando la distribución marginal y después

especificar la conjunta.

2. especificando distribuciones condicionales.

La diferencia fundamental entre ellas, es que en el primer caso la categoría de respuesta de una de las variables no afecta a la(s) otra(s), mientras que en el caso de las distribuciones condicionales esto sí sucede.

2.4.1 Modelos provenientes de distribuciones marginales

Suponemos dos variables aleatorias politómicas, Y_{1t}, Y_{2t} , de tal forma que los valores que pueden tomar las variables están asociados con la partición de un espacio de probabilidad, es decir $P(y_{it} = k) = P(S_{ik}^i)$ $i=1,2$, donde U, S_{ik}^i representa el espacio de probabilidad. Podemos entonces definir

$$P(y_{1t} = k, y_{2t} = h) = P(S_{k,h}^{1,2})$$

en donde la probabilidad del espacio producto debe ser apropiadamente elegida.

Obviamente la probabilidad marginal no determina en forma única la probabilidad en el espacio producto. En el caso especial en que cada distribución marginal es generada por una distribución normal la probabilidad conjunta se puede considerar en forma "natural" como normal multivariada, pero la extensión en general es difícil con otras distribuciones.

A diferencia del caso univariado, en este caso modelos multivariados normales son preferidos a los logísticos debido a

las propiedades de la distribución normal multivariada.

El modelo de análisis probit, tanto ordenado como no ordenado puede ser fácilmente extendido de esta manera. Ashford y Sowden(1970) presentan este modelo bajo el nombre de Análisis Probit Multivariado, en este caso un mismo estímulo puede afectar a más de un sistema y es posible registrar la respuesta de cada sistema por separado.

Para el caso particular en que la respuesta de cada sistema sea determinada en forma binaria :

$$P_i(x) = \frac{1}{1 + e^{-\beta_i x}}$$

va a representar la probabilidad de obtener respuesta en el sistema i , mientras que la probabilidad de no respuesta en dicho sistema es:

$$P_i^2 = 1 - \frac{1}{1 + e^{-\beta_i x}}$$

Cuando consideramos la respuesta conjunta de más de un sistema, es necesario especificar la correspondiente distribución conjunta de tolerancias. En este caso es natural tomar la distribución conjunta como normal bivariada estándar.

Si denotamos la distribución de una normal bivariada estándar con coeficiente de correlación ρ por $F(w_1, w_2, \rho)$ la probabilidad de que dos sistemas S_i, S_j manifiesten respuesta es

$$P_{ij}^{11}(x) = F(\beta_1 x, \beta_j x, \rho_{ij})$$

la probabilidad de que el sistema i manifieste la respuesta mientras que el sistema j no lo haga es

$$P_{ij}^{12}(x) = P_i^{11}(x) - P_{ij}^{11}(x).$$

analogamente, para el caso recíproco se tiene

$$P_{ij}^{21}(x) = P_j^{11}(x) - P_{ij}^{11}(x)$$

y la probabilidad de que ninguno de los sistemas responda es

$$P_{ij}^{22}(x) = 1 - P_{ij}^{12}(x) - P_{ij}^{21}(x) - P_{ij}^{11}(x)$$

Expresiones similares resultan para la respuesta conjunta de más de dos sistemas y más de dos categorías por sistemas. La forma de la función de los efectos $\beta_1 x$ está usualmente determinada por las correspondientes respuestas univariadas P_i , los únicos cuantiles adicionales involucrados en la probabilidad conjunta son los coeficientes de correlación $\{\rho_{ij}\}$, siendo que un ρ_{ij} particular es determinado por la respuesta conjunta a los sistemas S_i y S_j , el modelo está completamente definido considerando únicamente los distintos pares de sistemas.

Es importante hacer notar que a pesar de que las distribuciones marginales puedan ser tomadas como normales

estándar, no es posible el tomar una función de respuesta conjunta simplemente como normal multivariada estándar. La suposición de una distribución conjunta normal multivariada es plausible, sin embargo no es la única distribución multivariada con marginales normales y la validación de la suposición de modelo normal debe ser realizada, antes de concluir el análisis estadístico.

Si en este caso consideramos la función conjunta de tres o más sistemas, el costo computacional sería muy alto, además que en estos momentos no son practicables los métodos iterativos que solucionen este problema.

El modelo probit para datos ordenados también puede ser generalizado de esta manera, substituyendo β_i por $\beta_i^{()}$, donde $\beta_i^{()}$ representa el vector $(\alpha \ \beta)$.

Así también, el modelo Probit secuencial para datos no ordenados puede extenderse al caso multivariado, reescribiendo el modelo como sigue:

$$P(y=2) = \Phi\left(\frac{\beta_2 x}{2}\right)$$

$$= P(u < \frac{\beta_2 x}{2})$$

$$P(y=1) = \Phi\left(\frac{\beta_1 x}{1}\right) [1 - \Phi\left(\frac{\beta_2 x}{2}\right)]$$

$$= P(u > \frac{\beta_2 x}{2}, v < \frac{\beta_1 x}{1})$$

$$P(y=0) = [1 - \Phi(\frac{t}{2} \beta_1 x)] [1 - \Phi(\frac{t}{2} \beta_2 x)]$$

$$= P(u > \frac{t}{2} \beta_1 x, v > \frac{t}{2} \beta_2 x)$$

donde u, v son variables aleatorias independientes normalmente distribuidas con media cero y varianza 1. En el caso bivariado, definimos

$$P(y_i = 2) = P(u_i > \frac{t}{2} \beta_{i2} x)$$

$$P(y_i = 1) = P(u_i > \frac{t}{2} \beta_{i2} x, v_i < \frac{t}{2} \beta_{i1} x)$$

$$P(y_i = 0) = P(u_i > \frac{t}{2} \beta_{i2} x, v_i > \frac{t}{2} \beta_{i1} x)$$

y suponemos además que $E u_i = \rho_u$ y $E v_i = \rho_v$, donde ρ_u y ρ_v son parámetros a estimar.

Alternativo al modelo presentado por Ashford y Sowden se tiene el propuesto por Grizzle (1971), que puede ser visto como una aproximación del Análisis Probit Multivariado o bien como un método independiente. Este método tiene la ventaja de que no necesita el cálculo de normales acumulativas, por lo que la extensión a varias dimensiones puede hacerse sin necesidad de ajustar las respuestas por pares.

Este modelo se basa en el modelo propuesto por Grizzle, et. al. (1969) :

$$F(\Pi) = X\beta$$

donde Π es un vector de probabilidades multinomiales, β es un vector de parámetros a estimar y X es una matriz de rango completo.

Un caso particular de este modelo es:

$$K \ln A \Pi = X\beta$$

sujeto a la restricción de que la matriz de derivadas de $K \ln A \Pi$ tiene rango completo.

Este modelo puede extenderse al caso multivariado definiendo apropiadamente K y A . En particular, para el caso bivariado con variables dicótomicas y m grupos de individuos se tiene

$$A = \begin{matrix} * \\ A & \otimes I \\ 4 \times 4 & m \times m \end{matrix} .$$

$$K = \begin{matrix} * \\ K & \otimes I \\ 2 \times 4 & m \times m \end{matrix} .$$

$$\Pi = (\Pi_1^t, \dots, \Pi_m^t) .$$

en donde $\Pi = (\Pi_{11}^t, \dots, \Pi_{12}^t)$ y $\sum_{j,k} \sum_{ijk} \Pi_{ijk} = 1$

$$A = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{vmatrix}$$

$$K = \begin{vmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{vmatrix}$$

que produce los logits marginales

$$l_{11} = \ln\left(\frac{\pi_{11} + \pi_{12}}{\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22}}\right) \text{ y}$$

$$l_{12} = \ln\left(\frac{\pi_{11} + \pi_{21}}{\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22}}\right),$$

dado que el objetivo es calcular dos líneas asociadas con estos logits se elige el modelo

$$l_{ij} = \alpha_j + \beta_j x_{ji} \quad i=1, \dots, m, \quad j=1, 2$$

el cual se obtiene con la elección apropiada de β y X .

2.4.2 Modelos de distribución condicional

Dentro de este tipo de modelos se encuentra el logit condicional, el cual en el caso de dos variables dicótomicas es:

$$P(y_1 = 1 | y_2) = L(\beta_1 x_1 + \beta_2 y_2)$$

$$P(y_2 = 1 | y_1) = L(\beta_2 x_2 + \beta_1 y_1)$$

si definimos

$$P_{ij} = P(y_1 = i, y_2 = j).$$

Entonces evaluando

$$P_{11} = \exp(\beta_1 x + \beta_{12} y) P_{21}$$

$$P_{12} = \exp(\beta_1 x) P_{22}$$

o

$$P_{11} = \exp(\beta_2 x + \beta_{21} y) P_{21}$$

$$P_{21} = \exp(\beta_2 x) P_{22}$$

se tiene lo siguiente

$$P_{11} / P_{22} = \exp(\beta_1 x) \exp(\beta_{12} y)$$

o

$$P_{11} / P_{22} = \exp(\beta_2 x) \exp(\beta_{21} y)$$

por lo que se cumple que $\beta_{12} = \beta_{21}$.

De manera que la probabilidad conjunta se escribe como:

$$P(y_1, y_2) = D^{-1} \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} y_1 y_2)$$

donde $D = 1 + \exp(\beta_1 x) + \exp(\beta_2 x) + \exp(\beta_{12} y_1 y_2)$.

Este procedimiento se puede generalizar al caso de J variables dicótomicas considerando

$$P(y_j = 1 | y_{i \neq j}) = L(\beta_j x_j + \sum_{i \neq j} \beta_{ji} y_i + \sum_{i \neq j} \beta_{jh} y_h + \beta_{ji \dots j-1} y_{j-1} + \beta_{j+1} y_{j+1} + \dots + \beta_J y_J)$$

y la probabilidad conjunta

$$P(y_1, \dots, y_J) = D^{-1} \exp[\sum_j \beta_j x_j y_j + \sum_{j < i} \beta_{ji} y_j y_i + \sum_{j < i < h} \beta_{jih} y_j y_i y_h + \dots + \beta_{1 \dots J} y_1 \dots y_J]$$

$$\text{donde } D = 1 + \sum_j \exp(\beta_j x_j) + \sum_{j < i} \exp(\beta_{ij}) + \dots + \beta_{1 \dots J}$$

Una manera sencilla de generalizar este método al caso de J variables politómicas es el describir cada variable de n categorías como $n-1$ variables dicótomicas $y_j^i = 1$ si $y_j = i$ y 0 en otro caso. Entonces podemos emplear la generalización anterior añadiendo la restricción de que solamente una de las y_j^i sea uno para toda j .

Otra manera de tratar los modelos multivariados es el construir una sola variable respuesta politómica, creando una categoría por cada combinación de la respuesta multivariada. Así en el caso de un modelo bivariado dicotómico tendríamos una variable respuesta de cuatro categorías. Esto no es siempre lo adecuado, pero cuando no se cuenta con suficientes recursos de cómputo puede facilitar la estimación, sin embargo hay que tener en cuenta que si tenemos m variables cada una con l categorías, vamos a crear una variable de mxl categorías y esto puede producir problemas de cómputo.

3. ESTIMACION

Amemiya (1976) propone un modelo general que engloba muchos de los modelos descritos anteriormente. El modelo general es el siguiente:

Considerese la variable aleatoria y_t , $t=1,2,\dots,T$ la cual toma $K+1$ valores diferentes con probabilidades dadas por:

$$P_k(t) = F_k(\beta_{k1} x_{t1}, \dots, \beta_{kh} x_{th}) \quad k=1, \dots, K$$

$$P_{K+1}(t) = 1 - \sum_{k=1}^K P_k(t)$$

con la suposición $H < k$ y β_{ht} es un vector renglón de parámetros desconocidos de dimensión q , x_{ht} es un vector columna de constantes conocidas de dimensión q , con $h=1, \dots, H$ y suponemos además que para toda h la matriz X_h de $T \times q$, $X_h = (x_{h1}, \dots, x_{ht})$ tiene rango q .

Suponemos que F_k es tal que $P_k \geq 0$ $\forall k$ y que F_k tiene derivadas parciales continuas de primer y segundo orden.

Más aún suponemos que existe un subconjunto de H de las K ecuaciones para las cuales el Jacobiano no se hace cero en los verdaderos valores de β_{ht} , por esto se puede escribir

$$\beta_{h\bar{s}} = G_h [P_1(t), \dots, P_K(t)] \quad h=1, \dots, H$$

Si $H=K$ las G_h están determinadas univocamente por F_1, \dots, F_K alrededor del verdadero valor del parámetro, lo cual no es necesariamente cierto si $H < K$.

Lo que se pretende hacer es estimar las β en base a N observaciones independientes de y . N lo podemos expresar como $N = \sum_{k=1}^{K+1} n_k(t)$, en donde $n_k(t)$ es el número de veces que $y = k$.

La manera en que algunos de los modelos descritos anteriormente se expresan en la forma general es la siguiente:

El modelo de Ashford(1959) establece

$$P_1(t) = \frac{(\alpha + \beta x_1 - y_1)/\sigma}{(\alpha + \beta x_t - y_t)/\sigma}$$

donde las y son conocidas y por lo tanto α y βx pueden ser expresados como función de las probabilidades

El modelo de Gurland, Lee y Dahm(1960) en el caso de normit de k categorías queda como

$$P_{-(j)}(t) = \Phi\left(\frac{\alpha + \beta x_j}{\sigma}\right) - \Phi\left(\frac{\alpha + \beta x_{j-1}}{\sigma}\right) \quad j=2, \dots, k-1$$

$$P_1(t) = \Phi\left(\frac{\alpha + \beta x_1}{\sigma}\right)$$

$$P_k(t) = 1 - (P_1(t) + \dots + P_{k-1}(t)) = 1 - \Phi\left(\frac{\alpha + \beta x_{k-1}}{\sigma}\right)$$

lo cual inversamente se puede expresar como:

$$\beta x_t = G(P_1(t), \dots, P_{k-1}(t))$$

$$\alpha_{(j)} = \Phi^{-1}(P_j(t) + P_{j-1}(t)) - \beta x_t \quad j=2, \dots, k-1$$

$$\alpha_{(1)} = \Phi^{-1}(P_1(t)) - \beta x_t$$

En el caso de que exista una función de tolerancia logística el modelo se puede expresar inversamente como:

$$\beta x_t = G(P_1(t), \dots, P_{k-1}(t))$$

$$\alpha_{(1)} = -\ln((1-P_1(t))/P_1(t)) - \beta x_t$$

$$\alpha_{(j)} = -\ln((1-P_j(t)-P_{j-1}(t))/(P_j(t)-P_{j-1}(t))) - \beta x_t \quad j=2, \dots, k-1.$$

Ahora bien los modelos de Mc. Cullagh pueden expresarse inversamente como

$$P_{jt} = \exp(\alpha_j - \beta x_t) / (1 + \exp(\alpha_j - \beta x_t))$$

el modelo de momios proporcionales y

$$\sum_{j=1}^k P_{jt} = 1 - \exp(-\exp(\alpha_j - \beta x_t))$$

el de riesgos proporcionales.

En cuanto a los modelos de respuesta no ordenada, el logit independiente o condicional se expresa como:

$$\ln(P_k(t)/P_j(t)) = \exp(\beta x_t) \quad j=1, \dots, k-1$$

Obviamente los modelos secuenciales cumplen esta propiedad, por

la manera en que se definieron.

Dentro de los modelos de categoría de respuesta no ordenada, el probit politómico, generado por maximización de utilidad no cumple con esta propiedad, de hecho para la estimación de los parámetros es necesario integrar normales multivariadas, es decir se trata de un modelo especial el cual no es posible expresarlo dentro de un contexto general y que naturalmente cuenta con otro tipo de estimación.

Dentro de los modelos del caso multivariado, el probit multivariado se expresa, en el caso particular de dos variables dicotómicas como :

$$\beta_1^t = \Phi^{-1} [P_{1j}^{11}(t) + P_{1j}^{12}(t)]$$

$$\beta_j^t = \Phi^{-1} [P_{1j}^{11}(t) + P_{1j}^{21}(t)]$$

$$\rho = G(P_{1j}^{12}(t) + P_{1j}^{21}(t))$$

donde G es determinada resolviendo para ρ el sistema como función de $P_{1j}^{11}(t)$, $P_{1j}^{21}(t)$ y $P_{1j}^{12}(t)$.

El logit condicional multivariado, en el caso especial de dos variables dicotómicas se puede expresar como

$$\beta_2^t x = \ln(P_{21}^t / P_{22}^t)$$

$$\beta_1^t x = \ln(P_{12}^t / P_{11}^t)$$

$$\beta_{12} = \ln(P_{11}/P_{21}) - \ln(P_{12}/P_{22})$$

El modelo de Grizzle es diferente ya que los parámetros están definidos como

$$\beta_{h1} = G_h [P_1(t), \dots, P_k(t)] \quad h=1, \dots, H,$$

cuando $H=K$, y en el caso que $H < k$ el modelo se define por la expresión anterior y $(K-H)T$ identidades

$$\beta_{it} = P_i(t) \quad i=H+1, \dots, K \quad t=1, 2, \dots, T$$

3.1 Maxima Verosimilitud

El método utilizado o propuesto en la mayoría de los casos es el de máxima verosimilitud. De el modelo general es claro que la verosimilitud se expresa como:

$$L = \prod_{t=1}^T \prod_{k=1}^{K+1} (F_k(t))^{n_k(t)}$$

y por lo tanto el logaritmo de la función de verosimilitud es

$$l = \log(L) = \sum_{t=1}^T \sum_{k=1}^{K+1} n_k(t) \log(F_k(t))$$

maximizar esta función requiere de métodos numéricos.

Existen diferentes rutinas computacionales, para esta optimización que a continuación se discuten.

3.1.1 El Metodo de Newton-Raphson.

Este método está basado en la expansión del logaritmo de la función de verosimilitud alrededor del valor del estimador, w , del parámetro w . La derivada parcial del logaritmo de la función de verosimilitud con respecto al parámetro cumple con la ecuación

$$0 = \frac{\delta l}{\delta w} = \frac{\delta l}{\delta w}(w_1) + (w_2 - w_1) \frac{\delta^2 l}{\delta w^2}(w_1) + v(w_2 - w_1) \frac{\delta^3 l}{\delta w^3}(w_1)$$

para alguna v entre 0,1, donde w es una raíz de la ecuación de verosimilitud y w_1 es una solución inicial. En particular si $v=0$ una aproximación para w es:

$$w_2 = w_1 - \frac{\delta l(w_1)}{\delta w_1} / \frac{\delta^2 l(w_1)}{\delta w_1^2}$$

El valor de w_1 puede ser substituido por w_2 y obtener un nuevo valor w_3 , y así sucesivamente. En forma general comenzando con un valor inicial w_1 se genera una sucesión $\{w_k : k > 1\}$ la cual es determinada sucesivamente por la fórmula

$$w_{k+1} = w_k - \frac{\delta l(w_k)}{\delta w_k} / \frac{\delta^2 l(w_k)}{\delta w_k^2}$$

Este estimador de acuerdo a Kates(1962) es consistente.

3.1.2 El Metodo de Derivadas Fijas (aproximación de Newton)

En este método $\frac{\delta^2 l(w)}{\delta w^2}$ es reemplazado por $-n/a_k$, para evitar reevaluar la segunda derivada, donde $\{a_k\}$ es una sucesión de constantes elegida apropiadamente y n es el tamaño de la muestra. A partir de esta sucesión se genera la sucesión $\{w_k\}$ de la manera:

$$w_{k+1} = w_k + (a_k/n)(\delta l(w_k)/\delta w_k)$$

El utilizar esta sucesión, para una buena elección de $\{a_k\}$ puede ser más estable que el método de Newton-Raphson y converge a w_0 de una manera más regular. Sin embargo si la curva de l es casi vertical en la vecindad de un máximo local, con frecuencia falla la convergencia, pues continua en un ciclo alrededor del máximo relativo w_0 .

Existen casos especiales de este método, como el considerar $a_k = n$ para toda k entonces la sucesión de $\{w_k\}$ que se obtiene es la siguiente

$$w_{k+1} = w_k + (\delta l(w_k)/\delta w_k).$$

O el considerar $a_k = m$, donde m es una constante positiva arbitraria, en este caso el método no va a converger si para w_0 , un máximo local se cumple

$$-m(\delta l(w_0)/\delta w) > 2n.$$

Alternativamente se utiliza el sustituir a_k por $-n$ dividido entre el valor de la segunda derivada evaluada en el punto inicial del proceso.

Por último otro caso especial de este método es el conocido como método de puntajes, debido a Fisher. La sucesión especial sugerida por Fisher es $a_k = 1/I(w_k)$, donde $I(w)$ es la función de información sugerida por Fisher $I(w) = E(-\frac{\partial^2}{\partial w^2} \ln l(w))$. La sucesión generada por el método de puntajes es

$$w_{k+1} = w_k + 1/(nI(w_k)) \delta l(w_k) / \delta w_k$$

3.1.3 El Metodo de Falsas Posiciones.

Con este método se puede tener la seguridad de localizar al máximo relativo. Se comienza con la elección de dos valores de w ,

$w_0^{(1)}$ y $w_0^{(2)}$, tales que $w_0^{(1)} < w_0^{(2)}$ y

$$\frac{\delta l(w_0^{(1)})}{\delta w_0^{(1)}} > 0 \text{ y } \frac{\delta l(w_0^{(2)})}{\delta w_0^{(2)}} < 0$$

definimos

$$\xi = \frac{w_0^{(1)} \frac{\delta l(w_0^{(2)})}{\delta w_0^{(2)}} - w_0^{(2)} \frac{\delta l(w_0^{(1)})}{\delta w_0^{(1)}}}{\frac{\delta l(w_0^{(2)})}{\delta w_0^{(2)}} - \frac{\delta l(w_0^{(1)})}{\delta w_0^{(1)}}}$$

de donde $w_1^{(1)}$, $w_1^{(2)}$ son obtenidos como

$$(1) \quad w_1 = \xi_1 \quad \text{si } \frac{\delta l(\xi_1)}{\delta w} > 0$$

$$(2) \quad w_1 = w_{1-1}$$

o

$$(1) \quad w_1 = w_{1-1}$$

$$(2) \quad w_1 = \xi_1 \quad \text{si } \frac{\delta l(\xi_1)}{\delta w} < 0$$

Con este procedimiento $w_{(1)}$ y $w_{(2)}$ converge en un máximo relativo de l y cuando $w_{(1)} - w_{(2)}$ es suficientemente pequeño el proceso termina y el máximo se obtiene como $(1/2)(w_{(1)} + w_{(2)})$.

No es inmediatamente obvio como deben ser elegidos $w_{(1)}$ y $w_{(2)}$ en esta situación. Una posibilidad es considerar

$$x_j = m - je$$

$$y_j = m + je \quad j=1,2,\dots$$

donde m es algún estimador consistente de la mediana muestral y e es una cantidad pequeña positiva y elegir como $w_{(1)}$ y $w_{(2)}$ el primer par (x_j, y_j) que satisfaga el sistema anterior, sin embargo el método no garantiza que el máximo sea el más cercano a m .

Es muy simple el extender este método para localizar todos los máximos relativos y determinar el estimador máximo verosimil

w. Obviamente no puede ocurrir que un máximo relativo ocurra a una observación menor que el valor de la observación mínima $x_{(1)}$ o exceda a la máxima observación $x_{(n)}$, en consecuencia se tiene que considerar solamente valores iniciales $w_{(1)}$ y $w_{(2)}$ en ese rango.

Comenzando con $w_{(1)} = x_{(1)}$ y $w_{(2)} = x_{(1)} + e$, se utiliza el método para localizar el máximo relativo entre $w_{(1)}$ y $w_{(2)}$. Definimos nuevos $w_{(1)}$ y $w_{(2)}$ por medio de la relación

$$w_{(1)}^{(i)} = w_{(1)}^{(i-1)}$$

$$w_{(2)}^{(i)} = w_{(2)}^{(i-1)} + e.$$

y continuamos hasta que se cumpla $w_{(2)}^{(i)} > x_{(n)}$.

Invitablemente existe el riesgo de que el intervalo de longitud e contenga más de un máximo relativo. en ese caso sólo se detecta el primero. lo cual desde el punto de vista práctico no es un problema serio, ya que podemos elegir e suficientemente pequeño para asegurar que existe una probabilidad negligible de que esto ocurra. Además en el caso de no poder evitar por completo esta dificultad, la rara ocurrencia de ello, no ocasiona un sesgo apreciable en el estimador máximo verosímil, ya que dos máximos relativos a una distancia e (donde e es pequeña) deben tener valores asociados de l muy cercanos en valor.

3.1.4 El Metodo de Casi-verosimilitud.

Esencialmente este método es equivalente al método de mínimos cuadrados ponderados iterativos en el caso especial en que los pesos dependen solamente de los estimadores actuales de los parámetros de regresión.

El método en términos generales es el siguiente:

Suponemos Y un vector respuesta N -dimensional tal que:

$$E(Y) = \mu \quad \text{y} \quad V(Y) = \sigma^2 V(\mu)$$

donde $V(\mu)$ es una matriz positiva semidefinida cuyos elementos son funciones conocidas de μ .

Además $E(Y)$ puede ser expresado en términos de una ecuación de regresión

$$\mu = \mu(\beta)$$

donde β es un vector de parámetros y $\mu(\cdot)$ es un vector de funciones conocidas con terceras derivadas acotadas.

Para asegurar identificabilidad se supone que la matriz de derivadas $D = \delta\mu / \delta\beta$ tiene rango p para toda β , lo cual en el caso de que $\mu(\beta)$ involucre una matriz diseño X , como sucede para modelos lineales generalizados, es equivalente a considerar que la matriz tiene rango completo.

La función log-casi-verosimil es considerada inicialmente como una función de μ (de β) y esta definida por el sistema de

ecuaciones diferenciales parciales

$$\frac{\delta l(\mu, y)}{\delta \mu} = V^{-1}(y - \mu)$$

donde $V^{-1}(\mu)$ es una inversa generalizada de $V(\mu)$.

Usualmente es innecesario resolver este sistema de ecuaciones, sin embargo cuando una solución es requerida es suficiente encontrar funciones $\mu(\cdot)$ y $\beta(\cdot)$ que satisfagan :

$$l(\mu, y) = \sigma^{-2} \{ y^t \theta - \beta(\theta) - c(y, \sigma) \}$$

donde $c(y, \sigma)$ es arbitraria.

Las ecuaciones máximo casi-verosimiles para determinar el valor de :

$$\frac{\partial l(\beta, y)}{\partial \beta} = 0$$

pueden ser escritas como

$$D^t V^{-1}(y - \mu(\beta)) = 0$$

Estas ecuaciones reciben el nombre de mínimos cuadrados primeramente por la interpretación geométrica, la cual involucra proyecciones sucesivas del vector residual $y - \mu(\beta)$ en el espacio tangente de soluciones locales $\mu(\beta)$, donde β_0 es el estimador actual de β . Estas ecuaciones no dependen de σ , por lo tanto el

valor numerico es el mismo siendo σ conocida o no.

El metodo de Newton-Raphson con la matriz de segundas derivadas reemplazado por su valor esperado $D^t V D$, se convierte en

$$\hat{\beta}_1 - \beta_0 = (D^t V D)^{-1} D^t V (y - \mu_0)$$

donde el lado derecho de la ecuacion esta calculado en el valor β_0 .

Es claro que $\hat{\beta}$ puede no coincidir con el estimador maximo-verosimil ya que la funcion de verosimilitud correcta es generalmente diferente de $l(\mu, y)$. Ademas, es importante enfatizar que las propiedades de los estimadores casi-verosimiles no son en general las propiedades de todos los estimadores minimos cuadrados ponderados.

Las propiedades estadisticas de las funciones casi-verosimiles, son muy similares a las de las funciones de verosimilitud ordinarias, excepto que el parametro de estorbo (nuisance) es tratado separadamente de β cuando es desconocido y no es estimado por minimos cuadrados ponderados.

Algunas propiedades del estimador casi-verosimil, bajo la suposicion de que $\mu(\beta)$ tiene derivadas de tercer orden acotadas y que $N^{-1} D^t V^{-1} D$ tiene como limite una matriz positiva definida y los terceros momentos de Y son finitos son:

$$E(\hat{\beta} - \beta) = O(n^{-1})$$

$$n^{1/2} (\hat{\beta} - \beta) \sim N(0, n^{-1} (D' V D)^{-1}) + O_p(n^{-3/2})$$

Además si denotamos por $l(\hat{\beta}_0, y)$ y $l(\hat{\beta}_a, y)$, los máximos de la función log-casi-verosimilitud bajo H_0 y H_a respectivamente los espacios parametrales son de dimension q bajo H_0 y $p > q$ bajo H_a y están anidados, tenemos bajo H_0 que

$$2l(\hat{\beta}_0, y) - 2l(\hat{\beta}_a, y) \sim \sigma^2 X' X + O_p(n^{-1/2})$$

Para detalle de las demostraciones ver Mc. Cullagh(1983).

En el caso de modelos lineales generalizados la matriz de derivadas D es el producto de una matriz diagonal con $\{d\mu/d\eta\}$ en la diagonal y la matriz modelo X ; más aún W es una matriz diagonal de pesos cuyos elementos son $(d^2\mu/d\eta^2)/v$ y se define a z como la variable dependiente ajustada cuyos elementos estan dados por

$$z_i = \eta_i + (y_i - \mu_i) \left(\frac{d\eta_i}{d\mu_i} \right)$$

lo cual puede escribirse en forma computacionalmente más conveniente como

$$(X'WX)^{-1} X'WZ$$

donde W y Z son calculadas utilizando el estimador actual de β_0 , z es conocida como variable de trabajo dependiente.

En el caso de modelos lineales generalizados de liga compuesta, los cuales pueden ser utilizados para ajustar algunos de los modelos anteriores, Thomphson y Baker(1981) demostraron que una solución máximo casi-verosímil puede ser obtenida considerando como variables de trabajo dependiente e independiente

$$z_i = CH_i \eta_i + (y_i - \mu_i) \quad \text{y} \quad X = CHX.$$

donde H es una matriz diagonal con elementos $\delta \gamma / \delta \eta_k$ y con W una matriz diagonal de pesos $W_i = 1/t_i$ para z_i .

Ahora para cualquier solución

$$X^t WX = X^t Wz$$

tenemos que para cualquier matriz J no singular

$$(JX)^t (J^{-1} WJ^{-1}) (JX) = (JX)^t (J^{-1} WJ^{-1}).$$

Entonces si J es cualquier matriz diagonal podemos utilizar las variables de trabajo

$$z = JCH \eta + J(y - \mu)$$

$$X = JCHX \text{ y}$$

$$W = J^{-1} WJ^{-1}.$$

Una elección posible para J es aquella que deja inalterado un vector unitario en la matriz diseño, X , en la matriz X . Si elegimos $J = 1/\sum_{i,k} C_{ik} H_{kk}$ no vamos a tener problemas con la exclusión de la media general del modelo.

3.1.5 Escalamiento Iterativo Generalizado

Otro método de estimación aplicable a máxima verosimilitud es el de escalamiento iterativo generalizado.

El método de escalamiento iterativo generalizado es un método para encontrar una función de probabilidad de la forma

$$p_i = \pi_i \prod_{s=1}^d \mu_s^{b_{si}}$$

que satisfaga la restricción

$$\sum_{i \in I} b_{si} p_i = k_s, \quad s=1, \dots, d \quad \sum_{i \in I} p_i = 1$$

donde I es un conjunto finito y $P = \{p_i : i \in I, p_i \geq 0, \sum_{i \in I} p_i = 1\}$ es una función de probabilidad en I . Además $\pi = \{\pi_i : i \in I, \pi_i > 0, \sum_{i \in I} \pi_i < 1\}$ es una función de sub-probabilidad en I .

Para toda s existe i tal que $b_{si} \neq 0$. Las constantes π_i y b_{si} son dadas y μ_s son desconocidas. Este modelo es llamado log-lineal debido a que $\log(p_i / \pi_i) = \log \mu_s + \sum_{s=1}^d b_{si} \log \mu_s$.

La manera en que se relacionan el modelo y las restricciones es primero a través de propiedades de la función de información

discriminatoria y segundo a través de propiedades de estimación máximo verosímil.

La función de información discriminante esta definida como:

$$K[p, \pi] = \sum_{i \in I} p_i \log(p_i / \pi_i)$$

($0 \log 0 = 0$ por definición).

Además $K[p, \pi] = 0$ si y solamente si $p = \pi$.

Si definimos $\sigma = \sum_{i \in I} \pi_i$, entonces

$$K[p, \pi] = K[p, (1/\sigma)\pi] - \log \sigma,$$

lo cual es claramente no negativo e igual a cero si y solo si $p = \sigma^{-1}$ y $\sigma = 1$ i.e. $p = \pi$.

Esta función cumple además que una función de probabilidad de la forma log-lineal satisface las restricciones, minimiza $K[p, \pi]$ y es única. La demostración se anexa al final.

Las restricciones son consistentes si el conjunto de las funciones de probabilidad que las satisface es no vacío, y si las restricciones son consistentes podemos expresarlas como:

$$p_i = \pi_i \prod_{r=1}^c \lambda_{ir}^{a_r} \quad r=1,2,\dots$$

$$\sum_{i \in I} a_{ir} p_i = h_r$$

donde

$a_{ri} \geq 0$, $\sum_{r=1}^c a_{ri} = 1$, $h_r > 0$, $\sum_{r=1}^c h_r = 1$. La prueba se anexa al final.

Para expresar el modelo log-lineal en esta forma definimos

$$(0) \\ p_i =$$

$$p_i^{(n+1)} = p_i^{(n)} \prod_{r=1}^c \left(\frac{h_r^{(n)}}{h_r} \right)^{a_{ri}^{(n)}} \quad n=0,1,\dots$$

$$\text{donde } h_r^{(n)} = \sum_{i \in I} a_{ri} p_i^{(n)}$$

Siendo que las restricciones son consistentes la sucesión converge a una solución de

$$p_i = \prod_{r=1}^c h_r^{a_{ri}}$$

la cual es única y positiva.

Aplicación a la estimación máximo verosímil.

Si I es un conjunto finito y $\{p_i\}$ es una función de probabilidad desconocida en I de la forma log-lineal, dada la muestra de tamaño N con frecuencias positivas $\{f_i : i \in I, f_i > 0, \sum_{i \in I} f_i = N\}$, el estimador máximo verosímil $\{\hat{p}_i\}$ de $\{p_i\}$ es la única función de probabilidad de la forma log-lineal que satisface

$$\sum_{i \in I} b_{si} p_i = \sum_{i \in I} b_{si} (f/N) \quad s=1, \dots, d$$

con $\sum_{i \in I} D_i = 1$, por lo tanto $\{\beta\}$ puede ser obtenido por escalamiento iterativo.

En primer lugar las restricciones son consistentes ya que se satisfacen cuando $p_i = f/N$.

3.2 Estimadores de Minima Ji-cuadrada

Si definimos $P^k(t) = n(t)N^{-1}$ para $k=1, 2, \dots, K$ y $t=1, 2, \dots, T$.
El modelo general se puede expresar aproximadamente como:

$$G [P^1(t), P^2(t), \dots, P^K(t)] = \beta x_{ht} + \sum_{k=1}^K G^k(t) [P^k - P^k(t)].$$

Entonces el estimador de mínima Ji-cuadrada de se puede definir como el estimador de mínimos cuadrados generalizados de esta expresión. Ya que esta expresión es de la forma $Y = X + u$, donde $E u = 0$ y $E u u^t = R$.

Ahora en esta expresión $u = \sum_{k=1}^K G^k(t) [P^k - P^k(t)]$ y $E u u^t = E(\sum_{k=1}^K G^k(t) [P^k - P^k(t)])$.

Si definimos $P = (P^1, \dots, P^K)$, similarmente P y $G(t)$ la matriz de dimensión $H \times K$ cuyo k -ésimo elemento en la diagonal es igual a

$$G^k(t) \text{ y } A = N E(P-P)(P-P)^t = D(P) - PP^t.$$

$$\text{Entonces } E u u^t = (1/N) [G(t) A G(t)]^{-1}.$$

Por lo que es posible aplicar mínimos cuadrados generalizados.

Este método es sugerido en particular por Gurland et. al., los cuales dan los valores de los pesos para el logit y normit politómicos.

3.3 Otros Metodos de estimacion

Otro método de estimación es propuesto por Mc. Fadden y Domenech (1975) y publicado por Amemiya en (1977), el método es conocido como la estimación en dos pasos para el modelo logit multivariado. En particular en el logit bivariado, el cual está definido como:

$$P[y_1(t) = i, y_2(t) = j] = d_{ij}^{-1} \exp[\alpha x_{ij}(t) + \beta z_i(t)]$$

$$\text{donde } d_{ij} = \exp[\alpha x_{ij}(t) + \beta z_i(t)].$$

La función de verosimilitud esta definida por

$$L = \prod_{ij} \prod_{t} \frac{u_{ij}(t)}{d_{ij}(t)}$$

donde $u_{ij}(t) = 1$ si y sólo si $y_1(t) = i$ y $y_2(t) = j$.

Si y_1, y_2 son estocásticamente independientes, entonces pueden ser estimados separadamente considerando cada probabilidad marginal, sin embargo el caso de independencia es difícil que sea realista en muchas ocasiones. Suponiendo independencia (es decir αx_{ij}) puede ser escrito como αx_j

$$P\{y_2 = j \mid y_1 = i\} = \frac{\exp(\alpha x_{ij}^t)}{\sum_j \exp(\alpha x_{ij}^t)}$$

6

$$L = L_1(\alpha) L_2(\alpha, B)$$

donde $L_1 = \prod_{i=1}^t \prod_{j=1}^u P(y_2 = j \mid y_1 = i)$ y $L_2 = \prod_{i=1}^t \prod_{j=1}^u P(y_1 = i)$

El primer paso del método es estimar α , como el valor que minimiza a $L_1(\alpha)$, el cual se conoce como el estimador máximo verosímil condicional.

El segundo paso es estimar B como el valor que minimiza $L_2(\alpha, B)$.

Un estimador eficiente sólo se obtiene en el caso que $\alpha x_{ij}^t = \sigma x_j^t$.

Por otro lado este procedimiento puede extenderse fácilmente a un método de estimación da varios pasos para el proceso que involucre más de dos variables, sin embargo en este caso el número de funciones ue hay que minimizar crece y puede no ser ya tan conveniente.

3.4 Analisis de Residuales y Bondad de Ajuste

El análisis de residuales para modelos de respuesta multinomial es complicado por varias razones :

1. Los residuales no obstante de ser estandarizados, toman uno de un número limitado de posibles valores, lo cual causa problemas cuando el número de las celdas es pequeño;
2. En general no es claro que los residuales de las celdas sean las cantidades más relevantes para examinar, lo que es importante en los datos no es el número en la celda per se, sino como el conteo por celda o grupos de celdas varía relativo a otra celda o grupo de celdas.

Para datos binarios usariamos normalmente un residual asociado con 2 celdas o grupos de celdas, dado que los residuales están negativamente correlacionados en pares, solamente un residual es necesario. Para observaciones de datos binarios una observación es la proporción de éxitos o equivalentemente el número de éxitos relativo al número de fallas o fracasos. El residual estandarizado usual es $(p - p) / ((pq/n)^{1/2})$, lo cual es la raíz cuadrada de la contribución a la χ^2 -Pearson, un residual analogo puede ser definido utilizando la contribución al cociente de verosimilitud.

Mc. Cullagh(1980) propone utilizar como residual la

contribución a la estadística de cociente de verosimilitud. Este residual es siempre positivo y no indica la dirección de desvío de los valores observados de los ajustados.

En la gran mayoría de estos modelos, para medir la bondad de ajuste es posible utilizar como estadística la formada por el logaritmo del cociente de verosimilitudes llamada Devianza. La discrepancia de un ajuste es proporcional a 2 veces la diferencia entre el máximo log-verosimilitud posible y el conseguido por el modelo bajo estudio. Esta estadística tiene una distribución asintótica χ^2 con $N-p$ grados de libertad, donde N es el número de observaciones y p el número de parámetros.

4. APLICACIONES

En este capítulo se presentarán aplicaciones de algunos de los modelos descritos anteriormente. Estas aplicaciones requirieron el desarrollo de rutinas computacionales para llevarse a cabo, las cuales se encuentran disponibles para su futuro uso.

4.1 Aplicación en Modelos de Categoría de Respuesta Ordenada

Para los modelos de tipo umbral, los cuales puede considerarse que generalizan a los modelos de categorías ordenadas, se desarrolló una pequeña biblioteca de MACROS para su uso en el paquete GLIM (Generalized Linear Interactive Modelling), [Baker y Nelder, 1978].

Un macro es un conjunto de instrucciones de GLIM, definido por el usuario e identificado con un nombre. La manera de especificarlo es

```
$MACRO nombre de identificador instrucciones $ENDMAC
```

Las instrucciones pueden ser cualquiera de las permitidas en GLIM excepto \$MACRO, \$RETURN, \$FINISH, \$FORMAT y \$SUBFILE. Un espacio debe seguir al nombre del macro antes de cualquier instrucción. Para detalles sobre el funcionamiento de GLIM ver Baker & Nelder (1978) ó Ruiz Velasco (1982).

La manera de llevar a cabo estos ajustes es reescribir los modelos como modelos lineales generalizados de liga compuesta. Para la estimación de caso particulares de este tipo de modelos

Thompson y Baker (1981) propusieron una manera de hacerlo en GLIM, la cual presenta dificultades para su generalización.

Rogers(1983) propone una manera más general, la cual permite ajustar casi cualquier modelo de este tipo, utilizando también GLIM, esto se hace escribiendo cada componente del vector de esperanzas μ_i ($= \sum_{k=1}^p c_{ik}$) como $\mu_i = \sum_{k=1}^p \gamma_{ik} c_{ik}$ donde $p \leq p$ y $c_{ik} \neq 0$, es decir cada valor ajustado es una suma ponderada (por c_{ik}) de sus propias partes γ_{ik} , es claro que en algunos casos, p es igual a uno, en particular un modelo lineal generalizado es un modelo de liga compuesta con $p = 1 \forall i$. La manera de especificar esto a GLIM es como sigue:

1. Si el i -ésimo dato consta de un modelo con un sólo elemento γ_{i1} ambos el dato y el modelo γ_{i1} utilizan para su especificación una unidad en exactamente la misma manera que se haría normalmente en GLIM.
2. Si para el i -ésimo dato se tiene que $\mu_i = c_{i1} \gamma_{i1} + \dots + c_{ip} \gamma_{ip}$, el dato μ_i se especifica en una unidad y el modelo en las siguientes p unidades.

Para definir esta estructura se utiliza la variable auxiliar PT, la cual toma el valor 0 cuando la entrada es dato y el valor j para la j -ésima entrada de término del modelo. Cuando en una sola unidad (entrada) haya dato y término ($p=1$) PT toma el valor de cero.

El ajuste se lleva a cabo considerando cada dato, digamos el

i -ésimo, como p pseudodatos y el valor ajustado del i -ésimo dato será la suma \sum_i del valor ajustado de los p pseudodatos. El procedimiento de Rogers, consta de una parte general que consiste en identificar por medio de la variable PT esta estructura, es decir identifica los datos y el modelo, creando variables auxiliares (vectores), una de las cuales sirve como indicador para determinar que elementos corresponden a modelo unicamente y así asignarles peso cero en el ajuste y la otra contiene el número de dato para el cual es este modelo. Dentro de esta parte general se inicializa el predictor lineal (%LP).

Para especificar de que modelo se trata es necesario definir 5 macros, los tres primeros y el quinto de estos macros, deberán ser llamados M1, M2, M3 y M5 respectivamente, ya que son utilizados en otros macros especificados en la parte general. Por comodidad al cuarto de estos macros se le identifica como M4.

Lo que cada uno de estos macros deberá de contener es:

M1.- La expresión para los valores ajustados sin tomar en cuenta la matriz C; i.e.

$$\%FV = h(\eta), \quad \%FV = h(\%LP)$$

adicionalmente hay un macro, N1, que construye $\%FV = Ch(\eta)$, el cual es general para todos los modelos y no hay que especificarlo.

M2.- La expresión para la derivada

$$\%DR = 1 / (h(\%LP) / \%LP)$$

el cual es utilizado en el macro, N2, para construir

$$\%DR = 1 / C_{k ik} \left(\frac{\partial \gamma}{\partial \eta} \right)_{k k} (=J_{ii}) ,$$

$$\%LP = J_{iik} C_{ik} \left(\frac{\partial \gamma}{\partial \eta} \right)_{k k} y$$

$$\begin{matrix} * \\ X \\ ii \end{matrix}$$

M3.- Contiene la expresión de la varianza de

$$\%VA = W^{-1} \text{ donde } W_{ii} = 1/t_{ii}$$

este macro es utilizado en otro que asigna el valor arriba calculado se si trata de un dato y uno si se trata de modelo para evitar una posible varianza negativa.

M4.- Contiene el incremento de la función devianza por cada dato.

M5.- La expresión del predictor lineal

$$\%LP = g(\%FV)$$

Esto puede quedar más claro con algunos ejemplos:

4.1.1 Modelos de Mc. Cullagh

La manera de reescribir los modelos de Mc. Cullagh como modelos de liga compuesta es la siguiente:

El modelo de momios proporcionales, el cual esta expresado como

$$\log\left(\frac{y_{ji}}{1-y_{ji}}\right) = \alpha_j - \beta_j x_i$$

se puede expresar alternativamente como

$$y_{ji} = \frac{\exp(\alpha_j - \beta_j x_i)}{1 + \exp(\alpha_j + \beta_j x_i)}$$

donde $y_{ji} = p_{1i} + \dots + p_{ji}$ y p_{ki} es la probabilidad de que la observación pertenezca a la categoría k cuando la covariable toma el valor x_i .

Para expresar este modelo como un modelo de liga compuesta definimos

$$h(\eta) = \frac{\exp(\eta)}{(1 + \exp(\eta))} \quad \text{donde } \eta = X \underset{1 \ 1}{B}$$

y

$$X = \underset{1}{\begin{matrix} & k \\ \begin{matrix} 1 & 0 & 0 & \dots & 0 & x \\ 1 & 1 & 0 & \dots & 0 & x \\ \cdot & & & & \cdot & \cdot \\ \cdot & & & & \cdot & \cdot \\ 1 & 1 & 1 & \dots & 1 & x \\ 1 & 0 & 0 & \dots & 0 & x \\ \cdot & & & & \cdot & \cdot \\ \vdots & & & & \cdot & \cdot \\ i & 1 & 1 & \dots & 1 & x \end{matrix} \end{matrix}} \underset{m}{\begin{matrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \\ 2 \\ \cdot \\ \cdot \\ m \end{matrix}}$$

m es el número de muestras o subpoblaciones

C es una matriz diagonal por bloques con elementos en la diagonal de la forma $N E_j$, $j=1, \dots, m$, en donde E_j es una matriz cuadrada de orden $k \times k$, con unos en la diagonal y menos unos en la subdiagonal inferior.

$$y = \begin{matrix} \Omega_1 \\ \Omega_2 \\ \cdot \\ \Omega_k \end{matrix} \Bigg| \underset{B}{\quad}$$

$$\text{donde } \Omega_1 = \alpha_1, \Omega_2 = \alpha_1 + \alpha_2, \dots, \Omega_k = \sum_{i=1}^k \alpha_i.$$

Es claro que el estimar α_k es redundante, sin embargo es necesario especificar el modelo de esta manera, ya que el modelo que estamos ajustando no es un modelo de liga compuesta, sino que estamos fabricando su estructura como tal.

Los macros necesarios en este caso son :

$$\$M M1 \$CAL \%FV = \%EXP(\%LP)/(1 + \%EXP(\%LP)) \$E$$

$$\$M M2 \$CAL \%DR = (1 + \%EXP(\%LP)**2)/\%EXP(\%LP) \$E$$

$$\$M M3 \$CAL \%VA = \%FV \$E$$

$$\$M M4 \$CAL \%DI = 2*(\%YV*\%LOG(\%YV/\%FV) - (\%YV - \%FV)) \$E$$

$$\$M M5 \$CAL \%LP = \%LOG((\%FV + .5)/(1.5 - \%FV)) \$E$$

Aquí se considera que nuestro modelo es un modelo de respuesta multinomial, esto es equivalente, bajo algunas condiciones, a suponer que el componente aleatorio de nuestro modelo tiene una distribución Poisson (ver, Nelder y Mc. Cullagh 1983 pp 142).

En el caso del modelo de riesgos proporcionales el modelo es

$$-\log(1 - \gamma_{ji}) = \exp(\alpha_j - \beta_j x_{j1}^t),$$

que equivalentemente se puede expresar como:

$$\gamma_{ji} = 1 - \exp\{-\exp(\alpha_j - \beta_j x_{j1}^t)\}.$$

Entonces este modelo también se puede expresar como un modelo lineal generalizado de liga compuesta definiendo $h(\eta) = 1 - \exp(-\exp(\eta))$, donde $\eta = X\beta$, $X\beta$ iguales al caso anterior, C una matriz diagonal por bloques exactamente la misma que en el caso anterior.

Los macros que especifican este modelo son:

```
$M M1 $CAL %FV = 1 - %EXP(-%EXP(%LP)) $E
```

```
$M M2 $CAL %DR = 1/(%EXP(%LP)*%EXP(-%EXP(%LP))) $E
```

```
$M M3 $CAL %VA = %FV $E
```

```
$M M4 $CAL %DI = 2*(%YV*%LOG(%YV/%FV) - (%YV-%FV)) $E
```

```
$M M5 $CAL %LP = %LOG(-%LOG(.5 - %FV)) $E
```

Al igual que en el caso anterior estamos considerando un error Poisson por el tipo de datos que estamos trabajando.

4.1.2 Probit Político.

Este tipo de ajuste es posible llevarlo a cabo en modelos para probit político, en donde no suponemos que las categorías tienen fronteras definidas de antemano como es el caso de los modelos propuestos por Ashford. Al igual que los modelos anteriores, este tipo de modelos con o sin frontera definida de antemano puede expresarse como un modelo lineal generalizado de liga compuesta.

El proceso para ajustarlos es el mismo, con la excepción de que la inicialización es diferente debido a que la distribución de la

variable subyacente es simétrica y el procedimiento sugerido por Rogers no funciona bien en estos casos por la manera de llevar a cabo la inicialización, por este motivo se desarrollaron dos macros de inicialización para utilizarse en lugar de los sugeridos por Rogers, pero conservando la ideología de su procedimiento.

Los macros que debe definir el usuario en este caso son:

\$M M1 \$CAL %FV = %NP(%LP) \$E

\$M M2 \$CAL %DR = %SQRT(2*%PI)/%EXP(-(%LP**2)/2) \$E

\$M M3 \$CAL %VA = %FV \$E

\$M M4 \$CAL %DI = 2*(%YV*%LOG(%YV/%FV)-(%YV-%FV)) \$E

\$M M5 \$CAL %LP = %ND(%FV) \$E

Donde %NP y %ND representan la función de distribución normal y la función de distribución normal inversa respectivamente.

La definición de estos macros se encuentra ya dentro de la librería de macros, de tal forma que al ajustar un modelo unicamente es necesario llamar al subarchivo (subfile) que contine la definición de los macros. Si se desea ajustar otro tipo de modelo umbral es necesario definir estos cinco macros.

4.1.3 Primer Ejemplo

Los datos en este ejemplo corresponden a un estudio del Med. Mauricio Hernández A. llevado a cabo en la región del Chichonal.

Las variables en el estudio se midieron en 113 niños, todos menores de cinco años, las variables en el ejemplo corresponden a

1. Estado Nutricional .- la cual tiene tres categorías ordenadas: bueno, regular y malo, que se determinaron de acuerdo a la talla, peso y circunferencia del niño.
2. Educación de madre .- la cual tiene dos categorías: sabe leer y no sabe leer, es claro que las categorías guardan un orden, sin embargo como la variable es dicotómica el orden no importa.
3. Nivel Socioeconómico .- la cual tiene tres categorías: alto, medio y bajo, obviamente las categorías también guardan un orden y mientras mejor sea el nivel socioeconómico, mejor alimentados van a estar los hijos. Esta variable está determinada por el ingreso familiar.

Los datos son los siguientes:

Estado Nutricional

Nivel socio- economico	educ. de la madre	Malo	Regular	Bueno	Total
bajo	no sabe leer	3	5	0	8
	sabe leer	9	2	0	11
medio	no sabe leer	12	23	5	40
	sabe leer	8	11	2	21
alto	no sabe leer	1	15	9	25
	sabe leer	2	4	2	8
		35	60	18	113

A estos datos se le ajustaron el modelo de momios proporcionales (Mc. Cullagh 1980) y el de probit politómico (Gurland, et. al. 1960). En ambos casos se consideraron como variables explicativas

1. Nivel socioeconómico
2. Educación de la madre
3. Nivel socioeconómico y educación de la madre
4. Nivel socioeconómico, educación de la madre y la interacción entre ambas.

En el primer caso se mostraran todas la instrucciones para el ajuste.

Si tomamos en cuenta unicamente nivel-socioeconómico, la tabla de datos es la siguiente

Estado Nutricional

Nivel socio-económico	malo	regular	bueno	total
bajo	12	7	0	19
medio	20	34	7	61
alto	3	19	11	53
	35	60	18	113

El modelo que se analizará con detalle es

$$\log\left\{\frac{P_{ji}}{(1 - P_{ji})}\right\} = \alpha_j + \beta x_i$$

Por motivos numéricas la celda vacía se substituye por .5, pero sin afectar el marginal, es decir restando .5 de otra(s) celda(s). Esto es debido a problemas de frontera, ya que si no observamos individuos en alguna categoría los umbrales de las dos categorías serían muy semejantes o deberían al menos serlo para esa subpoblación.

En el caso de estos modelos existe un archivo que cuenta con las definiciones de MACROS de inicialización, así como las de ajuste, definidas en subarchivos. Entonces después de ejecutar GLIM habiendo declarado algún archivo de entrada el archivo (ISSB)GLIM/GENERAL ON IIMAS. Por ejemplo R*SERVICIO/GLIM;FILE FILE1(DISK,TITLE=(ISSB)GLIM/GENERAL ON IIMAS,FILETYPE=7)

Las instrucciones en GLIM para el ajuste del modelo de momios proporcionales son:

```

$INP 1 INICIA MOMIOS$
$COMMENT CON ESTA INSTRUCCION LEE DEL ARCHIVO 1 LA SUBROUTINAS
$COMMENT INICIA Y MOMIOS, DONDE INICIA PROCESA LA INFORMACION
$COMMENT EN PT Y MOMIOS CONTIENE LAS 5 MACROS NECESARIOS PARA
$COMMENT AJUSTAR EL MODELO.
$DATA Y PT C X1 X2 X3 NI$

```

```
$READ
```

```

11.5 0 20 1 0 0 1
7 0 0 0 0 0 0
0 1 20 1 1 0 1
0 2 -20 1 0 0 1
.5 0 0 0 0 0 0
0 1 20 1 1 1 1
0 2 -20 1 1 0 1
20 0 61 1 0 0 2
34 0 0 0 0 0 0
0 1 61 1 1 0 2
0 2 -61 1 0 0 2
7 0 0 0 0 0 0
0 1 61 1 1 1 2
0 2 -61 1 1 0 2
3 0 33 1 0 0 3
19 0 0 0 0 0 0
0 1 33 1 1 0 3
0 2 -33 1 0 0 3
11 0 0 0 0 0 0
0 1 33 1 1 1 3
0 2 -33 1 1 0 3

```

```
$YVAR Y$
```

```
$OWN N1 N2 N3 M4$
```

```
$USE STAR$
```

```
$COMMENT STAR Y INIT MACROS DE INICIALIZACION DENTRO DE INICIA
```

```
$ARG N2 X1 X2 X3 NI$CAL %X=4$
```

```
$COMMENT %X ES IGUAL AL NUMERO DE VARIABLES EXPLICATIVAS$
```

```
$USE INIT$
```

```
$FIT -%GM + X1 + X2 + X3 + NI$
```

```
$COMMENT AJUSTE DEL MODELO$
```

```
$DISP EW$
```

```
$COMMENT IMPRIME LOS ESTIMADORES Y LOS VALORES AJUSTADOS$
```

```
$STOP$
```

BIBLIOTECA
 JUAN A. ESCALANTE H.
 UNIDAD ACADEMICA DE
 LOS CICLOS PROFESIONAL
 Y DE POSGRADO / CCH
 UNAM

El parámetro que tiene interés para nosotros es el

correspondiente a NI debido a que mide el cambio en los momios debido a los diferentes niveles socioeconómicos, la conclusión que se puede obtener a partir de ese parámetro es que 2.9125 (= $\exp(1.069)$) veces los momios de tener mala nutrición son mayores para nivel bajo que medio y para medio que alto.

También se puede notar que el estimador correspondiente a X_3 (β_3) tiene una desviación estándar muy grande (alrededor de 8 veces su valor), esto es debido a que incluye la estimación de α , la cual habíamos supuesto . sin embargo como ya se había visto es necesario ajustar. Algo que es importante señalar es que en el caso de cambiar la inicialización, el valor de este parámetro cambia, al igual que el de todos los demás parámetros de estorbo, sin embargo los parámetros de interés, así como los valores ajustados conservan sus valores.

Los resultados obtenidos en el ajuste del modelo completo son :

$$\exp(\beta_1) = 4.7891$$

$$\exp(\beta_2) = 3.0509$$

$$\exp(\beta_3) = .8320$$

la interpretación considerando el modelo completo, ya no es tan sencilla como en el caso de una sola variable explicativa.

Ahora en el caso de ajustar un modelo probit politómico, a los parámetros obtenidos se les aplicó la transformación $g(\beta) = (\beta)/(1$

- B), para poder comparar los dos ajustes (probit politómico y modelos proporcionales).

En el caso del modelo completo los resultados son :

$$g(B)_1 = 4.4593$$

$$g(B)_2 = 3.0741$$

$$g(B)_3 = .8283.$$

Nota.- en el caso del modelo de Probit Politómico, el modelo es:

$$P_{ji} = (\alpha_j + Bx_i) - (\alpha_{j-1} + Bx_i) \quad j=2, \dots, k-1$$

$$P_{1i} = (\alpha_1 + Bx_i)$$

$$P_{ki} = 1 - (\alpha_{k-1} + Bx_i)$$

en realidad, debido a la forma en que GLIM maneja la función acumulativa normal se ajusta:

$$P_{ki} = .9995 - (\alpha_{k-1} + Bx_i)$$

4.2 Ajuste en el caso de categorías no ordenadas

El modelo logístico no ordenado, es posible ajustarlo utilizando el método propuesto por Goldstein (1980), el cual

consiste en expresar este modelo como uno log-lineal.

En el caso de dos subpoblaciones o muestras, un modelo recupera a otro, ya que el modelo logístico

$$\log \frac{P}{1-P_j} = \alpha + \beta_j$$

es posible generarlo a partir del modelo log-lineal para una tabla de $2 \times k$ con P_j la probabilidad de que la j -ésima columna de una observación caiga en el primer renglón de la tabla, $(1 - P_j)$ es entonces la probabilidad de que la j -ésima columna caiga en el segundo renglón de la tabla).

Si expresamos el modelo log-lineal como:

$$\log(m_{ij}) = \mu + \rho_i + \gamma_j + \zeta_{ij} \quad i=1,2 \quad j=1,\dots,k$$

entonces

$$\log \frac{P}{1-P_j} = \log \mu_{1j} - \log \mu_{2j}$$

$$= (\rho_1 - \rho_2) + (\zeta_{1j} - \zeta_{2j})$$

por lo que

$$\alpha = (\rho_1 - \rho_2)$$

$$B = \left(\begin{array}{cc} \zeta & -\zeta \\ j & 1j \quad 2j \end{array} \right)$$

La extensión al caso politómico, parte de considerar el modelo log-lineal para una tabla de $m \times k$.

$$\log(m_{ij}) = \mu + \rho_i + \gamma_j + \zeta_{ij} \quad i=1, \dots, m$$

6

$$\begin{aligned} m_{ij} &= \exp(\mu + \rho_i + \gamma_j + \zeta_{ij}) \\ &= \exp(\mu + \gamma_j) \exp(\rho_i + \zeta_{ij}). \end{aligned}$$

El vector $p_{ij} = m_{ij} / m_{.j}$, $i=1, \dots, m$ ($p_{i1} = 1$), es considerado como el vector de probabilidades respuesta de las m categorías para cada valor de j .

$$p_{ij} = \frac{\exp(\rho_i + \zeta_{ij})}{\sum_i \exp(\rho_i + \zeta_{ij})}$$

Entonces si ajustamos un modelo log-lineal y tomamos en cuenta únicamente los parámetros que intervienen en la definición de p_{ij} , vamos a tener los parámetros de un modelo logístico politómico.

La forma de ajustar estos modelos en GLIM quizá sea más conveniente verla con un ejemplo. El ejemplo que se presenta corresponde a un caso bivariado con variables dicotómicas, que se

va a trabajar como un modelo de respuesta politómica, con 4 categorías no ordenadas. El objetivo es ejemplificar además cómo, por medio de restricciones en los parámetros, se puede recuperar el modelo de Grizzle o equivalentemente el modelo de Probit Multivariado de Ashford y Sowden (1976).

Los datos corresponden a un estudio del Med. Mauricio Hernández, se trata de mediciones realizadas en 700 niños menores de 10 años en un campamento de Chiapas, el cual abriga niños mexicanos y guatemaltecos. A estos niños se les midió la presencia o ausencia de ciertos signos clínicos, en este ejemplo tomaremos únicamente dos signos, anemia y atrofia muscular, tomando en cuenta 7 grupos de edades. Los datos son los siguientes:

Anemia Edad	Atrofia Muscular				Total
	no	si	no	si	
Menos de 1 año	63	0	23	2	88
1-2 años	23	2	47	5	77
2-3 años	31	1	50	12	94
3-4 años	16	5	47	15	83
4-6 años	30	9	86	31	156
6-8 años	42	10	57	34	143
8-10 años	22	1	25	11	59
Total					700

Ajustando un modelo logit independiente con 4 categorías, es decir el modelo

$$p_{ji} = \frac{\exp(\beta_j x_i)}{1 + \exp(\beta_j x_i)}$$

$$1 + \sum_{k=1}^3 \exp(B_k x_i^t)$$

y

$$p_{1i} = \frac{1}{1 + \sum_{k=1}^3 \exp(B_k x_i^t)}$$

$$1 + \sum_{k=1}^3 \exp(B_k x_i^t)$$

donde $B_j x_i^t = \alpha_j + b_j x_i^t$ y $x_i^t = 1$. Este modelo sería equivalente a trabajar con

$$p_{1i} = \frac{\exp(B_j x_i^t)}{1 + \sum_{k=1}^3 \exp(B_k x_i^t)}$$

en donde con el objeto de obtener una única solución, imponemos la restricción $\alpha_1 = b_1 = 0$.

Este modelo lo vamos entonces a trabajar como uno log-lineal en donde vamos a considerar 2 factores, uno la edad, es decir un factor con 7 niveles y otro la categoría es decir un factor con 4 niveles, además tomamos como variable el grupo de edad y la interacción de la variable edad con el factor que define las categorías. Los parámetros de interés van a ser en este caso los que corresponden al factor de categorías y los que corresponden a la interacción de la variable edad con el factor categorías.

Realizando este ajuste con GLIM obtenemos los siguientes

estimadores:

$$\alpha_2 = 2.263$$

$$b_2 = -0.2509$$

$$\alpha_3 = -1.257$$

$$b_3 = -0.01906$$

$$\alpha_4 = 2.501$$

$$b_4 = -0.4118$$

con una devianza de 62.56 con 21 g.l..

Bajo este modelo la asociación de los dos signos (síntomas) es una función que varía con la edad e involucra un parámetro adicional a los de un probit bivariado. Entonces podríamos pensar en imponer restricciones adicionales, es decir estimar un modelo de cinco parámetros bajo el cual los signos tendrían una asociación constante con la edad. Esto lo hacemos imponiendo la restricción $b_1 + b_4 = b_2 + b_3$.

Para imponer este tipo de restricciones en GLIM con el mismo procedimiento que se utilizó anteriormente, lo que se hace es crear los 4 niveles del factor que define la categoría con 2 factores de dos niveles cada uno, de manera que obtenemos los cuatro niveles con las combinaciones 11,12,21,22. Es claro que si en lugar de ajustar la interacción del factor categorías con edad, ajustamos la interacción de cada una de estos nuevos factores con edad estamos imponiendo la restricción

$$b_1 + b_4 = b_2 + b_3$$

además como $b_1 = 0$ esta restricción es equivalente a

$$b_4 = b_2 + b_3$$

los parámetros obtenidos en este caso son :

$$\alpha_2 = 2.431$$

$$b_2 = -0.1439$$

$$\alpha_3 = -0.6566$$

$$b_3 = -0.2857$$

$$\alpha_4 = 2.602$$

con una devianza de 63.67 con 16 g.l., el ajuste no resulta bueno sin embargo el objetivo es ejemplificar la técnica más que encontrar un buen ajuste.

Ahora bien si además imponemos la restricción

$$\alpha_1 + \alpha_4 = \alpha_2 + \alpha_3$$

el modelo ajustado es equivalente al modelo por Grizzle y que éste llama Logit Multivariado, la restricción antes propuesta la podemos imponer ajustando los nuevos factores con dos niveles cada uno y su interacción con edad. Es decir estamos suponiendo que las probabilidades en las celdas en cualquier edad podrían darse por el producto de las marginales.

Los resultados obtenidos en este ajuste son:

$$\alpha_2 = 0.1536 \quad (0.1882) \quad b_2 = -0.1744 \quad (0.0462)$$

$$\alpha_3 = 2.787 \quad (0.2929) \quad b_3 = -0.3104 \quad (.05795)$$

en este caso anotamos los errores estándar de los estimadores para comparación, ya que se ajustará el modelo de Grizzle a estos mismos datos.

La devianzas obtenida en este caso es de 77.73 con 17 g.l..

4.3 Ajuste en el caso Multivariado.

En el caso multivariado unicamente se desarrolló un MACRO para el ajuste de el modelo propuesto por Grizzle.

$$K \ln A \Pi = X B + \epsilon \quad \text{donde } \text{cov}(\epsilon) = \Sigma$$

Este modelo se puede ajustar directamente por medio de minimos cuadrados generalizados o bien descomponiendo la matriz de varianzas y covarianzas, Σ , como LL^t , donde L es una matriz no singular y ajustar el modelo

$$L^{-1} K \ln A \Pi = L^{-1} X B + \eta \quad \text{cov}(\eta) = \sigma^2 I.$$

En el caso particular de este modelo un estimador de esta dado por

$$B = K D^{-1} A V(p) A^t D^{-1} K$$

donde D es una matriz diagonal con elementos en la diagonal a p_i , a_i es renglón de A, p_i es el estimador muestral de π_i , $V(p_i)$ es el estimador de $V(\pi_i)$, matriz diagonal por bloques con $V(p_i)$ en la diagonal. En el caso particular de dos variables dicotómicas $V(p_i)$ es:

$$\frac{1}{n} \begin{vmatrix} p_i(1-p_i) & -p_i & p_i & -p_i & p_i & -p_i & p_i \\ p_i(1-p_i) & -p_i & p_i & -p_i & p_i & -p_i & p_i \\ -p_i & p_i & p_i(1-p_i) & -p_i & p_i & -p_i & p_i \\ -p_i & p_i & p_i(1-p_i) & -p_i & p_i & -p_i & p_i \\ p_i(1-p_i) & -p_i & p_i & -p_i & p_i & -p_i & p_i \\ p_i(1-p_i) & -p_i & p_i & -p_i & p_i & -p_i & p_i \\ -p_i & p_i & p_i(1-p_i) & -p_i & p_i & -p_i & p_i \\ -p_i & p_i & p_i(1-p_i) & -p_i & p_i & -p_i & p_i \end{vmatrix}$$

Entonces Σ es una matriz diagonal por bloques, en donde el bloque i , si llamamos $C = \begin{pmatrix} p_i & -p_i \\ p_i & -p_i \end{pmatrix}$, $A = \begin{pmatrix} p_i & p_i \\ p_i & p_i \end{pmatrix}$, $B = \begin{pmatrix} p_i & p_i \\ p_i & p_i \end{pmatrix}$ es:

$$\frac{1}{n} \begin{vmatrix} 1/A & C/(A*B) \\ C/(A*B) & 1/B \end{vmatrix}$$

Descomponiendola en LL^t , L diagonal por bloques, cuyo bloque i -ésimo es:

$$\frac{1}{n} \begin{vmatrix} 1/A & 0 \\ C/A*B & D/B \end{vmatrix}$$

donde $D = 1 - (C^2) * A * B$ y el bloque i -ésimo de L^{-1} es:

$$\begin{array}{c|c} n & A \\ 1 & C/E * B \end{array} \quad \begin{array}{c|c} 0 & \\ B/E & E \end{array}$$

donde $E=1-C*A*B$.

Entonces lo que efectua el macro es encontrar $L^{-1} KlnAp$ y $L^{-1} X$ y ajustar el modelo directamente como un modelo de regresión en GLIM.

La manera de llevar a cabo el ajuste se discutirá con un ejemplo, el anteriormente visto de los signos clínicos.

Para utilizar este macro es necesario unicamente dar los datos y utilizar API el cual realiza todo el proceso.

La manera de dar los datos en el caso de dos variables dicotómicas es: Un vector $Y1$, el cual contiene los valores correspondientes a no presentar respuesta en la primera variable alternando el presentar respuesta en la segunda y no presentarla; un vector $Y2$, que contiene los valores correspondientes a presentar respuesta en la segunda variable alternando el no presentar respuesta en la primera y el presentarla; $2*m$ variables, donde m es el número de variables explicativas, $X1, \dots, X2*m$, donde cada par de variables representa una variable explicativa, alternando el valor de la variable explicativa con ceros de tal forma que en donde una de las variables tiene cero la otra no; por último un vector T que tiene el total de individuos en el grupo de edad correspondientes.

En este ejemplo:

Y1	Y2	X1	X2	X3	X4	T
62.5	.5	1	0	1	0	88
23	2	0	1	0	1	88
23	2	1	0	2	0	77
47	5	0	1	0	2	77
31	1	1	0	3	0	94
50	12	0	1	0	3	94
16	5	1	0	4	0	83
47	15	0	1	0	4	83
30	9	1	0	5	0	156
86	31	0	1	0	5	156
42	10	1	0	6	0	143
57	34	0	1	0	6	143
22	1	1	0	7	0	59
25	11	0	1	0	7	59

Los estimadores obtenidos en este caso son :

$$\alpha_1 = .1285(.4364) \quad \beta_1 = -.0.1604 (.09521)$$

$$\alpha_2 = 2.691(.7296) \quad \beta_2 = -0.3051(.1443)$$

Con una devianza de 46.64 con 10 g.l.. El valor de la devianza es menor, sin embargo no es comparable, ya que los grados de libertad también son menos.

Los cuales son comparables al ajuste llevado a cabo por medio del modelo logístico poliotómico.

5. Conclusiones y Generalizaciones

Las aportaciones principales de esta tesis, son el desarrollo de subrutinas que permiten ajustar sin problemas modelos complejos. En particular en el caso de modelos de respuesta politómica, el trabajo realizado permite ajustar cualquier modelo especificando unicamente cinco macros sencillos en GLIM.

En el caso de modelos categoricos multivariados, lo desarrollado permite también ajustar el modelo de Grizzle que se pueden considerar como aproximación al probit multivariado o bien como un modelo propio. Además se sugiere una forma de aproximación también al probit multivariado (o al modelo de Grizzle), que no requiere más que ajustar modelos simples con GLIM.

En base a los modelos presentados podemos hablar de distintas generalizaciones.

Una de ellas, quizá la más importante sería el crear una familia de transformaciones de las probabilidades en términos de funciones lineales de los parámetros. Para el caso de variables binarias Aranda(1980) propuso una familia de transformaciones(de probabilidades) en las que por medio de un parámetro ($0 < \lambda < 1$), se reproducen los modelos logístico y lineal en el límite $\lambda \rightarrow 0$ y $\lambda \rightarrow 1$ respectivamente, adicionalmete encontró que para $\lambda = .38$ y $\lambda = .67$ se podían aproximar el modelo Normit y arco seno, respectivamente. Algo importante de notar es que los valores de λ tiene que ver con la velocidad con que cada función alcanza su límite. Esta familia es

$$T_{\lambda}(\theta) = \frac{2\theta^{\lambda} + (1-\theta)^{\lambda}}{\lambda\theta^{\lambda-1} + (1-\theta)^{\lambda-1}}$$

Además esta familia puede ser vista como un modelo lineal generalizado, con una estructura de error Binomial, componente sistemático $\eta = X\beta$ y función liga $\mu = T_{\lambda}(\eta)$, donde:

$$\eta_i \left(\frac{1 + \lambda\eta_i/2}{1} \right)^{1/\lambda} \left(\left| \frac{1 - \lambda\eta_i/2}{1} \right|^{\lambda} + \left| \frac{1 - \lambda\eta_i}{1} \right|^{\lambda} \right)^{-1/\lambda}$$

0 $\lambda\eta_i/2 < -1$

1 $|\lambda\eta_i/2| < 1$

1 $\lambda\eta_i/2 > 1$

Esta familia incluye al caso simétrico. Aranda op.cit., propone una familia para el caso asimétrico, en la que se incluye el modelo doble logaritmo complementario y logaritmo complementario, además incluye al modelo logístico, esta familia consta de 2 parámetros

$$V_{\lambda, \gamma}(\theta) = (\log(\theta/(1-\theta)) + \gamma)^{\lambda} - 1 / \lambda$$

la cual reproduce el modelo logístico cuando $\gamma=0, \lambda=1$; el modelo doble logaritmo complementario cuando $\gamma=1, \lambda=0$; y el logaritmo complementario cuando $\gamma=1, \lambda=1$.

Esta función, al igual que la anterior, también se puede trabajar como un modelo lineal generalizado. Para evitar

problemas en el caso que $\lambda=0$ y $\gamma=0$, ya que si $\theta/(1-\theta)$ la familia anterior no se encontraría definida Aranda op.cit . sugiere la familia

$$V_{\zeta, \gamma, \xi} = [\log(\theta/(1-\theta))^{\zeta+\gamma} \zeta]^{-1/\zeta}$$

con $\zeta = -1, 1$, $-\infty < \zeta < \infty$, $0 < \gamma$.

Para el caso de respuesta politómica no ordenada Aranda op. cit. sugiere una generalización partiendo de la expresión de la probabilidad en términos de un componente sistemático de la generalización del logit a respuesta politómica no ordenada, llamado en esta tesis logit independiente, esta expresión es:

$$\theta_j = \frac{\exp(\eta_j)}{\sum_{g=1}^k \exp(\eta_g)}$$

generalizandola a

$$\theta_j = \frac{(1 + \zeta \eta_j)^{1/\zeta}}{\sum_{g=1}^k (1 + \zeta \eta_g)^{1/\zeta}}$$

con $-1 < \zeta$; que cuando $\zeta \rightarrow 0$ se reduce a la expresión anterior, es decir reproduce el modelo logístico y cuando $\zeta = 1$ reproduce el lineal. Esta misma familia puede reproducir el modelo probit politómico no ordenado para un valor apropiado de ζ , lo cual sería útil desde el punto de vista computacional, ya que para la

estimación de este modelo no existen métodos alternativos a la resolución de normales.

En el caso de categoría de respuesta ordenada, se podría pensar en una familia que contuviera algunos de los modelos sugeridos. en este caso se podría pensar en expresar la probabilidad acumulada en términos de un componente sistemático, en lugar de la probabilidad por categorías, probablemente sería necesario distinguir entre los modelos simétricos y los asimétricos.

Otro tipo de generalización sería lo que propone Mc. Cullagh op. cit. en los modelos para respuesta politémica ordenada que es el utilizar modelos no lineales, esto es debido a que en diferentes grupos pueden observarse distintos patrones, sin embargo en todos ellos se cumple el orden estocástico estricto, siendo la concentración en las diferentes categorías de respuesta diferente.

Mc. Cullagh propone el utilizar el modelo multiplicativo

$$P_{ji} = (\alpha_j - B x_i^t) / \tau_i$$

en donde B podría llamarse el parámetro de localización y el de escala.

Obviamente este modelo es adecuado cuando el número de categorías es tres o más. Como vamos a tener un parámetro de escala asociado a cada renglón de la tabla el modelo resulta saturado en parámetros de escala. Para hacer que los parámetros

de escala sean identificables es conveniente imponer una restricción como $\tau = 1$ o $\log \tau = 0$, en donde quizá sea más adecuada la segunda, debido a que los estimadores de $\log \tau$ son más cercanos a ser distribuidos simétricamente que los de τ .

El problema fundamental de esta generalización es el de estimación de los parámetros Mc. Cullagh op. ci. da expresiones para las derivadas de la función de log-verosimilitud, así como para las segundas derivadas y sus esperanzas, pero el implementar esto computacionalmente, no parece ni razonablemente sencillo.

Bartholomew (1980) sugiere, en la discusión del paper de Mc. Cullagh, el considerar que la variable subyacente que determina los umbrales, tiene la misma forma en todos los renglones, pero varía de renglón a renglón.

6. BIBLIOGRAFIA

Aitchison J., Bennet J. A. (1970) Polychotomuos quantal response by maximum indicant, *Biometrika*, 57, 233-262.

Aitchison J., Silvey S.D. (1959) The generalization of Probit Analysis to the case of multiple responses, *Biometrika*, 44, 131-143.

Amemiya T. (1975) Qualitative response models, *Annals of economic and Social Measurement*, 4, 363-371.

Amemiya T. (1976) The Maximum likelihood, the minimum Chi-square and the nonlinear Weighted least-squares estimator in the general qualitative response model. *J.A.S.A.*, 71, 347-351.

Amemiya T. (1978) On a two-step estimation of a multivariate logit model. *J. Econometrics*, 8, 13-21.

Amemiya T. (1981) Qualitative Response Models: A survey, *Journal of Economic Literature*, 19, 1483-1536.

Aranda F.J. (1980)

Ashford J.R. (1959) An approach to the analysis of data or semiquantal responses in biological assay, *Biometrics*, 15, 573-581.

Ashford J.R., Sowden (1976) Multivariate Probit Analysis, *Biometrics*, 26, 535-546.

Bock R.D. (1975) *Multivariate Statistical Methods in behavioral research*. Mc. Graw Hill-Inc.

Bishop Y. M., Fienberg S.E. Holland P.W. (1975) Discrete Multivariate Analysis. The Mil Press Cambridge, Massachusetts, and London, England.

Cox D.R. (1970) The analysis of Binary Data, London New York Chapman and Hall.

Darroch J. N., Ratcliff D. (1972) Generalized Iterative scaling for Log-linear models, The annals of Mathematical Statistics, 43, 1470-1480.

Fienberg S. E. (1970) An iterative Procedure for estimation in contingency tables, The annals of mathematical statistics, 41, 907-917.

Goldstein H. (1980) Specifying a Multivariate logit model using GLIM, Glimnewsletters, 1.

Grizzle J.E., Stomer C. F., Koch G.G. (1969) Analysis of categorical data by linear models, Biometrics, 25, 489-503.

Grizzle J.E. (19) Queries and Notes 317 Note: Multivariate Logit Analysis

Gurland J., Lee I., Dalm P.I., (1960) Polychotomous quantal response in biological assay, Biometrics, 16, 382-398.

Mantel N., Brown Ch. (1973) A logistic reanalysis of Ashford and Sowden's data on respiratory symptoms in British coal miners, Biometrics, 29, 649-665.

Mc. Cullagh P. (1980) Regression models for ordinal data J.R.

Statist. Soc., 42,109-142.

Mc. Cullagh P., Nelder J.A. (1983) Generalized Linear Models, Chapman and Hall.

Mc. Cullagh P. (1983) Quasi-likelihood function The annals of Statistics, 11,59-67.

Nelder J.A., Wedderburn W. M. (1972) Generalized Linear Models, J.R. Statistics Soc., A135,370-384.

Rogers J.H. (1983) Composite Link Functions with Linear Log Link and Poisson Error, Glimnewsletter, 7.

Thompson R., Baker J. (1981) Composite Link Functions in Generalized Linear Models, Appl. Statist, 30, 125-131.

Wedderburn R.W.M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method, Biometrika, 61,439-447.

7. APENDICE

7.1 Demostracion de identificabilidad de los parametros en el modelo de Aitchson y Benett

La prueba como se habia mencionad  es por inducci n en donde el primer paso de la inducci n esta dado por el caso de dos categor as. para el caso de k categor as. lo suponemos cierto para k-1. Entonces

$$\Delta_k = \begin{vmatrix} d - m & m & \dots & m \\ 1 & c=1 & ic & 12 & \dots & 1k \\ m & d - m & \dots & m & \dots & m \\ & 21 & 2 & c=2 & 2c & \dots & 2k \\ m & \dots & \dots & \dots & d - m & \dots & m \\ & k1 & \dots & \dots & k & c=k & kc \end{vmatrix}$$

sumando todas las columnas a la primera

$$\Delta = \begin{vmatrix} d & m & \dots & m \\ 1 & 12 & \dots & 1k \\ d & d - m & \dots & m \\ 2 & 2 & c=2 & 2c & \dots & 2k \\ d & \dots & d - m & m \\ k & \dots & k & c=k & kc \end{vmatrix}$$

resolviendo por menores

$$\Delta_k = d D_{k-1} + (-1)^m D_{12} + (-1)^2 D_{13} + \dots + (-1)^{k-1} D_{1k}$$

Ahora

$$D_{11} = \begin{vmatrix} d_{11} - m_{11} & m_{12} & m_{13} & \dots & m_{1k} \\ 2c_{21} & d_{22} - m_{22} & m_{23} & \dots & m_{2k} \\ m_{31} & m_{32} & d_{33} - m_{33} & \dots & m_{3k} \\ \dots & \dots & \dots & d_{kk} - m_{kk} & \dots \\ m_{k2} & \dots & \dots & \dots & d_{kk} - m_{kk} \end{vmatrix}$$

donde $d_{11} = d_{11} - m_{11}$, y por la suposición en la inducción $D_{11} > 0$.
 Los demás menores se pueden expresar en manera analoga con $i-2$
 intercambios de renglones adycentes, por lo que $(-1)^i D_{11} > 0$ y $\Delta_k > 0$.

7.2 logit independiente por maximización de utilidad

En particular par el caso de tres categorías

$$\begin{aligned} P(y_1=3) &= p(u_{13} > u_{12}, u_{13} > u_{11}) \\ &= P(e_3 + u_{13} - u_{12} > e_2, e_3 + u_{13} - u_{11} > e_1) \\ &= \int_{-\infty}^{\infty} f(e_3) \left[\int_{-\infty}^{\infty} \frac{e^{-u_{13}}}{3} \frac{e^{-u_{12}}}{2} f(e_2) de_2 \int_{-\infty}^{\infty} \frac{e^{-u_{13}}}{3} \frac{e^{-u_{11}}}{1} f(e_1) de_1 \right] de_3 \end{aligned}$$

si la función de distribución es $\exp(-\exp j)$

$$P(y_1=3) = \frac{\exp(\mu_{13})}{\exp(\mu_{13}) + \exp(\mu_{12}) + \exp(\mu_{11})}$$

de aquí el logit independiente se genera sustituyendo

$$x_{21} = \mu_{12} - \mu_{13} \quad y$$

$$B^t x = \mu - \mu$$

$$1 \quad 1 \quad 11 \quad 13$$

7.3 Demostración de la unicidad de la función discriminatoria (escalamiento iterativo)

La demostración se basa en la suposición de que existe otra

$$K(p, \pi) = \sum_{i \in I} p_i \left[\log \mu + \sum_{s=1}^d b_{si} \log \mu \right]$$

$$= \log \mu \left[\sum_{i \in I} p_i + \sum_{s=1}^d \log \mu \sum_{si \in I} b_{si} p_i \right]$$

$$= \log \mu \left[\sum_{i \in I} q_i + \sum_{s=1}^d \log \mu \sum_{si \in I} b_{si} q_i \right]$$

$$= \sum_{i \in I} q_i \left[\log \mu + \sum_{s=1}^d b_{si} \log \mu \right]$$

$$= \sum_{i \in I} q_i \log(p_i / \pi_i)$$

$$= K(q, \pi) - K(p, \pi) = K(q, p)$$

Lo cual es igual a cero $\iff q=p$.

BIBLIOTECA
 JUAN A. ESCALANTE
 UNIDAD ACADÉMICA DE
 LOS CICLOS PROFESIONAL
 Y DE POSGRADO / CCH
 UNAM