

Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS



MODELOS LINEALES GENERALIZADOS
Y SU RELACION CON GLIM

T E S I S
QUE PARA OBTENER EL TÍTULO DE:
A C T U A R I O
P R E S E N T A:

SILVIA RUIZ VELASCO ACOSTA

MEXICO, D. F.

1982



Universidad Nacional
Autónoma de México



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.

TESIS CON FALLA DE ORIGEN

I N D I C E

INTRODUCCION.

Importancia de los Modelos Lineales	1
Generalización a Modelos lineales generalizados....	5
Tablas de Contingencia	7
Probit Análisis	20

MODELO MATEMATICO

Planeamiento	29
Relación con Glim	40

TEOREMAS IMPORTANTES PARA ESTIMACION

Mínimos Cuadrados	48
Estimación Máxima Verosimilitud	52
Solución Iterativa a las Ecuaciones Máximo Verosimil .	56
Solución a Estimación Máximo Verosimil en Modelos Lineales Generalizados	58

EJEMPLOS DE APLICACIONES	64
--------------------------------	----

I N T R O D U C C I O N

Fundamentalmente el concepto de modelo estadístico para una variable aleatoria, es tal que la variable en investigación tiene una estructura definida que explica los valores actuales obtenidos y además predice valores futuros. La variable -- puede ser expresada en términos de otras variables básicas (los componentes de la estructura), si sus valores son fijos (posiblemente desconocidos) son llamados componentes sistemáticos y si hay también variables aleatorias son llamados componentes aleatorios.

Importancia de los modelos lineales.

Los modelos estadísticos lineales son la base de los métodos estadísticos más usuales en áreas tan importantes como regresión y diseño de experimentos.

El concepto básico de los modelos lineales es el que considera el estudio de varias poblaciones con un mismo grado de generalidad, en las que el modelo de las distribuciones de frecuencia es normal con varianza constante e independencia entre observaciones, además las medias de la población dependen, de manera conocida, de los factores que definen a las poblaciones. Se reconoce que la dependencia de la media sobre ciertos factores es de tipo lineal.

Entonces se tiene como modelo lineal general el siguiente

$$Y = \bar{X}\beta + \varepsilon$$

$Y \in \mathbb{R}^n$, $\beta \in \mathbb{R}^n$ constantes fijas pero desconocidas, \bar{X} matriz de constantes conocidas de dimensión $n \times p$ $\varepsilon \in \mathbb{R}^n$ vector aleatorio con distribución y media \mathcal{Q} y matriz de covarianza

$$\sigma^2 I (\Sigma \sim N(\mathcal{Q}, \sigma^2 I))$$

Este modelo tiene un grado máximo de flexibilidad.

Desde el punto de vista de la estadística el problema consiste en encontrar estimadores de β que minimizen los errores, la manera más conocida y eficiente es por medio de estimadores obtenidos por el método de mínimos cuadrados.

Entre las aplicaciones más importantes de regresión están :

1) Descripción y explicación.

Los modelos de regresión son valiosos para describir el tipo de asociación entre la variable y (respuesta) y las variables X_1, \dots, X_p (explicativos). En este caso, lo que se persigue es resumir las tendencias de los datos y encontrar la forma de asociación entre las variables. A través de los modelos de regresión es posible investigar si existe asociación entre las variables además, también es posible determinar cuál es la forma de dicha

asociación.

Un empleo interesante en los modelos de regresión se da en la determinación de las relaciones entre distintos factores que afectan el comportamiento de un fenómeno y que puedan pasar a ser establecidas como relaciones de causa efecto. Aunque es importante señalar que mediante los métodos estadísticos sólo se pueden establecer relaciones de asociación simple.

2) Predicción.

Un empleo sumamente importante de los modelos de regresión es la posibilidad de predecir el valor que tendrá Y_i o la media de la población del conjunto Y_i que se genera cuando se especifica las condiciones del proceso mediante los valores de las X_j 's $j=1, \dots, p$.

Un ejemplo son las llamadas series de tiempo, donde se trata de predecir cuál será el valor de una característica que depende del tiempo. Por ejemplo el precio de un cierto artículo dentro de un año.

Otro ejemplo, es el predecir o pronosticar la producción (Y) que se tendrá para ciertos costos, concentrados y calidades (X_j).

También se emplea para predecir el rendimiento (y) de -

un cultivo en una región bajo ciertas condiciones de lluvia, nitrógeno, fósforo, inudencias de plagas, etc. (X).

3) Control.

Un modelo de regresión puede servir para encontrar cuales son los valores de las $(X_{i,s})$ que pueden optimizar de acuerdo con algún criterio, los valores de la variable dependiente (Y). En la industria se pueden usar los modelos de regresión para controlar, es decir, para optimizar procesos de producción.

En términos generales si es posible ligar mediante un modelo de regresión una característica de interés económico Y con un conjunto de factores o variables independientes X_j susceptibles de modificarse según el arbitrio humano, se optimizará la característica Y variando los valores de los factores X_j . Este es el proceso estadístico conocido como control.

En Diseño de Experimentos se encuentran las siguientes aplicaciones :

comparación de medias de poblaciones distintas.

Los modelos estadísticos lineales de diseño de experimentos tienen como objetivo principal comparar las medias de las poblaciones estudiadas.

En medicina es frecuente la comparación de las medias de poblaciones de ciertos caracteres fisiológicos y morfológicos, en este caso las poblaciones que se generan al considerar cierto tipo de drogas, dosis de una misma droga o tipos de tratamientos médico en general.

Estudio de efectos.

Un diseño experimental puede planearse para estudiar las relaciones entre varios factores cualitativos o cuantitativos que se sirven como criterios para definir las poblaciones bajo estudio con sus respectivas medias. Una hipótesis de gran importancia práctica es la de no interacción entre dos o más factores cualitativos o cuantitativos que sirven como criterios para definir las poblaciones bajo estudio. Cuando la hipótesis de "no interacción" es rechazada, los resultados de un factor se ven modificados por los niveles del otro(s) factor(es).

Los modelos estadísticos lineales se han empleado para generar algunas técnicas estadísticas muy especiales, entre las que se encuentran las de estadística multivariada; gran parte de las técnicas de este campo son consideradas como extensión de los modelos lineales.

Modelos Lineales Generalizados.

Los modelos lineales están constituidos, como vimos en, por una parte sistemática y una aleatoria (error); donde usualmente se asume una distribución normal para los errores; la téc

nica de análisis como anteriormente se mencionó es la obtención de estimadores mediante el método de mínimos cuadrados, el cual asume solamente un componente del error; una extensión para errores múltiples fué desarrollada primeramente para análisis de diseño de experimentos.

Las técnicas desarrolladas para datos cuya distribución del error es no normal incluyen probit analysis, donde una variable binomial cuenta con un parámetro relacionado con una distribución de tolerancia. También para tablas de contingencia donde la distribución es multinomial y las partes sistemáticas de los efectos usualmente son multiplicativos. Ambos ejemplos tienen un aspecto lineal en el modelo de Probit el parámetro p es una función de tolerancia, la cual es por sí misma lineal en la dosis (o en una función de ella). En tablas de contingencia con un modelo multiplicativo, el logaritmo produce un modelo lineal.

Aparte de estos modelos asociados con distribuciones normal, binomial y multinomial (el cual puede pensarse como una familia de distribución Poisson con contrastes), el problema se puede plantear. En general mediante los modelos lineales generalizados que pueden considerarse de la forma $Y_i = u_i + \epsilon_i$,

donde (u_i) representa una parte de componentes sistemáticos y ϵ_i representa un componente aleatorio.

La distribución e_i es un miembro de una familia exponencial

Esta familia incluye las distribuciones Normales, Binomial, Poissos, X^2 y Gama entre otros.

A continuación presentamos algunos ejemplos de modelos lineales generalizados.

Tablas de contingencia.

El uso más importante es determinar el tipo de asociación que se da entre variables categóricas.

Cuando la población es clasificada con respecto a dos o más variables cualitativas, formamos una tabla de contingencia. La entrada en las celdas son frecuencias, cuando la tabla sólo se refiere a dos variables la llamamos tabla de contingencia de dos dimensiones y cuando involucra tres o más variables la llamamos multidimensional.

Una tabla de 2 x 2 es la tabla más simple de dos dimensiones y un ejemplo de ella sería :

Insidencia de Tuberculosis por Sexo

	Mujeres	Hombres	Total
Tub del Sist. Resp.	3534	1219	4853
Otras	270	252	522
todas las f.	3804	1571	5375

Como la población se encuentra clasificada mediante variables cualitativas cada miembro de la población debe pertenecer a una categoría y solo a una.

En el caso de tablas de 2×2 determinar independencia o no independencia de las variables es relativamente fácil, por ejemplo en la tabla 1.1 se pretende probar independencia entre el sexo y la forma de tuberculosis.

Nosotros esperamos entonces que la proporción de mujeres que murieron de tuberculosis del aparato respiratorio es igual a la proporción de hombres que murieron por lo mismo. Si las proporciones son diferentes entonces las muertes por tuberculosis en el sistema respiratorio tienden a ser más frecuentes en un sexo que en el otro. Es claro que las proporciones pueden ser diferentes, debido a mala recopilación de información u otras razones que pueden ser atribuidas a causas aleatorias. Necesitamos tener un criterio para saber cuando la diferencia es demasiado grande para atribuirla a dichas causas, por lo cual necesitamos realizar una prueba que nos de esa medida, la prueba comunmente utilizada es la de X^2 . (1).

En general el concepto de independencia en tablas de se puede explicar como sigue:

Supongamos que la probabilidad de una observación en la i -ésima categoría de la variable 1 (renglón) y la j -ésima catego

ría de la variable 2(columna) esta representada por P_{ij} ; por lo tanto, la frecuencia F_{ij} esperada en la ij -ésima celda de la tabla es $F_{ij} = N p_{ij}$

Si $P_{i.}$ representa la probabilidad "en la población", de que una observación corresponda a la i -ésima categoría de la "primera variable (no importando la categoría de la segunda) y $P_{.j}$ - representa la probabilidad correspondiente a la j -ésima categoría de la segunda. Entonces la hipótesis de independencia se puede plantear mediante la siguiente proposición.

$$P_{ij} = P_{i.} P_{.j}$$

Sin embargo, los valores de $P_{i.}$ y $P_{.j}$ son desconocidos, pero pueden ser estimados por medio de las frecuencias observadas; los estimadores son :

$$\hat{P}_{i.} = \frac{n_{i.}}{N} \text{ y } \hat{P}_{.j} = \frac{n_{.j}}{N}$$

Por lo tanto, de (1) se espera que bajo la hipótesis de independencia F_{ij} se estime mediante

$$E_{ij} = N \hat{P}_{i.} \hat{P}_{.j} = N \frac{n_{i.}}{N} \frac{n_{.j}}{N} = n_{i.} \frac{n_{.j}}{N}$$

(E_{ij} frecuencia estimada)

Cuando las variables son independientes las frecuencias estimadas y las observadas difieren por cantidades pequeñas que se pueden atribuir al azar, por el contrario si no son indepen

dientes esperamos una diferencia grande.

La prueba de la X^2 consiste en calcular la estadística

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

La magnitud de esta estadística depende de los valores de las diferencias $(n_{ij} - E_{ij})$; por lo que la X^2 es pequeña cuando es cierta la hipótesis de independencia. Buscando en tablas de X^2 a un cierto nivel de significancia rechazamos o aceptamos la hipótesis.

Tablas de rxc

El análisis de tabla de contingencia de rxc, con r ó c mayores que 2, presenta a diferencia de las tablas de 2 x 2, cierta dificultad en la interpretación cuando no se cumple la independencia.

En las tablas de rxc las categorías de las variables, pueden guardar un orden por ejemplo: bajo, medio, alto.

Si la prueba de X^2 para una tabla de rxc indica no-independencia, no tenemos suficiente información para saber si

es en toda la tabla o en una parte específica donde se rompe la independencia. La situación es semejante a técnicas de análisis de varianza, que nos determinan la existencia de diferencias en las medias, pero no nos dice cuales son las que difieren. De la misma manera que en análisis de varianza - existen técnicas o métodos para identificar los elementos de la tabla de contingencia rompen la hipótesis de independencia. Un método propuesto con este objetivo es el de Lancaster y Ferwin que se presenta a continuación.

Método de Lancaster y Ferwin.

Consiste en partir la tabla en tantas tablas como grados de libertad. Cada componente de la X^2 corresponde a una tabla particular de 2×2 proveniente de la original y cada componente es independiente de los otros.

En el caso de 2×3 , por ejemplo, necesitamos construir 2 tablas y estas serían :

a_1	a_2	$(a_1 + a_2)$	a_3
b_1	b_2	$(b_1 + b_2)$	b_3
a_1	a_2	$(a_1 + a_2)$	a_3
b_1	b_2	$(b_1 + b_2)$	b_3

Kimball desarrolló una fórmula para obtener los valores de la X^2 correspondientes al método de Lancaster y Ferwin.

En el caso de tablas de $2 \times c$ es:

$$X_t^2 = \frac{N^2 [b_{t+1} S_t^{(a)} - a_{t+1} S_t^{(b)}]^2}{A B n_{t+1} S_t^{(n)} S_{t+1}^{(n)}}$$

donde $S_t^{(a)} = \sum_{i=1}^t a_i$ $S_t^{(b)} = \sum_{i=1}^t b_i$ $S_t^{(n)} = \sum_{i=1}^n n_i$

considerando la tabla

n_1	a_2	a_c	A
b_1			b_c	B
n_1	n_2		n_c	N

de manera que los valores de la X^2 se obtienen al ir sustituyendo t por $1, 2, \dots, C-1$ de tal manera que

$$X^2 = X_1^2 + \dots + X_{C-1}^2$$

Existen otros métodos que consisten en partir tablas de 2×2 en tablas no-independientes de 2×2 , por ej. cuando se aplican diferentes tratamientos y un placebo, el principal interés es ver independencia entre los diferentes tratamientos y el placebo. Este procedimiento afecta seriamente el nivel de significancia, y podemos encontrar más diferencias entre placebo y tráfico que realmente existen. Broden demuestra que un método razonable es obtener una α' para probar el método de la X^2 , esa

$$\alpha' = \alpha/2(C-1).$$

Otra técnica que se utiliza con el mismo objetivo de identificar las celdas que contribuyen a que no se cumpla la hipótesis de independencia es el de ajustar modelos logaritmos lineales, como se presenta más adelante, y posteriormente analizar los parámetros obtenidos en el modelo.

Tablas Multidimensionales.

Los métodos para análisis de tablas de contingencia provenientes de 2 variables son en general bien conocidos, pero aquellos para analizar tablas provenientes de 3 ó más variables son muy poco conocidos.

El análisis de tablas de tres dimensiones presenta más problemas conceptuales comparado con el análisis de tablas de dos dimensiones. La extensión de tablas de tres dimensiones a cuatro ó más no presenta nuevos problemas conceptuales. Por lo que presentaremos con detalle el caso de 3 variables.

Un ejemplo del caso de 3 variables sería los datos que se presentan a continuación fueron citados por Cochran (1954). Madres que tienen hijos que han sido clasificados por los maestros como con problemas de comportamiento son comparados con madres con hijos que no han sido puestos en esta clasificación. Los dos

grupos mencionados se referirán como problemas y controles. También se han clasificado las madres en las que han sufrido de abortos y las que no han sufrido de abortos y con el orden de nacimiento del niño.

Orden de Nacimiento

	2		3-4		5+	
	Problemas	Controles	Problemas	Controles	Problemas	Controles
Con Abortos	20	10	26	16	27	14
Sin Abortos	82	54	41	30	22	23

La hipótesis de independencia mutua de las variables en una tabla de 3 dimensiones puede ser formulada como sigue:

$$H_0: P_{ijk} = P_{i..} P_{.j.} P_{..k}$$

donde P_{ijk} representa la probabilidad de que una observación ocurra en ijk -ésima celda y $P_{i..}$, $P_{.j.}$, $P_{..k}$ son las probabilidades marginales del renglón, columna y estrato de la variable respectivamente. En el ejemplo existen dos renglones representados por haber sufrido o no abortos, dos columnas representadas por niños con problemas o controles y 3 estratos representados por 2, 3-4 y 5+.

El procedimiento es igual al de tablas de 2 dimensiones: calculamos los estimadores de las frecuencias cuando H_0 es cierta; comparamos esos estimadores con las frecuencias cuando H_0 es cierta; comparamos esos estimadores con las frecuencias observadas por medio de la estadística usual X^2 y comparamos con la X^2 en tablas con $r-1-c-1+z$ grados de libertad, donde r , c y i representan el número de renglones, columnas y estratos respectivamente.

Si la prueba es no significativa no es necesario un análisis detallado de la tabla y concluimos que existe independencia entre las variables.

Cuando es significativa no podemos asumir que la dependencia es entre las tres variables. Puede darse el caso por ejemplo que dos de las variables estén asociados, pero la tercera sea completamente independiente; en ese caso la hipótesis de independencia puede plantearse. Puede haber situaciones en que dos variables son independientes en cada nivel de la tercera, pero pueden estar asociadas con ella, es decir las primeras 2 variables son condicionalmente independientes dado el nivel de la tercera.

Las hipótesis de independencia parcial en términos de probabilidad serían:

$$H_0^{(1)}: P_{ijk} = P_{i..} P_{.jk}$$

renglón independiente de columna y estrato.

$$H_0^{(2)}: P_{ijk} = P_{.j} P_{i.k}$$

columna independiente de renglón y estrato

$$H_0^{(3)}: P_{ijk} = P_{..k} P_{ij.}$$

estrato independiente de renglón y columna.

La prueba que se realiza es la de X^2 y el procedimiento es el ya anteriormente descrito donde para el primer caso

$$E_{ijk} = N \hat{P}_{i..} \hat{P}_{.jk}$$

$$y \quad \hat{P}_{i..} = \frac{N_{i..}}{N} \quad \hat{P}_{.jk} = \frac{N_{.jk}}{N}$$

$$\text{entonces} \quad E_{ijk} = \frac{N_{i..} N_{.jk}}{N}$$

y se compara con un X^2 de tablas con $c - 1 - r + 1$ grados de libertad.

Interacciones.

En tablas de 2x2 cuando el efecto de la columna j-ésima depende del renglón en que nos encontramos, decimos que existe interacción entre las 2 variables.

De este modo en tablas multidimensionales podemos hablar de interacción de segundo orden (entre tres variables) y de ordenes superiores.

Modelos Logarítmicos-lineales para tablas de contingencia

Comenzaremos por la tabla de 2 dimensiones, la hipótesis de independencia o de no interacción de primer orden entre las 2 variables es:

$$P_{ij} = P_{i.} P_{.j} \dots\dots(1)$$

lo que nosotros queremos es encontrar un modelo tal que P_{ij} o alguna función de ella pueda ser expresado como la suma de las probabilidades marginales ó alguna función de ellas. De modo tal que el modelo sea similar a los modelos de análisis de varianza.

Si aplicamos logaritmo natural en ambos lados de (1)

tenemos que

$$\text{Log } P_{ij} = \text{Log } P_{i.} + \text{Log } P_{.j}$$

y si lo escribimos en términos de las frecuencias tenemos que
 $\text{en } F_{ij} = \text{en } F_{i.} + \text{en } F_{.j} - \text{en } N$

sumados sobre i , y sobre j tenemos que:

$$\sum_{i=1}^r \sum_{j=1}^c \text{en } F_{ij} = C \sum_{i=1}^n \text{en } F_{i.} + r \sum_{j=1}^c \text{en } F_{.j} - rc \ln N$$

y la ecuación 1 se puede escribir

$$\text{en } F_{ij} = u + u_1(i) + u_2(j)$$

$$u = \frac{\sum_i \sum_j \log F_{ij}}{rc} - u_1(i) = \frac{\sum_j \log F_{ij}}{i} - \frac{\sum_i \sum_j \text{en } F_{ij}}{rc}$$

$$u_2(j) = \frac{\sum_i \text{en } F_{ij}}{r} - \frac{\sum_{i=1}^r \sum_j \text{en } F_{ij}}{rc}$$

$$u_1(.) = \sum_{i=1}^r u_1(j) = 0, \quad \sum_{j=1}^c (u_2(j)) = 0 = u_2(.)$$

Los parámetros se encuentran dados en términos de F_{ij} , por lo cual en la práctica es necesario estimarlos.

Es necesario, introducir un término extra que represente la interacción entre las dos variables,

$$\text{en } F_{ij} = u + u_1(i) + u_2(j) + u_{12}(ij)$$

donde $u_{12}(ij)$ representa la interacción entre los efectos de los niveles i y j de las variables 1,2 respectivamente.

En términos de la interacción de los parámetros la hipótesis de independencia sería $u_{12}(ij) = 0$ i, j , es decir todos los términos de interacción son cero, que es equivalente que el primer modelo especificado ajusta adecuadamente a los datos.

El modelo logaritmico lineal para una tabla de tres-dimensiones sería

$$\begin{aligned} \log F_{ijk} = & u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) \\ & + u_{23}(jk) + u_{123}(ijk) \end{aligned}$$

Este modelo incluye el efecto para cada variable, interacción de primer orden para cada par de variables, e interacción de segundo orden entre las tres variables.

Para estimación se debe de considerar el principio de jerarquía el cual pide que para estimación se debe considerar el prin

cipio de jerarquía lo cual cuando un término de alto orden es incluido en el modelo, los efectos de menor orden compuestos por las variables incluidas en el efecto mayor deben de ser también incluidas. Otro ejemplo sería de:

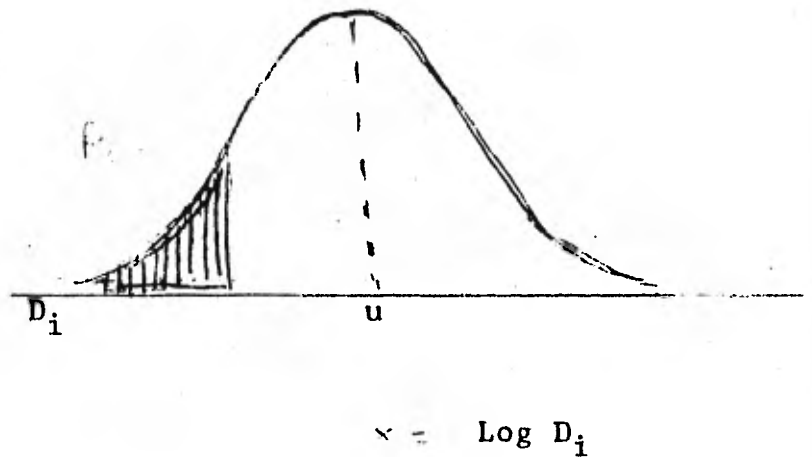
Probit Analisis.

Suponemos que se tienen K dosis de un estímulo y cada dosis se aplica a n_i individuos ($i=1, \dots, k$). Al final del experimento se observa en cada individuo si se presentó o no una respuesta predeterminada. El objetivo del experimento es estimar la dosis necesaria para que una proporción dada de los individuos de la población presente la respuesta.

Por ejemplo aplicar fertilizante con varias dosis a ciertas plantas y observar después de 8 días cuantas florecieron. Otro ejemplo sería dar veneno en varias concentraciones a roedores y ver después de ciertos días cuantos mueren.

Si D_1, D_2, \dots, D_k son las dosis aplicadas y sean r_1, \dots, r_k los individuos que muestran la respuesta a cada dosis. La distribución probabilística de las tolerancias a las dosis del estímulo generalmente es asimétrica, ya que se ve afectada por los individuos que tiene una tolerancia muy grande al estímulo. Por esta razón y con el objeto de volver simétrica la distribución se ha vuelto práctica común trabajar con los logaritmos de las

dosis. Sea $X_i = \log D_i$ y sea P_i la proporción de individuos en la población cuyo logaritmo es X_i . Si el cambio de variable efectuado mediante la transformación logarítmica ha vuelto simétrica la distribución, tenemos



es claro entonces que P_i es una área de la Distribución probabilística de las tolerancias.

Si se supone como es usual, que la distribución de tolerancias a los logaritmos de las dosis es normal se tiene como expresión para P_i

$$P_i = \int_{-\infty}^{X_i} \frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{1}{2\sigma^2} (t_1)^2 \right) dt_1 \dots (1)$$

En particular si $P_i = \frac{1}{2}$

$$\frac{1}{2} = \int_{-\infty}^{k_0} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (t - u)^2\right) dt$$

Por lo que $X_0 = u$ y $D_0 = e^u$

Por influencia de los experimentos en toxicología, a D_0 se le llama la dosis letal 50 ó DL50, es decir la dosis necesaria para que el 50% de los individuos sujetos al estímulo mueran.

El problema entonces es estimar u , la media de la distribución del logaritmo de las tolerancias. Se requiere también estimar σ^2 para la obtención de intervalos de confianza. La forma tradicional de hacerlo es a través de la transformación Probit.

La primera transformación a la ecuación (1) fue propuesta por Gaddum quién definió la Derivación Normal Equivalente (D.N.E.) estandarizando la variable bajo la integral. Entonces.

$$P_i = \int_{-\infty}^{\frac{X_i - u}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Si $Z_i = \frac{X_i - u}{\sigma}$ entonces Z_i es la D.N.E. propuesta por Gaddum.

Bills propuso la llamada transformación Probit, la cual se obtiene simplemente sumando 5 a cada D.N.E. El propósito es en la practica los numeros negativos. Entonces si Y es el probit de una dosis

$$Y = 5 + Z = 5 + \frac{X - u}{\sigma}$$

Mediante el uso de la tabla normal estandar cualquier proporción puede transformarse a probits. Como ejemplo suponga que se tiene una proporción $V_c .67$.

$$.67 = \int_{-\infty}^{\frac{X_i - u}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$Z_i = \frac{X_i - u}{\sigma} = .44$$

y por lo tanto $Y_i = 5 + .44 = 5.44$

Estimación.

En un experimento en el que se va a usar análisis probit,

se tiene el siguiente conjunto de datos (D_i, N_i, r_i) $i=1, \dots, K$ Los cuales, mediante la transformación logarítmica; se convierten en

$$(X_i, N_i, r_i) \quad i=1, \dots, k$$

La proporción de individuos que responden a cada dosis está dada por

$$\hat{P}_i = \frac{r_i}{n} \quad i=1, \dots, k$$

Si obtenemos el probit de cada P_i , tendremos los valores Y_1, \dots, Y_k , los cuales tienen como expresión

$$Y_i = 5 + \frac{X_i - u}{\sigma} \quad i=1, \dots, k$$

lo que puede escribirse en la forma más conocida de una ecuación de regresión lineal simple como sigue

$$Y_i = \alpha + \beta X_i \quad ; \quad i=1, 2, \dots, K$$

donde $\alpha = 5 - \frac{u}{\sigma}$ $\beta = \frac{1}{\sigma}$

El problema es estimar u y σ

Estimación por máxima verosimilitud

Considerese la teoría (D_i, N_i, r_i) donde, como antes:

iésima. D_i = dosis iésima

N_i = número de individuos que reciben la dosis iésima

r_i = número de individuos que responden a la dosis iésima

P_i = proporción (desconocida) de individuos que son susceptibles a la dosis iésima

$Q_i = 1 - P_i$

Entonces la función de probabilidades r_i está dada por la distribución binomial

$$F(r_i) = \begin{cases} \binom{N_i}{r_i} P_i^{r_i} Q_i^{N_i - r_i} & ; r_i = 0, 1, 2, \dots, N_i \\ 0 & \text{de otra forma} \end{cases}$$

Por lo tanto

$$F(r_1, \dots, r_k) = \prod_{i=1}^k \left[\binom{N_i}{r_i} P_i^{r_i} Q_i^{N_i - r_i} \right]$$

El logaritmo de la función de verosimilitud de las observaciones es proporcional a

$$l(P_1, \dots, P_k) = \sum_{i=1}^k r_i \log P_i + \sum_{i=1}^k (N_i - r_i) \log Q_i$$

Entonces para cualquier θ , función de P_i

$$\begin{aligned} \frac{\delta \ell}{\delta \theta} &= \sum \frac{k}{P_i} \frac{r_i}{P_i} \frac{\delta P_i}{\delta \theta} + \sum \frac{R}{\theta_i} \frac{N_i - r_i}{\theta_i} \frac{\delta \theta_1}{\delta \theta} \\ &= \sum \frac{r_i}{P_i} \frac{\delta P_i}{\delta \theta} + \sum \frac{N_i - r_i}{\theta_i} (-1) \frac{\delta P_i}{\delta \theta} \\ &= \sum \frac{\delta P_i}{\delta \theta} \left[\frac{r_i - N_i P_i}{P_i Q_i} \right] \\ \frac{\delta \ell}{\delta \theta} &= \sum \frac{N_i (\hat{P}_i - P_i)}{P_i Q_i} \frac{\delta P_i}{\delta \theta} \quad \dots (9) \end{aligned}$$

además

$$P_i = \int_{-\infty}^{\frac{X_i - u}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi\left(\frac{(X_i - u)}{\sigma}\right)$$

$$P_i = \Phi(\alpha + \beta X_i - 5)$$

$$\frac{\delta P_i}{\delta \alpha} = \phi(\alpha + \beta X_i - 5)$$

$$\frac{\delta P_i}{\delta \beta} = \theta(\alpha + \beta X_i - 5) X_i$$

usando estos resultados en (9)

$$\frac{\delta \ell}{\delta \alpha} = \sum \frac{N_i (\hat{P}_i - P_i)}{P_i Q_i} \phi(Z_i)$$

$$\frac{\delta \ell}{\delta \beta} = \sum \frac{N_i (\hat{P}_i - P_i)}{P_i Q_i} \phi(Z_i) X_i$$

Igualando a cero las ecuaciones anteriores y resolvieron para α y β se obtendrían los estimadores de máxima verosimilitud. Es decir los estimadores son la solución a

$$\frac{\delta \ell}{\delta \hat{\alpha}} = 0 \quad \frac{\delta \ell}{\delta \hat{\beta}} = 0$$

El sistema de ecuaciones no tiene una solución explícita, sin embargo existe un método iterativo, el cual será tratado posteriormente.

II. MODELO MATEMATICO

Planteamiento.

Los modelos lineales generalizados son una combinación de componentes sistemáticos y componentes aleatorios. Caracterizado por:

- 1) Una variable dependiente Z proveniente de una distribución con función de densidad

$$\Pi(Z, \theta, \phi) = \exp [\alpha(\phi) \{Z\theta - \gamma(\theta) + h(Z)\} + B(\phi, Z)]$$

donde $\alpha(\phi) > 0$, para ϕ fijo se trata de una familia exponencial. El parametro ϕ puede ser un parametro tal como la varianza σ^2 de una distribución normal o el parámetro P de una distribución gamma. La media de Z es denotada por μ .

Es posible encontrar expresiones para la primera y segunda derivada del logaritmo de la función de verosimilitud en términos de la media y varianza de Z y el factor escalar $\alpha(\phi)$.

Usando los resultados:

$$E(\delta L / \delta \theta) = 0 \quad \text{y} \quad E(\delta^2 L / \delta \theta^2) = -E(\delta L / \delta \theta)^2$$

donde $L(Z; \theta, \phi) = \text{Log } \Pi(Z; \theta, \phi)$

$$\Pi(Z, \theta, \phi) = \exp [\alpha(\phi) \{Z\theta - g(\theta) + h(Z)\} + B(\phi, Z)]$$

$$L = \alpha(\phi) (Z\theta - g(\theta) + h(Z)) + B(\theta, Z)$$

$$\frac{\delta L}{\delta \theta} = \alpha(\phi) (Z - g'(\theta))$$

$$E \frac{\delta L}{\delta \theta} = E(\alpha(\phi) (Z - g'(\theta))) = 0$$

$$\alpha(\phi) (u - g'(\theta)) = 0$$

Entonces $u = E(Z) = g'(\theta)$

$$y \quad \frac{\delta L}{\delta \theta} = \alpha(\phi) (Z - u)$$

Ahora $\frac{\delta^2 L}{\delta \theta^2} = -\alpha(\phi) g''(\theta)$

$$E \frac{\delta^2 L}{\delta \theta^2} = -\alpha(\phi) g''(\theta)$$

por otro lado

$$-E \left(\frac{\delta L}{\delta \theta} \right)^2 = -E [\alpha(\phi) (Z - u)]^2$$

$$-E \left(\frac{\delta L}{\delta \theta} \right)^2 = -[\alpha(\phi)]^2 E(Z - u)^2 = -[\alpha(\phi)]^2 \text{Var}(Z)$$

Entonces

$$\alpha(\phi) g''(\theta) = [\alpha(\phi)]^2 \text{Var}(Z)$$

por lo que

$$g''(\theta) = \alpha(\phi) \text{Var}(Z) = V$$

es decir V es la varianza de Z cuando el factor escalar es unitario.

$$\frac{\delta^2 L}{\delta \theta^2} = -\alpha(\phi) V$$

se puede notar también que

$$V = g''(\theta) = \frac{\partial \mu}{\partial \theta}$$

lo anterior constituye la parte aleatoria del modelo

2) Un conjunto de variables independientes X_1, \dots, X_m y predictor

$$Y = \sum B_i X_i$$

donde suponemos conocidos los valores de las X_i y las B_i representan parametros, cuyos valores pueden ser fijos conocidos o pueden ser desconocidos y requieren estimación.

Una variable independiente puede ser cuantitativa y producir una variable X - cuantitativa en el modelo, o cualitativa y producir un conjunto de X - variantes cuyos valores son 0 y 1, ó puede ser un modelo con mezcla de los 2 tipos de variable.

Considerese el modelo

$$Y_{ij} = \alpha_i + \beta u_{ij} + \gamma_j v_{ij} \quad (i=1, \dots, n, j=1, \dots, P)$$

los datos son "acomodados" por factores cuyos niveles son denotados por i y j . El término α_i incluye n parámetros asociados con una variación cualitativa representada por n indicadores, (mudos ó componentes dummy) que toman el valor de 1 para un nivel y 0 para el resto; βu_{ij} represente una variación cuantitativa, y $\gamma_j v_{ij}$ demuestra P parámetros asociados con variables mixtas independientes cuyos P componentes toman los valores v_{ij} para un nivel de j y cero por el resto.

- 3) Una función de enlace $\theta = F(Y)$ que relaciona el parámetro θ de la distribución de Z con las Y 's del modelo lineal cuando Z esta normalmente distribuída con media 0 y varianza σ^2 y cuando $\theta = Y$, se trata de un modelo lineal ordinario con errores normal.

Análisis de Desviación (Deviance).

Un modelo lineal se dice ordenado si el ajuste de la β 's es realizado en la misma secuencia como son declarados en el modelo. La ordenación (o ordenación parcial) puede ser producida por la estructura del modelo, por ejemplo no tiene sentido ajustar el término de interacción $(ab)_{ij}$ antes de ajustar los correspondientes efectos a_i y b_j . Es posible también que el orden sea implicado por los objetivos del ajuste es decir una tendencia debe ser removida antes del ajuste de efectos adicionales. Más comunmente, el ordenamiento es en alguna parte arbitrario, y eso ocasiona un problema de inferencia el cual no es tratado. Para la exposición de las ideas básicas se supone que el modelo está ordenado y los términos se van a ajustar secuencialmente. Los objetivos del ajuste son determinar cuantos términos son requeridos para una descripción adecuada de los datos y derivar de ahí la estimación asociada de los parámetros y la matriz de información.

Dos modelos extremos son concebibles para un conjunto de datos, el modelo mínimo el cual contiene el conjunto más pequeño de términos que el problema permite y el modelo completo en el cual todas las Y's son diferentes y el ajuste a los datos es perfecto, entonces $\hat{u} = \mathbf{z}$.

El caso extremo del modelo mínimo es el modelo nulo, el cual es equivalente a ajustar solamente la gran media y asignar : efectivamente toda la variación en los datos al componente aleatorio del modelo, lo mismo que el modelo completo que ajusta exactamente y asigna toda la variación con los datos a la parte sistemática. El proceso de ajuste del modelo con un modelo ordenado consiste en un proceder a ajustar el modelo partiendo de un modelo mínimo hacia el completo. A cada paso se trata de incrementar la bondad de ajuste incrementando la complejidad del modelo. El ajuste de los parámetros en cada paso es hecho maximizando la función de verosimilitud para el modelo correspondiente y el ajuste del modelo a los datos se medirá cuantitativamente a través de la estadística $-2L$ max. La cual es llamada "Darvance". Para las cuatro distribuciones mencionadas la desviación toma la forma

$$\text{Normal} \quad \Sigma (Z - \hat{u})^2 / \sigma^2$$

$$\text{Poisson} \quad 2 \{ \Sigma Z \log (Z/\hat{u}) - \Sigma (Z - \hat{u}) \}$$

$$\text{Binomial} \quad 2 \{ \Sigma Z \log (Z/\hat{u}) + \Sigma (N-Z) \log \{ (N-Z) / (N-\hat{u}) \}$$

$$\text{Gamma} \quad 2P \{ - \Sigma \log (Z/\hat{u}) + \Sigma (Z - \hat{u}) / \hat{u} \}$$

Asociado con cada modelo está el término r , los grados de libertad, el cual es dado por el rango de la matriz X , o equivalentemente el número de parámetros linealmente independientes

que se estimaron. Para una muestra de n observaciones independientes la "desviación" del modelo tiene grados de libertad residuales $(n-r)$. Los grados de libertad multiplicados, cuando es necesario, por el factor escalar formar una escala para un conjunto de modelos secuenciales, en el cual la desviación puede ser comparada. El factor escalar puede ser conocido (por ejemplo en la distribución Poisson) o desconocido (para la distribución normal con varianza desconocida). Si es desconocida puede ser estimada directamente por las observaciones o indirectamente por la desviación después de que un modelo adecuado fue ajustado. Lo adecuado del modelo puede ser determinado en algunas ocasiones al realizar una gráfica.

Ejemplos de algunos modelos lineales generalizados.

Distribución Normal

En el caso normal se tiene

$$L = \sigma^{-2} \left(Z\mu - \frac{1}{2} \mu^2 - \frac{1}{2} Z^2 \right) - \frac{1}{2} \log \sigma^2$$

en donde al comparar con el $L = \alpha(\phi) (Z\theta - \gamma(\theta) + h(Z)) + \beta(\phi, Z)$

$$u = 0, \quad \alpha(\phi) = \sigma^{-2}, \quad \gamma(\theta) = \frac{1}{2} \theta^2 = \frac{1}{2} \mu^2$$

$$h(Z) = -\frac{1}{2} Z^2, \quad \beta(\theta, Z) = -\frac{1}{2} \log \sigma^2 \text{ y por lo tanto}$$

$$V = \frac{d\mu}{d\theta} = 1$$

Los Polinomios inversos proporcionan un ejemplo donde las observaciones u son normales en la escala logarítmica y los efectos sistemáticos aditivos en la escala inversa.

$$Z = \log u \quad \text{y} \quad Y = e^u = \sum \beta_i X_i$$

Más generalmente, se pueden considerar modelos en los cuales hay una transformación para linealizar F y una transformación para normalizar g , tal que si las observaciones son denotadas por u entonces $g(u)$ está normalmente distribuido con media u y varianza constante σ^2 y $F\{g^{-1}(\mu)\} = \sum \beta_i X_i = Y$

entonces se tiene

$$V = 1 \quad \text{y} \quad Y = F\{Y^{-1}(\mu)\}$$

$$\text{y} \quad Y = Y + \{g(\sigma) - \mu\} / \{ (d/dY) g\{F^{-1}(Y)\}\}$$

$$E(Y) = \sum \beta_i X_i$$

La distribución Poisson

En este caso se tiene

$$L = Z \log \mu - \mu$$

en donde $\theta = \log \mu$ y $v = \frac{d\mu}{d\theta} = \mu$

Cuando $Y = \sum \beta_i X_i \log \mu$ hay estadística suficiente y un único estimador máximo verosímil de β , si es finito. Siempre será finito si no hay observaciones cero.

Si $Y = \sum \beta_i X_i \mu^\lambda$ ($0 < \lambda < 1$), $L \rightarrow \infty$ cuando $|\beta| \rightarrow \infty$ y por lo tanto L debe tener un máximo para un valor finito. En
tonces

$$\frac{\delta^2 L}{\delta \beta_i \delta \beta_j} = \frac{\delta^2 L}{\delta Y^2} X_i X_j$$

Es fácilmente verificable que $\delta^2 L / \delta Y^2 < 0$ y por esa razón L es negativa definida. Por lo tanto $\hat{\beta}$ es determinado uni
vocamente. Cuando $Y = \sum \beta_i X_i = \mu$ los mismos resultados se man
tienen con tal que la X 's sean lineales cuando las unidades con $Z=0$ son excluidas.

La principal de aplicación de los modelos lineales generalizados con errores Poisson son las tablas de contingencia. Los modelos probabilísticos para tablas de contingencia son constru
idos suponiendo una distribución multinomial o un conjunto de dis
tribuciones multinomiales. (Birch demostró que la estimación de un conjunto independiente de distribuciones multinomiales es equi

valente a la estimación de un conjunto de distribuciones Poisson independientes, y por esa razón se estima las tablas de contingencia como un conjunto de distribuciones Poisson in dependientes).

La distribución Binomial

Si reescribimos la forma usual

$$L = N \log p + (N - r) \log q$$

como

$$\begin{aligned} L &= Z \log (\mu/N) + (N-Z) \log \{1-(\mu/N)\} \\ &= Z \log (\mu/(N-\mu)) + N \log (N-\mu) + \text{terminos en } N \end{aligned}$$

es decir, substituyendo Z por r y $\mu = E(Z) = NP$.

Así

$$\theta = \text{en } \{\mu/(N-\mu)\} \quad , \quad g(\theta) = N \text{ en } (N-\mu)$$

$$V = \mu (N-\mu)/N.$$

Estadísticas suficientes son obtenidas por la transformación logarítmica dada por

$$\mu = Ne^Y / (1 + e^Y)$$

Probit Analysis

En probit analysis se tiene

$$\mu = N \Phi(Y)$$

donde Φ es la distribución normal acumulativa. En este caso no hay estadísticas suficientes.

La distribución Gama

$$L = -P (Z/\mu + \log \mu - \log Z) - \log Z$$

$$\alpha(\phi) = -P \quad g(\sigma) = -\log \mu, \quad h(Z) = -\log Z \quad \text{y} \quad \beta(\phi, Z) = -\log Z$$

Es decir $\theta = -1/\mu$ y $V = du/d\theta = \mu^2$. Hay estadísticas suficientes cuando $Y = 1/\mu$

Una aplicación de los modelos lineales que involucran la distribución Gama es la estimación de los componentes de varianza. La suma de cuadrados son proporcionales a X y la esperanza es una combinación de varias varianzas las cuales son estimadas.

Se puede escribir:

$$L = - (Z/2\mu) - (V/2) \log \mu + (V/2 + 1) \log Z$$

donde V son los grados de libertad de Z

$$\theta = - V/2\mu \quad V - du/d\theta = 2\mu^2/V \quad Y = Z$$

$$Y = \mu \quad y \quad W = V/2\mu^2$$

La desviación toma la forma

$$\Sigma V [-en (Z/\mu) + \{ (Z \cdot \hat{\mu}) / \hat{\mu} Y \}$$

y cuando $\Sigma V (Z - \hat{\mu}) / \hat{\mu} \} = 0$ se simplifica a $\Sigma V \log (\hat{\mu}/Z)$

Relación con Glim

Glim es un paquete diseñado para realizar el ajuste de modelos lineales generalizados como los descritos anteriormente. Consta de una parte central concerniente con la especificación y ajuste de los modelos y facilidades auxiliares tales como la lectura de los datos transformaciones de los datos así como la posibilidad de hacer varios tipos de gráficas para auxiliar al usuario en la interpretación de los datos. Glim es probablemente más útil cuando es usado de manera exploratoria, como herramienta para experimentar con los ajustes de varios modelos para un conjunto de datos dados, ya que se puede utilizar el paquete en forma interactiva.

La manera de ajustar un modelo lineal generalizado es por medio de las siguientes directivas.

YVARIATE

ERROR

LINK

WEIGHT

OFFSET

SCALE

En donde

YVARIATE corresponde a la declaración de la variable dependiente.

Su sintaxis es

\$ YVARIATE identificador

en donde el identificador se refiere a una variable declarada anteriormente. Para datos con distribución Binomial, donde se debe especificar la variable Y en la observación Y_i en la forma r_i/N_i donde r_i es el total de éxitos y N_i el total de ensayos realizados. La YVARIATE consistirá del numerador r_i ; la especificación de los valores N_i se hace en la directiva del error.

ERROR en esta parte se especifica la distribución del

componente aleatorio. Su sintaxis es

```
$ ERROR nombre [identificador]
```

donde el nombre puede ser:

NORMAL, POISSON, BINOMIAL ó GAMMA

Si es omitida la directiva por default la distribución se considera normal. El identificador es necesario solamente con distribución Binomial, este consiste de un vector donde se especifica el total de ensayos N_i realizados en cada observación Y_i .

LINK, se refiere a la función de enlace que relaciona al predictor y el parametro de la distribución del componente aleatorio. Su sintaxis es

```
$ LINK letra [número]
```

G	Logit
R	Reciproca
P	Probit
C	Log-Log
S	Raíz Cuadrada
E	Exponenciación
I	Identica
L	Log

Solo es necesario una letra y en el caso de la exponencia ción es necesario dar el número al cual se desea elevar el para metro, el cual debe ser diferente de cero. La omisión de la di rectiva LINK causa por default que las siguientes funciones se usen:

Error	link
N (Normal)	I (Idéntica)
P (Poisson)	L (Log)
G (Gamma)	R (Reciproca)

Estas funciones de enlace proporcionan estadística sufi cientes para sus respectivas distribuciones. La declaración LINK debe aparecer después (no necesariamente inmediatamente) a la declaración ERROR. Las funciones de enlace G, P y C son apropiadas para datos con distribución binomial solamente, y no pueden ser usados con otra distribución.

SCALE

De las cuatro distribuciones standard de las cuales GLIM permite describir la estructura del error se dividen en dos pares. Un par comprende la binomial y Poisson, las cuales tienen una relación entre media y varianza conocida completamente, para el otro par, la Normal y Gamma, la relación contiene un

parámetro escalar ϕ , el cual es generalmente desconocido. Para la distribución Normal, se tiene $\phi = \sigma^2$, la varianza, mientras que para la distribución Gamma

$$\text{Var}(Y_i) = K \mu^2$$

y se tiene $\phi = K$, el cuadrado del coeficiente de variación.

Como la matriz de varianza de $\hat{\beta}$ es función de ϕ , entonces el valor de ϕ es substituido para calcular la matriz y otros valores como los errores standard derivados en él.

Entonces la sintaxis de la directiva es.

\$ SCALE [numbers]

donde el número es no-negativo. Si el número es positivo entonces se asume ese valor para ϕ . Si el número es cero ó no existe entonces ϕ es estimado por la desviación de la media del modelo correspondiente.

La ausencia de la declaración de SCALE en el caso de Binomial y Poisson, implica por default el valor de ϕ igual a 1, mientras que en Normal y Poisson implica la estimación ϕ .

FIT esta directiva define la estructura lineal del modelo que se va a ajustar. Su sintaxis es.

\$ FIT formula del modelo

La formula del modelo asemeja una expresión aritmética ya que consta de operandos, operadores y paréntesis. Los operandos son factores o variables, los operadores comprenden el conjunto siguiente de símbolos con las jerarquías indicadas.

	.	/	*	,	+	-	-/	-*	
mayor	1°	2°	3°	4°	4°	4°	4°	4°	menos

El significado de los operadores, es diferentes al que tendría en una expresión aritmética común.

Los términos básicos en la fórmula del modelo corresponden a las tres posibles clases de términos en el modelo lineal, es decir cuantitativo, cualitativo y mixto.

En la fórmula del modelo los tres términos se describen como sigue:

En el modelo lineal	En la fórmula del modelo
β_i	X
α_j	A
$\beta_j X$	A.X

Los operadores + y , son sinónimos en este contexto, y son usados para unir términos en la fórmula del modelo. Así un modelo de regresión con tres variables aparece simplemente como

$$X_1, X_2, X_3$$

o equivalentemente

$$X_1 + X_2 + X_3$$

Para el caso de modelos anidados, con dos factores A y B donde A indica un grupo de unidades y B indica los elementos dentro del grupo, el modelo lineal correspondiente puede escribirse como A/B donde / es el operador de anidamiento. Los modelos anidados se expanden en términos básicos como sigue:

A/B	equivale a	A + A.B
A/B/C	equivale a	A + A.B + A.B.C

Cuando A.B no está precedido por B, su interpretación es B dentro de A. El término B no ocurre ya que no hay relación entre el elemento j del grupo 1 y el elemento j de el grupo 2.

El operador * define el modelo lineal para una clasificación "cruzada" de 2 ó más factores. Así,

$A*B$ es equivalente a $A + B + A.B$
 $A*B*D$ es equivalente a $A + B + B + A.B$
 $+ A.C + B.C + A.B.C.$

cuando el término $A.B$ es precedido por A y B , su interpretación es AxB interacción. Similarmente $A.B.C$ es la interacción de tres factores.

El uso de los paréntesis es similar al uso común en expresiones aritméticas. Así

$(A*B).(C+D)$ es equivalente a $(A+B+A.B).(C+D)$

que a su vez es equivalente a

$A.C + A.D + B.C + B.D + A.B.C + A.B.P$

Existe también el operador $-$, el cual retira el término siguiente a él, Así.

$A*B - A.B$ es $A + B$

$A*B*B - A,B(B*C)$ es $A + B + C + B.C$

El operador $-*$ remueve el siguiente término y todos los términos presentes en el modelo que lo contengan, Así.

$A*B*C - * B.C.$ se convierte $A+B+C+A.B+A.C$

El operador $- /$ remueve todos los términos que contienen al término siguiente pero no el término en si mismo. Así.

$A*B*C - /A$ se convierte en $A+B+C+B.C$

El modelo especificado normalmente incluye la gran media implícitamente como el primer término. El modelo nulo es permitido e implica el ajuste de la gran media solamente. Para el manejo de la gran media es necesario referirse explícitamente (y puede ser removida o re-incluída en el modelo) usando el término especial % GM. La manera de removerla es anteponiendo el operador $-$ y reincluírla por medio de $+$. La gran media no puede ser removida de el modelo nulo.

Estimación por Mínimos Cuadrados

Sea $\chi = (X_1, \dots, X_n)'$ un vector aleatorio

$$\chi = A\beta + \varepsilon$$

donde A es una matriz conocida de orden $n \times p$ con $p < n$

$\beta = (\beta_1, \dots, \beta_p)'$ es un vector de parámetros desconocidos

$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ es un vector de desviaciones de la media, o errores, cuyo valor esperado es cero.

Si estimamos β minimizando la suma de cuadrados

$$\sum \varepsilon_i^2 = \varepsilon' \varepsilon = (X - A\beta)' (X - A\beta)$$

El valor mínimo $\hat{\beta}$ es llamado el estimador mínimos-cuadrados de β

El método para la obtención del estimador mínimo-cuadrado es obteniendo las derivadas de la forma cuadrática, con respecto a β_1, \dots, β_p , se obtienen así llamadas ecuaciones normales

$$(A' A) \hat{\beta} = A' X$$

y una solución de esas ecuaciones en efecto minimiza

$$(X - A\hat{\beta})' (X - A\hat{\beta})$$

Prueba

$$\begin{aligned} (X - A\hat{\beta})' (X - A\hat{\beta}) &= [X - A\hat{\beta} + A(\hat{\beta} - \beta)]' [X - A\hat{\beta} + A(\hat{\beta} - \beta)] \\ &= [(X - A\hat{\beta})' + (\hat{\beta} - \beta)' A'] [X - A\hat{\beta} + A(\hat{\beta} - \beta)] \end{aligned}$$

$$= (X-A\hat{\beta})^1 (X-A\hat{\beta}) + (\hat{\beta}-\beta)^1 A^1 (X-A\hat{\beta}) + (X-A\hat{\beta})^1 A (\hat{\beta}-\beta) + (\hat{\beta}-\beta)^1 A^1 A (\hat{\beta}-\beta)$$

y como $A^1 A \hat{\beta} = A^1 X$

entonces $A^1 (X-A\hat{\beta}) = 0$

y $(X-A\hat{\beta})^1 A = 0$

Entonces

$$(X-A\beta)^1 (X-A\beta) = (X-A\hat{\beta})^1 (X-A\hat{\beta}) + (\hat{\beta}-\beta)^1 A^1 A (\hat{\beta}-\beta)$$

y como $(\hat{\beta}-\beta)^1 A^1 A (\hat{\beta}-\beta) \geq 0$

Entonces para toda β

$$(X-A\hat{\beta})^1 (X-A\hat{\beta}) \leq (X-A\beta)^1 (X-A\beta)$$

Si A es de rango completo entonces $A^1 A$ es de rango completo i.e no-singular y existe un único estimador mínimos cuadrados.

$$\hat{\beta} = (A^1 A)^{-1} A^1 X$$

Si el rango de $A < p$, entonces $A^1 A$ es singular, las ecuaciones normales no tienen solución única, es decir existe una familia de estimadores mínimos cuadrados la cual puede ser determinada en un caso particular por los métodos usuales de resolución de sistemas de ecuaciones.

Teorema de Gauss-Markov

Sea X un vector aleatorio en \mathbb{R}^n expresado en la forma

$$X = A\beta + \epsilon$$

donde A es una matriz conocida $n \times p$ β es un vector en \mathbb{R}^p desconocido y ϵ es un vector de errores en $E(\epsilon) = 0$ y $\text{Var}(\epsilon) = \sigma^2 I$ donde σ^2 es desconocida.

Es decir los componentes de ϵ tienen la misma varianza desconocida y son no correlacionados. Sea $\hat{\beta}$ un estimador mínimo cuadrado de β y sea $\phi = c^1\beta$ una función lineal paramétrica. Entonces $c^1\hat{\beta}$ es un estimador de ϕ y si $\tilde{\phi}$ es algún otro estimador lineal insesgado de ϕ y $\text{Var}(c^1\hat{\beta}) \leq \text{var}(\tilde{\phi})$.

El teorema de Gauss-Markov asegura que el estimador mínimo cuadrado de β es el mejor cuando los componentes no están correlacionados, sin embargo cuando los componentes de ϵ están correlacionados el estimador mínimos cuadrados puede no ser el mejor, ya que al suponer que la matriz de varianza del error es $\sigma^2 I$, estamos suponiendo que los componentes tienen la misma varianza y que no están correlacionados.

Existe una modificación al método de mínimos cuadrados cuando los errores tienen diferente varianza ó están correlacionados.

Consideremos el modelo

$$X = A\beta + \epsilon$$

Con las mismas suposiciones que antes excepto que ver $\epsilon = \sigma^2 \Sigma$, donde Σ es una matrix conocida, positiva definida.

Ahora bien si Σ es positiva definida, puede ser expresada de la forma PP^1 donde P es no-singular y si $\eta = P^{-1}\epsilon$, $Y = P^{-1}X$. Podemos escribir el modelo de la forma

$$P_y = A\beta + P\eta$$

$$X = P_y \quad \epsilon = P\eta$$

o equivalentemente

$$Y = P^{-1} A\beta + P^{-1} P\eta$$

es decir

$$Y = B\beta + \eta$$

donde

$$\text{Var } \eta = P^{-1} (\text{Var } \epsilon) P^{1-1} = \sigma^2 P^{-1} P P^1 P^{1-1} = \sigma^2 I$$

$$\text{y } B = P^{-1}A$$

Es decir en términos de Y el modelo es el ya discutido, entonces de acuerdo al Teorema de Gauss-Markov se obtiene un mejor estimador de β minimizado

$$(Y-B\beta)' (Y-B\beta)$$

donde

$$\begin{aligned} (Y-B\beta)' (Y-B\beta) &= (X-A\beta)' (P'P)^{-1} (X-A\beta) \\ &= (X-A\beta)' \Sigma^{-1} (X-A\beta) \end{aligned}$$

Si A es de rango completo

$$\begin{aligned} \hat{\beta} &= (B'B)^{-1} B'Y = (A'(P'P)^{-1} A)^{-1} P'P^{-1}Y \\ &= (A'\Sigma^{-1} A)^{-1} A' (P^{-1}) (P^{-1} X) \\ &= (A' \Sigma^{-1} A)^{-1} A' \Sigma^{-1} X \end{aligned}$$

Estimadores de Máxima Verosimilitud.

Mientras que la aplicación del método de mínimos cuadrados no requiere el conocimiento de la distribución del vector de errores, aparte de la media y matriz de varianza, el método de máxima verosimilitud es aplicado principalmente en las situaciones donde la verdadera distribución en el espacio muestral es conocida, excepto por los valores de un número finito de parámetros reales desconocidos.

El método de máxima verosimilitud es usualmente aplicado cuando la familia de posibles distribuciones en el espacio muestral puede ser identificado por un parámetro θ tomando valores en un espacio euclidiano de dimensión finita. En suma,

está aplicación es generalmente restringida en el caso donde está familia

$$\{P_{\theta} : \theta \in \Omega\} \quad (\Omega \subset \mathbb{R}^j)$$

pose función de densidad, $\{P_{\theta} : \theta \in \Omega\}$

En el caso discreto $P_{\theta}(X, \theta)$ es la probabilidad en el punto X cuando θ es el parámetro verdadero; en el caso continuo $P_{\theta}(X, \theta)$ es la densidad de probabilidad en X cuando θ es el parámetro verdadero.

Es conveniente hacer una distinción entre $P(\cdot, \theta)$ la cual es la función de densidad en el espacio muestral y la función $P(X, \cdot)$ la cual es una función en el espacio parametral. La función $P(X, \cdot)$ es llamada la función de verosimilitud.

El método de máxima verosimilitud, como su nombre lo indica consiste en estimar el verdadero valor del parámetro θ como el valor que maximiza la función de verosimilitud; este parámetro resulta ser el más plausible después de haber observado X .

En la mayoría de los casos, existe un único valor que maximiza el cual es el más plausible y es el estimador máximo verosimil de θ .

Definición.

El estimador máximo verosímil $\hat{\theta}(X)$ es un elemento de Θ dado por

$$P\{X, \hat{\theta}(X)\} = \max_{\theta \in \Theta} P(X, \theta)$$

Si esto es posible, si por el contrario Θ es un conjunto abierto, el estimador máximo verosímil no existe. En la práctica por lo general no existen inconvenientes.

Obtención.

Por lo general no es posible la obtención directa del estimador máximo verosímil es necesario recurrir a métodos numéricos para determinarlos. Es usualmente posible asumir que el estimador máximo verosímil emerge como una solución de las ecuaciones normales de verosimilitud

$$\frac{\delta}{\delta \theta_i} \log(X, \theta) = 0 \quad (i=1, \dots, S)$$

Estas ecuaciones en general pueden resolverse numéricamente.

Propiedades del estimador máximo verosímil.

El teorema de Gauss Markov da una justificación para el método de mínimos cuadrados en términos de estructuras lineales insesgados de mínima varianza. En cuanto al estimador de máxima verosimilitud no siempre es insesgado, sin embargo existen argumentos que justifican parcialmente el método.

Primero en la situación donde existe un estimador insesgado cuya varianza alcanza la frontera inferior de Cramer Rao el estimador máximo verosimil coincide con el. Es posible demostrar que el estimador máximo verosimil es eficiente y función de la estadística mínima suficiente. Sin embargo la justificación más importante del método de máxima verosimilitud es cuando se trata de una muestra grande.

Propiedades con muestras grandes

Se hablará de una muestra grande, cuando X es la forma $X=(X_1, \dots, X_n)$ donde n es grande y las X_j son independientes e idénticamente distribuidas.

Si $P^*(X, \theta)$ es la función de densidad correspondiente al valor del parámetro θ en el espacio de una observación simple. Entonces,

$$l(X, \theta) = \log P(X, \theta) = \sum_{Z=1}^n \log P^*(X, \theta)$$

es una variable aleatoria, ya que es la suma de variables aleatorias idénticamente distribuidas.

La justificación del método de máxima verosimilitud en el término del criterio de insesgados de mínima varianza, es posible demostrar que para muestras grandes, el estimador máximo verosimil es aproximadamente insesgado y tiene varianza aproximadamente igual a la frontera inferior de Cramer-Rao. Es más, se puede demostrar que $\hat{\theta}$ tiene aproximadamente una distribución

$$N(\theta, i)$$

donde i es la cota inferior de Cramer-Rao.

Solución Iterativa a las ecuaciones M.V.

En caso de que las ecuaciones máximo verosimiles no tengan solución explícita, es necesario utilizar métodos numéricos esto es posible al asumir que el estimador máximo verosimil es la solución de las ecuaciones máximo verosimiles es el método de Newton o una adaptación de él.

Las ecuaciones verosimiles son:

$$\frac{\delta}{\delta \theta_j} \log P(X, \theta) = 0 \quad (j=1, \dots, 5)$$

Las cuales podemos escribir simbólicamente como:

$$D_{\theta}L(X, \theta) = 0$$

Donde $l(X, \theta) = \log P(X, \theta)$ y D_{θ} es un operador, diferencial cuyo i -ésima componente es $\delta/\delta\theta_i$. Por la exploración de una característica especial acerca de la investigación (o por método gráfico) es posible obtener una buena aproximación inicial al valor $\hat{\theta}$ solución de las ecuaciones, esa aproximación la denotamos por $\theta^{(0)}$. Desarrollando el Teorema de Taylor alrededor de $\hat{\theta}$ hasta los términos del primer orden se obtiene como $D_{\theta}L(X, \theta) = 0$

$$0 = D_{\theta}L(X, \theta) \approx D_{\theta}L(X, \theta^{(0)}) + \{D_{\theta}^2 L(X, \theta^{(0)})\}(\hat{\theta} - \theta^{(0)})$$

Donde D_{θ}^2 es la matrix operador

$$\left[\frac{\delta^2}{\delta\theta_i \delta\theta_j} \right]$$

Entonces

$$\hat{\theta} \approx \theta^{(0)} - \{D_{\theta}^2 L(X, \theta^{(0)})\}^{-1} D_{\theta}L(X, \theta^{(0)})$$

Después se repite el proceso, usando $\theta^{(1)}$ obtenido de $\theta^{(0)}$, obteniendo una nueva aproximación $\theta^{(2)}$ y así sucesivamente. Entonces se establece un procedimiento iterativo del que se obtiene una sucesión $(\theta^{(n)})$ la cual converge a $\hat{\theta}$.

El aspecto laborioso del proceso iterativo es la inversión de la matrix $D_{\theta}^2 l(X, \theta^{(i)})$ en el paso i -ésimo. Si muestra

aproximación inicial $\theta^{(0)}$ es buena, entonces $D_{\theta}^2 L(X, \theta^{(0)})$ debe ser cercana $D_{\theta}^2 L(X, \theta^{(i)})$, entonces es posible usar la primera matriz en cada paso de el proceso y eliminar la necesidad de una nueva matriz de inversión en cada paso. Este proceso modificado da una nueva sucesión de aproximaciones a $\hat{\theta}$, la cual converge a $\hat{\theta}$ posiblemente más lentamente que la sucesión $(\theta^{(n)})$.

Una modificación adicional, es suponer que la matriz $D_{\theta}^2 L(X, \theta^{(0)})$ es relativamente cercana a su esperanza $E_{\theta}^{(0)} D_{\theta}^2 L(X, \theta^{(0)})$, la sucesión de aproximación de $\hat{\theta}$ basada en la esperanza de $D_{\theta}^2 L(X, \theta^{(0)})$, lo mismo que en $D_{\theta}^2 L(X, \theta^{(0)})$ converge a $\hat{\theta}$.

El proceso iterativo será definido como sigue

$$\theta^{(n+1)} = \theta^{(n)} + B_{\theta}^{-1}(n) \{D_{\theta} L(v, \theta^{(n)})\}$$

donde $-B_{\theta}^{(0)} = E_{\theta}^{(0)} \{D_{\theta}^2 L(X, \theta^{(0)})\}$

Solución a Estimación máximo verosímil con modelos lineales.

La solución de las ecuaciones máximo verosímiles es equivalente a un proceso iterativo de estimadores ponderados por mínimos cuadrados cuya función de ponderación es

$$W = (du/dy)^2 / V$$

Consideremos a Y la variable dependiente y la expresión

$$y = Y + (Z-u)(du/dy)$$

donde u , Y y V son basadas en estimaciones comunes.

Considerando el modelo

$$\Pi(Z, \theta, \phi) = \exp [\alpha(\phi) (Z\theta - g(\theta) + h(Z)Y + B(\phi, Z))]$$

Prueba

Si L es el logaritmo de la función de verosimilitud de una observación.

$$L = \alpha(\phi) [Z\theta - g(\theta) + h(Z)] + B(\theta, Z)$$

Por la regla de la cadena y ya que θ es función de u , u de Y y g de las B_i 's.

$$\frac{\delta L}{\delta B_i} = \alpha(\phi) (Z - g'(\theta)) \frac{\delta \theta}{\delta u} \frac{\delta u}{\delta y} \frac{\delta y}{B_i}$$

$$= \alpha(\phi) (Z - g'(\theta)) \frac{\delta \theta}{\delta u} \frac{\delta u}{\delta y} X_j$$

$$= \alpha(\phi) (Z - u) \frac{1}{V} \frac{\delta u}{\delta y} X_j$$

$$y \quad \frac{\delta^2 L}{\delta \beta_i \delta \beta_j} = \frac{\delta^2 L}{\delta Y^2} \frac{\delta Y}{\delta \beta} \frac{\delta Y}{\delta \beta}$$

$$= \frac{\delta^2 L}{\delta Y^2} X_j X_j$$

donde

$$\begin{aligned} \frac{\delta^2 L}{\delta Y^2} &= \frac{\delta^2 L}{\delta \theta^2} \left(\frac{\delta \theta}{\delta Y} \right)^2 + \frac{\delta L}{\delta \theta} \frac{\delta^2 \theta}{\delta Y^2} \\ &= \alpha(\phi) \left\{ -V \frac{(\delta \theta)^2}{\delta u} \left(\frac{\delta u}{\delta y} \right)^2 (Z-u) \frac{\delta^2 \theta}{\delta y^2} \right\} \\ &= \alpha(\phi) \left\{ -\frac{(\delta \theta)^2}{\delta u} V + (Z-u) \frac{\delta^2 \theta}{\delta y^2} \right\} \end{aligned}$$

Entonces

$$\begin{aligned} E \left(\frac{\delta^2 L}{\delta B_i \delta B_j} \right) &= E \left(\alpha(\phi) \left\{ -\frac{(\delta u)}{\delta y} V + (Z-u) \frac{\delta^2 \theta}{\delta y^2} \right\} \right) \\ X_i X_j &= -\alpha(\phi) \left\{ \frac{(\delta u)^2}{\delta y} V \right\} X_i X_j \end{aligned}$$

Escribiendo ω como la función de ponderación, entonces

$$\begin{aligned} \frac{\delta L}{\delta B_i} &= \alpha(\phi) \frac{(Z-u)}{V} \frac{\delta u}{\delta y} X_i \omega + (V - \frac{(\delta u)^2}{\delta y}) \\ \frac{\delta L}{\delta B_i} &= \alpha(\phi) (Z-u) X_i \omega + \frac{(\delta u)}{\delta y} \end{aligned}$$

Así el proceso Newton-Raphson reemplazando el valor por el valor esperado, para una muestra de tamaño n , se tiene

$$A\delta B = C$$

donde A es una matriz de $m \times m$ con

$$A_{ij} = \sum_{k=1}^n \omega_k X_{ik} X_{jk}$$

y C es un vector de $m \times 1$ con

$$C_i = \sum_{k=1}^n \omega_k X_{ik} (Z-u) \quad (du/dy)$$

Finalmente se tiene

$$(AB)_i = [A_{ij} B_j] = \sum \omega_k X_{ik} Y_k$$

entonces $A\delta B = C$ puede ser escrito $AB^* = \gamma$

donde $\gamma_i = \sum \omega_k X_{ik} Y_k$

$$Y_k = Y_k^* = (Z_k - u_k) / (du_k/dy)$$

$$\beta^* = \beta \delta \beta$$

En la práctica se puede obtener un buen proceso de iteración de la siguiente manera: Se toma como una primera aproximación $u = Z$ y se calcula Y para esa u , se calcula ω y se hace $y=Y$. Entonces se obtiene una primera aproximación de las β 's por regresión. El método puede necesitar modificaciones al tratarse de valores extremos de Z . Por ejemplo cuando la distribución binomial es probable que sea más adecuado reemplazar ensayos de $Z=0$ ó $Z=n$ con $Z=\frac{1}{2}$ y $Z=n-\frac{1}{2}$, es decir con el análisis probit o transformaciones logit, $u=0$ ó $u=n$ produce valor infinito para Y .

Un caso importante especial ocurre cuando θ , el parámetro de la distribución de Z , y Y el predictor de modelo lineal coinciden. Entonces

$$L = Z Y - g(Y) + h(Z)$$

$$y \quad \frac{\partial L}{\partial \beta_i} = \alpha(\theta) (Z \cdot u) X_i$$

Las ecuaciones máximo verosímiles son entonces de la forma

$$\sum (Z \cdot \hat{u}) X_{ik} = 0$$

Per lo tanto

$$\sum_k Z_k X_{ik} = \sum_k \hat{u}_k X_{ik}$$

Ejemplos de Aplicaciones

En este capítulo se trataran problemas reales, cada uno de los cuales se ajustan a un modelo logaritmico lineal de los anteriormente descritos, los ajustes para este modelo se realizaron por medio del programa GLIM. (Generalized Linear Interactive Modelling)

En los problemas que se resolveran se encuentran el ajuste a las curvas que definen el porcentaje de germinación de semillas de cebada seca Caryopses, en relación a la dosis de radiación de 1.0 Me Velectrons y de 60 Co Gama, a las que han sido expuestas encontrandose en ambos casos la dosis mediana, es decir aquella dosis que permite la germinación del 50% de las semillas, así como otras dosis como la 70 es decir la que permite que germinen el 70% de semillas, e intervalos de confianza para estas.

Otro problema presentado en este capítulo se refiere a una investigación acerca de la eficiencia de un cierto detergente comparada con la de un detergente estandar.

El último problema se refiere a un estudio sobre una cierta variedad de gusanos, en el cual se pretende investigar si las diversas formas y colores que adquieren depende o no del sexo o de la localidad en la que se encuentren.

Ejemplo 1.

El objetivo del experimento es obtener las curvas del porcentaje de germinación de cebada seca *Carypses* que se obtiene al exponer las semillas a radiaciones de 1.0 Mev electrones y 60 Co gamma, bajo condiciones similares de experimentación. Posteriormente se determinó para los dos tipos de radiación la dosis mediana, es decir aquella dosis que permite que el 50% de las semillas germinen, con sus respectivos intervalos de confianza. También se determinaron otras dosis de interés con sus respectivos intervalos de confianza.

Este experimento fue realizado por G. Palomino, F. Nepamuceno y R. Villalobos-Pietrini en el laboratorio de Genética y Radiobiología, Departamento de Biología Experimental, Instituto de Biología, Universidad Nacional Autónoma de México.

Los datos con los que se cuenta es el número de semillas germinadas de 100 semillas que fueron expuestas, para cada dosis aplicada. De acuerdo al tipo de datos y al objetivo del problema, surgió la idea de ajustar un modelo Probit, sólo que en lugar de tratar de obtener la dosis mínima letal es decir aquella que permite que mueran el 50% de los animales, como tradicionalmente se hace se trata de obtener la dosis máxima que permite que no se afecten el 50% de las semillas, es decir que germinen el 50% de las semillas.

Los datos obtenidos fueron:

Dosis (Kr)	1.0 MeV electrons	⁶⁰ Co gamma radiation
0	100	100
50	100	97
100	99	99
125	100	95
150	100	95
175	99	96
200	100	91
300	98	97
400	94	92
500	94	93
600	85	87
700	59	83
800	50	76
900	41	59
1000	34	*
1100	24	*
1200	0	45
1400	0	23
1600	0	4
1800	0	0

No existe medida

Los datos fueron analizados por medio de GLIM considerando la función de enlace Probit y error Binomial.

Para el primer caso, 1.0 Mev electrons, se tiene al considerar al ajuste de solamente a la gran media tenemos que la desviación es de 1752 con 19 grados de libertad la cual es altamente significativa.

Si consideramos el ajuste incluyendo la dosis la desviación baja considerablemente siendo de 42.12 con 18 g.l. la cual resulta ser significativa al 5%.

Si consideramos el ajuste en relación a el logaritmo natural de la dosis de desviación se vuelve a incrementar resultando de 101.2 con 18 g.l. lo que indica que el ajuste es peor que sin aplicar el logaritmo a la dosis.

Si consideramos el ajuste de la exponencial de la dosis encontramos que el ajuste no converge en el ciclo 10, y la desviación vuelve a subir a 425.8 con 18 g.l.

Si calculamos ahora dosis al cuadrado y realizamos el ajuste, encontramos una desviación de 71.14 con 18 g.l.

Posteriormente considerando todas las observaciones en las que germinaron 100 semillas como 99.5 semillas germinadas, y las observaciones en que germinaron 0 semillas como .5 semillas germinadas.

Los ajustes que encontramos son los siguientes:

Al ajustar dosis la desviación es de 36.86 con 18 g.l.

Al ajustar log dosis la desviación es de 144.4 con 18 g.l.

Al ajustar exp(dosis) la desviación es de 1294 con 18 g.l.

Al ajustar dosis al cuadrado la desviación es de 81.90 con 18 g.l.

De estos ajustes todos con excepción de el ajuste de dosis, resultan significativos a cualquier nivel. El ajuste de dosis resulta significativo al 5% pero no significativo al 1%. No obstante que el ajuste resulta ser todavía significativo al 5%, la reducción en el valor de la desviación es tan considerable: de 1757 a 36.86, para solo un grado de libertad que se considera que se tiene ya un ajuste razonable.

Entonces el mejor ajuste es aquel en el que tomamos en cuenta los valores extremos 99.5 en el caso de 100 y .5 en el caso de cero y ajustamos dosis. Para este caso calcularemos las dosis 50, 70 y 90% así como intervalos de confianza correspondientes a estas dosis.

En primer lugar se obtendrá de los resultados proporcionados por GLIM estimadores de parámetros adicionales.

GLIM nos proporciona los valores correspondientes a $\hat{\alpha}$ y $\hat{\beta}$; $\hat{\alpha}$ es el parámetro correspondiente al ajuste de la gran

media o media general del modelo, $\hat{\beta}$ es el valor correspondiente al ajuste de la dosis, tenemos también sus respectivos errores standard. Directamente del paquete podemos obtener el coeficiente de correlación, la covarianza entre α y β y los valores ajustados de acuerdo al modelo, así como los residuales de dichos ajustes.

El primer parámetro a estimar es la media 0 dosis 50 (DL50) la cual está dada por

$$DL50 = \hat{\mu} = - \frac{\hat{\alpha}}{\hat{\beta}}$$

y el intervalo de confianza para μ va a estar dado considerando el teorema de Fieller.

Teorema de Fieller. ~ Supongamos que $\hat{\theta}_1$ y $\hat{\theta}_2$ son estimadores tales que:

- i) $\hat{\theta}_1 \sim N(\theta_1, \sigma_{11}), \hat{\theta}_2 \sim N(\theta_2, \sigma_{22})$
- ii) $Cov(\theta_1, \theta_2) = \sigma_{12}$
- iii) V_{11}, V_{22}, V_{11} son estimadores de $\sigma_{11}, \sigma_{22}, \sigma_{12}$

respectivamente, cada uno con f grados de libertad

Sean $R = \frac{O_1}{O_2}$ y $\hat{R} = \frac{\hat{\theta}_1}{\hat{\theta}_2}$

Entonces

Los limites del intervalo de confianza $(1-\alpha)$ para R estan dados por

$$\hat{R} - g \frac{V_{12}}{V_{22}} \pm \frac{t}{\hat{\theta}_2} \left(V_{11} - 2\hat{R}V_{12} + \hat{R}^2V_{22} - g \left(V_{11} - \frac{V_{12}^2}{V_{22}} \right) \right)^{1/2}$$

(1 - g)

donde $g = \frac{t^2 V_{22}}{\hat{\theta}_2^2}$ y $t = t_{\alpha/2}(f)$ es una t-student con f g.l.

Aplicando el teorema para obtener un intervalo de confianza para μ se tiene

$$\begin{array}{ll} O_1 = \alpha & O_2 = \beta \\ \hat{O}_1 = \hat{\alpha} & \hat{O}_2 = \hat{\beta} \end{array}$$

$$V_{11} = S_{\hat{\alpha}}^2 = S_{\hat{\alpha}}^2 \qquad V_{22} = S_{\hat{\beta}}^2 = S_{\hat{\beta}}^2$$

$$V_{12} = S_{\hat{\alpha}, \hat{\beta}} = -S_{\hat{\alpha}, \hat{\beta}}$$

Entonces el intervalo se convierte en:

$$\begin{aligned}
 & - \frac{\hat{\alpha}}{\hat{\beta}} + \frac{t^2 (S_{\hat{\alpha}, \hat{\beta}}^2)}{\hat{\beta}} + \frac{t}{\hat{\beta}} S_{\hat{\alpha}}^2 + 2 \left(-\frac{\hat{\alpha}}{\hat{\beta}} \right) S_{\hat{\alpha}, \hat{\beta}} + \left(-\frac{\hat{\alpha}}{\hat{\beta}} \right)^2 S_{\hat{\beta}}^2 - \\
 & - \frac{t S_{\hat{\beta}}^2}{\hat{\beta}^2} (S_{\hat{\alpha}}^2 - S_{\hat{\alpha}_1, \hat{\beta}}^2) / \left(1 - \frac{t^2 (S_{\hat{\beta}}^2)}{\hat{\beta}^2} \right)
 \end{aligned}$$

Para obtener otras dosis el mecanismo a seguir es el siguiente:

Se encuentra el valor en la distribución de una normal que equivale a tener el área acumulada a la dosis deseada, por ejemplo para la dosis 70, se busca el valor de la normal que da un 70% de área acumulada el cual es .84, el cual equivale a

$$\begin{aligned}
 .84 &= \quad + \quad X \\
 X &= \frac{.84 - \hat{\alpha}}{\hat{\beta}}
 \end{aligned}$$

y finalmente aplicando a x la transformación inversa de la transformación aplicada a la dosis, en este caso particular se obtiene directamente ya que la dosis no fue transformada.

Y para la obtención del intervalo de confianza basta substituir este valor en el lugar correspondiente a $-\frac{\hat{\alpha}}{\hat{\beta}}$

Entonces para este caso

$$\hat{\beta} = -.3679 \text{ E-02}$$

$$\hat{\alpha} = 3.065$$

$$S_{\hat{\beta}} = .152 \text{ E-03}$$

$$S_{\hat{\alpha}} = .1274$$

$$p = .9330$$

$$S_{2, \hat{\beta}} = -.1812 \text{ E-05}$$

La dosis media o DL50

$$DL50 = - \frac{3.065}{-.023679} = 833.107$$

y el intervalo de confianza correspondiente al 95% de confianza es

$$(807.654, 859.352)$$

La dosis letal 70

$$DL70 = 689.046$$

y el intervalo de confianza

(630.857, 745.831)

DL90=482.468

y el intervalo

(402.465, 557.918)

Es decir la dosis máxima que se puede aplicar para que germinen el 50% de las semillas se encuentra entre 807.654 y 859.352 Kr; para que germinen el 70% de las semillas la dosis se encuentra entre 630.857 y 745.831 Kr y para que germinen el 90% de las semillas la dosis se encuentra entre 402.465 y 557.918 Kr.

Para el caso de 60 Co Mev radiación gama se realizaron los mismos ajustes encontrándose:

- 1) Al ajustar únicamente la media general se obtiene una desviación de 1025 con 17 g.l. la cual resulta también altamente significativa.

Incluyendo en el ajuste la dosis la desviación resulta ser de 44.32 con 16 g.l lo cual resulta ser significativa al 1%.

Al ajustar log dosis la desviación resulta ser de 216 con 16 g.l.

Al ajustar dosis al cuadrado la desviación es de 33.25 con 16 g.l. lo cual resulta significativa al 5% y apenas significativa al 1%.

Cambiando los valores extremos 100 por 99.5 y 0 por .5 los resultados de los ajustes son los siguientes:

Al ajustar dosis la desviación es igual a 39.34 con 16 g.l.

Al ajustar log (dosis) la desviación resulta ser de 231.6 con 16 g.l.

Al ajustar dosis al cuadrado la desviación resulta ser de 29.16 con 16 g.l. lo que resulta significativa al 5% y no significativa al 1%. En este caso al igual que el anterior se logra un ajuste razonable en el que ahora la desviación se reduce de 1025 a 29.16 con la pérdida de solo un grado de libertad.

Por lo que el mejor ajuste resulta ser aquel en el que consideramos los valores extremos 99.5 por 100 y .5 por cero y ajustamos dosis al cuadrado, con ese modelo se calcularon los intervalos de confianza para estimar las dosis 50, 70 y 90.

Los valores de los parámetros de interés proporcionados por GLIM para este caso son:

$$\begin{aligned}\hat{\beta} &= .1326 \text{ E-05} \\ \hat{\alpha} &= 1.603 \\ S_{\hat{\alpha}} &= .5728 \text{ E-01} \\ S_{\hat{\beta}} &= .5717 \text{ E-07} \\ p &= .6721 \\ S2, \hat{\beta} &= 2.201 \text{ E-09}\end{aligned}$$

La media general estará dada por

$$DL50 = \hat{\mu} = \left(\frac{1.603}{.1326 \text{ E-05}} \right)^{1/2} = 1099.499$$

y el intervalo de confianza

$$(1018.996, 1187.923)$$

$\frac{1}{2}$

$$DL75 = 834.313 = \left(\frac{.68 + 1.603}{.1326 \text{ E-}05} \right)$$

y el intervalo de confianza al 95 %

(750.963, 922.652)

A continuación se presentan los listados correspondientes de los programas realizados para los análisis anteriores.



 R#SERVICIO/GLIM
 #RUNNING 3543
 #?

GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON
 #UNITS 20
 #DATA DOSIS Y
 #READ

0 100
 50 100
 100 99
 125 100
 150 100
 1175 99
 200 100
 300 98
 400 94
 500 94
 600 85
 700 59
 800 50
 900 41
 1000 34
 1100 24
 1200 0
 1400 0
 1600 0
 1800 0

#YVARAITE Y
 #CALC NI = 100
 #ERROR B NI
 #LINK P
 #FIT#DISPLAY E C M#

CYCLE	SCALED DEVIANCE	DF
4	1752.	19

ESTIMATE	S.E.	PARAMETER
1 0.3545	.2867E-01	ZGM
SCALE PARAMETER TAKEN AS		1.000

CORRELATIONS OF ESTIMATES
 1 1.0000
 1

Y-VARJATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI



LINEAR PREDICTOR

%GM

%FIT + DOSIS\$DISPLAY E C M\$

CYCLE	SCALED DEVIANCE	DF
5	42.12	18

	ESTIMATE	S.E.	PARAMETER
1	3.269	0.1418	%GM
2	-.3933E-02	.1692E-03	DOSI
SCALE PARAMETER TAKEN AS			1.000

CORRELATIONS OF ESTIMATES

1	1.0000	
2	-0.9430	1.0000
1		2

Y-VARIATE Y

ERROR BINOMIAL LINK PROBIT

BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR

%GM DOSI

%CALC Z = %LOG(DOSIS)

%CALC X = DOSIS/100

----- INVALID FUNCTION/OPERATOR ARGUMENT(S)

%CALC X = %EXP(X)

%CALC K = DOSIS**2

%FIT Z\$DISPLAY E C M\$

CYCLE	SCALED DEVIANCE	DF
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 2	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 2	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 2	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 2	HELD AT LIMIT
/	101.2	18

	ESTIMATE	S.E.	PARAMETER
1	17.70	0.8528	%GM
2	-2.658	0.1266	Z
SCALE PARAMETER TAKEN AS			1.000

AA JMSRAC

101-10-15

SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES

1	1.0000	
2	-0.9986	1.0000
	1	2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
ZGM Z
\$FIT X\$DISPLAY E C M\$



CYCLE	SCALED DEVIANCE	DF
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 19 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 19 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 19 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 18 HELD AT LIMIT	
-----	UNIT 19 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 18 HELD AT LIMIT	
-----	UNIT 19 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
-----	UNIT 18 HELD AT LIMIT	
-----	UNIT 19 HELD AT LIMIT	
-----	UNIT 20 HELD AT LIMIT	
10	425.8	18

NO CONVERGENCE BY CYCLE 10

	ESTIMATE	S.E.	PARAMETER
1	1.139	.4085E-01	ZGM
2	-.3783E-04	.2094E-05	X
	SCALE PARAMETER TAKEN AS		1.000

CORRELATIONS OF ESTIMATES

1	1.0000	
2	-0.4005	1.0000
	1	2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

M. J. DEB...



LINEAR PREDICTOR
 ZGM X
 \$FIT K\$DISPLAY E C M\$

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S.E.	PARAMETER
5	71.14	18			
1			2.124	.7857E-01	ZGM
2			-.2782E-05	.1127E-06	K

SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES
 1 1.0000
 2 -.08093 1.0000
 1 2

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 ZGM K
 \$CALC Y(1) = 99.5 ; Y(2)=99.5 ; Y(3)=99.5 ; Y(5)=99.5 ; Y(7)=99.5

z----- CURRENT DISPLAY INHIBITED
 \$CALC Y(17)=.5 ; Y(18)=.5 ; Y(19)=.5 ; Y(20)=.5
 \$YVARIATE Y
 \$FIT\$DISPLAY E C M\$

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S.E.	PARAMETER
4	1696.	19			
1			0.3538	.2867E-01	ZGM

SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES
 1 1.0000
 1

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 ZGM
 \$FIT HOSIS\$DISPLAY E C M R V \$



SCALED
CYCLE DEVIANCE DF
4 36.86 18

ESTIMATE S.E. PARAMETER
1 3.065 0.1274 %GM
2 -.3679E-02 .1523E-03 DOSI
SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES
1 1.0000
2 -0.9338 1.0000
1 2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
%GM DOSI

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	100	100	99.89	-1.187
2	100	100	99.80	-0.6789
3	99	100	99.65	-1.102
4	100	100	99.54	-.6060E-01
5	100	100	99.40	0.1272
6	99	100	99.23	-0.2589
7	100	100	99.01	0.4965
8	98	100	97.51	0.3152
9	94	100	94.45	-0.1960
10	94	100	88.98	1.602
11	85	100	80.45	1.147
12	59	100	68.79	-2.113
13	50	100	54.86	-0.9765
14	41	100	40.29	0.1441
15	34	100	26.97	1.583
16	24	100	16.32	2.079
17	1	100	8.861	-2.942
18	1	100	1.853	-1.003
19	1	100	0.2395	0.5330
20	1	100	.1878E-01	3.512

(CO)VARIANCE MATRIX
1 1.6226E-02
2 -1.0118E-05 2.3203E-08
1 2
SCALE PARAMETER TAKEN AS 1.000

S.D. OF DIFFERENCES
1 0.

2 0.1275 0.
 1 2
 SCALE PARAMETER TAKEN AS 1.000
 \$FIT Z\$DISPLAY E C M\$

SCALED
 CYCLE DEVIANCE DF
 ----- UNIT 1 HELD AT LIMIT
 ----- UNIT 1 HELD AT LIMIT
 ----- UNIT 1 HELD AT LIMIT
 ----- UNIT 1 HELD AT LIMIT
 ----- UNIT 1 HELD AT LIMIT
 ----- UNIT 1 HELD AT LIMIT
 ----- UNIT 1 HELD AT LIMIT
 B 144.4 18

	ESTIMATE	S.E.	PARAMETER
1	14.14	0.6681	ZGM
2	-2.130	.9931E-01	Z

SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES
 1 1.0000
 2 -0.9980 1.0000
 1 2

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 ZGM Z
 \$FIT X\$DISPLAY E C M\$

SCALED
 CYCLE DEVIANCE DF
 ----- UNIT 20 HELD AT LIMIT
 6 1294. 18

	ESTIMATE	S.E.	PARAMETER
1	0.5485	.3145E-01	ZGM
2	-.1812E-06	.1766E-07	X

SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES
 1 1.0000
 2 -0.2413 1.0000
 1 2

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT



ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
ZGM X
*FIT K*DISPLAY E C M*

	SCALED	
CYCLE	DEVIANCE	DF
5	81.90	18

	ESTIMATE	S.E.	PARAMETER
1	1.992	.7232E-01	ZGM
2	-.2545E-05	.1027E-06	K

SCALE PARAMETER TAKEN AS 1.000

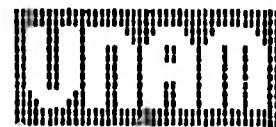
CORRELATIONS OF ESTIMATES

1	1.0000	
2	-0.7922	1.0000
	1	2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
ZGM K
STOP

*ET=14128.4 PT=40.1 IO=1.0



ZGM
\$FIT DOSIS\$DISPLAY E C M\$

CYCLE	SCALED DEVIANCE	DF
4	44.32	16

	ESTIMATE	S.E.	PARAMETER
1	2.316	.0367E-01	ZGM
2	-.2234E-02	.8927E-04	DOSI
	SCALE PARAMETER TAKEN AS		1.000

CORRELATIONS OF ESTIMATES
1 1.0000
2 -0.8617 1.0000
1 2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
ZGM DOSI
\$CALC Z = ZLOG(DOSIS)
\$CALC X = DOSIS/100

----- INVALID FUNCTION/OPERATOR ARGUMENT(S)
\$CALC X = ZEXP(X)
\$CALC K = DOSIS**2
\$FIT Z\$DISPLAY E C M\$

CYCLE	SCALED DEVIANCE	DF
-----	UNIT 1 HELD AT LIMIT	
-----	UNIT 1 HELD AT LIMIT	
-----	UNIT 1 HELD AT LIMIT	
4	216.0	16

	ESTIMATE	S.E.	PARAMETER
1	0.310	0.4086	ZGM
2	-1.210	.6126E-01	Z
	SCALE PARAMETER TAKEN AS		1.000

CORRELATIONS OF ESTIMATES
1 1.0000
2 -0.9953 1.0000
1 2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI



M. D. STRAC

BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
ZGM Z
\$CADEL
\$FIT X\$DISPLAY E C M\$



CYCLE	SCALED DEVIANCE	DF
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
----	UNIT 18 HELD AT LIMIT	
10	364.5	16
----	NO CONVERGENCE BY CYCLE 10	

	ESTIMATE	S.E.	PARAMETER
1	1.026	.3868E-01	ZGM
2	-.4184E-06	.6690E-07	X
	SCALE PARAMETER TAKEN AS		1.000

CORRELATIONS OF ESTIMATES

1	1.0000
2	-0.2322 1.0000
	1 2

Y-VARIATE Y
ERROR BINOMIAL LINK PROBIT
BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
ZGM X
\$FIT K\$DISPLAY E CM\$

CYCLE	SCALED DEVIANCE	DF
4	33.25	16

	ESTIMATE	S.E.	PARAMETER
1	1.693	.5761E-01	ZGM
2	-.1341E-05	.5786E-07	K
	SCALE PARAMETER TAKEN AS		1.000

CORRELATIONS OF ESTIMATES

1	1.0000
2	-0.6734 1.0000
	1 2



+

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI
 LINEAR PREDICTOR
 %GM K
 %CALC Y(1)=99.5 : Y(18)=.5
 %YVARIATE Y

----- CURRENT DISPLAY INHIBITED
 %FIT%DISPLAY E C M%

SCALED			
CYCLE	DEVIANCE	DF	
3	1012.	17	
ESTIMATE	S.E.	PARAMETER	
1 0.6433	.3187E-01	%GM	
SCALE PARAMETER TAKEN AS		1.000	

CORRELATIONS OF ESTIMATES
 1 1.0000
 1

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 %GM
 %FIT DOSIS%DISPLAY E C M%

SCALED			
CYCLE	DEVIANCE	DF	
4	39.34	16	
ESTIMATE	S.E.	PARAMETER	
1 2.302	.8308E-01	%GM	
2 -.2217E-02	.0862E-04	DOSI	
SCALE PARAMETER TAKEN AS		1.000	

CORRELATIONS OF ESTIMATES
 1 1.0000
 2 -0.8607 1.0000
 1 2

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

IBM

LINEAR PREDICTOR
 ZGM POST
 \$FIT Z\$DISPLAY E C M\$



SCALED		
CYCLE	DEVIANCE	DF
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
-----	UNIT 1	HELD AT LIMIT
7	231.1	16

	ESTIMATE	S.E.	PARAMETER
1	7.832	0.3843	ZGM
2	-1.138	.5783E-01	Z
SCALE PARAMETER TAKEN AS			1.000

CORRELATIONS OF ESTIMATES

1	1.0000	
2	-0.9948	1.0000
1		?

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 ZGM 7
 \$FIT X\$DISPLAY E C M\$

SCALED		
CYCLE	DEVIANCE	DF
-----	UNIT 18	HELD AT LIMIT
-----	UNIT 18	HELD AT LIMIT
-----	UNIT 18	HELD AT LIMIT
-----	UNIT 18	HELD AT LIMIT
-----	UNIT 18	HELD AT LIMIT
6	403.3	16

	ESTIMATE	S.E.	PARAMETER
1	0.9468	.3721E-01	ZGM
2	-.1988E-04	.1652E-07	X
SCALE PARAMETER TAKEN AS			1.000

CORRELATIONS OF ESTIMATES

1	1.0000	
2	-0.2891	1.0000
1		?

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 %GM X
 *FIT K*DISPLAY E C M R V S*

SCALED
 CYCLE DEVIANCE DF
 4 29.16 16

	ESTIMATE	S.E.	PARAMETER
1	1.683	.5728E-01	%GM
2	-.1326E-05	.5717E-07	K

SCALE PARAMETER TAKEN AS 1.000

CORRELATIONS OF ESTIMATES

1	1.0000
2	-0.6721 1.0000
1	2

Y-VARIATE Y
 ERROR BINOMIAL LINK PROBIT
 BINOMIAL DENOMINATOR NI

LINEAR PREDICTOR
 %GM K

UNIT	OBSERVED	OUT OF	FITTED	RESIDUAL
1	100	100	95.38	1.961
2	97	100	95.35	0.7826
3	99	100	95.26	1.762
4	95	100	95.18	-.8441E-01
5	95	100	95.09	-.4109E-01
6	96	100	94.98	0.4678
7	91	100	94.85	-1.741
8	97	100	94.11	1.228
9	92	100	92.94	-0.3663
10	93	100	91.18	0.6420
11	87	100	88.61	-0.5068
12	83	100	84.94	-0.5411
13	76	100	79.81	-0.9487
14	59	100	72.89	-3.124
15	45	100	41.06	0.8001
16	23	100	18.00	1.301
17	4	100	4.355	-0.1740
18	1	100	0.4494	.7558E-01

(CO)VARIANCE MATRIX



M. J. D. B. P. C. 1988

1 3.2681E-15
2 -2.2010E-09 3.2681E-15

1 2
SCALE PARAMETER TAKEN AS 1.000

S.E. OF DIFFERENCES

1 0.
2 5.7281E-02 0.

1 2
SCALE PARAMETER TAKEN AS 1.000

STOP

#E1=9108.8 PT=31.2 IO=1.3



Ejemplo 2.

Como se mencionó anteriormente este ejemplo se refiere a una investigación realizada para estudiar la eficiencia de un nuevo detergente X comparada con la de un detergente estándar M.

Se eligió una muestra de 1008 individuos y se les pidió que compararan los dos detergentes y dijeran cual de los dos preferían, también se les preguntó que temperatura usaban (alta, baja); que tipo de agua utilizaron (suave, de mediana dureza, dura) y finalmente si anteriormente usaban o no usaban el detergente de tipo M.

Los datos con que se cuenta, por lo tanto, son frecuencias clasificadas de acuerdo a las cuatro variables mencionadas anteriormente, es decir nuestros datos se pueden representar en una tabla de contingencia de 4 dimensiones.

Tipo de Agua	Previo uso de M		Previo no uso de M		
	Preferencia	Temperatura Alta	Temperatura Baja	Temperatura Alta	Temperatura Baja
suave	X	19	57	29	63
	M	29	49	27	53
mediana	X	23	47	33	66
	M	47	55	23	50
Dura	X	24	37	42	68
	M	43	52	30	42

Los resultados de los ajustes realizados se muestran a continuación, en donde Tem, representa temperaturas PRE, preferencia; USO, uso; AGUA, agua;.

R\$SERVICIO/GLIM
\$WAITING FOR AVAILABLE TASK
\$RUNNING 9225
\$?



GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON
\$UNITS 24
\$DATA Y
\$READ

19	57	29	63
29	49	27	53
23	47	33	66
47	55	23	50
24	37	42	60
43	52	30	42

\$YVARIATE Y

\$FACTOR PRE 2: TEM 2: AGUA 3 : USO 2

\$CALC PRE = %GL(2,4) : TEM = %GL(2,1) : AGUA = %GL(3,8) : USO = %GL(2,2)

\$C SE DEFINIERON LAS CUATRO VARIABLES CON LOS VALORES QUE TOMA CADA UNA
EN LAS 24 CELDAS DE LA TABLA

\$ERROR P
\$LINK 1

\$C SE VA AJUSTAR EL MODELO QUE SOLO INCLUYE COMO PARAMETRO LA MEDIA
GENERAL . ES DECIR SE PRUEBA LA HIPOTESIS DE QUE LA PROBABILIDAD
DE TENER UNA OBSERVACION EN CUALQUIER CELDA ES LA MISMA. ES DECIR
SE PRUEBA LA HIPOTESIS DE INDEPENDENCIA DE LAS VARIABLES MAS
IGUALDAD DE MARGINALES. EN CADA VARIABLE.

\$F111

	SCALED	
CYCLE	BEVIANCE	DF
3	111.6	23

\$DISPLAY 14

\$MODEL PREDICTOR

ADIRAC

ZGM



!C LA DESVIACION RESULTA SER ALTAMENTE SIGNIFICATIVA
!C AHORA SE AJUSTAN LOS EFECTOS PRINCIPALES PARA PROBAR LA HIPOTESIS
DE INDEPENDENCIA ENTRE LAS 4 VARIABLES.

\$FIT + PRE + AGUA + USO+ TEM\$

	SCALED	
CYCLE	DEVIANCE	DF
3	40.55	18

!C EL MODELO AJUSTADO ES EL SIGUIENTE
!DISPLAY L\$

LINEAR PREDICTOR
ZGM+PRE+AGUA+USO+TEM

!C LA DESVIACION RESULTA SER ALTAMENTE SIGNIFICATIVA .ES DECIR NO EXISTE
INDEPENDENCIA ENTRE LAS CUATRO VARIABLES.

!C EL SIGUIENTE MODELO QUE SE AJUSTARA CONTIENE TODOS LOS EFECTOS PRIN-
CIPALES E INTERACCIONES DE PRIMER ORDEN

\$FIT + TEM.AGUA + TEM.PRE + TEM.USO + USO.PRE + USO.AGUA + AGUA.PRE

!DISPLAY 04

	SCALED	
CYCLE	DEVIANCE	DF
3	9.568	9

SCALED
DEVIANCE = 9.568 DF = 9



4C EL MODELO AJUSTADO ES EL SIGUIENTE
 †DISPLAY L‡

LINEAR PREDICTOR
 ZOM PRE AGUA USO TEM PRE. AGUA PRE. USO AGUA, USO PRE. TEM AGUA, TEM USO.
 TEM

4C ESTE MODELO RESULTA TEMER UNA DESVIACION NO SIGNIFICATIVA POR LO
 QUE VEREMOS SI EXISTE UN MODELO CON UN NUMERO MENOR DE TERMINOS *parametros*
 QUE SE AJUSTE A LOS DATOS. PARA ESO VEREMOS QUE INTERACCIONES DE
 PRIMER ORDEN SON LAS QUE SE ENCUENTRAN MENOS REPRESENTADAS, ES
 DECIR CUALES DE LOS PARAMETROS QUE REPRESENTAN LA INTERACCION
 DE PRIMER ORDEN ESTAN CERCANOS A CERO

†CALC A1 = ZEQ(AGUA,1) - ZEQR(AGUA,2)
 † A2= ZEQ(AGUA,1) - ZEQ(AGUA,3)
 † P = ZEQ(PRE,1) - ZEQ(PRE,2)
 † U = ZEQ(TEM,1) - ZEQ(TEM,2)
 † U = ZEQ(USO,1) - ZEQ(USO,2)
 † PT= P*T ; UT = U*T ; PU = P*U ; P1= P*A1 ; P2 = P*A2; U1 = U*A1
 † U2 = U*A2 ; T*A1 = T*A1 ; DEL
 † U2= U*A2 ; T1 = T*A1 ; T2 = T*A2

†FIT A1+A2+P+U+T+PT+PU+UT+T1+T2+P1+P2+U1+U2
 †DISPLAY E‡

CYCLE	SCALER DEVIANCE	DF	ESTIMATE	S.E.	PARAMETER
3	9.568	9			
1			3.677	.3333E-01	ZOM
2			-.3511E-01	.4618E-01	A1
3			-.1203E-01	.4612E-01	A2
4			-.1996E-01	.3519E-01	P
5			-.3119E-01	.3313E-01	U
6			-.022710	.3304E-01	T
7			-.5023E-01	.3327E-01	PT
8			-.01356	.3201E-01	PU
9			-.2321E-01	.3327E-01	UT
10			-.5301E-02	.4625E-01	T1
11			-.01120	.4616E-01	T2

```

12 .1348E-01 .4489E-01 P1
13 .9121E-02 .4548E-01 P2
14 -.3500E-01 .4483E-01 U1
15 .2247E-01 .4547E-01 U2
SCALE PARAMETER TAKEN AS 1.000

```



A
 DEBRAC

10 COMO SE PUEDE VER SI ORDENAMOS LAS INTERACCION EN ORDEN DE REPRESENTACION EN EL MODELO EL PRIMER LUGAR LO OCUPA LA INTERACCION ENTRE PREFERENCIA Y USO, DESPUES TEMPERATURA Y AGUA, PREFERENCIA Y TEMPERATURA, USO Y AGUA, USO Y TEMPERATURAR Y POR ULTIMO PREFERENCIA Y AGUA. EN SEGUIDA VAMOS A IR QUITANDO TERMINOS MIENTRAS EL MODELO RESULTE SIGNIFICATIVO. PRIMERAMENTE QUITAREMOS LA INTERACCION ENTRE PREFERENCIA Y AGUA.

1FIT = P1 - P2

DISPLAY L

CYCLE	SCALED DEVIANCE	DF
3	9.817	11

LINEAR PREDICTOR
 ZGR A1 A2 P U T PT PU UT T1 T2 U1 U2

10 LA RESUJACION RESULTA SER NO SIGNIFICATIVA, COMO SE PUEDE VER AUMENTA MUY POCO Y SE GANAN DOS GRADOS DE LIBERTAD.

10 EL SIGUIENTE MODELO ES EL AJUSTADO

*DISPLAY L

LINEAR PREDICTOR
 ZGR A1 A2 P U T PT PU UT T1 T2 U1 U2

10 AHORA SE QUITARA DEL MODELO LA INTERACCION ENTRE USO Y TEMPERATURA

1FIT = UT

\$DISPLAY L\$

CYCLE	SCALED DEVIANCE	DF
3	10.30	12

LINEAR PREDICTOR
ZGM A1 A2 P U T PT PU T1 T2 U1 U2

*C LA DESVIACION RESULTA SER NO SIGNIFICATIVA POR LO QUE QUITAREMOS DEL
MODELO LA INTERACCION ENTRE USO Y AGUA

\$FIT - U1 - U2

*DISPLAY L\$

CYCLE	SCALED DEVIANCE	DF
4	10.98	14

*C LA DESVIACION RESULTA SER NO SIGNIFICATIVA, EL MODELO AJUSTADO ES

LINEAR PREDICTOR
ZGM A1 A2 P U T PT PU T1 T2

*C AHORA AJUSTANDO EL MODELO SIN LA INTERACCION ENTRE PREFERENCIA Y TEMPERATURA

\$FIT - PT

*DISPLAY E\$

CYCLE	SCALED DEVIANCE	DF
5	14.59	15

ESTIMATE	COEF.	PARAMETER
1	0.679	AX2RF=01 ZGM



1	.0000	.0000	A1
2	-.3462E-01	.4614E-01	A1
3	-.1761E-01	.4606E-01	A2
4	-.2138E-01	.3315E-01	P
5	-.3670E-01	.3194E-01	U
6	-0.2717	.3302E-01	T
7	-0.1370	.3194E-01	PU
8	.3390E-02	.4614E-01	TJ
9	-0.1118	.4606E-01	T2
10	-.6232E-01	.3284E-01	PT

SCALE PARAMETER TAKEN AS 1.000

STOP

!QUEUED

!ET=3124:27.7 P1=1128.2 IO=1.0

!

!BYE

!END SESSION 9216 ET=3125:35.6 PT=1128.2 IO=1.0

!USER = A082 12:19:59 05/06/82



Considerando

- como la media general
- A_i el i -ésimo nivel de Agua ($i=1,2,3$)
- T_j el j -ésimo nivel de Temperatura ($j=1,2$)
- P_k el k -ésimo nivel de Preferencia ($k=1,2$)
- U_l el l -ésimo nivel de Uso ($l=1,2$)
- PU_{kl} la interacción entre el k -ésimo nivel de Preferencia y el l -ésimo nivel de uso.
- AT_{ij} la interacción entre el i -ésimo nivel de Agua y el j -ésimo nivel de temperatura
- TP_{jk} la interacción entre el j -ésimo nivel de Temperatura y k -ésimo nivel de preferencia

Podemos observar que el modelo

$$\text{Log } F_{ijkl} = \mu + A_i + T_j + P_k + U_l + PU_{kl} + AT_{ij} + TP_{jk}$$

tiene una desviación de 10.98 con 14 g.l., lo que resulta ser no significativa y al retirar el término de interacción entre temperatura y preferencia la desviación se incrementa a 14.59 con 15 g.l., a pesar de que la desviación se incrementa en 3.61 ganándose un grado de libertad, 3,61 para un grado de libertad no es significativo al 5%, lo mismo que 14.59 no lo es para 15 grados de libertad y como lo que se pretende es encontrar el modelo más sencillo, que explique el comportamiento de los datos, el modelo con el que nos quedaremos es el siguiente ,

$$\text{Log } F_{ijkl} = \mu + A_i + T_j + P_k + U_l + PU_{kl} + AT_{ij}$$

En este modelo podemos observar que no existe asociación de los variables preferencia y uso con los variables tipo de agua y temperatura, es decir la preferencia es independiente del tipo de agua y la temperatura, sin embargo está relacionada con el uso previo de M; al igual que la variable temperatura no esta asociada con la preferencia y del uso previo de detergente M, sin embargo esta relacionada con el tipo de agua que usan las personas.

Los parámetros obtenidos son los siguientes

	E.S.
$u = 3.679$	0.033
$A_1 = -0.0346$	0.046
$A_2 = -0.0176$	0.046
$A_3 = 0.0428$	0.046
$P_1 = -0.0049$	0.031
$P_2 = 0.0049$	0.031
$U_1 = -0.0367$	0.031
$U_2 = 0.0367$	0.031
$T_1 = -0.270$	0.032
$T_2 = 0.270$	0.032
$PU_{11} = -0.1370$	0.031
$PU_{12} = 0.1370$	0.031
$PU_{21} = 0.1370$	0.031
$PU_{22} = -0.1370$	0.031
$AT_{21} = .0033$	0.046
$AT_{11} = -0.118$	0.046
$AT_{31} = 0.1085$	0.046
$AT_{22} = 0.0033$	0.046
$AT_{21} = 0.1118$	0.046
$AT_{23} = -0.1085$	0.046

De los parámetros estimados para la interacción entre preferencia y uso previo de M como PU_{11} es negativo las celdas

correspondientes a esta combinación de valores de las variables tienen menos observaciones de los que se esperaba si fueran independientes es decir, si la gente previamente había usado el detergente M era más factible que prefiriera X y si la gente no había usado el detergente M era menos factible que prefiriera X.

Esto se puede ver más claro si colapsamos la tabla, hacer esto es válido debido a que hay independencia entre 2 grupos de variables, considerando la independencia entre los dos grupos de variables.

Preferencia	Uso Previo de M	Nr. Uso Previo de M
X	207	301
M	275	225

En relación a los parámetros estimados para la interacción entre temperatura y tipo de agua se observa que $AT_{11} = -0.1118$, $AT_{12} = 0.0033$, $AT_{13} = 0.1085$ es decir que en las celdas que tienen la combinación (1,1) de las variables tienen menos observaciones de las esperadas mientras que en las celdas que tienen la combinación (1,3) tienen más observaciones de las esperadas. Es decir que existe una tendencia de las personas a utilizar temperaturas más altas entre más dura sea el agua.

Al igual que en el caso anterior la tabla se puede presentar en forma colapsada como a continuación

Tipo de Agua	Temperatura	
	Alta	Baja
Suave	104	222
Mediana	126	218
Dura	139	199

En esta tabla se observa como la proporción de personas que utilizan temperaturas altas en su lavado es mayor en el caso en que tienen aguas más duras. Esto último es explicable ya que se sabe que el detergente tiene menor eficacia en aguas duras y por consiguiente se requiere de temperaturas más altas para realizar un buen lavado.

Es importante hacer notar que en este ejemplo se podría considerar una variable respuesta y 3 variables explicativas, ajustando un modelo logístico, ya que este modelo permitiera interpretar las asociaciones como efectos sobre la respuesta y ciertas interacciones. Es decir se consideraría la preferencia como variable respuesta y el uso, agua y temperatura como variables explicativas, es decir se explicaría la preferencia de acuerdo a si es usado o no previamente el detergente, el tipo de agua y la temperatura con que se lava.

Ejemplo 3.

Se refiere como ya se había mencionado, a un estudio para investigar si las diversas formas y colores que adquieren cierta variedad de gusanos depende o no del sexo y localidad en que se encuentren.

Esta variedad de gusanos puede tomar 7 diferentes variedades de formas y colores las cuales son:

Populi (POP), Typica (TYP), Trilineata (TRI), Marginella (MAR), Lateralis (LAT), Flauicollis (FLA), y Albomaculota (ALB). Las formas MAR, LAT y FLA están ausentes en el sexo masculino. Estos gusanos se recolectaron en dos sitios diferentes denominados A y B respectivamente.

Los datos con los que se cuenta es el número de gusanos de cada forma clasificada por sexo y localidad. Es decir contamos con datos clasificados de acuerdo a tres variables y lo que se pretende es probar si existe relación o no entre esas variables, por lo que nuestros datos se pueden representar en una tabla de continuidad de tres dimensiones.

Los datos que se obtuvieron son los siguientes:

Sexo	Localidad	POP	TYP	F O R M A				
				TRI	MAR	LAT	FLA	ALB
M	A	51	123	18	-	-	-	-
M	B	116	303	44	-	-	-	-
F	A	58	91	15	7	5	1	2
F	B	133	289	44	31	13	2	7

Al conocer que las formas MAR, LAT y FLA no se presentan en el sexo masculino, se sabe que existe una relación entre el sexo y la forma y color de los gusanos, por lo que al ajustar el modelo de antemano se sabe que debe existir interacción entre sexo y color-forma, por lo que se realizaron dos tipos de ajustes. Los primeros considerando un peso igual a todas las celdas y posteriormente dando un peso de cero a las celdas correspondientes a sexo masculino y MAR, LAT y FLA. Lo que equivaldrá a probar la independencia entre los variables una vez que se ha eliminado la dependencia que se conoce en forma a-priori de que ciertas formas y colores no se presentan en el sexo masculino.

Los resultados obtenidos se muestran a continuación.

RESERVICIO ZGLIN
WAITING FOR AVAILABLE TASK
RUNNING 7653
17



GLIM 3.11 (C)1977 ROYAL STATISTICAL SOCIETY, LONDON
UNITS 28
DATA Y
\$READ

51	123	18	0	0	0	1
114	303	44	0	0	0	1
50	21	15	7	5	1	2
133	399	44	31	13	2	7

\$FACTOR FC 71SEX 21LOC 2

\$CALC FC = %GL(2,1) * SEX = %GL(2,14) : LOC = %GL(2,7)

*C SE DEFINIERON LAS TRES VARIABLES CON LOS VALORES QUE TOMA CADA UNA
EN LAS 28 CELDAS.

*ECHO E
\$WRITE Y

*C SE VA AJUSTAR EL MODELO QUE INCLUYE COMO PARAMETRO LA MEDIA
GENERAL. ES DECIR SE PRUEBA LA HIPOTESIS QUE LA PROBABILIDAD DE TENER UNA OBSERVACION EN CUALQUIER CELDA
ES LA MISMA. ES DECIR LA HIPOTESIS DE INDEPENDENCIA MAS IGUALDAD
DE MARGINALES EN LAS TRES VARIABLES.

SCALE		
CELL	DEVIANCE	SE

*C EL MODELO AJUSTADO ES EL SIGUIENTE
*ECHO E

*ECHO E
\$END

*C EL RESULTADO PUEDE SER ALTERNAMENTE SIGNIFICATIVO.

2	-.3462E-01	.4614E-01	A1
3	-.1761E-01	.4606E-01	A2
4	-.4995E-02	.3194E-01	P
5	-.3670E-01	.3194E-01	U
6	-0.2707	.3294E-01	T
7	-0.1370	.3194E-01	PU
8	.3390E-02	.4614E-01	T1
9	-0.1118	.4606E-01	T2

SCALE PARAMETER TAKEN AS 1.000



A2 DSPAC

4C EL MODELO AJUSTADO ES EL SIGUIENTE
 \$DISPLAY L

LINEAR PREDICTOR
 ZGM A1 A2 P U T PU T1 T2

4C LA DESVIACION RESULTA SER NO SIGNIFICATIVA

4C SI QUITAMOS LA INTERACCION ENTRE TEMPERATURA Y AGUA

4FIT = T1-T2

\$DISPLAY L4

CYCLE	SCALED DEVIANCE	DF
3	21.92	12

LINEAR PREDICTOR
 ZGM A1 A2 P U T PU

4C LA DESVIACION RESULTA SER SIGNIFICATIVA POR LO TANTO EL
 MODELO QUE MEJOR SE AJUSTA A LOS DATOS PUEDE SER CUALQUIERA DE LOS DOS ANTERIORES

4FIT = T1-T2 \$FIT \$DISPLAY E4

CYCLE	SCALED DEVIANCE	DF
5	10.98	14

ESTIMATE	S.E.	PARAMETER
1	2.622	0.4614

40 AHORA SE AJUSTAN LOS EFECTOS PRINCIPALES PARA PROBAR LA HIPOTESIS DE INDEPENDENCIA ENTRE LAS 3 VARIABLES.

\$FIT + FC + SEX + LOC

\$DISPLAY L\$

CYCLE	SCALED DEVIANCE	DF
5	96.62	19

LINEAR PREDICTOR
Z0M FC SEX LOC

40 LA DESVIACION RESULTA SER ALTAMENTE SIGNIFICATIVA. ES DECIR NO EXISTE INDEPENDENCIA ENTRE LAS TRES VARIABLES.

40 EL SIGUIENTE MODELO QUE SE AJUSTARA CONTIENE TODOS LOS EFECTOS PRINCIPALES E INTERACCIONES DE PRIMER ORDEN

\$FIT +FC.SEX + SEX.LOC +FC.LOC\$

CYCLE	SCALED DEVIANCE	DF
10	1.201	6

40 ESTE MODELO RESULTA TENER UNA DESVIACION NO SIGNIFICATIVA POR LO QUE VEREMOS SI EXISTE UN MODELO CON UN NUMERO MENOR DE TERMINOS QUE SE AJUSTE A LOS DATOS. PARA ESO VEREMOS QUE INTERACCIONES DE PRIMER ORDEN SON LAS QUE SE ENCUENTRAN MENOS REPRESENTADAS. ES DECIR CUALES DE LOS PARAMETROS QUE REPRESENTAN LA INTERACCION DE PRIMER ORDEN ESTAN CERCANOS A CERO

\$SCALE F1 = ZEQ(FC,1) - ZEQ(FC,2)
F2 = ZEQ(FC,1) - ZEQ(FC,3)
F3 = ZEQ(FC,1) - ZEQ(FC,4)
F4 = ZEQ(FC,1) - ZEQ(FC,5)





```

: F5 = ZEQ(FC,1) - ZEQ(FC,6)
: F6 = ZEQ(FC,1) - ZEQ(FC,7)
: S = ZEQ(SEX,1) - ZEQ(SEX,2)
: L = ZEQ(LOC,1) - ZEQ(LOC,2)
: SL = S*L: S1 = S*F1 : S2 = S*F2 : S3 = S*F3 : S4 = S*F4 :
: L1 = L*F1 : L2 = L*F2: L3= L*F3
: L4 = L*F4 : L5 = L*F5 : L6 = L*F6

```

```

*FIT(F1+F2+F3+F4+F5+F6+S+L+SL+S1+S2+S3+S4+L1+L2+L3+L4+L5+L6

```

```

*DISPLAY E*

```

CYCLE	SCALED DEVIANCE	DF	ESTIMATE	S.E.	PARAMETER
10	1.201	6			
1	.3847E-02	10.67			ZGH
2	-5.174	10.67			F1
3	-3.288	10.67			F2
4	4.014	36.33			F3
5	4.312	39.38			F4
6	5.184	39.18			F5
7	-0.6441	10.68			F6
8	-2.754	10.67			S
9	-0.4730	0.1134			L
10	.4511E-01	.3124E-01			SL
11	-2.833	10.67			S1
12	-2.800	10.67			S2
13	3.946	36.33			S3
14	3.641	38.30			S4
15	2.773	39.18			S5
16	-2.022	10.68			S6
17	.3932E-01	0.1187			L1
18	.1948E-01	0.1427			L2
19	0.2259	0.2101			L3
20	-1.4034E-01	0.2484			L4
21	-0.1715	0.2998			L5
22	-1.1073E-01	0.3075			L6
	SCALE PARAMETER TGLH AS	1.000			

11 COMO PODEROS VER LOS PARAMETROS QUE REPRESENTAN LA INTERACCION MAS
 CERCANOS A CERO SON LOS DE INTERACCION ENTRE SEXO Y LOCALIZACION,
 DE PUES ENTRE LOCALIZACION Y FORMA-COLOR Y POR ULTIMO LOS DE SEXO Y FORMA-COLOR,
 POR LO QUE IPRESO OMITIENDO DICHS TERMINOS HASTA QUE LA INTERACCION
 SEA SIGNIFICATIVA.

\$C SEA SIGNIFICATIVA

\$FIT F1+F2+F3+F4+F5+F6+S1+S2+S3+S4+S5+S6+L1+L2+L3+L4+L5+L6

\$DISPLAY L3

	SCALED	
CYCLE	DEVIANCE	DF
10	3.291	7

LINEAR PREDICTOR

%M F1 F2 F3 F4 F5 F6 S 1 S1 S2 S3 S4 S5 S6 L1 L2 L3 L4 L5 L6

\$CC DEL

\$C COMO LA DESVIACION RESULTA SER NO SIGNIFICATIVA QUITAREMOS DEL MODELO LOS TERMINOS CORRESPONDIENTES A LA INTERACCION ENTRE FORMA-COLOR Y LOCALIZACION.

\$FIT -L1-L2-L3-L4-L5-L6

	SCALED	
CYCLE	DEVIANCE	DF
10	6.942	13

\$C LA DESVIACION RESULTA SER NO SIGNIFICATIVA Y POR LO TANTO ESTE ES EL MODELO QUE MEJOR AJUSTA A LOS DATOS, YA QUE SOLO CONTIENE LA INTERACCION ENTRE SEXO Y FORMA-COLOR, LO CUAL SE ESPERABA DESDE UN PRINCIPIO AL SABER QUE CIERTAS FORMAS NO SE DAN EN EL SEXO MASCULINO.

\$C CONSIDERANDO LAS CELDAS EN QUE DE ANTEMANO SABEMOS QUE NO VA A HABER OBSERVACIONES CON UN PESO DE CERO, SE OBTIENE LO SIGUIENTE.

\$CALC WT = 1

\$CALC WT(4) = WT(5) = WT(6) = WT(11) = WT(12) = WT(13) = 0

\$MPLIST M

\$FIT 4



*C EL MODELO QUE SOLO INCLUYE A LA MEDIS GENERAL COMO PARAMETRO

----- CURRENT DISPLAY INHIBITED

CYCLE	SCALED DEVIANCE	DF
5	2059.	21

* L DEL

*C LA DESVIACION RESULTA ALTAMENTE SIGNIFICATIVA.

*C AHORA AJUSTANDO LOS EFECTOS PRINCIPALES PARA PROBAR LA HIPOTESIS
DE INDEPENDENCIA DE LAS VARIABLES

*FIT *FC* SEX * LOC*

CYCLE	SCALED DEVIANCE	DF
4	15.82	13

*C LA DESVIACION RESULTA SIEM NO SIGNIFICATIVA ES DECIR LAS VARIABLES
SON INDEPENDIENTES

*C EL MODELO AJUSTADO ES EL SIGUIENTES

DISPLAY L

LINEAR PREDICTOR
G *C* *SEX* *LOC*

*L1 *L1E24* *L1E54* *L1E54L* *DISPLAY L*

CYCLE	SCALED DEVIANCE	DF
4	15.82	13

ESTIMATE	SE	PARAMETER
0.1044	0.1044	%HH
0.1044	0.1044	EA





3	-0.5795	0.1290	F2
4	-0.1284	0.1721	F3
5	0.6188	0.2248	F4
6	2.411	0.4989	F5
7	1.810	0.2750	F6
8	.1389E-01	.2778E-01	S
9	-0.4859	.3044E-01	L

SCALE PARAMETER TAKEN AS 1.000

STOP

#ET=1159107.2 FT=3150.8 IO=0.9

Como se puede observar cuando no consideramos que las formas MAR, LAT y FLA no se dan en el sexo masculino, es decir todas las celdas tienen el mismo peso, el modelo obtenido es el siguiente

$$\text{Log } F_{ijk} = u + F_i + S_j + L_k + FS_{ij}$$

se confirma la dependencia conocida entre sexo y color y forma.

Mientras que cuando a las celdas anteriormente mencionadas les damos un peso de cero el modelo que mejor ajusta a los datos resulta ser

$$\text{Log } F_{ijk} = u + F_i + S_j + L_k$$

Es decir, al eliminar la parte de la interacción entre forma y sexo esperada, las variables resultan ser independientes. Al asignar peso cero a las celdas en las que no puede haber observaciones se esta dando estimación a-priori de cero para esas celdas. Es decir, el no asignar a-priori un peso cero a dichas celdas, nos conduciría a un modelo ^{''}enóneo, en el sentido que estaría reflejando una dependencia entre sexo y forma que de hecho no existe. De ahí la importancia de considerar explícitamente los ceros estructurales.

B I B L I O G R A F I A

- SILVEY, S.D., Statistical Inference., 1a. edición 1970.
U.S.A.
- BAKER, R.J., & NELDER, J.A., The Glim System., 1st. Edition 1978,
U.K.
- MENDEZ, IGNACIO., Modelos Estadísticos Lineales, FOCCA VI-CONACYT,
Primera Edición 1976, México.
- INFANTE, S. y CALDERON, L.C., Manual de Análisis Probit, Colegio
de Postgraduados 1980, México.
- NELDER, J.A. and WEDDERBURN, R.W.M., (1972). Generalized linear
models., J.R..statist.Soc., A, 135, 370-384
- WEDDERBURN, R.W.M., (1974). Quasi-likelihood functions, generalised
linear models and the Gauss-Newton method.
Biometrika, 64, 439-447.
- WEDDERBURN, R.W.M., (1976). On the existence and uniqueness of
the maximum likelihood for certain generalised linear
models.
Biometrika, 63, 27-32.