UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADEMICA DE LOS CICLOS PROFESIONALES Y DE POSGRADO DEL COLEGIO DE CIENCIAS Y HUMANIDADES

INSTITUTO DE INVESTIGACIONES EN MATEMATICAS APLICADAS Y EN SISTEMAS

"RESIDUALES EN MODELOS PARA DATOS CATEGORICOS"

T E S I S

Que para obtener el grado de

MAESTRO EN ESTADISTICA E
INVESTIGACION DE OPERACIONES

Presenta el Actuario

Jorge Manuel Olguin Uribe

México, D. F. Octubre, 1986

Para mis padres Mercedes y Francisco

Para Erika e Iván

Con agradecimiento al Dr. Francisco J. Aranda quien sugirió el tema y me asesoró durante el desarrollo, a la M. en C. Belem Trejo que revisó el trabajo final e hizo valiosas sugerencias para mejorarlo, a Adriana Castellanos que mecanografió gran parte y finalmente a los compañeros y amigos que me apoyaron en su realización.

INDICE

1. INTRODUCCION

McCullagh y Nelder (1983, p.24) señalan que el ajuste de un modelo estadístico a un conjunto de datos puede verse como una manera de reemplazar un conjunto de valores y por un conjunto de valores $\hat{y}$ derivados de un modelo que usualmente contiene un número relativamente pequeño de parámetros. En general las $\hat{y}$'s no son iguales a las y's por lo que surge la pregunta de què tanto difieren unas de otras, ya que mientras pequeñas diferencias son tolerables, no sucede lo mismo si las diferencias son grandes.

Méndez (1976, p.109) indica que al ajustar un modelo estadístico "se debe proceder con criterio científico, o sea, mediante

modelos provisionales que deben ser confrontados con las observaciones prácticas del mundo real; el modelo se va modificando hasta ser satisfactorio desde el punto de vista práctico. En él las suposiciones deberán cumplirse aproximadamente y producir errores lo suficientemente pequeños para garantizar su uso práctico".

El presente trabajo se centra en el estudio de estos errores o residuales que según se les defina o el trato que se les dé pueden ser de gran ayuda en el ajuste de modelos estadísticos. El objetivo del trabajo es el análisis de residuales en el ajuste de modelos estadísticos con datos categóricos.

En el capítulo dos se hace una revisión del uso de residuales en el ajuste de modelos de regresión lineal y de diseño de experimentos o análisis de varianza, a los cuales se les denomina modelos lineales "clásicos" o modelos lineales normales.

En el capítulo tres se hace una exposición de los diferentes tipos de variables categóricas, el arreglo de datos en tablas de contingencia multidimensionales y se hace una exposición breve de de dos tipos de modelos para datos categóricos: los modelos logarítmicos lineales y los modelos logísticos lineales, así como de la relación que existe entre ellos. Algunas características especiales de estos modelos se presentan en las aplicaciones del capítulo seis.

La presentación de residuales para datos categóricos se hace desde el punto de vista de los modelos lineales generalizados que incluyen, entre otros, a los modelos lineales normales y a los modelos para datos categóricos mencionados.

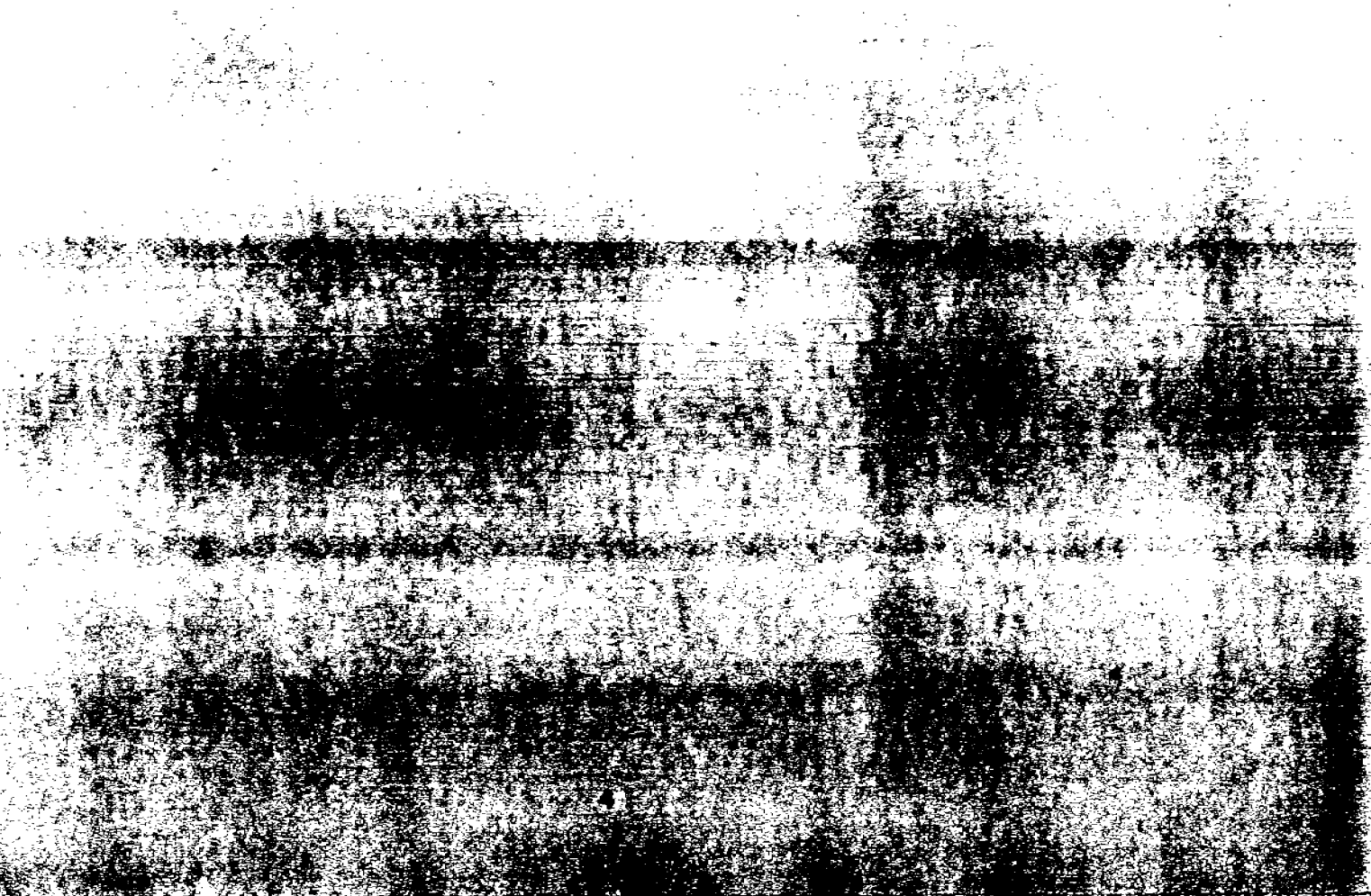
Por lo anterior, en el capitulo cuatro se hace una exposición de las características principales de los modelos lineales generalizados y en el capitulo cinco se definen diferentes tipos de residuales aplicables a estos modelos o a casos particulares de ellos y se revisan algunas recomendaciones generales para su análisis

En el capitulo seis se realizan aplicaciones que consideran diferentes aspectos en el ajuste de modelos para datos categóricos, el cálculo de diferente tipo de residuales y la forma de interpretarlos ya sea para ver si su comportamiento es el esperado o para realizar las modificaciones que este sugiera. En este último caso se utilizan algunos de los métodos revisados para encontrar un modelo adecuado. Para estas aplicaciones se utilizan datos presentados en diferentes artículos o libros, sin embargo, se va más allá del problema planteado en ellos con el propósito de ilustrar el análisis de residuales principalmente y, en su caso, encontrar una solución.

Al final se presentan las concluciones del trabajo y un anexo con los programas utilizados para el uso del paquete de cómputo "Generalized Linear Interactive Modelling" (GLIM) elaborado por Baker y Nelder (1978) para el ajuste de modelos lineales

generalizados. Con este paquete se realizaron la mayor parte de los cálculos para las aplicaciones presentadas en el capítulo seis.

Se considera importante señalar que al emplear criterios de estadística matemática en el ajuste de modelos no se pretende establecer relaciones causa-efecto que normalmente son determinadas por otro tipo de conocimientos científicos. Esto se menciona porque ocasionalmente se encuentran relaciones espúreas, es decir, modelos ajustados que producen errores pequeños y que aparentemente cumplen con las suposiciones básicas pero que no son válidos para explicar la situación del mundo real.



2.- ANALISIS DE RESIDUALES EN MODELOS LINEALES CLASICOS

El modelo estadístico lineal que se utiliza tanto en regresión como en diseño de experimentos se representa en forma matricial como sigue:

$$y = Xb + e \qquad (2.1)$$

donde: y es un vector de n observaciones.

X es una matriz n x p de variables explicativas con valores conocidos ($p<n$).

b es un vector de p parámetros desconocidos.

e es un vector aleatorio (nxp) comunmente conocido como vector de errores.

Generalmente se consideran las siguientes suposiciones:

a) $E(e_i) = 0 \quad i=1,2,\ldots,n$

b) $V(e_i) = \sigma^2 \quad i=1,2,\ldots,n$

c) $Cov(e_i,e_j) = 0 \quad i=j$

d) $e_i \sim N(0,\sigma^2) \quad i=1,2,\ldots,n$

Méndez (1976, p.42) señala que en la metodología estadística surgen dos grandes divisiones al considerar los renglones de la matriz X: si los elementos de los renglones de esta matriz se introducen como "variables indicadoras de la presencia o ausencia de los efectos que definen a las poblaciones (que se estudian), se tienen los modelos de diseños experimentales. En este caso (los elementos de los renglones de X) toman los valores cero para indicar la ausencia y uno para indicar la presencia de dichos efectos."

"Si son variables reales con valores irrestrictos dentro de ciertos intervalos...se obtienen los modelos de regresión."

"Se pueden obtener ambos tipos de comportamiento...en los llamados modelos de covarianza."

El vector $\hat{b}$ de parámetros estimados se obtiene al minimizar la suma de cuadrados de los errores en (2.1)

$$SCE = e'e = (y - Xb)'(y - Xb)$$

El proceso de minimizaciòn de esta suma de cuadrados conduce al sistema de "ecuaciones normales"

$$X'X\hat{b} = X'y$$

En los modelos de regresiòn X'X es de rango completo por lo que tiene inversa. En los modelos de diseño de experimentos se emplean diversos mètodos para solucionar las ecuaciones normales. Para facilitar la exposicion, cuando se traten los modelos de diseños de experimentos en este trabajo, se supondrà que se ha hecho una previa reparametrizaciòn en el modelo original de manera que X'X tenga inversa.

Aceptada esta suposiciòn, se tiene que en ambos casos el estimador de minimos cuadrados b es:*

$$\hat{b} = (X'X)^{-1}X'y$$

El vector de valores ajustados (estimados) y correspondiente al vector de observaciones y se obtiene como sigue:

$$\hat{y} = X\hat{b}$$

La diferencia entre la i-ésima observaciòn y_i y el i-ésimo valor ajustado $\hat{y}_i$ es el i-ésimo residual simple $\hat{e}_i$, los n residuales simples se pueden expresar en forma matricial como:

---

* En el caso de los modelos de diseño de experimentos, los componentes de b serán funciones lineales estimables (linealmente independientes) de los parámetros originales.

$$\hat{e} = y - \hat{y}$$

que es un estimador del vector de errores en (2.1). Este vector de residuales puede expresarse de la siguiente manera:

$$\hat{e} = (I - X(X'X)^{-1}X')y = (I - R)y$$

donde $R = X(X'X)^{-1}X'$ es una importante matriz que aparece en repetidas ocasiones durante el estudio de los modelos lineales.

Se puede probar también que:

$$E(\hat{e}) = (I - R)Xb = 0$$

y

$$V(\hat{e}) = (I - R)\tau^2 \qquad (2.2)$$

Entonces $V(\hat{e}_i)$ está dada por el i-ésimo elemento de la diagonal de (2.2) y $Cov(\hat{e}_i,\hat{e}_j)$ por el elemento (i,j) de esa misma matriz, así, la correlación entre $\hat{e}_i$ y $\hat{e}_j$ es

$$corr(\hat{e}_i,\hat{e}_j) = cov(\hat{e}_i,\hat{e}_j)/[V(\hat{e}_i)V(\hat{e}_j)]^{1/2}$$

Otro elemento importante para el análisis es el estimador de τ^2 que, si el modelo es correcto está dado por

$$\hat{\tau}^2 = s^2 = \sum\hat{e}_i^2/(n-p)$$

2.1 ANALISIS DE RESIDUALES EN MODELOS DE REGRESION

Este apartado se basa principalmente en el capitulo 3 del libro de Draper y Smith "Applied Regression Analysis" (segunda edición). Lo que se presenta no solo es vàlido para modelos de regresiòn sino tambièn para modelos de diseño de experimentos (conocidos tambien como modelos de anàlisis de varianza), por lo que en el apartado 2.2 donde se trata el anàlisis de residuales para estos ùltimos, se revisan principalmente los trabajos relacionados con observaciones discrepantes.

Estos autores señalan que el anàlisis de residuales debe permitir al investigador llegar a una de las siguientes concluciones:

a) Las suposiciones bàsicas sobre el modelo parecen ser violadas (de una manera que pueda ser especificada).

b) Las suposiciones no parecen ser violadas.

Presentan diferentes tècnicas para el anàlisis de residuales en modelos de regresiòn, algunas se refieren a la observaciòn visual de diferentes tipos de graficacion de èstos y otras a la obtenciòn de estadisticas que reflejan algùn aspecto en el comportamiento de los mismos.

2.1.1 Métodos gráficos

Dentro de los métodos gráficos se encuentran los siguientes:

i) La graficación de todos los residuales para observar si parecen provenir de una población Normal.

ii) Gráficas de los residuales contra el tiempo. Estas pueden hacerse cuando el tiempo interviene como una variable explicativa y las observaciones se han hecho en intervalos de tiempo conocidos.

iii) Gráficas de los residuales contra los valores ajustados $\hat{y}_i$.

iv) Gráficas de los residuales contra las variables explicativas o predictoras x_{ji}, $i=1,2,\ldots,n$.

Para el inciso i), se pueden utilizar los residuales estandarizados $\hat{e}_i/s$, por ejemplo en un histograma de frecuencias y compararlos con el modelo N(0,1). También puede utilizarse el llamado "papel Normal".

El "papel Normal" es una hoja de graficación elaborada especialmente de modo que la función de distribución N(0,1) en lugar de tener la apariencia usual, aparece como una linea recta con una pendiente positiva impresa en la hoja. Frecuentemente la escala del eje vertical (algunos autores prefieren el horizontal) corresponde a los valores que toma una variable con distribución

N(0,1), en esta escala, a diferencias iguales entre nûmeros corresponden distancias iguales en la hoja. En cambio, el eje horizontal tiene una escala especial; esta escala va del percentil 0.01 al 99.99 pero de tal forma que la distancia en el papel entre un percentil y otro se va ampliando conforme se avanza del percentil 50 al 99.99 y se va reduciendo cuando se baja del percentil 50 al 0.01. Esta escala està dada por $x=F^{-1}(y)$ donde F(y) es la funciòn de distribuciòn de una N(0,1). Esta hoja de graficaciòn es ûtil cuando se tiene una muestra independiente $x_1,x_2,\ldots,x_m$ que que se supone proviene de una N(0,1). El primer paso es ordenar los datos en forma ascendente. Supòngase que ya estàn ordenados, entonces el dato x_i se grafica sobre un valor del eje horizontal. Draper y Smith (1981) proponen que este valor sea (i-1/2)/m. Sin embargo Tukey (1962) analiza ademàs otros valores como i/(m+1) y (3i-1)/(3n+1); de este ûltimo afirma que "es simple y con seguridad uno de los màs adecuados."

Al seguir este mètodo para graficar los residuales estandarizados se podrìa concluir que èstos provienen de una N(0,1) si se observa que estàn pegados a la linea recta que representa a la funciòn de distribuciòn N(0,1). Este es un mètodo visual y por tanto subjetivo, Draper y Smith (op.cit.) recomiendan que para adquirir experiencia en la visualizaciòn de estas ghràficas conviene hacer ensayos graficando nûmeros aleatorios con distribuciòn Normal.

Por otra parte, Draper y Smith advierten que aunque los errores teòricos e_i se suponen independientes y con varianza constante,

esto no sucede con los residuales para los que $V(\hat{e}_i) = (1 - r_{ii})$, donde r_{ii} es el i-ésimo elemento de la diagonal de la matriz $R = X(X'X)^{-1}X'$, por lo que dependen de la forma de la matriz X, entonces, si hay grandes variaciones en las $V(\hat{e}_i)$ es más adecuado graficar los $\hat{e}_i/[(1 - r_{ii})s^2]^{1/2}$ (conocidos como residuales studentizados).

Sin embargo Behnken y Draper (1972) afirman que en la mayoría de los conjuntos de datos la graficación de los $\hat{e}_i$ o de los $\hat{e}_i/s$ tiende a reflejar las mismas características generales, buenas o malas, que los residuales studentizados por lo que, en general, se pierde muy poco al usar las formas más simples. En conclusión, señalan que aunque es más correcto el uso de los residuales studentizados, en casi todos los problemas prácticos es el uso de los $\hat{e}_i$ o de los $\hat{e}_i/s$ es perfectamente adecuado.

Para los incisos ii), iii) y iv), Draper y Smith presentan cuatro formas de comportamiento "tipo" de los residuales. En la figura 2.1 se presentan estos patrones de comportamiento. Aunque desde luego, se aclara que pueden ocurrir combinaciones o variaciones de éstos, por ejemplo, tendencias en dirección inversa etc.

La figura 2.1(a) (los residuales dentro de una banda horizontal) sugiere que las suposiciones sobre el modelo no han sido violadas, sin embargo, si se trata del inciso ii), una observación más detallada puede reflejar un comportamiento de

tipo cíclico en intervalos de tiempo más o menos iguales. Esto se puede corregir introduciendo al modelo variables "dummy" que corrijan el incremento uniforme del tiempo.

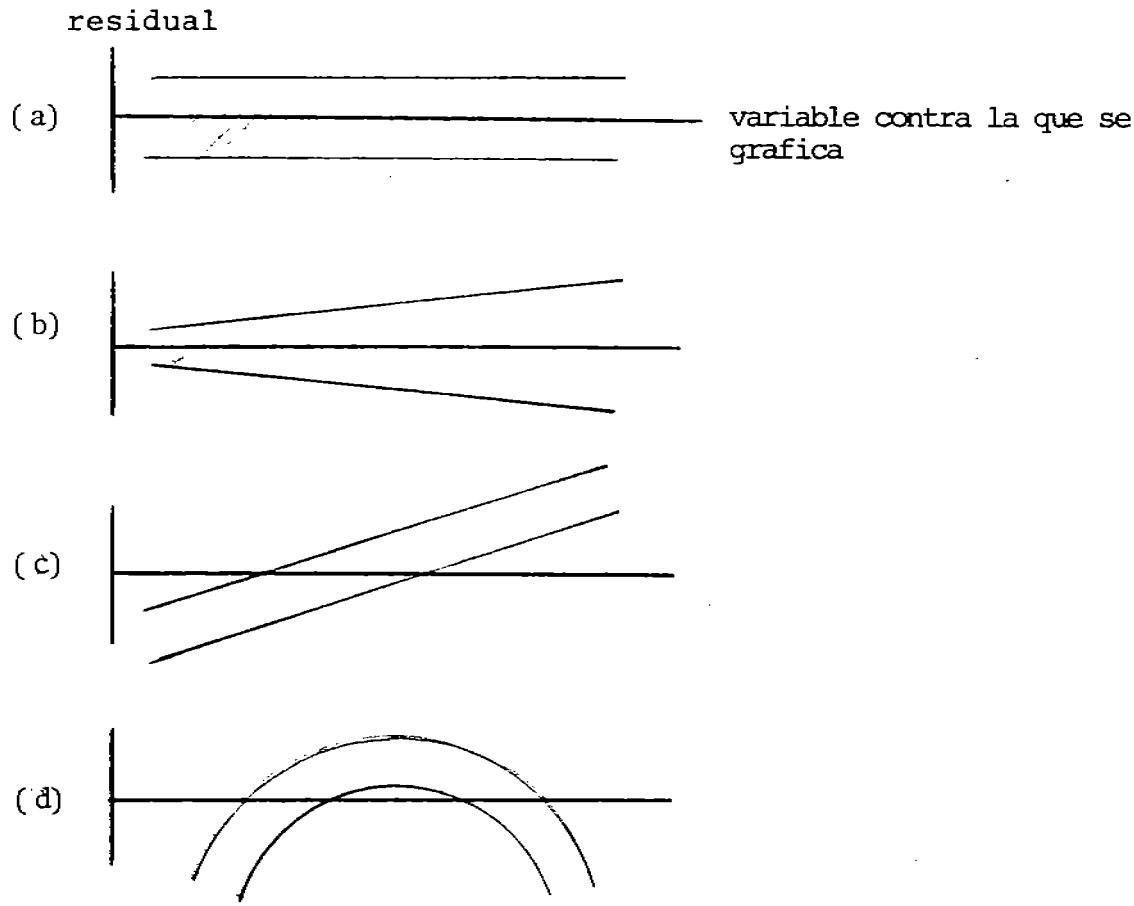


Figura 2.1

Una gráfica como la de la figura 2.1(b) en cualquiera de las graficaciones ii), iii) y iv) sugiere que las $V(e_i)$ no son iguales (lo que se conoce como heterocedasticidad) y que posiblemente los e_i están correlacionados. Para tratar estas

situaciones es necesario usar la tècnica de "minimos cuadrados ponderados", u otro tipo de transformaciòn de las observaciones y_i antes de la regresiòn. La tècnica de minimos cuadrados ponderados consiste en efectuar una transformaciòn lineal en las observaciones y_i que logre que èstas sean independientes y de varianza constante. Sin embargo, como lo señala Mèndez (op. cit. p. 116) "para encontrar la transformaciòn adecuada es necesario conocer el tipo de dependencia y heterocedasticidad, cosa que en la pràctica no es muy frecuente. Se puede sustituir el conocimiento de la estructura real de la dependencia y heterocedasticidad por estimadores de varianza y covarianza, utilizando al efecto un mètodo aproximado. Para efectuar la estimaciòn es necesario tener repeticiones de algunos de los puntos muestrales." Conviene señalar, sin embargo, que aùn en los casos en que se tiene dependencia y heterocedasticidad, los estimadores obtenidos por el mètodo de minimos cuadrados son insesgados aunque la varianza de las estimaciones se incrementa lo cual distorciona las pruebas de hipòtesis sobre los paràmetros.

Con respecto a un comportamiento de los residuales similar al de la gràfica 2.1(c) se indica, para la graficaciòn ii), que se podria corregir incluyendo un tèrmino lineal en el tiempo; para la graficaciòn iii), que existe un error en el anàlisis ya que la desviaciòn de la ecuaciòn ajustada es sistemàtica, este efecto puede ser causado por omitir errôneamente el paràmetro b_o en el modelo (un vector de unos como primera columna de la matriz X);

finalmente en la graficación iv), este patrón de comportamiento indica que posiblemente hay error en los cálculos.

Un patrón de comportamiento de los residuales como el que se muestra en la figura 2.1(d) indica: para la graficación i), que hace falta incluir en el modelo términos lineales y cuadráticos; para las graficaciones iii) y iv), que se necesitan algunos términos extra, por ejemplo, se podría incluir alguna variable explicativa o alguna función de ésta o de alguna de las ya incluidas o incluso, el producto de dos o más de éllas; también podría ser que lo que hiciera falta fuera alguna transformación en las observaciones y_i.

Estos tipos de graficación son los básicos, sin embargo, el investigador puede efectuar cualquier otro tipo de gráfica que le sugiera la estructura de los datos con los que está trabajando. No se recomienda la graficación de los residuales contra las observaciones y_i ya que frecuentemente se encuentran correlacionados, mientras que la correlación entre los residuales y los valores ajustados es nula.

2.1.2 Observaciones discrepantes

Una observación discrepante es aquella que se aleja mucho del patrón de comportamiento o tendencia que siguen las otras

observaciones.

Se han propuesto reglas para eliminar este tipo de observaciones, sin embargo, la eliminaciòn automàtica no siempre es adecuada ya que pueden proporcionar informaciòn valiosa que por alguna inusual combinaciòn de circunstancias, las otras observaciones no lo hicieron; esto podrìa ser de vital interès, por lo que es conveniente una reflexiòn profunda antes de rechazarla.

Frecuentemente los residuales ayudan en la detecciòn de observaciones discrepantes; cuando solo existe una, probablemente sea a la que corresponda el residual de mayor magnitud despuès del ajuste. En otras palabras, cuando se observa que un residual se aleja demasiado del resto (talvez 3 o 4 desviaciones estàndar de la media del conjunto de residuales en el anàlisis), este residual muy probablemente corresponda a una observaciòn discrepante.

Cuando se tienen dos o màs observaciones discrepantes, èstas son dificiles de detectar, sobre todo en los modelos de diseño de experimentos, ya que unas pueden enmascarar a otras y lo mismo puede suceder con los residuales. En el paràgrafo 2.2, que trata del anàlisis de residuales en modelos de diseño de experimentos, se hace una revisiòn de los principales mètodos que se han desarrollado para tratar con este tipo de observaciones.

Por ahora baste con señalar que, como regla general, las

observaciones discrepantes solo deben ser rechazadas definitivamente cuando se comprueba un error de registro, medición u otro como puede ser el que no se hayan mantenido las condiciones requeridas en la muestra o experimento.

2.1.3 Observaciones influyentes

Una observación influyente es aquella que tiene una influencia importante en el ajuste de un modelo, es decir, que el modelo ajustado sería significativamente diferente si esa observación fuera eliminada previamente.

Una observación de este tipo puede o puede no producir un residual de gran magnitud y no necesariamente es una observación discrepante.

Frecuentemente el problema de observaciones de este tipo puede surgir cuando la estimación de uno o varios parámetros depende fuertemente de muy pocas observaciones.

En los casos en que se tienen observaciones influyentes, las concluciones son dudosas y para resolver el problema es necesario obtener mayor número de datos.

Cook (1977) propone la siguiente estadística para evaluar la

influencia de la i-ésima observación en el ajuste de un modelo:

$$D_i = [\hat{b}-\hat{b}(i)]'X'X[\hat{b}-\hat{b}(i)]/ps^2$$

donde $\hat{b}$ es el estimador usual de mínimos cuadrados y $\hat{b}(i)$ es el estimador de mínimos cuadrados después de omitir la i-ésima observación. D_i se compara con el valor de $F_{(p,n-p,1-\alpha)}$, donde n y p son el número de renglones y columnas de la matriz X respectivamente y α es el nivel de significancia. Si el valor de D_i resulta significativo, la i-ésima observación se considera influyente.

Se puede facilitar la interpretación de esta estadística si se expresa de la siguiente forma equivalente:

$$D_i = [\hat{e}_i/s(1-r_{ii})^{1/2}]^2[r_{ii}/(1-r_{ii})](1/p)$$

El primer factor es el cuadrado del residual studentizado visto en 2.1.1, mientras que el segundo factor es el cociente de la varianza del i-ésimo predictor y la varianza del i-ésimo residual.

Draper y Smith citan a Andrews y Pregibon para presentar otra estadística útil para detectar observaciones discrepantes y/o influyentes. La estadística llamada AP consiste en calcular el siguiente cociente de determinantes:

$$R^{k}_{i,j,\ldots}(X,y)=[D^{k}_{i,j,\ldots}|(X,y)'(X,y)|\ ]/\ |(X,y)'(X,y)| \qquad (2,3)$$

Donde (X,y) es la usual matriz X añadida por la columna de observaciones y y el operador $D^{k}_{i,j,\ldots}$ significa "llevar a cabo la operación que se indica inmediatamente, después de haber eliminado los k renglones i,j,...". Valores pequeños de (2.3) son de interès ya que posiblemente estèn asociados con observaciones discrepantes o influyentes.

Si se define $R^{o,k}$ como el menor valor de (2.3) para todas las posibles eliminaciones de k renglones y se hace una gràfica horizontal del log de (2.3) entre $R^{o,k}$ para los menores valores de k, esta gràfica frecuentemente es reveladora ya que, dado un valor de k, en ella se pueden observar los valores de este logaritmo que estàn aislados del resto y entonces k es designado como el número de observaciones que deben ser reexaminadas y son precisamente las que estàn relacionadas con los valores aislados en la gràfica.

2.2 ANALISIS DE RESIDUALES EN MODELOS PARA DISENO DE EXPERIMENTOS

En el caso de modelos para diseño de experimentos (frecuentemente llamados modelos de Anàlisis de Varianza "ANOVA"), la mayor parte de los mètodos vistos en 2.1 pueden ser aplicados. Los mètodos gràficos para tratar de verificar la suposición de normalidad son

los màs sencillos y ademàs permiten identificar residuales sospechosos que pueden corresponder a observaciones discrepantes. Son comunes ademàs, las gràficas de residuales para los diferentes niveles de cada factor o los distintos tratamientos en un bloque para inspeccionar si en alguno de ellos se detecta una falla importante que estè afectando el ajuste, por ejemplo, si hay indicios de que la varianza es mayor en un bloque o aumenta al variar los niveles de un factor etc.

Sin embargo, en una revisiòn de tìtulos de trabajos sobre residuales en modelos de Anàlisis de Varianza, se observa que los autores que tratan este tema se centran principalmente en la detecciòn y tratamiento de observaciones discrepantes. Queda claro que aunque èste ha sido un tema en el que se ha trabajado mucho en años recientes, la mayor parte de los problemas que se presentan no han sido completamente resueltos en la actualidad.

John (1978) en un trabajo sobre este tipo de observaciones en diseños factoriales concluye que la necesidad de examinar los residuales se considera cada vez màs importante, pues aparte de ayudar a detectar observaciones posiblemente discrepantes, sus diferentes tipos de graficaciòn proveen informaciòn acerca de lo adecuado del modelo y las suposiciones subyacentes al anàlisis.

Indica ademàs, que si solamente està presente una observaciòn discrepante en los datos, los residuales necesariamente proporcionaràn informaciòn para detectarla.

Por otra parte, Beckman y Cook (1983) hacen una revisiòn muy completa sobre la literatura acerca de las observaciones discrepantes y muestran el papel que juegan los residuales en su detecciòn.

Al hablar en general sobre las observaciones discrepantes, señalan que una revisiòn de la historia de èstas muestra que aunque las tècnicas especificas para tratarlas han cambiado con el tiempo, no ha sucedido asi con la actitud bàsica hacia ellas.

Mencionan que la mayor parte de las definiciones que se dan sobre observaciones discrepantes son subjetivas pues en esencia casi siempre se dice que son aquellas que a juicio del investigador se apartan considerablemente del resto de los datos o de lo que se esperaria de ellos.

Citan a Collet y Lewis quienes reportan un experimento para detectar la subjetividad para identificar una observaciòn como discrepante y concluyen que la inclinaciòn individual para percibir una observaciòn como discrepante depende del mètodo de presentaciòn, de la experiencia del investigador y de la escala de los datos.

Señalan, sin embargo, que en grandes conjuntos de datos o en modelos complicados ya sea de regresiòn mùltiple, diseños experimentales o casos de datos multivariados, una inspecciòn visual de los datos puede resultar imposible por lo que es

necesario aplicar algún criterio objetivo.

Una idea general detrás de los métodos para identificar una observación discrepante consiste en transformar los datos en un conjunto de n estadísticas univariadas, una para cada observación, y luego inspeccionarlas visualmente o construir una prueba de significancia. Se han sugerido muchas transformaciones diferentes, algunas están basadas en modelos específicos alternativos y algunas son designadas para reflejar características específicas de las observaciones individuales.

Por ejemplo, los residuales studentizados son usados frecuentemente para identificar observaciones discrepantes en el análisis de modelos lineales. A la observación con el mayor residual studentizado en valor absoluto se le da especial atención y es tomada como la observación con mayores posibilidades de ser discrepante.

A continuación se examinan algunas formas que se han propuesto para tratar las observaciones discrepantes en modelos lineales normales, principalmente en diseños factoriales.

El desarrollo que sigue, si bien se basa en la revisión de Beckman y Cook (op. cit.), se emplea la notación que usan John y Draper (1978) y de Draper y John (1980) en los métodos que ellos proponen ya que esta notación coincide con la que se usó en el parágrafo anterior al examinar el análisis de residuales en

modelos de regresión.

2.2.1 Tratamiento de observaciones discrepantes en modelos de diseño de experimentos.

Considérese el modelo lineal básico (2.1) de n observaciones y p parámetros pero escrito (posiblemente rearreglando los renglones) de la siguiente manera:

$$E(y) = E\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} b \qquad (2.4)$$

donde y_2 corresponde a k observaciones consideradas como posibles observaciones discrepantes.

Después de ajustar este modelo por mínimos cuadrados, el vector de residuales puede expresarse como

$$e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = (I-R)y = \begin{bmatrix} I-R_{11} & -R_{12} \\ -R_{21} & I-R_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

donde $R_{ij} = X_i(X'X)^{-1}X_j'$ es una submatriz de $X(X'X)^{-1}X'$

En este caso, una alternativa para resolver el problema de las observaciones discrepantes sería rechazarlas y ajustar el modelo

$E(y_1)=X_1b$, obtener $\hat{b}=(X_1'X_1)^{-1}y_1$ y entonces estimar $\hat{y}_2=X_2\hat{b}$

Otra forma es ajustando el siguiente modelo:

$$E\begin{bmatrix}y_1\\y_2\end{bmatrix} = \begin{bmatrix}X_1 & 0\\X_2 & I\end{bmatrix}\begin{bmatrix}b\\g\end{bmatrix} \qquad (2.5)$$

donde g es un vector k x 1 de paràmetros adicionales cuyo estimador $\hat{g}$ permite una revisiòn del vector de observaciones discrepantes y_2. Los estimadores de $\hat{b}$ y $\hat{g}$ son

$$\hat{b} = (X_1'X)^{-1}X_1y$$

y

$$\hat{g} = y_2-(I-R_{22})^{-1}R_{21}y_1$$

Si se reemplaza y_2 por $y_2 - \hat{g}$ y se reajusta (2.4) se obtienen nuevos residuales:

$$u_1 = (I-R_{11}-R_{12}(I-R_{22})^{-1}R_{21})y_1$$
$$u_2 = 0$$

donde las dimensiones de u_1 son las mismas de y_1 en (2.4).

Este último procedimiento para ajustar y_2 necesita que $u_2 = 0$ mientras que u_1 son los residuales que resultan de ajustar $E(y_1)=X_1b$. Los u_1 se denominan "residuales revisados" y la suma de cuadrados "extra", debida al ajuste de g en el modelo (2.5) comparado con el ajuste del modelo (2.4) està dada por

$$Q_{(i)} = \hat{e}_{\underline{a}}'(I-R_{aa})^{-1}\hat{e}_{\underline{a}} \qquad (2.6)$$

Esta estadística es muy citada en trabajos que tratan con la detección e identificación de observaciones discrepantes; Beckman y Cook (op. cit.) citan a Barnett y Lewis quienes distinguen entre la versión "etiquetada" de (2.5) que es aquella en la que se tienen identificadas las observaciones discrepantes antes de la inspección de los datos y la versión "no etiquetada" en la cual son desconocidas.

Sin embargo, la revisión que hacen Beckman y Cook lleva a la conclusión de que cuando existen varias observaciones discrepantes, hasta la fecha de su artículo no estaba completamente resuelto el problema de identificarlas. Aunque es importante reconocer que se han propuesto algunas pruebas aproximadas y que muchos investigadores están trabajando en esto. Para profundizar en este tema consúltese a Beckman y Cook (1983).

3. MODELOS PARA VARIABLES (DATOS) CATEGORICAS

En los procesos de medición y/o experimentación que se realizan en diversas disciplinas, se presentan las variables categóricas. Estas corresponden a las observaciones de alguna característica de un fenómeno cuando ésta puede tomar varias modalidades que caen dentro de dos tipos de escalas de medición: nominal y ordinal.

a) Escala de medición nominal. En esta escala únicamente se les dan nombres a las modalidades de una característica o propiedad que presenta el fenómeno bajo estudio. Por ejemplo, cuatro marcas diferentes de un producto; las diferentes variedades de las plantas de café; el sexo de un insecto etc.

b) Escala de medición ordinal. En ésta también se les dan nombres a las modalidades que presentan las propiedades de algún fenómeno, pero estos nombres mantienen una relación de orden de acuerdo con la intensidad de la propiedad. Por ejemplo, la actitud de una persona hacia alguna circunstancia puede ser favorable, indiferente o contraria; la calidad de un vino podría catalogarse como excelente, buena, regular o mala; una quemadura, de acuerdo con la clasificación médica, puede ser de primer, segundo o tercer grado etc.

Para cualquiera de estas dos escalas, en muchas ocasiones conviene distinguir a las variables dicotómicas, es decir, aquellas en que la observacion solo puede presentar dos modalidades. Por ejemplo, el sexo solo puede ser masculino o femenino; se puede pensar en algún tipo de evento en el que solo se puede tener éxito o fracaso etc. A este tipo de variables frecuentemente se les llama variables binarias.

En cualquiera de las escalas mencionadas, frecuentemente se les llama categorías a los nombres que reciben las diferentes modalidades, de aquí el nombre de variables categóricas o datos categóricos.

3.1 TABLAS DE CONTINGENCIA

Cuando se estudian varias variables categòricas simultàneamente se forman las llamadas tablas de contingencia, èstas pueden ser de dos o màs dimensiones, el nùmero de dimensiones corresponde al nùmero de variables bajo estudio, cuando se tienen tablas de màs de dos dimensiones se les suele llamar tablas de contingencia multidimensionales.

Las tablas de contingencia estàn formadas por celdas que quedan determinadas por las combinaciones de las categorías de las diferentes variables bajo estudio. De modo que al tomar una muestra de objetos o sujetos donde se observan las diferentes variables de una tabla, a cada celda le corresponde el conteo o la frecuencia con que aparece en la muestra la combinaciòn de categorías que la determinan.

Una tabla de contingencia bidimencional r x c (de r renglones y c columnas) en cuyas celdas aparecen registrados los valores observados de las frecuencias, es como la que aparece en la figura 3.1

La variable A tiene r categorías y la variable B tiene c. En los totales marginales se reemplaza el subíndice sobre el que se suma por un signo "+", de modo que

$$x_{i+} = \sum_{j=1}^{c} x_{ij} \quad , \qquad x_{+j} = \sum_{i=1}^{r} x_{ij}$$

y $\qquad n = x_{++} = \sum_{i=1}^{r} \sum_{j=1}^{c} x_{ij}$ es el tamaño de la muestra

		VARIABLE B				
		Categ. 1	Categ. 2	. . .	Categ. c	Totales
VARIABLE A	Categ. 1	x_{11}	x_{12}	. . .	x_{1c}	x_{1+}
	Categ. 2	x_{21}	x_{22}	. . .	x_{2+}	x_{2+}
	. . .	. . .	. . .	. . .	. . .	. . .
	. . .	. . .	. . .	. . .	. . .	. . .
	. . .	. . .	. . .	. . .	. . .	. . .
	Categ. r	x_{r1}	x_{r2}	. . .	x_{rc}	x_{r+}
	Totales	X_{+1}	X_{+2}	. . .	x_{+c}	$x_{++} = n$

FIGURA 3.1

Si se considera una nueva variable C con d categorías se tendrá una tabla tridimensional r x c x d, y así se puede extender esta idea hasta una tabla n-dimensional, donde cada dimensión corresponde a una de n variables.

Fienberg (1977) señala que las tablas de contingencia de más de dos dimensiones presentan problemas especiales de análisis e interpretación. Indica que este tipo de problemas ha ocupado un importante lugar en artículos de revistas científicas de estadística después de que apareció un artículo de Bartlett en

1935 sobre pruebas en una tabla 2 x 2 x 2.

El mismo Fienberg señala que hasta principios de los años 70's casi todos los investigadores manejaban las tablas de contingencia multidimensionales analizando varias tablas de dos dimensiones utilizando los totales marginales de la tabla multidimensional y aunque este tipo de tratamiento a veces proporciona una visiòn relativamente buena de la relaciòn entre las variables, se presentaban algunos problemas:

a) Se confunde la relaciòn marginal entre un par de variables, con la relaciòn entre ellas cuando otras estàn presentes.

b) No permite el anàlisis simultàneo de la relaciòn entre varios pares de variables.

c) Ignora la posibilidad de asociaciòn (interacciòn) entre tres o màs variables.

Para el anàlisis de tablas de contingencia multidimensionales actualmente son ampliamente aceptados los modelos logaritmicos lineales que como su nombre lo dice, son modelos lineales en los logaritmos de las frecuencias esperadas en las celdas. Existen algunas analogias entre los tèrminos de interacciòn en estos modelos y los modelos de anàlisis de varianza, pero en los modelos logarimicos lineales estos tèrminos son usados para describir relaciones estructurales entre las variables categòricas.

Fienberg indica también, que cuando se distingue entre variables exploratorias y variables respuesta los modelos logaritmicos lineales pueden ser convertidos en modelos logísticos o logísticos lineales.

Sin embargo, esto último tiene sus limitaciones ya que, como se verà màs adelante, los modelos logísticos para respuesta binaria, tienen un desarrollo propio y no siempre pueden verse como una transformaciòn de un modelo logaritmico lineal. Al final de este capitulo se verà esto con mayor claridad.

3.2 MODELOS LOGARITMICOS LINEALES.

La exposiciòn de este tema se basa principalmente en el libro de Fienberg (1977) "The Analysis of Cross-Classified Categorical Data".

Para facilitar la exposiciòn, considèrese una tabla de tres dimensiones I x J x K. Sea x_{ijk} la frecuencia observada en las categorias i de la primera variable "A", j de la segunda variable "B", y k de la tercera variable "C"; es decir, el conteo en la celda (i,j,k) de la tabla. Y sea m_{ijk} el valor esperado de x_{ijk}.

Si las tres variables de la tabla son todas independientes entre sí, se tiene que

$$P(A=i,B=j,C=k) = P(A=i)P(B=j)P(C=k)$$

El estimador del valor esperado m_{ijk} sería en este caso

$$\hat{m}_{ijk} = (x_{i++}/n)(x_{+j+}/n)(x_{++k}/n)n$$

Tomando logaritmos se tiene

$$\log \hat{m}_{ijk} = \log x_{i++} + \log x_{+j+} + \log x_{++k} - 2 \log n$$

Esta forma aditiva sugiere que $\log m_{ijk}$ sea expresado de la siguiente manera (lo que puede ser probado sin mucha dificultad)

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \qquad (3.1)$$

donde:

$$u = (1/IJK)\sum_i\sum_j\sum_k \log m_{ijk}$$

$$u_{1(i)} = (1/JK)\sum_j\sum_k \log m_{ijk} - u$$

$$u_{2(j)} = (1/IK)\sum_i\sum_k \log m_{ijk} - u$$

$$u_{3(k)} = (1/IJ)\sum_i\sum_j \log m_{ijk} - u$$

Como se puede observar (3.1) se asemeja a un modelo de anàlisis de varianza; el primer tèrmino es la media general de los logaritmos; los tèrminos segundo, tercero y cuarto, representan desviaciones de esta gran media y estàn en funciòn de los valores o categorias de las variables A, B y C respectivamente. Se puede abreviar el modelo (3.1) de completa independencia, con la siguiente notaciòn: [A] [B] [C]

Supòngase ahora, que las tres variables no son todas independientes entre si, se pueden tener entonces cuatro tipos de situaciones:

a) Independencia de una variable con las otras dos conjuntamente. En este caso se pueden construir tres modelos: independencia de A con B y C conjuntamente, independencia de B con A y C conjuntamente, e independencia de C con A y B conjuntamente. La notaciòn abreviada para estos tres modelos en el mismo orden es

[A] [BC] , [B] [AC] , [C] [AB]

En forma explicita el modelo representado por [A] [BC] es

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)} \qquad (3.2)$$

Como se puede ver, este modelo es el (3.1) màs un tèrmino, èste corresponde a la relaciòn estructural entre las variables B y C.

b) Independencia condicional de dos variables dada la otra. A esta situaciòn corresponden tambièn tres modelos cuya notaciòn

abreviada es:

[AB] [AC] , [AC] [BC] , [AB] [BC]

El segundo de èstos en forma màs explicita seŕa (3.2) aŕadido por el tèrmino $u_{13(ik)}$ que representa la relaciòn estructural entre las variables A y C.

c) Relaciòn entre los tres pares de variables, sin que la relaciòn entre cada par estè afectada por la otra variable. En este caso se tiene un solo modelo cuya notaciòn abreviada es:

[AB] [AC] [BC]

Para ponerlo explicitamente, se debe aŕadir al modelo descrito en el inciso b) el tèrmino correspondiente a la relaciòn entre las variables A y B.

d) Relaciòn entre las tres variables (lo que corresponderia a una interacciòn de segundo orden en un modelo de Anàlisis de Varianza). Su notaciòn abreviada es [ABC], y èste es el modelo logaritmico lineal saturado para tres dimensiones, es decir, se tiene un paràmetro para cada observaciòn (celda) y todos los modelos derivados de las situaciones descritas en los incisos a), b) y c) son casos particulares de èste. Explicitamente este modelo se expresa como sigue:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \qquad (3.3)$$

El modelo logaritmico lineal saturado para cuatro dimensiones tendrà entonces, un tèrmino que representa a la gran media de los

logaritmos, cuatro para las variables individuales, seis para las relaciones de primer orden, cuatro para las relaciones de segundo orden y uno para la relaciòn entre las cuatro variables.

De este modo, no es dificil la generalizaciòn para n variables, sin embargo, la cantidad de situaciones que se pueden presentar (o hipòtesis que se pueden establecer) aumenta considerablemente al aumentar el nùmero de variables.

Fienberg solamente considera modelos logaritmicos lineales jeràrquicos, es decir, aquellos en los que si aparece un tèrmino representando la relacion estructural entre varias variables, necesariamente deben incluirse los tèrminos que representan relaciones de menor grado entre las variables involucradas en el primero. Por ejemplo, el tèrmino u_{123} no puede estar en un modelo si no estàn los tèrminos u_{12}, u_{13} y u_{23} y, por supuesto, los tèrminos u_1, u_2 y u_3.

En una tabla de cuatro dimensiones se pueden considerar 113 modelos logaritmicos lineales jeràrquicos, todos ellos incluyendo los tèrminos del modelo de completa independencia entre las cuatro variables, de ahí que muchos autores se han preocupado por desarrollar procedimientos para la selecciòn de un modelo. En los ejemplos del capitulo 6 se presentaràa la aplicaciòn de uno de estos procedimientos.

3.2.1 Estadisticas suficientes y estimadores de màxima verosimilitud

Fienberg (op.cit.), Haberman(1973) y otros autores señalan que comunmente los datos en las celdas de una tabla de contingencia de cualquier dimensiòn surgen de alguno de los siguientes tres modelos de muestreo: Poisson (no se fija el tamaño de muestra), Multinomial (se fija el tamaño de muestra de toda la tabla), Producto-Multinomial (se fijan totales marginales para alguna o algunas variables). Señalan ademàs, que para cualquiera de estos tipos de muestreo, los estimadores de màxima verosimilitud (EMV) son los mismos para el modelo logaritmico lineal particular que se considere. Solamente se requiere que los tèrminos "u" correspondientes a los totales marginales fijados en el esquema producto-multinomial, aparezcan en el modelo logaritmico lineal bajo consideraciòn.

A continuaciòn se presentan algunos resultados bàsicos sobre estadisticas suficientes y estimaciòn de valores esperados en modelos logaritmicos lineales.

Considèrese una tabla multidimensional. Para abreviar (siguiendo a Fienberg) se usarà un solo subindice para describir las frecuencias, ya que todas las tablas pueden verse como un conjunto de frecuencias observadas $x = \{x_i : i \in F\}$, indexada por un conjunto F que contiene t elementos, las cuales son la

realización de un conjunto de variables aleatorias X similarmente indexadas.

Al conjunto de frecuencias observadas le corresponde un conjunto de valores esperados $m = \{m_i = E(x_i) : i \in F\}$ y de logaritmos de valores esperados $\lambda = \{\lambda_i = \log m_i : i \in F\}$.

Considérese primero el caso en que las frecuencias observadas provienen de distribuciones Poisson independientes. La función de verosimilitud en este caso es proporcional a:

$$\prod_{i \in F} m_i^{x_i} \exp(-m_i)$$

y el kernel del logaritmo de esta función es:

$$\sum x_i \log m_i - m_i = \sum x_i \lambda_i - \sum m_i$$

Estamos interesados en situaciones en las que m es descrito por un modelo logaritmico lineal (un modelo lineal de λ) el cual se denotará por M, se tiene el siguente resultado:

<u>Bajo muestreo Poisson, para cada parámetro en M, hay una estadistica suficiente minimal que es combinación lineal de (x_i)</u>

La estadistica suficiente minimal será denotada por $P_M x$

Ejemplos:

a) Considérese una tabla de contingencia I x J con frecuencias observadas $\{x_{ij}\}$ y sea M el modelo logaritmico lineal que especifica la independencia entre las dos variables. Entonces: $P_M x = \{x_{i+}\ i = 1,2,\ldots,I;\ x_{+j}\ j = 1,2,\ldots,J\}$

b) Considérese que en el modelo (3.3) se desea probar la hipótesis: $u_{13(ik)} = u_{23(jk)} = u_{123(ijk)} = 0$, es decir, independencia de la variable tres con la uno y dos respectivamente. Entonces: $P_M x = \{x_{ij+}:\ i=1,2,\ldots,I;\ j= 1,2,\ldots,J.\ x_{++k}:\ k=1,2,\ldots,K\}$

Otro resultado importante es el siguiente:

Los estimadores de màxima verorosimilitud (EMV), $\hat{m}$ de m bajo el modelo logarítmico lineal M existen y son ónicos si y sólo si las ecuaciones de verosimilitud.

$$P_M \hat{m} = P_M x$$

tienen una solucion que está en el interior del subespacio correspondiente a M.

Continuando con los ejemplos anteriores, se tiene:

a) Las ecuaciones de verosimilitud para el ejemplo de una tabla IxJ donde se considera la hipótesis de independencia son:

$$\hat{m}_{i+} = x_{i+} \quad i= 1,2,\ldots,I$$

$$\hat{m}_{+j} = x_{+j} \quad j= 1,2,\ldots,J$$

las cuales tienen solución única

$$\hat{m}_{ij} = x_{i+}x_{+j}/x_{++}$$

b) Para el ejemplo de tres variables donde se desea probar la hipótesis de independencia de la variable C con la A y la B respectivamente (lo que correspondería a ajustar el modelo [C][AB]), las ecuaciones de verosimilitud son:

$$\hat{m}_{ij+} = x_{ij+} \qquad i=1,2,\ldots,I;\ j=1,2,\ldots,J$$
$$\hat{m}_{++k} = X_{++k} \qquad k=1,2,\ldots,k$$

las que tienen solución única:

$$\hat{m}_{ijk} = x_{ij+}x_{++k}/x_{++k}$$

Sin embargo, no siempre se puede encontrar en forma directa la solución, por ejemplo, si en el modelo (3.3) se quiere probar la hipótesis $u_{123(ijk)} = 0$. Por las reglas dadas por Fienberg (op.cit p.31), los estimadores $\{m_{ijk}\}$ son funciones de $\{x_{ij+}\}$, $\{x_{i+j}\}$, $\{x_{+jk}\}$.

Usando el método de máxima versimilitud se tiene que $\{m_{ijk}\}$ deben satisfacer:

$$\hat{m}_{ij+} = x_{ij+} \quad i=1,2,\ldots,I;\quad j=1,2,\ldots,J$$
$$\hat{m}_{i+k} = x_{i+k} \quad i=1,2,\ldots,I;\quad k=1,2,\ldots,K$$

$$\hat{m}_{+jk} = x_{+jk} \quad j=1,2,\ldots,j; \quad k=1,2,\ldots,k$$

para las cuales no existe una solución en forma cerrada. El problema, sin embargo, puede ser resuelto usando el método conocido como "Ajuste Proporcional Iterativo" (API) o bien el algoritmo "Newton-Raphson" (NR). Para mayores detalles sobre API ver Fienberg (op. cit. p-33) y sobre NR ver Haberman (1974b). Estos algoritmos se pueden usar también en las situaciones anteriores.

Un tercer resultado importante sobre estimación en modelos logaritmico lineales es el siguiente:

Bajo el esquema de muestreo multinominal y producto-multinomial los estimadores de máxima verosimilitud de los valores esperados m son los mismos que bajo el esquema de muestreo Poisson.

3.2.2 Pruebas de hipótesis

Teniendo los valores esperados para cualquiera de los modelos logaritmicos lineales se puede verificar la bondad de ajuste del modelo con alguna de las siguientes estadisticas:

$$X^2 = \sum (\text{observados} - \text{estimados})^2/\text{estimados}$$

$$G^2 = 2 \sum(\text{observados}) \log (\text{observados/estimados})$$

Donde los valores estimados son los que se esperarian bajo la suposiciòn de que la hipòtesis en consideraciòn es cierta.

Para muestras grandes se tiene que tanto X^2 como G^2 se distribuyen como Ji-cuadrada con grados de libertad igual al nùmero de celdas menos el nùmero de paràmetros estimados en el modelo logaritmico lineal.

3.3 MODELOS LOGISTICOS PARA RESPUESTA BINARIA

Considèrese un experimento en que a cada observaciòn le corresponde una de dos posibles categorias y a cada una de estas categorias se le asignan valores cero o uno, de tal manera que si la variable aleatoria Y_i representa a la i-èsima observaciòn, èsta solamente podrà tomar uno de estos valores con alguna probabilidad, es decir,

$$Pr\{Y_i = 1\} = p_i \; ; \qquad Pr\{Y_i = 0\} = 1 - p_i$$

A este tipo de observaciones frecuentemente se les llama observaciones binarias.

Siguiendo la terminologia màs usual en teoria de probabilidades,

cuando $Y_i = 1$ se dice que en la i-èsima observaciòn ha ocurrido un "èxito" y si $Y_i = 0$, un "fracaso".

En muchas investigaciones con respuesta binaria, ya sea en diseño de experimientos o muestreo de poblaciones, existe, asociado con cada observaciòn, un vector de covariables o variables explicativas. El objetivo de los modelos logísticos lineales que aquí se presentan es evaluar la forma en que la probabilidad de respuesta p_i depende del conjunto de variables explicativas (tambièn llamadas covariables o condiciones experimentales).

Para hacer la presentaciòn en forma màs general, supòngase que para la i-èsima combinaciòn de condiciones experimentales $(X_{i1},\ldots,X_{ip})$, $i = 1,2,\ldots,N$ se hacen observaciones binarias en n_i individuos u objetos. En este caso, se dice que las observaciones son agrupadas en grupos de tamaños $n_1,\ldots,n_N$; y tienen la forma r_1/n_1, $r_2/n_2,\ldots,r_N/n_N$; donde r_i el número de èxitos en n_i observaciones, por supuesto $0 \leq r_i \leq n_i$. En el caso de $n_1=\ldots=n_N=1$, se dice que los datos estan desagrupados. De acuerdo con Mc Cullagh y Nelder (1983,p.24) la distinciòn entre datos agrupados o desagrupados es importante por dos razones:

a) Algunos mètodos de anàlisis para datos agrupados no son aplicables a datos desagrupados.

b) De acuerdo con la teoria asintòtica, para datos agrupados se pueden hacer inferencias, ya sea si $n \to \infty$ o si $N \to \infty$. Pero para

datos desagrupados solamente cuando $N \to \infty$.

Si todas las observaciones de un mismo grupo son independientes y tienen las mismas probabilidades de éxito, entonces la distribución de r_i dado n_i es binomial, comunmente denotada por $B(n_i; p_i)$ con esperanza y varianza:

$$E(r_i) = n_i p_i \ , \qquad Var(r_i) = n_i p_i (1-p_i)$$

Cox (1983 cap.1) muestra ocho ejemplos en los que se puede apreciar la gran variedad de situaciones en que se tienen respuestas binarias y diferentes tipos de variables explicativas.

Estos ejemplos van, desde una simple comparación de probabilidades de éxito en dos poblaciones diferentes, hasta diseños factoriales y problemas de regresión con respuesta binaria; pasando por el análisis de una tabla de contingencia de 2 x 2, varias tablas de contingencia de 2 x 2 y otros.

3.3.1 El modelo logistico lineal.

De acuerdo con Cox (op. cit. cap. 2) en muchos aspectos la forma más simple de representar la dependencia de una probabilidad p_i respecto de un conjunto de variables explicativas $(X_1, X_2, \ldots,$

X_p) es postular una dependencia para i= 1,...,N como:

$$p_i = \exp(b_o + \sum_{j=1}^{p} x_{ij} b_j) / [1 + \exp(b_o + \sum_{j=1}^{p} x_{ij} b_j] \qquad (3.4)$$

$$1 - p_i = 1/[1 + \exp(b_o + \sum_{j=1}^{p} x_{ij} b_j)] \qquad (3.5)$$

donde $b_o, b_1, \ldots, b_p$ es un conjunto de paràmetros desconocidos.

Nòtese que (3.4) y (3.5) satisfacen la condiciòn de que $0 \leq p_i \leq 1$ Estas ecuaciones se pueden expresar en forma equivalente como:

$$\lambda_i = \log p_i/(1-p_i) = b_o + \sum_{j=1}^{p} x_{ij} b_j \qquad (3.6)$$

que es el modelo logistico lineal. $\lambda_i = p_i/(1 - p_i)$ es conocida como la transformaciòn logistica o logaritmo de "momios" .

Con objeto de analizar las ventajas de la transformaciòn λ sobre otras, y para mostrar dos situaciones diferentes en que los modelos logisticos lineales pueden ser usados, a continuaciòn se presentan dos ejemplos.

3.3.2 Estudios prospectivos y retrospectivos.

Supòngase que se tiene una poblaciòn en la cual cada individuo es clasifiacado de acuerdo a dos variables aleatorias binarias S y T

donde:

$$S = \begin{cases} 1 & \text{si el individuo ha estado expuesto a ciertas condiciones que se piensa que propician una enfermedad.} \\ 0 & \text{si el individuo no ha estado expuesto.} \end{cases}$$

$$T = \begin{cases} 1 & \text{si el individuo padece la enfermedad.} \\ 0 & \text{si no la padece.} \end{cases}$$

La distribución de probabilidades de clasificación de los sujetos de acuerdo con estas variables se puede expresar en la siguiente tabla:

S \ T	0	1
0	p_{00}	p_{01}
1	p_{10}	p_{11}

Donde $p_{i,j}$ es la probabiliad de que un individuo esté en el nivel i de la variable S y en el nivel j de la variable T (i,j = 0,1)

Supóngase ahora que se quiere comparar la probabilidad de que un individuo padezca de enfermedad (T = 1) dado que ha estado expuesto (S = 1), con la probabilidad de que el individuo padezca la enfermedad (T = 1) dado que no ha estado expuesto (S = 0).

Existen por lo menos tres procedimientos de muestreo para este

estudio.

a) Una muestra al azar de toda la población.

b) Una muestra de individuos que han estado expuestos y otra de individuos que no han estado expuestos.

c) Una muesta de individuos que padecen la enfermedad y otra de individuos que no la padecen.

En el caso (a) se puede estimar cualquiera de las cuatro p_{ij}'s. Las probailidades condicionales de que T = 1 dado que S = 0 y S = 1 respectivamente son:

$$P(T=1|S=0) = p_{01}/(p_{00}+p_{01}) \quad y \quad P(T=1|S=1) = p_{11}/(p_{10}+p_{11}) \qquad (3.7)$$

O bien, utilizando (3.4) y (3.5)

$$P(T=1|S=1) = \exp(a+Dx_1)/[1+\exp(a+Dx_1) \qquad (3.8)$$

con $x_1=0$ si S=0 y $x_1=1$ si S=1

La transformación logistica (3.6) de las probabilidades en (3.7) es:

$$\lambda_{10} = \log[P(T=1|S=0)/(1-P(T=1|S=0))] = \log(p_{01}/p_{00})$$

y $$\lambda_{11} = \log[P(T=1|S=1)/(1-P(T=1|S=0))] = \log(p_{11}/p_{10})$$

O bien, usando (3.8):

$$\lambda_{11} = a + Dx_1 \ , \quad \text{con } x_1=0 \text{ si } S=0 \text{ y } x_1=1 \text{ si } S=1$$

y las diferencia de las transformaciones logísticas de estas probabilidades, utilizando ambas expresiones es:

$$\lambda_{11} - \lambda_{10} = D = \log(p_{11}p_{00}/p_{10}p_{01}) \qquad (3.9)$$

En el Caso (b) los tamaños de muestra para S = 0 y S = 1 son fijos, por lo que no se pueden calcular las probabilidades marginales $p_{00} + p_{01}$ y $p_{10} + p_{11}$. Este sería el caso de un estudio prospectivo, ya que se toman muestras fijas de individuos que han y que no han estado expuestos a condiciones que supuestamente propician la enfermedad y posteriormente se observan los valores de T, es decir, cuantos de cada situación padecen la enfermedas y cuantos no la padecen.

En este caso, las probabilidades p_{ij}'s no pueden ser estimadas individualmente, sin embargo, al observar los valores de T sí se pueden estimar las probabilidades condicionales en (3.7) y (3.8), y en particular se puede estimar la diferencia logística (3.9).

En el procedimiento (c), los tamaños de muestra de individuos que padecen y no padecen la enfermedad son fijos (se fijan para T=0 y para T=1) y posteriormente a la selección de las muestras, se averigua si los individuos han estado expuestos o no a las

condiciones que podrían propiciar la enfermedad, por lo que se trataría de un estudio retrospectivo. En este caso se pueden obtener estadísticas que están en función de:

$$p_{10}/(p_{00}+p_{10}) \quad y \quad p_{11}/(p_{01}+p_{11})$$

que son precisamente las probabilidades condicionales expresadas en (3.7) y (3.8) y por lo tanto la diferencia logística es la misma que en (3.9) y puede ser estimada.

Este ejemplo es presentado por Cox (1983,op.cit) y McCullagh y Nelder (1983,op.cit. cap. 4) con algunas diferencias. Pero la conclusión es básicamente la misma: si el interés es considerar T como variable respuesta y S como variable explicativa, pero el esquema de muestreo está basado en el procedimiento inverso de fijar los tamaños de muestra para T y luego observar S (que, dicho sea de paso, es lo más frecuente ya que en la mayoría de los problemas prácticos es más fácil localizar antes a los enfermos) la diferencia logística puede ser estimada. Esta es una de las ventajas que tiene la escala logística con respecto a otras escalas.

Fienberg (1977,op.cit.) muestra una prueba de hipótesis de independencia en una tabla 2 x 2 basada en la estimación de esta diferencia logística.

3.3.3 Relación entre un estímulo (variable continua) y una respuesta binaria.

De acuerdo con Cox (op.cit.) esta situación se presenta con mucha frecuencia. Existe un estímulo bajo control del investigador; cada individuo es asignado a un nivel del estímulo y una respuesta binaria es observada. Un importante campo de aplicación se encuentra en experimentos biológicos donde por ejemplo, diferentes niveles del estímulo pueden ser representados por diferentes dosis de veneno y la respuesta binaria de cada sujeto es muerte o supervivencia.

En tales aplicaciones muchas veces es posible seleccionar una medida x de la intensidad del estímulo de tal manera que la probabilidad de "éxito" es cero para x negativa y uno a partir de algún valor grande y positivo de x y la probabilidad es una función estrictamente creciente de x. De hecho esta función tiene las propiedades matemáticas de una función de distribución continua. Si la escala de x se selecciona adecuadamente, la función de distribución será simétrica; por ejemplo, en la aplicación mencionada frecuentemente es útil tomar x como el logaritmo de la dosis de veneno.

Haberman (1973) presenta un ejemplo de este tipo y el cual se abordará en diferentes partes de este trabajo, concretamente en el capítulo 5, donde se presentan los residuales de Haberman y en

el capitulo 6, donde se hace una aplicaciòn sobre detecciòn y tratamiento de observaciones discrepantes.

En este ejemplo N_j sujetos reciben una log dosis t_j de un veneno, $1 \leq j \leq r$. Para cada sujeto existen dos respuestas posibles al estimulo: supervivencia o muerte. Sea n_{jk} el nùmero de sujetos que habiendo recibido la dosis t_j tienen respuesta k; k=0,1. Para emplear un modelo logistico, la probabilidad de respuesta cero dada la dosis t_j se puede expresar como:

$$Pr\{Y = 0|j\} = 1/[1+\exp(-a-bt_j)] \quad 1 \leq j \leq r$$

para a y b desconocidos.

Al aplicar la transformaciòn logistica se tiene:

$$\log[Pr\{Y=0|j\}/Pr\{Y=1|j\}] =$$

$$= \log\{[1/(1+\exp(-a-bt_j))]/[\exp(-a-bt_j)/(1+\exp(-a-bt_j))]\}$$

$$= a + bt_j \qquad 1 \leq j \leq r$$

3.3.4 Estimaciòn y pruebas de bondad de ajuste.

Considèrese elmodelo logistico lineal (3.6). De acuerdo con

Fienberg (1979) para estimar los paràametros b's en este modelo, se usa el mètodo de màxima verosimilitud, sin embargo, las p+1 ecuaciones de verosimilitud resultantes para los p+i paràmetros b's no tienen una soluciòn cerrada, por lo que para llegar a su soluciòn numèrica es necesario usar algùn procedimiento computacional iterativo.

Haberman(1974), por ejemplo sugiere el uso de un procedimiento Newton-Raphson modificado que tiene propiedades cuadràticas de convergencia.

Para las aplicaciones que se presentan en el capitulo 6 del presente trabajo se utiliza el paquete de còmputo "Generalized Linear Interactive Modelling" (GLIM) desarrollado por Beckman y Nelder (1978) y que para la estimaciòn de paràmetros en modelos para datos categòricos utiliza el algoritmo Newton-Raphson modificado de diversas formas segùn el tipo de modelo que se trate.

Para probar la bondad de ajuste de un modelo logistico lineal, se usan las estadisticas G^2 y/o X^2 vistas en 3.2, las cuales tienen una distribuciòn aproximada Ji-cuadrada.

Cabe señalar que conforme los tamaños de muestra son mayores, las estadisticas G^2 y X^2 tienden a tomar los mismos valores para la prueba de un modelo determinado.

3.1 RELACION ENTRE LOS MODELOS LOGARITMICOS LINEALES Y LOS MODELOS LOGISTICOS LINEALES

En determinadas circunstancias en que se tiene respuesta binaria, el modelo logistico puede ser tratado desde el enfoque de los modelos logaritmico lineales. Fienberg (op.cit.) muestra este procedimiento con unos ejemplos.

En el primer ejemplo, toma unos datos reportados por Bliss (1967) acerca de una investigación que consistió en plantar en el invierno retoños de dos variedades de pino; para cada variedad se experimentaron dos tipos de plantación: demasiado superficial (ds) y demasiado profunda (dp). Para cada tipo de plantación se usaron 100 retoños de cada variedad de pino, y en el otoño del suguiente año se observó cuantos pinos habían sobrevivido. En el cuadro 3.1 se presentan los datos.

Para ajustar un modelo logaritmico lineal a estos datos, se debe incluir el término u_{13} en todos los modelos que se consideren, con objeto de que al aplicar el método de ajuste proporcional iterativo, los valores esperados estimados no alteran los totales marginales de las variables 1 y 3 (profundidad de plantación y variedad de pino) fijados por el diseño.

C U A D R O 3.1

Retoños "Longleaf"

Tipo de plant.	muer-tos	vivos	tot.
dp	41	59	100
ds	11	89	100
tot.	52	148	200

Retoños "Slash"

Tipo de plant.	muer-tos	vivos	tot.
dp	12	88	100
ds	5	95	100
tot.	17	183	200

En el cuadro 3.2 se encuentran en notación abreviada, los modelos jerárquicos permitidos para este conjunto de datos y los valores de las correspondientes estadísticas de bondad de ajuste.

C U A D R O 3.2

Modelo	X^2	G^2	g.l.
[12] [13] [23]	1.37	1.28	1
[12] [23]	26.54	27.79	2
[12] [13]	24.03	25.03	2
[13] [2]	54.7	50.1	3

Es claro que solamente el modelo de "no interacción de tres factores" se ajusta a los datos, sin embargo, se reconoce que tomando otros modelos que consideren $u_{13} = 0$ se podrían obtener valores pequeños de X^2 y G^2 pero los valores esperados estimados podrían no corresponder a los tamaños de muestra fijados por el diseño.

Puesto que el interés está en los efectos de la profundidad de plantación y la variedad de pino sobre la mortalidad, es razonable observar la razón de mortalidad (muertos/vivos) para cada combinación de variedad de pino y tipo de plantación, esto es: m_{i1k}/m_{i2k}

Puesto que el modelo ajustado es

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

se tiene que el modelo logístico lineal es:

$$\log(m_{i1k}/m_{i2k}) = [u_{2(1)} - u_{2(2)}] + [u_{12(i1)} - u_{12(i2)}] + [u_{23(ik)} - u_{23(2k)}] =$$

$$= 2[u_{2(1)} + u_{12(i1)} + u_{23(ik)}]$$

$$= w + w_{1(i)} + w_{3(k)} \qquad (3.10)$$

Nótese que el primer miembro de (3.10) es equivalente al primer miembro del modelo logístico (3.6)

En el último miembro de (3.10) no aparece el subíndice que indica la variable 2 (mortalidad) ya que el primer miembro es el logaritmo de la razon de mortalidad.

Fienberg (op.cit.) señala que este procedimiento se usa con preferencia cuando los tamaños de muestra de las variables explicativas son controladas por el diseño pero que muchos investigadores ajustan este tipo de modelos en conjuntos de datos en los cuales algunas variables explicativas no son controladas por el diseño por lo que pierden información sobre las relaciones entre ellas.

Como se ha visto, con este procedimiento los modelos logísticos contienen términos que corresponden a los de los modelos logarítmicos lineales.

Sin embargo este procedimiento tiene algunas limitaciones y desventajas que se pueden resumir en los siguientes puntos.

a) No se puede aplicar cuando alguna de las variables explicativas es continua.

b) Cuando se utiliza este procedimiento los modelos logarítmicos deben ser jerárquicos por lo que en ocaciones se estiman parámetros que no son de interés, lo que puede hacer decrecer la precisión en la estimación de los otros parámetros.

c) Pueden existir otros modelos logísticos desde el enfoque de 3.3 que no tienen su equivalente en modelos logarítmicos lineales jerárquicos.

d) Como ya se mencionó, se pierde información cuando las variables no son controladas por el diseño.

4. MODELOS LINEALES GENERALIZADOS

Tanto los modelos logaritmicos lineales como los modelos logisticos son "modelos lineales generalizados", este término debido a Nelder y Wedderburn (1972), se refiere a una generalización de los modelos lineales clásicos al introducir dos conceptos básicos: a) la distribución de las observaciones puede ser de una familia exponencial (no necesariamente normal) y b) El concepto de la función liga.

Muchos resultados válidos para los modelos logaritmicos lineales y para los modelos logisticos, son válidos para otros modelos lineales generalizados, en particular la definición de residuales que se analizará con detalle. Por otra parte, existen resultados

propios de los modelos lineales generalizados que al ser tratados para el caso especial de datos categòricos enriquecen el anàlisis de los modelos especificos para este tipo de datos. Entonces la presentaciòn del anàlisis de residuales desde el punto de vista de los modelos lineales generalizados, ademàs de enriquecerla, permite dar una mayor homogeneidad a la discusiòn y se puede apreciar la amplitud de aplicaciones de los resultados independientemente de las que para datos categòricos se hacen aquí.

La exposiciòn que se hace aquí està basada principalmente en el libro "Generalized Linear Models" de Mc Cullagh y Nelder (1983). Estos autores señalan que los modelos lineales generalizados son una extensiòn de los modelos lineales clàsicos por lo que èstos constituyen un adecuado punto de partida.

Sea $\mathbf{y}$ un vector de observaciones de longitud N que se supone es la realizaciòn de un vector de variables aleatorias $\mathbf{Y}$ independientemente distribuidas con medias $\boldsymbol{\mu}$. El vector de medias $\boldsymbol{\mu}$ constituye la parte sistemàtica del modelo y se supone la existencia de variables explicativas (covariables) x_1, $x_2,\ldots,x_p$ con valores conocidos tales que:

$$\mu = \sum_{j=1}^{p} b_j x_j$$

donde las b's son paràmetros usualmente desconocidos y tienen que ser estimados de los datos. Si se indexan por i las

observaciones, la parte sistemàtica puede expresarse como sigue:

$$E(Y_i) = \mu_i = \sum_{j=1}^{p} b_j x_{ij} \qquad i=1,2,\ldots,N$$

donde x_{ij} es el valor de la j-èsima variable para la observaciòn i. En notaciòn matricial se puede escribir

$$\mu_{n\times 1} = X_{n\times p} b_{p\times 1}$$

donde X es la matriz del modelo y b el vector de paràmetros.

Esto completa la especificaciòn de la parte sistemàtica del modelo.

Las suposiciones de independencia y varianza constante para el error son importantes y es necesario confirmarlas. La estructura de la parte sistemàtica supone que las covariables que influyen en la media son conocidas y pueden medirse sin error.

Otra importante suposiciòn es que los errores tienen una distribuciòn Normal con varianza constante σ^2.

Entonces el modelo lineal clàsico se puede resumir de la siguiente manera:

Y es un vector de variables aleatorias independientes normalmente distribuidas con:

$$E(Y) = \mu \quad \text{donde} \quad \mu = Xb \qquad (4.1)$$

4.1 LA GENERALIZACION.

Para simplificar la transición a los modelos lineales generalizados se rearreglará ligeramente (4.1) para producir los siguientes tres enunciados:

a) El componente aleatorio: Y tiene una distribución normal independiente con varianza constante σ^2 y $E(Y) = \mu$.

b) El componente sistemàtico: Las covariables $x_1, x_2, \ldots, x_p$ producen un "predictor lineal" dado por:

$$\eta = \sum_{j=1}^{p} b_j x_j$$

c) La "liga" entre los componentes sistemàtico y aleatorio:

$$\mu = \eta$$

Esta forma de representar el modelo lineal clàsico introduce un nuevo símbolo η para el predictor lineal, y el inciso (c) indica que μ y η son de hecho idénticas. Si se escribe

$$\eta_i = g(\mu_i)$$

a $\eta(\cdot)$ se le llamará "función liga". De este modo, se puede decir que los modelos lineales clásicos tienen una distribución Normal en el inciso (a) y una función liga idéntica en el inciso (c).

Los modelos lineales generalizados permiten dos extensiones:

1o. La distribución en el inciso (a) puede ser de una familia exponecial.

2o. La función liga del inciso (c) puede ser cualquier función monótona diferenciable.

4.2 EL COMPONENTE ALEATORIO.

En los modelos lineales generalizados la forma que se puede usar para la función de densidad de probabilidad de una observación y es:

$$f_Y(y;\theta,\phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y,\phi)\} \qquad (4.2)$$

para algunas funciones a( ), b( ) y c( ). Si ϕ es conocida (4.2) es una familia exponencial con paraámetro canónico θ . Si ϕ es desconocida (4.2) puede o puede no ser una familia exponencial.

Con el propósito de ver esto con mayor claridad e introducir otros conceptos considérese primero la distribución normal.

$$f_y(y;\theta,\phi) = (2\pi\sigma^2)^{-1/2}\exp[-(y-\mu)^2/2\sigma^2$$

$$= \exp\{(y\mu-\mu^2/2)/\sigma^2 - 1/2[y^2/\sigma^2+\ln(2\pi\sigma^2)]\}$$

tal que $\theta=\mu$, $b(\theta)=\mu^2/2$, $\phi=\sigma^2$, $a(\phi)=\phi$,

y $c(y,\phi)=-1/2[y^2/\sigma^2+\ln(2\pi\sigma^2)]$

Sea $l(\theta,\phi;y) = \ln f_y(y;\theta,\phi)$ la función de log-verosimilitud considerada como una función de θ y ϕ dado y. La media y la varianza de y pueden entonces ser derivadas de las conocidas relaciones.

$$E(\partial l/\partial\theta) = 0 \qquad (4.3)$$

$$E(\partial^2 l/\partial\theta^2) + E(\partial l/\partial\theta)^2 = 0 \qquad (4.4)$$

De (4.2) se obtiene que $l = [y\theta - b(\theta)]/a(\phi) + c(y,\phi)$ entonces

$$\partial l/\partial\theta = [y - b'(\theta)]/a(\phi) \qquad (4.5)$$

y $$\partial^2 l/\partial\theta^2 = b''(\theta)/a(\phi) \qquad (4.6)$$

de (4.3) y (4.5) se tiene

$$0 = E(\partial l/\partial\theta) = [Y-b'(\theta)]/a(\phi)$$

por lo que $E(Y) = \mu = b'(\theta)$

Similarmente, de (4.4), (4.5) y (4.6) se tiene que

$$0 = - b''(\theta)/a(\phi) + Var(Y)/a^2(\phi)$$

de modo que $$Var(Y) = b''(\theta)a(\phi) \qquad (4.7)$$

por lo que la varianza de Y es producto de dos funciones: $b''(\theta)$ que depende solamente del parámetro canónico θ (y por lo tanto, de la media) y será llamada "funcion varianza", por otro lado, $a(\phi)$ es independiente de θ y solamente depende del parámetro de dispersión ϕ.

La funcion $a(\phi)$ es comunmente de la forma

$$a(\phi) = \phi/w$$

donde "w" es un peso previo que se supone conocido y es llamado el parámetro de dispersión (a veces se denota por τ^2). Por ejemplo, para un modelo normal en el que cada obsevación es la media de n registros independientes, se tiene

$$a(\phi) = \sigma^2/n$$

de modo que en este caso $w = n$.

En el cuadro 4.1 se presentan algunas de las más importantes distribuciones de este tipo.

Cuadro 4.1*

	Normal	Poisson	Binomial	Gamma	Inversa Gaussiana
Rango de y	$(-\infty,\infty)$	$0(1)\infty$	$\frac{0(1)n}{n}$	$(0,\infty)$	$(0,\infty)$
$a(\ )$	ϕ	1	$1/n$	ϕ	ϕ
$b(\ )$	$\frac{1}{2}\theta^2$	e^θ	$\ln(1+e^\theta)$	$-\ln(-\theta)$	$-(-2\theta)^{\frac{1}{2}}$
$c(\ )$	$-\frac{1}{2}\left(\frac{y^2}{\phi}+\ln(2\pi\phi)\right)$	$-\ln y!$	$\ln\left[\binom{n}{ny}\right]$	$(\phi-1)\ln(y/\phi)+\ln\phi-\ln\Gamma(\phi)$	$\frac{1}{\phi y}-\frac{1}{2}\ln(-\pi\phi y^3)$
$\mu = E(Y)$	θ	e^θ	$e^\theta/(1+e^\theta)$	$-1/\theta$	$(-2\theta)^{-\frac{1}{2}}$
Función varianza	1	μ	$\mu(1-\mu)$	μ^2	μ^3

Puesto que en los modelos logaritmicos lineales y los modelos logisticos que se aplican en este trabajo se consideran los componentes aleatorios Poisson y Binomial respectivamente, èstos se ven a continuaciòn.

a).- La distribucion Poisson como miembro de la familia exponencial.- Si "Y" es una variable aleatoria Poisson, su funciòn de densidad de probabilidades (f.d.p.) comunmente se escribe:

$$Pr[Y=y] = f_Y(y;\lambda) = \exp(-\lambda)\,\lambda^y/y! \qquad y=0,1,\ldots$$

* Tomado de McCullagh y Nelder (op. cit. 1983)

esta puede reescribirse de la forma (4.2) como sigue:

$$f_Y(y;\theta,\phi) = \exp\{[y\theta - b(\theta)]/a(\phi) + c(y,\phi)\}$$

donde $b(\theta) = \exp(\theta) = \lambda$, $a(\phi) = 1$, $c(y,\phi) = -\ln y!$

de (4.5) y (4.6) se obtiene:

$$E(y) = \lambda \equiv \mu = b'(\theta) = e^{\theta}$$

$$Var(y) = b''(\theta)a(\phi) = e^{\theta}(1) = e^{\theta}$$

Como se puede apreciar, la varianza y la función varianza en la Poisson son idênticas.

b).- La distribución Binominal como miembro de la familia exponencial .- Si "X" es una variable aleatoria Binominal, su f.d.p. comunmente se escribe:

$$Pr[X=x] = f_x(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x=0,1,\ldots,n$$

Mc. Cullagh y Nelder, hacen una pequeña modificación que no altera la esencia de esta distribución. Ellos consideran la variable aleatoria $Y = X/n$, entonces Y, tiene una f.d.p. binominal :

$$Pr[Y=y] = f_Y(y;n,p) = \binom{n}{ny} p^{ny}(1-p)^{n-ny} \qquad y=0,1/n,\ldots,n/n=1$$

Para presentar esta f.d.p. como miembro de la familia exponencial en la forma (4.2), se puede expresar:

$$Pr[Y=y] = f_Y(y;\theta,\phi) = \exp\{[(y\theta-\ln(1+e^{\theta}))/(1/n)]+\ln[\binom{n}{ny}]\}$$

donde $1/n = a(\phi)$, $\ln(1+e^{\theta}) = b(\theta)$, $\ln[\binom{n}{ny}] = c(y,\phi)$

y p se ha sustituido de tal modo que $p = e^{\theta}/(1+e^{\theta})$

por lo que $1 - p = 1/(1+e^{\theta})$

De (4.3) y (4.4) se tiene que :

$$E(y) = b'(\theta) = e^{\theta}/(1+e^{\theta}) = p \equiv \mu$$

y

$$Var(y) = b''(\theta)a(\phi) = e^{\theta}/(1+e^{\theta})^2(1/n)$$

$$= [e^{\theta}/(1+e^{\theta})][1/(1+e^{\theta})](1/n) = \mu(1-\mu)/n$$

por lo que la función varianza es: $\mu(1-\mu) \equiv p(1-p)$

4.3 LA FUNCION LIGA

La función liga relaciona al predictor lineal η con el valor esperado μ de un dato y. En los modelos lineales clàsicos, el predictor lineal y el valor esperado de una observaciòn son idênticas y η y μ pueden tomar cualquier valor en la recta real.

Cuando las observaciònes son conteos y la distribuciòn es Poisson, se tiene que $\mu > 0$, de modo que en este caso la liga idêntica es poco atractiva en parte porque η debe poder tomar valores negativos. Los modelos basados en independencia de probabilidades como en las tablas de contingencia, de manera natural consideran efectos multiplicativos y estos son expresados por una liga logaritmica $\eta = \log\mu$ con inversa $\mu = e^{\eta}$. Entonces los efectos aditivos en η son efectos multiplicatios en μ.

Para la distribuciòn binomial se tiene que $0 \leq \mu \leq 1$ por lo que la funciòn liga debe de satisfacer la condiciòn de mapear el intervalo (0,1) en toda la recta real. Hay varias funciones que satisfacen esto y que son ampliamente usadas como la "logit", la "probit" y la "log-log complementaria", sin embargo, para este caso aquí solamente se trata la "logit" o logística que se vio en el capitulo anterior y que es la siguiente:

$$\eta = \log[\mu/(1-\mu)]$$

Cada una de las distribuciones del cuadro 4.1 tiene una funciòn liga especial para las cuales existen estadisticas suficentes para el predictor lineal $\eta = \sum b_i x_i$. Estas son llamadas ligas canònicas y ocurren cuando

$$\theta = \eta$$

donde θ es el parâmetro canònico como se muestra en el cuadro 4.1. Las ligas canònicas para las distribuciones del cuadro 4.1 son:

Normal	$\eta = \mu$	
Piosson	$\eta = \ln \mu$	(4.8)
Binomial	$\eta = \ln[\mu(1-\mu)]$	(4.9)
Gamma	$\eta = \mu^{-1}$	
Inversa Gaussiana	$\eta = \mu^{-2}$	

Para las ligas canònicas, las estadisticas suficientes estàn dadas por $\sum y x_j$ $j = 1,2,\ldots,p$. Las ligas canònicas tienen propiedades estadisticas deseables particularmente en muestras pequeñas y facilitan el càlculo de los estimadores.

En los modelos logaritmicos lineales y logisticos lineales se pueden usar las ligas canònicas (4.8) y (4.9) respectivamente, por lo que se son modelos canònicos; ademàs, como se viò en el capitulo 3, los efectos sistemàticos en estos modelos son aditivos por lo que se facilita su manejo e interpretaciòn.

5. DEFINICION Y ANALISIS DE RESIDUALES EN MODELOS LINEALES GENERALIZADOS.

Mc. Cullagh y Nelder (op. cit. cap 2) señalan que los residuales pueden ser usados para explorar que tan adecuado es el ajuste de un modelo con respecto a la selección de la función varianza y los términos en el predictor lineal, además de que también pueden indicar la presencia de valores anómalos, los cuales requieren una mayor investigación.

También apuntan que para modelos lineales generalizados se requiere de una generalización de los residuales aplicable a todas las distribuciones que puedan reemplazar a la Normal y que puedan ser usados con los mismo propósitos que los residuales

Normales estandarizados en los modelos lineales normales. Definen tres tipos de residuales (residuales de Pearson, Devianza y Anscombe). En el presente trabajo se consideran además, los residuales de Haberman.

5.1 RESIDUALES DE PEARSON.

El residual de Pearson se define como:

$$r_p = (y - \hat{\mu}) / \hat{V}(y)$$

El numerador es el residual simple mostrado en el capítulo 2 y el denominador es el estimador de la desviación estándar de Y, de modo que si Y tuviera distribución Normal, éste sería el residual estandarizado visto en 2.1.1.

El nombre se debe a que para la distribución Poisson el residual de Pearson es la raiz cuadrada (con signo) de los componentes de la estadística X^2 de Pearson para bondad de ajuste, de modo que

$$\sum r_p^2 = X^2$$

Al analizar estos residuales en modelos para datos categóricos, Agresti (1984 p. 62) señala, que si el modelo es correcto, estos residuales son asintóticamente normales con media cero y con una

varianza promedio de los (r_p) igual al cociente del número de grados de libertad entre el número de celdas en una tabla de contingencia por lo que su varianza asintótica es menor que 1.0 . En modelos complejos puede ser bastante menor por lo que en esos casos no pueden considerarse aproximadamente procedentes de una N(0,1).

Ejemplos:

a) Haberman (1973) muestra que si x_{ij} es una observación con distribución Poisson o Multinomial correspondiente a una tabla de contingencia bidimensional, el residual de Pearson para el modelo logarítmico lineal de independencia entre las dos variables es

$$r_p = (x_{ij} - x_{i+}x_{+j}/x_{++})/(x_{i+}x_{+j}/x_{++})^{1/2}$$

b) Si y_i es una observación con distribución Binomial $B(n_i,p_i)$, el residual de Pearson esta dado por

$$r_p = (y_i - n_i\hat{p}_i)/[n_i\hat{p}_i(1-\hat{p}_i)]^{1/2}$$

5.2 RESIDUALES DE DEVIANZA

La devianza se define como la medida de discrepancia establecida

por el logaritmo de una razón de verosimilitudes .

Dadas N observaciones es posible ajustar modelos que contengan desde uno hasta N parámetros; al modelo con N parámetros (uno por observación) se le suele llamar modelo saturado, no tiene componente aleatorio (toda la variación de las y's es explicada por el componente sistemático) y no es informativo ya que no resume los datos sino que simplemente es una repetición de ellos. Sin embargo, el modelo saturado sirve de base para medir la discrepancia entre éste y un modelo intermedio con p<N parámetros.

Para definir esta discrepancia, es conveniente expresar la log-verosimilitud en términos del parámetro μ en lugar del parámetro canónico θ. Sea $l(\mu,\phi;y)$ la log-verosimilitud maximizada sobre b mas no sobre el parámetro de estorbo ϕ.

La máxima log-versomilitud factible en el modelo completo con N paráetros es $l(y,\phi;y)$ la cual es generalmente finita. La discrepancia de un ajuste es proporcional al doble de la diferencia entre la máxima log-verosimilitud posible (modelo saturado) y la máxima log-verosimilitud obtenida para el modelo bajo investigación. Si $\hat{\theta} = \theta(\hat{\mu})$ y $\tilde{\theta} = \theta(y)$ son los estimadores del parámetro canónico bajo los dos modelos, la discrepancia puede ser escrita como sigue:

$$\sum 2w_i[y_i(\tilde{\theta}_i-\hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)]/\phi = D(y;\hat{\mu})/\phi$$

donde $D(y;\hat{\mu})$ es conocida como la devianza del modelo bajo estudio y es función de los datos solamente.

Las formas de las devianzas para las distribuciones del cuadro 4.1 son las siguientes:

Normal	$\sum(y-\hat{\mu})^2$
Poisson	$2\{\sum[y\ln(y/\hat{\mu}) - (y-\hat{\mu})]\}$
Binomial	$2\sum\{[y\ln(y/\hat{\mu})] + (n-y)\ln[(n-y)/(n-\hat{\mu})]\}$
Gamma	$2\sum[-\ln(y/\hat{\mu}) + (y-\hat{\mu})/\hat{\mu}]$
Inversa Gaussiana	$(y-\hat{\mu})^2/(\hat{\mu}^2 y)$

donde todas las sumas son sobre $i = 1,2,\ldots, N$

Para la distribución Normal la devianza es la suma de cuadrados de los residuales simples, mientras que para la Poisson es la estadística G^2 vista en el capítulo 3. Nótese que los segundos términos de las expresiones de la devianza para la Poisson y la Gamma son usualmente iguales a cero.

La devianza es entonces la suma de N términos, si a cada uno de ellos se le llama d_i, el residual de Devianza se define como

$$r_D = \text{sgn}(y-\hat{\mu})(d_i)^{1/2}$$

Esta es una cantidad que crece o decrece con $y-\hat{\mu}$ y para la cual

$$\sum r_D^2 = D$$

Así, para la distribución Poisson se tiene

$$r_D = \text{sgn}(y - \hat{\mu})[2(y\ln(y/\hat{\mu}) - y + \hat{\mu}]^{1/2}$$

y para la distribución Binomial es

$$r_D = \text{sgn}(y - \hat{\mu})(2[y\ln(y/\hat{\mu}) + (n - y)\ln((n - y)/(n - \hat{\mu}))])^{1/2}$$

5.3 RESIDUALES DE ANSCOMBE

Una desventaja del residual de Pearson es que la distribución de r_p para distribuciones no normales es marcadamente sesgada por lo que puede fallar en cuanto a que no tendrá propiedades similares a las de los residual normales.

Anscombe propuso la definición de un residual usando una función A(y) en lugar de y, donde A(y) es escogida de modo que su distribución sea "tan Normal como sea posible". Se ha demostrado que para las funciones de verosimilitud en Modelos Lineales Generalizados, $A(\hat{\mu})$ está dada por

$$A(\hat{\mu}) = \int V^{-1/3}(\hat{\mu})\, d\hat{\mu}$$

por lo que el residual de Anscombe se basa en la diferencia $A(y)-A(\hat{\mu})$. Sin embargo, la transformación $A(\cdot)$ que normaliza la función de probalidad no estabiliza la varianza por lo que para definir el residual de Anscombe la diferencia anterior debe dividirse entre la raiz cuadrada de la varianza de $A(y)$, la cual se puede aproximar por $A'(\hat{\mu})[V(\hat{\mu})]^{1/2}$. Entonces el residual de Anscombe se define como

$$r_A = [A(y) - A(\hat{\mu})]/[V(A(y))]^{1/2} \cong [A(y) - A(\hat{\mu})]/A'(\hat{\mu})[V(\hat{\mu})]^{1/2}$$

y para la distribución Poisson se tiene

$$A(\hat{\mu}) = \int \hat{\mu}^{-1/3} d\hat{\mu} = 3/2\hat{\mu}^{2/3}$$

por lo que para esta distribución el residual de Anscombe se obtiene por la siguiente expresión

$$r_A = [3/2(Y^{2/3} - \hat{\mu}^{2/3})]/\hat{\mu}^{1/6} \qquad (5.1)$$

Cox y Snell (1968) presentan también el residual de Anscombe para la Poisson, para lo cual citan a Anscombe (1953) quien indica que si Y es una variable aleatoria Poisson, la esperanza de $Y^{2/3}$ es $(-1/6)^{2/3}$ por lo que esta versión del residual de Anscombe presenta una pequeña diferencia con el de la expresión (5.1) que es la que dan McCullagh y Nelder (op. cit.), ya que se tiene

$$r_A = 3/2[y^{2/3} - (\hat{\mu} - 1/6)^{2/3}]/\hat{\mu}^{1/6} \qquad (5.2)$$

La pequeña diferencia (el término - 1/6) entre las expresiones (5.1) y (5.2) es que en esta última se procura que la distribución sea insesgada. Sin embargo en la mayoría de los casos prácticos la diferencia que se obtiene al aplicar (5.1) y (5.2) es irrelevante.

Cox y Snell (op. cit.) tratan también el residual de Anscombe para la distribución Binomial definiéndolo de la siguiente manera

$$r_A = [h(y/n) - h(\hat{p}-1/6(1-2\hat{p})/n]/\hat{p}^{1/6}(1-\hat{p})^{1/6}/n^{1/2} \qquad (5.3)$$

$$\text{donde } h(u) = \int_0^u t^{-1/3}(1-t)^{-1/3}\,dt, \quad 0 \leq u \leq 1 \qquad (5.4)$$

Antes de ver la solución de h(u) es importante hacer dos observaciones:

a) Si p es pequeña este residual se reduce a (5.2), es decir, el definido para distribución Poisson.

b) La correción del sesgo $-1/6(1-2\hat{P})/n$ frecuentemente puede ser omitida.

Para el cálculo de h(u) Cox y Shell (op. cit.) elaboraron el cuadro 5.1 donde se encuentran tabulados los valores de la función beta incompleta $I_u(2/3,2/3)$, la cual es simètrica alrededor de u = 0.5; estos valores multiplicados por la función

beta b(m,n) valuada en m=n=2/3, es decir, por b(2/3,2/3) = 2.0533 da el valor de (5.4). Por ejemplo h(0.2) = (2.0533)(0.257) = 0.528, h(0.8)=(2.0533)(1-0.257) = 1.526.

CUADRO 5.1

Valores de la funcion Beta incompleta b(2/3,2/3)

x	0-000	0-001	0-002	0-003	0-004	0-005	0-006	0-007	0-008	0-009
0-00	0	0-007	0-012	0-015	0-018	0-021	0-024	0-027	0-029	0-032
0-01	0-034	0-036	0-038	0-040	0-043	0-045	0-046	0-048	0-050	0-052
0-02	0-054	0-056	0-058	0-059	0-061	0-063	0-064	0-066	0-068	0-069
0-03	0-071	0-072	0-074	0-075	0-077	0-079	0-080	0-082	0-083	0-084
0-04	0-086	0-087	0-089	0-090	0-092	0-093	0-094	0-096	0-097	0-098
0-05	0-100	0-101	0-102	0-104	0-105	0-106	0-108	0-109	0-110	0-112
0-06	0-113	0-114	0-115	0-117	0-118	0-119	0-120	0-122	0-123	0-124
0-07	0-125	0-126	0-128	0-129	0-130	0-131	0-132	0-134	0-135	0-136
0-08	0-137	0-138	0-139	0-141	0-142	0-143	0-144	0-145	0-146	0-147
0-09	0-149	0-150	0-151	0-152	0-153	0-154	0-155	0-156	0-157	0-158
0-10	0-160	0-161	0-162	0-163	0-164	0-165	0-166	0-167	0-168	0-169
0-11	0-170	0-171	0-172	0-173	0-174	0-176	0-177	0-178	0-179	0-180
0-12	0-181	0-182	0-183	0-184	0-185	0-186	0-187	0-188	0-189	0-190
0-13	0-191	0-192	0-193	0-194	0-195	0-196	0-197	0-198	0-199	0-200
0-14	0-201	0-202	0-203	0-204	0-205	0-206	0-207	0-208	0-209	0-210
0-15	0-211	0-212	0-213	0-214	0-214	0-215	0-216	0-217	0-218	0-219
0-16	0-220	0-221	0-222	0-223	0-224	0-225	0-226	0-227	0-228	0-229
0-17	0-230	0-231	0-232	0-232	0-233	0-234	0-235	0-236	0-237	0-238
0-18	0-239	0-240	0-241	0-242	0-243	0-244	0-244	0-245	0-246	0-247
0-19	0-248	0-249	0-250	0-251	0-252	0-253	0-254	0-254	0-255	0-256
0-20	0-257	0-258	0-259	0-260	0-261	0-262	0-262	0-263	0-264	0-265
0-21	0-266	0-267	0-268	0-269	0-270	0-270	0-271	0-272	0-273	0-274
0-22	0-275	0-276	0-277	0-277	0-278	0-279	0-280	0-281	0-282	0-283
0-23	0-284	0-284	0-285	0-286	0-287	0-288	0-289	0-290	0-290	0-291
0-24	0-292	0-293	0-294	0-295	0-296	0-296	0-297	0-298	0-299	0-300
0-25	0-301	0-302	0-302	0-303	0-304	0-305	0-306	0-307	0-308	0-308
0-26	0-309	0-310	0-311	0-312	0-313	0-313	0-314	0-315	0-316	0-317
0-27	0-318	0-318	0-319	0-320	0-321	0-322	0-323	0-323	0-324	0-325
0-28	0-326	0-327	0-328	0-328	0-329	0-330	0-331	0-332	0-333	0-333
0-29	0-334	0-335	0-336	0-337	0-338	0-338	0-339	0-340	0-341	0-342
0-30	0-342	0-343	0-344	0-345	0-346	0-347	0-347	0-348	0-349	0-350
0-31	0-351	0-351	0-352	0-353	0-354	0-355	0-355	0-356	0-357	0-358
0-32	0-359	0-360	0-360	0-361	0-362	0-363	0-364	0-364	0-365	0-366
0-33	0-367	0-368	0-368	0-369	0-370	0-371	0-372	0-372	0-373	0-374
0-34	0-375	0-376	0-376	0-377	0-378	0-379	0-380	0-380	0-381	0-382
0-35	0-383	0-384	0-384	0-385	0-386	0-387	0-388	0-388	0-389	0-390
0-36	0-391	0-392	0-392	0-393	0-394	0-395	0-396	0-396	0-397	0-398
0-37	0-399	0-400	0-400	0-401	0-402	0-403	0-403	0-404	0-405	0-406
0-38	0-407	0-407	0-408	0-409	0-410	0-411	0-411	0-412	0-413	0-414
0-39	0-414	0-415	0-416	0-417	0-418	0-418	0-419	0-420	0-421	0-422
0-40	0-422	0-423	0-424	0-425	0-425	0-426	0-427	0-428	0-429	0-429
0-41	0-430	0-431	0-432	0-433	0-433	0-434	0-435	0-436	0-436	0-437
0-42	0-438	0-439	0-440	0-440	0-441	0-442	0-443	0-443	0-444	0-445
0-43	0-446	0-447	0-447	0-448	0-449	0-450	0-450	0-451	0-452	0-453
0-44	0-454	0-454	0-455	0-456	0-457	0-457	0-458	0-459	0-460	0-461
0-45	0-461	0-462	0-463	0-464	0-464	0-465	0-466	0-467	0-468	0-468
0-46	0-469	0-470	0-471	0-471	0-472	0-473	0-474	0-474	0-475	0-476
0-47	0-477	0-478	0-478	0-479	0-480	0-481	0-481	0-482	0-483	0-484
0-48	0-485	0-485	0-486	0-487	0-488	0-488	0-489	0-490	0-491	0-491
0-49	0-492	0-493	0-494	0-495	0-495	0-496	0-497	0-498	0-498	0-499

Cox y Snell señalan que gráficas sobre papel de probabilidad sugieren que la transformación es muy efectiva incluso para valores tan pequeños como $p = 0.04$ y $n = 5$.

5.4 RESIDUALES DE HABERMAN

Aunque McCullagh y Nelder (op.cit.) no los presentan, estos residuales ocupan un lugar importante en el ajuste de modelos para datos categóricos. Haberman (1972) los propuso y les dió el nombre de "residuales ajustados" (aunque en este trabajo se les denomina residuales de Haberman o r_H). Haberman (1973) ilustra el uso de estos residuales tanto en modelos logarítmicos lineales como en modelos logísticos lineales.

Estos residuales tienen una analogía con los residuales studentizados en los modelos lineales normales, en el sentido de que se refieren al cociente de los residuales simples $y_i - \hat{y}_i$ entre su desviación estándar real, con la diferencia de que la varianza de los residuales en los modelos para variables categóricas es asintótica.

Una dificultad que presentan los residuales de Haberman es en su cálculo, ya que la fórmula para cada modelo es diferente y suele ser complicada. Haberman (1978 vo. 1) presenta algunos ejemplos y

fórmulas para su cálculo.

La principal ventaja de los residuales de Haberman es que tienen una distribución asintótica N(0,1). Para más detalles ver Haberman (1978, vol.1).

Considérese primero, una tabla de contingencia I x J, y sean $\{x_{ij}\}$ los valores observados, $1 \leq i \leq I$, $1 \leq j \leq J$. Para el modelo logarítmico lineal de independencia entre las dos variables, el residual de Pearson (también llamado residual estandarizado) visto en 5.1 es:

$$r_P = (x_{ij} - x_{i+}x_{+j}/x_{++}) / (x_{i+}x_{+j}/x_{++})^{1/2} \qquad (5.5)$$

donde x_{i+}, x_{+j} y x_{++} son los totales definidos en 3.1. Haberman (1972, op.cit.) muestra que el estimador de la varianza asintótica de (5.5) es

$$\hat{v}_{ij} = (1-x_{i+}/x_{++})(1-x_{+j}/x_{++})$$

por lo que, para el modelo logarítmico lineal de independencia entre dos variables categóricas, el residual de Haberman es:

$$r_H = r_P/\hat{v}_{ij}$$

Considérese ahora una tabla de contingencia de tres dimensiones y sean $\{x_{ijk}\}$ las frecuencias observadas en las celdas, $1 \leq i \leq I$, $1 \leq j \leq J$, $1 \leq j \leq J$ $1 \leq k \leq K$; y sean x_{i++}, x_{+j+}, x_{++k} x_{ij+}, x_{i+k} y x_{+jk} los

totales marginales vistos en 3.2 y $x_{+++} = n$ es la suma total de las x_{ijk}; Haberman (1978, op.cit.) presenta para este caso, las fórmulas de los residuales para los posibles modelos logarítmicos lineales jerárquicos. En el cuadro 5.2 se presentan los posibles modelos con notación abreviada y los residuales de Haberman para cada uno.

Haberman indica, al presentar este cuadro, que se omite el modelo saturado [ABC] porque los residuales son nulos y se omite también el modelo de "no interacción de tres factores " [AB] [AC] [BC] porque el cálculo de los r_H generalmente requiere de la invesión de la matriz $\hat{S}$ correspondiente a la versión del algoritmo Newton-Raphson empleado para encontrar los estimadores de máxima versomilitud. Sin embargo, señala que para resultados aproximados se podría usar en este caso el residual standarizado (r_p).

Haberman (1973, op. cit.) utiliza sus residuales en el modelo logístico lineal bosquejado en 3.3 y que se verá con más detalle en la aplicación 6.4. Como se recordará, en ese ejemplo N_j sujetos reciben una log dosis de veneno t_j, $i \leq j \leq r$; y n_{jk} es el número de sujetos que habiendo recibido la log dosis t_j tienen respuesta k=1,2 (sobrevivencia o muerte).

Al emplear en este ejemplo un modelo logístico lineal se tiene que

$$\log[P(1|j)/P(2|j)] = a + bt_j$$

donde P(k|j) es la probabilidad de que un sujeto tenga respuesta k dado que recibió la log dosis t_j.

CUADRO 5.2

Modelo [A] [B] [C]

$$r_M = [x_{ijk} - (x_{i\cdot\cdot}x_{\cdot j\cdot}x_{\cdot\cdot k}/n^2)]/[x_{i\cdot\cdot}x_{\cdot j\cdot}x_{\cdot\cdot k}/n^2)(1 - x_{i\cdot\cdot}x_{\cdot j\cdot}/n^2 - x_{i\cdot\cdot}x_{\cdot\cdot k}/n^2 - x_{\cdot j\cdot}x_{\cdot\cdot k}/n^2 + 2x_{i\cdot\cdot}x_{\cdot j\cdot}x_{\cdot\cdot k}/n^3)]^{1/2}$$

Modelo [AB] [C]

$$r_M = [x_{ijk} - (x_{ij\cdot}x_{\cdot\cdot k}/n)]/[x_{ij\cdot}x_{\cdot\cdot k}/n(1 - x_{ij\cdot}/n)(1 - x_{\cdot\cdot k}/n)]^{1/2}$$

Modelo [AB] [AC]

$$r_M = [x_{ijk} - (x_{ij\cdot}x_{i\cdot k}/x_{i\cdot\cdot})]/[x_{ij\cdot}x_{i\cdot k}/x_{i\cdot\cdot}(1 - x_{ij\cdot}/x_{i\cdot\cdot})(1 - x_{i\cdot k}/x_{i\cdot\cdot})]^{1/2}$$

Para este modelo simple (ya que solamente tiene una variable explicativa) Haberman (1973, op. cit) muestra que la fórmula para el cálculo de los r_{Mi} es

$$r_H = (n_{jk} - \hat{m}_{jk}) / [w_j[1 - (w_j/\sum_{j=1}^{r} w_j) - (w_j(t_j-\bar{t}) / \sum_{j=1}^{r} w_j(t_j-\bar{t})^2)]] \quad \ldots(5.6)$$

donde:

$w_j = \hat{m}_{j1}\hat{m}_{j2}/N_j$

$\bar{t} = (\sum_{j=1}^{r} w_j t_j) / \sum w_j$

y $\hat{m}_{jk}$ son los valores esperados estimados al ajustar el modelo.

Es importante señalar sin embargo, que para mdelos logísticos màs complejos, la expresiòn para el cálculo de los r_H no es mucho màs complicada que èsta.

Una ventaja de ver estos residuales desde el enfoque de los modelos lineales generalizados es que su representaciòn y cálculo en ocaciones se facilitan.

Por ejemplo, Pregibon (1982) muestra que (5.6) se puede expresar de la siguiente forma equivalente que puede ser fácilmente calculada usando el paquete de cómputo GLIM.

$$r_H = (n_{jk} - \hat{m}_{jk}) / \{(1 - o_j\hat{v}_j)[\hat{m}_{jk}(N_j - \hat{m}_{jk})]/N_j\}^{1/2} \quad (5.7)$$

donde:

v_j es la varianza del predictor lineal η_j

o_j son las últimas ponderaciones en el algoritmo iterativo en el ajuste del modelo.

Haberman (1976) propone e ilustra el uso de residuales generalizados ajustados (que aquí llamaríamos residuales generalizados de Haberman), los cuales pueden expresarse en forma general como sigue:

$$r_{H,a} = \sum a_i (n_i - m_i)/h(a)$$

donde:
$h(a)$ es el estimador de la varianza asintótica de $\sum a_i (n_i - \hat{m}_i)$ y los a_i son números reales adecuadamente seleccionados para los fines que se persigan.

La principal dificultad que presentan estos residuales generalizados de Haberman está en imaginar e interpretar su aplicación en situaciones concretas. El mismo Haberman indica que para ciertas selecciones de las a_i no se puede garantizar la distrubución asintótica Normal de dichos residuales.

5.5 ANALISIS DE RESIDUALES EN MODELOS LINEALES GENERALIZADOS

Como se ha señalado anteriormente, el análisis de residuales tiene una gran importancia en la verificación de las suposiciones

bàsicas en el ajuste de modelos estadísticos y en las posibles causas cuando se tienen evidencias de las fallas.

Las definiciones de residuales vistos en los parágrafos anteriores buscan que èstos tengan una dostribuciòn aproximadamente Normal de modo que para modelos lineales generalizados puedan ser usados con los mismos propòsitos que en los modelos lineales clàsicos. Es importante señalar sin embargo, que algunos autores no se limitan a procurar que la definiciòn de residuales tenga una distribuciòn aproximadamente Normal, sino cualquier distribuciòn conocida (ver Cox y Snell, 1968).

Al tratar modelos para datos categòricos Haberman (1978, vol.1) indica que dados los conocimientos actuales no es posible dar recomendaciones precisas sobre lo que indican los residuales, aunque sí hay algunos aspectos aceptados en general sobre todo en lo que se refiere a inspecciones visuales cuando, como en el presente trabajo, se espera que los residuales sean aproximadamente normales.

Para modelos lineales generalizados McCullagh y Nelder (1983, cap. 11) dan algunas recomendaciones generales, unas de ellas consisten en extensiones de las propuestas originalmente para modelos de regresiòn y anàlisis de varianza (revisadas en el capítulo 2 de este trabajo) y otras que son novedosas como son las concernientes a la funciòn liga, a la funciòn varianza y algunas sobre la escala de las variables explicativas. Subrayan

la importancia de los mètodos gràficos al indicar que una pràctica importante despuès del ajuste de un modelo, es graficar los residuales contra los valores estimados (èstos últimos posiblemente transformados) y que este tipo de gràficas pueden mostrar puntos aislados con grandes residuales o una curvatura general indicadora de que la escala de las covariables o la funciòn liga, no son satisfactorias; o bien, una tendencia creciente de los valores ajustados reveladora de una funciòn varianza no satisfactoria.

Otras recomendaciones generales muchas veces son difíciles de entender si no estan aplicadas a problemas concretos donde se requiere de alguna experiencia o cierto conocimiento del fenòmeno bajo estudio que no son resueltos por las recomendaciones generales.

En el siguiente capítulo se presentan ejemplos de ajuste de modelos logarítmicos lineales y logístico lineales, el comportamiento y anàlisis de diferentes tipos de residuales, y se detectan y resuelven algunos problemas de diferente índole con ayuda de los mètodos y recomendaciones revisadas.

4. APLICACIONES

Se presentan cinco aplicaciones de modelos para datos categòricos, el càlculo de diferentes tipos de residuales segùn se considere conveniente en cada caso y su anàlisis. Las dos primeras aplicaciones no presentan comportamientos extraños en los residuales una vez ajustado el modelo, sin embargo, en ellas se presentan algunas caracterìsticas del ajuste de modelos para datos categòricos que no se mencionaron antes y que son de interès. En la primera se muestra ademàs, còmo el anàlisis de los residuales de Pearson para esa aplicaciòn en particular, conduce a la misma selecciòn del "mejor" modelo que al aplicar el mètodo que consiste en particionar la estadìstica G^2. En la segunda, se calculan todos los tipos de residuales definidos y se hace una

comparación de los valores que toman.

En las otras tres aplicaciones el comportamiento de los residuales en un primer intento de ajuste de un modelo es indicador de alguna falla. Su análisis permite proponer algún tipo de solución, por lo que se ilustra la aplicación de métodos diversos para llegar a modelos que se ajustan adecuadamente y donde el nuevo comportamiento de los residuales resulta aceptable. Además se presentan algunas características de los modelos que no fueron discutidas en los capítulos anteriores.

6.1 SELECCION DE UN MODELO LOGARITMICO LINEAL PARA DATOS DE UNA TABLA DE CONTINGENCIA DE CUATRO DIMENSIONES.

Como se señaló en 3.2, para una tabla de cuatro dimensiones existen 113 modelos logarítmicos lineales jerárquicos, todos ellos conteniendo los términos principales, es decir, el caso de independencia total entre las cuatro variables

$$\log m_{ijkl} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} \qquad (6.1)$$

siempre será un caso particular de esos 113 modelos.

Sería realmente fastidioso obtener estadísticas de bondad de ajuste para todos los posibles modelos, pero si éstas se

calcularan, seria también difícil seleccionar uno entre todos aquellos cuyo ajuste haya sido adecuado.

Por supuesto que esto casi nunca se hace, lo común es que se tengan definidos previamente, al menos de una manera aproximada, los modelos que se desea explorar.

Por otra parte, casi todos los autores consultados consideran que lo más frecuente en la práctica es preferir un modelo simple que uno complicado, aunque el modelo complicado tenga un ajuste ligeramente mejor que el modelo simple (Esto resulta obvio si se piensa en el caso extremo de que el modelo saturado, siempre tendrá un ajuste perfecto). De cualquier manera, exite cierto conflicto entre bondad de ajuste y simplicidad del modelo.

Fienberg (1979) presenta un método basado en la partición de la estadística de bondad de ajuste G^2 en varias partes aditivas, las cuales tienen distribución asintótica Ji-cuadrada. Este se puede usar como criterio para escoger el "mejor" modelo dentro de un grupo de modelos para los que se tiene un ajuste adecuado.

Para llevar a cabo este método es necesario escoger un grupo de modelos logararítmicos lineales jerárquico "anidados", es decir, un grupo de modelos tales que, ordenados, los términos del primer modelo estén presentes en el segundo, de mayor complejidad que el primero; y los términos del segundo estén presentes en el tercero, de mayor complejidad, etc.

Sean A, B, C y D las variables que definen una tabla de cuatro dimensiones; un ejemplo de grupo anidado de modelos logaritmico lineales expresados en la forma abreviada es:

(a) [A] [B] [C] [D]

(b) [A] [B] [CD]

(c) [B] [AC] [CD]

(d) [AC] [BC] [CD]

(e) [BC] [ACD]

(f) [ACD] [BCD]

(g) [ACD] [BCD] [ABD]

Otro grupo podria ser definido añadiendo en otro orden los tèrminos de un modelo a otro pero respetando la caracteristica de que deben ser anidados, por ejemplo si en (b) se cambia [CD] por [AC] se tendrà otro grupo. Para una tabla de tres dimensiones podrian incluso explorarse todos los grupos anidades y seleccionar el màs convincente, para cuatro dimensiones, sin embargo, el nûmero posible de grupos es demasiado grande.

Volviendo a los 7 modelos (a),(b),...,(g) en el ejemplo, si denotamos la estadistica G^2 como $G^2(a)$, para el modelo (a); $G^2(b)$ para el modelo (b); ...; G^2 (g) para el (g) se tiene que

$$G^2(a) \geq G^2(b) \geq \ldots \geq G^2(g) \qquad (6.2)$$

Una razòn por la que no se considera aquí la estadistica X^2 de Pearson es que (6.2) no es necesariamente cierto para cualquier

grupo anidado si se reemplaza G^2 por X^2.

Otro resutlado importante es que si se toman del grupo dos modelos digamos (a) y (c), entonces la estadistica de razòn de versomilitud

$$2 \sum (\text{observados}) \log[(\text{Esparados})_{(a)}/(\text{Esperados})_{(a)}] \qquad (6.3)$$

puede ser usada para probar si la diferencia entre los modelos se debe a una simple variaciòn aleatoria dado que los valores esperados satisfacen el modelo (c). Esta prueba estadistica condicional tienen una distribuciòn asintòtica Ji-cuadrada (bajo la hipòtesis nula) con grados de libertad (g.l.) igual a la diferencia de g.l. de los dos modelos.

Fienberg cita a Goodman (1969) para afirmar que, dada la forma multiplicativa de la estimaciòn de los valores esperados en modelos logaritmicos lineales jeràrquicos, el valor de (6.3) es el mismo si se cambia (observados) por $(\text{esperados})_{(a)}$.

Nòtese entonces, que en nuestro grupo anidado $G^2(a) - G^2(b)$, $G_2(b) - G^2(c),\ldots,G^2(g) - G^2(f)$; son estadisticas de la forma (6.3) y pueden ser usadas para probar diferencias entre cada par de modelos. Entonces la estadistica de razòn de verosimilitud de bondad de ajuste para el modelo de completa independencia se puede expresar como sigue:

$$G^2(a)=[G^2(a)-G^2(b)]+[G^2(b)-G^2(c)]+\ldots+[G^2(f)-G^2(g)]+G^2(g) \qquad (6.4)$$

Un procedimiento para la selección de un "mejor" modelo consiste en recorrer de derecha a izquierda los términos compuestos por diferencias en la expresión (6.4) y observar, para cada término, dos valores: el del el de G^2 para el modelo más complejo dentro del término y el del término. La regla sería : detenerse en este recorrido de derecha a izquierda cuando alguno de los valores sea significativo (considerando la distribución Ji-cuadrada apropiada para cada uno) y considerar como mejor modelo el del paso previo.

Para ilustrar este procedimiento de selección considérense los datos del cuadro 6.1, los cuales han sido analizados por varios autores, entre ellos Bishop, Fienberg y Holland (1975) y Fienberg (op.cit.). Los datos proceden de una nuestra de 1008 personas entrevistadas con objeto de comparar dos clases de detergente: un producto nuevo "R" y otro de uso común "S". Además de la preferencia por alguno de estos dos productos, se les preguntó si habían utilizado antes el "S", el tipo de agua (suave, regular o dura) que usaron y la temperatura de ésta.

Se tienen entonces, cuatro variables:

A: tipo de agua (S, M, D) ó (1, 2, 3)

B: uso previo de S (si, no) ó (1,2)

C: temperatura (caliente, fría) ó (1,2)

D: producto que se prefiere (R, S) ó (1,2)

De modo que cada celda se puede representar por un arreglo (i,j,k,l); i = 1, 2, 3; j = 1, 2; k = 1, 2; l = 1, 2.

Al tratar estos datos Fienberg escoge un grupo de 6 modelos aninados, obtiene la estadística G^2 para cada uno y luego la particiona para seleccionar el mejor modelo. Para complementar a Fienberg, en esta aplicación se presentan además los valores estimados y los residuales de Pearson para cada modelo.

CUADRO 6.1

		Usaron antes el producto S			
		si		no	
Tipo de agua	Producto que prefiere	Temperatura alta	Temperatura baja	Temperatura alta	Temperatura baja
Suave	R	19	57	29	63
	S	29	49	27	53
Mediana	R	23	47	33	66
	S	47	55	23	50
Dura	R	24	37	42	68
	S	43	52	30	42

El propósito principal de presentar los residuales es para observar si al analizarlos existe alguna relación aparente entre el procedimiento para la selección del mejor modelo por la

tècnica de la Ji-cuadrada particionada y el comportamiento de los residuales.

Para escoger el grupo de modelos logarítmicos lineales anidados Fienberg razona del siguiente modo (los modelos se presentan en notaciòn abreviada): selecciona primero el modelo simple de independencia total entre las cuatro variables.

(a) [a] [b] [c] [d]

Luego dice que sería natural que la preferencia estuviera relacionada con el uso previo, por lo que incluye el tèrmino que representa esta relaciòn:

(b) [A] [C] [BD]

El siguiente tèrmino es escogido en funciòn de que algunos detergentes son elaborados para agua caliente o fria por lo que podría existir una relaciòn entre C y D

(c) [A] [BD] [CD]

Luego introduce la relaciòn entre tipo de agua y temperatura ya que algunos fabricantes de detergentes hacen recomendaciones al respecto.

(d) [AC] [BD] [CD]

y finalmente incluye dos modelos que incorporan relaciones de segundo orden

(e) [AC] [BCD]

(f) [ABC] [BCD]

En el cuadro 6.2 se presentan los valores esperados estimados, las estadísticas X^2 y G^2 y los grados de libertad (g.l.) para los modelos (a), (b), (c), (d), (e) y (f) y en el cuadro 6.3 se presentan los valores observados y los los residuales de Pearson para cada modelo.

Antes de analizar el comportamiento de los residuales, se procede a seleccionar el modelo por la tècnica de la Ji-cuadrada particionada.

Sea $G^2(a)$ la estadística de razòn de verosimilitud y supngase que se particiona en componentes aditivos como en (6.4) pero considerando el grupo de modelos anidados que se està analizando para los datos del cuadro 6.1

De acuerdo con el cuadro 6.2 $G^2(a) = 42.93$. En el cuadro 6.4 se muestran los resultados de esta estadística particionada por diferencias y su valor para cada modelo, así como los grados de libertad correspondientes.

CUADRO 6.2

Celda (A,B,C,D)	Valor. obser.	Estimación de valores esperados (a)	(b)	(c)	(d)	(e)	(f)
(1,1,1,1)	19	28.76	24.51	22.40	19.52	18.60	17.12
(2,1,1,1)	23	30.35	25.86	23.64	23.65	22.54	24.97
(3,1,1,1)	24	29.82	25.41	23.23	26.09	24.86	23.90
(1,2,1,1)	29	31.38	35.64	32.58	28.39	29.31	31.65
(2,2,1,1)	33	33.12	37.60	34.38	34.40	35.51	31.65
(3,2,1,1)	42	32.54	36.95	33.78	37.94	39.18	40.70
(1,1,2,1)	57	49.80	42.44	44.54	47.85	48.99	50.32
(2,1,2,1)	47	52.55	44.78	47.00	46.99	48.10	48.42
(3,1,2,1)	37	51.64	44.00	46.18	42.89	43.91	42.25
(1,2,2,1)	63	54.35	61.71	64.77	69.58	68.44	66.82
(2,2,2,1)	66	57.35	65.12	68.35	68.32	67.21	66.82
(3,2,2,1)	68	56.35	63.98	67.15	62.37	61.35	63.36
(1,1,1,2)	29	28.31	22.56	35.40	30.85	33.54	30.88
(2,1,1,2)	47	29.87	34.36	37.35	37.37	40.63	45.03
(3,1,1,2)	43	29.35	33.76	36.70	41.33	44.83	43.10
(1,2,1,2)	27	30.89	26.64	28.96	25.24	22.55	24.35
(2,2,1,2)	23	32.60	28.11	30.56	30.58	27.32	24.35
(3,2,1,2)	30	32.03	27.62	30.03	33.73	30.14	31.30
(1,1,2,2)	49	49.02	56.38	53.54	57.52	54.20	55.68
(2,1,2,2)	55	51.72	59.49	56.50	56.48	53.22	53.58
(3,1,2,2)	52	50.82	58.46	55.51	51.56	48.58	46.75
(1,2,2,2)	53	53.49	46.13	43.81	47.06	50.38	49.18
(2,2,2,2)	50	56.45	48.68	46.22	46.21	49.47	49.18
(3,2,2,2)	42	55.46	47.83	45.42	42.18	45.16	46.64
X^2		43.88	23.13	18.33	11.92	8.44	5.65
G^2		42.93	22.35	17.99	11.89	8.41	5.66
g.l.		18	17	16	14	12	8

CUADRO 6.3

Celda (A,B,C,D)	Valor. obser.	Residuales de Pearson para cada modelo (a)	(b)	(c)	(d)	(e)	(f)
(1,1,1,1)	19	-1.82	-1.11	-0.72	-0.12	-0.09	0.45
(2,1,1,1)	23	-1.33	-0.56	-0.13	-0.13	-0.10	-0.39
(3,1,1,1)	24	-1.06	-0.28	0.16	-0.41	-0.17	0.02
(1,2,1,1)	29	-0.43	-1.11	-0.63	0.11	-00.06	0.47
(2,2,1,1)	33	-0.02	-0.75	-0.23	-0.24	-0.42	-0.24
(3,2,1,1)	42	1.66	0.83	1.41	0.66	0.45	0.20
(1,1,2,1)	57	1.02	2.23	1.87	1.32	1.14	0.94
(2,1,2,1)	47	-0.77	0.33	-0.00	-0.00	-0.16	-0.20
(3,1,2,1)	37	-2.04	-1.05	-1.35	-0.90	-1.04	-0.81
(1,2,2,1)	63	1.17	0.16	-0.22	-0.79	-0.66	-0.47
(2,2,2,1)	66	1.14	0.11	-0.28	-0.28	-0.15	-0.10
(3,2,2,1)	68	1.55	0.50	0.10	0.71	0.85	0.58
(1,1,1,2)	29	0.13	-0.62	-1.07	-0.33	-0.78	-0.34
(2,1,1,2)	47	3.13	2.16	1.58	1.57	1.00	0.29
(3,1,1,2)	43	2.52	1.59	1.04	0.28	-0.27	-0.01
(1,2,1,2)	27	-0.70	-0.07	-0.36	0.35	0.94	0.54
(2,2,1,2)	23	-1.68	-0.96	-1.37	-1.37	0.83	-0.27
(3,2,1,2)	30	-0.36	0.45	-0.01	-0.64	-0.02	-0.23
(1,1,2,2)	49	-0.00	-0.98	-0.62	-1.12	-0.71	-0.89
(2,1,2,2)	55	0.45	-0.58	-0.20	-0.20	0.24	0.19
(3,1,2,2)	52	0.16	-0.84	-0.47	-0.06	0.49	0.77
(1,2,2,2)	53	-0.07	1.01	1.39	0.87	0.37	0.54
(2,2,2,2)	50	-0.86	0.19	0.55	0.56	-0.08	0.12
(3,2,2,2)	42	-1.81	0.84	-0.51	-0.03	-0.47	-0.68

Existen ciertas diferencias entre estos valores calculados con el paquete GLIM y los obtenidos por Fienberg, sin embargo èstos son a nivel de centèsimas. El asterisco indica que el valor rebasa el valor de la Distribuciòn Ji-cuadrada correspondiente tomando un nivel de significancia del 0.05.

Como se puede observar en los valores G^2 en el cuadro 6.4, los modelos (b) (c) (d) (e) y (f) se ajustan adecuadamente y sòlo es

rechazado el modelo (a) de completa independencia. Y como es de esperarse, mientras el modelo es más complejo, el ajuste es mejor.

CUADRO 6.4

Componente debido a	G^2	g.l.
Modelo (a)	42.93*	18
Diferencia entre modelos (b) y (a)	20.58*	1
Modelo (b)	22.35	17
Diferencia entre modelos (c) y (b)	4.36*	1
Modelo (c)	17.99	16
Diferencia entre modelos (d) y (c)	6.1*	2
Modelo (d)	11.89	14
Diferencia entre modelos (e) y (d)	3.48	2
Modelo (e)	8.41	12
Diferencia entre modelo (f) y (e)	2.75	2
Modelo (f)	5.66	8

Si se emplea la técnica de la G^2 particionada, en el cuadro 6.4 se puede ver que el primer valor que resulta significativo (recorriendo de abajo hacia arriba) es el correspondiente a la

diferencia entre los modelos (d) y (c). Esto sugiere que el modelo (d) es el mejor modelo del grupo anidado.

En este caso no ha sido necesario recurrir al anàlisis de residuales para encontrar el modelo màs adecuado; sin embargo es ilustrativo analizar el comportamiento de èstos en el proceso seguido.

En la figura 6.1 se muestran las gràficas de los residuales de Pearson, que aparacen en el cuadro 6.3, contra los valores estimados para cada uno de los modelos probados.

Como se muestra en el cuadro 6.4 solamente el modelo (a) es rechazado; en la gràfica correspondiente se observa que los residuales estàn bastante dispersos pero sin ninguna tendencia clara y su comportamiento no corresponde a ninguno de los patrones de la figura 2.1 del capitulo 2. Sin embargo, es claro que lo que debe hacerse es considerar otros tèrminos en el modelo.

Los modelos (b), (c), (d) y (f) son aceptados como modelos que se ajustan adecuadamente a los datos al considerar las estadisticas X^2 y G^2 y contrastarlas con los valores de la distribuciòn Ji-cuadrada correspondiente en cada caso.

En la gràfica correspondiente al modelo (b), los residuales positivos se observan bastante dispersos por lo que se advierte un sesgo pronunciado en su distribuciòn. Esto de acuerdo a los comentarios del capitulo dos, seria suficiente para no

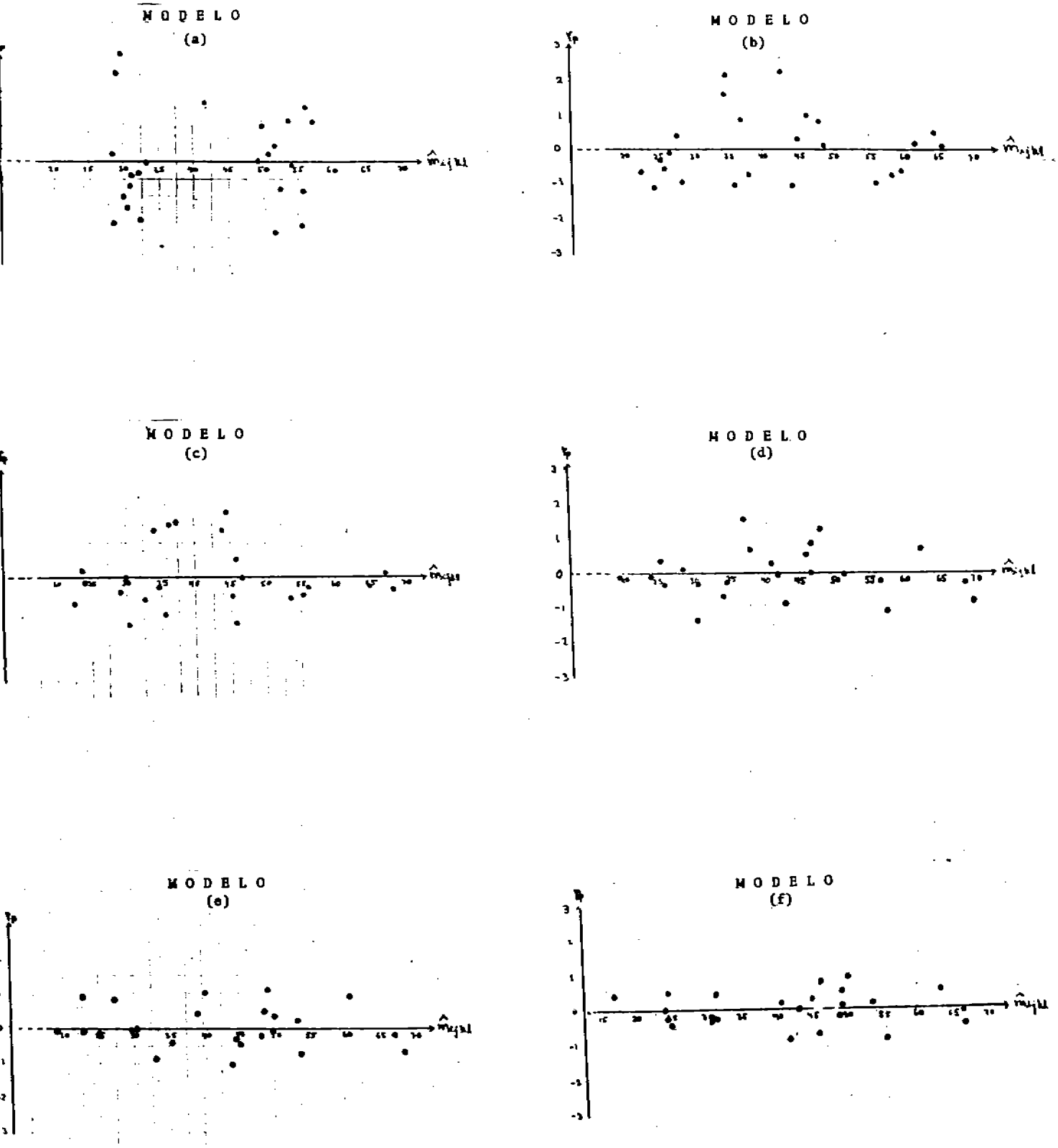


Figura 6.1

considerarlo como el mejor de los modelos del (b) al (f).

En los modelos (c), (d), (e) y (f) la gràfica de los residuales parece comportarse como la de la figura 2.1 (a) del capitulo 2, es decir, los residuales dentro de una banda horizontal con fronteras alrededor del cero. Esta banda, como es natural, es màs estrecha cuando el modelo es màs complejo. Esto no nos lleva a ninguna conclusiòn sobre cual es el mejor modelo entre el (c), (d), (e) y (f).

Para utilizar los residuales de Pearson para decidir entre alguno de estos modelos, una sugerencia podria ser escoger aquel para el cual los residuales tengan una distribuciòn màs parecida a una N(0,1).

Para ver esto, a continuaciòn se realizan pruebas Ji-cuadrada de bondad de ajuste sobre los residuales para cada uno de estos modelos.

Hoel (1966 p. 247) al considerar las limitaciones de esta prueba señala que tanto la experiencia como investigaciones teòricas indican que es usualmente satisfactoria cuando los valores esperados en cada intervalo, asi como el nûmero k de intervalos es mayor o igual que 5, pero que si alguna de estas dos condiciones no se puede mantener es mejor tener k<5 intervalos, conservando el nûmero esperado de observaciones en cada uno, mayor o igual que 5.

Como en este caso se tienen 24 residuales para cada modelo se consideran solamente 4 intervalos: $I_1=(-\infty,-0.8)$, $I_2=(-0.8,0)$, $I_3=(0,0.8)$, $I_4=(0.8,\infty)$.

Bajo la suposición de que los residuales provienen de una N(0,1), el valor esperado para los intervalos I_1 e I_4 es 5.08 y para los intervalos I_2 e I_3 es 6.92

La prueba para cada modelo consiste en contar cuantos residuales están dentro de cada intervalo y calcular la estadística X^2 vista en 3.2.2

Los valores de la estadística X^2 para los residuales en cada uno de los modelos considerados son:

Modelo	Valores observados (I_1, I_2, I_3, I_4)	X^2
(c)	(3, 13, 3, 5)	8.41
(d)	(3, 12, 6, 3)	5.55
(e)	(1, 14, 4, 5)	11.75
(f)	(2, 10, 11, 1)	8.92

Como no se estiman parámetros se tienen tres grados de libertad. El valor en tablas de una variable Ji-cuadrada con tres grados de libertad y considerando un nivel de significancia de 0.05 es 7.81, por lo que para el único modelo que no se rechaza la hipótesis de que los residuales de Pearson provengan de una

N(0,1) es el modelo (d) y por tanto, es el que se considera màs adecuado.

Es interesante observar cómo, en este caso, el anàlisis de residuales nos conduce a seleccionar como el modelo màs adecuado al mismo modelo que se consideró como el mejor, empleando el mètodo de particionar la estadistica G^2.

En las figuras 6.2, 6.3, 6.4, 6.5, 6.6 y 6.7 se presentan en "papel de graficaciòn normal" las gràficas de los residuales de Pearson correspondientes a los modelos (a), (b), (c), (d), (e) y (f). Para elaborarlas se sigue el procedimiento visto en 2.1.1 y se toma la recomendaciòn de Tukey (1962) de asignar el k-èsimo menor residual a la coordenada correspondiente a la probabilidad $(3k-1)/(3n+1) = (3k-1)/73$. Junto a cada punto se marca entre parèntesis el lugar que ocupa el residual contando de arriba hacia abajo en el cuadro 6.3. Cuando los residuales provienen de una distribuciòn N(0,1) se espera que estèn pegados a la raya continua que representa a la funciòn de distribuciòn N(0,1).
En este ejemplo solamente para el modelo (d) se aceptò la hipòtesis de que los r_p provienen de una N(0,1) por lo que su gràfica sirve como referencia para analizar las otras; se puede apreciar que en ella los puntos estan màs pegados a la raya continua que en las otras gràficas.

Es especialmente interesante observar las gràficas de los modelos (d), (e) y (f). Conviene recordar que en los modelos (e) y (f) se

MODELO (a)

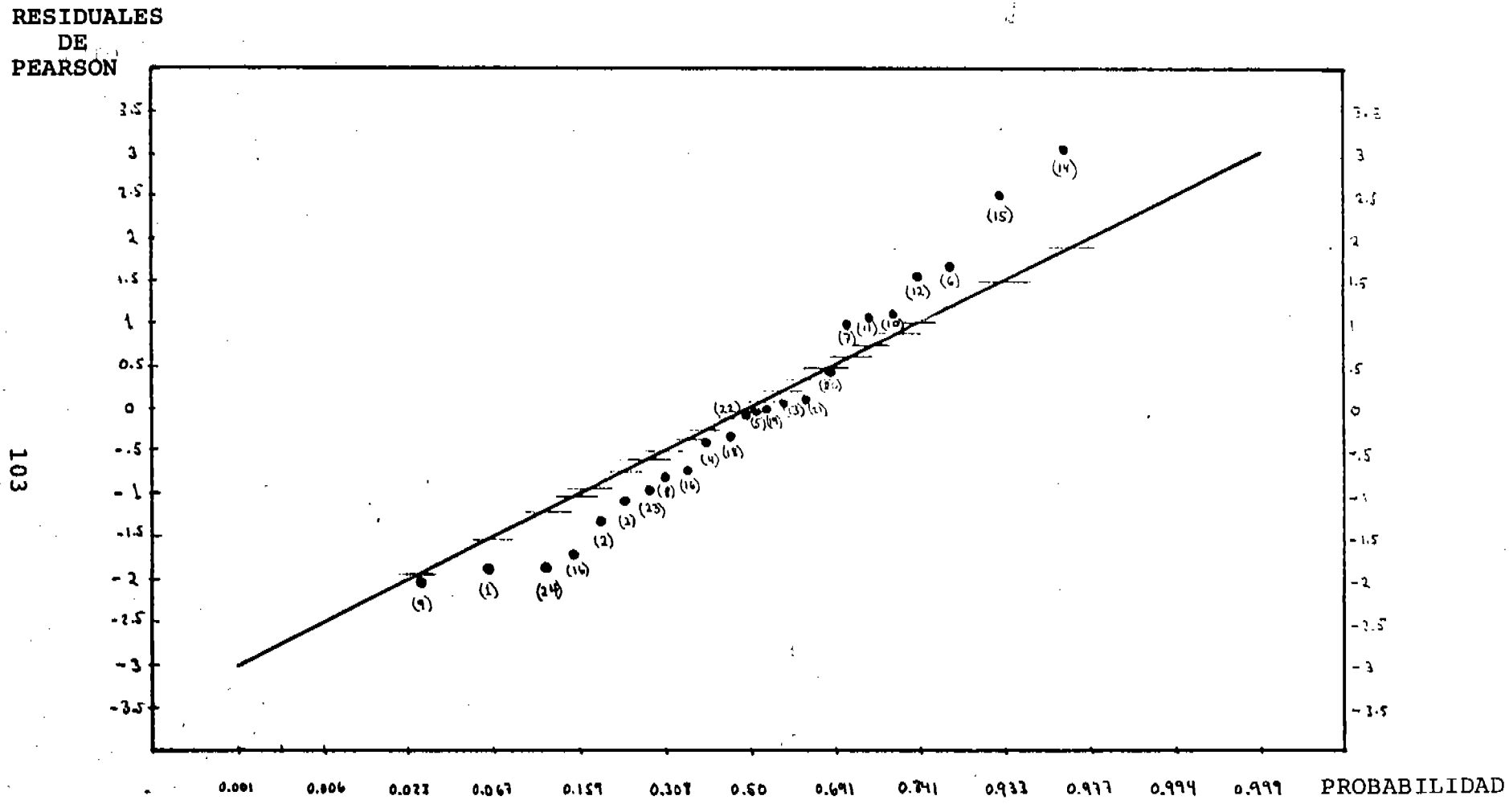


Figura 6.2

MODELO (b)

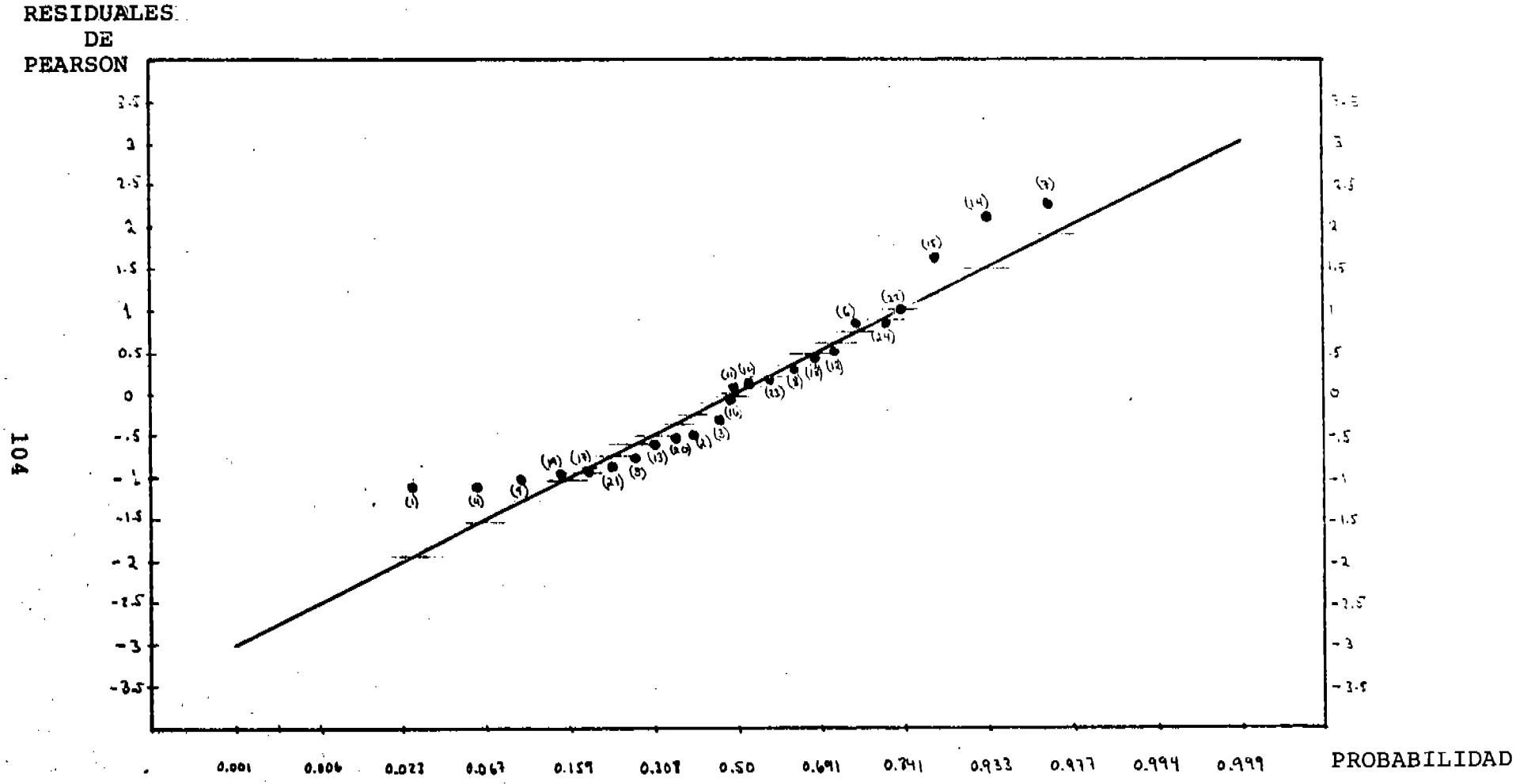


Figura 6.3

MODELO (c)

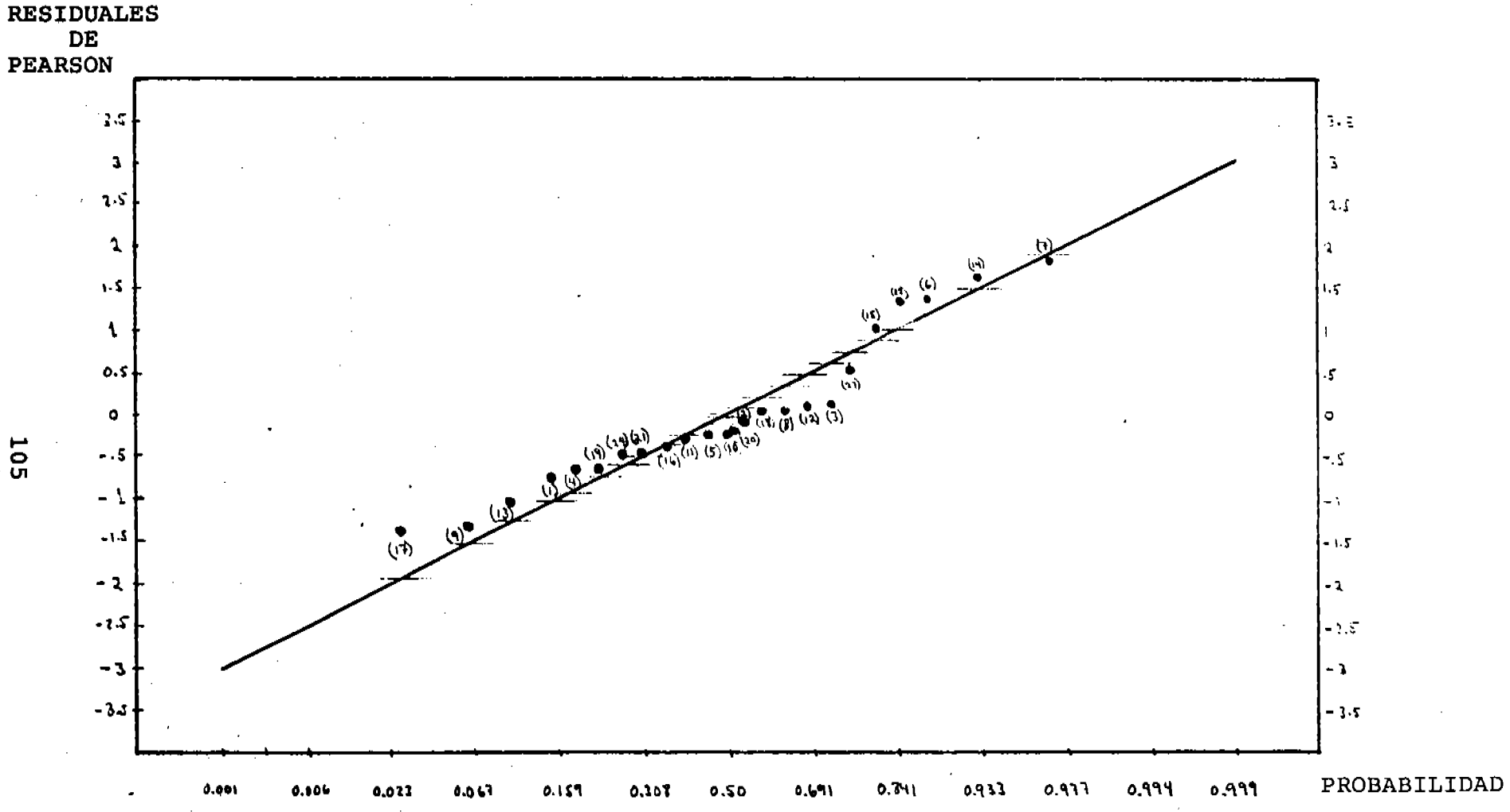


Figura 6.4

MODELO (d)

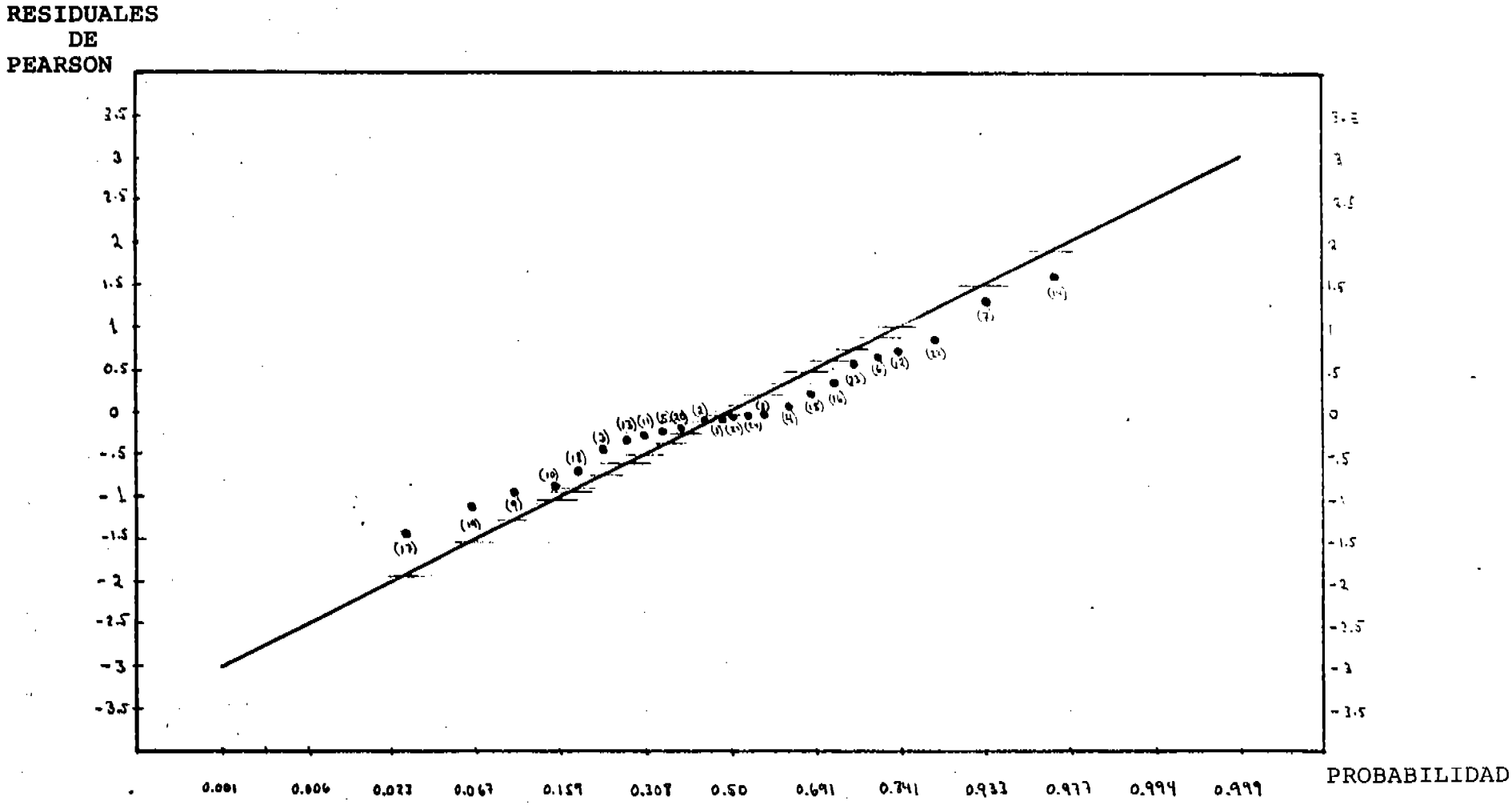


Figura 6.5

MODELO (e)

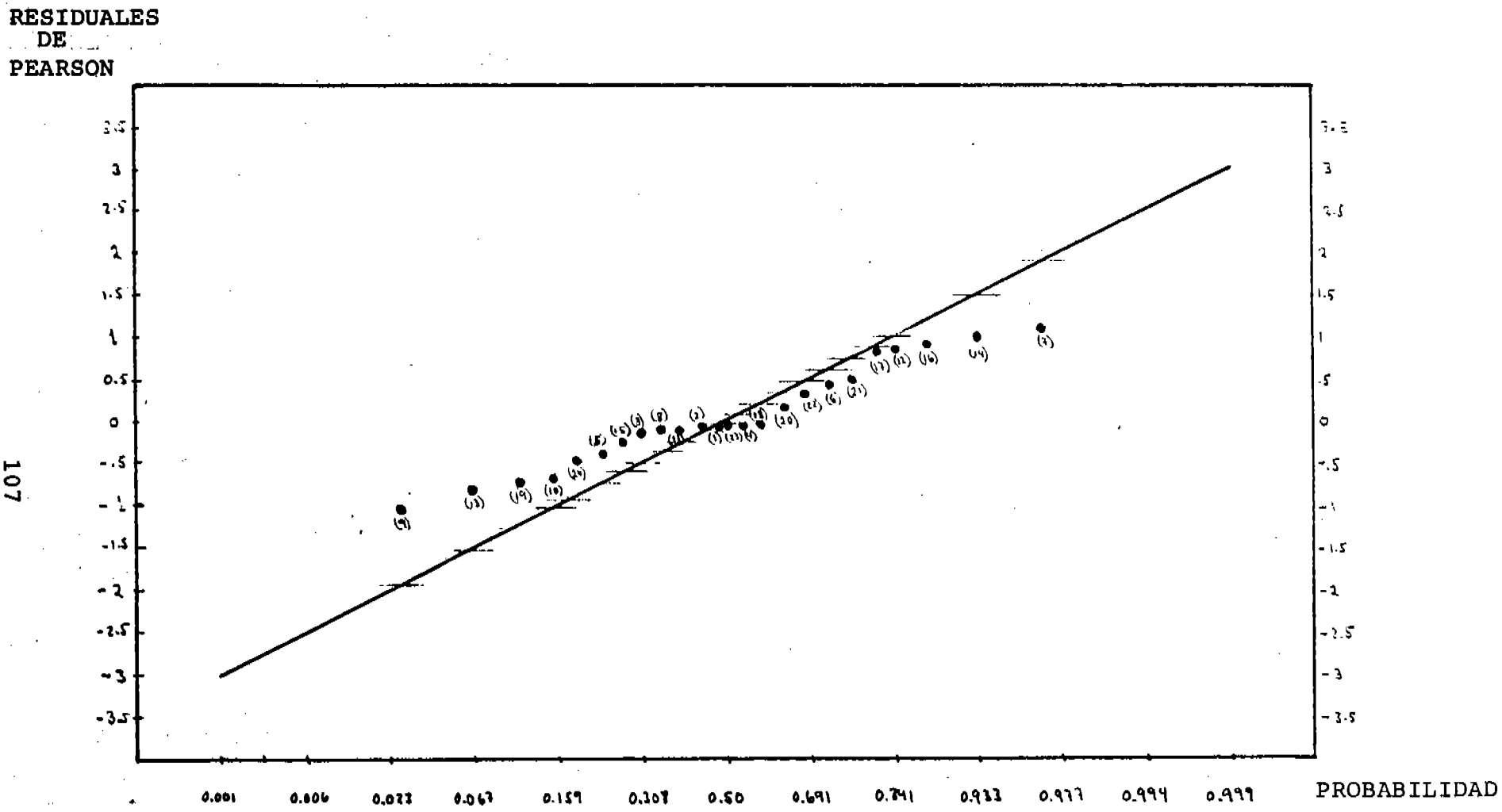


Figura 6.6

MODELO (f)

RESIDUALES
DE
PEARSON

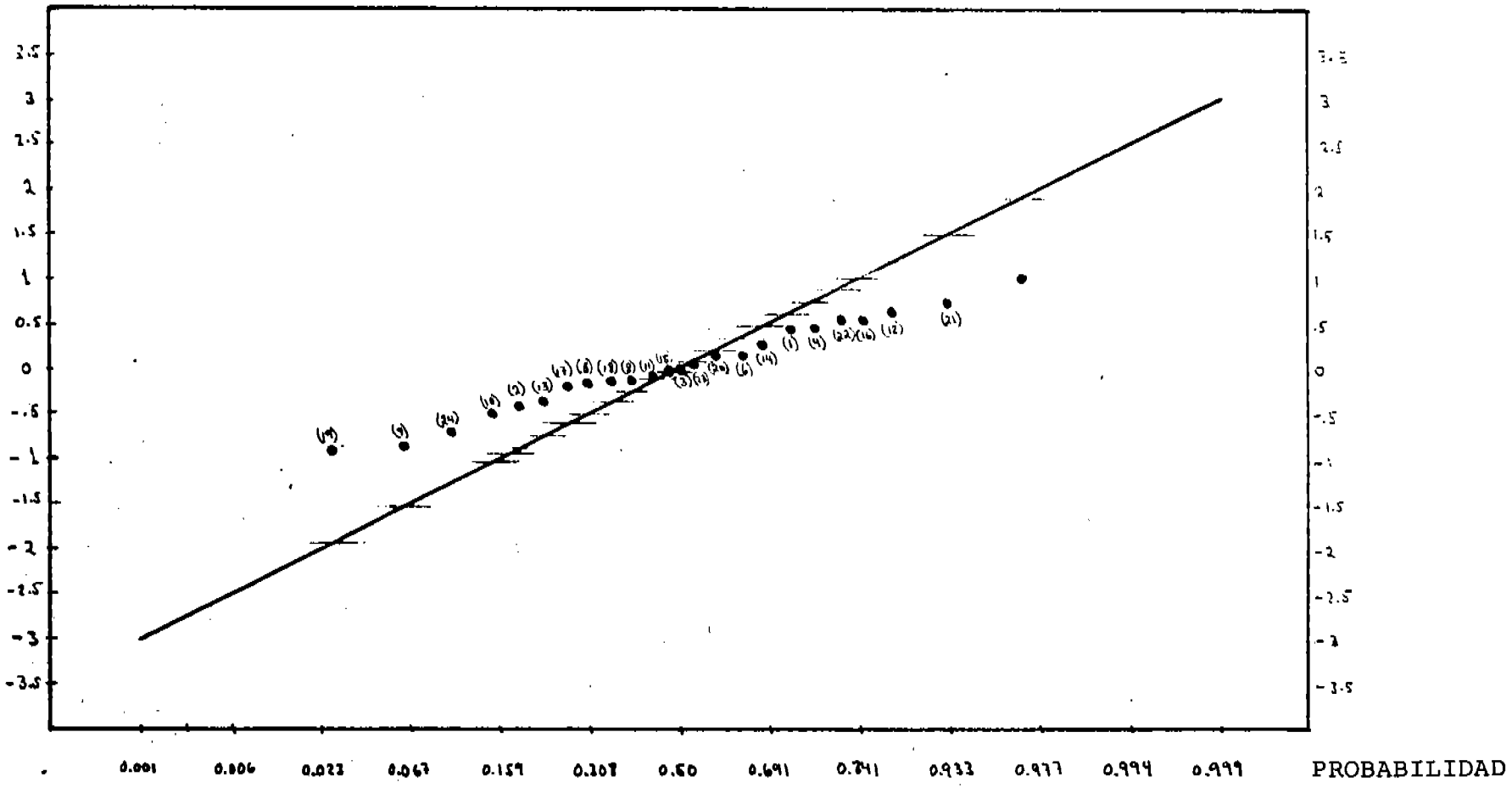


Figura 6.7

logra un mejor ajuste al observar la estadística G^2, sin embargo, se eligió el modelo (d) como el más adecuado; en las gráficas correspondientes se observa que los residuales se alejan más de la raya continua mientras el modelo es más complejo a partir del (d) y la tendencia es de acercarse a una raya horizontal a la altura del cero que es la que se obtendría al ajustar el modelo saturado en que todos los residuales serían nulos.

6.2 AJUSTE DE UN MODELO LOGISTICO LINEAL EN UNA TABLA DE CONTINGENCIA 2 x 2 x 2

Haberman (1978, vol. 1 pp. 163-181) presenta unos datos de víctimas de homicidios cometidos en Estados Unidos en 1970. Fueron excluidos algunos datos, pues sólo se tomaron aquellos que pudieran clasificarse de acuerdo con tres variables de dos categorías cada una: sexo, raza (negra o blanca) y tipo de homicidio para el que las categorías consideradas son a) armas de fuego o explosivos y b) instrumentos punzocortantes. A pesar de estas restricciones los datos abarcan al 82% de los homicidios reportados ese año.

Estos datos se pueden presentar en un tabla de contingencia 2 x 2 x 2 como se muestra en el cuadro 6.5.

Con el propòsito de simplificar la notaciòn, sea la variable A el tipo de homicidio, B el sexo y C la raza.

Haberman (idem) ajusta el modelo logaritmico lineal

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

es decir, el modelo de "no relaciòn de segundo orden". Este modelo, como se viò en 3.2 puede abreviarse [AB] [AC] [BC].

CUADRO 6.5

Raza	Sexo	TIPO DE ATAQUE: Armas de fuego y explosivos	TIPO DE ATAQUE: Instrumentos punzocortantes	Totales
Blanca	Masc.	3 910	808	4 718
	Fem.	1 050	234	1 284
Negra	Masc.	5 218	1 385	6 603
	Fem.	929	298	1 227
Totales		11 107	2 725	13 832

Los valores observados y ajustados para cada una de las 8 celdas, obtenidos por Haberman se presentan en el cuadro 6.6.

CUADRO 6.6

Celda (A,B,C,)	Valores ovservados	Valores Ajustados
(1,1,1)	3 910	3 919.50
(2,1,1)	808	798.54
(1,2,1)	1 050	1 040.54
(2,2,1)	234	243.46
(1,1,2)	5 218	5208.50
(2,1,2)	1 385	1394.50
(1,2,2)	929	938.46
(2,2,2)	298	288.54

Sin embargo, en el mismo problema se menciona que el interès principal de la investigaciòn es conocer el grado en que el tipo de ataque depende del sexo o la raza de la victima. Entonces de acuerdo a la parte 3.3 de este trabajo, el anàlisis de los datos puede hacerse considerando un modelo logistico para respuesta binaria, donde el tipo de homicidio es la respuesta o variable dependiente.

Para utilizar un modelo logistico sin emplear el recurso visto en 3.4, el problema puede ser replanteado del siguiente modo:

Sea Y una variable aleatoria binaria que toma el valor cero si el tipo de homicidio en una victima fue por arma de fuego o explosivos y uno si fue con un objeto punzocortante.

Los datos necesarios solamente son las victimas por uno de los dos tipos de homicidio y el total, clasificados por raza y sexo como se muestra en el cuadro 6.7.

CUADRO 6.7

Raza (r_i)	Sexo (s_j)	Victimas por armas de fuego o explos. n_{jk}	Total de Victimas N_{jk}
Blanca (r_1)	Masculino (s_1)	3 910	4 718
	Femenino (s_2)	1 050	1 284
Negra (r_2)	Masculino (s_1)	5 218	6 603
	Femenino (S_2)	929	1 227

El número de victimas por armas de fuego o explosivos puede considerarse como proveniente de una distribución Binomial $B(p_{jk},N_{jk})$.

La probabilidad de que la victima haya sido atacada por arma de fuego y explosivos puede ser representada por:

$$P[Y=1|j,k] = \exp(u + s_i + r_j)/[1 + \exp(u + s_i + r_j)] \quad i,j = 1,2$$

Al aplicar la transformación logística

$$\lambda = \log (P[Y=1|j,k] / P[Y=0|j,k]) = u + r_i + s_j$$

Por medio del paquete GLIM, utilizando el componente aleatorio binomial, la liga logistica y los datos del cuadro 6.7 se ajusta el modelo logistico. En el cuadro 6.8 se encuentran los valores

observados, estimados y los residuales de Pearson, Devianza, Anscombe y Haberman.

Los valores estimados para el tipo de ataque por arma punzocortante son $N_{ij} - \hat{n}_{ij}$. Como se puede observar, los valores estimados son casi idènticos a los obtenidos por Haberman usando el modelo logaritmico lineal. El valor X^2 es de 1.077 con un grado de libertad por lo que el ajuste resulta ser muy adecuado.

Los residuales de Pearson y Devianza toman valores pequeños y muy parecidos y se podria decir que tienen una magnitud similar a los residuales de Anscombe; en estos tres casos se advierte un ligero sesgo en la distribuciòn de los residuales.

CUADRO 6.8

(s_i, r_j)	n_{ij}	N_{ij}	Valores estim. $\hat{n}_{ij}$	r_P	r_D	r_A	r_H
(s_1, r_1)	3 910	4 718	3 919	-0.37	-0.37	-0.38	-1.037
(s_2, r_1)	1 050	1 284	1 041	0.67	0.68	0.60	1.037
(s_1, r_2)	5 218	6 603	5 209	0.28	0.29	0.22	1.037
(s_2, r_2)	929	1 227	938.5	-0.64	-0.63	-0.67	-1.037

Para el cálculo de los residuales de Ascombe no es necesario usar el tèrmino para correcciòn de sesgo visto en 5.3 ya que su valor es insignificante al tenerse muestras grandes y probabilidades

estimadas suficientemente alejadas de los valores 0 y 1.

En este problema son de especial interès los residuales de Haberman ya que el nûmero de observaciones es grande y se podrìa decir que tienen una distribuciòn muy aproximada a una N(0,1); como se puede observar, estos residuales no son de una magnitud preocupante pues ninguno de ellos sale del intervalo (-1.96,1.96) y su suma es igual a cero. Para su obtenciòn se aplicô la expresiòn (5.7)

6.3 USO DE LA INFORMACION ACERCA DE LA ESCALA ORDINAL EN UN MODELO LOGARITMICO LINEAL.

En muchas ocasiones las categorìas de una variable representan puntos en una escala ordinal. Esta caracterìstica puede ser importante por lo que se debe tener presente cuando se ajusta un modelo.

Fienberg (1979 op.cit.) señala que varios autores han propuesto mètodos para tratar con categorìas ordinales en tablas de contingencia multidimensionales. El tratamiento que èl describe y que aquì se aplica, es bàsicamente el mismo que presentan Haberman (1974), Nerlove y Press (1973) y Simon (1974).

Considèrese una tabla de contingencia I x J con valores

observados $\{x_{ij}\}$ y supòngase que las J columnas estàn medidas en una escala ordinal y que se pueden ponderar previamente con coeficientes $\{v_j\}$. Si se decide que el modelo de independencia no es adecuado, en lugar de aceptar directamente el modelo saturado, conviene explorar algùn modelo alternativo que considere la relaciòn entre las columnas y sus ponderaciones, por ejemplo, se puede considerar el modelo

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + (v_j - \bar{v})u'_{1(i)} \qquad (6.5)$$

donde $\bar{v}$ es la media de las $\{v_j\}$. Para la soluciòn de las ecuaciones de verosimilitud es necesario que $0 \leq v_j \leq 1$ para $j = 1,2,\ldots, J$. Estas ecuaciones para el modelo (6.5) son:

$$\hat{m}_{i+} = x_{i+} \quad i = 1,2,\ldots,I$$

$$\hat{m}_{+j} = x_{+j} \quad j = 1,2,\ldots,J$$

y $$\sum v_j \hat{m}_{ij} = \sum v_j x_{ij} \quad i = 1,\ldots,I$$

Las cuales pueden ser solucionadas por el algoritmo Newton-Raphson, o bien por el mètodo de ajuste proporcional iterativo.

Fienberg (op.cit) cita a otros autores al presentar una tabla de contingencias 3 X 3 con datos cuya finalidad es relacionar la frecuencia de visitas con el tiempo de estancia de 132 internos esquizofrènicos en dos hospitales mentales de Londres. En el cuadro 6.9 se muestran los datos

El modelo de independencia entre renglones y columnas es:

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)}$$

Al ajustar este modelo se tiene que las estadísticas X^2 y G^2 tienen valores de 35.15 y 38.35 respectivamente, y se tienen 4 grados de libertad.

CUADRO 6.9

	Período de estancia en el hospital (t: años)			
Frecuencia de visitas	$2 \leq t < 10$	$10 \leq t < 20$	$t \geq 20$	Totales
Regularmente	43	16	3	62
Menos de una al mes	6	11	10	27
Nunca	9	18	16	43
Totales	58	45	29	132

Al comparar estos valores con el de la distribución Ji-cuadrada con cuatro grados de libertad y considerándose un nivel de significancia de 0.05 se rechaza la hipótesis de independencia.

Antes de tomar una decisión definitiva conviente observar el comportamiento de los residuales. En el cuadro 6.10 se presentan los valores observados, estimados y los residuales de Pearson, Haberman, Devianza y Anscombe.

Agresti (1984 p.63) muestra un ejemplo de una tabla bidimensional con variables ordinales donde calcula los residuales de Haberman para el modelo de independencia y observa que se obtienen residuales positivos donde el nivel es alto o bajo para ambas variables, mientras que cuando el nivel es bajo para una y alto para otra, se obtienen residuales negativos. Este comportamiento de los residuales sugiere ajustar un modelo que considere las escalas ordinales en las variables para tener bien representada su influencia.

El ejemplo de Agresti nos enseña la conveniencia de observar los residuales correspondientes a las celdas cercanas a las esquinas de la tabla cuando se tiene alguna variable ordinal y de este modo obtener información de la posible influencia de esta característica en el comportamiento de los datos.

En el cuadro 6.10 se observa que los residuales mayores en valor absoluto son los de Haberman y que los de Anscombe y Devianza toman prácticamente los misмo valores.

CUADRO 6.10

Celda	Valores observ.	Valores estima.	Residuales de Pearson	Residuales Haberman	Residuales Devianza	Residuales Anscombe
(1,1)	43	27.24	3.02	5.54	2.78	2.78
(2,1)	6	11.86	-1.70	-2.54	-1.88	-1.89
(3,1)	9	18.89	-2.28	-3.71	-2.54	-2.54
(1,2)	16	21.14	-1.12	-1.89	-1.17	-1.17
(2,2)	11	9.20	0.59	0.81	0.57	0.57
(3,2)	18	14.66	0.87	1.30	0.84	0.84
(1,3)	3	13.62	-2.88	-4.48	-3.49	-3.52
(2,3)	10	5.93	1.67	2.12	1.52	1.52
(3,3)	16	9.45	2.13	2.94	1.94	1.94

En el cuadro 6.11 se presentan los residuales de Haberman, Pearson y Devianza acomodados de acuerdo con la forma de la tabla para su anàlisis.

Se puede observar que, como en el ejemplo de Agresti, en las esquinas superior izquierda e inferior derecha los residuales son positivos y en las esquinas superior derecha e inferior izquierda, son negativos.

Tambièn se puede observar que los mayores residuales en valor absoluto se encuentran precisamente en las esquinas de la tabla. Esto sucede para los tres tipos de residuales considerados.

CUADRO 6.11

Frecuencia de Visitas	$2 \leq t < 10$	$10 \leq t < 20$	$t \geq 20$	
Regularmente	3.02	-1.12	-2.88	r_P
	5.54	-1.89	-4.48	r_H
	2.78	-1.17	-3.49	r_D
Menos de una vez al mes	-1.70	0.59	1.67	r_P
	-2.54	0.81	2.12	r_H
	-1.88	0.57	1.52	r_D
Nunca	-2.28	0.87	2.13	r_P
	-3.71	1.30	2.94	r_H
	-2.54	0.84	1.94	r_D

Se podría pensar en un modelo alternativo al sataurado y al de independencia (que es rechazado por las estadísticas X^2 y G^2), que considere la escala ordinal de las variables.

En este ejemplo Fienberg considera solamente la variable "tiempo de estancia como ordinal y propone el ajuste del modelo (6.5) con $v_1=-1$, $v_2=0$ y $v_3=1$; y añadir la restricción $u_{2(j)} = (v_j-\bar{v})u'$ con lo que quedaría

$$\log m_{ij} = u + u_{ij} = (v_j - \bar{v})u' + (v_j - \bar{v})\hat{u}_{2(i)} \qquad (6.6)$$

Al ajustar este modelo con el paquete GLIM se obtiene que $X^2 = 3.264$ y $G^2 = 3.195$ y ahora se tiene 3 g.l. por lo que el modelo se ajusta bastante bien.

En el cuadro 6.12 se presentan entre parèntesis los valores estimados al ajustar el modelo (6.6) y abajo de ellos los residuales de Pearson que se obtienen.

En los residuales ya no se observa ninguna tendencia y, en valor absoluto, sôlo uno excede la unidad y por muy poco.

CUADRO 6.12

Frecuencia de Visitas	$2 \leq t < 10$	$10 \leq t < 20$	$t \geq 20$
Regularmente	(44.19) -0.179	(13.61) -0.404	(4.19) -0.6
Menos de una vez al mes	(7.07) -0.647	(8.85) 0.722	(11.07) 1.055
Nunca	(10.98) -0.583	(14.05) -0.323	(17.98) -0.466

6.4 DETECCION Y TRATAMIENTO DE OBSERVACIONES DISCREPANTES EN UN MODELO LOGISTICO ESTIMULO-RESPUESTA.

Como se mencionó en 3.3 y en 5.4, Haberman (1973) ilustra el uso de sus residuales ajustados en modelos logistico lineales al considerar los datos del cuadro 6.13. Estos datos se refieren a un experimento sobre efecto tóxico de diferentes dosis de óxido de etileno aplicadas al insecto "Calandra granaria". La variable explicativa es el logaritmo de la dosis aplicada a diferentes grupos de insectos. Se emplea un modelo logistico para evaluar la forma en que el logaritmo de la dosis afecta la probabilidad muerte de los insectos.

Para plantear el modelo logistico considérese la columna del centro del cuadro 6.13 como el número N_j de insectos que reciben la log dosis t_j $1 \leq j \leq 10$. Para cada insecto se tienen dos posibles respuestas: sobrevive o muere. Sea n_{jk} es el número de insectos que habiendo recibido la dosis t_j tienen respuesta k, donde $1 \leq k \leq 2$. Entonces se puede considerar que la probabilidad de respuesta 1 dada la dosis t_j es

$$P(1|j) = 1/[\ 1 + \exp(-\ a - bt_j)]\quad 1 \leq j \leq 10$$

para alguna a y b.

se tiene entonces que:

$$\log [\ P(1|j)/P(2|j)] = a + bt_j$$

CUADRO 6.13

Log Dosis	Número de Sujetos	Número de Muertos
0.394	30	23
0.391	30	30
0.362	31	29
0.322	30	22
0.314	26	23
0.260	27	7
0.225	31	12
0.199	30	17
0.167	31	10
0.033	24	0

Para el anàlisis de residuales en el ajuste de este modelo, Haberman propone su càlculo por medio de la expresiòn (5.6) que como se viò en 5.4, es equivalente a la expresiòn (5.7) que puede ser calculada fàcilmente con el paquete GLIM.

En el cuadro 6.14 se añaden los valores ajustados y los residuales de Haberman, Pearson, Devianza y Anscombe.

Las estadísticas G^2 y X^2 son 36.25 y 33.22 respectivamente, ambas con 8 grados de libertad. El valor de la variable con distribuciòn Ji-cuadrada en el percentil 95 (es decir al considerar un nivel de significancia de $\alpha = 0.05$) y 8 grados de libertad es 15.51 por lo que el modelo falla claramente en el

ajuste. Los estimadores son $\hat{a} = -3.443$ y $\hat{b} = 14.44$.

CUADRO 6.14

log dosis	N_j	n_{j2}	Val. estim.	r_H	r_P	r_D	r_A
0.394	20	23	27.13	-2.86	-2.56	-2.21	-2.21
0.391	30	30	27.02	2.03	1.82	2.51	2.68
0.362	31	29	26.54	1.40	1.26	1.39	1.40
0.322	30	22	23.09	-0.52	-0.47	-0.47	-0.46
0.314	26	23	19.46	1.73	1.60	1.73	1.74
0.260	27	7	15.59	-3.60	-3.34	-3.35	-3.40
0.225	31	12	14.00	-0.81	-0.72	-0.73	-0.74
0.199	30	17	10.84	2.68	2.34	2.28	2.29
0.167	31	10	8.15	0.91	0.76	0.74	0.73
0.033	24	0	1.17	-1.25	-1.11	-1.55	-1.64

En el experimento se esperaría que el aumentar la dosis de veneno, aumentara paulatinamente la proporción de insectos muertos, es cierto que se deben presentar fluctuaciones debidas al azar pero ¿pueden ser éstas tan grandes como las que se presentan entre las observaciones 1 y 2 o entre las observaciones 6,7 y 8 ?

En las últimas cuatro columnas del cuadro 6.14 se presentan los residuales de Haberman, Pearson, Devianza y Anscombe. Igual que en los ejemplos anteriores, los residuales de Devianza y Anscombe toman valores casi idénticos.

Los residuales de Pearson correspondientes a las observaciones 1, 6 y 8 tienen magnitudes mayores de 2 en valor absoluto. Los

residuales de las otras tres columnas que exceden al 2 en valor absoluto son los correspondientes a las observaciones 1, 2, 6 y 8. El residual correspondiente a la observaciòn 6 (log dosis 0.260) en todos los casos es menor que -3.30 por lo que es sumamente improbable que èstos pudieran surgir de una distribuciòn N(0,1).

Haberman concluye que se deben haber presentado variaciones en las condiciones experimentales que no fueron registradas y que esto es fuertemente soportado por la diferencia tan grande en los resultados al aplicar dos log dosis de tòxico tan cercanos como .394 y .391 correspondientes a las observaciones 1 y 2.

Sin embargo, si se quiere profundizar en el anàlisis, se puede aceptar la conclusiòn de Haberman, pero por otro lado se puede observar que de cualquier forma, existe una tendencia de que al aumentar la log dosis de tòxico se incrementa la proporciòn de insectos muertos, por lo que los descuidos al mantener las condiciones constantes en el experimento pueden haber ocasionado observaciones discrepantes que pueden ser eliminados del anàlisis y estimadas posteriormente.

En 2.2 se viò que cuando se tiene un modelo lineal con observaciones discrepantes, este puede ser rearreglado del siguiente modo:

$$E\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} b$$

Donde Y_2 corresponde a k observaciones consideradas como posibles observaciones discrepantes. Y que una alternativa para tratar el problema es rechazar las posibles observaciones discrepantes, ajustar el modelo $E(Y_1) = X_1 b$, obtener $\hat{b} = (X_1' X_1)^{-1} X_1' Y_1$ y estimar $\hat{Y}_2 = X_2 \hat{b}$.

Por otro lado se revisaron articulos cuyo propòsito es la identificaciòn de 1,2,3,... observaciones discrepantes y se viò que mientras màs son èstas, el problema es mucho màs complejo.

En este problema concreto se considerarà primero como observaciòn discrepante a la correspondiente al residual de Pearson de mayor magnitud en valor absoluto. despuès se procederà como se acaba de exponer (ver 2.2) y asi sucecivamente hasta un buen ajuste y un comportamiento aceptable de los residuales.

Antes de proceder es importante tener mayor evidencia de que ese residual no corresponde al patrònde comportamiento esperado, con este propòsito se elaborò la gràfica en "papel de graficaciòn Normal" que aparece en la figura 6.8. Para elaborarla se siguiò el procedimiento visto en 2.1.1 y se tomò la recomendaciòn de Tukey (1962) de asignar el k-èsimo menor $r_{(k)}$ a la coordenada correspondiente a la probabilidad $(3k-1)/(3n+1) = (3k-1)/31$; idealmente los residuales deberian caer en la raya continua que

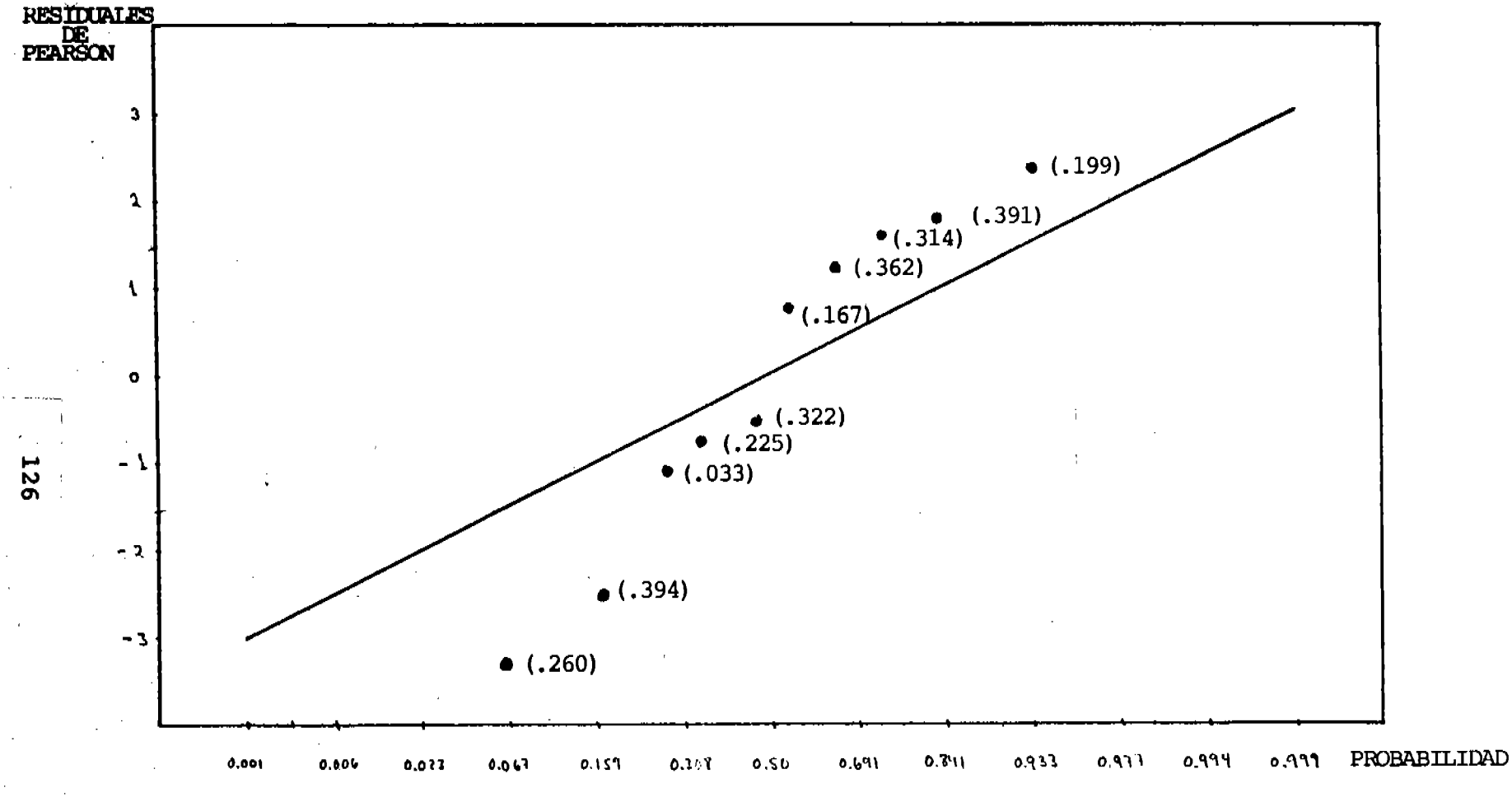

RESIDUALES DE PEARSON
3
2
1
0
-1
-2
-3
(.199)
(.391)
(.314)
(.362)
(.167)
(.322)
(.225)
(.033)
(.394)
(.260)
0.001
0.006
0.022
0.067
0.159
0.317
0.50
0.691
0.841
0.933
0.977
0.994
0.999
PROBABILIDAD

Figura 6.8

representa la Función de distribución N(0,1).

En la gráfica, junto a cada punto se encuentra entre paréntesis la log dosis a la que corresponde. Se observa que los residuales que se alejan más de la raya son los correspondientes a las log dosis 0.260 y 0.394.

Se podrían considerar de una vez, como observaciones discrepantes a las corresondientes a estas dos log dosis, sin embargo esto es peligroso ya que al eliminar una, el comportamiento de los demás residuales puede cambiar considerablemente.

Entonces, como primer paso, eliminamos la observación corespondiente a la log dosis .260 que tiene el residual de mayor magnitud y ajustamos el modelo con los otros nueve renglones.

En el cuadro 6.15 se presentan los nuevos valores ajustados y los residuales de Pearson.

Las estadísticas G^2 y X^2 son 23.2 y 22.14 respectivamente y se tienen 7 grados de libertad por lo que el modelo nuevamente falla al tratar de explicar los datos. El residual correspondiente a la log dosis .394 ahora es -3.14.

En la figura 6.9 se presenta la gráfica en "papel Normal" elaborada de manera similar a la anterior pero considerando los nueve residuales que ahora se tienen.

Es notorio el alejamiento de la raya continua del residual correspondiente a la log dosis 0.394 al compararlo con los demàs residuales. Entonces se elimina ahora el renglòn correspondiente a ese residual y se ajusta el modelo con los 8 restantes.

CUADRO 6.15

Log dosis	Observados	Estimados	Residuales de Pearson
.394	23	27.6	-3.14
.391	30	27.5	1.64
.362	29	27.3	0.94
.322	22	24.2	-0.99
.314	23	20.4	1.22
.225	12	15.6	-1.31
.199	17	12.3	1.72
.167	10	9.5	0.20
.033	0	1.4	-1.24

En el cuadro 6.16 se presentan las mismas columnas que en el cuadro 6.15 pero solamente con los 8 renglones considerados ahora.

La estadistica X^2 es ahora 12.2 y se tienen 6 grados de libertad por lo que se tiene un ajuste aceptable y todos los residuales estan dentro del intervalo (-1.96, 1.96). Los estimadores de a y b son: $\hat{a} = -3.858$ y $\hat{b} = 17.56$.

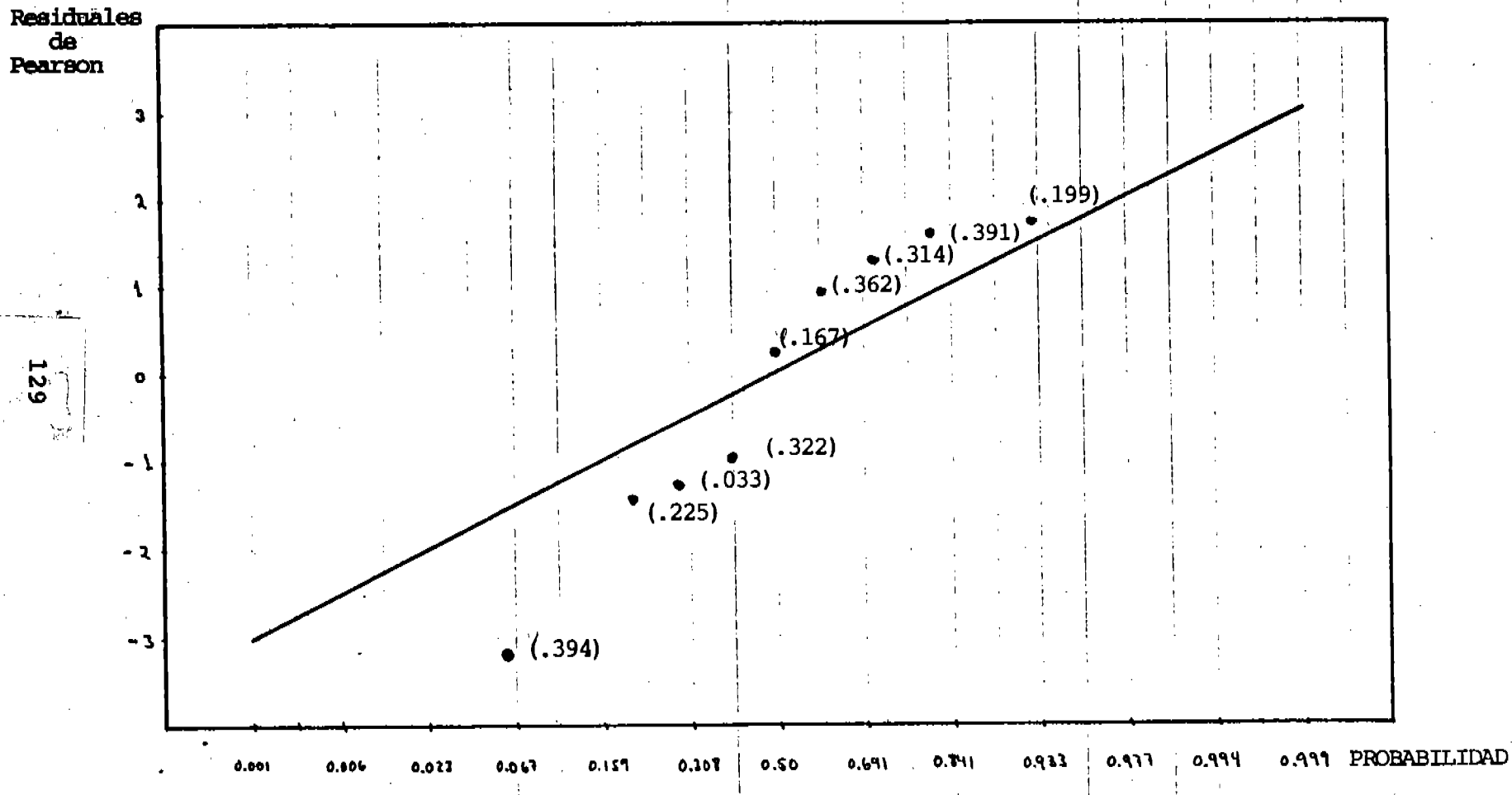

Residuales de Pearson
3
2
1
0
-1
-2
-3
(.199)
(.391)
(.314)
(.362)
(.167)
(.322)
(.033)
(.225)
(.394)
0.001
0.006
0.022
0.067
0.159
0.309
0.50
0.691
0.841
0.933
0.977
0.994
0.999
PROBABILIDAD

Figura 6.9

Para tener mayor claridad sobre el comportamiento de los residuales en este último ajuste, en la figura 6.10 se presenta la gráfica en "papel normal" de los ocho residuales de Pearson obtenidos. Como se puede ver, ningún residual proporsiona evidencias de que el modelo falle fuertemente en algún valor de la variable explicativa.

CUADRO 6.16

Log dosis	Observados	Estimados	Residuales
.391	30	28.6	1.22
.362	29	28.6	0.24
.322	22	25.7	-1.95
.314	23	21.8	0.62
.225	12	16.2	-1.52
.199	17	12.3	1.74
.167	10	8.8	0.48
.033	0	0.9	-0.95

Los estimadores del número de insectos muertos para las dosis 0.394 y 0.260 son:

$$\hat{Y}(.394) = N_1\hat{P}_1 = 30/[1 + \exp(3.852 - 17.56(.394))] = 28.67$$

$$\hat{Y}(.260) = N_4\hat{P}_4 = 27/[1 + \exp(3.852 - 17.56\ (.260))] = 18.09$$

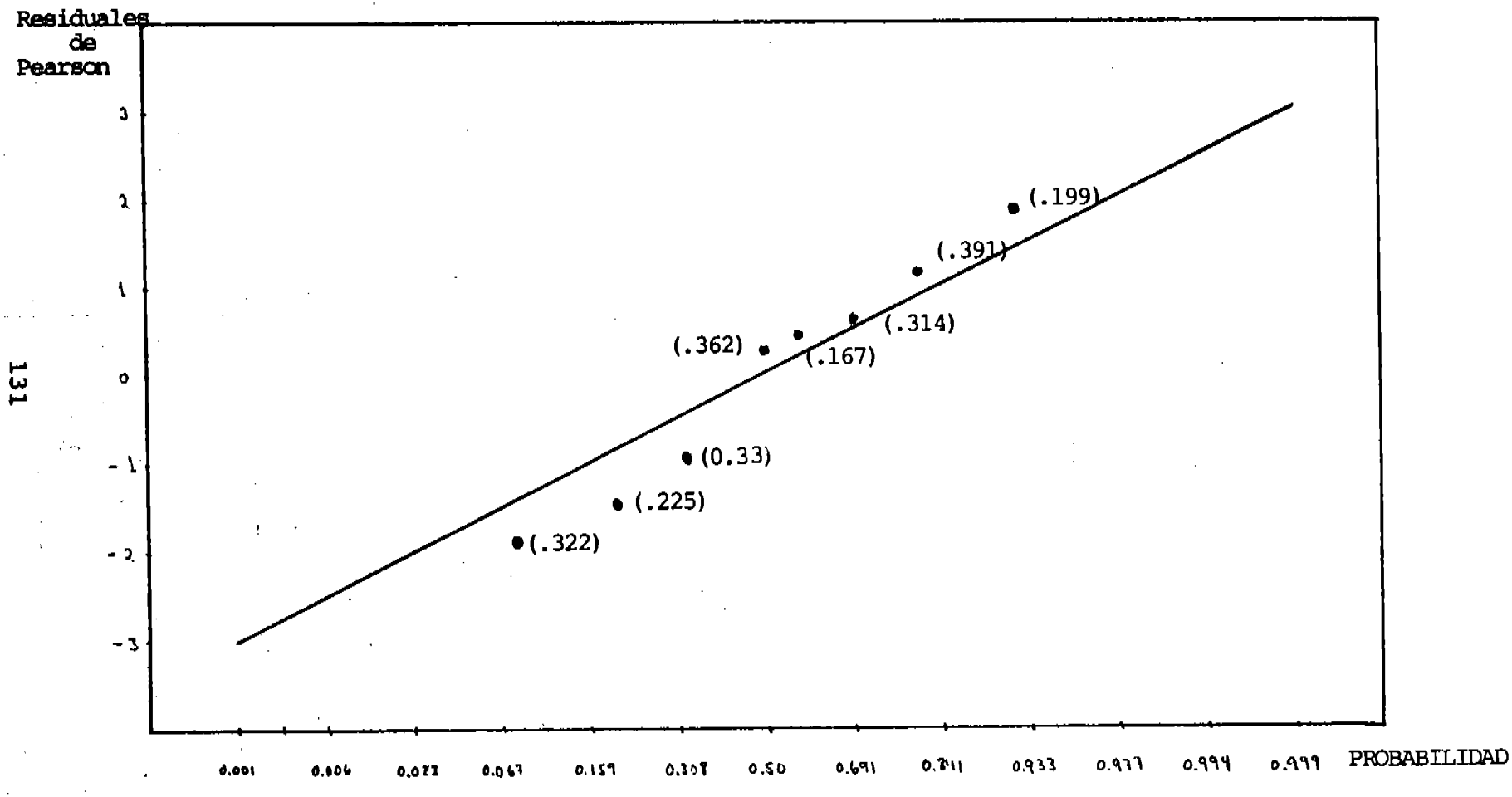

Residuales de Pearson
3
2
1
0
-1
-2
-3
(.199)
(.391)
(.314)
(.362)
(.167)
(0.33)
(.225)
(.322)
0.001
0.006
0.023
0.067
0.159
0.309
0.50
0.691
0.841
0.933
0.977
0.994
0.999
PROBABILIDAD

Figura 6.10

6.5 UN CASO DE SOBREDISPERSION EN UN MODELO LOGARITMICO LINEAL

Haberman (1973) ilustra el uso de sus residuales simples ajustados para modelos logaritmico lineales vistos en el capitulo 5, al considerar los datos que se muestran en el cuadro 6.17

CUADRO 6.17

Fallas en anillos de pistòn en cuatro compresoras

Número de compresora	Localizacion de la falla			Totales
	Parte superior	Parte media	Parte inferior	
1	17	17	12	46
2	11	9	13	33
3	11	8	19	38
4	14	7	28	49
Totales	53	41	72	166

Los datos del cuadro 6.17 corresponden a la frecuencia observada de fallas de anillos de pistòn en tres partes de cuatro compresoras en una planta quimica industrial.

El modelo logaritmico lineal que satisface la hiòtesis de independencia entre la probabilidad de falla en una compresora y la localizaciòn de aquella es:

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)}$$

La estadística X^2 de Pearson para esta tabla es 11.7 y se tienen 6 grados de libertad. El valor crítico en tablas de la distribución Ji-cuadrada para un nivel de significancia de $\alpha = 0.05$ es 12.6 y para $\alpha = .10$ es 10.6, por lo que se tiene una evidencia relativamente débil de la asosiación entre las dos variables.

Al tomar un nivel de significancia $\alpha = .05$ se podría concluir simplemente que el modelo es aceptado y que las variables son independientes.

Sin embargo, el investigador podría considerar que debe haber una relación entre las dos variables y entonces seleccionar $\alpha = 0.1$ con lo que el modelo de independencia sería rechazado.

A continuación se verá cómo el análisis de residuales contribuye de manera considerable a resolver la ambiguedad que provoca, en este caso, un valor de X^2 tan cercano al límite de la región de rechazo (ya sea $\alpha = 0.05$ ó $\alpha = 0.1$)

Aunque Haberman toma estos datos para ilustrar el uso de sus residuales, aquí se presentan diferentes residuales vistos en el capítulo anterior con el propósito de compararlos en un caso concreto y realizar un análisis más completo.

En el cuadro 6.18 se presentan los valores observados, estimados (considerando el modelo de independencia) y los residuales de Pearson, Haberman, Anscombe y Devianza.

Como se puede observar en ese cuadro, el conjunto de residuales de Haberman tiene cuatro elementos cuya magnitud en valor absoluto es mayor que 2.0, mientras que los otros tres conjuntos de residuales no tienen ningún elemento de esa magnitud y no parecen alejarse de un patrón de comportamiento de acuerdo con la N(0,1). Sin embargo, como se vió en el capitulo anterior la varianza asintótica de r_P es menor que 1.0 mientras que la distribución asintótica de los residuales de Haberman es N(0,1).

CUADRO 6.18

Celda	m_{ij}	$\hat{m}_{ij}$	r_P	r_H	r_A	r_D
(1,1)	17	14.69	0.604	0.86	0.58	0.59
(2,1)	11	10.54	0.143	0.19	0.14	0.14
(3,1)	11	12.13	-0.325	-0.45	-0.33	-0.33
(4,1)	14	15.64	-0.416	-0.60	-0.42	-0.42
(1,2)	17	11.36	1.673	2.27	1.56	1.56
(2,2)	9	8.151	0.297	0.38	0.29	0.29
(3,2)	8	9.386	-0.452	-0.59	-0.47	-0.46
(4,2)	7	12.10	-1.467	-2.01	-1.59	-1.59
(1,3)	12	19.95	-1.78	-2.78	-1.93	-1.92
(2,3)	13	14.31	-0.347	-0.52	-0.35	-0.35
(3,3)	19	16.48	0.620	0.94	0.61	0.61
(4,3)	28	21.25	1.464	2.32	1.40	1.39

También puede observarse que $\{r_A\}$ y $\{r_D\}$ tienen valores prácticamente iguales.

En su análisis Haberman presenta los r_H de acuerdo con la estructura de la tabla, como se muestra en el Cuadro 6.19

CUADRO 6.19

Número de compresora	Localización de la falla		
	P. Superior	P. Media	P. Inferior
1	0.86	2.27	-2.78
2	0.19	0.38	-0.52
3	-0.45	-0.59	0.94
4	-0.60	-2.01	2.32

Haberman señala que la inspección de esta tabla muestra que los residuales cuya magnitud es mayor que 2.0 en valor absoluto, son los correspondientes a las celdas (1,2), (1,3), (4,2) y (4,3) y que la probabilidad de que una extracción al azar de una N(0,1) exceda en valor absoluto al mayor de ellos (-2.78) es 0.0054. Indica que como el residual es sólo una selección de las 12 no se puede concluir tajantemente que la parte inferior de la compresora 1 no obedezca al modelo de independencia pero que hay razones obvias para sospecharlo.

Muestra una gràfica de los r_H sobre papel de probabilidad normal y comenta que èsta arroja todavìa màs evidencias de que el modelo puede no ser satisfactorio. Para elaborar la gràfica (fig. 6.11) se construye "papel normal" y se sigue el procedimiento visto en 2.1.1: se asigna el k-èsimo menor r_H a la coordenada correspondiente a la probabilidad $(3k-1)/(3rc+1) = (3k-1)/37$ donde r es el nûmero de renglones y c el nûmero de columnas. Idealmente los residuales deberìan caer en la raya continua de la figura que representa los percentiles para la distribuciòn N(0,1). Junto a cada punto se marca la celda a la que corresponde el residual.

Es claro que los r_H correspondientes a las celdas (1,2), (1,3), (4.2) y (4,3) estàn notoriamente alejados de la linea continua.

Haberman concluye que hay amplias razones para sospechar desviaciones del modelo en las compresoras 1 y 4, y que estas desviaciones pueden representar falta de independencia o bien, pueden indicar que el nûmero de fallas en anillos de pistòn tiene una varianza mayor de la debida si se acepta que tienen una distribuciòn Poisson.

Hasta aquì llega Haberman en su anàlisis, sin embargo, se puede tomar el señalamiento subrayado y pensar en datos provenientes de una distribuciòn Poisson con una sobredispersiòn.

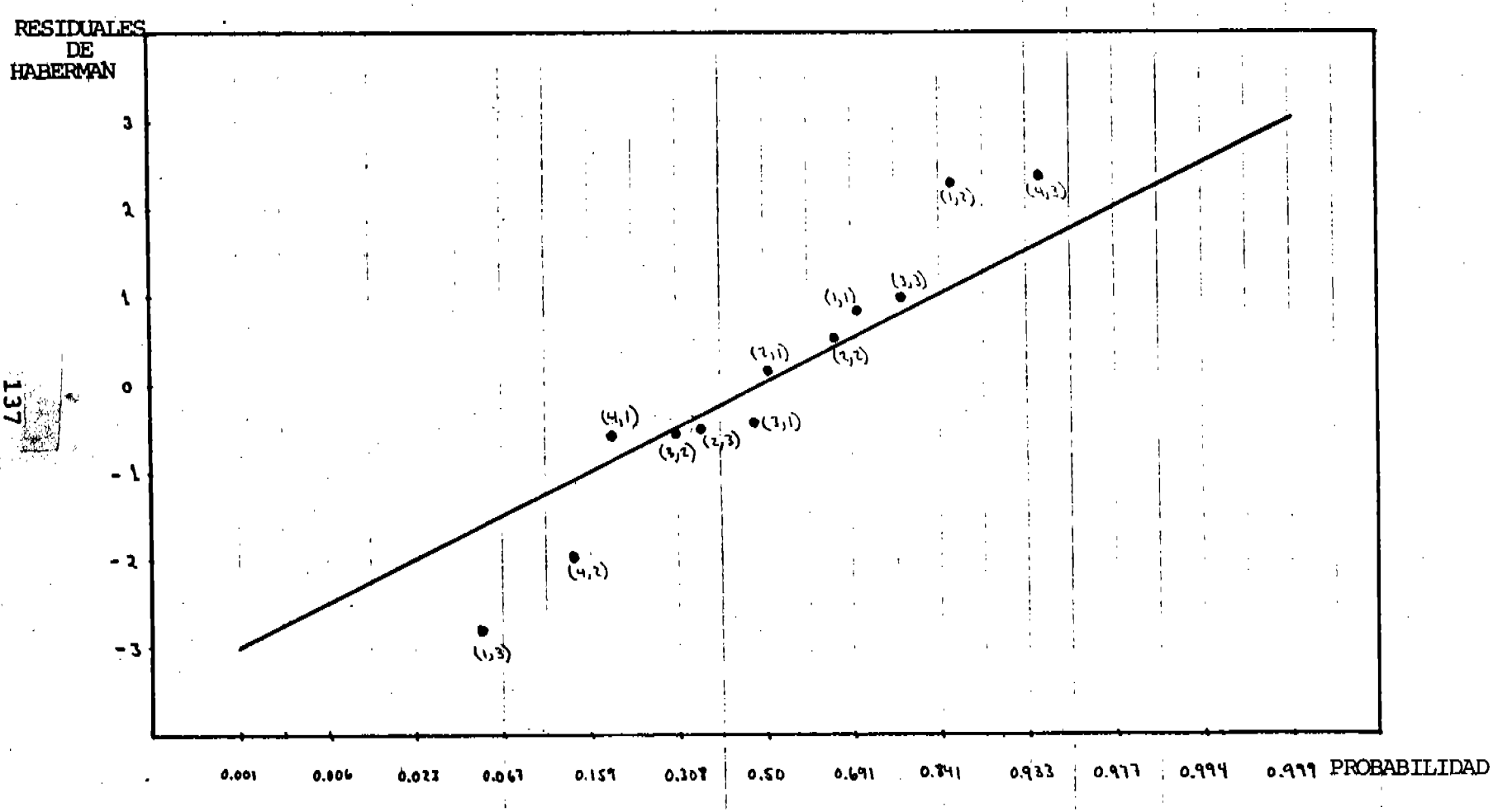

RESIDUALES DE HABERMAN
3
2
1
0
-1
-2
-3
(1,2)
(4,2)
(3,3)
(1,1)
(2,1)
(2,2)
(4,1)
(3,2)
(2,3)
(3,1)
(4,2)
(1,3)
0.001
0.006
0.022
0.067
0.159
0.308
0.50
0.691
0.841
0.933
0.977
0.994
0.999
PROBABILIDAD

Figura 6.11

McCullagh y Nelder (op.cit. pp 131-133) estudian la situaciòn de sobredispersiòn en modelos logaritmico lineales.

Señalan que existen casos en que la dispersiòn de los datos es mayor que la que se experaria del modelo Poisson, es decir, que si Y es la variable, se tiene $V(Y)= \varphi E(Y) > E(Y)$; que este fenòmeno puede surgir por diversas razones, por ejemplo, cuando se observa un proceso Poisson sobre intervalos cuya longitud en lugar de ser fija es aleatoria.

Despuès de dar otros ejemplos en que este fenòmeno puede ocurrir, proporcionan una forma sencilla para estimar el paràmetro de dispersiòn φ.

En el caso presente el paràmetro de dispersiòn puede ser estimado por el cociente $X^2/(N-p)$. Donde X^2 es la obtenida para el modelo de independencia, N el tamaño de la muestra y p el nùmerode paràmetros estimados.

En nuestro caso tenemos $N = 4 \times 3 = 12$ y $p = 6$ y el valor de $X^2_{(4)}$ es 11.7 por lo que $\hat{\varphi} = 11.7/6 = 1.95$

El paquete GLIM permite incorporar un paràmetro de dispersiòn. Al incluir $\varphi = 1.95$, el ajuste del modelo de independencia arroja un valor de $X^2 = 6.01$ que conduce a la aceptaciòn de la hipòtesis de independencia entre las dos variables. Y aunque los valores ajustados permanecen iguales, la magnitud de los residuales se

reduce considerablemente.

En el cuadro 6.20 se muestran los residuales de Haberman obtenidos al ajustar el modelo de independencia antes y después de incorporar el parámetro de sobredispersión. Entre paréntesis se encuentran los residuales antes de incluir el parámetro de dispersión y abajo de éstos, los valores después de incluirlo.

Como se puede apreciar, después de considerar el parámetro de sobredispersión en el ajuste del modelo, el mayor de los residuales de Haberman en valor absoluto toma un valor ligeramente menor que 2.0, lo que no es raro en una muestra de tamaño 12 de una N(0,1).

CUADRO 6.20

Número de Comp.	Localización de la falla		
	P. Superior	P. Media	P. Inferior
1	(0.86) 0.62	(2.27) 1.63	(-2.78) -1.99
2	(0.19) 0.136	(0.38) 0.27	(-0.52) -0.37
3	(-0.45) -0.32	(-0.59) -0.42	(0.94) 0.67
4	(-0.60) -0.43	(-2.01) -1.44	(2.32) 1.66

Para apreciar con mayor claridad el comportamiento de estos nuevos residuales de Haberman, en la figura 6.12 se presentan graficados en "papel de graficación Normal". Nuevamente junto a cada punto se señala la celda a la que corresponde.

Ahora que el modelo de independencia con sobredispersión se ajusta bien, los residuales aparecen bastante más pegados a la raya continua.

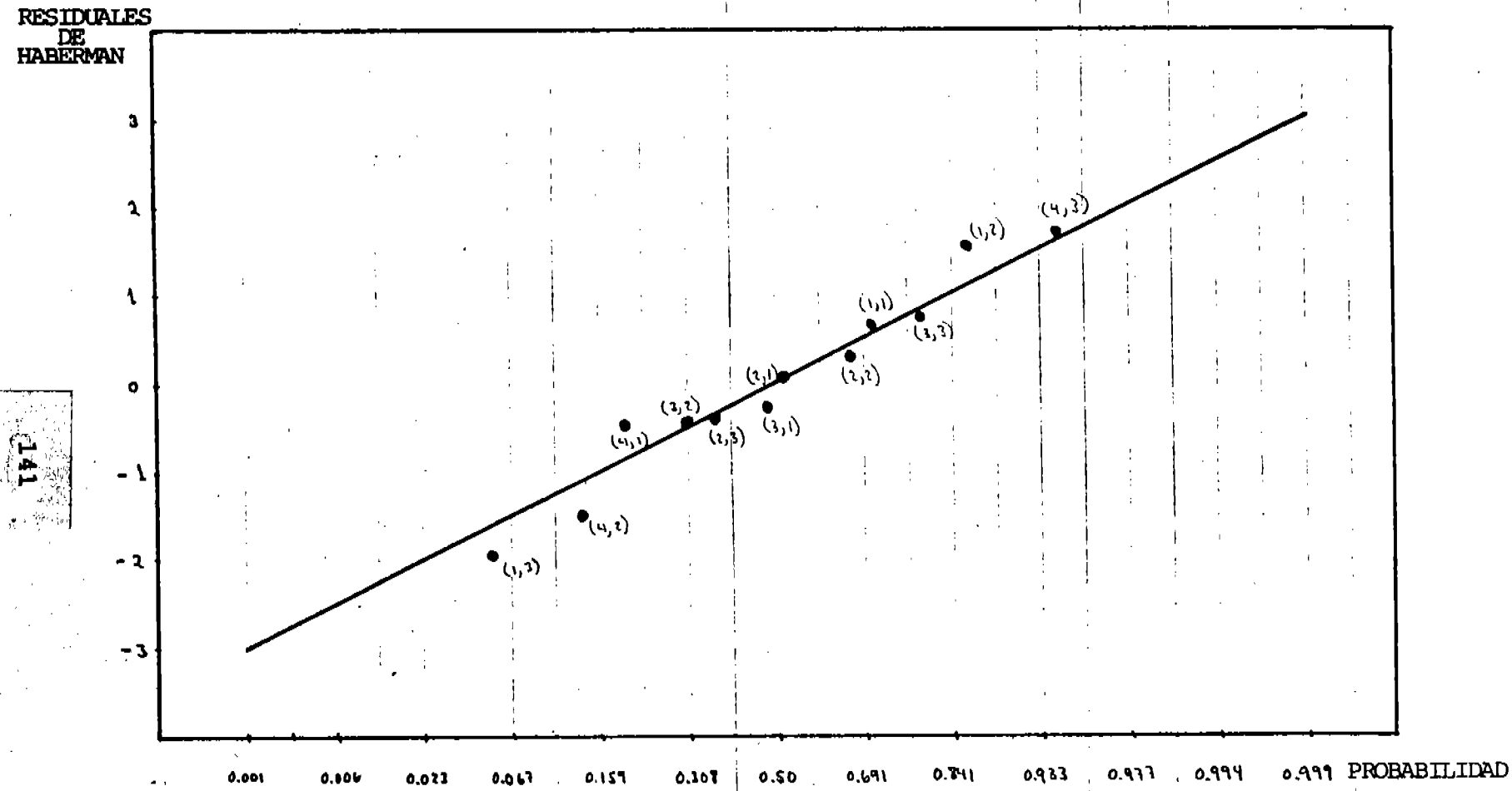

RESIDUALES
DE
HABERMAN
3
2
1
0
-1
-2
-3
(4,3)
(1,2)
(1,1)
(3,2)
(2,2)
(3,1)
(2,1)
(2,2)
(4,1)
(2,3)
(3,1)
(4,2)
(1,3)
0.001
0.006
0.022
0.067
0.159
0.309
0.50
0.691
0.841
0.933
0.977
0.994
0.999
PROBABILIDAD

Figura 6.12

7. CONCLUSIONES

7.1 La necesidad del anàlisis de residuales en el ajuste de modelos estadísticos en general, se considera cada vez màs importante, pues sus diferentes tipos de graficaciòn y mètodos de anàlisis proveen informaciòn acerca de lo adecuado de los modelos y del cumplimiento de las suposiciones subyacentes al anàlisis, ademàs de orientar respecto de las modificaciones que pueden intentarse para corregir las fallas que se presenten.

7.2 Algunas de las tècnicas para el anàlisis de residuales en modelos lineales clàsicos pueden ser utilizadas en modelos lineales generalizados y concretamente en la subclase de èstos que corresponden a datos categòricos. Esto sucede principalmente

cuando las situaciones guardan una analogía y cuando en los modelos lineales generalizados se utilizan residuales con una distribución aproximadamente Normal. Un caso típico de la traslación de los métodos de análisis de residuales se da en el caso de los modelos logísticos lineales con variables explicativas continuas ya que la interpretación de los parámetros en éstos es muy similar a la interpretación que se les da en los modelos de regresión.

7.3 Sobre el difícil caso de las observaciones discrepantes se puede concluir lo siguiente: en modelos lineales normales, el uso del mayor residual studentizado en valor absoluto es un criterio ampliamente aceptado para identificar una sola observación discrepante. Este criterio probablemente pueda trasladarse a los modelos para datos categóricos, con la diferencia de que en estos últimos este criterio se basaría en el mayor residual de Haberman en valor absoluto.

En la aplicación 6.4 se utilizó este último criterio incluso para la detección por pasos de dos observaciones discrepantes utilizando el método frecuentemente usado en modelos de análisis de varianza que consiste en reemplazar las observaciones identificadas como discrepantes por estimaciones de ellas, para lo cual el primer paso consiste en eliminar las observaciones discrepantes y los renglones correspondientes de la matriz X.

7.4 En modelos para datos categòricos, cuando alguna o algunas variables se miden es escalas ordinales, los residuales pueden dar evidencias de la necesidad de hacer explícito el orden en el modelo, a diferencia de los modelos lineales clàsicos en que el orden està implícito.

Como se pudo apreciar en la aplicaciòn 6.3 en los modelos para datos categòricos esta evidencia se puede observar en las celdas de las esquinas en una tabla de contingencia ya que si los residuales en estas celdas son de mayor magnitud en valor absoluto que el resto de residuales, esto indicarà que las fallas en el ajuste del modelo se presentan en las categorías mayores o menores de una o varias variables y/o en las categorias mayores de una o varias variables que se cruzan con categorías menores de otra u otras.

Al afinar el modelo incluyendo tèrminos que representen este tipo de informaciòn sobre categorías ordenadas, se lograrà un mejor ajuste y los residuales tendràn un comportamiento màs adecuado.

Este es uno de los aspectos que surgen en el anàlisis de residuales en modelos para datos categòricos que no corresponde a la adaptaciòn de alguna tècnica empleada en modelos lineales clàsicos.

7.5 Otras situaciones nuevas se presentan cuando en el anàlisis de residuales en modelos lineales generalizados (esto se facilita màs con los mètodos gràficos) permiten apreciar la necesidad de

cambiar la función liga o de la función varianza. Por ejemplo, en la aplicación 6.5 se vió la necesidad de incluir un parámetro de sobredispersión en un modelo logarítmico lineal para una tabla de contingencia; lo cual se pudo hacer gracias a la visión de los modelos para datos categóricos como modelos lineales generalizados.

7.6 Otro señalamiento con respecto a las observaciones discrepantes es que la revisión histórica de ellas muestra que, aunque las técnicas han cambiado con el tiempo, la actitud básica hacia ellas no ha variado considerablemente en el estudio de los modelos lineales clásicos y se podría decir que esta actitud se observa también en los modelos lineales generalizados ya que la mayor parte de definiciones que se dan acerca de ellas (cuando son varias) son subjetivas pues en esencia casi siempre se dice que son aquellas que a juicio del investigador se apartan considerablemente del resto de los datos o de lo que se esperaría de ellos.

7.7 Finalmente, se considera que se podría establecer un paralelismo en el estudio de las últimas técnicas desarrolladas o en desarrollo para observaciones influyentes y/o discrepantes para los modelos lineales clásicos y los modelos lineales generalizados con el propósito de lograr en lo posible un avance simultáneo.

A N E X O

PROGRAMAS DEL PAQUETE GLIM PARA LAS APLICACIONES DEL CAPITULO 6.
(*)

INSTRUCCIONES PARA LA APLICASCION 6.1

```
R *SERVICIO/GLIM
$UNITS 24
$DATA M
$READ M
19 23 24 29 33 42 57 47 37 63 66 68 29 47 43 27 43 27 23 30 49 55
52 53 50 42
$YVAR M
$ERROR P
$LINK L
$CAL X1=%GL(3,1) : X2%GL(2,3) : X3=%GL(2,6) : X4=%(2,12)
$FIT +X1 +X2 +X3 +X4
$
$LOOK %X2
$
$DISPLAY E R M T
$FIT $X2.X4
$
$LOOK %X2
$
$DISPLAY E R M T
$LOOK %X2
$
$DISPLAY E R M T
$FIT +X1.X3
$
$LOOK %X2
$
$FIT +X2.X3.X4
$
$LOOK %X2
$
$DISPLAY E R M T
$FIT +X1.X2.X3
$
$LOOK %X2
$
$DIDPLAY E R M T
$STOP
```

* El paquete GLIM està disponible en la computadora Burroughs 7-800 de la Direcciòn General de Servicios de Cômputo Acadèmico de la UNAM.

INSTRUCCIONES PARA LA APLICACION 6.2

```
R *SERVICIO/GLIM
UNITS 4
$DATA X1 X2 R N
$READ
1 1 3910 4718
1 2 1050 1284
2 1 5218 6603
2 2 929 1227
$YVAR R
$ERROR B N
$FIT +X1 +X2
$
$DISPLAY E R M T
$LOOK %X2
$
$CAL      D=%SQRT(2*(R*%LOG(R/%FV)+(N-R)*%LOG((N-R)*%LOG((N-R)/(N-
%FV))))
$LOOK D
$
$EXT %VL
$CAL V=N/(%FV*(N-%FV))
$CAL H=%WT*%VL
$CAL A=(R-%FV)/%SQRT((1-H)/V)
$LOOK A
$
$STOP
```

INSTRUCCIONES PARA LA APLICACION 6.3

```
R *SERVICIO/GLIM
$UNITS 9
$DATA FREC
$READ 43 6 9 16 11 18 3 10 16
$FAC X1 3 X2 3
$CAL X1=%GL(3,1) : X2=%GL(3,3)
$YVAR FREC
$ERROR P
$LINK L
$FIT +X1 +X2
$
$LOOK %X2
$DISPLAY E R M T
$CAL RD = %SQRT(2*FREC*%LOG(FREC/%FV)-FREC+%FV))
```

```
$LOOK RD
$
$CALL RA=(3/2)*FREC**(2/3)-%FV**(2/3)/%FV**(1/6)
$LOOK RA
$
$STOP
```

(ahora introduciendo un término que considere la relacion de la escala ordinal)

```
R *SERVICIO/GLIM
$UNITS 9
$DATA FREC
$READ 43 16 3 6 11 10 9 18 16
$FAC R 3
$CAL R=%GL(3,3)
$CAL V=%GL(3,1)
$CAL V=V-2
$YVAR FREC
$ERROR P
$LINK L
$FIT +R.V
$
$LOOK %X2
$
$DISPLAY E R M T
$STOP
```

INSTRUCCIONES PARA LA APLICACION 6.4

```
R *SERVICIO/GLIM
$UNITS 10
$DATA T N R
$READ
$ERROR V N
$YVAR R
$LINK G
$FIT T
$
$LOOK %X2
$
$CAL D =%SQRT(2*(R*%LOG(R/%FV)+(N-R)*%LOG((N-R)/N-%FV)))
$LOOK D
$
$EXT %VL
$CAL V=N/(%FV*(N-%FV))
$CAL H=%WT*%VL
$CAL A=(R-%FV)/%SQRT((1-H)/V)
$LOOK A
```

```
$
$STOP
```

(se repiten todas las instrucciones, primero eliminando el sexto renglon de datos y despuès el primero)

NSTRUCCIONES PARA LA APLICACION 6.5

```
R *SERVICIO/GLIM
$UNITS 12
$DATA FREC
$READ 17 11 11 14 17 9 8 7 12 13 19 28
$FAC X1 4 X2 3
$CAL X1=%GL(4,1) : X2=%GL(3,4)
$YVAR FREC
$LINK L
$ERROR P
$FIT +X1 +X2
$
$LOOK %X2
$
$DISPLAY E R M T
$EXT %VL
$OWN M1 M2 M3 M4
$MAC M1 $CAL %FV=%EXP(%LP) $ ENDMAC
$MAC M2 $CAL %PR=1/%FV $ENDMAC
$MAC M3 $CAL %VA=1.95*%FV $ENDMAC
$MAC M4 $CAL %DI=2*(%YV*%LOG(%YV/%FV)-(%YV-%FV) $ENDMAC
$LOOK %X2
$DISPLAY E R M T
```

BIBLIOGRAFIA

Agresti, A. 1983a. "A Survey of Strategies for Modeling Cross-Classificatons Having Ordinal Variables". J. Amer. Statist. Assoc. 78, pp. 184-198.

Agresti, A. 1984 "Analysis of Ordinal Categorical Data". John Wiley & Sons, New York.

Anscombe, F.J. 1953 "Contribution to the discussion of H. Hotellings paper". J.R. Statist. Soc., B, 30, 248-275.

Backer, R.J. and Nelder, J.A. 1978. "The Glim System. Release 3. Generalized Linear Interactive Modelling Manual". Numerical Algorithms Group, Oxford.

Beckman, R.J. and Cook, R.D. 1983. "Outlier...s" Technometrics Vol. 25, No. 2 pp. 119-149.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. 1975. "Discrete multivariate analysis: theory and practice". MIT Press, Cambridge, Massachusetts.

Cox, D.R. and Snell, E.J. 1968. "A General Definition of Residuals". J.R. Statist. Soc., B 30. pp 248-275.

Cox, D.R. 1970. "The Analysis of Binary Data" Chapman and Hall, London.

Draper, N.R. and John, J.A."Testing for Three or Fewer Outliers in two-Way tables". Technometrics Vol. 22 No. 1 pp. 9-15.

Draper, N.R. and Smith, H. 1981. "Applied Regression Analysis" 2d. ed. Wiley, New York.

Fienberg, S.E. 1979. "The Analysis of Cross-Classified Categorical Data" Mit Press, Cambridge.

Haberman, S.J. 1973. "The Analysis of Residuals in Cross-Classified Tables". Biometrics 29, pp. 205-220.

Haberman, S.J. 1976. "Generalized Residuals for Log-Linear Models" Proc. Ninth Inter., Biometrics, Conf. 1, pp. 104-122.

Haberman, S.J. 1978. "Analysis of Qualitative Data. Vol. 1: Introductory Topics". Academic Press, New York.

Haberman, S.J. 1978. "Analysis of Qualitative Data Vol. 2: New Developments". Academic Press, New York.

Hoel, P. G. 1966. "Introduction to Mathematical Statistics" 3d. ed. Wiley, New York.

John, J.A. and Draper, N.R. 1978. "One Testing for Two Outliers in Two-Way Tables". Technometrics, Vol. 20, No.1, pp. 69-78.

Lindgren, B.W. 1968. "Statistical Theory". Third ed. Macmillan, New York .

McCullagh, P. and Nelder, J. 1983 "Generalized Linear Models". Chapman and Hall, London.

Méndez, R. I. 1976. "Modelos estadisticos lineales". Foccavi/Conacyt; Mexico, D. F.

Montgomery, D. C. and Peck, E. A. 1982. "Introduction Regression Analysis" Wiley, New York.

Nelder, J. A. 1974 "Log Linear Models for Contingency Tables: A Generalization of Classical Least Squares". Appl. Statist. 13, No.3, pp. 323-329.

Nelder, J. A. and Wederburn, R. W. M. 1972. "Generalized Linear Models". J. Roy. Statist. Soc. A 135, pp. 370-384.

Nerlove, M. and Press, S.J. 1973. "Univariate and Multivariate Log-linear and Logistic Models." Rand Corporation Tech. Report R-1306-EDA/NIH, Santa Monica, Calif.

Pregibon, D. 1982. "Score test in GLIM" First Symposium of GLIM Applications. Lectures Notes in Statistics No. 12. Springer Verlag, London.

Simon, G. 1979 "Alternative Analysis for the Singly-Ordered contingency Table" J. Amer. Statist. Assoc. 69, 971-976.

Tukey, J.W. 1962. "The Future of Data Analysis" Ann. Math. Statist. 33, 1-67.