

03061

1
2ej.



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UACP y P del CCH

Técnicas de diagnóstico para el
modelo logístico

FALA DE ORIGEN

T E S I S

QUE PARA OBTENER EL TITULO DE

MAESTRA EN CIENCIAS

P R E S E N T A

REBECA AGUIRRE HERNANDEZ

MEXICO, D. F.

1991



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Indice.

1. Introducción	1
2. Modelos	3
2.1 Modelos lineales generalizados	3
2.2 Modelos para respuestas bionomiales	4
2.2.1 El modelo logístico	5
2.2.2 Otros modelos	5
2.2.3 Comparación de modelos	6
Bibliografía abreviada	6
3. Estimación	7
3.1 Modelos lineales generalizados	7
3.2 Modelo logístico	10
Bibliografía abreviada	11
4. Bondad de ajuste, residuos y palancas	12
4.1 Bondad de ajuste	12
4.1.1 Estadísticas de bondad de ajuste	12
4.1.2 Propiedades	13
4.1.3 Razones por las que un modelo es inapropiado	14
4.2 Residuos	14
4.2.1 Modelos lineales generalizados	14
4.2.2 Modelo logístico	16
4.3 Palancas	16
Bibliografía abreviada	17
5. Inferencia	18
5.1 Pruebas de hipótesis	18
5.2 Intervalos de verosimilitud	19
Bibliografía abreviada	20

6. Diagnósticos	21
6.1 Respuestas sin suavizar	21
6.1.1 Datos no agrupados. Gráfica de la devianza media local	22
6.1.2 Datos agrupados o no agrupados	26
6.1.2.1 Gráfica de probabilidad empírica	26
6.1.2.2 Gráfica de la variable agregada	31
6.1.2.3 Gráfica de la variable construida	35
6.1.2.4 Colinealidad	39
6.1.2.5 Influencia	41
6.1.2.5.1 Influencia global	41
· Influencia global en el nivel de colinealidad	41
· Estimadores a un paso	44
a. Influencia global de un dato en $\hat{\beta}$	46
b. Influencia global de un dato en las medidas de bondad de ajuste	47
c. Influencia global de un dato en los residuos de la devianza	48
6.1.2.5.2 Influencia local	56
· Influencia local en todos los parámetros	56
a. Influencia local del peso de los datos en δ	59
b. Influencia local de las variables explicativas en δ	59
· Influencia local en un subconjunto de parámetros	60
a. Influencia local del peso de los datos en δ_1	60
· Influencia local en las predicciones puntuales	64
a. Perturbación de la variable de respuesta	66
b. Perturbación de un variable explicativa	67
c. Perturbación de un dato	67
d. Perturbación del peso de los datos	68
6.1.2.6 Datos discrepantes	69
6.1.3 Datos agrupados	72
6.1.3.1 Prueba de bondad de liga	72
6.1.3.2 Variabilidad extra-binomial	77
6.1.3.3 Influencia global en pruebas de hipótesis	82
6.2 Respuestas suavizadas	85

7. Conclusiones	90
Bibliografía	92
Apéndices	95

1. Introducción.

Esta tesis trata de las técnicas estadísticas cuya finalidad es examinar el ajuste del modelo logístico con variables explicativas continuas.

El objetivo es dar a conocer las técnicas post-ajuste para el modelo mencionado y propiciar su aplicación. Con este fin se realizó una recopilación de las técnicas, se homogeneizó la notación, se crearon subrutinas de cómputo y se aplicaron y compararon las técnicas estudiadas. Adicionalmente se identificaron problemas que precisan de mayor estudio. Aproximadamente la mitad de las técnicas se presentan en el contexto de los modelos lineales generalizados y no exclusivamente en términos del modelo logístico. Esto le da a la tesis un alcance mayor al planeado.

El proceso de modelaje estadístico no solo incluye la recolección de datos y la estimación de parámetros sino también la evaluación del modelo ajustado. Esta última etapa pocas veces se efectúa en la práctica. Las razones son diversas: las técnicas post-ajuste para el modelo logístico son recientes y están dispersas en la literatura estadística, se carece de algoritmos de cómputo, no siempre se dispone de tiempo para realizar este tipo de análisis o no existe el hábito de hacerlo. Se espera que la tesis contribuya a facilitar la solución de estos problemas.

La tesis no es teórica. Las demostraciones se omitieron. Se adoptó un enfoque práctico porque: 1). El objetivo es difundir las técnicas post-ajuste entre las distintas personas que trabajan con el modelo logístico; 2). No es sencillo aplicar las técnicas post-ajuste; se requieren guías y experiencia en la interpretación de índices y gráficas.

Las aplicaciones se hicieron principalmente con GLIM3. Este es un paquete estadístico interactivo que ajusta modelos lineales generalizados y que permite la creación de subrutinas. Algunos autores lo usan para ilustrar la aplicación de las técnicas que proponen. También se usaron BMDPLR e IMSL.

En los capítulos 2 a 5 se describen someramente los modelos lineales generalizados, el modelo logístico, el procedimiento de ajuste, las estadísticas de bondad de ajuste, los residuos, las palancas y las pruebas de hipótesis para modelos anidados. Se sugiere que las personas que conocen el modelo logístico les den un vistazo para familiarizarse con la notación. Las secciones 4.2 y 4.3 deben leerse más cuidadosamente porque los residuos y las palancas son las herramientas básicas del análisis post-ajuste. El capítulo

6, que trata de las técnicas post-ajuste (o diagnósticos), se divide en dos partes. La primera engloba a las técnicas que se aplican sin necesidad de suavizar las respuestas y se subdivide en tres secciones. La sección 6.1.1 (datos no agrupados) reúne a los diagnósticos que sólo deben aplicarse cuando la respuesta es binaria; la sección 6.1.2 (datos no agrupados y agrupados) comprende a las técnicas aplicables a cualquier modelo logístico y la sección 6.1.3 (datos agrupados) incluye a los diagnósticos para respuestas Bernoulli con repeticiones. En la segunda parte del capítulo 6 se describe un procedimiento para suavizar las respuestas binarias y algunas técnicas post-ajuste basadas en respuestas suavizadas. La interpretación de los diagnósticos se ilustra con ejemplos. Al final de cada sección, bajo el encabezado de: Bibliografía abreviada, se mencionan los autores y el año de las publicaciones consultadas. La bibliografía completa está al final de la tesis. El capítulo 7 contiene las conclusiones. En el apéndice se encuentran los datos y las subrutinas creadas.

Agradezco al Departamento de Probabilidad y Estadística del IIMAS, U.N.A.M., el espacio físico que me proporcionaron a petición del Dr. Francisco Javier Aranda Ordaz. Esto me permitió participar en diversas actividades académicas y convivir con los integrantes de dicho departameto y con los alumnos de la Maestría en Estadística e Investigación de Operaciones. También deseo expresar mi gratitud a la Dra. Silvia Ruíz-Velasco Acosta por asesorarme al morir el Dr. Francisco J. Aranda Ordaz. Asimismo, agradezco a los sinodales sus valiosos comentarios que contribuyeron a mejorar la tesis, especialmente al Dr. Ignacio Méndez Ramírez y a la Dr. Silvia Ruíz-Velasco Acosta. A las personas que tuvieron la paciencia de enseñarme a usar el procesador de texto TeX les doy las gracias. Además, agradezco al CONACYT su apoyo.

2. Modelos.

Existen diversos modelos estadísticos que permiten estudiar cómo un conjunto de variables, llamadas explicativas, afectan el comportamiento de otra variable, denominada "la respuesta". El modelo que se escoja debe ser consistente con leyes conocidas (físicas, biológicas, etc.). Estrictamente hablando, cualquier modelo que se seleccione es incorrecto, porque todos simplifican la realidad. Sin embargo, algunos modelos son mejores que otros y el problema es identificarlos.

En este capítulo se describen los modelos lineales generalizados y algunos modelos para respuestas con distribución binomial.

2.1 Modelos lineales generalizados.

Los modelos lineales generalizados tienen tres componentes:

- a. Aleatoria. Suponga que Y_1, Y_2, \dots, Y_n es una muestra de n variables independientes o no correlacionadas. La función de densidad de la respuesta Y_i debe tener la forma:

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}$$

con $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ funciones conocidas. El parámetro de dispersión ϕ puede ser conocido o desconocido; θ_i es el parámetro canónico y y_i el valor observado de Y_i . Si ϕ es conocido, la función de densidad de Y_i pertenece a la familia exponencial.

- b. Sistemática. También se le llama predictor lineal y se define como

$$\eta_i = x_i' \beta = \beta_0 + \sum_{j=1}^{m-1} x_{ij} \beta_j; \quad i = 1, 2, \dots, n.$$

En forma matricial: $\eta = X \cdot \beta$,

en donde $\eta' = (\eta_1, \eta_2, \dots, \eta_n)$,

$X' = (x_1, x_2, \dots, x_n)$ con $x_i' = (x_{i1}, x_{i2}, \dots, x_{im})$,

$\beta' = (\beta_0, \beta_1, \dots, \beta_{m-1})$.

Las variables explicativas: x_1, x_2, \dots, x_{m-1} pueden ser cuantitativas o cualitativas y deben medirse sin error. El vector de parámetros β , de $m \times 1$, es desconocido.

- c. Función liga. En la literatura se manejan dos ligas: $g(\cdot)$ y $k(\cdot)$. La primera relaciona a $\mu_i = E(Y_i)$ con el predictor lineal: $g(\mu_i) = \eta_i$. Es decir, $g(\cdot)$ es una transformación de la media expresada como una función lineal de los parámetros $\beta_0, \beta_1, \dots, \beta_{m-1}$. Se dice que $g(\cdot)$ es una liga canónica si $g(\mu) = \theta$, esto implica que $\eta = \theta$. La liga "natural" se define como: $k(\eta_i) = \theta_i$. Si $g(\cdot)$ es una liga canónica entonces $k(\cdot)$ es la identidad.

2.2 Modelos para respuestas binomiales.

En esta sección se mencionan varios modelos que describen la relación entre una respuesta binomial y un conjunto de variables explicativas. Los datos serán:

$$(y_1, x'_1), (y_2, x'_2), \dots, (y_n, x'_n)$$

en donde y_i es el valor observado de $Y_i \sim \text{Bin}(n_i, p_i)$ con $i = 1, 2, \dots, n$.

Se dice que los datos no están agrupados si $n_i = 1$ para toda i . Esto implica que (y_i, x'_i) contiene la información de un sólo individuo. Estos datos son comunes en estudios observacionales. Un conjunto de datos desagrupados tiene repeticiones si hay dos o más individuos con el mismo vector de variables explicativas (i.e si por lo menos existe una $i \neq i'$ tal que $x_i = x_{i'}$).

Los datos están agrupados si existe alguna $n_i > 1$. En este caso, y_i es el número de éxitos observados en n_i individuos con vector de variables explicativas x_i . Esto quiere decir que (y_i, x'_i) resume la información de un subconjunto de datos desagrupados con repeticiones. Los datos agrupados son frecuentes en estudios experimentales en los que el investigador controla las variables explicativas.

Es importante distinguir entre datos agrupados y no agrupados porque: (a) algunas técnicas estadísticas sólo deben aplicarse a datos no agrupados y (b) la teoría asintótica es distinta en cada caso.

2.2.1 El modelo logístico.

El modelo logístico pertenece a la familia de los modelos lineales generalizados. La variable de respuesta Y_i es una binomial con media $\mu_i = n_i p_i$. El parámetro desconocido p_i es la probabilidad de observar un éxito en un ensayo Bernoulli que se repite n_i veces de manera independiente. La función que relaciona a μ_i con el predictor lineal es la logística:

$$g(\mu_i) = \ln \left(\frac{\mu_i}{n_i - \mu_i} \right) = \ln \left(\frac{p_i}{1 - p_i} \right).$$

Esta liga es canónica y se denota como $\text{logit}(p_i)$. En otras palabras, el modelo logístico se define como:

$$\ln \left(\frac{p_i}{1 - p_i} \right) = \eta_i = \beta_0 + \sum_{j=1}^{m-1} x_{ij} \beta_j = x_i' \beta; \quad i = 1, 2, \dots, n$$

lo que implica que

$$p_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} \quad \text{con } i = 1, 2, \dots, n.$$

A este modelo se le llama el modelo de regresión logística cuando η_i está formado por variables explicativas cuantitativas y posiblemente algunas variables indicadoras — como en regresión lineal normal. El modelo de regresión logística ha sido usado en bioensayos (para estudiar el efecto que tienen distintas dosis de veneno en la supervivencia o muerte de un animal), observaciones seriales ($x_i = i$) y datos de supervivencia.

2.2.2 Otros modelos.

Existen otros modelos, además del logístico, que describen el comportamiento de la probabilidad de éxito en términos de x . Cox y Snell (1989) mencionan dos modelos lineales. La transformación angular de p_i , definida como

$$p_i = \begin{cases} 1, & \text{si } x_i > \frac{\pi}{4\delta}; \\ \text{sen}^2(\delta x_i + \frac{\pi}{4}), & \text{si } |x_i| \leq \frac{\pi}{4\delta}; \\ 0, & \text{si } x_i < -\frac{\pi}{4\delta}; \end{cases}$$

con x univariada, induce simetría en la distribución de Y_i y estabiliza su varianza cuando $n_i = k$ para toda i (δ es un parámetro). Los modelos probit y doble-log complementario también son apropiados para respuestas binomiales y pertenecen a los modelos lineales generalizados. La liga probit se define como: $g(\mu_i) = \Phi^{-1}(\mu_i)$ (Φ^{-1} es la inversa de la función de distribución normal) y la doble-log complementaria como: $g(\mu_i) = \log(-\log(1 - \mu_i))$. Esta última es apropiada cuando se desea tratar asimétricamente los éxitos y los fracasos.

2.2.3 Comparación de modelos.

La transformación angular de p_i , el modelo logístico y el probit son similares cuando $p_i \in (0.1, 0.9)$. El segundo tiene el atractivo de que sus parámetros se interpretan en términos del logaritmo del momio de p_i .

Bibliografía abreviada.

Cox, D.R. y Snell (1989).

Chatterjee, S. y Hadi, A.S. (1988).

McCullagh P. y Nelder J.A. (1983).

Thomas y Cook (1990).

3. Estimación.

Para poder ajustar un modelo es necesario que el número de parámetros, m , no exceda al número de observaciones, n . Generalmente, el ajuste de un modelo mejora a medida que el número de parámetros aumenta, sin embargo la interpretación se complica. Los modelos lineales generalizados más simples son de la forma $g(\mu) = \beta_0$. En estos, la componente aleatoria describe toda la variabilidad de la respuesta. En cambio, los modelos saturados, $g(\mu) = \beta_0 + \sum_{j=1}^{n-1} x_j \beta_j$, tienen n parámetros y la componente sistemática explica toda la variabilidad de Y . Los valores observados (y_i) y ajustados ($\hat{\mu}_i$) de los modelos saturados son iguales.

Cuando se trabaja con modelos lineales generalizados, el estimador de máxima verosimilitud de β frecuentemente se obtiene con el método de Newton-Raphson. Este método, y una variante de él, se describen a continuación.

3.1 Modelos lineales generalizados.

El vector de coeficientes β de un modelo lineal generalizado puede estimarse por máxima verosimilitud. Como $\theta_i = k(\eta_i) = k(x_i' \beta)$, el logaritmo de la función de verosimilitud de un modelo lineal generalizado es:

$$L(\beta) = \sum_{i=1}^n l_i = \sum_{i=1}^n \{a^{-1}(\phi)[y_i k(x_i' \beta) - b\{k(x_i' \beta)\}] + c(y_i, \phi)\} \quad (3.1.1)$$

con $l_i = \ln f_{Y_i}(y_i; \theta_i, \phi)$. Generalmente, ϕ y β se estiman por separado. Por esto, se supondrá que el parámetro de dispersión es conocido o que se tiene un estimador de él. La función de puntaje ("score") para β es:

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{1}{a(\phi)} \cdot \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta} = \sum_{i=1}^n \frac{1}{a(\phi)} x_i' \dot{\epsilon}_i(\beta)$$

$$\text{con } \dot{\epsilon}_i(\beta) = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \eta_i} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial k(\eta_i)}{\partial \eta_i} = [y_i - b\{k(x_i' \beta)\}] k'(x_i' \beta),$$

$$b\{k(x_i' \beta)\} = \frac{\partial b(\theta_i)}{\partial \theta_i} \quad \text{y} \quad k'(x_i' \beta) = \frac{\partial k(\eta_i)}{\partial \eta_i}.$$

Las ecuaciones de verosimilitud para β son:

$$\left. \frac{\partial L(\beta)}{\partial \beta} \right|_{\beta=\hat{\beta}} = \frac{X' \dot{\epsilon}(\beta)}{a(\phi)} \Big|_{\beta=\hat{\beta}} = 0, \quad (3.1.2)$$

en donde X es la matriz de diseño de $n \times m$, $\hat{\beta}$ es el estimador de máxima verosimilitud de β y $\dot{\epsilon}(\beta)$ es un vector de $n \times 1$ cuyos elementos son $\dot{\epsilon}_i(\beta)$.

Generalmente, las ecuaciones (3.1.2) no son una combinación lineal de $\hat{\beta}$, pero pueden resolverse con el método de Newton-Raphson, cuando puede calcularse la matriz de información observada:¹

$$-\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} = \frac{1}{a(\phi)} \cdot X' \bar{E}(\beta) X;$$

$$\bar{E}(\beta) = \text{diag}\{\bar{e}_i(\beta)\},$$

$$\bar{e}_i(\beta) = \bar{b}\{k(x'_i\beta)\} \{k'(x'_i\beta)\}^2 - \{y_i - \bar{b}\{k(x'_i\beta)\}\} \bar{k}(x'_i\beta),$$

$$\bar{b}\{k(x'_i\beta)\} = \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} \quad \text{y} \quad \bar{k}(x'_i\beta) = \frac{\partial^2 k(\eta_i)}{\partial \eta_i^2}.$$

El método de Newton-Raphson es iterativo; el estimador de β en la iteración $t+1$, β^{t+1} , satisface el sistema de ecuaciones:

$$\left\{ a^{-1}(\phi) X' \dot{\epsilon}(\beta) - a^{-1}(\phi) X' \bar{E}(\beta) X \cdot (\beta^{t+1} - \beta) \right\} \Big|_{\beta=\beta^t} \approx 0 \quad (3.1.3).$$

Estas ecuaciones son una aproximación lineal a (3.1.2) y se obtienen desarrollando en series de Taylor la función de puntaje para β alrededor de β^t . El estimador β^{t+1} converge a $\hat{\beta}$ si

$$\left. \frac{\partial L(\beta)}{\partial \beta} \right|_{\beta=\beta^{t+1}}$$

es suficientemente cercana a cero. Aparentemente, β^0 , el estimador inicial de β , no tiene un efecto relevante en la convergencia del método.

A menudo, el método de Newton-Raphson se modifica con la técnica de puntajes de Fisher. Esta técnica consiste en aproximar la matriz de información observada con su esperanza:

¹ El hessiano es el negativo de la matriz de información observada.

$$E \left(\frac{-\partial^2 L(\beta)}{\partial \beta \partial \beta'} \right) = a^{-1}(\phi) X' W(\beta) X \quad \text{con} \quad W(\beta) = E[\bar{E}(\beta)].$$

(Como $E(Y_i) = \bar{b}\{k(x'_i; \beta)\}$ entonces $W(\beta) = \text{diag}\{\bar{b}\{k(x'_i; \beta)\}\{k(x'_i; \beta)\}^2\}; \bar{b}\{k(x'_i; \beta)\}$ es la función de varianza de Y_i y $\text{Var}(Y_i) = \bar{b}\{k(x'_i; \beta)\}a(\phi)$.)

Si se usa esta aproximación en (3.1.3) se obtiene:

$$\left. \left\{ a^{-1}(\phi) X' \dot{e}(\beta) - a^{-1}(\phi) X' W(\beta) X \cdot (\beta^{t+1} - \beta) \right\} \right|_{\beta=\beta^t} \simeq 0$$

o equivalentemente:

$$X' W(\beta^t) X \cdot \beta^{t+1} \simeq X' \{ \dot{e}(\beta^t) + W(\beta^t) X \beta^t \} \quad (3.1.4)$$

$$\text{con} \quad W(\beta^t) = W(\beta) \Big|_{\beta=\beta^t} \quad \text{y} \quad \dot{e}(\beta^t) = \dot{e}(\beta) \Big|_{\beta=\beta^t}.$$

Nótese que éstas son las mismas ecuaciones que resultan cuando el modelo:

$$W^{-1}(\beta^t) \dot{e}(\beta^t) + X \beta^t = X \beta^{t+1} + \varepsilon_t \quad \text{con} \quad E(\varepsilon_t) = 0 \quad \text{y} \quad \text{Var}(\varepsilon_t) = W^{-1}(\beta^t) \quad (3.1.5)$$

se ajusta con el método de mínimos cuadrados ponderados. De esto se concluye que:

1. La solución a (3.1.2) puede calcularse aplicando iterativamente mínimos cuadrados ponderados; y
2. Este método es equivalente al de Newton-Raphson modificado con la técnica de puntajes de Fisher.

En la expresión (3.1.5), $W(\beta^t)$ y $z(\beta^t) = W^{-1}(\beta^t) \dot{e}(\beta^t) + X \beta^t$ se conocen como la matriz de pesos y la variable de trabajo, respectivamente. Se denotarán como W y z a la matriz $W(\beta)$ y al vector $z(\beta) = W^{-1}(\beta) \dot{e}(\beta) + X \beta$ evaluados en $\hat{\beta}$ respectivamente. Las componentes del vector \dot{e} serán los residuos de puntaje:

$$\dot{e}_i = \dot{e}_i(\beta) \Big|_{\beta=\hat{\beta}}.$$

Asintóticamente, el estimador de máxima verosimilitud de β se distribuye como una $N_m(\beta, (X' W X)^{-1})$.

Cuando la liga es canónica (como en el modelo logístico), $k(x'_i\beta) = 1$ y $\bar{k}(x'_i\beta) = 0$. Como consecuencia, la matriz de información observada es igual a su esperanza y la técnica de puntajes de Fisher no altera al método de Newton-Raphson. Además, en este caso existen estadísticas suficientes para β .

3.2 Modelo logístico.

El logaritmo de la función de verosimilitud para el modelo logístico es:

$$L(\beta) = \sum_{i=1}^n \left\{ y_i(x'_i\beta) - n_i \log(1 + \exp(x'_i\beta)) + \log \binom{n_i}{y_i} \right\}$$

Es decir, en la expresión (3.1.1), $\phi = a(\phi) = 1$, $\theta_i = \eta_i = x'_i\beta$, $b\{k(x'_i\beta)\} = -n_i \log(1 + \exp(x'_i\beta))$ y $c(y_i, \phi) = \log \binom{n_i}{y_i}$. Las ecuaciones de verosimilitud son:

$$X'\dot{\epsilon} = 0,$$

los elementos de $\dot{\epsilon}$ son los residuos de puntaje: $\dot{\epsilon}_i = y_i - n_i \hat{p}_i$ con $\hat{p}_i = \exp(x'_i\beta) / \{1 + \exp(x'_i\beta)\}$. Obsérvese que las ecuaciones de verosimilitud no son una combinación lineal de $\hat{\beta}$, pero pueden resolverse con el método de Newton-Raphson. Como la liga logística es canónica:

$$-\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} = E \left(\frac{-\partial^2 L(\beta)}{\partial \beta \partial \beta'} \right) = X'W(\beta)X$$

con $W(\beta) = \text{diag}\{n_i p_i (1 - p_i)\}$. De acuerdo con las expresiones (3.1.3) y (3.1.4) de la sección anterior, el estimador de β en la iteración $t + 1$ puede expresarse como:

$$\beta^{t+1} \simeq \beta^t + (X'W(\beta^t)X)^{-1} X'\dot{\epsilon}(\beta^t).$$

Una expresión equivalente para β^{t+1} es:

$$\beta^{t+1} \simeq (X'W(\beta^t)X)^{-1} X'W(\beta^t)z(\beta^t),$$

$W(\beta^t)$ es la matriz diagonal de pesos con elementos $w_i^t = n_i p_i^t (1 - p_i^t)$; $z(\beta^t)$ es la variable de trabajo, definida como $W^{-1}(\beta^t)\dot{\epsilon}(\beta^t) + X\beta^t$ y $\dot{\epsilon}(\beta^t)$ es un vector de $n \times 1$,

sus elementos son $\dot{e}_i(\beta^t) = y_i - n_i p_i^t$ con $p_i^t = \exp(x_i^t \beta^t) / \{1 + \exp(x_i^t \beta^t)\}$. La matriz de pesos y la variable de trabajo evaluadas en el estimador de máxima verosimilitud de β serán W y z respectivamente.

La estadística suficiente para β es $s = X'Y$. Cuando los datos están agrupados, los supuestos para calcular la distribución asintótica de $\hat{\beta}$ son que $n_i \rightarrow \infty$ para $i = 1, 2, \dots, n$ y que n y m permanecen fijos. En este caso, si el modelo logístico es correcto,

$$\sqrt{n}(\hat{\beta} - \beta) \sim N_m(0, n(X'WX)^{-1} + O_m(n^{-\frac{1}{2}})).$$

Cuando los datos no están agrupados y $n \rightarrow \infty$, la distribución asintótica de $\hat{\beta}$ también es una normal si el modelo es correcto.

Bibliografía abreviada.

McCullagh, P. y Nelder, J.A. (1983).

Green, P.J. (1984).

Chatterjee, S. y Hadi, A.S. (1988).

Pregibon, D. (1981).

Thomas, W. y Cook, D.R. (1990).

4. Bondad de ajuste, residuos y palancas.

Las medidas de bondad de ajuste indican si en términos generales el modelo ajustado describe apropiadamente la relación entre Y y x . Si ellas indican que el modelo es satisfactorio se procede a examinar los residuos y las palancas. Los residuos permiten determinar qué datos están bien representados por el modelo y cuáles no. Las palancas, en combinación con los residuos, reflejan la influencia potencial de cada dato en el ajuste del modelo. Diversas técnicas de diagnóstico están basadas en los residuos y las palancas.

4.1 Bondad de ajuste.

La bondad de ajuste de los modelos lineales generalizados se mide con las estadísticas ji-cuadrada de Pearson generalizada y de devianza. Se definirán estas dos estadísticas que miden la distancia entre Y y $\hat{\mu}$ y se mencionarán algunas de sus propiedades.

4.1.1 Estadísticas de bondad de ajuste.

La ji-cuadrada de Pearson generalizada y la devianza miden qué tan bien se ajusta, en términos globales, un modelo lineal generalizado. Estas estadísticas resumen las discrepancias entre los valores observados (y_i) y ajustados ($\hat{\mu}_i$).

La ji-cuadrada de Pearson generalizada es:

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\widehat{Var}(Y_i)}$$

en donde $\hat{\mu}_i = \hat{E}(Y_i) = \hat{b}\{k(x_i^T \hat{\beta})\}$. En el modelo logístico, la media y la varianza estimadas de Y_i son, respectivamente, $\hat{\mu}_i = n_i \hat{p}_i$ y $\widehat{Var}(Y_i) = n_i \hat{p}_i (1 - \hat{p}_i)$ con $\hat{p}_i = \exp(x_i^T \hat{\beta}) / \{1 + \exp(x_i^T \hat{\beta})\}$; por lo tanto,

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{p}_i)^2}{n_i \hat{p}_i (1 - \hat{p}_i)}$$

La devianza, D^2 , de cualquier modelo lineal generalizado es proporcional a:

$$-2[L(\hat{\beta}) - L(\bar{\beta})] = -2 \sum_{i=1}^n a^{-1}(\phi) [y_i(\hat{\theta}_i - \bar{\theta}_i) + b(\bar{\theta}_i) - b(\hat{\theta}_i)]$$

$$\text{con } L(\hat{\beta}) = \sup_{\beta} L(\beta), \quad \beta \text{ de mx1} ;$$

$$L(\bar{\beta}) = \sup_{\beta^*} L(\beta^*), \quad \beta^* \text{ de nx1} .$$

Obsérvese que $\bar{\beta}$ estima al vector de coeficientes del modelo saturado: β^* . Cuando $a(\phi) = \phi$ (el peso inicial asignado a cada dato es unitario),

$$D^2 = -2 \sum_{i=1}^n [y_i(\hat{\theta}_i - \bar{\theta}_i) + b(\bar{\theta}_i) - b(\hat{\theta}_i)]. \quad (4.1.1.1)$$

La devianza del modelo logístico es:

$$D^2 = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) + (n_i - y_i) \ln \left(\frac{n_i - y_i}{n_i - \hat{\mu}_i} \right) \right\}, \quad (4.1.1.2)$$

$$\text{con } \bar{\theta}_i = \ln \left(\frac{y_i}{n_i - y_i} \right) \quad \text{y} \quad \hat{\theta}_i = \ln \left(\frac{\hat{\mu}_i}{n_i - \hat{\mu}_i} \right).$$

Si los datos no están agrupados, la expresión (4.1.1.2) se reduce a:

$$D^2 = -2 \sum_{i=1}^n \left\{ y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i) \right\} \quad (4.1.1.3).$$

Existen otras medidas de bondad de ajuste para el modelo logístico — consúltese: Hosmer y Lemeshow (1989, pp. 139 -149) —, pero no son de uso general.

4.1.2 Propiedades.

El valor de la devianza decrece conforme el número de parámetros aumenta; esto no sucede con la X^2 . Por otro lado, es más sencillo interpretar la ji-cuadrada de Pearson generalizada que la devianza. Jennings (1986) y Davison (1989) señalan que D^2 no es un indicador de la bondad de ajuste del modelo logístico. La razón es que la estadística

suficiente, $s = X'Y$, es una función uno a uno de $\hat{\beta}$ y D^2 sólo depende de los datos a través de $\hat{\beta}$. En otras palabras, en D^2 no se compara a los datos con el modelo. La estadística X^2 está exenta de este problema.

Sólo cuando los datos están agrupados y el modelo logístico es correcto, la distribución asintótica de X^2 y D^2 es una ji-cuadrada con $n - m$ grados de libertad. Empíricamente, se considera que el modelo logístico se ajusta a un conjunto de datos no agrupados si X^2 y D^2 son aproximadamente iguales a $n - m$.

4.1.3 Razones por las que un modelo es inapropiado.

Las causas de que un modelo lineal generalizado no se ajuste a los datos pueden ser varias: la distribución de la respuesta es incorrecta, el predictor lineal está mal definido o la función liga es inapropiada. En particular, en el modelo logístico es factible que la distribución binomial no refleje la variabilidad de la respuesta (variabilidad extra-binomial). También es posible que falte incluir en el predictor lineal alguna variable explicativa o que sea necesario transformar alguna ya incluida. Asimismo, puede suceder que la liga logística sea inapropiada.

4.2 Residuos.

Los residuos reflejan qué tan bien se ajusta el modelo a cada uno de los datos. Existen varios residuos para los modelos lineales generalizados. En la sección siguiente se mencionan tres de ellos: los de Pearson, Anscombe y de devianza.

4.2.1 Modelos lineales generalizados.

El residuo de Pearson para la observación i de un modelo lineal generalizado es:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(Y_i)}} \quad \text{con} \quad \hat{\mu}_i = \hat{E}(Y_i) = b\{k(x_i' \hat{\beta})\}.$$

La suma de estos residuos elevados al cuadrado es igual a la estadística ji-cuadrada de Pearson generalizada, i.e. $X^2 = \sum_{i=1}^n r_i^2$. En general, la distribución de r_i es

asimétrica.

Los residuos de Anscombe involucran a la función $A(\cdot)$ que transforma la distribución de Y_i a una que sea aproximadamente normal. La función $A(\cdot)$ se define como:

$$A(\cdot) = \int \frac{d\mu}{V^{\frac{1}{3}}(\mu)}.$$

El i -ésimo residuo de Anscombe es:

$$r_{A_i} = \frac{A(Y_i) - A(\hat{\mu}_i)}{\dot{A}(\hat{\mu}_i)\sqrt{\widehat{Var}(Y_i)}} \quad \text{para } i = 1, 2, \dots, n;$$

$\dot{A}(\hat{\mu}_i)$ es la derivada de $A(\mu_i)$ respecto a μ_i evaluada en $\hat{\mu}_i = \hat{\mu}_i$.

Según la expresión (4.1.1.1), la contribución del dato i a la devianza es:

$$d_i = -2[y_i(\hat{\theta}_i - \bar{\theta}_i) + b(\bar{\theta}_i) - b(\hat{\theta}_i)].$$

El residuo de la devianza correspondiente al dato i es:

$$dev_i = \sqrt{d_i} \operatorname{sgn}(y_i - \hat{\mu}_i)$$

$$\text{en donde } \operatorname{sgn}(y_i - \hat{\mu}_i) = \begin{cases} -1, & \text{si } y_i - \hat{\mu}_i < 0, \\ 1, & \text{si } y_i - \hat{\mu}_i > 0. \end{cases}$$

Cuando los residuos de Pearson y de la devianza se dividen por $\sqrt{1 - h_i}$ su varianza asintótica es igual a uno; $h_i = w_i x_i' (X'WX)^{-1} x_i$ con w_i el i -ésimo elemento en la diagonal de la matriz de pesos W . A

$$r_{P_i} = \frac{Y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(Y_i)(1 - h_i)}} = \frac{r_i}{\sqrt{1 - h_i}}$$

se le llama el residuo de Pearson estandarizado y a

$$r_{D_i} = \frac{\sqrt{d_i} \operatorname{sgn}(y_i - \hat{\mu}_i)}{\sqrt{1 - h_i}} = \frac{dev_i}{\sqrt{1 - h_i}}$$

el residuo de la devianza estandarizado.

4.2.2 Modelo logístico.

El i -ésimo residuo de Pearson para el modelo logístico es:

$$r_i = \frac{Y_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}}$$

y el residuo de la devianza es: $dev_i = \sqrt{d_i} \operatorname{sgn}(y_i - n_i \hat{p}_i)$ con

$$d_i = 2 \left[y_i \ln \left(\frac{y_i}{n_i \hat{p}_i} \right) + (n_i - y_i) \ln \left(\frac{n_i - y_i}{n_i - n_i \hat{p}_i} \right) \right].$$

Si los datos no están agrupados,

$$r_i = \frac{Y_i - \hat{p}_i}{\sqrt{\hat{p}_i (1 - \hat{p}_i)}} \quad \text{y} \quad dev_i = \{-2[y_i \ln \hat{p}_i + (1 - y_i) \ln (1 - \hat{p}_i)]\}^{\frac{1}{2}} \operatorname{sgn}(y_i - \hat{p}_i).$$

La distribución de los residuos anteriores depende de p_i y por lo tanto de x_i . McCullagh y Nelder (1983) no recomiendan usar el residuo de Pearson cuando n_i es pequeña ó p_i está próxima a cero o uno. Los residuos estandarizados de Pearson y de la devianza son, respectivamente:

$$r_{P_i} = \frac{r_i}{\sqrt{1 - h_i}} \quad \text{y} \quad r_{D_i} = \frac{dev_i}{\sqrt{1 - h_i}} \quad \text{con} \quad h_i = n_i \hat{p}_i (1 - \hat{p}_i) x_i' (X' W X)^{-1} x_i.$$

Una manera de examinar la magnitud de los residuos es graficándolos contra i , el número de dato.

4.3 Palancas.

En el modelo de regresión lineal normal, los datos influyentes tienen las características siguientes: a) su palanca ("leverage") es relativamente grande; b) su valor ajustado, $\hat{\mu}_i$, está básicamente determinado por y_i ; y c) su vector de variables explicativas, x_i , es extremo en el espacio generado por las columnas de X . Los datos influyentes de los

modelos lineales generalizados no poseen las tres características anteriores. Además, hay dos tipos de palanca para estos modelos.

Thomas y Cook (1990) propusieron que las palancas de los modelos lineales generalizados sean:

$$h_i^* = \bar{h}_i k(x_i' \hat{\beta}) \sqrt{\widehat{Var}(Y_i)}, \quad (4.3.1)$$

con \bar{h}_i el i -ésimo elemento en la diagonal de $X(X'EX)^{-1}X'$. Los autores señalan que entre mayor sea el valor de la expresión (4.3.1) mayor será la influencia (local) de Y_i en $\hat{\mu}_i$. En particular, para el modelo logístico la expresión (4.3.1) es igual a:

$$h_i^* = x_i'(X'WX)^{-1}x_i \sqrt{n_i \hat{p}_i(1 - \hat{p}_i)}; \quad (4.3.2)$$

W es una matriz diagonal con i -ésimo elemento: $w_i = n_i \hat{p}_i(1 - \hat{p}_i)$.

Basado en el algoritmo de mínimos cuadrados ponderados, Pregibon definió las palancas del modelo logístico como en regresión lineal normal ponderada:

$$h_i = w_i x_i'(X'WX)^{-1}x_i. \quad (4.3.3)$$

Estas palancas suman a m , a diferencia de lo que sucede con las de Thomas y Cook. Además, $0 < h_i < 1$. Los datos con una palanca relativamente grande (h_i ó h_i^*) no son extremos en el espacio generado por las columnas de X - como sucede en regresión lineal normal. Además, los datos con los mayores valores de h_i no siempre tienen una h_i^* "grande".

Bibliografía abreviada.

- Davison, A.C. (1989).
- Hosmer, D.W. y Lemeshow, S. (1989).
- Jennings, D.E. (1986).
- McCullagh, P. y Nelder, J.A. (1983).
- Pregibon, D. (1981).
- Thomas, W. y Cook, D.R. (1990).

5. Inferencia.

La teoría de este capítulo es aplicable a cualquier modelo lineal generalizado. Se explicará cómo probar la significancia de un subconjunto de variables explicativas a partir de resultados asintóticos. También se indicará cómo construir un intervalo de verosimilitud para una función real de β . La aproximación ji-cuadrada que se usa convierte al intervalo de verosimilitud en un intervalo de confianza — exacto o asintótico.

5.1 Pruebas de hipótesis.

En esta sección se explica cómo probar la significancia de un subconjunto de variables explicativas de un modelo lineal generalizado. Sin pérdida de generalidad, se supondrá que dichas variables son las últimas q de X . Entonces, dado el modelo:

$$g(\mu_1) = X\beta = X_p\beta_p + X_q\beta_q,$$

con X_p de $n \times p$, X_q de $n \times q$, β_p de $p \times 1$, β_q de $q \times 1$ y $p + q = m$, se probará la hipótesis:

$$H_0 : \beta_q = 0 \quad \text{vs.} \quad H_1 : \beta_q \neq 0.$$

La estadística de prueba puede ser la de puntaje ó menos dos veces el logaritmo del cociente de verosimilitudes generalizado: $\Omega = -2 \ln \Lambda$. En ambos casos, el valor crítico se calcula a partir de la distribución asintótica de la estadística de prueba.

Puede demostrarse que $\Omega = \phi^{-1}(D_0^2 - D_1^2)$; D_0^2 es la devianza del modelo reducido: $g(\mu_0) = X_p\beta_p$ y D_1^2 es la devianza del modelo completo: $g(\mu_1) = X_p\beta_p + X_q\beta_q$. Entre mayor sea la diferencia entre $\hat{\mu}_0$ y $\hat{\mu}_1$ mayor será el valor de Ω . Bajo ciertas condiciones, Jennings(1986), Ω se distribuye asintóticamente como una ji-cuadrada con q grados de libertad.

La estadística de puntaje se define como:

$$S^2 = \frac{[\dot{e}(\beta)]' X}{a(\phi)} \left\{ \frac{X' \ddot{E}(\beta) X}{a(\phi)} \right\}^{-1} \frac{X' \dot{e}(\beta)}{a(\phi)} \Bigg|_{\beta = \hat{\beta}_{H_0}}$$

$$\frac{1}{a(\phi)} [\dot{e}(\beta)]' X \{X' \bar{E}(\beta) X\}^{-1} X' \dot{e}(\beta) \Big|_{\beta = \hat{\beta}_{H_0}} \quad \text{con } \hat{\beta}_{H_0} = \begin{pmatrix} \hat{\beta}_{p_0} \\ 0 \end{pmatrix};$$

$\hat{\beta}_{p_0}$ es el estimador de β_p obtenido al ajustar el modelo reducido. Obsérvese que $a^{-1}(\phi) X' \dot{e}(\beta)$ es la función de puntaje para β y que $a^{-1}(\phi) X' \bar{E}(\beta) X$ es la matriz de información observada. Si H_0 es cierta, S^2 se distribuye asintóticamente como una ji-cuadrada con q grados de libertad. En el paquete GLIM3, la estadística de puntaje se calcula como (Pregibon, D.; 1982):

$$X_{\hat{\mu}_0}^2 - X_{\hat{\mu}_1}^2;$$

$X_{\hat{\mu}_0}^2$ es la ji-cuadrada de Pearson generalizada del modelo reducido y $X_{\hat{\mu}_1}^2$ es una aproximación a la ji-cuadrada de Pearson generalizada del modelo completo. Para calcular $X_{\hat{\mu}_1}^2$ se usa a $\hat{\beta}_{H_0}$ como el estimador inicial de β y se efectúa una iteración.

5.2 Intervalos de verosimilitud.

Sea $\lambda \equiv \lambda(\beta)$ una función real de β , i.e $\lambda: \mathbb{R}^m \rightarrow \mathbb{R}$. La función de verosimilitud relativa de λ se define como

$$R(\lambda) = \frac{L(\lambda)}{\sup_{\lambda} L(\lambda)}$$

en donde $L(\lambda)$ es la función de verosimilitud de λ . El intervalo de verosimilitud para λ al $100 \cdot \alpha$ % se define como:

$$\{\lambda; R(\lambda) \geq \alpha\}. \quad (5.2.1)$$

En la práctica, estos intervalos pueden calcularse graficando $R(\lambda)$ ó su logaritmo, $r(\lambda) = \log R(\lambda)$, contra λ .

El verdadero valor de λ , λ_0 , pertenece al intervalo (5.2.1) si y sólo si

$$-2r(\lambda_0) \leq -2\log \alpha.$$

La probabilidad de que suceda esto se conoce como la probabilidad de cobertura:

$$PC(\lambda_0) = P(D \leq -2\log\alpha) \quad \text{con} \quad D \equiv -2r(\lambda_0).$$

Para calcular esta probabilidad es necesario conocer la distribución de D (D es la estadística de cociente de verosimilitudes para probar $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$.) Si λ_0 es un punto interior del espacio parametral, la distribución asintótica de D es una ji-cuadrada con un grado de libertad. Por lo tanto,

$$PC(\lambda_0) = P(D \leq -2\log\alpha) \simeq P(\chi_{(1)}^2 \leq -2\log\alpha).$$

Bibliografía abreviada.

Kalbfleisch, J. G. (1985).

Pregibon, D. (1982).

6. Diagnósticos.

El proceso de ajustar un modelo a un conjunto de datos es iterativo. Primero, con base en la teoría y en estudios previos, se selecciona un modelo. Después, con ayuda de un paquete estadístico, se estima el vector de coeficientes y se calculan las medidas de bondad de ajuste. Si estas medidas indican que, en términos generales, el modelo es bueno se procede a realizar el análisis post-ajuste. El objetivo de este análisis es determinar si los supuestos del modelo son válidos e identificar a los datos que afectan a las conclusiones o que tienen características inesperadas. El análisis post-ajuste indica qué correcciones deben hacerse al modelo. Esto redundará en una mejor descripción de los datos y en predicciones más certeras.

En la práctica, el análisis post-ajuste no siempre se lleva a cabo. Quizás esto se debe al desconocimiento de las técnicas de diagnóstico post-ajuste y a que los paquetes estadísticos no las calculan automáticamente, excepto en regresión lineal normal. El objetivo de esta tesis es ayudar a subsanar estos problemas en el caso del modelo logístico.

6.1 Respuestas sin suavizar.

Los diagnósticos post-ajuste basados en variables de respuestas sin suavizar se clasificaron en tres grupos:

1. Diagnósticos para datos no agrupados, aplicables a modelos con respuestas binarias.
2. Diagnósticos para datos agrupados o no agrupados.
3. Diagnósticos para datos agrupados; en este caso la variable de respuesta es una binomial con $n_i > 1$.

La mayoría de las técnicas post-ajuste para el modelo logístico pertenecen al grupo dos.

6.1.1 Datos no agrupados.

Gráfica de la devianza media local.

Daniel y Wood(1971) propusieron una prueba de carencia de ajuste para el modelo de regresión lineal normal sin repeticiones. En 1984, Landwher *et. al.* la generalizaron al modelo de regresión logística para datos no agrupados.

Procedimiento. La prueba de carencia de ajuste para el modelo de regresión logística es gráfica. La idea es dividir la devianza en dos componentes: carencia de ajuste y error puro. El modelo presenta carencia de ajuste si el puro es pequeño comparado con la devianza. El error puro se calcula como sigue:

1. Formar G grupos ajenos con los n datos. Los autores recomiendan que los grupos sean chicos e internamente homogéneos. El tamaño del grupo g se denotará como n_g ; $g = 1, 2, \dots, G$.
2. Definir la matriz $Z = \{z_{ig}\}$ de $n \times G$ con

$$z_{ig} = \begin{cases} 1, & \text{si el dato } i \text{ pertenece al grupo } g; \\ 0, & \text{en otro caso} \end{cases}$$

3. Ajustar el modelo $\text{logit}(p^L) = X\beta + Z\gamma$; el vector p^L contiene las probabilidades de éxito asociadas a Y .
4. Calcular la devianza de cada grupo: $D_g^2 = \sum_{l=1}^{n_g} d_{lg}$; d_{lg} es el residuo de la devianza, elevado al cuadrado, del individuo l del grupo g . D_g^2 tiene $n_g - 1$ grados de libertad.
5. Ordenar los grupos de acuerdo a alguna medida de homogeneidad interna. La medida depende del método usado para agrupar a los individuos.
6. Calcular los estimadores aproximados de error puro:

$$\bar{D}^L(t) = \frac{\sum_{g=1}^t D_g^2}{\sum_{g=1}^t (n_g - 1)} \quad t = 1, 2, \dots, G.$$

A $\bar{D}^L(t)$ se le llama la devianza media local porque es la devianza promedio de los t grupos más homogéneos.

Después, para comparar el error puro con la devianza,:

1. Graficar $\bar{D}^L(t)$ contra sus grados de libertad: $\sum_{g=1}^t (n_g - 1)$.

2. Trazar, paralela al eje X , la recta $\bar{D} = \frac{D^2}{n-m}$. Esta es la devianza promedio del modelo: $\text{logit}(p) = X\beta$ y se le llama la devianza media global.

Generalmente, en los datos no agrupados existen pocos individuos con el mismo vector de variables explicativas. Como consecuencia, los grupos formados por individuos con la misma x_i tendrán pocos grados de libertad como para estimar la devianza media local. Por esto, los autores sugieren formar los grupos con datos próximos en el espacio de variables explicativas. Así, los grupos no son totalmente homogéneos pero se ganan grados de libertad para estimar a $\bar{D}^L(t)$. Existe el riesgo, con este método, de que $\bar{D}^L(t)$ no refleje únicamente la variabilidad local. Por lo tanto, debe haber un balance entre los grados de libertad y la homogeneidad interna de los grupos.

La métrica para medir la distancia entre las x_i y el método para agrupar a los individuos se determinan con base en los datos. Las recomendaciones que dan los autores son la siguientes:

1. Medir la distancia entre los individuos a partir de las variables explicativas que tengan un efecto lineal en $\text{logit}(p)$. Por ejemplo, usar únicamente a x_1 y x_2 si $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \log(x_1) + \beta_4 (x_1 x_2) + \beta_5 x_2^2$. Esto se hace con el fin de que la relación entre los grupos y el modelo sea mínima.
2. Estandarizar las variables explicativas antes de medir la distancia entre los individuos.
3. Usar la métrica Euclidiana y el método jerárquico de liga completa (del vecino más lejano).

El número de grupos depende de a qué nivel se corte el dendograma. Los autores sugieren que se corte cuando 5% de los individuos formen grupos unitarios.

Características. El modelo de regresión logística es adecuado si \bar{D} es aproximadamente igual al valor en el que se estabiliza $\bar{D}^L(t)$. Hay carencia de ajuste si \bar{D} es mayor que dicho valor y la causa puede ser la omisión de algún término de interacción.

El programa. En GLIM3, el macro DEVM (apéndice B1) hace la gráfica de la devianza media local. Antes de usarlo se realiza el agrupamiento con algún paquete estadístico adecuado y se ajusta el modelo global: $\text{logit}(p) = X\beta$. DEVM no tiene argumentos y se divide en tres submacros. En el primero, DATO, se teclean el número de grupos formados (%g) y el grupo al que pertenece cada individuo (g). Se considera

que el grupo más homogéneo (el primero en formarse) es el número 1, el segundo más homogéneo es el 2, etc. En el submacro LOCAL se ajusta el modelo $\text{logit}(p) = X\beta + Z\gamma$ y se calcula la devianza de cada grupo. Finalmente, en el submacro GRAF se obtienen las devianzas medias locales y se grafican contra sus respectivos grados de libertad.

Ejemplo. Al final de esta sección se presenta la gráfica de la devianza media local para los datos del apéndice A1. La única variable explicativa que se consideró en el modelo global fue 'CAL'. Se formaron 24 grupos con el método de liga completa, cuatro de ellos unitarios.

Los últimos (ocho) puntos en la gráfica muestran un valor estable de la devianza media local. La línea punteada, correspondiente a la devianza media global, pasa aproximadamente entre esos puntos. Entonces, aparentemente no existe carencia de ajuste, aunque la devianza media local se estabiliza hasta el final — cuando la homogeneidad interna de los grupos es menor.

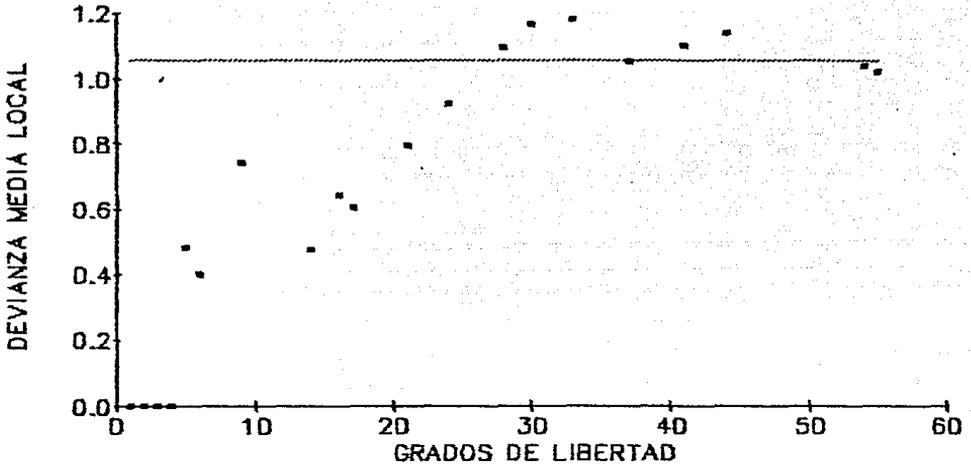
Comentarios. $\bar{D}^L(t)$ no estima insesgadamente al verdadero valor de la devianza porque depende de las probabilidades de éxito de los t grupos más homogéneos. Si estas probabilidades son cercanas a cero o uno, $\bar{D}^L(t)$ subestima al verdadero valor de la devianza. Esto puede provocar que, con base en la gráfica, se concluya erróneamente que existe carencia de ajuste. Cuando dichas probabilidades son próximas a 0.5, la devianza media local sobreestima al valor de la devianza. En este caso, la gráfica puede indicar incorrectamente que no existe carencia de ajuste. Para evitar este problema, Jennings (1986) recomienda usar los residuos $(y_{l_g} - \hat{p}_{l_g})^2$ en lugar de d_{l_g} ; y_{l_g} es la respuesta observada para el individuo l del grupo g y \hat{p}_{l_g} es su correspondiente probabilidad de éxito estimada.

El procedimiento analítico para determinar si existe carencia de ajuste en el modelo de regresión logística consiste en probar las hipótesis:

$$H_0 : \text{logit}(p) = X\beta \quad \text{vs.} \quad H_1 : \text{logit}(p) = X\beta + Z\gamma.$$

La estadística de prueba puede ser la de cociente de verosimilitudes o la de puntaje (véase la sección 5.1). Ambas se distribuyen asintóticamente como una $\chi^2_{(G'-1)}$; G' es el número de grupos con dos o más individuos.

GRAFICA DE LA DEVIANZA
MEDIA LOCAL



Bibliografía abreviada.

Daniel y Wood (1971).

Jennings, Dennis E. (1986).

Landwehr, J.M.; Pregibon, D. y Shoemaker, A. C. (1984).

Pregibon, D. (1982).

6.1.2 Datos agrupados o no agrupados.

Las técnicas post-ajuste descritas en esta sección son aplicables a cualquier modelo logístico, los datos pueden estar agrupados o no agrupados. Algunas técnicas se presentan en el contexto de los modelos lineales generalizados porque no son exclusivas del modelo logístico. Además varios diagnósticos son una generalización de los propuestos en regresión lineal normal.

Se explicará cómo hacer la gráfica de probabilidad empírica para el modelo logístico. Se indicará cómo determinar si una variable explicativa es significativa o si es necesario transformar una variable explicativa de un modelo lineal generalizado. Se verá qué es la colinealidad en un modelo lineal generalizado y cómo medirla. Se presentarán las medidas de influencia global propuestas por Pregibon (1981) para el modelo logístico. Se expondrá el método de influencia local desarrollado por Cook. Se explicará cómo probar la hipótesis de que el i -ésimo dato de un modelo lineal generalizado es discrepante. Asimismo, se indicará cómo probar si existe algún dato discrepante en la muestra.

6.1.2.1 Gráfica de probabilidad empírica.

En regresión lineal normal, las gráficas de probabilidad empírica y probabilidad normal indican si es válido el supuesto de normalidad para los errores. Landwehr, Pregibon y Shoemaker (1984) generalizaron la gráfica de probabilidad empírica al modelo logístico.

Procedimiento. El procedimiento para hacer la gráfica de probabilidad empírica es el siguiente:

1. Seleccionar algún residuo. Los autores trabajaron con los residuos de puntaje: $\hat{\epsilon}_i = y_i - n_i \hat{p}_i$. Estos residuos, ordenados de manera creciente se denotarán como: $\hat{\epsilon}_{(1)} \leq \hat{\epsilon}_{(2)} \leq \dots \leq \hat{\epsilon}_{(n)}$.
2. Simular la distribución de los residuos suponiendo que el modelo ajustado es correcto. La distribución de los residuos es desconocida y por esto se simula de la forma siguiente: para cada valor de k , con $k = 1, 2, \dots, K$,
 - a. Simular y_{ki}^* de la distribución Binomial(n_i, \hat{p}_i); $i = 1, 2, \dots, n$.
 - b. Ajustar el modelo logístico a $(y_{k1}^*, x'_{11}), (y_{k2}^*, x'_{21}), \dots, (y_{kn}^*, x'_{n1})$.

- c. Obtener los residuos $\hat{e}_{ki}^* = y_{ki}^* - n_i \hat{p}_{ki}^*$ y ordenarlos.
3. Graficar $\hat{e}_{(i)}$ contra los fractiles de la distribución simulada. Los pasos a seguir para hacer la gráfica son:
 - a. Calcular la mediana $T_{(i)}$ o algún otro valor típico de $\{\hat{e}_{1(i)}^*, \hat{e}_{2(i)}^*, \dots, \hat{e}_{K(i)}^*\}$; $\hat{e}_{k(i)}^*$ es la i -ésima estadística de orden de los residuos: $\hat{e}_{k1}^*, \hat{e}_{k2}^*, \dots, \hat{e}_{kn}^*$.
 - b. Graficar los puntos $(T_{(i)}, \hat{e}_{(i)})$; $i = 1, 2, \dots, n$.
 - c. Formar una envoltura para cada $(T_{(i)}, \hat{e}_{(i)})$. Para esto se seleccionan dos fractiles, uno superior y otro inferior, de $\{\hat{e}_{1(i)}^*, \hat{e}_{2(i)}^*, \dots, \hat{e}_{K(i)}^*\}$ y se grafican contra $T_{(i)}$.

A veces, el signo de $T_{(i)}$ es distinto al de $\hat{e}_{(i)}$. En este caso, $T_{(i)}$ se reemplaza por la mediana de aquellos residuos $\hat{e}_{k(i)}^*$ que tengan el mismo signo que $\hat{e}_{(i)}$. Landwher *et. al.* señalan que — de acuerdo con su experiencia — se requieren por lo menos $K = 25$ simulaciones para que las envolturas sean estables. La etapa que mayor tiempo consume es la de ajustar el modelo logístico a cada una de las K muestras simuladas. Para ahorrar tiempo puede usarse a $\hat{\beta}$ como el estimador inicial de β^* y detener el proceso de estimación después de una o dos iteraciones.

Características. El modelo logístico se ajusta a los datos cuando los puntos $(T_{(i)}, \hat{e}_{(i)})$ se aproximan a la recta $y = x$. En general, la envoltura no es simétrica respecto a $y = x$. Además, no tiene el mismo ancho para toda $T_{(i)}$. Aunque el modelo describa apropiadamente la relación entre Y y x puede haber puntos cerca de los límites de la envoltura. Es factible que el dato i sea discrepante cuando $(T_{(i)}, \hat{e}_{(i)})$ está fuera de su envoltura. (Los datos discrepantes pueden ser el resultado de registrar erróneamente el valor de y_i). Cuando los datos no están agrupados, la gráfica de probabilidad empírica tiene un hueco alrededor del origen. En este caso, la ausencia de puntos en otra región es indicio de que el modelo es inadecuado — posiblemente porque falta incluir alguna variable explicativa.

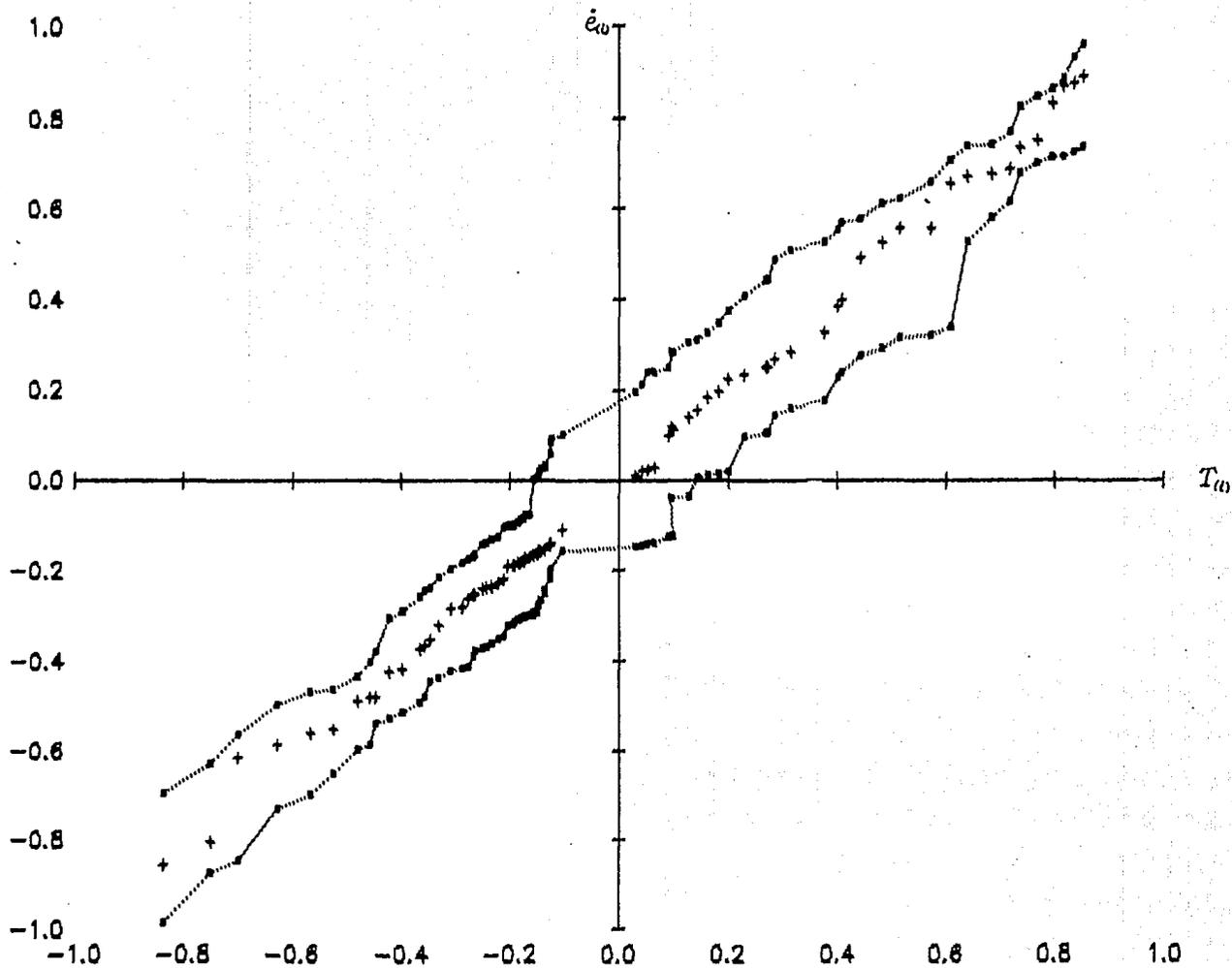
Programa. El macro ENVOL (apéndice B2), que sólo funciona en GLIM3, hace la gráfica de probabilidad empírica para datos no agrupados. Antes de usar este macro deben simularse 25 muestras independientes, cada una con n variables uniformes en el intervalo $(0, 1)$. Estas variables deben tener seis dígitos decimales y deben grabarse

verticalmente en un archivo para que ENVOL las lea correctamente. El modelo logístico debe ajustarse a los datos inmediatamente antes de usar el macro. Los argumentos de ENVOL son las probabilidades estimadas (\hat{p}_i) y los residuos ordenados ($\hat{e}_{(i)}$), en este orden.

A partir de las variables uniformes, ENVOL genera respuestas con distribución Bernoulli; después ajusta el modelo logístico a los datos simulados. El ajuste es lento porque no se aplicó la modificación mencionada arriba. Las envolturas se calculan con el máximo y mínimo de $\{\hat{e}_{1(i)}^*, \hat{e}_{2(i)}^*, \dots, \hat{e}_{K(i)}^*\}$ $i = 1, 2, \dots, n$.

Ejemplo. A partir de los datos del apéndice A1 se hizo la gráfica de probabilidad empírica que está al final de esta sección. La única variable explicativa incluida en el modelo fue 'CAL'. Los puntos marcados con '+' se aproximan a la recta $y = x$ y ninguno de ellos está fuera de su envoltura. Esto hace pensar que el modelo logístico es adecuado.

Comentarios. También hay gráficas de probabilidad normal para el modelo logístico, véase Pregibon (1981) y Davison, A. C. (1989). Estas gráficas se hacen ordenando los residuos de Pearson o de devianza y graficándolos contra los fractiles de la distribución normal estandarizada.



Bibliografía abreviada.

Davison, A. C. (1989).

Landwehr, J. M.; Pregibon, D. y Shoemaker, A. C. (1984).

Pregibon, D. (1981).

Wichmann, B. A. y Hill, I. D. (1982).

6.1.2.2 Gráfica de la variable agregada.

Hay dos maneras de saber si una variable explicativa es estadísticamente significativa. Una consiste en realizar una prueba de hipótesis y la otra en hacer la gráfica de la variable agregada. Estos dos procedimientos están relacionados. Sin embargo, la gráfica, además de indicar si la variable es significativa, revela qué datos influyen en el resultado de la prueba de hipótesis.

Suponga que en el modelo lineal generalizado: $g(\mu_1) = X\beta + v\gamma$ se desea probar si v es significativa, es decir:

$$H_0 : g(\mu_0) = X\beta \quad \text{vs.} \quad H_1 : g(\mu_1) = X\beta + v\gamma.$$

Esta hipótesis es un caso particular de la mencionada en la sección 5.1. La estadística de prueba puede ser la de puntaje o la de cociente de verosimilitudes. En términos del vector de trabajo y de la matriz de pesos, las hipótesis anteriores son equivalentes a:

$$H_0 : W_{H_0}^{\frac{1}{2}} z_{H_0} = W_{H_0}^{\frac{1}{2}} X\beta + \varepsilon_{H_0}$$

vs.

$$H_1 : W_{H_1}^{\frac{1}{2}} z_{H_1} = W_{H_1}^{\frac{1}{2}} X\beta + W_{H_1}^{\frac{1}{2}} v\gamma + \varepsilon_{H_1};$$

$W_{H_0}^{\frac{1}{2}}$ y z_{H_0} dependen de $\hat{\beta}_{H_0}$ mientras que $W_{H_1}^{\frac{1}{2}}$ y z_{H_1} dependen de $\hat{\beta}_{H_1}$ y $\hat{\gamma}$.

Una expresión equivalente para el modelo postulado en la última hipótesis alternativa es:

$$W_{H_1}^{\frac{1}{2}} z_{H_1} = W_{H_1}^{\frac{1}{2}} X_a \beta_a + \varepsilon_{H_1}, \quad (6.1.2.2.1)$$

$$\text{con } X_a = (X, v) \text{ y } \beta_a = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}.$$

Si la expresión (6.1.2.2.1) se premultiplica por $(I - H_a)$ con $H_a = W_{H_1}^{\frac{1}{2}} X_a (X_a' W_{H_1} X_a)^{-1} X_a' W_{H_1}^{\frac{1}{2}}$ se obtiene el modelo:

$$(I - H_a)W_{H_1}^{\frac{1}{2}}z_{H_1} = (I - H_a)\varepsilon_{H_1},$$

que carece de componente sistemática. En cambio, si (6.1.2.2.1) se premultiplica por $(I - H_n)$ con $H_n = W_{H_0}^{\frac{1}{2}}X(X'W_{H_0}X)^{-1}X'W_{H_0}^{\frac{1}{2}}$ se obtiene el modelo:

$$(I - H_n)W_{H_1}^{\frac{1}{2}}z_{H_1} = (I - H_n)W_{H_1}^{\frac{1}{2}}X_a\beta_a + (I - H_n)\varepsilon_{H_1},$$

que contiene más de una variable explicativa. Entonces, a partir de (6.1.2.2.1) es imposible reproducir el resultado obtenido en regresión lineal normal: un modelo sin ordenada al origen y con una variable explicativa.

Wang(1985), en lugar de usar la expresión (6.1.2.2.1) trabajó con la aproximación:

$$W_{H_0}^{\frac{1}{2}}z_{H_0} \simeq W_{H_0}^{\frac{1}{2}}X\beta + W_{H_0}^{\frac{1}{2}}v\gamma + \varepsilon. \quad (6.1.2.2.2)$$

Esta expresión premultiplicada por $(I - H_n)$ es igual a:

$$(I - H_n)W_{H_0}^{\frac{1}{2}}z_{H_0} \simeq (I - H_n)W_{H_0}^{\frac{1}{2}}v\gamma + (I - H_n)\varepsilon.$$

De aquí que Wang sugiriera graficar en el eje vertical las componentes del vector $(I - H_n)W_{H_0}^{\frac{1}{2}}z_{H_0}$ y en el horizontal las de $(I - H_n)W_{H_0}^{\frac{1}{2}}v$. Sin embargo, la información proporcionada por esta gráfica de la variable agregada puede ser errónea. La causa es que, en general, la aproximación (6.1.2.2.2) no es válida.

O'Hara y Carter (1990) propusieron otra gráfica de la variable agregada. Esta gráfica está basada en la expresión del estimador de γ para el modelo: $g(\mu_1) = X\beta + v\gamma$:

$$\hat{\gamma} = \frac{v'W_{H_1}^{\frac{1}{2}}(I - H_a)W_{H_1}^{\frac{1}{2}}z_{H_1}}{v'W_{H_1}^{\frac{1}{2}}(I - H_a)W_{H_1}^{\frac{1}{2}}v}$$

Los autores mencionados sugirieron graficar:

$$(I - H_a)W_{H_1}^{\frac{1}{2}}z_{H_1} \quad \text{contra} \quad (I - H_a)W_{H_1}^{\frac{1}{2}}v.$$

Esta gráfica indica si v es significativa y si existen datos que influyen en la prueba de hipótesis mencionada al principio de la sección.

Interpretación. Existe evidencia de que v es significativa cuando los puntos graficados tienen un comportamiento lineal con pendiente distinta de cero.

Cómputo. Las entradas de $(I - H_a)W_{H_1}^{\frac{1}{2}}z_{H_1}$ son los residuos de Pearson del modelo postulado en H_1 porque $W_{H_1}^{\frac{1}{2}}z_{H_1} = W_{H_1}^{-\frac{1}{2}}\hat{\epsilon}_{H_1} + W_{H_1}^{\frac{1}{2}}X\hat{\beta} = W_{H_1}^{-\frac{1}{2}}\hat{\epsilon}_{H_1} + H_aW_{H_1}^{\frac{1}{2}}z_{H_1}$. El vector $(I - H_a)W_{H_1}^{\frac{1}{2}}v$ contiene los residuos del modelo de regresión lineal normal:

$$W_{H_1}^{\frac{1}{2}}v = W_{H_1}^{\frac{1}{2}}X\beta + \epsilon \quad \text{con } \epsilon \sim N_n(0, \sigma^2 I).$$

Ejemplo. Los datos analizados en esta sección se encuentran en el apéndice A2. El modelo propuesto originalmente fue:

$$\text{logit}(p) = \beta_0 + \beta_1 \cdot \log(\text{vol}) + \beta_2 \cdot \log(\text{tas})$$

que tiene una devianza de 29.227 con 36 grados de libertad. A continuación se probó si $\log(\text{vol})$ es significativa es decir:

$$H_0 : \text{logit}(p) = \alpha_0 + \alpha_1 \cdot \log(\text{tas})$$

vs.

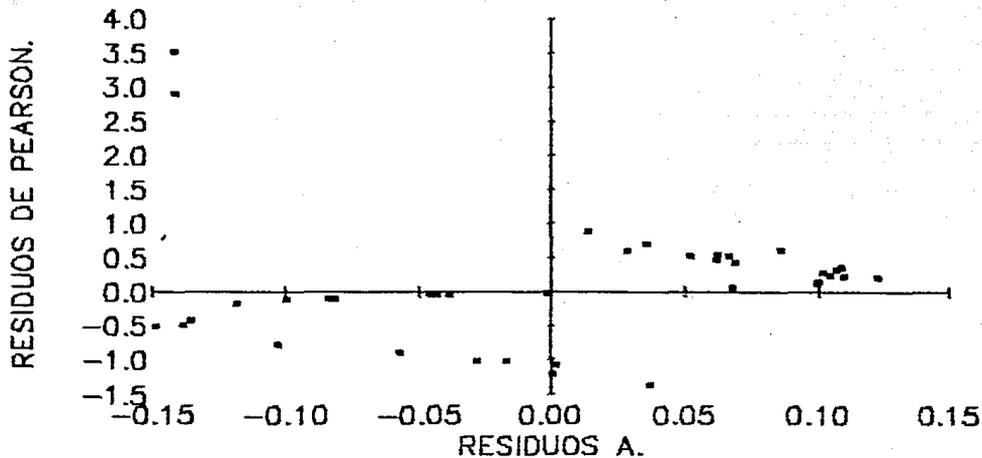
$$H_1 : \text{logit}(p) = \beta_0 + \beta_1 \cdot \log(\text{vol}) + \beta_2 \cdot \log(\text{tas}).$$

Las estadísticas de prueba: $S^2 = 14.64$ y $\Omega = -2 \ln \Lambda = 19.63$ son mayores que el valor en tablas de una ji-cuadrada con un grado de libertad. Por lo tanto, $\log(\text{vol})$ es significativa. La gráfica de la variable agregada para $\log(\text{vol})$ está al final de la sección. En ella puede observarse que todos los puntos, excepto dos, se encuentran alrededor del eje X. Esto quiere decir que la significancia de $\log(\text{vol})$ está determinada por esos dos puntos: datos 4 y 18. Lo mismo sucede al probar las hipótesis:

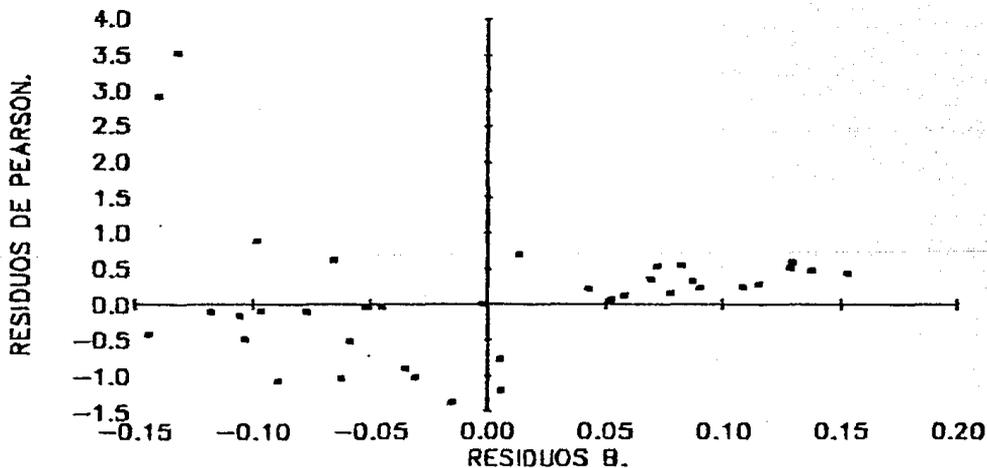
$$H_0 : \text{logit}(p) = \delta_0 + \delta_1 \cdot \log(\text{vol}) \quad \text{vs.} \quad H_1 : \text{logit}(p) = \beta_0 + \beta_1 \cdot \log(\text{vol}) + \beta_2 \cdot \log(\text{tas})$$

— véase la segunda gráfica.

G. VARIABLE AGREGADA.
LOG(VOL).



G. VARIABLE AGREGADA.
LOG(TAS).



Bibliografía abreviada.

O'Hara Hines, R. J. y Carter, E. M. (1990).

Wang, P. C. (1985).

6.1.2.3 Gráfica de la variable construida.

Las gráficas de residuos parciales y de la variable construida ayudan a determinar si es necesario transformar alguna variable explicativa. En esta sección sólo se explica cómo hacer la segunda gráfica. La primera se comenta en la sección correspondiente a los diagnósticos para respuestas suavizadas.

Considere el modelo lineal generalizado:

$$g(\mu) = \beta_0 + \sum_{j=1}^{m-2} \beta_j x_j + \beta_{m-1} x_{m-1}^{(\lambda_{m-1})} \quad (6.1.2.3.1)$$

en donde $x_{m-1}^{(\lambda_{m-1})}$ es una transformación de x_{m-1} . Se supondrá que $x_{m-1}^{(\lambda_{m-1})} = x_{m-1}$ si $\lambda_{m-1} = 1$. La gráfica de la variable construida está relacionada con el juego de hipótesis:

$$H_0 : \lambda_{m-1} = 1 \quad \text{vs.} \quad H_1 : \lambda_{m-1} \neq 1.$$

En los modelos lineales generalizados, el procedimiento para hacer la gráfica de la variable construida es similar al que se usa en regresión lineal normal. La expansión en series de Taylor de (6.1.2.3.1) alrededor de $\lambda_{m-1} = 1$ es, hasta el término de primer orden,

$$g(\mu) \simeq \beta_0 + \sum_{j=1}^{m-1} \beta_j x_j + \alpha v_0 \quad (6.1.2.3.2)$$

$$\text{con} \quad \alpha = \beta_{m-1}(\lambda_{m-1} - 1) \quad \text{y} \quad v_0 = \left. \frac{\partial x_{m-1}^{(\lambda_{m-1})}}{\partial \lambda_{m-1}} \right|_{\lambda_{m-1}=1}$$

Si H_0 es cierta, la variable construida v_0 no es significativa y $g(\mu)$ depende linealmente de x_{m-1} . En términos del modelo aproximado (6.1.2.3.2), las hipótesis mencionadas son equivalentes a:

$$H_0 : g(\mu) = \beta_0 + \sum_{j=1}^{m-1} \beta_j x_j \quad (\alpha = 0)$$

vs.

$$H_1 : g(\mu) \simeq \beta_0 + \sum_{j=1}^{m-1} \beta_j x_j + \alpha \cdot v_0 \quad (\alpha \neq 0).$$

Entonces, el problema de determinar si es necesario transformar a x_{m-1} se redujo a probar la significancia de v_0 . Siguiendo a O'Hara y Carter (1990), en la gráfica de la variable construida, el eje vertical corresponde a $(I - H_a)W_{H_1}^{\frac{1}{2}} z_{H_1}$ y el horizontal a $(I - H_a)W_{H_1}^{\frac{1}{2}} v_0$; W_{H_1} y z_{H_1} estiman, respectivamente, a la matriz de pesos y a la variable de trabajo del modelo (6.1.2.3.2).

Cuando se rechaza H_0 , el estimador

$$\hat{\lambda}_{m-1} = 1 + \frac{\hat{\alpha}}{\hat{\beta}_{m-1}}$$

ayuda a establecer cuál es la transformación más apropiada para x_{m-1} . En la expresión anterior, $\hat{\beta}_{m-1}$ y $\hat{\alpha}$ estiman a los parámetros β_{m-1} y α de los modelos reducido y completo respectivamente. En la práctica, no tiene sentido usar estimadores de λ_{m-1} fuera del intervalo $[-2, 2]$.

A menudo, $x_{m-1}^{(\lambda_{m-1})}$ es la familia de transformaciones de Box y Cox que se define como:

$$x_{m-1}^{(\lambda_{m-1})} = \begin{cases} \frac{x_{m-1}^{(\lambda_{m-1})} - 1}{x_{m-1}}, & \text{si } \lambda_{m-1} \neq 0; \\ \log(x_{m-1}), & \text{si } \lambda_{m-1} = 0. \end{cases}$$

La variable construida es, en este caso, $v_0 = x_{m-1} \log(x_{m-1})$ (el término x_{m-1} se omite deliberadamente de v_0 porque ya aparece en la expresión (6.1.2.3.2)).

Ejemplo. El objetivo de este ejercicio es determinar qué relación existe entre la edad y la probabilidad de que una varsoviaña haya tenido la menarquía. El análisis se llevó a cabo con los datos del apéndice A3 que están agrupados. Originalmente se ajustó el modelo:

$$\text{logit}(p_i) = \beta_0 + \beta_1 \cdot \text{edad}_i \quad \text{con} \quad i = 1, 2, \dots, n.$$

La devianza obtenida: 26.703 tiene 23 grados de libertad. Su nivel de significancia descriptivo, 0.7 aproximadamente, indica que en general el modelo es adecuado. Sin embargo, se decidió investigar si era necesario transformar la variable explicativa. La hipótesis:

$$H_0 : \lambda = 1 \quad \text{vs.} \quad H_1 : \lambda \neq 1$$

se probó para la familia de transformaciones de Box y Cox. Otra manera de escribir las hipótesis anteriores es:

$$H_0 : \text{logit}(p) = \beta_0 + \beta_1 \cdot \text{edad}$$

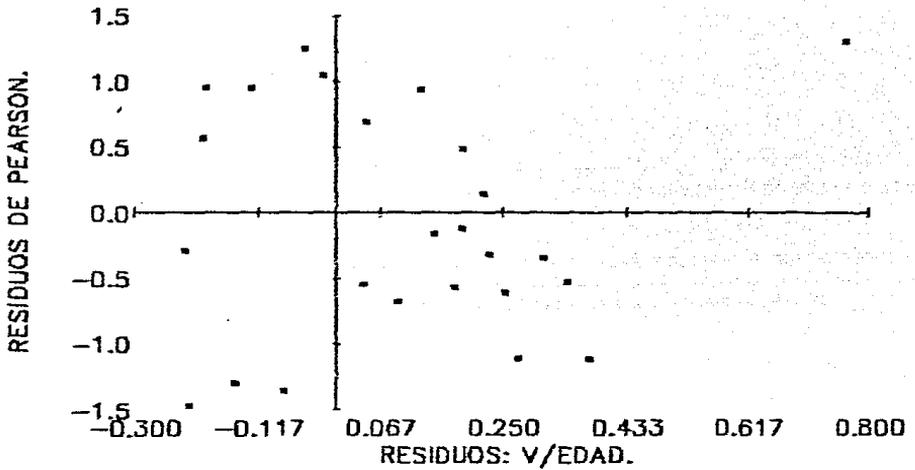
vs.

$$H_1 : \text{logit}(p) = \delta_0 + \delta_1 \cdot \text{edad} + \phi \cdot x_2$$

con $\phi = \beta_1(\lambda - 1)$ y $x_2 = \text{edad} \cdot \log(\text{edad})$. El valor de la estadística de puntaje fue 4.196 y el del cociente de verosimilitudes 4.076. El correspondiente nivel de significancia descriptivo es ligeramente menor a 0.05. La gráfica de la variable construida está al final de la sección. En ella se observa que todos los puntos, excepto uno, tienen la misma tendencia: lineal y decreciente. Sin considerar el punto situado en el extremo superior dserecho de la gráfica (dato 25), la variable construida no parece ser significativa. Sin embargo, esta conclusión puede variar cuando se toma en cuenta al dato 25. En la sección 6.1.3.3 se analizará la influencia de dicho dato en la prueba de hipótesis.

Como $\hat{\phi} = -1.656$ y $\hat{\beta}_1 = 1.632$ entonces $\hat{\lambda} \simeq -0.015$. Este valor sugiere que posiblemente la transformación logarítmica sea la adecuada (teniendo en mente la posible influencia del dato 25).

GRAFICA DE LA VARIABLE
CONSTRUIDA.



Bibliografía abreviada.

O'Hara Hines, R. J. y Carter, E. M. (1990).

Wang, P. C. (1987).

6.1.2.4 Colinealidad.

Se dice que hay colinealidad en un modelo lineal generalizado si existe una asociación lineal fuerte entre las columnas de $W^{\frac{1}{2}}X$. Los valores propios de $X'WX$ son la herramienta básica para estudiar dicha asociación. Si $W^{\frac{1}{2}}X$ es de rango completo (m), la matriz $X'WX$ tiene m valores propios distintos de cero: $\lambda_1, \lambda_2, \dots, \lambda_m$. Entre más chico sea el menor de ellos mayor será la asociación lineal entre las columnas de $W^{\frac{1}{2}}X$. Sin embargo, la magnitud de λ_j también depende de las unidades de medición de las variables explicativas. Esta dependencia desaparece — dicen Mackinnon y Puterman (1989) — al estandarizar las columnas de X . Los autores no indican cómo estandarizar la j -ésima variable explicativa; mi recomendación es:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_{jw}}{S_{jw}}, \quad j = 0, 1, 2, \dots, m-1;$$

$$\bar{x}_{jw} = \frac{\sum_{i=1}^n w_i x_{ij}}{\sum_{i=1}^n w_i} \quad \text{es la media ponderada de las } x_{ij},$$

$$S_{jw}^2 = \sum_{i=1}^n w_i (x_{ij} - \bar{x}_{jw})^2$$

y w_i es el i -ésimo elemento en la diagonal de W , la matriz de pesos.

Si las columnas de la matriz X^* , de $n \times (m-1)$, son $(x_1^*, x_2^*, \dots, x_{m-1}^*)$ y $W^{\frac{1}{2}}$ es la matriz de pesos del modelo $g(\mu) = X\beta$, los valores propios de $W^{\frac{1}{2}}X^*$ se denotarán como: $\lambda_1^* > \lambda_2^* > \dots > \lambda_{m-1}^*$. El k -ésimo índice de condición de $W^{\frac{1}{2}}X^*$ se define como:

$$\eta_k^* = \frac{\lambda_1^*}{\lambda_k^*} \quad k = 2, 3, \dots, m-1.$$

A $\eta_{m-1}^* = \lambda_1^*/\lambda_{m-1}^*$ se le llama el número de condición de $W^{\frac{1}{2}}X^*$ e indica qué tan grande es λ_1^* en relación a λ_{m-1}^* . En regresión lineal normal se dice que la asociación lineal entre las variables explicativas es fuerte o moderada si $100 < \eta_{m-1}^* < 1000$ y grave si $\eta_{m-1}^* > 1000$.

El efecto que tiene la colinealidad en el estimador de β_j se mide con el j -ésimo factor de inflación de la varianza:

$$FIV_j^* = \frac{\sum_{r=1}^{m-1} v_{jr}^2}{\lambda_r^*} \quad j = 1, 2, \dots, m-1;$$

$v_j^* = (v_{j1}, v_{j2}, \dots, v_{jm-1})$ es el vector propio asociado a λ_j^* . Los factores de inflación de la varianza son los elementos en la diagonal de $(X^*W X^*)^{-1}$.

Si se desea, los índices de condición y los factores de inflación de la varianza pueden calcularse con los valores y vectores propios de $X^*W X^*$.

Ejemplo. El modelo ajustado a los datos del apéndice A2 fue

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{6.1.2.4}$$

con x_1 y x_2 el logaritmo del volumen y de la proporción de aire aspirado respectivamente. El número de condición de $W^{\frac{1}{2}} X^* : \eta_1^* = 0.107233/0.016903 = 6.344$ indica que no existe colinealidad entre x_1 y x_2 . Los factores de inflación de la varianza fueron $FIV_1^* = 26.22$ y $FIV_2^* = 42.21$. El número de condición de $W^{\frac{1}{2}} X$ es $\eta_1 = 6.569307/0.129794 = 50.613$ y los factores de inflación de la varianza son: $FIV_1 = 1.739$, $FIV_2 = 3.466$ y $FIV_3 = 3.368$.

Bibliografía abreviada.

Mackinnon, M. J. y Puterman, M. L. (1989).
Walker, Esteban (1989).

6.1.2.5 Influencia.

Es conveniente evaluar la influencia de cada dato en el modelo ajustado cuando el método de estimación no es robusto. Hay dos maneras de evaluar la influencia del dato i . Una es excluirlo de la muestra y la otra es perturbarlo por una cantidad infinitesimal. Ambos casos pueden tratarse de manera similar: modificando el peso inicial del dato. Normalmente, el peso inicial es unitario. Cuando el peso es nulo, el dato queda fuera del análisis y si el peso pertenece al intervalo $(0, 1)$ sólo se le perturba. Se dice que un dato tiene influencia global en el modelo si su exclusión afecta el ajuste. Un dato es localmente influyente si la perturbación inducida altera el ajuste.

La sección 6.1.2.5.1 trata de la influencia global de un dato en: 1). el nivel de colinealidad de un modelo lineal generalizado y 2). el estimador de β , las medidas de bondad de ajuste y los residuos de la devianza del modelo logístico. En la sección 6.1.2.5.2 se verá el tema de influencia local.

6.1.2.5.1 Influencia global.

Influencia global en el nivel de colinealidad.

Esteban Walker (1989) definió lo que significa que x_i influya en la colinealidad de un modelo de regresión lineal normal. También indicó cómo medir el efecto de x_i en la colinealidad de dicho modelo. La propuesta de Walker es aplicable a los modelos lineales generalizados si la matriz X se reemplaza por $W^{\frac{1}{2}}X$ ó $W^{\frac{1}{2}}X^*$.

El j -ésimo valor propio de $W^{\frac{1}{2}}X$ será λ_j y el correspondiente vector propio será v_j . El k -ésimo índice de condición de $W^{\frac{1}{2}}X$ se denotará como η_k . Se supondrá que $X_{(i)}$, de $(n-1) \times m$, es la matriz de diseño sin el renglón i . $W_{(i)}^{\frac{1}{2}}$ denotará a la matriz de pesos obtenida al ajustar el modelo $g(\mu_{(i)}) = X_{(i)}\beta$. Finalmente, $\eta_{k(i)}$ será el k -ésimo índice de condición de $W_{(i)}^{\frac{1}{2}}X_{(i)}$.

Definición. Se dirá que $\sqrt{w_i}x_i$ influye en la colinealidad de un modelo lineal gene-

realizado si, para δ fija, se cumple que:

$$|\eta_k - \eta_{k(i)}| > \delta \eta_k \quad \text{para alguna } k = 2, 3, \dots, m-1.$$

Un punto que influye en la colinealidad la enmascara si $\eta_k \ll \eta_{k(i)}$ y la induce si $\eta_k \gg \eta_{k(i)}$.

A partir de la descomposición en valor singular de $W^{\frac{1}{2}}X$, UDV' , puede demostrarse que:

1. Si $\sqrt{w_i}x_i = \alpha v_j$ con α que pertenece a los reales, entonces $|\eta_k - \eta_{k(i)}|$ es una función creciente de las palancas de Pregibon.
2. Las palancas de Pregibon pueden expresarse como $h_i = \|u_i\|^2 = \sum_{j=1}^m u_{ij}^2$; u_i es el i -ésimo renglón de U .

La igualdad $h_i = u_{ij}^2$ se cumple cuando $\sqrt{w_i}x_i = \alpha v_j$.

El supuesto de que $\sqrt{w_i}x_i = \alpha v_j$ para alguna $\alpha \in \mathbb{R}$ es irreal. Sin embargo, en la práctica puede suceder que la dirección de los vectores: $\sqrt{w_i}x_i$ y v_j ; sea casi igual. En este caso, el valor de h_i estará básicamente determinado por u_{ij}^2 . Entonces, las u_{ij}^2 miden la influencia de $\sqrt{w_i}x_i$ en la colinealidad de los modelos lineales generalizados. Se considerará, empíricamente, que u_{ij}^2 es grande si su valor es mayor que m/n (una cota empírica para h_i es $2m/n$; recuérdese que $\sum_{i=1}^n h_i = m$).

A veces, la influencia individual de $\sqrt{w_i}x_i$ es irrelevante, pero conjuntamente con otros puntos puede tener un efecto importante en la colinealidad del modelo. Esto muestra la necesidad de generalizar los resultados anteriores. Como primer paso puede suponerse — como lo hizo Walker en regresión lineal normal — que

$$\sqrt{w_i}x_i = \alpha v_j \quad \text{y} \quad \sqrt{w_i}x_{i'} = \beta v_j \quad \text{con } \alpha \text{ y } \beta \in \mathbb{R}$$

6

$$\sqrt{w_i}x_i = \alpha v_j \quad \text{y} \quad \sqrt{w_{i'}}x_{i'} = \beta v_j \quad \text{con } \alpha \text{ y } \beta \in \mathbb{R}.$$

Los índices u_{ij}^2 también pueden calcularse a partir de la descomposición en valor singular de $W^{\frac{1}{2}}X^*$. En este caso debe investigarse si la cota m/n es útil (las cantidades $w_i(x_i^*)'(X^{*'}WX^*)^{-1}x_i^*$ reemplazan a las palancas de Pregibon).

Cómputo. El programa "SINGULAR" (apéndice B3) utiliza la subrutina "LVSDF" de IMSL para calcular la descomposición en valor singular de una matriz A y las u_{ij}^2 . Los datos, que son el número de renglones y columnas de A así como sus entradas, deben estar en un archivo llamado SING.DAT. SINGULAR calcula los valores singulares de A , los valores propios, los índices de condición y las u_{ij}^2 . Estas cantidades, junto con los datos, se graban en el archivo SING.SAL.

Ejemplo. Se analizará si existen puntos que influyen en la colinealidad del modelo (6.1.2.4). Los índices u_{ij}^2 se calcularon con la descomposición en valor singular de $W^{\frac{1}{2}}X$. Varios de ellos exceden la cota empírica: $m/n = 0.0769$; los mayores son $u_{29,2}^2 = 0.129184$ y $u_{31,2}^2 = 0.229566$. Sin embargo, la exclusión individual o conjunta de los datos 29 y 31 no tiene un efecto relevante en el estimador de β . En cada uno de estos casos,

$$\max_k |\eta_k - \eta_{k(29)}| = 3.06,$$

$$\max_k |\eta_k - \eta_{k(31)}| = 1.01,$$

$$\max_k |\eta_k - \eta_{k(29,31)}| = 5.5.$$

Por otro lado, las palancas de Pregibon de los puntos 4 y 18 son $h_4 = 0.0854$ y $h_{18} = 0.094$ respectivamente; además $u_{4,3}^2 = 0.074349$ y $u_{18,3}^2 = 0.80448$. El paquete GLIM3 tiene problemas para ajustar el modelo sin los datos 4 y 18, aparentemente porque el modelo está sobreparametrizado. Esto no sucede cuando se elimina sólo uno de dichos datos.

También se calcularon las $(u_{ij}^*)^2$ con la descomposición en valor singular de $W^{\frac{1}{2}}X^*$. Como en el caso anterior, el mayor índice corresponde al dato 31. De acuerdo a las $(u_{ij}^*)^2$, el efecto de los puntos 4, 18 y 29 es menor de lo que señalaban las u_{ij}^2 .

Bibliografía abreviada.

Mackinnon, M. J. y Puterman, M. L. (1989).

Walker, Esteban (1989).

Estimadores a un paso.

Hay varias maneras de medir la influencia global del dato i en los estimadores de β y de la devianza. Una de ellas es estimar a β con los n datos, eliminar el i -ésimo y estimar a $\beta_{(i)}$ (el vector de coeficientes del modelo sin el dato (y_i, x_i)). Este procedimiento requiere de dos ajustes con el método de Newton-Raphson. Si el análisis se realiza para cada uno de los n datos, el número total de ajustes (iterativos) es $n+1$.

Cuando existe una expresión explícita para $\hat{\beta}_{(i)}$ en términos de $\hat{\beta}$ es suficiente con estimar a β ; es decir, basta hacer un ajuste. Sin embargo, esto sólo es factible cuando el método para estimar al modelo con $n-1$ datos converge en una iteración.

El método propuesto por Pregibon (1981) consiste en usar a $\hat{\beta}_{(i)}^1$ en lugar de $\hat{\beta}_{(i)}$. A $\hat{\beta}_{(i)}^1$ se le llama: "el estimador a un paso de $\beta_{(i)}$ " y es una aproximación a $\hat{\beta}_{(i)}$. $\hat{\beta}_{(i)}^1$ se calcula como sigue:

1. Eliminar el dato i ;
2. Usar como estimador inicial de $\beta_{(i)}$ a $\hat{\beta}$; y
3. Detener el procedimiento de ajuste cuando finalice la primera iteración.

En GLIM3, $\hat{\beta}_{(i)}^1$ se calcula mediante los comandos \$WEI y \$REC 1. Por ejemplo, suponga que se desea estudiar la influencia global del dato 2 en el estimador de β . Si la variable de respuesta del modelo logístico es Y , el denominador binomial es N y las variables explicativas son X_1, X_2, \dots, X_M , las instrucciones para calcular el estimador a un paso de $\beta_{(2)}$ son (en GLIM3):

```
$YVAR Y
$ERROR B N
$LINK G
$FIT X1 + X2 + ... + XM
$DIS M E
$CAL NUM = %CU(1): PESO = 1 - %EQ(NUM,2)
$WEI PESO
$REC 1
$FIT .
$DIS M E
```

Todas las entradas del vector "PESO" son iguales a uno excepto la segunda que es

cero. El estimador de máxima verosimilitud de $\beta_{(2)}$, i.e. $\hat{\beta}_{(2)}$, se obtiene omitiendo la instrucción §REC 1.

El estimador a un paso de $\beta_{(i)}$ puede expresarse explícitamente en términos de $\hat{\beta}$. Esto permite evaluar la influencia de (y_i, x_i') para $i = 1, 2, \dots, n$ a partir del modelo ajustado a los n datos.

El logaritmo de la función de verosimilitud de $\beta_{(i)}$ es:

$$L(\beta_{(i)}) = \sum_{l=1}^n v_l l_l; \quad (6.1.2.5.1.2.1)$$

$$\text{con } l_l = \log f_{Y_l}(y_l; \theta_l, \phi) \text{ y } v_l = \begin{cases} 1, & \text{si } l \neq i; \\ 0, & \text{si } l = i. \end{cases}$$

En forma matricial, las ecuaciones de verosimilitud son:

$$0 = \frac{1}{a(\phi)} X' V \dot{\epsilon}(\beta_{(i)}) \Big|_{\beta_{(i)} = \hat{\beta}_{(i)}} \quad \text{con } V = \text{diag}\{v_1, v_2, \dots, v_n\}.$$

(Compárese este sistema de ecuaciones con (3.1.2)). Si se usa el método modificado de Newton-Raphson, el estimador de $\beta_{(i)}$ en la iteración $t + 1$ es:

$$\beta_{(i)}^{t+1} = \beta_{(i)}^t + (X' W^{\frac{1}{2}}(\beta_{(i)}^t) V W^{\frac{1}{2}}(\beta_{(i)}^t) X)^{-1} X' V \dot{\epsilon}(\beta_{(i)}^t).$$

El estimador a un paso de $\beta_{(i)}$ se obtiene haciendo $t = 0$ y $\beta_{(i)}^0 = \hat{\beta}$; es decir:

$$\hat{\beta}_{(i)}^1 = \hat{\beta} + (X' W^{\frac{1}{2}}(\hat{\beta}) V W^{\frac{1}{2}}(\hat{\beta}) X)^{-1} X' V \dot{\epsilon}(\hat{\beta}).$$

Otra expresión para $\hat{\beta}_{(i)}^1$ es (véase Pregibon (1981)):

$$\hat{\beta}_{(i)}^1 = \hat{\beta} - \frac{(X' W(\hat{\beta}) X)^{-1} x_i \dot{\epsilon}_i(\hat{\beta})}{1 - h_i}; \quad (6.1.2.5.1.2.2)$$

en donde h_i es la i -ésima palanca de Pregibon. En el modelo logístico, $W(\hat{\beta})$ es una matriz diagonal con elementos $w_i = n_i \hat{p}_i (1 - \hat{p}_i)$, $\dot{\epsilon}_i(\hat{\beta}) = y_i - n_i \hat{p}_i$ y $h_i = w_i x_i' (X' W(\hat{\beta}) X)^{-1} x_i$.

A continuación se verá cómo medir, en el modelo logístico, la influencia global del dato i en $\hat{\beta}$, D^2 , X^2 y los residuos de la devianza. Estas medidas, basadas en $\hat{\beta}_{(i)}^1$, subestiman a los datos influyentes y pueden generalizarse a otros modelos lineales generalizados (véase Williams, D. A. (1987)).

a. Influencia global de un dato en $\hat{\beta}$. El vector

$$\Delta_i \hat{\beta}^1 \equiv \hat{\beta} - \hat{\beta}_{(i)}^1 = \frac{(X'W(\hat{\beta})X)^{-1}x_i \hat{e}_i(\hat{\beta})}{1 - h_i}, \tag{6.1.2.5.1.2.3}$$

obtenido a partir de la expresión (6.1.2.5.1.2.2), mide las discrepancias entre $\hat{\beta}$ y $\hat{\beta}_{(i)}^1$. La j -ésima componente de este vector se denotará como $\Delta_i \hat{\beta}_j^1$. La gráfica de los puntos $(i, \Delta_i \hat{\beta}_j^1 / \sqrt{\widehat{Var}(\hat{\beta}_j)})$ con $i = 1, 2, \dots, n$ sirve para determinar qué tan sensible es $\hat{\beta}$; a cada uno de los n datos. Sin embargo, analizar simultáneamente m gráficas como la anterior es impráctico cuando m es relativamente grande. Además, las componentes de $\hat{\beta}$ están correlacionadas por lo que la influencia del dato i en $\hat{\beta}_j$ también afecta a otros coeficientes.

La ecuación $-2\{L(\beta) - L(\hat{\beta})\} = c$ es el contorno de una región de confianza para β . Cuando $\beta = \hat{\beta}_{(i)}^1$ entonces $c = c_i^1$ para alguna $c_i^1 \in \mathbb{R}$. Entre más grande sea c_i^1 mayor es la influencia de (y_i, x_i') en $\hat{\beta}$. Una expresión aproximada para c_i^1 se obtiene desarrollando en series de Taylor, y hasta el término de segundo orden, la función $F(\beta) = -2\{L(\beta) - L(\hat{\beta})\}$ alrededor de $\beta = \hat{\beta}_{(i)}^1$. De esta forma se obtiene que:

$$c_i^1 \simeq (\hat{\beta}_{(i)}^1 - \hat{\beta})' X'W(\hat{\beta})X(\hat{\beta}_{(i)}^1 - \hat{\beta})$$

y por (6.1.2.5.1.2.3):

$$c_i^1 \simeq \frac{h_i r_i^2}{(1 - h_i)^2};$$

r_i es el i -ésimo residuo de Pearson. Es conveniente graficar los puntos (i, c_i^1) con $i = 1, 2, \dots, n$ para determinar qué datos son los que mayor influencia tienen en $\hat{\beta}$. Pregibon también recomienda graficar $(\hat{\eta}_i, c_i^1)$ para $i = 1, 2, \dots, n$ y Hosmer y Lemeshow (1989) sugieren graficar c_i^1 contra \hat{p}_i y h_i .

El índice c_i^1 es análogo a la distancia de Cook. En lugar de c_i^1 , algunos autores — como Cook y Weisberg (1982) y Williams (1987) — usan la medida:

$$D_i^1 = \frac{h_i r_i^2}{m(1 - h_i)^2}.$$

El factor $1/m$ no aparece en c_i^1 .

b. Influencia global de un dato en las medidas de bondad de ajuste.

Pregibon también desarrolló medidas que reflejan la influencia global de un dato en D^2 y X^2 . El procedimiento es similar al que se usó para obtener c_i^1 .

La influencia del dato i en la devianza se estudia mediante la función:

$$D^2(\hat{\beta}_{(i)}^1) = 2\{L(\bar{\beta}; v) - L(\hat{\beta}_{(i)}^1; v)\};$$

en donde $\bar{\beta}$ estima al vector de coeficientes del modelo saturado,

$$L(\bar{\beta}; v) = \sum_{l=1}^n v_l \ln\{f_{Y_l}(y_l; k(x_l' \bar{\beta}), \phi)\},$$

$$L(\hat{\beta}_{(i)}^1; v) = \sum_{l=1}^n v_l \ln\{f_{Y_l}(y_l; k(x_l' \hat{\beta}_{(i)}^1), \phi)\},$$

$$v' = (v_1, v_2, \dots, v_n) \quad y \quad v_l = \begin{cases} 1, & \text{si } l \neq i; \\ 0, & \text{si } l = i. \end{cases}$$

$D^2(\hat{\beta}_{(i)}^1)$ es la devianza del modelo $logit(\hat{p}_{(i)}^1) = X_{(i)}' \hat{\beta}_{(i)}^1$ y es función de $\hat{\beta}_{(i)}^1$. El desarrollo en series de Taylor de $D^2(\hat{\beta}_{(i)}^1)$ alrededor de $\bar{\beta}$ es, hasta el término de segundo orden,

$$D^2(\hat{\beta}_{(i)}^1) \simeq D^2(\bar{\beta}) - \left(dev_i^2 + \frac{r_i^2 h_i}{1 - h_i} \right);$$

con

$$D^2(\bar{\beta}) = 2\{L(\bar{\beta}; v) - L(\hat{\beta}; v)\},$$

$$L(\hat{\beta}; v) = \sum_{l=1}^n v_l \ln\{f_{Y_l}(y_l; k(x_l' \hat{\beta}), \phi)\}.$$

y dev_i es el i -ésimo residuo de la devianza del modelo $logit(\hat{p}) = X\hat{\beta}$. De la expresión anterior se deduce que la devianza decrece cuando se elimina el dato i . El decremento producido en la devianza es:

$$\Delta_i D \equiv D^2(\hat{\beta}) - D^2(\hat{\beta}_{(i)}^1) \approx dev_i^2 + \frac{r_i^2 h_i}{1 - h_i}.$$

La gráfica de $(i, \Delta_i D)$ con $i = 1, 2, \dots, n$ indica qué datos son los que tienen la mayor influencia en la devianza del modelo $logit(\hat{p}) = X\hat{\beta}$. Si en $\Delta_i D$, el residuo de Pearson (r_i) se reemplaza por dev_i se obtiene:

$$\Delta_i D \approx \frac{dev_i^2}{1 - h_i} = \frac{d_i}{1 - h_i}.$$

Este es el i -ésimo residuo de la devianza estandarizado y elevado al cuadrado (sección 4.2.2).

El estimador de máxima verosimilitud de β no minimiza el valor de la estadística ji-cuadrada de Pearson generalizada. Por esto, no es sencillo determinar qué efecto tiene el dato i en X^2 . A veces, el valor de esta medida de bondad de ajuste crece cuando (y_i, x_i') se elimina de la muestra. El cuadrado del i -ésimo residuo de Pearson estandarizado, $r_i^2/(1 - h_i)$, es el índice propuesto por Pregibon para medir la influencia global del dato i en X^2 .

La gráfica de $\Delta_i D$ ó $r_i^2/(1 - h_i)$ contra \hat{p}_i está formada por dos curvas simétricas respecto a $\hat{p} = 0.5$. Los datos con los residuos más grandes se localizan en la parte superior de dichas curvas.

c. Influencia global de un dato en los residuos de la devianza.

El cambio producido en la devianza del dato j cuando (y_i, x_i') se excluye de la muestra es:

$$\Delta_i d_j \equiv d_{j(i)} - d_j,$$

$d_{j(i)}$ y d_j son el cuadrado de la devianza del dato j en los modelos $logit(\hat{p}_{(i)}) = X_{(i)}\hat{\beta}_{(i)}$ y $logit(\hat{p}) = X\hat{\beta}$ respectivamente. Cuando se usa el estimador a un paso de $\beta_{(i)}$, $\Delta_i d_j$ se aproxima con:

$$\Delta_i d_j^1 = \frac{2r_j h_{ij} r_i}{1 - h_i} + \frac{r_i^2 h_{ij}^2}{(1 - h_i)^2};$$

h_{ij} es el elemento en el renglón i y columna j de la matriz

$$H = W^{\frac{1}{2}}(\hat{\beta})X(X'W(\hat{\beta})X)^{-1}X'W^{\frac{1}{2}}(\hat{\beta}).$$

Para facilitar el análisis de la medida anterior conviene graficar $(i, \Delta_i d_j^1)$ para $i = 1, 2, \dots, n$.

Cuando $\Delta_i d_j^1$ es igual a cero, el dato i no afecta el ajuste de (y_j, x_j') . Si $\Delta_i d_j^1$ es mayor que cero entonces $d_{j(i)}^1 > d_j$ y el ajuste del dato j empeora cuando se ignora a (y_i, x_i') . En cambio, si $\Delta_i d_j^1 < 0$, el ajuste de (y_j, x_j') mejora al suprimir el dato i . Además, $\Delta_i d_j^1 \neq \Delta_j d_i^1$; esto quiere decir que: la influencia de (y_i, x_i') en la devianza del dato j es distinta a la influencia del dato j en dev_i . $\Delta_i d_j^1$ mide el efecto que tiene (y_i, x_i') en su devianza. Como

$$\Delta_i d_i^1 = \frac{r_i^2 h_i (2 - h_i)}{(1 - h_i)^2} \geq 0,$$

la devianza del dato i aumenta cuando se le excluye del modelo. Sin embargo, si (y_i, x_i') se omite del análisis, el ajuste global de los $n - 1$ datos restantes mejora porque:

$$\sum_{j \neq i} \Delta_i d_j^1 = \frac{-r_i^2 h_i}{1 - h_i} \leq 0.$$

Por último, la suma de las n devianzas elevadas al cuadrado es mayor para el modelo $logit(\hat{\beta}_{(i)}) = X_{(i)}\hat{\beta}_{(i)}$ que para $logit(\hat{\beta}) = X\hat{\beta}$; es decir:

$$\sum_{j=1}^n \Delta_i d_j^1 = \sum_{j \neq i} \Delta_j d_j^1 + \Delta_i d_i^1 = \frac{r_i^2 h_i}{(1 - h_i)^2} \geq 0.$$

Ejemplo. Las gráficas que presenta Pregibon en su artículo están basadas en los datos del apéndice A2 y el modelo:

$$logit(\hat{p}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot \log(vol) + \hat{\beta}_2 \cdot \log(tas).$$

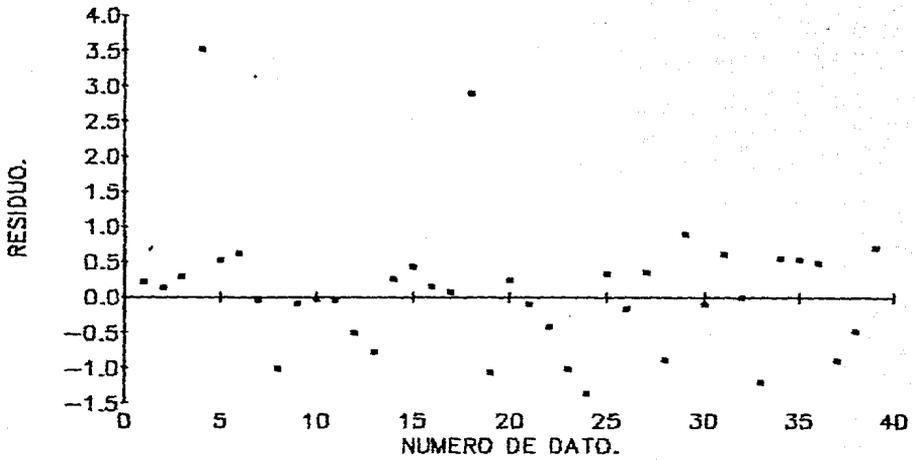
Aquí se analizará la influencia global que tiene cada dato de este modelo en $\hat{\beta}$ y D^2 . El análisis está basado en c_i^1 y $\Delta_i D$ que dependen de los residuos y las palancas. Las dos primeras gráficas al final de esta sección muestran que, en valor absoluto, los residuos más grandes corresponden a los datos 4 y 18. Nótese, además, que en las dos gráficas el comportamiento de los residuos es similar. A continuación se graficaron las palancas de Thomas y Pregibon contra el número de dato. El rango de las palancas de Thomas es mayor que el de las de Pregibon. Las primeras suman a 10.18 y las segundas a 3 (el número de parámetros en el modelo). En ambas gráficas, el dato 31 es el que tiene la palanca más grande. Sin embargo, el dato 30 tiene una palanca grande de Thomas y una palanca pequeña de Pregibon. La quinta gráfica muestra los residuos estandarizados de la devianza contra el número de dato. En valor absoluto, los mayores residuos estandarizados corresponden a los datos 4 y 18 que tienen una palanca de Pregibon relativamente pequeña. Abajo se graficó la distancia de Cook, c_i^1 , contra i . Esta gráfica indica que los datos 4 y 18 son los que mayor influencia global tienen en $\hat{\beta}$. La gráfica siguiente, de los puntos $(i, \Delta_i D)$, revela que (y_4, x'_4) y (y_{18}, x'_{18}) también influyen en el valor de la devianza. En la última gráfica, los puntos $(\hat{p}_i, \Delta_i D)$ forman dos curvas que se intersectan en $\hat{p} = 0.5$. Una curva se extiende de la esquina superior izquierda a la esquina inferior derecha. En ella se observa que la probabilidad de éxito estimada de los datos influyentes (cuatro y dieciocho) es menor a 0.11.

De acuerdo con c_i^1 y $\Delta_i D$, el dato 4 es el que mayor influencia global tiene en $\hat{\beta}$ y D^2 . Cuando este dato se excluye de la muestra, la devianza del modelo decrece de 29.227 a 22.426. Los estimadores de β y $\beta_{(4)}$ son respectivamente:

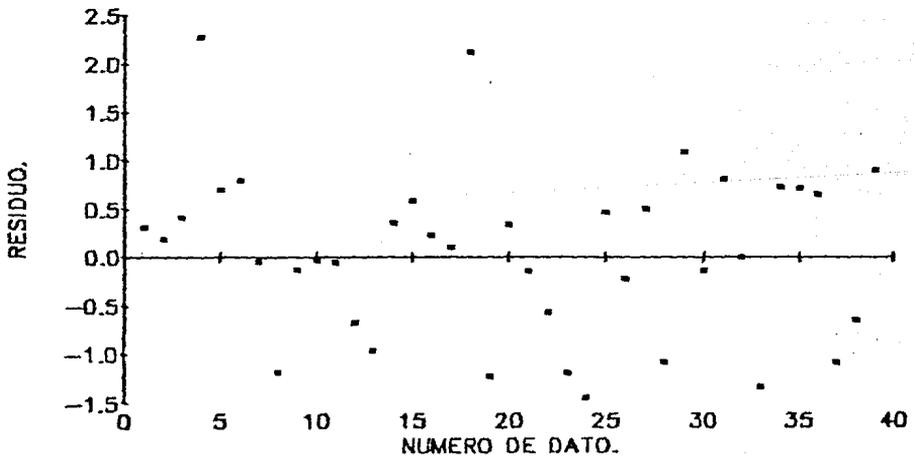
$$\hat{\beta} = (-2.875, 5.179, 4.562),$$

$$\hat{\beta}_{(4)} = (-5.206, 8.468, 7.455).$$

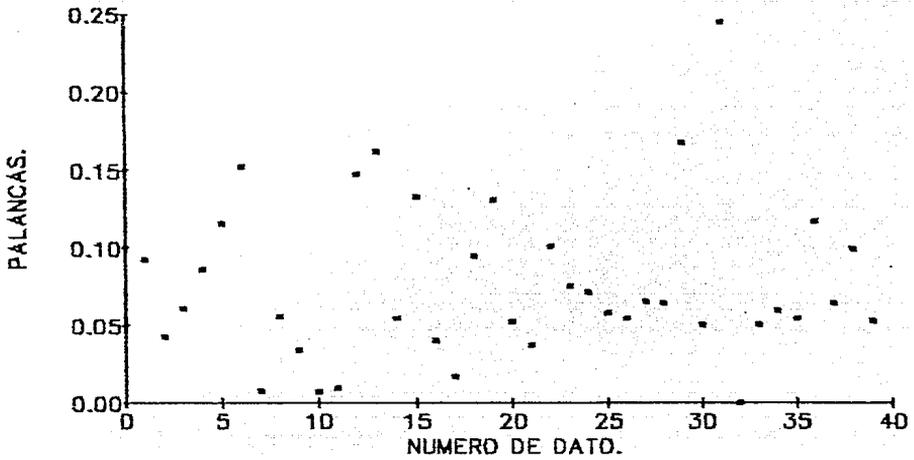
RESIDUOS DE PEARSON.



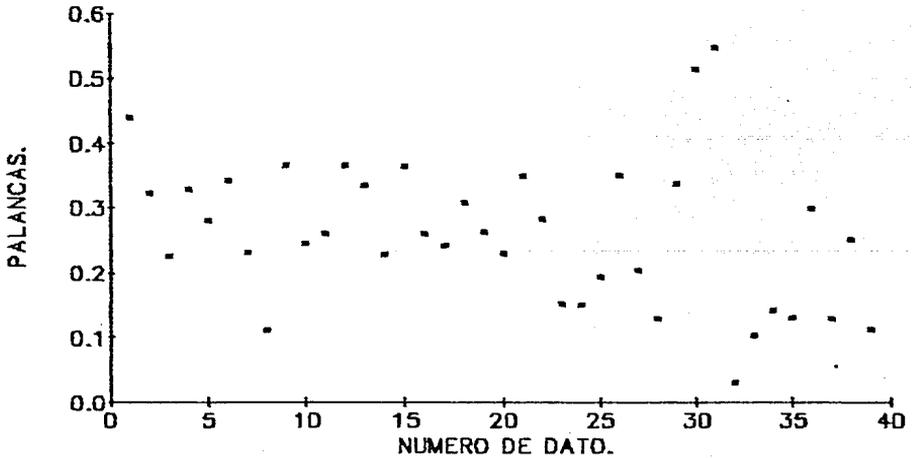
RESIDUOS DE LA DEVIANZA.



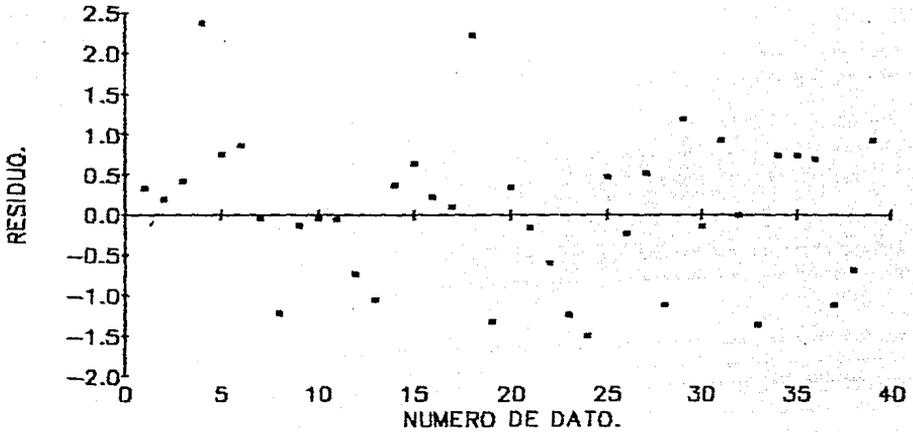
PALANCAS DE PREGIBON.



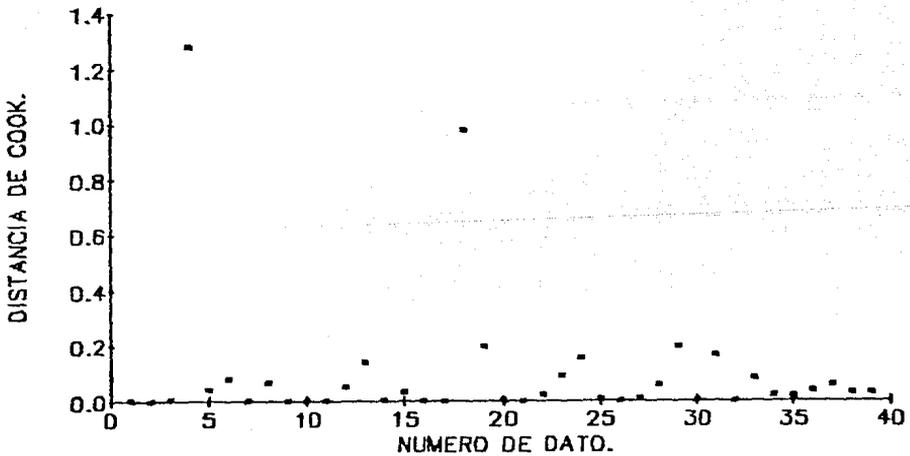
PALANCAS DE THOMAS.



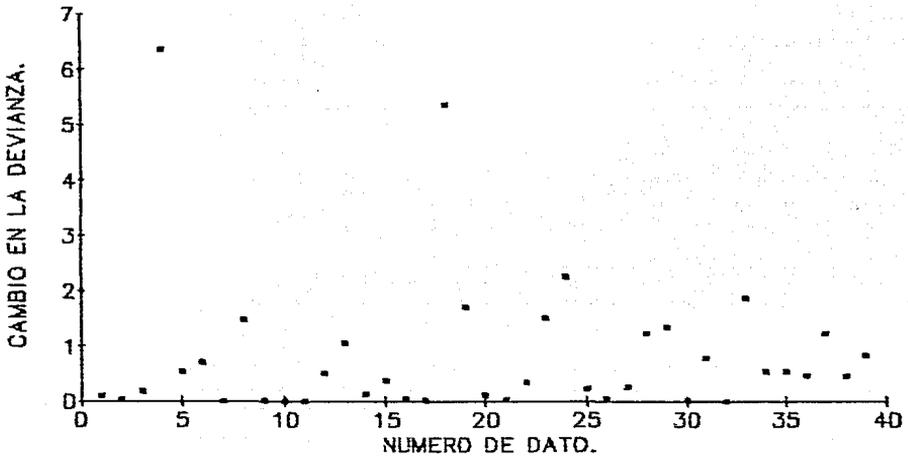
RESIDUOS DE LA DEVIANZA
ESTANDARIZADOS.



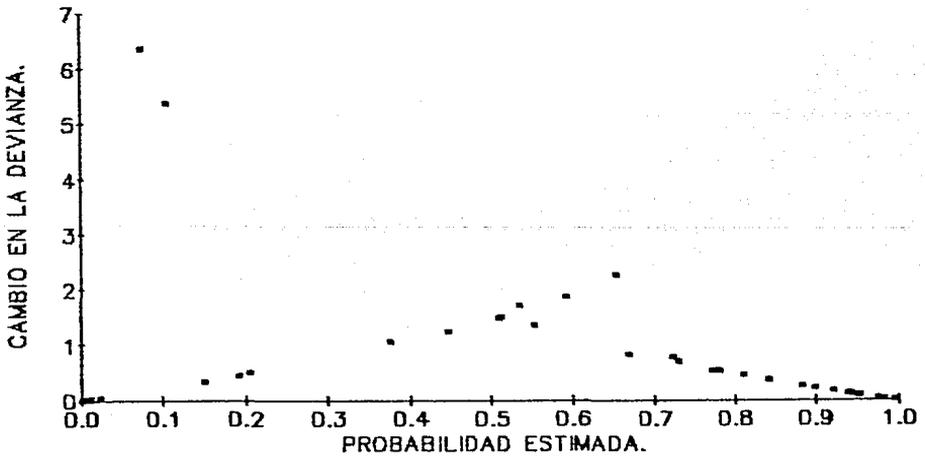
DISTANCIA DE COOK.



CAMBIO EN LA DEVIANZA.



CAMBIO EN LA DEVIANZA.



Bibliografía abreviada.

Cook, R. D. y Weisberg, S. (1982).

Hosmer, David W. y Lemeshow, Stanley (1989).

Pregibon, D. (1981).

Williams, D. A. (1987).

6.1.2.5.2 Influencia local.

Cook (1986) desarrolló un método para evaluar el efecto que tienen las perturbaciones infinitesimales en el modelo de regresión lineal normal. El método, basado en geometría diferencial, ha sido aplicado a los modelos lineales generalizados. La perturbación inducida en uno o más datos, las variables explicativas, el vector de respuestas o el peso de los datos se representará con un vector o una matriz. El método permite estudiar el efecto que tiene la perturbación en:

1. Todos los parámetros del modelo;
2. Un subconjunto de parámetros; y
3. Las predicciones puntuales.

Influencia local en todos los parámetros.

El efecto que tienen las perturbaciones infinitesimales en los parámetros de un modelo lineal generalizado se estudia mediante la curva de influencia. El logaritmo de la función de verosimilitud de un modelo lineal generalizado se denotará como $L(\delta)$ con

$$\delta = \begin{pmatrix} \beta \\ \phi \end{pmatrix},$$

ϕ puede ser desconocido. Las perturbaciones se representarán mediante el vector ω de dimensión $q \times 1$ que pertenece al conjunto abierto $\Omega \subset \mathbb{R}^q$. $L(\delta; \omega)$ será el logaritmo de la función de verosimilitud del modelo perturbado. Se supondrá que :

- a. Existe una $\omega_0 \in \Omega$ tal que $L(\delta; \omega_0) = L(\delta) \quad \forall \delta$ en el espacio parametral.
- b. Las dos primeras derivadas de $L(\delta; \omega)$ con respecto a δ y ω existen y son continuas.

Cuando el objetivo es evaluar el efecto de la perturbación en δ , la curva o gráfica de influencia es la superficie geométrica generada por:

$$\alpha(\omega) = \begin{pmatrix} \omega \\ DV(\omega) \end{pmatrix};$$

ω varía en Ω . $DV(\omega)$ es el desplazamiento de la verosimilitud y se define como:

$$DV(\omega) = 2[L(\hat{\delta}) - L(\hat{\delta}_\omega)].$$

$\hat{\delta}$ y $\hat{\delta}_\omega$ son los estimadores de máxima verosimilitud de δ para los modelos sin perturbar y perturbado respectivamente, es decir:

$$L(\hat{\delta}) = \sup_{\delta} L(\delta) \quad \text{y} \quad L(\hat{\delta}_\omega; \omega) = \sup_{\delta} L(\delta; \omega).$$

$\alpha(\omega)$ mide, con base en $L(\delta)$, la discrepancia entre $\hat{\delta}$ y $\hat{\delta}_\omega$.

La diferencia entre los estimadores $\hat{\delta}$ y $\hat{\delta}_\omega$ también puede medirse con base en $L(\delta; \omega)$. $L(\delta)$ es un "marco de referencia fijo" porque no depende de ω ; $L(\delta; \omega)$ es un "marco de referencia variable". En general, los diagnósticos de influencia basados en marcos de referencia fijos son más sencillos de motivar e interpretar.

La superficie generada por la gráfica de influencia está contenida en \mathbb{R}^{q+1} por lo que generalmente no es posible visualizarla. Sin embargo, existen medidas que describen el comportamiento global (en Ω) y local (en $V(\omega_0)$, una vecindad de ω_0) de dicha superficie. Así, hay medidas de influencia global y medidas de influencia local.

El comportamiento de $\alpha(\omega)$ alrededor de ω_0 se estudiará mediante las curvaturas de las secciones normales de la gráfica de influencia en $V(\omega_0)$. En geometría diferencial, el vector $\alpha(\omega(a)) = \omega_0 + a \cdot l$, con $a \in \mathbb{R}$ y l un vector fijo de norma unitaria, es una sección normal de $\alpha(\omega)$. La dirección de $\alpha(\omega(a))$ está determinada por la dirección de l y $\alpha(\omega(a)) = \alpha(\omega_0)$ si $a = 0$. Calcular la curvatura de $\alpha(\omega(a))$ en $a = 0$ es equivalente a rotar la sección normal y calcular la curvatura del vector resultante:

$$\rho(a) = \left(DV(\omega(a)) \right)^a.$$

En la dirección del vector l , la curvatura de $\rho(a)$ en el punto $a = 0$ es:

$$c_l = |\bar{D}V(\omega_0)| = 2|l^T \bar{F} l|,$$

$$\text{en donde} \quad \bar{D}V(\omega_0) = \left. \frac{\partial^2 DV(\omega(a))}{\partial a^2} \right|_{a=0},$$

y \bar{F} es una matriz simétrica de qxq formada por

$$\left. \frac{\partial^2 L(\hat{\delta}_\omega)}{\partial \omega_k \partial \omega_j} \right|_{\omega=\omega_0} \quad \text{para} \quad k, j = 1, 2, \dots, q.$$

A c_l , la curvatura de $\alpha(\omega(a))$, se le llamará, en forma abreviada, la curvatura normal. Cuando $\alpha(\omega)$ está en \mathbb{R}^3 , c_l indica qué tanto se desvía la gráfica de influencia del plano tangente al punto ω_0 . No hay un valor de referencia para juzgar la magnitud de la curvatura normal máxima: $c_{max} = \max\{c_l\}$. Lo que se recomienda es comparar la máxima curvatura normal con las curvaturas obtenidas a partir de varios conjuntos de datos simulados. A la postre, la experiencia adquirida en los análisis de influencia local permitirá hacer una apreciación correcta de c_{max} — dicen Thomas y Cook (1990).

La expresión $l'\tilde{F}l$ denota un valor propio de \tilde{F} , su respectivo vector propio es l . Esto quiere decir la dirección de los vectores propios de \tilde{F} determina la dirección de las secciones normales. El diagnóstico más valioso — dice Cook — lo proporciona l_{max} , el vector propio asociado a c_{max} . Sus componentes indican qué perturbación, en $V(\omega_0)$, produce el máximo cambio en el desplazamiento de la verosimilitud.

La matriz \tilde{F} puede expresarse como $\tilde{F} = J'\tilde{L}J$, \tilde{L} es el hessiano del modelo sin perturbar evaluado en $\delta = \hat{\delta}$ y $\omega = \omega_0$; $J = -(\tilde{L})^{-1}\Delta$ con

$$\Delta = \{\Delta_{rk}\} \quad \text{y} \quad \Delta_{rk} = \left. \frac{\partial^2 L(\delta; \omega)}{\partial \delta_r \partial \omega_k} \right|_{\substack{\delta=\hat{\delta} \\ \omega=\omega_0}} \quad \text{para} \quad \begin{array}{l} r = 1, 2, \dots, m \\ j = 1, 2, \dots, q. \end{array}$$

Por lo tanto,

$$\tilde{F} = \Delta'(\tilde{L})^{-1}\Delta \quad (6.1.2.5.2.1)$$

y

$$c_l = 2|l'\Delta'(\tilde{L})^{-1}\Delta l|. \quad (6.1.2.5.2.2)$$

A veces existen expresiones explícitas para las medidas de influencia local (c_{max} , l_{max} , ..) que evitan el cálculo de \tilde{F} .

Por último, la solución a un problema de influencia local depende del estudio; no pueden darse recomendaciones universales.

a. Influencia local del peso de los datos en δ .

Suponga que ϕ es conocido, es decir que la variable de respuesta del modelo lineal generalizado pertenece a la familia exponencial. En términos de la liga natural, el logaritmo de la función de verosimilitud del modelo sin perturbar es:

$$L(\delta) = \sum_{i=1}^n \{a^{-1}(\phi)[y_i k(x'_i \beta) - b\{k(x'_i \beta)\}] + c(y_i, \phi)\}.$$

Las entradas del vector $\omega' = (\omega_1, \omega_2, \dots, \omega_n)$ representarán la perturbación inducida en el peso de los datos (nótese que $q = n$). Usualmente, $0 < \omega_i < 1$ con $i = 1, 2, \dots, n$. El logaritmo de la función de verosimilitud del modelo perturbado es:

$$L(\delta; \omega) = \sum_{i=1}^n \omega_i \{a^{-1}(\phi)[y_i k(x'_i \beta) - b\{k(x'_i \beta)\}] + c(y_i, \phi)\}.$$

(esta expresión es similar a (6.1.2.5.1.2.1)). Como se puede ver, $L(\delta; \omega) = L(\delta)$ si ω es un vector unitario, es decir: $\omega_0 = 1$.

Para estudiar el efecto que tiene ω en δ (ó en β porque ϕ es conocido) deben calcularse los valores y vectores propios de $\bar{F} = J' \bar{L} J$, más explícitamente,

$$\bar{F} = J' \bar{L} J = -\frac{1}{a(\phi)} \dot{E}' \dot{K} X (X' \bar{E} X) X' \dot{K}' \dot{E},$$

con $\dot{E} = \text{diag}\{\dot{e}_i\}$, \dot{e}_i es el i -ésimo residuo de puntaje del modelo sin perturbar (véase la sección 3.1); $\dot{K} = \text{diag}\{\dot{k}(x'_i \hat{\beta})\}$; X es la matriz de diseño; $J = (\dot{K} X)' \dot{E}$; $\bar{E} = \bar{E}(\beta)|_{\beta=\hat{\beta}}$ y $\bar{L} = a^{-1}(\phi) X' \bar{E} X$.

b. Influencia local de las variables explicativas en δ .

A veces existen errores de medición o de redondeo en las variables explicativas. Su efecto en el estimador de δ puede estudiarse mediante el método desarrollado por Cook. La matriz de diseño perturbada de un modelo lineal generalizado es:

$$X_\omega = X + US;$$

$U = \{\omega_{ij}\}$ con ω_{ij} una perturbación cualquiera y $S = \text{diag}\{s_j\}$ con s_j un factor de escala que transforma a ω_{ij} para que sea compatible con x_{ij} . Beckman propuso que s_j sea el rango o la desviación estándar de X_j .

El logaritmo de la función de verosimilitud del modelo perturbado es:

$$L(\delta; \omega) = \sum_{i=1}^n \{a^{-1}(\phi) |y_i k(x'_{i\omega} \beta) - b\{k(x'_{i\omega} \beta)\} + c(y_i, \phi)\},$$

$x'_{i\omega}$ es el i -ésimo renglón de X_ω . Se supondrá que ϕ es conocido. Por la expresión (6.1.2.5.2.1), $\bar{F} = \Delta'(\bar{L})^{-1}\Delta$ con $\bar{L} = a^{-1}(\phi)X'E'X$, $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_m)$, $\Delta_j = s_j\{d_j \hat{e}'\bar{K} + \hat{\delta}_j X'E'\bar{E}\}$ y d_j el j -ésimo vector canónico de dimensión $nx1$.

Influencia local en un subconjunto de parámetros.

Cuando se desea estudiar el efecto que tiene ω en δ_1 , un subconjunto de parámetros: $\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}$, la gráfica de influencia es la superficie generada por

$$\alpha_s(\omega) = \begin{pmatrix} \omega \\ DV_s(\omega) \end{pmatrix},$$

con $\omega \in \Omega$ y $DV_s(\omega) = 2[L(\hat{\delta}) - L(\hat{\delta}_\omega)]$. Aquí, $L(\hat{\delta}_\omega)$ es el perfil de la verosimilitud de δ_1 evaluado en $\hat{\delta}_{1\omega}$, el estimador de δ_1 obtenido al maximizar $L(\delta; \omega)$. El perfil de la verosimilitud de δ_1 es $L(\delta_\omega) = L((\delta_1, g(\delta_1))$; $g(\delta_1)$ maximiza a $L(\delta)$ cuando δ_1 está fijo. En $\alpha_s(\omega)$, igual que en $\alpha(\omega)$, $L(\delta)$ se usó para comparar a $\hat{\delta}$ con $\hat{\delta}_\omega$.

Las curvaturas de $\alpha_s(\omega)$ son:

$$c_1(\delta_1) = 2|l' \Delta' [(\bar{L})^{-1} - B_{22}] \Delta l|$$

$$\text{con } \bar{L} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix},$$

$$B_{22} = \begin{pmatrix} 0 & 0 \\ 0 & L_{22}^{-1} \end{pmatrix};$$

Δ se definió arriba de la expresión (6.1.2.5.2.1). $c_1(\delta_1)$ indica cómo afectan a $\hat{\delta}_1$ las perturbaciones infinitesimales.

a. Influencia local del peso de los datos en δ_1 .

Considere el problema de medir las discrepancias entre $\hat{\beta}_j$ y $\hat{\beta}_{j\omega}$ causadas por una perturbación en el peso de los datos del modelo logístico. Las medidas de influencia local propuestas por Cook se calculan a partir de

$$\bar{F} = \dot{E}X(X'WX)^{-1}X'\dot{E}$$

con $\dot{E} = \text{diag}\{\dot{\epsilon}_i = y_i - n_i\hat{p}_i\}$ y $W = \text{diag}\{w_i = n_i\hat{p}_i(1 - \hat{p}_i)\}$. El cambio producido en el desplazamiento de la verosimilitud es máximo en la dirección señalada por

$$l'_{max} \propto (r_1\epsilon_1, r_2\epsilon_2, \dots, r_n\epsilon_n);$$

r_i es el i -ésimo residuo de Pearson del modelo sin perturbar y ϵ_i es el i -ésimo residuo obtenido al efectuar la regresión de la j -ésima columna de $W^{\frac{1}{2}}X$ en las columnas restantes.

Ejemplo. El ejemplo está basado en los datos del apéndice A2 y el modelo:

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2},$$

con $x_{i1} = \log(\text{vol}_i)$, $x_{i2} = \log(\text{tas}_i)$ e $i = 1, 2, \dots, n$.

La primera gráfica corresponde a los puntos

$$\left(i, \frac{r_i\epsilon_i}{\sqrt{\sum_{i=1}^n (r_i\epsilon_i)^2}} \right) \quad \text{con} \quad i = 1, 2, \dots, n.$$

r_i es el i -ésimo residuo de Pearson del modelo mencionado anteriormente. El residuo ϵ_i se obtiene al ajustar el modelo de regresión lineal normal:

$$E(\sqrt{w_i}X_{i1}) = \delta\sqrt{w_i} + \delta_2\sqrt{w_i}x_{i2} \quad \text{con} \quad w_i = n_i\hat{p}_i(1 - \hat{p}_i).$$

La gráfica muestra que una pequeña perturbación en el peso inicial de los datos 4 y 18 tiene un efecto relevante en el estimador de β_1 . En este caso, la perturbación consiste en reducir el peso inicial de (y_4, x'_4) y (y_{18}, x'_{18}) .

La segunda gráfica es similar a la primera; se obtiene graficando las entradas del vector

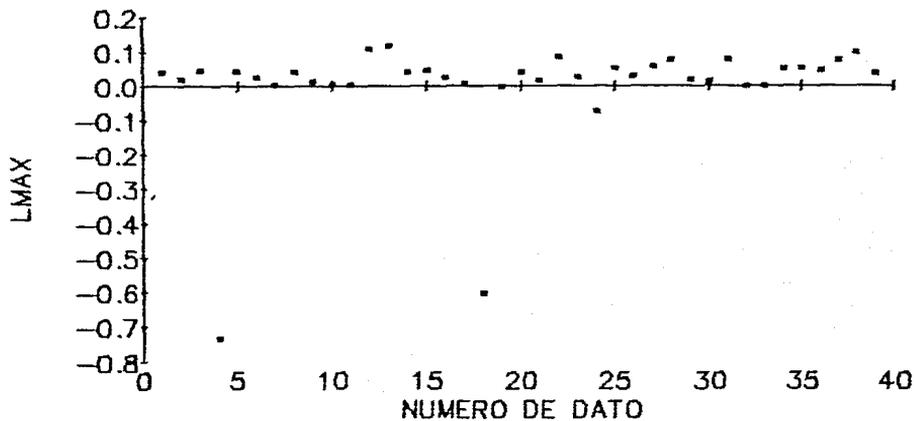
$$l_{max} = \frac{1}{\sqrt{\sum_{i=1}^n (r_i\epsilon_i)^2}} (r_1\epsilon_1, r_2\epsilon_2, \dots, r_n\epsilon_n)$$

contra el número de dato, i . Igual que en el caso anterior, r_i es el i -ésimo residuo de Pearson del modelo logístico. En cambio, ε_i es el i -ésimo residuo del modelo de regresión lineal normal:

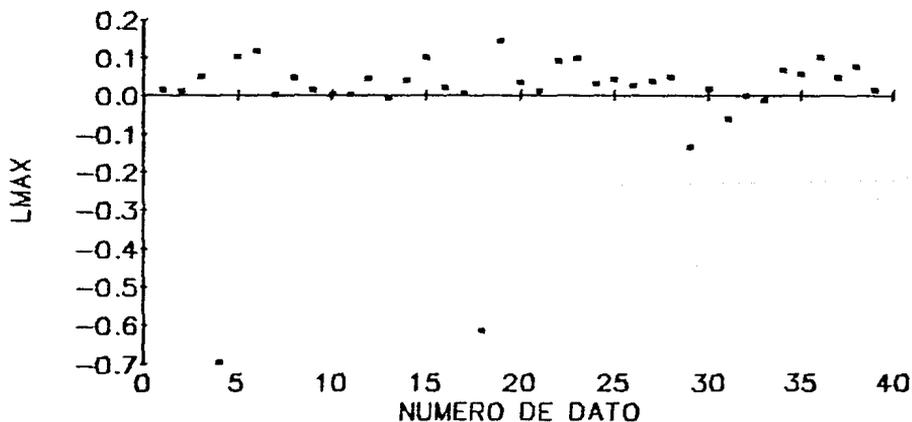
$$E(\sqrt{w_i}X_{i2}) = \alpha_1\sqrt{w_i} + \alpha_2\sqrt{w_i}x_{i1} \quad \text{con} \quad w_i = n_i\hat{p}_i(1 - \hat{p}_i).$$

La gráfica indica que los pesos de los datos 4 y 18 son los que mayor influencia local tienen en $\hat{\beta}_2$.

INFLUENCIA LOCAL DEL PESO
DE LOS DATOS EN LOG(VOL).



INFLUENCIA LOCAL DEL PESO
DE LOS DATOS EN LOG(TAS).



Influencia local en las predicciones puntuales.

Una predicción puntual es, para Thomas y Cook (1990), el estimador de la media de una respuesta observada o futura de un modelo lineal generalizado. Se supondrá que z es el vector de variables explicativas de dicha respuesta y que ϕ es conocido o que se reemplazó por un estimador. La predicción puntual de la respuesta asociada a z es:

$$\hat{y}(z) = \hat{b}\{k(z'\hat{\beta})\}.$$

Generalmente, $\hat{y}(z)$ sólo depende de $\hat{\theta}$, el estimador del parámetro canónico. La varianza de $\hat{y}(z)$ se estimará con:

$$\text{var}\{\hat{y}(z)\} = a(\hat{\phi})\bar{b}\{k(z'\hat{\beta})\}.$$

Cuando el modelo se perturba por ω , la predicción puntual de la respuesta con vector de variables explicativas z se denotará como:

$$\hat{y}(z, \omega) = \hat{b}\{k(z'\hat{\beta}_\omega)\};$$

$\hat{\beta}_\omega$ maximiza $L(\beta; \omega)$, el logaritmo de la función de verosimilitud del modelo perturbado.

El desplazamiento de la verosimilitud no es útil para comparar a $\hat{y}(z)$ con $\hat{y}(z, \omega)$ cuando $\hat{y}(z)$ estima a una respuesta futura. Por esto, la influencia local en las predicciones se estudiará con "las funciones objetivo". Estas funciones se definen a partir de los residuos abajo mencionados que miden las discrepancias entre $\hat{y}(z)$ y $\hat{y}(z, \omega)$:

$$r_P(z, \omega) = \frac{\hat{y}(z) - \hat{y}(z, \omega)}{\sqrt{\text{Var}\{\hat{y}(z)\}}}, \quad \text{el residuo de Pearson};$$

$$r_A(z, \omega) = \frac{A(\hat{y}(z)) - A(\hat{y}(z, \omega))}{\hat{A}(\hat{y}(z))\sqrt{\text{Var}\{\hat{y}(z)\}}}, \quad \text{el residuo de Anscombe}$$

$$\text{y } r_D(z, \omega) = 2\{\hat{y}(z)[k(z'\hat{\beta}) - k(z'\hat{\beta}_\omega)] - b\{k(z'\hat{\beta})\} + b\{k(z'\hat{\beta}_\omega)\}\},$$

el residuo de la devianza. Las funciones objetivo son:

$$f_P(z, \omega) = r_P^2(z, \omega), \quad f_A(z, \omega) = r_A^2(z, \omega) \quad \text{y} \quad f_D(z, \omega) = \frac{r_D(z, \omega)}{a(\phi)}$$

y tienen en común las propiedades siguientes:

- No las afecta el reescalamiento o la combinación lineal de variables explicativas. Es decir, son invariantes cuando X se posmultiplica por una matriz no singular.
- Como función de ω , alcanzan su mínimo en ω_0 (la perturbación que deja intactos a los datos).
- Si la variable de respuesta se distribuye como una normal y el peso asignado al i -ésimo dato es cero, es decir

$$\omega = \{\omega_l\} \quad \text{con} \quad \omega_l = \begin{cases} 1, & \text{si } l \neq i; \\ 0, & \text{si } l = i, \end{cases}$$

$$\text{entonces} \quad \sum_{l=1}^n f_P(x_l, \omega) = \sum_{l=1}^n f_A(x_l, \omega) = \sum_{l=1}^n f_D(x_l, \omega) = \frac{1}{\sigma^2} \|\hat{Y} - \hat{Y}_{(i)}\|^2.$$

Esta es la distancia de Cook para el dato i multiplicada por m .

Si z es un vector fijo, las funciones objetivo representan una superficie en Ω . Se determinará en qué dirección se produce el máximo cambio en la superficie cuando $\omega \in V(\omega_0)$. En lugar de trabajar con la superficie generada por la función objetivo se usará una aproximación de ella. El plano tangente a la superficie en el punto $f(\omega_0)$ no es útil por ser horizontal. Por lo tanto, se trabajará con la aproximación de segundo orden: la curvatura normal de la superficie en el punto $f(\omega_0)$. Esta curvatura es, en la dirección del vector unitario l ,

$$c(l, z) = |l' \tilde{f} l| \quad \text{con} \quad \tilde{f} = \left. \frac{\partial^2 f}{\partial \omega \partial \omega'} \right|_{\omega = \omega_0}.$$

Más explícitamente:

$$c(l, z) = 2a(\hat{\phi})q(z) |l' \Delta' (X' \bar{E} X)^{-1} z z' (X' \bar{E} X)^{-1} \Delta l|$$

$$\text{con} \quad \Delta = \left. \frac{\partial^2 L(\beta; \omega)}{\partial \beta \partial \omega'} \right|_{\substack{\beta = \hat{\beta} \\ \omega = \omega_0}} \quad \text{y} \quad q(z) = \bar{b}\{k(z' \hat{\beta})\} [k(z' \hat{\beta})]^2.$$

Como la curvatura normal es la misma para f_P , f_A y f_D , los autores llaman a la función objetivo la diferencia predictiva.

La dirección de máxima curvatura normal está dada por el vector

$$l_{max} \propto \Delta'(X'\bar{E}X)^{-1}z. \quad (6.1.2.5.2.3)$$

El signo y la magnitud de las componentes de l_{max} indican cómo definir a ω para que se produzca el máximo cambio local en la diferencia predictiva.

a. Perturbación de la variable de respuesta.

El vector de respuestas perturbadas se define como:

$$Y_\omega = Y + S\omega; \quad \text{con} \quad S = \text{diag}\{[a(\hat{\phi})\bar{b}\{k(x'_i\hat{\beta})\}]^{\frac{1}{2}}\}.$$

con $S = \text{diag}\{s_i\}$ y $s_i = \{a(\hat{\phi})\bar{b}\{k(x'_i\hat{\beta})\}\}^{\frac{1}{2}}$. El término s_i estima a la desviación estándar de Y_i y modifica la escala de ω_i para que coincida con la de Y_i . Si $\omega = 0$, $Y_\omega = Y$. El vector Y_ω carece de una interpretación práctica cuando Y es discreta (como en el modelo logístico). Si $Y_\omega = y_\omega$, el logaritmo de la función de verosimilitud de β es:

$$L(\beta; \omega) = \sum_{i=1}^n \left\{ \frac{1}{a(\hat{\phi})} [y_{i\omega} k(x'_i\beta) - b\{k(x'_i\beta)\}] + c(y_{i\omega}, \phi) \right\}; \quad y_{i\omega} = y_i + \omega_i s_i.$$

La dirección de máxima curvatura normal la proporciona el vector:

$$l_{max} \propto S\dot{K}X(X'\bar{E}X)^{-1}z$$

con $\dot{K} = \text{diag}\{\dot{k}(x'_i\hat{\beta})\}$ (en la expresión (6.1.2.5.2.3), $\Delta = a^{-1}(\hat{\phi})X'\dot{K}S$). Cuando $z = x_i$, la i -ésima componente del vector $S\dot{K}X(X'\bar{E}X)^{-1}x_i$ es

$$h_i^* = x'_i(X'\bar{E}X)^{-1}x_i \dot{k}(x'_i\hat{\beta}) \sqrt{\text{Var}(Y_i)}.$$

Esta es la palanca de Thomas y Cook para el i -ésimo dato y mide la influencia local de Y_i en su valor ajustado.

b. Perturbación de un variable explicativa.

Thomas y Cook sólo consideran el caso de medir la influencia local de una variable explicativa (la j -ésima) en las predicciones. El cómputo de las medidas de influencia local se simplifica si la perturbación inducida en X_j es aditiva. Sin embargo, debe tenerse presente que estas perturbaciones sólo son aplicables a variables explicativas continuas. Entonces, la j -ésima variables explicativa perturbada será :

$$X_{j\omega} = X_j + s_j\omega;$$

ω es el vector de perturbaciones y está escalado por s_j , una norma de X_j (por ejemplo la Euclidiana). El término s_j permite comparar la máxima curvatura normal que se obtiene cuando distintas variables explicativas (continuas) se perturban por separado.

La función $L(\beta; \omega)$ sólo depende de $X_{j\omega}$ a través del parámetro canónico:

$$\theta_\omega = k(x_{j\omega}\beta_j + \sum_{t \neq j} x_t\beta_t).$$

Además, $\Delta = \frac{1}{a(\hat{\phi})} s_j A_j$ con $A_j = u_j e' - \hat{\beta}_j X' \bar{E}$ y u_j el j -ésimo vector canónico de dimensión $n \times 1$. Por lo tanto,

$$l_{max} \propto A_j' (X' \bar{E} X)^{-1} z$$

$$y \quad c(l_{max}, z) = 2s_j^2 q(z) a^{-1}(\hat{\phi}) \|A_j' (X' \bar{E} X)^{-1} z\|^2.$$

Si $i = 1, 2, \dots, n$, la gráfica de los puntos $(i, c(l_{max}, x_i))$ con

$$c(l_{max}, x_i) = 2s_j^2 q(x_i) a^{-1}(\hat{\phi}) \|A_j' (X' \bar{E} X)^{-1} x_i\|^2$$

mide la influencia local del vector X_j en $\hat{\mu}$.

c. Perturbación de un dato.

El vector $(x_{r1} + s_1\omega_1, \dots, x_{r(m-1)} + s_{m-1}\omega_{m-1}, y_r + s_m\omega_m)$ es el resultado de perturbar el r -ésimo dato: (x_r, y_r) ; s_j es la norma de X_j y $s_m = \sqrt{\widehat{Var}(Y_r)}$. En este caso,

ω es un vector de dimensión $m \times 1$ que no altera a (x_r', y_r) cuando todas sus componentes son iguales a cero. La matriz Δ es:

$$\Delta = \frac{B_r}{a(\hat{\phi})};$$

B_r es una matriz particionada definida como:

$$B_r = [(\dot{e}_r S - \bar{e}_r x_r \hat{\beta}' S) | s_m k(x_r' \hat{\beta}) x_r] \quad \text{con} \quad S = \text{diag}\{s_j\}.$$

La dirección en la que se presenta el máximo cambio local en la diferencia predictiva está dado por el vector:

$$l_{max} \propto B_r'(X' \bar{E} X)^{-1} z.$$

d. Perturbación del peso de los datos.

Cuando se perturba el peso de los datos de un modelo lineal generalizado,

$$\Delta = a^{-1}(\hat{\phi}) X' \dot{E}.$$

Como consecuencia, $l_{max} \propto \dot{E} X (X' \bar{E} X)^{-1} z$

$$\text{y} \quad c(l_{max}, z) = 2q(z) a^{-1}(\hat{\phi}) \|\dot{E} X (X' \bar{E} X)^{-1} z\|^2.$$

Bibliografía abreviada.

Cook, D. R. (1986).

Thomas, W. y Cook, D. R. (1990).

6.1.2.6 Datos discrepantes.

Un dato es discrepante cuando el modelo lo describe inadecuadamente. Como consecuencia, los residuos de los datos discrepantes son grandes en valor absoluto.

La componente sistemática de un modelo lineal generalizado cuyo i -ésimo dato es discrepante se define como:

$$\eta = X\beta + d_i \xi;$$

d_i es el i -ésimo vector canónico de dimensión $n \times 1$. El estimador de máxima verosimilitud de β es $\hat{\beta}_{(i)}$ — consúltese la sección de influencia global. El estimador máximo verosimil de ξ satisface la ecuación:

$$\dot{e}_i(x_i' \hat{\beta}_{(i)} + \hat{\xi}) = 0 \quad \text{con} \quad \dot{e}_i = [y_i - \dot{b}\{k(x_i' \hat{\beta})\}] k(x_i' \hat{\beta}).$$

La estadística de puntaje para probar:

$$H_0 : \text{El dato } i \text{ no es discrepante} \quad (\text{i.e. } \xi = 0)$$

vs.

$$H_1 : \text{El dato } i \text{ es discrepante} \quad (\text{i.e. } \xi \neq 0)$$

es $r_{P_i}^2$, el cuadrado del i -ésimo residuo de Pearson estandarizado, cuando la matriz de información observada es igual a su esperanza. Jennings (1986) afirma que la distribución asintótica de $r_{P_i}^2$ no es una $\chi_{(1)}^2$ como señala Pregibon (1982). Otro método para probar las hipótesis anteriores es el de cociente de verosimilitudes; la estadística de prueba es $\Omega = -2 \ln \Lambda = \phi^{-1}(D_0^2 - D_1^2)$ si ϕ es conocido. Williams (1987), basado en las medidas de influencia global desarrolladas por Pregibon (1981), obtuvo la aproximación siguiente para Ω :

$$\Omega = -2 \ln \Lambda = \phi^{-1}(D_0^2 - D_1^2) \simeq \phi^{-1}(\Delta_i D) = \phi^{-1} dev_i^2 + r_{P_i}^2 h_i.$$

Como $r_{D_i}^2 = dev_i^2 / \{\phi(1 - h_i)\}$, entonces $\Omega \simeq r_{D_i}^2(1 - h_i) + r_{P_i}^2 h_i \equiv r_{G_i}^2$.

Williams menciona que $r_{G_i}^2$ subestima a la estadística de cociente de verosimilitudes cuando el dato i es discrepante.

A $r_{G_i} = \sqrt{(1 - h_i)r_{D_i}^2 + h_i r_{P_i}^2}$ se le llama el i -ésimo residuo verosimil. La esperanza de r_{G_i} es distinta de cero, su varianza generalmente es mayor que uno y su sesgo es pequeño. Además, $\min\{r_{D_i}, r_{P_i}\} \leq r_{G_i} \leq \max\{r_{D_i}, r_{P_i}\}$. Williams sugiere graficar los residuos verosimiles contra i , h_i ó η_i . También puede hacerse una gráfica de probabilidad empírica para r_{G_i} con $i = 1, 2, \dots, n$.

Una generalización de las hipótesis mencionadas arriba es:

$$H_0 : \eta = X\beta \quad \text{vs.} \quad H_1 : \eta = X\beta + \lambda_i d_i \quad \text{con} \quad \lambda_i \neq 0 \quad \text{para alguna } i;$$

d_i es el i -ésimo vector canónico. El método de Monte-Carlo consiste en comparar la estadística de prueba: $\max_i\{r_{G_i}^2\}$ con el correspondiente valor obtenido al simular datos del modelo $\hat{\eta} = X\hat{\beta}$.

Cómputo. En GLIM3, el macro RESV (apéndice B4) calcula el cuadrado de los residuos verosimiles de cualquier modelo logístico. El macro no tiene argumentos y antes de usarlo debe ajustarse el modelo logístico. Las cantidades $r_{G_i}^2$ se almacenan en el vector "rg2". Adicionalmente se obtienen el cuadrado de los residuos de Pearson (rp2) y de la devianza (rd2). Cuando alguna y_i es igual a n_i ó 0, GLIM3 presenta el mensaje:

"- - invalid function/operator argument(s)"

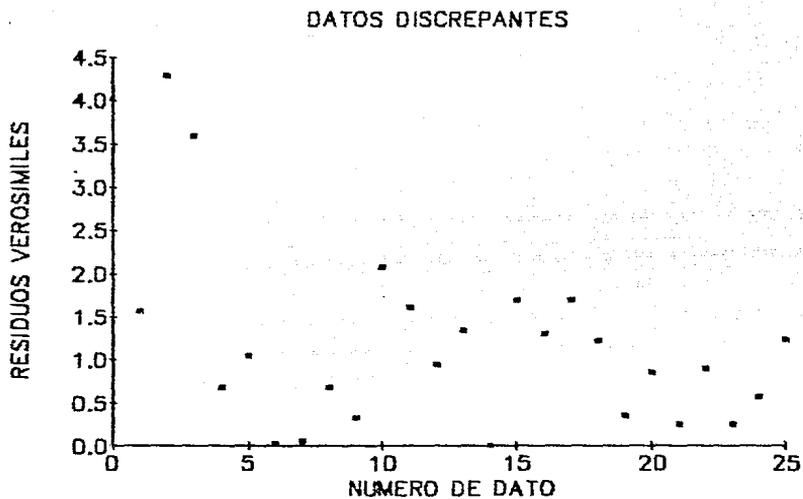
Este mensaje debe ignorarse si los dos valores que aparecen en la pantalla para X^2 y D^2 son iguales.

Ejemplo. Se calculó el cuadrado de los residuos verosimiles para los datos del apéndice A3 y el modelo:

$$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{edad.}$$

La gráfica de los puntos ($i, r_{G_i}^2$) muestra que los datos 2 y 3 están pobremente representados por el modelo. El residuo verosimil más grande es del dato 2.

Gráfica.



Bibliografía abreviada.

Cook, D. R. y Weisberg, S. (1982).

Jennings, Dennis E. (1986).

Pregibon, D. (1981, 1982).

Williams, D. A. (1987).

6.1.3 Datos agrupados.

Las técnicas post-ajuste de esta sección son — cada una por una razón distinta — más apropiadas para datos agrupados que no agrupados.

Por ejemplo, la prueba de bondad de liga permite determinar si la transformación logística para p_i es adecuada. El procedimiento consiste en elegir una familia de funciones liga que incluya a la logística y probar la hipótesis de que ésta es la más apropiada. La mayoría de los autores que han propuesto estas familias suponen que $n_i > 1$.

La variabilidad extra-binomial generalmente se presenta en respuestas asociadas a variables explicativas categóricas. Los datos frecuentemente se reportan en una tabla de contingencia, por lo que están agrupados.

Williams (1987) propuso un índice para medir la influencia global de un dato en el cociente de verosimilitudes de un prueba de hipótesis para modelos lineales generalizados anidados. Sin embargo, el autor primordialmente tuvo en mente a respuestas con distribución Poisson o binomial con $n_i > 1$.

6.1.3.1 Prueba de bondad de liga.

Pregibon (1980) diseñó una prueba de hipótesis para determinar si, en un modelo lineal generalizado, la liga hipotetizada describe apropiadamente la relación entre $\mu = E(Y)$ y $\eta = x'\beta$. Adicionalmente, la prueba produce un estimador de la liga correcta.

Primero se verá el desarrollo teórico de la prueba de bondad de liga — que es similar al de la variable construida. Después se verá cómo aplicar la prueba al modelo logístico.

Los supuestos de la prueba de bondad de liga son que:

1. La distribución de Y es correcta; y
2. Las variables explicativas relevantes están incluidas en el predictor lineal.

La familia de funciones liga definida por los parámetros α y δ se denotará como $g(\mu; \alpha, \delta)$. Se supondrá que la liga hipotetizada, $g_0(\mu) = g(\mu; \alpha_0, \delta_0)$, y la liga correcta, $g_*(\mu) = g(\mu; \alpha_*, \delta_*)$, pertenecen a dicha familia. Las hipótesis a probar son:

$$H_0 : g_*(\mu) = g_0(\mu) \quad \text{i.e.} \quad \alpha_* = \alpha_0 \quad \text{y} \quad \delta_* = \delta_0$$

vs.

$$H_1 : g_*(\mu) \neq g_0(\mu) \quad \text{i.e.} \quad \alpha_* \neq \alpha_0 \quad \text{ó} \quad \delta_* \neq \delta_0.$$

El desarrollo en series de Taylor de $g_*(\mu)$ alrededor de (α_0, δ_0) produce la aproximación lineal siguiente para $g_*(\mu)$:

$$g_*(\mu) \simeq g_0(\mu) + (D_\alpha(g_0(\mu)), D_\delta(g_0(\mu))) \begin{pmatrix} \alpha_* - \alpha_0 \\ \delta_* - \delta_0 \end{pmatrix}.$$

$$\text{con } D_\alpha(g_0(\mu)) = \left. \frac{\partial g(\mu; \alpha, \delta)}{\partial \alpha} \right|_{\substack{\alpha=\alpha_0 \\ \delta=\delta_0}} \quad \text{y} \quad D_\delta(g_0(\mu)) = \left. \frac{\partial g(\mu; \alpha, \delta)}{\partial \delta} \right|_{\substack{\alpha=\alpha_0 \\ \delta=\delta_0}}.$$

Si $z' \equiv (D_\alpha(g_0(\mu)), D_\delta(g_0(\mu)))$, $\gamma' \equiv -(\alpha_* - \alpha_0, \delta_* - \delta_0)$ y $g_*(\mu) = X\beta$, entonces

$$g_0(\mu) \simeq X\beta + z'\gamma.$$

El término $z'\gamma$ describe las discrepancias entre la liga hipotetizada y la liga correcta. La matriz z' tiene tantas columnas (variables construidas) como parámetros tenga la familia de funciones liga. En términos del vector γ , las hipótesis planteadas arriba son:

$$H_0 : \gamma = 0 \quad \text{vs.} \quad H_1 : \gamma \neq 0.$$

En GLIM3, la prueba de bondad de liga se realiza como sigue. Primero se ajusta el modelo con función liga $g_0(\mu)$ y componente sistemática $X\beta$. Con los valores ajustados se calcula la matriz $\hat{z}' = (D_\alpha(g_0(\hat{\mu})), D_\delta(g_0(\hat{\mu})))$. Después se ajusta el modelo: $g_0(\mu) = X\beta + \hat{z}'\gamma$. Si el cambio producido en la devianza es significativo, la liga hipotetizada es incorrecta. Un estimador de la liga correcta es: $g(\mu; \hat{\alpha}_*, \hat{\delta}_*)$; $\hat{\alpha}_*$ y $\hat{\delta}_*$ se obtienen a partir del vector: $\hat{\gamma}' = -(\hat{\alpha}_* - \alpha_0, \hat{\delta}_* - \delta_0)$.

A continuación se mencionan varias familias de funciones liga para variables de respuesta con distribución binomial. La liga logística es un caso particular de ellas.

Pregibon propuso una familia de ligas con dos parámetros para respuestas binomiales:

$$g(\mu; \alpha, \delta) = \frac{\left(\frac{\mu}{n}\right)^{\alpha-\delta} - 1}{\alpha - \delta} - \frac{\left(1 - \frac{\mu}{n}\right)^{\alpha+\delta} - 1}{\alpha + \delta} \quad \text{con } \mu = np.$$

La liga logística se obtiene cuando $(\alpha, \delta) \rightarrow 0$. Las variables construidas para probar si la liga logística es correcta son:

$$z_1 = D_\alpha(g(\mu)) \Big|_{(\alpha, \delta)=0} = \frac{1}{2} \left[\log^2 \left(\frac{\mu}{n} \right) - \log^2 \left(1 - \frac{\mu}{n} \right) \right];$$

$$z_2 = D_\delta(g(\mu)) \Big|_{(\alpha, \delta)=0} = -\frac{1}{2} \left[\log^2 \left(\frac{\mu}{n} \right) + \log^2 \left(1 - \frac{\mu}{n} \right) \right].$$

Aranda-Ordaz (1981) definió una familia de transformaciones simétrica y otra asimétrica para p_i . La familia simétrica es:

$$T_\lambda(p) = \frac{2}{\lambda} \left[\frac{p^\lambda - (1-p)^\lambda}{p^\lambda + (1-p)^\lambda} \right].$$

$T_\lambda(p)$ es simétrica porque los resultados no se alteran, excepto por el signo, al codificar los éxitos como fracasos o viceversa; es decir: $T_\lambda(p) = -T_\lambda(1-p)$ y $T_\lambda(p) = T_{-\lambda}(p)$. La liga logística se obtiene cuando $\lambda \rightarrow 0$. Cuando la hipótesis a probar es:

$$H_0 : g_*(\mu) = \text{logit}(p) \quad \text{vs.} \quad H_1 : g_*(\mu) \neq \text{logit}(p),$$

la variable construida es la derivada de $T_\lambda(p)$ respecto a λ evaluada en $\lambda = 0$.

En los problemas relacionados con valores extremos es conveniente tratar asimétricamente los éxitos y los fracasos. La familia asimétrica de Aranda-Ordaz es:

$$g(\mu; \alpha) = \log \left\{ \frac{(1-p)^{-\alpha} - 1}{\alpha} \right\}.$$

El parámetro α describe la curtosis de la distribución de Y . Dos funciones liga que pertenecen a esta familia son la logística ($\alpha = 1$) y la doble log-complementaria ($\alpha \rightarrow 0$). La variable construida para probar si la liga logística es correcta es:

$$z = \frac{\partial g(\mu; \alpha)}{\partial \alpha} \Big|_{\alpha=1} = - \left[1 + \frac{1}{p} \log(1-p) \right].$$

La familia de Guerrero y Johnson (1982) se obtiene aplicando la transformación de potencias de Box y Cox al cociente de momios:

$$\left(\frac{p_i}{1-p_i} \right)^{(\lambda)} = \begin{cases} \log\left(\frac{p_i}{1-p_i}\right) & \text{si } \lambda = 0; \\ \frac{1}{\lambda} \left[\left(\frac{p_i}{1-p_i}\right)^\lambda - 1 \right] & \text{si } \lambda \neq 0. \end{cases}$$

La variable construida para probar si la liga logística es correcta es:

$$z = \frac{1}{2} \log^2 \left(\frac{p}{1-p} \right).$$

La liga sugerida por la prueba no es óptima cuando $g_*(\mu)$ no pertenece a la familia de ligas seleccionada. Además, cuando los valores ajustados del modelo $g_0(\mu) = X\beta$ son casi constantes, la aproximación lineal a $g_*(\mu)$ no es válida y la prueba es inservible.

No se recomienda comparar cada coeficiente estimado con su desviación estándar porque la correlación entre $\hat{\beta}$ y $\hat{\gamma}$ es alta. Esto se debe a que la muestra no proporciona suficiente información para estimar simultáneamente la liga correcta y el vector de coeficientes del modelo.

Ejemplo. El ajuste del modelo $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{edad}$ a los datos del apéndice A3 produjo una devianza de 26.703 con 23 grados de libertad. En términos generales, el modelo es adecuado. Sin embargo, se analizará si en particular la liga hipotetizada — la logística — es correcta. La prueba de bondad de liga se realizó para distintas familias de funciones liga.

Primero se trabajó con la familia de Pregibon. La devianza del modelo:

$$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{edad} + z_1 \gamma_1 + z_2 \gamma_2.$$

fue 15.234 con 21 grados de libertad. Es decir, cuando z_1 y z_2 se incluyen en η , el decremento producido en la devianza es 11.469. El nivel de significancia descriptivo es menor que 0.005 lo que indica que la liga logística es incorrecta. A partir de los

estimadores de γ_1 y γ_2 se obtiene que $\hat{\alpha}_* = -0.1038 \approx -0.1$ y $\hat{\delta}_* = 0.1977 \approx 0.2$.
Entonces, el estimador de la liga correcta es:

$$g(\mu; \hat{\alpha}_*, \hat{\delta}_*) = \frac{p^{-0.3} - 1}{-0.3} - \frac{(1-p)^{0.1} - 1}{0.1} = -\frac{10}{3} [3(1-p)^{0.1} + p^{-0.3} - 4].$$

Esta función no se parece a ninguna liga conocida lo que limita su aplicación.

Cuando se usa la familia de transformaciones asimétrica de Aranda-Ordaz y de Guerrero y Johnson, el cambio producido en la devianza es 3.078 y 3.501 respectivamente. El nivel de significancia descriptivo es ligeramente menor a 0.05 y la conclusión es ambigua: no es claro si la liga debe modificarse.

Adicionalmente se ajustó el modelo probit:

$$\Phi^{-1}(p) = \delta_1 + \delta_2 \cdot \text{edad}.$$

La devianza obtenida fue 22.887 con 23 grados de libertad. Los datos se ajustan al modelo probit que tiene el atractivo de que depende de una liga conocida.

Bibliografía abreviada.

- Aranda-Ordaz, F. J. (1981).
- Guerrero, V. M. y Johnson, R. A. (1982).
- Pregibon, D. (1980).

6.1.3.2 Variabilidad extra-binomial.

Un modelo logístico tiene variabilidad extra-binomial cuando el parámetro de dispersión es distinto de uno: $\phi \neq 1$. Esto quiere decir que la distribución binomial no describe apropiadamente la variabilidad de la respuesta. A la variabilidad extra-binomial se le llama sobredispersión si $\phi > 1$ y, en la práctica, esto es más común que la situación contraria: $\phi < 1$. La variabilidad extra-binomial puede ser causada por la omisión de alguna variable explicativa relevante o porque la distribución de la respuesta es incorrecta. Los modelos para datos con variabilidad extra-binomial sólo deben usarse cuando, después de incluir a todas las variables explicativas en η , $\phi \neq 1$. Generalmente, esto sólo sucede cuando se trabaja con variables explicativas categóricas.

Williams (1982) modeló la variabilidad extra-binomial con n variables aleatorias independientes y no observables: S_1, S_2, \dots, S_n . Estas variables deben ser continuas en $(0, 1)$; además, $E(S_i) = p_i$ y $Var(S_i) = \alpha p_i(1 - p_i)$. La distribución de la respuesta, Y_i , condicionada a que $S_i = s_i$ es una $Bin(n_i, p_i)$ con

$$p_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$$

Este es el modelo logístico para datos con variabilidad extra-binomial. Los parámetros, α y β , se estiman a partir de la cuasi-verosimilitud de Y_i . El método de máxima verosimilitud no es aplicable porque lo único que se sabe de la distribución incondicional de Y_i es que $E(Y_i) = n_i p_i$ y $Var(Y_i) = n_i p_i(1 - p_i) v_i^{-1}$ con $v_i^{-1} = 1 + \alpha(n_i - 1)$.

Suponga por el momento que α es conocido. $E(Y_i) = n_i p_i$ es una función de β y su desarrollo en series de Taylor alrededor del estimador β^* produce el modelo lineal aproximado:

$$E(Y_i) = n_i p_i \approx n_i p_i^* + w_i^* \sum_{j=1}^m x_{ij} (\beta_j - \beta_j^*)$$

$$\text{con } Var(Y_i) = v_i^{-1} w_i^*, \quad w_i^* = n_i p_i^* (1 - p_i^*) \quad \text{y} \quad p_i^* = \frac{\exp(x_i' \beta^*)}{1 + \exp(x_i' \beta^*)}$$

En términos de $Y_i^* = \sum_{j=1}^m x_{ij} \beta_j^* + \frac{1}{w_i^*} (Y_i - n_i p_i^*)$, el modelo anterior es equivalente a:

$$E(Y^*) \simeq X\beta \quad \text{con} \quad \text{Var}(Y^*) \simeq (VW^*)^{-1};$$

$$V = \text{diag}\{v_i\} \quad \text{y} \quad W^* = \text{diag}\{n_i p_i^*(1 - p_i^*)\}.$$

El estimador de β , $\hat{\beta}_{CV}$, es la solución a las ecuaciones de cuasi-verosimilitud:

$$X'VW^*X\hat{\beta}_{CV} = X'VW^*Y^*.$$

Si $n_i = k$ para $i = 1, 2, \dots, n$ entonces V es proporcional a la matriz identidad. Como consecuencia, el estimador de β es igual al del modelo logístico para datos sin variabilidad extra-binomial. Sin embargo, la varianza de los estimadores es distinta, la de $\hat{\beta}_{CV}$ debe escalarse por $1 + \alpha(k - 1)$, el factor de heterogeneidad (es decir, el parámetro de dispersión ϕ).

Una estadística de bondad de ajuste para el modelo logístico con variabilidad extra-binomial es:

$$X^2 = \sum_{i=1}^n \frac{v_i(Y_i - n_i\hat{p}_i)^2}{n_i\hat{p}_i(1 - \hat{p}_i)}.$$

X^2 es aproximadamente igual a $(Y^* - X\hat{\beta}_{CV})'VW^*(Y^* - X\hat{\beta}_{CV})$, la suma de cuadrados de los residuos ponderados. Si estos residuos se distribuyen asintóticamente como una normal, entonces

$$E(X^2) \simeq \begin{cases} n - m + \alpha \sum_{i=1}^n (n_i - 1)(1 - w_i^* q_i) & \text{si } V = I; \\ \sum_{i=1}^n v_i(1 - v_i w_i^* q_i)[1 + \alpha(n_i - 1)] & \text{si } V \neq I; \end{cases}$$

q_i es el i -ésimo elemento en la diagonal de $X(X'VW^*X)^{-1}X'$, la matriz de varianzas y covarianzas del predictor lineal.

Si α es desconocido, los pesos (v_i) se calculan con un estimador inicial de él. Los pesos sirven para estimar a β . Para obtener un estimador revisado de α se iguala la estadística X^2 con su esperanza aproximada. Con el nuevo estimador de α se recalculan los pesos y se reestima a β . Cuando $n_i = k$ para $i = 1, 2, \dots, n$, la expresión

$$E(X^2) \simeq n - m + \alpha \sum_{i=1}^n (n_i - 1)(1 - w_i^* q_i)$$

se reduce a: $E(X^2) \simeq (n - m)[1 + \alpha(k - 1)]$. Esto implica que $E(X^2)/(n - m)$ estima a $\phi = 1 + \alpha(k - 1)$.

El procedimiento sugerido por Williams para ajustar el modelo logístico con variabilidad extra-binomial es el siguiente:

1. Suponga que $\alpha = 0$ (o equivalentemente que $\phi = 1$). Use el método de máxima verosimilitud para ajustar el modelo logístico sin variabilidad extra-binomial. Calcule la estadística ji-cuadrada de Pearson generalizada.
2. Compare la X^2 con la distribución $X^2_{(n-m)}$. Si X^2 es inusualmente grande concluya que $\alpha \neq 0$ y estímelo con:

$$\hat{\alpha} = \frac{X^2 - (n - m)}{\sum_{i=1}^n (n_i - 1)(1 - w_i^* q_i)}$$

3. Use los pesos: $v_i = [1 + \hat{\alpha}(n_i - 1)]^{-1}$ para reestimar a β y recalculé X^2 .
4. El estimador de α es satisfactorio si $X^2 \simeq n - m$. En caso contrario reestime a α con

$$\hat{\alpha} = \frac{X^2 - \sum_{i=1}^n v_i(1 - v_i w_i^* q_i)}{\sum_{i=1}^n v_i(n_i - 1)(1 - v_i w_i^* q_i)}$$

y vuelva al paso tres.

Estos pasos pueden llevarse a cabo en GLIM3. Los pesos iniciales (v_i) se almacenan en el vector %PW y los pesos iterativos (w_i^*) en %WT. La varianza del predictor lineal, q_i , y n_i están contenidas en %VL y %BD respectivamente. Además, los escalares %X2 y %DF son la estadística X^2 y sus grados de libertad respectivamente. Suponga que las Y_i están almacenadas en Y, las n_i en M, que los pesos iniciales son unitarios y que las variables explicativas son X1, X2, ..., XM. Así, en GLIM3, el procedimiento para ajustar el modelo logístico con variabilidad extra-binomial es:

```

$YVAR Y
$ERROR B N
$LINK G
    
```

\$WEI V

\$FIT X1 + X2 + ...+ XM

\$EXTRACT %VL

\$CAL VWQ = %PW * %WT * %VL:

\$%P = (%X2 - %CU(%PW * (1 - VWQ)))/(%CU((%BD - 1) * %PW * (1 - VWQ))):

V = 1/(1 + %P * (%BD - 1))

Esta secuencia de instrucciones se repite dos o tres veces hasta que %X2 sea aproximadamente igual a %DF. GLIM3 calcula correctamente la desviación estándar de los coeficientes estimados. El estimador de α obtenido a partir del modelo que incluye a todas las variables explicativas puede usarse para ajustar modelos con menos variables.

Existen otros modelos para datos con variabilidad extra-binomial. Williams menciona uno apropiado para respuestas con probabilidad de éxito cercana a cero o uno.

Cómputo. Los macros "VEB1" y "VEB2" (apéndice B5) fueron creados para ajustar el modelo logístico para datos con variabilidad extra-binomial descrito anteriormente. Ninguno de ellos tiene argumentos, pero su aplicación requiere del ajuste previo de un modelo. Antes de usar VEB1 debe ajustarse el modelo logístico para datos sin variabilidad extra-binomial. En VEB1 se calcula el estimador inicial de α . Con él se obtienen los pesos iniciales para estimar a β . Si la estadística ji-cuadrada indica que el modelo es inadecuado se procede a usar el macro VEB2. En éste se obtiene un nuevo estimador de α para reestimar al vector de coeficientes del modelo para datos con variabilidad extra-binomial. Si el modelo sigue siendo incorrecto, se realiza otro ajuste con VEB2 hasta que la X^2 indique que el modelo describe apropiadamente a los datos.

Ejemplo. El modelo originalmente propuesto para los datos del apéndice A4 fue:

$$\text{logit}(p_{i,rs}) = \alpha + \beta_{(s)} + \gamma_{(r)} + \delta_{(sr)};$$

$p_{i,rs}$ es la probabilidad de que germine una semilla del tipo "s" cultivada en el r-ésimo extracto de raíz;

α es la media general;

$\beta_{(s)}$ es el efecto producido por el tipo de semilla ($s = 1, 2$);

$\gamma_{(r)}$ es el efecto producido por el extracto de raíz ($r = 1, 2$);

$\delta_{(sr)}$ es el efecto de interacción entre el tipo de semilla y el extracto de raíz.

A pesar de que el modelo incluye a todos los parámetros, la estadística ji-cuadrada de Pearson fue igual a 31.65 con 17 grados de libertad. A continuación se ajustó el modelo logístico con variabilidad extra-binomial descrito en esta sección. El ajuste mejoró notablemente: la X^2 fue igual a 17.34 con 17 grados de libertad. El estimador inicial de α fue 0.02371. Estos resultados son similares a los que se obtienen cuando se realiza un segundo ajuste con VEB2.

Bibliografía abreviada.

McCullagh, P. y Nelder, J.A. (1983).

Williams, D. A. (1982).

6.1.3.3 Influencia global en pruebas de hipótesis.

La sección 6.1.2.5 trató de datos que influyen en $\hat{\beta}$, en las estadísticas de bondad de ajuste, los residuos y las predicciones. Ahora se hablará de datos que influyen en la estadística de cociente de verosimilitudes de una prueba de hipótesis para modelos lineales generalizados anidados.

La estadística de cociente de verosimilitudes de una prueba de hipótesis en la que el modelo definido en H_0 es un caso particular del propuesto en H_1 es:

$$-2\ln\Lambda = \phi^{-1}\{D_0^2(\hat{\beta}) - D_1^2(\hat{\beta})\};$$

se supondrá que ϕ es conocido. Si se elimina el dato i , la estadística de prueba para las hipótesis mencionadas es:

$$-2\ln\Lambda_{(i)} = \phi^{-1}\{D_0^2(\hat{\beta}_{(i)}) - D_1^2(\hat{\beta}_{(i)})\}.$$

El dato i influye en el cociente de verosimilitudes si en un caso la prueba es significativa y en el otro no. En otras palabras, el dato i influye en el cociente de verosimilitudes si el cambio producido en la estadística de prueba es "grande". Este cambio es:

$$-2\ln\Lambda_{(i)} + 2\ln\Lambda = \phi^{-1}\{D_0^2(\hat{\beta}_{(i)}) - D_1^2(\hat{\beta}_{(i)})\} - \phi^{-1}\{D_0^2(\hat{\beta}) - D_1^2(\hat{\beta})\}$$

o bien:
$$-2\ln\Lambda_{(i)} + 2\ln\Lambda = \phi^{-1}\{D_1^2(\hat{\beta}) - D_1^2(\hat{\beta}_{(i)})\} - \phi^{-1}\{D_0^2(\hat{\beta}) - D_0^2(\hat{\beta}_{(i)})\}.$$

El término $\phi^{-1}\{D_1^2(\hat{\beta}) - D_1^2(\hat{\beta}_{(i)})\}$ es el cociente de verosimilitudes para probar si, en el modelo completo, el dato i es discrepante. Por lo tanto, una aproximación a dicho término es el i -ésimo residuo verosimil del modelo completo (véase la sección 6.1.2.6). Análogamente, el segundo término de la expresión de arriba es el cociente de verosimilitudes para probar si, en el modelo reducido, el dato i es discrepante. Esto quiere decir que $\phi^{-1}\{D_0^2(\hat{\beta}) - D_0^2(\hat{\beta}_{(i)})\} \simeq r_{G_i}^2(H_0)$. Entonces,

$$-2\ln\Lambda_{(i)} + 2\ln\Lambda \approx r_{G_i}^2(H_1) - r_{G_i}^2(H_0).$$

Si esta discrepancia es mayor que un valor crítico calculado mediante simulaciones, el dato i influye en el cociente de verosimilitudes generalizado.

Las hipótesis relacionadas con la gráfica de la variable agregada, la gráfica de la variable construida y la prueba de bondad de liga son de la forma:

$$H_0 : g(\mu) = X\beta \quad (\gamma = 0)$$

vs.

$$H_1 : g(\mu) = X\beta + z\gamma \quad (\gamma \neq 0).$$

Obsérvese que el modelo nulo está anidado en el modelo alternativo; por lo tanto, la diferencia $r_{G_i}^2(H_1) - r_{G_i}^2(H_0)$ puede usarse para medir la influencia global del dato i en el cociente de verosimilitudes.

El intervalo de verosimilitud para $\lambda(\beta)$, una función real de β , está relacionado con la hipótesis:

$$H_0 : \lambda(\beta) = t \quad \text{vs.} \quad H_1 : \lambda(\beta) \neq t.$$

La influencia del dato i en dicho intervalo puede estudiarse mediante $r_{G_i}^2(H_1) - r_{G_i}^2(H_0)$. Si el dato i se omite de la muestra, el intervalo de verosimilitud para $\lambda(\beta_{(i)})$ es:

$$\{t; -2\ln\Lambda_{(i)} \leq \chi_{(1)}^{2,1-\alpha}\} = \{t; 2\ln\Lambda + r_{G_i}^2(H_1) - r_{G_i}^2(H_0) \leq \chi_{(1)}^{2,1-\alpha}\}.$$

Este es el conjunto de valores de t para los cuales no se rechaza $H_0 : \lambda(\beta_{(i)}) = t$. El intervalo anterior puede compararse gráficamente con el obtenido a partir de los n datos.

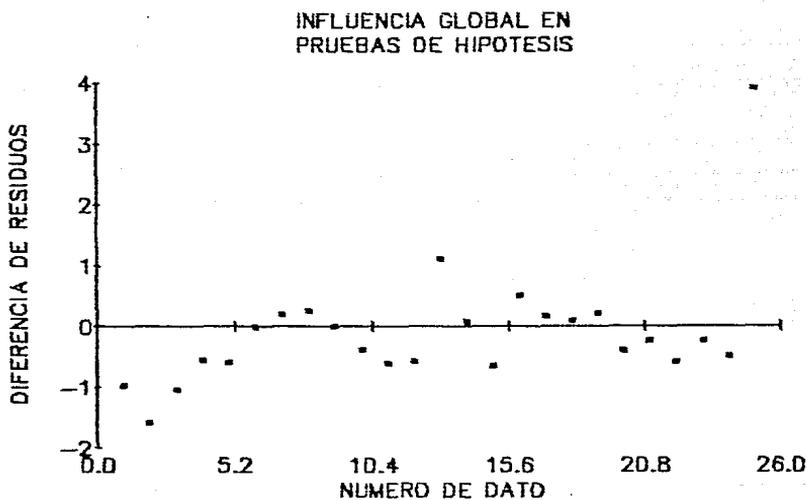
Ejemplo. En la sección 6.1.2.3 se usaron los datos del apéndice A3 para probar las hipótesis:

$$H_0 : \text{logit}(p) = \beta_0 + \beta_1 \cdot \text{edad} \quad (\phi = 0)$$

vs.

$$H_1 : \text{logit}(p) = \beta_0 + \beta_1 \cdot \text{edad} + \phi \cdot \text{edad} \cdot \log(\text{edad}) \quad (\phi \neq 0).$$

Estas hipótesis están anidadas porque el modelo propuesto en H_0 se obtiene imponiendo la restricción de que $\phi = 0$ en el modelo alternativo. La gráfica de la variable construida mostró que posiblemente el dato 25 influye en la prueba de hipótesis. Para corroborar esta afirmación se calculó la diferencia entre los residuos verosímiles de los modelos alternativo y nulo, es decir: $r_{G_i}^2(H_1) - r_{G_i}^2(H_0)$. Estas diferencias se graficaron contra el número de dato. La gráfica muestra que el dato 25 es el más influyente.



Bibliografía abreviada.

Williams, D. A. (1987).

6.2 Respuestas suavizadas.

Cuando la respuesta del modelo logístico es binaria, las gráficas de residuos están formadas por dos nubes de puntos: una para los éxitos y otra para los fracasos. Esto impide realizar un análisis post-ajuste como el que se efectúa en regresión lineal normal. La naturaleza binaria de los residuos puede eliminarse suavizando la variable de respuesta del modelo logístico.

Fowlkes (1987) propuso calcular la i -ésima respuesta suavizada, Y_{i_s} , con el promedio ponderado:

$$Y_{i_s} = \sum_k \left(\frac{w_k Y_k}{\sum_k w_k} \right) \quad \text{para } i = 1, 2, \dots, n.$$

El número de términos en la sumatoria depende del tamaño de la ventana de suavizamiento. Fowlkes recomienda usar entre 10 y 15% de los datos para suavizar cada respuesta. Las Y_k deben ser las más cercanas a Y_i en el espacio de variables explicativas. El peso, w_k , es una función decreciente de la distancia entre x_i y x_k . Generalmente, w_k varía entre cero y uno. Los pesos pueden definirse como:

$$w_k = 1 - \left(\frac{d(x_k, x_i)}{d_{i_s}} \right)^3$$

en donde $d(x_k, x_i)$ es la distancia, en alguna métrica, entre x_k y x_i y $d_{i_s} = \max_k \{d(x_k, x_i)\}$. Por ejemplo, si sólo se tiene una variable explicativa:

$$w_k = 1 - \left(\frac{|x_k - x_i|}{d_{i_s}} \right)^3 \quad \text{con } d_{i_s} = \max_k \{|x_k - x_i|\}.$$

Cuando se tienen varias variables explicativas correlacionadas es conveniente usar la distancia de Mahalanobis entre x_i y x_k . La distancia Euclidiana entre las variables explicativas estandarizadas es recomendable cuando se ignora cuál es la métrica más apropiada.

Fowlkes definió el i -ésimo residuo de Pearson suavizado como:

$$X_{i_s}^s = \frac{\hat{p}_i^s - \hat{p}_i}{\hat{\sigma}(\hat{p}_i^s)}$$

en donde

$$\hat{p}_i = E(Y_i),$$

$$\hat{p}_i^e = \hat{E}(Y_{i\cdot}) = \sum_k \frac{w_k \hat{p}_k}{\sum_k w_k};$$

$$y \quad \hat{\sigma}(\hat{p}_i^e) = \sqrt{\widehat{\text{Var}}(Y_{i\cdot})} = \sqrt{\frac{\sum_k w_k^2 \hat{p}_k (1 - \hat{p}_k)}{(\sum_k w_k)^2}}.$$

Para calcular $\hat{\sigma}(\hat{p}_i^e)$ se usó el supuesto de que la correlación entre las respuestas binarias es nula. El modelo ajustado es adecuado si la gráfica de los residuos suavizados contra los valores ajustados, (\hat{p}_i, X_i^e) , muestra un comportamiento aleatorio. También puede hacerse una gráfica de probabilidad normal para los residuos suavizados porque X_i^e se distribuye asintóticamente como una normal estandarizada.

Los residuos suavizados permiten generalizar la gráfica de residuos parciales al modelo logístico. Esta gráfica indica si la variable explicativa z debe incluirse linealmente en el predictor lineal. Varios autores han propuesto distintos residuos parciales para el modelo logístico. Fowlkes (1987) los definió a partir del desarrollo en series de Taylor de $\log(\hat{p}_i^e/(1 - \hat{p}_i^e))$ alrededor de p_i :

$$\log \left(\frac{\hat{p}_i^e}{1 - \hat{p}_i^e} \right) \simeq \log \left(\frac{p_i}{1 - p_i} \right) + \frac{\hat{p}_i^e - p_i}{p_i(1 - p_i)} = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \frac{\hat{p}_i^e - p_i}{p_i(1 - p_i)}.$$

Si $x_j = z$ y $\beta_j = \tau$ entonces:

$$\log \left(\frac{\hat{p}_i^e}{1 - \hat{p}_i^e} \right) \simeq \beta_0 + \sum_{l \neq j} \beta_l x_{il} + \tau z_i + \frac{\hat{p}_i^e - p_i}{p_i(1 - p_i)}.$$

Los estimadores de máxima verosimilitud de β_l , τ y p_i ($\hat{\beta}_l$, $\hat{\tau}$ y \hat{p}_i respectivamente) se obtienen al ajustar el modelo:

$$\text{logit}(p_i) = \beta_0 + \sum_{l \neq j} \beta_l x_{il} + \tau z_i \quad \text{para } i = 1, 2, \dots, n.$$

con ellos se calculan los residuos parciales suavizados:

$$r_i^* = \frac{\hat{p}_i^* - \hat{p}_i}{\hat{p}_i(1 - \hat{p}_i)} + \hat{r}z_i \quad \text{para } i = 1, 2, \dots, n.$$

Landwehr *et. al.* (1984) definieron los residuos parciales como:

$$r_i = \frac{y_i - \hat{p}_i}{\hat{p}_i(1 - \hat{p}_i)} + \hat{r}z_i \quad \text{para } i = 1, 2, \dots, n;$$

después los suavizaron. Los autores se basaron en el algoritmo de mínimos cuadrados ponderados; por esto la analogía entre r_i y los residuos parciales de regresión lineal normal.

La gráfica de residuos parciales está constituida por los puntos (z_i, r_i^*) . Si los puntos muestran una tendencia lineal con pendiente aproximadamente igual a \hat{r} , entonces z debe incluirse linealmente en η . De lo contrario es necesario transformar a z .

Fowlkes usó $\hat{E}(Y_{i,s}) = \hat{p}_i^*$ para elaborar una gráfica similar a la de la devianza media local que llamó: "la gráfica de la devianza delta local". En ésta:

$$d_i^* = -2\{y_i \log(\hat{p}_i^*) + (1 - y_i) \log(1 - \hat{p}_i^*)\}$$

es un estimador local de la devianza del dato i . La devianza del i -ésimo dato calculada con las respuestas sin suavizar es:

$$d_i^L = -2\{y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)\}.$$

Para hacer la gráfica de la devianza delta local, d_i^* y d_i^L se ordenan de acuerdo a la magnitud de $d_{i_s} = \max_k \{d(x_k, x_i)\}$. De esta forma se obtienen las estadísticas de orden:

$$d_{(1)}^* \leq d_{(2)}^* \leq \dots \leq d_{(n)}^* \quad \text{y} \quad d_{(1)}^L \leq d_{(2)}^L \leq \dots \leq d_{(n)}^L.$$

Después se calculan las devianzas acumuladas:

$$D_i^* = \sum_{r=1}^i d_{(r)}^* \quad \text{y} \quad D_i^L = \sum_{r=1}^i d_{(r)}^L$$

y, con el algoritmo MLOW (Fowlkes 1987), sus respectivos grados de libertad:

$$N_i^s = i \left(\frac{GL^s}{n} \right) \quad \text{y} \quad N_i^L = i \left(\frac{GL^L}{n} \right).$$

Posteriormente se grafican los puntos $(i, \frac{D_i^s}{N_i^s} - \frac{D_i^L}{N_i^L})$ para $i = 1, 2, \dots, n$. Existe carencia de ajuste en el modelo logístico cuando los puntos anteriores no se estabilizan alrededor de la recta $y = 0$. Obsérvese que esta gráfica, a diferencia de la que propusieron Landwehr *et. al.*, está formada por n puntos y la recta de comparación ($y = 0$) no varía de un problema a otro.

Las simulaciones realizadas por Fowlkes para estudiar el efecto de n y m en \hat{p}_i^s y $\hat{\sigma}(\hat{p}_i^s)$ indican que:

1. En el caso de variables explicativas uniformes, el sesgo de \hat{p}_i^s aumenta a medida que m crece.
2. El valor de $\hat{\sigma}(\hat{p}_i^s)$ se altera cuando el número de variables explicativas (m) varía y el tamaño de la muestra (n) es constante. Además, los residuos de Pearson suavizados se distribuyen aproximadamente como una $N(0, 1)$ cuando la distribución de Y_i es simétrica respecto a p_i . El sesgo de la devianza local, d_i^L y de la devianza delta local, $d_i^s - d_i^L$, es pequeño y $\widehat{Var}(d_i^s - d_i^L) < \widehat{Var}(d_i^s)$.

El análisis post-ajuste basado en respuestas binarias suavizadas debe hacerse con precaución: las $Y_{i,s}$ están correlacionadas — a diferencia de las variables originales. Además, el tamaño de la ventana de suavizamiento afecta el sesgo y la varianza de \hat{p}_i^s . Si la ventana es pequeña, el sesgo de \hat{p}_i^s es pequeño pero su varianza es grande. Las variables explicativa irrelevantes también afectan al estimador de $p_i^s = E(Y_{i,s})$.

Cómputo. El macro "SUAVIZA" calcula los residuos de Pearson suavizados, X_i^s , de los modelos logísticos con un máximo de seis variables explicativas. El modelo debe ajustarse antes de usar el macro. SUAVIZA tiene varios argumentos. El primero es el tamaño de la ventana de suavizamiento (expresado como un porcentaje de n), el segundo es el número de variables explicativas en η , el tercero es el vector de respuestas (%YV), y el resto son los vectores de variables explicativas (estandarizadas si se quiere). El macro calcula la distancia Euclidiana entre las x_i .

Aparentemente, SUAVIZA calcula correctamente las respuestas suavizadas, $(Y_{i,s})$, y

el estimador de p_i^* , pero no los residuos de Pearson suavizados. Quizás el problema es el cómputo de $\hat{\sigma}(\hat{p}_i^*)$.

Bibliografía abreviada.

Fowlkes, Edward, B. (1987).

Landwehr, J. M.; Pregibon, D. y Shoemaker, A. C. (1984).

7. Conclusiones.

Cada técnica post-ajuste permite identificar, si existe, un problema específico, una distinta anomalía del modelo. Por esto, la mayoría de las técnicas estudiadas se complementan; sólo algunas son equivalentes. Un ejemplo de las últimas son las medidas de influencia global en las pruebas de hipótesis para modelos lineales generalizados anidados y las gráficas para la variable agregada y construida.

Varias de las técnicas descritas son aplicables a cualquier modelo lineal generalizado; no fueron diseñadas exclusivamente para el modelo logístico. Las expresiones para $\partial L(\beta)/\partial\beta$ y $\partial^2 L(\beta)/\partial\beta\partial\beta'$ son más explícitas cuando se usa la liga "natural" en lugar de la usual: $g(\mu) = \eta$. Esto permitió desarrollar algunas fórmulas explícitas para las medidas de influencia local.

El método iterativo de mínimos cuadrados ponderados es la base para generalizar al modelo logístico las técnicas post-ajuste de regresión lineal normal. La generalización no es inmediata ni sencilla porque la matriz de pesos es distinta de una iteración a otra y de un modelo a otro. El estimador a un paso de β permitió generalizar la distancia de Cook al modelo logístico y agilizar la construcción de la gráfica de probabilidad empírica.

También se ha propuesto suavizar la respuesta del modelo logístico para poder aplicar las técnicas post-ajuste de regresión lineal normal. Este fue el procedimiento usado para elaborar la gráfica de residuos parciales.

Algunas técnicas post-ajuste sólo identifican un problema, otras lo identifican y sugieren un remedio. Entre las primeras se encuentran las medidas de influencia (global y local). A las segundas pertenecen la prueba de bondad de liga y la prueba para determinar si la variable construida es significativa.

El tema de variabilidad extra-binomial se trató parcialmente porque, salvo casos excepcionales, este problema no se presenta en los modelos de regresión logística.

En la literatura existe poco acerca de colinealidad en los modelos lineales generalizados o en el modelo logístico. Con respecto a las palancas, conviene hacer un análisis comparativo entre las de Pregibon y las de Thomas y Cook. Asimismo, se carece de un algoritmo que suavice correctamente las respuestas binarias. Finalmente, la gráfica de la variable construida es aplicable a la prueba de bondad de liga cuando la familia

de funciones liga depende de un parámetro.

Bibliografía.

1. **Aitkin, Murray; Anderson, Dorothy; Francis, Brian y Hinde, John:** "Statistical Modelling in GLIM." Oxford University Press, New York, 1989.
2. **Aranda-Ordaz, F. J. (1981):** "On two families of transformations to additivity for binary response models." *Biometrika*; vol. 68, pp.357 - 363.
3. **Andrews, David, F. y Herzberg, M.:** "A Collection of Problems From Many Fields for the Student and Research Worker." Springer Verlag, New York, 1985.
4. **Chatterjee, Samprit y Hadi, A. S.:** "Sensitivity Analysis in Linear Regression." J. Wiley, New York, 1988.
5. **Cook, D. R. (1986):** "Assessment of Local Influence." *J. R. Statist. Soc. B*; vol. 48. pp. 133 - 169.
6. **Cook, D. R. y Welsberg, S.:** "Residuals and Influence in Regression." Chapman & Hall, New York, 1982.
7. **Cox, D. R. y Snell:** "Analysis of Binary Data." Chapman & Hall, New York, 2a. ed., 1989.
8. **Daniel, C. y Wood, F. S.:** "Fitting Equations to Data." John Wiley, New York, 1971.
9. **Davison, A. C. (1989):** "Model-Checking II: Binary Data." *Revista Brasileira de Probabilidade e Estatística*; vol. 3, pp. 87 - 96.
10. **Fowlkes, Edward, B. (1987):** "Some diagnostics for binary logistic regression via smoothing." *Biometrika*; vol. 74, pp. 503 - 515.
11. "Generalised Linear Interactive Modelling Manual. The GLIM System, Release 3.77." Editor: Payne, C. D., Numerical Algorithms Group Ltd., 1986.
12. **Green, P. J. (1984):** "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and Some Robust and Resistant Alternatives." *J. R. Statist. Soc. B*; vol. 46, pp. 149 - 192.
13. **Guerrero, V. M. y Johnson, R. A. (1982):** "Use of the Box-Cox transformation with binary response models." *Biometrika*; vol. 69, pp. 309 - 314.

14. Hosmer, D. W. y Lemeshow, S.: "Applied Logistic Regression." John Wiley & Sons, New York, 1989.
15. "IMSL Library, User's Manual. Fortran Subroutines for Mathematics and Statistics." IMSL, Inc., U.S.A., 1984.
16. Jennings, D. E. (1986): "Outliers and Residual Distributions in Logistic Regression." Journal of the American Statistical Association; vol. 81, pp. 987 - 990.
17. Johnson, Wesley (1985): "Influence measures for logistic regression: Another point of view." Biometrika; vol. 72, pp. 59 - 65.
18. Kalbfleisch, J. G.: "Probability and Statistical Inference", vol. 2. Springer Verlag, New York, 2a. ed., 1985.
19. Landwehr, J. M.; Pregibon, D. y Shoemaker A. C. (1984): "Graphical Methods for Assessing Logistic Regression Models." Journal of the American Statistical Association; vol. 79, pp. 61 - 83.
20. Mackinnon, M. J. y Puterman, M. L. (1989): "Collinearity in Generalized Linear Models." Commun. Statist.- Theory Meth.; 18(9), pp. 3463 - 3472.
21. McCullagh P. y Nelder, J. A.: "Generalized Linear Models." Chapman & Hall, New York, 2a. ed., 1989.
22. O'Hara Hines, R. J. y Carter, E. M. (1990): "Improved Added Variable and Partial Residual Plots for the Detection of Influential Observations in Generalized Linear Models." Statistics Technical Report Series, 14. University of Waterloo; Department of Statistics and Actuarial Science.
23. Pregibon, D. (1980): "Goodness of Link Tests for Generalized Linear Models." Applied Statistics, vol. 29, pp. 15 - 24.
24. Pregibon D. (1981): "Logistic Regression Diagnostics." The Annals of Statistics, vol. 9, pp. 705 - 724.
25. Pregibon D.: "Score Tests in GLIM with Applications. GLIM 82: Proceeding of the International Conference." Springer Verlag, New York, 1982.
26. Thomas, W. y Cook, D. R. (1990): "Assessing Influence on Predictions From Generalized Linear Models." Technometrics; vol. 32, pp. 59 -65.

27. **Walker, Esteban** (1989): "Detection of Collinearity-Influential Observations." *Commun. Statist.- Theory Meth.*; 18(5), pp. 1675 - 1690.
28. **Wang, P. C.** (1985): "Adding a Variable in Generalized Linear Models." *Technometrics*; vol. 27, pp. 273 - 276.
29. **Wang, P. C.** (1987): "Residual Plots for Detecting Nonlinearity in Generalized Linear Models." *Technometrics*; vol 29, pp. 435 - 438.
30. **Wichmann, B. A. y Hill, I. D.** (1982): "An Efficient and Portable Pseudo-random Number Generator. Algorithm AS 183." *Applied Statistics*; pp. 188 - 190.
31. **Williams, D. A.** (1982): "Extra-binomial Variation in Logistic Linear Models." *Applied Statistics*; vol. 31, pp. 144 - 148.
32. **Williams, D. A.** (1987): "Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions." *Applied Statistics*; vol. 36, pp. 181 - 191.

Apéndice A: Datos.

Apéndice B: Subrutinas.

Apéndice A1.

Los datos listados fueron recabados para determinar si ciertas propiedades físicas de la orina están relacionadas con la formación de cristales. Las propiedades físicas estudiadas son:

1. Gravedad específica (*Grave*), es la densidad de la orina con respecto al agua.
2. *ph* (*ph*).
3. Osmolaridad (*Osm*), unidad usada en medicina y biología.
4. Conductividad (*Cond*).
5. Concentración de urea (*Urea*), milimoles/litro.
6. Concentración de calcio (*Cal*), milimoles/litro.

La variable de respuesta, *Cris*, vale uno si no se encontraron cristales en la orina y dos si los hubo. La muestra es de tamaño 79. Dos individuos, el 1 y el 55, tienen un valor faltante.

Referencia. Andrews, David F. y Herzberg, M. : "Problems From Many Fields for the Student and Research Worker"; Springer Verlag, New York, 1985, pp. 249 - 251.

<i>Cris</i>	<i>Grave</i>	<i>ph</i>	<i>Osm</i>	<i>Cond</i>	<i>Urea</i>	<i>Cal</i>
1	1.021	4.91	725	-	443	2.45
1	1.017	5.74	577	20.0	296	4.49
1	1.008	7.20	321	14.9	101	2.36
1	1.011	5.51	408	12.6	224	2.15
1	1.005	6.52	187	7.5	91	1.16
1	1.020	5.27	668	25.3	252	3.34
1	1.012	5.62	461	17.4	195	1.40
1	1.029	5.67	1107	35.9	550	8.48
1	1.015	5.41	543	21.9	170	1.16
1	1.021	6.13	779	25.7	382	2.21
1	1.011	6.19	345	11.5	152	1.93
1	1.025	5.53	907	28.4	448	1.27
1	1.006	7.12	242	11.3	64	1.03
1	1.007	5.35	283	9.9	147	1.47
1	1.011	5.21	450	17.9	161	1.53
1	1.018	4.90	684	26.1	284	5.09
1	1.007	6.63	253	8.4	133	1.05
1	1.025	6.81	947	32.6	395	2.03
1	1.008	6.88	395	26.1	95	7.68
1	1.014	6.14	565	23.6	214	1.45
1	1.024	6.30	874	29.9	380	5.16
1	1.019	5.47	760	33.8	199	0.81
1	1.014	7.38	577	30.1	87	1.32
1	1.020	5.96	631	11.2	422	1.55
1	1.023	5.68	749	29.0	239	1.52

1	1.017	6.76	455	8.8	270	0.77
1	1.017	7.61	527	25.8	75	2.17
1	1.010	6.61	225	9.8	72	0.17
1	1.008	5.87	241	5.1	159	0.83
1	1.020	5.44	781	29.0	349	3.04
1	1.017	7.92	680	25.3	282	1.06
1	1.019	5.98	579	15.5	297	3.93
1	1.017	6.56	559	15.8	317	5.38
1	1.008	5.94	256	8.1	130	3.53
1	1.023	5.85	970	38.0	362	4.54
1	1.020	5.66	702	23.6	330	3.98
1	1.008	6.40	341	14.6	125	1.02
1	1.020	6.35	704	24.5	260	3.46
1	1.009	6.37	325	12.2	97	1.19
1	1.018	6.18	694	23.3	311	5.64
1	1.021	5.33	815	26.0	385	2.66
1	1.009	5.64	386	17.7	104	1.22
1	1.015	6.79	541	20.9	187	2.64
1	1.010	5.97	343	13.4	126	2.31
1	1.020	5.68	876	35.8	308	4.49
2	1.021	5.94	774	27.9	325	6.96
2	1.024	5.77	698	19.5	354	13.00
2	1.024	5.60	866	29.5	360	5.54
2	1.021	5.53	775	31.2	302	6.19
2	1.024	5.36	853	27.6	364	7.31
2	1.026	5.16	822	26.0	301	14.34
2	1.013	5.86	531	21.4	197	4.74
2	1.010	6.27	371	11.2	188	2.50
2	1.011	7.01	443	21.4	124	1.27
2	1.022	6.21	-	20.6	398	4.18
2	1.011	6.13	364	10.9	159	3.10
2	1.031	5.73	874	17.4	516	3.01
2	1.020	7.94	567	19.7	212	6.81
2	1.040	6.28	838	14.3	486	8.28
2	1.021	5.56	658	23.6	224	2.33
2	1.025	5.71	854	27.0	385	7.18
2	1.026	6.19	956	27.6	473	5.67
2	1.034	5.24	1236	27.3	620	12.68
2	1.033	5.58	1032	29.1	430	8.94
2	1.015	5.98	487	14.8	198	3.16
2	1.013	5.58	516	20.8	184	3.30
2	1.014	5.90	456	17.8	164	6.99
2	1.012	6.75	251	5.1	141	0.65
2	1.025	6.90	945	33.6	396	4.18

2	1.026	6.29	833	22.2	457	4.45
2	1.028	4.76	312	12.4	10	0.27
2	1.027	5.40	840	24.5	395	7.64
2	1.018	5.14	703	29.0	272	6.63
2	1.022	5.09	736	19.8	418	8.53
2	1.025	7.90	721	23.6	301	9.04
2	1.017	4.81	410	13.3	195	0.58
2	1.024	5.40	803	21.8	394	7.82
2	1.016	6.81	594	21.4	255	12.20
2	1.015	6.03	416	12.8	178	9.39

Apéndice A2.

Estos datos se obtuvieron con el fin de estudiar qué efecto tiene, en la vasoconstricción sanguínea, el volumen (*Vol*) y la proporción (*Tas*) de aire aspirado. La variable de respuesta, *Resp*, vale uno si hubo vasoconstricción y cero en caso contrario. Treinta y nueve individuos integran la muestra.

Referencia. Pregibon, D. (1981): "Logistic Regression Diagnostics"; The Annals of Statistics, vol. 9, pp. 705 - 724.

<i>Vol</i>	<i>Tas</i>	<i>Resp</i>
3.7	0.825	1
3.5	1.09	1
1.25	2.50	1
0.75	1.50	1
0.8	3.20	1
0.7	3.50	1
0.6	0.75	0
1.1	1.70	0
0.9	0.75	0
0.9	0.45	0
0.8	0.57	0
0.55	2.75	0
0.6	3.00	0
1.4	2.33	1
0.75	3.75	1
2.3	1.64	1
3.2	1.60	1
0.85	1.415	1
1.7	1.06	0
1.8	1.80	1
0.4	2.00	0
0.95	1.36	0
1.35	1.35	0
1.5	1.36	0
1.6	1.78	1
0.6	1.50	0
1.8	1.50	1
0.95	1.90	0
1.9	0.95	1
1.6	0.40	0
2.7	0.75	1
2.35	0.03	0
1.1	1.83	0
1.1	2.20	1
1.2	2.00	1

0.8	3.33	1
0.95	1.90	0
0.75	1.90	0
1.3	1.625	1

Apéndice A3.

Se presentan los datos de un estudio realizado para conocer la relación que existe entre la edad de la muchachas varsovianas y la probabilidad de que hayan menstruado. Se formaron 25 grupos (o clases) de edad. La marca de clase de cada grupo se indica en la columna titulada "Edad". En la columna siguiente ("Tot") se muestra cuál fue el número de muchachas en cada grupo de edad. Finalmente, en la última columna ("Exit") se encuentra el número de varsovianas que, en cada grupo, dijeron haber menstruado.

Referencia. Guerrero, V. M. y Johnson, R. A. (1982): "Use of the Box-Cox transformation with binary response models"; *Biometrika*, vol. 69, pp. 309 - 314.

<i>Edad</i>	<i>Tot</i>	<i>Exit</i>
9.21	376	0
10.21	200	0
10.58	93	0
10.83	120	2
11.08	90	2
11.33	88	5
11.58	105	10
11.83	111	17
12.08	100	16
12.33	93	29
12.58	100	39
12.83	108	51
13.08	99	47
13.33	106	67
13.58	105	81
13.83	117	88
14.08	98	79
14.33	97	90
14.58	120	113
14.83	102	95
15.08	122	117
15.33	111	107
15.58	94	92
15.83	114	112
17.58	1049	1049

Apéndice A4.

Los datos mostrados en esta página provienen de un experimento factorial de 2x2 con repeticiones. Se estudiaron dos tipos de semilla: *O. aegyptiaca* 75 y *O. aegyptiaca* 79. Algunas semillas del primer tipo se cultivaron en un extracto de raíz y las restantes en otro. Lo mismo se hizo para el segundo tipo de semillas. A cada combinación de semilla y extracto de raíz se le llama tratamiento. En total se tienen cuatro tratamientos y 5 ó 6 repeticiones para cada uno de ellos. El objetivo del estudio era comparar el número de semillas que germinaban en cada medio de cultivo.

Referencia. Crowder, Martin J. (1978): "Beta-binomial Anova for Proportions"; Applied Statistics, vol. 27, pp. 34 - 37.

No. germinaron	Total	Raíz	Semilla
10	39	1	1
23	62	1	1
23	81	1	1
26	51	1	1
17	39	1	1
5	6	2	1
53	74	2	1
55	72	2	1
32	51	2	1
46	79	2	1
10	13	2	1
8	16	1	2
10	30	1	2
8	28	1	2
23	45	1	2
0	4	1	2
3	12	2	2
22	41	2	2
15	30	2	2
32	51	2	2
3	7	2	2

Apéndice B1.

! Este macro hace la gráfica de la devianza media local
! para un conjunto de observaciones clasificadas en %g
! grupos.
! Precauciones: 1. Las respuestas deben ser variables aleatorias
! Bernoullis.
! 2. Debe ajustarse el modelo global antes de usar
! este macro.
! Referencia: Landwehr, J.M.; Pregibon, D. y Shoemaker, A.C.
! (1984): "Graphical Methods for Assessing Logistic
! Regression Models." JASA; vol. 79, pp. 61 - 83.
!

\$mac dato!

! Notación:
! %g = número de grupos.
! g = vector que indica a qué grupo pertenece cada
! dato. Se supondrá que el grupo 1 fue el primero
! en formarse, el grupo 2 el segundo, etc.
!

\$cal %g = 24\$!

\$dat g\$!

\$fac g %g\$!

\$read 11 17 11 16 19 12 8 14 19 16 16 19 19 8 8 3 19 16 9 8

3 13 19 8 8 13 16 5 13 7 19 15 18 12 17 15 19 12 19 18

1 19 1 11 17 2 20 18 21 6 22 17 11 19 15 7 7 10 14 11

6 18 20 4 7 12 2 13 15 17 5 9 10 14 4 13 9 24 23\$!

\$endmac!

!

\$mac local!

! Ajuste del modelo local, cálculo de las devianzas
! locales y de sus grados de libertad.
! Notación:
! devi = vector con la devianza local de cada individuo.
! dg = vector con la devianza de cada grupo.

```

!           gl = vector con los grados de libertad de cada grupo.
$scal %z2 = %g:!  

w = 1$!  

$while %z2 ciclo$!  

$wei w$!  

$print : 'Ten paciencia, el ajuste del modelo requiere tiempo.' :$!  

$fit + g$!  

$dis e$  

$scal devi = 2 * w * %if(%yv == 0, -%log((1 - %fv) * w), - %log(%fv * w)):  

%z1 = %cu(devi):$!  

$print : 'Hay error si los números : ' *r%z1 ' y ' *r%dv ' son distintos.'$!  

$var %g dg gl$!  

$scal %c = %g$!  

$while %c devlo$!  

$endmac!  

!  

$mac ciclo!  

$scal %z3 = %cu(g == %z2):!  

%z4 = %if(%z3 == 1, 1, -1)$!  

$switch %z4 peso$!  

$scal %z2 = %z2 - 1$!  

$endmac!  

!  

$mac peso!  

$scal w = %if(g == %z2, 0, w)$!  

$endmac!  

!  

$mac devlo!  

$scal z = %eq(g,%c):!  

dg(%c) = %cu(z * devi):!  

gl(%c) = %cu(z) - 1:  

%c = %c - 1$!  

$del z$!  

$endmac!  

!

```

```

$mac graf!
!           Gráfica de la devianza media local.
!           La j-ésima entrada del vector DLP contiene la
!           devianza promedio de los primeros j grupos.
!
$cal %d = %cu(%gt(g1,0))$!
$var %d j dlp$!
$cal j = %cu(1):!
dl2 = dg(j):!
gl2 = gl(j):!
glac = %cu(gl2):!
dlp = %cu(dl2)/glac$!
$print : 'Gráfica de la devianza media local.' :$!
$plot dlp glac '**'$!
$endmac!
!
$mac devm!
!
!           Macro ( o programa principal).
!
$use dato$!
$use local$!
$use graf$!
$print : 'FIN del macro DEVM.' :$!
$endmac!
!
$return$

```

Apéndice B2.

!
!
! Gráfica de probabilidad empírica. Referencia:
! Landwher, J.M.; Pregibon, D. y Shoemaker, A. C.
! (1984): "Graphical Methods for Assessing Logistic
! Regression Models". JASA; vol. 79, pp. 61 - 83.
!
! Argumentos: %1, el vector de valores ajustados (%fv) y
! %2, los residuos (%yv - %fv) ordenados
! del modelo ajustado.

\$mac inicia!

\$print : 'Empieza la INICIALIZACION.' \$!

\$cal r1 = 0:!

r2 = 0:!

r3 = 0:!

r4 = 0:!

r5 = 0:!

r6 = 0:!

r7 = 0:!

r8 = 0:!

r9 = 0:!

r10 = 0:!

r11 = 0:!

r12 = 0:!

r13 = 0:!

r14 = 0:!

r15 = 0:!

r16 = 0:!

r17 = 0:!

r18 = 0:!

r19 = 0:!

r20 = 0:!

r21 = 0:!

r22 = 0:!

```

r23 = 0:!
r24 = 0:!
r25 = 0$!
$print : 'Fin de la INICIALIZACION.' :$!
$endmac!
!
$mac ajusta!
!
!           , En cada iteración genera un vector de respuestas
!           binarias. Después usa este vector para ajustar un
!           modelo. Los residuos del modelo ajustado son
!           ordenados en forma creciente.
!
$dat u$!
$format f8.6$!
$dinput 11$!
$cal y = %le(u,%1)$!
$yvar y$!
$fit .3$!
$cal r = %yv - %fv$!
$sort r$!
$cal r1 = %if(%z1 == 1,r,r1):!
r2 = %if(%z1 == 2,r,r2):!
r3 = %if(%z1 == 3,r,r3):!
r4 = %if(%z1 == 4,r,r4):!
r5 = %if(%z1 == 5,r,r5):!
r6 = %if(%z1 == 6,r,r6):!
r7 = %if(%z1 == 7,r,r7):!
r8 = %if(%z1 == 8,r,r8):!
r9 = %if(%z1 == 9,r,r9):!
r10 = %if(%z1 == 10,r,r10):!
r11 = %if(%z1 == 11,r,r11):!
r12 = %if(%z1 == 12,r,r12):!
r13 = %if(%z1 == 13,r,r13):!
r14 = %if(%z1 == 14,r,r14):!

```

```

r15 = %if(%z1 == 15,r,r15):!
r16 = %if(%z1 == 16,r,r16):!
r17 = %if(%z1 == 17,r,r17):!
r18 = %if(%z1 == 18,r,r18):!
r19 = %if(%z1 == 19,r,r19):!
r20 = %if(%z1 == 20,r,r20):!
r21 = %if(%z1 == 21,r,r21):!
r22 = %if(%z1 == 22,r,r22):!
r23 = %if(%z1 == 23,r,r23):!
r24 = %if(%z1 == 24,r,r24):!
r25 = %if(%z1 == 25,r,r25):!
$print 'Cuenta regresiva. Fin de la iteracion ' *i%z1 :$!
$cal %z1 = %z1 - 1:!
u = 0$!
$endmac!
!
$mac estad!
!
!           En cada iteración: almacena en el vector V la
!           estadística de orden %z1 de los 25 vectores de
!           residuos. Posteriormente calcula el mínimo, el
!           máximo y la mediana de V.
!
$var 25 v$!
$var %nu med min max$!
$cal v(1) = r1(%z1):!
v(2) = r2(%z1):!
v(3) = r3(%z1):!
v(4) = r4(%z1):!
v(5) = r5(%z1):!
v(6) = r6(%z1):!
v(7) = r7(%z1):!
v(8) = r8(%z1):!
v(9) = r9(%z1):!
v(10) = r10(%z1):!

```

```

v(11) = r11(%z1):!
v(12) = r12(%z1):!
v(13) = r13(%z1):!
v(14) = r14(%z1):!
v(15) = r15(%z1):!
v(16) = r16(%z1):!
v(17) = r17(%z1):!
v(18) = r18(%z1):!
v(19) = r19(%z1):!
v(20) = r20(%z1):!
v(21) = r21(%z1):!
v(22) = r22(%z1):!
v(23) = r23(%z1):!
v(24) = r24(%z1):!
v(25) = r25(%z1):!
bin = %ge(v,0):!
%z2 = %cu(bin)$!
$tab the v smallest into min1:!
the v largest into max1$!
$cal min(%z1) = min1:!
max(%z1) = max1:!
%z3 = %if((%z2 == 25) ? (%z2 == 0),1,2)$!
$arg dos %1$!
$switch %z3 uno dos$!
$print 'Cuenta regresiva. Terminó la iteración ' *i%z1$!
$cal %z1 = %z1 - 1$!
$endmac!
$mac uno!
$tab the v fifty into med1$!
$cal med(%z1) = med1$!
$endmac!
!
$mac dos!
$tab the v fifty for bin into med2$!
$cal med(%z1) = %if(%1(%z1) < 0, med2(1), med2(2))$!

```

```

$endmac!
!
$mac grafica!
!
!           Gráfica de probabilidad empírica.
!
$plot %1 %2 %3 %4 '-*-'$!
$print : 'Fin de GRAFICA.' :$!
$endmac!
!
$mac envol!
!
!           Macro ( o programa) principal.
!
$use inicia$!
$cal %z1 = 25$!
$arg ajusta %1$!
$print : 'Comienza el macro AJUSTE.' :$!
$print ' Archivo con los números pseudoaleatorios y uniformes.' :$
$while %z1 ajusta$!
$print : 'Fin del AJUSTE.' :$!
$cal %z1 = %nu$!
$arg estad %2$
$print : 'Comienza el macro ESTADISTICAS.' :$!
$while %z1 estad$!
$print : 'Fin de ESTADISTICAS' :$!
$arg grafica max %2 min med$!
$use grafica$!
$endmac!
$return$

```

Apéndice B3.

program singular

C

C **Introducción.** El objetivo de este programa es cal-
 C cular las medidas propuestas por Esteban
 C Walker para determinar qué puntos in-
 C fluyen en el nivel de colinealidad de
 C la matriz de diseño (X).

C **Método.** Se usa la subrutina LSVDF de IMSL para
 C obtener la descomposición en valor singular de X.
 C Los datos se encuentran en el archivo
 C SING.DAT y los resultados se almacenan
 C en SING.SAL.

C **Referencias.** 1. Walker, Esteban (1989):
 C "Detection of collinearity-influential
 C observations". Commun. Statist.-Theory
 C Meth., 1675-1690. Department of
 C Statistics. The University of Tennessee
 C Knoxville, TN 37996.

C 2. Capítulo L del manual de IMSL.

C **Notación.** Descomposición en valor singular de a:

$$C \quad a = u * D * v'$$

C a: matriz de diseño de mxn (real).

C m,ia: número de renglones de a.

C n: número de columnas de a.

C ut: matriz de mxm (de entrada y salida).

C s: vector que contiene los (n) valores
 C singulares en orden decreciente.

C wk: vector de trabajo de dimension $2n + 1$.

C ier: parámetro de error.

C eta: índices de condición.

C v: se guarda en los primeros n renglones de a.

C **Programó:** Rebeca Aguirre Hernández.

C **Lugar:** IIMAS, UNAM.

C Fecha: Diciembre de 1989.

C

```

implicit double precision(a-h,o-z)
integer ia,iu,m,n,ier
dimension a(37,3),ut(37,37),s(3),wk(7)
dimension cut(37,37),cv(3,3),cs(3),eta(3)
open(unit=6,file='sing.dat',status='old')
open(unit=8,file='sing.sal',status='new')
read(6,*) m
read(6,*) n
read(6,*) ((a(i,j),j=1,n),i=1,m)
close (6)
write(8,999) m,n
write(8,998)
write(8,997) ((a(i,j),j=1,n),i=1,m)
do 10 i=1,m
do 10 j=1,m
10      ut(i,j) = 0.0
do 20 i=1,m
20      ut(i,i) = 1.0
ia = max0(m,n)
iu = m
write(8,996) ia,iu
call lsvdf(a,ia,m,n,ut,iu,m,s,wk,ier)
do 30 i=1,m
do 30 j=1,m
30      cut(i,j) = ut(i,j)**2
do 40 i=1,n
do 40 j=1,n
40      cv(i,j) = a(i,j)**2
do 50 i=1,n
50      cs(i) = s(i)**2
do 60 i=1,n
60      eta(i) = cs(1)/cs(i)
write(8,995)

```

```

write(8,994) (s(i),i=1,n)
write(8,993)
write(8,992) (cs(i),i=1,n)
write(8,991)
write(8,990) ((cv(i,j),j=1,n),i=1,n)
write(8,989)
write(8,988) ((cut(i,j),i=1,n),j=1,m)
write(8,987) ier
close(8)
stop
999 format(3x,'Las dimensiones de la matriz son : ',i4,2x,'x',i4,/)
998 format(3x,'La matriz es :',/)
997 format(3(3x,f16.5,1x))
996 format(/,3x,'Las constantes ia e iu : ',3(i4,2x))
995 format(/,3x,'Los valores singulares :',/)
994 format(3(3x,f12.6,1x))
993 format(/,3x,'Los valores propios son:',/)
992 format(3(3x,f12.6,1x))
991 format(/,3x,'La matriz CV :',/)
990 format(3(3x,f12.6,1x))
989 format(/,3x,'La matriz CU :',/)
988 format(3(3x,f8.6,2x))
987 format(/,3x,'Error # : ',i5,/)
end

```

Apéndice B4.

\$com Este macro calcula los residuos verosimiles de Williams. Se presupone que se ha ajustado un modelo para respuestas binomiales.

Notación:

h = palanca de las observaciones.

m = 1 - h.

rp2 = residuos de Pearson, estandarizados y elevados al cuadrado.

rd2 = residuos de la devianza, estandarizados y elevados al cuadrado.

rg = residuos verosimiles.

sig = signo de los residuos verosimiles.\$

\$

\$macro resv!

\$extract %v1\$!

\$cal h = %wt * %v1:!

m = 1 - h:!

rp2 = ((%yv - %fv)**2)/(%wt * m):!

sig = %if(%gt(%yv - %fv,0),1,-1):!

prop = %yv/%bd:!

a = %if(%eq(%yv,0),-%bd * %log(1 - %fv/%bd),0):!

b = %if(%eq(%yv,%bd),-%bd * %log(%fv/%bd),0):!

c = %ne(%yv,0) * %ne(%yv,%bd)\$!

\$warn\$

\$cal c1 = %if(c, %yv * (%log(prop) - %log(1 - prop) - %lp),0):!

c2 = %if(c,%bd * (%log(1 - prop) - %log(1 - %fv/%bd)),0)\$!

\$warn\$

\$cal devi = 2 * (a + b + c1 + c2):!

rd2 = devi/m:!

rg2 = m * rd2 + h * rp2:!

%d = %cu(rd2 * m):!

%p = %cu(rp2 * m)\$

\$print : 'La devianza del modelo es: ' *r%d %dv :\$

\$print : 'La ji-cuadrada de Pearson es: ' *r%p %x2 :\$

\$del prop a b c c1 c2\$

\$endmac!

Apéndice B5.

\$com Este archivo tiene dos macros que ajustan un modelo logístico con un parámetro de variabilidad extra-binomial.

Antes de usar cualquiera de los macros debe ajustarse el modelo sin el parámetro mencionado.\$

\$

\$subfile mac1\$

\$mac veb1!

\$extract %v1\$!

\$cal %p = (%x2 - %df)/%cu((%bd - 1)*(1 - %wt * %v1))!:

peso = 1/(1 + %p * (%bd - 1))\$!

\$wei peso\$!

\$fit . \$!

\$print : 'La estadística ji-cuadrada para el modelo II es: ' %x2 :\$!

\$print 'Estimador del parámetro de variabilidad extra-binomial: ' *r%p :\$!

\$endmac!

\$return\$

\$subfile mac2\$

\$mac veb2!

\$extract %v1\$!

\$cal wvq = %pw * %wt * %v1!:

%p = (%x2 - %cu(%pw*(1 - wvq)))/(%cu(%pw * (%bd - 1) * (1 - wvq)))!:

peso = 1/(1 + %p * (%bd - 1))\$!

\$wei peso\$!

\$fit . \$!

\$print : 'La estadística ji-cuadrada para el modelo II es: ' %x2 :\$!

\$print 'Estimador del parámetro de variabilidad extra-binomial: ' *r%p :\$!

\$endmac!

\$return\$

\$finish\$

Apéndice B6.

```
$com Objetivo. Este macro calcula los residuos de Pearson suavi-
$ zados de acuerdo a como los definió Fowlkes.
$ Precaución. Antes de usar este macro debe ajustarse el
$ modelo logístico apropiado.
$ Argumentos del macro SUAVIZA.
$ %1 = tamaño de la ventana de suavizamiento (en %).
$ %2 = número de variables explicativas.
$ %3 = variable de respuesta.
$ %4 a %9 = variables explicativas.
$ Constantes y vectores de SUAVIZA.
$ %z1 = contador, controla el número de veces que debe
$ ejecutarse el submacro REPITE.
$ %z4 = número de observaciones que se usarán para
$ suavizar cada respuesta.
$ PESO = vector con el peso asociado a los datos
$ que suavizarán una determinada respuesta.
$ YSUAV = vector de respuestas suavizadas.
$ MEDS = estimador de la media de YSUAV.
$ VARS = estimador de la varianza de YSUAV.
$ RESS = vector de residuos de Pearson suavizados.$
$
$macro suaviza!
$cal %z1 = %nu!
%z2 = (%nu * %1)/100!
%z3 = %z2 -%tr(%z2):!
%z4 = %if(%ge(%z3,0.5),%tr(%z2) + 2, %tr(%z2) + 1)!
$var %z4 peso j $!
$cal j = %gl(%z4,1)$!
$var %nu ysuav meds vars ress$!
$arg repite %2 %3 %4 %5 %6 %7 %8 %9$!
$while %z1 repite$!
$cal ress = (meds - %fv)/%sqrt(vars)$!
$del peso j $!
```

```

$print : 'FIN DEL MACRO SUAVIZA.' :$!
$endmac!
!
$macro repite!
$cal %z5 = %nu - (%z1 - 1):!
suma = 0:!
%w = %1$!
$arg euc %1 %3 %4 %5 %6 %7 %8$!
$while %w euc$
$cal dist = %sqrt(suma)$!
$sort yord %2 dist:!
prob %fv dist:!
dist$!
$cal %a = %eq(dist(%z4),0)$!
$switch %a frase$!
$cal peso = (1 - (dist(j)/dist(%z4))**3)**3:!
%z6 = %cu(peso * yord(j)) - peso(1) * %2(%z5):!
%z7 = %cu(peso) - peso(1):!
ysuav(%z5) = %z6/%z7:!
%z8 = %cu(peso * prob(j)) - peso(1) * prob(1):!
%b = %cu(peso * peso * prob(j) * (1 - prob(j))):!
%c = peso(1) * peso(1) * prob(1) * (1 - prob(1)):!
%z9 = %b - %c:!
meds(%z5) = %z8/%z7:!
vars(%z5) = %z9/(%z7 * %z7):!
%z1 = %z1 - 1$!
$del dist$!
$print 'Respuesta suavizada: ' *i%z5 $!
$endmac!
!
$macro euc!
$arg x1 %2$!
$arg x2 %2 %3$!
$arg x3 %2 %3 %4$!
$arg x4 %2 %3 %4 %5$!

```

\$arg x5 %2 %3 %4 %5 %6\$!

\$arg x6 %2 %3 %4 %5 %6 %7\$!

\$switch %1 x1 x2 x3 x4 x5 x6\$!

\$cal suma = suma + (x(%z5) - x)**2:!

%w = %w - 1\$!

\$endmac!

!

\$macro x1!

\$cal x = %1\$!

\$endmac!

!

\$macro x2!

\$cal x = %1 * %eq(%w,1) + %2 * %eq(%w,2)\$!

\$endmac!

!

\$macro x3!

\$cal x = %1 * %eq(%w,1) + %2 * %eq(%w,2) + %3 * %eq(%w,3)\$!

\$endmac!

!

\$macro x4!

\$cal x = %1 * %eq(%w,1) + %2 * %eq(%w,2) + %3 * %eq(%w,3) + %4 * %eq(%w,4)\$!

\$endmac!

!

\$macro x5!

\$cal x = %1 * %eq(%w,1) + %2 * %eq(%w,2) + %3 * %eq(%w,3) + %4 * %eq(%w,4) + %5 * %eq(%w,5)\$!

\$endmac!

!

\$macro x6!

\$cal x = %1 * %eq(%w,1) + %2 * %eq(%w,2) + %3 * %eq(%w,3) + %4 * %eq(%w,4) + %5 * %eq(%w,5) + %6 * %eq(%w,6)\$!

\$endmac!

!

\$macro frase!

\$print : 'Division por cero. Intenta con una ventana mayor.' :\$!

\$del %a peso j\$!

\$exit 3\$!

\$endmac!

!

\$return\$